

続・建築生産のデジタル化に向けた AI 活用

－複数カメラのデータ連携と体幹前傾角推定手法の構築－

石岡 宏晃

(技術研究所)

AI for Digitization of Construction Management Part II

－Multiple Cameras Data Linkage and Trunk Inclination Estimation of Construction Worker－

Hiroaki Ishioka

建築生産のデジタル化に向けた AI 活用の事例として、現場管理者が作業を振り返ることが可能となるデータ取得手法を示す。画像処理 AI によるデータ取得手法を発展させ、複数のカメラ間の同一人物の推定とカメラの相互位置関係の推定からなるデータ連携手法を構築した。我々は精度低下の発生する条件を明らかにし、実用化には他の技術との組み合わせ等による解決が望ましいと結論付けた。また、映像を安全衛生管理に活用する手法として、姿勢推定 AI を用いた体幹前傾角推定手法の精度確認実験をおこなった。撮影条件によって AI の誤検出に起因するノイズが多く発生することを明らかにし、適切なノイズ除去の設計に課題が残っていると結論付けた。

This paper shows an example of AI for digitization of construction management. We built a method to acquire digital data of the site work for the site manager's review. As improvement of the data acquisition method using deep object detection, we made data linkage method consisting of multi-camera tracking method and multi-camera auto calibration method. We indicated some conditions of false positive occurrence and we concluded the combination with other methods will be needed for satisfy as practical application. For health and safety management, we made trunk inclination capturing method using deep pose estimation and we confirmed its performance in laboratory experiments. As a result, we need improve noise removal function on data interpretation step.

1. 緒言

建築生産における就業者数は依然として減少しており¹⁾、2024年度より残業規制が適用対象となるなど、生産性向上が引続き強く求められている。筆者は生産性向上実現に向けて重要な点として、新規入職者からベテランまで、現場内で作業するすべての職人が、自身の技能を発揮するうえで受ける種々の制約を最小化することが、現場管理者の業務時間を適正化するうえでも重要であると捉えている。そのために必要なのは2点あり、一つは作業の前提となる工程計画と作業計画に、当該作業空間である建築現場の実状を適切に反映すること、もう一つは作業実施中に発生した事象が他の作業に与える影響を適切に配分すること、である。いずれも作業の状況を最適化計算に乗せるために当該事象と相互作用をモデル化する必要がある、そのためには作業の状況をデジタルデータとして取得するところから実現しなければならない。

建築現場内の作業のデータ取得は、全世界の建設業界で共通の課題であり、多くの研究者が建築

現場内の作業のデータ取得に取り組んできた。このことはその課題とともに、前報²⁾で複数の文献^{3)~5)}を参照し示した。

本報で課題とするのは建築物が大規模かつ現地一品生産であるためにおこる、測定データの分割という状況の解決である。図-1に示す通り、画像処理 AI によりカメラの映像から作業データと身体負荷データを取得することでそれぞれ工程管理と安全衛生管理に使用できる。カメラでは遮蔽物の向こうをデータにできないため、複数のカメラを配置して死角を減らす必要があり、現場管理という観点では、複数のカメラから出力された複数のデータを適切に同期・連携させて現場全体の情報として把握する必要がある。

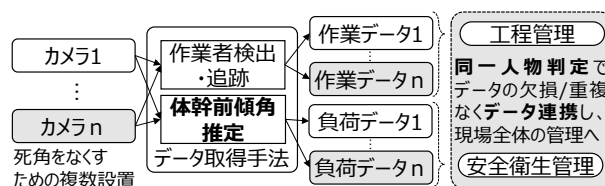


図-1 データ取得手法と管理の概要

本報では、複数カメラによるデータ取得に向けて、複数の映像に映る作業者を同一と判定する画像処理 AI によるデータ連携と、カメラ相互の位置関係を映像から推定するアルゴリズムによる盛り替え(あるいは意図しない衝突による移動)への校正工数の削減を目的とした検討を行うとともに、安全衛生管理に向けて、姿勢推定 AI により得られる映像中の作業者の骨格情報から体幹前傾角を推定する仕組みを実装し、精度検証を行ったので、その結果と考察を述べる。

2. 既往の研究

2.1 複数カメラ間のデータ連携

複数カメラ間のデータ連携において今回取り扱うのは、複数の映像に映り込んだ同一人物を同一人物であると判定するアルゴリズムの研究である。

1) **データセット**：複数カメラ間で人物を同定し追跡する研究における代表的なデータセットが存在する。例えば、DukeMTMC⁶⁾は大学キャンパス内の教室への往来を8台の監視カメラで撮影した14時間・約2000人・200万超の人物枠を付与した映像データセットである。MARS⁷⁾は清華大学キャンパス内の歩行者を6台のカメラで撮影した1261人のデータセットである。EPFL⁸⁾は4台のカメラで取り囲んだ4シーン(実験室内・バスケットボール・通路・屋外テラス)のデータセットである。WILDTRACK⁹⁾はスイスETHの本館前広場で晴天時に撮影された7台のGoPro映像群によるデータセットである。いずれも映像に映り込んだ人物は私服の歩行者であるため、建築作業者のデータセットが必要となる。

2) **歩行者の追跡**：複数の映像間の人物を同定する研究は再識別 Re-Identification という研究分野である。オーバーラップのない(別の場所を映した)映像間で同一人物を検出することに主眼が置かれている。アプローチは大きく二つに分かれている。画像単体の特徴量を用いる Image-based Re-ID と、画像の時系列を用いる Video-based Re-ID とである。前者の Image-based は例えば、Cam-Style¹⁰⁾の手法はGANを用いてより多くの学習サンプルを生成することで精度向上を図っている。SPreID¹¹⁾やPCB¹²⁾やPart-aligned¹³⁾といった近年の手法は人体の個別の部位に関連した確率を用いて精度向上を図っている。調査時点では Bag-of-tricks¹⁴⁾の手法がベストパフォーマンスを示していた。後者の Video-based は2016年に最初のRNNを適用した手法¹⁵⁾が示された。以降時系列を学習するのに Attention を導入する例

