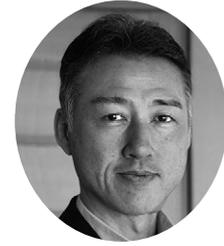


〈一般研究課題〉 全周映像を用いた人の行動計測と
VRを用いた行動支援に関する研究
助成研究者 中京大学 瀧 剛志



全周映像を用いた人の行動計測とVRを用いた 行動支援に関する研究

瀧 剛志
(中京大学)

Study on human behavior measurement and behavior support using 360-degree camera and VR device

Tsuyoshi Taki
(Chukyo University)

Abstract :

When people take an intentional action in a certain space (indoor/outdoor), the moving direction and line-of-sight direction of a person vary depending not only on the person's interest but also on the purpose and skill level. The final goal of this research is to estimate a person's movement and line of sight, and conversely to estimate the person's purpose and skill level from their behavior. For that purpose, it is necessary to track the position of the person and the direction of the line of sight. In this research, we will realize this only with a commercially available 360-degree camera and a PC. This research is especially aimed at cooperative or competitive behaviors of multiple people, and is expected to be applied to collaborative work in factories and analysis of team sports competitions. In many cases, the skilled person sensuously (or unconsciously) decides the remarkable point or the next movement. If the difference in the remarkable place and timing between skilled and unskilled people can be clarified, it can be an effective tool in succession of skills and in instruction.

This paper describes, as applications, the development of a tracking and VR viewing system for compulsory figures in skating, and an attempt of line-of-sight measurement in American football.

1. はじめに

複数の人々がある空間内(屋内/屋外)において、意図的な行動をとるとき、人の移動方向や視線方向は、その人の興味だけでなく、目的や熟練度によっても変わってくる。本研究の最終ゴールは、人の移動や視線を推定し、そこから逆にその人の目的や熟練度を推定しようとするものである。そのためには、まず人物位置や視線方向をトラッキングする必要がある。一般に人の移動方向や視線方向を計測するためには、被検者に何からのデバイスを装着する必要がある、人の自然な動きを阻害してしまう場合がある。また、外部から非接触で行動を計測するには、周囲にセンサを配置する必要があり、行動範囲が制限される、屋外で安定利用できない、導入コストが高いといった問題がある。そのため本研究ではシステムの可搬性を考慮し市販のビデオカメラを用いることにする。デスクワークや工場内作業など大きな移動を伴わない人の行動を計測する際は、光軸方向やズーム率を固定した1台のカメラで撮影可能だが、室内での撮影ではカメラと計測対象との距離が近く、少しの移動でもカメラの撮影範囲から外れてしまったり、屋外での大きな移動を伴う場合、1台のカメラでは対象となる人物の解像度が低下してしまう。そのため、複数台のカメラを利用したり、パン、チルト、ズーム制御可能なカメラを利用することになり、システム構成が複雑になってしまう。本研究では、これを市販の360度カメラとPCのみで実現することにする。360度カメラ映像を用いたトラッキングに関しては、8面キューブマップを利用して各面で物体検出し後に各面の結果を統合する方法[1]や、あらかじめレンズで歪んだ形状を学習させることでリアルタイムに物体検出する方法[2]など、現在様々な研究が取り組まれている。本研究は、特に複数人の協調的あるいは競合的な行動を対象としたもので、工場内の共同作業やチームスポーツ競技の分析等への応用が想定される。熟練者は感覚的(無意識)に注目すべき場所や次の動きを決めている場合が多い。熟練者と非熟練者で注目する場所やタイミングの違いを発見するシステムが実現できれば、技術の伝承や指導の場において有効なツールとなりえる。

2. 全周映像を用いた人の行動計測

2.1 全周映像からの人物検出

本研究では小型で比較的解像度の高い360度カメラであるRICOH社のTheta VまたはTheta Z1[3]を用いる。Thetaは図1に示すようにカメラの両面にレンズが搭載され、それぞれのレンズで撮影された画像が横に並んだ映像が記録される。この映像を同社の専用アプリで変換することで、図2のように正距円筒図法という手法で1つの連続する画像として合成される。

