

# RGB 画像からの距離画像推定による土上異物検出

寺西陸人（名城大学） 亀田陽平（昭和鋼機株式会社） 田崎豪（名城大学）

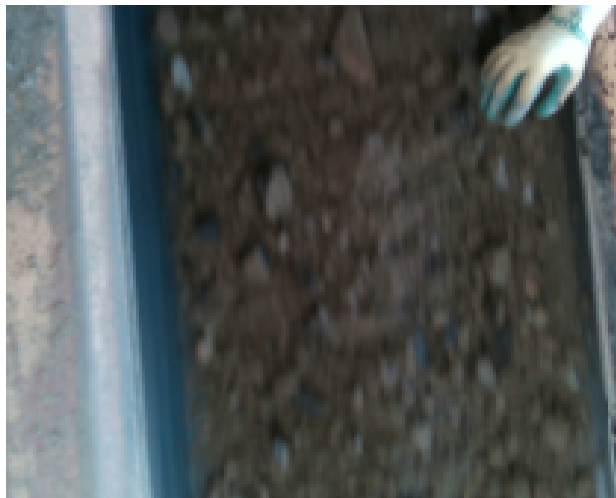


図 1 土からの異物の除去作業

## 1. はじめに

現在，工事現場などで生じた建設廃棄土を再資源化するために人間の手による異物の除去作業が行われている（図 1）．近年の除去作業員不足や長時間の単純作業による精神的苦痛といった問題からロボットによる異物除去の自動化が求められている．ここで異物とは，土として利用できない木の根や枝，ゴミのことを指し，石と同程度の数 cm サイズの物体である．ロボットによる異物除去の自動化のためには異物の形状・位置を正確に検出する必要がある．ロボットの物体検出には RGB 画像と距離画像を同期取得できる RGBD カメラがよく用いられる．RGBD カメラを用いた単純な異物検出手法として周囲より高い位置を検出する凸検出法がある．しかし，凸検出法では異物以外にも石や土の凹凸を検出してしまいロボットが異物を正確に検出できない．また，検出対象である異物はふるいで大まかに自動除去され，ふるいを通過した異物だけが人手で除去されている．したがって，人手作業場における異物のデータが少ないため大量のデータから学習を行う物体検出手法である MASK R-CNN[1] による検出は困難である．学習データに検出対象を必要としない異物検出手法として GAN (Generative Adversarial Networks) が用いられる．GAN を用いた異物検出手法としては，AnoGAN[2] や EfficientGAN[3] による手法がある．しかし，これらの手法は異物を含む画像を検出するだけで，画像のどこに異物があるのかをピクセルレベルで検出することができない．ピクセルレベルでの検出ができない場合，異物の正確な形状・位置が分からず，ロボットで把持することが困難になる．そこで，本研究では，ピクセルレベルでの異物の検出という課題に取り組む．課題解決にあたり，RGBD カメラならば，真の距離画像と，RGB 画像から GAN によって推定した

推定距離画像の二つの距離画像が取得できることに着目した．大量に学習することができる土や石のデータであれば，GAN の学習により RGB 画像から距離画像を推定することができる．逆に GAN が全く学習をしていない土や石以外の異物は RGB 画像から距離画像を正しく推定することができない．そこで，真の距離画像と推定距離画像の差分を取ることで，異物の領域を過不足なくピクセルレベルで検出する新手法を提案する．本研究の目的はピクセルレベルの異物検出を行うことでロボットに物体を把持させることである．本研究における学術的な貢献を以下に示す．

- ピクセルレベルでの異物検出を可能にした
- 一般的な異物検出法と比較して 1.3 倍高精度な把持位置推定を実現した

## 2. 関連研究

### 2.1 GAN を用いた異物検出

GAN を用いた手法として AnoGAN や EfficientGAN があるが，これらの手法はピクセルレベルでの異物検出が困難である．AnoGAN や EfficientGAN による異物検出では，まず学習用画像と類似した画像を生成するノイズを学習によって獲得する．獲得したノイズから画像を生成し，生成時の異常度によってどの画像に異物が含まれているかを判断する．この手法は，画像の生成時の異常度により，入力された画像が異常かどうかを判断する手法なので画像内のどこに異物があるかはわからない．

### 2.2 GAN による画像変換

入力としてノイズだけではなく画像を用いることで変換画像を生成する研究が盛んである．Pix2Pix[4] は画像を用いる GAN の一種である．入力された 2 枚のペアの画像間の関係を学習することで，片方の画像が入力されたとき，学習した関係からペアのもう一方の画像を生成することができる．逆に学習していない関係を持つ画像は生成できない．本研究では，一般的には失敗とされる生成できない画像を利用する新手法を開発する．

