映像中の動作の躍動感を評価する観察者から計測された視線分布の 三次元ヒートマップ可視化

岩崎芙由子*日置翔太*米田駿介*井上路子*西山正志*

3D heatmap visualization of gaze distribution measured from observers predicting dynamism of subject's movement

Fuyuko IWASAKI, Shouta HIOKI, Shunsuke YONEDA, Michiko INOUE, and Masashi NISHIYAMA

We propose a method to visualize the gaze distribution measured from an observer who perceives dynamism from the subject's movement in a movie using a three-dimensional (3D) heat map in a manner that is independent of time series changes in the subject's posture. In the previous method, the gaze distribution measured from an observer is represented using a heat map on a 3D human body model estimated from the subject in a still image. However, the previous method cannot handle time-series changes in the subject's posture in a movie. Furthermore, the previous method cannot visualize the gaze distribution measured from the surface region of the subject's body. Our method introduces an angle between the gaze distribution weetor and the vertex position vector to visualize the gaze distribution measured in the surface regions. From the experimental results, we confirmed that our method can visualize not only the gaze distribution measured in the surface region but also that measured in the surrounding region compared to the previous method. We also confirmed that it is possible to visualize the gaze distribution of a subject's movements in a movie without depending on time series changes using a standard posture human body model.

Key words: Gaze distribution, 3D heat map, Body model, Visualization, Vertex attention probability

1. はじめに

近年, HMD (Head Mounted Display) 技術の発達により, 仮 想空間でのスポーツ観戦が期待されている.本論文では、選手 個人が行うスポーツの動画を、観客が HMD を用いて観戦する 場面について考えていく.その場面の例を図1に示す.仮想 空間内に配置された大型ディスプレイに表示されるスポーツ動 画を, 観客は目視した上で躍動感を受けている. その時に観客 は、動画中でスポーツを行う選手の動作に注目していると考 えられる.本論文では、観客が選手の動作から躍動感を受ける 時,その手掛かりとなる身体部位に,観察者の視線が集まると 考えていく.様々な動画の間で、どの身体部位に注目していた かを比較するため, 観察者から視線分布を計測し, 分析者が調 査する状況について考えていく. 分析者は、例えばスポーツ選 手へ躍動感を高めるフィードバックを与えることを目的とし, 選手の身体のどこへ観客の視線が集まるかを調査する.以下で は、動画中に存在する選手を被写体と呼び、動画中の被写体の 動作を観戦する観客を観察者と呼ぶ.

分析者が視線分布を調査する状況において,観察者の視線が 被写体のどの身体部位に集まるかを,分かりやすく可視化する 手法が求められている.観察者から計測された視線分布を空間 的に可視化するため,一般的には二次元ヒートマップを静止画 へ重ね合わせる手法が多く使用されている.例えば,既存の分 析研究^{1,2,3)}において,静止画中の被写体を刺激とした場合に おける視線分布の調査で使用されている.ただし,動画中の被



Fig. 1 When an observer looks at a subject in a movie, the subject's movement gives the observer dynamism. We consider investigating where the observer focuses on the subject's body.

写体を刺激とした場合,スポーツの動作を行う被写体の姿勢 は、時々刻々と変化することに注意を払う必要がある.具体的 には、各時刻の二次元ヒートマップを用いて視線分布を調査す る際、被写体の姿勢の違いを、動画の時刻間で考慮しなければ ならない.

被写体の姿勢が時間方向に変化したとしても、分析者はその 姿勢の違いを意識せずに、視線分布を容易に調査できる可視化 手法が必要である.ここで、我々が開発してきた従来手法⁴⁾を 活用することを考える.この可視化手法では、被写体間で姿勢 が共通化された三次元人体形状を表す標準人体モデルを、静止 画中の被写体から推定し、その人体モデルの表面上で視線分布 を三次元ヒートマップを用いて可視化している.これにより、 静止画において被写体間の姿勢の違いを考慮せずとも視線分布 を調査できる利点をもつ.ただし、従来手法は、静止画中の被 写体のみを対象としているため、時々刻々と姿勢が変化する動

^{*} 原稿受付 令和*年*月*日

^{*} 掲載決定 令和*年*月*日

^{**} 鳥取大学大学院工学研究科 (鳥取市湖山町南 4 丁目 101)



Fig. 2 Examples of surface and surrounding regions of the subject's body.

Surface region

on the subject's body

Surrounding region

A set of pixels

A set of pixels in the vicinity of the body

画中の被写体については考慮していなかった. さらに従来手法 には大きな課題が存在する. 被写体の表面領域で計測される視 線のみを距離に基づき可視化していたため,その周囲領域で計 測された視線を可視化できなかった. ここで述べる表面領域と は,動画から推定された被写体に対応する画素の集合を指し, 周囲領域とは,被写体の近傍に存在する画素の集合を指す. こ れらの領域の例を図2に示す.

