

令和 5 年度
修士学位論文

歩行型 LiDAR から得られる
園地点群データを用いた
ユズ樹木の葉数推定に関する研究

A Study on Estimating the Number of Yuzu Leaves
from Orchard Point Cloud Data
Using Backpack LiDAR

1265103 末廣 いのり

指導教員 栗原 徹

2024 年 2 月 28 日

高知工科大学大学院 工学研究科 基盤工学専攻
情報学コース

要旨

歩行型 LiDAR から得られる 園地点群データを用いた ユズ樹木の葉数推定に関する研究

末廣 いのり

ユズ栽培の課題として労働量の多さがある。ユズ栽培は山間部で行われることが多く、機械の導入が難しい。また、機械化が難しい作業も多く、必要な労働量は膨大である。しかし、高齢化によって農家数や耕作面積は減少傾向にあり、栽培の効率化による労働量の削減が求められる。果樹栽培の効率化を実現するにあたって、葉果比の管理が大きな役割を持つ。適切な葉果比を保たないと隔年結果が生じてしまう。したがって、安定した生産を行うためには樹木ごとに適切な葉果比率を維持する必要がある。しかし、手作業で果実と葉の数を数えるのは膨大な時間がかかり管理も難しい。

本研究では、樹木の葉数を自動で推定することを目的として、LiDAR から得られる点群データを用いた葉数推定法について提案した。園地データから各樹木データを分離させた後、樹木ごとに特徴量を抽出し、主成分回帰分析にて推定モデルを算出した。検証は一つ抜き交差検証にて行い、評価には MAPE を用いた。結果として、最も良い MAPE は 2 つの園地でそれぞれ 13.4%、9.5%となった。特徴量をそのまますべて使用した重み付き重回帰分析の MAPE よりも値が小さくなったことから、主成分の形成が多重共線性および次元の軽減に繋がり、より推定精度が高い推定モデルの算出が可能になったと考えられる。また、MAPE の結果と主成分係数から樹高の情報が推定モデルの推定精度の低下に繋がっていることが確認できた。このことから、対象の園地の特徴に応じて葉数推定に効果的な特徴量を取捨選択することがよりよい推定モデルに繋がることが分かった。

キーワード 葉果比, LiDAR, 点群データ, 主成分回帰分析, MAPE

Abstract

A Study on Estimating the Number of Yuzu Leaves from Orchard Point Cloud Data Using Backpack LiDAR

SUEHIRO, Inori

One of the problems of yuzu cultivation is the large amount of labor involved. Yuzu cultivation is often done in mountainous areas, where it is difficult to introduce machinery. In addition, there are many tasks that are difficult to mechanize, and the amount of labor required is enormous. However, the number of farmers and the area under cultivation are decreasing due to the aging of the population, and there is a need to reduce the amount of labor by improving the efficiency of cultivation. In order to realize efficient fruit cultivation, the management of the leaf-fruit ratio plays a major role. Failure to maintain an appropriate leaf-fruit ratio can result in biennial results. Therefore, it is necessary to maintain an appropriate leaf-fruit ratio for each tree to ensure stable production. However, manually counting the number of fruits and leaves is time-consuming and difficult to manage.

In this study, we proposed a method for estimating the number of leaves on trees using point cloud data obtained from LiDAR, with the aim of automatically estimating the number of leaves on trees. After separating tree data from orchard data, features were extracted for each tree, and an estimation model was calculated using principal component regression analysis. The model was validated by one-by-one cross-validation, and MAPE was used for evaluation. The best MAPE was 13.4% and 9.5% for the two

parks, respectively. The fact that the value was smaller than the MAPE of the weighted multiple regression analysis in which all the features were used as they were, suggests that the formation of principal components leads to multicollinearity and dimensionality reduction, enabling the calculation of an estimation model with higher estimation accuracy. The MAPE results and the principal component coefficients confirmed that the information on tree height led to a decrease in the estimation accuracy of the estimation model. Therefore, it was found that the selection of effective features for estimating the number of leaves according to the characteristics of the target parkland would lead to a better estimation model.

key words Leaf-fruit ratio, LiDAR, point cloud data, principal component regression(PCR), MAPE

目次

| | | |
|--------------|---|-----------|
| 第 1 章 | 序論 | 1 |
| 1.1 | 背景 | 1 |
| 1.2 | 関連研究 | 3 |
| 1.3 | 目的 | 4 |
| 1.4 | 本論文の構成 | 4 |
| 第 2 章 | 関連技術 | 5 |
| 2.1 | LiDAR | 5 |
| 2.2 | SLAM | 5 |
| 2.3 | RANSAC | 6 |
| 2.4 | Cloth Simulation Filter (CSF) | 6 |
| 2.5 | DBSCAN | 8 |
| 2.6 | k-means 法 | 9 |
| 2.7 | watershed 法 | 10 |
| 2.8 | 主成分分析 | 10 |
| 2.8.1 | 累積寄与率 | 10 |
| 2.8.2 | 主成分係数 | 11 |
| 2.9 | 回帰分析 | 11 |
| 第 3 章 | 提案手法 | 12 |
| 3.1 | クラスタリング手法を用いた樹木分離法 | 13 |
| 3.1.1 | 園地点群データの前処理 | 13 |
| 3.1.2 | DBSCAN による樹木分離 | 15 |
| 3.1.3 | k-means 法による樹木分離 | 16 |

目次

| | | |
|--------------|---------------------------------|-----------|
| 3.2 | watershed 法を用いた樹木分離法 | 18 |
| 3.2.1 | 樹高マップの作成 | 19 |
| 3.2.2 | watershed 法の実行 | 20 |
| 3.3 | 樹木ごとの特徴量の算出 | 26 |
| 3.3.1 | 樹高 | 26 |
| 3.3.2 | 楕円の長軸方向・短軸方向の分散 | 27 |
| 3.3.3 | 高さごとの楕円長軸・短軸方向の広がりの分布 | 28 |
| 3.4 | 主成分回帰分析による葉数推定モデルの算出 | 29 |
| 第 4 章 | 検証 | 31 |
| 4.1 | 検証データ | 31 |
| 4.1.1 | 点群データの取得 | 31 |
| 4.1.2 | 葉数実測値の取得 | 32 |
| 4.1.3 | 対象園地・対象樹木 | 33 |
| 4.2 | 検証パターン | 34 |
| 4.2.1 | 特徴量数の検討 | 34 |
| 4.2.2 | 成分数の検討 | 34 |
| 4.2.3 | 推定モデルの精度検証 | 36 |
| 第 5 章 | 結果 | 38 |
| 第 6 章 | 考察 | 40 |
| 6.1 | 推定モデルの MAPE の結果について | 40 |
| 6.2 | 特徴量数の違いについて | 40 |
| 第 7 章 | 結論 | 42 |
| | 謝辞 | 43 |

目次

| | |
|--------------------|----|
| 参考文献 | 44 |
| 付録 A 特徴量と葉数の相関図 | 48 |
| 付録 B 各データにおける主成分係数 | 50 |

目次

| | |
|---------------------------------------|----|
| 2.1 CSF の概要 [20] | 7 |
| 2.2 DBSCAN のイメージ | 9 |
| 3.1 提案手法の流れ | 12 |
| 3.2 生の園地点群データ | 14 |
| 3.3 園地部分を切り出したデータ | 14 |
| 3.4 CSF にて地面を削除したデータ | 14 |
| 3.5 DBSCAN のみを実行した結果 | 15 |
| 3.6 k-means 法のみを実行した結果 | 17 |
| 3.7 DBSCAN と k-means 法を実行した結果 | 17 |
| 3.8 watershed 法の流れ | 18 |
| 3.9 樹高マップの作成のイメージ | 19 |
| 3.10 樹高マップ | 20 |
| 3.11 二値化処理の結果 | 21 |
| 3.12 オープニング処理の結果 | 22 |
| 3.13 膨張処理の結果（黒領域は確実に背景の部分） | 23 |
| 3.14 閾値処理の結果（白領域は確実に物体の部分であり，シード値となる） | 24 |
| 3.15 watershed 法の入力となる各画素の情報 | 25 |
| 3.16 watershed 法の実行結果 | 25 |
| 3.17 樹高算出のイメージ | 26 |
| 3.18 地面平面への射影と楕円の長軸方向・短軸方向の分散の算出のイメージ | 27 |
| 3.19 各層における楕円の標準偏差を算出する | 28 |
| 4.1 LiDAR によるデータ計測の様子 | 32 |

図目次

| | | |
|-----|---------------------|----|
| 4.2 | 葉数実測値データ計測の様子 | 33 |
| 4.3 | すべてのデータ | 35 |
| 4.4 | 北川村データ | 35 |
| 4.5 | 三原村データ | 36 |
| 4.6 | LOOCVによる精度検証のイメージ | 37 |
| 5.1 | MAPE (すべてのデータ) | 39 |
| 5.2 | MAPE (北川村データ) | 39 |
| 5.3 | MAPE (三原村データ) | 39 |
| 6.1 | PCRの主成分係数 (三原村データ) | 41 |
| A.1 | ボクセル数 | 48 |
| A.2 | 樹高 | 48 |
| A.3 | 楕円の長軸方向標準偏差 | 48 |
| A.4 | 楕円の短軸方向標準偏差 | 48 |
| A.5 | 各層の長軸方向標準偏差の歪度 | 49 |
| A.6 | 樹高各層の長軸方向標準偏差の尖度 | 49 |
| A.7 | 各層の短軸方向標準偏差の歪度 | 49 |
| A.8 | 各層の短軸方向標準偏差の尖度 | 49 |
| B.1 | PCRの主成分係数 (すべてのデータ) | 50 |
| B.2 | PCRの主成分係数 (北川村データ) | 51 |
| B.3 | PCRの主成分係数 (三原村データ) | 52 |

表目次

| | | |
|-----|---|----|
| 3.1 | CSF のパラメータ設定値 | 13 |
| 3.2 | 特徴量間の VIF | 30 |
| 4.1 | Target orchards and trees used for verification | 33 |
| 4.2 | 葉数と各特徴量の決定係数 | 34 |

第 1 章

序論

1.1 背景

高知県のユズは国内シェア 52.7%を占めており、高知県の重要な生産品目である。ユズの需要は青果用や加工用など多岐にわたるが、いずれにおいても果皮を利用するため味だけでなく見た目の品質も高いことが求められる [1]。高品質なユズの需要がある一方で、ユズ栽培の課題として労働量の多さが挙げられる。ユズ栽培は稲作や畑作といった他の作物の栽培が難しい山間部で行われることが多く、機械の導入が難しい。また、収穫や剪定など高度な技術を要する機械化が難しい作業も多く、単位面積当たりの必要な労働者数や労働時間が膨大である。しかし、高齢化によって農家数や耕作面積は減少傾向にあり、農林水産省による調査では高知県の農業経営体数は 2020 年の時点で 2005 年の 6 割にまで減少していることが報告されている [2]。そのため、継続的かつ安定的なユズ栽培を行うためには、栽培の効率化による作業量の軽減や作業員数・労働時間の削減が求められる。

柚子に限らず果樹栽培の効率化を実現するにあたって、葉と果実の比率の管理が大きな役割を持つ。葉数が果実の着果数や成長率に影響を及ぼすことはこれまでの研究で明らかになっており [3]、植物の成長や果実の実りに欠かせない光合成は主に葉で行われていることから想像がつく。このことから各樹種ごとに葉果比と呼ばれる栽培に最適な葉と果実の比が調べられており、適切な葉果比を保たないと豊作と不作が交互に訪れる隔年結果が生じてしまう。したがって、安定した生産を行うためには樹木ごとに適切な葉果比率を維持する必要がある。しかし、手作業で果実と葉の数を数えるのは膨大な時間がかかり管理も難しい。

近年では、深層学習を用いた画像認識の技術開発が進んでおり、果実栽培においても果実

1.1 背景

の検出や識別などが盛んに検討されている [4]. 葉についても種類の識別 [5] や病害検出 [6] などの研究が行われているが, 葉数推定の研究の報告は少ない. J.Weyler らは移動ロボットで記録したカメラデータから植物固有の葉のキーポイントを検出することで葉数を推定する方法 [7] を提案しているが, 圃場整備されたテンサイ畑が対象であり, 多少の重なりはあるが葉が隠れることなく映っている. 樹木は圃場の植物に比べ茂りが多く画像に映らない部分が出てくる為, 同様の手法は難しいと考えられる.

そこで著者らは, 3次元情報を持つ LiDAR データに着目した. LiDAR (Light Detection and Ranging) は光を使用して距離を測定するリモートセンシング技術の一種で, レーザー光を照射し対象までの距離を測定, その情報を地図や 3D モデルの作成などに活用する. LiDAR を用いた森林の研究は 1990 年代中ごろから行われており [8], 航空機 LiDAR を用いた樹冠体積および断面積の計測 [9] や地上型 LiDAR を用いた樹木形状の計測 [10] など数多くの研究が報告されている. LiDAR は航空機搭載型や地上設置型など測定方法の種類によって得られるデータ異なり, 樹木を対象にする場合は点群の疎密度から取得できる情報の範囲といった様々な情報に差異が出てくる為 [11], 目的に応じた LiDAR の選択や組み合わせが必要がある.

これまでに著者らは, 点群情報の精度や計測のしやすさなどの観点から地上型の 1 種である歩行型 LiDAR を用いて点群データを取得し, 特徴量を抽出することで葉数を推定する方法の検討を行ってきた [12]. この研究の過程で, 同じ樹種でも園地によって樹体の特徴の傾向が異なることが確認できた. この違いは剪定や摘葉, 誘引といった農家の介入が主な原因であり, この違いにも対応できる推定方法を検討する必要があると考えた. 本論文では, 先行研究からさらに抽出する特徴量を増やし, 樹木の特徴が異なる 2 つの地域の園地データを準備することで, 園地の特徴に応じた葉数推定モデルを算出する手法を提案する.