が出始め¹⁶⁾、人体部位を考慮する Part-Attention¹⁷⁾や調査時点でベストパフォーマンスを示す NVAN¹⁸⁾が提案されている。いずれの手法も高画質映像を前提とし、歩行中の人物を対象としており、しゃがむなど別の姿勢の検出精度を議論の対象としていない点ため、建築現場への応用時に注意が必要である。一方、全く異なる映像間で同一人物を判定する研究もすすめられており、Qingqiu^ら¹⁹⁾は、時系列の人物画像群同士を比較する際に、最も類似性の高い画像同士を互いの判定対象として学習に使用する手法(PPCC)を提案している。また Zhong^ら²⁰⁾は一度同一人物であるかを判定した結果を再度双方向に確認することで精度を高める手法を提案している。これらは本研究が対象とする建築現場の複数カメラのように映像が大きく異なる対象に有効と考えられる。

3) **カメラの姿勢推定**：複数の映像からカメラの相対的な姿勢を推定するタスクに対して、深層学習アルゴリズムを用いようとする研究は比較的新しい分野(Deep-Regression)である。2017-2018年にいくつかの手法²¹⁾⁻²⁵⁾が提案されているが、それらの推定精度は、以前から行われてきた研究者が手動で設計する特徴量検出アルゴリズムに対して低い結果しか得られていない²⁶⁾。カメラの姿勢推定手法には、アルゴリズムは重いものの安定した結果を得られる構造ベースの手法(Structure-based method)²⁷⁾も用いることができる。例えば SIFT²⁸⁾、SURF²⁹⁾、Superpoint³⁰⁾、といった信頼性の高い手法の組合せによる特徴量抽出ののち、三角測量(Triangulation)とバンドル調整(Bundle Adjustment)により姿勢を推定するものであり、本研究におけるカメラ姿勢推定に用いる。

2.2 体幹前傾角推定手法

建築生産において建築作業者の作業状況を分析する重要性が高いことは認識されており、製造業において生産工学が発展するのに合わせて、生産工学における作業測定手法や生産計画手法は建設業に応用されてきた^{31),32)}。モーションキャプチャシステムを用いた作業姿勢の測定と身体負荷推定³³⁾も含め、建築作業者の身体負荷を測定しようとする取り組みがある。日本建築学会の作業能率測定指針³⁴⁾において取りまとめられた作業測定手法のうち、作業者の姿勢を含む詳細な測定はヒューマンファクターの測定手法として位置付けられている。これらは測定員による現場測定を前提とする手法が多く、必要となる工数が大きい点に課題がある。

一方、身体負荷を取り扱う学問である人間工学の分野では、人体を複数の剛体が連結されたモデルとして簡易化し、各関節にかかるモーメント量から身体負荷を計算する手法が複数存在する³⁵⁾。人体の関節角の扱いは ISO11226 に定められている。また、2D 上の計算から発展させ 3D 空間上でデジタルヒューマンによる身体負荷計算をおこなう手法が提案されている³⁶⁾。深層学習手法の登場により画像から多くの情報を抽出することが可能となり、2017 年の Zhe ら³⁷⁾が示した姿勢推定アルゴリズムにより映像中の人物の骨格情報の推定が容易におこなえることが広く知られるようになったことを受け、筆者らは人間工学分野の知見を得ながら、体幹前傾角推定手法を提案している³⁸⁾。

3. 複数カメラ間のデータ連携手法

3.1 データセット構築

前報²⁾と同様に、映像に映り込んだ作業者の個人情報保護への配慮として、測定対象とする現場の選定ののちに許可の交渉、映像使用目的の掲出、映像撮影、データアノテーション(映像に正答とする情報を付与する作業)の手順でデータセットを構築する。

複数の映像に映り込んだ同一人物を判定するために必要な学習用データセットは、作業者検出の正答としてのバウンディングボックス(作業者を囲む矩形領域、以下 Bbox)と、作業者追跡の正答として複数映像間で共通の作業者 ID の付与である。

3.2 研究手法

本研究実施時の制約条件^{注1)}を踏まえ、いくつかの手法を比較検証した結果、本研究における手法は以下に示すものである。

1) **前提とする入力情報**：本手法は、建築現場内に設置された複数のカメラによって取得された映像と、その映像内に映り込んだ人物の、時系列に連続な矩形領域(同一人物の追跡がおこなわれているものとする。以下、Tracklet)とを入力情報とする。映像の解像度は特に指定しないが、人物の矩形領域が十分に特徴量を抽出可能なものとなる指標として“人物の矩形領域の縦に 100 ピクセルが確保されている”ことが目安とされており、設置されているカメラの選定時には、当該カメラが担当する撮影区域内の人物との距離を考慮して十分な解像度を有するカメラを選定する必要がある。またカメラの姿勢推定と作業者の実空間上の位置計算には同時刻の映像を用いる必要があるため、映像は相互に時間を同期する。

なおそれぞれのカメラは焦点距離や歪みといった固有パラメータが既知であるものとする。

2) **同一人物の推定**：同一人物の推定には、NVAN¹⁸⁾のアルゴリズムをベースに、大きく異なる映像の類似性を判定可能な PPCC¹⁹⁾を導入し、判定制度を高めるための Re-Ranking²⁰⁾を用いる。学習は Tracklet 単位で入力するため、用意したデータセットのうち、学習用と評価用の分割は映像単位でおこなう^{注2)}。

3) **カメラ姿勢の推定**：カメラ姿勢の推定は、対象とするカメラ群が捉えた映像中の、同時刻の同一人物に対する矩形領域の組を計算に用いる。個々の矩形領域の中心のピクセル座標を計算対象とし、対象とカメラとの相対関係を推定する。この相対座標系はいずれかのカメラを基準とし、そのカメラの中心点を原点、向きを座標軸として計算される。

推定結果を実空間上の寸法に合わせるために、映像中の寸法既知の線分情報を与える。カメラ群に共通して映り込んでいる線分を、その両端のピクセル座標と実空間上の線分の長さとのペアで与えることで、推定結果を実空間上の寸法に変換する。

4) **作業者の移動軌跡の可視化**：推定されたカメラ姿勢を用いて、作業者の 3 次元座標が計算可能となる。これは前報²⁾で示した座標計算と共通であるため割愛する。本手法においてはカメラに対する相対座標である点に注意が必要である。データ利用時に、適切に絶対座標系に変換する必要がある。絶対座標系として変換することで、作業者の移動軌跡を平面上に投影した可視化が可能となる。画像処理における一般的な射影変換であるため説明は割愛する。

