

## 7. 機械学習と画像処理を用いた重機の自動カメラ追従システム

芝浦工業大学  
芝浦工業大学  
芝浦工業大学

○ 井上 慎人  
藤武 将人  
吉見 卓

### 1. はじめに

地震や台風、火山などの災害の多い日本では建造物の倒壊、河川の氾濫や土砂災害などが多く発生する。これらの災害現場では早急な復旧工事が求められるが、危険性が高いため人が立ち入ることができない。そのため近年、重機を安全な場所から遠隔操作により復旧作業を行う、無人化施工の技術開発・導入がなされている。重機操縦者は現場に設置した監視カメラの映像を頼りに施工作业を行っている。現状、広範囲内を作業する重機を監視カメラで映し続けるため、カメラ操縦者が遠隔操作により監視カメラの向きを変えている（図-1）。

このカメラ操縦者の業務軽減及び効率化を図るため、カメラ追従の自動化が求められている。

一部の重機は高精度 GPS が搭載されており、その位置情報を元にカメラ追従を行うことは技術的に可能であるが、実際の現場ではコストの問題からすべての重機に GPS を搭載することは難しい。一方で監視カメラは遠隔操作を行うため、常に現場に設置される。そこで本研究では監視カメラ映像をもとに、画像処理を用いて重機を追従するシステムを開発し、カメラ操縦者作業の自動化を行うことを研究目的とした。



図-1 無人化施工システム

### 2. 提案手法

ここでは、目的達成に対する現状での問題点を述べた後、それら問題点を解決するための提案システムの全体像および具体的手法を順に述べる。

#### 2.1 技術的問題

まず重機を追従するためには、画面内の重機の位置を検出する必要があるが、既存の画像処理手法による追従では現場環境にある重機を検出し続けることは、以下の理由により難しい。

- ①重機は屋外で使用されるが、天候や時間帯などの環境の変化を受け、カメラに映る色合いが変化しやすい屋外での物体追従は一般に難しい。
- ②重機は旋回動作を行うが、その際、カメラに写っている面が変化するため、見た目が劇的に変化する。
- ③重機は広い範囲を移動するため、カメラとの距離が変化することで、画面サイズに対する重機のサイズが大きく変化する。

#### 2.2 システム構成

本研究では、上記問題を解決できるロバストなカメラシステムの実現のため、開発した二手法の追従システム (ATM3D、POLO) を統合しロバストなカメラ追従システムを作成した。

提案する追従システムである ATCS(Automatic Tracking Camera System)は、大別して二種類の画像処理による検出手法を統合することでロバストなカメラ追従を実現した。一方は古くから用いられるテンプレートマッチングに動体検出を加え改良した ATM3D(Automated Template Matching meets motion detection)<sup>1)</sup>。もう一方は、部位に着目し深層学習を用いた POLO(Parts-based YOLO)<sup>2)</sup>である。

①ATM3D の特徴：処理が軽快であり、通常のラップトップ PC でリアルタイムに動作させることが出来る。追従初期ほど検出精度が高い。ある程度のロバスト性が有るが、一度追従が失敗すると再び検出をすることが難しい。最初に重機の場所を教示する必要がある。

②POLO の特徴：検出精度が高く、一度検出に失敗しても再び検出することが可能でロバストである。データ処理量が多く高速に処理するためには GPU を要する。

これらを用いた ATCS の構成全体を図解したものが図-2 である。リアルタイムに追従を行うため、

現場環境に設置されたカメラ制御用 PC 内で常に ATM3D による追従を行い、かつロバスト性を高めるため、研究室に設置された GPU 搭載 PC にカメラ画像を送り POLO 検出結果を受け取る。図-3 はこれらのシステムフローを図示したものである。

