

# 全方位カメラ画像における歪みを考慮した ブドウの房数推定

赤井 亮太<sup>1,a)</sup> 内海 ゆづ子<sup>1,b)</sup> 三輪 由佳<sup>2,c)</sup> 岩村 雅一<sup>1,d)</sup> 黄瀬 浩一<sup>1,e)</sup>

**概要:** ブドウ栽培の作業の一つとして摘房がある。摘房は、房を間引く作業であり、単位面積辺りの房数で定められる摘房基準を満たすよう房を切り落とす。栽培初心者は、摘房基準を満たしているかを確認するために房を数える必要があるが、これが大きな負担になる。そこで本研究では、栽培初心者の負担を軽減することを目的とし、画像認識技術を用いた房の計数の自動化手法を提案する。提案手法では、ブドウ棚の撮影に全方位カメラを用い、Stereographic 投影画像に対して object counting の手法を適用する。しかし、Stereographic 投影画像には特有のゆがみがあるため、透視投影画像を前提とした手法をそのまま用いることはできない。そこで、Stereographic 投影特有の歪みに合わせた切り出し画像のアラインメントを行うことで、精度の高い推定を実現する。房数推定用のデータセットを作成し実験を行った結果、アラインメントを行わないものと比較して平均絶対誤差で 9.6 % 精度が向上した。

## 1. はじめに

近年、日本のブドウ栽培は衰退傾向にある。図 1 は平成 26 年から平成 30 年までのブドウの結果樹面積、収穫量、出荷量の推移を示す。結果樹面積とは、果樹が栽培されている面積のうち、農家が意図して結実させた面積のことである。図 1 を見ると、最近 5 年で収穫量は 7.4 パーセント減、出荷量は 6.3 パーセント減、結果樹面積は 3.4 パーセント減であり、ブドウ栽培が衰退傾向にあることがわかる。原因は農業従事者の高齢化による廃園である。このような状況下で日本がブドウ栽培を続けていくには新規参入者が必要である。

しかし、ブドウ栽培には他の果樹にはない特有の作業があり、それが新規参入の障壁となっている。ブドウ栽培特有の作業で難しいものの一つに摘房がある。摘房とは、房を間引くことである。ブドウをはじめとする果樹は実を多くつけがちであるが、光合成によって得られる糖の生産能力は限りがあるため、実を出荷できる品質の糖度を保つためには実を間引いて減らす必要がある。ブドウの適正着果量は研究が進められており、新梢長・新梢数の値に対する着果量の基準の一例を表 1 に示す。この表は新梢長・新梢

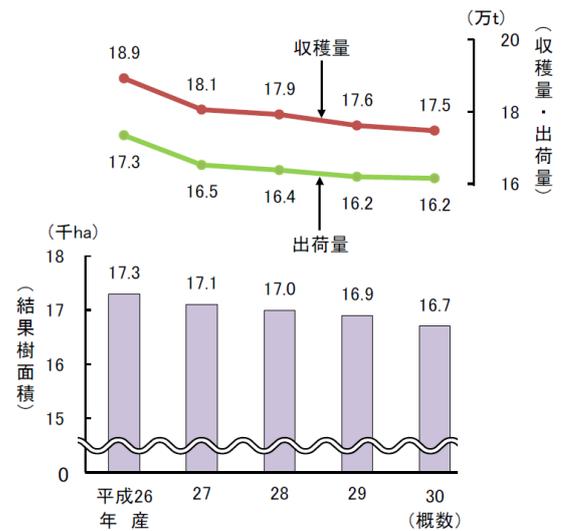


図 1 ブドウの結果樹面積及び収穫量の推移 [1]

数の値に対する着果量を基準として示している。実際にはこのような基準は簡略化され、単位面積当たりの房数が摘房の基準として用いられる。摘房では、摘房基準を満たすように房を切り落として減らしていく。摘房が新規参入者にとって難しいのは、房の数を数えながらブドウを切るマルチタスクであることが理由である。摘房作業に慣れている篤農家の場合、圃場を見ると房の大体の密度が分かるため、房を数えることなく作業ができる。しかし、初心者の場合は、房を切るのに躊躇し摘房基準を上回る量の房を残してしまう傾向にあるので、房の数を数えながら作業を