1.2 関連研究

植物の葉の量を把握する際に、農業研究者や試験場は葉面積指数 (LAI: Leaf Area Index) と呼ばれる指数を用いることが多い。これは単位土地面積当たり存在する葉面積 (片面) の総和であり、1960年代から植物に関する研究で幅広く用いられているパラメータである [13]。LAI の直接的な計測方法は正確である一方非常に時間がかかる為、様々な機材を用いてより効率的に高精度な計測が可能となる方法が検討されている [14]。LiDAR データを用いた LAI の計測は 2004 年から行われており [15]、2024 年には Y.Li らが UAV LiDAR 技術を用いて、山岳森林での LAI の推定における地形要因の影響を調査している [16]。地上でプラントキャノピーアナライザー (PCA) を用いて LAI を計測した後、上空から UAV を用いて LiDAR データを取得している。この時、山岳森林における LAI 推定の精度を高めることを目的として、UAV の飛行軌道を単一および複数の飛行経路の両方を含む 3 つの場合で実験を行っている。取得した LiDAR の離散点群データからベール・ランバートの法則 [17] に基づいて LAI を推定した結果、土地の傾斜角が推定値と測定値の間の推定誤差に影響を及ぼすことが確認された。また、UAV の飛行経路の設計によって点群の品質も決まり、計測経路やスキャン角度・範囲など細かい設定の調査が必要であるとされている。

LAI が果実の収穫量に影響を及ぼすことはこれまでに明らかになっている [13] が、農家がプラントキャノピーアナライザー (PCA) や魚眼カメラといった機材を用いて 1 樹ごとに計測するのは手間がかかる。UAV を用いた LiDAR による推定も 1 回の計測飛行のコストが高いため、定期的に樹体情報を更新する必要がある果実栽培には適していない。また、LAI は単位土地面積当たりの葉面積の指数であるため、どのくらい摘葉する必要があるのかが不明確である。本研究では、歩行型 LiDAR を用いて葉数を推定することで、農業現場で容易かつ継続的に樹体情報を管理できることが期待できる。

1.3 目的

1.3 目的

1.1 節および 1.2 節から本報告では、果樹の葉数推定の自動化を目的として、LiDAR から得られる点群データを用いた樹木推定法について提案し、MAPE を算出することで推定モデルの精度検証を行った。

1.4 本論文の構成

本論文では、はじめに本研究にて扱った関連技術について 2 章で述べる。その後、提案手法として LiDAR データから葉数を推定する方法について 3 章で述べる。提案手法における推定モデルの推定精度の検証のために使用したデータや、検証条件などを 4 章で述べ、検証によって得られた結果を 5 章で、考察を 6 で述べる。最後に 6 章にて本研究における結論について述べる。

第 2 章

関連技術

2.1 LiDAR

LiDAR は Light Detection And Ranging または Laser Imaging Detection And Ranging の略称である。パルス状のレーザー光を照射し、対象物に当たって跳ね返ってくるまでの時間差を計測することで、反射点までの距離を検知する技術のことを指す。LiDAR で用いるレーザー光は主に不可視光線であり、密度が高く波長が短いことから誤差が発生しづらく、遠距離からでも高精度で小さい物体の計測ができるという特徴を持つ。また、360° 観測できる LiDAR は内部に搭載されている反射ミラーを高速で回転させることで 1 秒で約 30 万のポイント情報を取得する [18]。しかし、レーザー光を用いることで広範囲にわたって状況を捉えることができる一方で、天候に左右されやすいことや高価であるという課題もある。

2.2 SLAM

SLAM は Simultaneous Localization and Mapping の略称である。自己位置推定と環境地図作成を同時に行う技術のことを指す。センサー機材から得られた情報を用いてマッピングを行う。SLAM システムは、センサーデータの処理と解釈、自己位置の推定、地図の更新などのステップから構成される。SLAM の種類は入力されるセンサーデータによって異なり、LiDAR の場合は点群データとなる。SLAM の応用は広範で、自律走行車や無人航空機（ドローン）、AR、VR などの分野で使用されている。

2.3 RANSAC

RANSAC はロバスト推定の一環で、直線や平面を推定する場合において外れ値の影響を無視してパラメータを推定する手法である。アルゴリズムは以下の通りである [19].

1. データの集合からパラメータを求める際に必要なサンプルをランダムに選択する.
2. 最小二乗法を用いてサンプルから臨時モデルを推定する.
3. 推定した臨時モデルを用いてすべてのデータの誤差を算出する.
4. 誤差が閾値以下であるデータの個数を数えて保存する.
5. 1. から 4. を繰り返し行い 4. のデータの個数が最も多い場合のデータ集合を保存する.
6. 5. で求めたデータ集合において最小二乗法によりモデルを推定し結果として出力する.

ステップ 1 からステップ 4 までの流れは指定された回数または収束条件が満たされるまで続けられる。また、ステップ 1 で求めるサンプル数は求めたいパラメータ数以上であればよいが、なるべく小さい方が外れ値が紛れ込みにくくなる。例えば、直線の方程式は $y = ax + b$ であることから、サンプル数は 2 以上あればよい。また、ステップ 4 の閾値は外れ値に応じて適宜変える必要がある。

2.4 Cloth Simulation Filter (CSF)

Cloth Simulation Filter (CSF) は 2016 年に W.Zhang らによって提案された LiDAR 点群内の地上点を抽出するツールである [20]。CSF は 3DCG アルゴリズムである布シミュレーションに基づいた手法であり、図 2.1 に示すように反転させた LiDAR 点群データの表面を硬い布で覆い、クロスノードと対応する点群間の相互作用を分析することでクロスノードの位置を特定し、地表の近似値を生成する。その後、地表情報をベースとして地上部分と非地表部分に分類する。

CSF のパラメータは、プログラム実行に必須な一般パラメータとニーズに応じて設定する詳細パラメータがあり、以下の通りである。

2.4 Cloth Simulation Filter (CSF)

- 一般パラメータ

- Scenes

対象の点群データの形状に応じて「Steep slope (急勾配)」、「Relief (起伏)」、「Flat (平坦)」の3つから選択するパラメータ。このパラメータによって布の剛性が決定される。

- Slope post processing for disconnected terrain

対象に急な斜面がある場合に設定するオプションのパラメータ。急な斜面の影響で大きな誤差が生じる場合はこのオプションを選択することで解決できる。急な斜面がない場合は無視する。

- 詳細パラメータ

- Cloth resolution (布の解像度)

地形を覆うために使用される布のグリッドサイズを指す。設定したクロス解像度が大きくなるほどDTMは粗くなる。

- Max iterations (最大反復回数)

地形シミュレーションの最大反復回数を指す。基本的には500で十分である。

- Classification threshold (分類の閾値)

点群とシミュレートされた地形との距離に基づいて、点群を地上部と非地上部に分類するための閾値を指す。基本的には0.5が適応される。

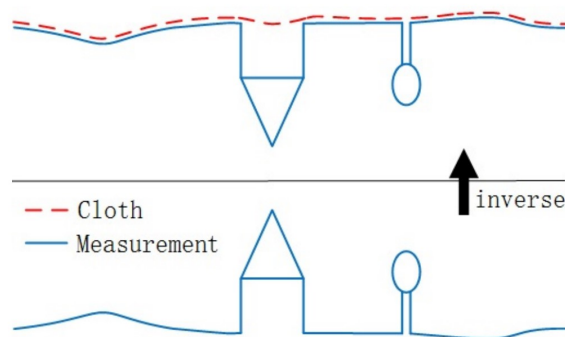


図 2.1: CSF の概要 [20]

2.5 DBSCAN

DBSCAN は Density-Based Spatial Clustering of Applications with Noise の略称である [21]. 半径内に存在する点群数に応じてクラスタが成長する密度準拠クラスタリングアルゴリズムであり, 利点としてクラスタの形状やクラスタの数を事前に指定しなくてよい点, 更にはノイズを除外できる点などが挙げられる. 一方で, 適切なパラメータを選択することが重要であり, データの密度に大きく依存する. アルゴリズムは以下の通りである.

1. ランダムにデータ点を選択する.
2. 選択したデータ点の ϵ -近傍内のデータ点の数を数える. ϵ -近傍は指定された半径 ϵ 内に存在するデータ点の集合を指す. ϵ -近傍内のデータ点の数が閾値未満である場合は, 選択した点をノイズとする.
3. 1. で選択したデータ点が核点であり, その ϵ -近傍内に十分な数のデータ点が含まれる場合はそのデータ点を新しいクラスタの一部としてマークする. このとき, ϵ -近傍内のデータ点も同じクラスタに追加される. ϵ -近傍の範囲外の点はノイズとし, 同じクラスタにはないとみなす.
4. 1. から 3. をすべての点にクラスタが割り振られるまで繰り返す.

DBSCAN の実行時に必要なパラメータは距離の閾値 ϵ と最小のデータ点の閾値の 2 つである. 距離の閾値は選択した点からどこまでを範囲とするかを決定するパラメータで, 最小のデータ点の閾値は範囲内に何個のデータ点があればクラスタとするのかを決定するパラメータである. DBSCAN のイメージを図 2.2 に示す. DBSCAN は画像処理や地理情報システム, 異常検出などのさまざまな分野で利用されている.

2.6 k-means 法

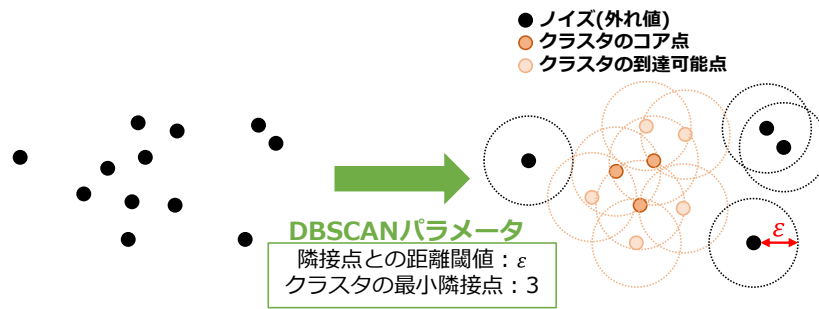


図 2.2: DBSCAN のイメージ

2.6 k-means 法

k-means 法は k 点平均法とも呼ばれ、データセット内のデータ点を事前に指定された数のクラスタ (k 個) にグループ化する非階層的なクラスタリング手法である [22]. 各データ点を最も近い中心点に割り当てることによってクラスタを形成し、中心点を更新することでクラスタリングを繰り返す. アルゴリズムは以下の通りである.

1. 最初にクラスタ数 k を選択し、各データ点をランダムに k 個のクラスタに割り振る.
2. 各クラスタの平均ベクトル (重心) を求める.
3. 各点について k 個のクラスタの重心からの距離を算出し、距離が一番近いクラスタに割り振る.
4. 2. から 3. を重心の更新によるクラスタの変化がなくなるまで繰り返す.

k-means 法は、クラスタの数が事前にわかっている場合やデータが均一な密度で分布している場合に有効である. また、計算効率が高いため大規模なデータセットにも適用できる. ただし、初期の中心点の選び方によって最終的なクラスタリング結果が変わることがあるため、異なる初期化を試行し最も適切なクラスタリング結果を見つけることが重要である. k-means 法はデータマイニング、パターン認識、画像処理などのさまざまな分野で広く利用されている.

2.7 watershed 法

watershed 法は 1979 年に S.Beucher と C.Lantuéjoul によって導入されたセグメンテーション手法の一種である [23]. 画像処理やコンピュータビジョンの分野で使用されるこの手法は, 画像の輝度勾配を山の稜線に見立て, 高い位置から流れ込む水によってできる領域を 1 つのエリアと見なし, 領域に分割する. watershed 法は, 画像中の物体の形状が不規則である場合や物体同士が密接に接している場合に有用である.

2.8 主成分分析

主成分分析は, 複数の変数を少ない変数に合成し要約することで, データを理解しやすくする分析手法の一つである. 主成分分析を一般化したものを次に示す [24].

変数 x, y, z, \dots であるとき, 合成変数 (主成分)

$$u = ax + by + cz + \dots \quad (a, b, c, \dots \text{は } a^2 + b^2 + c^2 + \dots = 1 \text{ を満たす定数})$$

を考える. u の分散 S_u^2 を最大にする係数 a, b, c, \dots を探す.

この u を利用して分析を行う.

新しく一つに合成された変数を主成分と呼び, 最初に求められる主成分から順に第一主成分, 第二主成分 \dots と呼ばれる.

2.8.1 累積寄与率

寄与率は主成分が全データのどれ位の割合を説明しているかを示す量である. 次の式で求めることができる.

$$\text{寄与率 } C = \frac{\text{主成分の分散}}{\text{分散の総和}} \quad (0 \leq C \leq 1) \quad (2.1)$$

累積寄与率は各主成分の寄与率を大きい順に足したもので, それまでの主成分でデータをどれだけ説明しているかを示している.

2.9 回帰分析

2.8.2 主成分係数

主成分係数は主成分を構成する各変数の重みを表しており、主成分は係数が大きい変数を中心に合成されていると言える。

2.9 回帰分析

回帰分析は、説明変数が目的変数に及ぼす影響を調べ、説明変数の各値に対応する目的変数の値の予測モデルを算出する方法である。説明変数が1つの場合は単回帰分析、2つ以上の場合には重回帰分析となる。一般的には直線の回帰分析が用いられ、説明変数と目的変数を回帰直線と呼ばれる式で表現する。目的変数 y と説明変数 x_1, x_2 の場合の回帰方程式を次に示す。

$$\hat{y} = ax_1 + bx_2 + c \quad (a, b, c \text{ は定数}) \quad (2.2)$$

回帰方程式は最小二乗法という技法で求められる。目的変数の実測値 y と予測値 \hat{y} の誤差の二乗の総和を残差平方和と呼び、最小二乗法はこれを最小にするように回帰方程式のパラメータを決定する技法である。

$$\text{残差平方和 } Q_e = (y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + \cdots + (y_n - \hat{y}_n)^2 \quad (2.3)$$

第 3 章

提案手法

提案手法の処理の流れを図 3.1 に示す。提案手法では、園地全体の点群データから各樹木の点群データを分離した後、各樹木の特徴量を抽出し主成分回帰分析（PCR）にて推定モデルを算出する。樹木分離の方法については、DBSCAN と k-means 法の 2 つのクラスタリング手法を用いた方法と watershed 法を用いた方法の 2 通りを検討してきた。本章では、まず 2 通りの樹木分離法を述べた後、特徴量の算出方法及び主成分回帰分析について提案する。

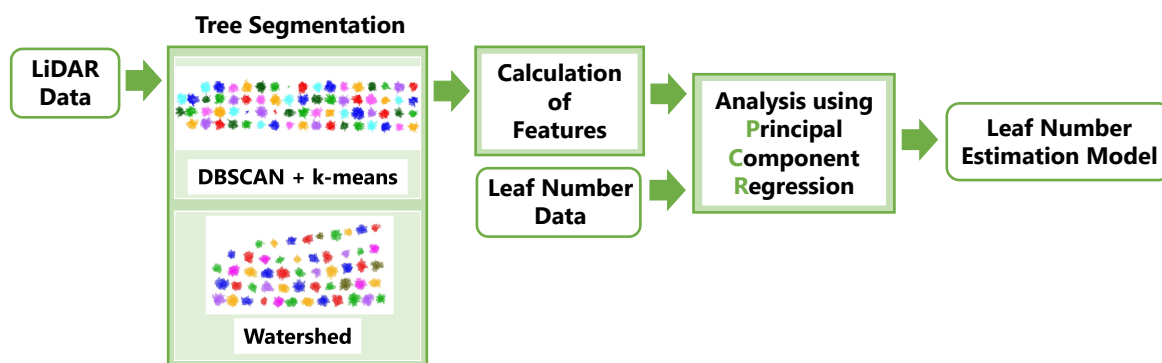


図 3.1: 提案手法の流れ

3.1 クラスタリング手法を用いた樹木分離法

3.1 クラスタリング手法を用いた樹木分離法

点群データの3次元座標情報を活かし、2種類のクラスタリング手法を用いて樹木を分離する。最初に、LiDAR データから園地データを切り出し地面部分を削除した後、DBSCAN と k-means 法を順に実行することで樹木の分離を図る。

