Vol.2014-CVIM-191 No.16
2014/3/3

ブドウ摘みロボットのための RGBD画像認識手法の基礎検討

川口 達也[1,a)]　川上 玲[2]　池内 克史[1]

概要：代替的な労働力として期待される果実の収穫ロボットは枝や葉などが入り乱れた複雑な環境を適切に認識する必要がある．本稿では，ブドウ摘みロボットのセンシング機構のために，RGBD 画像を利用した認識手法について検討をおこなう．RGB 画像の一般物体認識手法を適用した結果から問題点を調べた．また奥行き情報を加えた場合の大局的特徴とセグメンテーション手法について提案を行い，局所的特徴について検討をおこなった．

1. はじめに

農業従事者の急速な減少・高齢化に伴い，代替的な労働力として農業用ロボットへの期待が高まってきている．とりわけ果実の収穫は複雑で細かい作業が要求されるため，通常の農業用機械ではなくロボットを用いる意義が大きい．農林水産省による 2007 年の調査 [23] によれば，ブドウ収穫作業に要する農家一戸当たりの労働時間は年間 287 時間にも上り，収穫ロボットの導入は農家の負担軽減に大きく貢献すると考えられる．収穫ロボットには様々な技術的課題があるが，果実や葉や枝などの遮蔽物の認識はその一つである．

ブドウの房は形状が個々で大きく異なり，表面色が周囲に溶け込みやすいため，リンゴやイチゴなどの果物に比べ認識が容易ではない．画像中からブドウを検出する手法では，色，テクスチャ，形状などを利用したものがこれまで提案されてきた [5], [16], [18]．しかしこれらはブドウの房のみの検出にとどまり，葉や枝などの認識はおこなっていない．収穫ロボットは，房を遮蔽する葉を取り除く，枝を切り取る，といった幅広いタスクを要求されるため，房以外の物体も認識することが望ましい．

Dey らは，Structure from Motion を用いて取得されたブドウの樹の 3 次元形状から房・葉・枝をそれぞれ認識する手法を提案した [6]．色情報と Saliency Feature 特徴量 [11] を用いて認識をおこない高い精度を得ている．しかし葉が果実を遮蔽しない，房はほぼ水平に並ぶ，といった仮定を敷いていた．

一般物体認識の分野では，RGB 画像を用いた研究が数多くなされてきた．しかし近年，レンジセンサとして安価な Kinect の登場により容易に奥行画像が取得可能となったため，RGB に奥行値という強力な情報を加えた RGBD 画像を用いた認識手法が増加している [2], [3]．しかし RGBD 画像認識の分野は若いため，研究例も RGB 画像のものほど多くはない．

我々は，ブドウ農園の画像において上述した様な仮定は用いず，また房だけでない多クラスの認識を目的とする．そのため，まず多くの先行研究がなされている RGB のみを用いた認識手法を適用し，問題点を探る．その後 RGBD 画像に応用した際の手法について提案と検討をおこなう．

近年の一般物体認識手法には，

- 大局的特徴量を用い，画像データを選択
- Superpixel ごとに特徴量を計算
- Superpixlel のクラスを推定
- コンテクストを考慮したラベリング

といったステップを踏む手法がある [8], [14]．Superpixel とは，似た特徴を持つピクセルをまとめた領域を指す．

このような手法では，大量の訓練画像を用いて SVM や Adaboost といった識別器を生成することが多いが，多クラスの識別器の生成では，クラス数の増加に伴い計算コストが増大し，また訓練画像が追加されるごとに逐一生成し直す必要があるという問題がある．

Tighe らにより提案された Super-parsing[20] は上記の問題を解決する手法として注目されている．この手法では学習データを用いた訓練はせず，Superpixel そのものを辞書として，Superpixel の特徴空間において近傍探索をおこ

[1] 東京大学
The University of Tokyo
[2] 大阪大学
Osaka University
[a)] kawa@cvl.iis.u-tokyo.ac.jp

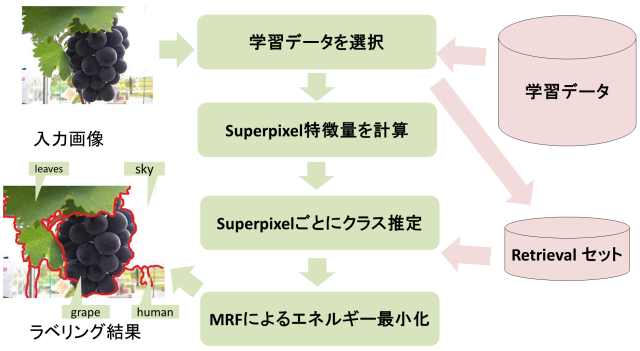

図 1　Super-parsing による学習認識手法の流れ
Fig. 1　The Flow of Leaning and Recognition Method based on Super-parsing

