

# 骨格推定を用いた人型視線ヒートマップの作成

中村 智喜<sup>†</sup> 三上 弾<sup>‡</sup>

<sup>†</sup> 工学院大学情報学部コンピュータ科学科

〒163-8677 東京都新宿区西新宿 1-24-2

E-mail: <sup>†</sup> j119215@g.kogakuin.jp, <sup>‡</sup> mikami.dan@cc.kogakuin.ac.jp

**あらまし** スポーツの練習ではしばしば、自身のフォームや熟練者のフォームを動画で分析することがある。トレーニングや指導の場面でこのように動画で分析する際、注視している場所は非常に重要である。注視している場所を可視化する手法としてヒートマップがあるが、被写体が動いてしまう場合、どの関節を注視しているかという観点で適切な表示をすることが困難である。さらに、より効果的なフィードバックを行うためには、視線の分析結果を素早く提示する必要がある。そこで本研究では、骨格推定を用いて人型視線ヒートマップを作成し、それをリアルタイムに提示する手法を提案する。評価実験では、人体画像上に視聴部位に基づいたヒートマップを適切に作成できていて、ヒートマップを提示するまでの時間が早いことを確認した。

**キーワード** アイトラッキング, 視線, 骨格推定, ヒートマップ

## Creation of human gaze heat maps using skeletal estimation

Satoki Nakamura<sup>†</sup> and Dan Mikami<sup>‡</sup>

<sup>†</sup> Department of Computer Science, Faculty of Informatics, Kogakuin University

1-24-2 Nishi-Shinjuku, Shinjuku-ku, Tokyo 163-8677, Japan

E-mail: <sup>†</sup> j119215@g.kogakuin.jp, <sup>‡</sup> mikami.dan@cc.kogakuin.ac.jp

**Abstract** In sports, one may analyze one's own form or the form of a skilled athlete on video. When analyzing such video in training or instructional situations, the location of the gazing point is very important. A heat map is a method to visualize the location of gazing. However, when the subject moves, it is difficult to properly display the location of gazing in terms of which joints are being gazed at. Furthermore, for more effective feedback, the results of gaze analysis need to be presented quickly. Therefore, we propose a method to create a human gaze heat map using skeletal estimation and present it in real time. In evaluation experiments, we confirmed that the method was able to properly create a heatmap based on the viewing area on the human body image and that the time to present the heatmap was fast.

**Keywords** Eye tracking, eye gaze, bone structure estimation, heat map

### 1. はじめに

スマートフォンやタブレットの普及で、スポーツにおいて自身の動作や、熟練者の動作を撮影し、動画で確認するという事は非常に一般的になってきている。映像によって自身の動作や熟練者との違いが確認できることは上達において大きなメリットがあると考えられる。

一方、本研究における著者らの着眼点はその時の視線である。「目は口ほどにものをいう」ということわざがあるほど、視線には多くの情報が含まれる。視線情報を取得することで動作の振り返り時にチェックポイントをしっかりと確認できているか、コーチと選手とで同じポイントを見て議論しているかを確認可能になる。

それだけにとどまらず、選手が気になっているところに無意識に多く視線を向けることも考えられ、課題点をあぶり出すなど、コミュニケーションの質を高める手助けとなる可能性も考えられる。

人間の視線情報を取得する技術はアイトラッキングと呼ばれ、アイトラッキングのための装置はアイトラッカーと呼ばれる。多くのアイトラッカーでは、人の眼球を検出し、その位置などから、視線位置を計測する。中でもメガネ式アイトラッカーは、頭部姿勢に対する制約が低いことから多くの利用例がある。山品らは、マンモグラフィの検査現場における熟練者と初學者の視線を比較し[1]、Pohl らは、グラフィックアウトの読みやすさの比較においてアイトラッカーを利用している[2]。

