

논문 2009-46CI-2-11

모바일 증강 현실 및 항공사진을 이용한 건물의 3차원 모델링

(Towards 3D Modeling of Buildings using Mobile Augmented Reality and Aerial Photographs)

김 세 환*, Jonathan Ventura*, 장 재 식**, 이 태 회***, Tobias Höllerer*

(Sehwan Kim, Jonathan Ventura, Jae Sik Chang, Taehee Lee, and Tobias Höllerer)

요 약

본 논문에서는 모바일 증강 현실 시스템 및 항공사진을 이용하여 건물의 부분적 3D 모델을 생성하고, 이를 비디오 영상과 비교하여 사용자의 위치를 실시간으로 추적하는 방법을 제안한다. 제안된 시스템은 미리 생성된 모델을 사용하는 대신, 시스템 동작 중에 사용자 뷰와 항공 뷰를 결합하여 3D 모델을 생성한다. 우선 GPS의 위치에 따라 데이터베이스로부터 검색된 항공사진과, 피치를 추정하는 관성 센서를 이용하여 사용자의 초기 자세를 계산한다. 그리고 그래프 컷을 이용하여 건물의 상단의 에지를 검출하고, 제안된 비용 함수를 최소화함으로써 하단의 에지와 모퉁이 위치를 찾는다. 실시간으로 사용자의 자세를 추적하기 위해, 사용자가 관찰 중인 건물의 에지 및 벽면에서의 특이점을 이용하여 추적을 수행한다. 본 논문에서는 최소 자승 추정법과 언센티드 칼만 필터를 사용하여 카메라 자세 추정 방법을 구현하고 비교하였다. 또한 두 방법에 대한 속도와 정확도를 비교하고, Anywhere Augmentation 시나리오에 대한 중요한 기본 구성 요소들로서 실험결과와 유용성을 보였다.

Abstract

This paper presents an online partial 3D modeling methodology that uses a mobile augmented reality system and aerial photographs, and a tracking methodology that compares the 3D model with a video image. Instead of relying on models which are created in advance, the system generates a 3D model for a real building on the fly by combining frontal and aerial views. A user's initial pose is estimated using an aerial photograph, which is retrieved from a database according to the user's GPS coordinates, and an inertial sensor which measures pitch. We detect edges of the rooftop based on Graph cut, and find edges and a corner of the bottom by minimizing the proposed cost function. To track the user's position and orientation in real-time, feature-based tracking is carried out based on salient points on the edges and the sides of a building the user is keeping in view. We implemented camera pose estimators using both a least squares estimator and an unscented Kalman filter (UKF). We evaluated the speed and accuracy of both approaches, and we demonstrated the usefulness of our computations as important building blocks for an Anywhere Augmentation scenario.

Keywords: Outdoor augmented reality, online modeling, feature-based tracking, UKF, camera pose estimation

* 정회원, 산타바바라 캘리포니아대학 전산과
(Computer Science Department, University of California, Santa Barbara)

** 정회원, 삼성전자 정보통신연구소
(Telecommunication R&D Center, Samsung Electronics)

*** 정회원, 로스앤젤레스 캘리포니아대학 전산과
(Computer Science Department, University of California, Los Angeles)

※ 이 논문은 2006년 정부(교육인적자원부)의 재원으로 한국학술진흥재단(KRF-2006-352-D00165)의 지원과 한국과학기술연구원(KIST) 실감형 차세대 웹 기술개발 과제에 지원을 받아 수행된 연구임

접수일자: 2008년9월16일, 수정완료일: 2009년2월26일

I. 서론

최근, 모델링, 추적 및 렌더링 기술 등의 진보로 증강 현실 시스템들이 우수한 결과를 보이고 있다. 반면, 고가의 장비, 정밀한 측정, 추적을 위한 환경의 사전 준비 등을 요구하는데, 이는 초기 셋업 시간 증가와 함께 사용자에게 부담으로 작용한다. Anywhere Augmentation은 모바일 증강 현실을 확장한 개념으로, 일반적으로 활용 가능한 데이터를 이용하여 셋업 시간을 줄이거나 제거하며, 위치 기반 컴퓨팅 서비스를 물리적 공간과 링크시킴으로써 준비되지 않은 환경에서도 언제, 어디서나 서비스를 사용 가능하게 한다^[1]. 이를 실현시키기 위해서는 이러한 준비되지 않은 환경에서도 실시간으로 사용자의 정확한 자세를 추정하는 것이 중요하다.

다양한 센서를 이용한 추적 기술들이 AR 시스템을 위해 연구되고 있고^[2], 실시간 자세 추정을 위해 다양한 영상 특징들을 사용하여 왔다^[3]. Rosten 등은 빠른 카메라 움직임에 강인하기 위해 에지와 점을 결합한 방법을 제안하였다^[4]. 반면, Reitmayr 등은 텍스처 3D 모델을 사용한 모델 기반 복합 추적 시스템을 개발하고 정확한 정합을 수행했다^[2]. 이러한 대부분의 방법들은 2D 영상과 3D 좌표간의 대응으로 카메라 자세를 계산한다. 그러나 Davison 등은 Simultaneous Localization And Mapping (SLAM)을 로봇 영역에서 핸드헬드 (hand-held) 또는 헤드윈 (head-worn) 카메라 응용분야로 가져왔다^[5]. 본 논문에서는 Mono-SLAM과 모델 기반 AR 추적 방법을 결합하고자 한다.