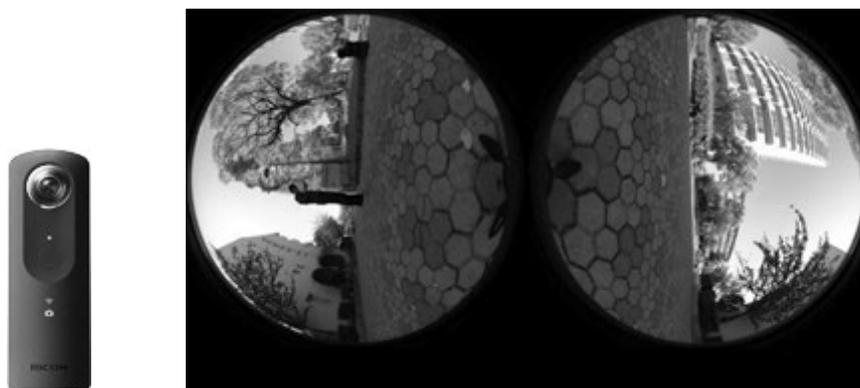


図1. 360度カメラTheta V(左)と撮影された映像の一例(右)



図2. 合成された映像の一例

人物の検出には、Deep Learning (深層学習)により単一画像のみから複数人物の姿勢推定が可能な骨格検出ライブラリOpenPose[4]を用いる。OpenPoseでは通常のカメラで撮影された一般的な人の姿勢を主に学習しているため、図1に示すような魚眼レンズで撮影されたような画像では人物の骨格が歪み安定した人物検出が行えない。よって、図2に示すような画像を入力とする。OpenPoseでは人物の骨格が25点のキーポイント(関節点)で構成され、各キーポイントは画像上での2次元座標 (x, y) と信頼度の3つの値で表現される。オクルージョン等で関節位置が不明瞭な場合はこの信頼度が低くなり、未検出の場合は0となる。OpenPoseによって人物が検出された場合でも、その人物に含まれる関節点(キーポイント)が極端に少ない場合は誤検出の可能性もあるため人物と見なさないこととする。

2.2 人物のトラッキング

複数人物が映る映像の中から、ある特定の人物の移動を計測するためには、フレーム間で同一人物かどうかを判定する必要がある。本研究では図2のような画像を対象とするため、人物がカメラの周りを周回するような行動をとると、例えば画像の右端で消えた人物は、画像の左端に現れることになる。また、画像の端で人物領域が見切れてしまうと人物の姿勢を正しく推定することができない。そこで、画像の右端部分の矩形領域を複製し、それを画像の左端部分に付加する、同様に、画像の左部分の領域を画像の右端部分に付加することで、画像の両端を拡張し、画像端での人物検出精度の安定を図る。図3はオリジナル画像の右半分を左端に、左半分を右端に付加してできた画像の例である。なお、図中のオリジナル画像と付加画像の境界線は説明のため書き込んだもので実際の画像には存在しない。



図3. 両端を拡張した画像

はじめに、映像中に人の密集が見られず、計測対象が一人のみの場合のトラッキング手法について述べる。まず、画像全域を対象にOpenPoseを適用する。複数人物が検出された場合は、人物領域の面積が大きいものを初期対象人物とする。ただし、ここでいう人物領域とは、その人物に対して得られたキーポイントすべてを囲う最小矩形とする。次に、得られた人物領域の中心座標を中心に人物領域を縦横に定数倍した矩形領域内を探索範囲と設定し、次フレームにおいては、その範囲内に限定して人物検出を適用する。これによって処理の高速化と誤追跡の抑制を図る。図4に検出された骨格情報、人物領域(内側に描かれた白色の矩形)および探索範囲(外側の矩形)の一例を示す。また、360度カメラでは全周を1枚の画像に収めるため、一般のビデオカメラと比べると人物領域の解像度はそれほど高くない。対象人物がカメラ位置から遠ざかると、その人物領域の解像度は更に低下することになり、キーポイント検出の精度も低下する。そこで、この探索範囲内の画像を拡大(アップサンプリング)し、この拡大画像において人物検出を行うことで骨格情報の高精度化も図る。



図4. 人物領域と探索範囲作業

次フレームにおいて人物が検出できなかった場合は数段階に分け探索範囲を広げながら人物検出を試み、一定時間内に検出ができなかった場合は追跡処理を停止する。これらの処理を毎フレーム繰り返し適用することで対象人物のトラッキングを実現する。なお、初期フレームで設定された人物が目的の対象人物と異なる場合や、途中で対象人物を見失ったり、別の人物の追跡を始めた場合は、画像上で目的の対象人物をマウスクリックすることによって手動で追跡対象を指定する機能も実装する。

計測対象が複数の場合においても上記手法をベースとし、探索範囲内に複数の人物領域が検出された際には、現フレームと1つ前のフレームで検出された人物に対して、それぞれの骨格情報の類似度により同一人物かどうかを推定する。骨格情報の類似度は次式で定義する。

$$S = \alpha S_{len} + (1 - \alpha) S_{ang} \quad \text{ただし、} 0 \leq \alpha \leq 1$$