3.3 連携に対する精度検証

同一人物の推定における精度検証は、推定された候補の第何位に正答が含まれているかによる指標として、上位 k までの結果の中に正解がある場合を True=1 とする「ランク k」の指標を用いる。

例えば 3 台のカメラで、映像の長さ 100 フレームの映像を撮影し、その初めから最後まで 3 人の作業員(ID=0,1,2)が映り込んでいる場合があるとする。この時 3 つのカメラに対して 3 つの Tracklet(それぞれ 100 フレーム)があるため、計 9 つの Tracklet がある。ここでは各 Tracklet を判定対象とし、他の 8 つを判定候補として 9 回繰り返し判定が行われることになる。cam1 で ID=0 の作業員の際、類似性の判定結果として順に[ID1cam1、ID0cam2、ID0cam3、ID1cam2、ID2cam2、ID2cam3、ID1cam3、ID2cam1]が得られた場合、Rank1 は False=0(上位 1 に正しい ID=0 がない)、ランク 2 は True=1(上位

2 に ID=0 がある)となり、ランク 3 以降はすべて True=1(上位 3 以上に正答がある)となる。ただし、通常、判定候補に同じカメラは考慮されないため、あらかじめランキング結果は、[ID0cam2, ID0cam3, ID1cam2, ID2cam2, ID2cam3, ID1cam3]のように間引くことができ、この場合ランク 1 が True=1 となり、ランク 6 まですべて True=1 となる。また、平均精度 mAP(Mean Average Precision)は、下記式(1)で求める。

$$mAP = \frac{1}{Q} \sum_{q=1}^Q \frac{1}{m_q} \sum_{k=1}^{n_q} P_q(k) rel_q(k) \quad (1)$$

Q: 判定対象の総数

m_q : 判定対象と同じIDを持つ判定候補の総数

n_q : 判定候補の総数

P_q : q 番目の判定候補のランク k での精度

rel_q : q 番目の判定候補と予測の関連性。k 番目の予測が正しいなら 1, それ以外は 0

カメラ姿勢の推定の精度検証に関しては、現地で手計測により取得したカメラ設置位置座標を用い、相対距離の比較により行うこととする。

3.4 現場計測の実施

現場映像の撮影には、防塵防水性能や広角レンズを有する GoPro Hero7 Black を採用し、4 台での同時撮影を行った。撮影対象とした現場は RC 在来工法によるマンションの躯体工事を実施中の都内の 1 現場とした。最上階の矩形の 1 工区を囲う外周足場の 3 点に 4 台を設置(内 1 つの角では上下に約 1m の間隔をあけて設置)し、2 日間にわたって映像を取得した。映像取得の際には、実寸法取得のガイドとして、90mm 球を用いた治具を作成し、現場外周足場に設置した。治具は単クランプとボルトナットを用いて自作した。カメラの設置状況ならびに使用した治具の外観を図-2 に示す。

撮影した動画のうち、建築作業に特徴的な「部材を持ち上げて運搬する」動作が含まれるタイミングに着目して抽出し、同期された 9 シーン 22 ファイル計約 65 分間(カメラ 4 台で 88 ファイル計約 260 分間)の映像をデータアノテーション対象とした。22 の映像群が取得された天候と時刻については後述する。動画の切り出しはフリープログラムの「ffmpeg」を使用した。4 台のカメラの同時刻の映像に対して、同一人物に同一の作業者 ID を付与する必要があるため、アノテーション作業の効率化を図るために 4 つの映像を割り付けた大きな 1 つの映像として結合させた。アノテーション作業は前報²⁾と同様に当社から国内企業への業務委託により実施した。人物領域の矩形座標と作業者 ID が付与された映像の例を

図-3 に示す。図中の矩形枠は作業者 ID ごとの彩色をしており、同一作業者に同一 ID が付与されていることがわかる。

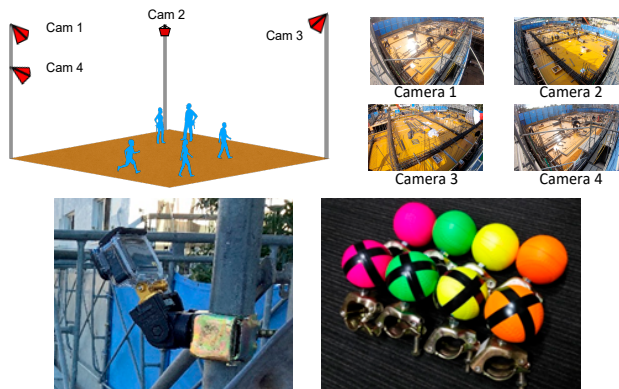


図-2 カメラ設置状況と治具外観

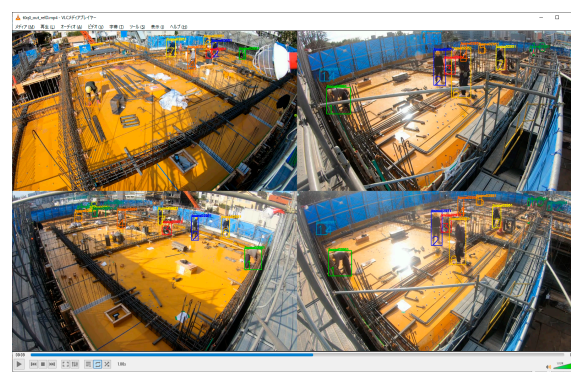


図-3 アノテーション後の映像例

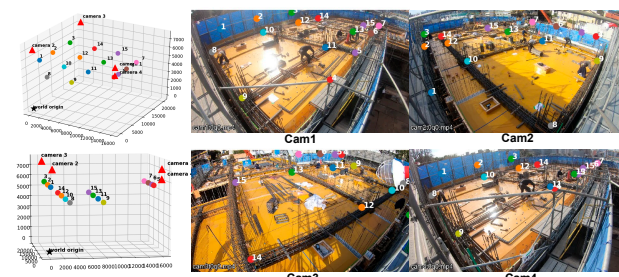


図-4 入力情報として用意した寸法既知点

アノテーションにはツール「CVAT」を使用しており、PASCAL/VOC 形式の XML ファイルが出力された。出力されたデータは 4 つの映像を結合した動画の 픽셀座標となっているため、左上の映像はそのまの 픽셀座標で良いが、ほかの 3 つの映像は適切に 픽셀座標を減じる必要がある。Python により変換プログラムを作成し処理した。Python プログラムの構築と処理実施環境には無償のツール「Jupyter notebook」を使用した。結果として 88 個の映像に対して約 470 万行のアノテー

