

THE IEICE TRANSACTIONS ON FUNDAMENTALS OF ELECTRONICS, COMMUNICATIONS AND COMPUTER SCIENCES (JAPANESE EDITION)

**IEICE** **電子情報通信学会**  
**A** **論文誌** 基礎・境界

VOL. J106-A NO. 3

MARCH 2023

本PDFの扱いは、電子情報通信学会著作権規定に従うこと。  
なお、本PDFは研究教育目的（非営利）に限り、著者が第三者に直接配布することができる。著者以外からの配布は禁じられている。

**基礎・境界ソサイエティ**

一般社団法人 **電子情報通信学会**

THE ENGINEERING SCIENCES SOCIETY

THE INSTITUTE OF ELECTRONICS, INFORMATION AND COMMUNICATION ENGINEERS

観衆の顔向きの時空間統合による注目対象の位置及び被注目度の推定

武田 一馬<sup>†a)</sup> 川西 康友<sup>††,†b)</sup> 平山 高嗣<sup>†††,†c)</sup> 出口 大輔<sup>†d)</sup>  
井手 一郎<sup>†e)</sup> 村瀬 洋<sup>†f)</sup> 柏野 邦夫<sup>††††g)</sup>

Estimation of Targets' Locations and Attention Degrees by Spatio-temporal Integration of Audiences' Facial Orientations

Kazuma TAKEDA<sup>†a)</sup>, Yasutomo KAWANISHI<sup>††,†b)</sup>, Takatsugu HIRAYAMA<sup>†††,†c)</sup>,  
Daisuke DEGUCHI<sup>†d)</sup>, Ichiro IDE<sup>†e)</sup>, Hiroshi MURASE<sup>†f)</sup>, and Kunio KASHINO<sup>††††g)</sup>

あらまし 本研究の目的は、多数の人物の視行動を分析することで、観衆が注目している複数の注目対象の位置の推定と、それらが注目されている度合（被注目度）を定量化することである。被注目度を推定する典型的な方法として、観衆の視線を推定し、その視線と物体の位置を対応付けることで、被注目度を推定することが考えられる。その場合、機器を設置するコストや手間をふまえると、観衆全体を一度に撮影した映像から視線を推定することが望ましい。しかし、このようにして撮影した映像から切り出した顔画像の解像度は観客ごとに撮影した場合と比べて小さく、視線推定精度は低い。そこで本論文では、低解像度でも比較的推定しやすい顔向きの時系列データを入力とし、これらを時空間的に統合することで、観衆が複数の注目対象を注視する状況下で注目対象の位置と被注目度を同時に推定する手法を提案する。提案手法の有効性を確認するため、アイドルのライブ公演を模したデータセットを構築し、注目対象の位置及び被注目度の推定精度を評価した。実験結果から、提案手法により比較手法と比べて被注目度の推定精度が向上することを確認した。

キーワード 群衆, 被注目度推定, 被注目対象推定, 時系列フィルタ

1. ま え が き

「目は口ほどに物を言う」ということわざにあるように、人間は無意識のうちに他者の視行動などから情報を取り入れている。例えば他者の視線に反応して同じ方向を注視する共同注視、自身に注目している人物

に注意を向ける相互注視などは、他者の意思を推測する上で重要な視行動である [1]。このことから、カメラなどを利用して人間の目がつ情報解析することで、その人が興味をもっている対象や思惑などを推測できる可能性がある。また、異なる時刻に同じ映像を注視している人の視線を重ね合わせることで、映像中の注目がより集まっている場所を推定する研究 [2] がなされている。このように、多数の人に対する視行動分析の結果を統合することで、映像コンテンツの評価など、より幅広い分析も可能になる。例えば、複数人の視線から注目対象となる物体がある位置を推定する手法 [3] や円卓に座る人物への注目度推定手法 [4] はその例である。

本研究では多数の人に対する視行動分析として、多数の観客（観衆）が複数の人物や物体（以下、注目対象）に注目している際の視行動を分析し、各注目対象が観衆から注目されている度合（被注目度；Attention Degree）を定量化することを目指す。このような状況の例として、ライブ公演や講演会などのイベント会場が挙げられる。例として図 1 に示すようなアイドルの

