

移動ロボットによる変化検知のための多視点ステレオに基づく 座標整合性を考慮した参照領域抽出

Reference Region Extraction Preserving Positional Consistency

Based on Multi-View Stereo for Change Detection with a Mobile Robot

星井智仁*, 荻原佑介*, 伊賀上卓也*, ルイ笠原純ユネス*,
木下将嘉**, 笠原清司**, 伊藤裕之**, 甲田梨沙**, 田村直**, 加藤俊哉**, 野中史彦**,
神田真司*, 永谷圭司*, 浅間一*, 安琪*, 山下淳*

Tomohito Hoshii*, Yusuke Ogihara*, Takuya Igaue*, Jun Younes Louhi Kasahara*,
Masayoshi Kinoshita**, Seiji Kasahara**, Hiroyuki Ito**, Risa Koda**, Sunao Tamura**,
Toshiya Kato**, Fumihiko Nonaka**,
Shinji Kanda*, Keiji Nagatani*, Hajime Asama*, Qi An* and Atsushi Yamashita*

*東京大学, hoshii@robot.t.u-tokyo.ac.jp

**ENEOS 株式会社

概要: 本研究は, 移動ロボットを用いた石油精製プラント点検において, 変化検知のために画像対からクロップ画像対を抽出する手法を提案する. プラントでは異常領域が小さいため, 検知対象画像と参照画像から小領域を抽出し, 変化検知を行うことが有用である. そこで, 多視点ステレオに基づく 3 次元座標の整合性を考慮したクロップ画像対の抽出を行った. その結果, 画像全域でペア決めを行う手法より, クロップ画像対が同じ物体を撮像することができ, これにより高精度な変化検知が期待される.

Keywords: プラント点検, 変化検知, 深度マップ

1. 序論

燃料油やプラスチックをはじめとする, 石油製品・石油化学製品は我々の日々の生活に欠かせないものである. これらは図 1 に示すような石油精製・石油化学プラント (以下, プラント) で精製, 製造される. プラント内で装置・機器の経年劣化などにより発生する様々な異常は, 運転効率の低下や事故を誘発する. そのため, 定期的な点検によりプラント内の異常を早期に発見し, 対策を講じることが不可欠である.

異常の発見のために, 現在は運転員による巡回点検が行われている. その際, 運転員の目視により, プラント内で生じる配管の腐食, 内部流体の漏洩といった異常を発見する. しかし, 熟練者の高齢化や人手不足, 巡回点検の技術の継承が困難であるといった課題があるため, 視覚を用いた点検の自動化が期待されている.



図 1 石油精製プラント (ENEOS 株式会社)

視覚的な点検を自動化するために, これまで固定カメラを用いた異常検知手法が数多く提案されてきた. 例えば Fahimipirehgalin らは固定された赤外線カメラを用いた, フレーム間の差分画像を利用した漏洩検知手法を提案した [1]. しかし, 点検すべきプラントエリアは

