

## 수직이착륙 무인항공기 자동 착륙을 위한 영상기반 항법

이상훈<sup>\*,1)</sup> · 송진모<sup>1)</sup> · 배종수<sup>1)</sup>

<sup>1)</sup> (주)한화 종합연구소 핵심기술 2팀

### Vision-based Navigation for VTOL Unmanned Aerial Vehicle Landing

Sang-Hoon Lee<sup>\*,1)</sup> · Jin-Mo Song<sup>1)</sup> · Jong-Sue Bae<sup>1)</sup>

<sup>1)</sup> Sensor Systems Department, Defence R&D Center, Hanwha Corporation, Korea

(Received 1 December 2014 / Revised 6 April 2015 / Accepted 22 May 2015)

#### ABSTRACT

Pose estimation is an important operation for many vision tasks. This paper presents a method of estimating the camera pose, using a known landmark for the purpose of autonomous vertical takeoff and landing(VTOL) unmanned aerial vehicle(UAV) landing. The proposed method uses a distinctive methodology to solve the pose estimation problem. We propose to combine extrinsic parameters from known and unknown 3-D(three-dimensional) feature points, and inertial estimation of camera 6-DOF(Degree Of Freedom) into one linear inhomogeneous equation. This allows us to use singular value decomposition(SVD) to neatly solve the given optimization problem. We present experimental results that demonstrate the ability of the proposed method to estimate camera 6DOF with the ease of implementation.

Key Words : VTOL(수직이착륙), Navigation(항법), Computer Vision(컴퓨터 비전), Homography(호모그래피), Epipolar Constraint(에피폴라 제약조건)

#### 1. 서론

영상 기반의 위치 인식 방법은 다양한 연구 분야에서 관심을 받아왔다. 재난 상황 또는 인간이 접근하기 힘든 지역에 투입되는 로봇의 위치 인식 및 지도 작성(SLAM)을 위한 분야, 인식된 위치를 이용하여 3차원

형상복원(3D reconstruction)을 위한 분야에도 활발한 연구가 진행되고 있다. 군사적인 목적으로는 무기체계에 탑재되어 GPS를 사용하기 어려운 상황에서 더욱 정밀한 위치 인식을 위해 활용되고 있다.

영상 기반의 위치 인식 방법은 크게 두 가지 연구 방향으로 나누어 볼 수 있는데, 그 첫 번째는 A. J. Davison이 제안한 확장 칼만 필터(EKF)기반의 방법이다<sup>[1]</sup>. 이 방법은 단안 카메라에서 카메라의 위치와 자세 그리고 움직이지 않는 3차원 공간의 특징점을 칼만

\* Corresponding author, E-mail: shleebot@hanwha.com  
Copyright © The Korea Institute of Military Science and Technology

필터의 상태벡터( $x, y, z, q_0, q_1, q_2, q_3, f_1, f_2, \dots$ )로 정의하고 영상에서 측정된 특징점을 이용하여, 상태벡터를 추정하는 과정으로 이루어진다. 이 방법은 주변 환경을 모르는 상태에서도 상태벡터와 3차원 특징점의 대응관계 해석을 통해 일관된 위치 인식과 지도 작성을 할 수 있지만, 특징점의 개수에 따라 연산량이 기하급수적으로 증가한다는 단점이 있다. 다른 방법의 연구는 인접 영상의 두 시점이 가지는 제약사항(epipolar constraint)을 이용하여 카메라의 위치와 자세 3차원 특징점을 복원하는 방법(Structure from Motion : SfM)이다. 투영기하학(projective geometry)를 컴퓨터 비전에 적용을 시도하는 여러 연구들이 있었으며 대표적으로 R. Hartley, A. Zisserman이 다시점 영상 기하학에 관한 이론을 정립하였다<sup>[2]</sup>. D. Nister는 3차원의 3개 이상의 점과 그것의 영상에 투영된 점을 이용하여 카메라 위치와 자세를 복원하는 방법을 제안하였고<sup>[3]</sup>, N. Snavely는 불특정 다수의 사용자가 업로드한 영상을 기반으로 피사체의 3차원 형상을 복원하는 방법을 제안하였다(photo tourism)<sup>[4]</sup>. 최근에는 두 가지 방법론을 혼용하여 적은 수의 특징점으로 상태벡터를 추정하고 많은 수의 특징점으로는 더욱 정확한 카메라 위치, 자세를 계산하는 방법도 제안되었다<sup>[5,6]</sup>.