<sup>†</sup>名古屋大学, 名古屋市  
Nagoya University, Furo-cho, Chikusa-ku, Nagoya-shi, 464-8601 Japan  
<sup>††</sup>理化学研究所, 京都府  
RIKEN, 2-2-2 Hilaridai, Seika-cho, Soraku-gun, Kyoto, 619-0288 Japan  
<sup>†††</sup>人間環境大学, 岡崎市  
University of Human Environments, 6-2 Kami Sanbonmatsu, Motojuku-cho, Okazaki-shi, 444-3505 Japan  
<sup>††††</sup>日本電信電話株式会社, 厚木市  
Nippon Telegraph and Telephone Corporation, 3-1 Morinosato Wakamiya, Atsugi-shi, 243-0198 Japan  
a) E-mail: takedak@murase.is.i.nagoya-u.ac.jp  
b) E-mail: yasutomo.kawanishi@riken.jp  
c) E-mail: takatsugu.hirayama@nagoya-u.jp  
d) E-mail: ddeguchi@nagoya-u.jp  
e) E-mail: ide@i.nagoya-u.ac.jp  
f) E-mail: murase@nagoya-u.jp  
g) E-mail: kunio.kashino.me@hco.ntt.co.jp  
DOI:10.14923/transfunj.2022HAP0006

ライブ公演を考えた場合、観衆の前方にあるステージ上に複数のアイドルとバックスクリーンが存在することから、観衆の多くはこのいずれかに注目していると考えられる。このようなイベントにおいて、各注目対象の被注目度を定量化できれば、その高さに応じて特別な演出をすることもできるようになると考えられる。

複数の注目対象に対する観衆の被注目度を推定する場合の最も単純な方法は、各観客の視線と注目対象の位置を対応付けることで、各注目対象を注視している観客の人数を数える方法である。装着型の専用機器を用いた視線推定手法[5]を利用することで高精度に視線を得ることができるが、多数の観客に対してこのような手法を適用するには非常に高いコストがかかる。それに対し、観衆全体を単一のカメラで撮影した映像から個々の観客の顔を検出し、その顔画像から視線を推定することができれば、コストを抑えて視線を推定することが可能になる。しかし、この場合には観客を個別に撮影した場合に比べて顔画像の解像度が小さくなるため、視線推定精度は低くなる。一方で、顔向きは視線方向とある程度相関があるといわれている[6],[7]。このうち Funatsu らの研究[6]によれば、体の向きと顔向きが一致しているほど、視線方向と顔向きのずれ

は小さくなることが示されている。また、Stiefelhagen らの研究[7]によると、顔向きと眼球姿勢を足し合わせた値を視線方向とした場合、視線方向に顔向きが占める割合は 68.9% であることが示されており、視線方向は顔向きと強い関係があることが分かる。更に、Slaney らの研究[8]では、顔向きは視線方向の移動に対して平均 0.3 秒ほど遅れて追従することが示されている。このように、顔向きは視線方向をよく表す特徴であるといえる。このことから、本研究では比較的解像度の顔画像からでも安定して推定できると考えられる顔向きを、視線方向の代わりに用いることとした。

また、本研究で想定している状況下では、図 2 の例に示すように、注目対象がステージ上に複数存在し、時刻とともに位置が動的に変化する。そのため、注目対象の位置を事前に知ることは簡単ではない。一方で、注目対象の位置を直接知り得なくても、そこに多数の視線が集まることを仮定することで、視線の分布から近似的に推定することもできると考えられる。実際に、多数の視線から平面上での注視点を計算し、注視点の分布から注目対象の位置を高精度に推定する手法[9]や、共同注視点を用いた注視方向推定手法[10]が提案されている。

これらをふまえ、本研究では視線方向の代わりに顔向きを入力とし、多数の顔向きを時空間的に統合することで注目対象の位置と観客の注目対象を同時に推定し、これらから位置が未知である注目対象の被注目度を算出する手法を提案する。

本研究と最も近い手法として、注目対象の位置を考慮した視線推定手法[11]がある。この手法では観測した顔向きと注目対象の位置を用いて時系列フィルタにより推定することで高精度な推定を可能にしている。その一方で、注目対象の位置は既知として入力されるため、注目対象の位置が未知の場合には用いることができない。また、映像中の各人物の注目対象や視線方向の推定は人物ごとに独立して行われている。本研究



図 1 音楽のライブ公演の例  
Fig. 1 Example of live music performance.