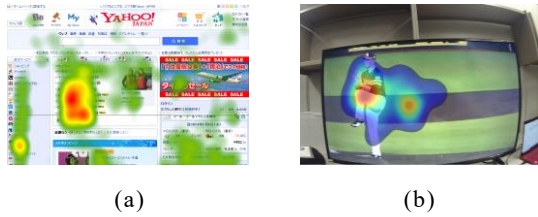


図1 ヒートマップの例

ところで、アイトラッキング結果の可視化方法としてヒートマップが広く使われている。Zamani らは、医療画像を用いた視線分析の際にヒートマップを利用し[3], Nakano らは、マーケティング製品における消費者行動の調査の際にヒートマップを利用している[4]。ヒートマップは注視点位置毎の注視頻度を色によって可視化したものである。これは、一般的に静止画像を対象として視線の集まり方を可視化するものである。図1(a)に Web サイト視聴時のヒートマップを示す。ヒートマップでは、赤が視聴頻度が高く、黄色、緑、塗りつぶしなしと頻度が低下する。Web サイトでは、しばしば左上から右下へ視線が移動すると言われるが、図1(a)の例でもそのような傾向が見て取れる。加えて、ニューストピックが一覧となっている部分に注視が集まっていることも確認できる。

図1(b)にもう1つのヒートマップの例を示す。これは、練習の振り返りシーンを想定したものであり、画像中央にディスプレイがあり、そのディスプレイ内で選手の動作が再生される。図1(b)は、これを視聴した際のヒートマップである。図1(b)を確認すると、左ひざを注視しているように感じられる。しかしながら実際の視聴においては、左ひざを中心に注視していることはなく、実態とはかけ離れたヒートマップとなってしまう。

このような背景から本研究では、動作する人物を撮影した動画を対象として、人物のどこに注目していたかを分かりやすく提示可能なヒートマップを提案する。以降これを人型視線ヒートマップと呼ぶ。従来のヒートマップが、暗に頭部姿勢の固定と、静止した対象を仮定していたのに対して、提案手法では、人物の動きのみならず、視聴者の頭部姿勢の変化を許容し、注視点のヒートマップを作成する。

本研究での主な想定シーンは、スポーツにおいて、練習風景を撮影し、その後のミーティングにおいて動画をモニターなどに映して確認・議論を行う場面である(図2参照)。このようなシーンを想定すると、人型視線ヒートマップは、映像視聴後すぐに提示可能なものであることが望ましい。

本稿は第1に人型視線ヒートマップを提案し、加えて、人型視線ヒートマップ生成のリアルタイム化を目

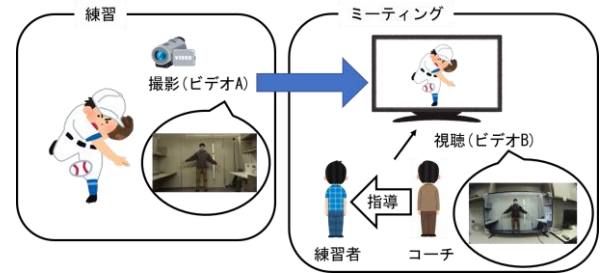


図2 想定する利用シーン

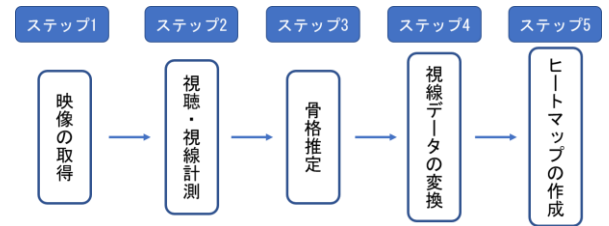


図3 人型視線ヒートマップの処理の流れ

指した拡張の提案を行うものである。その構成は以下の通りである。2において、人型視線ヒートマップを提案する。3において、そのリアルタイム化および高精度化を提案する。4で実験を述べた後に、5でまとめる。

## 2. 人型視線ヒートマップの提案

メガネ式アイトラッカーを用いて頭部姿勢変化を伴いながら、経時的に姿勢が変化する動作を視聴した場合、頭部姿勢の変化に伴う視界の変化と、視聴対象の変化の双方から、アイトラッカーを通じて得られる視線位置座標に対して直接の意味づけをすることが難しい。そのためその頻度マップであるヒートマップも直感的な解釈が難しくなる。

著者らはこの問題を解決するために、視聴対象が人物であることに注目し、「人のどの部分を見ていたか?」という情報に基づいた表示を実現することで、直感的な理解が容易なヒートマップの合成が可能になると考えている。