## 3. 提案手法

### 3.1 提案手法概要

提案手法の概要を図 2 に示す．入力は RGBD カメラから同期取得した RGB 画像と距離画像である．Pix2Pix による距離画像推定では土の状態によっては，土のデータでも距離推定が出来ないことがある．例えば，未学習の特徴を持つ石がある場合，石が異常として検出されてしまう場合がある．そこで，Discriminator の損失関数に正則化項を加えることでわずかに汎化性を高め，土データの距離推定を行う bCR-Pix2Pix という GAN モデルを新たに構築した．推定した距離画像と実距離

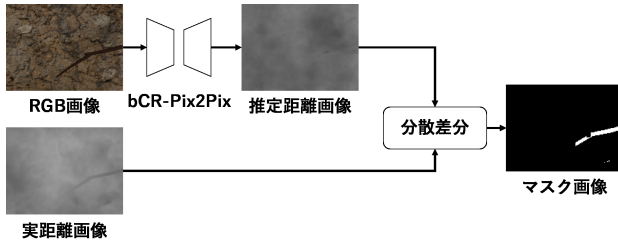


図 2 提案手法の概要図

画像の差分を取ることで異物の位置をピクセルレベルで示したマスク画像を出力する。

### 3.2 bCR-Pix2Pix

異物の距離は推定しない程度に Pix2Pix による距離推定の汎化性を向上させるため、Discriminator の損失関数に正則化項を加える bCR 構造 [5] を追加した bCR 構造を加えたものを bCR-Pix2Pix と名付ける。bCR 構造とは、Discriminator の損失関数に式 (1)、(2) で示す正則化項を加えた GAN である。

$$L_{real} = \|D(x) - D(T(x))\|_2 \quad (1)$$

$$L_{fake} = \|D(G(x)) - D(T(G(x)))\|_2 \quad (2)$$

ここで  $D(X)$  は図 3 の Discriminator のネットワーク構造に  $X$  を入力した際に最終層から取得できる  $30 \times 30$  サイズの特徴マップを示す。 $X$  は Discriminator に入力する RGB 画像と距離画像のペアであり以下の①から④の 4 種類があり、それぞれが式 (1)、(2) の  $x, G(x), T(x), T(G(x))$  を示す。

- ① RGBD カメラで取得した RGB 画像と、実距離画像のペア
- ② RGBD カメラで取得した RGB 画像と、Generator で推定した推定距離画像のペア
- ③ ①の RGB 画像と実距離画像のそれぞれに Data Augmentation (DA) を行って得られた RGB 画像と距離画像
- ④ ②の RGB 画像と推定距離画像のそれぞれに DA を行って得られた RGB 画像と推定距離画像

一般的な Pix2Pix が①と②だけを使用するのに対し、bCR-Pix2Pix では、DA を行った入力に対しても類似した特徴マップを出力するように学習させることで汎化性能を向上させる。

本研究で DA には、上下反転・グレースケール変換・欠落処理・ガウシアンフィルタ・エッジ検出・切り抜き・アフィン変換 (x 軸方向, y 軸方向, 両方)・色反転の 10 種類を用いた。bCR 構造を加えた Discriminator の損失関数を式 (3) に示す。この時、 $L_D$  は式 (4) で示すように、一般的な Pix2Pix の Discriminator で用いる損失関数を示す。 $y_n$  は入力データが実距離画像の時は 1 で推定距離画像の時は 0 である。はシグモイド関数で出力を 0 から 1 の値にする。 $a_{ij}$  は特徴マップの  $i$  行  $j$  列目の値である。

