

특집
05

증강현실을 위한 인식 및 3D 추적 기술 동향과 전망

| 목 차 | |
|-----|-----------------|
| 1 | 초 록 |
| 2 | 증강현실에서의 인식 및 추적 |
| 3 | 추적 기술 동향 |
| 4 | 증강현실의 전망 |
| 5 | 결 론 |

박영민 · 우운택
(광주과학기술원)

1. 서론

증강현실(Augmented Reality)은 컴퓨터에서 생성된 가상의 영상(imagery)을 실시간으로 실제 세계에 덧씌우는 기술로, 사용자를 컴퓨터의 가상공간에 완전히 몰입시키는 가상현실(Virtual Reality)과는 달리 사용자가 실제 객체를 통해 가상의 영상과 상호작용할 수 있도록 한다. 일반적인 증강현실은 시각적 증강뿐만 아니라 청각, 촉각 등의 증강을 통틀어서 의미하지만, 본 논문에서는 시각적 증강에 관련된 기술을 중점적으로 다룬다.

시각적 증강은 카메라를 통해 사용자가 보고 있는 현실의 세계에 가상의 영상을 정확히 덧씌워 보여주는 형태로 구현된다. 일반적으로 증강현실에서의 인식(Detection & recognition) 기술은 카메라의 영상에서 증강하고자 하는 특정 대상을 발견하고 인지하는 일련의 기술을 의미하고, 추적(Tracking) 기술은 이어지는 카메라의 영상에서 카메라 또는 대상 객체의 움직임, 즉, 자세(Pose)의 변화를 정확히 추정해내는 기술이다. 이러한 두 기술은 가상의 영상을 실제 대상

에 정확히 겹쳐서 보여주기 위해 필수적이다.

본 논문은 증강현실 분야에서 소개되고 활용 가능성이 있는 자연 영상 기반 인식, 추적 기술 및 영상 외 센서를 결합한 하이브리드(Hybrid) 기술의 최근 동향을 소개하고 향후 발전 방향에 대해 논의한다. 기존에도 증강현실 기술 관련 동향을 소개한 논문이 있었으나 [1] 일반적인 증강현실 관련 기술 전체를 다루고 있어 인식 및 추적 기술은 간략하게만 소개되어 있다. 또한, 관련된 기반 기술을 소개한 논문이 있었고 [2], 최근의 개발된 기술도 다루게으로는 많은 기술적 부분에서 여기에 소개된 알고리즘을 사용하고 있다. 본 논문에서는 증강현실 분야에서 인식 및 추적 기술의 일반적인 처리 과정과 추고하게 사용되는 기술을 정리하고, 최근에 제안된 기술과 시스템을 소개하여 최신 동향과 전망을 분석한다.

본 논문의 2장에서는 배경 기반 인식 및 추적 기술을 단계별로 소개하고, 3장에서는 최근 발표된 기술 및 시스템을 소개한다. 4장에서는 발전 전망에 대해 논의하고, 5장에서 결론을 맺는다.

증강현실에서의 인식 및 추적

일반적으로 증강현실에서의 인식 기술은 독립 영상에서 데이터베이스에 저장되어 있는 '알려진' 대상을 찾아내는 기술을 말한다. 이와 함께 대상 객체의 3D 중심을 원의 카메라와 객체간의 관계를 알아내는 단축식 알고리즘 수행된다. 추적 기술은 카메라를 통해 입력되는 연속적인 영상에서 대상 객체의 3D 자세 변화를 추적하는 기술이다. 이러한 인식과 추적 기술은 주로 상호 보완적으로 동작한다. 즉, 인식 기술은 영상에서 특정 대상을 찾아낼 수 있는 반면, 이를 통해 획득된 대상의 자세는 상대적으로 불안정하며, 관찰량이 추적 기술에 비해 높고 추적 속도가 제한적이다. 반면, 추적 기술은 대상의 자세 변화를 안정적이고 빠르게 추적하여 실시간 추적이 용이하지만, 추적에 실패할 경우 다시 인식 방법이 필요하다.

