



J. Korean Soc. Aeronaut. Space Sci. 48(10), 783-789(2020)

DOI:https://doi.org/10.5139/JKSAS.2020.48.10.783

ISSN 1225-1348(print), 2287-6871(online)

## 의미론적 분할된 항공 사진을 활용한 영상 기반 항법

홍경우<sup>1</sup>, 김성중<sup>2</sup>, 박준우<sup>3</sup>, 방효충<sup>4</sup>, 허준희<sup>5</sup>, 김진원<sup>6</sup>, 박장호<sup>7</sup>, 서송원<sup>8</sup>

## Vision-based Navigation using Semantically Segmented Aerial Images

Kyungwoo Hong<sup>1</sup>, Sungjoong Kim<sup>2</sup>, Junwoo Park<sup>3</sup>, Hyochoong Bang<sup>4</sup>, Junhoe Heo<sup>5</sup>,

Jin-Won Kim<sup>6</sup>, Chang-Ho Pak<sup>7</sup> and Songwon Seo<sup>8</sup>

Korea Advanced Institute of Science and Technology<sup>1,2,3,4</sup>, Poongsan R&D Institute<sup>5,6,7,8</sup>

### ABSTRACT

This paper proposes a new method for vision-based navigation using semantically segmented aerial images. Vision-based navigation can reinforce the vulnerability of the GPS/INS integrated navigation system. However, due to the visual and temporal difference between the aerial image and the database image, the existing image matching algorithms have difficulties being applied to aerial navigation problems. For this reason, this paper proposes a suitable matching method for the flight composed of navigational feature extraction through semantic segmentation followed by template matching. The proposed method shows excellent performance in simulation and even flight situations.

### 초 록

영상 기반 항법은 GPS/INS 통합 항법 시스템의 취약점을 보강할 수 있는 보조 항법 기술로 비행체에서 촬영한 항공 영상과 기존의 데이터베이스를 비교하여 비행체의 위치를 구한다. 하지만 데이터베이스가 생성된 시점은 항공 영상 촬영 시점과 다를 수밖에 없으며, 이러한 시점 차이로 인해 두 영상 간의 다른 특징점들이 생성된다. 즉, 유사하지만 다른 두 영상이므로 일반적인 영상 대조 알고리즘을 항법 문제에 적용하기 힘들다. 따라서 본 논문에서는 인공지능 기법인 의미론적 분할을 활용하여 항공 영상에서 항법에 필요한 정보를 분류한 후 영상 대조를 수행하는 방법을 제안한다. 의미론적 분할로 시점 변화, 촬영 조건 변화가 있더라도 강건하게 두 영상이 정합 되도록 한다. 제안한 방법은 시뮬레이션과 비행 실험을 통해 성능을 확인하며, 일반적인 영상 대조 알고리즘을 이용하여 항법을 수행한 결과와 비교한다.

**Key Words** : Vision-Based Navigation(영상 기반 항법), Semantic Segmentation(의미론적 분할), Template Matching(템플릿 정합)

### 1. 서 론

매우 강력한 성능을 보여주는 GPS / INS (Global Positioning System/Inertial Navigation System) 통합

항법 시스템은 GPS 재밍(Jamming) 등의 문제점을 가지고 있다. 이에, 이를 보완하기 위한 보조 항법 시스템에 관한 연구가 많은 관심을 받고 있다. 영상 기반 항법 또한 보조 항법 중 하나로 랜드마크

† Received : July 20, 2020 Revised : September 3, 2020 Accepted : September 14, 2020

<sup>1,2,3</sup> Graduate Student, <sup>4</sup> Professor, <sup>5,6,7,8</sup> Researcher

<sup>4</sup> Corresponding author, E-mail : hcbang@ascl.kaist.ac.kr, ORCID 0000-0001-6016-8102

© 2020 The Korean Society for Aeronautical and Space Sciences

(Landmark) 기반 방식[1,2]과 그리고 비행체에서 촬영한 영상과 데이터베이스를 비교하는 방식이 있다. 데이터베이스는 항공 영상의 형태로 또는 항공 영상에서 얻을 수 있는 특징들 형태로 위치 정보와 함께 미리 제작되며, 비행체에서 촬영한 항공 영상 또는 그로부터 얻은 특징들과 비교하여 비행체의 현재 위치를 추정한다[3-6].

영상 기반 방법은 두 사진을 비교하는 측면에서, 기존의 영상 대조(Image Matching) 문제와 유사하다. 영상 대조 문제는 일반적으로 특징점들을 추출하여 정합한다. 대표적인 특징점 기술자(descriptor)로는 DoG(Difference of Gaussian) 기반으로 크기 변화에 강건한 SIFT(Scale Invariant Feature Transform)[7], SIFT의 수행 속도 문제를 보완한 SURF(Speed Up Robust Feature)[8] 등이 있으며 해당 기법들의 대안으로 FAST 특징점 추출 및 BRIEF(Binary Robust Independent Elementary Features) 기술자를 이용한 ORB(Oriented FAST and Rotated BRIEF)[9,10] 등이 있다.

