

# 증강현실 요소기술과 동향 - 비전기반 추적 기술을 중심으로 -

박 한 훈\* · 박 종 일\*\*

## 1. 서 론

최근 증강현실(augmented reality) 기술이 큰 주목을 받고 있다. 스마트폰, HMD 등과 같은 모바일 기기의 기능과 성능 향상 및 위치기반 서비스의 보편화로 인해 그 유용성을 입증하는 다양한 응용 분야들이 소개되고 있으며 [1, 2], 포켓몬고 게임 출시와 함께 대중들에게서 폭발적인 관심을 받고 있다[3].

증강현실은 기본적으로 현실세계에 가상의 정보나 콘텐츠를 자연스럽게 병치하여 인간의 감각과 인식을 확장시키는 실감미디어 기술이다. 증강현실은 가상현실과 달리 현실공간을 미디어로 활용하기 때문에 실감있는 가상세계 구성을 위한 3차원 모델링/렌더링의 부담을 줄인 반면, 사용자의 움직임이나 의도, 주변 환경 변화에 대응하여 가상 정보나 콘텐츠를 적응적으로 제공해야 하므로 새로운 기술적인 어려움을 해결해야 한다[4].

증강현실의 가장 기본적이고 핵심적인 기술

은 객체(object)나 장면(scene)을 카메라로 비추어 인식하고 그 움직임을 추적하는 것이다. 본고에서는 객체 인식 및 추적을 위한 비전 기반 기술을 중심으로 설명하고, 최근 동향을 살펴본다. 또 증강현실 구현을 위해서는 카메라를 통해 들어온 실사영상과 컴퓨터가 만들어낸 그래픽 영상(CG)을 자연스럽게 합성하는 기술이 필요하며, 실시간 상호작용을 위한 다양한 인식기술도 요구되므로 이에 대해서도 살펴보기로 한다.

본고는 다음과 같이 구성된다. 2장에서는 객체 인식 및 추적을 위한 요소 기술의 개념과 원리를 설명하고, 대표적인 비전 기반 객체 인식 및 추적 기술을 분류하여 각각의 기술에 대해 설명한 후, 최근에 발표된 논문들을 통해 동향을 살펴본다. 3장에서는 실사와 CG의 합성 기술을 설명하고, 4장에서는 증강현실의 실시간 상호작용에 필요한 동작인식 기술의 개념과 동향을 소개한다. 마지막으로, 5장에서는 결론을 제시한다.

## 2. 객체 인식/추적 기술

### 2.1 객체 인식/추적의 기반기술

증강현실에서의 객체 인식(detection & recog-

※ 교신저자(Corresponding Author): 박한훈, 주소: 부산광역시 남구 용소로 45 부경대학교 누리관, 전화: 051-629-6225, E-mail: hanhoon\_park@pknu.ac.kr

\* 부경대학교 전자공학과  
\*\* 한양대학교 컴퓨터공학부

ation)은 카메라 영상에서 현실세계와 가상세계를 정합(registration)하기 위한 관심 객체(object of interest or region of interest)를 발견하고 인지하는 기술을 의미한다. 객체 인식을 위해 객체의 고유한 크기, 색상, 히스토그램, 텍스처 등 다양한 정보를 활용할 수 있으며, 증강현실의 경우 일반적으로 객체 인식은 추적을 통해 동시에 이루어진다. 객체 추적(tracking)은 연속된 카메라 영상에서 카메라나 관심 객체의 움직임 또는 카메라와 관심 객체 사이의 상대적인 포즈 변화를 추정하는 기술을 말한다. 이를 위해 카메라(비전 센서), 가속도 센서, GPS, 자이로스코프, RFID 등의 다양한 센서를 활용할 수 있으나, 본고에서는 카메라를 이용하는 비전 기반 기술을 중심으로 논한다. 비전 기반 객체 인식 및 추적을 위해서는 카메라 투영(projection) 행렬 계산, 영상 특징(image feature) 검출(detection)/서술(description)/매칭(matching), 3차원 모델링과 같은 요소 기술을 필요로 한다.

카메라 투영 행렬은 그림 1에서 보는 것처럼 객체(장면)의 3차원 좌표( $M$ )와 이에 대응하는 카메라 영상에서의 2차원 좌표( $m$ ) 사이의 관계를 표현한다. 일반적으로 핀홀(pinhole) 카메라 모델을 사용하며 크게 내부(intrinsic) 행렬  $K$ 와 외부(extrinsic) 행렬  $E$ 로 나누어진다. 내부 행렬은 카메라의 고유한 광학 특성과 관련된 파라미터를 포함하며, 여러 시점에서 촬영된 2차원 패턴 영상으로부터 계산될 수 있다(이러한 과정을 카메라 보정(calibration)이라고 함[10]).

외부 행렬은 객체와 카메라 좌표계 사이의 상대적인 이동( $t$ ) 및 회전( $R$ ) 변환과 관련된 파라미터를 포함하며, 식 (1)을 통해 알 수 있듯이 일정한 양의 3차원 좌표와 이에 대응하는 2차원 좌표 쌍이 주어지면 구할 수 있다.

$$m = \lambda PM, \quad P = KE = \begin{bmatrix} f_x & \alpha & c_x \\ 0 & f_x & c_y \\ 0 & 0 & 1 \end{bmatrix} [R \quad t]. \quad (1)$$

이처럼 카메라 투영 행렬이 결정되면, 임의의 가상 콘텐츠를 카메라 투영 행렬을 이용하여 투영함으로써 카메라 영상에 삽입할 수 있다. 결국 증강현실을 구현하기 위해서는 카메라나 객체의 움직임으로 인해 실시간으로 변하는 외부 행렬을 추정해야 하는데, 이를 위해서는 관심 객체의 3차원 좌표를 획득하는 기술과 카메라 영상으로부터 관심 객체에 대응하는 2차원 좌표를 획득하는 기술, 주어진 3차원-2차원 좌표를 이용하여 외부 행렬을 계산하는 기술이 필요하다.