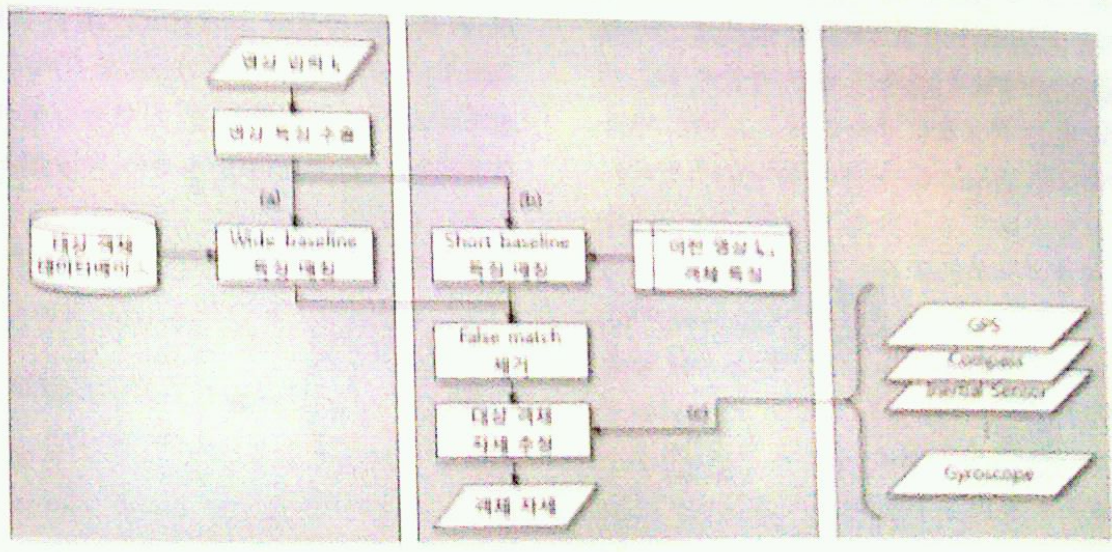
(그림 1)은 일반적으로 증강현실 시스템에서 인식 및 추적 기술이 동작하는 처리 과정을 보여준다. 인식 및 추적은 독립적으로 동작할 수 있고, 혼합되어 사용되기도 한다.

최근에는 멀티-코어 CPU를 이용하여 인식과 추적을 병렬로 동시에 수행하기도 한다[10].

두 기술은 전체적으로 비슷한 역할을 하는 단계를 필요로 하나 영상 특징(Feature) 추출 이후 인식은 데이터베이스와 추출된 특징의 Wide-baseline 매칭을 수행하며, 추적은 이전 영상에서 추출된 특징과의 Short-baseline 매칭을 수행한다. Wide-baseline 매칭은 시점 변화가 매우 큰 두 영상 간 매칭을 수행하며, Short-baseline 매칭은 시점 변화가 크지 않은 경우의 매칭을 의미한다. 두 경우 활용되는 영상 특징이 다를 수 있으며, 이에 따라 각각의 동작 속도가 차이가 나게 된다. 단계별 세부 기술은 아래에 소개되어 있다.

2.1 영상 특징 추출 및 Wide-baseline 매칭

영상 특징 추출과 매칭은 입력된 카메라 영상에서 두드러진 특징을 찾아내고 동일한 대상을 촬영한 다른 시점의 영상에서 일치되는 특징을 찾아내는 기술이다. 따라서 매칭 방법에 따라 필요로 하는 특징이 다르며, 서로 연관되어 개발된 경우가 많다. Wide-baseline 매칭에 많이 사용되



(그림 1) 증강현실을 위한 인식 기반 인식 및 추적 기술의 일반적인 처리 과정. 전체적으로 비슷한 과정을 필요로 하나 인식은 (a) 추적은 (b) 로 구분된다. (c)는 카메라 외 센서를 결합한 하이브리드 추적의 경우를 보여준다

는 알고리즘은 크게 특징점(Feature point) 분류(Classification) 방식과 특징점 서술자(Descriptor) 방식이 있다.