그러나 영상 기반 방법을 위해 기존의 영상 대조 알고리즘을 그대로 적용하기에 몇 가지 문제점을 가진다. 무엇보다도, 방법에서는 데이터베이스를 생성하는 시점과 비행 시점이 다를 수밖에 없다. 이로 인해 촬영 조건도 달라지며 항공 영상 속 물체들 또한 변하게 된다. 즉, 같은 지역을 촬영한 사진이므로 유사한 사진이지만 동일하지 않은 두 사진이다. 이러한 두 영상간의 차이는 기존의 영상 대조의 정합 실패로 이어진다. Fig. 1은 널리 알려진 영상 대조 알고리즘 SURF를 이용하여 항공 영상에 적용한 결과를 보여준다. Fig. 1의 (a)는 무인기에서 촬영한 항공사진과 웹 지도에서 제공하는 항공사진을 정합한 결과이다. 해당 그림에서, 같은 지역의 사진이지만 다른 사진이므로 같은 특징을 추출하지 못하는 것을 확인할 수 있다. Fig. 1의 (b)는 10초 뒤 촬영한 사진과 정합한 결과이다. 이 경우 두 영상 촬영 시점이 거의 같아 우수한 정합 결과를 얻는 것을 알 수 있다. 하지만 영상 기반 방법을 수행하기 위해서는 미리 데이터베이스를 생성해야 하고, 따라서, 기존의 영상 대조 알고리즘을 방법을 위해 그대로 적용하기에는 문제가 있다.

본 논문에서는 기존의 영상 대조 알고리즘인 템플릿 정합(Template Matching)을 방법에 적용하기 위해 인공지능 기법을 적용한다. 최근, 인공지능의 발달로 인해 사진 속 의미론적 분류의 성능이 대폭 향상되었다[11-13]. 또한, 이러한 기법들을 항공 영상에 적용하거나[14,15], 방법 문제에 적용하려는 시도가 있다[16]. 이를 활용하여 시점, 조명 등 조건들이 바뀌더라도 강건하게 특징들을 추출할 수 있다. 또한, 시 불변의 특징을 추출하는 것이 가능하며, 해당 특징들만을 방법에 이용하는 것이 가능하다. 따라서 의

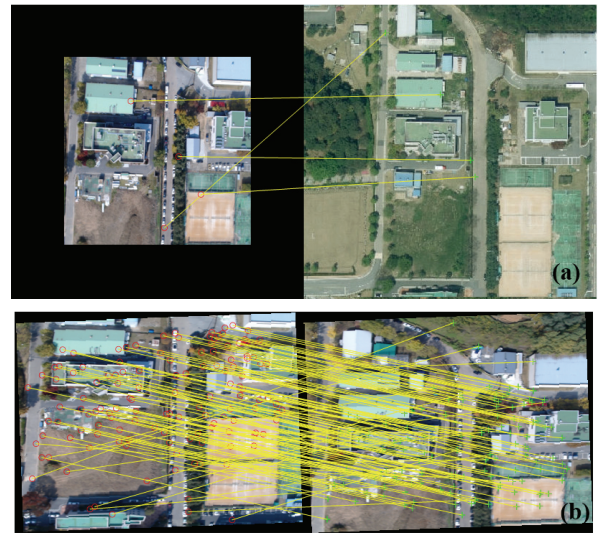


Fig. 1. The matching result using SURF between the aerial image and the database image(a), and the aerial image after 10sec(b)

미론적 분류를 이용하여 방법에 이용할 수 있는 정보를 추출하고, 해당 특징들을 데이터베이스와 정합한다.

본 논문의 구성은 다음과 같다. 2.1장에서는 항공 영상의 특징 추출에 사용한 의미론적 분할(Semantic Segmentation)에 대해 설명한다. 2.2장에서는 템플릿 정합 방법을 간략히 요약하고, 의미론적 분할된 항공 영상 기반의 템플릿 정합 방법에 대해 설명한다. 2.3장에서는 시뮬레이션을 통해 제안한 방법 성능을 확인한다. 2.4장에서는 실제 비행실험을 통하여 촬영한 영상을 기반으로 방법 수행 결과를 거리 오차 및 정합 성공률 측면에서 확인한다. 또한, 일반적인 회색조 영상의 템플릿 정합 결과와 비교한다. 마지막으로 3장에 결론을 서술한다.

