

의미론적 분할된 항공 영상을 이용한 시각적 의미 문맥의 오차 특성 분석

박준우*, 김성중, 홍경우, 방효충
한국과학기술원

Error Characteristics of Visual Semantic Context using Semantically Segmented Aerial Image

Junwoo Park*, Sungjoong Kim, Kyungwoo Hong, Hyochoong Bang

Key Words : Visual Semantic Context(시각적 의미 문맥), Semantic Segmentation(의미론적 분할), Aerial Image(항공 영상), Vision-based Navigation(영상 기반 항법)

서 론

관성 항법은 항공기의 위치, 속도, 자세 등의 항법 해 도출에서 가장 먼저 가용한 방법이며 위성항법 시스템과 같이 절대 측위가 가능한 보조항법 수단과 결합된 형태로 널리 사용된다. 위성항법 시스템은 상한이 정해진 오차 수준을 가짐으로써 효과적인 보조항법 수단을 제공하지만, 군사적인 목적으로 사용 시 상대의 의도적인 신호 방해 및 교란, 스푸핑 등의 공격에 취약하다는 단점이 있다. 이는 위성과 같은 외부 시스템 및 신호에 의존하는 위성항법 시스템의 특성에 기인하는 것으로, 자족적인 특징을 가진 강건한 절대 측위 대체 수단이 필요하다. 데이터베이스 기반 항법 시스템은 위성항법 시스템의 대표적인 대체 수단이며 최근에는 그중에서 특히, 저렴하고 자족적인 특징을 가진 영상 센서와 상용 영상 지도 데이터베이스를 활용한 절대 측위에 관한 연구가 활발히 이루어지고 있다. Google Maps, OpenStreetMap, VWorld 등과 같이 영상과 위치를 결합한 상용 지도는 누구나 사용 가능하다는 점에서 군사적인 목적에서 벗어나 범용적인 영상 기반 항법 시스템 구축을 가능케 한다는 점에서 특히 매력적이다.

영상 지도와 항공 영상을 이용 및 대조를 통한 항공기의 위치 추정 은 공통적으로 항공 영상 내의 특징점을 추출해 영상 지도의 그것들과 비교하는 과정을 거쳐 이루어진다. 여기서 특징점들이란 종래의 영상 특징 추출기들이 제공하는 모서리, 직선, 가장자리 등을 포함하며, 이들은 계절이나 항공기의 비행 조건 등 지속적으로 변화하는 촬영 조건에는 강건하지 않다. 따라서 이를 직접적으로 영상 지도와 대조하는 방식은 보조항법 시스템으로는 부적합하다. 최근에는 심층 신경망 기술의 발전으로⁽¹⁾ 주어진 영상 내에서 픽셀별 상위 수준의 의미를 추출할 수 있는 의미론적 분할(Semantic Segmentation)이 가능하다. 항공 영상에서 항법 해 도출에 특히 도움이 되는 의미는 도로, 건물, 숲, 논밭 등과 같이 긴 시간 간격 동안 위치변화가 적어서 일관된 특징을 보이는 것들이 있다. 주어진 항공 영상 내에서 시간적 변화가 적은 상위 수준의 의미를 추출할 수 있다면 이들 상호 간의 위치 관계, 패턴 등을 이용해 상용 영상 지도와의 대조를 통해 강건한 항

법해 도출이 가능하다. 홍경우 등⁽²⁾의 연구결과에서는 항공 영상 내 도로와 건물을 추출 후 가우시안 혼합 모델로 근사하여 파티클 필터를 이용한 영상 지도와의 대조 항법을 수행하였다. 김영주⁽³⁾의 연구는 영상 내 인접한 같은 의미들을 클러스터링할 수 있다면 이들 사이의 기하 관계를 이용하여 영상의 스케일 정보를 몰라도 활용 가능한 지도 항법을 소개하였다. 이는 위성의 자세 결정 시스템에 활용되는 별 센서 측정치와 성표 사이의 대조 알고리즘과 유사한 형태이다.

