

증강 현실 구현을 위한 기본 기술과 카메라 모델링

윤용인* · 김진태** · 최종수*

1. 서 론

증강 현실(augmented reality: AR)은 일반적으로 가상 환경(virtual environment: VE) 혹은 가상 현실(virtual reality: VR)에서 파생한 것으로 실제 환경(real environment)에 컴퓨터 그래픽스 영상을 삽입한 실제와 가상 영상의 혼합을 뜻한다 [1]. 실세계는 컴퓨터로 만들기 힘든 많은 정보들이 존재한다. 그렇지만 그 정보들 중에 필요하지 않은 많은 정보도 있고, 때로는 필요로 하는 정보가 부족할 수도 있다. 컴퓨터로 재생한 가상 환경을 이용하면, 필요하지 않은 정보들을 단순하게 만들거나 보이지 않는 정보들을 만들어 낼 수 있다.

사람들은 실제와 가상의 혼합체를 통해 주위 환경을 쉽게 인식하고, 이를 보다 더 잘 이용할 수 있다. 효율적인 AR 시스템은 실세계와 가상세계를 합성하며, 실시간으로 사람과 상호 작용이 이루어 져야 한다. 따라서 AR 시스템은 삼차원 세계에 잘 정합되어야 좋은 시스템이다. 실세계와 가상세계 간에 공존하는 세계를 이른바 혼합 현실(mixed reality)라 한다. 증강 현실 시스템(augmented reality system)은 사용자가 실세계 환경하에서 가상물체를 추가되어 증강된 실세계(augmenting real world)를 말한다. 증강된 가상(aug-

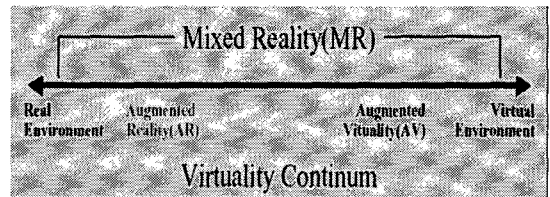


그림 1. 현실과 가상의 연속 관계

mented virtuality: AV)은 컴퓨터가 만들어낸 완전한 가상세계를 기반으로 실제 환경을 합성한 것을 말한다. 실세계와 가상세계 사이의 연속관계는 그림 1과 같다. 그림 1에서 AR은 실세계와 가까이 있으며, AV는 가상세계와 가까이 있다. 그렇지만 둘 관계를 확실히 구별하는 것은 애매모호하다.

2. AR 시스템과 기본 기술

2.1 AR 기본 시스템

AR이 사용되는 곳은 여러 곳이며, 이에 사용되는 시스템도 다르다. AR 시스템은 사람이 보는 기준 프레임에 가상 프레임을 잘 정합해야 한다. 그렇지만 일반적으로 가상환경에서 발생하는 시각-운동 오류보다 시각적 부정합에 더 민감하기 때문에, 정합이 AR 시스템에서 매우 어렵다. 즉, 가상물체는 실세계와 같이 물체가 움직이지 않는 것으로 생기는 어색함보다 환경과 일치하지 않아 발생하는 어색함이 사람에게 더 크다. 그림 2는