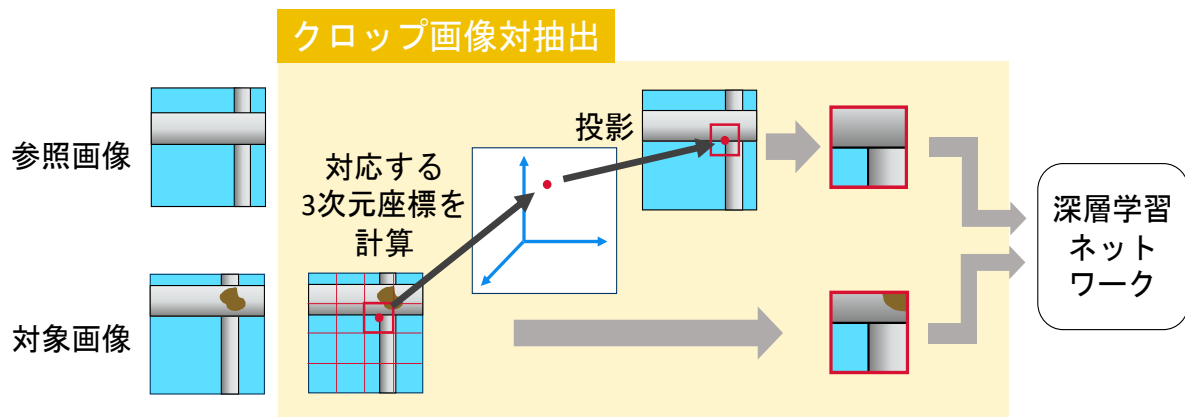


図2 提案手法の概要

広大であるため、点検に固定カメラを用いると多数のカメラが必要となり、維持管理が困難である。

Lawson ら [2] は、カメラを搭載した移動ロボットを巡回させることで、用いるカメラの台数を1台に制限した。そして、撮影された画像から正常画像を生成する Generative Adversarial Networks (GAN) を訓練し、対象画像の再構成誤差に基づき異常を検知した。Lawson らは屋内環境で実証を行っており、靴の有無やドアの開放といった異常を検知している。しかし、プラント内の配管に生じる異常は正常画像に含まれる汚れや錆と類似しており、プラント内でこの手法を用いると、異常も再現されてしまう懸念がある。

Lawson らが1枚の画像から異常の有無を検知する手法をとった一方、Shimizu らは移動ロボットを用いた変化検知による画像点検手法を提案した [3]。変化検知とは、同じ領域について異なる時点で撮影された画像を比較し、変化があった部分を検知する手法である。したがって、ある画像内に異常が含まれているか判定するために、同じ箇所を異なる時点で撮影した画像を用いる。Shimizu らは過去に撮影した異常を含まない参照動画と、現在撮影した異常を含む可能性のある対象動画を撮影し、これらから異常の判定をニューラルネットワークに基づき行った。そして、対象画像に対応する参照画像として、移動ロボットに搭載されたエンコーダに基づくカメラ位置姿勢が近く、マッチングする特徴点が多い画像を選択した。しかしながら、プラントで生じる異常により変化する画

像内の領域は小さく、画像単位で変化を検知するのは困難である。

以上から、本研究では、移動ロボットを用いたプラント内巡回点検による変化検知において、小さな異常に対応するための手法の構築を目的とする。

2. 提案手法

2.1. 提案手法の概要

プラントで生じる異常により変化する画像内の領域は小さいため、ペア決めした画像対から小さな領域を抽出し、変化検知を行うことが有用である。Shimizu らの手法による画像単位でのペア決めに基づき、同一ピクセル座標の窓枠についてクロップ画像を抽出すると、抽出したクロップ画像対に含まれる物体に大きなずれが生じてしまう。そこで本研究では図2に示すように、対象画像のクロップ画像に対応する3次元座標を計算し、参照画像内の領域を抽出する。具体的には、多視点ステレオ (Multi-View Stereo: MVS) [4] に基づくピクセルレベルの深度マップを生成し、クロップ画像が撮像している物体の3次元座標が対応するようにクロップ画像を抽出する手法を提案する。

2.2. 深度マップに基づくクロップ画像対の抽出

参照動画および対象動画からフレームを抽出し、それらのフレームが撮影されたカメラの内部パラメータと位置姿勢を Structure from

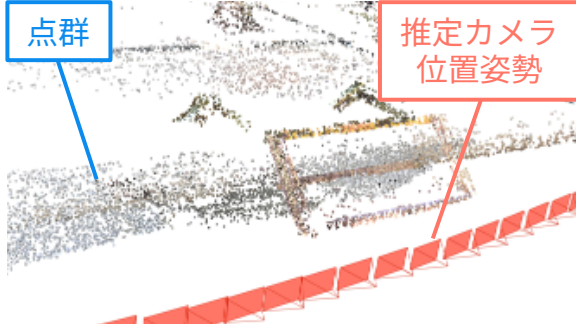


図3 Structure from Motion により推定した撮影位置姿勢と3次元点群

Motion (SfM) [5] を用いて推定する．これにより，図3のように推定したカメラの撮影位置姿勢および特徴点の3次元点群が取得される．次に，SfM を用いて得た情報をもとに，MVS を用いて深度情報を推定する．

