

キーポイント検出技術を用いた定量的行動観察：ビデオ映像による新しい行動観察手法の開発に向けて

西村，英伍
九州大学大学院芸術工学研究院人間生活デザイン部門

<https://doi.org/10.15017/7170833>

出版情報：芸術工学研究. 39, pp.19-22, 2024-03-11. Faculty of Design, Kyushu University
バージョン：
権利関係：



キーポイント検出技術を用いた定量的行動観察 ビデオ映像による新しい行動観察手法の開発に向けて

Quantitative Behavioral Observation Using Keypoint Detection Technology Towards the Development of a New Behavioral Observation Method through Video Imagery

西村英伍¹
NISHIMURA Eigo

Abstract

In behavioral observation within videos using keypoint detection technology, the movement of specific points (keypoints) is tracked to quantify behavior. The use of video cameras enables long-term observation and simultaneous observation of multiple people, but this increases the burden on observers. Therefore, the efficiency and automation of behavioral recording using video analysis technology have become important challenges for the future. The author is conducting research to expand quantitative methods of behavioral observation by combining video cameras with keypoint detection technology. There are two main challenges being addressed in this research. The first is improving the accessibility of keypoint detection technology. Despite technological advancements, the latest technologies are difficult to use for non-information technology students and researchers. To address this, the author is developing an intuitive application designed to handle various image recognition libraries. The second challenge is converting keypoints into meaningful behavioral data. An application is being developed to transform keypoint coordinate data into behavioral data, allowing for analysis of movement speed and direction of keypoints. Currently, these technologies are being implemented into applications to verify their practical applicability in behavioral observation.

1. 背景

動画内の人の行動を観察するとき、観察者は対象の人物の手や顔の動きに注目し、その様子を記録する。このとき注目する身体上の点を、画像処理の分野ではキーポイント(keypoint)と呼ぶ^{*1)}。ひとつのキーポイントは独立して以下の情報を有する。

- ・ 動画の各フレームの左上を原点とした x,y 座標
- ・ 動画の開始時点を原点とした時刻
- ・ フレーム間でキーポイントを同定するための ID

行動観察の方法の中には行動を定量的に分析するものがあり、応用行動分析(ABA)や生産技術の分野で発展してきた。本稿ではそのような行動観察手法を定量的行動観察と呼ぶ。定量的行動観察においては、まず注目すべき観察対象の行動を定義し、その発生回数や継続時間、発生順序を記録することで行動を定量化することができる。さらに近年のカメラや記録デバイスの性能向上に伴い、直接目視だけでなくビデオカメラを用いた観察も可能になった。

ビデオカメラを導入することで、長時間にわたる連続観察や複数人の行動の同時観察が可能になるほか、同じシーンを繰り返し観察することや、行動の関係を過去にさかのぼって見直すといったことも容易になる。このほかにも動画に対する行動観察の方法には大きな発展の余地がある一方で観察者の負担は大きくなる。そこで、動画解析技術を用いて行動回数の記録を効率化、自動化し、従来目視では捉えにくかった複雑な行動を分析する手法が今後重要視されると考えられる。

連絡先：西村英伍, eigo@design.kyushu-u.ac.jp

¹ 九州大学大学院芸術工学研究院人間生活デザイン部門
Department of Human Life Design and Science, Faculty of Design,
Kyushu University

2. 目的

筆者はビデオカメラによる観察とキーポイント検出技術を用いて、これまで目視が中心であった人の定量的行動観察の方法を研究している。本節ではキーポイント検出技術を行動観察に応用するにあたっての課題を2つ挙げる。

2.1. キーポイント検出技術のアクセシビリティ

近年においてキーポイント検出技術は性能向上を続けており、複数種類の画像認識ライブラリや学習用データセットがオープンソースで公開されている。一方でこれらのライブラリや学習用データセットを行動観察に利用したアプリケーションは少なく、情報系以外の分野の学生や研究者、デザイナーが最新の技術にアクセスすることが困難な状況にある。そこで筆者は、複数の画像認識ライブラリの仕様の差異を統合しながら、情報系以外の学生でも少しの訓練でキーポイント検出技術を行動観察に応用できるようなアプリケーションを開発している。