<sup>1</sup> 大阪府立大学 大学院工学研究科  
<sup>2</sup> 大阪府立環境農林水産総合研究所  
a) akai@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

行う必要がある。基準を超える房を残すと、樹全体の房の精度が落ちるため、基準に沿って房を切ることは品質管理上重要である。

そこで、本研究では摘房のうち、房を数える作業に着目し、画像認識技術を用いたブドウの房の計数作業の自動化を目的とする。摘房基準は、単位面積あたりの面積であるため、画像中の房数を数えるような技術が必要となる。また、摘房基準はある程度広い範囲に対して定められるため、広い範囲を撮影する必要がある。

画像中の物体を数えるタスクは object counting という名前前で知られており、中でも、顕微鏡画像中の細胞を数える cell counting [3], [4] や群衆を数える crowd counting [5] といった分野で研究が盛んに行われている。細胞や人ほどではないものの、収量予測のためにブドウの実を数えた研究も存在する [6] が、本研究ではブドウの房を対象とするので、この手法を適用することはできない。そのため、新たにブドウの房数を推定する手法を検討する必要がある。

Object counting には大きく分けて物体検出に基づく手法と回帰に基づく手法の2つのアプローチが取られている。物体検出に基づく手法は、対象物体のオクルージョンが少ない場合に推定精度が良くなるが、対象物体の密度が高く、オクルージョンが多い場合に精度が悪くなるのが知られている [7]。一方、回帰に基づく手法は、物体検出系の手法と対照的に、対象物体の数が少ない場合には精度が低いが、対象物体の密度が高い場合やオクルージョンが多い場合に精度が良いということが示されている [7]。本研究で数える対象となるブドウは柵状に生育されているため、房が葉の裏に隠れていることが多かったり、房の数が多といった特徴があり、密度が高くオクルージョンが多い。そのため、本研究の推定手法としては後者の回帰ベースの手法が相応しい。

また、本研究では、広範囲のブドウを撮影するために全方位カメラを用いる。しかし、全方位カメラで撮影された画像には透視投影画像と異なる歪みが存在する。物体検出分野では、全方位画像に伴う歪みに対処する研究が行われている [8], [9] が、object counting においては研究がされていない。そのため、全方位画像を用いて object counting を行うためには工夫が必要となる。

そこで、本研究では、全方位画像で撮影された画像の歪みに対応した回帰に基づくブドウの房数推定手法を提案する。提案手法では、全方位カメラで撮影された画像を Stereographic 投影画像に変換することで歪みを扱いやすくした後、歪みが画像の位置によって同じになるようにすることで房数推定モデルを学習し精度を向上させる。

## 2. 関連研究

本章では、画像中の物体を数える object counting と全方位画像に対する画像認識の研究について紹介する。

## 2.1 object counting

Object counting とは、画像中における特定物体の数を数えるタスクである。応用例としては、群衆 [10], [7], [11], [12], [13] や顕微鏡中の細胞 [3], [4], 車両 [14], [15], [16], 植物 [17], [18], [19], [20], [21] など、多くの物体に対して研究が行われている。

ブドウに対しても object counting の応用例があり、[22] や [6] などがある。[22] では、夜間に人工照明下で収穫期の赤ブドウを撮影し、果実表面に対する鏡面反射のピーク点を検出した結果を元にブドウの実を数えている。この研究では、人工的な照明環境を用いていることや数える対象が実であるため、本研究で対象とする摘房期のブドウに適用することはできない。[6] は、radial symmetry transform などを用いてブドウの実を数え、収量予測をした研究である。こちらの研究では緑色のブドウを対象としているため、摘房期のブドウと見た目は似ている。しかし、数える対象が実であるため、房の計数を目的とした本研究で用いることはできない。ブドウの房を数える方法は今まで研究されていないため、新しく方法を考える必要がある。

object counting は、主として物体検出に基づく手法と、回帰に基づく手法の2つのアプローチが取られている。物体検出系の手法は画像に対して物体検出を行い、検出できた物体の数を推定値とする手法である。物体検出系の手法は、密度が低い状況では精度がいいが、物体検出が苦手とする高密度やオクルージョンが多い状況では精度が悪くなることが知られている。もうひとつの手法である回帰系の手法では、画像と対照物体の数の関係を回帰問題とする手法である。こちらの手法は、物体検出系とは対照的に、物体の数が少ない場合には誤った推定を行う傾向にあり精度が低いが、物体の数が多く密度が大きい場合やオクルージョンが多い場合に推定精度が良いということが示されている。