参照動画から抽出した参照画像と，対象動画から抽出した対象画像，そしてそれぞれのカメラの撮影位置姿勢の情報から，以下の手続きでクロップ画像対を抽出する．

(1) 対象画像の抽出領域決定

対象画像内から，抽出したい長方形領域を選択する．選択した領域を対象クロップ画像として抽出する．

(2) 対象クロップ画像の代表値の計算

外れ値の影響を受けにくくするため，対象クロップ画像の深度の中央値を代表的な深度の値として用いる．この値を d_t とする．また，対象画像内における対象クロップ画像の中心のピクセル座標を (u_t, v_t) とする．

(3) 投影

図4に示すように，対象画像上のピクセル座標 (u_t, v_t) から，ワールド座標系の座標 (x_w, y_w, z_w) を経由して，対応する参照画像上のピクセル座標 (u_r, v_r) を計算する．

対象画像について，ピクセル座標 (u_t, v_t) ，深さ d_t である点について，この点の対象座標のカメラ座標系における座標を (x_t, y_t, z_t) とおく．カメラの焦点距離を f とするとき，この点について式 (1) の関係がある．

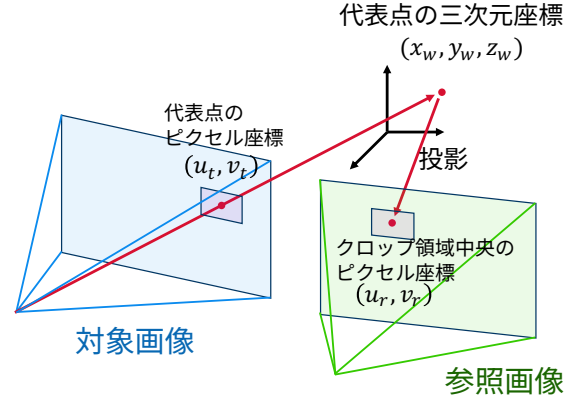


図4 投影によるクロップ領域の決定

$$s \begin{pmatrix} u_t \\ v_t \\ 1 \end{pmatrix} = \begin{pmatrix} f & 0 & c_x \\ 0 & f & c_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_t \\ y_t \\ z_t \end{pmatrix}, \quad (1)$$

ここで， $s \in \mathbb{R}$ は定数であり， (c_x, c_y) は対象画像の中央のピクセル座標を表す．また，深度 d_t は，次式 (2) で表される．

$$d_t = \left\| \begin{pmatrix} x_t \\ y_t \\ z_t \end{pmatrix} \right\|. \quad (2)$$

カメラは $z_t > 0$ の物体を撮像する．そのため， (x_t, y_t, z_t) が求められる．

次に，対象画像のカメラ座標系における座標 (x_t, y_t, z_t) をワールド座標系の座標 (x_w, y_w, z_w) に変換する．これらの間には式 (3) の関係がある．

$$\begin{pmatrix} x_t \\ y_t \\ z_t \end{pmatrix} = \mathbf{R}_t \begin{pmatrix} x_w \\ y_w \\ z_w \end{pmatrix} + \mathbf{t}_t, \quad (3)$$

ここで， $\mathbf{R}_t \in \mathbb{R}^{3 \times 3}$ ， $\mathbf{t}_t \in \mathbb{R}^{3 \times 1}$ はそれぞれ，対象画像の撮影位置姿勢を表す回転行列と平行移動ベクトルである．

次に，この関係を用いて求めたワールド座標系の座標 (x_w, y_w, z_w) を参照画像のカメラ座標系における座標 (x_r, y_r, z_r) に変換する．これらの間には式 (4) の関係がある．

$$\begin{pmatrix} x_r \\ y_r \\ z_r \end{pmatrix} = \mathbf{R}_r \begin{pmatrix} x_w \\ y_w \\ z_w \end{pmatrix} + \mathbf{t}_r, \quad (4)$$

ここで, $\mathbf{R}_r \in \mathbb{R}^{3 \times 3}$, $\mathbf{t}_r \in \mathbb{R}^{3 \times 1}$ はそれぞれ, 参照画像の撮影位置姿勢を表す回転行列と平行移動ベクトルである.

