

# 建設現場の安全性と生産性向上を支える AI 画像・自然言語解析技術

## ヒューマンセンシング (Human Sensing) AI の適用

園 田 亜斗夢

建設現場の安全性と生産性向上には、データから得られる情報を必要な人に必要な形で提供する必要がある。各社に蓄積されている過去データの活用には、目的に応じた構造化が重要である。また、撮影したカメラ映像は目視確認するには膨大な時間がかかるため、そのデータを効率的に活用するためには、画像解析技術を用いて構造化する必要がある。本稿では、自然言語処理や画像解析技術を活用し、建設現場のデータを構造化している事例を紹介する。

キーワード：人工知能、画像解析、距離推定、顔向き、動作解析、自然言語処理、安全性、生産性、労働災害

### 1. はじめに

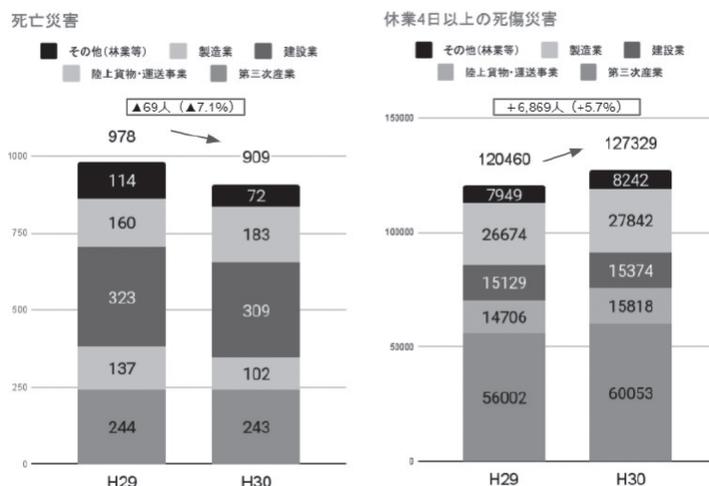
経済産業省が2018年12月に「DX推進ガイドライン」を発表し、産業界でDX化が急務とされている。建設業界でもAIの活用やDXの推進が進められており、自動化や省力化、モニタリングなど様々な取り組みが行われている。一方で、完全な自動化には多くの障壁が残されており、人間による作業が多く残されている。

日本における平成30年の労働災害による休業4日以上の死傷者数は127,329人であり、死亡者数は909人である。その内、建設業界における労災が占める件数は、休業4日以上の死傷者数は15,374人、死亡者数は309人である(図-1)<sup>1)</sup>。

近年、技術革新により死亡災害は減少傾向にあるが、休業4日以上の死傷災害は増加傾向にある。国土交通省が2016年から取り組みを行っている建設現場のICT推進策であるi-Constructionでも、安全性の向上を目指すべきものとして指定している(図-2)<sup>2)</sup>。

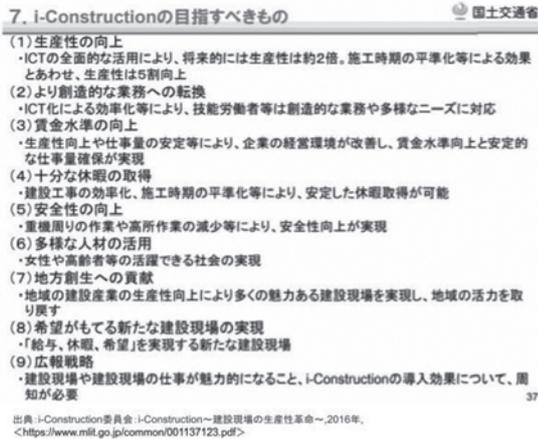
建設現場の安全性向上と生産性向上を支援するにあたり、実際の業務に携わっている建設事業者複数社にヒアリングした結果、転倒・転落による事故と機械と作業員の接触による事故に対する課題意識が強いことがわかった。また、業務効率化では作業の可視化の要望が大きいことがわかった。

