

製造現場に向けた階層的な詳細行動認識

ファン チョンフィ

OKIでは、製造現場における人やモノ、道路における交通事象などのさまざまなミッションクリティカルな現場の課題解決に向けた画像センシングの研究開発を行っている。人手作業が欠かせない製造現場においても、より高度な品質確保や工程管理などの実現を目指し、画像センシングを用いた作業者の行動認識技術を開発している。この技術は、例えば、製造ラインにおける正しい作業手順の判定や作業時間の分析などの実現に非常に有用である^{1),2)}。この技術により、対象人物の骨格などの用途に応じた有効な情報を動画像から抽出し、時系列ディープラーニングモデルの処理により一連の細かな行動の識別が可能となる。

本稿では、行動認識性能を向上させるための一つの工夫である階層的な識別技術を紹介する。また、有効性を確認するために、製造分野を想定したデスクトップPCを組み立てるデータセット³⁾に対する検証結果も述べる。

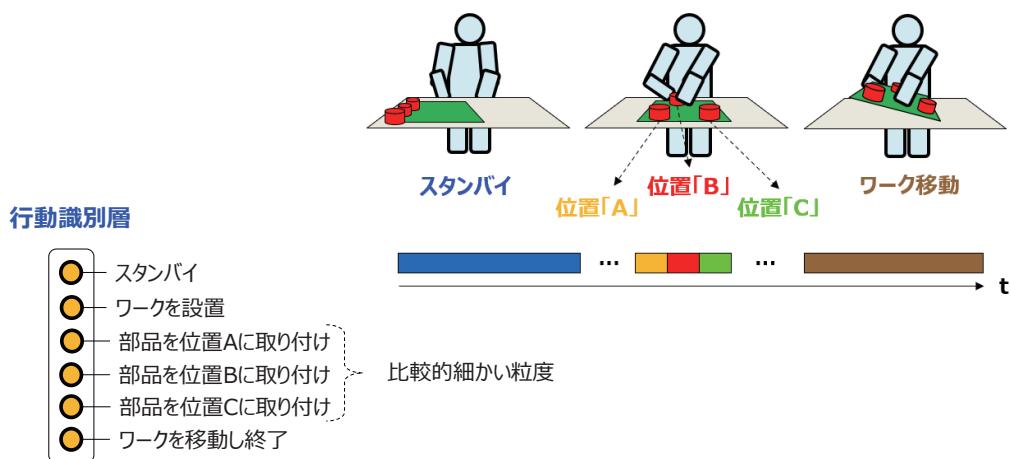
階層的な識別技術

●現行技術課題

動画像における人物行動の詳細な識別を実現するための研究が注目されており、ヘルスケア、製造などのさまざまな分野の実課題への適用が進んでいる。

画像センシングを用いた人物の行動認識は「『粗い』行動認識」(Coarse-grained Action Recognition)と「『詳細な』行動認識」(Fine-grained Action Recognition)に分類できる。「『粗い』行動認識」では、人物の体全体が映像中に映り、「走る」や「ジャンプする」や「座る」などの一般的な行動を認識する。このような行動の識別は、人物や背景などの「見え」情報及び全体の「動き」情報を用いることで実現できる。一方、「『詳細な』行動認識」は、主に腕などの体の一部に注目し「モノを取る」や「モノを置く」などの何かの作業に関する一連の細かな動作を識別するという比較的難しいタスクである²⁾。

日常生活にある「ジャンプする」、「走る」、「スポーツをする」などの背景も人物の動きもかなり異なるという粗い粒度の行動を識別するには、行動の数を分類するタスクとして扱うことができる。一方、製造現場のように一連に起きた「スタンバイ」、「テーブルに脚を取り付ける」、「脚を回す」などのテーブルを組み立てる際の行動は識別レベルを示す粒度が異なったり境界線が曇昧だったりするためフラットな学習が難しくお互いの誤識別も発生する。このように、対象の行動数が多く連續して行動が起きると境界線が曇昧になるため、従来のように全行動を同様にフラットに学習・識別することは困難になるという課題がある。



