

教師なしクラスタリング手法による ブドウ葉検出用疑似ラベル生成手法の考察

小嶋研究室
0312023156 山本倅也

2026/07/27

令和8年度夏季中間発表

1

研究背景

- ・ 農業分野において、AIを活用した画像解析が進められている
- ・ 物体検出モデルの学習には、大量のラベル付きデータが必要である
- ・ 物体の位置を示すラベルは、一般的に人手で作成される
- ・ 大量の画像に対するラベル作成には、時間と労力を要する



教師なしクラスタリングを用いた疑似ラベル生成により、ラベル作成の効率化を目指す

2026/07/27

令和8年度夏季中間発表

2

疑似ラベル

ラベル

人が画像内の物体の位置や種類を示した正解情報

疑似ラベル

人手で付与した正解ラベルではなく、モデルやアルゴリズムによって自動生成されたラベル

2026/07/27

令和8年度夏季中間発表

3

研究目的

ブドウ葉画像に教師なしクラスタリングを適用し、葉検出に利用可能な疑似ラベル生成方法を検討する

検討内容

- ・ ブドウ葉画像に対する教師なしクラスタリングの適用
- ・ クラスタリング結果を改善する特徴量と前処理の検討
- ・ 複数の教師なしクラスタリング手法の比較
- ・ 疑似ラベル生成に適したクラスタリング結果の評価

最終目標：人手によるラベル作成負担の軽減

2026/07/27

令和8年度夏季中間発表

4

研究全体の流れ

ブドウ葉画像

↓
特徴量抽出・前処理

↓
【現在の段階】
教師なしクラスタリング
DBSCAN・K-means・その他の手法

結果から葉領域を推定

↓
物体検出用疑似ラベルの生成

↓
物体検出モデルの学習

↓
人手ラベルを使用した場合との比較

2026/07/27

令和8年度夏季中間発表

5

これまでの研究内容

段階	実施内容
画像前処理	画像サイズの統一、HSV変換
特徴量検討	HSV、位置情報、RGB、ExG
葉領域抽出	HSVを利用した葉マスク
DBSCAN	パラメータ・特徴量・前処理の比較
K-means	複数のKによるクラスタリング
数値評価	3種類の評価指標による最適Kの選択
自動処理	複数画像の一括処理、Kの自動選択

2026/07/27

令和8年度夏季中間発表

6

現在使用している特徴量

特徴量	内容
H	色相
S	彩度
V	明度
x, y	画像上の画素位置
RGB	赤・緑・青の色情報
ExG	緑色の強調指標

主に使用している特徴量

xyHSV = (x, y, H, S, V)

位置情報を加える理由

色が似ていても、画像上で離れた領域が同一クラスタになることを抑える

2026/07/27

令和8年度夏季中間発表

7

葉領域マスク

HSVの値を用いて、緑色と考えられる画素を抽出

処理内容

1. 入力画像をHSV色空間へ変換
2. 設定したHSV範囲に含まれる画素を抽出
3. 抽出範囲外の画素をクラスタリング対象から除外し、クラスタリング精度向上を図る

目的

背景画素を事前に除外し、クラスタリングへの影響を低減する

使用範囲

H : 25~90
S : 30~255
V : 30~255

2026/07/27

令和8年度夏季中間発表

8

DBSCAN

データの密度に基づいてクラスタを形成する教師なしクラスタリング手法

基本的な処理

- 基準点を選択
 - eps以内の近傍点を探索
 - 近傍点数がmin_samples以上か判定
 - クラスタを拡張し、全データの処理が終了するまで繰り返す
- 特徴
- ・クラスタ数を指定不要
 - ・ノイズを検出可能
 - ・結果はパラメータ設定の影響を受けやすい
- ・ eps : 近傍と判定する距離
・ min_samples : クラスタ形成に必要な最小点数

2026/07/27

令和8年度夏季中間発表

9

DBSCANで行った実験

実験項目	検討内容
パラメータ比較	eps, min_samples
特徴量比較	HSV, xyHSV, xyRGB, xyExGなど
前処理比較	葉マスクなし／あり
画像比較	一葉, 複数葉, 背景あり画像
数値評価	クラスタ数, ノイズ率, 最大クラスタ率を用いて評価

基準条件の例

eps = 0.2, min_samples = 50, sample_step = 2, 特徴量 = xyHSV

比較条件 : 特徴量, 前処理, パラメータ, 複数画像

2026/07/27

令和8年度夏季中間発表

10

DBSCANの実験結果（背景あり・複数葉画像）



元画像



前処理なし



葉マスクあり

結果

- ・葉マスクを適用しても、背景の緑色領域が残る場合があった
- ・複数の葉が同一クラスタとなる場合が確認された
- ・一枚の葉が複数のクラスタへ分割される場合があった