또한 지도와 항공사진들이 위치 추적을 위해 모바일 시스템에서 사용되고 있다. ARVino는 추상적 정보를 물리적 환경에 매핑하는데 GIS 데이터의 가상현실 뷰에서 항공사진을 사용했다^[6]. Wither 등은 활용 가능한 데이터와 항공사진을 사용하여 옥외 환경에서 주해가능한 모바일 AR을 제안했다^[7]. 본 논문에서는 3D 모델을 생성하는데 있어 항공사진을 사용함으로써 위에 기술된 시스템들의 기능성을 확장시키고자 한다.

제안된 방법은 준비되지 않은 환경에서 건물에 대한 온라인 3D 모델링 기법을 사용한다. 이로써 Davison의 Mono-SLAM과 동일한 목표를 공유한다^[5]. 하지만 보다 원활한 시스템 동작을 위해 간단한 사용자 입력과 일반적으로 사용 가능한 정보를 활용한다. 시스템 동작 중에 (on the fly) 3D 가상 건물 생성을 위해 비디오 영상과 항공사진을 Global Positioning System (GPS)과

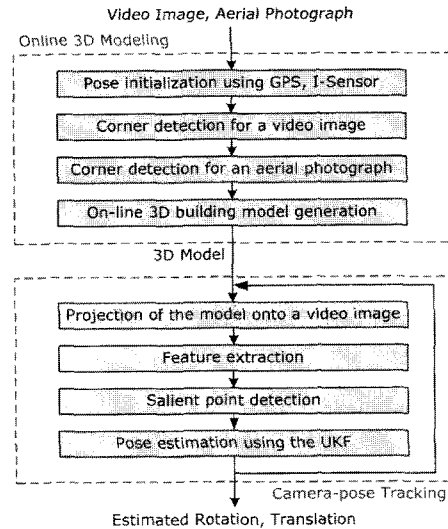


그림 1. 전체 흐름도
Fig. 1. Overall flow diagram.

관성 센서 (inertial sensor)에서 추정된 값들과 사용한다. 이를 위해 비디오 영상과 항공사진에 간단한 마우스 클릭만을 요한다. 그리고 건물 에지와 벽면에서 특이점 (salient point)을 실시간으로 추출하며, 언센티드 칼만 필터 (Unscented Kalman Filter: UKF)를 이용하여 카메라 자세를 추적한다. 최종적으로 보다 넓은 범위의 추적을 위해 랜드마크 (landmark) 기반 추적 방법을 제안하고 에지 기반 추적 방법과 비교한다.

본 논문에서는 첫째, 항공사진과 사용자 뷰 영상에 간단한 마우스 클릭으로 건물의 모델 생성 초기화를 수행한다. 둘째, 건물의 벽면과 에지를 이용하여 실시간 자세 추적이 가능하며, 사용자 뷰에서 건물의 특징들에 대해 실시간 주해 (annotation)가 가능하다. 마지막으로, 비전 기반 추적에 대해 최소 자승 추정법 (Least Squares estimator: LS estimator)과 UKF를 구현, 그 결과를 비교하였고, 또한 에지 기반 추정 방법을 랜드마크 기반 추정 방법과도 비교하였다.

본 논문은 다음과 같이 구성된다. II장에서 온라인 3D 모델 생성 과정을 기술하고 III장에서 생성된 3D 모델과 UKF를 이용한 카메라 자세 추적 모듈을 설명한다. IV장에서 시스템 정확도 및 성능에 대한 실험 결과를 보인 후, V장에서 결과 및 추후 연구에 대해 언급한다.

II. 온라인 3D 모델 생성

본 논문에서는 [8]을 확장하여 사전 정보 없이 사용

자가 건물 주위를 걸으며, 그 위치를 추적하기 위해 시스템 동작 중에 건물을 모델링하는 방법을 제안한다. 그리고 영상의 특이점을 추출하고, UKF 기반 자세 추적 방법을 제안한다. 전체 과정은 그림 1과 같으며, 즉석에서 환경 모델을 생성하고, 모델의 지리적 좌표와 함께 저장하여 추후에 사용하거나 개선 가능하도록 한다.

1. 그래프 컷을 이용한 영상 분할 및 에지 검출

우선, 입방체라고 가정된 건물의 3D 모델 생성을 위해 폭과 깊이, 높이를 추정하는데, 이를 위해 GPS와 항공사진을 이용하여 요 (Yaw)를 추정하고, 관성 센서를 이용하여 건물의 높이를 추정한다^[8].

제안된 시스템의 주요 부분 중 하나는 항공사진에서 건물의 자동 검출이다. 건물 외곽선을 찾기 위해 영상 분할 (segmentation)을 위한 기계 학습을 수행한다. 그리고 사용자 시야 (Field Of View: FOV) 내에서 사용자에게 가장 가까운 건물 코너를 찾고, 그 코너에 관계된 건물 상단에서의 두 에지를 찾는다. 그림 3에서 사용자의 위치와 FOV를 살펴볼 수 있다.