●階層的な識別

一連に起きる行動はアノテーションの定義によって粒度がさまざまである。図1に示す例では、より粒度が細かい三つの「位置別の部品を取り付け」行動は、互いに識別する必要があるだけでなく「スタンバイ」などのより粗い粒度の行動との識別も必要であるため、学習が困難である。そこで、識別性能を向上させるために、我々は対象行動の粒度をもとに適切にグルーピングを行い、従来の一つの難しい識別タスクを複数のより簡単な識別タスクに階層化して識別する技術を開発した。

行動の粒度は、上位の階層で粗く、下位の階層で細かく分類する。人物の状態のみを表す(例:スタンバイ中)は、上位の階層に粗く、物体に対する同じ動作を示す(例:部品Xを取ると部品Yを取るなど)は下位の階層でより細かく分類する。具体的な動作がどのように行われるかは、さらに下位の階層で詳細なグループにする(例:ネジを位置Aに締めるとネジを位置Bに締める)。また、この仕組みでは、アノテーションされた行動の粒度が細かい場合はより上位で粗い階層に属する代表的な行動クラス(例:作業中やネジを締めるなどを新たに設ける工夫が必要となる。

例にした部品組立てシーンに対して、もとのフラットな対象行動を、図2のように粒度が異なる三つの階層にして識別の学習及び推論を行う。最も粒度が細かく動きが似たような「部品を位置A/B/Cに取り付ける」という三つの行動を一つのグループにまとめ、それらの代表である「部品を取り付け」をより粗い粒度の層に設定する。このように、各グループでの識別対象を2~3と少数の行動とすることで、よりバランスよく学習を行える。また、行動を推論するときは階層的に行うことで、別の階層や別のグループにある行動との誤識別を防ぐことができる。例えば、フラットな識別

では「部品を取り付け」が次の行動である「ワークを移動」などに間違えられることがあり得るが、階層的な識別ではA/Bだけと他の位置への誤識別しか発生しない。そのため、位置別の部品取付けのような詳細な判定が困難な場合は「部品取付け」という適切な推論結果を出力可能なメリットがある。さらに非標準的な作業やその他の行動の発見にもつながる(図3)。