* 중앙대학교 첨단영상대학원 영상공학과

** 한서대학교 컴퓨터정보학과

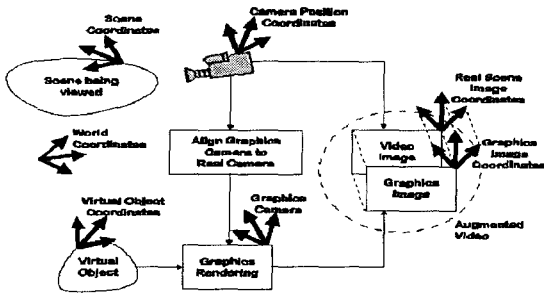


그림 2. AR 기본 시스템

AR의 기본 시스템을 보여 준다[2]. 장면은 영상 장치(비디오 카메라)로부터 구한다. 카메라는 3D 세계에서 2D 영상으로 투영된다. 이때 카메라의 내부(초점거리), 외부(위치) 변수 값들은 정확히 계산된다. 가상 영상은 컴퓨터 그래픽스 시스템에서 만들어 지고, 가상 물체는 물체의 기준 프레임에서 모델링 되어 만들어진다. 그래픽스 시스템은 가상 물체를 정확히 렌더링하기 위해서 실제 영상에 관한 정보가 반드시 필요하다. 렌더링된 데이터는 가상 물체 영상을 가지고 카메라로 제어된다. 마지막으로 가상의 영상은 실제 영상에 합성되어 증강 현실 영상을 형성한다.

AR 시스템은 사용자가 자유롭게 돌아다니면서 적절히 렌더링된 증강 영상을 보기 위해서 실시간으로 실행되는 것을 목표로 한다. 우선 증강된 영상을 만드는 것에 대하여 갱신하고, 실제와 가상 영상 간에 정합이 정확하게 되어야한다. 물체가 갑작스럽게 흔들려 보이는 것을 방지하기 위해 최소 10frame/sec 정도로 계속 갱신되어야 한다. 따라서 AR 시스템은 실제 영상과 가상 영상에서 잡음이 시간적인 지연으로 인한 부정합 현상이 발생된다.

2.2 정합

증강 현실은 가상 현실보다 더욱 정확한 정합

(registration)이 필요하다. AR 응용을 제한하는 가장 큰 문제 중의 하나가 정합이다. 실세계와 가상 물체가 잘 정합되어 적절히 정렬이 되어야 합성된 영상이 어색해 보이지 않기 때문에 매우 중요하다. 정합에는 굉장한 정밀도가 요구되는데, 수많은 종류의 오류가 있기 때문에 정합 오류는 적절히 처리하기가 매우 어렵다. 정합 오류는 크게 정적인 것과 동적인 오류의 두 가지 타입이 있다. 정적인 오류는 광학적인 왜곡, 추적 시스템에서 오류, 기계적인 불일치, 부정확한 시점 변수 등이 있다.[2] 동적인 오류는 시스템 지연에서 일어난다. 추적 시스템의 지연, 통신상의 지연, 프레임 버퍼로부터 특정 영상을 만들고 디스플레이 하는데 걸리는 지연 등이 정합의 동적인 오류에 해당된다.[3]

2.3 추적

물체에 대한 추적(tracking)이나 카메라 위치 결정은 AR을 형성하는데 매우 필요한 기술이다. 비디오 카메라, 자이로, 가속기, RF(GPS, Impulse Radio), 나침반, 보수계, 고도계 센서 등 여러 장비가 물체의 추적에 사용된다. 그렇지만 하나의 기술로써 강건하고 정확한 추적은 수행하기에 많은 어려움이 있다. 결국에 이러한 장비들을 하나가 아닌 여러 개의 복합적으로 이용하는 형태의 추적이 사용된다. 추적하는 방법을 공간적으로 분리하면, 실내에서 추적하는 것과 실외에서 추적하는 것이 있다.[2-3]

2.3.1 실내 추적

실내에서 추적(indoor tracking)하는 방법 중 많이 사용되는 기준점 탐색기에 대해 알아보자. 원하는 물체에 칼라로 원형 같은 특정 형태의 스티커를 붙여서 추적한다. 각각의 색에 대한 표식

의 집합을 가지고 추적하므로, 정확한 모서리를 구하거나 건물의 구조 등은 몰라도 혹은 GPS같은 센서의 노이즈에도 강건하게 추적한다. 그림 3은 처음에 기준점 추적을 하면서, 자연적 특징점(코너)의 추적을 보여준다.