우선, 건물 상단 검출을 위해 항공사진에서 훈련 셋 (training set)을 획득하고, 다변수 가우시안 혼합 모델 (Gaussian Mixture Model: GMM)을 훈련 데이터의 각 그룹에 적용한다. 건물 상단, 풀밭, 포장 도로, 그리고 그 외의 영역으로 구분하고, 각각에 5개의 요소를 갖는 혼합 모델을 사용한다. 여러 그룹에 대한 GMM 학습 후, 그래프 컷 (Graph Cut) 영상 분할 방법을 사용한다^[9]. 최소화하고자 하는 에너지 함수는 다음과 같다.

$$U(\alpha, \theta, z) = U(\alpha, \theta, z) + V(\alpha, z) \quad (1)$$

여기서, α 는 알려지지 않은 이진 영상 분할을 의미하며, θ 는 색 모델 파라미터를 포함한다. 그리고 z 는 영상화소를 나타낸다. α 가 색 분포 θ 에 따라 z 에 가장 적합할 때 U 가 최소화되도록 다음과 같이 정의한다.

$$U(\alpha, \theta, z) = \sum_n -\log p(z_n | \alpha_n) \quad (2)$$

여기서, $p(x; a)$ 는 $a = 1/a = 0$ 일때 x 에서 전경/배경 혼합 모델 값이다. V 는 영상 분할이 에지를 가로지르기 보다는, 그 에지를 따라 수행될 때 최소화된다.

$$V(\alpha, z) = \lambda \sum_{(m,n) \in C} dis(m,n)^{-1} [\alpha_n \neq \alpha_m] \exp\{-\beta \|z_m - z_n\|^2\} \quad (3)$$

여기서, C 는 이웃 화소 쌍들의 집합, $dis(m,n)$ 는 점 m 과 n 간 유클리디언 거리, $[\phi]$ 는 ϕ 가 참일 때는 1, 그 외에는 0이다. $\lambda = 50$, $\beta = (2 \langle (z_m - z_n)^2 \rangle)^{-1}$ 이다. 그리고 $\langle \cdot \rangle$ 는 영상 표본의 기댓값이다. 본 논문에서는 α 에 대한 에너지 최소화 문제를 위해 표준 최소화 컷 알고리즘을 사용하였다^[10].

그 후에는 건물의 모퉁이를 찾기 위해 이진 분할 영상의 컨투어 (contour)들을 검출한다^[11]. 특별히, 잡음을 줄이기 위해 Douglas-Peucker 알고리즘을 사용하여 컨투어를 단순화한다^[12]. 그리고 100개의 화소를 포함하지 못하는 컨투어는 제외시킨다.

따라서 최종적인 건물 모퉁이들은 컨투어 폴리곤에 있는 연속된 선들의 교차점으로 주어지고, GPS로부터의 사용자 위치, 방향, FOV에 의해 주어진 사용자 뷰잉 프러스텀 (viewing frustum)에 대해 대부분의 코너들은 제외된다. 그러면 사용자 위치에 가장 근접한 코너를 찾을 수 있으며, 이를 지나는 두 선을 최종적인 지붕의 에지로 선택한다. 결과를 그림 3에서 확인할 수 있다.

2. 건물 하단에 대한 에지 및 모퉁이 검출

위의 과정들을 수행하기 위해서는 정확한 에지와 건물 아래 모퉁이 검출이 요구된다. 실시간 처리가 가능하도록 마우스 클릭의 주변으로 건물의 모퉁이 위치를 추출하였다. 여기서 건물 하단의 두 선이 건물 상단의 선들과 비슷한 평행선을 유지한다고 가정한다. 최종적으로 다음 에너지 함수를 최소화하는 두 선 $l_1 : y_1 = a_1 x_1 + b_1$ 와 $l_2 : y_2 = a_2 x_2 + b_2$ 를 찾는다.

$$E(\theta_1, \theta_2, x_c, y_c) = - \int_S (\nabla I(x_1 + x_c, y_1(x_1 + x_c))) dx_1 - \int_S (\nabla I(x_2(y_2 + y_c), y_2 + y_c)) dy_2 - \int_{x_c}^{upper} (\nabla I(x_3, y_3(x_3))) dx_3 \quad (4)$$

여기서 $\theta = atan(a)$, S 는 상수, (x_c, y_c) 는 건물 하단의 l_1 과 l_2 의 교차점, (x_{upper}, y_{upper}) 는 건물 상단의 두 선의 교차점이다. 그리고 l_3 는 그 두 교차점간의 선분이다. $\nabla(\cdot)$ 는 주어진 화소의 기울기로서 캐니 에지 검출기 (Canny Edge Detector)에 의해 계산된다. 그림 2(a)는 식 (4)에 사용된 표기법을 보인 것이다.

식 (4)의 최소화값을 찾기 위해, x_c, y_c, θ_1 그리고 θ_2 의 4D 공간을 검색한다. 첫째, 보다 큰 영역인 $x_c \in (x_u - \sigma_L, x_u + \sigma_L)$, $y_c \in (y_u - \sigma_L, y_u + \sigma_L)$ 안에서 2D

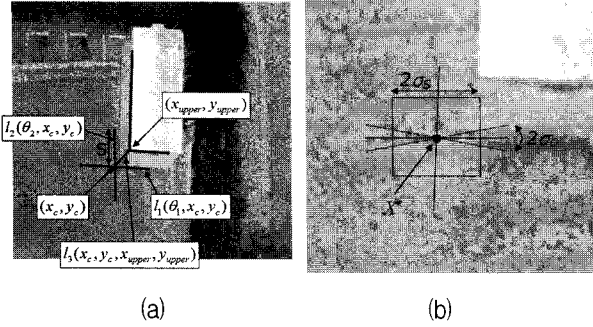


그림 2. 상위 뷰에서의 에지 및 모퉁이 검출 (a) 표기법 (b) 둘째 검색을 위한 영역
 Fig. 2. Edge and corner detection in a frontal view (a) notations (b) search area of the second search.