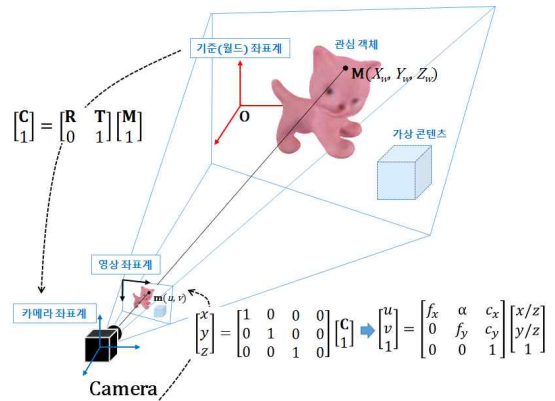


그림 1. 객체의 3차원 좌표와 이에 대응하는 카메라 영상에서의 2차원 좌표 사이의 관계.

앞에서 설명한 대로 관심 객체의 인식 및 추적을 위해서는 카메라 영상으로부터 객체의 2차원 좌표를 찾아야 한다. 이를 위해서 모든 객체 영역의 2차원 좌표를 찾지 않고 일부 특징적인 점, 에지(edge), 영역(region) 등을 찾음으로써 보다 효

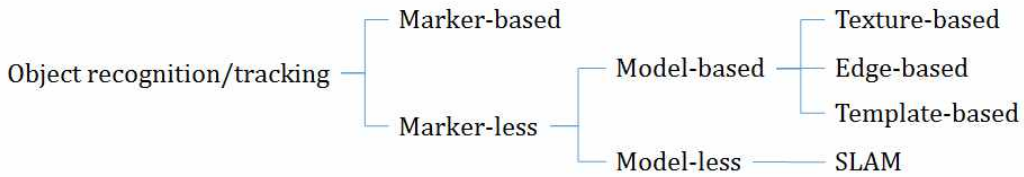


그림 2. 객체 인식 및 추적 기술 분류.

울적인 인식 및 추적이 가능하다. 이러한 특징적인 코너, 에지, 영역 등을 영상 특징 또는 짧게 특징이라고 하며, 영상 내에서 주변과 크게 구별되면서 반복되지 않는 특성을 가진다. 특징을 검출하기 위해서는 주로 픽셀 값의 변화가 큰 지점을 찾으며, 영상의 크기 및 회전 변환, 조명 변화 등에 강인하도록 해야 한다. 일반적으로 각 픽셀에서의 gradient 정보를 이용하여 가로, 세로 방향으로 높은 픽셀 값 변화를 가지는 점 혹은 영역을 찾는다[21-24].

특징 서술은 일반적으로 검출된 특징 주변의 텍스처 정보를 정량적으로 표현하는 특징 벡터를 구하는 과정으로, 특징 검출과 마찬가지로 영상 변환이나 조명 변화에 강인하도록 서술해야 한다. 이를 위해 특징 주변의 일정 영역 내 픽셀 위치에서의 밝기 변화의 히스토그램[22, 23]이나 임의의 두 픽셀 값 사이의 차[25, 26], 특징 위치에서의 픽셀 값 히스토그램[27] 등을 활용한다. 텍스처가 부족한 객체나 장면의 경우, 특징 사이의 기하 관계를 활용하여 특징 벡터를 구할 수도 있다[28].

특징 매칭은 서로 다른 조건에서 촬영된 영상에서 특징 벡터 사이의 차이가 작은 대응(corresponding) 특징들을 찾는 과정으로, 해싱(hashing)이나 트리(tree) 기반 검색 등을 통해 매칭 속도를 향상시킬 수 있다. 대부분의 특징 검출,

서술, 매칭 방법들은 OpenCV 라이브러리[29]에서 오픈 소스로 제공된다.

3차원 모델링은 카메라 영상으로부터 현실세계의 3차원 좌표를 추정하는 것을 말하며, 스테레오 영상 혹은 다시점 영상으로부터 특징을 검출하고, 특징 사이의 대응 관계를 계산하여 대응되는 특징들에 삼각 측량(triangulation)을 적용한다[10]. 다시점 영상은 여러 대의 카메라를 이용하거나 단일 카메라를 움직이면서 촬영함으로써 얻을 수 있으며, 비선형 최적화(bundle adjustment)를 통해 향상된 3차원 모델을 얻을 수 있다[17].

또한, 조명을 이용하여 현실세계에 선 모양의 그림자를 투영하거나 코드 패턴을 프로젝터를 이용하여 현실세계에 투사한 후 카메라로 촬영함으로써 삼각 측량을 위한 대응 특징 쌍을 얻을 수도 있다[15]. 최근에는 깊이 정보를 함께 얻을 수 있는 3차원 카메라[12]가 상용화됨으로써 각 시점에서 현실세계의 부분적인 3차원 모델을 바로 얻을 수 있기 때문에, ICP(iterative closest point)와 같은 알고리즘을 이용하여 각 시점에서의 3차원 모델을 결합한다[16].

## 2.2 객체 인식/추적 기술의 분류

앞서 설명한대로 객체를 인식하고 추적하는 기술은 관심 객체를 찾아 추적함으로써 카메라 투영 행렬의 외부 행렬을 추정하는 과정으로, 다양한 형태로 분류할 수 있으나 본고에서는 그림 3에서 보는 것처럼 마커 사용 유무, 객체의 3차원 모델 보유 여부에 따라 분류하고 각각에 대해 설명한다.