本論文では、観察者が動画中の被写体の動作から躍動感を受 ける時、被写体の表面領域とその周囲領域において計測された 視線分布を、被写体の姿勢が時間方向に統一された三次元ヒー トマップを用いて可視化する手法について述べる.提案手法 は,表面領域とその周囲領域との両方に対応するため,仮想空 間における観察者の視線方向ベクトルと人体モデルの頂点位置 ベクトルとがなす角度を算出する. これらのベクトルの間の角 度を用いることで、各時刻の被写体の姿勢を表す人体モデルの 各頂点において, 観察者の視線が集まる割合を表す頂点注目確 率を求める. 姿勢が時刻間で共通化された標準人体モデルを導 入し,頂点注目確率を全時刻で周辺化することで,三次元ヒー トマップとして視線分布を可視化する.評価実験において、動 画中の被写体の動作から受ける躍動感の度合いを評価する設問 を観察者に与え、その評価の最中に観察者から視線分布を計測 し可視化した.実験結果において,従来手法では,表面領域で 計測された視線分布のみしか可視化できなかったが、提案手法 では、表面領域と周囲領域との両方で計測された視線分布を三 次元ヒートマップで可視化できることを確認した. 更に, 従来 手法では、動画中で時系列に変化する被写体の姿勢へ対応でき ていなかったが、提案手法では、動画の全時刻を通して計測さ れた視線分布を可視化できることを確認した.以下, 2. で表面 領域と周囲領域の扱いについて述べ、3. で姿勢の時系列変化へ の対応について述べ、4. で評価実験の条件について述べ、5. で 評価実験の結果について述べ、6. でまとめる.

2. 表面領域と周囲領域とで計測される視線分布の扱い

2.1 従来手法の概要

従来手法⁴⁾には, 1. で述べたように, 動画に対応できない課 題と, 人体の周囲領域で計測された視線分布に対応できない課 題の二つが存在する.本論文では, 各時刻において発生する後 者の課題をまず解決し, その上で前者の時間方向の課題につい て対応していくアプローチをとる.以下では, 従来手法の後者 の課題である各時刻における周囲領域の取り扱いについて議論 を進めていく.

従来手法の課題を明確にするため、従来手法の流れについ て概要を述べる.従来手法は大きく分けて次の 2 つのステッ プで構成される.まずステップ1 では、時刻 t において、観 察者 o が静止画を見る際の視線方向ベクトル g(t,o) をアイト ラッカーを用いて計測する.g(t,o) を単位ベクトルとし、そ



Fig. 3 Comparison of the calculation of the attention degree in the surface and surrounding regions between the previous method and our method.

の始点を目の位置とし仮想空間の原点とする.次にステップ2 では、定数 α (> 0) 倍された視線方向ベクトル $\alpha g(t, o)$ を用い て、静止画から推定された人体モデルの各頂点 v(t) へ、視線 がどれだけ集まるかを表す注目度合を求める.なお、人体モデ ルとして SMPL⁵⁾ を用いており、このモデルの内部は姿勢パラ メータと体型パラメータから表現されている.出力される人体 モデルは、人体表面における三次元頂点位置ベクトル v(t) の 集合 V(t) と、頂点同士の隣接関係とから構成されている.定 数 α は、 $\alpha g(t, o)$ が人体モデル表面に衝突することで定まる. 従来手法では、 $\alpha g(t, o)$ から v(t) までの距離 ϵ に基づき、人体 モデルの頂点へ視線が集まる注目度合 d(v(t), g(t, o)) を式 (1) で計算している.

$$d(\boldsymbol{v}(t), \boldsymbol{g}(t, o)) = \exp(-\frac{\epsilon^2}{2\sigma^2})$$
(1)

ここで,距離は $\epsilon = \|\boldsymbol{v}(t) - \alpha \boldsymbol{g}(t, o)\|_2$ で計算される.距離 ϵ が小さいほど $\boldsymbol{v}(t)$ への注目度合は大きくなり,距離 ϵ が大き いほど注目度合は小さくなる.従来手法が距離に基づいている 理由としては,一般的な二次元ヒートマップで視線分布を可視 化する場合,画像平面上で視線が計測された位置からの距離に 基づいていることが挙げられる.

2.2 周囲領域における従来手法の課題

従来手法の課題の一つである周囲領域で計測された視線分布 を取り扱えない原因について述べる.従来手法において,定数 倍された視線方向ベクトル $\alpha g(t, o)$ は,静止画から推定された 人体モデルの表面に必ず衝突するという強い制約条件の下で注 目度合を計算している.この制約条件を満たす例を図 3(a) に 示す.人体モデルに衝突する場合,視線方向ベクトル $\alpha g(t, o)$ と人体モデルの頂点位置ベクトル v(t) との間で距離 ϵ が求め られるため,注目度合 d(g(t, o), v(t))は計算可能である.一 方,制約条件を満たさない場合の例を図 3(b) に示す.人体モ デルに衝突しない場合, $\alpha g(t, o)$ とv(t) との間で距離 ϵ が求 められないため,注目度合 d(g(t, o), v(t))は計算不可能であ る.周囲領域で視線が計測される場合,従来手法の三次元ヒー トマップは視線分布を十分に可視化できているとは言えない.

ここで、表面領域で計測される視線分布だけでなく、周囲領

域で計測される視線分布も考慮する必要がある状況について説 明する.観察者の視線方向ベクトル $\alpha g(t, o)$ が周囲領域に向 く状況として,次の二つが考えられる.一つ目の状況は,観察 者が,被写体の輪郭の近傍に存在する背景などの領域を見てい るときである.二つ目の状況は,観察者が,既存の人体モデル では表現できていない被写体の髪型や衣服などの領域を見て いるときである.観察者は,上記で述べた二つの状況が発生し た時,それらの周囲領域だけなく,その近くに存在する表面領 域の人体も同時に見ていると考えられる.具体的には,視線の g(t, o) が指し示す方向に存在する中心点のみを見ている訳で はなく,その点の近傍に存在する人体モデルの頂点v(t)も同 時に見ていると考えられる.このため,視線分布を三次元ヒー トマップを用いて可視化する際,表面領域だけでなく周囲領域 も同時に取り扱う必要があると言える.