2.3.2 실외 추적

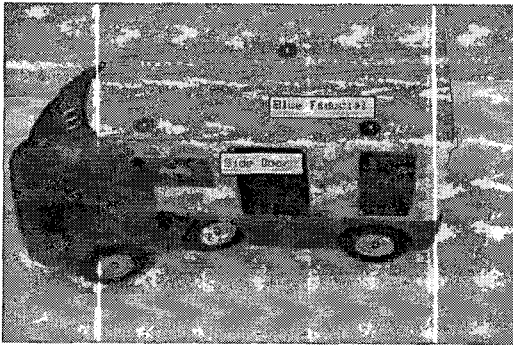
실외에서의 추적(outdoor tracking)은 주위 환경에 대한 어떤 준비나 변형 없이 구성된다. 추적할 물체가 어떤 형태이던 간에 카메라 모션을 계산하기 전에, 이 특징 등은 감지되고 추적되어야 한다. 모서리나 점들이 실외 영상에서 많이 추적된다. 그렇지만 이들은 서로 유사한 것들이 많아

서 추적하기 힘들다. 각각의 특징점들의 배열 상태나 추적된 영역의 전후 관계를 잘 살펴서, 가장 잘 맞는 특징점을 선택해야 한다.

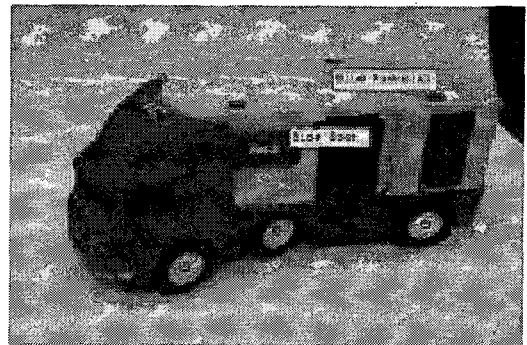
폐 루프의 접근 방법은 자연 환경에서 자연 특징 점을 정확하고 강건하게 감지하고, 추적에 유용하다. 그림 4는 실외에서 특징 점을 고른 뒤 이를 추적한 결과를 보여준다.

3. 카메라 모델링

카메라 모델링은 카메라의 내부 변수와 외부 변수를 계산하는 것이다. 이것은 두 가지 과정으로 진행되는데, 우선 투영 행렬 P 를 구한다. 다음으

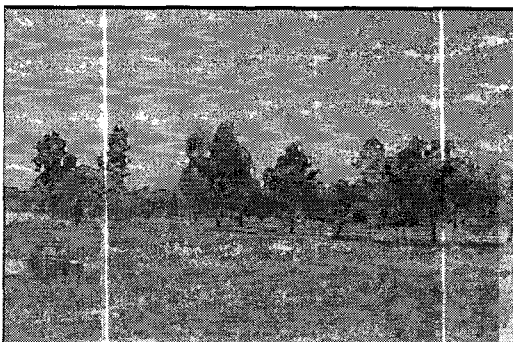


(a) 기준점 카메라 추적을 가진 초기 영상

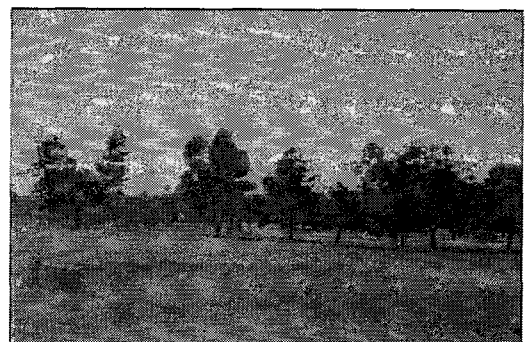


(b) 추적과 교정된 자연적 특징 점에 기반한 카메라 추적을 가진 영상

그림 3. 실내의 기준점 추적



(a) 실외에서의 특징점



(b) 각 특징점들의 추적 결과

그림 4. 실외의 기준점 추적

로 계산된 P 로부터 내부 및 외부 변수를 구한다.