본 논문에서는 적은 수의 알고 있는 3차원 특징점과 많은 수의 인접영상의 대응점을 모두 활용하는 방안을 제시한다. 카메라가 장착되어 있는 물체는 짧은 시간의 간격에서 상대 변위가 유지된다고 가정할 수 있으므로 측정 가능한 모든 조건을 만족시키는 해를 찾는다면 그 순간의 카메라 위치를 가장 잘 찾는 결과를 얻게 된다. 이 과정에서 제안한 방법은 행렬 해법을 통해 구현의 용이성과 풀이의 간결성을 유지한다.

본 논문은 다음 순서로 구성되어 있다. 2장에서는 다시점 영상기하학에서 알고 있는 3차원 특징점과 인접한 영상의 대응점 탐색으로 카메라의 위치와 자세를 얻는 방법을 기술한다. 3장에서는 수직이착륙 무인항공기에 적용하기 위한 랜딩마크의 설계와 제안하는 방법을 기술하고, 4장에서 모사 생성 데이터와 실제 영상을 기반으로 수행된 실험 결과를 보이고, 5장에서 결론과 추후 연구 과제를 기술한다.

## 2. 이론 해석

### 2.1 알고 있는 3차원 특징점과 카메라

투영기하학에서 사용되는 핀홀 카메라 모델(pin-hole camera)모델에서 3차원 공간의 특징점  $X$ 를 2차원 영상 평면의 특징점  $x$ 로 투영시키는 행렬을  $P$ 로 정의하면 식 (1)의 관계식이 성립한다.

$$x = PX \tag{1}$$

여기서 투영 행렬(projection matrix)  $P$ 는  $3 \times 4$  행렬이며, 카메라 내부 변수  $K$ 와 외부 변수인 회전 행렬  $R$ 과 병진 이동 행렬  $t$ 의 결합  $K[R|t]$ 로 이루어진 행렬이다. 투영 행렬의 자유도(DOF)는 11개 이므로 6개의 대응점으로부터 정의할 수 있는  $2 \times 6$ 개의 식을 이용하면 DLT(Direct Linear Transform)을 통해 그 해를 구할 수 있다.

만약 여기서 알고 있는 점들의 집합이 모두 하나의 평면에 있다고 가정할 수 있다면 식 (2)와 식 (3)의 관계식이 만족한다. 여기서,  $p_1 \sim p_4$ 는 투영 행렬의 각 열을 요소를 의미한다. 이 경우에는 3차원 공간의 평면과 2차원 영상 평면과의 호모그래피(Homography)를 구하는 것이 곧 투영 행렬을 구하는 것과 같다는 것을 알 수 있다.

$$[p_1 p_2 p_3 p_4] \begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix} = [p_1 p_2 p_4] \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \tag{2}$$

$$P = H = K[R|t] = K[r_1 r_2 t], r_3 = r_1 \times r_2 \tag{3}$$

### 2.2 인접 영상의 대응 관계 해석과 카메라

인접 영상의 대응 관계 해석을 통해서도 카메라 외부 변수를 얻을 수 있는데, 두 장의 영상이 만족해야 하는 에피폴라 제약조건(epipolar constraint)을 이용한다. 카메라 내부 변수를 알고 있다고 가정할 수 있는 경우 펀더멘탈 행렬(fundamental matrix)과 에센셜 행렬(essential matrix)은 식 (4)와 식 (5)의 관계식을 만족한다. 단,  $E$ 는 에센셜 행렬,  $F$ 는 펀더멘탈 행렬이다. 에센셜 행렬은 두 장의 영상으로부터 4개 이상의 대응점이 존재하는 경우 찾을 수 있다<sup>[2]</sup>.

$$\hat{x}'^T E \hat{x} = 0 \tag{4}$$

$$E = K^T FK \tag{5}$$

에센셜 행렬은 영상 평면에서 해석하는 펀더멘털 행렬과 달리 영상 좌표에 카메라 내부 변수의 역행렬을 영상 좌표에 곱하여 사용하므로 정규화된 형태로 생각할 수 있으며, 에센셜 행렬은  $E = [t]_{\times} R$  이므로 첫 번째 카메라  $P = [I|0]$ 에 대해 다음 네 개의 가능한 행렬 분해를 얻을 수 있다.