なっている．そのため識別器は使用せず，学習データの追加も容易であるため，多クラスの高速認識に向いている．またこの手法は特徴量の入れ替えが容易であるため RGBD 画像を用いた認識への拡張性が大きい．本研究では Super-parsing の手法を紹介しつつ，ブドウ農園の RGB 画像の認識に向けた改良手法を説明し，また認識実験について述べる．

Super-parsing は上述した理由から RGBD 画像認識への応用が容易である．我々は Super-parsing の枠組みをベースとしたブドウ農園の RGBD 画像認識の手法を提案する．RGBD 画像認識は分野の若さゆえ，上記の一般物体認識の各ステップそれぞれで研究が浅い．例えば一枚の RGB 画像を表す大局的特徴量はこれまで様々な手法が提案されてきたが，RGBD 画像の大局的特徴量は少ない．我々はブドウ農園の RGBD 画像の選択という目的の元，新しい大局的特徴量を複数提案し，比較実験を実施した．また RGBD 画像のセグメンテーション手法においては，奥行値は強力な情報となるが，単純に奥行値を追加した場合，RGB 画像と奥行画像のずれが悪影響を及ぼす，遠方の距離値が支配的になる，などの問題により精度の良い分割が困難となる．我々はそれらの問題を解決するセグメンテーション手法を提案し，実験によりその有効性を示した．

2. Super-parsing に基づく認識

本節では，Super-parsing による認識手法について紹介し，その改良点について述べる．

2.1 Super-parsing

Super-parsing の手法の流れを図 1 に示す．最初に，入力画像と似た学習用画像を大局的特徴量を用いて選択し，これを Retrieval セットとする．次に入力画像にセグメンテーションをおこない，生成された Superpixel ごとに特徴量を計算する．この特徴量を本稿では Superpixel 特徴量と呼ぶことにする．そして各 Superpixel が所属すべきクラスについてベイズ理論を元に予測をおこなう．その後，隣接する Superpixel のクラス間の不自然さをもとに MRF のエネルギー関数を生成し，それを最小化することでコンテクストに基づいたラベリングを施す．最後に，幾何的クラス（垂直，水平など）と意味的クラス（川，建物など）の整合性にペナルティを課して最終的なラベリングを施す．以下，各段階における手法の概略について述べる．詳しくは [20] を参照されたい．

以下「学習データ」とは，後述する大局的特徴量・Superpixel 特徴量・Superpixel 隣接関係が前処理によって計算され，あらかじめ人間によってラベリングが施されている画像セットのことを指す．

学習データの画像を全て用いると計算コストが増大し，また関係の薄いデータが認識に悪影響を及ぼすため，あらかじめ入力画像に近い画像を学習データから選択する．一枚一枚の画像の特徴量として，Spatial pyramid[12]，Gist[17]，Tiny image[21]，カラーヒストグラムの 4 つの大局的特徴量を用いている．その 4 つの特徴量ごとに，入力画像と特徴量が近い学習データの画像を kNN 探索を用いて選ぶ．[20] では一つの特徴量につき 50 枚としているため，$50 \times 4 = 200$ 枚の画像が選ばれる．これを Retrieval セットと定義する．

Retrieval セットを元に入力画像の認識をおこなうが，画像にセグメンテーションを施して得られた Superpixel ごとに認識をおこなう．Superpixel をベースとした認識は，ピクセル単位や長方形単位で認識をおこなう従来手法 [13] に比べ高速処理が可能であるだけでなく，単一物体の特徴を計算しやすいため精度の向上にもつながる．[20] では，graph-based segmentation アルゴリズム (GS04)[7] を用いて Superpixel を生成している．各 Superpixel において，形状，色，画像中の位置，テクスチャといった様々な特徴に基づく 20 個の特徴量 [20] を計算する．

得られた Superpixel 特徴量と，Retrieval セット内の Superpixel 特徴量を比較することで，各 Superpixel が所属すべきクラスについて推測をおこなう．s_i を i 番目の Superpixel，c_j を j 番目のクラスとすると，公算比 $L(s_i, c_j)$ を全ての i，j に対し計算する．$L(s_i, c_j)$ は以下の式で表される．

$$L(s_i, c_j) = \prod_m \left(\frac{n(c_j, N_i^m)}{n(\overline{c_j}, N_i^m)} \times \frac{n(\overline{c_j}, D)}{n(c_j, D)} \right) \tag{1}$$

ここで $n(a, B)$ は Superpixel の集合 B の中でクラス a にラベル付けをされた要素の個数を表す．D は学習データ内の全 Superpixel，N_i^m は m 番目の特徴量空間において s_i との距離が閾値 t_m 以下の Retrieval セット内の Superpixel の集合である．t_m は，学習データの全 Superpixel 間の m 番目の特徴量空間における 20 近傍の中央値距離を用いている．学習データ内の値がゼロとなるのを防ぐため，$n(c_j, N_i^m)$ と $n(\overline{c_j}, N_i^m)$ には 1 が加算されている．