공간 (x_c, y_c) 를 검색한다. 이 때, 두 선의 기울기 θ_1 과 θ_2 는 건물 상단의 $\theta_{1,upper}$ 과 $\theta_{2,upper}$ 에 의해 고정되며, 검색 영역의 중심은 마우스 입력 값이 된다. 두 번째, 4D 공간은 첫 번째 과정에서 선택된 점 주변의 작은 영역에서 수행되며 그 영역은 $x_c \in (x^* - \sigma_s, x^* + \sigma_s)$, $y_c \in (y^* - \sigma_s, y^* + \sigma_s)$, $\theta_1 \in (\theta_{1,upper} - \sigma_\theta, \theta_{1,upper} + \sigma_\theta)$, $\theta_2 \in (\theta_{2,upper} - \sigma_\theta, \theta_{2,upper} + \sigma_\theta)$ 이다. 여기서, (x^*, y^*) 는 처음 검색으로 선택된 점, σ_s 는 상수이다. x_c, y_c, θ_1 그리고 θ_2 의 검색 범위는 각각 1, 1, 0.5° 그리고 0.5° 이다. 그림 2(b)에 두 번째 단계의 검출 영역을 보였다. 하지만, 때로는 항공사진에서 l_2 가 보이지 않는 경우가 발생하는데, 이때 식 (4)에서 두 번째 선을 고려하지 않는다. 이와 비슷하게 캐니 에지 검출기와 확률적 허프 (Hough) 변환을 이용하여 사용자 뷰에 대해 대응되는 건물의 상단 코너와 에지를 검출할 수 있다^[8].

IV. 카메라 자세 추적

1. 움직임 모델 및 특이점 검출

본 논문에서는 카메라 자세에 대해 6개, 속도에 대해 6개로 총 12개 파라미터를 갖는 등속 모델을 사용하며, 시간 T_k 에서의 상태 벡터는 다음과 같다.

$$\mathbf{x}_k = \begin{bmatrix} t_k \\ r_k \\ v_k \\ \omega_k \end{bmatrix} = \begin{bmatrix} t_{k-1} + v_{k-1} \Delta T \\ r_{k-1} + w_{k-1} \Delta T \\ v_{k-1} + n_{v,k-1} \\ \omega_{k-1} + n_{\omega,k-1} \end{bmatrix} \quad (5)$$

여기서, t_k 및 r_k 는 이동과 회전 벡터, v_k 및 w_k 는 대응되는 벡터의 속도이다. 짧은 샘플링 시간

($\Delta T = T_k - T_{k-1}$) 동안에 위와 같은 시스템 역학을 위한 등속 모델을 사용할 수 있다. n_v 및 n_w 는 이동 속도와 회전 속도의 랜덤 분포 잡음 성분을 의미한다.

건물 벽면의 특징들은 Shi와 Tomasi 알고리즘을 사용하여 추출한다^[13]. 이미 선택된 특징들로부터 일정 거리 이상의 점들만을 선택하는데, 역투영을 통해 3D 공간의 제어점으로 변환된다. 반대로, 에지에서의 특이점들은 카메라 자세를 이용하여 렌더링된 에지의 제어점을 비디오 영상의 에지와 비교함으로써 검출된다.

2. 언센티드 칼만 필터 이용한 에지 기반 자세 추정

3D 모델의 투영을 비디오 영상과 비교함으로써 프레임간의 추적 수행 과정은 비선형 추정 문제로 귀결된다. 우선, 상태 벡터 \mathbf{x}_k 는 식 (5)에 보인 것과 같이 정의하고, 프로세스 모델은 다음과 같이 표현된다.

$$\mathbf{x}_k = F \mathbf{x}_{k-1} + \mathbf{w}_k \quad F = \begin{bmatrix} I_6 & (\Delta T)I_6 \\ 0 & I_6 \end{bmatrix} \quad (6)$$

여기서, \mathbf{w}_k 는 공분산 행렬 Q 에 대한 가우시안 프로세스 (process) 잡음으로, 관찰된 측정 (measurement)을 기반으로 실험적으로 결정된다. 따라서 선형 프로세스 모델에 대해 시간 업데이트 식은 다음과 같다.

$$\hat{\mathbf{x}}_{k|k-1} = F \hat{\mathbf{x}}_{k-1|k-1} \quad P_{k|k-1} = F P_{k-1|k-1} F^T + Q \quad (7)$$

여기서, $P_{k|k}$ 는 T_k 에서 필터의 상태 오류 공분산, Q 는 프로세스 잡음 \mathbf{w}_k 의 공분산 행렬을 나타낸다.

언센티드 변환 (Unscented Transform: UT)에 기반하여, $2L + 1$ 개의 스igma 점 (sigma point)들이 생성되며^[15-16], 매 프레임마다 가변될 수 있는 N_s 개의 특이점들을 다루므로 측정 함수는 다음과 같이 쓸 수 있다.

$$\begin{bmatrix} x_k^i \\ y_k^i \end{bmatrix} = \begin{bmatrix} f_x \frac{R_j(r_k) \cdot [X - t_k]}{R_j(r_k) \cdot [X - t_k]} + x_0 + n_x \\ f_y \frac{R_j(r_k) \cdot [X - t_k]}{R_j(r_k) \cdot [X - t_k]} + y_0 + n_y \end{bmatrix} \quad (8)$$

$(i=0, \dots, N_s - 1)$

여기서, R_j 는 회전 행렬 $R(r_k)$ 의 j 번째 행벡터이다. n_x 와 n_y 는 특징 검출의 측정 잡음, i 는 0부터 $N_s - 1$ 까지의 i 번째 특징을 의미한다.