● 마커 기반 인식/추적

마커 사용은 관심 객체의 인식과 추적을 위한 정보 획득을 용이하게 한다. 즉, 그림 4에서 보는 것처럼, 마커는 내부에 고유한 이진 패턴이나 모양을 가지기 때문에 쉽게 인식될 수 있고, 3차원 좌표를 알고 있는 코너(corner)를 가지며, 흑백으로 구성되어 카메라 영상 내에서 간단한 영상처리를 통해 강인하게 대응 코너를 검출할 수 있다. 이러한 장점으로 초기 증강현실 시스템들은 마커를 이용하였으며, ARToolkit[13]과 같은 SDK가 보급되면서 증강현실 구현을 위한 기술 진입 장벽을 크게 낮추었다. 그러나, 가려짐에 대한 취약성, 인위적인 마커의 시각적 불편함(몰입감 저해)이나 유지, 보수에서의 어려움 등으로 인해 최근에는 마커 사용이 지양되고 있다. 다만, 이러한 마커 사용의 문제가 큰 영향을 주지 않는 응용 분야(주로 오락, 교육 분야)에서는 여전히 마커를 사용한 시스템들이 개발, 사용되고 있으며, 일부 가려짐에 대한 취약성을 개선하거나 마커의 시각적 불편함을 해결하기 위한 방법들도 개발되어 왔다[14, 20].

● 텍스처 기반 인식/추적

텍스처 기반 방법은 관심 객체의 3차원 모델이 텍스처 정보를 가져야 하며, 일반적으로 특징 검

출/서술 방법을 활용한다. 3차원 모델과 카메라 영상으로부터 특징을 검출하고 서술하여 대응 관계를 구함으로써 각 특징에 대해 3차원-2차원 대응 쌍을 획득할 수 있고, 이로부터 PnP(Perspective-n-point) 방법을 통해 카메라 포즈를 계산할 수 있다[11]. 관심 객체가 2차원 평면일 경우 3차원 모델 정보를 필요로 하지 않으며, 임의 시점에서의 카메라 영상을 참조 영상( $Z_w$ 가 0인 3차원 모델로 볼 수 있음)으로 하여 현재 시점에서의 카메라 영상과의 특징 대응 쌍을 획득한 후, 호모그래피(homography) 관계를 이용하여 보다 간단하면서도 정확하게 카메라 포즈를 계산할 수 있다[10].

● 에지 기반 인식/추적

관심 객체 자체가 텍스처가 부족하거나 주어진 관심 객체의 3차원 모델이 텍스처 정보를 가지지 않을 경우, 기존 특징 검출/서술을 통한 카메라 포즈 추정은 불가능하기 때문에, 다른 접근 방법이 필요하다. 텍스처가 부족한 객체의 경우 일반적으로 에지를 특징으로 이용하며, 그림 6에서 보는 것처럼 이전 프레임에서의 카메라 포즈( $R, t$ )를 이용하여 3차원 모델을 현재 프레임에 투영한 후, 식 (2)를 이용하여 투영된 모델 에지와 카메라 영상에서 검출된 에지 사이의 오차를 최소화함으로써 이전 프레임과 현재 프레임 사이의 미소 카메라 포즈 변화( $\Delta R, \Delta t$ )를 추정할 수 있다. 추정된 포즈 변화를 이전 프레임의 카메라 포즈에 반영하여 현재 프레임에서의 카메라 포즈를 얻을 수 있다[8]. 복잡한 배경의 경우 많은 에지 성분을 포함하기 때문에, 객체의 경계선을 정확하게 추정해서 추적해야 할 필요가 있다. 이를 위해 객체의 고유한 색상이나 텍스처 정보가 활용될 수 있

다[33].

$$[\Delta \hat{\mathbf{R}} \quad \Delta \hat{\mathbf{t}}] = \arg \min_{\Delta \mathbf{R}, \Delta \mathbf{t}} \sum_i \{ \text{dist} \left( \mathbf{K}[\mathbf{R} \quad \mathbf{t}] \begin{bmatrix} \Delta \mathbf{R} & \Delta \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} \mathbf{m}_i, \mathbf{m}_i \right) \}. \quad (2)$$

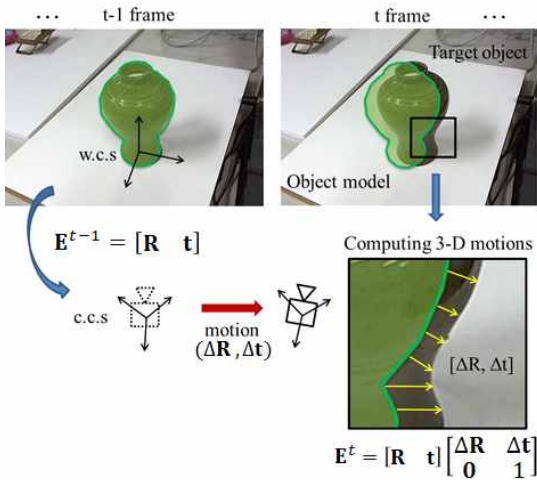


그림 3. 에지 기반 카메라 추적 원리.

● 템플릿 기반 인식/추적

한편, 모델 에지를 추적하지 않고, 식 (3)을 이용하여 투영된 모델과 영상의 픽셀 밝기 차를 반복적으로 최소화하는 영상 정렬(image alignment) 기법을 활용하는 방법들도 있다. 최근에 단순히 픽셀 밝기만을 이용하지 않고 가우시안 미분 값을 이용하거나[36] 카메라 움직임으로 인한 픽셀 밝기 변화를 보다 정교하게 모델링하여 강건하게 카메라 포즈를 추정하는 방법이 제안되었다[34].

$$\Delta \mathbf{E} = \arg \min_{\Delta \mathbf{E}} \sum_{\mathbf{M} \in \mathcal{S}} \|I(\mathbf{P}\mathbf{M}) - J(\mathbf{P}\Delta \mathbf{E}\mathbf{M})\|^2. \quad (3)$$

여기서,  $I(\mathbf{m})$ ,  $J(\mathbf{m})$ 는 이전 프레임과 현재 프레임에서의  $\mathbf{m}$  위치에서의 픽셀 밝기를 나타낸다.