Superpixel ごとに独立したラベリングでは，水が空に浮

かんでいるといった現実的に不自然なラベリングが施されることがあるため，Superpixel の隣接関係に基づくラベリングが施される．学習データからクラス同士が接する確率を取得し，MRF のエネルギー最小化問題を α-β swap アルゴリズム [4], [10] を用いて解く．最小化すべき式は

$$J(\mathbf{c}) = \sum_{s_i \in SP} g(s_i, c_i) + \lambda \sum_{(s_i, s_j) \in A} h(c_i, c_j) \qquad (2)$$

と表せる．ここでは $\mathbf{c}$ をクラスの集合 ($\mathbf{c} = \{c_i\}$)，λ を平滑化定数, SP を Superpixel の集合，A を隣接する Superpixel のペアの集合とする．g，h はそれぞれデータ項・平滑化項を意味し，それぞれ以下の式で表される．

$$g(s_i, c_i) = -w_i \log L(s_i, c_i) \qquad (3)$$

$$h(c_i, c_j) = -\log\left((P(c_i \mid c_j) + P(c_j \mid c_i))/2\right) \times \delta(c_i, c_j) \qquad (4)$$

ここで w_i は s_i の面積を Superpixel の平均面積で割った値である．$P(c_i \mid c_j)$ は c_j の隣に c_i が存在する確率である．また $\delta_{i,j}$ は

$$\delta(c_i, c_j) = \begin{cases} 0 & (i = i) \\ 1 & (i \neq j) \end{cases} \qquad (5)$$

を表す．以上の手法により各 Superpixel が一つのクラスに対応付けられる．これをラベリング結果として出力する．

[20] ではこの後，建物は垂直であり，川は水平である，といった幾何的クラスと意味的クラスの整合性に基づくラベリングをおこなうが，これは屋外の生活空間位おける画像の認識に特化しているため，本稿では触れない．

2.2 改良手法

Super-parsing の手法は屋内と屋外の生活空間における画像の認識を目的としているため，ブドウ農園の画像認識のために改良をおこなう．

提案手法では，セグメンテーションの手法で，GS04[7] ではなく SLIC アルゴリズム [1] を用いる．SLIC は Superpixel の個数や大きさの安定性などを調節でき，かつ高速で高精度なセグメンテーションのアルゴリズムである [15]．

また Superpixel 特徴量は，[20] で用いた特徴量を全て用いるのではなく，tab:tab1 で示される 9 個の特徴量を計算する．Superixel の特徴量を計算する際に，領域内だけでなく周囲の情報も重要だと考え，Superpixel から 4 連結成分で 10 ピクセル膨張させた領域を膨張 Superpixel とした．また Texton，SIFT 特徴量についてはあらかじめ Bag of Features を用いて 100 個の辞書を作成しておき，最近傍の辞書に投票したヒストグラムを特徴量とした．

3. RGBD 画像を用いた手法

本節では，RGB 画像に奥行画像を加えた RGBD 画像を用いて認識をおこなう場合について，大局的特徴量を用いた学習データの選択とセグメンテーションのアルゴリズムのそれぞれについて，提案手法を説明する．

表 1　Superpixel 特徴量
Table 1　Superpixel Features

特徴量	次元
bounding box のマスク画像の形状 (8 × 8)	64
画像の高さに対する y 座標	1
膨張 Superpixel 内における Texton のヒストグラム	100
膨張 Superpixel 内における SIFT のヒストグラム	100
RGB 平均	3
RGB 標準偏差	3
カラーヒストグラム (11 bin)	33
サムネイル RGB 画像 (8 × 8)	192
グレー画像の Gist	320

図 2　隙間の多い画像の例
Fig. 2　Example of an Image Containing Many Gaps

3.1 大局的特徴量

一枚の RGB 画像を表す大局的特徴量はこれまで様々な手法が提案されてきたが，分野の新しさゆえ RGBD 画像の大局的特徴量についてはあまり研究がなされていない．ブドウ農園の RGBD 画像の選択という目的の元，新しい大局的特徴量の提案をおこなう．

本研究では，RGB 画像の大局的特徴量である Tiny image[21] への奥行情報を追加を試みる．Tiny image とは，画像を $n \times m$ ブロックに分割し各ブロックの RGB それぞれの平均値を特徴量としたものである．本研究では 16×16 ブロックに分割する．この場合 RGB のみの Tiny image の次元は $3 \times 16 \times 16 = 768$ となる．

奥行情報を追加する最も単純な手法は，RGB と同様に各ブロックの平均をとるものである．しかし図 **2** のようにブドウ農園の画像は隙間が散在し，平均値はその隙間に影響され不安定となる．また画像の選択をおこなう際，近方は細かく遠方は粗く評価するのが妥当である．奥行値をそのまま用いず，画像選択という目的に即した特徴量を設計する必要がある．

