CAD モデルを用いないばら積みシーンからの物体検出

福地 伸晃* 千葉 直也* 橋本 浩一*

Object Detection in Bin Scene without CAD Model

Nobuaki Fukuchi*, Naoya Chiba* and Koichi Hashimoto*

Bin picking robot plays an important role in automated manufacturing in industries. One of the most important techniques of bin picking robot is object detection or pose estimation in bin scene, and recently various methods reported good performance. Some industrial parts have its 3D CAD model that can represent whole shape of those parts; hence these methods are based on comparing the model of target object with the scene measured by camera or 3D sensors. However, almost all foods handled at factory do not have fixed shape against the recent demand of automating pick and place task. Therefore model-based techniques cannot be applied. In this paper, we propose a method of object detection without its 3D CAD model by utilizing target surface area information and concavity of boundaries between respective individuals. We evaluate proposed method against some foods and show the effectiveness of the method.

Key Words: Object Detection, Segmentation, 3D Point Cloud Processing, Graph Cut

1. はじめに

人件費の削減や今後の人手不足への対応として、製造業やサービス業において作業工程の自動化が求められている。とりわけ、ロボットによる自動化はソフトウェアの変更によって様々な形状の対象物に対応することができるため、その需要は大きい。例えば、製造業において乱雑に置かれた部品を一つずつとり出す「ビンピッキング作業」を行うロボットの開発が盛んに行われてきた。これはビンピッキングロボットとよばれ、ビジョンによるばら積みからの物体検出や位置・姿勢推定、把持計画や衝突回避動作計画などロボット分野における重要な要素技術の組み合わせで実現される。

ばら積み状態の物体の認識は困難な課題であり、主に工業部品を対象物にこれを解決するための研究が行われてきた [1]~[6]。具体的には、物体同士が重なっていることによる物体の部分的な隠れ（オクルージョン）や影の発生 [1] が挙げられる、また、同じ物体が多く置かれることにより物体間の境界があいまいになることや、点の周囲のテクスチャや形状情報によって計算される特徴点や特徴量 [7]~[9] の計算の際、別個体のテクスチャや形状情報が含まれモデル単体での計算と差が生じることが起きる。

これらの問題に対し Liu らは Multi Flash Camera とよばれ

る多方面から光を照射できる光源をもつカメラを用いることで、画像中で物体間の境界を際立たせ信頼性の高いエッジ（Depth Edge）を取得し、対象物体の三次元 CAD モデルからあらかじめ計算できる各姿勢におけるエッジとの比較を効率的に行うことで位置姿勢推定を達成した [1]。また、Rodrigues らは Multi Flash Camera によって得られる対象物体の様々な姿勢における各部位の反射を Random Fern [10] に学習させることで、金属部品の位置姿勢推定を達成した [2]。二次元画像ではなく三次元点群を用いた手法では、Drost らが三次元点群中の任意の 2 点の法線を用いた幾何関係の特徴量として記述する手法を提案し、三次元点群モデル中の 2 点と対応付けすることで物体検出を達成した [3]。さらに Choi らは計算する 2 点を物体の境界など限定的な領域から選ぶことで、ばら積みからの位置姿勢推定を達成した [4]。

一方で食品が対象物の場合、工業製品とは異なり対象の形状が様々、かつサイズのばらつきもあるため、従来のモデルベース位置姿勢推定や物体検出を適用することはできない。ばら積みされることによるオクルージョンや背景の煩雑化などの困難さは工業部品の際と同様に存在し、いっそうチャレンジングな課題である。

形状モデルを用いないピッキングアプローチとして堂前らの例がある [5]。これはばら積みからのピッキングを目的として個々の物体の検出そのものにこだわらず把持点を導出する手法である。野田らはこの手法を用いてばら積みからのピッキングを行い、その後持ち替えを行うことで一般形状部品の供給のロボットによる自動化を製品化した例を報告した [6]。しかし、唐揚げ

原稿受付 2017 年 10 月 3 日

*東北大学

*Tohoku University

■ 本論文は新規性（要素分野）で評価されました。

などのように一個体内に細かい凹凸やそれによるエッジを含み、さらに、単純なパラメータの曲面で表せない対象では専用のセグメンテーション手法を要することとなり、前述の手法を単純に適用することができない。また、Gori らは形状モデルを必要とせずに片面点群から精密把持する把持点を導出する手法を提案した [11]。しかし、入力する点群が把持したい目標物体のみであることが前提であり、ばら積みシーンには適用できない。適用には本論文で提案する個体のセグメンテーションが必要となる。