## II. 본 론

### 2.1 의미론적 분할(Semantic Segmentation)

항공 영상은 풍부한 정보를 담고 있으며, 이러한 정보들은 방법 문제에 많은 도움을 줄 수 있다. 그러나 수많은 정보 중에서 방법 문제에 방해되는 정보들도 함께 포함되어있다. 예를 들어, 항공 영상 속 사람 또는 차와 같은 물체들은 계속해서 움직이며 데이터베이스 영상과 비행체에서 촬영한 항공 영상의 차이를 만들어낸다. 따라서, 방법 문제에 활용하기 위해서는 항공 영상에서 필요한 특징들만 추출하는 것이 중요하다. 게다가, 서론에서 언급했듯이, 방법 문제 특성상 항공 영상과 데이터베이스 영상의 시점 차이가 불가피하게 발생한다. SURF 등 일반적인 영상처리 방법은 방법 문제에 적용 시, 이러한 시

점 변화, 조명 변화, 크기 변화 등에 취약하다는 문제점을 가진다(Fig. 1). 따라서, 이러한 시점 변화 상황에서 항법에 필요한 특징들을 추출하기 위해 의미론적 분할(Semantic Segmentation)을 수행한다.

의미론적 분할은 객체 검출 기법의 하나이다. 특히 의미가 있는 물체들을 같은 영역으로 분할하여 표시하며, 한 영상 내에 다수의 특징을 구분하는 것이 가능하다. 특징 분류는 픽셀 단위로 이루어지며, 즉, 영상 내 각 픽셀이 분류된 특징 정보를 가진다. 따라서, 이를 이용하여 항공 영상 중 필요한 정보들과 방해가 되는 정보들과 도움이 되는 정보들을 분리할 수 있다. 여러 의미론적 분할 방법이 있지만, 본 논문에서는 그 중에 널리 사용되고 있는 FCN(Fully Convolutional Network)를 사용한다[13].

항법에 사용하기 위한 특징은 움직이지 않고 고정되어 있으며 시간이 지나도 변하지 않아야 한다. 예를 들어, 사람과 차는 계속해서 움직이며, 나무의 경우 계절에 따라 그 모습이 변한다. 따라서, 이러한 특징들은 정합에 고려하지 말아야 하며 시간이 지나도 추출이 가능한 물체 정보를 이용하여 정합해야 한다. 본 논문에서는 도심지 비행을 가정하며 시물변의 특징으로 건물과 도로 2가지를 사용한다. 즉, FCN을 이용하여 사진에서 건물과 도로인 픽셀 정보들을 얻어낸다.

FCN을 학습하기 위해서는 항공 영상과 함께 해당이 표시된 영상이 필요하다. 즉, 항공 영상과 함께 해당 항공 영상에 건물과 도로, 특징들이 표시된 해당 영상도 함께 필요하다. 이는 Vworld 공간 정보 오픈 플랫폼에서 제공하는 항공 영상과 GIS(Geographic Information System) 데이터를 이용하여 해당 영상을 만들었다. 시물레이션에서도 Vworld에서 제공하는 항공 영상을 측정치로 활용한다. 그러나 시물레이션에서 측정치로 사용하는 항공 영상은 FCN의 학습 데이터로 사용하지 않는다. 학습은 NVIDIA Digits를 활용하였으며 Caffe frameworks를 이용하였다.

## 2.2 템플릿 정합(Template Matching)

본 장에서는 템플릿 정합을 소개하며 다음으로 의미론적 분할인 FCN 결과 영상을 이용한 템플릿 정합을 소개한다. 템플릿 정합은 작은 영상(template)을 참조(reference) 영상과 비교하여 참조 영상 내에서 가장 유사한 위치를 찾는 과정이다. 항법 문제에서는 참조 영상은 미리 제작된 항공 영상 또는 항공 영상으로부터 얻을 수 있는 특징과 위치 정보를 포함하고 있는 데이터베이스가 된다. 또한, 템플릿 영상은 현재 항공기에서 촬영한 영상이 된다. 즉, 항공기에서 촬영한 항공 영상을 데이터베이스와 정합함으로써 현재 항공기의 위치를 얻는다.

템플릿과 참조 영상, 두 사진을 비교하기 위해서는 비교 기준이 되는 값이 필요하다. 템플릿 정합에서는

대표적으로 영상 픽셀들의 SSD(Sum of Squared Difference), SAD(Sum of Absolute Difference), NCC(Normalized Cross Correlation) 그리고 정규화된 상관 계수(Normalized Correlation Coefficient) 값을 이용한다. 작은 템플릿 영상을 큰 참조 영상 상에서 옮겨가며 영상 간의 차이 또는 상관 계수를 계산한다. 그리고 차이가 가장 적은 위치 또는 상관 계수가 가장 높은 위치를 정합 위치, 즉 항공기의 위치로 본다. 템플릿 영상을 T, 참조 영상을 I라고 할 때 각 방법은 아래와 같이 계산한다[17].