특징점 분류 방식은 미리 학습된 분류기를 통해 입력된 특징점을 여러 개의 클래스 중 하나로 분류하는 방식이다. 이 방식의 대표적인 예는 (그림 2(a))의 Randomized Tree를 이용한 알고리즘과 Ferns가 있다[4][5]. 이를 위해선 인식하고자 하는 특징점을 다양한 시점에서 획득하여 특징점 주변의 영상 패치를 분류기에 입력함으로써 분류기를 학습시켜야 한다. 이 학습 단계는 긴 시간이 소요되므로 실제 증강현실 시스템을 동작하기 이전에 수행된다. 반면, 시스템 수행 중에는 분류기 안에서의 단순한 픽셀 연산을 통해 빠르게 특징점 매칭을 수행할 수 있다. 또한 학습 단계에서 충분한 데이터를 제공할 경우 노이즈, 시점변화, 조명변화 등에 강건하게 매칭을 수행한다.

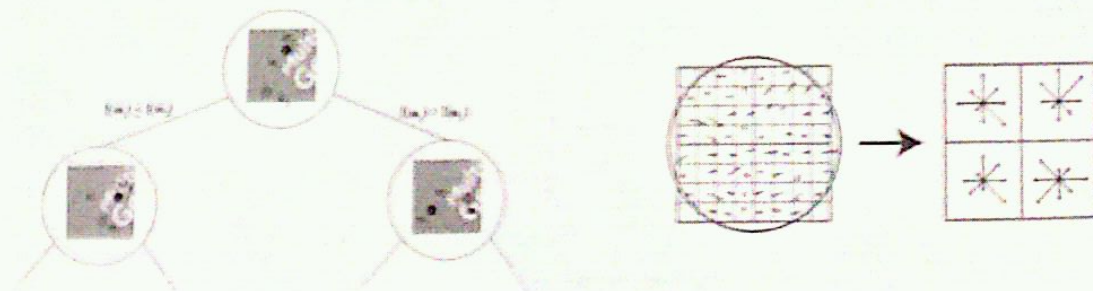
특징점 서술자 방식은 특징점 주변의 영상 패치를 서로 다른 시점에서도 강건하게 매칭할 수 있는 일련의 숫자열, 즉, 서술자로 변환하여 저장한 후 매칭에 사용하는 방식이다. 대표적인 예로는 (그림 2(b))의 SIFT(Scale Invariant Feature Transform) 알고리즘이 있다[3]. 이 알고리즘은 스케일의 변화가 있는 영상에서도 반복성이 좋은 특징점을 추출한 후, 그 주변 패치의 서술자를 만드는 방식이다. 그러나 이 방식은

연산량이 많아 기본 알고리즘의 구현이 아직까지 실시간으로 동작하기 어렵다. 따라서, SIFT 알고리즘에서 시간이 오래 걸리는 특징점 추출 단계를 빠르게 수행하도록 개량한 알고리즘으로 SURF(Speeded Up Robust Features) [26]와 GPU(Graphics Processing Unit)를 이용해 빠르게 동작하도록 구현한 GPUSIFT, SIFTGPU 등이 있다[9][27].

2.2 영상 특징 추출 및 Short-baseline 매칭

Short-baseline 매칭은 Wide-baseline 매칭과 달리 시점 변화가 작은 두 영상에서 동일한 특징점을 찾아내는 것을 목표로 하며, 특징점 추출 방식에 따른 의존도가 상대적으로 낮다. 일반적으로 영상에서의 코너(Corner)를 활용할 수 있으며, Wide-baseline 매칭에서 사용된 포인트 또한 영상에서의 코너인 경우가 많으므로 대부분 활용할 수 있다.

Short-baseline 매칭에는 고전적인 KLT [8] 류의 알고리즘의 알고리즘이 여전히 활용되고 있다. 이들은 영상 패치의 픽셀 강도의 경사도(gradient)의 변화를 통해 연속 영상에서 매칭되는 패치의 위치를 찾아내는 알고리즘이다. 영상이 커질수록 속도가 느려진다는 점을 보완하기 위해 GPU로 구현한 알고리즘으로 GPU KLT도 소개되었다[9]. 그 외에 두 영상에서 인접한 특징점들을 서로 비교하여 오차가 가장 작은 특징



