H-027

マスクを用いた堅牢な自己位置推定
Robust Self-Localization Leveraging Image Masks

厳 海旻† Haimin Yan　石原 章翔† Akito Ishihara　嶋田 知泰† Tomoyasu Shimada　孔 祥博‡ Xiangbo Kong　冨山 宏之† Hiroyuki Tomiyama

1. はじめに

近年、深層学習の進展により、視覚的自己位置推定は、さまざまな応用分野において重要な役割を果たしています。しかし、季節変動や植生などの環境要因により、位置推定精度が低下するという課題はあります。本研究では、この問題に対処するために、色検出および画像セグメンテーションを用いたマスキング手法を提案し、高精度な視覚的自己位置推定の実現を目指します。マスキングの有無による位置推定性能を、異なる 2 種類の撮影条件下で比較・分析し、位置推定誤差を定量的に評価しました。その結果、特定の環境条件において、マスク処理を適用し、位置および姿勢推定の精度向上が確認されました。

2. 提案手法

本手法は色検出と深層学習に基づくマスキング技術を組み合わせ、視覚的自己位置推定精度の向上を目的としています。図 1 に示すように、本研究では 3 段階のマスキング処理を適用し、画像から植生を段階的に除去、自然要素が位置推定精度に与える影響を低減します。

図 1 :マスキング処理の流れ

具体的には、まず第 1 段階として、色検出によって植生領域を識別し、バウンディングボックスを生成します。その後、Segment Anything Model (SAM) [1]を用いて大部分の植生を除去します。第 2 段階では、前段階でマスクされなかった残存領域に対して再度色検出を行い、新たなバウンディングボックスを生成し、再度 SAM による処理を行うことで、マスキングの精度と網羅性をさらに向上させます。第 3 段階では、まだ完全にマスクされていない植生領域に対して追加の色検出を行い、補足的なマスキングを適用することで、植生情報を徹底的に除去します。最終的に得られた画像は、HLoc [2]への入力として使用され、マスキングを施していない元画像と位置推定性能を比較しました。この比較実験により、本手法が晴天や曇天といったさまざまな環境条件下において、視覚的自己位置推定精度を向上させる効果があることが実証されました。

2.1 色検出と SAM を用いたマスキング処理

画像内の植生の色と背景の他の要素（建物や道路など）の色が類似しているため、緑の植生が画像内でノイズとなり、自己位置推定アルゴリズムがターゲットを正確に識別することを困難にし、安定性と精度に悪影響を与えます。この課題に対応するため、本研究では OpenCV による色検出と SAM を用いたマスキング処理を提案します。

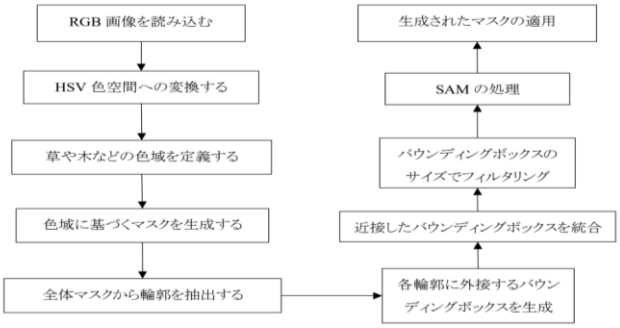

図 2:色検出からマスク処理までのフローチャート

図 2 に示すとおり、まず RGB 画像を読み込んだ後、色検出処理の前段階として、RGB 色空間から HSV 色空間への変換を行います。次に、HSV 空間において定義された特定の色域範囲に基づき、対象となる植生領域を抽出するためのバイナリマスク画像を生成します。このマスク画像により、背景と植生との明確な識別が可能となり、画像内の不要な要素を排除する基盤が整います。その後、生成されたマスク画像から対象物の輪郭を抽出し、各輪郭の外接領域を囲む矩形を設定し、対象領域を定義します。得られた矩形情報は近接する領域を統合しつつ、過度に小さいものを除外します。最後に、抽出された領域情報を SAM に入力し、画像内の植生領域に対する精密なセグメンテーションを実施し、生成されたマスクを元画像に適用し、植生を除去した画像を得ることができます。

2.2 追加の色検出のマスク処理

色検出と SAM を用いたマスキング処理を行った後の画像には、依然として一部にマスクされていない領域が残る場合があります。このような未処理の領域が存在することにより、後続の処理精度に影響を及ぼす可能性があります。この課題に対応するために、本研究では追加の色検出処理を適用し、マスク後に残存する細部領域を検出したうえで、より精緻なマスキングを行う手法を導入します。

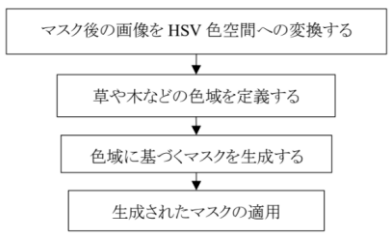

図 3：追加の色検出からマスク処理までのフローチャート

† 立命館大学　Ritsumeikan University
‡ 富山県立大学　Toyama Prefectural University

図 3 に示すように、まずマスキング処理後の画像を再度 HSV 色空間に変換し、対象となる未マスク領域の色域を定義します。次に、定義された色域に基づいて新たなバイナリマスクを生成し、これをマスキング後の画像に適用することで、初回処理では対応できなかった領域の補完を行います。