$$P' = [UWV^T | +u_3] \text{ 또는 } [UWV^T | -u_3], \tag{6}$$

$$[UW^T V^T | +u_3] \text{ 또는 } [UW^T V^T | -u_3]$$

여기서  $W = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ ,  $Z = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ 는 각각 직교행

렬(orthogonal matrix), 왜대칭행렬(skew symmetric matrix)이고,  $U, V$ 는 각각 특이값 분해(SVD)의 행렬 분해값이다. 또한,  $u_3$ 는  $U$ 의 3번째 열 벡터를 의미한다.

### 2.3 관성력

충분히 연속적인 운동을 하는 카메라에 대한 해석을 하는 경우에 충분히 작은 시간 내에서는 등속으로 운동한다고 생각할 수 있다. 왜냐하면, 병진 이동과 회전 운동에 대한 매우 작은 시간 변화량에 대한 미분치는 변화량에 크게 영향을 미치지 않는 고차항을 소거하면 선형적으로 수렴할 것이 예측되기 때문이다. 그러므로, 바로 이전 시간  $t-1$ 의 병진 이동 속도와 회전 각속도를 이용하면 다음 시간  $t$ 의 카메라의 위치와 자세를 예측하는 것이 가능하다. 병진 이동 속도 벡터를  $v$ , 회전 각속도 벡터를  $\omega$ 라 하면, 식 (7)로부터 이전에 구한 호모그래피의 형태로 변환할 수 있다.

$$R_t = R_{t-1} + \text{rod}(\omega)$$

$$t_t = t_{t-1} + v \tag{7}$$

$$H_t = [r_1 \ r_2 \ t]$$

여기서  $\text{rod}()$ 는 3축 오일러 각속도 성분을  $3 \times 3$  회전 행렬로 변환하는 연산이다.

## 3. 제안하는 방법

### 3.1 랜딩 마크의 설계

수직 이착륙 무인 비행체가 관측할 수 있도록 랜딩 마크를 설계하기 위해서는 외부 환경(조명, 가림)에 강인하도록 해야 한다. Fig. 1은 설계된 랜딩 마크와 원점, 실측치를 기록한 그림이다.

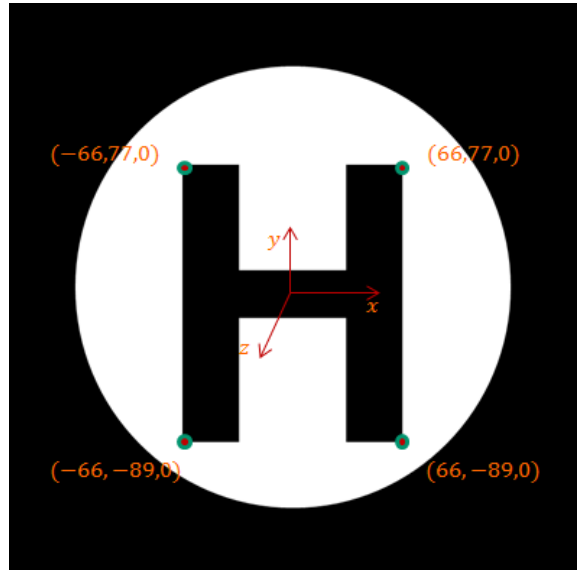


Fig. 1. Landing mark

### 3.2 최적화 함수 정의

앞서 2장에서 기술한 내용을 바탕으로 세 가지 조건을 모두 만족하는 최적화 함수는 다음 식 (8)과 같이 정의할 수 있다.

$$E(P) = (PX_g - x_g)^2 + \lambda_e (x_1 E x_2)^2 + \lambda_i (P_i - P)^2 \tag{8}$$

여기서  $X_g$ 와  $x_g$ 는 알고 있는 4개 이상의 점에 대한 3차원 좌표와 영상 평면에 투영된 점을 의미하고,  $\lambda_e, \lambda_i$ 는 인접영상으로부터 얻은 것과 관성력에 의해 예측된 카메라 외부 변수 각각의 스케일 인자이다. 제시된 최적화 함수의 파라미터는 경사 하강법(gradient descent), 가우스-뉴턴(Gauss-Newton), LM(Levenberg-Marquardt) 방법과 같은 최적화 기법을 적용하여 구할 수 있다. 본 논문에서는 다음절에서 제시하는 행렬 연산을 통해 구했다.