### 2.2. 処理の流れ

人型視線ヒートマップ作成の処理全体の流れを図3に示す。ここでは図2に示した想定シーンと対比させながら説明する。

- [1] 提案手法はまず練習などにおいて動作を**撮影**する。
- [2] ミーティングの場面で、撮影された映像を**視聴**する。その際、メガネ式アイトラッカーを装着し**視線計測**を行う。なお、視線計測データには、アイ

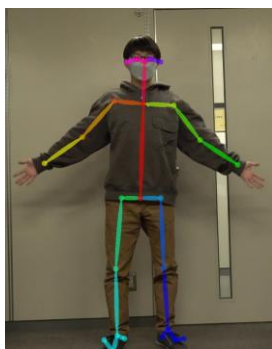


図 4 OpenPose の例

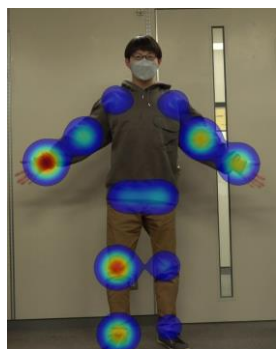


図 5 人型視線ヒートマップの例

トラッカー装着者の視界を撮影するワールドカメラの映像と、視線位置が含まれる。

- [3] ワールドカメラ映像に対して、**骨格推定**を適用する。なお、本研究において骨格推定には OpenPose[5][6]を用いた。図 4 に OpenPose による骨格推定結果の例を示す。
- [4] 視線位置と骨格推定結果に基づいて**注視対象**を推定する。
- [5] 最後に、**人の形状に合わせたヒートマップ**を作成する。人型視線ヒートマップの例を図 5 に示す。このように、人物のどの場所に注視があてられていたかを直感的に把握可能な表現となっている。

## 2.1. 視線位置と骨格推定結果に基づく注視対象の推定 (ステップ 4)

本研究では、スポーツにおけるフォームの確認を対象としており、視点は人体のどこかと想定している。しかし、骨格推定で求まる関節は、体全体に対してわずか 25 点である。そのため、視点位置と関節位置とはズレを含む。

本手法では、注視対象を便宜上、いずれかの関節とすることとする。しかし、前述の通り視点と関節とはズレがある。そこで、以下の処理により、注視を、複数の関節に割り振ることとする。

- [1] 視線位置からの各関節位置の距離を求める。
- [2] 視線位置からの距離が近い順に 3 つの関節を求める。
- [3] 視線位置からの 3 つの関節の距離に応じて重みづけをし、それぞれの関節に視線を割り当てる。

距離に応じた重み  $W$  は、式(1)により計算する。

$$W = \frac{\frac{1}{j_d}}{\text{SUM}_d}, \quad (1)$$

ここで、 $j_d$  は視線位置から関節  $d$  までの距離、 $\text{SUM}_d$  は

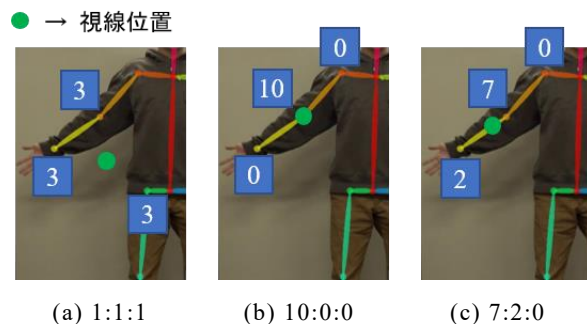


図 6 視線データの変換例

視線位置から 3 つの関節までの距離の逆数の総和とする。

すなわち、視線位置から近い距離ほど大きい重みづけをし、遠い距離ほど小さい重みとなる。重みは、それぞれの関節に 0 から 1 の間で割り振られ、視線位置から近いほど 1 に近い重みになり、遠いほど 0 に近い重みになる。また 3 つの重みの合計は 1 である。

これらの処理の概念を、図 6 を用いて説明する。図 6(a)のように視線位置から 3 つの関節の距離が同じ場合、このタイミングの注視は、右手首、右ひじ、右腰の 3 つの関節に均等に配分される。一方、図 6(b)のように右ひじと一致する場合には、注視を右ひじに集中させる。また、図 6(c)のように、3 つの関節が直線上に存在し、視点位置が 2 つの関節の内分点の場合には、最も近い 2 つの関節に配分する。