$$R_{SSD}(x,y) = \sum_{i,j} \{T(i,j) - I(x+i,y+j)\}^2 \quad (1)$$

$$R_{SAD}(x,y) = \sum_{i,j} |T(i,j) - I(x+i,y+j)| \quad (2)$$

$$R_{NCC}(x,y) = \frac{\sum_{i,j} \{T(i,j) \times I(x+i,y+j)\}}{\sqrt{\sum_{i,j} \{T(i,j)\}^2 \times \sum_{i,j} \{I(x+i,y+j)\}^2}} \quad (3)$$

$$R_{NCC_{eff}}(x,y) = \frac{\sum_{i,j} \{T'(i,j) \times I'(x+i,y+j)\}}{\sqrt{\sum_{i,j} \{T'(i,j)\}^2 \times \sum_{i,j} \{I'(x+i,y+j)\}^2}} \quad (4)$$

$$T'(i,j) = T(i,j) - \bar{T}$$

$$I'(x+i,y+j) = I(x+i,y+j) - \bar{I}_{x,y}$$

여기서  $\bar{T}, \bar{I}_{x,y}$ 는 각각 템플릿의 평균 및 x, y 위치에서 템플릿 크기의 참조 영상 평균을 의미한다. 본 논문에서는 밝기 및 대조 변화에 강건하다고 알려진 정규화된 상관 계수를 이용한다[18].

의미론적 분할 방법인 FCN의 결과는 각 픽셀의 분류 결과이다. 따라서 두 영상 간의 비교는 차이 및 상관관계 계산 대신 간단히 같은 특징으로 분류되었는지를 비교함으로써 확인할 수 있다. 마찬가지로 템플릿 영상을 T, 참조 영상을 I로 할 때, 두 영상 비교는 아래와 같이 수행한다.

$$R_{FCN}(x,y) = \sum_{i,j} \{T(i,j) \div I(x+i,y+j)\} \quad (5)$$

템플릿 영상 및 참조 영상은 모두 특징들의 라벨로 이루어져 있다. 또한, 위  $\div$  연산자는 특징이 같으면 1, 다르면 0을 계산한다. 두 영상이 비슷할수록 큰 값을 가지며 가장 높은 곳을 비행체의 위치로 본다. 참고로 FCN 수행 시 특징들의 모서리 및 변 부분이 보존이 안 되기 때문에 SIFT, ORB 등의 특징점 기반 대조 알고리즘 대신 템플릿 기반의 영상 대조 알고리즘을 사용하였다.



## 2.3 시뮬레이션

본 장에서는 FCN을 이용한 템플릿 정합의 시뮬레이션 결과를 보여준다. 시뮬레이션에서는 실 비행 없이 Vworld에서 제공하는 항공 영상을 측정치로 이용한다. 먼저 시뮬레이션 환경에 대해 설명하고 시뮬레이션에서의 항법 성능을 확인한다.

### 2.3.1 시뮬레이션 환경

비행체는 장착된 카메라를 통해 직하방의 영상을 얻는다고 가정한다. 본 시뮬레이션 상에서의 직하방 영상은 Vworld에서 제공하는 항공 영상을 사용한다. 해당 직하방의 영상을 의미론적 분할되어 특징들로 분류된 영상이 된다. 또한, 데이터베이스도 분류할 특징, 건물과 도로가 표시된 영상으로 구성한다. 여기서 데이터베이스는 FCN 학습에 사용되었던 Vworld의 GIS 데이터를 이용한다. 마지막으로, 특징들이 분류된 항공 영상은 데이터베이스와 대조하여 비행체의 위치를 추출한다. 앞서 언급했듯이, 본 시뮬레이션에서 비행할 지역은 FCN 학습 데이터에서 제외하였다.

비행체는 일정한 고도, 자세, 그리고 속도를 유지한다고 가정한다. 고도는 200m이며, 동쪽으로 10m/s 속도로 비행한다. 두 영상을 템플릿 정합하기 위해서는 크기와 방향이 같도록 정렬해야 한다. 항공 영상의 크기는 비행 고도와 연관되며, 카메라가 gimbal에 장착되었다고 가정하면 방향은 비행체의 yaw 자세 각과 연관되어 있다. 여기서 고도는 기압 고도계 등의 센서를 통해 얻어진다고 가정하며, 자세 센서를 통해 비행체의 yaw 값이 얻어진다고 가정한다. 이산 시간  $k$  때의 고도는 실제 값에  $\sigma_h$ 만큼의 표준편차를 가지는 가우시안 오차를 추가하여 사용한다.

$$h_k = h_{k,true} + w_k \quad (6)$$

$$w_k \sim N(0, \sigma_h^2) \quad (7)$$

여기서  $\sigma_h$ 은 4m를 사용하였다. 마찬가지로, 이산 시간  $k$  때의 yaw 방향 자세 각은 실제 값에 가우시안 오차를 추가하여 사용한다.

$$\psi_k = \psi_{k,true} + \zeta_k \quad (8)$$

$$\zeta_k \sim N(0, \sigma_\psi^2) \quad (9)$$

$\sigma_\psi$ 는 yaw 자세 각 오차의 표준편차를 의미하며 0.5°를 사용하였다. 시작 위치는 위도 36.3379°, 경도 127.3735°이며 총 200초간 비행한다.

영상 갱신 주기는 1Hz이며, 마찬가지로 템플릿 정합을 1초마다 수행한다. 카메라의 AFOV(Angular Field Of View)는 가로 53.26°로 가정하였으며 영상 해상도는 640, 480으로 가정하였다.