3.1.1 園地点群データの前処理

LiDAR から得られる点群の生データから樹木分離法を実行できるデータに処理する。LiDAR から得られる点群データは図 3.2 のように、園地周辺の地形や建物まで入っている。そのため、手作業にて園地データを大まかに切り出した後、地面部分と樹木部分の分離を行う。地面の分離には W.Zhang ら [20] が 2016 年に発表した Cloth Simulation Filter (CSF) と呼ばれるフィルタリング手法を用いる。本論文では園地が中山間地域であることを考慮し、各パラメータを表 3.1 の通りに設定した。手作業にて園地を切り出した結果を図 3.3, CSF 処理後の結果を図 3.4 に示す。

表 3.1: CSF のパラメータ設定値

| Parameters | value |
|--------------------------|--------|
| Scenes | Relief |
| Cloth resolution | 1.0 |
| Max iterations | 500 |
| Classification threshold | 0.5 |

3.1 クラスタリング手法を用いた樹木分離法

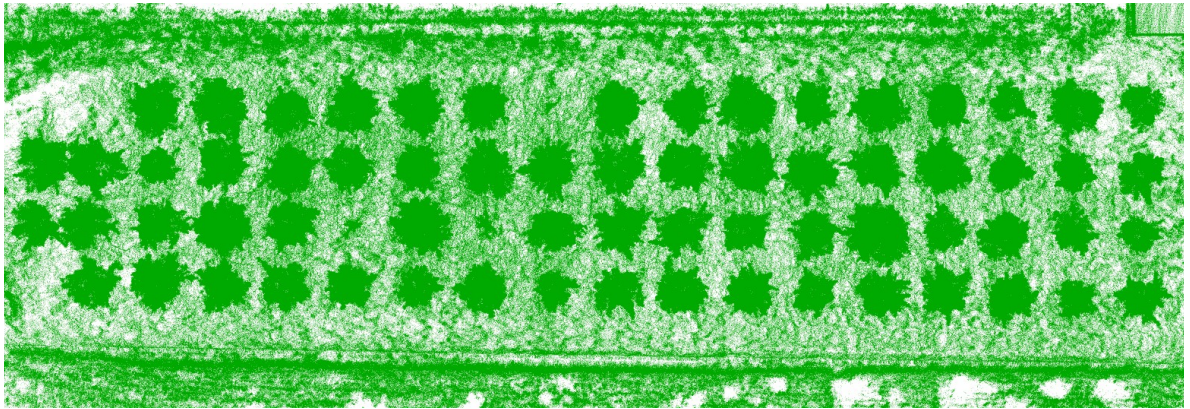


図 3.2: 生の園地点群データ

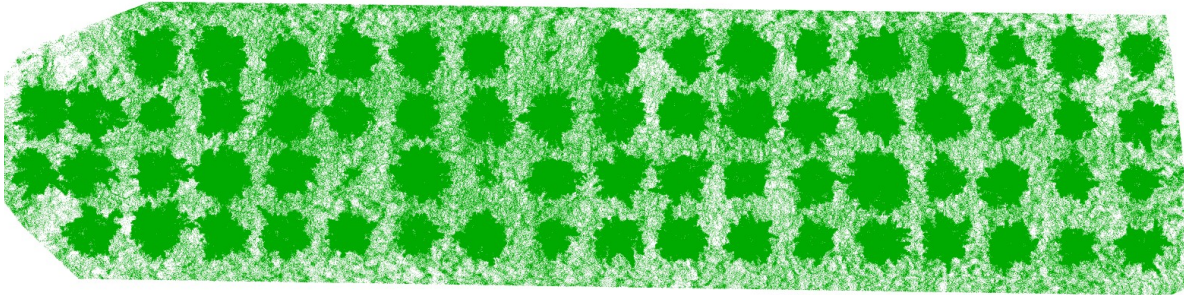


図 3.3: 園地部分を切り出したデータ

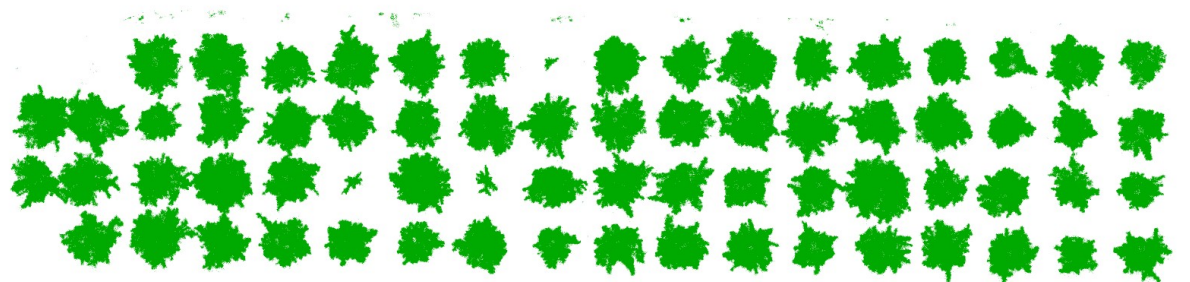


図 3.4: CSF にて地面を削除したデータ

3.1 クラスタリング手法を用いた樹木分離法

3.1.2 DBSCAN による樹木分離

前処理を行った点群データは地面付近のノイズは削除されているが、背丈の高い雑草ノイズや空中ノイズが存在する。その為、ノイズの影響を受けにくい手法を用いる必要がある。本論文では DBSCAN と呼ばれるクラスタリング手法の特徴に着目し本工程に用いる。DBSCAN は M.Ester ら [21] によって提唱された密度準拠クラスタリングアルゴリズムである。任意の半径以内に存在する点群数に応じて領域を判断し、近傍の点群密度が閾値を下回るまでクラスタが成長する。DBSCAN の利点としてクラスタ数を事前に指定しなくてよい点やノイズの影響を受けにくく、ノイズ自体を除外できる点が挙げられる。本研究では園地の樹木同士の間隔が 1m 程度であることを考慮し半径パラメータを 0.2m、最小点数パラメータを 30 と設定した。その際、測定条件による点群密度のばらつきや計算量を削減することを目的として、点群データをボクセル化した。ボクセルサイズは先行研究 [12] で算出した決定係数の比較から、推定モデルの算出に影響が少ない 0.05m のサイズで行った。DBSCAN を実行した結果を Fig.3.5 に示す。DBSCAN によって隣の樹木と離れている樹木はクラスタリングされていることが確認できる一方で、隣の樹木と枝が触れ合っている樹木同士は同じクラスタとしてクラスタリングされていることが確認できる。

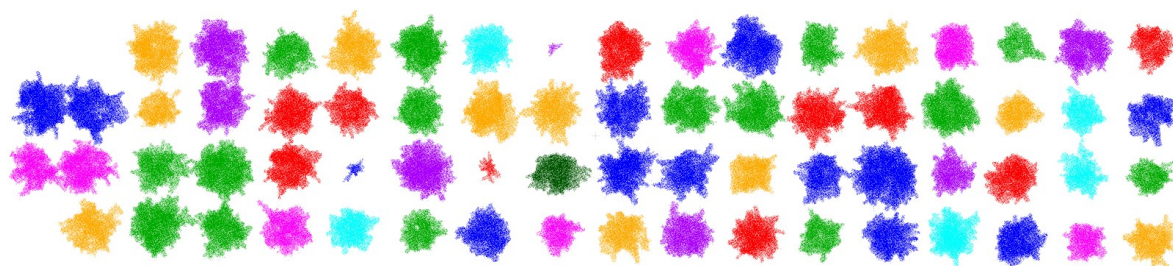


図 3.5: DBSCAN のみを実行した結果

3.1 クラスタリング手法を用いた樹木分離法

3.1.3 k-means 法による樹木分離

DBSCAN では分離できない樹木の対処に k-means 法を用いた。k-means 法は J.B.MacQueen[22] によって提唱された非階層的クラスタリングの代表的手法である。クラスタ数 k を任意に指定し、クラスタの平均を用いてデータを k 個のクラスに分割する [25]。この手法は実装が簡単であらかじめクラスタ数を指定することができる一方で、すべてのデータがクラスタのいずれかに入る為、雑草ノイズなども樹木クラスタに分類される問題がある。ノイズは DBSCAN の段階で除外している為、ここではノイズの影響を受けることなく連結した樹木データを任意のクラスタに割り振ることができる。各クラスタのデータが 2 樹以上の樹木データか否かの判断をする為に地面平行方向の分散に着目した。クラスタの地面平行方向の分散が閾値を超えている場合は複数の樹木が繋がっていると判断し、k-means 法を用いて樹木分離を試みた。閾値は先行研究の際に算出した各樹木の地面平行方向の分散から 1.5 に設定した。k-means 法のクラスタ数は $k=2$ から開始し、クラス分けされたそれぞれのデータの地面平行方向の分散が閾値よりも小さくなるまで k を増やす。このとき、k-means 法の再帰が終了しない場合を考慮し、 $k=20$ になった時点で処理を中止するように設定した。k-means 法のみを実行した結果を Fig.3.2(e)、DBSCAN 後に k-means 法を実行した結果を Fig.3.2(f) に示す。DBSCAN 後に k-means 法を実行することで、DBSCAN では分離できなかった樹木同士や k-means 法では 2 つのクラスタに分割されていた樹木も正しく分離できていることが確認できる。

3.1 クラスタリング手法を用いた樹木分離法

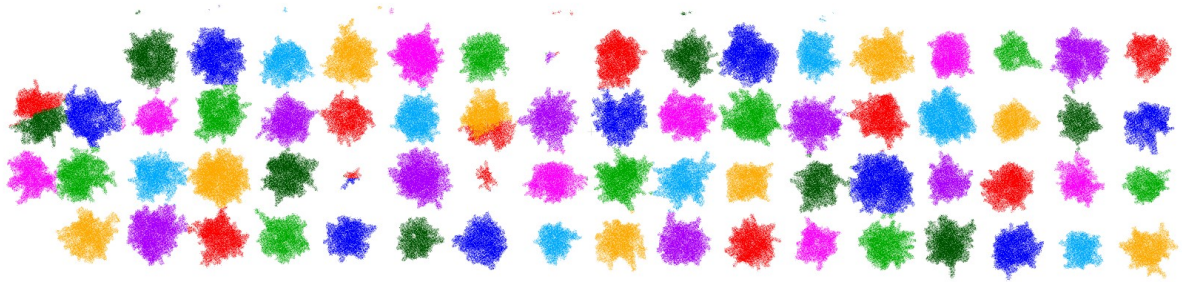


図 3.6: k-means 法のみを実行した結果

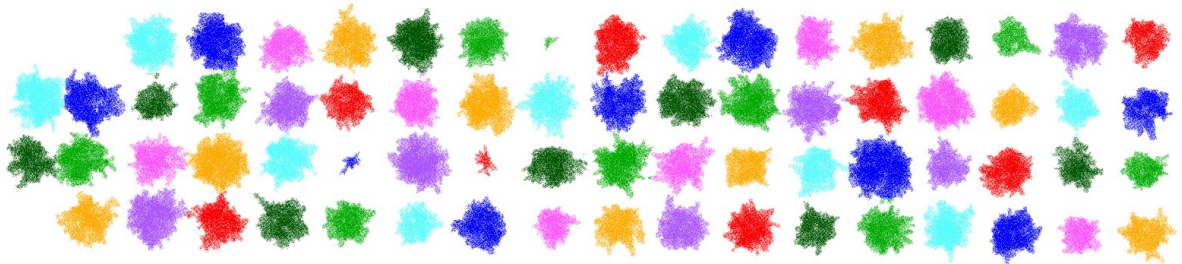


図 3.7: DBSCAN と k-means 法を実行した結果

3.2 watershed 法を用いた樹木分離法

3次元クラスタリング手法を用いて樹木を分離する法は、データセットのサイズに大きく依存し、園地が広ければ広い程計算量が増えて実行時間が長くなる。この問題を解決する方法として、watershed 法を用いた点群データからの樹木セグメンテーション手法が提案されている [26]。J.Yang[27] らは、marker-controlled watershed 法と LiDAR 点群からの 3次元空間分布解析を用いた樹木セグメンテーション手法を提案し、その有用性を示した。

本節では、実行時間を短縮することを目的に watershed 法を用いた樹木分離法を検討した。提案手法の流れを図 3.8 に示す。3次元点群データから 2次元画像を生成し、各ピクセルごとにマーカーを設定することで watershed 法を実行する。

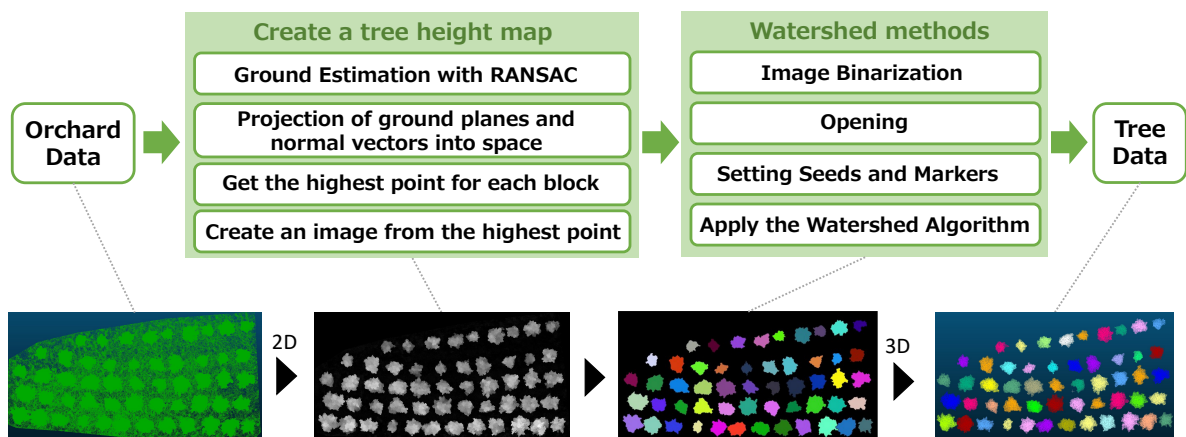


図 3.8: watershed 法の流れ

3.2 watershed 法を用いた樹木分離法

3.2.1 樹高マップの作成

watershed 法を用いるためには、まず 3 次元点群データから 2 次元画像を作成する必要がある。2 次元画像は、3 次元点群データを地面平面とその法線ベクトルからなる空間に投影した後、平面ブロック内の最も高い点を抽出することで樹木の高さを表す画像として生成する。樹高マップの作成のイメージを図 3.9 に示す。