ションデータが得られ、映像に映り込んだ作業者数は最大で21名、最小で8名だった。

カメラ姿勢推定に対する寸法既知点として、カメラ4つに共通して映り込む15点を選定し、画像中のピクセル座標と、図面から読み取った実空間上の3次元座標とを組にしたCSVファイルを作成した。寸法既知点の位置関係と画像上に描画した結果を図-4に示す。

3.5 複数カメラ間のデータ連携

1) 同一人物の推定：構築したアルゴリズムに対して、図-5に示す4パターンでの学習を実施した。図中の彩色部が学習に使用された映像群、無彩色部が評価に使用された映像群を示している。22の映像群は、図中上部に示す通り「晴天・日中(天頂付近に太陽があり一部の床面に反射光によるホワイトアウトが発生・影が発生)」「晴天・夕刻(一部に夕日の映り込みあるいは反射光によるホワイトアウトが発生・影が発生)」「曇天・日中(太陽が雲に隠れており全体に薄暗い・影がない)」という3種に分かれている。Split1～Split3はファイル名順に機械的に8/6/8に分けた分割方法であるため、Split1は晴天・日中を学習しておらず、Split3は曇天・日中の学習が少ない。対してSplit4は学習と評価と交互に採用することで、いずれの天候条件も学習している。

天候	晴天・日中						晴天・夕刻						曇天・日中						晴天・夕刻					
name	000	001	002	003	004	005	006	007	008	009	010	011	012	013	014	015	016	017	018	019	020	021	022	023
max人数	14	13	10	11	14	10	14	11	14	10	11	14	11	13	10	11	14	11	14	10	11	14	11	14
min人数	8	10	13	10	11	14	10	11	14	10	11	14	11	13	10	11	14	11	14	10	11	14	11	14
Split1																								
Split2																								
Split3																								
Split4																								

図-5 学習対象データと天候の関係

表-1 同一人物の推定精度

学習データの分割	Rank1	Rank5	Rank10	mAP
Split1	85.0	95.4	98.6	78.0
Split2	89.0	98.8	99.4	85.5
Split3	78.2	95.6	98.2	75.1
Split4	93.2	98.7	99.4	90.3
Split4(RCNN_bbox)	84.1	90.5	96.6	83.5

精度検証の結果を表-1に示す。最も高い精度を示したSplit4はRank1で93.2%、mAPで90.3%となった。学習と評価とをほぼ同じ条件の映像群で実施した場合に高い値となる一方、最も低いSplit3では70%台に低下している。今回、cam1とcam4を近くに設置したため、それらの映像は非常に似通っており、推定精度がRank1でも高い数値が得られた

ものと考えられる。なお、表の最下段の「Split4(RCNN_bbox)」は、判定するTrackletを前報²⁾で示した作業者検出アルゴリズムにより得た場合を示している。Trackletが未検出等により途切れる場合や矩形領域のズレなどの影響により、正答率が下がっていることがわかる。

2) 考察：同一人物の推定についてRank1で最大93.2%という高い値が得られたものの、今回の撮影条件においてカメラ1とカメラ4とが1m程しか離隔していないことが影響していることが考えられた。

誤答をした例を観察すると、原因は下記二つに大別された。

- ① 低解像度による誤答：判定対象とした人物領域が遠方の人物のものであり、十分な特徴量が得られなかったと想定されるもの
- ② 似通った服装による誤答：判定対象とした人物が着用している作業着がチームで共通のユニフォームであり、同じ服装をしている人物と混同したと想定されるもの

誤答の観察により得た原因の内、①の低解像度については、ピクセル数で判定対象を限定するなどにより誤答を避ける方法が考えられる。②の「似通った服装」については日本の建設現場では一般的な状況であり、それを混同してしまう今回のアルゴリズムは実用性が不足していると判断される。

今回は映像のみを用いた同一人物の推定であるが、今後は取得された映像の位置関係を考慮に入れるなど、画像判定のみに頼らない手法として工夫する必要がある。

3) カメラ姿勢の推定：同期された人物のピクセル座標群を用いて推定されたカメラの相対姿勢について、カメラ相互の距離を表-2に示す。

表-2 推定されたカメラ位置の相互距離 単位[mm]

	Cam1-2	Cam1-3	Cam1-4	Cam2-3	Cam2-4	Cam3-4
真値(図面)	15510	25318	900	20010	15540	25340
Cam1基準	15816	26262	1350	13846	16075	26280
Cam2基準	15816	29381	662	20722	15465	29014
Cam3基準	20036	26262	1767	20722	20227	25835
Cam4基準	15213	25732	1350	13792	15465	25835
平均	15701	26168	1142	20668	15637	26039
差分(真値-平均)	+191	+850	+242	+658	+97	+699

表には各カメラを基準として推定を行った場合のそれぞれの結果を示している。全ての平均値をとると、真値に対して最大で850mmの誤差に収まる結果となった。基準とするカメラを比較するとcam2基準が他よりも安定している。これはcam2が他のカメラの中間地点に設置されているためではないかと推察される。ただし、各値を見ると最大で約5m

の誤差を示す例もあり、人物検出枠の情報のみを使用して、20m 近く離隔したカメラ姿勢を推定するのは不安定であると明らかになった。

なお、ここまで示した2つの手法は建築現場における実用性に着眼して考察を行っている。機械学習分野への発表としては既存手法との精度比較が実施されており、結果について共同研究先から論文が投稿されている。^{41),42)}

4. 体幹前傾角推定手法

4.1 研究手法

- 1) 前提とする入力情報：体幹前傾角推定には、測定対象となる作業者の全身が映り込んだ映像を入力情報として用いる。また当該映像を撮影したカメラの内部・外部パラメータは取得されているものとする。
- 2) 関節点の推定：映像に映り込んだ作業者の全身に対し、深層学習を用いた姿勢推定アルゴリズムによる関節点の推定を行う。本研究においては、「OpenPose³⁹⁾」と「Alphapose⁴⁰⁾」とを用いて試作した^{注3)}。それぞれの姿勢推定アルゴリズムが出力する関節点を図-6に示す。