에지 기반이나 템플릿 기반 방법의 경우, 이전

프레임에서의 카메라 포즈를 필요로 하기 때문에, 시작 프레임에서의 카메라 포즈를 구하는 방법이 필요한데, 그래디언트 방위 정보를 이용하여 텍스처가 부족한 객체를 인식하는 방법을 활용할 수 있다[19].

● SLAM(Simultaneous Localization and Mapping)

일반적으로 관심 객체의 3차원 모델을 얻는 것은 쉽지 않고, 미리 모델링을 할 수도 없는 상황일 경우, 3차원 모델링(지도 생성)과 카메라 포즈 추정을 동시에 수행할 수 있다. 이를 위한 방법으로, 현실 세계를 여러 시점에서 촬영한 영상으로부터 추출된 특징을 매칭하고 이를 3차원으로 복원함으로써 3차원 특징점 지도를 생성하고, 생성된 지도를 이용하여 지도상의 카메라 위치와 자세를 계산하는 방법인 Visual SLAM이 제안되었다[30]. 이 방법은 실시간으로 추적이 가능하지만 모든 영상 프레임에 대해 지도 갱신과 추적이 이루어지기 때문에 중복되는 데이터를 처리하는데 많은 연산 시간을 소모하게 되어 정확하고 강건한 지도 갱신 방법을 적용하기 어려운 단점이 있다. 이러한 단점을 개선하기 위한 방법으로, 계산량이 많은 지도 생성에 걸리는 시간을 줄이기 위해 지도 생성을 키 프레임에서만 수행하고(모든 프레임에 대해 지도 생성을 하지 않기 때문에 sparse bundle adjustment를 통해 고속으로 최적화 가능), 추적과 지도 생성을 분리하여 두 개의 쓰레드에서 병렬처리 하도록 함으로써 보다 정확하고 빠르게 SLAM을 수행할 수 있다[18]. 그러나, 초기의 Visual SLAM 방법은 텍스처가 풍부하고 작은 공간에서만 활용할 수 있다는 제약을 가지고 있었으나 최근에는 특징점이 아닌 영상의

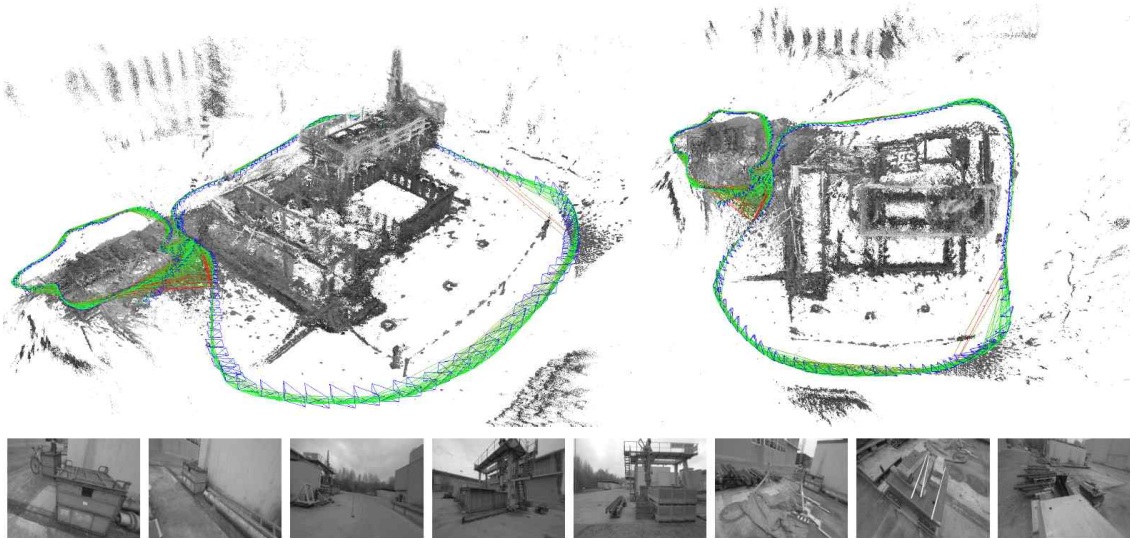


그림 4. 실외 규모가 큰 객체를 대상으로 SLAM 적용 결과. 아래 영상은 일부 키 프레임임.

모든 화소를 사용함으로써 텍스처가 부족한 공간에서 활용 가능한 방법[31, 37]이나 넓은 공간에서 활용 가능한 방법[32]도 제안되었다.

### 2.3 최신 객체 인식/추적 기술 동향

객체 인식 및 추적 기술은 꾸준히 개발, 발전되어 왔으며, 높은 기술 성숙도로 인해 증강현실 응용 분야의 확대에 크게 기여하고 있다. 그러나, 앞서 설명한 비전 기반 방법들은 여전히 기술적 한계를 가진다. 대표적인 예로, 반짝이거나 투명하거나 가변적이거나 빠르게 움직이는 객체를 인식하거나 추적하는 것은 어려우며, 역광이나 야간 환경처럼 아주 밝거나 어두운 조명 환경에서 제약된 동적 범위(dynamic range)를 가진 카메라를 이용하여 객체를 인식, 추적하는 것은 쉽지 않다. 또한, 최근 증강현실 구현을 위한 주요 플랫폼인 모바일 폰이나 HMD의 경우 컴퓨팅 리소스가 부족하기 때문에 정교한 객체 인식 및 추적을 위한 많은 계산량을 실시간에 처리하는 것은 여전

히 불가능하다. 최근에 발표된 논문들은 이러한 기술적 한계를 극복하기 위한 다양한 방법들을 제시하고 있다.

