

UAV 画像のあいまい性・取得時期の違いを考慮した機械学習による 河川地被分類手法の精度検討

芝浦工業大学大学院 学生員 ○百瀬 文人
元芝浦工業大学 近藤 大樹
(株)建設技術研究所 正会員 永矢 貴之

(株)建設技術研究所 正会員 佐藤 拓也
(株)建設技術研究所 正会員 岩見 収二
芝浦工業大学 正会員 宮本 仁志

1. はじめに

近年、砂州や高水敷に安定的な植生域が拡大繁茂し、治水と生態系の双方の観点から河川管理上の問題となっている。これに対して既往研究¹⁾では、UAV(Unmanned Aerial Vehicle)で得られる河川空間情報に機械学習を適用することで、河道内土地被覆の自動判別手法を検討した。しかし、影や木本・草本の混在など、画像のもつ「あいまい」性が介在する部分で誤分類となることが課題となっていた。さらに、**図-1(a)**の赤丸で例示される植生域では、**図-1(b)**でより深い緑となる。このように時期の異なる画像間ではデータの質に違いがあるため、このような場合の精度検討も課題であった。そこで本稿では黒部川 4-6km 区間を対象として以下の 2 つの検討を行った: 1) 2017 年の UAV 画像について、あいまい性を陽にした新しい土地被覆分類を定義し、その有無が機械学習の精度に及ぼす影響、2) 2018 年の UAV 画像により機械学習を行い、2017 年を地被分類することで時期の違いが精度に及ぼす影響。



(a) 2017年11月28日撮影



(b) 2018年11月3日撮影

図-1 UAVのRGBオルソ画像

2. 解析手法

2.1 現地観測・SfM 処理と画像の前処理

2017・2018 年において黒部川の対象区間を UAV 搭載のマルチスペクトルカメラで撮影し、PIX4D による SfM 処理により RGB 画像と正規化植生指数 NDVI(Normalized Difference Vegetation Index)、数値表層モデル DSM(Digital Surface Model)を取得した。DSM はさらにウェーブレット変換を用いて河川縦断勾配を除いた標高データに加工し、ラプラシアンフィルタを用いて DSM のエッジを抽出した¹⁾。

2.2 あいまい性を考慮した土地被覆分類の定義

既往研究¹⁾では土地被覆の真値は RGB を元にして手動で水面・裸地・草本・木本に分類・同定していた。本稿ではそれらに加えて、あいまい部分を考慮した新しい分類を定義した。分類は基本的に RGB の目視によって行った。表-1 に目視判断基準を示す。目視判断基準には、当該オブジェクト(後述)内で明確に土地被覆分類が可能な面積占有率を用いた: 80%以上;明確な分類, 60~80%;あいまいな分類, 60%未満;モザイク。さらに目視判断が困難な場合は、機械学習で使用する特徴量 (NDVI/DSM 高波数) のしきい値を定義し、それを判断基準とした。その上で、データセット作成に際してこれらの組み合わせにより正確真値・中間真値・あいまい真値を設定した: 正確真値;表-1 の①~

表-1 RGB 画像の目視判断基準

| 占有率 | 80%以上 | 60%~80% | 60%未満 |
|--------|-------|---------|------------|
| 土地被覆分類 | ①水面 | ⑥あいまい水面 | ⑩モザイク |
| | ②裸地 | ⑦あいまい裸地 | ⑫その他、河川構造物 |
| | ③草本 | ⑧あいまい草本 | ⑬定義外 |
| | ④木本 | ⑨あいまい木本 | |
| | ⑤影 | ⑩あいまい影 | |

表-2 あいまい性の検討におけるデータセット

| | セットA | セットB | セットC |
|--------|--------|------|------|
| 教師データ | あいまい真値 | 中間真値 | 正確真値 |
| テストデータ | 正確真値 | 正確真値 | 正確真値 |

目視判断基準には、当該オブジェクト(後述)内で明確に土地被覆分類が可能な面積占有率を用いた: 80%以上;明確な分類, 60~80%;あいまいな分類, 60%未満;モザイク。さらに目視判断が困難な場合は、機械学習で使用する特徴量 (NDVI/DSM 高波数) のしきい値を定義し、それを判断基準とした。その上で、データセット作成に際してこれらの組み合わせにより正確真値・中間真値・あいまい真値を設定した: 正確真値;表-1 の①~

④, 中間真値;これに⑥~⑨を追加したもの, あいまい真値;さらに⑤および⑩~⑬を追加したもの. なお, 機械学習を行う際には, 例えば「あいまい水面」は「水面」として学習させた.

