

ブドウ摘房支援のための画像における 房計数領域の推定

西田 龍平^{1,a)} 内海 ゆづ子^{1,b)} 三輪 由佳^{2,c)} 岩村 雅一^{1,d)} 黄瀬 浩一^{1,e)}

概要:ブドウ栽培の作業の1つである摘房は、単位面積あたりのブドウの房数を数えながら間引くため重労働である。画像を用いた単位面積あたりのブドウの計数手法は既に確立している。しかし、この手法では、房を計数する対象の領域が分かるよう、ブドウ棚に目印をつけ、それをもとに対象領域のブドウの房数を推定している。目印をつける作業は非常に手間がかかるため、できるだけそのような作業が必要のない手法が望ましい。そこで本研究では、目印に依存せずに房を計数することを目的として、計数の対象領域を切り出す手法を提案する。決まった大きさの平面領域を切り出すためには、その平面とカメラの位置関係と、スケールを推定する必要がある。そこで、提案手法では、ブドウの圃場内を Visual SLAM を用いて 3 次元復元し、長さが既知であるブドウ棚と地面からの距離を用いて推定された復元結果のスケールを推定し、切り出し領域を決定する。実験の結果、動画画像からマーカを用いずに計数領域を切り出すことに成功した。

1. はじめに

ブドウ栽培の作業の1つに摘房がある。これは、ブドウの房を間引くことによりブドウの品質を保つための作業である。ブドウは光合成の能力上、棚の面積により生成できる糖の量が決まっている。そのため、1房あたりの糖度を出荷可能な品質に保つための摘房基準として、単位面積あたりの房数が決められている。そのため作業者は、ブドウの房を数えながら間引くこととなり、摘房作業は2種類の作業を同時に行うことから重労働となる。我々の研究グループでは、摘房における作業者の負担軽減のため、ブドウの房の数を自動で計数するシステムを開発している。

現在、画像を用いた単位面積あたりの房の計数手法 [1] は既に確立されている。しかし、この計数手法はブドウ棚に計数したい正方形の領域の頂点に目安となるマーカを人の手で付け、そのマーカをもとに画像から計数領域を切り出して推定をしている。この手法を適用するには、作業者にブドウ棚の全ての計数したい領域の頂点にマーカを付けてもらい、そのマーカが写るよう、計数領域をカメラで撮

影しなければならず負担が大きい。そこで、本研究ではブドウ計数手法をマーカを使わずに適用することを目的とし、ブドウ棚を撮影した任意の視点の画像から、計数領域を切り出す手法を提案する。

任意の画像から房計数領域を切り出すためには、ブドウ棚とカメラの位置の把握が必要である。既存手法として、動画内の 3 次元の位置関係とカメラの姿勢を同時に得る手法である Visual Simultaneously Localization and Mapping (Visual SLAM) により、周りの環境の 3 次元形状の復元とカメラ姿勢を取得し、ブドウ棚の位置を見つけることは可能である。しかし、Visual SLAM で得られる 3 次元復元では、画像内にあるものの相対的な位置関係しかわからず、得られた 3 次元復元結果からは、スケールが不明である。そのため、スケール推定を行わなければ、本研究のような特定の大きさの領域を切り出すことはできない。

空間を 3 次元復元した結果に対するスケール推定手法には、Global Positioning System (GPS) などの衛星測位システムを使い実際の距離を測ることで 3 次元復元結果と実際の距離を比較したり、慣性計測装置 (inertial measurement unit; IMU) [2] や遠赤外線 (far-infrared; FIR) カメラ [3]、透明平板 [4] などカメラ以外に他の装置を使って得られた情報を元にスケール推定するものが提案されている。これらのように、復元結果内で長さの既知のものから、スケールを推定をする手法が多く提案されている。そこで、本研究では、ブドウ棚の地面から高さをを用いて復元結果のスケールを推定する。ブドウ棚の地面からの高さは、圃場で

¹ 大阪府立大学 大学院工学研究科
Graduate School of Engineering, Osaka Prefecture University

² 大阪府立環境農林水産総合研究所

a) nishida@m.cs.osakafu-u.ac.jp

b) yuzuko@cs.osakafu-u.ac.jp

c) MiwaY@mbox.kannousuiken-osaka.or.jp

d) masa@cs.osakafu-u.ac.jp

e) kise@cs.osakafu-u.ac.jp

の実測が容易である。また、ブドウ棚と地面は平面であり、復元結果の3次元点群から容易に抽出可能である。これらの理由から、本研究ではブドウ棚の高さをを用いて3次元復元結果のスケール推定を行い、決まった面積である摘房領域を切り出す。実験の結果、画像内にブドウの房を計数する一定の領域を切り出すことができた。