## 2.2 全方位画像に対する画像認識

全方位画像の画角が広いという特徴を生かして、様々な研究が行われている。[9] や [8] では全方位画像を用いた物体検出、[23] では場所認識、[24] では顕著性マップの推定が提案されている。この中では物体検出が本研究と近い研究になるが、全方位画像に対して object counting が行われた例は我々の知る限り存在しない。

## 3. 提案手法

本章では、全方位画像を用いてブドウの房数推定を行う方法を述べる。提案手法の概要を図2に示す。提案手法では、Stereographic 投影画像を入力とする。まず、入力画像を data augmentation として画像のクロッピングを行う。クロップされた画像に対し、歪みのアラインメントを行う。最後に、アラインメント後の画像を用いて房数推定モデル

表 1 露地栽培デラウェアの 10a 当たりにおける着果量の基準 (kg)[2]

平均 新梢長 (cm)	10 m <sup>2</sup> 枠内の新梢数 (本)							
	80	100	120	140	160	180	200	220
80	1.163	1.302	1.438	1.573	1.705	1.835	1.963	2.100
100	1.228	1.406	1.546	1.701	1.851	2.000		
120	1.281	1.478	1.634	1.805	1.969			
140	1.311	1.524	1.689	1.875				
160	1.324	1.545	1.717					

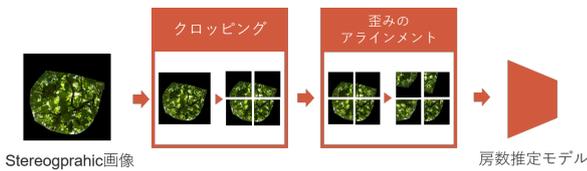


図 2 提案手法の概要

を学習し、推定結果を得るという流れである。

### 3.1 Stereographic 投影

本論文のブドウの計数には、広範囲の撮影が可能な全方位カメラを用いる。本節では、全方位カメラで撮影されたシーンがどのように 2 次元画像へ変換されるかについて説明した後、本研究で用いた投影法である Stereographic 投影について説明する。

全方位カメラで撮影されたシーンは、一般的に 2 つの投影を経て 2 次元画像に投影されるようモデル化される。まずはじめに、3 次元空間の空間の任意の点が、spherical 投影によりカメラ中心を中心とする半径 1 の単位球面に投影される。続いて、単位球面を任意の平面に投影することで、2 次元画像として投影される。単位球面から任意の平面への投影には様々な種類があり、目的によって適宜選択される。本研究では、この単位球面から平面への投影に Stereographic 投影を用いる。

本論文で用いる Stereographic 投影について、3 次元空間上の点からの 2 次元画像平面までの投影を定式化する。図 3 のように、カメラ中心  $C$  を原点とする 3 次元空間を仮定し、spherical 投影、Stereographic 投影を通じて 3 次元空間がどのように画像平面上に投影されるかを考える。はじめに、spherical 投影では、3 次元空間の任意の 3 次元点  $X = (x, y, z)$  が、カメラ中心  $C$  を中心とする単位球面上の点  $X_s$  に投影される。投影点の座標  $X_s$  は、 $r = \sqrt{x^2 + y^2 + z^2}$  とすると以下の通り表現できる。

$$X_s = \left( \frac{x}{r}, \frac{y}{r}, \frac{z}{r} \right) \quad (1)$$

続いて、単位球面上に投影された 3 次元点を任意の面に投影することで、全天球カメラの撮影シーンを 2 次元画像として表現する。Stereographic 投影は、一般的に北極  $N = (0, 0, 1)$  を投影中心とし、平面  $z = 0$  もしくは南極  $S = (0, 0, -1)$  の接平面  $z = -1$  に対して投影すると定義されている。本研究では、この定義を拡張し、図 3 に示す通り、単位球の