転倒・転落による事故は階段や足場など危険エリアが特定されており、危険予知活動など作業員の意識



出典：平成30年の労働災害発生状況 [https://www.mhlw.go.jp/stf/newpage\\_04685.html](https://www.mhlw.go.jp/stf/newpage_04685.html)

図-1 平成30年の労働災害発生状況(厚生労働省統計より当社作成)



図一 2 i-Construction の目指すべきもの (国土交通省資料より抜粋)

向上策が実施されているが、該当エリアでの作業は今でも実施されており、経験が浅い作業員によるヒヤリハット事象が多く発生していることがわかった。また、機械と作業員の接触事故は、機械のオペレータが作業員に気づかず接触する例や、機械で運搬中の資材が落下することで発生する事故があることがわかった。このような接触事故は、事故発生件数は転倒・転落より少ないが、発生した際の影響は大きくなることが多い。

作業の可視化では、現場ごとの作業サイクルの可視化と作業員ごとの生産性の可視化の要望があることがわかった。

これらの安全性向上と生産性向上の要望を実現するために、危険情報を構造化し、その情報を適切に届ける技術と、機械と人の動きを把握し危険状態を把握し、また作業内容を可視化する技術の提供を目指している。具体的には、危険情報の構造化のために過去の事故事例に対し言語処理による情報抽出を行い、類似の現場での事故情報の提供を目指している。また、画像解析により危険な状態の検知や作業内容の可視化を行い、リアルタイムのアラート発報、情報の一元管理・提供を目指している。

## 2. 過去の事故事例の構造化・管理

建設現場での事故は休業4日以上災害の場合は労働基準監督署に報告がなされ、また、それより小さい事故であっても社内で蓄積されていることが多い。一方で、これらの情報は現場や事業部、支社ごとに形式が異なっていることも多く、また、自由書式で記述されていることも多い。そのため、危険予知活動を行う際に現場ごとの担当者の経験に依存している場合が多く、過去の経験を活用できていない例が多い。

そこで、過去の事故事例を構造化し、類似の工事の際にそれらの事故を検索、予防策を立てることができれば、事故防止につながるだけでなく、効率的な工事計画の立案が行えると期待される。

自然言語領域では2018年にBERTと呼ばれる自然言語処理のモデルをGoogleが発表して以降、決められたタスク設定であれば人間を超える精度が出ている(図一3)<sup>3)</sup>。本取り組みでは、建設現場の事故事例の構造化に対し、質問応答としてタスク設定することで、これらの技術を適用し、必要な情報を管理・検索が行えるようになった。

BERTモデルやその後継技術の特徴として大量の言語データによる事前学習と、タスクごとのファインチューニングが行えることがある。本取り組みで用いたモデルは、過去の大量のニュースデータとwikipediaのデータをもとに事前学習を行い、独自に用意した日本語のQAデータセットによりファインチューニングを行った。これらのモデルを事故データに適用することで、構造化されていなかった事故発生状況を構造化することが出来る。以下に質問と抽出結果の例を示す。

どのような状況で事故が発生しましたか？

- a. 確認作業中に〇〇ステップを登っている最中に
- b. 〇〇組立作業中に、〇〇車両から転落してなぜ切創しましたか？(切りましたか?)
- c. 〇〇を切断中勢い余って
- d. カッターの刃がしまわれていないの気づかず

このように構造化することで、傷病部位や建設現場名だけでなく、利用する設備や車両、作業内容などの軸で検索を行うことが可能になる。また、この質問応答モデルは、自由書式での検索が行えるため、構造化だけでなく検索を直接行うことも可能である。

