

# セマンティックセグメンテーションによる衛星画像を用いた UAV の自己位置推定の精度向上

Improving The Accuracy of Visual Localization for UAV  
Using Satellite Imagery by Semantic Segmentation

○ 徳尾野 陸 (中央大学) Sarthak Pathak (中央大学)  
正 梅田 和昇 (中央大学)

Riku TOKUONO, Chuo University, toku.riku.0317@gmail.com

Sarthak PATHAK, Chuo University

Kazunori UMEDA, Chuo University

In this paper, we propose a method to accurately estimate the position of a UAV on satellite images by using the region information of buildings extracted by semantic segmentation for UAV pose estimation. The localization of UAV using satellite images is based on the similarity between images features using intensity information. However, the estimation accuracy of intensity information alone is degraded by changes in the seasons and moving objects. Therefore, in addition to the intensity information, the proposed method uses information on building areas extracted by semantic segmentation from overhead images to improve the accuracy of position estimation.

**Key Words:** UAV, Localization, Semantic Segmentation, Satellite image

## 1 序論

現在, 無人航空機 (unmanned aerial vehicle, UAV) は農業や橋梁の点検, 測量など幅広い分野で活用されており, 将来的にはさらなる分野での活用が期待されている。UAV は地上を移動するロボットに比べ, 障害物や道路の状況に左右されず移動することや上空で移動または作業が出来る点で優れている。この点を生かし高層ビルや橋梁の点検作業や市街地での巡回警備や配送業務への利用が考えられる。これらのような UAV は人による操縦ではなく無人での周囲の環境を考慮した自律的な飛行が求められるが, このような自律飛行では自己位置推定や障害物を考慮した経路計画, 環境地図の作成が必要であり, 特に自己位置推定は自律飛行に必要不可欠な技術である。

UAV の自己位置推定において最も幅広く使われている技術として GPS に代表される全球測位衛星システム (GNSS) がある。GPS は上空にある複数の衛星からの信号を GPS 受信機で受け取り, 受信者の位置を衛星からの相対位置によって取得するシステムである。GPS は平地など開けた場所においては比較的高精度な自己位置推定が実現可能であるが高い建造物付近などの GPS 信号の受信しづらい環境においてや GPS 信号の受信機が故障した場合正確に自己位置を推定が出来なくなり, 自律飛行システムにおいて致命的なトラブルを起こす可能性がある。

そのため GPS による自己位置推定が不可能になった場合他の手段で自己位置推定を行う必要がある。自己位置推定を行う用途のみに新しいセンサを UAV に搭載することはコスト面や搭載性を考えたうえで望ましくないため UAV に標準的に搭載されているセンサであるカメラの情報のみで自己位置推定を完結させることが望ましい。そのためカメラから得られた画像のみで精度の高いグローバルな位置推定を行うシステムが必要である。

衛星画像を用いた UAV の自己位置推定に関する研究は, 現在数多く行われている。古賀ら [1] は, 衛星画像と UAV の撮影画像を SURF 特徴量を用いた特徴点マッチングによって衛星写真上から UAV の位置を特定することで UAV の位置を推定する手法を提案した。しかし衛星画像と撮影画像の画角の違いなどによりマッチングすることが出来ないことや特徴量の疎な画像では特徴点の抽出が困難であり適用できる環境に限られるという問題点がある。Yol ら [2] は衛星画像と UAV の撮影画像との変換行列を相互情報量を用いて求めることによって UAV の位置を推定する手法を提案した。しかし変換行列を求めるための計算コストの大きさと前フレームからのカメラの移動が微小であるという前提により成り立っているため大きな UAV の移動に対応できないという問題点がある。そして衛星画像と撮影画像を用いた UAV の自

己位置推定には衛星画像が撮影された時期と UAV が撮影した時間帯や季節が異なることによる照度や移動物体により自己位置推定の精度が低下するという問題がある。

本論文では, セマンティックセグメンテーションによって抽出された建物の領域情報を用いた衛星画像を事前情報とするパーティクルフィルタに基づく UAV の自己位置推定手法を提案する。セマンティックセグメンテーションは機械学習を用いた画像のラベル付けアルゴリズムであり, 事前に学習させるデータによって幅広い時間帯や環境に対してロボストに物体の抽出を行うことが可能である。この手法を用いて時間によって外観の変化しにくい建物領域を抽出し利用することによって位置推定の精度を向上させる。