2.3 データセットの作成とランダムフォレストによる土地被覆分類

機械学習のデータセットは, GIS で RGB 画像からオブジェクトを作成し, そこに含まれる特徴量の平均・標準偏差を用いたり. 表-2 にあいまい性の検討を行うデータセット A~C の内訳を示す. あいまい性の検討における教師データは, あいまい真値・中間真値・正確真値と変化させた. テストデータは一貫して正確真値を用いた. 一方, 取得時期の違いによる影響を検討するデータセットは D とし, 教師データに 2018 年の正確真値を, テストデータには 2017 年のものを用いた. 機械学習アルゴリズムには RF(Random Forest)を用いた. 精度評価には正解率と再現率の調和平均である F 値に加え, セット D の検討では ArcGIS での画像出力も参照した.

3. 結果と考察

3.1 あいまい性の有無による影響

図-2 に各データセットを用いた学習アルゴリズム別の分類結果を示す. この図の縦軸は F 値であり, それぞれの棒グラフは土地被覆の分類精度である. 図-2 より, F 値はセット A から C で草本 4%, 裸地・木本で 1%増加する. 一方, セット B と C の比較では全ての土地被覆で F 値の変化はない. 以上より, 精度向上に寄与した要因は, ⑥~⑨で定義されたあいまい分類ではなく,モザイクや影などを分けて学習させたことによると判断される.

3.2 撮影時期の違いによる影響

図-2 のセット D の結果では裸地の F 値が 90%と比較的高精度となった. これは, 時期によらず対象地点で裸地の RGB 値が高かったため, 特徴が効果的に学習できたと考えられる. これより, 別時期に取得されたデータを用いても, 特徴量の分布特性によってはある程度精度よく地被分類できる可能性が示唆された.

図-3 にセット D における真値と分類結果の誤分類箇所を示す. 図-3(b)の誤分類の分布と図-3(a)の真値との比較より, 裸地が水面に, また草本が木本に誤分類される頻度が高い. これら誤分類が起こった土地被覆は互いに似た特徴量をもっていた. さらに, 本来真値と同じ分類に同定されるべき土地被覆では, 取得時期の違いによって特徴量の分布特性も歪むため, 誤分類になったと推察される. 今後は, このような誤分類の原因を詳細に調べ, 時期による特徴量の変質が土地被覆の同定精度に与える影響を検討する予定である.

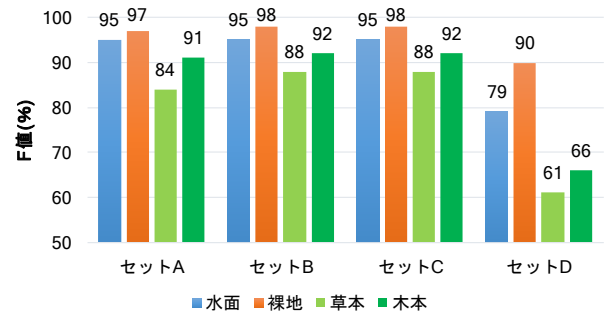
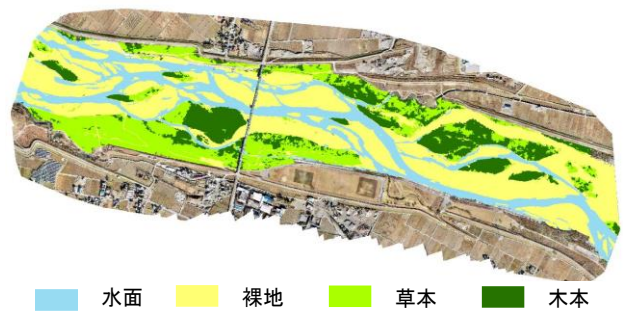


図-2 各データセットにおける土地被覆分類の精度 (F 値)



(a) テストデータの真値



(b) 誤分類

図-3 セット D におけるテストデータの真値と土地被覆分類結果の誤分類箇所