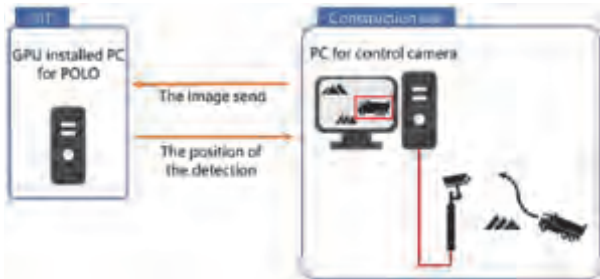


図-2 ATCSシステム図

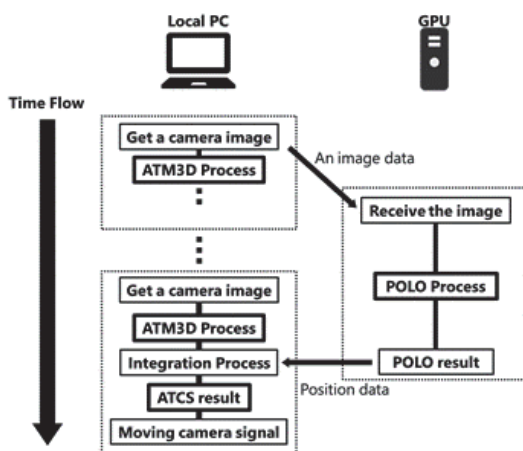


図-3 ATCSシステムフロー

### 2.3 ATM3D

ATM3D はテンプレートマッチングベースによる追従と動き検出を組み合わせたカメラ追従手法である。

まず、テンプレートマッチングに基づくカメラ追跡手法の有効性を確認した。この方法では重機を安定して追従することができるが、外観が大きく変わると失敗することがあった。これは、画像処理における物体追従手法の共通の問題である。しかし、この方法は、カメラの移動を行っている際にカメラ映像内の重機を検出し続けることを得意とする。次に、背景差分法を用いて動体検出に基づく重機の検出性能を確認した。重機がカメラ映像内を移動しているとき、重機を検出出来る。しかし、重機が停止したとき、またはカメラが動いたときに当然、重機を検出することができなくなる。これらの 2 つの方法の長所と短所を確認し、これらの検出方法を組み合わせて重機のカメラ追従を行うことにした。重機が動いているときは動体検出の結果を使い、カメラが動いているときはテンプレートマッチングベースの検出結果を使い、重機を追跡する (図-4)。

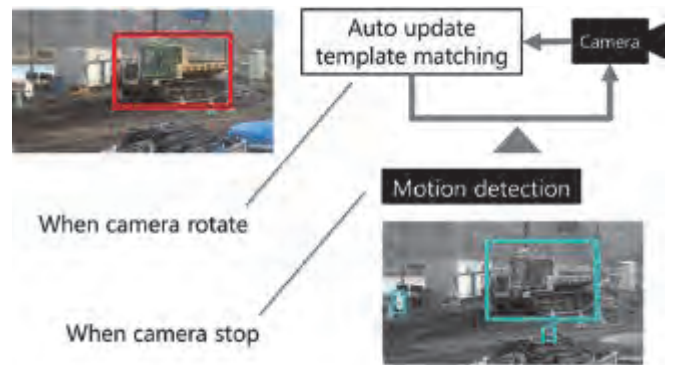


図-4 ATM3D

### 2.4 POLO

POLO は部位に着目した、機械学習による物体検出手法である。

現場では図-5 のように、カメラから重機の一部が隠れてしまうことが考えられる。一般に機械学習を用いた検出では、検出対象の一部が隠されることで検出精度が低下する。そこで POLO では図-6 のように重機の部位及び重機全体をデータセットとして学習を行っている。

POLO では機械学習による物体検出で有名な YOLOv2<sup>3)</sup> をバックボーンとし、部位情報を有効的に活用することで、オクリュージョンに強い検出を実現した。



図-5 検出が難しい場合の例



図-6 データセット例

### 2.5 ATM3D と POLO の統合

ATM3D と POLO の検出結果を統合する手法について述べる。

ATM3D は初期テンプレートを設定してから、重機の見え方の変化が少ないほど追従精度は高い。一方で、POLO は時系列データを用いていない

め変化に依存しない。また POLO は他の重機の検出や突発的な誤検出が発生する可能性がある一方で、ATM3D の検出結果は追従対象から突如大きく離れることは少ない。ゆえに ATM3D の検出結果に POLO の検出結果が類似している場合のみ POLO の検出結果を採用する。

ATM3D および POLO はいずれも重機の検出結果として画面内の重機を囲む矩形で表現されるため、この2つの矩形の類似度は IoU を用いて算出し、閾値によって ATM3D と POLO の検出結果のどちらを採用するかを決定する。その後どちらを採用したかに関わらず採用した矩形位置に応じてカメラ雲台の制御と ATM3D で用いるテンプレート画像の更新を行う。