## 2 提案手法

### 2.1 提案手法の概要

本手法では, UAV から得られた鉛直下向き画像を高解像度衛星画像中で画像の類似度を用いて探索と追跡を行うことによって Fig.1 のように衛星画像上で自己位置の推定を行う。衛星画像の全領域の探索には多大な計算コストがかかるため探索と追跡にはパーティクルフィルタを用いる。UAV から得られた画像と衛星写真にセマンティックセグメンテーションを適用し, 建物領域のラベル画像を作成する。このとき得られた建物のラベル画像と元画像の両方の類似度をパーティクルフィルタ内の尤度計算の部分に適用する。提案手法のフローチャートを Fig.2 に示す。

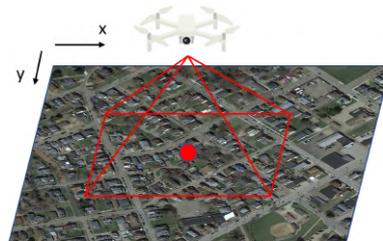


Fig.1 Exploring UAV captured image on pre-existing satellite image

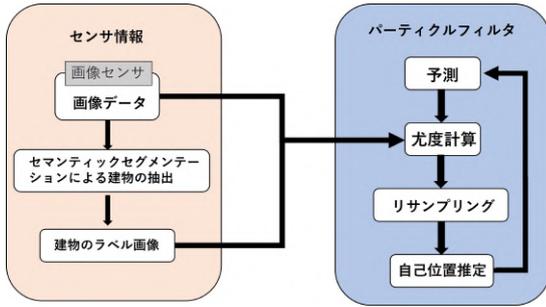
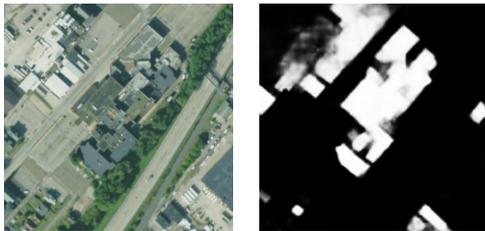


Fig.2 Flowchart of proposed framework

## 2.2 建物領域の抽出

提案手法では、衛星画像と撮影画像からの建物の抽出に U-Net を用いた建物のセマンティックセグメンテーションモデルである building-footprint-segmentation[3] を用いる。学習には、元衛星画像と建物のアノテーション済み衛星画像のデータセットである Massachusetts Building Dataset[4] を用いた。このモデルに UAV の撮影画像と衛星画像の双方を入力し建物領域を抽出したラベル画像を出力する。入力例と出力例を Fig.3 に示す。また衛星画像の入力例と出力例を Fig.4 に示す。出力例は輝度値が高いほどその画素が建物領域である可能性が高いことを示している。



(a) Example of input UAV captured images (b) Example of semantic segmentation result

Fig.3 Extraction of building area from UAV captured images by semantic segmentation



(a) Input of satellite image



(b) Semantic segmentation result

Fig.4 Extraction of building area from satellite image by semantic segmentation

## 2.3 建物の領域情報を考慮した尤度計算

パーティクルフィルタは、時系列フィルタリングの一種であり、過去の状態から未来の状態を逐次的に推定できる手法である。この手法では、有限個のパーティクルの分布によって UAV の位置の確率分布の近似を行う。パーティクルフィルタにおける計算処理は、予測、尤度計算、リサンプリング、自己位置推定の4つに分かれており、これを逐次的に繰り返すことで自己位置推定を行う。本手法では尤度計算の部分に着目する。