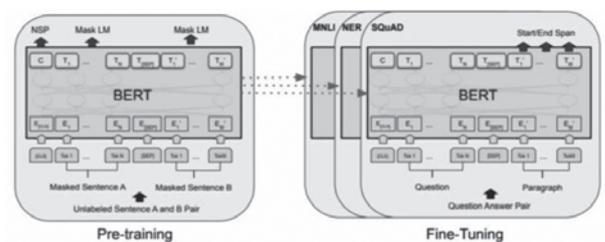


Figure 1: Overall pre-training and fine-tuning procedures for BERT. Apart from output layers, the same architectures are used in both pre-training and fine-tuning. The same pre-trained model parameters are used to initialize models for different down-stream tasks. During fine-tuning, all parameters are fine-tuned. [CLS] is a special symbol added in front of every input example, and [SEP] is a special separator token (e.g. separating questions/answers).

出典: Jacob Devlin et al.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, 2018

図一 3 BERT の学習モデル図

### 3. 建設現場の状況を可視化する画像解析技術

建設現場における危険な状態の検知や作業内容の可視化には、カメラを設置して画像解析を行うことが有効である。解析対象は、建造物、資材、機械、作業員である。特に人が関係する場所で事故が発生するため、人の解析は重要である。一方で、建設現場で見られる資材や機械は、日常生活では見られないものが多く、公開データ等の学習用データが入手しづらいという特徴がある。人にフォーカスした画像解析技術 Human Sensing 技術と、建設現場における画像解析の特徴を以下に紹介する。

#### (1) 画像解析技術による構造化

近年、建設現場へのカメラの導入が進んでいる。防犯用途や記録用途など複数の用途があるが、カメラの低廉化やビデオマネジメントシステムの普及や建設事業者の意識の変化が要因として考えられる。一方で、現状のカメラ映像は現場の事務所でのみ確認できるものや、録画を行っていないものなど十分に管理されているとは言えないものも多く、また、クラウド等に記録されているものも、問題が起こった際の確認用という位置づけのものも多い。これは、映像データという性質上、倍速等で見たとしても確認に時間を要することが要因である。そこで、映像データを画像解析技術により構造化し、現場の状況を人間が瞬時に理解できる形で提供することは、データ活用や業務効率化の観点で非常に重要である。

このような目的には、解釈性の高い分析が重要であ

る。表—1のように、現場ごとにほしい情報を定義した際に、それを実現するための画像解析モジュールを組み合わせることでシステムを実現することで、最終的なアウトプットの基準を変更することが可能になり、目的に応じた情報の収集を行うことが出来ると期待される。

表—1 アプリケーション例

撮影対象	解析対象	出力・判定事項
トンネル工事映像	使用機材・作業分類	作業進捗 ○/×
建設機械周辺映像	人物検出・動作解析	安全/不安全
戸建住宅建設映像	人物検出・位置推定	感染症リスク ○/×

#### (2) Human Sensing の特長

Human Sensing 技術とは、人にフォーカスした画像解析技術であり、姿勢推定、物体検出、セグメンテーション等の画像解析技術を組み合わせ、時系列解析を行うことで、人の状態の把握を実現する技術である(図—4)。

技術領域としては画像解析、対象は人にフォーカスすることで、建設現場だけでなく製造業・インフラ・小売等の分野での開発ノウハウやデータ蓄積を行うことが出来る。これは、良品不良品の検出などと異なり、対象を人にフォーカスすることで、異なるフィールドのデータであっても、画像的な特徴が大きく変化しないことを担保できるからである。そのため、個別のフィールドに対する人の行動の調査・評価について、少ない教師データで短期間に行うことが可能である。