2. 関連研究

ここではブドウ栽培支援、Visual SLAM、スケール推定についての関連研究について述べる。

2.1 ブドウ栽培支援

ブドウ栽培を支援する手法は既に存在する。NellithimaruらはRobust Object-Level SLAM for grape counting (ROLS) [5]と呼ばれる、動画を使用しブドウの房を計数するシステムを提案した。ROLSは垣根状にブドウ果樹が植えられた圃場内をロボットが通り過ぎながら動画を撮影し、ブドウの房を計数する手法である。そのため、本研究で想定している棚状にブドウ果樹が植えられた圃場とは想定している環境が異なる。また、垣根状に果樹が植えられた圃場での摘房は枝あたりの房数が基準としてあるので、棚状に植えられた圃場の場合のように領域ごとの計数が不必要である。そのため、ROLSを本研究で想定している地面と平行のブドウ棚の房計数に使用することはできない。

2.2 Visual SLAM

Visual SLAMは動画内の特徴点の位置合わせにより3次元環境地図とカメラの自己位置推定を同時に行う技術のことである。各フレーム間での画像のマッチング手法により多種多様なVisual SLAMが存在する。例えばORB-SLAM [6]は動画のフレームから特徴点を検出しその点の局所特徴量を表すORB特徴量で特徴点を抽出し再投影誤差を最小化することで推定するVisual SLAMである。それに対し、LSD-SLAM [7]は輝度勾配の大きな画素を抽出し輝度値を最小化し推定するVisual SLAMである。

また、自然環境などでは類似物、移動物体、光の明暗、カメラで撮影しづらいなどの影響によりVisual SLAMに悪影響を及ぼすことがあるため、その場所や環境によって合ったVisual SLAMが使われる。例として道路 [8]、畑 [9]、森林 [10] そして水中 [11] など多岐にわたる。本研究では全天球動画を用いるため、全天球動画を容易に使用できるOpenVSLAM [12]を使用した。

2.3 スケール推定

3次元復元でのスケールを推定する手法と画像からスケールを推定する手法について紹介する。3次元復元においてスケール推定する手法には、実際の距離が既知であるものと比較する手法 [13] やRGBカメラ以外の装置を使用

して得られる情報で推定する手法 [2], [3], [4] が挙げられる。GPSなどの衛星測位システムを使う場合、3次元復元での特定の距離をGPSで得られた実際の距離と比較することでスケール推定ができる。Kittら [13]は地面を検出しカメラの高さを固定することで、カメラの高さを比較しスケールの推定を行った。Nütziら [2]は慣性計測装置 (inertial measurement unit: IMU) から得られる加速度と角速度、複数の画像から得られるカメラ姿勢を拡張カルマンフィルタ (extended Kalman filter: EKF) を用いて統合することでスケールを推定した。角倉ら [3]は、1台のカメラを移動させ取得した画像のみを用いて3次元復元を行う手法であるStructure from Motion (SfM) において、遠赤外線 (far-infrared: FIR) カメラとRGBカメラを組み合わせたマルチモーダルステレオカメラを用い、2視点でのFIR画像のエピポーラ拘束の残差の最小化によって得られるFIR画像間での対応点とFIRカメラとRGBカメラそれぞれの相対的な姿勢によりスケールを推定した。Chenら [4]は、SfMにおいて透明平板の有無による視差からスケール推定をした。

果樹栽培に関連した画像からのスケール推定の手法としては、Mominら [14]のマンゴーの大きさを推定する手法が挙げられる。カメラとマンゴーの距離を固定して撮影できる装置を使い、画素あたりの実際の距離を求めることで、マンゴーが写っている画素の領域を数え大きさを推定した。

本研究では、ブドウ棚と地面が水平であるため、計測が簡単な既知の距離であるブドウ棚の高さを比較することでスケール推定を行う。

3. 提案手法

提案手法では、カメラで撮影した圃場を3次元復元しながらスケールを推定し、ブドウの計数領域が画像上のどの範囲を占めるかをカメラ位置と3次元復元結果から計算し、撮影画像上から計数領域を切り出す。提案手法の流れを図1に示す。ブドウ圃場の撮影には、全天球カメラを用いる。撮影されたブドウ圃場を撮影した動画をもとに、OpenVSLAMを用いて空間の3次元復元を行う。続いて、3次元復元結果からブドウ棚と地面を表す平面を抽出する。次に、平面どうしの距離をブドウ棚の高さと比較することで房計数領域のスケールを推定する。最後に、任意の画像からマーカに頼らずブドウ棚の房を計数する領域を表示させる。以下、圃場を撮影した画像である全天球画像、圃場の3次元復元を行うOpenVSLAM、スケール推定と、計数領域の切り出しについて詳細を説明する。