한편, 2016년 ISMAR에서는 비전 기반으로 수 kHz의 빠른 속도로 카메라 추적하는 방법이 발표되었다. 헬멧에 120Hz의 프레임률을 가지는 롤링 셔터 카메라 10대를 전후좌우상 방향으로 스테레오 형태로 배치하고 롤링셔터 문제를 역으로 이용함으로써, 머리의 움직임을 수 kHz의 속도를 가지고 추적할 수 있는 획기적 방법이 제안되었다[51].

### 3. 영상 합성 기술

증강현실은 현실세계 위에 가상세계를 겹쳐놓는 기술이기 때문에 카메라로 받아들인 실사 영상에 컴퓨터로 제작한 영상을 자연스럽게 합성하는 기술이 꼭 필요하다.

자연스런 영상합성을 위해서는 실사 객체를 잘 분할하거나 추출할 수 있어야한다. 그림 6의 좌측

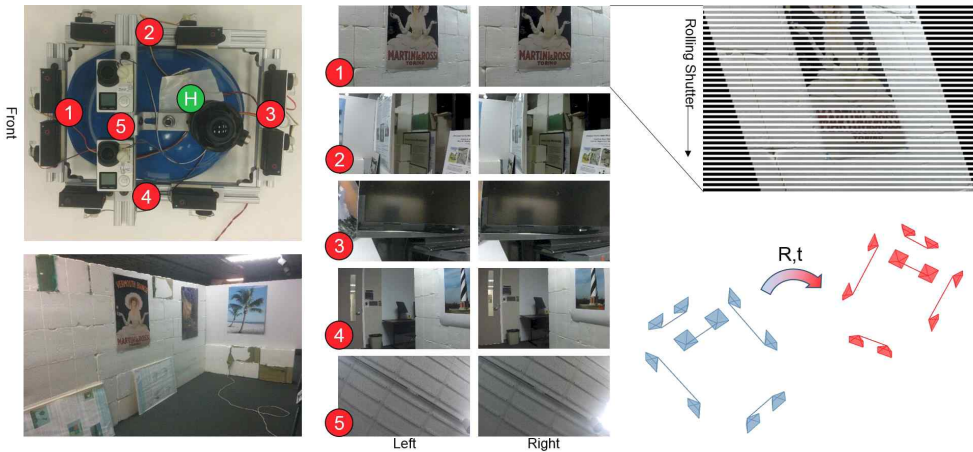


그림 5. 초고속 객체 추적을 위한 시스템 프로토타입.

그림에 스포츠경기에 적용된 증강현실의 사례를 표시하였다. 축구장에서 잔디영역에만 가상객체를 블렌딩하여 합성함으로써 마치 가상객체가 그곳에 실제로 깔려있는 것처럼 표현한 것이다. 기술적으로는 물체영역 추출과 영상블렌딩 기술이 적용되었다.

증강현실은 실사와 CG의 합성 뿐만아니라 실사와 실사의 합성을 요하기도 한다. 하나의 실사에서 객체를 뽑아내서 다른 실사 속에, 즉 다른 환경에 집어넣을 경우에는 새로운 환경에 맞게 실사 객체를 재조명해야한다. 그림 6의 우측에 나타난 예처럼 환경에 맞게 실사 객체를 재조명하거나 물성을 바꿔 재조명하는 등의 기술이 필요하다.

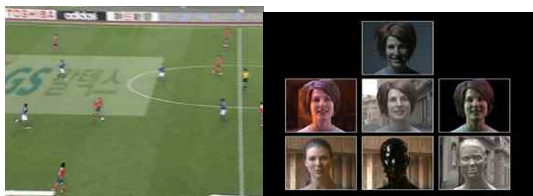


그림 6. 영상합성의 예(좌), 영상기반재조명(우). 어떤 객체가 새로운 환경에 삽입될 경우에는

가장자리의 화소가 배경과 적절히 블렌딩되지 않으면 부자연스런 결과를 얻게 된다. 경계면의 화소에 대해 배경 성분과 어느 정도의 비율로 섞어 줄지를 적절히 결정하면 매우 자연스런 영상을 얻을 수 있다. 그림 7에 나타난 비디오 매팅의 예를 보면, 인물의 가장자리 화소에서 전경 (foreground)  $F$ 의 비율  $\alpha$ 를 구한 후, 새로운 배경  $B$ 과 합성할 때

$$I = \alpha F + (1 - \alpha) B \quad (4)$$

과 같이 적절히 블렌딩함으로써 자연스런 결과를 얻을 수 있다.



그림 7. 비디오 매팅.

영상의 재조명 기술은 최근에 많은 발전을 이루어, 물체 상호간 반사(inter-reflection)나 가상 객체의 그림자가 실사에 반영되는 등의 고차원적 표현이 가능하게 되었다. 그림 8에 ISMAR 2014에서 최우수논문상을 받은 Rohmer et al.의 결과를 제시하였다.



그림 8. 최신 영상기반 재조명 (52)

#### 4. 실시간 상호작용 기술

증강현실 시스템에서는 사용자의 관찰위치에 맞게 적절한 위치에 가상객체가 제시되도록 설계되어 있는데, 단순히 보여주는 차원을 넘어서 사용자가 가상객체와 상호작용을 하는 것은 실감을 구현하는데 필수적인 요소이다. 종래의 컴퓨터 인터페이스처럼 키보드나 마우스를 사용하는데 그치지 않고, 실환경에서 사용자가 자연스런 동작으로 상호작용을 할 수 있도록 만들어주는 다양한 사용자 체험이 제안되어 이용되고 있다.