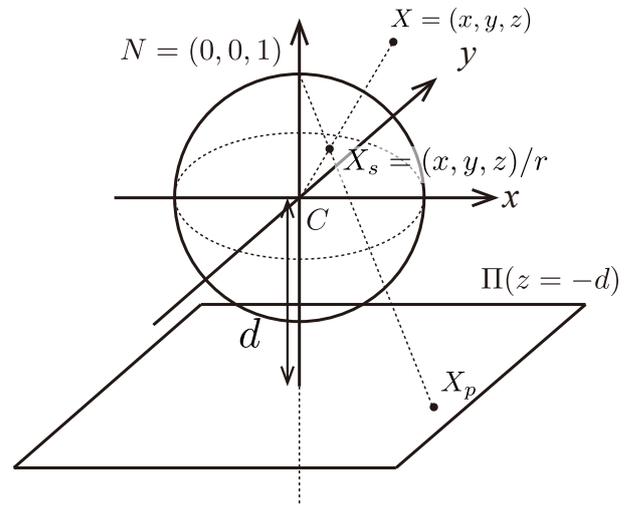


図 3 Stereographic 投影のモデル

北極  $N = (0, 0, 1)$  から、単位球面の点  $X_s$  を  $z$  軸に垂直な投影面  $\Pi(z = -d, d \geq 1)$  に投影するものとする。このとき、単位球面の点  $X_s$  を画像平面へ投影した点  $X_p$  は、

$$X_p = \left( \frac{(1+d)}{r-z}x, \frac{(1+d)}{r-z}y, -d \right) \quad (2)$$

と表される。式 (2) から、投影面と平行な面が投影された場合、 $x^2 + y^2$  の値に依存して投影の係数が決定する。そのため、平行な面と  $z$  軸との交点からの距離に応じて、ゆがみが増える画像が得られることが分かる。ブドウ棚は平面と近似できるため、この性質より、ブドウ棚と平行になるような平面に Stereographic 投影を行うと、画像は中心からの距離によってゆがみが増える。そのため、画像中心からの距離に応じて画像をそろえると、画像上の位置に応じてゆがみの程度をそろえて学習・推定が可能となり、ゆがみによる影響を抑えることが可能となる。ブドウ棚は地面とほぼ平行になっているとみなせることから、本研究では鉛直下向きを 3 次元座標の  $z$  軸の方向として Stereographic 投影を行った画像を推定に用いる。

### 3.2 Augmented データに対するアライメント

crowd counting では通常、アノテーションの困難さからデータセットが小規模になってしまうという問題がある。そのため、data augmentation の一つとして画像のクロッピングを用いて画像枚数を増やすことが良く行われている。

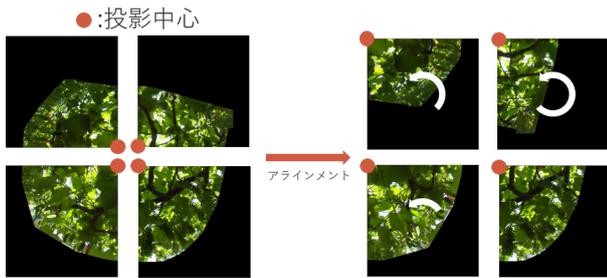


図 4 歪みのアラインメント

透視投影画像においては、画像の位置によって歪みの大きさが変わらないため、どの位置でクロップしても画像の歪みは変わらない。しかし、Stereographic 投影画像は投影中心からの距離によって歪みの大きさが変化するため、どの位置でクロップするかによって歪み方が変わり、クロップ後の画像それぞれで画像の歪み方が異なってしまうという問題が発生する。

そこで、本研究では、クロッピング後の画像に対し、歪みのアラインメントを行うことでこの問題に対処する。3.1 節で述べたように、Stereographic 画像には歪みの大きさが投影中心からの距離によって変わるという特徴がある。その特徴を利用すると、画像を回転することで歪みの大きさが変化する方向を揃えることができる。

図 4 はアラインメントを示したものである。図の例では、投影中心が画像の左上に来るように回転を行っている。回転によって、投影中心が左上に、投影中心から最も遠い点が右下になり、歪みが大きくなる方向が揃う。この一種の正規化によって、歪みが学習を妨げるのを防ぐ効果が期待できる。

### 3.3 房数推定モデル

本研究で扱うデータセットの特徴として、物体の密度が高い点、オクルージョンが多い点が挙げられる。そこで、本研究では房数推定モデルとして、現在 crowd counting において state-of-the-art であり植物の計数 [20] でも成果を出している S-DCNet を用いる。