본 연구는 김기섭 등⁽⁴⁾의 연구결과에 영감을 받아 의미론적 분할된 항공 영상에서의 시각적 의미 문맥을 제시하며 이를 이용한 영상 지도 대조 항법으로의 활용, 주어진 영상 측정치의 유일성 지수로의 활용 가능성을 확인하기 위해 제시하는 개념의 수치적인 특성을 분석한다.

본 론

제안하는 개념인 시각적 의미 문맥은 항공기의 직하방을 촬영한 항공 영상의 의미론적 분할을 수행하는 것으로 시작한다. 본 연구가 중점을 두는 영상 내 의미는 도로와 건물인데, 기학습된⁽¹⁾ 네트워크를 Google Maps가 제공하는 한국의 항공 영상과 국토교통부에서 운영하는 공간정보 오픈 플랫폼 지도인 VWorld가 제공하는 픽셀별 의미 라벨을 이용해 추가 학습한 네트워크를 사용한다. 자세한 내용은 선행연구⁽⁵⁾에 소개되어있으며, 영상 속 의미 추출 결과는 Fig. 1에 나타내었다.

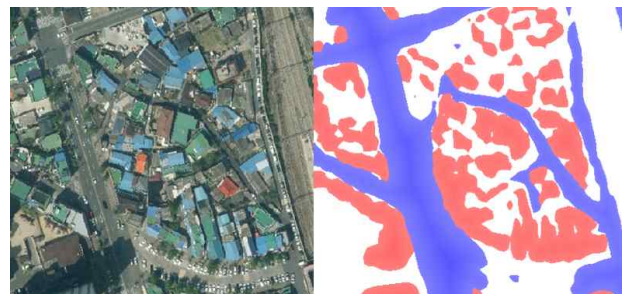


Fig. 1. Aerial image and its semantically segmented version

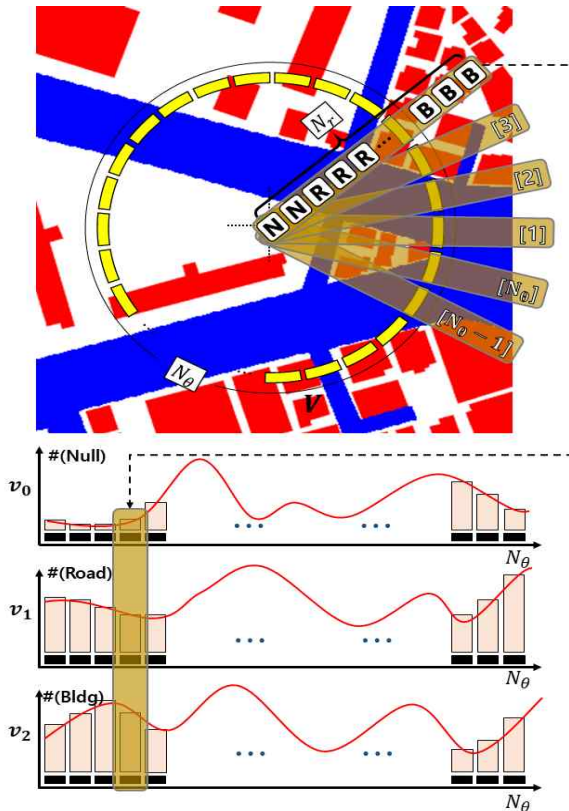


Fig. 2. Calculating visual semantic context from semantically segmented image

이후 의미론적으로 분할된 영상에서 문맥을 해석하고자 하는 지점을 중심으로 방위방향과 방사방향에 놓인 의미들을 모아 Fig. 2.와 같이 수집한다. 특히, j 번째 방향으로 k 픽셀 거리 만큼 떨어진 픽셀을 p_{jk} , 해당 픽셀의 의미를 s_{jk} 라 했을 때, i 번째 시각적 의미 문맥을 아래와 같이 정의한다.

$$v_i = \begin{bmatrix} \sum_{k=1}^{N_r} \delta_i(s_{1k}) \\ \sum_{k=1}^{N_r} \delta_i(s_{2k}) \\ \vdots \\ \sum_{k=1}^{N_r} \delta_i(s_{N_\theta k}) \end{bmatrix}^T + \vec{1}_{N_\theta} \quad (1)$$