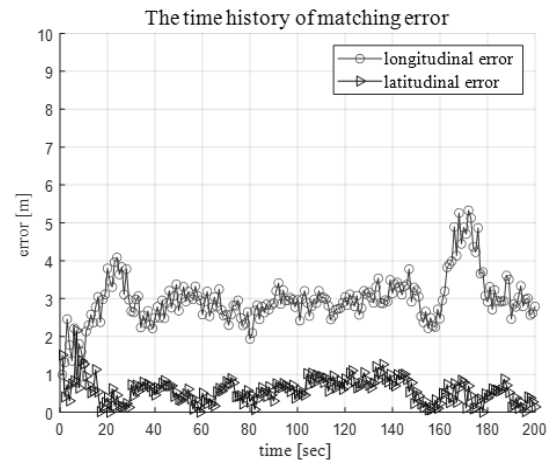


Fig. 2. Time history of the template matching error using the FCN result as the measurement

### 2.3.2 시뮬레이션 결과

FCN 결과를 측정치로 사용한 템플릿 정합의 성능을 확인하기 위해 Vworld에서 제공하는 항공 영상을 사용하여 정합을 수행한다. Fig. 2는 시간에 따른 템플릿 정합의 결과를 보여준다. 위도 방향의 오차는 삼각형으로 표시하였으며, 경도 방향의 오차는 원으로 표시하였다. 160초에서 180초 부분을 제외하고 모든 시간에서 5m 이하의 오차로 우수한 성능을 보여준다. 160초에서 180초 부분은 건물들이 밀집되어 있어 FCN 성능이 다른 시간대와 비교하여 좋지 않은 것을 확인하였다. 그러나 이를 고려해도 위, 경도 방향의 평균 오차가 각각 약 0.566m, 2.958m로 높은 정확도를 보여준다.

그러나, 그림에서 알 수 있듯이, 경도 방향의 오차가 위도 방향의 오차보다 일관성 있게 크게 나타난다. 본 원인은 Vworld 데이터와 FCN 차이로 보인다. Fig. 3은 시뮬레이션 상의 측정치 중 하나이다. 왼쪽 사진은 항공 영상과 Vworld의 GIS 데이터를 결합한 것이며, 오른쪽 사진은 항공 영상과 FCN 결과를 결합한 것이다. 해당 사진에서 붉은색 영역은 건물, 파란색 영역은 도로를 의미한다. FCN의 경우

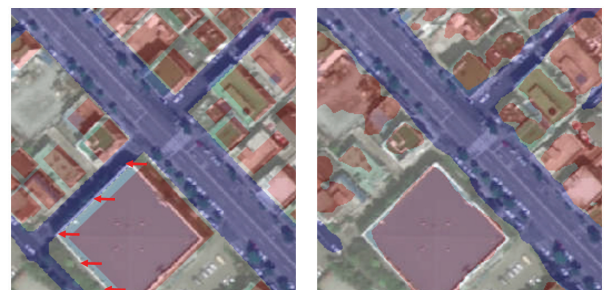


Fig. 3. The feature image from Vworld and the result of the FCN

항공 영상에 건물의 윗부분을 인식하지만, Vworld의 경우 해당 건물의 바닥을 표시하고 있다. 이는 정확히 직하방의 항공 영상을 촬영한다면 같은 결과가 된다. 그러나, 그렇지 않으면 편향되게 오차가 발생하게 된다. 해당 시뮬레이션 지역은 이러한 편향된 사례가 많았으며, 이러한 오차는 사전 교정으로 줄일 수 있다.

## 2.4 비행실험

앞선 시뮬레이션 결과로 FCN을 이용한 템플릿 정합의 가능성을 확인하였다. 그러나, 시뮬레이션의 경우 FCN 학습에 사용되었던 Vworld의 항공 영상을 측정치로 사용하였다. 물론 시뮬레이션 지역을 학습 데이터로 사용하지 않았지만 같은 고도 및 조건으로 촬영된 영상이므로 실제 비행에서 얻어진 항공 영상을 통해 확인할 필요가 있다. 따라서, 실제 무인기에서 촬영한 항공 영상을 측정치로 사용하여 템플릿 정합을 수행한다. 또한, 일반 회색조 영상의 템플릿 정합 결과와 제안하는 FCN을 이용한 항법 결과를 비교한다.

### 2.4.1 비행실험 환경

무인기는 김발 및 카메라, 항법 센서를 탑재하고 비행한다. 비행 중 김발에 장착된 카메라에서 직하방의 항공 영상을 수집한다. 또한, 김발이 정확히 직하방을 바라보지 못할 경우를 대비하여 카메라의 지향각을 측정할 수 있는 별도의 자세 센서를 부착한다. 해당 센서값을 이용하여 얻어진 영상을 직하방 영상으로 교정한다. 항법 센서로는 RTK GNSS 수신기를 가진 GPS/INS를 탑재하여 템플릿 정합 결과와 비교할 비행경로를 저장한다. 탑재한 GPS/INS는 Spatial Dual로 RTK 사용 시 수평 위치 정확도가 0.008m로 알려져 있다.