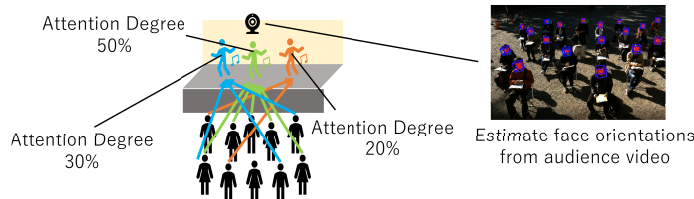


図 2 複数注目対象への被注目度推定の概念  
Fig. 2 Concept of Attention Degrees estimation to multiple targets.

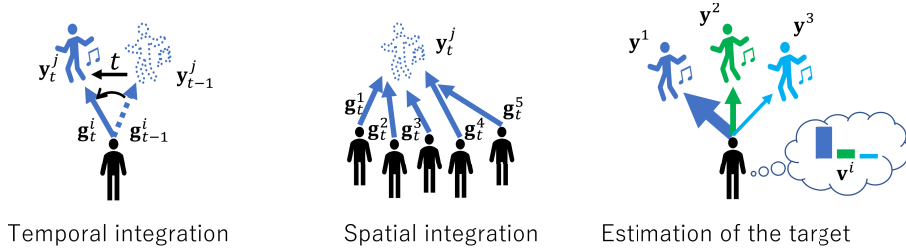


図3 顔向きの時空間統合による被注目度推定概念

Fig. 3 Concept of Attention Degrees estimation by spatiotemporal integration of face orientation.

のように幾つかの対象に注目が集まるような状況下では、ある観客が注目している対象は、他の多くの観客が見ている対象と一致することが多いと考えられる。そのため、視線方向の推定時に他の多くの観客が注目している対象の位置を考慮することで、より高精度な推定ができるようになる可能性がある。

そこで本研究では、注目対象の位置と各観客の視線及び注視している注目対象を潜在変数と考え、顔向きの時系列データから時系列フィルタを用いて各潜在変数の値を推定することで、注目対象の位置と被注目度を推定する。個々の観客の顔向きの時系列データを用いて注目対象を推定する時間的統合に加え、各時刻における多数の観客の顔向きを空間的に統合して同時に推定することで、視線方向に比べて曖昧な顔向きからでも高精度に注目対象の位置を推定する。これにより、複数の注目対象の位置が未知である場合であっても、それらの位置と被注目度を推定することを目指す。

## 2. 顔向きの時空間統合による注目対象の位置及び被注目度推定手法

### 2.1 提案手法の枠組み

本研究では、映像中の  $N$  人の観客  $i = 1, \dots, N$  からなる観衆のうち、ある注目対象  $j$  を注視している人数が  $n_j$  人のときに、その注目対象の被注目度を  $n_j/N$  と定義する。実際には観客が注目する可能性がある位置は特定の注目対象には限られないと考えられる。しかし、本研究で想定する状況では、多くの観客は、不特定の物体や領域を眺めているのではなく、ステージ上の演者のような特定の注目対象を見ていることを想定する。この想定により、本研究では特定の対象のみを注目対象として定義する。また、映像中で観客は必ず一つの注目対象を注視しており、その注目対象を注視し続けると仮定する。本手法では、入力である各観客の顔向きから各観客の視線方向を推定する。更に、視

線から注目対象の位置を推定すると同時に観客の注目対象を推定することで、各注目対象の被注目度を算出する。本節では、注目対象の位置が未知である場合と、誤差を含む注目対象の位置が観測できる場合に分けてそれぞれ手法を提案する。

#### 2.1.1 提案手法 1：注目対象の位置が未知である場合

本項では、注目対象の位置が未知である場合に、時刻  $t$  における  $N$  人の観客の顔向き集合  $\mathbf{h}_t = \{\mathbf{h}_t^1, \dots, \mathbf{h}_t^N\}$  を観測値とし、 $M$  個の注目対象の位置  $\{\mathbf{y}_t^1, \dots, \mathbf{y}_t^M\}$  をまとめた確率変数  $\mathbf{y}_t$  及び、観客それぞれの注目対象の確率分布  $\{\mathbf{v}^1, \dots, \mathbf{v}^N\}$  をまとめた確率変数  $\mathbf{v}$  を推定する手法を提案する。ここで、顔向きは Euler 角を表す 3 次元のベクトル、注目対象の位置は空間中の 3 次元位置で表される。また、映像中では観客の注目対象は変わらないことを仮定しているため、 $\mathbf{v}$  は時刻にかかわらず一定とする。図 3 に本手法の概念図を示す。顔向きの時空間統合により注目対象の位置と視線方向を推定すると同時に、それらを用いて観客の注目対象を推定する。以下で本手法の詳細について述べる。