2026/07/27

令和8年度夏季中間発表

11

DBSCANの課題

- ・画像ごとに適切なパラメータが異なる
- ・背景に緑色領域がある場合、葉マスクだけでは除外できない
- ・複数の葉が近接する場合、同じクラスタになることがある
- ・同じ葉が複数のクラスタに分割される場合がある
- ・パラメータ探索の計算量が大きい

今後の対応

- ・K-meansとの比較（実施中）
- ・境界情報の追加
- ・他クラスタリング手法との比較

2026/07/27

令和8年度夏季中間発表

12

K-means法

データを、指定したK個のクラスタに分類する教師なしクラスタリング手法

基本的な処理

- K個のクラスタ中心を初期化
 - 各データを最も近いクラスタ中心へ割り当て
 - クラスタ中心を再計算
 - 割り当てが安定するまで繰り返す
- 特徴
- ・代表的なクラスタリング手法
 - ・処理が比較的単純
 - ・クラスタ数Kを事前に設定する必要がある

2026/07/27

令和8年度夏季中間発表

13

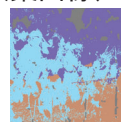
K-meansの実験結果（背景あり・複数葉画像）



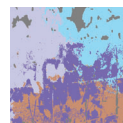
元画像



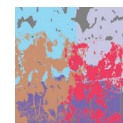
K=2



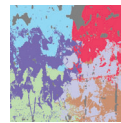
K=3



K=4



K=5



K=6

2026/07/27

令和8年度夏季中間発表

14

K-meansの評価指標

Kを決めるために3種類の評価指標を使用

Silhouette Score(S) : クラスタ内のまとまりとクラスタ間の分離を評価（高いほど良い）

Davies-Bouldin Index(DB) : クラスタ同士の近さ・重なりを評価（低いほど良い）

Calinski-Harabasz Index(CH) : クラスタリング全体の性能を評価（高いほど良い）

2026/07/27

令和8年度夏季中間発表

15

Kの自動選択方法

画像ごとに最適なクラスタ数Kを自動選択

↓
K = 2~10について
K-meansを実行

↓
各Kについて評価

↓
評価指標を統合

↓
最適なクラスタ数Kを決定

評価指標	良い結果
Silhouette Score	高い
Davies-Bouldin Index	低い
Calinski-Harabasz Index	高い

2026/07/27

令和8年度夏季中間発表

16

評価指標の統合方法

正規化による統合

Silhouette (高いほど良い)
Davies-Bouldin (低いほど良い)
Calinski-Harabasz (高いほど良い)

↓
各指標を0～1に正規化し、
DBのみ反転 (1-DB)

↓
3指標の平均値を計算

↓
平均が最大のKを選択

総合スコア $= (S + (1-DB) + CH) / 3$

順位による統合

Silhouette順位
DB順位
CH順位

↓
3指標の順位を合計

↓
順位合計が最小

↓
最適なK

順位合計 = S順位 + DB順位 + CH順位

2026/07/27

令和8年度夏季中間発表

17

複数画像に対する数値評価

対象画像：67枚

K候補：2～10

統合方法

- ・ 正規化
- ・ 順位

評価指標

- ・ Silhouette
- ・ Davies-Bouldin
- ・ Calinski-Harabasz

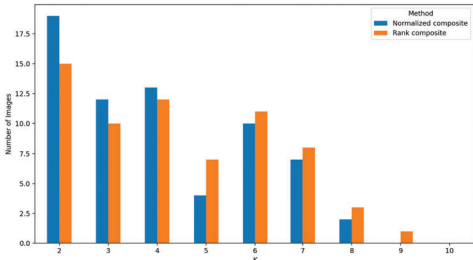


図1. 統合方法による最終選択Kの分布

2026/07/27

令和8年度夏季中間発表

18

現在の結果と考察

現在の結果

- ・ DBSCANとK-meansを同一の特徴量・前処理で実行可能となった
- ・ 複数画像に対する一括処理を実装した
- ・ K-meansでは3指標によるKの自動選択を実装した
- ・ 目視評価だけでなく、数値評価も可能となった

考察

- ・ 評価指標が良くても、葉領域を正しく抽出できるとは限らない
- ・ 現在の評価指標はクラスタ構造を評価しており、疑似ラベルの精度を直接評価していない
- ・ 色と位置情報だけでは、葉と背景の境界を十分に表現できない場合がある

2026/07/27

令和8年度夏季中間発表

19

今後の予定

1. 葉マスク処理の改善
2. Kの自動選択結果とKの選択範囲の拡張
3. 正解ラベルを用いた数値評価方法の検討
4. Sobel等による境界情報の追加
5. その他手法の実装・比較

2026/07/27

令和8年度夏季中間発表

20