S_{len} は骨格の長さに関する類似度で、各関節間の長さを要素とするベクトル間の正規化相互相関、 S_{ang} は姿勢に関する類似度で、各関節角度の大きさを要素とするベクトル間の正規化相互相関である。最終的な類似度 S は、この両者の値を重み係数 α で加算したもので、この値が大きいものから順に同一人物として対応付ける。なお、対象人物がカメラ位置から見て急に離れたり近づいたりするような場合は骨格の長さ変化が大きくなるため、 α の値は小さく設定する。

2.3 骨格情報の補正

OpenPose自体は映像中の各フレームを独立に処理するため、各キーポイントの座標を時系列で扱うことで、信頼度が低い、あるいは、未検出のキーポイント位置をオフラインで補正・補間する。ここではキーポイントの信頼度があるしきい値以下の場合、次の方法でそのキーポイントの位置を再計算する。

- (1) あるキーポイントを時系列にスキャンし、信頼度がしきい値以下のもの、および、しきい値を超えるキーポイントが2つ以上連続しないもの(その前後が共にしきい値以下のもの)は未

検出扱いとし除去する。

- (2) (x 座標値の補間)：キーポイントを、その x 座標とフレーム番号 t の2次元ベクトル (x, t) の時系列データと考え、未検出区間の前後に存在するキーポイントの位置ベクトルをそれぞれ P_0 、 P_1 、そこでの接ベクトルをそれぞれ V_0 、 V_1 とし、次式で未検出区間のデータを補間する。

$$C(p) = P_0 H_0^3(p) + V_0 H_1^3(p) + V_1 H_2^3(p) + P_1 H_3^3(p) \quad \text{ただし、} 0 \leq p \leq 1$$

ここで、 $H^3(p)$ は3次エルミート関数であり次のように定義される。

$$H_0^3(p) = (2p+1)(1-p)^2$$

$$H_1^3(p) = p(1-p)^2$$

$$H_2^3(p) = -p^2(1-p)$$

$$H_3^3(p) = p^2(3-2p)$$

なお、 $p=0$ のとき P_0 、 $p=1$ のとき P_1 となり、未検出区間のフレーム番号 t に応じてパラメータ p の値を指定することで t フレーム目のキーポイントの x 座標を得る。

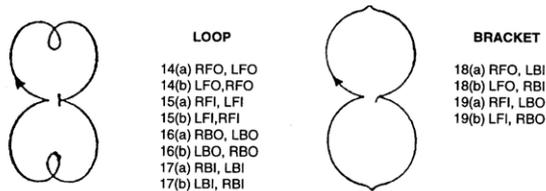
- (3) (y 座標値の補間)：上記(2)の x を y として同様に y 座標値を算出する。
(4) 未検出区間の数だけ上記(2)と(3)を繰り返す。ただし、未検出区間が長く、その間に関節位置が不規則に移動すると自然な補間結果が得られない場合が多いため注意が必要である。

3. コンパルソリーフィギュアのVR視聴システム

前章で述べた360度カメラ映像からの人物トラッキング手法の有効性評価と、その応用可能性を示すため、コンパルソリーフィギュアという競技への適用例を示す。コンパルソリーフィギュアとは、氷上を滑走し与えられた課題図形を描くときの正確さや滑走姿勢を競うスケート競技の一種である。現在はフリースケーティングが主流となり見る機会はほとんどないが、本システムの開発はスポーツ文化や歴史の継承という目的も併せ持っている。図5は課題図形の一例であり、その輪郭線に沿って前向き／後ろ向き、右足／左足、使用するエッジ(スケートの刃の内側か外側かなど)を使い分けて滑走する。本章では、コンパルソリーフィギュア映像からのスケータートラッキング結果および滑走軌跡の可視化結果を示す。

3.1 トラッキング結果

本学アイスアリーナにおいてスケート部員2名の協力のもとコンパルソリーフィギュアの撮影を実施した。図6に示すように、アイスリンク上に三脚を立て、その上にリコー社製360度カメラThetaを配置した。なお、部員2名はコンパルソリーフィギュアの経験はなく、撮影前に課題図形を提示し、その後数分間練習した後に撮影を行った。まず、2.2節で述べた探索範囲の拡大処理がない場合とある場合で結果を比較するため、同一シーンに対して、図7では拡大処理なし、図8では拡大処理ありの結果を示す。各図の8シーンそれぞれには探索範囲を示す矩形と検出された骨格が描画されているが、拡大処理なしの結果である図7の2段目左の2つのシーンについては骨格情報が現れていない。それに対して拡大処理ありの図8では拡大処理によって十分な解像度が確保され問題なく骨格情報が検出され人物トラッキングが行えていることが分かる。また、検出された骨格(関節)の位置を比較すると、図8の方がより人物のシルエットに合致した結果が得られていることも分かる。これらの結果から、拡大処理が骨格検出の精度向上に有効であることを確認した。