### 3.3 최적화 인자 계산식

선형 비동차 방정식(linear inhomogeneous equation)으로 주어진 문제를 풀기 위해서 다음 식 (9)와 같은 형태로 변형한다.

$$\begin{bmatrix} 0^T & -w'_i x_i^T & y'_i x_i^T \\ w'_i x_i^T & 0^T & -x'_i x_i^T \\ & \vdots & \\ & \lambda_c I & \\ & \lambda_i I & \\ A & & \end{bmatrix} \begin{bmatrix} h^1 \\ h^2 \\ h^3 \end{bmatrix} = \begin{bmatrix} 0 \\ h_e \\ h_i \end{bmatrix} \quad (9)$$

$$(26 \times 9) \quad X = b \quad (9 \times 1) \quad (26 \times 1)$$

식 (9)의 좌변 행렬  $A$ 의 1~8행은 알고 있는 3차원 점과 영상간의 대응점으로부터 유도된 식과 두 개의  $9 \times 9$  단위행렬로 구성된다. 우변은  $8 \times 1$ 의 영행렬과  $h_e, h_i$ 로 구성되어 있는데, 각각 에센셜 행렬과 관성력에 의해 예측된 카메라 외부변수를 호모그래피 형태로 변환한 식의 1~3열을  $9 \times 1$ 의 행벡터로 쌓은 형태이다.

식 (9)는  $AX=b$ 형태의 식이므로  $A$ 행렬의 의사 역행렬(left pseudo inverse)  $A^+ = (A^T A)^{-1} A^T$ 를 양변에 곱함으로써  $X$ 를 구할 수 있다. 이때 SVD를 사용하면  $A^{-1} = V^T D^* U^T$ 이므로 한 번의 연산으로 해를 구할 수 있다.

## 4. 실험 결과

### 4.1 모사 데이터 생성 및 시뮬레이션 결과

제안된 방법의 성능을 평가하여 보기 위해 카메라의 위치와 자세를 알고 있는 모사 데이터를 생성하였다. Table 1에 보인 입력으로 만들어진 카메라 위치의 궤적은 Fig. 2와 같다. 모사 카메라의 위치는 나선형의 궤적을 그리게 된다. 카메라의 최초 위치는 (0, 100, 800)으로 설정되었고, 3차원 위치를 알고 있는 점 4개와 인접 영상에서 찾은 대응점을 10개로 설정하였다. 카메라의 내부변수는 알고 있는 것으로 가정 하였다. 설정된 점은 모사카메라의 자세를 알고 있으므로  $720 \times 480$  영상에 투영시킬 수 있고, 영상에서 측정된 점은 3픽셀의 분산을 가지는 정규분포 노이즈(자체 실험치 기준)를 더하였다.

Fig. 3 ~ Fig. 6은 1) 참값, 2) 알고 있는 점으로 추정된 위치, 3) 인접영상으로부터 추정된 위치, 4) 2)와 3) 방법만으로 추정된 결과, 5) 제안한 방법의 결과를 보

인 것이다. Fig. 7은 Fig. 3의 궤적 부분을 확대하여 보인 것이다. 알고 있는 점을 통해 추정된 위치는 드리프트 없지만, 측정 노이즈로 인하여 참값 주변에서 가장 큰 오차를 가지고 있으며, 인접영상으로부터 얻은 값은 누적 오차로 인하여 드리프트를 가지고 있는 것을 확인할 수 있다. 반면에 제안한 방법은 참값을 다른 방법에 비해 우월하게 추종하고 있다.

Fig. 8과 Fig. 9에서 보인 yaw 축의 각도 추정 결과도 마찬가지로의 경향을 보였으며, 제안한 방법이 가장 우월하게 참값을 추종하고 있다.