3. 랜드마크를 이용한 자세 추정

에지 기반 방법은 한 쪽 벽면의 특징 검출이 어려우

면 추적이 실패하는데, 이를 위해 랜드마크 기반 방법을 제안한다. 본 논문에서는 [17]을 확장하여 원거리의 물체 추적을 위해 옥외 환경에 적용한다. 그리고 대상을 두 개 이상의 평면으로 확장하고, 장애물에 대해 강건하기 위해 이상값 (outlier) 제거 부분을 강화한다.

랜드마크를 이용한 실시간 자세 추정을 위해, 멀티쓰레딩을 이용하여 랜드마크 특징점을 추출하는 과정과 추출된 랜드마크들을 추적하는 과정을 분리한다^[17]. 이로써 상대적으로 긴 시간이 소모되는 랜드마크 추출 과정은 느린 속도로 동작하고, 계산 시간이 짧게 걸리는 특징점 추적 과정은 실시간 주기로 동작하게 된다.

제안된 방법에서, 랜드마크 특징점의 추출은 SIFT 변환을 이용하여 키 포인트를 찾는다^[18]. 시점과 크기 변화에 강건한 SIFT 특징점들이 이용되어 자세 추정 과정에서 정확성이 높아지고, 시야를 돌려 다른 곳을 본 이후에 다시 추적 중인 건물을 볼 때에도 기존의 랜드마크를 다시 추출할 수 있게 된다.

추출된 랜드마크 특징점들을 매 영상마다 실시간으로 추적하기 위해서는 피라미드 영상을 통해 효과적으로 optical flow를 계산한다^[14]. SIFT 특징점들은 기존의 optical flow를 계산하는 방법으로 추적하기에 가능할 정도의 텍스처를 가지고 있는 경우가 많아 제안된 방법으로 랜드마크를 실시간으로 추적할 수 있다.

또한 SIFT 특징점과 같은 랜드마크를 추출하는 시간이 영상을 받아들여 처리하는 시간보다 더 길기 때문에, 실시간으로 랜드마크를 추적하기 위해서는 멀티쓰레딩 구조에서 추가적인 처리가 필요하다. 즉 랜드마크 추출이 완료된 시점에서는 이미 기존의 특징점들을 실시간으로 추적하여 상당한 영상이 처리된 상태이기 때문에, 과거의 영상과 현재의 영상 사이에 추출된 랜드마크를 추적하는 과정을 추가하게 된다.

V. 실험

실험을 위해 2.0 GHz CPU IBM T43, PointGrey Dragonfly2 IEEE 1394 카메라, SVGA Sony Glasstron PLM-S700E 디스플레이, InterSense InertiaCube2 회전 추적기, Garmin GPS 18을 사용하였다. 카메라 내부 파라미터는 Zhang의 방법을 이용해 획득하였고^[19], UKF의 파라미터 $L = 12$, $\alpha = 10^{-4}$, $\beta = 2$, $\kappa = -9$ 이다.

그림 3(a)는 제안된 영상 분할 알고리즘 적용 결과이며, 영상 분할된 다른 건물들은 FOV에 의해 제외되었

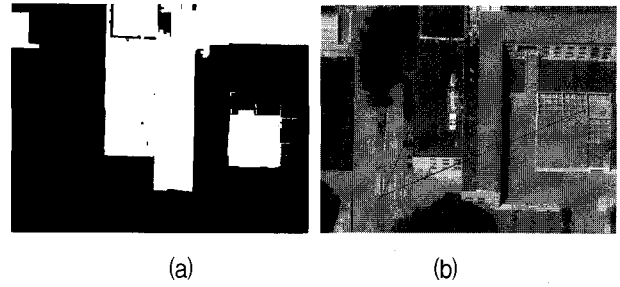


그림 3. (a) 그래프 컷을 이용한 영상 분할 결과와 (b) 에지 및 코너 검출 결과
Fig. 3. (a) Graph cut segmentation result and (b) Result of edge and top corner finding procedure.

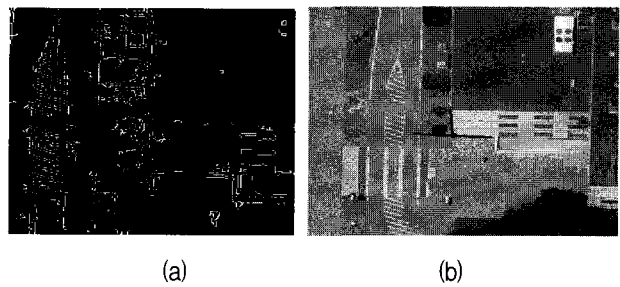


그림 4. 하단의 에지 및 모퉁이 검출 (a) 캐니 에지 및 확률적 허프 선들 (c) 하단 모퉁이 검출 결과
Fig. 4. Edge and corner detection in the bottom part: (a) canny edges and probabilistic Hough lines (b) result of the bottom corner.