2.2. キーポイントから行動データへの変換

検出したキーポイントは、あくまで動画の各フレームにおける x, y 座標でしかないため、観察の目的に応じて意味のある行動データに変換する必要がある。たとえばフレーム間のキーポイントの移動量を計算することで、そのキーポイントの移動速度を推定することができる。また後述のように、複数のキーポイントの位置関係からその人が向いている方向を推定することができる。筆者は、キーポイントの時系列データをそういった意味のある行動データに変換するアプリケーションを開発している¹⁾。

3. 近年の有力なキーポイント検出技術

本節では、近年の有力なキーポイント検出として、YOLOv8²⁾とMediaPipe³⁾ Holisticを挙げる。これらは2023年以降に発表、またはアップデートが継続されているライブラリであり、追跡性能や検出速度の点で優れている。

3.1. YOLOv8

2023年にUltralytics社が発表した画像認識ライブラリであり、オープンソース(AGPL-3.0)で公開されている。複数人の全身キーポイント検出と追跡が可能である点が特徴である。YOLOv8では動画内に複数の人が写っている場合でも入れ違いや見落としが発生しにくいいため、撮影条件のコントロールが困難なフィールド調査や複数人が参加する実験にも使用が可能である。

YOLOv8をはじめ、多くの全身キーポイント検出では図1に示す17のキーポイントが使用される。複数人を検出する場合を想定して、各部位のIDのほかに、個人を識別するコードを別途割り当てる必要がある。



図1 YOLOv8の全身のキーポイントID

3.2. MediaPipe Holistic

Google社が2019年に発表し、2024年現在もメンテナンスが継続されている画像認識ライブラリであり、オープンソース(Apache License 2.0)で公開されている。MediaPipe Holisticは全身および右手と左手のキーポイントと顔のランドマーク^{*1)}検出が可能である。右手と左手を識別できることから、人の表情の分析のほか机上での作業の詳細な分析が可能である。

MediaPipe Holisticをはじめとする多くの手のキーポイント検出では、図2に示す21のキーポイントが使用される。MediaPipe Holisticでは手の各部位のIDとは別に、右手と左手を識別するコードが別途付与される。なお顔のランドマーク数は478点である。

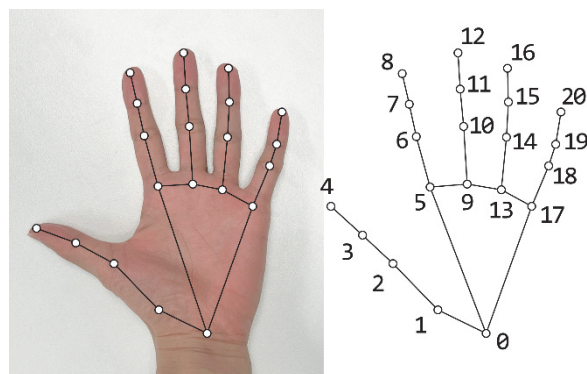


図2 MediaPipe Holisticの手のキーポイントID

4. キーポイントを用いた行動観察方法

上に挙げた画像認識ライブラリはいずれも高い検出・追跡性能を有するが、行動観察を目的に開発されているわけではない。そこで検出したキーポイントを観察にとって意味のあるデータに変換する機能やデータを可視化する機能を実装する必要がある。本節では検出したキーポイントデータ群を観察に有用な時系列データに変換するための手法について述べる。

4.1. 方向推定

動画内の人が、いつどの方向に向きを変えたか、どれくらいの間同じ方向を向いていたかといった情報は行動観察において重要である。これまでは、人のシルエット形状から向いている方向を推定する方法や、深層学習による画像分類を用いた方法などが考案されているが、両肩と左右の股関節の座標のキーポイント座標を使用したベクトル計算で推定することも可能である。

二次元平面に投影された四角形の裏表は、隣り合う二辺がつくるベクトルの外積から判別できることから、両肩と左右の股関節のキーポイント座標からその人が正面を向いているのか、背を向けているのかを推定することが可能である。たとえば図3において、キーポイントID=6(右肩)からID=5(左肩)へのベクトルを $\mathbf{v}$ 、ID=6からID=12(右股関節)へのベクトルを $\mathbf{w}$ としたとき、左手系においてその外積 $\mathbf{v} \times \mathbf{w}$ の符号は対象者がカメラに対して正面を向いているときに正、背を向けているときに負となる。