上記の処理をすべてのフレームで行うことで、関節に基づいた注視情報を取得することが可能となる。

## 3. 人型視線ヒートマップ作成の高速化および高精度化

前節で示した手順により人型視線ヒートマップを作成することが可能である。しかし実利用を考えた場合、いくつかの課題が存在する。ここでは、まず課題を明らかにした上で、改良に向けた新しい提案を行う。

### 3.1. 課題の確認

以下 2 つの課題を有すると考えている。第 1 に関節位置推定の精度が悪いことである。アイトラッカーのワールドカメラの映像に対して関節位置推定を行う。そのため、ディスプレイの鏡面反射による映り込みや、視聴角度による歪み、またワールドカメラのレンズ歪みなど様々な要因で関節位置推定が困難なケースが発生し得る。図 7 にその例を示す。図 7 左の例は映り込みが大きく、またレンズ歪みの影響も受けている。右の例では画面左側からの視聴のために歪みが発生していることが分かる。その結果どちらの例においても正確な関節位置推定ができていない。



図 7 OpenPose の精度が悪い例

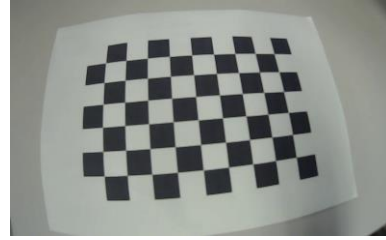


図 9 撮影したチェスボード画像

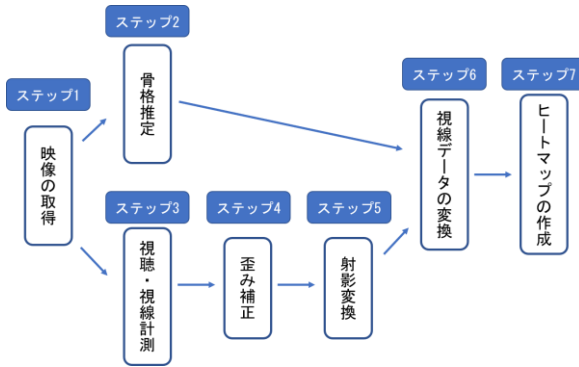


図 8 高精度化・リアルタイム化の処理の流れ

第 2 に処理速度が遅いことである．2 で提案した処理の流れでは，ミーティングなどで映像を視聴したあとの映像に対して関節推定の処理を行う．処理手順の中で関節推定は比較的負荷の高い処理であり，これがボトルネックとなり人型視線ヒートマップを提示するまでの時間が長くなってしまう．

### 3.2. 新しい処理の流れ

以上の 2 点の問題を同時に解決する方法として，新しい処理の流れを図 8 に示す．

- [1] 提案手法はまず練習などにおいて動作を撮影する．
- [2] 映像に対して**骨格推定**を行う．なおこの処理は，ミーティングの前にあらかじめ行っておくことを想定する．
- [3] ミーティングの場面で，撮影された映像を**視聴**する．その際，メガネ式アイトラッカーを装着し**視線計測**を行う．
- [4] ワールドカメラ画像に対してレンズ歪みの補正を行う．視線計測位置も同時に補正する．
- [5] ディスプレイの四隅を検出し，ステップ 2 の骨格推定結果を，ステップ 4 実施後の座標へと変換する．
- [6] 視線位置と骨格推定結果に基づいて**注視対象**を推定する．



図 10 歪み補正を行った結果  
(左：補正前 右：補正後)

- [7] 最後に，人の形状に合わせたヒートマップを作成する．人型視線ヒートマップの例を図 5 に示す．このように，人物のどの場所注視が当てられているかを直感的に把握可能な表現となっている．

以上の新しい処理では，骨格推定の対象が練習映像そのものであり，ディスプレイの映り込みや歪みの影響がなくなる．このことから骨格推定の精度向上が期待できる．さらに重要なことは，アイトラッカーによる視聴後に，ワールドカメラ映像に対して関節推定する必要がないという点である．視線計測後の処理は，あらかじめ実施してある骨格推定結果の変換のみとなるため大きな高速化が可能となる．以降，実装に関する詳細を述べる．