다. OpenCV와 maxflow 라이브러리를 사용해서 구현하였는데, 640×480 테스트 영상에서 멀티쓰레드 없이, 영상 분할 및 코너 검출에 대략 3.142초가 소요되었고, 더욱 빠르게 수행하기 위한 연구를 진행 중이다. 그림 3(b)는 그 결과를 보인 것인데, 붉은 색 점은 빌딩 코너, 파란색 선은 지붕의 에지를 나타낸다.

그림 4(a)는 항공사진의 캐니 에지 검출 결과와 확률적 허프 선들, 그리고 그림 4(b)는 모퉁이 검출 결과를 보여준다. 하단 모퉁이 주위의 세 선 중에 한 선이 보이지 않기 때문에, l_2 는 추출 과정에서 제외되고, l_1 만이 그림에 표시되었다. 본 실험에서 사용된 파라미터 값은 $\alpha_L = 10$, $\alpha_S = 3$ 그리고 $\sigma_\theta = 3$ 이다.

1. 정확도

위치 추정의 정확도 평가를 위해, 항공사진으로 획득된 그라운드 트루스 (ground truth) 데이터와 제안된 방법의 결과를 비교하였다. 3D 점 (x, y, z) 에서 x 와 z 좌표는 항공사진의 동쪽과 북쪽, y 는 지표면 위로의 높이를 의미한다. 원점은 그림 4(c)에 보이는 건물의 하단

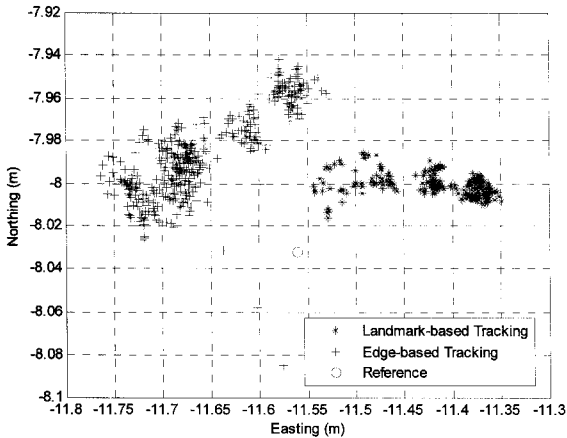


그림 5. 정적 카메라에 대한 추적의 정확도 비교: 원은 항공사진에서 기준 점을 나타낸다.

Fig. 5. Tracking accuracy comparison for a stationary camera pose: Circle denotes reference point from aerial photograph.

모퉁이 부분이다. 그림 5는 이미 알고 있는 위치에서 카메라를 삼각대 위에 놓고, 자세를 추정하는 동안에 카메라의 (x, z) 좌표를 추정한 것이다.

에지 기반 추적과 랜드마크 기반 추적 방법에 대한 평균과 표준편차를 이용하여 실제로 측정된 값인 $(-11.5620\text{ m}, 1.25\text{ m}, -8.0320\text{ m})$ 과 비교하였다. 에지 기반 추적 방법의 경우, 평균 $(\hat{x}, \hat{y}, \hat{z}) = (-11.6568\text{ m}, 1.1290\text{ m}, -7.9862\text{ m})$ 그리고 표준 편차 $(s_x, s_y, s_z) = (0.0591\text{ m}, 0.0952\text{ m}, 0.0207\text{ m})$ 였으며, 랜드마크 기반 추적 방법의 경우, 평균 $(\hat{x}, \hat{y}, \hat{z}) = (-11.4214\text{ m}, 0.9176\text{ m}, -8.0007\text{ m})$ 그리고 표준 편차 $(s_x, s_y, s_z) = (0.0559\text{ m}, 0.0189\text{ m}, 0.0051\text{ m})$ 였다. 측정 결과와 차이를 보인 이유는 캘리브레이션과 센싱 오류 때문이다. 그리고 에지 기반 방법의 오차 분포가 더 큰 것은 SIFT에 의해 검출된 특징들이 상대적으로 안정된 반면, 에지들은 옥외 환경의 외부 조건들에 의해 큰 오차가 발생되기 때문이다.

또한, 이동 중의 정확도 평가를 위해 사용자 자전거 길을 일정하게 걷게 하고, 추정된 카메라 자세를 그림 6(a)에 보였다. 그림 6(b)는 UKF를 사용하여 일정 시간 동안 추정된 위치를 자전거 길과의 거리 오차를 보인 것이다. 이를 통해, UKF가 LS 추정법과 비교해 오차가 작음을 알 수 있으며, LS 추정법은 UKF와 비교해 높은 피크 (peak)를 발생시키는데 이는 LS 추정법이 빠른 움직임에 취약하고, 잡음에 민감함을 의미한다.

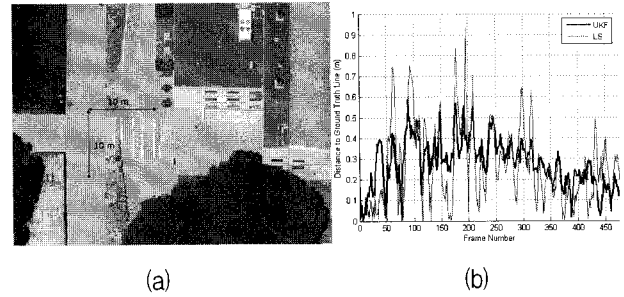


그림 6. 자세 추적에 의해 생성된 카메라 위치의 궤적 (a) 추적 결과 (b) LS 추정법과 UKF의 비교

Fig. 6. A trajectory on an aerial photograph as produced by the pose tracking: (a) tracking results (b) comparison plot of the UKF against the LS estimator results.