3.1 全天球画像での圃場の撮影

本研究ではまず、圃場の地面とブドウ棚の3次元位置情報を得るために、圃場をカメラで撮影する。撮影には、ブドウ圃場内のブドウ棚部分と圃場内の地面を同時に撮影

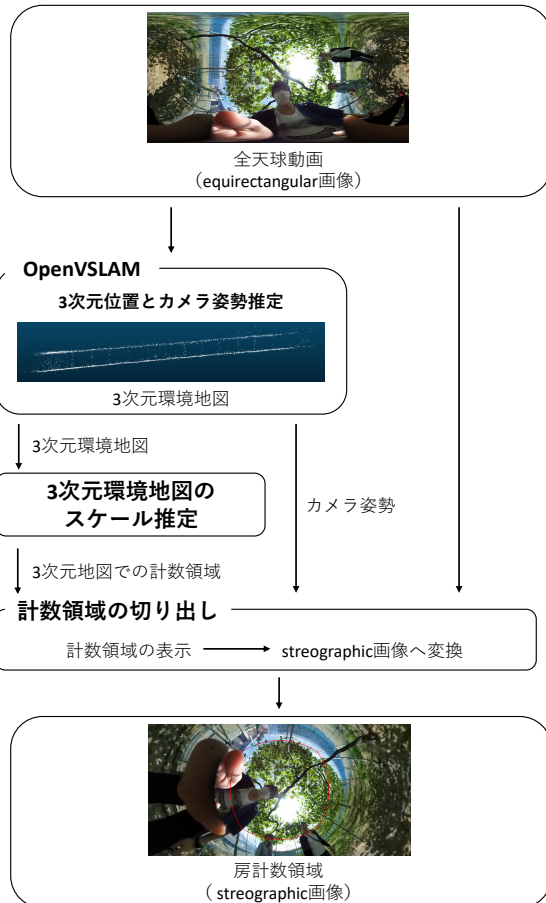


図 1 提案手法の概要

出来る全地球カメラを用いる。得られた全地球動画をもとに、圃場と地面の位置推定をする。

全地球動画を使う理由は 2 点挙げられる。1 点目は、ブドウ棚の計数領域を全体が写るように撮影するためには広角なカメラが必要であるからである。ブドウ棚の高さが約 180 センチメートル、計数領域は $2\text{ m} \times 2\text{ m}$ である。計数のための画像はブドウ棚の下部から、作業者の目線程度の高さで撮影するため、計数領域を 1 枚の画像に収めるには、一般的なデジタルカメラの画角では不十分である。そのため、画角の広いカメラが必要であることから、全地球カメラを使用する。

全地球動画を用いる 2 点目の理由は、ブドウ棚と地面両側の 3 次元復元結果を効率的に得ることができるためである。本研究でのスケール推定には、ブドウ棚と地面の双方が含まれる必要がある。ブドウ棚と地面の両方を同時に撮影できる全地球カメラを用いることで、1 つの視点でその周り全てに空間の撮影が可能となり、カメラを不用意に動かすことなくブドウ棚と地面の双方の復元が可能となる。

3.2 圃場の 3 次元復元とスケールの推定

全地球動画に対し OpenVSLAM [12] を使用し、ブドウ

圃場内の 3 次元形状とカメラ姿勢を得る。Visual SLAM はカメラで撮影した動画を入力することで、カメラの自己位置推定と撮影した周辺環境の 3 次元復元を同時に行う技術のことである。Visual SLAM の一種である OpenVSLAM は ORB 特徴量を使用する特徴点ベースの Visual SLAM で、動画のフレームどうしでの特徴点の相対的な位置合わせを最適化することでカメラ位置の推定と特徴点の集合である 3 次元復元ができる。

OpenVSLAM では、圃場の 3 次元形状が復元できるものの、そのスケールは不明である。計数領域を切り出すためには、復元した結果から既知の情報を用いて、スケールを推定する必要がある。そこで本研究では、ブドウ棚の高さ、つまりブドウ棚と地面の距離を用いて比較することでスケールを推定する。ブドウ棚の高さでスケール推定する理由としては、ブドウ棚と地面の 3 次元復元が安定してできるため 3 次元復元からの取得が容易であり、実際のブドウ棚の高さの計測が簡単である。また棚状にブドウ果樹が植えられた圃場であれば必ず計測できるものであるからである。そこで、3 次元復元からブドウ棚と地面を表す平面を抽出し、ブドウ棚平面と地面平面との距離を求めることで 3 次元復元でのブドウ棚の高さを求める。