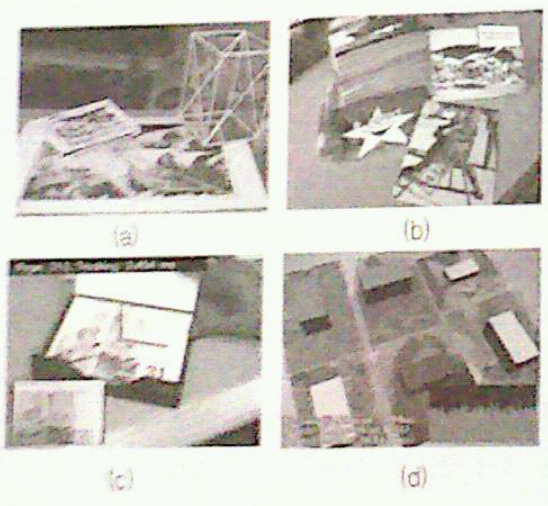
(a) Randomized Tree를 이용한 영상 특징 분류[4]

(b) SIFT에서 영상 패치로부터 서술자 생성[3]

(그림 2) 특징 분류 방식과 특징 서술자 방식의 예

제안되었다[17][18]. 이미지 검색은 수 만 개의 데이터베이스에서 입력 영상과 가장 매칭 되는 키프레임을 짧은 시간 안에 찾아낼 수 있으므로, 영상 내에 존재하는 객체의 키프레임을 객체 인식 이전에 미리 찾아낼 수 있다[18]. [17]은 인식된 특징점을 유지하면서 추적을 수행하여 다수의 평면형 객체를 성능저하 없이 추적하였다. 이들은 이미지 검색 기법의 효과로 300 여 개의 키프레임 데이터베이스를 실시간으로 처리할 수 있다.

[15]는 단일 코어의 저성능 CPU를 내장한 모바일 폰에서 다수의 평면 객체 추적을 구현하기 위하여 객체 인식과 추적을 순차적으로 수행하되 각각이 소요할 수 있는 시간을 한정시켜 프레임 속도를 유지하도록 하였다. 또한, 추적하고 있는 특징점 영역은 다음 영상에서의 객체 인식에서 제외시키고, 패치 투영을 이용하여 강건한 객체 추적 방식을 구현하였다.



(그림 3) 다수 객체 인식 및 추적

- (a) 다수의 3D 객체 인식 및 추적[10]
- (b) 이미지 검색 기반 큰 데이터베이스에서의 특징점 인식 및 추적[17]
- (c) 이미지 검색에 기반 키프레임 선택 및 인식[18]
- (d) 모바일 폰에서 다수의 평면 객체 추적[15]

3.2 병렬 추적 및 맵핑

증강현실에서 일반적으로 사용되는 모델 기반 객체 추적 알고리즘은 추적하고자 하는 객체의 정보를 시스템 동작 이전에 알고 있어야 한다. PTAM(Parallel Tracking and Mapping)은 이와 달리 SLAM(Simultaneous Localization and Mapping)과 마찬가지로 특징점 추적과 동시에 특징점의 3차원 구조를 복원하여 준비되지 않은 환경에서 카메라 자세 추적과 동시에 환경 맵을 생성하는 기술이다[21].

PTAM은 SLAM과 달리 멀티-코어를 활용하여 추적을 수행하는 동시에 백그라운드에서 환경 맵 복원을 동시에 수행하여 빠르고 안정적으로 카메라 추적을 하면서도 정확도가 높은 환경 모델을 복원한다. 단일 코어의 모바일 폰에서 동작하는 경량화된 버전이 개발되기도 했다[19].

특징점만을 활용한 PTAM이 모션 블러에 취약하다는 점을 보완하기 위하여 에지(Edge) 정보를 결합한 알고리즘이 개발되었다[20]. 영상 에지는 모션 블러가 강하게 나타나는 영상에서도 움직임의 방향과 일치하는 방향의 에지는 선명하게 나타나는 점을 이용하였다. 이를 위해 환경 맵 복원 과정에서 특징점 뿐만 아니라 단편적인 에지, 즉, 에젤(Edgel)의 복원을 수행하였다.