本研究では CAD モデルを用意することができない対象の中で特に困難である「唐揚げ」についてばら積みから一個体を検出する方法を提案する。唐揚げのみがばら積みされているという環境を利用すると、「ばら積みからの一個体の検出」という問題を「ばら積みからの一個体の領域の適切なセグメンテーション」という問題に帰着し達成できる。唐揚げは個体内で色がほぼ一様でテクスチャ情報に乏しく、かつ凹凸によるエッジを多数含むため画像処理による個体の検出は容易ではない。画像ではなく三次元点群を利用すると、ばら積みであいまになる個体間の境界を三次元形状として効果的に取得できる。しかし、個体間の局所形状のみを利用する従来手法 [12]～[14] では唐揚げ個体内の凹凸の影響で、複数個体をまとめてセグメンテーションすることや、個体の一部をセグメンテーションする失敗が起きる (Fig. 1)。そこで提案手法は、個体間の境界に現れる形状

特徴と一個体の表面積がある一定の大きさでその平均が既知であることをそれぞれエネルギー関数として定式化し、同時に最適化することでセグメンテーションを達成する。

2. 提案手法概要

ばら積みにされるのは唐揚げのみであることを考えると、あらかじめ容器などの背景を取得し除去することで唐揚げのみ含む計測データが得られる。また、ロボットによるピッキングを想定するため、ばら積みの中から上に積まれている個体（他の個体によるオクルージョンが存在しない個体）を検出することが望ましい。以上より、個体の検出という問題を「ばら積みシーンを計測した三次元点群を適切にセグメンテーションし、一番上の個体に属する部分点群を得る」という問題に帰着させる。これは、得られた三次元点群の各点について一番上の個体に属する点を前景、それ以外の点を背景とする 2 値のラベル付けをする問題となる。

唐揚げのような食品は個体ごとに形状が異なるため、基準となる三次元形状モデルを用意することができない。そのためセグメンテーションの際、下記の仮定をおく。

- (1) サイズ制約：個体の大きさはある一定範囲の大きさである。
- (2) エッジ制約：個体間の境界には空間的に大きな断絶（エッジ）が現れる。
- (3) 凹制約：個体間の境界がエッジではない場合、凹形状が現れることが多い。

三つの唐揚げが隣接して置かれており、それぞれエッジと凹形状をとっている点群を Fig. 2 に示す。提案手法ではサイズ制約と凹制約を同時に満たすようにそれぞれをエネルギー関数として設計し、グラフカット [15] によるエネルギー最小化で一個体のセグメンテーションを行う。

Fig. 3 に提案手法による個体検出のアウトラインを示す。

Fig. 1 Point cloud segmentation result (Fried chicken). For ease of viewing, three highest segmented clusters are selected and highlighted. The intensity level of points represents different clusters. (a) Locally Convex Connected Patches (LCCP) [12] [13] Segmentation. The cluster no.1 in image represents partially surface of fried chicken. The other two are segmented as individuals appropriately. (b) Constrained Planar Cuts (CPC) [14] Segmentation. The cluster no. 2 in the image are segmented appropriately. The cluster no. 1 and the cluster no. 3 in the image consists of multiple pieces of fried chicken in each cluster

Fig. 2 Edge and concave shape of three adjacent fried chicken. Left: Top view. Right: Side view

Fig. 3 Proposed detection method's pipeline. Left: Point cloud of fried chicken. Middle: Supervoxel Clusters and adjacent edge. Each point represents Supervoxel Cluster's centroid and lines represents adjacent edges. Right: Detection result. White points are detected points (foreground) and dark points are the others (background)

3. セグメンテーションによる個体検出

3.1 グラフカット

グラフカットは頂点集合 V とその頂点間の隣接関係を表す辺集合 E で構成されるグラフが与えられたとき、その頂点に対応するラベルベクトル $L = (L_1, L_2, \dots, L_{|V|})$ について、エネルギー関数

$$E(L) = \sum_{v \in V} D_v(L_v) + \sum_{(u,v) \in E} S_{uv}(L_u, L_v) \quad (1)$$

を最小化する手法である[16]。ここで $|V|$ は集合 V の要素数である。普通、各頂点 $u, v \in V$ は画像における各ピクセルや点群における各点などに対応している。また $(u, v) \in E$ のとき頂点 u, v は隣接しているとする。したがって各頂点に理想的なラベルが与えられたとき、式(1)が最小となるような問題として定式化できる。提案手法では一番上の個体を意味する「前景」とそれ以外の「背景」の二値ラベルを考えることで、セグメンテーションすることを目指す。