OpenVSLAM で得られる 3 次元復元結果は 3 次元点群データであり、ブドウ棚と地面の推定結果は誤差が含まれているものの、おおよそ平面として復元されている。この点群データから、地面とブドウ棚の平面を抽出するために Random Sample Consensus (RANSAC) [15] を用いる。RANSAC は、データ集合から少数のサンプルをランダムに選び最小 2 乗法によりモデル化したのち外れ値が少ないものを選ぶ一連の作業を繰り返すことでより外れ値が少ないモデルを選ぶ手法である。今回は、3 次元での平面の式をモデルとして RANSAC を用いる。抽出できた平面と平面の距離を 3 次元復元のブドウ棚の高さとし、実際のブドウ棚の高さを使うことでスケールを推定する。

3.3 計数領域の切り出し

スケール推定の結果をもとに、ブドウ棚上で計数を行う領域を決定し、撮影された画像からその領域を切り出す。ブドウの計数手法 [1] では、全地球画像のうち、3 次元視野を投影した際に歪みが少ない Stereographic 画像を用いている。そのため、本研究でも、Stereographic 画像から計数領域を切り出すこととする。ここでは、切り出しの方法について、ブドウ棚のどの領域を計数領域として決定するか、また、撮影動画の各フレームの Stereographic 画像への変換と、計数領域を画像平面上に投影する方法について詳細に述べる。

3.3.1 計数領域の決定

ブドウの計数手法 [1] では、カメラ中心が計数領域の中心に位置するよう、撮影を行っている。ブドウの計数手法

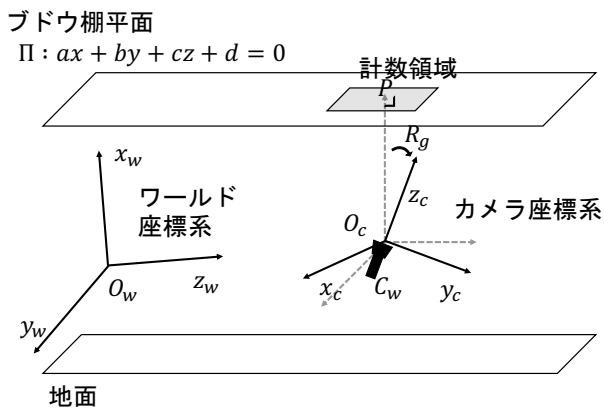


図 2 カメラ位置とブドウ棚の計数領域の関係

を適用することを考慮すると、この際の撮影条件と同様な計数領域を設定する必要がある。そこで、本研究では、カメラ中心が計数領域の中心に位置するよう、切り出し領域を設定する。

切り出し領域の決定方法の詳細について説明する。図 2 は、圃場の 3 次元復元結果から推定した、3 次元空間上でのブドウ棚 Π と地面 G と、あるフレームでのカメラの中心位置 O_c を示す。計数領域は、 O_c から、 Π に向けて垂線を引き、その交点 P を中心とし、その周囲の $N \times m$ 四方の領域と決定する。

3.3.2 Stereographic 画像変換

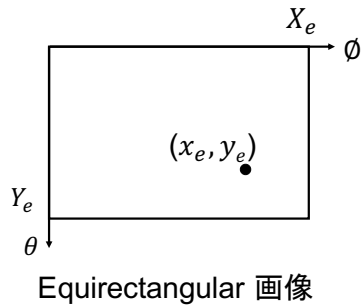
図 1 の全天球動画で示した通り、圃場を撮影した動画の各フレームは、Equirectangular 画像として保存されている。そのため、まず、Equirectangular 画像を Stereographic 画像に変換したのち、計数領域を Stereographic 画像に投影することで、切り出しを行う。ここでは、Equirectangular 画像から Stereographic 画像への変換方法について説明をする。

まずは、Equirectangular 画像と Stereographic 画像が、全天球視野をどのように 2 次元画像に変換しているかを説明する。図 3 は、全天球の視野を Equirectangular 画像、Stereographic 画像それぞれに投影するモデルである。3 次元座標の原点はカメラ中心 C であり、視野である単位円上の点 Q は、極座標を用いて $(1, \theta, \phi)$ と表すことができる。この点 Q が、幅 X_e pixels、高さ Y_e pixels の Equirectangular 画像の点 (x_e, y_e) に投影されたとすると、 θ, ϕ は x_e, y_e を用いて以下のように表現される。