### 3. 実験と結果

無人化施工実験場において実際の現場で使用されている重機を用いたカメラ追従実験を行った。その結果、実験環境において重機をカメラ追従できることが確認できた。この映像を用いて、既存の追従方法などと比較するため、手動で正解データを用意し、比較を行った。また、他の有名な追従手法とである Boosting<sup>4)</sup>、MIL<sup>5)</sup>、MedianFlow<sup>6)</sup>、KCF<sup>7)</sup> による追従との比較を行った。

図-7、図-8 は追従開始数秒後のフレームと最終フレームである。矩形はそれぞれの手法による検出結果を表しており、正解データ：赤色、MedianFlow：水色、KCF：青色、MIL：緑色、Boosting：白色、ATCS：黄色である。評価映像において重機を追従することが出来たのは、提案手法と Boosting のみであったことが確認できる。また、最終フレームの検出結果矩形の中心位置に着目すると、Boosting は重機から外れており、提案手法は重機の中心付近にあることがわかる。



図-7 検出結果(初期フレーム)

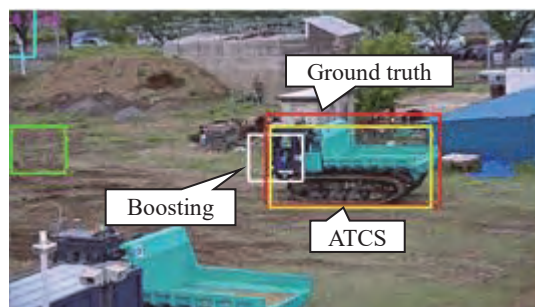


図-8 検出結果(最終フレーム)

また、これらを定量的に評価するために、評価用映像全体での「追従時間率」と「検出率」を示す二つの評価値を定義した。

二つの評価値の具体的な算出方法を述べる。正解の矩形と、検出結果として得られる矩形の、重なり度合いを示す IoU 値を、正解データのあるすべてのフレームで算出し、IoU 値が 0.7 以上のフレーム数をカウントする。これを全体のフレーム数で割ることで、「検出率」として数値化した。

この数値が高いほど、検出精度が高い状態を、維持できていた時間が長いことを意味するが、IoU 値が低くても追従することは可能なため、この数値は追従能力を完全に反映するものではない。故に、他の評価基準として重機の追従を開始してから追従ができなくなったフレームまでのフレーム数をカウントし、全体のフレーム数で割ることで、どれくらいの時間、重機を追っていたかを「追従時間率」として数値化した。追従ができなくなったフレームは正解データと検出結果の IoU 値がはじめて 0 になったフレームとした。評価値と各種法の値は表-1 のようになった。

他の手法と比較して、ATCS の2つの評価値はどちらも最も高く、提案手法の追従性能が最も高いことがわかる。しかし ATCS は通信速度に依存した遅延による影響で、検出精度が低下する箇所があることが分かった。

表-1 検出精度の比較

Method	Value	The trackable time ratio [%]	The detection ratio [%]
MEDIANFLOW		37.22	10.38
KCF		35.65	7.86
MIL		56.78	9.43
Boosting		100	23.27
ATCS		100	77.67

### 4. 統合方法の改善

ATCS の遅延問題を解決するため、POLO と ATM3D の統合手法を改良し、この問題を解決した ATCS+を開発した。図-8 が ATCS と ATCS+のシステムを比較したものである。ATCS では POLO の検出位置のみを受け取り、ATCS に統合していた。しかし ATCS+では、POLO に送信したフレーム画

像を記憶しておき、POLO から返ってきた検出位置をもとに送信時のフレーム画像から重機の画像を切り出す。次のフレームにおいてこの画像をテンプレートとしてテンプレートマッチングを行う。つまりカメラの移動により重機の位置座標は急激に変化する可能性があるのに対し、カメラが動いても重機の見た目は急激に変化しないことを利用し、遅れの問題を解決した。

カメラが動き、ATCS と ATCS+の検出結果の差が顕著なフレームを図-9、図-10に示した。青矩形が ATCS、黄矩形が ATCS+を表している。また評価値を比較したものを表-2 に示した。これらの結果から、黄矩形の方が重機全体を正しく検出しており、遅れの問題を解決できたことが確認できた。