(그림 4) 템플릿 기반 추적 알고리즘

- (a) PTAM [21], (b) 모바일 폰에서 동작하는 PTAM[19]
- (c) 에젤을 활용하여 모션 블러에 강건하게 향상[20]

3.3 모바일 및 하이브리드 인식 추적

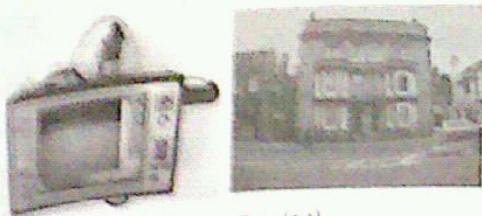
하이브리드 인식 및 추적 기술은 카메라 외에 센서를 함께 사용하여 비전 기반 알고리즘의 한계를 극복하기 위한 방법이다. 최근에는 모바일

가속도 센서 등이 탑재되어 있으나, 추적만 해도 이러한 센서를 직접 사용하지 않아도 된다. [14]은 UMPC에 GPS, 가속도 센서를 장착하여 실외 환경에서 추적하였다. 비전 기반 추적에는 역모션을 적용하였다. 따라서 실외 환경에서 사람이나 자동차 등에 의해 카메라의 움직임이 크거나, 또는 UMPC를 들고 이동하는 사용자 움직임 등에 의해 영상 기반 알고리즘을 사용하는 경우가 있다. 탑재된 센서는 이러한 움직임을 구분하고, 추적의 안정성을 향상시킨다. 추적은 사용자와 초기 위치와 카메라의 자세 정보를 알고 있음을 가정하는데, [24]에 소개된 추적 알고리즘은 빠르게 획득할 수 있는 데이터를 생산하였다. [25]는 이와 유사한 접근법을 사용하여 센서를 사용하고, 센서 간 취약점을 보완한 추적 영상 기법을 제시하였다.

의 블러(Blur)로 인해 특징점 추출이 어려운 경우에도 최적의 객체 자세를 추정할 수 있다.

ESM(Efficient Second-order Minimization)은 템플릿 기반 추적 알고리즘의 자세 최적화 단계에서 2차 미분계수를 고려하여 연산 복잡도를 줄이는 Hessian의 계산을 대체하여 효과적으로 자세 최적화의 수행도를 높인 알고리즘이다[23]. 이 알고리즘은 기존의 방식에 비해 잘못된 자세로 수렴될 경우를 줄여준다.

ESM-Blur는 모션 블러를 고려한 영상 모델을 사용한 모션 블러에 더욱 강건한 템플릿 추적 알고리즘이다[11]. 카메라에서 한 영상을 캡처하는 동안의 객체 움직임으로 인해 모션 블러가 생긴 영상의 모델을 가정하고, 이 모델을 ESM과 동일한 최적화 모델로 유도함으로써 모션 블러에 대한 강건성 뿐만 아니라 추적 속도를 향상시켰다.



(a) Going Out [14]



(b) 멀티-센서 퓨전 [25]



(그림 6) 템플릿 기반 추적. (a) Inverse-compositional 알고리즘 [22], (b) ESM [23], (c) ESM-Blur [11]

4. 양우 발전 전망

앞에서 소개한 바와 같이 최근에는 일정 수준에 도달한 자연 영상 기반 인식 및 추적 기술을 실제 시스템 구현에 적용하기 위한 요소를 고려하고, 많은 사용자들이 접근 가능한 형태로 구현하는데 관심이 집중되고 있다. 이를 정리하면 크게 두 가지 방향을 확인할 수 있다. 첫째는 모바일 플랫폼에서의 구현 및 하이브리드 기술의 구현, 둘째는 추적 시스템의 확장성과 실질적 어려움에 대한 강건성의 향상이다.

최근에는 출시되는 모바일 폰에는 GPS, 가속도 센서, 각속도 센서 및 나침반 센서 등이 탑재

3.4 템플릿 기반 추적

세부론의 추적 알고리즘이 특징점을 이용하는 반면, 템플릿 기반 추적 알고리즘은 영상에서 객체 해당하는 영역 전체를 추적한다[22][23]. 이 방식의 장점은 특징점 추출이 어려운 움직임도 동작할 수 있다는 것이다. 즉, 영상

그림 6 하이브리드 추적 시스템의 예

되어 있다. 따라서 하이브리드 추적 시스템으로써 모바일 폰이 최적의 플랫폼으로 부상하고 있다. 뿐만 아니라 그래픽 가속 기능이 있는 GPU는 이미 많은 제품에 내장되어 있고, 멀티-코어의 CPU를 내장한 제품들이 출시될 예정에 있어 기존의 추적 시스템의 성능 향상은 물론 콘텐츠 렌더링 등이 용이해 질 것으로 예상된다.