종합적으로 RMS error를 분석한 도표를 Table 2와 Fig. 10으로 제시한다. Fig. 10에서 각 행은 비교 대상 4가지 방법이고, 각 열은 각 축(x, y, z)과 그 합을 의미한다. 붉은색으로 표현될수록 큰 오차를 가지고 푸른색으로 표현될수록 작은 오차를 가지게 되는 컬러 맵으로 표현되었으며, 4행의 오차(제안한 방법)가 가장 작은 것으로 분석되었다.

Table 1. Simulation setup

axis	func.	amp.(mm)	freq.(rad/s)
x	sin	100	1
y	cos	100	1
z	ramp	-10(slope)	
yaw	sin	$\pi/8$	1
roll	sin	$\pi/8$	1
pitch	sin	$\pi/8$	1

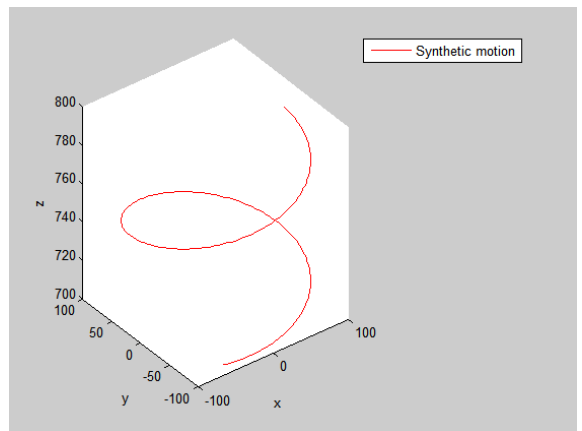


Fig. 2. Pseudo camera trajectory(ground truth)

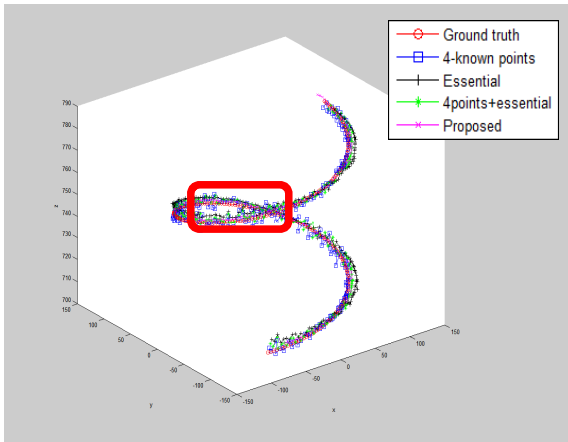


Fig. 3. Pseudo camera trajectory estimation

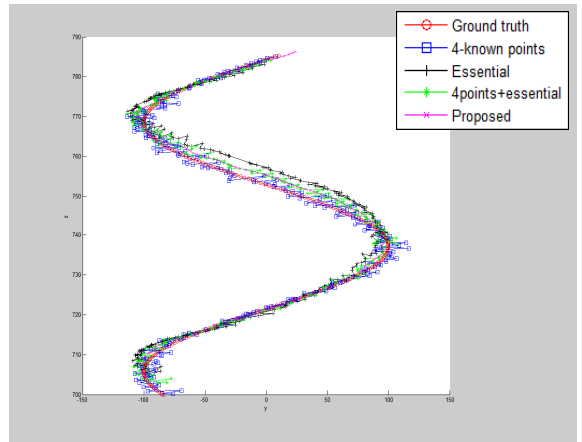


Fig. 6. Pseudo camera trajectory estimation(y-z)

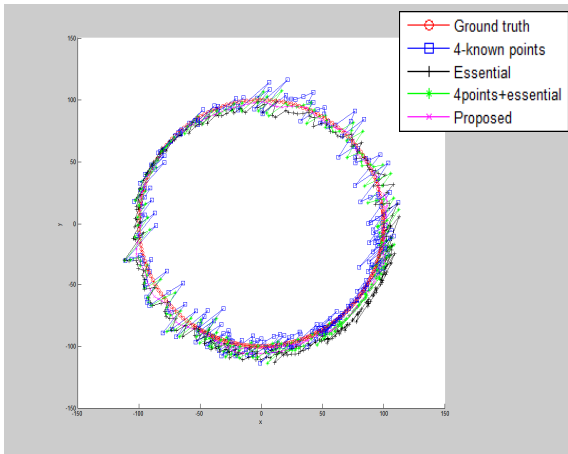


Fig. 4. Pseudo camera trajectory estimation(x-y)