<https://allaboutwork.org/2015/11/04/compulsory-figures-technical-guidelines-diagrams-and-tests/>

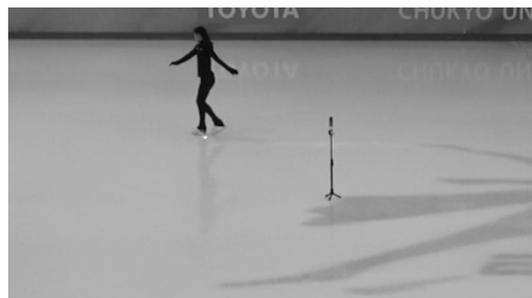


図5. 課題図形の一例

図6. 撮影風景

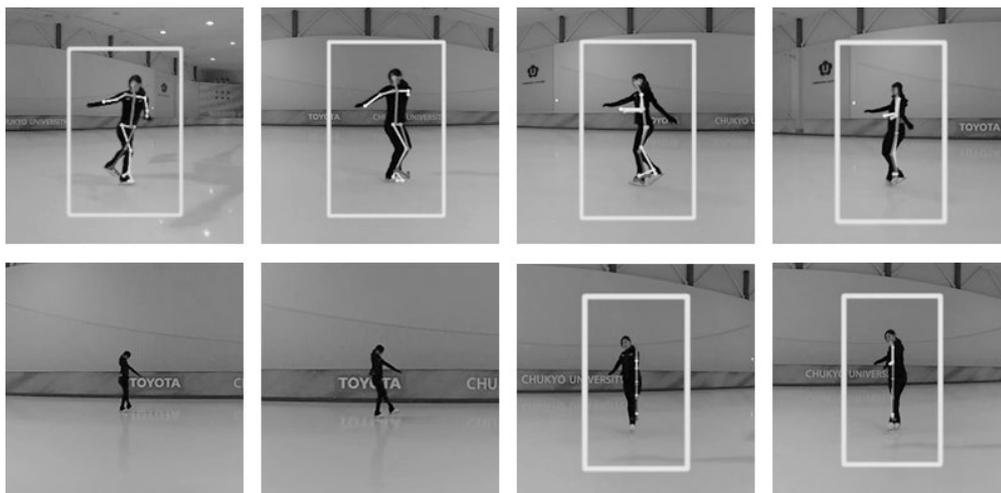


図7. 骨格検出とトラッキング結果(拡大処理なし)