3.1 가상 카메라

가상 카메라(virtual camera)란 일반적으로 사용하는 카메라가 아닌 수학적으로 모델링된 카메라를 의미한다. 그림 5와 같이 모델링된 카메라를 핀홀 카메라(pinhole camera)라고 부른다.[4]

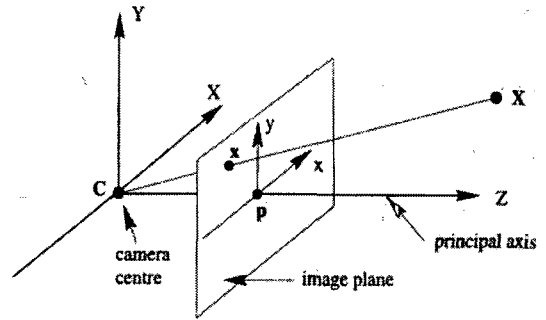


그림 5. 핀홀 카메라 모델

3.2 핀홀 카메라

그림 5에서 보는 바와 같이 하나의 핀홀 카메라는 공간상의 한 점 C (카메라 렌즈의 중심)와 하나의 평면 E (image plane)에 의해서 식 (1)과 (2)와 같이 결정지을 수 있다

$$Cam = (C, E) \tag{1}$$

$$f = d(C, E) \tag{2}$$

먼저 주어진 카메라에서 어떻게 공간 좌표계를 설정할 수 있는지 알아보자. 그림 5와 같이 C 는 카메라 렌즈의 중심이고, C 가 3차원 공간의 원점 $C = (0, 0, 0)$ 이라 하면, 기준 축(principal axis)을 따라 $f = d(C, E)$ 을 만족하는 점 p (principal point)가 영상 평면 E 에 유일하게 존재한다. 이때 $p = (0, 0)$ 을 영상 평면 E 의 원점으로 생각하자. 그러면 영상 평면 E 는 $p = (0, 0)$ 을 중심으로 하나의 직교 좌표계 (x, y) 라고 취급할 수 있다. 그림 4에서 \overline{Cp} 을 공간 좌표계의 Z 축으로 생각하고, \overline{px} 와 평행한 \overline{CX} 을 X 축으로 하고, \overline{py} 와 평행한 \overline{CY} 을 Y 축으로 하면, $C = (0, 0, 0)$ 을 원점으로 하는 하나의 공간 좌표계 $(C: X, Y, Z)$ 가 형성 된다. 이때 공간 좌표계 $(C: X, Y, Z)$ 에 대한 영상 평면 E_f 는 식 (3)과 같다.

$$E_f = \{(X, Y, f) | X, Y \in R\} \tag{3}$$

공간상의 한 점 $P = (X, Y, Z)$ 가 렌즈 중심 C

을 통하여 영상 평면 E_f 에 도달하는 관계를 살펴보자. 한 점 $P = (X, Y, Z)$ 가 영상 평면 E_f 에 맺히는 관계는 그림 5와 같이 직선 Cp 과 영상평면 E 와의 교점을 의미한다. 우리는 카메라 초점 C 을 좌표계의 원점으로 잡았으므로 직선 Cp 의 방정식은 식 (4)와 같다.

$$Cp: t(X, Y, Z) \quad t \in R \tag{4}$$

또한 영상 평면 E_f 의 방정식은 $Z = f$ 이므로, 점 $t(X, Y, Z)$ 을 $Z = f$ 에 대입하면 $tZ = f$ 라는 방정식을 얻게 되고, t 는 $t = \frac{f}{Z}$ 라는 값을 구할 수 있게 된다. 따라서 직선 Cp 과 영상 평면 E_f 의 교점은 $(\frac{fX}{Z}, \frac{fY}{Z}, f)$ 가 된다. 여기서 초점거리 f 는 상수이므로, 다음과 같은 비선형 방정식이 성립한다.

$$R^3 \rightarrow E_f \rightarrow R^2: (X, Y, Z)^T \rightarrow (\frac{fX}{Z}, \frac{fY}{Z}, f)^T \rightarrow (\frac{fX}{Z}, \frac{fY}{Z})^T \tag{5}$$

이것은 유클리드 공간 R^3 에서 유클리드 평면 R^2 로 가는 하나의 함수이다. 식 (5)를 핀홀 카메라의 비선형 방정식이라고 한다.