この関係を用いて求めた参照画像のカメラ座標系における座標を, 参照画像上の 픽セル座標 (u_r, v_r) に射影する. ここで, 式 (5) の関係が成り立つ.

$$k \begin{pmatrix} u_r \\ v_r \\ 1 \end{pmatrix} = \begin{pmatrix} f & 0 & c_x \\ 0 & f & c_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_r \\ y_r \\ z_r \end{pmatrix}, \quad (5)$$

ここで, $k \in \mathbb{R}$ は定数であり, (c_x, c_y) は対象画像の中央の 픽セル座標を表す.

以上より, 参照画像上の 픽セル座標 (u_r, v_r) が求まる.

(4) 参照画像のクロップ画像抽出

求めた参照画像上の 픽セル座標 (u_r, v_r) を中心とした, 対象画像のクロップ画像と同じ大きさの領域を参照画像から抽出する.

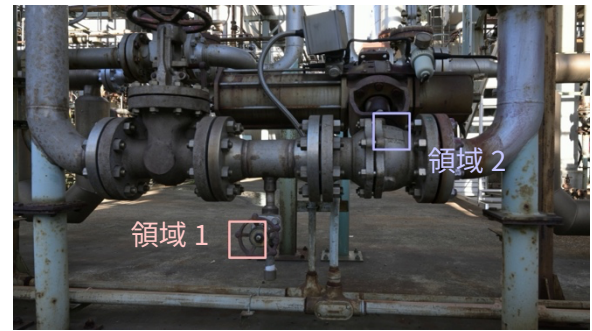
3. 検証実験

3. 1. 方法

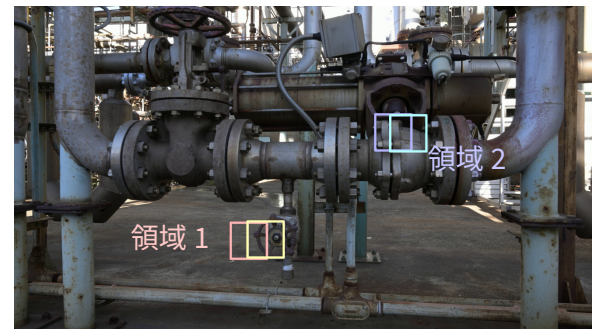
提案手法の評価のために, ENEOS 株式会社のプラントで取得した参照動画と対象動画の組を用いて手法を検証した.

参照動画と対象動画は, プラント内の同一の配管を同一の経路で撮影した動画である. 動画のフレームレートは 30 fps, 画素数は 7680×4320 픽セルであった. 対象動画から 0.5 秒おきに, 参照動画から 0.2 秒おきに画像を抽出した. 対象動画からは 100 枚の対象画像が, 参照動画からは 217 枚の参照画像の候補が抽出された.

抽出したそれぞれの画像について, COLMAP[5] を用いて SfM を行い, カメラの内部パラメータと撮影位置姿勢および 3 次元点群を推定した. また, 提案手法において, OpenMVS [4] を用いて MVS による深度情報の推定を行った. それぞれの対象画像に対して, 撮影位置間の距離が最も短い画像を参照画像とした. これにより, 100 対の画像対を決定した. これらの画像対に対して同一 픽セル座標の窓枠によるクロップおよび提案手法を用いてクロップ画像対を抽出した.



(a) 対象画像



(b) 参照画像

図 5 決定された画像対

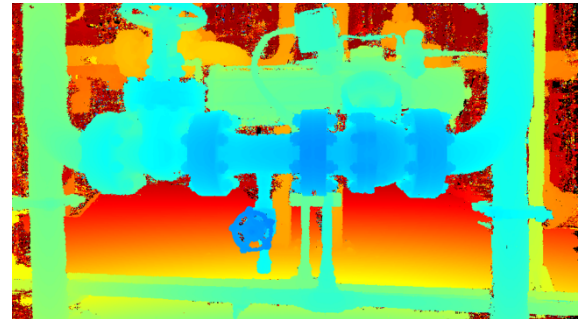


図 6 対象画像の深度マップ

対象画像から, 480×480 픽セルの窓枠によって, 対象クロップ画像を抽出した. 提案手法において, 画像の外側の領域が抽出する領域として含まれる場合がある. これらの場合を除外した, 8885 対のクロップ画像対について, それぞれ MSE および SSIM [6] を用いた類似度評価を行った.