2.3 提案手法における表面領域と周囲領域の扱い方

従来手法で扱うことができなかった周囲領域を,提案手法 で表面領域と同時に扱うための考え方について述べる.従 来手法では、2.1 で述べたように、距離 ϵ に基づき注目度合 d(v(t), g(t, o))を求めていた.一方,提案手法では,視線方向 ベクトル g(t,o) と頂点位置ベクトル v(t) とがなす角度 θ を用 いて、視線がどれだけ集まるかを表す注目度合a(v(t), g(t, o))を求めることを考えていく. 角度を用いる利点を例を用いて以 下で説明する.まず、表面領域で視線分布が計測される場合の 例を図 3(c) に示す. 視線方向ベクトル g(t,o) と人体モデルの 頂点位置ベクトルv(t)との間で角度 θ が求められるため,注 目度合 a(g(t, o), v(t)) は計算可能である.また、周囲領域で 視線分布が計測される場合の例を図3(d)に示す.表面領域の 場合と同様に、g(t,o)とv(t)との間の角度 θ が求められるた め、注目度合 a(g(t,o), v(t)) は計算可能である.以上より、角 度が導入された提案手法では、視線分布を三次元ヒートマップ を用いて可視化する際,表面領域だけでなく周囲領域も同時に 取り扱うことができる.

2.4 動画の各時刻における注目度合の算出方法

提案手法において,時刻 t の注目度合を算出する方法につい て詳細を述べる.人体モデルを構成する v(t) に,観察者 o の 視線が集まることを表す注目度合 a(v(t), g(t, o)) を,視線方 向ベクトル g(t, o) と頂点位置ベクトル v(t) とがなす角度 θ を 用いて式 (2) で計算する.

$$a(\boldsymbol{v}(t), \boldsymbol{g}(t, o)) = \exp(-\frac{\theta^2}{2\sigma^2})$$
(2)

ここで,角度θは内積を用いて式(3)で表される.

$$\cos \theta = \frac{(\boldsymbol{v}(t), \boldsymbol{g}(t, o))}{\|\boldsymbol{v}(t)\|_2 \|\boldsymbol{g}(t, o)\|_2}$$
(3)

なお,角度 θ が小さいほどv(t)への注目度合は大きくなり, 角度 θ が大きいほど注目度合は小さくなる.人体モデルにおい て,観察者の目の位置から直接見ることができない裏面に存在 するv(t)については,注目度合を0とする.

次に、動画の時刻 t におけるフレーム S(t) を見た時、人 体モデルを構成する頂点 v(t) に視線が集まる割合を表す 頂点注目確率 p(v(t)|S(t)) を、上記で算出された注目度合 a(v(t), g(t, o)) を用いて求める。具体的には、各時刻の頂点注 目確率 p(v(t)|S(t)) を式 (4) で求める。

$$p(\boldsymbol{v}(t)|\mathcal{S}(t)) = \frac{1}{W} \sum_{o \in \mathcal{O}} a(\boldsymbol{v}(t), \boldsymbol{g}(t, o))$$
(4)

ここで, O 名からなる観察者の集合を O とする. なお, 係 数 W は式 (5) で表される.

$$W = \sum_{\boldsymbol{v}(t)\in\mathcal{V}(t)}\sum_{o\in\mathcal{O}}a(\boldsymbol{v}(t),\boldsymbol{g}(t,o))$$
(5)

また, $p(\boldsymbol{v}(t)|\mathcal{S}(t))$ は式 (6)を満たすものとする.

$$\sum_{\boldsymbol{v}(t)\in\mathcal{V}(t)} p(\boldsymbol{v}(t)|\mathcal{S}(t)) = 1$$
(6)

以上より,提案手法は,各時刻で求めた注目度合を用いて頂点 注目確率を算出することで,人体の表面領域で計測された視線 に加え,その周囲領域で計測された視線にも対応できるように なることが期待される.

3. 動画中の被写体の姿勢の時系列変化への対応

3.1 概要

従来手法⁴⁾では動画に対応できなかった残りの課題を解決 する方法について議論を進めていく. 被写体の姿勢は、1. で述 べたように、動画中で時間方向に連続して変化している.動画 の全時刻を通して、観察者は被写体の身体のどこを見ていたか を調査することを考える. 2.4 で述べた式 (4) の時刻 t の頂点 注目確率 p(v(t)|S(t)) を三次元ヒートマップで可視化し、各 時刻の可視化結果をそのまま時間方向につなぎ合わせて三次元 ヒートマップの動画を生成することは簡単にできる.ただし, 時系列に姿勢が変化する三次元ヒートマップ動画では、各時刻 でどこに視線が集まっているかを、視線の分析者が記憶しなが ら調査する手間が発生する.本論文では、各時刻の姿勢が共通 化された標準人体モデルを用いることで、時刻間で異なる姿勢 を統一し、頂点注目確率を全時刻で周辺化することを考えてい く. これにより,視線の分析者は,標準人体モデル上で可視化 された単一の三次元ヒートマップを見るだけで、動画の全時刻 を通した視線分布を調査できると考える.以下では、その具体 的な計算方法について述べる.