#### 3.2.1. 歪み補正

レンズ歪みの補正には Zhang らの手法 [7]を用いた．チェスボード画像（図 9）を様々な距離や角度で撮影し，その画像から直線の交点などの特徴点を検出し，カメラの内部パラメータを計算する．今回はチェスボード画像を 20 枚撮影し，パラメータ算出を行った．これらのパラメータを用いて歪み補正を行った結果を図 10 に示す．図 10 から歪みが補正されていることがわかる．

#### 3.2.2. 射影変換による各関節座標の変換

本提案では，アイトラッカーのワールドカメラの映像ではなく，元の映像に対して関節位置推定の処理を行う．しかし，その後の処理においては，同一画像内における，視線の位置と関節位置とが必要である（図 11 参照）．

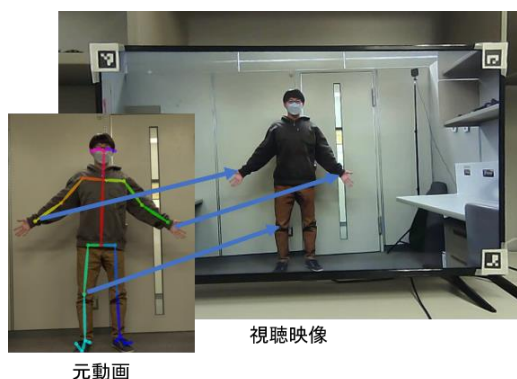


図 11 各関節座標の変換例



図 12 ArUco マーカー  
(左から id=0,1,2,3)

そのために、今回は、ディスプレイの四隅を検出し、それらを用いた射影変換を行った。ディスプレイでは元動画が全画面で再生されている。したがって、元動画の四隅の座標を、ディスプレイの四隅の座標に合わせて射影変換する。なお、ディスプレイの四隅の検出は簡単のために、ArUco マーカー[8]を利用する。図 12 に ArUco マーカーの例を示す。このマーカーをディスプレイの四隅に貼ることで簡易かつ比較的頑健に四隅を検出することが可能である。四隅を検出し、射影変換を行った結果を図 13 に示す。図 13 からマーカーの上に元動画が合成されていることがわかる。

## 4. 実験

### 4.1. 実験概要

提案手法の有用性を評価するために評価実験を行った。利用したアイトラッカーは Pupil Labs 社の Pupil Core[9]であり、図 14 に示す通り、装着者にとって負荷の低いものとなっている。

実験では、以下の 3 つの観点の評価を行った。第 1 は、注視点を反映した人型視線ヒートマップが合成されることの確認である。第 2, 3 は、3 記載の追加提案部分の評価である。まず、元映像と、歪み補正済みワールドカメラ映像との間の射影変換が正しく実施されていることの検証である。これに失敗していると人型視線ヒートマップも不適切なものとなるため検証を行った。第 3 に提案手法による、ヒートマップを提示までの時間の変化である。



図 13 射影変換した結果



図 14 使用したアイトラッカー

### 4.2. 人型視線ヒートマップの生成

図 15 に一般的なヒートマップと、人型視線ヒートマップの例を示す。通常のヒートマップでは、右手を中心に視聴していたように感じる。一方、提案した人型視線ヒートマップでは、右手だけでなく、右ひざや右足首にも多くの注視があることが分かりやすく提示されている。

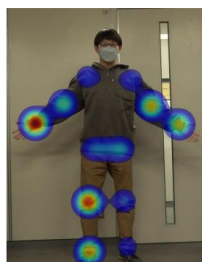
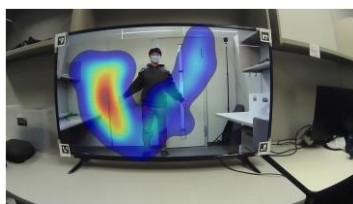
なお、この視聴は、視聴者に指示した視聴状況と比較して適切なものであり、定性的なものではあるが、人型視線ヒートマップを適切に生成することができたと考える。