式(1)の第1項は各頂点 v とそのラベル L_v のみに依存する値を全頂点について足し合わせたものであり、入力画像や入力点群のデータの直接的な影響が現れるため「データ項」とよばれる[16]。提案手法ではサイズ制約を考えた際、各頂点について前景(上に積まれている個体の一部)とすべきか背景とすべきかをエネルギーとして数値化する。また、第2項は隣接する頂点間でラベルの変化を少なくする項であり、「平滑化項」と呼ばれる。我々の手法では、前景の頂点と隣接する頂点は前景である確率が高いという事前知識に基づき辺で接続された頂点間のラベルを同一にする効果を与える。ただし、凹形状をもつ頂点間は個体間をつなぐ辺であると考えラベル変化を許容するように設計する。この平滑化項が劣モジュラ性を持ち、かつ、ラベルが二値のとき式(1)についてグラフカットによって大域的最適解 $\arg \min_L E(L)$ が必ず得られることが証明されている[17]。後述するように本手法における平滑化項は劣モジュラ性をもつため、Boykov と Kolmogorov による Min-Cut/Max-Flow アルゴリズム[18]を用いて最適化を行う。

3.2 三次元点群のグラフ表現

点群を頂点集合 V と辺集合 E によって構成されるグラフで表現する。そのため、一度点群を対象物体よりも小さい領域に過分割して、その各小領域(Supervoxel Cluster)を頂点とする。Supervoxel Cluster が個体間の境界をまたいだ場合、一つの Supervoxel Cluster 内に前景と背景の点がともに存在し得る。したがって、Supervoxel Cluster の検出には、個体間の境界をまたがないよう設計されている Voxel Cloud Connectivity Segmentation (VCCS) [19]を用いた。Fig.3 中央に VCCS による Supervoxel Clusters の検出結果を示す。VCCS により点群からグラフの頂点とそれをつなぐ辺が得られる。Supervoxel Cluster とグラフの頂点は一対一対応しており、本論文では以後三次元空間を考える際「Supervoxel Cluster」、グラフを考える際「頂点」とよび分ける。また、頂点 v に対応する Supervoxel Cluster の重心座標を $\mathbf{p}_v$ 、その Supervoxel Cluster 内の点群を平面近似した際の法線(代表法線)を $\mathbf{n}_v$ と表記する。

検出された Supervoxel Cluster は元の点群との対応関係を保持しているため、各頂点について前景か背景かを判定すればすべての点の前景/背景を判定できる。

3.3 サイズ制約(データ項)

一番上に積まれている個体を前景としてセグメンテーションする際、検出する対象一つの大きさはある程度既知であり、その表面積の平均が分かっていると仮定する。これにより平均表面積を大きさの制約として利用する。各 Supervoxel Cluster の面積は同一と近似し、表面積を Supervoxel Cluster の個数を単位として考え、個体一つの表面積に相当する個数の頂点を前景として検出する。

3.3.1 グラフ距離による前景/背景らしさ

一番上に積まれている個体を検出するため、三次元空間中で重心位置が最も高い Supervoxel Cluster の頂点 v_{high} を前景としてその周囲の頂点が前景/背景のどちらかを考える。 v_{high} との距離が近い頂点は前景らしさが大きく、距離が遠い頂点は背景らしさが大きい。この距離としては三次元空間中のユークリッド距離ではなく、グラフ距離を利用する。ここで、グラフ距離 $d(u, v)$ は頂点 u, v の最短経路における辺の数である。

Fig.4(a)に示すように計測して得られる三次元点群には大きな断絶があり VCCS によって検出されるグラフの辺が接続されない場合がある。この断絶があるとき個体間であると考えられるため(エッジ制約)、グラフ距離を用いることでこの接続関係の頂点間の距離を大きくする。Fig.4(b)の点線矢印は v_{high}

(a) Illustration of point cloud and detected Supervoxel Clusters

(b) Graph distance from the v_{high} and the number of nodes H_i in each distance group

Fig.4 Detected graph from point cloud and the number of nodes H_i in each distance group

とのユークリッド距離は小さいがグラフ距離は 6 となり大きくなることが分かる．以上より，各頂点について v_{high} からのグラフ距離を考え，その距離ごとに頂点が前景／背景となるときエネルギーの増加（ペナルティ）を考える．