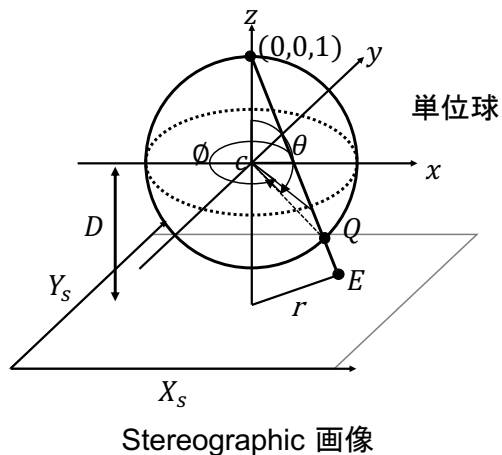
$$\phi = 2\pi \frac{x_e}{X_e} \quad (1)$$

$$\theta = \pi \frac{y_e}{Y_e} \quad (2)$$

続いて Stereographic 画像の解像度を幅 X_s pixels、高さ Y_s pixels とし、カメラ中心から投影面までの距離を D 、点 Q を Stereographic 画像に投影した時の点を $E = (x_s, y_s)$ とすると、 E は次式で与えられる。



Equirectangular 画像



Stereographic 画像

図 3 Equirectangular 投影, Stereographic 投影のモデル

$$x_s = r \cos \phi + \frac{X_s}{2} \quad (3)$$

$$y_s = r \sin \phi + \frac{Y_s}{2} \quad (4)$$

ただし、 $r = (1 + D) \tan(\frac{\theta}{2})$ である。

ブドウ計数手法 [1] では、Stereographic 画像はブドウ棚がカメラ座標系の z 軸の負の位置に、かつ、 z 軸がブドウ棚に垂直になるように投影がされている。そこで、本研究でも同じようになるよう、Equirectangular 画像を単位円上に投影した後、座標系を回転させて補正を行った後、Stereographic 画像に投影をする。つまり、式 (1), (2) で Equirectangular 画像上の点を単位円上の点に変換し、回転の補正を行った後、式 (3), (4) を用いて Stereographic 画像へと変換する。補正の回転角と回転行列の求め方の詳細は A.1 で説明する。

3.3.3 Stereographic 画像への計数領域の投影

3.3.1 で決定した 3 次元空間上での切り出し領域を、3.3.2 で変換した Stereographic 画像に対して投影し、計数領域を切り出す。切り出しの具体的な状況を図 4 に示す。3.3.2 で補正したように、カメラ座標の z 軸の負の方向に計数領域があり、計数領域は z 軸と直交している。ブドウ棚平面の法線とブドウ棚平面との交点とカメラ位置の距離 F とすると、 F は、3.2 のカメラ位置推定と平面推定から得られたブドウ棚の平面までの距離として計算が可能である。

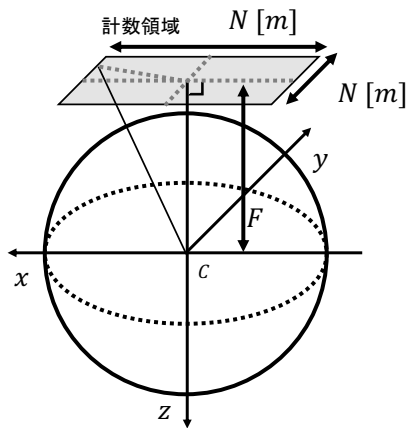


図 4 計数領域の切り出し

そして、図 4 に示すように 3.3.1 で決定した計数領域の N [m] \times N [m] の頂点をそれぞれをカメラ視野の単位球上に中心投影し、式 (3), (4) に基づいて Stereographic 画像上に投影することで、計数領域を表示させる。

4. 実験

OpenVSLAM で得られた 3 次元復元結果から、提案手法によりスケールの推定を行い、計数領域の切り出しが可能であるか、評価を行った。ここでは、実験での用いた動画データと実験上の設定、実験結果について述べる。

4.1 データ

実験に使用した動画は、2017 年 6 月 9 日に大阪府立環境農林水産総合研究所のブドウ園場で撮影されたものである。当日の天候は晴れであった。撮影には RICOH THETA S *1 を使用した。撮影は、ブドウ園場内をカメラを持ちながら歩いて撮影した。撮影した動画は解像度が 3840×1920 pixels, 30 fps であり、長さは 1 分 18 秒ある。

4.2 実験設定

今回、房を計数する領域は実際に作業者が房を数える際の基準となる $2\text{ m} \times 2\text{ m}$ の平面とした。またスケールを推定する際に使用する実際のブドウ園場での地面とブドウ棚との距離は 1.8 m であった。全天球動画はフレームレートに従い、1 秒間に 30 枚の静止画になるように画像を生成した。OpenVSLAM [12] は、著者らが公開していた実装を用いた。OpenVSLAM で探索してくる特徴点の上限を 5000 個とし、その他の設定は [12] に従った。