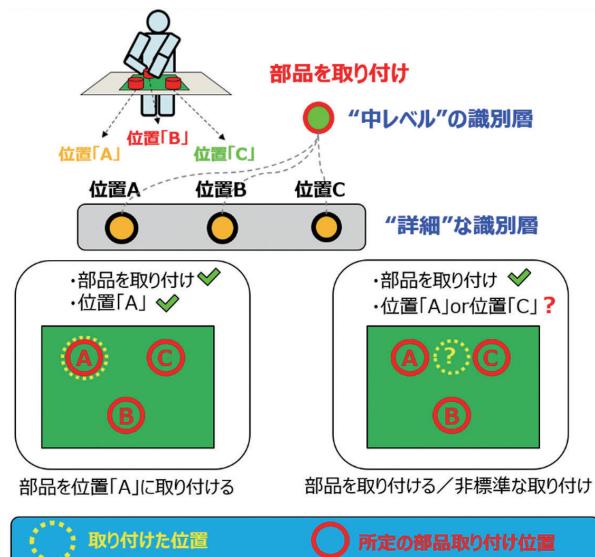


図3 階層的な識別結果を活用する例

検証内容

●検証データセット³⁾

検証対象のデスクトップPC組立てタスクは、被験者が机の上に置いてあるファン、RAM、SSDなどの複数の部品

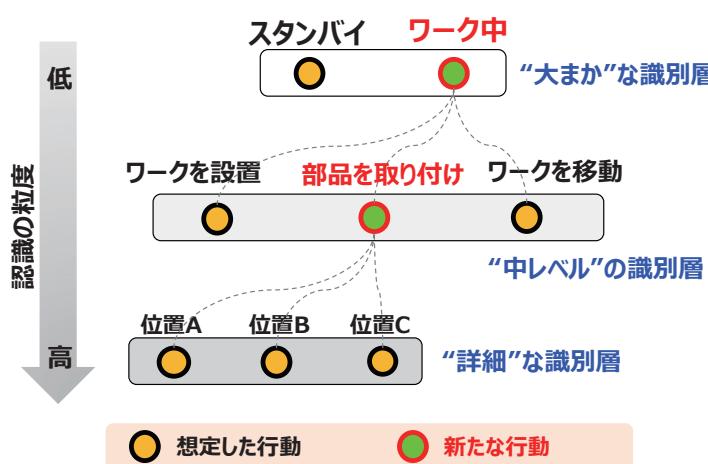


図2 部品取付けシーンの例にある一連の行動での階層的な識別

を本体に取り付ける作業である(図4)。取り付ける際には部品によってドライバーとネジという道具も利用する。識別対象の行動は、「Pick_up_RAM／Install_RAM」(動作と対象物体)や「Tighten_screw_1／Tighten_screw_2」(動作、対象物体、位置)のように詳細な情報までアノテーションされている。

また、本稿では階層的な識別を実施時は図5に示すように三つの階層で全6グループに行動ラベルの整理を行い、学習及び推論を行う。

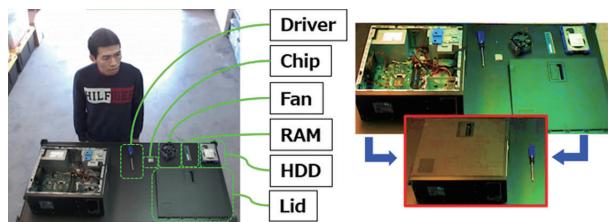


図4 デスクトップPC組立てタスク

検証・分析結果

我々は時系列のディープラーニングモデルと骨格情報を用いて検証を行い、有効性を確認した。表1に示すとおり、フラットな識別と比べて、階層的な識別のほうが全体的に性能が向上している。これは、階層ごとに識別対象の行動が同レベルで少数であるため、学習がより適切に行われていることが要因だと考えられる。全20クラスのフラットな識別では、「部品を取る」や「部品を取り付ける」(Pick_up／Place行動など)という同じ動きで対象部品の場所が近いとお互いの誤識別が起こりやすい。また、「より詳細にネジ締め」(Tighten_screw行動)のような位置別の識別が必要な場合でも同じの誤識別の傾向がみられる。一方、階層

的な識別では、図6に示すように上位の階層で大まかな分類を担い、下位の階層で詳細な識別を担うように役割分担をすることで誤識別が回避できる。例として、Pick_up_chip行動に対しては、作業中(In_progress)で何かを取っており(Pick_up)対象部品がChipであることと識別する。このように、従来と比べデスクトップPCの組立ての細かな行動をより高精度に認識できるため、手順間違の判定や各ステップの作業時間の計測などの想定アプリケーションに活用可能である。

表1 フレーム単位の正解率(%)での評価結果

階層／グループ	行動クラス	階層的	フラット
大まかな識別 (2クラス)	In Progress	99.5	—
	Background	92.6	92.7
中レベルの識別 (6クラス)	Pick_up	89.4	—
	Place/Install	91.2	—
	Tighten_screw	96.8	—
	Put_down (Screwdriver)	82.9	76.9
	Connect	92.6	—
	Close (Lid)	98.6	96.8
詳細な識別 グループ① (取る行動 6クラス)	Pick_up_chip	96.9	81.8
	Pick_up_screw_and_driver	97.3	79.3
	Pick_up_fan	90.3	74.7
	Pick_up_RAM	90.4	74.8
	Pick_up_HDD	88.7	77.8
	Pick_up_lid	95.1	82.3
詳細な識別 グループ② (置く行動 4クラス)	Place_chip_on_motherboard	95.9	88.3
	Place_fan_on_motherboard	94	68.8
	Install_RAM	95.4	90.7
	Install_HDD	99.3	86.8
詳細な識別 グループ③ (ネジ締め 4クラス)	Tighten_screw_1	92	86.1
	Tighten_screw_2	94	88
	Tighten_screw_3	93.1	93
	Tighten_screw_4	97.8	93.2
詳細な識別 グループ④ (接続する行動 3クラス)	Connect_wire_to_motherboard	98.2	84.3
	Connect_wire_1_to_HDD	91.7	83.1
	Connect_wire_2_to_HDD	89.9	76