3.3 핀홀 카메라 방정식의 선형화

핀홀 카메라 방정식을 선형 방정식으로 변환하

기 위해, 사영 공간에 대한 지식이 필요하다. 공간상의 점 $P = (X, Y, Z)$ 은 다음과 같이 하나의 4차원 벡터 좌표로 표시할 수 있다.

$$P = [X, Y, Z, 1] \tag{6}$$

또한 평면상의 점은 하나의 3차원 벡터 좌표로 표시할 수 있다.

$$p = [x, y, 1] \tag{7}$$

만약 $z \neq 0$ 인 3차원 벡터 좌표 $[x, y, z]$ 는 $[\frac{x}{z}, \frac{y}{z}, 1]$ 로 생각하면, 식 (6)의 방정식은 다음과 같이 3x4 크기의 행렬을 사용할 수 있고, 이것을 카메라의 선형 방정식이라고 한다.

$$\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \rightarrow \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} fX \\ fY \\ Z \end{bmatrix} = \begin{bmatrix} \frac{fX}{Z} \\ \frac{fY}{Z} \\ 1 \end{bmatrix} \tag{8}$$

이때, $P = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$ 을 사영 행렬(projection

matrix)이라 부른다. $H_p = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix}$

하면 사영 행렬 P 는 다음과 같다

$$P = [H_p, O] \tag{9}$$

3.4 좌표계의 변환

3.4.1 카메라 내부변수

CCD 카메라로 취득하는 영상은 식 (9)처럼 얻어지는 형태가 아니고, CCD 칩의 영향에 의하여 실제로 얻는 영상의 형태는 다음과 같은 변형(distortion)이 이루어진다.

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} p_x & (\tan\alpha)p_y & c_x \\ 0 & p_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_R \\ y_R \\ 1 \end{bmatrix} \tag{10}$$

여기서, 행렬 $[x_R, y_R, 1]^T$ 는 식 (8)에서 얻어진

좌표이고, $[x, y, 1]^T$ 는 CCD 칩의 변형 식으로 얻어지는 영상의 좌표 행렬이다. 식 (10)은 식 (8)을 이용하여 다음과 같이 표현 가능하다.

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} p_x & (\tan\alpha)p_y & c_x \\ 0 & p_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{11}$$

식 (11)에서

$$\begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{12}$$

로 분리할 수 있고, 또한

$$\begin{bmatrix} p_x & (\tan\alpha)p_y & c_x \\ 0 & p_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} fp_x & (\tan\alpha)p_y f & c_x \\ 0 & p_y f & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{13}$$

이므로, $fp_x = f_x, p_y(\tan\alpha)f = s, p_y f = f_y$ 라 놓으면, 식 (14)와 같이 행렬 K 를 만들 수 있다.

$$K = \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{14}$$

이때, 행렬 K 의 인자 f_x, f_y, s, c_x, c_y 을 카메라 내부변수(intrinsic parameters)라 부른다. 이 카메라 내부변수는 $z = 1$ 인 카메라를 기준으로 결정되는 변수이다. 변수 s 을 비뒀어짐 변수(skew parameter)라고 하는데, 일반적인 카메라인 경우는 $s = 0$ 으로 가정한다.