本研究では，様々な特徴量を提案して比較をおこなう．各ブロックの距離の平均だけでなく，中央値，十分位数，平方根などを組み合わせた特徴量を候補とした．候補一覧

表 2　大局的特徴量と次元
Table 2　The Global Features and their dimensions

記号	各ブロックの値 (16 × 16 ブロック)	次元
a	RGB の平均	786
b	奥行値の平均	256
c	奥行値の中央値	256
d	$\sqrt{\text{奥行値}}$ の平均	256
e	$\sqrt{\text{奥行値}}$ の中央値	256
f	奥行値の十分位数	256
g	$\sqrt{\text{奥行値}}$ の十分位数	256
h	RGB の平均 と 奥行値の平均	1024
i	RGB の平均 と奥行値の中央値	1024
j	RGB の平均 と $\sqrt{\text{奥行値}}$ の平均	1024
k	RGB の平均 と $\sqrt{\text{奥行値}}$ の中央値	1024
l	RGB の平均 と奥行値の十分位数	1024
m	RGB の平均 と $\sqrt{\text{奥行値}}$ の十分位数	1024

とその次元を**表 2** に示す．中央値や十分位数を用いたのは隙間による影響を軽減するため，平方根を用いたのは遠方を粗く評価するためである．本研究ではどの特徴量が最も適切にブドウ農園の画像を選択するか検証すべく実験をおこなった．

3.2　セグメンテーション

本研究の手法のような Superpixel 単位で認識をおこなう手法の場合，Superpixel が複数クラスに跨ると最終的なラベリングの精度が落ちる．物体認識を目的としたセグメンテーションでは各 Superpixel が単一のクラスしか含有しないようなアルゴリズムが必要となる．また Superpixel は認識の計算コストを減らす役割を担うが，セグメンテーションに時間を要すれば全体としての実行時間も伸びる．

植物の画像は物体が入り乱れ，また細かい枝などが混在しているため，セグメンテーションが容易ではない．Superpixel の個数を増やせば必然的にセグメンテーションの精度は伸びるが，それに伴い個々の Superpixel の面積が小さくなるため，認識において特徴量が算出しにくい．また RGB 画像と奥行画像の位置合わせをおこなう際にずれが生じるため，その悪影響を軽減する必要がある．

提案手法では SLIC[1] アルゴリズムに基づき，特徴空間において k-means 法を適用しピクセルをクラスタリングする手法を用いる．特徴空間は，Lab 色空間・xy 座標・奥行値 z を合わせた 6 次元空間を用いる．以下アルゴリズムの概略を述べる．

Superpixel の個数を K，画像サイズを N とする．K 個の seed-point を用意し，これを画像中に均等に配置する．この seed-point を k-means 法における重心とする．このとき Superpixel の一辺の長さの平均は $S = \sqrt{N/K}$ となる．その後は k-means 法と同様に seed-point の移動とクラスタの再構成を反復しておこなう．通常の k-means 法は収束するまで反復されるが，計算量を減らす目的で反復回数の上限を指定する．本研究では 10 回としている．また画像の全ピクセルと全 seed-point の距離は計算せず，seed-point の近傍の $2S \times 2S$ の領域のみ計算する．このアルゴリズムでは最終的に，分断された細かな領域が取り残される可能性があるため，連結成分ラベリング法 [19] を用いて小さい領域は近くの Superpixel に取り込む．

クラスタの再構成時に計算する seed-point と各ピクセル間の距離は，Lab 空間・xy 空間・z 空間で個別に計算された距離の和を用いる．その際単位の違いを考慮した重みづけを施す必要がある．3.1 で述べたものと同様の理由から，奥行値 z をそのまま用いず，特殊な扱いを施す．本手法では，seed-point sp_i の座標を $(l_i^s, a_i^s, b_i^s, x_i^s, y_i^s, z_i^s)$，ピクセル p_j の座標を $(l_j^p, a_j^p, b_j^p, x_j^p, y_j^p, z_j^p)$ としたとき，sp_i と p_j の距離 $distance$ を

$$d_{lab} = \sqrt{(l_i^s - l_j^p)^2 + (a_i^s - a_j^p)^2 + (b_i^s - b_j^p)^2} \tag{6}$$

$$d_{xy} = \sqrt{(x_i^s - x_j^p)^2 + (y_i^s - y_j^p)^2} \tag{7}$$

$$d_z = \left| \frac{z_i^s - z_j^p}{z_i^s + z_j^p} \right| \tag{8}$$

$$distance = d_{lab} + \frac{m_1}{S} d_{xy} + m_2 d_z \tag{9}$$

と計算する．ここで m_1，m_2 はそれぞれ d_{xy}，d_z の影響力を決める係数である．本研究では $m_1 = 7$，$m_2 = 15$ を用いた．このアルゴリズムの計算量は，反復回数が定数であり，探索範囲が限定されるため，$O(N)$ となる．