領域 1



領域 2



対象クロップ画像

提案手法

同一座標の窓枠

図 7 抽出されたクロップ画像

3. 2. 結果と考察

決定された対象画像と参照画像の対の例を図 5 に示す. 図 5(a)に対応する対象画像の例の深度マップを図 6 に示す. 図 6 より, 対象画像から深度マップが推定できていることがわかる. 図 5 の画像から, 同一ピクセル座標の窓枠によるクロップおよび提案手法により抽出したクロップ画像の例を図 7 に示す. 図 5 のそれぞれの画像中の四角形は図 7 に示す各クロップ画像の領域を表す. 図 7 の右に示す同一ピクセル座標の窓枠によるクロップの例では対象画像の同一座標の領域を参照画像から抽出している. 同一ピクセル座標の窓枠によるクロップでは, 物体の異なる場所がクロップ画像に撮像された. その一方, 提案手法では, 参照領域の位置整合性を考慮することにより, 同じ物体が過不足なく撮影されている領域を抽出することに成功した.

抽出された領域の対象領域との一致度を定量的に評価するために, 各手法を用いて抽出した 8885 対のクロップ画像対について, MSE お

表 1 類似度評価

	MSE	SSIM
同一座標の窓枠	3574±3324	0.6019±0.1328
提案手法	1510±1970	0.7248±0.1219

よび SSIM を用いて評価を行った. SSIM について, 計算のためのガウスウィンドウの大きさは 11 とした. MSE および SSIM それぞれについて, ピクセルの R, G, B それぞれの要素について評価を行い, R, G, B それぞれについての類似度の平均値を画像間の類似度とした.

類似度評価を行った結果を表 1 に示す. 表 1 内の±の前に書かれている値は各画像間の類似度の平均値, 後に書かれている値は各画像間の類似度の標準偏差を表す. 提案手法について, MSE が小さく, SSIM が大きいことから, 空間的に近いピクセル領域を抽出したことが示された. 提案手法を用いて作成したクロップ画像対に変化検知を行うことで, 変化検知精度の向上が見込まれる.

4. 結論

本研究では多視点ステレオに基づく深度マップから、クロップ画像単位で3次元座標の整合性をもとにペア決めを行う手法を提案した。実験では、カメラ位置姿勢とピクセルレベルでの3次元座標を利用した対応付けによりクロップ画像間のずれを抑制した。今後の課題は、提案手法によるクロップ画像対抽出に基づく変化検知ネットワークの実装および有効性検証である。

参考文献

- [1] M. Fahimipirehgalin, E. Trunzer, M. Odenweller and B. Vogel-Heuser: Automatic Visual Leakage Detection and Localization from Pipelines in Chemical Process Plants Using Machine Vision Techniques, *Engineering*, vol. 7, no. 6, pp. 758–776, 2021.
- [2] W. Lawson, E. Bekele and K. Sullivan: Finding Anomalies with Generative Adversarial Networks for a Patrolbot, *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 484–485, 2017.
- [3] S. Shimizu, T. Igaue, J. Y. Louhi Kasahara, N. Yamato, S. Kasahara, H. Ito, T. Daito, S. Tamura, A. Sasamura, T. Kato, S. Kanda, K. Nagatani, H. Asama, Q. An and A. Yamashita: Change Detection in Pipe Image Pairs Extracted from Inspection Videos by Sequential Filtering, *Proceedings of the 2024 IEEE/SICE International Symposium on System Integration*, pp. 629–636, 2024.
- [4] D. Cernea: OpenMVS: Multi-View Stereo Reconstruction Library, 2020, [online] Available: <https://github.com/cdcseacave/openMVS>
- [5] J. L. Schönberger and J.-M. Frahm: Structure-from-Motion Revisited, *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4104–4113, 2016.
- [6] Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli: Image Quality Assessment: from Error Visibility to Structural Similarity, *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.