따라서,

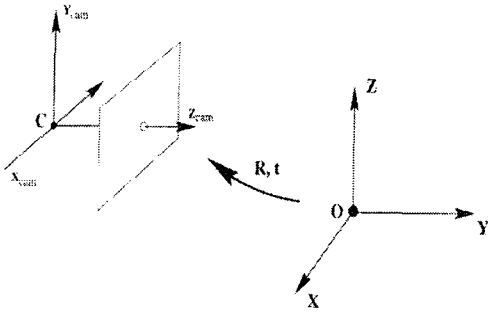
$$\det K = f_x f_y \tag{15}$$

이 되고, 카메라 행렬은 다음과 같다.

$$x = K [I_{3 \times 3} | O] X_{cam} \tag{16}$$

3.4.2 카메라의 이동

카메라 좌표계 프레임에서 공간 좌표계 프레임으로의 카메라 이동(camera motion)은 현실 세계를 가상의 공간으로 맵핑하는 것인데, 그림 6에



The Euclidean transformation between the world and camera coordinate frames.

그림 6. 공간 프레임과 카메라 좌표 프레임 간의 관계

두 프레임 간의 관계를 그림으로 나타내었다.

X_{cam} 은 카메라 중심 C 를 원점으로 하는 3차원 공간에 대한 좌표계의 표현이다. 또, 공간 상의 한 점 O 을 중심으로 하는 좌표계(world coordinate)를 생각해 보자. 공간 상의 한 점 X 는 카메라 좌표계 X_{cam} 으로 표시할 수도 있고, O 을 중심으로 하는 공간 좌표계로 X_w 으로 도 표시 가능하다. 수학적으로 한 점 X 의 두 좌표계에서 표현된 X_{cam} 와

$$X_{cam} = R(X_w - C_w), \quad R \in SO_3(R) \quad (17)$$

식 (17)을 유사 좌표계를 사용하여 표현하면 다음과 같다.

$$X_{cam} = \begin{bmatrix} R & -RC_w \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}_w = \begin{bmatrix} R & -RC_w \\ 0 & 1 \end{bmatrix} X_w \quad (18)$$

따라서 $x = K[I_{3 \times 3} | O]X_{cam}$ 은 공간 좌표계로 다음과 같다.

$$\begin{aligned} x &= K[R | -RC_w]X_w = KR[I | -C_w]X_w \\ &= [KR | -KRC_w]X_w, \quad R \in SO_3(R) \end{aligned} \quad (19)$$

여기서, $t = -RC_w$ 라 놓으면,

$$\begin{aligned} P &= K[R | t] = [KR | Kt], \quad R \in SO(3), \\ t &= -RC_w \end{aligned} \quad (20)$$

이 되고,

$$\det KR = \det K \det R = f_x f_y \neq 0 \quad (21)$$

을 만족한다.

4. AR의 응용 분야

AR이 실생활에서 응용될 수 있는 여러 분야를 그림을 통해 알아본다. 먼저, 의학 분야의 예로써 그림 7은 임신한 환자의 자궁 안의 가상 태아를 보여주고 있다.



그림 7. 임신한 자궁 속에 가상의 태아

그림 8은 로봇에 사용된 예로써, 가상 영상을 사용하여 로봇을 조종할 수 있다. 실제로 로봇을 움직이기 이전에 가상의 로봇을 이용하여 그 움직임을 보고 움직임에 대한 신뢰도를 제공받을 수 있다.

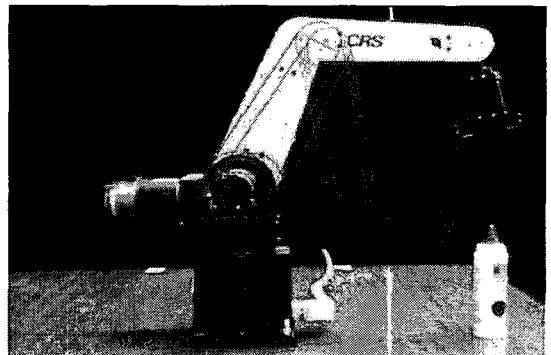


그림 8. 가상 선들에 의한 로봇 팔의 평면 움직임

그림 9는 레이저 프린터 수리 응용이다. HMD (Head-Mounted Displays)를 착용하여 종이 트레이를 제거하는 과정을 보여준다.