4.3 実験結果

OpenVSLAM を用いて得た 3 次元復元結果を図 5 に示す。本研究で必要であるブドウ棚と地面は復元できていることがわかる。また、ブドウ棚を支える柱が一部復元され

ていることがわかる。

続いて、この 3 次元復元について RANSAC を用いて平面抽出し、平面以外の復元点を除いた結果を図 6 に示す。

ただし、ブドウ棚を緑で、地面を茶色で表示させた。平面の抽出に成功しており、図 5 の復元結果には存在していたブドウ棚の柱の 3 次元位置を復元した点が図 6 では消去されていることがわかる。

最後に、提案手法でスケールを推定し、切り出し領域を Stereographic 画像に重畳させた。切り出し領域を赤い短形で囲った stereographic 画像を図 7 に示す。今回の結果について、ブドウ園場にある柱と柱の間隔が約 2 m であるので、それを参考にすると概ね成功していると言える。

5. まとめ

ブドウ栽培の作業の 1 つである摘房は、単位面積あたりのブドウの房数を数えながら間引くため重労働である。そこで、我々の研究グループでは摘房支援のための房を計数するシステムを開発しており、計数手法は既に確立している。しかしこの計数手法では、房を計数する領域を得るために領域ごとの頂点に印となるマーカを人の手で付けている。そこで本研究では、マーカに依存せずに房を計数する領域の切り出しを目的とする。計数領域を得るためにブドウ園場内の動画に対して Visual SLAM を用いた 3 次元復元を行ったが、相対的な位置関係しかわからずスケールが不明である問題があった。そこで、長さが既知のブドウ棚と地面の高さを用いてスケールを推定し、Stereographic 画像から房計数領域を切り出す手法を提案した。実験の結果、動画からマーカを用いずに計数領域を切り出すことに成功した。

今後の課題として、実験結果の定量的誤差評価、実際に計数手法に入力しての房計数の評価、最終的には房計数提案手法と計数手法を合わせた実用的なシステムの開発が挙げられる。

謝辞 本研究は 2019 年度電気普及財団研究調査助成、大阪府信用農業協同組合連合会令和 2 年度産学連携研究支援事業による研究成果に基づく。

参考文献

- [1] Akai, R., Utsumi, Y., Miwa, Y., Iwamura, M. and Kise, K.: Distortion-Adaptive Grape Bunch Counting for Omnidirectional Images, *Proceedings of the 25th International Conference on Pattern Recognition*, pp. 599–606 (2021).
- [2] Nützi, G., Weiss, S., Scaramuzza, D. and Siegwart, R.: Fusion of IMU and vision for absolute scale estimation in monocular SLAM, *Journal of intelligent & robotic systems*, pp. 287–299 (2011).
- [3] 角倉慎弥, 櫻田健, 河口信夫, 中村良介: マルチモーダルステレオカメラにおける単眼 SfM のスケール推定, 情報処理学会研究報告コンピュータビジョンとイメージメディア (CVIM), Vol. 2018, No. 29, pp. 1–8 (2018).
- [4] Chen, Z., Wong, K.-Y. K., Matsushita, Y., Zhu, X. and