### 4.3. 各関節座標位置の比較

元映像からワールドカメラ映像への変換精度の検証として以下を行った。なお、本実験においては、映り込みのない環境および関節位置推定の失敗の起きにくい動作を選定している。

まず、ワールドカメラ映像に対して関節位置推定を行った。次に、ワールドカメラ映像の四隅を検出し、元映像における関節位置を、ワールドカメラ映像における位置へと変換する。最後に、各関節点について、誤差を求める。

その結果を図 16 に示す。図 16 において各円の中心は、検出された関節の位置を表し、円の大きさは平均誤差を表している。この結果から分かる通り、この変換による誤差は極めて小さく、人型視線ヒートマップの精度に影響を及ぼすものではないことが確認できる。



(a) (b)  
図 15 通常のヒートマップ(a) と  
人型視線ヒートマップ(b)

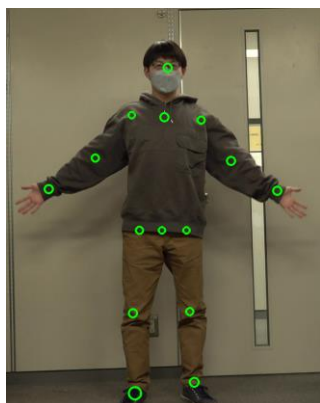


図 16 各関節の誤差円

表 1 処理時間の比較

|      | 比較手法     | 提案手法     |
|------|----------|----------|
| 所要時間 | 175.9[s] | 117.6[s] |

#### 4.4. 処理時間の比較

提案手法でのヒートマップを提示するまでの時間が早い確認するために、視線の計測が終わってからヒートマップを提示するまでの処理時間を計測した。ここで比較対象は、2 記載の提案手法である。

表 1 に処理時間を計測した結果を示す。表 1 からわかるように、提案手法では視線の計測が終わってから約 2 分でヒートマップを提示できている。

### 3. おわりに

本研究では、骨格推定を用いた人型視線ヒートマップを作成し、その結果をリアルタイムに提示する手法を提案した。評価実験から射影変換することで各関節座標が適切に変換されていることを確認した。また、正しく視聴部位に基づいたヒートマップを作成し、リアルタイムに結果を提示できたことを確認した。

これまでの処理はすべて Python で行っていた。その

ため、歪み補正など動画のフレーム数が多いほど処理時間が長くなっていた。そこで今後として、処理速度が速いと知られている C 言語で実装し、さらに処理時間を短くし、よりリアルタイム化していきたい。また、実際にスポーツに運用していき、上手な人がどこを注視しているのか、選手とコーチが同じところを見てコミュニケーション取れているか、無意識に気にしているところの調査など、検証していきたい。

### 文 献

- [1] 山品ら, "アイトラッキングによるマンモグラフィ撮影時の注視点検出: 熟練者と初学者の比較", 日本放射線技術学会雑誌, 2019 年 75 巻 11 号 pp.1316-1324
- [2] M.Pohl et al., "Comparing the Readability of Graph Layouts using Eyetracking and Task-oriented Analysis", The Eurographics Association, 2009.
- [3] H. Zamani et al., "EyeTracking Application on Emotion Analysis for Marketing Strategy", Artificial intelligence and robotics, vol. 8, No.11, pp.87-91, 2016.
- [4] Nakano et al., "Evaluation of 3D Markerless Motion Capture Accuracy Using OpenPose With Multiple Video Cameras", Brief research report article, May 2020.
- [5] Z. Cao et al., "Realtime multiperson 2d pose estimation using part affinity fields", IEEE, pp.7291-7299, May 2017.
- [6] Z. Cao et al., "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 43, no.1, pp.172-186, July 2019
- [7] Zhengyou Zhang, "A Flexible New Technique for Camera Calibration", IEEE Trans. Pattern Anal. Mach. Intell., vol.22, pp.1330-1334, 2000.
- [8] S. Garrido-Jurado et al., "Automatic generation and detection of highly reliable fiducial markers under occlusion", Pattern Recognition, Vol. 47, Issue. 6, pp.2280-2292, June 2014.
- [9] Pupil Labs, "Pupil core"  
<https://pupil-labs.com/products/core/>, (参照 Feb.9, 2023)