$t$  フレームにおいて UAV が取得したサイズ  $(u, v)$  の画像を画像  $I$  とする。また UAV の高度に応じた係数  $k$  とし各パーティクルを中心にサイズ  $(ku, kv)$  で衛星画像から切り取り、サイズ  $(u, v)$  に縮小した画像を画像  $I'$  とする。これらの画像  $I, I'$  に対して画像間相互情報量を用い画像間の類似度  $MI(I, I')$  を求める。UAV の回転方向による影響をなくすために画像  $I, I'$  の双方に対して Fig.5 のように円形にマスクを行う。また画像  $I, I'$  はグレースケール画像であるとする。画像  $I$  の持つ情報量  $H(I)$  は式 1 によって定義され、結合情報量  $H(I, I')$  は、式 2 で表される。

$$H(I) = - \sum_{i=0}^{255} P_I(i) \log(P_I(i)) \quad (1)$$

$$H(I, I') = - \sum_{i=0}^{255} \sum_{j=0}^{255} P_{II'} \log(P_{II'}(i, j)) \quad (2)$$

これらを用いて画像間相互情報量  $MI(I, I')$  は式 3 で表される。

$$MI(I, I') = H(I) + H(I') - H(I, I') \quad (3)$$

画像  $I$  に対してセマンティックセグメンテーションによる建物の抽出を行った画像を画像  $J$  とする。また、セマンティックセグメンテーションによってラベル化した衛星画像から画像  $I'$  と同様の方法で切り取った画像を画像  $J'$  とする。画像  $J, J'$  に対しても同様にマスクを行う。これらの画像  $J, J'$  に対して節 2.2 で述べた SAD を用い画像間の類似度  $SAD(I, I')$  を求める。画像内の画素を  $(i, j)$  とし、画像  $J$  の各画素の画素値を  $I(i, j)$ 、画像  $J'$  の各画素の画素値を  $J(i, j)$  とすると SAD の値  $SAD(J, J')$  は式 4 のように求められる。

$$SAD(J, J') = \sum_{i=0}^{u-1} \sum_{j=0}^{v-1} |J(i, j) - J'(i, j)| \quad (4)$$

これらの値を用いて各パーティクルの尤度の計算を行う。画像間の類似度が大きくなればなるほど  $MI(I, I')$  は大きく、 $SAD(J, J')$  は小さくなるため各パーティクルの尤度  $L_t^{(i)}$  は式 5 で表す。

$$L_t^{(i)} = \frac{MI(I, I')}{SAD(J, J')} \quad (5)$$

求めた各パーティクルの尤度を尤度の総和で割ることによって尤度を正規化し重みを求める。パーティクルの総数を  $n$  とすると各パーティクルの重み  $\epsilon_t^{(i)}$  は式 6 によって求められる。

$$\epsilon_t^{(i)} = \frac{L_t^{(i)}}{\sum_{i=0}^N L_t^{(i)}} \quad (6)$$

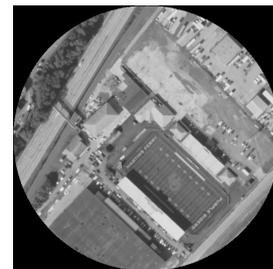


Fig.5 Example of Mask image

### 3 実験

#### 3.1 実験条件

提案手法により, UAV が衛星画像を用いた自己位置推定において輝度情報のみを類似度を用いた場合と比較しより高い精度が得られることを検証するために, シミュレーション実験を行った. UAV の空撮画像のデータセットを用いて, 相互情報量のみのパーティクルフィルタと提案手法のパーティクルフィルタのそれぞれで自己位置推定をし, 評価を行う. Fig.3 に例を示す UAV の撮影画像のデータセット [5]113 フレーム分と Fig.4 に示すアメリカ, オハイオ州マーティンズフェリー周辺のサイズ (8192,4320) の衛星画像を用いる. UAV の撮影画像と衛星画像は異なる季節のものを使用する. 真値にはデータセットに含まれる GPS の値を用いる. 初期位置の探索に用いるパーティクルを 1000, 追跡に用いるパーティクルを 50, 100, 200 としてそれぞれ実験を行った. また, ロバスト性を示すため Fig.6 に示す  $\sigma = 40$  のガウシアンノイズを加えた衛星画像でも同様に実験を行った.