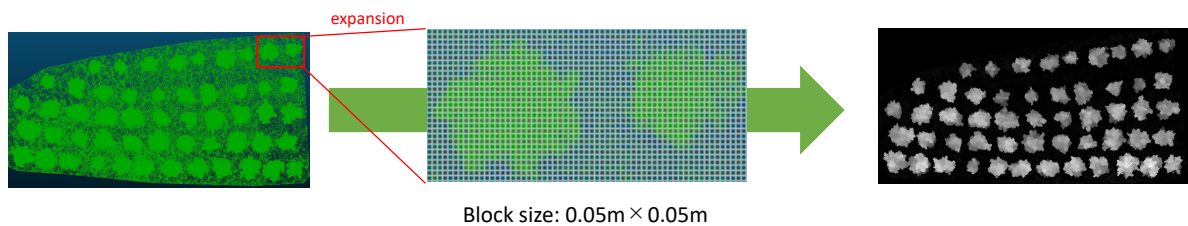


図 3.9: 樹高マップの作成のイメージ

上空から俯瞰した果樹園の画像を得るために、点群データを地上平面とその法線ベクトルを軸とする空間に投影する。本研究では、法線ベクトルから正規直交ベクトル e_1, e_2, e_3 を作成する。まず、直交ベクトル q_1, q_2, q_3 を求める。地面方程式を式 (3.1) とすると、法線ベクトルは式 (3.2) となる。

$$ax + by + cz + d = 0 \quad (3.1)$$

$$\omega = (a, b, c) \quad (3.2)$$

直交ベクトル q_1 は $q_1 = \omega$ として受け付ける。次に、法線ベクトルと一次独立ベクトル関係にあるベクトルを $v_1 = (1, 0, 0)$ とする。この ω と v_1 の一次独立関係から、直交ベクトル q_2 は式 (3.3) で求められる。

$$q_2 = v_1 - \frac{q_1 \cdot v_1}{q_1 \cdot q_1} q_1 \quad (3.3)$$

直交ベクトル q_3 は、直交関係とベクトル q_1 と q_2 の外積を用いて求められる。外積の計算を式 (3.4) に示す。

$$q_3 = q_1 \times q_2 \quad (3.4)$$

3.2 watershed 法を用いた樹木分離法

直交ベクトルは式 (3.5) のようにそれぞれの長さで割ることができ、正規直交ベクトル e_1 , e_2 , e_3 を得ることができる。

$$e_1 = \frac{q_1}{\|q_1\|}, \quad e_2 = \frac{q_2}{\|q_2\|}, \quad e_3 = \frac{q_3}{\|q_3\|} \quad (3.5)$$

最後に、得られた正規直交ベクトル空間上の点群の座標を求める。点 p の場合、正規直交ベクトル空間上の座標は式 (3.6) のようになる。このとき、法線ベクトルは z 軸方向であるため、 z' は直交ベクトル e_1 を用いて計算される。

$$x' = p \cdot e_2, \quad y' = p \cdot e_3, \quad z' = p \cdot e_1 \quad (3.6)$$

上記で算出した x' , y' 平面を任意のブロックサイズに分割した後、各ブロックに含まれる点群から地上から最も高い点のみを抽出する。本研究では $0.05\text{m} \times 0.05\text{m}$ のブロックに分割している。全ブロックの最高点を算出した後、最高点の情報を用いて輝度画像を生成する。このとき、画素値は座標の最小値と最大値で正規化し、各ブロックに 1 画素とする。画像生成結果を図 3.10 に示す。

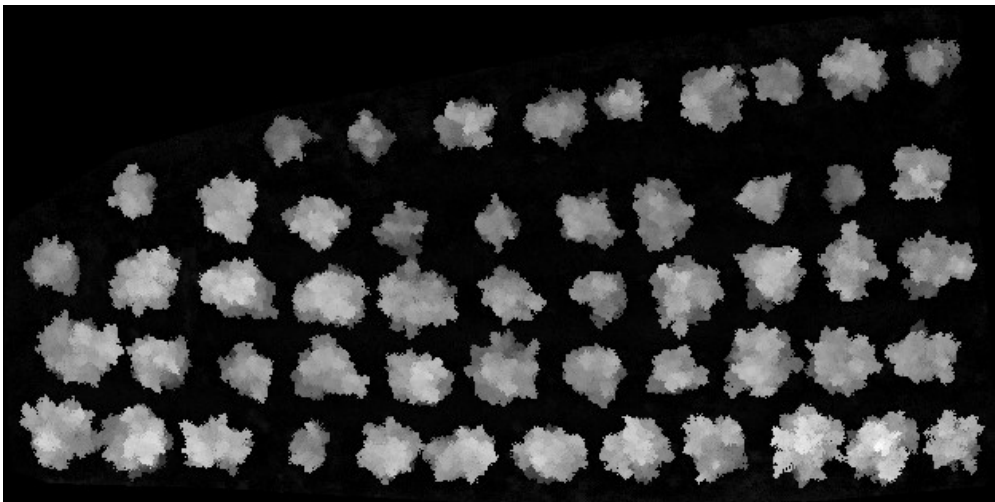


図 3.10: 樹高マップ

3.2.2 watershed 法の実行

本研究では、watershed アルゴリズムのうち J.Yang[27] らが使用していた marker-controlled watershed 法を用いて樹木の分離を試みる。まず、各画素ごとにシード値を含む

3.2 watershed 法を用いた樹木分離法

マーカー情報を設定する。次に、画像の輝度勾配を勾配とし、シードから水が流れてくるようにラベリングして領域を分割する。マーカーには背景であることが確実な領域、物体であることが確実な領域、どの領域に属するかわからない領域の情報が含まれる。

最初のステップはメディアンフィルターを使って画像をぼかすことである。この工程でノイズの影響を軽減する。本研究では、輪郭部分の情報が失われることを考慮し、カーネル3で処理を行う。次に、大津二値化処理により地面と樹木の領域を区分する。2つの処理の結果を図 3.11 に示す。

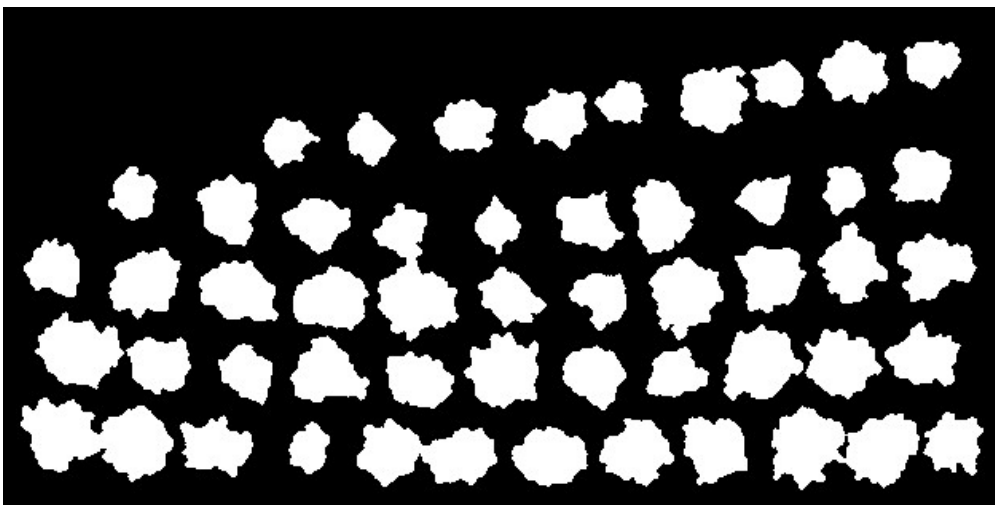


図 3.11: 二値化処理の結果

画像の縮小と拡大を繰り返すオープニング処理を行う。この処理により、背景部分に残るノイズを除去すると同時に、樹木がつながっている部分を分離することができる。本研究ではカーネルを3としてオープニング処理を1回行っている。図 3.12 は処理の実行回数を変更した結果を示す。

3.2 watershed 法を用いた樹木分離法

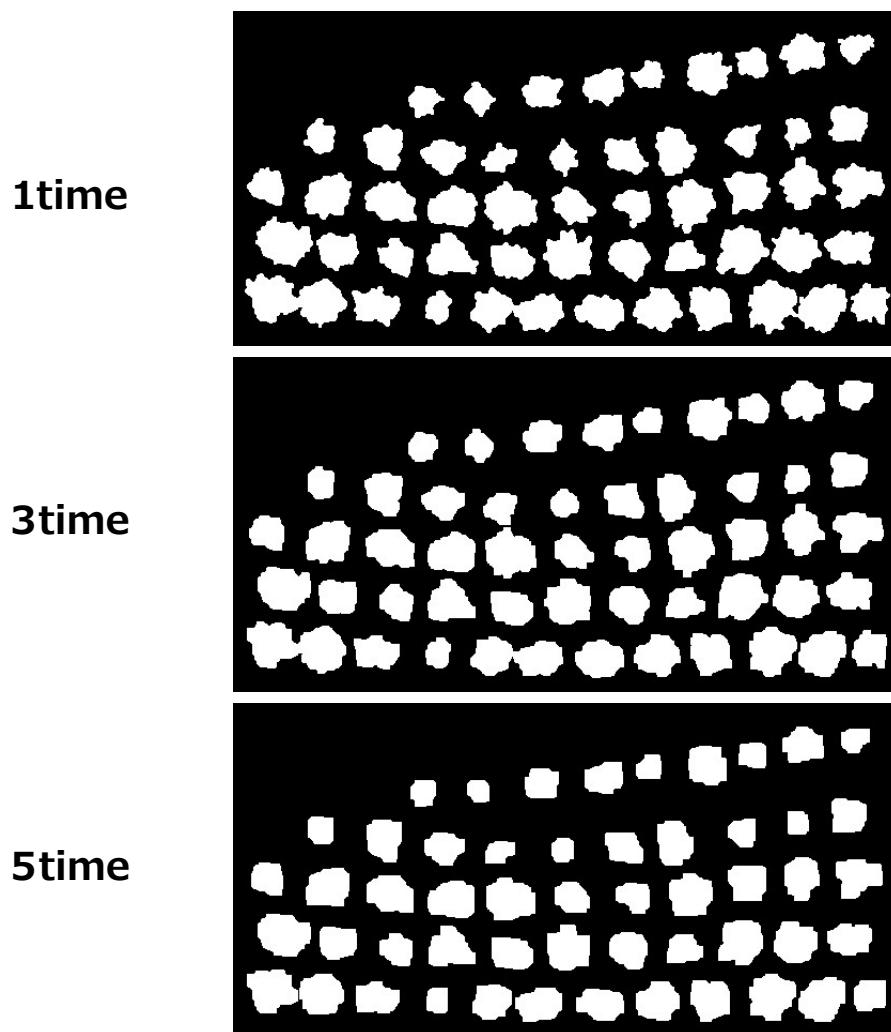


図 3.12: オープニング処理の結果

シードを得るために確実に背景となる部分を検出する。ノイズを除去するために開いただけのデータに拡張処理を施し，確実に背景となる部分を検出する。本研究では，枝部分を背景と判定しないようにカーネル 3 で 3 回拡張処理を行った。図 3.13 は，処理の回数を変更した結果を示す。

3.2 watershed 法を用いた樹木分離法

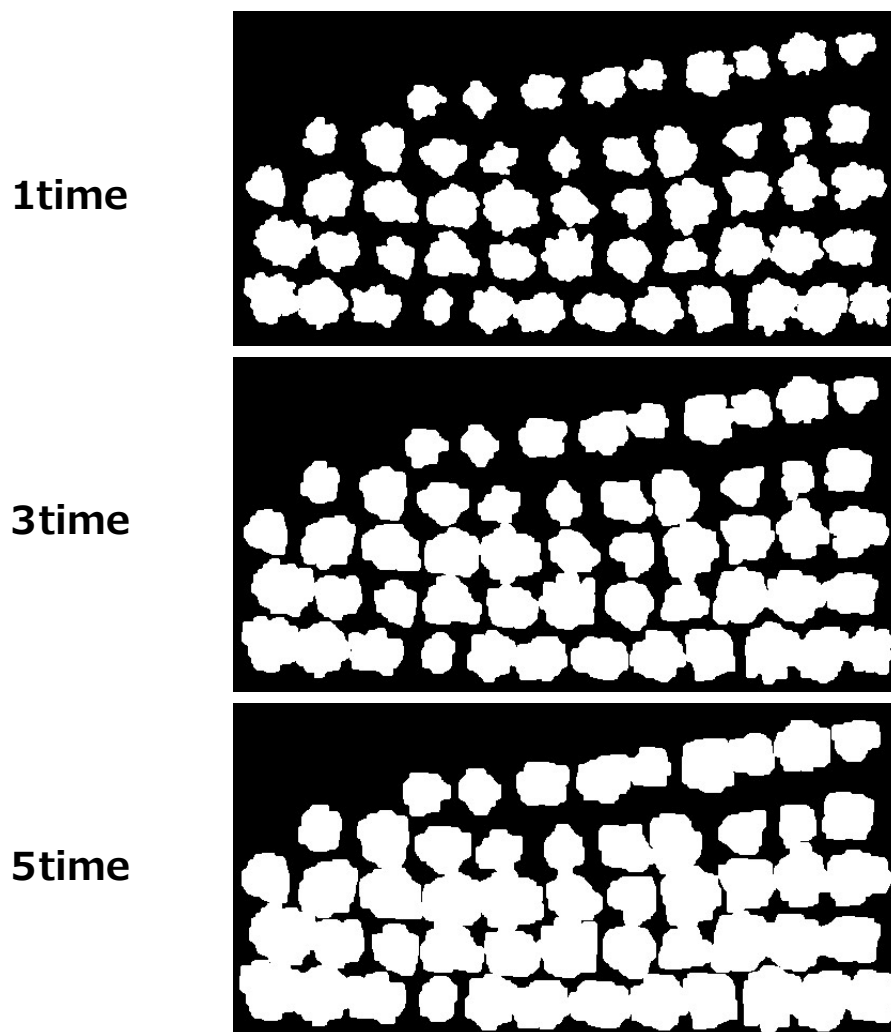


図 3.13: 膨張処理の結果（黒領域は確実に背景の部分）

確実に物体である領域を検出する．領域検出には背景領域からの距離を計算する距離変換を用いる．本研究では，背景領域から最も遠い距離を 100%としたとき，最大値の 30%までの距離範囲を物体領域とみなすように閾値を設定している．閾値を変えて処理を実行した結果を図 3.14 に示す．これが watershed 法を実行するためのシード値となる．