여기서 $\vec{1}_n$ 은 모든 요소가 1이고 길이가 n 인 행이며, $\delta_i(s)$ 는 다음을 의미한다.

$$\delta_i(s) = \begin{cases} 1 & \text{if } s = s_i \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

아울러, s_i 는 i 번째 의미를 나타내는데($0 \leq i \leq N_s$) 본 연구에서는 건물과 도로를 사용하였고($N_s = 2$), s_0 , s_1 , s_2 는 각각 의미 없음, 건물, 그리고 도로를 나타낸다. N_r 및 N_θ 는 각각 문맥에 포함할 방사방향의 범위와 방위방향의 이산화 개수로 본 연구에서는 각각 40과 240의 값을 이용하였다. N_r 의 경우 큰 값을 사용하면

특정 지점의 지역적 특징이 아닌 전역적인 통계를 기술하는 것에 가까워지기 때문에 영상의 스케일(고도), 영상 센서의 화각 및 해상도에 따른 적절한 조정이 필요하다. N_θ 의 경우 큰 값을 사용하면 방위방향의 해상도가 높아지고 세밀하게 문맥을 기술할 수 있지만, 특정 값 이상으로는 성능 향상을 기대하기 어렵고 연산량만 증가하여서 다수의 시도를 통해 도출된 해당 값을 사용하였다.

(1)을 이용해 영상 내에서 해석하고자 하는 지점에서의 시각적 의미 문맥 V 를 아래와 같이 정의한다.

$$V = \begin{bmatrix} v_0 \\ v_1 \\ \vdots \\ v_{N_s} \end{bmatrix} \quad (3)$$

여기서 V 의 앞 m 개의 열을 뒤쪽으로 옮겨 (4)와 같이 치환 변환한 V' 은 항공 영상을 $2\pi m/N_\theta$ rad 회전하여 새로이 문맥을 계산한 것과 같다.

$$V' = V \begin{bmatrix} 0 \cdots 0 & I_m \\ 0 \\ I_{N_\theta - m} & \vdots \\ 0 \end{bmatrix} \quad (4)$$

따라서 제안하는 시각적 의미 문맥 V' 는 특정 위치를 기술하는 기술자로서 역할함과 동시에 치환 변환을 통해 항공기의 heading을 기술하는 것도 가능하다.

두 문맥 간의 차이는 쿨백-라이블러 발산, 카이-제곱 거리 등을 이용해 정량적으로 계산한다. 영상 지도 내 넓은 관심 영역에 대해 문맥을 미리 확보하고, 주어진 영상 속 문맥과의 차이를 계산하면 제안하는 개념의 오차 특성 분석이 가능하다. 나아가서는 현재 영상 측정치의 유일성 및 정보량을 정량화할 수 있는 지수로써도 활용할 수 있다.

결론

본 연구는 의미론적 분할된 항공 영상을 이용해 항공기의 위치 및 heading을 기술할 수 있는 시각적 의미 문맥을 제안하였으며, 영상 지도와의 비교를 통해 제안하는 개념의 오차 특성을 제시한다.

참고문헌

- 1) Long, J., Shelhamer, E., Darrell, T., "Fully convolutional networks for semantic segmentation," *In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, 2015, pp. 3431~3440.
- 2) Hong, K., Kim, S., Bang, H., "Vision-based Navigation using Gaussian Mixture Model of Terrain Features," *In AIAA Scitech 2020 Forum*, 2020, p. 1344.
- 3) Kim, Y., "Aerial Map-Based Navigation Using Semantic Segmentation and Pattern Matching," *arXiv preprint*, 2021,

arXiv:2107.00689.

4) Kim, G., Kim, A, "Scan context: Egocentric spatial descriptor for place recognition within 3d point cloud map," *In 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018, pp. 4802~4809.

5) Hong, K., Kimg, S., Bang, H., "Aerial Image Segmentation and Template Matching for Vision-aided UAV Navigation," *Proceeding of the 2018 KSAS Fall Conference*, 2018, pp. 599~600.