d_z では，遠距離になるほど奥行差を減らすような重みづけをおこなっている．これは近方ほど細かく遠方ほど粗く評価するためである．

4.　実験

本節では，2 節，3 節で説明した手法を評価するため，ブドウ農園の RGB 画像の認識と，RGBD 画像の選択とセグメンテーションについて実験をおこなった．

4.1　RGB 画像の認識

2 節で述べた RGB 画像認識手法について実験をおこなった．学習データとして 15 枚の画像を用意し，LabelMe[24] によってあらかじめラベリングをおこなう．その後，大局的特徴量，Superpixel 特徴量，Superpixel の隣接関係について前処理をおこなう．入力画像として**図 3** に示す 4 枚のブドウ農園の画像を用いた．使用した画像は 700 × 525 ピクセルのサイズを持つ 8bit の JPEG 画像である．

認識結果を図 3 に示す．左から，元画像，Superpixel 特徴量のみを用いたラベリング結果，コンテクストを考慮したラベリング結果となる．

画像 1 は枝の背後に房があるという構造であり，房・葉・枝・地面の領域の大まかな認識に成功している．しかし画

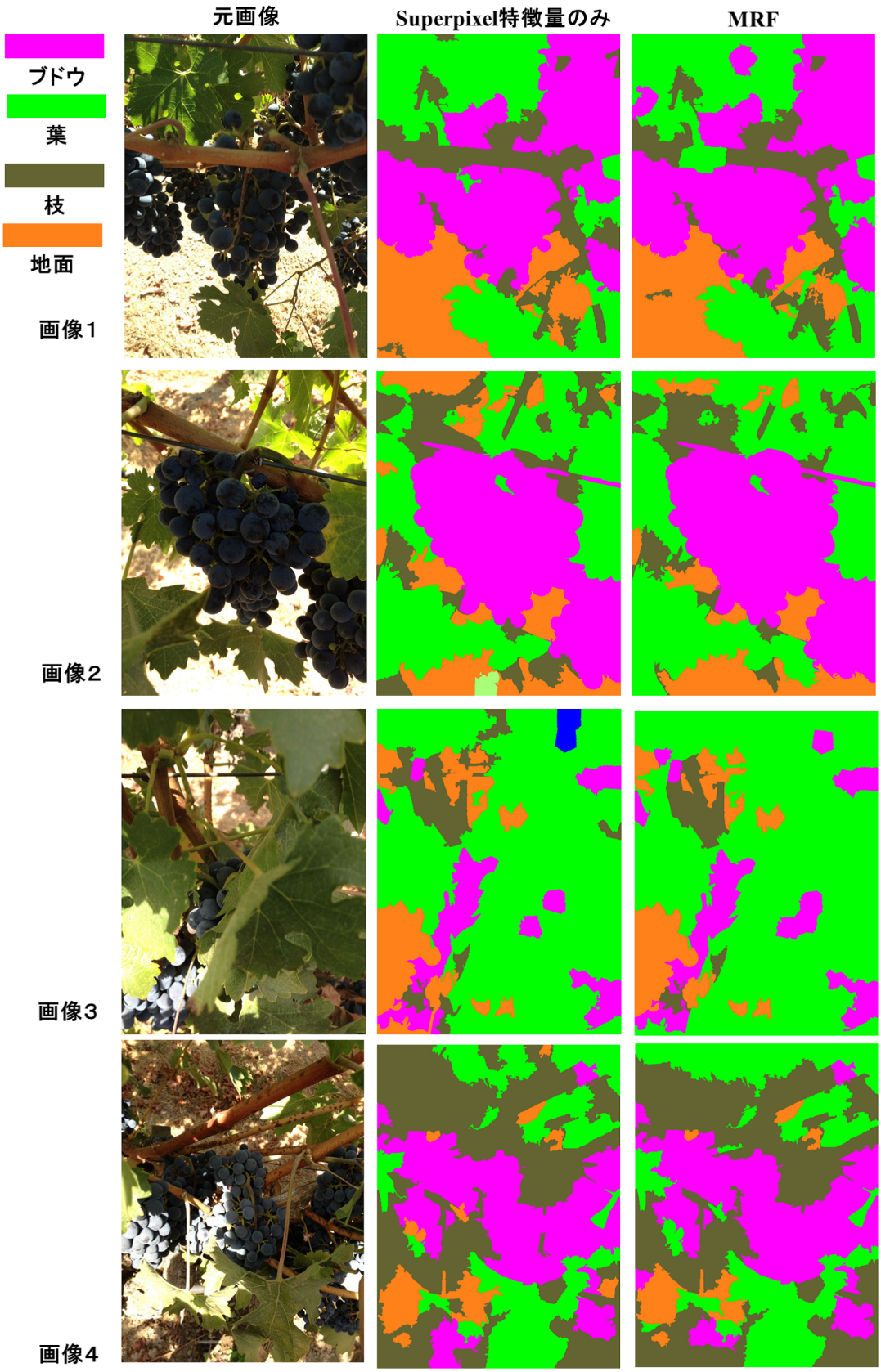

図 3　RGB 画像認識の結果
Fig. 3　The Result of RGB Image Recognition

像上部の葉の領域では影となった部位が房の色と似ているため誤認識を起こしている．右下の細かい枝ではセグメンテーションの精度の問題か認識が粗い．画像 2 は房を中心とした画像であり，その領域を精度よくラベリングしている．しかし地面と枝は，色とテクスチャが似ているためか誤認識を起こしている．また金属棒をブドウと誤認識しているが，これは学習データに金属の領域が不足していたため十分な Superpixel が確保できていないことが原因だと考えられる．画像 3 は房が葉に遮蔽されている画像である．隠れたブドウもよく検出しているが，他の領域をブドウだと誤認識を起こしている．また左上の葉・枝・地面が入り乱れる領域では誤認識の率が高い．画像 4 は日光と影の影響により同じクラスの物体でも色に大きな差が生じている．そのため大まかな認識はしているものの，複数のクラスが入り乱れる画像上部では誤認識が多い．