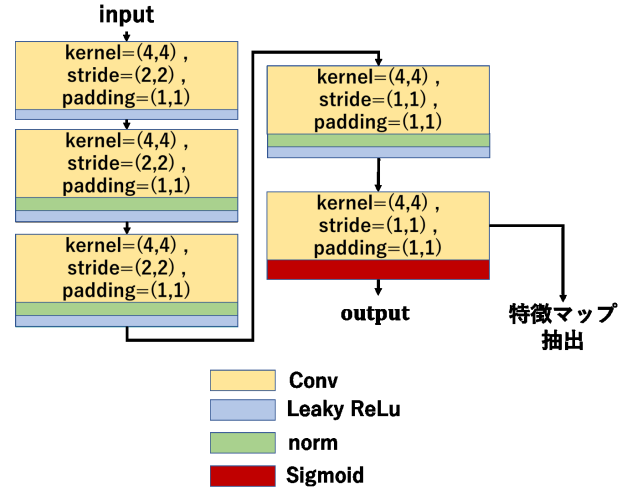


図 3 Discriminator のネットワーク概要図

$$L_{bCR} = L_D + L_{fake} + L_{real} \quad (3)$$

$$L_D = \sum_{j=1}^{30} \sum_{i=1}^{30} (-[y_n \log(a_{ij}) + (1-y_n) \log(1-a_{ij})]) \quad (4)$$

Generator の損失関数を式 (5) に示す。このとき、 $M$  は画像の総ピクセル数、 $t_n$  は推定距離画像の距離データ、 $s_n$  は実距離画像の距離データを示す。

$$L_G = \sum_{i=1}^{30} \sum_{j=1}^{30} [-\log(a_{ij})] + \sum_{n=1}^M (|t_n - s_n|) \quad (5)$$

### 3.3 実距離画像と推定距離画像の比較

異物を含まないデータで学習した bCR-Pix2Pix は、学習データに含まれていない異物の距離画像を推定することができない。そこで、RGB 画像と同期して取得した実距離画像とピクセルレベルで比較することによって異物検出を行う。各ピクセルの距離を直接使用した差分では、画像全体の距離スケールの変化に対応できず、画像全てが異物として検出されてしまう。そこで、距離の変化で比較できるよう、比較対象とするピクセル周囲の距離の分散の差分を用いた。分散差分後、モルフォロジー変換を用いてノイズ除去を行う。分散差分は  $k \times k$  ピクセルのブロック単位で行った。各ブロックが異物か否かは式 (8) によって決定する。ここで  $S_{real}, S_{fake}$  はそれぞれ実距離画像と推定距離画像のブロック内の距離データの分散値であり、それぞれ式 (6)、(7) で示される。 $P(p, q), E(p, q)$  は実距離画像と推定距離画像の各ブロックに含まれる距離データである。 $\bar{P}, \bar{E}$  は実距離画像と推定距離画像の各ブロックに含まれる距離データの平均である。 $T$  は分散差分時の閾値である。

$$S_{real} = \frac{1}{k \times k} \sum_{p=1}^k \sum_{q=1}^k (P(p, q) - \bar{P})^2 \quad (6)$$

$$S_{fake} = \frac{1}{k \times k} \sum_{p=1}^k \sum_{q=1}^k (E(p, q) - \bar{E})^2 \quad (7)$$

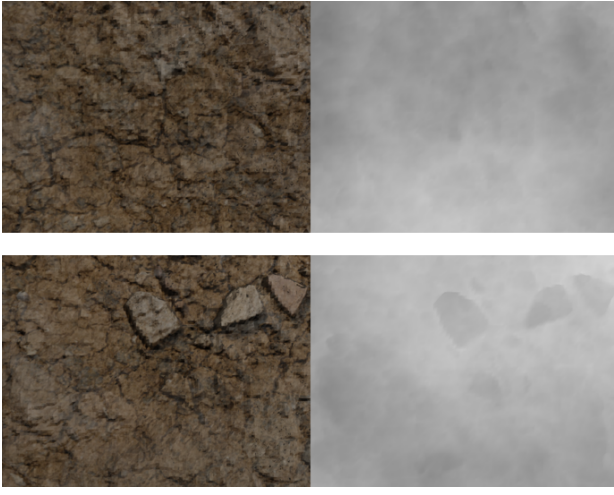


図 4 学習データ例

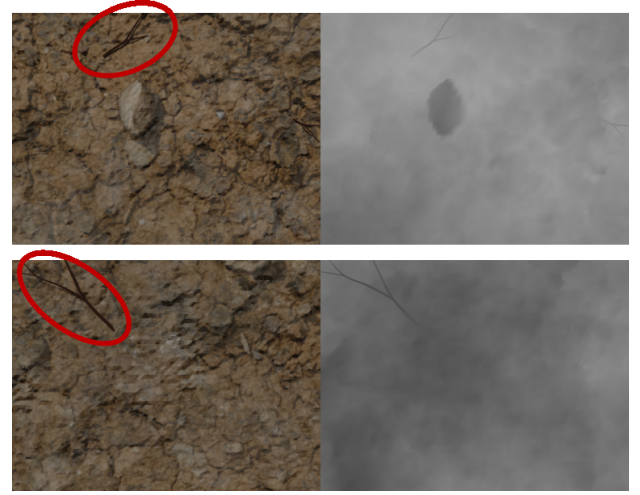


図 5 テストデータ例

$$F_{Block} = \begin{cases} 255, & |S_{real} - S_{fake}| > T \\ 0, & otherwise \end{cases} \quad (8)$$