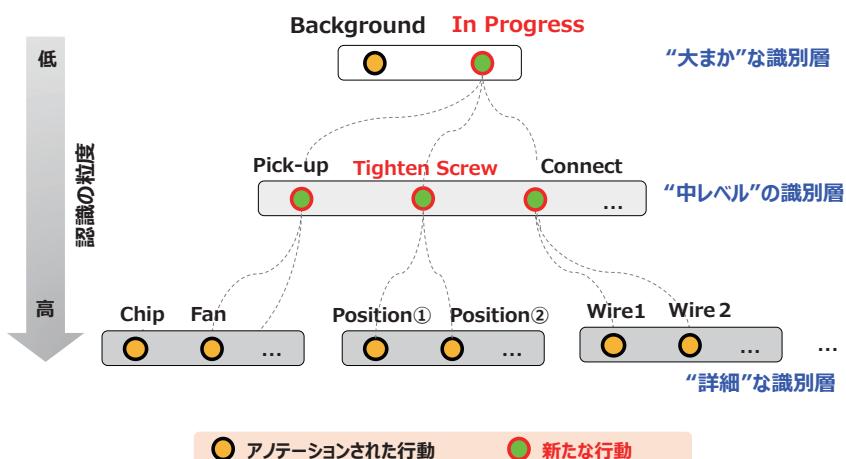


図5 デスクトップPC組立てデータでの階層的な識別



※緑字:正解、赤字:推論
 ※階層的な識別では三段階で表示
 (Coarse_Level:大きな識別、Mid_Level:中レベルの識別、Fine_Level:詳細な識別)

図6
 (上) 部品を取る行動に対する識別結果
 (下) 位置別のネジ締めに対する識別結果

●筆者紹介

ファンチョンフィ:Phan Trong Huy. 技術本部 研究開発センター AI研究開発部

TiPO【基本用語解説】

アノテーション
 テキスト、音声、画像、動画などのデータに、情報を分かりやすくするための「タグ」や「メタデータ」を付加する作業。特にAI分野では、AIに学習させるための「教師データ(正解データ)」を作成するプロセスであり、AIの性能を向上させるために不可欠。

まとめと今後の展望

本稿では、作業者の行動認識性能を向上させるための一つの工夫である階層的な識別技術を紹介した。また、製造分野を想定したデスクトップPC組立てデータセットを用いた検証を行い有効性の確認ができた。

現在、人手作業が欠かせない製造現場において、さまざまなタスク・アプリケーションに適用できるように詳細な行動認識技術の高度化に向けた取組みを継続している。今後、ムダな作業の発見につながる「作業分析」のような高度なモデリングの実現のために研究開発を行っていく予定である。



■参考文献

- 1) 川面怜哉、浅野将仁、蘭浩二、山本一真、小林司:作業内容や作業手順の正しさを判定する行為判定システム、OKIテクニカルレビュー第238号、Vol.88 No.2、pp.46-49、2021年11月
- 2) ファンチョンフィ、上野鷹幸、山本、一真:製造現場向け作業者の詳細な行動認識技術、OKIテクニカルレビュー第239号、Vol.89 No.1、pp.16-19、2022年5月
- 3) The Desktop Assembly Dataset
<https://retrocausal.ai/>