데이터베이스는 FCN 결과를 위한 것 과 회색조 영상을 위한 것 2가지를 준비한다. FCN 결과 정합용은 시뮬레이션과 마찬가지로 Vworld에서 제공하는 GIS 데이터를 활용하여 만들었으며, 회색조 영상정합용은 Vworld에서 제공하는 항공 영상을 회색조로 변환하여 만들었다(Fig. 4). 데이터베이스의 중심은 위도 36.3916°, 경도 127.3984°이며 가로, 세로 약 630m로 만들었다.

템플릿 정합에서는 만들어진 데이터베이스 전체와 비교하는 것이 아닌 실제 위치 기준으로 일부분을 추출하여 비교한다. 해당 비교 영상의 크기는 사용자가 설정할 수 있으며 좁은 위치를 비교할 경우 수행 시간이 빠르다는 장점이 있지만, 해당 위치에 참값이 없을 확률이 높다는 단점이 있다. 마찬가지로 넓은 위치를 비교할 경우 수행 시간이 오래 걸린다는 단점이 있다. 본 실험에서는 비교 영상의 크기를 가로, 세로 300, 400, 500픽셀(pixel)로 설정하여 크기에 따른 결과를 보여준다.



Fig. 4. The databases for the feature image and the grayscale image

템플릿 정합은 고도가 150m 이상인 구간에서만 수행하였다. 약 10분간 비행하였으며 고도가 150m 이상인 구간은 약 350초로 350초간의 결과를 보여준다. 항공 영상과 데이터베이스의 비율 및 방향을 맞추기 위해서 고도 및 yaw 자세 값이 필요하며 이는 GPS/INS의 고도와 yaw 자세 값을 이용하였다.

실험기체는 대각선 길이 약 1.4m의 X8 기체를 사용하였으며, 영상처리 컴퓨터는 Jetson AGX Xavier를 사용하였다. 카메라는 logitech c920을 사용하였으며 김발에 장착되어 계속해서 직하방을 주시하도록 만들었다. 김발은 Infinity MR-S2를 사용하였다.

### 2.4.2 비행실험 결과

Figure 5는 시간에 따른 정합 성능을 보여준다. 오차는 각 방법의 정합 위치와 GPS/INS의 위치의 차이의 절대값으로 계산하였다. 위치 오차가 50m 이상인 부분은 발산했다고 판단하여 제외하고 표시하였다. 발산한 위치는 회색조 영상의 경우 원으로 FCN 영상의 경우 'X' 표시로 나타내었다. 20m 이상의 오차를 보인 결과는 정합에 실패했다고 판단하여 오차 20m 기준으로 수렴률을 계산하였다. Table 1은 수렴률 및 평균 오차를 보여준다. 여기서 평균 오차는 발산한 경우는 제외하고 수렴한 결과에 대해서만 계산하였다. 즉, 오차가 20m 미만인 결과들의 평균을 의미한다.

회색조 영상의 성공률은 크기가 큰 순으로 98.29%, 82.62%, 79.77%를 보였다. 크기가 300 x 300 픽셀인 경우, 수렴률이 98%로 매우 높은 것을 확인할 수 있다. 그러나 비교 데이터베이스 크기가 커질수록 수렴

Table 1. The converge rate and mean error using the FCN result and the grayscale image as the measurement

Database size(pixel)		300 x 300	400 x 400	500 x 500
Converge rate	Grayscale	98.29%	82.62%	79.77%
	FCN	100%	99.15%	95.16%
Mean error	Grayscale	6.48m	6.41m	6.39m
	FCN	9.41m	9.44m	9.31m