## 4. 評価実験

### 4.1 実験設定

提案手法により，ピクセルレベルの異物検出ができることと，距離画像を推定することによる有用性を確認する．確認のために以下の 4 つの手法で比較を行う．

- 1 凸検出
- 2 Pix2Pix で RGB 画像から RGB 画像を推定した場合の異物検出
- 3 Pix2Pix で RGB 画像から距離画像を推定した場合の異物検出 (提案手法 Pix2Pix)
- 4 bCR-Pix2Pix で RGB 画像から距離画像を推定した場合の異物検出 (提案手法 bCR-Pix2Pix)

学習とテストに用いるデータは，CG を用いて作成した．学習用データセットには，土と石を 1m の高さから見下ろした際の RGB 画像と距離画像が含まれている．テストデータには異物として実際によく混入している枝を配置した．枝は土から平均 0.038m，最大 0.122m 突出している．石は土から平均 0.039m，最大 0.129m 突出している．学習用画像として 640 × 480 ピクセルの解像度を持つ 4800 枚の RGB 画像と距離画像を作成した．テスト用画像として 640 × 480 ピクセルの解像度を持つ 100 枚の RGB 画像と距離画像を作成した．また，異物の領域を白く，その他の領域を黒く塗りつぶした正解マスク画像を同時に作成した．学習データとテストデータの例をそれぞれ図 4，図 5 に示す．分散差分に用いたブロックサイズ  $k$ ，閾値  $T$  はそれぞれ実験的に 4，10 とした．

### 4.2 評価指標

本実験ではピクセルレベルでの検出の評価のため， $Recall$ ， $Precision$ ， $F$  値を用いた．また，得られた検出精度で把持可能かどうかを確認するため，把持位置推定手法 [6] による把持推定の成功率も評価指標とし

た． $Recall$ ， $Precision$ ， $F$  値はそれぞれ式 (9)，(10)，(11) によって示される．把持推定の成功率は式 (12) によって示される． $TP$  は正解画像と検出結果画像のマスク領域の重なったピクセル数を示す． $FN$  は正解画像マスク領域のうち，検出結果画像中のマスクと重なっていないピクセル数を示す． $FP$  は検出結果画像のマスク領域のうち，正解画像のマスク領域と重なっていないピクセル数を示す． $N$  はテスト画像のうち枝が撮像されている画像枚数を示し，本実験では 52 枚である． $Suc$  は把持が成功した画像枚数を示す．

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

$$Precision = \frac{TP}{TP + FP} \quad (10)$$

$$F \text{ 値} = \frac{TP}{TP + 0.5 \times (FN + FP)} \quad (11)$$

$$Auc = \frac{Suc}{N} \quad (12)$$

把持位置推定は [5] と同様に，以下の条件式 (13)(14) を満たした場合に成功したと判断する．この時， $g_p$  は推定された把持矩形， $g_t$  を正解把持矩形とする．また， $\angle g_p$ ， $\angle g_t$  はそれぞれ画像の横方向を 0 度としたときの矩形の角度である．

$$|\angle g_p - \angle g_t| < \frac{\pi}{6} \quad (13)$$

$$\left( \frac{g_p \cap g_t}{g_p \cup g_t} \right) > 0.25 \quad (14)$$

### 4.3 実験結果

実験結果を表 1 に，それぞれの手法の検出結果を図 6 に示す．また，枝の高さと把持成功率の関係性を図 7 に示す．図 7 より 60mm 以上の高さを持つ枝ならば高い精度で把持ができていることが分かる．図 6 では，検出したい枝が四角で囲まれている．図 6 右の入力には枝が存在しない．図 6 より RGB から RGB を推定しても異物は検出できず，距離画像を推定させることが



表 1 Result

	Recall[%]	Precision[%]	F 値 [%]
凸検出	46.0	16.5	24.3
RGB 推定	37.6	0.65	1.29
提案手法 (Pix2Pix)	<b>66.0</b>	45.1	53.6
提案手法 (bCR-Pix2Pix)	65.8	<b>54.1</b>	<b>59.3</b>