S-DCNet[13] は Xiong らによって提案された crowd counting モデルである。crowd counting は通常回帰問題として解かれるため、学習用データセットに存在しない範囲の値を推定するのが難しいという問題があった。S-DCNet では、counting 問題は画像を分割して解けるという特徴を生かして、物体の密度が高いところは分割して物体数を推定するような手法を提案した。crowd counting における state-of-the-art であり、トウモロコシ房を数えるデータセット [20] の state-of-the-art でもある。

## 4. 評価用データセット

ブドウの房数推定の評価のためにデータセットを作成した。完成したデータセットの一例が図 5 である。データ

表 2 データセットの詳細

ビニールハウス	平均	分散	枚数
A	46.3	14.2	268
B	34.6	12.0	259

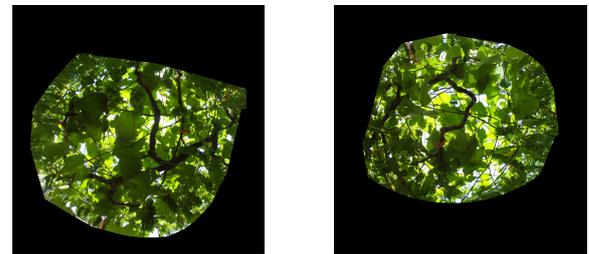


図 5 データセットの一例

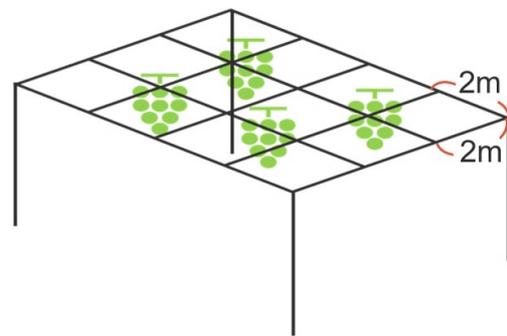


図 6 ブドウ棚の領域分割

セットは  $2,688 \times 2,688$  pixels の画像 527 枚から成り、房数の平均・分散は表 2 の通りである。以下では、画像の撮影方法と撮影後の処理について述べる。

### 4.1 撮影

撮影は大阪府立環境農林水産総合研究所の果樹園場にて行った。ブドウの品種はデラウェアである。栽培環境はビニールハウスであり、棚状に仕立てられている。2019 年 5 月 13 日に摘房前、同月 20 日に摘房後のブドウ棚を撮影した。撮影機材としては、全方位カメラの RICOH THETA S\*1 を用いた。天候は両日とも晴れだった。

撮影方法について述べる。図 6 のようにブドウ棚を  $2 \times 2 m^2$  の格子状の小領域に区切り、それぞれの小領域において数回ずつ撮影を行った。ブドウ棚上の 2 m 間隔の格子点上に赤いリボンを付けた。2 m  $\times$  2 m の小領域のなるべく中心を目測で測り、その場で撮影を行った。撮影時のブドウ棚との距離はおおよそ 0.5 – 1.5 m の間で高さを変えて撮影した。撮影は 2 つのビニールハウスについて行い、それぞれ 268 枚、259 枚の合計 527 枚画像が集まった。解像度は  $5376 \times 2688$  pixels、投影方法は equirectangular を用い、画像フォーマットは JPEG で記録した。

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

## 4.2 撮影後の処理

Theta S では、撮影結果を equirectangular 投影を用いて 2次元画像に変換し、保存する。そのため、Stereographic 投影画像を得るためには、equirectangular 投影画像からの変換が必要となる。また、3.1節で述べたように、画像のゆがみを画像中心からの距離に応じたものにするため、投影面とブドウ棚を平行にする必要がある。Theta S にはジャイロセンサが内蔵されており、撮影画像の Exif タグに撮影時のセンサの傾きの情報が記録されている。その傾きの情報をもとに鉛直下向きが  $z$  軸の正の方向になるよう回転させた後、Stereographic 投影を行い画像を得る。Equirectangular 投影画像は図 7(a) に示すように、単位球面の点  $p$  を球面座標  $p = (r, \theta, \phi)$  で表現し、図 7(b) に示すように  $\theta$  を画像の縦方向、 $\phi$  を横方向に写像することで、画像を表現する。