3.3.2 距離によるグループ作成とグループ内の表面積

グラフ距離は辺の数なので整数値として得られる． v_{high} からの各距離ごとにグループを作成し，距離 i におけるグループ内の頂点の個数を h_i として，表面積による制約を考える．前述したように v_{high} からのグラフ距離が小さい頂点はより前景らしきが大きいため，距離 i におけるグループの頂点よりも距離 $i-1$ の頂点のほうが前景らしきは大きい．したがって，距離 i の頂点を前景として検出する場合，距離 i 以下の頂点がすべて前景として検出されることを想定する．この検出される頂点数 $H_i \left(= \sum_{j=0}^i h_j \right)$ よりも，目標とする個体一個の平均表面積 μ_o のほうが大きいとき，その頂点は前景と推測できる．反対に検出される頂点数 H_i よりも目標とする個体一個の平均表面積 μ_o のほうが小さいときその頂点は背景と推測できる．

3.3.3 表面積を用いたエネルギーの設計

頂点 v を前景としたときに前景として検出される頂点数 $H_{d(v, v_{\text{high}})}$ が検出目標平均表面積 μ_o よりも小さいとき，頂点 v を前景（‘obj’）とするときのペナルティ（エネルギー増加）は 0， $H_{d(v, v_{\text{high}})} > \mu_o$ のとき，差に応じて前景とするときのペナルティを増加させる．同様に頂点 v を背景（‘bkg’）とするペナルティは， $H_{d(v, v_{\text{high}})} \leq \mu_o$ のとき差に応じて増加し， $H_{d(v, v_{\text{high}})} > \mu_o$ のとき 0 とする．ペナルティの増加は個体ごとの検出目標個体のばらつき σ_o^2 を考慮できるようにガウス分布を用いて，その対数に -1 をかけたものに従うとする．以上を数式で表現すると，データ項は

$$D_v(\text{'obj'}) = \begin{cases} 0 & (H_{d(v, v_{\text{high}})} \leq \mu_o) \\ -\log \left(\mathcal{N} \left(H_{d(v, v_{\text{high}}} \mid \mu_o, \sigma_o^2 \right) \right) & (H_{d(v, v_{\text{high}})} > \mu_o) \end{cases} \quad (2)$$

$$D_v(\text{'bkg'}) = \begin{cases} -\log \left(\mathcal{N} \left(H_{d(v, v_{\text{high}}} \mid \mu_o, \sigma_o^2 \right) \right) & (H_{d(v, v_{\text{high}})} \leq \mu_o) \\ 0 & (H_{d(v, v_{\text{high}})} > \mu_o) \end{cases} \quad (3)$$

$$\mathcal{N}(x \mid \mu, \sigma^2) = \exp \left(-\frac{(x - \mu)^2}{2\sigma^2} \right) \quad (4)$$

と定義する．ここで $\mathcal{N}(x \mid \mu, \sigma^2)$ は平均 μ ，分散 σ^2 のガウス分布である．

3.3.4 検証

ラベルが前景の場合と背景の場合のエネルギーの差を

$$G_v = D_v(\text{'obj'}) - D_v(\text{'bkg'}) \quad (5)$$

とし，その例を **Fig. 5** に示す．エネルギー最小のときに最適なラベルをとるので，式 (5) が小さいほど前景になりやすく，大きいほど背景になりやすい．また，0 のときは前景／背景でエネルギー増加は等しいため，データ項だけではラベルは定まらない．

Fig. 5 Difference of energy between forward and backward

Fig. 6 Illustration of smooth term's convexity in 2D side view. Each line indicates supervoxel cluster surface

Fig. 5 より， $\mu_o = 30$ ではある頂点 v について $H_{d(v, v_{\text{high}})} < 30$ のとき G が負の値をとり v は前景になりやすいことが分かる．反対に $H_{d(v, v_{\text{high}})} > 30$ のときに正の値をとり， v は背景になりやすい． $H_{d(v, v_{\text{high}})} = 30$ と等しいとき， $G = 0$ となるため，その頂点のラベルは隣接した頂点との平滑化項に委ねられる． $\mu_o = 40$ では， $G = 0$ となる $H_{d(v, v_{\text{high}})}$ が $H_{d(v, v_{\text{high}})} = 40$ となり，仮定した平均表面積付近の数の頂点を前景とすることが分かる．また，Fig. 5 より $\sigma_o = 5$ のときと比べて $\sigma_o = 10$ のとき $H_{d(v, v_{\text{high}})}$ と μ_o の差に対する G の増減が小さくなっている．このことは， $H_{d(v, v_{\text{high}})} > \mu_o$ の頂点が前景となることや $H_{d(v, v_{\text{high}})} < \mu_o$ の頂点が背景となったときのデータ項のエネルギー増加が小さくなることを導く．