Human Sensing 技術は複数のアルゴリズムを組み合わせることで、機能が実現される。各機能はレベルごとに



図—4 Human Sensing の主な機能と特長

モジュール化されており、適用するフィールドごとに適切なモジュールを選択するが、大きなカスタマイズが必要なのは発見したい動作の定義部分のみである。

まず、最も基礎的なモジュールとして、物体検出や姿勢推定、顔認識、顔向き推定、セグメンテーションなどの機能をもったモジュールが存在する。これらのモジュールは各機能ごとに複数のアルゴリズムと独自の学習データにより学習済みモデルが存在する。これらのモデルの学習には数千～数十万枚と行った大規模なアノテーション付きの教師データが必要であるが、人物の形や姿勢はフィールドごとに大きく変化しないため、汎用的なモデルとして、創業以来蓄積したデータによって学習を行っている。各モジュールは、利用する環境（クラウドの高性能のコンピュータかエッジマシンのような性能が限定されるコンピュータか）や、同時に処理するカメラ台数によって、精度と処理速度のパフォーマンスのバランスを考慮しアルゴリズムが選定される（図—5）。

次に、基礎的な機能をもとにした応用的なモジュールとして、速度計測や入退場検出、トラッキング、汎用的な姿勢推定（しゃがんでいるか立っているか等）などの機能を持ったモジュールが存在する。これらは、適用するフィールドやカメラの画角（高さ、角度）ごとにパラメータ等はセミカスタマイズを行う。

そして、フィールドごとにこれらのモジュールを組み合わせ、必要な機能を持ったアルゴリズムを開発する。この最終的な組み合わせの課程では、フィールドごとの教師データをもとに、人手でパターンを作成したり、統計的機械学習やディープラーニングにより判定パラメータを決定する。

以上のように、「基礎アルゴリズム」「応用アルゴリズム」「カスタマイズアルゴリズム」と分けることに

より、フィールドごとの新規開発要素を減らし、既存のモジュールを活用した開発を行うことで、短納期・教師データが少ないといった条件で精度の高いアルゴリズムの作成を可能としている。

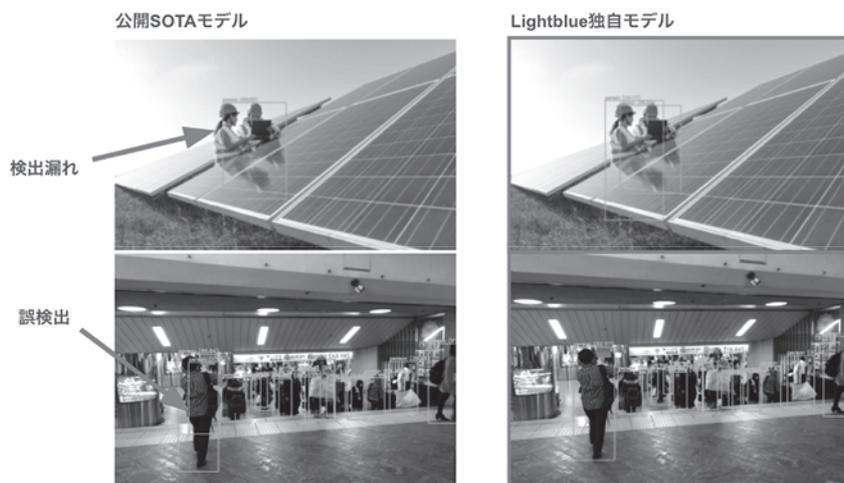
### (3) 建設現場における画像解析技術の特徴

建設現場には通常の社会生活では見られない資材や建設機械が多くある。建設現場で画像解析を行うためには、これらの物体検出やセグメンテーションモデル構築のためのデータが必要であるが、その特殊性から公開データ等の学習用データが入手しづらいという特徴がある。また、トンネル等の土木現場では屋外や室内と比較して薄暗いといった特徴もある。その他に、工場や商業施設等と大きく異なるポイントとして、日々、状況が変わるといった点がある。これは建設という事業の性質上避けられない課題であり、画像解析を導入するに当たり重要な観点である。これらの課題の説明と対策を以下に紹介する。