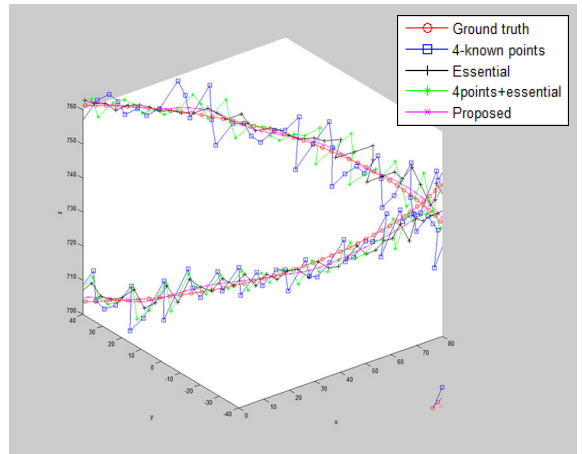


Fig. 7. Pseudo camera trajectory estimation(enlarged)

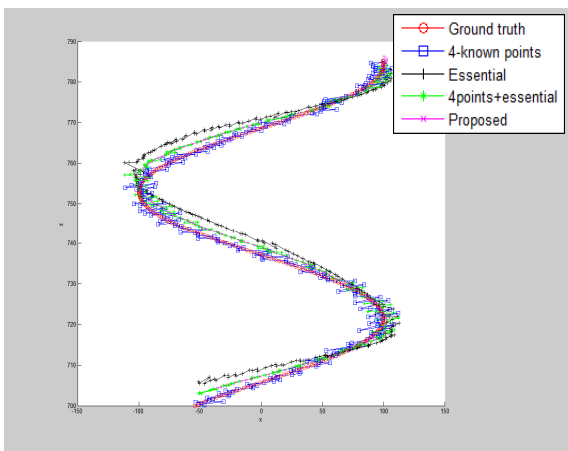


Fig. 5. Pseudo camera trajectory estimation(x-z)

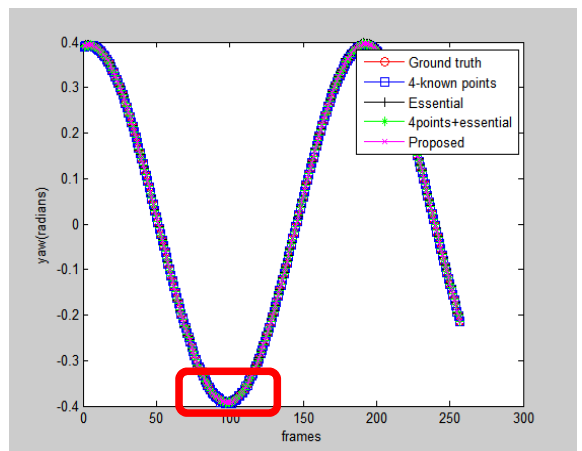


Fig. 8. Pseudo camera rotation estimation(yaw angle)

Table 2. RMS error

	x	y	z	sum
4points	7.35	7.35	0.29	15.00
essential	5.01	8.30	3.08	16.40
4points + essential	5.19	6.66	1.56	13.41
proposed	2.56	4.40	1.50	8.46

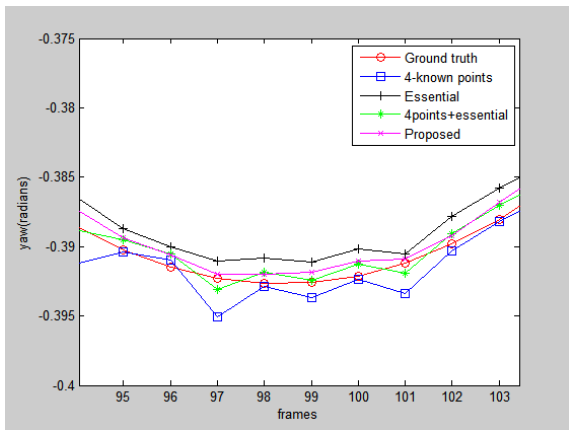


Fig. 9. Pseudo camera rotation estimation(yaw angle, enlarged)