3.4 凹制約（平滑化項）

通常ばら積みされた個体同士の境界では「凹形状」やエッジ（三次元実空間における高低差）が現れることが想定される．本手法でグラフ作成に用いる VCCS は点群が連続的に分布していない頂点間は辺で接続しないため，エッジが生じている頂点間はその辺が検出されない．よって，個体間として分離したい辺，すなわちラベル変化を許容したい辺は凹形状をもつ辺である．したがって，平滑化項は i) 頂点間にラベルの差がない場合増加するエネルギーが 0，ii) 凸形状であればラベル変化によるエネルギー増加を大きく，iii) 凹形状であれば小さくなるよう設計する．

3.4.1 凹凸形状の判定

凹凸形状は Christoph らが用いている判定法 [12] を利用する．頂点 u の Supervoxel Cluster の重心座標と代表法線をそれぞれ $\mathbf{p}_u, \mathbf{n}_u$ として，重心 $\mathbf{p}_u$ から重心 $\mathbf{p}_v$ へ方向ベクトルを $\mathbf{a}_{uv} = (\mathbf{p}_u - \mathbf{p}_v) / \|\mathbf{p}_u - \mathbf{p}_v\|$ とおく．**Fig. 6** に示すように隣接する Supervoxel Cluster 間の関係が凸であるとき，各代表法線と方向ベクトルの内積の差が大きくなり，反対に凹の

Fig. 7 Illustration of orthogonal relationship. left: Normals and direction vector are on same plane. Convexity can be computed. Normals and direction vector are orthogonal. Convexity cannot be computed

とき小さくなる。よって、隣接する Supervoxel Cluster 間の凸度を

$$f(u, v) = \mathbf{a}_{uv} \cdot \mathbf{n}_u - \mathbf{a}_{uv} \cdot \mathbf{n}_v \quad (6)$$

と定義する。これにより、Supervoxel Cluster u, v 間が凸のときは、 $f(u, v)$ が大きくなり凸度の指標として適切である。しかし、**Fig. 7** の右に示すような重心間の方向ベクトル $\mathbf{a}_{uv}$ と法線 $\mathbf{n}_u, \mathbf{n}_v$ が同一平面上にないとき、凹凸では Supervoxel Cluster 間の関係を表現できない。このように 3 ベクトルの直交度合いが大きいとき、**Fig. 7** に示すように代表法線 $\mathbf{n}_u$ と方向ベクトル $\mathbf{a}_{uv}$ のなす平面と代表法線 $\mathbf{n}_v$ の角度差が大きいので、直交度合いを

$$o(u, v) = |\mathbf{a}_{uv} \cdot (\mathbf{n}_u \times \mathbf{n}_v)| \quad (7)$$

によって評価する。隣接する Supervoxel Cluster の代表法線 $\mathbf{n}_u, \mathbf{n}_v$ と、方向ベクトル $\mathbf{a}_{uv}$ がそれぞれ垂直に近いほど式 (7) は大きくなり、直交度合いが評価できている。直交度合いが大きい頂点間も凹形状と同様に別個体であることが多いため、エネルギーの増加を小さくしラベル差を許容するよう設計したい。式 (6) と等しく評価するために、値域を $[-1, 1]$ に揃えて、かつ切断したい辺で値が小さくなるよう変換し、

$$g(u, v) = 1 - 2o(u, v) \quad (8)$$

とする。

3.4.2 凹凸形状を用いたエネルギーの設計

式 (6)、式 (8) のどちらかが小さいとき (u, v) は切断したい辺なので、このときラベル変化による平滑化項のエネルギー増加を小さくする。さらにラベルが前景の頂点と隣接する頂点は通常前景であるという事前知識を反映させるため、ラベルが同一の場合に増加するエネルギーが 0 とし、平滑化項は

$$s_{uv}(L_u, L_v) = [L_u \neq L_v] \min(f(u, v), g(u, v)) \quad (9)$$

とする。 $[\cdot]$ は条件 $\cdot$ が真のときに 1 を返し、偽のときに 0 を返す indicator function である。これによりエネルギー最小化によるラベル付けは、隣接頂点のラベルが同一になりやすく、また同一にならない辺として凹形状や直交度合いが大きい頂点間が選ばれる。したがって、データ項による「一番高い Supervoxel

Fig. 8 Soft threshold function

Cluster を中心に一定表面積の領域をセグメンテーションするだけでなく、個体間の形状特徴に沿ってセグメンテーションできる。このようなラベル変化があるときのみ平滑化項のエネルギーが増加するモデルは Potts モデルとよばれる。Potts モデルは劣モジュラ性をもつため、グラフカットにより大域的最適解が得られる [17]。