#### (a) 教師データ作成

建設現場特有の物体として建設機械がある。建設機械の中には幅広い現場で用いられるものと、特定の現場でしか使われないものがある。幅広い現場で見られる機械としては、バックホウやトラック、小型のクレーン装置、ミキサー車などがある。これらの機械は大きく見目が変わることも少なく、導入数も多いため、汎用的なデータとして蓄積がある（写真—1）。一方で、例えばトンネル現場でしか見られない吹き付け機やジャンボなど導入数が少ないため、プロジェクトごとに新たにデータを用意する必要がある。

教師データの作成には、人手によるラベル付けが必要であるが、オンライン上で利用できるラベリングツールを用いて多くの作業者に同時に作業してもらう



図—5 蓄積したデータとモデルのチューニングにより人物検出の性能が優れている例



写真—1 幅広い現場で用いられる建設機械の例（バックホウ）

ことで高速なラベル付けを実現できる（図—6）。また、一般的な車両を検出するモデルを適用し、検出された矩形に対しラベル名を変更するだけの作業とすることや、新たな教師データを作成途中の段階で暫定的なモデルを構築し、それにより推定されたラベルの修正を人間が行うことで効率的な教師データ作成を実現できる。このように人間参加型で機械学習モデルの精度向上を目指すアプローチは Human in the loop と呼ばれ、機械学習を用いたアプリケーション開発の効率的な進め方として注目されている。

#### (b) 環境への対応

建設現場は工程の違いにより使用する機材が異なるだけでなく、工事の進捗に伴い建造物の様子も変化する。建物の建築の場合は、基礎工事の際は地面が見えるが、躯体工事では足場や養生等で外から中の確認が難しくなる。また、仕上げ工事では部屋ごとに壁ができ、俯瞰での撮影が困難となる。そのため、基礎工事の際はトンネル等の土木工事と同様に俯瞰での撮影が可能であるが、工程が進むにつれて作業員自身にセンサーやカメラを装着する方が効率的な場合が増える。このように目的と環境に応じて撮影方法を検討する必要がある。

また、トンネル現場などでは光源が乏しく、常に薄暗い環境である。ただし、トラックや作業員のライトによる局所的な光源への対応が求められる場合もある。このような環境下では、ソフトウェアでの対応の他にカメラセンサーやレンズ、フィルターといったハードウェアでの対応が必要な場合もある。画像解析のための映像取得の際は、単純に人間がきれいと感じる撮影ではなく、解析がしやすい撮影であることを確認する必要がある。そのため、ハードウェアの特性と画像解析の特性に通じた専門家としてのノウハウが重要である。

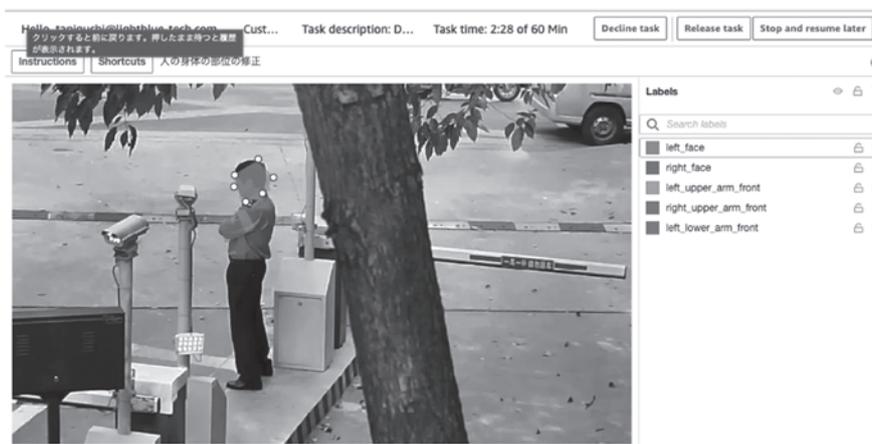
## 4. Human Sensing/画像解析技術の提供方法・想定利用シーン

### (1) 提供方法

建設現場への画像解析技術の提供にあたっては、安全性向上用途か業務改善用途かで提供方法は異なる。安全性向上用途では、危険状態を検出しアラートの発報や機械の制御等をリアルタイムに行う必要がある場合が多い。安全教育のために過去動画の解析を行う場合もあるが、そのような場合でも過去動画の解析とは別にリアルタイムの検出が求められることが多い。一方で、業務改善用途の場合は、リアルタイム性より情報の構造化や検索性、アクセス性が求められることが多い。