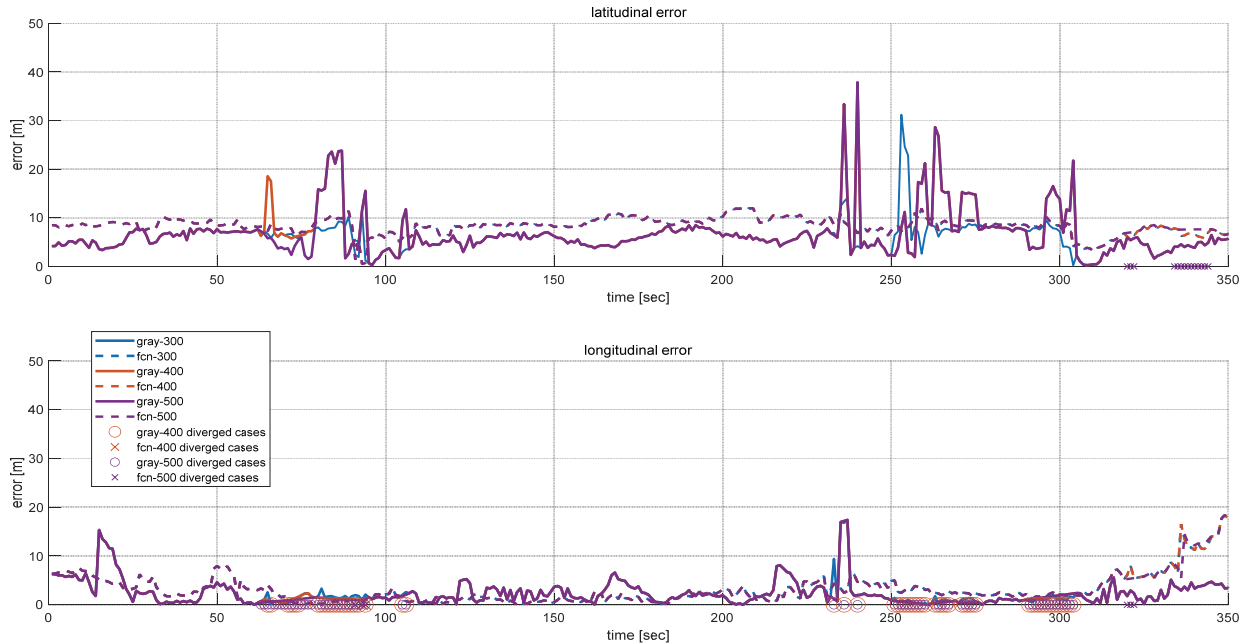


Fig. 5. Time history of the template matching error using the FCN result and the grayscale image as the measurement

률이 현저하게 낮아지는 것을 확인할 수 있다. 반면, FCN 영상의 경우 수렴률이 100%, 99.15%, 95.16%로 회색조 영상보다 높은 수렴률을 보였으며, 특히 비교 데이터베이스 크기가 클 때도 강건하게 작용하는 것을 알 수 있다.

회색조 영상의 낮은 수렴률은 데이터베이스와 항공 영상의 차이로 여겨진다. 데이터베이스와 항공 영상은 같은 지역을 촬영하여 유사한 영상이지만 시점, 조명 등 촬영 조건이 다르므로 같은 영상으로 보기 힘들다. 또한, 템플릿 정합의 경우 크기 비율과 회전에 취약하다는 단점이 있다. 따라서, 항공 영상과 데이터베이스 간의 정렬에 필요한 yaw 자세 각과 고도 측정치가 정확하지 않으면 정합이 실패할 수 있다. 하지만 FCN을 이용한 정합의 경우 항법에 필요한 정보들만 추출하므로 자세와 고도 측정치의 오차에 더욱 강건한 모습을 보인다.

회색조 영상의 평균 오차는 각각 6.48m, 6.41m, 6.39m이며, FCN 영상의 평균 오차는 각각 9.41m, 9.44m, 9.31m이다. 평균 오차 측면에서는 회색조의 결과가 더 좋은 것을 확인할 수 있다. FCN 영상의 오차 원인을 분석하자면, 해당 오차가 무작위적이지 않으며 일정하게 편향된 모습을 보여주는 것을 알 수 있다. 또한, 위도 방향의 오차가 일정하고 경도 방향의 오차가 매우 낮은 점을 고려하였을 때(Fig. 5), 비교 영상 크기가 달라져도 비슷한 평균 오차를 보이는 점을 고려하였을 때, 이는 정합의 오차보다 카메라 촬영 각도의 문제 또는 데이터베이스 영상과 실제 위치 간의 오프셋(offset) 문제로 여겨진다. 즉, 올바른 위치에 정합이 이루어졌지만, 해당 문제들로 인해 오

차가 발생할 수 있다. 카메라가 직하방을 주시하지 않을 때를 대비하여 센서를 부착하여 카메라 각을 측정하였지만, 카메라 장착 및 센서 부착에서 서로 정렬되지 않는다면 편향된 오차를 보일 수 있다. 따라서, 정밀한 김발 또는 정교한 기기 부착, 데이터베이스 교정 등으로 오차를 줄일 수 있을 것이다.

### III. 결 론

본 논문은 실제 비행 상황에 적합한 영상 기반 항법을 제안하였다. 데이터베이스 영상과 항공 영상의 차이에도 시불변의 특징들의 추출하여 대조함으로써 더욱 강건하게 작동되는 것을 확인하였다. 시뮬레이션뿐만 아니라 비행실험을 통하여 항법을 검증하였으며, 대조 수렴률과 위치 오차 측면에서 우수한 성능을 보였다.

그러나, 템플릿 정합의 경우 수행 시간이 오래 걸린다는 단점이 있다. 이는 항법을 수행하기에 치명적인 단점이다. 각 연산과정이 독립적이므로 병렬 처리 장치를 이용한 병렬화[19], 또는 템플릿이 데이터베이스 상에서 움직이는 이동 거리를 확장[20]하는 등의 방법으로 해결할 수 있을 것이다. 현재 결과는 단순히 정합 결과를 위치로 계산한 것으로 정합 실패로 순간적으로 오차가 커질 수 있다. 이는 실패를 인식하는 로직을 추가하여 해결할 수 있을 것이다. 게다가, INS와 함께 정합 결과를 측정치로 사용하는 필터를 설계한다면 더욱 높은 성능을 보여줄 것으로 기대된다. 수행 시간 측면에서도, 수행 시간은 참조 영상의 크기와 관계가 있어 필터의 공분산 값으로



참조 영상의 크기를 결정한다면 불필요한 계산을 획기적으로 줄일 수 있을 것이다.