図 3 が示すように，コンテクストを考慮したラベリングをおこなった場合，Superpixel 特徴量のみを用いた場合よりも，誤認識が増加する傾向にある．実験に用いたような植物の画像は，従来の一般物体認識で用いられてきたような屋内や屋外の画像と違い，複数のクラスが複雑に入り乱れているため，単純な隣接関係を考慮したラベリング手法が精度を落とす要因になったと考えられる．

4.2 RGBD 画像の選択

3.1 で述べた大局的特徴量による画像選択の手法について比較実験を行った．以下，実験で使用する画像は全て，レンジセンサとして Kinect，RGB カメラとして Point Grey Research 社製の Chameleon を用いて撮影し，[22] の手法を用いて画像データの位置合わせをおこなったものである．画像サイズは Chameleon の規格である 1296 × 964 ピクセルに統一し，Kinect の奥行画像はそれに合わせ引き伸ばされている．

本研究では，ロボットがブドウの房を収穫する状況を想定し，ブドウ農園の画像 143 枚をブドウの房への距離に応じ，

- 近距離：収獲すべき房に手が届く
- 中距離：複数の房が見えるため収穫すべき房を選ぶ
- 遠距離：遠くの房を探す

の 3 クラスに分類した．これをデータセットとする．また入力画像として，3 クラスそれぞれに対応する画像を 2 枚ずつ，計 6 枚のブドウ農園の画像を用意した．これら全ての画像について，表 2 の 13 種類の大局的特徴量を計算した．

画像選別の精度を特徴量ごとに評価するため，各入力画像と特徴量が近い画像をデータセットから 20 枚ずつ，13 種類の特徴量ごとに選択した．そしてクラスが一致する適合率，つまり入力画像のクラスと選択された画像のクラスが一致した個数を選択された画像枚数で割った値を求めた．特徴量間の距離については，L1 ノルム，L2 ノルムの

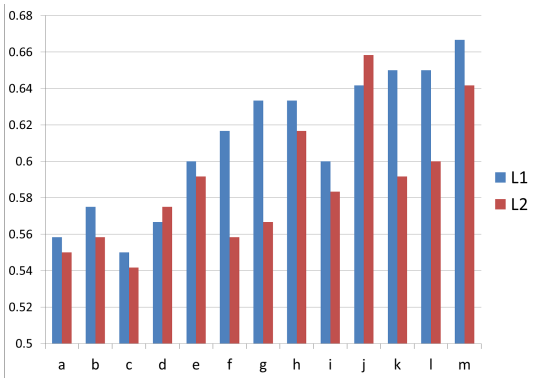

図 4　大局的特徴量による画像選択の適合率の比較

Fig. 4　The Comparison of the Precisions of the Image Selection by Global Features

2 種類を用いた．

図 4 にその結果を示す．横軸が各特徴量（記号は表 2 の記号に対応），縦軸が適合率を表す．a は従来の Tiny image であるが，L1 ノルム，L2 ノルムともに低い値をとっている．全体の傾向といて距離計算の手法として L1 ノルムを用いた方が適合率が高く，また奥行値だけを用いた特徴量よりも，RGB と奥行値の両者を用いた特徴量の方が適合率が高い．