3.2 頂点注目確率を全時刻で周辺化

提案手法は、頂点注目確率を全時刻で周辺化する際、各時刻 で推定された人体モデルを標準人体モデルに変換する.以下 で、その具体的な方法について述べる.時刻 t のフレームS(t)から推定された人体モデルでは、2.1 で述べたように、時刻に応 じて姿勢が時間方向に変化するため、そのモデルは時刻 t の頂 点v(t)の集合V(t)で構成されている.一方、標準人体モデル では、時刻に関係なく姿勢が一定であるとし、そのモデルは頂 点 \hat{v} の集合 \hat{V} で構成されているとする.なお、標準人体モデ ルの \hat{v} と各時刻の人体モデルのv(t)との間で、三次元位置ベ クトルは異なるものの人体表面上での隣接関係は等しい.よっ て、各時刻の人体モデルのv(t)での頂点注目確率p(v(t)|S(t))は、標準人体モデルの \hat{v} での頂点注目確率 $p(\hat{v}|S(t))$ に式(7) で変換できる.

$$p(\tilde{\boldsymbol{v}}|\mathcal{S}(t)) = p(\boldsymbol{v}(t)|\mathcal{S}(t))$$
(7)

なお、様々な人体モデル^{5.6.7)} が提案されているが、提案手 法では従来手法と同様に SMPL⁵⁾ を用いる.標準人体モデ ルとして、時刻間で共通化された特定の姿勢を設定し、時 刻間で SMPL の姿勢パラメータを一定値にする.具体的に は、直立姿勢を保ちつつ両腕を横に水平に伸ばした姿勢に 統一する.以下では動画の各フレームに対する時刻の集合を $\mathcal{T} = \{t_1, \dots, t_T\}$ と表記し、1 個の動画におけるフレームの集



Fig. 4 Overview of our method.

合を $S = \{S(t_1), \dots, S(t_T)\}$ と表記する.なお、Tの要素数 はSの要素数と等しいとする.

各時刻の視線分布を全時刻での視線分布に統合するため, 式 (7) の頂点注目確率 $p(\hat{v}|S(t))$ を,動画を構成するフレーム の集合 S で周辺化する.具体的には,全時刻での頂点注目確 率 $p(\hat{v}|S)$ を式 (8) で計算する.

$$p(\tilde{\boldsymbol{v}}|\mathcal{S}) = \frac{1}{n(\mathcal{S})} \sum_{\mathcal{S}(t) \in \mathcal{S}} p(\tilde{\boldsymbol{v}}|\mathcal{S}(t))$$
(8)

ここで、n()は集合の要素数を返す関数であり、具体的には動 画の総フレーム数を返す.なお、 $p(\hat{v}|S)$ は式(9)を満たす.

$$\sum_{\tilde{\boldsymbol{v}}\in\tilde{\mathcal{V}}} p(\tilde{\boldsymbol{v}}|\mathcal{S}) = 1$$
(9)

観察者は動画の全時刻を通して被写体の身体のどこを見ていたかを、視線の分析者が調査できるよう提案手法は可視化する.具体的には、全時刻で周辺化された頂点注目確率 $p(\hat{v}|S)$ を三次元ヒートマップで表し、標準人体モデルに重ね合わせる.視線が集まる確率が高い頂点ほど赤色に近くなり、視線が集まる確率が低い頂点ほど青色に近くなるよう標準人体モデル表面に色付ける.

3.3 提案手法の全体の流れ

提案手法の流れを図4に示し、各ステップについて説明する.

S1. 各時刻の視線計測

動画の時刻 t のフレーム S(t) を,観察者 o が見ている時 の視線方向ベクトル g(t,o) について,HMD に搭載され たアイトラッカーで計測する.

S2. 各時刻の人体モデル推定

フレーム S(t) の被写体から人体モデルを推定し, 各時刻

の頂点位置ベクトル v(t) を取得する.人体モデルは各時 刻の姿勢を表しており、これまでに述べたように SMPL⁵⁾ で定義されている.動画の各時刻において SMPL の姿勢 パラメータを推定するため、多くの手法^{8,9,10,11,12)} が提案 されているが、提案手法では 4D Humans⁸⁾ を用いる.

S3. 各時刻の頂点注目確率の算出

S1 で得た視線方向ベクトル g(t, o) と S2 で得た頂点位置 ベクトル v(t) とがなす角度 θ を用いて,各時刻の人体モ デルを構成する頂点へ,観察者の視線がどの程度集まる かを表す注目度合 a(v(t), g(t, o)) を求める.具体的には, 2.4 で述べた式 (2) を計算する.次に,観察者の集合 O に 属する全ての観察者から得られた注目度合を用いて,人 体モデルの頂点に全観察者の視線が集まる割合を表す頂 点注目確率 p(v(t)|S(t)) を各時刻で求める.具体的には, 2.4 で述べた式 (4) を計算する.

S4. 各時刻の視線分布の可視化

各時刻において, 観察者の視線は被写体の身体のどこに集 まるかを調査するため, S3 で得た各時刻の頂点注目確率 を表す三次元ヒートマップを, 各時刻の人体モデルへ重ね 合わせて可視化する.

S5. 頂点注目確率を全時刻で周辺化

動画の全時刻において,観察者の視線は被写体の身体の どこに集まっていたかを調査するため,S3で得た各時刻 の頂点注目確率を周辺化する.具体的には,3.2で述べた 式(8)を計算する.

S6. 全時刻での視線分布の可視化

動画の全時刻を通して観察者の視線は被写体の身体のど こに集まるかを表す視線分布を可視化するため, S5 で得 た全時刻の頂点注目確率を表す三次元ヒートマップを,標



Fig. 5 Examples of stimulus movies for gaze measurement.