*1 <https://theta360.com/ja/about/theta/s.html>

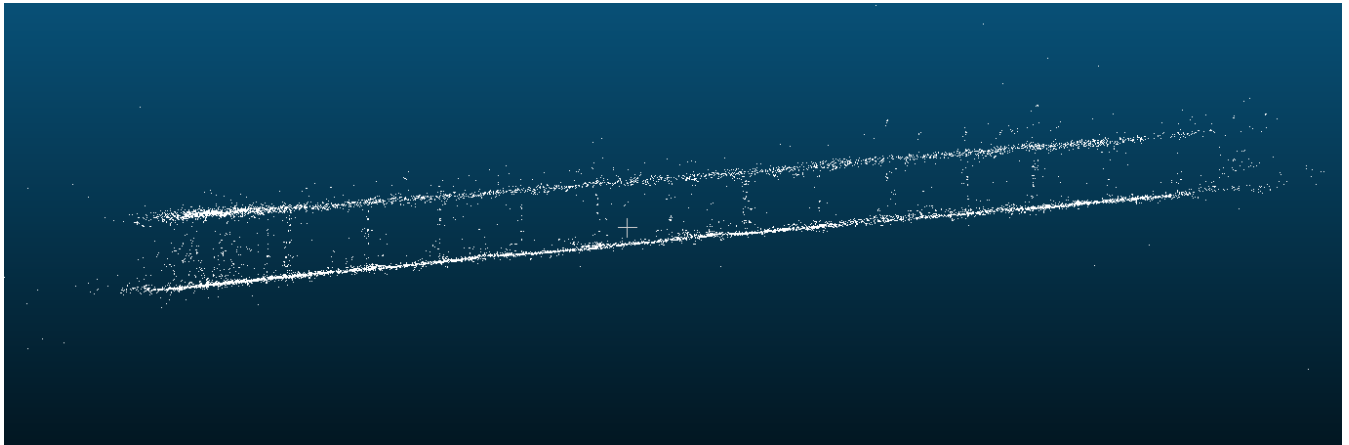


図 5 3次元復元結果

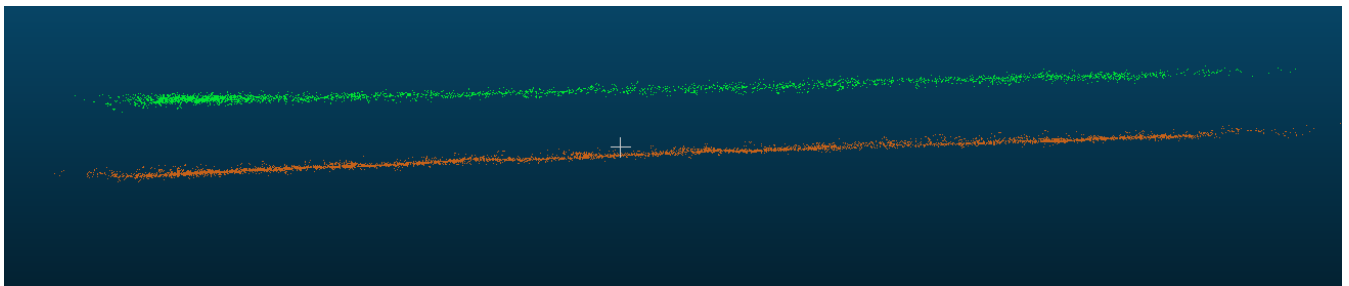


図 6 3次元復元の平面抽出の結果 (ブドウ棚を緑, 地面を茶色で表示)

- Liu, M.: Self-calibrating depth from refraction, *Proceedings of International Conference on Computer Vision*, pp. 635–642 (2011).
- [5] Nellithimaru, A. K. and Kantor, G. A.: ROLS: Robust Object-level SLAM for grape counting, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (2019).
- [6] Mur-Artal, R. and Tardós, J. D.: ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras, *IEEE Transactions on Robotics*, Vol. 33, No. 5, pp. 1255–1262 (2017).
- [7] Engel, J., Schöps, T. and Cremers, D.: LSD-SLAM: Large-scale direct monocular SLAM, *Proceedings of European conference on computer vision*, pp. 834–849 (2014).
- [8] Singh, G., Wu, M. and Lam, S. K.: Fusing semantics and motion state detection for robust visual SLAM, *Proceedings of Winter Conference on Applications of Computer Vision*, pp. 2753–2762 (2020).
- [9] Chebroly, N., Lottes, P., Läbe, T. and Stachniss, C.: Robot localization based on aerial images for precision agriculture tasks in crop fields, *Proceedings of International Conference on Robotics and Automation*, pp. 1787–1793 (2019).
- [10] Garforth, J. and Webb, B.: Visual appearance analysis of forest scenes for monocular SLAM, *Proceedings of International Conference on Robotics and Automation*, pp. 1794–1800 (2019).
- [11] Cheung, M. Y., Fourie, D., Rypkema, N. R., Teixeira, P. V., Schmidt, H. and Leonard, J.: Non-Gaussian SLAM utilizing synthetic aperture sonar, *Proceedings of International Conference on Robotics and Automation*, pp. 3457–3463 (2019).
- [12] Sumikura, S., Shibuya, M. and Sakurada, K.: OpenVS-LAM: A Versatile Visual SLAM Framework, *Proceedings of the 27th ACM International Conference on Multimedia*, pp. 2292–2295 (2019).
- [13] Kitt, B. M., Rehder, J., Chambers, A. D., Schonbein, M., Lategahn, H. and Singh, S.: Monocular visual odometry using a planar road model to solve scale ambiguity, *Proceedings of European Conference on Mobile Robots* (2011).
- [14] Momin, M., Rahman, M., Sultana, M., Igathinathane, C., Ziauddin, A. and Grift, T.: Geometry-based mass grading of mango fruits using image processing, *Information processing in agriculture*, pp. 150–160 (2017).
- [15] Fischler, M. A. and Bolles, R. C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, *Communications of the ACM*, pp. 381–395 (1981).

付 録

A.1 補正用回転行列の計算

ブドウの計数手法 [1] と同様に, Stereographic 画像の画像中心にブドウ棚を投影させるため, 回転補正を行う. ここでは, その回転補正の回転行列の推定方法について述べる.