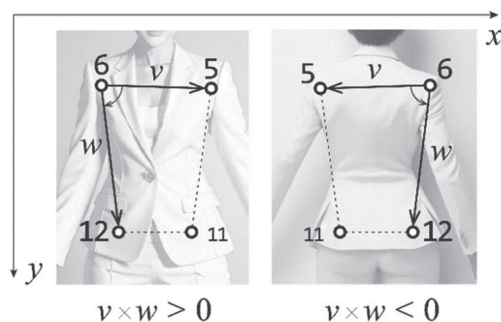


図3 外積を用いた正面/背面推定

また、二次元平面に投影された四角形の鈍角と鋭角は隣り合う二辺が作るベクトルの内積から判別できることから、遠近法を用いて両肩と左右の股関節のキーポイント座標からその人が右を向いているのか、左を向いているのかを推定することが可能である。

たとえば図4左のように、対象者が右を向いている場

合4つのキーポイントは平行四辺形に歪み、 $\mathbf{v}_5 \cdot \mathbf{w}_5 < 0$ かつ $\mathbf{v}_{11} \cdot \mathbf{w}_{11} > 0$ となる。あるいは、カメラの取り付け高さや絞りの設定によっては図4右のように台形に歪み、 $\mathbf{v}_6 \cdot \mathbf{w}_6 < 0$ かつ $\mathbf{v}_{11} \cdot \mathbf{w}_{11} > 0$ となる。

これらのキーポイントは、正面を向いているときには $\mathbf{v}_{11} \cdot \mathbf{w}_{11} < 0$ かつ $\mathbf{v}_6 \cdot \mathbf{w}_6 > 0$ かつ $\mathbf{v}_5 \cdot \mathbf{w}_5 > 0$ となることから、内積の正負の組み合わせによって右を向いているか左を向いているかを推定することが可能となる。

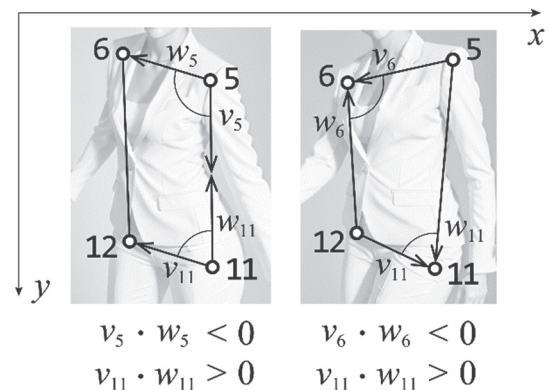


図4 内積を用いた右向き推定

以上のように、両肩と股関節の4つのキーポイントが作るベクトルの外積と内積の正負判別による正面背面推定と左右推定を組み合わせることで、4方向の身体の向きの推定が可能となる。

さらに外積による正面背面推定において、外積の値が十分に0に近いときは正面でも背面でもなく、真横を向いていると推定できる。これはベクトル $\mathbf{v}$ と $\mathbf{w}$ がなす角を θ としたとき、以下の式より θ が 0° または 180° に十分に近いかを考えると同じである。すなわち

$$\frac{\mathbf{v} \times \mathbf{w}}{\|\mathbf{v}\| \|\mathbf{w}\|} = \sin \theta \approx 0$$

のとき真横を向いている。

また内積による左右推定において、内積の値が十分に0に近いときは、右でも左でもなく、真正面または真背面を向いていると推定できる^{*2)}。これはベクトル $\mathbf{v}$ と $\mathbf{w}$ がなす角 θ が 90° に十分近いかを考えると同じである。すなわち

$$\frac{\mathbf{v} \cdot \mathbf{w}}{\|\mathbf{v}\| \|\mathbf{w}\|} = \cos \theta \approx 0$$