그러나 현재 모바일 폰에 구현된 인식 및 추적 시스템은 대부분 속도 향상을 위해 기존 알고리즘의 경량화에 중점을 두고 있으므로 현 시점에서는 매력적으로 보일 수 있지만, 근본적으로 더 나은 성능을 기대하기 힘들다. 또한, 현재의 알고리즘도 여전히 확장성과 실제 동작 환경에 대한 강건성 측면에서 향상될 여지가 많다. 더불어 모바일 폰 자체의 성능이 급격한 향상을 보이고 있으므로 점차적으로는 속도 향상을 위해 다른 중요한 측면을 희생시킨 알고리즘은 활용도가 낮아질 수밖에 없다. 따라서 인식 및 추적의 확장성 및 강건성 향상을 위한 연구는 앞으로도 활발히 이루어 질 것으로 예상할 수 있다.

5. 결론

본 논문에서는 증강현실 시스템에서의 인식 및 추적 기술의 개요와 함께 최근 동향을 소개하고, 향후 전망을 기술하였다. 모바일 폰이 다양한 센서를 탑재하고 점차 고성능의 CPU가 내장됨에 따라 하이브리드 추적 기술의 최적의 플랫폼으로 부상하고 있고, 비전 기반 알고리즘은 확장성 및 실제 환경에 대한 안정성 측면에서 지속적인 연구가 진행 될 전망이다.

참고문헌

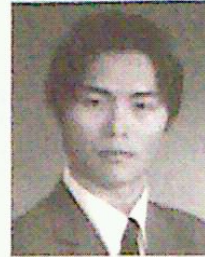
- [1] Zhou, F., Henry B.L. Duh, Billingham, M., "Trends in Augmented Reality Tracking, Interaction and Display: A Review of Ten Years of ISMAR", ISMAR, 2008.
- [2] V. Lepetit and P. Fua, "Monocular Model-Based 3D Tracking of Rigid Objects: A Survey", *Foundations and Trends in Computer Graphics and Vision*, Vol. 1, Nr. 1, pp. 1-89, Oct. 2005.
- [3] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", *International Journal of Computer Vision*, 60, 2, pp. 91-110, 2004.
- [4] V. Lepetit, P. Fua, "Keypoint Recognition using Randomized Trees", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, Nr. 9, pp. 1465-1479, 2006.
- [5] M. Ozuysal, P. Fua, and V. Lepetit, "Fast Keypoint Recognition in Ten Lines of Code", *Computer Vision and Pattern Recognition*, 2007.
- [6] M. A. Fischler and R. C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography", *Comm. Of the ACM* 24: 381 - 395.
- [7] R. Hartley and A. Zisserman, "Multiple View Geometry in Computer Vision (2nd ed.)", Cambridge University Press, 2003.
- [8] B. D. Lucas and T. Kanade., "An Iterative Image Registration Technique with an Application to Stereo Vision", *International Joint Conference on Artificial Intelligence*, pp. 674-679, 1981.
- [9] S. N Sinha, Jan-Michael Frahm, and M. Pollefeys, Y. Genc, "Feature Tracking and Matching in Video Using Programmable