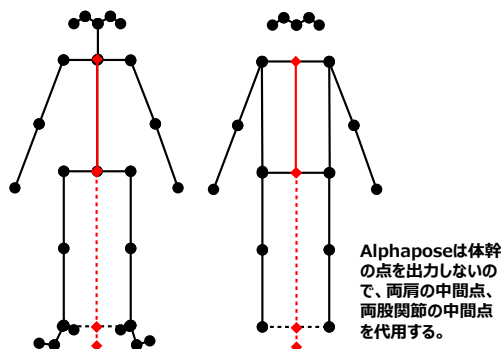


図-6 出力される関節点(左 Openpose, 右 Alphapose)

- 3) 必要な身体寸法の推定：体幹前傾角の推定に必要なとなる「腰部高さ[mm]」と「体幹長さ[mm]」を推定する手法を構築した。

腰部高さ z_h の入力を不要とする $P_{hip}(x_h, y_h, z_h)$ は、腰部は足の重心 $F_0(x_f, y_f, 0)$ の直上にあると仮定して求める。足の重心は、姿勢推定により得られる“くるぶし”の2点の中間直下にあると仮定し、くるぶしが床からの高さ $z = 63[mm]$ にある^{注4)}ものとして、カメラ座標からの射影計算により $F_0(x_f, y_f, 0)$ を求める。そのうえでカメラ座標 $C(x_c, y_c, z_c)$ の直下の点 $C_0(x_c, y_c, 0)$ から足の重心 $F_0(x_f, y_f, 0)$ と、 $H_0(x_h, y_h, 0)$ とが同じ距離にあることから、三角形の相似により z_c から z_h が求まる。

また直立時において、胸部は腰部の直上にあるため、同様の考え方で胸部 $P_{breast}(x_b, y_b, z_b)$ を求め、体幹長を求めることができる。ただしこの計算は直立時にのみ適用でき、部材の把持等で腰を曲げる姿勢においてはその直前に推定された直立時の体幹長を継承して用いることで解決を図る。姿勢推定アルゴリズムを変更する場合には、適宜使用する関節点を調整し対応が可能であるものと考えている。

- 4) 体幹前傾角の計算：既報³⁸⁾で示した通り、体幹前傾角は図-7の通り求める。

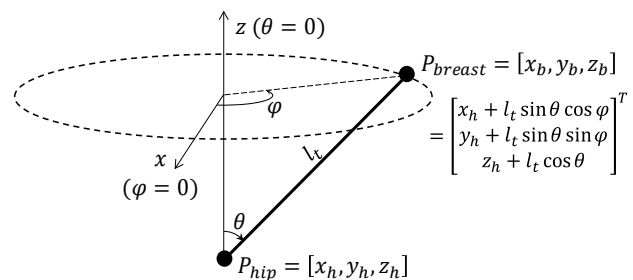


図-7 体幹前傾角計算の概要図

- 5) 腰部負荷の計算：既報³⁸⁾で示した通り、体幹前傾角に対して、身体寸法と各部位の重量を用いた剛体モデル(図-8)のモーメント計算により腰部負荷を求めることができる。

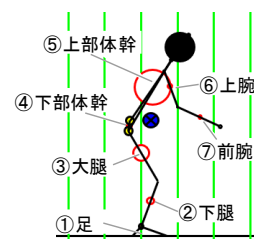


図-8 7セグメントによる剛体モデル

- 6) 体幹前傾角推定に対する精度検証：体幹前傾角推定の精度検証は、人間工学分野でも使用されるより正確な測定手法との出力値の比較により行う。本研究では、建築作業を模擬した実験において光学式のモーションキャプチャによる測定を比較する^{注5)}。模擬的に実施する作業を、提案手法と光学式のモーションキャプチャとで同時に計測し、双方の体幹前傾角を比較し提案手法の推定精度を検証する。

4.2 体幹前傾角推定手法の性能検証

- 1) システムの実装：システムは将来的な安全衛生管理への活用を想定し、動画に対してリアルタイムで推定結果を出力できるように、映像の入力に対し、フレームごとに「姿勢推定→体幹前傾角推定→腰部

「姿勢推定」のフローを回すことができるように設計した。システムは姿勢推定アルゴリズムの推奨動作環境に従いGPU搭載のUbuntuOSのPCを使用し、Pythonによるコーディングとした。

2) 実験参加者：実験参加者は22歳から24歳の健康な男子学生3名とした。各実験参加者はこれまでに大きな外傷や疾病の罹患歴が特になく、今回の実験を行うにあたり支障はない。各実験参加者の基本データを表-3に示す。実験参加者には実験同意書への記入ののち、体調等のアンケートに回答のうえ、モーションキャプチャスーツを着用させた。

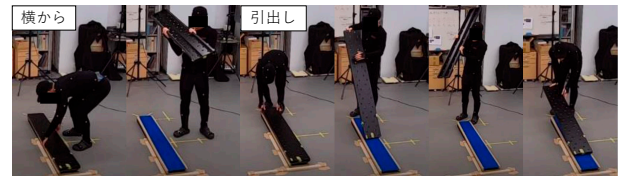
表-3 実験参加者データ

ID	Age	Height [cm]	Weight [kg]
A	24	169.0	66.4
B	24	173.8	59.5
C	22	173.5	79.7
Mean	23.3	172.1	68.5
Std. Dev.	1.2	2.7	10.3

3) 実験条件：本実験では、重さ3.0kgのアルミ製の足場板の持ち上げ作業を行った。光学式モーションキャプチャシステムを使用するにあたり、足場板表面の反射を防ぐために塗装を行った。塗装したことによる持ち上げ方への影響はない。3枚重ねて床に置いてある足場板のうち一番上の板のみを持ち上げる。なお塗料の剥がれ防止のため床に置いてある足場板に、持ち上げる板と接する面に保護シートを貼り付けた。