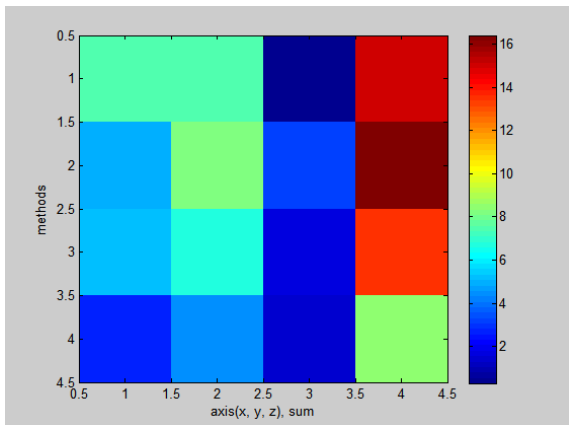


Fig. 10. Analysis of RMS error

본 논문에서 제시된 결과의 시뮬레이션은 모두 모델 기반 설계 기법을 적용하였으며, Fig. 11은 개발된 환경에서 실시간으로 추정된 카메라 위치, 자세를 관찰할 수 있도록 구성된 전시 화면이다. 중앙에 위치한

좌표축으로부터 사각뿔 형태의 회색 도형(카메라)이 움직이는 것을 통해 실제 움직임을 추종하는지 확인할 수 있다.

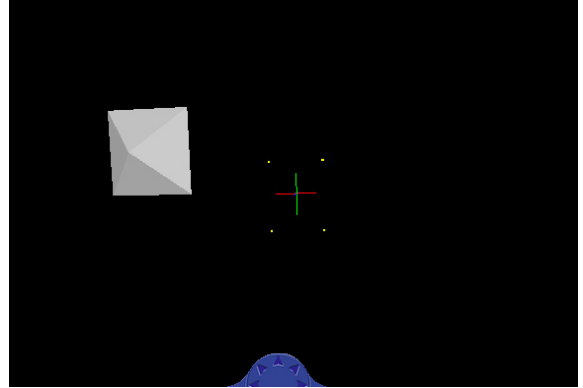


Fig. 11. Real-time 3D simulation environment

#### 4.2 실제 영상 적용 결과

본 절에서는 제안한 기법을 실제 영상에 적용한 결과를 보인다. Fig. 12에서 제시된 결과는 입력된 영상으로부터 특징점을 추출하고 알고 있는 점(H 마커의 최외곽 네 점)으로부터 추정된 것과 인접 영상과의 대응점으로부터 추정된 카메라 위치, 자세 파라미터를 이용하여 화면에 재 투영한 결과이다. 2장 이론 해석에서 이미 제시한 바와 같이 카메라 내, 외부변수를 알고 3차원 특징점의 위치를 아는 경우 식 (1)을 이용하여 영상평면에 재 투영하는 것이 가능하다. 붉은색 점은 추출된 특징점, 푸른색 점은 알고 있는 점만을 이용하여 추정된 카메라 위치, 자세로부터 재 투영된 점, 초록색 점은 제안한 방법으로 재 투영된 점이다. 제안한 방법이 최초로 추출된 특징점의 위치를 보다 더욱 정확하게 복원하는 것을 확인할 수 있다.

단안 카메라에서 3차원 공간과 영상 평면은 비례적인 관계(up to scale)에 있으므로 재투영 오차가 가지는 실제 카메라 위치, 자세 오차는 얼마나 멀리서 관측하는가 하는 것과 비례적이다. 이론적으로는 영상 평면을 무한한 화소의 관측량으로 해석하는 것이 가능하지만, 영상을 얻기 위해 일반적으로 사용되는 영상 센서는 화소 개수가 유한하기 때문에 특징점 관측 시 양자화 오차를 수반한다.

Fig. 13은 추정된 카메라 위치 궤적을 x-z평면에 대해 보인 것이다. 실제 영상에 대한 실험에서는 참값을 알기가 힘들지만 개략적인 경향 분석은 가능하다. 실

제 촬영된 궤적과 유사하게 카메라의 위치를 추정할 것을 확인할 수 있다. 알고 있는 4점을 통해서 추정된 위치는 대응점 탐색이 잘못될 경우 전혀 다른 위치를 추정하는 것에 비해 제안된 기법은 보다 안정적으로 카메라 위치 추정을 수행하고 있는 것을 확인할 수 있다.