3.2 watershed 法を用いた樹木分離法

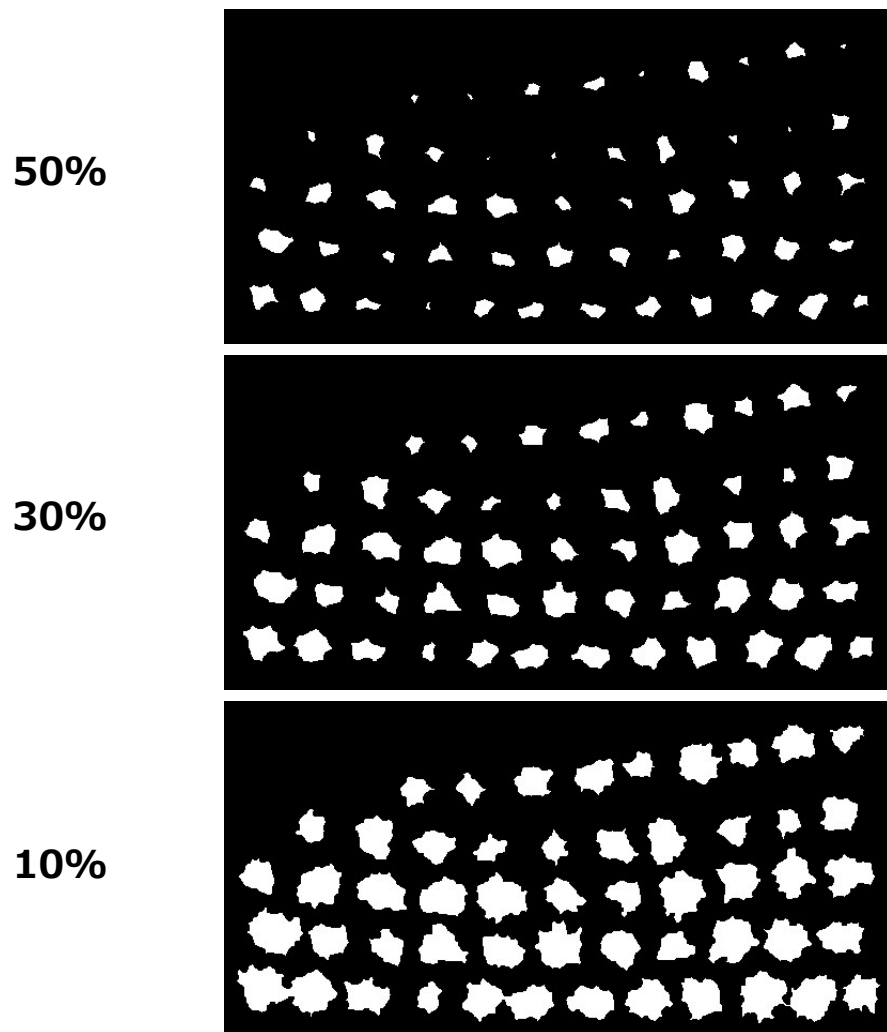


図 3.14: 閾値処理の結果（白領域は確実に物体の部分であり，シード値となる）

これまでに検出された背景領域と物体領域の間の不定領域を検出する．図 3.15 は，背景領域を画素値 0，物体領域を画素値 255，不確かな領域を画素値 128 で示している．

3.2 watershed 法を用いた樹木分離法

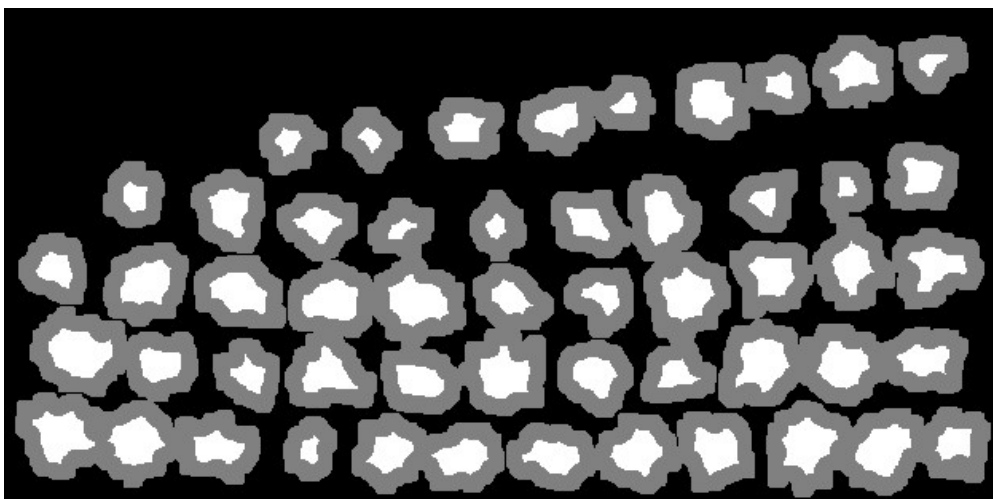


図 3.15: watershed 法の入力となる各画素の情報

マーカーはこれまでに検出された各領域の情報に基づいて作成される。まず、物体か背景であることが確実な領域は正の値でラベルを指定する。本研究では、背景領域のラベルを 1、物体領域のラベルを 2 以降とする。次にどの領域かわからない不定領域についてはラベルを 0 とする。最後に watershed 法を実行する。実行結果を図 3.16 に示す。それぞれの木が分割されていることが確認できる。

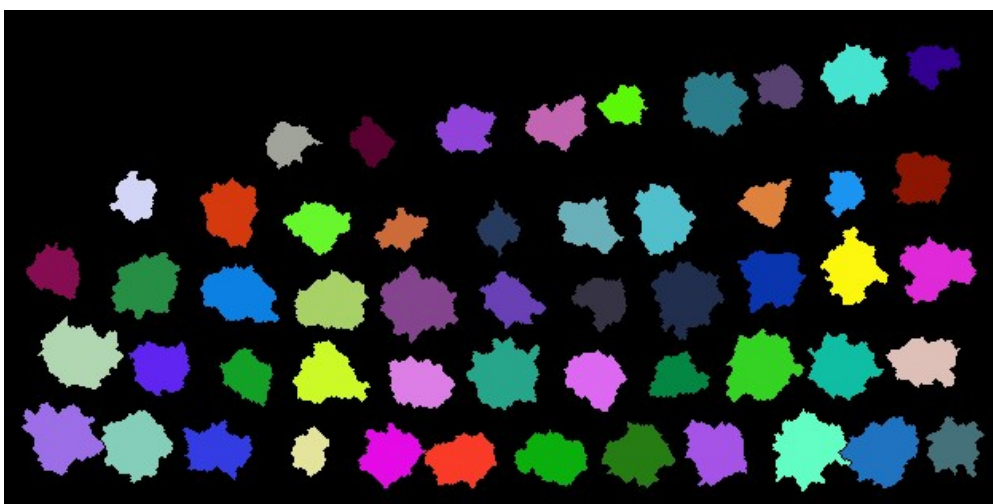


図 3.16: watershed 法の実行結果

3.3 樹木ごとの特徴量の算出

3.3 樹木ごとの特徴量の算出

樹木葉数の計測に最も効果的な特徴量は明らかではないため、樹木データから複数の特徴量を抽出し葉数を推定する多変量解析を用いる。本論文では、樹木データの各点の座標情報を用いて、ボクセル数を含む8つの特徴量を算出する。

3.3.1 樹高

樹木の特徴量として最も代表的な樹高を取得する。はじめに園地データから RANSAC を用いて園地の地面平面の方程式を算出する。平面方程式は次の (3.7) 式で表される。

$$ax + by + cz + d = 0 \quad (3.7)$$

RANSAC で得られた地面平面と各点 (x, y, z) の距離を (3.8) 式にて算出し、最も大きい値を樹高とする。イメージを図 3.17 に示す。

$$L = \frac{|ax + by + cz + d|}{\sqrt{a^2 + b^2 + c^2}} \quad (3.8)$$

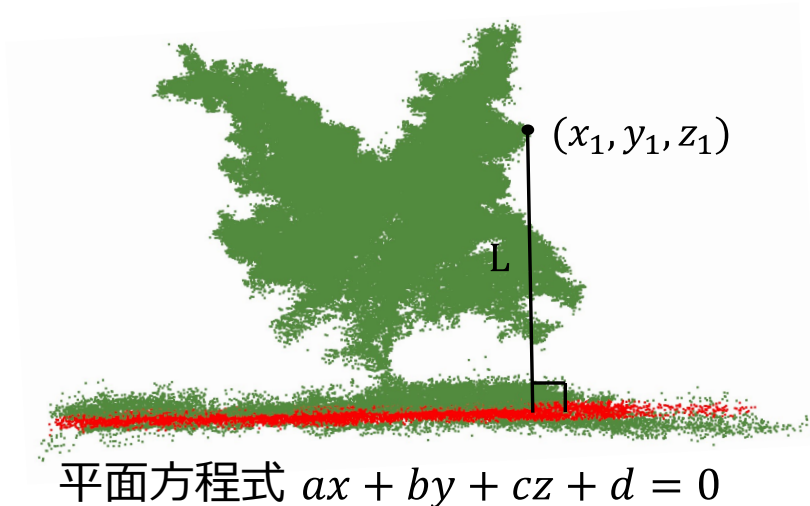


図 3.17: 樹高算出のイメージ

3.3 樹木ごとの特徴量の算出

3.3.2 楕円の長軸方向・短軸方向の分散

樹木の地面平行方向に対する広がりはずしも正円ではない。そのため広がりをも楕円と仮定し、楕円の長軸方向・短軸方向の分散を算出することで、樹木の広がりをも特徴量化する。長軸方向と短軸方向の分散は地面平面における共分散行列 (3.9) から固有値を算出することで求めることができる。この時、 C_{xx} は x の分散、 C_{yy} は y の分散、 C_{xy} は xy の共分散である。

$$\begin{pmatrix} C_{xx} & C_{xy} \\ C_{xy} & C_{yy} \end{pmatrix} \quad (3.9)$$

本論文では、3次元空間において樹木の点群データを地面平面に射影した後、2次元座標を取得する。座標変換の方法は watershed 法の樹高マップの算出の際に使用した方法と同様である。その後、共分散行列を求め、固有値から長軸方向の分散 λ_{max} および短軸方向の分散 λ_{min} を求める。楕円の長軸方向・短軸方向の分散の算出のイメージを Fig.3.18 に示す。本研究では、ばらつきの指標が本来のデータと同じ単位になるよう、標準偏差 $\sqrt{\lambda_{max}}$ 、 $\sqrt{\lambda_{min}}$ を求める。

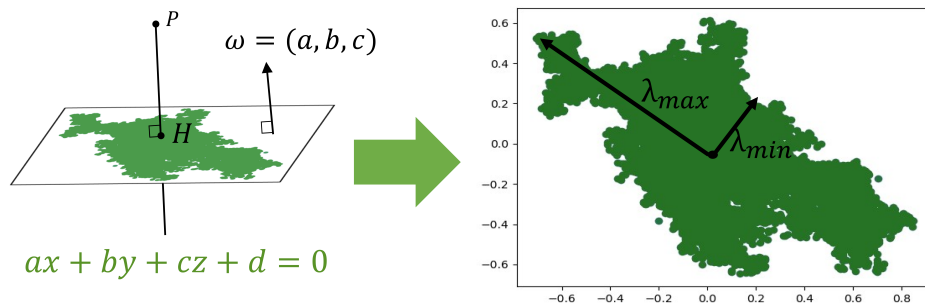


図 3.18: 地面平面への射影と楕円の長軸方向・短軸方向の分散の算出のイメージ

3.3 樹木ごとの特徴量の算出

3.3.3 高さごとの楕円長軸・短軸方向の広がりの分布

これまで算出してきた特徴量は、高さ方向や地面平面方向など1方向のみの情報を用いている。ここでは樹木データを一定の高さごとに分割し、各分割層での地面平面方向の広がり分布を算出することで3次元的な特徴を抽出する。

はじめに、樹木データを地面垂直方向に任意の数に分割する。本論文では樹高が1.5mから3.0mであることから、分割数を60と設定した。樹木データを分割した後、各層での地面平行方向に対する標準偏差を3.3.2の手法で算出する。その後、各層の標準偏差の分布を図3.19のようにまとめる。

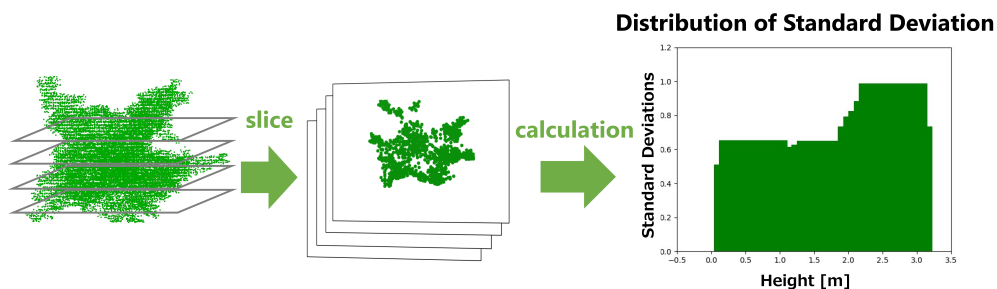


図 3.19: 各層における楕円の標準偏差を算出する

次に、各層の標準偏差の分布から歪度と尖度を用いて特徴を抽出する。平均値 μ は (3.10) 式で求めることができ、この時の x は各層の高さ、 $f(x)$ は相対度数である。また、分散 σ^2 は偏差の2乗の平均であることから、(3.11) 式にて算出することができる。

$$\mu = \sum_x x f(x) \quad (3.10)$$

$$\sigma^2 = \sum_x (x - \mu)^2 f(x) \quad (3.11)$$

3.4 主成分回帰分析による葉数推定モデルの算出

歪度は分布が正規分布からどれだけ歪んでいるかを表す指標であり、正規分布の場合が0、右に歪んでいる場合が負の値、左に歪んでいる場合が正の値になる。歪度は平均値まわりの3次モーメントを σ^3 で割ったものであることから、式(3.12)を用いて算出することができる。

$$skewness = \frac{\sum_x (x - \mu)^3 f(x)}{\sigma^3} \quad (3.12)$$

尖度は分布が正規分布からどれだけ尖っているかを表す指標であり、正規分布の場合が0、尖っていない場合が負の値、尖っている場合が正の値になる。尖度は平均値まわりの4次モーメントを σ^4 で割ったものであることから、式(3.13)を用いて算出することができる。

$$kurtosis = \frac{\sum_x (x - \mu)^4 f(x)}{\sigma^4} \quad (3.13)$$

3.4 主成分回帰分析による葉数推定モデルの算出

各樹木の特徴量を用いて葉数推定モデルを算出する。複数の特徴量を用いた多変量解析を行う際、説明変数間に高い関連性が存在すると解析時の計算が不安定となり、分析精度が悪くなる場合がある。この現象は多重共線性と呼ばれている。本研究で算出した特徴量間に多重共線性が生じているか判定するために分散拡大係数(VIF:Variance Inflation Factor)を算出した。VIFの結果を表3.2に示す。一般的にVIFが10以上あると多重共線性が生じていると判断される。各特徴量のVIFを確認すると、8つの特徴量のうち5つの特徴量でVIFが10を超える結果となった。そのため、本研究では多重共線性による分析精度の低下を防ぐことを目的として、主成分分析を行った後に回帰分析を行う主成分回帰分析(PCR:Principal Component Regression) [28]を行う。このとき、各特徴量はスケールが異なる為、標準化を行う。標準化 z は \bar{x} を特徴量 x の平均、 S_x を特徴量 x の標準偏差とすると、式(3.14)で求めることができる。

$$z = \frac{x - \bar{x}}{S_x} \quad (3.14)$$

また、葉数推定誤差の平均が小さくなるよう、葉数実測値の小さい樹木データの重みが大

3.4 主成分回帰分析による葉数推定モデルの算出

きくなるように重みをつける．重みは葉数 y と平均 \bar{y} から (3.15) 式で算出する．

$$\omega = \frac{\bar{y}}{y} \quad (3.15)$$

表 3.2: 特徴量間の VIF

| | | | | |
|-----|--------|--------|--------|--------|
| | ボクセル数 | 樹高 | 長軸標準偏差 | 短軸標準偏差 |
| VIF | 9.31 | 3.87 | 5.68 | 10.26 |
| | 長軸歪度 | 長軸尖度 | 短軸歪度 | 短軸尖度 |
| VIF | 564.84 | 523.88 | 581.35 | 458.52 |