자연스런 실시간 상호작용을 가능하게 하는

NUI(natural user interface)로 몸동작 인식, 손동작 인식, 얼굴 인식/추적 기술이 많이 사용된다. 몸과 손동작 인식은 과거에 상당히 어려운 기술로 여겨졌으나 Microsoft사에서 Kinect를 내놓은 이후로 비약적인 발전을 이루었다. Kinect의 요체는 실시간 깊이맵 추출과 이를 이용한 사용자 자세 추정 기술이다. 2011년에 Kinect 하드웨어가 나오고, 거기서 나오는 실시간 깊이맵을 사용하여 학습기반의 실시간 신체 자세 추정기술이 제시되면서 몸동작을 이용한 자연스런 상호작용이 가능하게 되었다(그림 9). 일반적으로 RGB영상에서는 사용자의 몸을 추출하거나 손영역을 뽑아내는게 극히 어려운 과제이지만, 깊이맵을 사용하면 비교적 용이하게 배경으로부터 몸과 손을 분리해낼 수 있고, 적절한 불규칙 결정 트리를 이용한 학습을 통해 몸이나 손의 동작을 실시간에 인식할 수 있다 [53].



그림 9. Kinect(좌)에 의한 몸 자세 추정(우).

증강현실 시스템에서 가상객체를 조작하는데 가장 중요한 인터페이스는 손동작이다. 마치 실물을 만지거나 조작하는 것과 똑같이 가상객체를 다룰 수 있도록 하려면 실시간에 정확한 손동작을 얻어내야 한다.

손으로 객체를 만지고, 돌리고, 이동시키는 동작은 비교적 간단한 처리로도 구현할 수 있으며, HMD를 이용하는 증강현실 시스템에 매우 유용



하게 적용될 수 있다. 그림 10에 RGBD카메라를 이용한 효과적인 손동작 인식과 응용의 예를 보여주고 있다 [54].

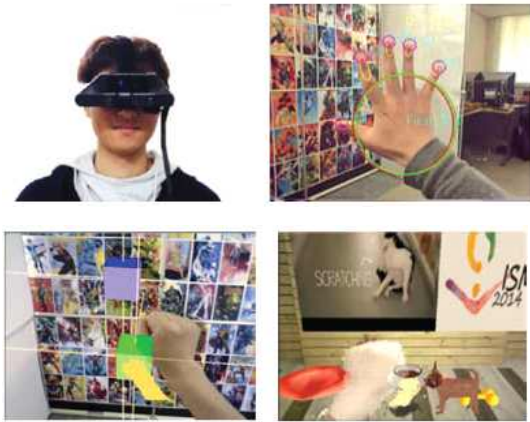


그림 10. HMD환경에서 손동작 인식과 응용.

최근에는 깊이카메라 영상을 이용하여 손가락마다 하나하나의 동작을 실시간에 정확하게 추출하는 기술이 개발되었다. 복잡한 손동작도 실시간에 정확히 추출하는 기술이 나날이 발전하고 있으며, RGBD카메라의 소형화와 함께 다양한 솔루션이 개발되고 있다 [55-56].

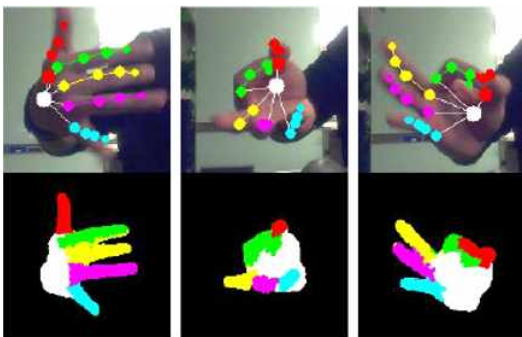


그림 11. 깊이카메라에 의한 복잡한 손동작인식 [55].



그림 12. 딥러닝에 의한 복잡한 손동작인식 [56].

실사 영상으로부터 실시간 모델링을 한 후에 가상객체를 조작하듯이 실사 객체를 조작, 변형하는 기술도 제안되었다 [57]. 그림 13에 보인 것처럼, 이제 현실 공간에 존재하는 물체를 자유롭게 조작, 변형해서 보여주는 것이 가능하게 된 것이다.

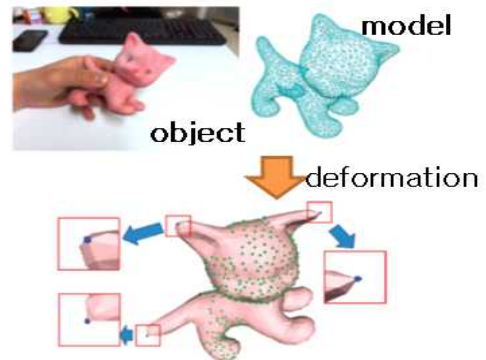


그림 13. 라이브 실사 객체 변형 [57]

## 5. 결 론

본고에서는 증강현실 요소기술의 개념과 원리를 설명하고 최신 동향을 살펴보았다. 증강현실 구현의 핵심기술이라고 할 수 있는 비전 기반의 카메라 추적 기술을 중심으로 자세히 살펴보았으며, 영상 합성과 실시간 상호작용 기술의 원리와 최신 동향에 대해서도 다루었다.

최근에 증강현실이 본격적으로 하나의 산업으로 자리매김하면서 점차 콘텐츠의 중요성이 부각되고 있다. 이제 단순히 객체를 인식하고 추적하는 수준을 넘어서, 실감있는 콘텐츠를 자유자재로 향유하는데 필요한 실감콘텐츠 제작 기술, 자연스러운 실시간 상호작용 기술, 고품질 증강현실 디스플레이 기술 개발에 가속도가 붙어야 할 시점에 도달했다고 판단된다. 본고에서 다루지는 않았지만 앞으로는 현실 공간을 직접 변신시키는 공간증강현실 기술의 중요성도 증대될 전망이다.