Equirectangular 投影画像と Stereographic 投影画像の対応を導出するため、まず、Stereographic 投影画像上の任意の点に対応する単位球面上の点を球面座標で表現する。Stereographic 画像の任意の点を  $p_s = (u_s, v_s)$  とし、 $r_s = \sqrt{u_s^2 + v_s^2}$  とする。  $p_s$  に対応する単位球面上の点  $p$  を球面座標で表すと、

$$p = (r, \theta, \phi) = \left( 1, \pi - 2 \arctan \frac{r}{d+1}, \arctan 2 \left( \frac{v_s}{r_s}, \frac{u_s}{r_s} \right) \right) \quad (3)$$

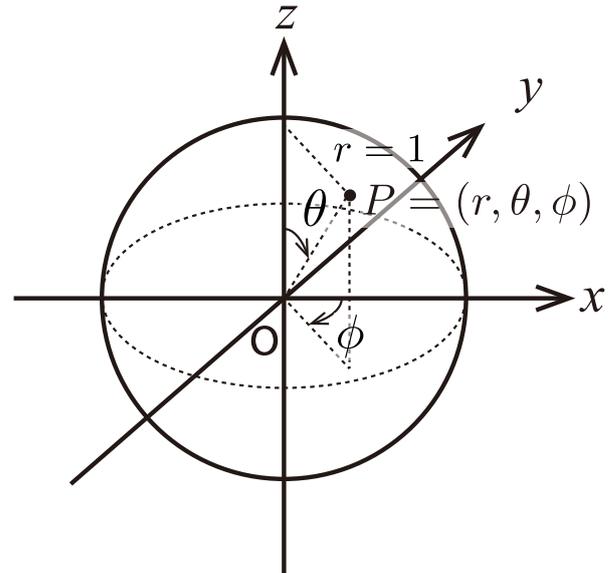
となる。ただし、 $\arctan 2(y, x)$  は値域を  $0 \leq \arctan 2(y, x) < 2\pi$  とし、以下の通り定義する。

$$\arctan 2(y, x) = \begin{cases} \arctan \frac{y}{x} & (x > 0, y \geq 0) \\ \arctan \frac{y}{x} + \pi & (x < 0) \\ \arctan \frac{y}{x} + 2\pi & (x > 0, y < 0) \\ \frac{\pi}{2} & (x = 0, y > 0) \\ \frac{3\pi}{2} & (x = 0, y < 0) \\ \text{undefined} & (x = 0, y = 0) \end{cases} \quad (4)$$

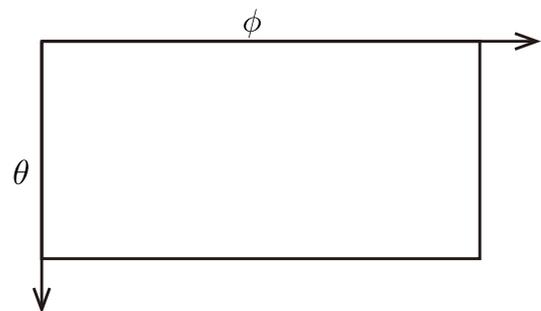
そして、これを equirectangular 投影画像の点  $p_e$  に変換する。  $w_e, h_e$  を equirectangular 投影画像の幅、高さとし、球面座標において、 $\theta, \phi$  の取り得る値がそれぞれ  $0 \leq \theta \leq \pi, 0 \leq \phi < 2\pi$  であることを考慮すると、 $p_e$  は、

$$p_e = \left( \frac{w_e}{2\pi} \phi, \frac{h_e}{\pi} \theta \right) \quad (5)$$

と表すことができる。このように対応する点を計算し、対応する画素値を写像することで、Stereographic 投影画像を得る。Equirectangular 投影画像から Stereographic 投影画像への写像の際の画素値の補完には bicubic 法を用いる。Stereographic 投影画像への変換後は、房に対してアノテーションを付与した。まず、変換後の画像に対し、 $2m \times 2m$  の格子点を表すリボンのポイントアノテーションを行っ