- Graphics Hardware", MVA, Nov. 2007.
- [10] Y. Park, V. Lepetit, and W. Woo, "Multiple 3D Object Tracking for Augmented Reality", International Symposium on Mixed and Augmented Reality, 2008.
- [11] Y. Park, V. Lepetit, and W. Woo, "ESM-Blur: Handling and Rendering Blur in 3D Tracking and Augmentation", International Symposium on Mixed and Augmented Reality, 2009.
- [12] R.T. Azuma, B.R. Hoff, I. Howard E. Neely, R. Sarfaty, M. J. Daily, G. Bishop, L. Vicci, G. Welch, U. Neumann, S. You, R. Nichols, and J. Cannon, "Making augmented reality work outdoors requires hybrid tracking", IWAR, pp. 219-224, 1998.
- [13] G. Klein and T. Drummond, "Robust visual tracking for noninstrumented augmented reality", ISMAR, pp. 113-122, 2003.
- [14] G. Reitmayr and T. Drummond, "Going out: robust model-based tracking for outdoor augmented reality", ISMAR, pp. 109-118, 2006.
- [15] W. Daniel, S. Dieter, B. Horst, "Multiple Target Detection and Tracking with Guaranteed Framerates on Mobile Phones", ISMAR, 2009.
- [16] D. Nister, H. Stewenius, "Scalable Recognition with a Vocabulary Tree", Computer Vision and Pattern Recognition, Vol. 2, pp. 2161-2168, 2006.
- [17] J. Pilet and H. Saito, "Virtually augmenting hundreds of real pictures: An approach based on learning, retrieval, and tracking", IEEE Virtual Reality, Mar. 2010.
- [18] K. Kim, V. Lepetit and W. Woo, "Scalable real-time planar targets tracking for diglog books", The Visual Computer, pp. 1145-1154, 2010.
- [19] G. Klein and D. Murray, "Parallel Tracking and Mapping on a Camera Phone", International Symposium on Mixed and Augmented Reality, 2009.
- [20] G. Klein and D. Murray, "Improving the Agility of Keyframe-based SLAM", European Conference on Computer Vision, 2008.
- [21] G. Klein and D. Murray, "Parallel Tracking and Mapping for Small AR Workspaces", International Symposium on Mixed and Augmented Reality, 2007.
- [22] S. Baker and I. Matthews, "Lucas-Kanade 20 years on: A unifying framework", IJCV, 56(3):221-255, Mar. 2004.
- [23] E. Malis, "Improving vision-based control using efficient second-order minimization techniques", ICRA, 2004.
- [24] G. Reitmayr and Tom Drummond, "Initialisation for Visual Tracking in Urban Environments", ISMAR, 2007.
- [25] G. Schall, D. Wagner, G. Reitmayr, E. Taichmann, M. Wieser, D. Schmalstieg, and B. Hofmann-Wellenhof, "Global pose estimation using multi-sensor fusion for outdoor Augmented Reality", ISMAR, 2009.

- [26] H. Bay, A. Ess, T. Tuytelaars, L. V. Gool, "SURF: Speeded Up Robust Features", Computer Vision and Image Understanding, Vol. 110, No. 3, pp. 346-359, 2008.
- [27] C. Wu, "SiftGPU: A GPU Implementation of Scale Invariant Feature Transform (SIFT)", <http://www.cs.unc.edu/~ccwu/siftgpu/>
- [28] V. Lepetit, F. Moreno-Noguer, and P. Fua, "EPnP: An Accurate $O(n)$ Solution to the PnP Problem", International Journal of Computer Vision, 2008.

※ 본 연구는 문화체육관광부 및 한국콘텐츠진흥원으로 2010년도 문화콘텐츠산업기술지원사업의 연구결과로 수행되었음

저자약력



박영민

2004년 강원대학교 전기전자정보통신공학부(학사)
 2006년 광주과학기술원 정보통신공학과(석사)
 2006년~현재 광주과학기술원 정보통신공학과 박사과정
 관심분야 : 컴퓨터 비전, 증강현실, 유비쿼터스 컴퓨팅
 이 메 일 : ypark@gist.ac.kr



우운택

1989년 경북대학교 전자공학과(학사)
 1991년 포항공과대학교 전기전자공학과(석사)
 1998년 University of Southern California, Electrical Engineering-System 공학과(박사)
 2001년~현재 광주과학기술원 정보기전공학부 부교수
 2005년~2006년 문화기술연구센터 센터장
 2007년~현재 문화기술연구소 연구소장
 관심분야 : 3D 컴퓨터 비전, Attentive AR/MR, HCI, Affective sensing, 맥락인식 컴퓨팅, 유비쿼터스 컴퓨팅
 이 메 일 : wwoo@gist.ac.kr