그림 9. 프린터 보수 응용

그림 10은 제한된 공간에서 일정한 패턴을 가진 가상 물체가 공간에 존재하는 것처럼 하고, 두 사람이 HMD를 착용하여 각각의 관점으로 보여 공동 작업을 수행하는 것이다.

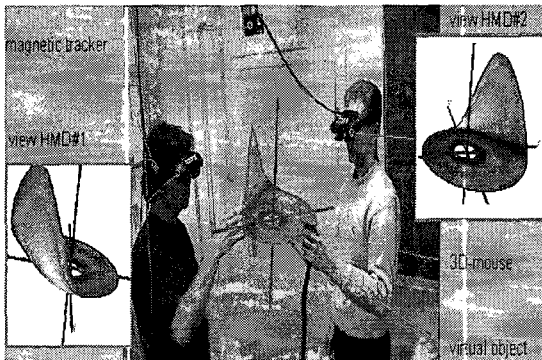


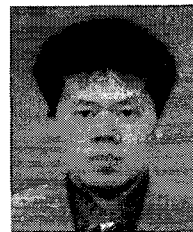
그림 10. 공동 작업

이상의 예로써 AR이 우리의 실생활에 사용될 수 있는 여지는 무궁무진하다. 이러한 AR 응용 제품을 개발하는 데에는 여러 가지 기본 기술들이 필요하지만 본고에서는 AR 구현에 가장 필요한

기본적이고 핵심적으로 요구되는 기술을 정리하였다.

참 고 문 헌

- [1] Yuchi Ohta, Hideyuki Tamura, "Mixed Reality-Merging Real and Virtual Worlds", Augmented Reality Magazine, 2003.
- [2] Augmented Reality Home Pages, <http://www.cs.rit.edu/~jrv/research/ar/introduction.html>.
- [3] R. T. Azuma, "A Survey of Augmented Reality," Teleoperators and Virtual Environments, vol. 6, no. 4, pp. 355-385, 1997.
- [4] R. Hartely and A. Zissermann, Multiple View Geometry in Computer Vision, Cambridge University Press, 2000.



윤 용 인

- 1986년 2월 중앙대학교 전자공학과 졸업(공학사)
- 1988년 2월 중앙대학교 전자공학과 졸업(공학석사)
- 2003년 8월 중앙대학교 첨단영상대학원 영상공학과 졸업(공학박사)
- 1991년 7월~1993년 11월 국제전자주 연구원
- 1994년 1월~1999년 2월 대우고등기술연구원 주임연구원
- 2003년 9월~2006년 2월 중앙대학교 첨단영상대학원 Post-Doc
- 2006년 3월~현재 중앙대학교 문화 콘텐츠 연구원 영상콘텐츠연구원
- 관심 분야: 컴퓨터 비전, 신호처리, 증강현실 등



김진태

- 1987년 2월 중앙대학교 전자공학과 졸업(공학사)
- 1989년 2월 중앙대학교 대학원 전자공학과 졸업(공학석사)
- 1993년 8월 중앙대학교 대학원 전자공학과 졸업(공학박사)
- 1995년 3월~현재 한서대학교 컴퓨터정보학과 부교수
- 관심분야: 영상처리, MPEG, 얼굴인식, 멀티미디어 등



최종수

- 1975년 2월 인하대학교 전기공학과 졸업(공학사)
- 1977년 2월 서울대학교 전자공학과 졸업(공학석사)
- 1981년 2월 일본 Keio University 전기공학과 졸업(공학박사)
- 1981년 9월~1999년 8월 중앙대학교 전자공학과 교수
- 1999년 9월~현재 중앙대학교 첨단영상대학원 교수
- 관심분야: 증강현실, 컴퓨터비전, 영상정보기술 등