## 후 기

본 논문은 풍산-KAIST 미래기술연구센터에서 시행 중인 “상용기반 보조센서에 의한 복합항법 시스템 개발” 과제의 지원으로 수행되었으며, 지원에 감사드립니다.

## References

- 1) Kim, J. M. and Leeghim, H., “INS/Multi-Vision Integrated Navigation System Based on Landmark,” *Journal of the Korean Society for Aeronautical and Space Sciences*, Vol. 45, No. 8, July 2017, pp. 671~677.
- 2) Won, D. H., Chun, S. B., Sung, S. K., Cho, J. S. and Lee, Y. J., “Observability Analysis of a Vision-INS Integrated Navigation System Using Landmark,” *Journal of the Korean Society for Aeronautical and Space Sciences*, Vol. 38, No. 3, March 2010, pp. 236~242.
- 3) Conte, G. and Doherty, P., “Vision-based unmanned aerial vehicle navigation using geo-referenced information,” *EURASIP Journal on Advances in Signal Processing*, 2009, 10.
- 4) Wu, A. D., “Vision-based navigation and mapping for flight in GPS-denied environments,” PhD Thesis, Georgia Institute of Technology, 2010.
- 5) Dumble, S. J. and Gibbens, P. W., “Airborne vision-aided navigation using road intersection features,” *Journal of Intelligent and Robotic Systems*, Vol. 78, No. 2, May. 2015, pp. 185~204.
- 6) Hong, K. W., Kim, S. J., Bang, H. C., Kim, J. W., Seo, I. W. and Pak, C. H., “Particle Filters using Gaussian Mixture Models for Vision-Based Navigation,” *Journal of the Korean Society for Aeronautical and Space Sciences*, Vol. 47, No. 4, April 2019, pp. 274~282.
- 7) Lowe, D. G., “Distinctive image features from scale-invariant keypoints,” *International journal of computer vision*, Vol. 60, No. 2, November 2004, pp. 91~110.
- 8) Bay, H., Ess, A., Tuytelaars, T. and Van Gool, L., “Speeded-up robust features (SURF),” *Computer vision and image understanding*, Vol. 110, No. 3, June 2008, pp. 346~359.
- 9) Calonder, M., Lepetit, V., Strecha, C. and Fua, P., “Brief: Binary robust independent elementary features,” *In European conference on computer vision*, Springer, Berlin, Heidelberg, September 2010, pp. 778~792.
- 10) Karami, E., Prasad, S. and Shehata, M., “Image matching using SIFT, SURF, BRIEF and ORB: performance comparison for distorted images,” *arXiv preprint arXiv:1710.02726*.
- 11) He, K., Gkioxari, G., Dollár, P. and Girshick, R., “Mask r-cnn,” *Proceedings of the IEEE international conference on computer vision* 2017, pp. 2961~2969.
- 12) Noh, H. W., Hong, S. H. and Han, B. H., “Learning deconvolution network for semantic segmentation,” *Proceedings of the IEEE international conference on computer vision* 2015, pp. 1520~1528.
- 13) Long, J., Shelhamer, E. and Darrell, T., “Fully convolutional networks for semantic segmentation,” *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431~3440.
- 14) Chai, D., Newsam, S. and Huang, J., “Aerial image semantic segmentation using DCNN predicted distance maps,” *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 161, March 2020, pp. 309~322.
- 15) Montoya-Zegarra, J. A., Wegner, J. D. and Schindler, K., “Semantic segmentation of aerial images in urban areas with class-specific higher-order cliques,” *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 2, 2015, pp. 127~133.
- 16) Hong, K. W., Kim, S. J. and Bang, H. C., “Vision-based Navigation using Gaussian Mixture Model of Terrain Features,” *AIAA Scitech 2020 Forum*, Jan 2020, p. 1344.
- 17) Briechle, K. and Hanebeck, U. D., “Template matching using fast normalized cross correlation,” *Optical Pattern Recognition XII*, International Society for Optics and Photonics, Vol. 4387, March 2001, pp. 95~102.
- 18) Mahmood, A. and Khan, S., “Exploiting transitivity of correlation for fast template matching,” *IEEE Transactions on Image Processing*, Vol. 19, No. 8, August 2010, pp. 2190~2200.
- 19) Uchida, A., Ito, Y. and Nakano, K., “Fast and accurate template matching using pixel rearrangement on the GPU,” *2011 Second International Conference on Networking and Computing IEEE*, December 2011, pp. 153~159.
- 20) Pele, O. and Werman, M., “Accelerating pattern matching or how much can you slide?,” *Asian Conference on Computer Vision*, November 2007, pp. 435~446.