#### (a) エッジ環境

リアルタイムの解析には、安定性や通信の遅延の影響を避けるという観点からエッジでの処理を行うことが多い。速度を重視したエッジ処理のためには、適切なカメラ制御と通信の速度、画像解析の高速化など様々な領域の技術が求められる。カメラから取得した映像のエンコード・デコードの処理の時間や、通信の



図—6 高速ラベリングツールの操作画面

速度、通信容量といった解析の前の部分の速度も非常に重要である。また、インターフェースの安定性も重要で、例えばUSBインターフェースは接続の安定性が低いため、M12コネクタなどへの変換を行うことも重要である。もちろん、解析の速度を向上させることも重要である。解析速度の向上にはアルゴリズムの面からのアプローチと計算機性能によるアプローチが可能である。アルゴリズムの高度化により軽量のアーキテクチャで高性能を出すことと、枝刈り、量子化、蒸留といったモデル圧縮技術によるモデルの軽量化は重要である。さらに、画像解析に用いるコンピュータの性能を最大限活かす工夫もある。GPUを使用したデコードやモデルの最適化を行うことで、CPUのみを使う場合と比較して高い性能を実現することが可能である。

エッジマシンとしては、NVIDIAの提供するJetsonがよく用いられるが、R&Dでよく用いられる開発キットではなく産業用筐体を用いることで上記のインターフェースの課題を解決可能であるが、開発キットとOperating Systemが一部異なることがあり、開発キットでの成果を再現するためには追加の開発等が必要な場合もある。そのため、R&Dの時点で産業用の筐体を用いるほうが良い場合も多い(図-7)。

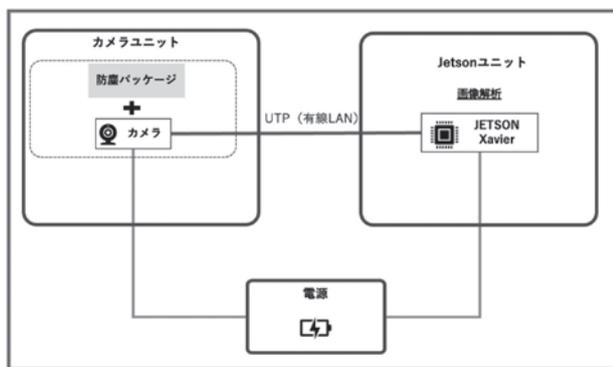


図-7 エッジ環境の構成図

### (b) クラウド環境

業務改善のために、情報の構造化や検索性、アクセス性を実現するには、解析をクラウド環境で実施する方法が考えられる。情報の収集・蓄積・提供にあたっては、施主、建設事業者、作業員等の利益を損ねないために万全なセキュリティ対策が必要だが、建設に関わる限られた関係者に必要な情報を提供することで、業務改善に資すると期待される。

現状、建設現場にカメラが設置されている場合でも、事務所での確認や防犯が主要な用途であり、録画機能さえない場合や録画していても数日で上書きさ

れ、共有されないという例も多い。これらのデータをクラウド上で管理できるようになることで、様々な効果が期待できる。

例えば、現場の作業内容を本社で管理したいという場合は、Human Sensing技術により作業内容を分類し、時刻ごとの作業内容を記録した情報をインターネット上で確認でき、必要に応じて該当箇所の映像を確認することができるといった仕組みの構築が可能である。ログ情報を確認できるユーザと、オリジナルの映像を確認できるユーザの権限を分けることで、必要な人に必要な情報を届けることも可能である。管理業務の効率化だけでなく、複数現場の比較により優れた現場の発見やその知見の共有が可能になる。