のとき真正面あるいは真背面を向いている。

以上より、図5に示す8方向の身体の向きを両肩と両股関節の4つのキーポイント座標から推定することが可能である。

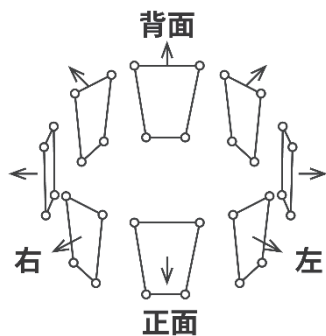


図5 身体のキーポイントによる8方向推定

4.2. 滞在時間推定

人が右から何番めのいすに座ったか、手が机の上に何秒間置かれたかなど、人やその人の部位の場所に関する情報は行動観察において有用である。

動画内の所定の領域に対し、キーポイントがその中にあるか外にあるかは、領域の各頂点とキーポイントがつくるベクトルの外積を計算することで判別できることが知られている。

たとえば図6において、領域ABCDの各頂点からあるキーポイントXへのベクトルを $\overrightarrow{AX}, \overrightarrow{BX}, \overrightarrow{CX}, \overrightarrow{DX}$ 、領域の頂点同士のベクトルを $\overrightarrow{AB}, \overrightarrow{BC}, \overrightarrow{CD}, \overrightarrow{DA}$ としたとき、

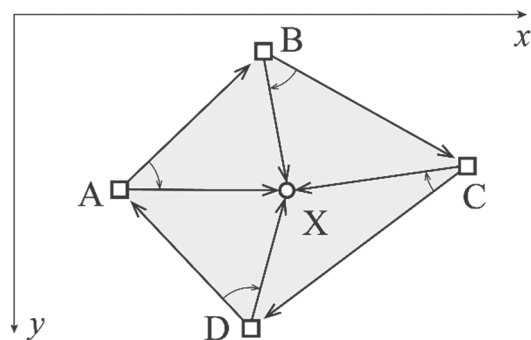


図6 外積によるキーポイントの領域内外推定

左手系において $\overrightarrow{AB} \times \overrightarrow{AX} > 0$ かつ、 $\overrightarrow{BC} \times \overrightarrow{BX} > 0$ かつ、 $\overrightarrow{CD} \times \overrightarrow{CX} > 0$ かつ、 $\overrightarrow{DA} \times \overrightarrow{DX} > 0$ であれば、キーポイントXは領域ABCDの内側にあると判別できる。

ただし、動画はあくまで二次元平面に投影された像であるため奥行き方向の領域内外推定には不向きである。

5. 定量的行動観察に向けて

本稿では近年の有力なキーポイント検出技術と、検出されたキーポイントの座標データを行動観察にとって有用な時系列データに変換するための方法について述べた。現在はこれらの技術を実アプリケーションに実装し、行動観察に実際に応用可能であるかを検証している。

注

- *1) キーポイントに似た語にランドマーク(landmark)というものがある。たとえば顔の画像認識においては“keypoint”ではなく“landmark”が使用されることが多い。キーポイントは「画像中の特徴的な点」というニュアンスを含み、画像処理による検出・識別を主眼に置いた言葉であるのに対して、ランドマークは「画像中の意味のある領域」というニュアンスを含み、画像の理解に主眼を置いた言葉である。近年の深層学習を用いた画像処理技術の進展により、画像中の意味のある領域を学習させることで、任意の領域あるいは点を高精度で検出・識別することが可能になった。このことから、キーポイントとランドマークの意味の違いは以前と比べて小さなものになってきている。なおMediaPipeのドキュメントでは全身や手のキーポイントに対しても“landmark”の語を使用している。
- *2) 実際には股関節のキーポイントにおける角度はやや鈍角、方のキーポイントにおける角度はやや鋭角となるため実装にあたっては閾値の調整が必要となる。

参考文献

- 1) Nishimura, E. (2024). Behavior Senpai (Version 1.0.0) [Computer software]. Kyushu University, <https://doi.org/10.48708/7160651>
- 2) Jocher, G., Chaurasia, A., & Qiu, J. (2023). YOLO by Ultralytics (Version 8.0.0) [Computer software]. <https://github.com/ultralytics/ultralytics>
- 3) Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., ... & Grundmann, M. (2019). Mediapipe: A framework for building perception pipelines. arXiv preprint arXiv:1906.08172.