回転補正では, カメラ座標系の z 軸と垂直にブドウ棚が存在し, かつ, z 軸の負の方向にブドウ棚が存在するよう, カメラ座標系を回転させる. 基本方針としては, 図 ?? のカメラ座標系の z 軸をまずブドウ棚平面と垂直にするための回転行列をロドリゲスの公式を元に求める.



図 7 $2m \times 2m$ の平面を切り出した stereographic 画像

スケール推定の際に得られたワールド座標系でのブドウ棚の平面 Π の式を $ax + by + cz + d = 0$ とする。また、OpenVSLAM で得られたワールド座標系でのカメラ位置 $C_w = (c_{w1}, c_{w2}, c_{w3})$ 、カメラ方向をクォータニオンで $Q = q_x i + q_y j + q_z k + q_w$ (i, j, k は虚数単位) とする。ワールド座標系でカメラ位置を通るブドウ棚平面の法線は B を媒介変数として次式のように表現できる。

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} c_{w1} \\ c_{w2} \\ c_{w3} \end{pmatrix} + B \begin{pmatrix} a \\ b \\ c \end{pmatrix} \quad (\text{A.1})$$

カメラ位置を通るブドウ棚平面の法線とブドウ棚平面との交点 P は次式ようになる。

$$P = \begin{pmatrix} c_{w1} \\ c_{w2} \\ c_{w3} \end{pmatrix} - \frac{ac_{w1} + bc_{w2} + cc_{w3} + d}{a^2 + b^2 + c^2} \begin{pmatrix} a \\ b \\ c \end{pmatrix} \quad (\text{A.2})$$

ここで、カメラ姿勢を使い交点 P をカメラ座標系に変換する。カメラ方向であるクォータニオンからカメラ座標系への回転行列 R_q とするとカメラ座標系での交点 P_c は次式で与えられる。

$$R_q = \begin{pmatrix} 1 - 2q_y^2 - 2q_z^2 & 2q_x q_y - q_w q_z & 2q_w q_y + 2q_x q_z \\ 2q_w q_z + 2q_x q_y & 1 - 2q_x^2 - 2q_z^2 & 2q_y q_z - 2q_w q_x \\ 2q_x q_z - 2q_w q_y & 2q_y q_z + 2q_w q_x & 1 - 2q_x^2 - 2q_y^2 \end{pmatrix} \quad (\text{A.3})$$

$$P_c = R_q \cdot (P - C_w) \quad (\text{A.3})$$

カメラ座標系の原点から交点 P_c への単位ベクトル P_p は次式で与えられる。

$$P_p = \frac{P_c}{\|P_c\|} \quad (\text{A.4})$$

ここで、カメラ座標系において単位ベクトルが投影方向である z 軸と重なるような回転行列を求める。これにより、ブドウ棚平面が投影方向と垂直になる。カメラ単位ベクトル P_p とカメラ座標系での $(0, 0, 1)$ との外積の単位ベクトル $n = (n_1, n_2, n_3)$ を回転軸、 P_p と $(0, 0, 1)$ とのなす角 γ を回転角として、ロドリゲスの回転公式より回転行列 R_γ は次式で与えられる。

$$n = \frac{(0, 0, 1) \times P_p}{\|(0, 0, 1) \times P_p\|}$$

$$\gamma = -\arccos P_p \cdot (0, 0, 1)$$

$$R_\gamma = \begin{pmatrix} n_1^2 \alpha + C_\gamma & n_1 n_2 \alpha - n_3 S_\gamma & n_1 n_3 \alpha + n_2 S_\gamma \\ n_1 n_2 \alpha + n_3 S_\gamma & n_2^2 \alpha + C_\gamma & n_2 n_3 \alpha - n_1 S_\gamma \\ n_1 n_3 \alpha - n_2 S_\gamma & n_2 n_3 \alpha + n_1 S_\gamma & n_3^2 \alpha + C_\gamma \end{pmatrix} \quad (\text{A.5})$$

ただし、 $C_\gamma = \cos \gamma$, $S_\gamma = \sin \gamma$, $\alpha = 1 - C_\gamma$ とする。この回転を equirectangular 画像から変換した 3次元極座標に行うことで、カメラ座標の z 軸がブドウ棚平面と垂直に交わることとなる。

また、ブドウ棚は z 軸の負の位置にあるため、ここから z 軸を 180 度回転させるよう、さらに回転を行う。この際の回転行列 R_c は、

$$R_c = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \quad (\text{A.6})$$

と表すことができる。 R_c をもって 3次元極座標上を回転させることでブドウ棚は z 軸の負の位置にあり、stereographic 画像に投影した際にブドウ棚が画像の中央に表示される。