## (2) 利用シーン・事例

### (a) 重機と人の接触事故防止

人工知能学会第34回全国大会で著者が発表した例として、建設機械による災害防止のための対人距離推定手法の研究がある<sup>4)</sup>。本取り組みでは、建設現場における建設機械と作業員の接触災害防止を目的とし、単眼カメラによる作業員の姿勢推定とそれを応用した距離推定を行っている。人物の矩形の底辺の中心座標、姿勢推定によって得られる足元の座標、姿勢推定によって得られる頭と腰の座標をもとに、足元の座標を推定し、二次元上の位置を幾何的に推定している。検出範囲は10m以内を対象としている(図-8)。

過去の発表の際は、アルゴリズムの検討段階であったが、現在ではコンピュータの筐体設計や建設機械からの電力供給の仕組みの開発なども完了し、実現場での実証実験が行われている。

### (b) 俯瞰カメラによる作業内容・危険行動監視

俯瞰カメラからの危険状態の把握や作業内容の推定を行った事例がある。俯瞰カメラで画像解析を行う際は、高解像度カメラを活用することで精度の向上が期

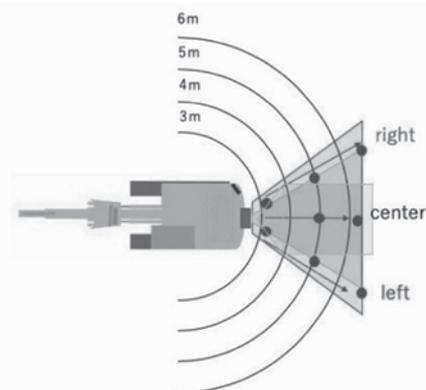


図-8 対人距離推定の実験位置概要(重機を上から見た図)



図一9 顔向き推定（よそ見検出）



図一10 階段での転倒防止・転倒検出

待できる。4Kや8Kの高解像度カメラを使う利点として、人物検出自体は低解像度に変換した状態で実行し、人物エリアに対し詳細な解析を行うというトップダウン型の解析の際に、作業員ごとのエリアを高解像度に解析できるため、姿勢推定などの精度が飛躍的に向上する。俯瞰で撮影した場合、大型機械も映像に収まるため、可能な解析の幅が広がることも利点として挙げられる。

高解像度カメラを用いることで、カメラ台数を減らし、位置関係のキャリブレーションの手間を省き、現場運用で毎回かかっていた手間も省略できる。また、台数が少ない分、メンテナンスの管理工数・コストも減少するという利点もある。

#### (c) 他業界での実績の転用

製造業等での現場では、作業の可視化の事例が多数ある。顔向き推定やセグメンテーションなど複数の解析を組み合わせ、より具体的に作業の内容を判定することが可能である（図一9）。この技術を建設現場に転用することで、重機を見ていなかったり、何かの作業をしていて危険に気づいていなかったりする人を検知するシステムを構築することが可能である。また、階段での転倒防止・転倒検出の事例では、手すりを持っているかどうかを手と手すりの位置関係で解析し、予防のアラートを鳴らすシステムを構築しているが、足場や立ち馬上での不安全行動の検出に転用が可能である（図一10）。