3. 実験

3.1 実験で使用するデータセット

本実験で使用したデータセットは立命館大学ローム記念館にて、Open Camera [3] を用いて屋外の曇天環境下で撮影されたものであり、合計 568 枚の画像から構成されています。これらの画像は 3D 再構築ツールである COLMAP [4] を用いて処理され、特徴抽出、マッチング、カメラ姿勢の推定、3D 点群の再構築を経て、ローム記念館の 3D 点群モデルが生成されました。本研究では、この点群モデルを基に、マスキング処理を施した画像と未処理画像に対する自己位置推定の精度を比較し、異なる環境条件下におけるマスキング手法の有効性を評価しました。

3.2 評価方法と指標

本実験では、作成したデータセット上での精度を検証するために、クエリ画像の撮影場所をデータセットと同様に立命館大学 BKC キャンパスのローム記念館とします。撮影環境は曇天時と晴天時の 2 種類があり、各環境において評価に使用したクエリ画像はそれぞれ 26 枚ずつです。本実験では、COLMAP により取得されたローカル座標系上の真値と HLoc による推定結果を比較し、位置および姿勢の誤差を算出することで、定量的な評価を行います。

誤差の標準的な評価指標として、本研究では位置の評価に対しては位置誤差である Relative Translation Error (RTE) と、姿勢の評価に対しては回転誤差である Relative Rotation Error (RRE) を採用します。これらの指標は、推定された位置および姿勢がグラウンドトゥルースと比較してどの程度の誤差を有するかを定量的に測定するためのものであり、本実験においては以下の式に基づいて計算を行います。

まず、RTE の計算式を式（1）に示します。式（1）は、推定されたカメラの位置ベクトルt_{est}と真のカメラ位置ベクトルt_{gt}の間の直線距離を、ユークリッド距離（2 ノルム）を用いて求めるものです。この計算により、空間的な位置のずれを定量的に評価することが可能となります。

$$RTE = \left\| t_{est} - t_{gt} \right\|_2 \tag{1}$$

次に、RRE の計算式を式（2）に示します。式（2）は推定された回転行列R_{est}と真の回転行列R_{gt}との間の角度差を、行列のトレース（対角要素の和）を用いて測定し、逆余弦関数を適用し、角度誤差（ラジアン単位）として算出するものです。

$$RRE = \left(\frac{trace\left(R_{gt}^{\mathrm{T}} \cdot R_{est}\right) - 1}{2} \right) \tag{2}$$

3.3 実験結果

表 1 には位置および姿勢に関して、真値に対するマスクなしとマスクありの誤差の差分（マスクあり − マスクなし）を算出し、その平均値、標準偏差、および中央値を、2 種類の撮影環境ごとにまとめた結果を示しています。なお、位置はローカル座標系に基づいて評価しているため、特定の単位は付与していません。位置誤差の評価には RTE、姿勢誤差の評価には RRE の差分を用いています。

表 1:真の位置に対するマスクなしとマスクありの誤差の差

指標	曇りの日	晴れの日
位置誤差：平均	-0.001758	0.001113
位置誤差：標準偏差	0.005651	0.005422
位置誤差：中央値	-0.000416	0.000107
回転誤差[度]：平均	-0.039871	-0.009182
回転誤差[度]：標準偏差	0.098141	0.126231
回転誤差[度]：中央値	-0.028690	-0.002236

表 5.1 の結果より、まず曇天環境においては、平均値および中央値の双方から、マスクを適用することにより位置および姿勢の両方において精度の向上が確認され、特に姿勢に関してはわずかながら精度が改善されていることが示されました。次に、晴天環境においては平均値および中央値より、マスクの適用によって位置の精度が低下する傾向が見られました。

一方で、回転誤差については平均値および中央値の観点からは精度の向上が示されているものの、標準偏差がこれらの統計値に比して非常に大きな値を示します。そのため、晴天環境においてはマスクの効果が限定的であり、マスク処理を行う必要性は低いと考えられます。

4. おわりに

本研究では、植生領域に対してマスキング処理を施し、さらに HLoc を用いた自己位置推定を行い、自己位置推定の精度向上を図るために、独自に構築したデータセットを活用しています。自己位置推定の精度は、定性的評価および定量的評価の両面から検証されました。定性的評価では、位置スケーリングの影響は小さい一方で、姿勢推定は撮影環境に大きく依存してばらつきが見られました。定量的評価では、データセット取得時に曇天条件で撮影された画像において、マスキング処理により位置および姿勢の精度がわずかに向上する傾向が見られました。しかしながら、他の撮影条件下では、マスキングの効果は限定的であることが確認されました。

謝辞

本研究の一部は公益財団法人スズキ財団の科学技術研究助成、および NEDO の 委託(JPNP22006)による。

参考文献

[1] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T, Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollar, and R. Girshick, ” Segment Anything, ” In IEEE/CVF International Conference on Computer Vision, pp. 4015-4026, 2023

[2] P.-E. Sarlin, C. Cadena, R. Siegwart, and M. Dymczyk, ” From Coarse to Fine: Robust Hierarchical Localization at Large Scale,” In IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 12716- 12725, 2019.

[3] Open Camera, [Online]. Available: https://opencamera.org.uk/ [Accessed on Jan 2025].

[4] Colmap, GitHub, [Online]. Available: https://github.com/colmap/colmap [Accessed on Jan 2025].