(a) 球面座標の定義



(b) equirectangle 画像の対応

図 7 球面座標表現と equirectangle 画像の対応座標

た。このアノテーションは株式会社 ABEJA\*2に依頼し作成した。

次に、4点のアノテーションを用いて、 $2 \times 2 m^2$  の領域を表す矩形を Stereographic 投影で描画した。領域内のブドウ房に対してバウンディングボックスのアノテーションを行った。このアノテーションも  $2 \times 2 m^2$  の領域を表す4点のアノテーションと同じく ABEJA に依頼し作成した。

全方位画像はブドウの房だけでなく地面なども映っているため、房数の推定に不要な部分が多い。そのため、房数推定に用いる部分を切り出すマスク処理を行った。切り出しに用いたマスクとしては、撮影した  $2 \times 2 m^2$  の領域と領域内の房のバウンディングボックスの凸包を用いた。

## 5. 実験

提案手法の有効性を確かめるため、作成したデータセットを用いて実験を行った。

\*2 <https://abejainc.com/ja/>

表 3 ブドウの房数推定結果

手法	MAE	MSE
ベースライン	3.41	4.54
ベースライン + 歪みのアラインメント	3.29	4.40

## 5.1 実験条件

本実験の推論用モデルとしては、S-DCNet [13] を用いた。学習に関する全てのパラメータは [13] に基づいて行った。density map 作成時の gaussian filter の  $\sigma$  は [20] に基づいて実験を行った。全画像 527 枚のうち、学習用の画像として 268 枚、テスト用として 258 枚用いた。撮影したビニールハウスによって分割した。

## 5.2 実験結果

ブドウの房数推定について実験を行った結果が表 3 である。この表においては、ベースライン手法を、歪みのアラインメントを行わずに房数推定する方法と定義する。精度を MAE で比較すると、歪みのアラインメントの導入で精度が 9.6% 向上した。このことから、歪みのアラインメントが精度向上に寄与したと言える。

## 6. まとめ

本論文では、全方位画像を用いて摘房時におけるブドウの房数を推定する手法を提案した。提案手法を評価するため、摘房前後のブドウ棚を撮影したデータセットを作成した。Stereographic 投影画像に変換し、歪みのアラインメントを行うことで、全方位画像を用いる上で避けることができない歪みの影響を低下させ、MAE で 9.6% 改善した。今後の予定としては、推定モデル自体の改善を行うことが課題である。

**謝辞** 本研究は大阪府信用農業協同組合連合会平成 31 年度産学連携研究支援事業、電気通信普及財団 2019 年度研究調査助成に基づく。

## 参考文献

[1] 農林水産省大臣官房統計部: 平成 30 年産日本なし、ぶどうの結果樹面積、収穫量及び出荷量 (2019).

[2] 高橋国昭: ブドウの適正収量に関する研究, 島根県農業試験場研究報告 = Bulletin of the Shimane Agricultural Experiment Station, No. 21, pp. 1–104 (1986).

[3] Xie, W., Noble, J. and Zisserman, A.: Microscopy Cell Counting and Detection with Fully Convolutional Regression Networks, *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, pp. 1–10 (2016).

[4] He, S., Minn, K. T., Solnica-Krezel, L., Li, H. and Anastasio, M.: Automatic Microscopic Cell Counting by Use of Unsupervised Adversarial Domain Adaptation and Supervised Density Regression, *Medical Imaging 2019: Digital Pathology*, Vol. 10956, pp. 1–8 (2019).

[5] Cenggoro, T. W.: Deep Learning for Crowd Counting: A Survey, *Engineering, Mathematics and Computer Science (EMACS) Journal*, Vol. 1, No. 1, pp. 17–28 (2019).

[6] Nuske, S., Achar, S., Bates, T., Narasimhan, S. and Singh, S.: Yield Estimation in Vineyards by Visual Grape Detection, *Proceedings of 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2352–2358 (2011).

[7] Liu, J., Gao, C., Meng, D. and Hauptmann, A. G.: DecideNet: Counting Varying Density Crowds Through Attention Guided Detection and Density Estimation, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5197–5206 (2018).

[8] Coors, B., Condurache, A. P. and Geiger, A.: SphereNet: Learning Spherical Representations for Detection and Classification in Omnidirectional Images, *Proceedings of European Conference on Computer Vision*, pp. 525–541 (2018).