持ち上げ方は「横から」と「引出し」の2種類とした。「横から」は板長辺の中間および板中心から20cm離れたところにつま先がくる位置からかがみ、右肩にのせる。「引出し」は板長辺の端に左足つま先をつけ、板に対して45度の向きから板の端を持ち、自分の身体に引き寄せて右肩にのせる。どちらも持ち上げたあとは実験参加者から見て奥側の辺を左手でつかみ裏返しながら持ち上げることとした。下ろし方は「横から」は持ち上げの逆の手順とし、「引出し」は裏返しながら板の短辺を先に下ろす場所につけ片手で押し込むようにして置くこととした。

1 試行のフローは5秒ごとに動作を切り替えるものとし、動きの大きな動作の間には静止時間を設け、データ上で判別しやすくした。実験では実際に歩く動作は行わず。持ち上げてその場で5秒間足踏みをしてもらい下ろす作業とした。どの実験参加者にも動作の練習を行わせ、個人ごとにやり方が大きく変わらないようにした。計測は「横から」を5試行ののち、「引出し」を5試行の順で実施した。図-9に実験条件の概要を示す。



○:作業開始、足踏み、終わり ○:持ち上げ、下ろし

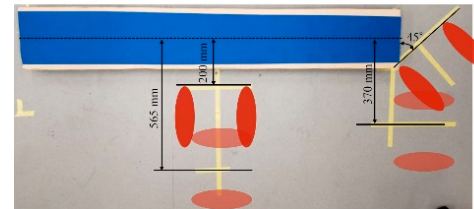


図-9 持ち上げ方の概要と動作フロー

4) データ取得方法：動作中の姿勢の真値として扱うデータを、光学式モーションキャプチャシステム(OptiTrack NaturalPoint 社製)を使用した姿勢計測により取得した。身体負荷については、計測したデータを用いて共同研究先研究室開発の生体力学解析ソフトで関節角度、関節トルク比を推定した。

体幹前傾角推定システムのための映像取得は、カメラパラメータ取得済みのGoProを用いて行った。システムの出力する画像の例を図-10に示す。図中左の直立時には、腰部高さ(h=838mm)、体幹長(L=497mm)、体幹前傾角(1degree)、腰部負荷(4Nm)、の各値が計算される。図中右のように体幹が前傾する場合にも、各種の値が出力されている。

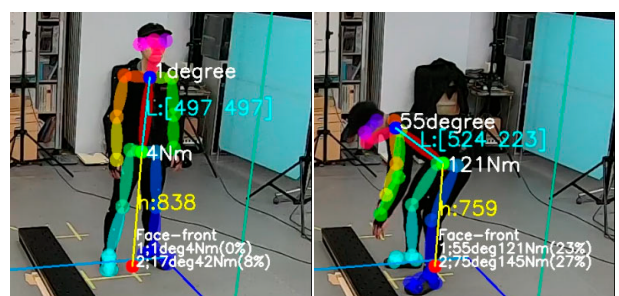


図-10 システムからの出力例

5) データの同期：モーションキャプチャシステムはスポーツ工学などの精密測定を目的とした装置であるため出力は100Hz(1秒間に100データ)であり、対する映像は60Hz(1秒間に60枚の静止画)である。静止画に対する姿勢推定は負荷の大きい処理であることから、作業効率を考慮すると処理回数を減らしたい。建築作業における人間の姿勢変化はゆっくりとしたものであるため、10Hz(1秒間に10枚の静止

画)で十分な推定がおこなえると観察から判断した。10Hzで双方のデータを抽出し、同期するタイミングには部材把持のしゃがみ込み開始時の急激な体幹前傾角変化を選定した。

4.3 結果と考察

同期後の各データを一つのグラフとした結果を図-11に示す。図は横軸に時間変化[秒]、縦軸に体幹前傾角[度]を示す。横軸には5秒ごとの動作の変化をテキストで付記した。オレンジの太線が光学式モーションキャプチャによる測定値(motion-cap.と表記)、緑が「Alphapose」による関節点を用いた体幹前傾角の推定値、青が「Openpose」による関節点を用いた体幹前傾角の推定値、を示している。

「横から」の場合は、被験者の動作は単純に足場板に手をかけて持ち上げる動作となるため体幹前傾角の変化もシンプルに増減する様子がわかる。また、足場板を肩に担いでいる15秒から35秒の間には背面を反らせて支持する状態となり、モーションキャプチャではマイナス10度の値として測定されているが、体幹前傾角推定手法では前傾のみを考慮した計算としているため0以下の値が出力されていない。ただしこの背面の反りは被験者に与える負荷が軽微であるため身体負荷推定上は問題ないと判断できる。前傾中の10-15秒、35-40秒を観察すると、推定値が真値とよく整合していることがわかる。今回の撮影条件においてはほぼ真横から被験者の前傾姿勢を撮影しているため、姿勢推定が正確におこなえていたことが要因であると考えられる。

「引出し」の場合は、被験者の動作は、持ち上げにおいては「端部の把持と引き起こし」と「肩に当てて持ち上げる」という2段階の動作、設置においては「端部を設置場所にあてる」と「滑らせながら設置する」という2段階の動作となっており、モーションキャプチャによる測定値から複雑な体幹前傾角の変化であることがわかる。また、体幹前傾角推定手法の推定値が、真値に対して10度程度の誤差をもって出力されていることがわかる。

7)ノイズ発生原因と考察：出力される前傾角にノイズが発生する原因を観察した結果を図-12に示す。今回の撮影条件においてはカメラに向かって頭を下げる映像であるため、姿勢推定アルゴリズムが誤った関節点を推定(あるいは姿勢推定を失敗)していることがノイズ発生の原因であった。人物であると判定できなかった「未検出」のタイミングや、足場板までを含めて人物であると判定する「誤検出」のタイミングが確認された。未検出の場合には出力値は