## 5. 可視化技術のさらなる発展・可能性・将来像

現在、構想検討を進めている画像解析アプリケーション開発について以下に紹介する。

### (1) 安全管理の高度化

前節までに紹介したような画像解析アプリケーションが各建設現場に導入されることで、可視化される作業が増加していき、現場の状況をより正確に中央管理できるようになると期待される。画像解析技術は単に現在の危険性を抑制し業務効率化に資するだけでなく、多くの現場のデータを収集し中央管理できるようになることで、事故のデータの蓄積も行える。カメラ、画像解析システムを用いて事故のデータを記録することで、現在行われている報告書に基づく記録と比べて、作業前後の周囲の環境や被災した作業員の動きなどを詳細かつ正確に記録できるようになる。これらの記録を全国の現場から収集することで、危険性の高い動きや、勤務パターン、現場の特徴が可視化されることが期待される。このようなデータに基づき、事故対策を行うことで、安全管理の優先順位や投資効果がわかるようになる。このようなデータは各ステークホルダーにとって非常に有益なデータとなる。施主や建設事業者は効率的に事故を減らすことで、事故対策費用や事故発生時の費用の抑制を実現でき、また工期の遅延リスクを低減できる。また、民間保険事業者もこのようなデータに基づき、正確な事故リスクを把握することで、適切な保険料の算定や保険請求の確認作業を軽減できると期待される。公共セクターにとっても、建設現場での労災事故の発生状況の正確な把握を行うことで、適切な補助事業や労災保険の運用が行え、建設業に従事する労働者の労働環境改善に資することが出来る。

このような事故の解析には、事故発生状況を直接撮影しているカメラからの情報だけでなく、工事サイク

ルの情報や工期が遅れているといったようなデータも合わせて分析することで実現される。

## (2) 人と機械が共存するための技術

建設業は、他の産業と比較して高齢化が進んでいる産業である。日本における2020年の建設業に従事する労働者の数は492万人であり、そのうち84万人が65歳以上である<sup>5)</sup>。このような環境を背景に、省力化が求められている。自動化技術は、危険な作業に従事する労働者を削減することができ、安全性の観点でも非常に有益である。一方で、土木工事の場合は地質により対応を変える必要があり、現場数も多くないため自動化は遅れると考えられる。また都市における建築の場合は、土地が狭く自動化のための装置が導入しづらい環境も存在する。工程によっても自動化しやすい作業とそうでない作業が存在する。このような環境では、一気に完全自動化に移行することは難しく、自動化された作業と人間が行う作業が混在すると考えられる。その際、従来の人間しかいなかった環境とは異なるため、場合によっては新たに危険な状況が発生することも考えられる。このような移行期を支える技術として、画像解析による安全管理はその効果を発揮すると期待される。人の動きを正確に捉え、自動化装置の安全のバックアップを行うことができるため、自動化装置の導入のハードルを下げることも可能である。

## 6. おわりに

今後、建設現場のデータ取得は進むと考えられる。その際に、単に映像を記録するだけでは、問題が発生した際の確認用という位置づけになってしまう。一方で、DXを推進するためにはデータを構造化し、必要な情報を必要な人に届ける技術が重要である。そのため、テキストデータからの情報抽出技術と画像解析による現場の動きの可視化が重要な役割を果たすと期待される。

JICMA

### 《参考文献》

- 1) 厚生労働省：平成30年における労働災害発生状況（確定）、2019年、<https://www.mhlw.go.jp/bunya/roudoukijun/anzeneisei11/rousaihassei/xls/18-kakutei.xls>
- 2) i-Construction委員会：i-Construction～建設現場の生産性革命～、2016年  
<https://www.mlit.go.jp/common/001137123.pdf>
- 3) Jacob Devlin et al. : BERT : Pre-training of Deep Bidirectional Transformers for Language Understanding, 2018年
- 4) 建設機械による災害防止のための対人距離推定手法の研究, 2020年
- 5) 総務省：労働力調査 基本集計 年齢階級、産業別就業者数（2007年～） - 第12・13回改定産業分類による、2021年  
<http://www.stat.go.jp/data/roudou/index.html>

### 【筆者紹介】

園田 亜斗夢（そのだ あとむ）  
 (株) Lightblue Technology  
 代表取締役社長