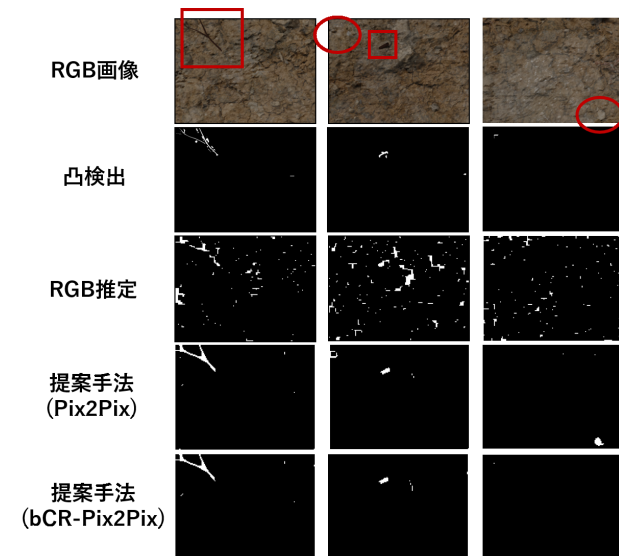


図 6 提案手法と比較手法の比較

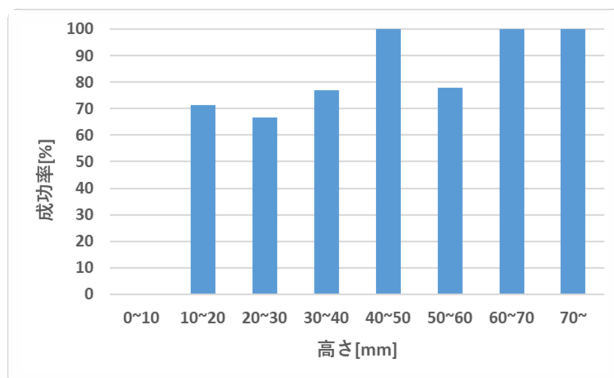


図 7 把持成功した枝の高さと把持成功率の推移

重要であることがわかる．また，表 1 より bCR の構造を加えることで，Pix2Pix で距離画像を推定するより Precision が 9 ポイント改善している．これは土部分の距離推定の精度が向上し，ノイズや誤検出による影響が減少したためだと考えられる．実際図 6 の円で示す部分の石などは bCR 構造で正しく距離が復元されるため，bCR-Pix2Pix では除去できていることがわかる．

## 5. 結論

本研究では，ロボットによる異物の把持を目的とし，ピクセルレベルでの異物検出という課題に取り組んだ．課題解決のために bCR-Pix2Pix により推定した距離画

像と実距離画像を比較することで異物検出を行う新手法を提案した．CG データによる実験の結果，提案手法は従来手法と比較し，異物検出率を示す F 値が 2.4 倍になり，把持成功率も 1.3 倍高まった．今後は，実データを用いて異物検出を行い，提案手法が実環境でも有効であることを確認する予定である．

65.4

## 参 考 文 献

- [1] K. He, G. Gkioxari, P. Dollr and R. G, "Mask R-NN," 2017 IEEE International Conference on Computer Vision, pp. 2980-2988, 2017.
- [2] S. Thomos, S. Philipp, S. W. Sebastian, E. Ursula, and L. Georg, "Unsupervised anomaly detection with generative adversarial networks to guide marker discovery." In International Conference on Information Processing in Medical Imaging, pp. 146-157. Springer, 2017.
- [3] H. Zenati, C. S. Foo, B. Lecouat, G. Manek, and V. R. Chandrasekhar, "Efficient GAN-based anomaly detection." In International Conference on Learning Representations. 2018.
- [4] P. Isola, J. Zhu, T. Zhou, A. A. (2017). "Image-to-Image Translation with Conditional Adversarial Networks." Proc. IEEE conference on Computer Vision and pattern Recognition, pp. 5967-5976, 2017.
- [5] Z. Zhao, S. Singh, H. Lee, Z. Zhang, A. Odena, and H. Zhang. Improved consistency regularization for GANs. Proceedings of the AAAI Conference on Artificial Intelligence, vol.35 No.12, pp. 11033-11041. 2021.
- [6] F. Chu, R. Xu and A. V. P., "Real-World Multi-object, Multigrasp Detection", IEEE Robotics and Automation Letters, vol. 3, no 4, pp. 3355-3362, 2018.