準人体モデルへ重ね合わせる.

提案手法を用いることで,動画中で被写体の姿勢が時系列変化 する場合,その時刻間の姿勢の違いを考慮せずとも,動画の全 時刻を通して,観察者が被写体の身体のどこを見ていたかを, 視線の分析者は調査できるようになることが期待される.

4. 評価実験の条件

4.1 刺激として用いる動画

提案手法は、1. で述べた従来手法の二つの課題に対応できた かを確認するため、動画中の被写体から躍動感を受けている観 察者の視線分布を計測し可視化する実験を行った. 観察者 1名 につき刺激として3個の動画を準備した. 各動画から抽出され たフレームの例を図5に示す. 被写体1の動画では被写体が壁 を登る動作を行い, 被写体2の動画では被写体が球を打つ動作 を行い, 被写体3の動画では被写体が足を蹴り上げる動作を行 うことにした. 動画の時間長は, 被写体1で10.4 秒, 被写体2 で 6.9 秒, 被写体3 で 7.3 秒とした. これらの動画を素材サイ ト^{*1}よりダウンロードした. 動画のフレームレートは30 fps と した. 動画の縦幅を 900 画素となるようにアスペクト比を維 持してリサイズした. 動画の横幅は被写体1で 506 画素, 被写 体2で 570 画素, 被写体3 で 730 画素とした. リサイズした 動画を灰色画像の上のラインダムな位置に配置した. 灰色画像 の解像度は 1920 × 1080 画素とした.

4.2 視線計測のセッティング

観察者の視線を計測するため,アイトラッカーが内蔵されて いる HMD (VIVE Pro Eye, HTC 社)を使用した. 観察者は裸 眼,もしくは、コンタクトの状態で HMD を装着した. 仮想空 間内の大型ディスプレイ上に表示される動画を観察者へ見せ た. 仮想空間において,観察者の視点から大型ディスプレイ までの距離を 9 m とし,大型ディスプレイの大きさを 16 × 9 = 144 m² とした. アイトラッカーのサンプリングレートを 90 Hz とした. なお,HMD の仕様により,計測される視線の時刻 間隔は一定でないため、等間隔になるよう線形補間を用いてリ サンプリングした. この際,観察者が瞬きした時刻や,視線が 大型ディスプレイから外れた時刻については,視線を補間しな かった.

4.3 観察者が取り組む設問

視線計測の際に以下の設問を観察者に与えた.

Q : 動画中の被写体の動作に躍動感があると感じるか.

この設問について、4 段階評価 (4:感じる、3:やや感じる、

2: やや感じない, 1: 感じない) で観察者に回答させた. 評価実 験に参加する観察者の人数を 24 名 (男性 12 名,女性 12 名, 平均年齢 23.6 ± 2.0 歳) とした. 観察者の平均瞳孔間距離は $63.2 \pm 1.9 \text{ mm}$ であった.なお,設問Qに対する4段階評価の 平均スコアは,被写体1の動画で 2.0 ± 1.0 ,被写体2の動画 で 2.7 ± 0.8 ,被写体3の動画で 2.1 ± 0.7 であり,どの動画で も躍動感があるとはやや感じない傾向であった.

4.4 視線計測の手順

観察者の視線を計測する手順を以下で述べる.

- *P*₁:観察者の集合 *O*から1名の観察者 *o* をランダムに選出 した.
- P₂:計測手順,4.3 で述べた設問Q,および,回答方法を観察者 o に説明した.
- P3:開始画像を大型ディスプレイ上に2秒間表示した.
- P4:観察者へ与える刺激として,4.1 で述べた3個の動画 の中から重複無しでランダムに選択された1個の動画 を大型ディスプレイ上に表示した.この時に観察者の視 線を計測した.
- *P*₅:終了画像を大型ディスプレイ上に3秒間表示した.また,設問Qについて,観察者に回答させた.
- *P*₆:全ての動画の計測が完了するまで,手順 *P*₃ から *P*₅ までを繰り返した.
- *P*₇:観察者の集合 *O* の全ての観察者の計測が完了するまで,手順 *P*₁ から *P*₆ までを繰り返した.

動画を表示する際,観察者の視線が,初期時刻にディスプレイ 中心に固定されるセンターバイアス¹³⁾に気をつける必要があ る.センターバイアスを避けるため,手順 *P*₃ で準備した開始 画像を表示する時,無地背景のランダムな位置に配置された白 い十字印の中心を観察者に注視させた.

4.5 可視化のための三次元ヒートマップ生成

計測された観察者の視線分布を三次元ヒートマップで可視 化した時の条件を述べる.提案手法で用いる SMPL⁵⁾の人体 モデルを WEB サイト^{*2}からダウンロードした.この SMPLの 頂点数は 6890 点であり,三つの頂点を結んだ三角形メッシュ で隣接関係が設定されていた.各時刻の人体モデルを得る際 には,3.3 の S2 で述べた 4D Humans⁸⁾を用いて,動画の各フ レームから姿勢パラメータ推定した.提案手法の S5 で用いる 標準人体モデルにおいて,姿勢パラメータを SMPL で設定さ れているデフォルト値とした.なお,各頂点の位置ベクトルを 求める際に必要となる体型パラメータを, SMPL における中性 の平均体型を表すデフォルト値とした.視線分布を三次元ヒー トマップで可視化するため, blender 3.5 を用いてレンダリング した.