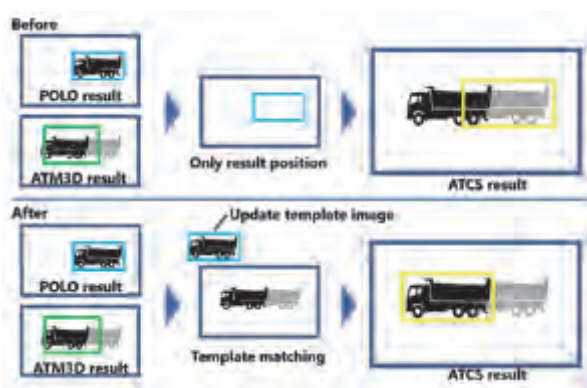


図-8 改善システム



図-9 ATCSとATCS+の比較例1

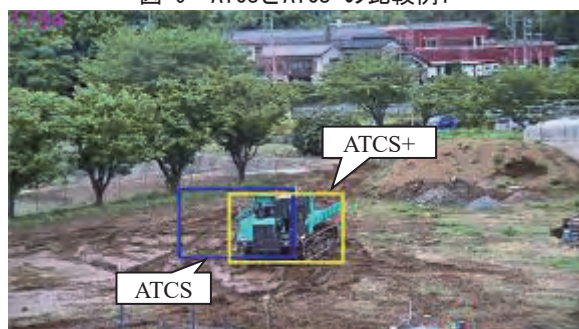


図-10 ATCSとATCS+の比較例2

表-2 改善手法の比較

Method	Value	The trackable time ratio [%]	The detection ratio [%]
ATCS		100	77.67
ATCS+		100	82.70

## 5. 結論

本稿では画像処理の適応が難しい、無人化施工現場での物体追従システムの構築に取り組み、通常の画像処理と機械学習に基づく二種類のアルゴリズム (ATM3D、POLO) を統合した新たなカメラ追従手法を提案した。画像処理を改良して提案した ATM3D ではテンプレートマッチングと動き検出を用いて、現場環境に適した、処理の軽い追従が可能である。また、機械学習側の検出アルゴリズムでは、既存のニューラルネットワークモデルを拡張した Part-based YOLO モデルが重機をよりロバストに検出可能である。実際の現場実験では、実用的なリアルタイム処理を維持し、かつ提案した統合手法がロバストであることを確認した。また、ネットワーク遅延とカメラの方向転換によって引き起こされる問題を解決し、よりロバストなカメラ追従を提供可能な改善手法とした。

## 参考文献

- 1) Makito Inoue and Takashi Yoshimi: Automatic Tracking Camera System for Construction Machines by Combined Image Processing, 35th International Symposium on Automation and Robotics in Construction (ISARC), 2018
- 2) Masato Fujitake, Takashi Yoshimi, "Estimation System of Construction Equipment from Field Image by Combination Learning of Its Parts", Asian Control Conference (ASCC), 2017
- 3) Erhan, D., Szegedy, C., Toshev, A., Anguelov, D.: Scalable object detection using deep neural networks. In: CVPR. (2014)
- 4) H Grabner, M Grabner, and H Bischof. Real-time tracking via on-line boosting. Proceedings of the 17th British Machine Vision Association 2006; September 4-7; Edinburgh, British.
- 5) Boris Babenko, Ming-Hsuan Yang, Serge Belongie. Visual Tracking with Online Multiple Instance Learning. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR) 2009; June 20-25; Miami Beach, Florida.
- 6) Z.Kalal, K. Mikolajczyk, J. Matas. Forward-Backward Error: Automatic Detection of Tracking Failures. Proceedings of the 20th International Conference on Pattern Recognition; 2010 Aug 23-26; Istanbul, Turkey.
- 7) João F. Henriques, Rui Caseiro, Pedro Martins, et al. High-Speed Tracking with Kernelized Correlation Filters. Journal of IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015;37(3):583-596.

※本研究は、一般社団法人日本建設機械施工協会の平成29年度研究開発助成を受けて実施したものです。また、実験の実施においては、(株)熊谷組に御協力をいただきました。