しかし、食品によっては個体内に小さな凹凸をもち、またその大きさはさまざまである。より多くの食品にも適用できるように平滑化項を **Fig. 8** に示す連続的に変化するゆるい閾値関数 [20]

$$t(x, \theta_{th}, c_{sharp}) = \frac{1}{2} \left(1 + \frac{(x - \theta_{th})c_{sharp}}{\sqrt{1 + (x - \theta_{th})^2 c_{sharp}^2}} \right) \quad (10)$$

により変換する。 θ_{th} は設定する閾値であり、 c_{sharp} は値変化率の急さを制御するパラメータである。データ項との重みも考慮し、最終的な平滑化項は

$$S_{uv}(L_u, L_v) = \beta t(s_{uv}(L_u, L_v), \theta_{th}, c_{sharp}) \quad (11)$$

と定義する。 β はデータ項に対する平滑化項の重みであり、経験的に $\beta = 10$ とし、式 (10) のパラメータは $\theta_{th} = 0, c_{sharp} = 3$ に固定した。

4. 検出評価実験

本章で提案アルゴリズムによるばら積みからの個体検出の評価実験を行う。提案手法はピックングのしやすさを考慮して、ばら積みの中で一番上に積まれている物体を検出する。よって対象物体を乱雑にばら積みにし、提案手法によって検出した結果と人手で作成した正解データを比較することによって検証する。

三次元点群の計測には **Fig. 9** (a) に示すプロジェクタ・カメラシステムによる Moreno と Taubin の手法 [21] を用いた。プロジェクタは EPSON 社製 EB-1776W (1,280 × 800 [pixels])、カメラは FLIR 社製 Flea3 (4,096 × 2,160 [pixels]) を使用した。

提案手法は唐揚げを検出するための手法として提案したため、表面積の制約を考える際にグラフ距離を用いている。したがって設計時点ではより球形に近い対象を想定しており、細長い対象を想定していない。その他の対象物に適用できるかを検証するため、弁当箱に配置の需要があり、かつ想定していなかった細長い形状のエビフライと魚フライを用いた。これら対象物はすべて形状、サイズが個体ごとに異なる。対象物を **Fig. 9** (b) に示す。

4.1 一物体検出評価実験

グラフを検出する VCCS のパラメータは $r_{\text{seed}} = 10$ [mm], $r_{\text{SV}} = 2.5$ [mm] とし, 提案手法の物体の大きさについてのパラ

(a) Experiment setting (b) Food example

Fig. 9 Experiment setting and target food

メータは唐揚げについて $\mu_o = 30, \sigma_o = 7.5$, 魚のフライについて $\mu_o = 45, \sigma_o = 10$, エビフライについて $\mu_o = 40, \sigma_o = 15$ とした. これら提案手法のサイズ制約のパラメータについては VCCS による Supervoxel Cluster 検出後, 各個体が平均して何個の Supervoxel Cluster により構成されているかで決定した. 魚フライはセンサからの見え方に応じて計測される表面積のばらつきが大きく, エビフライはサイズの個体差が大きいため σ_o を大きく設定している. 代表的な検出結果と正解データを **Fig. 10**, **Fig. 11**, **Fig. 12** に示す. それぞれ左が計測した三次元点群, 中央が検出結果, 右が人手で作成した正解データであり, 検出された個体の点群と正解点群の輝度値を大きくして表示している. 唐揚げ, 魚フライ, エビフライともに正解データとほぼ同一の点が一番上の個体として検出できていることが分かる. 正解データと比較して検出結果の点の数が少ないのは,

Fig. 10 Detection of the highest object surfaces and ground truth (Fried chicken). (a) Point cloud measured by a 3D sensor. (b) Detection of the highest object surfaces. (c) Ground truth. For (b) and (c), white points represent the highest object and dark points represent the others

Fig. 11 Detection of the highest object surfaces and ground truth (Fried fish). (a) Point cloud measured by a 3D sensor. (b) Detection of the highest object surfaces. (c) Ground truth. For (b) and (c), white points represent the highest object and dark points represent the others

Fig. 12 Detection of the highest object surfaces and ground truth (Fried shrimp). (a) Point cloud measured by a 3D sensor. (b) Detection of the highest object surfaces. (c) Ground truth. For (b) and (c), white points represent the highest object and dark points represent the others