Fig.6 Satellite image with gaussian noise( $\sigma = 40$ )

#### 3.2 実験結果

提案手法と従来手法のそれぞれで位置推定を行い, 推定された位置と真値を比較することによって評価を行った. 従来手法では尤度計算に相互情報量のみを用いた. 何も手を加えていない衛星画像での RMSE を Fig.7, Table.1 に, ノイズを加えた衛星画像での RMSE を Fig.8, Table.2 に示す. ノイズを加えた場合とくわえていない場合の双方において提案手法の RMSE は従来手法の RMSE より減少していることが分かる. ノイズを加えていない場合において RMSE は, パーティクル数が 50 のときに約 13%, パーティクル数が 100 のときに約 11%, パーティクル数が 200 のときに約 12% 減少している. ノイズを加えた場合において RMSE は, パーティクル数が 50 のときに約 13%, パーティクル数が 100 のときに約 10%, パーティクル数が 200 のときに約 12% 減少している. 全ての条件において提案手法の RMSE は従来手法に比べ 10% 程度の減少を確認することができた.

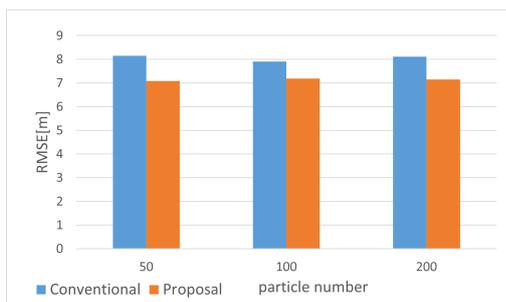


Fig.7 Comparison of RMSE

Table 1 Root Mean Squared Error[m]

	Conventional method	proposed method
50	8.10	7.07
100	7.95	7.09
200	8.07	7.13

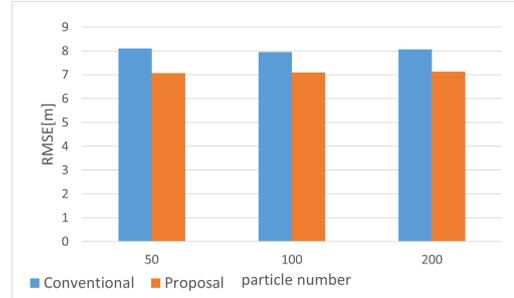


Fig.8 Comparison of RMSE with noise

Table 2 Root Mean Squared Error with noise[m]

	Conventional method	Proposed method
50	8.15	7.09
100	7.90	7.19
200	8.11	7.14

### 4 結論

本論文では, セマンティックセグメンテーションによって抽出した建物の領域情報をパーティクルフィルタの尤度計算に用い衛星画像上で位置推定する手法を提案した. 提案手法では季節の相違やノイズのある衛星画像上において従来手法と比べ位置推定の精度が向上した. 今後の展望として, 更に幅広い環境での実験や路面等の領域情報を適用することを考えている.

#### 参考文献

- [1] 古賀 丈晴, 小川 洋史, 松本 剛明, 米本 浩一, “火星の地図情報とカメラ撮影画像による自己位置推定のアルゴリズム”, 日本機械学会九州支部講演論文集 2014.67 巻 2014.
- [2] A. Yol, B. Delabarre, A. Dame, J. E. Dartois and E. Marchand, “Vision-based absolute localization for unmanned aerial vehicles,” *IEEE International Workshop on Intelligent Robots and Systems (IROS)*, pp. 3429-3434, 2014.
- [3] Palnak Fuzail A., building-footprint-segmentation.(v0.2.4),2022 [Source code].<https://github.com/fuzailpalnak/building-footprint-segmentation>.
- [4] Volodymyr Mnih, “Machine learning for aerial image labeling,” University of Toronto (Canada), 2013.
- [5] Ivan Cisneros et al., “A Large-Scale Dataset for UAV Visual Place Recognition and Localization”, *Computer Vision and Pattern Recognition*, arXiv:2207.1231 2022.