本研究で用いた特徴量の中では，m の各ブロックの RGB の平均と奥行値の平方根の十分位数をとった特徴量を用いて L1 ノルムで距離計算を行ったものが，最も適合率が高かった．これは 3.1 で述べた，隙間による影響を減らし，遠方のものは粗く評価すべき，という本研究の考えと一致する．また特徴量の距離計算の手法は L2 ノルムが一般的であることを考慮すると，j の RGB の平均と奥行値の平方根の平均を特徴量としたものが，最も適合率が良い．いずれにせよ遠距離の影響力を減らすことが重要であると考えられる．

4.3 RGBD 画像のセグメンテーション

3.2 で述べたセグメンテーションのアルゴリズムを**図 5** 左上の RGBD 画像に適用した．画像サイズ・形式などは 4.2 と同様である．提案手法と比較するため，SLIC，SLIC+Z を用いた結果も示す．SLIC+Z は，元の SLIC アルゴリズムに奥行値 z をそのまま追加したものである．つまり式 (8) における d_z を

$$d_z = \left| z_i^s - z_j^p \right| \tag{10}$$

としている．Superpixel の個数を 600 個としてセグメンテーションをおこなった．結果を図 5 に示す．

(a) 列はセグメンテーションに用いた元画像，(b)，(c)，(d) 列はその一部を拡大した画像である．画像中の紫色の線が Superpixel 同士の境界線を表す．

(b) ではそれぞれの手法がブドウの房の境界線を良く表しているが，SLIC は色に敏感であるため他手法よりも細か

く分割している．3.2 で述べたように，Superpixel が必要以上に細かくなるのは望ましくない．(c) では SLIC と提案手法は細い枝の領域を精度よく切り取っているが，SLIC+Z は奥行画像と RGB 画像のずれの影響により切り取りに失敗している．(d) では枝を，SLIC は細かく分割し，SLIC+Z は境界線のずれを生じさせている．だが提案手法は枝の領域のみを綺麗に切り取っていることがわかる．

このように提案手法は，色情報に過敏に反応した分割を防ぎつつ，RGB 画像と奥行画像のずれによる悪影響を軽減する，という特徴を持つことが示せた．

5. 今後の展望

本節では，RGBD 画像認識の手法について検討する．3.2 と 3.3 で述べた手法は学習データの選択とセグメンテーションにとどまっていたため，Superpixel 特徴量，コンテクストを考慮したラベリングについて，手法の検討をおこなう．

5.1 Superpixel 特徴量

奥行情報を利用すると，Superpixel 間の遮蔽関係を知ることができる．遮蔽関係により使用する特徴量を使い分ければ，ラベリング精度の向上が期待される．

例えばブドウの房の一部が葉に遮蔽される場合を考える．この画像をセグメンテーションした際，房と葉の境界で Superpixel の境界が現れたとする．このとき境界線の輪郭は葉の輪郭を表すが，ブドウ側の情報は含まれない．輪郭の特徴量として，Livarinen らによって提案された Chain code histogram(CCH)[9] などがある．Superpixel 特徴量に基づいてクラスを推測する際，遮蔽の奥側にある房の Superpixel においては CCH を推測に用いず，他の特徴量の重みを増し，逆に遮蔽の前側の葉の Superpixel では CCH は有力な情報となるため重みを増やす，といった手法が考えられる．

このように遮蔽関係に応じて特徴量を使い分けることで，より Superpixel の特徴を正しく評価できる可能性が高まる．

5.2 コンテクストを考慮したラベリング

RGB 画像認識実験で示したように，従来手法のような単純な隣接関係のみを考慮したモデルでは，植物の複雑な構造に対応できず，かえって認識精度を落とす．特に葉・枝・地面が入り乱れた領域では誤認識が多い．

奥行情報を利用した場合，地面と前景物体を選別できため，そのような誤認識の軽減は可能だと考えられる．また認識をより強固にするため，隣接関係と遮蔽関係の一貫性を用いる手法が考えられる．例えば地面に枝が遮蔽されるといった状況は考えにくいため，ラベリング時に可能性の低い遮蔽関係にペナルティを課すことにより，テクスチャの似ている枝と地面の精度の良い識別が可能だと期待される．従来の MRF のエネルギー関数の平滑化項において，学習データから得られたクラス間の遮蔽関係の不自然さ基づいたペナルティ項を追加する手法について，今後検討していきたい．

6. まとめ

本稿では，近年の RGB 画像を用いた一般物体認識における手法を応用したブドウ農園の RGB 画像を認識する手法と実験結果について述べた．また RGBD 画像に発展させた場合の大局的特徴量による画像選択，セグメンテーション手法について提案をおこない，実験によってその有効性を示した．最後には Superpixel 特徴量とコンテクストを考慮したラベリング手法について基礎検討をおこなった．

RGB のみを使った認識では，大まかな認識は可能であることを示したが，植物の複雑な構造が認識精度を悪化させた．従来の一般物体認識は屋外や屋内における人間の生活空間に重点が置かれ，植物の画像に関する研究はあまりなされていなかった．今後は複雑な構造を持つ植物の画像を精度よく認識する手法について研究を進めたい．