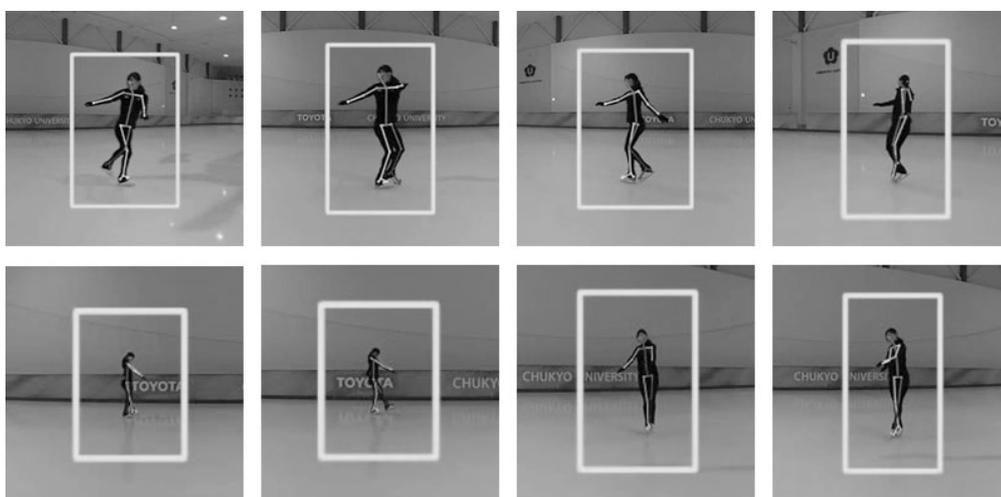


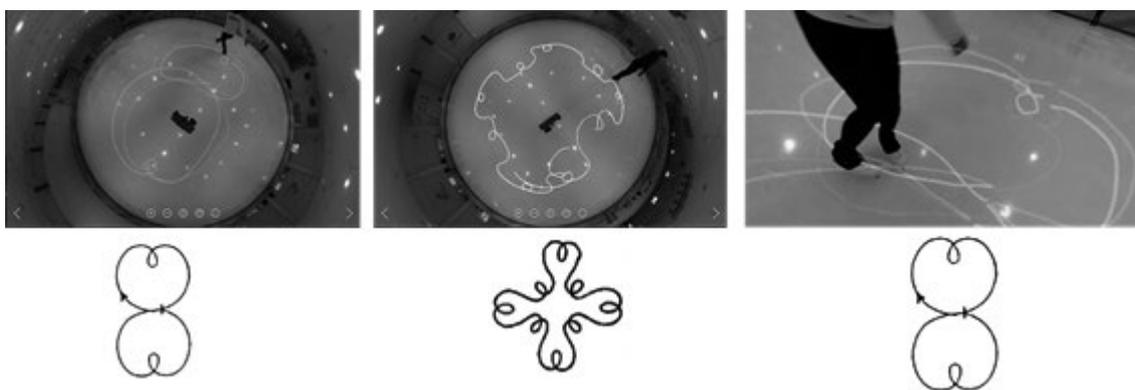
図8. 骨格検出とトラッキング結果(拡大処理あり)

次に、各フレームで得られた骨格情報から、その最下点となる関節位置を氷面との接点と見なし、その時系列変化を線分で結び移動軌跡を算出する。図9は同じ課題図形を3回連続で滑走した結果であり、その試行ごとに滑走軌跡を異なる色で描画したものである。いずれも線が途切れることなく滑らかな軌跡が得られ、目視による評価でもスケーターの足元を問題なくトラッキングできていることを確認した。氷上に課題図形や目印がある訳ではないため試行ごとにばらつきがあることも見て取れ、どれくらい正確に、かつ、再現性高く演技できたかを評価するツールとして有効で

あると考える。また、図10に示すように、得られた滑走軌跡を俯瞰的に見たり、VRゴーグルを装着して演技と軌跡を同時に見ることも可能となり、新しい競技の見せ方、トレーニングやコーチングのツールとしても期待できる。今回は軌跡を描画するところまでの実現に留まったが、その正確性や再現性を定量評価する機能を実装することで、より実用的なシステムへと発展させたい。



図9. 滑走軌跡の算出結果



上段: 実際の滑走軌跡 下段: 課題図形

図10. 滑走軌跡の表示とVR視聴

4. まとめと今後の展望

本稿では、人物の移動と行動を計測するために、360度カメラを用い、その映像から人物をトラッキングする方法について検討した。その中で、人物検出およびトラッキングの精度を向上させるため、人物が見切れてしまうことに対する画像端の処理、カメラ位置から離れることで小さく映る人物に対する処理、時系列情報を用いて関節位置を補間する処理について述べ、コンパルソリーフィギュアを対象に人物(スケーター)の移動軌跡が安定して検出できることを示した。また、本研究では人物の移動計測の他に、人物の視線計測も目的の1つに挙げていた。これに関しては具体的な対象としてアメリカンフットボール競技を取り上げ、司令塔のポジションであるクォーターバックの選手のヘルメットに360度カメラを取り付け、実戦形式の練習風景映像を収集した。その一例

を図11に示す。中央の図は視線が正面に向いた場合で、左右の図はその時の視線方向を360度映像上で左右に約45度傾けた図である。ヘルメットによって通常よりも視野が狭まり周囲を見るためには首を振る動作が必要となるが、今後は、この首振りのタイミングや角度をこの360度カメラ映像から検出し、パスする相手や敵選手の動きをどう見ているのかを分析する。また、熟練した選手とそうでない選手で、そのタイミングや角度がどれくらい異なるのかを明らかにするために、同じ映像をVRゴーグルで提示し、その際の首振りのタイミングや角度の差を計測する予定である。



図11. アメリカンフットボール選手の視点映像の一例

参考文献

- [1] 井上慶彦, 岩村雅一, 黄瀬浩一, “全方位カメラを用いた物体検出とトラッキングー|視覚障害者支援システムの実現に向けてー”, |情報処理学会研究報告CVIM, 2018
- [2] 林田和磨, 横山昌平, 全天球カメラにより配信される正距円筒図法動画からのリアルタイム人物検出, 第12回データ工学と情報マネジメントに関するフォーラムDEIM2020, C6-2, 2020
- [3] RICOH THETA公式サイト <https://theta360.com/ja/>
- [4] Zhe Cao, Tomas Simon, Shih-En Wei, Yaser Sheikh, “Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields”, Computer Vision and Pattern Recognition, arXiv:1611.08050, 2017