0となり、誤検出の場合には前後と不連続な出力値となる。

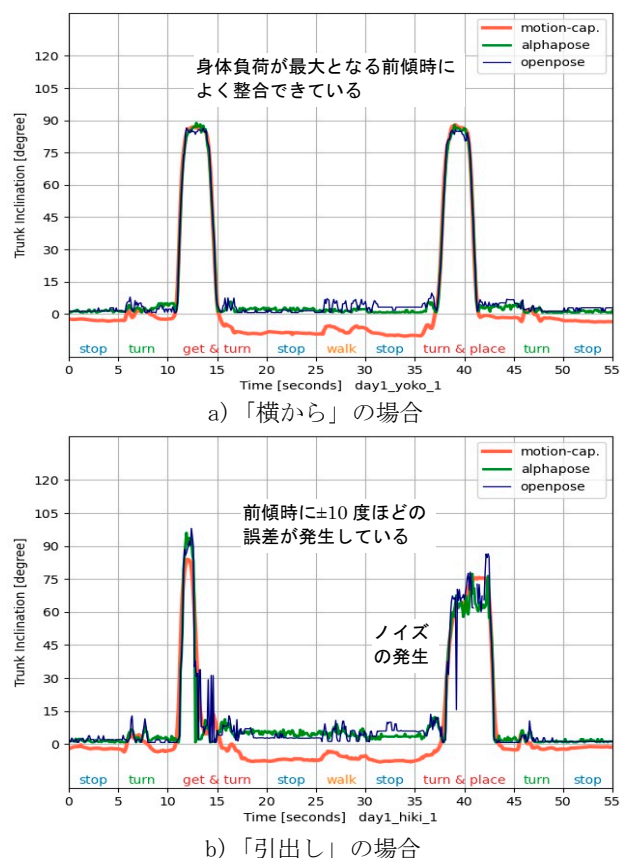


図-11 体幹前傾角の時系列変化(Aの1回目)

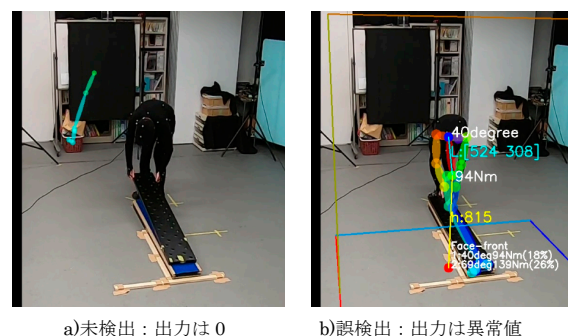


図-12 検出失敗の例

8)考察：今回の実験においては真値に光学式のモーションキャプチャによる測定値を用いた。そのため、実験参加者は黒の全身スーツを着用する必要があり、その影響で姿勢推定AIが背景の黒幕や黒の足場板と人物を混同し、未検出や誤検出を発生させることとなった。建築現場においては作業着であり、煩雑な背景のもとで作業することとなるため、今回の実験と同様に未検出や誤検出によるノイズの発生が想定される。適切なノイズ除去の工夫が必要となると考えられる。

5. 結言

本報では建築生産のデータ取得に資する AI 活用手法の検討として、データ連携と体幹前傾角推定の検討結果を示した。いずれも AI 特有の未検出・誤検出による精度低下が確認されており、画像のみを推定の対象とする手法では、出力値にノイズが含まれることは避けられない。ノイズをなくすには、映像のみでなく別のデータを合わせた推定や、他の計測技術との組み合わせなどの新しい設計が必要といえる。現場における実用には、課題解決に向けた先端技術の調査ならびにソリューションの設計について、継続的な取り組みが望まれる。

謝辞

本報の内容のうちデータ連携に関しては米国のカーネギーメロン大学との共同研究における検討内容を含みます。Kris Kitani 准教授、Xinshuo Weng、Yu-jhe Li、Yan Xu、Vivek Roy にここに感謝いたします。また、体幹前傾角推定に関しては東京都立大学との共同研究における検討内容を含みます。ご指導いただいた瀬尾明彦教授、倉元昭季助教(東京工業大)、平内和樹様、矢治美穂様、にここに感謝いたします。また、現場計測にご協力いただきました皆様にもここに感謝いたします。

<注釈>

- 注1) アルゴリズム構築は2019年10月から2020年9月にかけて実施した。2020年2月頃からCovid-19への社内の対策や緊急事態宣言の発令等を受け、出社停止や現場閉所により、現場計測による映像取得とデータセット拡充ならびに共同研究先の大学においてもラボでの測定実験等が実施できず、初期のデータセットのみで検討しなければならなかった。
- 注2) 画像判定 AI の場合はデータセット全体を8:2等の比率でランダムに学習用と評価用に分割する場合が一般的だが、映像の場合はTrackletごとに学習しなければならないため、比率でランダム分割し難い。
- 注3) ただしいずれのアルゴリズムも、試作当時から更新されており、本研究実施時と同様の動作をする保証はない。図-1の関節点を参照し調整されたい。
- 注4) 人間工学で用いられる標準身体寸法比から、足関節の高さ比0.0394(床から足関節までの鉛直距離の身長との比率)を用いて求めた。 $1600[mm] \times 0.0394 = 63.04[mm]$
- 注5) モーションキャプチャには、簡易な計測が可能な地磁気を利用した製品があるが、測定環境とする建物の床の磁性化、ならびに、実験に使用する資材の磁性化の影響を受けて、正確な測定がおこなえないことを事前検討により確認した。

<参考文献>

- 1) 日本建設業連合会：“建設業デジタルハンドブック”
<https://www.nikkenren.com/publication/handbook/chart6-4/index.html>
- 2) 石岡宏晃：“建築生産のデジタル化に向けた AI 活用—単眼カメラを用いた作業検出と追跡ならびに作業動作認識の検証—”，清水建設研究所報 vol.98, 2020.12
https://www.shimztechnonews.com/tw/sit/report/vol98/98_004.html
- 3) Tarek O. and Moncef N.: "Data Acquisition Technologies for Construction Progress Tracking", Automation in Construction 70, pp.143-155, 2016
- 4) Amin A. et al.: "Sensor-based Safety Management", Automation in Construction 113, 103128, 2020
- 5) 前田邦夫：“現代アメリカ建設学”，開発問題研究所, 1987.02
- 6) Ergys Ristani, Francesco Solera, Roger Zou, Rita Cucchiara, and Carlo Tomasi.: "Performance measures and a data set for multi-target, multi-camera tracking", In European Conference on Computer Vision workshop on Benchmarking Multi-Target Tracking, 2016
- 7) Zheng et al.: "MARS: A Video Benchmark for Large-Scale Person Re-identification", In ECCV, 2016
- 8) F. Fleuret, J. Berclaz, R. Lengagne, P. Fua.: "Multi-Camera People Tracking with a Probabilistic Occupancy Map", In IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008
- 9) Tatjana Chavdarova, et al.: "WILDTRACK: A Multi-Camera HD Dataset for Dense Unscripted Pedestrian Detection", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018
- 10) Zhun Zhong, Liang Zheng, Zhedong Zheng, Shaozi Li, Yi Yang: "Camera Style Adaptation for Person Re-identification", CVPR, 2018.
- 11) Mahdi M. Kalayeh, Emrah Basaran, Muhittin Gokmen, Mustafa E. Kamasak, Mubarak Shah: "Human Semantic Parsing for Person Re-identification", CVPR, 2018
- 12) Yifan Sun, Liang Zheng, Yi Yang, Qi Tian, Shengjin Wang: "Beyond Part Models: Person Retrieval with Refined Part Pooling (and A Strong Convolutional Baseline)", ECCV, 2018
- 13) Yumin Suh, Jingdong Wang, Siyu Tang, Tao Mei, and Kyoung Mu Lee: "Part-Aligned Bilinear Representations for Person Re-Identification", Proceedings of the European Conference on Computer Vision (ECCV), 2018.
- 14) Hao Luo, Wei Jiang, Youzhi Gu, Fuxu Liu, Xingyu Liao, Shenqi Lai, Jianyang Gu: "A Strong Baseline and Batch