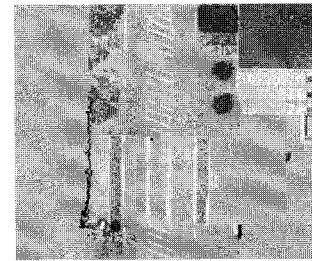


그림 7. 에지 기반 추적 (붉은 색) 및 랜드마크 기반 추적 (검은 색)에 의해 생성된 카메라 위치의 궤적

Fig. 7. Trajectories on an aerial photograph as produced by the edge-based tracking (red line) and the landmark-based tracking (black line).

표 1. 에지 기반 추적과 랜드마크 기반 추적의 비교
Table 1. Comparison between the edge-based tracking and the landmark-based tracking.

방 법	평균 RMS 오류 (m)	표준 편차 (m)
에지 기반 추정	0.42874	0.20940
랜드마크 기반 추적	0.18372	0.08774

그림 7은 두 방법을 적용하여 사용자 자세를 추정한 결과이다. 본 논문에서 제안된 에지 기반 추적 방법은 사용자가 건물 상단 코너를 볼 수 있다고 가정한다. 하지만 사용자가 건물의 코너 및 한 쪽 벽면의 특징 검출이 어려운 위치가 되면 추적이 실패하게 된다. 랜드마크 기반 방법은 이러한 단점을 보완하였으며, 건물의 한 면만을 이용한 추적이 가능하다.

표 1은 그림 7의 추적 결과에 대한 거리 오차의 평균과 표준 편차이다. 랜드마크 기반 추정 방법이 보다 안정되고 정확한 카메라의 위치 추정 결과를 보였다. 이는 이전 실험과는 달리 건물의 코너 주변에서 수행되었으며, 랜드마크 기반 추적 방법이 건물의 한 면만을 이

용해서도 추적 가능함을 보여준다.

2. 성능

현 시스템은 초당 14-17 프레임으로 동작한다. 표 2는 각 단계의 평균 처리 시간으로, 추적에 대한 전체 평균 수행률은 약 25 Hz 정도이다. 즉, 추적 과정만 수행하는 경우에는 약 25 Hz로 수행이 가능하지만, 입력 영상을 획득하고 이에 대한 전처리 과정 등으로 인한 오버헤드로 인해 프레임률이 14-17 Hz로 낮아진다. 반면, 랜드마크 기반 추적 방법에 대해서는 평균적으로 영상 획득에는 31.61 Hz, 렌더링은 30.57 Hz, SIFT는 1.83 Hz, 그리고 추적에는 16.13 Hz의 성능을 보였다.

표 2. 각 과정에 대한 평균 처리 시간
Table 2. Average processing times for individual steps.

각 단계	처리 시간 (ms)
건물 벽의 특징 추적	21.810
에지 추출	0.236
특징점 검색	1.283
언센티드 칼만 필터링	16.636
전체	39.965

VI. 결론 및 추후 연구

본 논문에서는 건물의 일부분에 대한 간단한 모델을 생성하기 위해 컴퓨터 비전 기반 추적을 항공사진, 간단한 사용자 상호작용과 함께 사용함으로써 온라인으로 3D AR 모델을 만드는 방법을 제안하였다. 이를 위해 UKF 프로임퓌를 기반으로 건물의 에지와 벽면에 있는 특이점을 사용하여 3D 모델을 추적하는 시스템을 구현하고 이를 증명하였다. 결과로부터 우리는 사용자가 추적을 위한 충분한 정밀도로 3D 모델을 시스템 동작 중에 생성하고, 그 3D 모델이 옥외 환경의 AR을 위해 실시간으로 추적되는 것을 보였다. 또한 랜드마크 기반 추적을 통해 넓은 영역에까지 추적이 성공적으로 수행됨을 보였다. 현재 여러 가지 추후 연구가 남아있다. 현재의 일반 GPS는 사용자의 위치에 대해 간단하지만 조금은 부정확한 추정치를 제공한다. 사용자의 위치 초기화를 위해서 다른 센서와의 결합을 통해 GPS의 정확도를 향상시켜야 한다. 그리고 사용자의 빠른 움직임에 적절히 반응하기 위해 비전 기반 추적기를 관성 센서와 함께 사용하기 위한 UKF를 실험 중에 있다. $2L+1$ 시그마 점을 사용하는 현재의 UKF 구현 대신에, $L+2$

시그마 점을 이용하는 방법을 고려하고 있는데, 이는 정확도는 유지하면서 계산량을 줄임으로써 성능의 개선을 보일 것으로 기대된다. 현재는 간단한 박스 모양의 건물을 대상으로 하고 있지만, 이를 보다 일반적인 건물에까지 확장할 예정이며, 이미 존재하는 부분적인 모델과 함께 새로운 빌딩 코너들을 결합시킴으로써 큰 건물에 대한 모델링도 가능할 것으로 보인다. 나무, 사람, 차, 가로등과 같은 여러 방해물들에 강인한 시스템을 현재 개발 중이다. 마지막으로, 텍스처 분석 (texture synthesis) 연구를 통해 사용자 움직임에 따라 건물의 부분적인 텍스처만으로도 건물의 부족한 텍스처를 자동으로 계산하는 방법을 연구할 예정이다.