Fig. 12. Real scene and re-projection of features

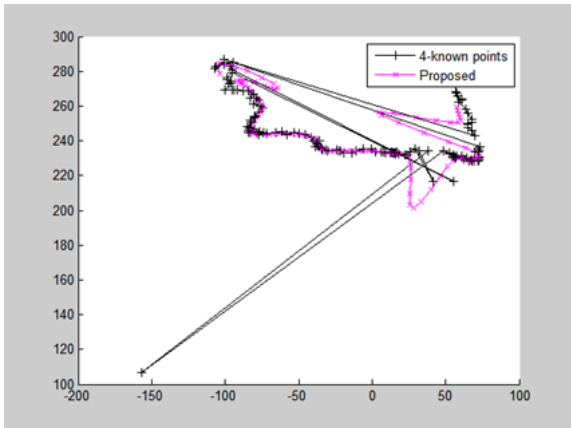


Fig. 13. Camera trajectory(x-z)

## 5. 결론 및 추후 연구 과제

본 논문에서는 영상 응용분야의 핵심인 카메라의 위치, 자세 추정에 관한 연구 결과로 알고 있는 점과 인접 영상의 대응점, 관성력에 의해 예측된 값을 모두 만족하는 최적의 카메라 위치 추정에 관한 기법을 제안하였다. 생성된 모의데이터와 실제 영상을 통해 검증한 결과 제안된 방법은 각각의 단일 입력을 통해

추정한 것보다 우수한 결과를 보여줄을 검증하였다.

실제 영상의 특징점을 추출하는 것(corner extraction)과 대응점을 탐색하는 과정(correspondence search), 노이즈가 수반된 자료들에서 가장 합리적인 자료를 추출하는 방법(RANdom SAmple Consensus: RANSAC)은 실제 영상에 적용하기 위해서 구현되었지만 본 논문에서 제시하는 기법의 성능 평가와는 직접적인 연관성을 갖지 않기 때문에 언급을 생략하였다.

추후 연구과제는 다음과 같다. 제안된 방법은 모사 생성 데이터와 실험실 내의 환경에서 알고 있는 특징점이 추출되었을 경우에만 동작한다. 불규칙한 조명조건과 대기조건이 주어지는 야외 환경에서 동작하기 위해서는 알고 있는 지도로부터 영상에 투영된 특징점을 정확하게 찾아낼 수 있어야 한다. 데이터베이스로 가지고 있는 영상과 현재 위치, 자세를 획득하고자 하는 카메라의 관측 자세와 시간은 상이할 수 있으므로 둘 사이에 대응관계 해석을 하는 것은 매우 도전적인 문제이다. 특징점의 추출과 대응점으로부터 얻을 수 있는 카메라의 위치, 자세는 정확하게 탐색된 대응점의 수와 비례하고 SVD의 연산 복잡도는  $m \times n$  행렬 분해에 대하여  $4m^2n + 8mn^2 + 9n^3$  이므로 많은 양의 데이터를 처리하기 위한 실시간 처리 방안 또한 연구되어야 할 것이다.

## References

- [1] Davison, Andrew J., and David W. Murray. "Simultaneous Localization and Map-building using Active Vision," Pattern Analysis and Machine Intelligence, IEEE Transactions on 24.7, 865-880, 2002.
- [2] Hartley, Richard, and Andrew Zisserman, "Multiple View Geometry in Computer Vision," Cambridge University Press, UK. 2003.
- [3] Nistér, David, and Henrik Stewénus, "A Minimal Solution to the Generalised 3-point Pose Problem," Journal of Mathematical Imaging and Vision 27.1, 67-79, 2007.
- [4] Snavely, Noah, Steven M. Seitz, and Richard Szeliski, "Photo Tourism : Exploring Photo Collections in 3D," ACM Transactions on Graphics(TOG) 25.3, 835-846, 2006.

- [5] Klein, Georg, and David Murray, "Parallel Tracking and Mapping for Small AR Workspaces," Mixed and Augmented Reality, ISMAR, 6th IEEE and ACM International Symposium on. IEEE, 2007.
- [6] Newcombe, Richard A., Steven J. Lovegrove, and Andrew J. Davison, "DTAM : Dense Tracking and Mapping in Real-time," Computer Vision(ICCV), 2011 IEEE International Conference on. IEEE, 2011.