- Normalization Neck for Deep Person Re-identification”, In IEEE Transactions on Multimedia, 2019.
- 15) McLaughlin, Niall, Jesus Martinez del Rincon, and Paul Miller: “Recurrent convolutional network for video-based person re-identification”, CVPR, 2016.
 - 16) Li, Shuang, et al.: “Diversity Regularized Spatiotemporal Attention for Video-based Person Re-identification”, CVPR, 2018.
 - 17) Fu, Yang, et al.: “STA: Spatial-Temporal Attention for Large-Scale Video-based Person Re-Identification.”, AAAI, 2019.
 - 18) Chih-Ting Liu, Chih-Wei Wu, Yu-Chiang Frank Wang and Shao-Yi Chien: “Spatially and Temporally Efficient Non-local Attention Network for Video-based Person Re-Identification”, In BMVC, 2019.
 - 19) Qingqiu Huang, Wentao Liu, Dahua Lin: “Person Search in Videos with One Portrait Through Visual and Temporal Links”, In ECCV, 2018.
 - 20) Zhong, Zhun, et al.: “Re-ranking person re-identification with k-reciprocal encoding.”, ICCV, 2017
 - 21) Melekhov, Iaroslav, et al.: “Relative camera pose estimation using convolutional neural networks.”, International Conference on Advanced Concepts for Intelligent Vision Systems, Springer, Cham, 2017.
 - 22) Charco, Jorge L., Boris X. Vintimilla, and Angel D. Sappa: “Deep learning based camera pose estimation in multi-view environment.”, 2018 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), IEEE, 2018.
 - 23) Melekhov, Iaroslav, et al.: “Camera relocation by computing pairwise relative poses using convolutional neural network.”, Proceedings of the IEEE International Conference on Computer Vision Workshops, 2017.
 - 24) Balntas, Vassileios, Shuda Li, and Victor Prisacariu: “Relocnet: Continuous metric learning localisation using neural nets.”, Proceedings of the European Conference on Computer Vision (ECCV), 2018.
 - 25) En, Sovann, Alexis Lechervy, and Frédéric Jurie: “Rpnet: An end-to-end network for relative camera pose estimation.”, Proceedings of the European Conference on Computer Vision (ECCV), 2018.
 - 26) Sattler, Torsten, et al.: “Understanding the limitations of cnn-based absolute camera pose regression.”, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019.
 - 27) Schonberger, Johannes L., and Jan - Michael Frahm.: “Structure - from - motion revisited.”, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016.
 - 28) Lowe, David G.: “Distinctive image features from scale-invariant keypoints.”, International journal of computer vision, 60.2 (2004), 91-110, 2004
 - 29) Bay, et al.: “Surf: Speeded up robust features.”, European conference on computer vision, Springer, Berlin, Heidelberg, 2006.
 - 30) Detone, et al. “Superpoint: Self-supervised interest point detection and description.”, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018.
 - 31) 田村恭ら: “建築現場施工の合理化に関する研究 第二報サーブリック記号の提案”, 日本建築学会関東支部第33回学術研究発表会, pp.49-52, 1963.01
 - 32) 小林謙二: “建築技能者の作業負荷に関する研究__建具作業の生態に与える負荷”, 日本建築学会大会学術講演梗概集(東北), pp.2129-2130, 1982.10
 - 33) 蔡成浩, 嘉納成男: “モーションキャプチャデータを用いたOWAS手法の姿勢コードの判定”, 日本建築学会大会学術講演梗概集(近畿), pp.1409-1410, 2005.09
 - 34) “作業能率測定指針”, 日本建築学会, 1990
 - 35) Mohammad Ali Rajaei, et al.: “Comparative evaluation of six quantitative lifting tools to estimate spine loads during static activities”, Applied Ergonomics, 48 (2015), 22-32
 - 36) 茅原崇徳, 坂本二郎, 瀬尾明彦: “デジタルヒューマンモデリングを用いたシミュレーションベースの人間工学設計”, 設計工学, 第55巻, 第4号, 2020
 - 37) Zhe Cao, et al.: “Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields”, CVPR, 2017
 - 38) 石岡宏晃: “建築作業者の安全衛生管理の実現に向けた単眼カメラを用いた腰部負荷推定手法の基礎的検討”, 日本建築学会大会学術講演梗概集(北陸), pp.129-130, 2019年9月
 - 39) [github] CMU-Perceptual-Computing-Lab/openpose
<https://github.com/CMU-Perceptual-Computing-Lab/openpose>
 - 40) [github] MVIG-SJTU/AlphaPose
<https://github.com/MVIG-SJTU/AlphaPose>
 - 41) Yu-Jhe Li, Xinshuo Weng, Yan Xu, Kris M. Kitani: “Visio-Temporal Attention for Multi-Camera Multi-Target Association”, Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 9834-9844, 2021
 - 42) Yan Xu, Yu-Jhe Li, Xinshuo Weng and Kris Kitani: “Wide-Baseline Multi-Camera Calibration using Person Re-Identification”, Conference Paper, Proceedings of Computer Vision and Pattern Recognition (CVPR), June, 2021