5. 評価実験の結果

5.1 各時刻において表面領域と周囲領域とから計測された 視線分布の可視化

視線を可視化する際,表面領域に加えて周囲領域も提案手法 は取り扱えるかどうかを評価した.この評価実験では,図6(a), (f),および,(k)のフレームS(t)を対象とした.各フレームに おいて,従来手法⁴⁾では,2.1で述べた式(1)の距離 ϵ に基づく

^{*1} https://www.pexels.com/ja-jp/

^{*2} https://smpl.is.tue.mpg.de



Fig. 6 Visualization of the gaze distribution measured from surface and surrounding regions in the frame S(t) of time t.

注目度合 $d(\mathbf{v}(t), \mathbf{g}(t, o))$ を利用した.また,提案手法では,2.4 で述べた式 (2) の角度 θ の基づく注目度合 $a(\mathbf{g}(t, o), \mathbf{v}(t))$ を使 用した.まず,距離に基づく従来手法を用いて可視化した結果 を図 **6**(b),(g),および,(l) に示す.比較のため,表面領域の みにおいて,角度に基づく提案手法を用いて可視化した結果を (c),(h),および,(m) に示す.(b) と(c) とを比べた時,可視化 された視線分布の見た目はほぼ同じであった.また,(g) と(h) とを比べた時,および,(l) と(m) とを比べた時も,可視化結果 はほぼ同じであった.表面領域から計測された視線分布の可視 化において,距離に基づく従来手法と角度に基づく提案手法と は同等の表現力があると考えられる.

次に、従来手法では取り扱えなかった周囲領域において、提 案手法を用いて可視化した結果を図 6(d), (i), および, (n) に 示す. 被写体1について,表面領域の(c)では視線が頭や右腕 に主に集まっていたが、周囲領域の (d) では視線が左肘に主に 集まっていた. 被写体 2 について,表面領域の (h) では視線が 右腕に主に集まっていたが、周囲領域の (i) では視線が頭に主 に集まっていた. 被写体 3 について,表面領域の (m) では視 線が左肩に主に集まっていたが、周囲領域の (n) では視線が頭 に主に集まっていた. これらの結果から、表面領域において計 測された視線分布のみでは見落とされていた傾向が,周囲領域 において計測された視線分布から発見できるようになると考え られる. なお, 図中において, 可視化結果の上部に示した数字 は、フレーム S(t) において計測された観察者 24 名全員分の視 線のサンプリング総数を100%とした時、表面領域、または、 周囲領域のそれぞれで計測された視線のサンプリング数の割合 を表す.

最後に,表面領域と周囲領域との両方から計測された視線分

布について,提案手法を用いて可視化した結果を図 6(e),(j), および,(o)に示す.これらの可視化結果では,表面領域で計 測された視線分布と周囲領域で計測された視線分布とが,前述 の割合で重み付け平均された見え方であることが分かる.以上 の結果より,提案手法は,各時刻において表面領域に加えて, 従来手法が取り扱えなかった周囲領域から計測された視線分布 を可視化できることを定性的に確認した.従来手法の表面領域 のみから得た調査結果に加えて,異なる身体部位にも視線が集 まる傾向も示していることから,観察者の視線が表面領域から 外れることが多い場合は,周囲領域も同時に取り扱える提案手 法は,視線分布の調査に重要であると考えられる.

5.2 動画の全時刻で周辺化された頂点注目確率による視線 分布の可視化

動画の全時刻で頂点注目確率を周辺化することで,提案手法 は動画に対応できるかどうかを評価した. 4.1 で述べた動画の 全フレームについて, 4.4 で述べた手順で計測された各時刻の 視線分布を用いて,全時刻を通した視線分布を可視化した.動 画から抽出された一部のフレーム *S*(*t*) を図 **7**(a), (e),および, (i) に示す.

まず,一般的な可視化手法である二次元ヒートマップを用い て,各時刻の視線分布を可視化した結果を図7(b),(f),および, (j)に示す.また,3.3で述べた図4のステップS4で生成され た三次元ヒートマップを用いて,各時刻の頂点注目確率を可 視化した結果を(c),(g),および,(k)に示す.各時刻の二次元 ヒートマップ,および,各時刻の三次元ヒートマップは共に, 観察者が被写体の身体のどこを各時刻で見ていたかを可視化し ている.ただし,動画の全時刻を通して観察者が被写体の身体 のどこを見ていたかを調査する際,それら各時刻のヒートマッ プでは,視線が集まる身体をそれぞれの時刻で細やかに記憶す る手間が発生すると言える.

次に, 3.3 で述べた図 4 のステップ S6 で生成された三次元 ヒートマップを用いて,頂点注目確率を全時刻で周辺化し,単 一の標準人体モデル上で可視化した結果を図 7(d),(h),およ び,(l)に示す.なお図中では,標準人体モデルの三次元ヒート マップを可視化する時,仮想カメラの視点位置を三つ設けてレ ンダリングした.これは,三次元ヒートマップを,複数の視点 から同時に確認するためである.提案手法を用いた三次元ヒー トマップの可視化により姿勢が全体で統一されているため,動 画の全時刻を通して観察者が被写体の身体のどこを見ていたか を,視線の分析者は一目で調査できると考えられる.