[9] Yang, W., Qian, Y., Kämäräinen, J.-K., Cricri, F. and Fan, L.: Object Detection in Equirectangular Panorama, *Proceedings of 2018 24th International Conference on Pattern Recognition*, pp. 2190–2195 (2018).

[10] Li, Y., Zhang, X. and Chen, D.: CSRNet: Dilated Convolutional Neural Networks for Understanding the Highly Congested Scenes, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1091–1100 (2018).

[11] Ma, Z., Wei, X., Hong, X. and Gong, Y.: Bayesian Loss for Crowd Count Estimation With Point Supervision, *Proceedings of the IEEE International Conference on Computer Vision*, pp. 6142–6151 (2019).

[12] Wan, J. and Chan, A.: Adaptive Density Map Generation for Crowd Counting, *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1130–1139 (2019).

[13] Xiong, H., Lu, H., Liu, C., Liu, L., Cao, Z. and Shen, C.: From Open Set to Closed Set: Counting Objects by Spatial Divide-and-Conquer, *Proceedings of the IEEE International Conference on Computer Vision*, pp. 8362–8371 (2019).

[14] Guerrero-Gómez-Olmedo, R., Torre-Jiménez, B., López-Sastre, R., Maldonado-Bascón, S. and Oñoro-Rubio, D.: Extremely Overlapping Vehicle Counting, *Pattern Recognition and Image Analysis*, pp. 423–431 (2015).

[15] Mundhenk, T. N., Konjevod, G., Sakla, W. A. and Boakye, K.: A Large Contextual Dataset for Classification, Detection and Counting of Cars with Deep Learning, *Proceedings of European Conference on Computer Vision*, pp. 785–800 (2016).

[16] Amato, G., Ciampi, L., Falchi, F. and Gennaro, C.: Counting Vehicles with Deep Learning in Onboard UAV Imagery, *Proceedings of 2019 IEEE Symposium on Computers and Communications (ISCC)*, pp. 1–6 (2019).

[17] Chen, S. W., Shivakumar, S. S., Dcunha, S., Das, J. and Okon, E.: Counting Apples and Oranges With Deep Learning: A Data-Driven Approach, *IEEE Robotics and Automation Letters*, Vol. 2, No. 2, pp. 781–788 (2017).

[18] Li, W., Fu, H., Yu, L. and Cracknell, A.: Deep Learning Based Oil Palm Tree Detection and Counting for High-Resolution Remote Sensing Images, *Remote Sensing*, Vol. 9, No. 1, p. 22 (2017).

[19] Rahmehoonfar, M. and Sheppard, C.: Deep Count: Fruit Counting Based on Deep Simulated Learning, *Sensors*, Vol. 17, No. 4, p. 905 (2017).

[20] Lu, H., Cao, Z., Xiao, Y., Zhuang, B. and Shen, C.: TasselNet: Counting Maize Tassels in the Wild via Local Counts Regression Network, *Plant Methods*, Vol. 13,

- No. 1, p. 79 (2017).
- [21] Fernandez-Gallego, J. A., Kefauver, S. C., Gutiérrez, N. A., Nieto-Taladriz, M. T. and Araus, J. L.: Wheat Ear Counting In-Field Conditions: High Throughput and Low-Cost Approach Using RGB Images, *Plant Methods*, Vol. 14, No. 1, p. 22 (2018).
- [22] Font, D., Pallejà, T., Tresanchez, M., Teixidó, M., Martínez, D., Moreno, J. and Palacín, J.: Counting Red Grapes in Vineyards by Detecting Specular Spherical Reflection Peaks in RGB Images Obtained at Night with Artificial Illumination, *Computers and Electronics in Agriculture*, Vol. 108, pp. 105–111 (2014).
- [23] Wang, T.-H., Huang, H.-J., Lin, J.-T., Hu, C.-W., Zeng, K.-H. and Sun, M.: Omnidirectional CNN for Visual Place Recognition and Navigation, *Proceedings of 2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2341–2348 (2018).
- [24] Monroy, R., Lutz, S., Chalasani, T. and Smolic, A.: SalNet360: Saliency Maps for Omni-Directional Images with CNN, *Signal Processing: Image Communication*, Vol. 69, pp. 26–34 (2018).