第 4 章

検証

主成分回帰分析に用いる主成分数は主成分分析の累積寄与率にて判断した。また、葉数推定モデルの推定精度を MAPE にて評価した。本章では、検証データの計測方法や検証を行う際の条件について述べる。

4.1 検証データ

検証には高知県の特徴が異なる 2 つの地域（北川村・三原村）のユズ園地で計測したデータを用いる。各データの計測方法を以下で述べる。

4.1.1 点群データの取得

点群データの計測には歩行型 LiDAR の 1 種である GreenValley 社製 LiBackpack C50 を使用した。LiBackpack C50 は、LiDAR 技術に加え、LiDAR 情報をもとに自己位置推定や環境マッピングを行う SLAM 技術を搭載しており、リアルタイムでの 3D マッピングが可能である。本機材のスキャンレートは 300,000p/s、データ誤差は 5cm 以下、最大スキャン範囲は 100m であり、計測時は背中に背負った状態で園地全体をゆっくりと歩行した。計測の様子を図 4.1 に示す。

4.1 検証データ



図 4.1: LiDAR によるデータ計測の様子

4.1.2 葉数実測値の取得

葉数の計測は数取器（カウンター）を用いて 1 樹ごと手作業で行った。本来であれば計測の人的ミスを防ぐため、1 樹を複数人で計測することが望ましい。しかし、本検証で用いた対象樹木は 1.5m から 3.0m の比較的低いまたは茂りが少ないユズ樹を選択したにも関わらず、木は枝が垂直方向だけでなく水平方向にも広く伸びている為、脚立を使用しながらの慎重な作業となり、1 樹の計測に半日程度を要した。そのため、1 樹 1 人ずつの計測となっている。計測の様子を図 4.2 に示す。

4.1 検証データ



図 4.2: 葉数実測値データ計測の様子

4.1.3 対象園地・対象樹木

計測は北川村，三原村それぞれの地域の複数の園地で行った．対象樹木は各園地からランダムに選択しており，条件を本格的な収穫が始まるとされる 5 年目を目安に葉数 3500 枚以上の樹木とした．その結果，表 4.1 に示す樹木データが集まった．本研究では比較として北川村データのみ，三原データのみ，北川村と三原村の両方のデータの 3 パターンで推定モデルの検証を行った．

表 4.1: Target orchards and trees used for verification

| location | orchards | sample trees | lineage |
|----------|----------|--------------|--|
| kitagawa | 4 | 19 | Kochi Fruit Trial Selection No. 3, 4, 5 |
| mihara | 3 | 15 | Kito Yuzu |

4.2 検証パターン

4.2.1 特徴量数の検討

本研究で算出した各特徴量と葉数実測値の決定係数を表 4.2 に示す。樹高に注目すると、三原村では決定係数が他の特徴量に比べ決定係数が小さいことが確認できる。この原因として、樹木の整枝による低樹高化が考えられる。果実栽培において樹形の管理が収量や品質に影響を及ぼすことはこれまでの研究で明らかになっており、樹種に応じた整枝や剪定の方法が検討されている [29]。カンキツ類の果実栽培においては、低樹高化が栽培管理の効率化や多収・高品質果実生産に繋がるということが報告されており、低樹高化のために垂主枝を水平方向に誘引することが推奨されている [30]。本研究で計測した三原の園地においてもワイヤーを使って垂主枝を水平方向に誘引しているため、葉数に関わらず樹高がおよそ 2m から 2.5m であることが確認できた。本研究では、樹高データが主成分回帰分析に及ぼす影響を考慮し、特徴量に樹高を含む場合と含まない場合の 2 パターンで検証を行う。

表 4.2: 葉数と各特徴量の決定係数

| | ボクセル | 樹高 | 長軸標準 | 短軸標準 | 長軸歪度 | 長軸尖度 | 短軸歪度 | 短軸尖度 |
|----|------|------|------|------|------|------|------|------|
| 全て | 0.74 | 0.29 | 0.56 | 0.65 | 0.57 | 0.55 | 0.66 | 0.63 |
| 北川 | 0.75 | 0.60 | 0.81 | 0.74 | 0.68 | 0.66 | 0.69 | 0.66 |
| 三原 | 0.87 | 0.05 | 0.73 | 0.72 | 0.50 | 0.48 | 0.73 | 0.71 |

4.2.2 成分数の検討

PCR の累積寄与率を図 4.3(a) から図 4.5(a) に示す。北川村のデータ以外での第 1 寄与率に注目すると、0.8 に届いていない場合があり、第 1 成分だけで全ての特徴量を十分に表現できると判断できない。そのため、推定モデルの構築に十分であると考えられる成分数をグリッドサーチで探索し、評価関数を用いて算出する主成分数を検討する。本研究は自然物が対象のため外れ値があることを考慮し、評価関数には誤差を平均的に評価できる平均絶対

4.2 検証パターン

誤差（MAE:Mean Absolute Error）を算出する。MAE は予測値と実測値の誤差の絶対値の平均値であり、この値が小さい程推定モデルに適した成分数であると言える。各成分数における推定モデルの MAE を図 4.3(b) から図 4.5(b) に示す。全てのデータの場合は 4 成分や 5 成分、北川村のデータの場合は 2 成分や 3 成分、三原村のデータでは 3 成分や 4 成分で MAE が小さくなっていることが確認できる。この結果から、本研究では各データごとに 5 成分までを算出し、主成分の個数が異なる推定モデルで精度比較を行う。

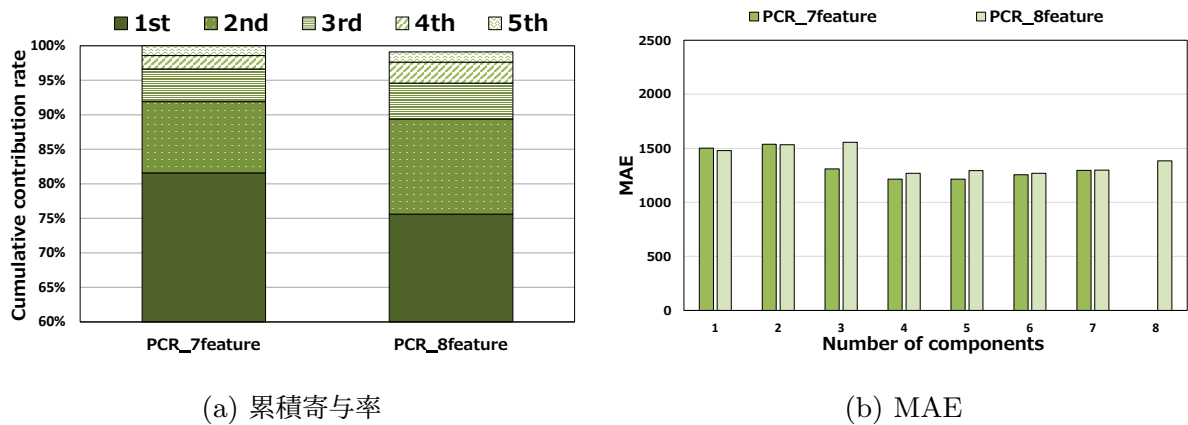


図 4.3: すべてのデータ

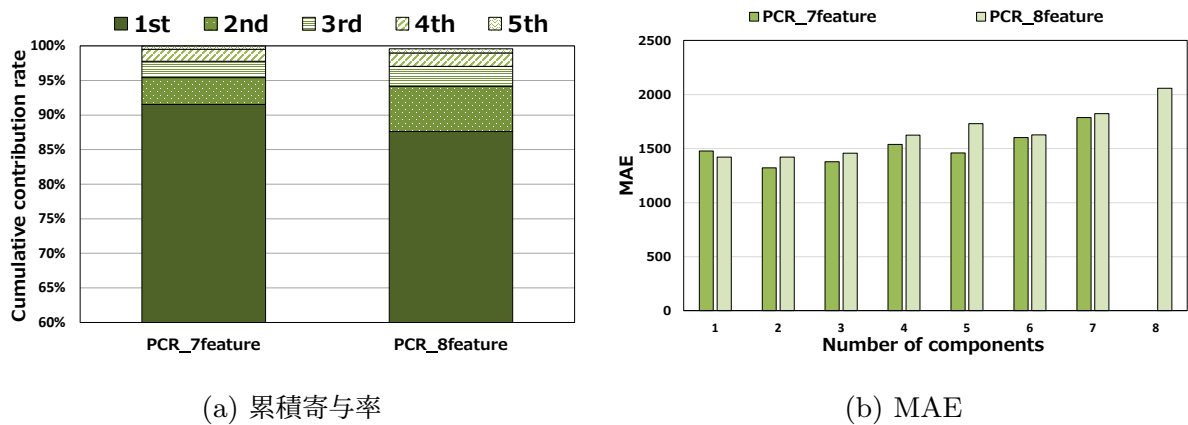


図 4.4: 北川村データ

4.2 検証パターン

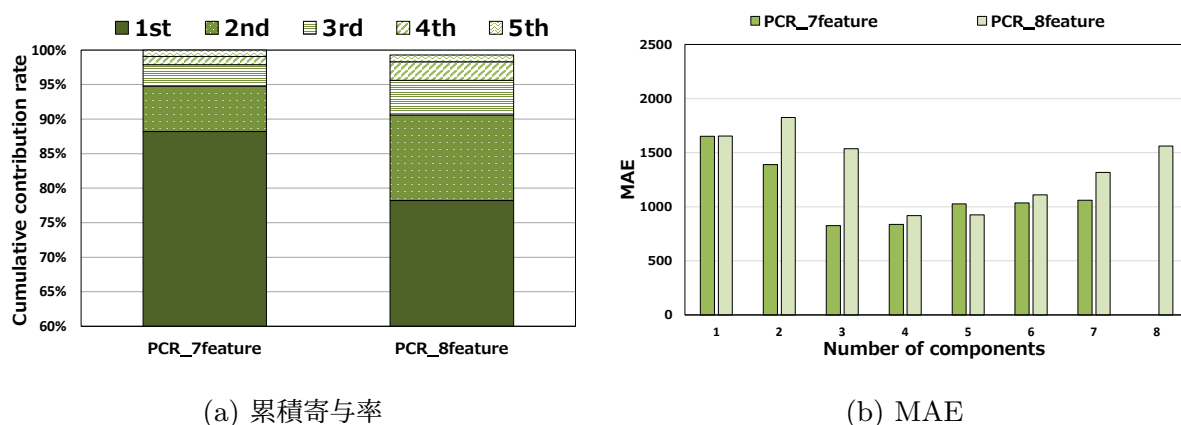


図 4.5: 三原村データ

4.2.3 推定モデルの精度検証

推定モデルの性能を正確に評価するために検証データを学習データとテストデータに分ける必要がある。しかし、本研究で検証に用いる樹木数は 19 樹および 15 樹であり、説明変数の数に対して十分とは言えない。そのため、一つ抜き交差検証 (LOOCV:leave-one-out cross-validation) によるモデル検証を行う。LOOCV は、1 個のデータのみをテスト用に利用し、残りを全て学習データとして利用する交差検証であり、小さなデータセットでも効果的なモデル構築が可能となる。また、推定モデルの評価には誤差率で評価を行う平均絶対パーセント誤差 (MAPE:Mean Absolute Percentage Error) を用いる。MAPE は、予測値と実測値との差を正解値で割った値の絶対値の平均値であり、モデルの推定値が正解値と何%ずれているかを算出する。MAPE は式 (4.1) で算出することができる。本研究では、LOOCV で葉数推定モデルを算出しテストデータの葉数を推定した後、実測値との MAPE を計算し、すべての交差検証の平均を算出する。イメージを図 4.6 に示す。

$$MAPE = \frac{100}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| \quad (4.1)$$

4.2 検証パターン

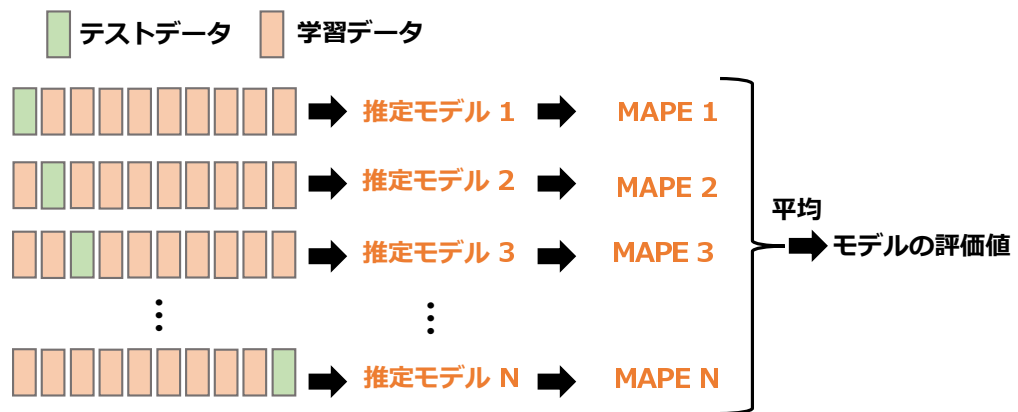


図 4.6: LOOCV による精度検証のイメージ

第 5 章

結果

各検証パターンの推定モデルの MAPE の比較を図 5.1 から図 5.3 に示す。比較用として算出した重み付き重回帰分析 (WLSR) では、すべてのデータの場合の 7 特徴量が 14.2%、8 特徴量が 15.9%、北川村のデータの場合は 7 特徴量が 19.2%、8 特徴量が 22.9%、三原村のデータの場合は 7 特徴量が 15.3%、8 特徴量が 20.8%となった。一方 PCR では、すべてのデータの場合は成分数を増やすごとに MAPE が減少傾向にあり、7 特徴量による成分数 4 つの MAPE が 13.2%と最も小さくなった。北川村のデータの場合は 7 特徴量による成分数 2 つの MAPE が 13.4%と最も小さくなり、成分数 3 つ以降は MAPE が増加傾向になった。三原村のデータの場合は 7 特徴量では減少傾向にあり、7 特徴量による成分数 3 つの MAPE が 9.5%と最も小さくなった。一方で、8 特徴量では成分数 2 つで一度 MAPE が増加し、その後は減少傾向になった。