観察者が動画中の被写体の動作から躍動感を受ける時に計測 された視線分布に関する具体的な調査結果として,以下が得ら れた.壁を登る動作を行う被写体1の(d)では,動画の全時刻 を通して,頭,両腕,かつ,右脇に観察者の視線が主に集まっ ていた.球を打つ動作を行う被写体2の(h)では,動画の全時 刻を通して,頭に観察者の視線が主に集まっていた.また,足 を蹴り上げる動作を行う被写体3の(l)では,動画の全時刻を 通して,頭に観察者の視線が主に集まっていた.調査結果よ り,これらの動画において被写体の動作は異なるものの,観察 者の視線は,全時刻を通して頭に共通して主に集まる傾向を示 すことが分かった.以上より,提案手法を用いて頂点注目確率 を全時刻で周辺化し標準人体モデル上で可視化することで,動 画中で時系列に変化していく被写体の姿勢に対応できることを 定性的に確認した.



Fig. 7 Visualization of gaze distribution by marginalized vertex attention probability in all time points of the movie.

5.3 周囲領域を人体モデルで表現する動機と利点

提案手法を開発するに至った動機として,動画の各時刻で視線分布を可視化した結果が,一般的な二次元ヒートマップと距離に基づく従来手法⁴⁾との間で異なっていたことが挙げられる.その具体例を図8に示す.図中(a)のフレームについて,二次元ヒートマップで可視化した結果を(b)に示し,従来手法で可視化した結果を(c)に示す.被写体1では左肘,被写体2では頭と左肩,被写体3では頭において,二次元ヒートマップと従来手法との間で可視化結果が大きく異なっていた.この差異は,人体の周囲領域で観測された視線に従来手法が対応し

ていなかったことに起因すると考え,表面領域に加え周囲領域 に対応した提案手法を開発した.提案手法で可視化した結果を (d) に示す.この結果(d)は、二次元ヒートマップで可視化した 結果(b)と近づいていると言える.提案手法で可視化した結果 は、二次元ヒートマップと三次元人体モデルとを関連付けて調 査できる利点があり、さらにカメラ視点を仮想的に変えながら 視線分布を調査できる利点があると考える.

提案手法では,動画中のシーンを三次元的に意味のあるモデ ルに再構成することで,動画の時間方向に同一の意味をもつ 人体の頂点ごとに,視線分布を可視化することを狙った.ただ



Fig. 8 Comparison of visualization results.

し,提案手法はシーン中に存在する人体のみを再構成してお り,他の物体の三次元形状を再構成していない.このため,人 体の周りにある物体との関係を調査できない限界がある.被写 体2が所持しているラケットやボールなど,動作に大きく関係 する道具は視線分布に影響を及ぼしていると想定される.道具 を手で持っている場合は周囲領域でカバーできる可能性がある が,例えばボールを高く投げるなど,道具が人体から遠く離れ た場合,その道具に視線が集まったとしても提案手法では対応 できない.今後はシーン全体を三次元再構成できる手法を導入 し,視線分布の可視化手法を改良していく必要がある.

5.4 周囲領域における全時刻を通した視線分布の可視化

表面領域に加えて周囲領域に対応することの利点を確認する ため、全時刻を通した視線分布を可視化した.ここでは5.2 で 扱った3個の動画を用いた.その結果を図9に示す.図中(a) では表面領域のみ、(b)では周囲領域のみ、(c)では表面領域と 周囲領域とから、全時刻で周辺化された頂点注目確率を三次 元ヒートマップで可視化した.図中において、可視化結果の左 下部に示した数字は、全時刻を通して計測された観察者24名 全員分の視線のサンプリング総数を100%とした時、各領域で 計測された視線のサンプリング数の割合を表す.なお、図9(c) は、図7(d)と(h)と(l)で示した三つの仮想視点の中で中央に あたる結果と同一のものである.

動画の全時刻を通して周囲領域で視線がサンプリングされた



Fig. 9 Visualization of gaze distribution measured from surface and surrounding regions in all times of the movies.

割合は,被写体1と被写体2の動画で約5割,被写体3の動画 で約4割であった.このため周囲領域でサンプリングされる視 線の影響も無視できないと考えられる.被写体1の周囲領域で は,全時刻を通して頭や両腕や右脇や右太腿や左足首に観察者 の視線がやや集まっていた.被写体2や被写体3の周囲領域で は,全時刻を通して頭や左肩に観察者の視線がやや集まってい た.動画の全時刻を通して人体の表面領域から視線が外れて計 測されることが多い場合,周囲領域に対応する利点があると考 える.

5.5 被写体がその他の動作を行った場合の評価

被写体の動作の種類を増やした時の評価を行った.被写体4 の動画では左足を軸に回転を繰り返す動作を行い,被写体5の 動画ではバットを振る動作を行い,被写体6の動画ではロープ の上下振りを繰り返す動作を行うことにした.各動画から抽出 されたフレームの例を図 10(a)に示す.図中(b)では表面領域 のみ,(c)では周囲領域のみ,(d)では表面領域と周囲領域とか ら,全時刻で周辺化された頂点注目確率を三次元ヒートマップ で可視化した.動画以外の評価条件は4.と同じとした.なお, 図中において,可視化結果の左下部に示した数字は,全時刻を 通して計測された視線のサンプリング数の割合を表す.周囲領 域における視線のサンプリング割合は,被写体4と被写体6に おいて約2割,被写体5において約4割であった.設問Qに対 する4段階評価の平均スコアは,被写体4の動画で 2.9 ± 0.6 , 被写体 5 の動画で 2.6 ± 0.6 , 被写体 6 の動画で 3.5 ± 0.5 で あった.