Fig. 13 Multiple object detection results (Fried chicken). Proposed method iterates three times and intensity level of points represents the detected clusters for each iterations. (a) Multiple detection results where size constraint parameters are $\mu_o = 30$ and $\sigma_o = 7.5$. Three pieces of fried chicken are detected appropriately. (b) Multiple detection results where size constraint parameters are $\mu_o = 45$ and $\sigma_o = 7.5$. The cluster no. 3 in the image consists of two pieces of fried chicken because of the large difference between size constraint parameters and actual fried chicken's surface area

本手法の前処理として用いる色によるフィルタで対象物候補ではないとして除かれている点が存在するためであり、個体形状の認識には影響しない。三次元点群取得後、一番上に積み重ねられている個体一つを検出するのに要する計算時間は平均して 0.38 [s] である。

4.2 複数個体検出と他手法との比較

本手法は一個体の検出のみではなく、複数個体の検出も可能である。個体一つを検出後、前景として検出されたグラフの頂点を除いて逐次的にグラフカットを適用することで数個の対象物を同時に検出できる。唐揚げについて前節と同じパラメータにおける検出結果を Fig. 13 (a) に示す。計算時間の多くは前処理の色フィルタや、法線検出、VCCS によるグラフ検出であり、グラフカットによるラベル付けは平均して 2.0×10^{-4} [s] なので、数個の対象物を同時に検出した際の計算時間は一個体とほぼ同様な 0.38 [s] であった。

4.3 サイズ制約パラメータ変更による検出性能影響評価

本手法では個体検出を行う際、サイズ制約を用いて対象物の平均の大きさを仮定している。代表表面積のばらつきが大きい対象物に対する性能を評価するため、唐揚げについて異なるサイズ制約パラメータでの検出を行う。前節の個体検出パラメータ $\mu_o = 30$, $\sigma_o = 7.5$ について検出目標表面積 μ_o を個体検出が失敗するまで 5 ずつ大きくしたところ、 $\mu_o = 45$, $\sigma_o = 7.5$ のときに初めて誤検出が起きた。検出結果を Fig. 13 (b) に示す。実際の唐揚げよりも大きな目標表面積のパラメータを用いているため、2 個の唐揚げを一つとして検出している。したがって、代表表面積のばらつきが大きくなり、検出目標表面積と実際の表面積に差異が生じた場合失敗する可能性がある。

5. お わ り に

食品産業の自動化に必要な形状モデルを使用しないばら積みからの物体検出を提案した。唐揚げなどの食品は一つ一つ形状や大きさが異なり三次元形状モデルをつくることができないため、既存のモデルベース物体検出技術を適用できない。本手法はピンピックにおける一種類の物体のみがばら積みされている状況を利用し、物体検出をばら積みシーンのセグメンテーション

に帰着した。セグメンテーションは対象物体の表面積がある程度既知で、ばら積みになっている個体間の境界には凹形状が現れることが多いことを利用し、グラフカットにより達成した。実験により高速・高精度にセグメンテーションしていることを示した。

残された課題として、検出目標として設定した表面積と対象物の表面積が大きく異なる場合に検出が失敗する。したがって、センサからの見え方によって代表表面積のばらつきが大きくなるとき、複数の検出目標平均表面積を用意し、最もエネルギー最小の結果を採用するなどの対処が必要になると考えられる。