RGBD 画像を使った大局的特徴量は，従来の RGB 画像用の特徴量に比べ，状況に応じた画像選択に適していることを示した．植物の画像は隙間が多いため，遠方の奥行値による影響を軽減することにより，画像選択をより適切におこなうことができた．適切な画像選択は認識の精度と計算速度の向上につながる．また RGBD 画像を用いたセグメンテーション手法を提案し，RGB のみを用いた従来手法や奥行情報を単純に加えた手法よりも，適切に画像を分割できることを示した．複雑な植物の画像をセグメンテーションすることは難しい問題であるためあまり研究されていないが，Superpixel を用いた手法は精度と時間の双方での効果が期待されるため，今後も引き続き Superpixel に基づいた認識手法について研究を進めていきたい．

Superpixel 特徴量やコンテクストを考慮したラベリング手法はまだ検討段階であるが，RGB 画像認識の手法の枠組みを下地に，奥行画像という強力な情報を生かすことで更なる認識力の向上が期待される．

今後は，まだ検討段階である手法について実装を進め，認識実験によりその有効性を検証したい．

謝辞 本研究は JSPS 科研費 24240034 の助成を受けたものである．

参考文献

[1] R. Achanta, et al. “Slic superpixels.” Ecole Polytechnique Federal de Lausssanne (2010).
[2] M. Blum, et al. “A learned feature descriptor for object recognition in rgb-d data.” ICRA (2012).
[3] L. Bo, X. Ren, and D. Fox. “Depth kernel descriptors for object recognition.” Intelligent Robots and Systems

図 5　セグメンテーション手法の比較
Fig. 5　Comparison of the Segmentation Methods

(2011).
[4] Y. Boykov, and V. Kolmogorov. “An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision.” Pattern Analysis and Machine Intelligence (2004).
[5] R. Chamelat, et al. “Grape detection by image processing.” IEEE Industrial Electronics, IECON 2006-32nd Annual Conference on. IEEE (2006).
[6] D. Dey, L. Mummert, and R. Sukthankar. “Classification of plant structures from uncalibrated image sequences.” Applications of Computer Vision (2012).
[7] PF. Felzenszwalb, and DP. Huttenlocher. “Efficient graph-based image segmentation.” International Journal of Computer Vision (2004).
[8] D. Hoiem, AA. Efros, and M. Hebert. “Recovering surface layout from an image.” International Journal of Computer Vision (2007).
[9] J. Iivarinen, and Ari J. E Visa. “Shape recognition of irregular objects.” Photonics East’96. International Society for Optics and Photonics (1996).
[10] V. Kolmogorov, and R. Zabin. “What energy functions can be minimized via graph cuts?.” Pattern Analysis and Machine Intelligence (2004).
[11] JF. Lalonde, et al. “Natural terrain classification using three - dimensional ladar data for ground robot mobility.” Journal of field robotics (2006).
[12] S. Lazebnik, C. Schmid, and J. Ponce.“Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories.” Computer Vision and Pattern Recognition (2006).
[13] C. Liu, J. Yuen, and A. Torralba. “Nonparametric scene parsing: Label transfer via dense scene alignment.” CVPR (2009).
[14] T. Malisiewicz, and AA. Efros. “Recognition by association via learning per-exemplar distances.” Computer Vision and Pattern Recognition (2008).
[15] P. Neubert, and P. Protzel. “Superpixel Benchmark and Comparison.”, Technical report (2012).
[16] S. Nuske, et al. “Yield estimation in vineyards by visual grape detection.” Intelligent Robots and Systems (2011).
[17] A. Oliva, and T. Antonio. “Building the gist of a scene: The role of global image features in recognition.” Progress in brain research (2006).
[18] MJCS. Reis, et al. “Automatic detection of bunches of grapes in natural environment from color images.” Journal of Applied Logic (2012).
[19] A. Rosenfeld, and JL. Pfaltz. “Sequential operations in digital picture processing.” Journal of the ACM (1966).
[20] J. Tighe, and S. Lazebnik. “Superparsing: scalable nonparametric image parsing with superpixels” Computer Vision-ECCV 2010. Springer Berlin Heidelberg (2010).
[21] A. Torralba, R. Fergus, and W. T. Freeman. “80 million tiny images: A large data set for nonparametric object and scene recognition.” Pattern Analysis and Machine Intelligence (2008).
[22] Z. Zhang. “A flexible new technique for camera calibration.” IEEE Transactions on Pattern Analysis and Machine Intelligence (2000).
[23] 農林水産省 品目別経営統計
http://www.maff.go.jp/j/tokei/kouhyou/noukei/hinmoku/
[24] LabelMe: http://labelme.csail.mit.edu/Release3.0/