同じ回転動作を繰り返す被写体4において,視線が頭や首に 主に集まっていた.バットを振る動作を1回行う被写体5にお いて,視線が頭に主に集まっていた.道具を用いながら同じ動 作を繰り返す被写体6において,視線が頭や胸元に主に集まっ ていた.なお,被写体6において,周囲領域で計測された視線



Fig. 10 Visualization of gaze distribution measured from the subjects performing other actions.

は、人体モデルの頂点注目確率にほぼ影響していなかった.人 体から離れた位置に存在する道具であるロープの先端に視線が 分散し、人体近くの周囲領域に集まる視線が少なかったためと 考える.提案手法による可視化は、人体のみを考慮しており周 囲に存在する道具を考慮していないため、5.3 で述べた限界が あると言える.以上の結果より、表面領域に加えて周囲領域に も対応している提案手法は、道具を使用しない動作や道具を身 体付近のみで使用する動作における視線分布を可視化できる有 効性があると考える.

6. まとめ

本論文では,観察者が動画中の被写体の動作から躍動感を受 ける時,被写体の表面領域とその周囲領域において計測された 視線分布を,被写体の姿勢を時間方向に統一した上で可視化す る手法について述べた.表面領域と周囲領域とを扱うため,仮 想空間における観察者の視線方向ベクトルと人体モデルの頂点 位置ベクトルとの間の角度を用いた.また,動画の視線分布に 対応するため,姿勢が時刻間で共通化された標準人体モデル を導入し,頂点注目確率を全時刻で周辺化した.実験結果にお いて,従来手法では,表面領域で計測された視線分布のみしか 可視化できなかったが,提案手法では,表面領域と周囲領域と の両方で計測された視線分布を可視化できることを確認した. また,従来手法では,動画中で時系列に変化する被写体の姿勢 へ対応できていなかったが,提案手法では,動画の全時刻を通 して姿勢が共通化された視線分布を可視化できることを確認 した.

今後の課題として,提案手法の有効性を示すためにユーザス タディを用いた定量的な評価を実施することが挙げられる.また,動画中に複数名の被写体が同時に存在する場合について, 視線分布をどのように可視化するかについて提案手法を発展さ せることが挙げられる.更に,被写体の動作に応じて観察者が 受ける躍動感が変化するかどうかについて,視線分布に基づき 調査を深めていくことが挙げられる.

謝辞

本研究の一部は, JSPS 科研費 JP23K11145 の助成を受けた ものである.

参考文献

- C. Piers, H. Peter, V. Kiviniemi, G. Hannah, and T. Martin. Patterns of eye movements when male and female observers judge female attractiveness, body fat and waist-to-hip ratio. *Evolution and Human Behavior*, Vol. 30, No. 6, pp. 417–428, 2009.
- L. Nummenmaa, H. Jari, P. Santtila, and J. Hyönä. Gender and visibility of sexual cues influence eye movements while viewing faces and bodies. *Archives of Sexual Behavior*, Vol. 41, No. 6, pp. 1439–1451, 2012.
- 3) K. R. Irvine, K. McCarty, T. V. Pollet, K. K. Cornelissen, M. J. Tovée, and P. L. Cornelissen. The visual cues that drive the self-assessment of body size: Dissociation between fixation patterns and the key areas of the body for accurate judgement. *Body Image*, Vol. 29, pp. 31–46, 2019.
- F. Inoue, M.and Iwasaki, M. Nishiyama, and Iwaim Y. Heatmap overlay using neutral body model for visualizing the measured gaze distributions of observers. In Asian Conference on Pattern Recognition, pp. 102–114, 2023.
- M. Matthew, N. Mahmood, J. Romero, G. Pons-Moll, and M.J. Black. SMPL: A skinned multi-person linear model. ACM Transactions on Graphics, Vol. 34, No. 6, pp. 1–16, 2015.
- 6) P. Georgios, V. Choutas, N. Ghorbani, T. Bolkart, A. A. A. Osman, D. Tzionas, and M. J. Black. Expressive body capture: 3D hands, face, and body from a single image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10975– 10985, 2019.
- A. A. A. Osman, T. Bolkart, D. Tzionas, and M. J. Black. SUPR: A sparse unified part-based human representation. In *Proceedings of European Conference on Computer Vision*, pp. 1–18, 2022.
- S. Goel, G. Pavlakos, J. Rajasegaran, A. Kanazawa, and J. Malik. Humans in 4D: Reconstructing and tracking humans with transformers. In *International Conference on Computer Vision*, pp. 1–12, 2023.
- R.A. Güler, N. Neverova, and I. Kokkinos. Densepose: Dense human pose estimation in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7297–7306, 2018.
- 10) Y. Xiu, J. Yang, D. Tzionas, and M. J. Black. Icon: Implicit clothed humans obtained from normals. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 13296– 13306, 2022.
- 11) V. Choutas, L. Müller, C. P. Huang, S. Tang, D. Tzionas, and M. J. Black. Accurate 3d body shape regression using metric and semantic attributes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2718–2728, 2022.
- 12) H. Zhang, Y. Tian, X. Zhou, W. Ouyang, Y. Liu, L. Wang, and Z. Sun. Pymaf: 3d human pose and shape regression with pyramidal mesh alignment feedback loop. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 11446–11456, 2021.
- M. Bindemann. Scene and screen center bias early eye movements in scene viewing. *Vision research*, Vol. 50, No. 23, pp. 2577–2587, 2010.