## 참 고 문 헌

- [1] 방준성, 이동춘, 서상현, 김용준, 이현주, 손옥호, "VR/AR 게임기술 동향," 전자통신동향분석, 31(1), pp. 146-156, 2016년 2월.
- [2] 김경호, 장중찬, 한상철, "자동차 증강현실 기술 현황과 전망," KEIT PD ISSUE REPORT, 15-3, pp. 13-29, 2015년 3월.
- [3] <http://www.pokemongo.com/>
- [4] 3D융합산업 기술 동향 분석, 경북대학교 3D융합기술지원센터, 2015년 6월
- [5] 엄기열, 김규진, 김문현, "증강현실 환경하에서 비마커 기반 객체 인식 및 추적 기술 동향," 정보과학회지, 28(8), pp. 54-66, 2010년 8월.
- [6] 박영민, 우운택, "증강현실을 위한 인식 및 3D 추적 기술 동향과 전망," 정보처리학회지, 17(4), pp. 40-48, 2010년 7월.
- [7] 박종일, 서병국, 최지윤, 박정식, "실제감과 상호작용의 영상 기술: 3D/UHD, VR, 그리고 AR," 정보과학회지, 28(8), pp. 26-37, 2010년 8월.
- [8] 박한훈, 서병국, 박종일, "모델 기반 카메라 추적 기술 동향," 전자공학회지, 39(2), pp. 66-75, 2012년 2월.
- [9] E. Marchand, H. Uchiyama, and F. Spindler, "Pose estimation for augmented reality: a hands-on survey," IEEE Transactions on Visualization and Computer Graphics, doi:10.1109/TVCG.2015.2513408
- [10] R. Hartley and A. Zisserman, Multiple View Geometry, 2<sup>nd</sup> Ed., Cambridge University Press, 2003.
- [11] V. Lepetit, F. Moreno-Noguer, and P. Fua, "EPnP: An accurate O(n) solution to the PnP problem," IJCV, vol. 81, pp. 155-166, 2009.
- [12] Kinect, <https://developer.microsoft.com/en-us/windows/kinect/>
- [13] ARTToolkit, <https://www.hitl.washington.edu/artoolkit/>
- [14] H. Park and J.-I. Park, "Invisible marker tracking for AR," Proc. of ISMAR, pp. 272-273, 2004.
- [15] 3D scanning using structured lighting, <http://mesh.brown.edu/byo3d/source.html>
- [16] R. A. Newcombe, et al., "KinectFusion: Real-time dense surface mapping and tracking," Proc. of ISMAR, pp. 127-136, 2011.
- [17] D. P. Robertson and R. Cipolla, "Structure from motion," Practical Image Processing and Computer Vision, John Wiley, 2009.

- [18] G. Klein and D. Murray, "Parallel tracking and mapping for small AR workspaces," Proc. of ISMAR, 2007.
- [19] S. Hinterstoisser, V. Lepetit, S. Ilic, P. Fua, and N. Navab, "Dominant orientation templates for real-time detection of texture-less objects," Proc. of CVPR, 2010.
- [20] 박한훈, 최준영, 문광석, "파티클 집단 최적화를 이용한 가려짐에 강인한 마커 기반 증강 현실," 한국HCI학회 논문지, 제11권, 1호, pp. 39-45, 2016년 5월.
- [21] C. Harris and M. Stephens, "A combined corner and edge detector," Proc. of Alvey Vision Conference, pp. 147-151, 1988.
- [22] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," IJCV, vol. 60, issue 2, pp. 91-110, 2004.
- [23] H. Bay, T. Tuytelaars, and L. V. Gool, "SURF: Speeded up robust features," Proc. of ECCV, 2006.
- [24] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," Proc. of ECCV, 2006.
- [25] S. Leutenegger, M. Chli, and R. Y. Siegwart, "BRISK: Binary robust invariant scalable keypoints," Proc. of ICCV, 2011.
- [26] M. Ozuysal, M. Calonder, V. Lepetit, and P. Fua, "Fast keypoint recognition using random ferns," IEEE Trans. on PAMI, vol. 32, no. 3, pp. 448-461, 2010.
- [27] S. Taylor, E. Rosten, and T. Drummond, "Robust feature matching in 2.3us," Proc. of CVPRW, pp. 15-22, 2009.
- [28] T. Nakai, K. Kise, and M. Iwamura, "Use of affine invariants in locally likely arrangement hashing for camera-based document image retrieval", LNCS 3872, pp. 541-552, 2006.
- [29] OpenCV, <http://opencv.org>
- [30] A. Davison, I. Reid, N. D. Molton, and O. Stasse, "MonoSLAM: Real-time single camera SLAM," IEEE Trans. on PAMI, vol. 29, issue 6, pp. 1052-1067, 2007.
- [31] R. A. Newcombe, S. J. Lovegrove, and A. J. Davison, "DTAM: Dense tracking and mapping in real-time," Proc. of ICCV, 2011.
- [32] J. Engel, T. Schöps, D. Cremers, "LSD-SLAM: Large-scale direct monocular SLAM," Proc. of ECCV, vol. 8690, pp. 834-849, 2014.
- [33] B.-K. Seo, H. Park, J.-I. Park, S. Hinterstoisser, and S. Ilic, "Optimal local searching for fast and robust textureless 3D object tracking in highly cluttered backgrounds," IEEE Trans. on Vis. and Comp. Grap., vol. 20, no. 1, pp. 99-110, 2014.
- [34] B.-K. Seo and H. Wuest, "Robust 3D object tracking using an elaborate motion model," Adjunct Proc. of ISMAR, pp. 70-71, 2016.
- [35] A. Torres-Gomez and W. Mayol-Cuevas, "Recognition and reconstruction of transparent objects for augmented reality," Proc. of ISMAR, pp. 129-134, 2014.
- [36] A. Crivellaro and V. Lepetit, "Robust 3D tracking with descriptor fields," Proc. of CVPR, 2014.
- [37] T. Schops, J. Engel, and D. Cremers, "Semi-dense visual odometry for AR on a smartphone," Proc. of ISMAR, 2014.
- [38] C. Paulus, N. Haouchine, D. Cazier, and S. Cotin, "Augmented reality during cutting and tearing of deformable objects," Proc. of ISMAR, 2015.
- [39] N. Haouchine, M.-O. Berger, and S. Cotin, "Simultaneous pose estimation and augmentation of elastic surfaces from a mov-