참고 문헌

- [1] T. Hollerer, J. Wither, and S. DiVerdi, Anywhere Augmentation: Towards Mobile Augmented Reality in Unprepared Environments, G. Gartner, M.P. Peterson, and W. Cartwright (Eds.), Location Based Services and TeleCartography, Springer Verlag, 2007.
- [2] G. Reitmayr, and T. Drummond, "Going out: Robust model-based tracking for outdoor augmented reality," Proc. of IEEE ISMAR'06, pp. 109-118, 2006.
- [3] G. Simon, A. Fitzgibbon, and A. Zisserman, "Markerless tracking using planar structures in the scene," Proc. of ISAR'00, pp. 120-128, 2000.
- [4] E. Rosten, and T. Drummond, T., "Fusing points and lines for high performance tracking," Proc. of ICCV'05, pp. 1508-1511, 2005.
- [5] A. Davison, W. Mayol, and D. Murray, "Real-time localisation and mapping with wearable active vision," Proc. of ISMAR'03, 2003.
- [6] G. King, W. Piekarski, and B. Thomas, "Arvino - outdoor augmented reality visualisation of viticulture gis data," Proc. of IEEE ISMAR'05, pp. 52-55, 2005.
- [7] J. Wither, S. DiVerdi, and T. Hollerer, "Using aerial photographs for improved mobile ar annotation," Proc. of IEEE ISMAR'06, pp. 159-162, 2006.
- [8] S. Kim, S. DiVerdi, J. S. Chang, T. Kang, R. Iltis, and T. Hollerer, "Implicit 3D Modeling and Tracking for Anywhere Augmentation," VRST'07, pp. 19-27, 2007.
- [9] C. Rother, V. Kolmogorov, and A. Blake,

“grabcut: interactive foreground extraction using iterated graph cuts,” ACM Trans. Graph, 23(3), pp. 309-314, 2004.

[10] Y. Boykov and M.-P. Jolly, “Interactive graph cuts for optimal boundary & region segmentation of objects in n-d images,” IEEE ICCV’01, pp. 105-112, 2001.

[11] Intel, OpenCV Library, <http://www.intel.com/technology/computing/opencv/>, 2007.

[12] J. Hershberger and J. Snoeyink, “Speeding up the douglas-peucker line simplification algorithm,” Spatial Data Handling. IGU Commission on GIS, pp. 134-143, 1992.

[13] J. Shi, and C. Tomasi, “Good features to track,” Proc. of CVPR’94, pp. 593-600, 1994.

[14] B. Lucas, and T. Kanade, “An iterative image registration technique with an application to stereo vision,” Proc. of the International Joint Conference on AI, pp. 674-679, 1981.

[15] S. Haykin, Kalman Filtering and Neural Networks, John Wiley & Sons, Inc., 2001.

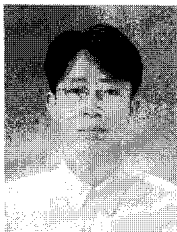
[16] S. Julier, and J. Uhlmann, “Unscented filtering and nonlinear estimation,” Proc. of the IEEE, vol. 92, pp. 401-422, 2004.

[17] T. Lee and T. Höllerer, “Hybrid Feature Tracking and User Interaction for Markerless Augmented Reality,” Proc. of IEEE VR’08, pp. 145-152, 2008.

[18] D. G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints,” IJCV, vol. 60, no. 2, pp.91 - 110, 2004.

[19] Z. Zhang, “A flexible new technique for camera calibration,” IEEE Transactions on PAMI, 22, 11, pp. 1330-1334, 2008.

저 자 소 개



김 세 환(정회원)
1998년 서울시립대 전자공학과
학사 졸업
2000년 광주과학기술원 정보통신
공학과 석사 졸업
2006년 광주과학기술원 정보통신
공학과 박사 졸업

현재 UCSB 전산과 박사후과정
<주관심분야 : 증강/혼합 현실, 3차원 컴퓨터 비
전, 컴퓨터 그래픽스, 휴먼-컴퓨터 상호작용>

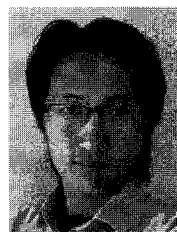


Jonathan Ventura
2005년 UCSB 전산과 학사 졸업
현재 UCSB 전산과 박사과정
<주관심분야 : 증강/혼합 현실, 컴
퓨터 비전, 3차원 복원>



장 재 식(정회원)
2001년 금오공과대학교 컴퓨터
공학과 학사 졸업
2003년 경북대학교 컴퓨터공학과
석사 졸업
2005년 경북대학교 컴퓨터공학과
박사 졸업

현재 삼성전자 정보통신연구소 연구원
<주관심분야 : 컴퓨터 비전, 휴먼-컴퓨터 상호작
용, 증강 현실>



이 태 희(정회원)
2001년 KAIST 전산과 학사 졸업
2007년 UCSB 전산과 석사 졸업
현재 UCLA 전산과 박사과정
<주관심분야 : 컴퓨터 비전, 3차원
복원, 비주얼 추적, 증강 현실>



Tobias Höllerer
1995년 Technical University
Berlin, Informatics
학사 졸업
1997년 Columbia University
전산과 석사 졸업
2004년 Columbia University
전산과 박사 졸업

현재 UCSB 전산과 부교수
<주관심분야 : 증강/가상 현실, 3차원 디스플레이
및 상호작용, 모바일 및 착용형 컴퓨팅, 적응형 사
용자 인터페이스>