謝 辞 本研究は JSPS 科研費 JP16H06536 の助成を受けたものである。

参 考 文 献

- [1] M.-Y. Liu, O. Tuzel, A. Veeraraghavan, Y. Taguchi, T.K. Marks and R. Chellappa: "Fast Object Localization and Pose Estimation in Heavy Clutter for Robotic Bin Picking," The International Journal of Robotics Research, vol.31, no.8, pp.951–973, 2012.
- [2] J.J. Rodrigues, J.-S. Kim, M. Furukawa, J. Xavier, P. Aguiar and T. Kanade: "6D Pose Estimation of Textureless Shiny Objects Using Random Ferns for Bin-picking," Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems, pp.3334–3341, 2012.
- [3] B. Drost, M. Ulrich, N. Navab and S. Ilic: "Model Globally, Match Locally: Efficient and Robust 3D Object Recognition," Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, pp.998–1005, 2010.
- [4] C. Choi, Y. Taguchi, O. Tuzel, M.-Y. Liu and S. Ramalingam: "Voting-based Pose Estimation for Robotic Assembly Using a 3D Sensor," Proc. of IEEE International Conference on Robotics and Automation, pp.1724–1731, 2012.
- [5] Y. Domae, H. Okuda, Y. Taguchi, K. Sumi and T. Hirai: "Fast Graspability Evaluation on Single Depth Maps for Bin Picking with General Grippers," Proc. of IEEE International Conference on Robotics and Automation, pp.1997–2004, 2014.
- [6] 野田哲男, 堂前幸康, 永谷達也, 奥田晴久, 田中健一: "一般形状部品の多品種供給の自動化", 日本ロボット学会誌, vol.33, no.5, pp.387–394, 2015.
- [7] S. Salti, F. Tombari and L. Di Stefano: "SHOT: Unique Signatures of Histograms for Surface and Texture Description," Computer Vision and Image Understanding, vol. 125, pp. 251–264, 2014.
- [8] R.B. Rusu, N. Blodow and M. Beetz: "Fast Point Feature Histograms (FPFH) for 3D Registration," Proc. of IEEE International Conference on Robotics and Automation, pp.3212–3217, 2009.
- [9] D.G. Lowe: "Distinctive Image Features from Scale-invariant Keypoints," International journal of computer vision, vol.60, no.2, pp.91–110, 2004.
- [10] M. Özyual, M. Calonder, V. Lepetit and P. Fua: "Fast Key-point Recognition Using Random Ferns," IEEE transactions on pattern analysis and machine intelligence, vol.32, no.3, pp.448–461, 2010.
- [11] I. Gori, U. Pattacini, V. Tikhonoff and G. Metta: "Three-finger Precision Grasp on Incomplete 3D Point Clouds" Proc. of IEEE International Conference on Robotics and Automation, pp.5366–5373, 2014.
- [12] S.C. Stein, M. Schoeler, J. Papon and F. Wörgötter: "Object Partitioning Using Local Convexity," Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, pp.304–311, 2014.

- [13] S.C. Stein, F. Wörgötter, M. Schoeler, J. Papon and T. Kulvicius: “Convexity Based Object Partitioning for Robot Applications,” Proc. of IEEE International Conference on Robotics and Automation, pp.3213–3220, 2014.
- [14] M. Schoeler, J. Papon and F. Wörgötter: “Constrained Planar Cuts-Object Partitioning for Point Clouds,” Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, pp.5207–5215, 2015.
- [15] Y.Y. Boykov and M.-P. Jolly: “Interactive Graph Cuts for Optimal Boundary & Region Segmentation of Objects in ND Images,” Proc. of IEEE International Conference on Computer Vision, vol. 1, pp.105–112, 2001.
- [16] 石川博: “グラフカット”, 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, vol.2007, no.31, pp.193–204, 2007.
- [17] V. Kolmogorov and R. Zabini: “What Energy Functions Can Be Minimized via Graph Cuts?,” IEEE transactions on pattern analysis and machine intelligence, vol.26, no.2, pp.147–159, 2004.
- [18] Y. Boykov and V. Kolmogorov: “An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision,” IEEE transactions on pattern analysis and machine intelligence, vol.26, no.9, pp.1124–1137, 2004.
- [19] J. Papon, A. Abramov, M. Schoeler and F. Wörgötter: “Voxel Cloud Connectivity Segmentation—Supervoxels for Point Clouds,” Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, pp.2027–2034, 2013.
- [20] F. Moosmann and C. Stiller: “Joint Self-localization and Tracking of Generic Objects in 3D Range Data,” Proc. of IEEE International Conference on Robotics and Automation, pp.1146–1152, 2013.
- [21] D. Moreno and G. Taubin: “Simple, Accurate, and Robust Projector-camera Calibration,” Proc. of Second International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission, pp.464–471, 2012.

福地伸晃 (Nobuaki Fukuchi)

2016 年東北大学工学部機械知能・航空工学科卒業, 現在東北大学大学院情報科学研究科システム情報科学専攻博士前期課程に在籍。ピンビッキングロボット, 三次元点群処理の研究に従事。
(日本ロボット学会学生会員)

千葉直也 (Naoya Chiba)

2016 年東北大学工学部機械知能・航空工学科卒業, 現在東北大学大学院情報科学研究科システム情報科学専攻博士前期課程に在籍。プロジェクト・カメラシステムによる三次元形状計測および三次元点群処理の研究に従事。
(日本ロボット学会学生会員)

橋本浩一 (Koichi Hashimoto)

1990 年大阪大学大学院工学研究科博士課程修了。同年大阪大学工学部講師。1997 年岡山大学工学部助教授。2000 年東京大学大学院工学研究科助教授。2004 年東北大学大学院情報科学研究科教授。ビジュアルサーボ, 並列処理, 微生物制御, 三次元形状計測, 三次元点群処理および生物ナビゲーションのシステム科学などの研究に従事。工学博士。日本機械学会, 計測自動制御学会, IEEE などの会員。
(日本ロボット学会正会員)