- ing monocular camera,” Proc. of ISMAR, 2016.
- [40] N. Haouchine, J. Dequidt, M.-O. Berger, and S. Cotin, “Single view augmentation of 3D elastic objects,” Proc. of ISMAR, 2014.
- [41] I. Leizea, H. Alvarez, I. Aguinaga, and D. Borro, “Real-time deformation, registration and tracking of solids based on physical simulation,” Proc. of ISMAR, pp. 165-170, 2014.
- [42] L. Carozza, F. Bosche, and M. Abdel-Wahab, “Visual-inertial 6-DOF localization for a wearable immersive VR/AR system,” Proc. of ISMAR, pp. 257-258, 2014.
- [43] C. Sweeney, J. Flynn, B. Nuernberger, M. Turk, and T. Hollerer, “Efficient computation of absolute pose for gravity-aware augmented reality,” Proc. of ISMAR, 2015.
- [44] X. Yang, X. Si, T. Xue, and K.-T. T. Cheng, “Fusion of vision and inertial sensing for accurate and efficient pose tracking on smartphones,” Proc. of ISMAR, 2015.
- [45] C. Poglitsch, C. Arth, D. Schmalstieg, and J. Ventura, “A particle filter approach to outdoor localization using image-based rendering,” Proc. of ISMAR, pp. 132-135, 2015.
- [46] J. R. Rambach, A. Tewari, A. Pagani, and D. Stricker, “Learning to fuse: a deep learning approach to visual-inertial camera pose estimation,” Proc. of ISMAR, pp. 71-76, 2016.
- [47] I. Ihm, Y. Kim, J. Lee, J. Jeong, and I. Park, “Low-cost depth camera pose tracking for mobile platforms,” Adjunct Proc. of ISMAR, pp. 123-126, 2016.
- [48] I. A. Kakadiaris, M. M. Islam, T. Xie, C. Nikou, and A. B. Lumsden, “iRay: Mobile AR using structure sensor,” Adjunct Proc. of ISMAR, 2016.
- [49] J. Park, B.-K. Seo, and J.-I. Park, “Binocular mobile augmented reality based on stereo camera tracking,” J. of Real-Time Image Processing, 2016. doi:10.1007/s11554-016-0640-9
- [50] P. Fleck, C. Arth, C. Pirchheim, and D. Schmalstieg, “Tracking and mapping with swarm of heterogeneous clients,” Proc. of ISMAR, pp. 136-139, 2015.
- [51] A. Bapat, E. Dunn, and J.-M. Frahm, “Towards kilo-hertz 6-DoF visual tracking using an egocentric cluster of rolling shutter cameras,” IEEE Trans. on Vis. and Comp. Grap., vol. 22, no. 11, 2016.
- [52] Kai Rohmer, Wolfgang Büschel, Raimund Dachselt, and Thorsten Grosch, “Interactive Near-Field Illumination for Photorealistic Augmented Reality on Mobile Devices” Proc. of ISMAR 2014. (IEEE TVCG DOI: 10.1109/TVCG.2015.2450717)
- [53] Shotten et al., “Real-Time Human Pose Recognition in Parts from a Single Depth Image,” Proc. of CVPR 2011.
- [54] Ha et al., “WeARHand: Head-Worn, RGB-D Camera-Based, Bare-Hand User Interface with Visually Enhanced Depth Perception,” Proc. of ISMAR 2014.
- [55] Qian et al., “Realtime and Robust Hand Tracking from Depth,” Proc. of CVPR 2014.
- [56] Sinha et al., “DeepHand: Robust Hand Pose Estimation by Completing a Matrix Imputed with Deep Features,” Proc. of CVPR 2016.
- [57] J. Park and J.-I. Park, “Interactive deformation of real objects,” Proc. of ISMAR 2014.



박 한 훈

- 2000년 한양대학교, 전자통신전파공학과, 공학사
  - 2002년 한양대학교, 전자통신전파공학과, 공학석사
  - 2007년 한양대학교, 전자통신전파공학과, 공학박사
  - 2008년~2011년 일본 NHK방송기술연구소 박사후연구원
  - 2012년~현재 부경대학교 전자공학과 부교수
  - 관심분야: 증강현실, 인간컴퓨터상호작용, 3차원 영상처리/비전 등
- 
- 



박 종 일

- 1987년 서울대학교 전자공학과 공학사
  - 1989년 서울대학교 전자공학과 공학석사
  - 1995년 서울대학교 전자공학과 공학박사
  - 1992년~1994년 일본 NHK방송기술연구소 객원연구원
  - 1995년~1996년 한국방송개발원 선임연구원
  - 1996년~1999년 일본 ATR지능영상통신연구소 연구원
  - 1999년~현재 한양대학교 공과대학 컴퓨터공학부 교수
  - 관심분야: 증강현실, 계산사진학, 3차원 컴퓨터비전, 인간컴퓨터상호작용
- 
-