また、北川村、三原村のどちらのデータにおいても樹高を特徴量として含まない場合の方が MAPE が小さくなった。このことから、樹高を除く 7 特徴量での主成分回帰分析がよりよい推定モデルを算出することができると言える。

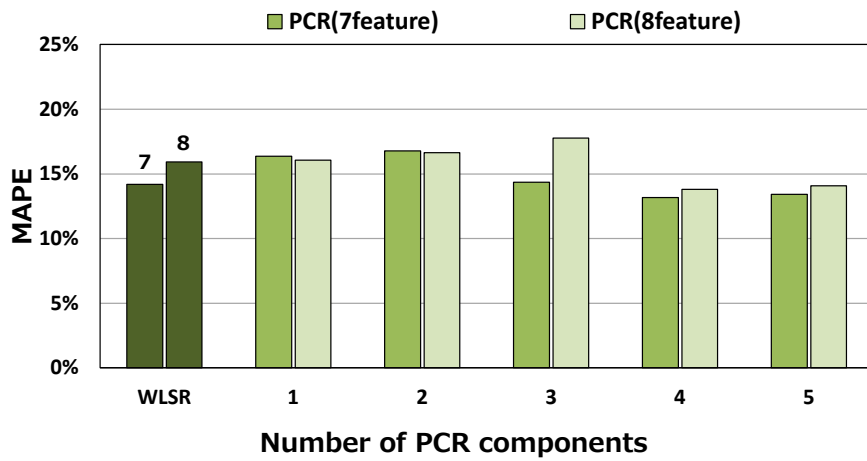


図 5.1: MAPE (すべてのデータ)

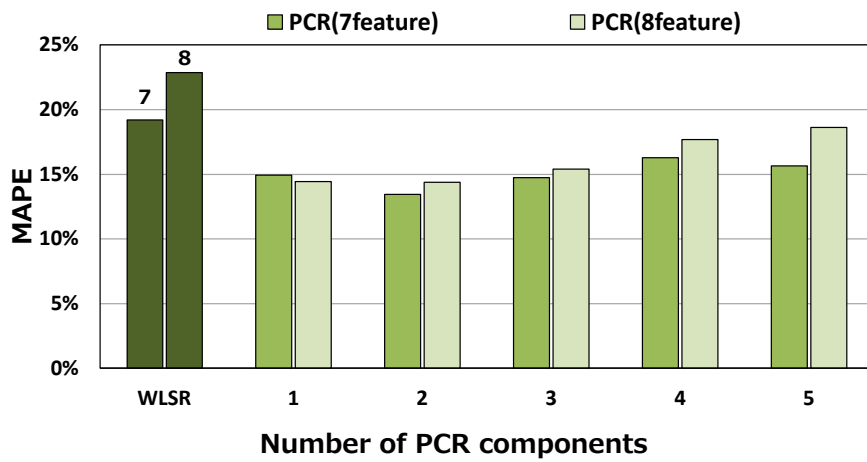


図 5.2: MAPE (北川村データ)

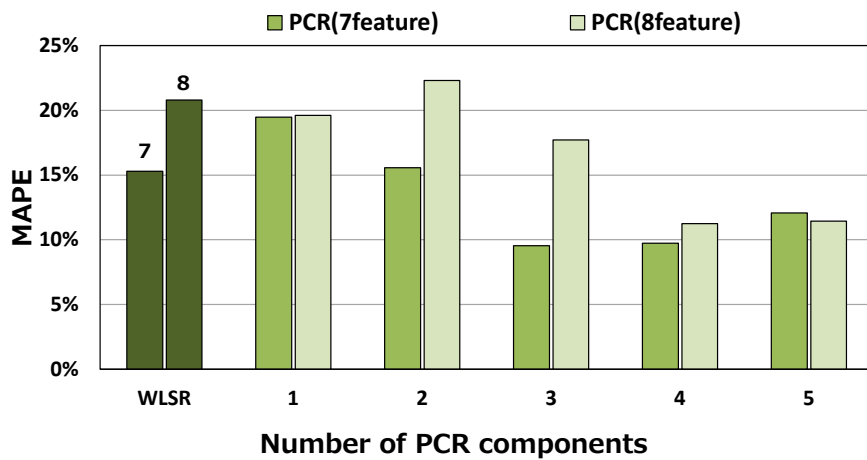


図 5.3: MAPE (三原村データ)

第 6 章

考察

6.1 推定モデルの MAPE の結果について

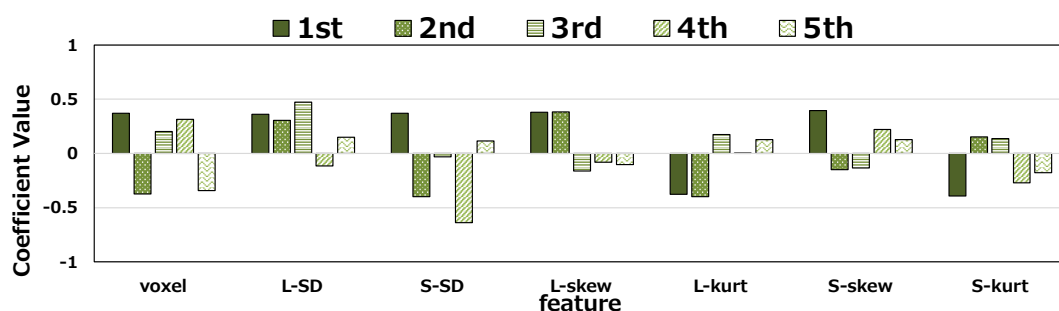
重回帰分析に比べて PCR を行うことによって MAPE が小さくなることが明らかとなった。PCR で MAPE が小さくなる要因として主成分分析による多重共線性および過学習の軽減が挙げられる。本研究で使用する特徴量は、表 3.2 に示した通り各特徴量間に強い相関があることから、重回帰分析では多重共線性が存在し推定モデルの予測精度が低下していると予想できる。一方、PCR では主成分分析によって得られる主成分を説明変数にすることで多重共線性を軽減することができ、更には次元が削減されることによって過学習のリスクも軽減することに繋がったと考えられる。

6.2 特徴量数の違いについて

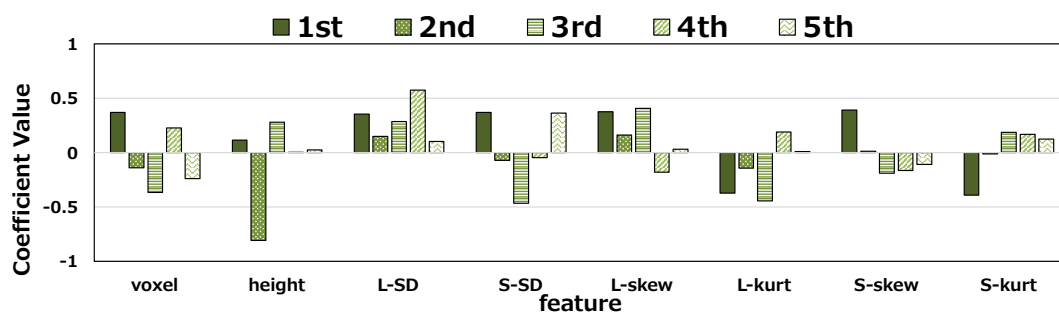
どの検証パターンにおいても 8 特徴量よりも 7 特徴量の方が MAPE が小さくなったことから、特徴量の取舍選択の重要性が分かる。図 5.3 に注目すると、三原村データに対する 8 特徴量の PCR において主成分を 1 つから 2 つに増やすと MAPE が増大していることが確認できる。この原因を主成分分析の主成分係数から考える。三原村データの主成分分析によって算出された主成分係数を図 6.1 に示す。図 6.1(b) に示す 8 特徴量による主成分分析の主成分係数を確認すると、第 1 主成分係数では樹高の係数が他の特徴量に比べて小さく、ほとんど使われていないことが分かる。そのため、7 特徴量と 8 特徴量でそれぞれ第 1 主成分

6.2 特徴量数の違いについて

のみを用いた推定モデルの MAPE に差がないことが Fig.5.3 から確認できる。一方で、第 2 主成分係数を確認すると樹高の係数が一番大きくなっている。葉数実測値との相関が低い樹高データが多く使われている第 2 主成分を回帰分析に用いたことで、推定モデルの精度が低下したと考えられる。このことから、推定モデルを算出する際には対象の園地の特徴に応じて特徴量を取捨選択する必要があることが分かる。



(a) 7 特徴量



(b) 8 特徴量

図 6.1: PCR の主成分係数 (三原村データ)

第7章

結論

本研究では、果実の安定生産に必要な情報である樹木の葉数を自動で推定することを目的として、LiDAR から得られる点群データを用いた葉数推定法について提案した。園地の点群データから各樹木データをクラスタリング手法や watershed 法を用いて分離させた後、樹木ごとに特徴量を抽出し、主成分回帰分析にて推定モデルを算出した。サンプル数が少ないことから、検証は一つ抜き交差検証にて行い、評価には MAPE を用いた。結果として、最も良い MAPE は2つの園地でそれぞれ 13.4%、9.5%となった。特徴量をそのまますべて使用した重み付き重回帰分析の MAPE よりも値が小さくなったことから、主成分の形成が多重共線性および次元の軽減に繋がり、より推定精度が高い推定モデルの算出が可能になったと考えられる。また、MAPE の結果と主成分係数から樹高の情報が推定モデルの推定精度の低下に繋がっていることが確認できた。このことから、対象の園地の特徴に応じて葉数推定に効果的な特徴量を取捨選択することがよりよい推定モデルに繋がることが分かった。

謝辞

本研究は、内閣府地方大学・地域産業創生交付金「IoP(Internet of Plants)が導く、「Society 5.0 型農業」への進化」及び「IoP(Internet of Plants)が導く、「Society 5.0 型農業」への進化」の助成を受けたものです。

また、本研究を進めるにあたり、ご指導いただきました栗原徹教授に心から感謝を申し上げます。そしてお忙しい中、副査を引き受けていただきました福本昌弘教授、松崎公紀教授にお礼申し上げます。

また、検証データの取得に当たり機材レンタルおよび測定のご協力を頂きました高知大学 浜田和俊准教授に感謝申し上げます。

参考文献

- [1] 杉本, 谷本, 矢野, 田中: ユズの系統選抜に関する研究 第3報 短刺系統の特性, 高知県農業技術センター研究報告 = Bulletin of the Kochi Agricultural Research Center, **29**, 39/44 (2020).
- [2] 高知県農業振興部: 高知県農業の動向, https://www.pref.kochi.lg.jp/soshiki/162201/files/2023042800191/file_20234285161830_1.pdf (2024年1月18日閲覧)
- [3] 清水, 鳥潟, 鳥居: 温州ミカンの着果負担に関する研究 (第3報), 園芸学会雑誌, **43-4**, 423/429 (1975).
- [4] G.Xuan, C.Gao, Y.Shao, M.Zhang, Y.Wang, J.Zhong, Q.Li and H.Peng: Apple Detection in Natural Environment Using Deep Learning Algorithms, IEEE Access, **8**, 216772/21678 (2020).
- [5] N.Goyal, K.Gupta and N.Kumar: Clustering-Based Hierarchical Framework for Multiclass Classification of Leaf Images, IEEE Transactions on Industry Applications, **58-3**, 4076/4085 (2022).
- [6] Z.Xiao, Y.Shi, G.Zhu, J.Xiong and J.Wu: Leaf Disease Detection Based on Lightweight Deep Residual Network and Attention Mechanism, IEEE Access, **11**, 2169/3536 (2023).
- [7] J.Weyler, A.Milioto, T.Falck, J.Behley and C.Stachniss: Joint Plant Instance Detection and Leaf Count Estimation for In-Field Plant Phenotyping, IEEE Robotics and Automation Letters, **6-2**, 3599/3606 (2021).
- [8] E.Næsset: Determination of mean tree height of forest stands using airborne laser scanner data, ISPRS Journal of Photogrammetry and Remote Sensing, **52-2**, 49/56 (1997).

参考文献

- [9] K.Omasa, F.Hosoi, T.M.Uenishi, Y.Shimizu and Y.Akiyama : Three-Dimensional Modeling of an Urban Park and Trees by Combined Airborne and Portable On-Ground Scanning LIDAR Remote Sensing, *Environmental Modeling & Assessment*, **13-4**, 473/481 (2008) .
- [10] M.Holopainen, V.Kankare, M.Vastaranta, X.Liang, Y.Lin, M.Vaaaja, X.Yu, J.Hyyppä, H.Hyyppä, H.Kaartinen, A.Kukko, T.Tanhuanpää and P.Alho : Tree mapping using airborne, terrestrial and mobile laser scanning – A case study in a heterogeneous urban forest, *Urban Forestry & Urban Greening*, **12-4**, 546/553 (2013) .
- [11] F.Hosoi, Y.Nakai, K.Omasa : Estimation and Error Analysis of Woody Canopy Leaf Area Density Profiles Using 3-D Airborne and Ground-Based Scanning Lidar Remote-Sensing Techniques, *IEEE Transactions on Geoscience and Remote Sensing*, **48-5**, 2215/2223 (2010) .
- [12] 末廣, 栗原, 浜田 : 歩行型 LiDAR による園地点群データからの樹木分離法の検討と葉数推定, 第 39 回センシングフォーラム講演論文集, 89/92 (2022) .
- [13] 橘, 中井 : 異なった栽植密度におけるワセウンシュウの樹冠密度と収量及び葉面積指数との関係, *園芸学会雑誌*, **58-1**, 91/96 (1989) .
- [14] 石井, 梨本, 下垣 : 衛星データによる葉面積指数 LAI の推定, *水分・水資源学会誌*, **12-3**, 210/220 (1999) .
- [15] J.L.R.Jensen, K.S.Humes, A.T.Hudak, L.A.Vierling, E.Delmelle : Evaluation of the MODIS LAI product using independent lidar-derived LAI:A case study in mixed conifer forest, *Remote Sensing of Environment*, **115-12**, 3625/3639 (2011) .
- [16] Y.Li, H.Zeng, J.Xiong, G.Miao : Influence of Topography on UAV LiDAR-Based LAI Estimation in Subtropical Mountainous Secondary Broadleaf Forests, *forests*, **15-1**, 17 (2024) .
- [17] J.J.Richardson, L.M.Moskal, S-H.Kim : Modeling approaches to estimate effective

参考文献

- leaf area index from aerial discrete-return LIDAR, *Agricultural and Forest Meteorology*, **149**-6-7-15, 1152/1160 (2009) .
- [18] Velodyne Lidar : VLP-16/VLP-16-LITE/VLP-16-HiRes 【Puck】 , <https://www.argocorp.com/cam/special/Velodyne/common/pdf/VLP-16.pdf> (2024年2月18日閲覧)
- [19] デジタル画像処理編集委員会：デジタル画像処理 [改訂新版], 公益財団法人画像情報教育振興協会, (2015).
- [20] W.Zhang, J.Qi, P.Wan, H.Wang, D.Xie, X.Wang and G.Yan : An Easy-to-Use Airborne LiDAR Data Filtering Method Based on Cloth Simulation, *Remote Sensing*, **8**-6, 501 (2016) .
- [21] M.Ester, H.Kriegel, J.Sander, X.Xu : A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise, *Proceedings of the Second International Conference on Knowledge Discovery and Data Miningg (KDD'96)*, 226/231 (1996) .
- [22] J.B.MacQueen : Some Methods for Classification and Analysis of Multivariate Observations, *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, 281-297 (1967) .
- [23] S.Beucher, C.Lantuéjoul : Use of Watershed in Contour Detection, *International Workshop on Image Processing, Real-Time Edge and Motion Detection Estimation France, Sep.* (1979) .
- [24] 涌井良幸, 涌井貞美：まなびのずかん統計学の図鑑, 株式会社技術評論社, (2015).
- [25] 宮本定明：クラスター分析入門, 森北出版株式会社, (1999).
- [26] Q.Chen, D.Baldocchi, P.Gong, M.Kelly : Isolating individual trees in a savanna woodland using small footprint LiDAR data, *Photogrammetric Engineering and Remote Sensing*, **72**-8, 923/932 (2006) .
- [27] J.Yang, Z.Kang, S.Cheng, Z.Yang, P.Akwensi : An Individual Tree Segmentation

参考文献

- Method Based on Watershed Algorithm and Three-Dimensional Spatial Distribution Analysis From Airborne LiDAR Point Clouds, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **13**, 1055/1067 (2020) .
- [28] W.F.Massy : Principal Components Regression in Exploratory Statistical Research, *Journal of the American Statistical Association*, **60**, 234/266 (1965) .
- [29] 倉橋孝夫 : リンゴの Y 字形棚整枝における物質生産と栽培法に関する研究, 島根県農業試験場研究報告 = *Bulletin of the Shimane Agricultural Experiment Station*, **31**, 1/59 (1997) .
- [30] 富田栄一 : ミカンの低樹高整枝, *農業および園芸* = *Agriculture and horticulture*, **84-9**, 902/914 (2009) .

付録 A

特徴量と葉数の相関図

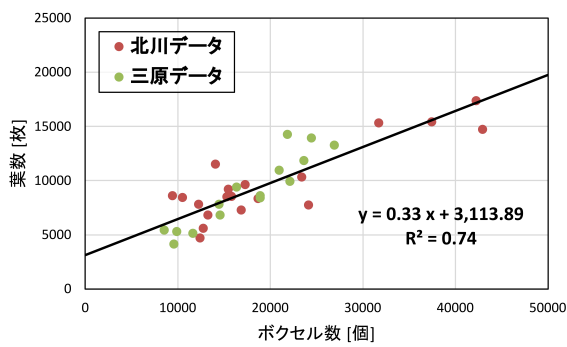


図 A.1: ボクセル数

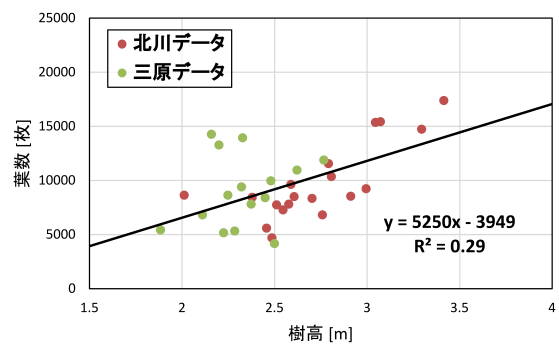


図 A.2: 樹高

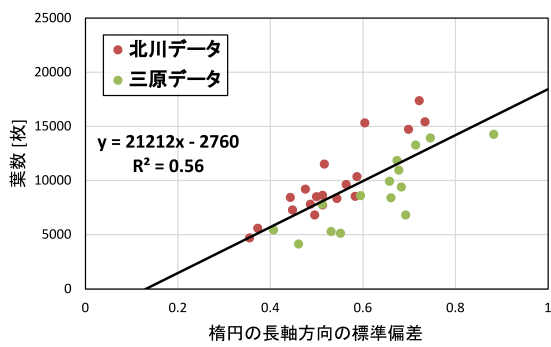


図 A.3: 楕円の長軸方向標準偏差

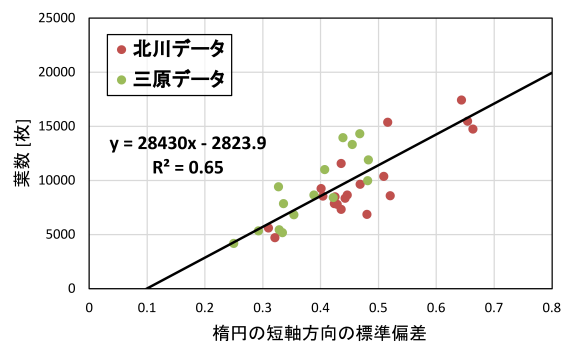


図 A.4: 楕円の短軸方向標準偏差

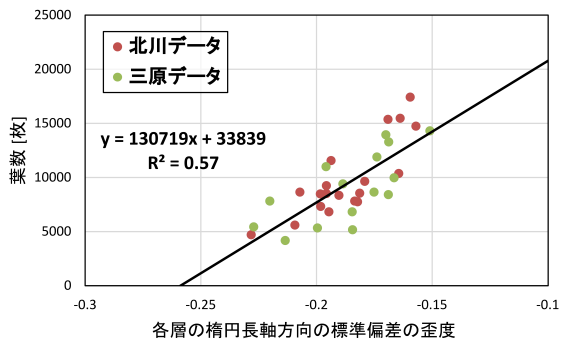


図 A.5: 各層の長軸方向標準偏差の歪度

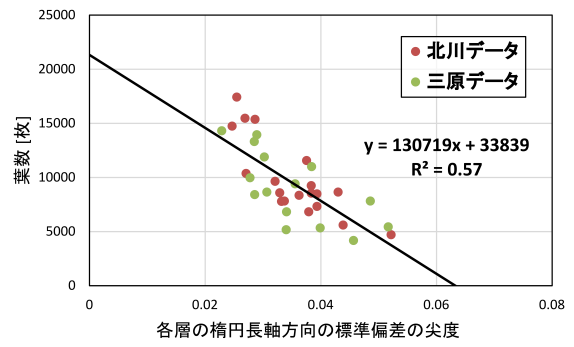


図 A.6: 樹高各層の長軸方向標準偏差の尖度

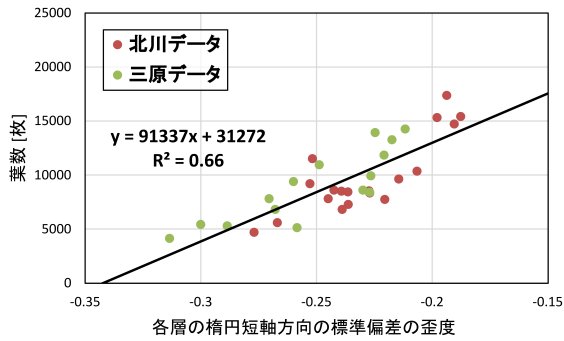


図 A.7: 各層の短軸方向標準偏差の歪度

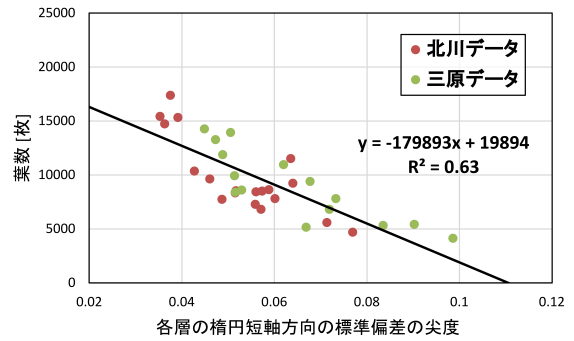
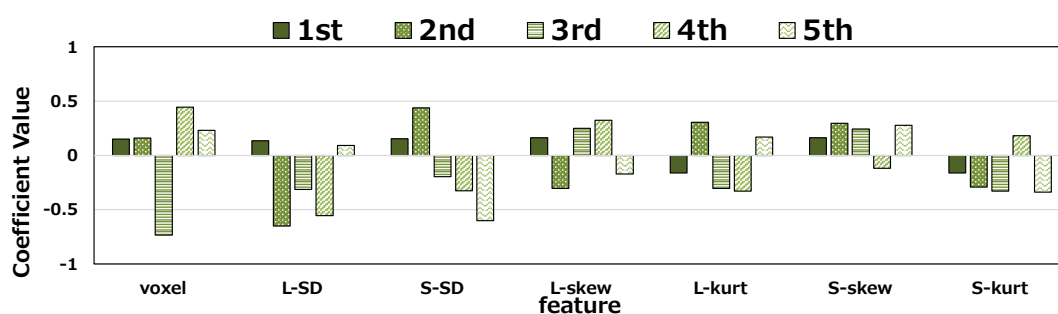


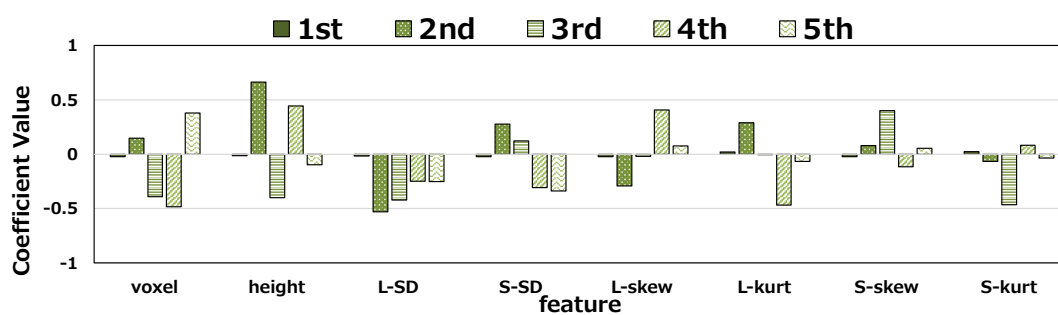
図 A.8: 各層の短軸方向標準偏差の尖度

付録 B

各データにおける主成分係数

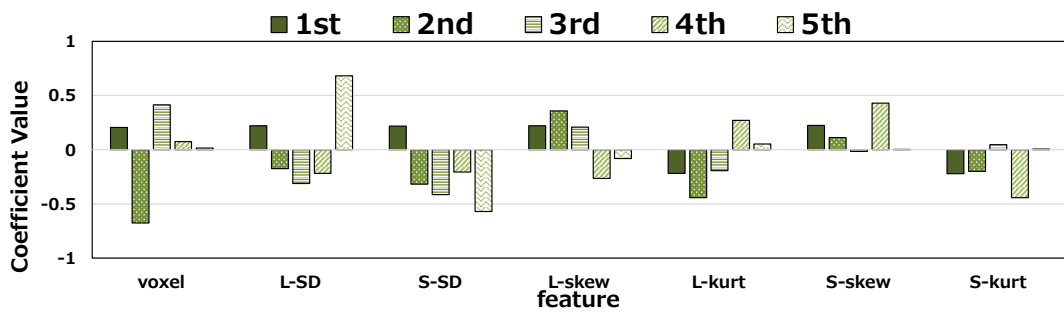


(a) 7 特徴量

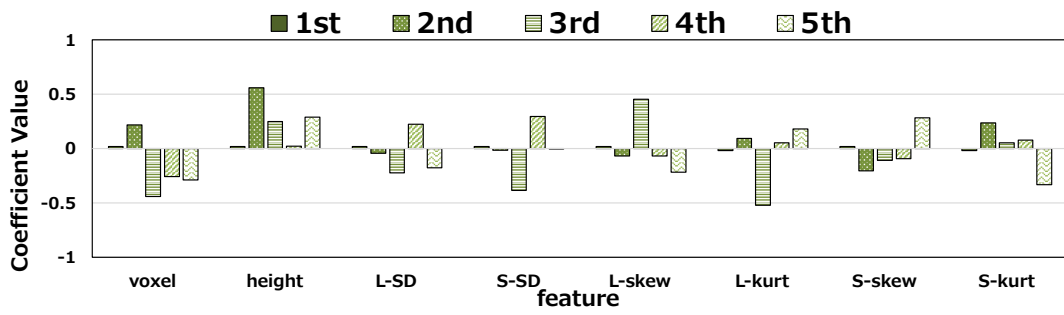


(b) 8 特徴量

図 B.1: PCR の主成分係数 (すべてのデータ)

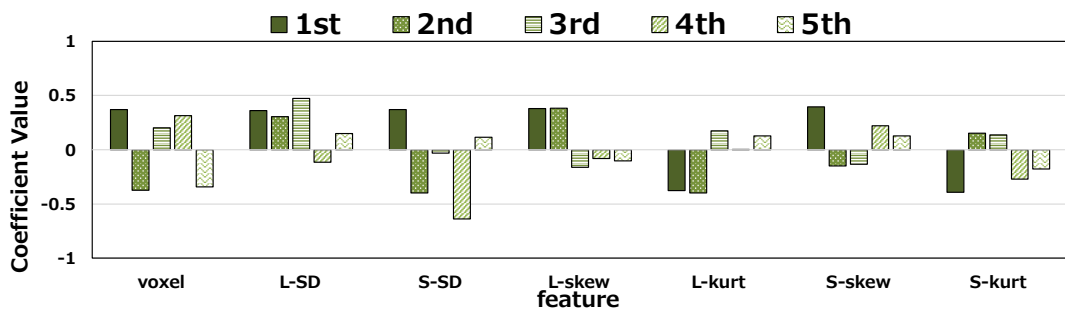


(a) 7 特徴量

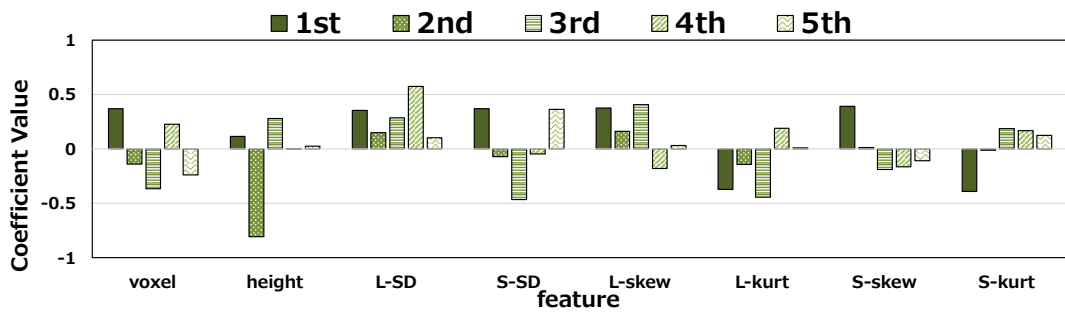


(b) 8 特徴量

図 B.2: PCR の主成分係数 (北川村データ)



(a) 7 特徴量



(b) 8 特徴量

図 B.3: PCR の主成分係数 (三原村データ)