ある時刻  $t$  における観衆の顔向き集合  $\mathbf{h}_t$  に対し、時刻  $t$  までの観衆の顔向き集合の列を  $\mathbf{H}_t = (\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_t)$  と表す。この  $\mathbf{H}_t$  を観測値として時系列フィルタを適用することで、時刻  $t$  における注目対象の位置  $\mathbf{y}_t$  及び各観客の注目対象に関する確率分布  $\mathbf{v}$  の期待値を、同時確率分布における条件付期待値として以下により求める。

$$\hat{\mathbf{y}}_t = E[\mathbf{y}_t | \mathbf{H}_t] = \iint \mathbf{y}_t p(\mathbf{y}_t, \mathbf{v} | \mathbf{H}_t) d\mathbf{y}_t d\mathbf{v} \quad (1)$$

$$\hat{\mathbf{v}} = E[\mathbf{v} | \mathbf{H}_t] = \iint \mathbf{v} p(\mathbf{y}_t, \mathbf{v} | \mathbf{H}_t) d\mathbf{y}_t d\mathbf{v} \quad (2)$$

なお、 $\mathbf{y}_t$  と  $\mathbf{v}$  は独立であると仮定する。ここで、Bayes の定理を用いることで  $p(\mathbf{y}_t, \mathbf{v} | \mathbf{H}_t)$  は、

$$\begin{aligned}
 p(\mathbf{y}_t, \mathbf{v} | \mathbf{H}_t) &= p(\mathbf{y}_t, \mathbf{v} | \mathbf{h}_t, \mathbf{H}_{t-1}) \\
 &\propto p(\mathbf{h}_t | \mathbf{y}_t, \mathbf{v}) p(\mathbf{y}_t, \mathbf{v} | \mathbf{H}_{t-1})
 \end{aligned} \quad (3)$$

と書き換えることができる。ここで、 $\mathbf{y}_t$  に対する 1 重 Markov 性と、 $\mathbf{y}_t$  と  $\mathbf{v}$ 、 $\mathbf{H}_{t-1}$  が独立であることを仮定することにより、

$$\begin{aligned}
 p(\mathbf{y}_t, \mathbf{v} | \mathbf{H}_{t-1}) &= \int p(\mathbf{y}_t, \mathbf{y}_{t-1}, \mathbf{v} | \mathbf{H}_{t-1}) d\mathbf{y}_{t-1} \\
 &\propto \int p(\mathbf{y}_t | \mathbf{y}_{t-1}) p(\mathbf{y}_{t-1}, \mathbf{v} | \mathbf{H}_{t-1}) d\mathbf{y}_{t-1}
 \end{aligned} \quad (4)$$

が導かれる。

ここで、本研究では特定の注目対象を注視しており、その場合視線方向と注目対象のある方向が一致すると考えられることから、観客の頭部位置から対象を見たときの真の視線方向を 3 次元単位ベクトルで表現し、観衆全員分のこの単位ベクトルをまとめた集合  $\mathbf{g}_t$  を潜在変数として導入すると、

$$\begin{aligned}
 p(\mathbf{h}_t | \mathbf{y}_t, \mathbf{v}) &= \int p(\mathbf{h}_t | \mathbf{g}_t) p(\mathbf{g}_t | \mathbf{y}_t, \mathbf{v}) d\mathbf{g}_t \\
 &= \int p(\mathbf{h}_t | \mathbf{g}_t) \int p(\mathbf{g}_t | \mathbf{g}_{t-1}, \mathbf{y}_t, \mathbf{v}) d\mathbf{g}_{t-1} d\mathbf{g}_t
 \end{aligned} \quad (5)$$

と表せる。ここで、 $p(\mathbf{y}_t | \mathbf{y}_{t-1})$  は注目対象の運動モデルを表し、 $p(\mathbf{g}_t | \mathbf{g}_{t-1}, \mathbf{y}_t, \mathbf{v})$  は視線の運動モデルを表している。また、視線が与えられたときの  $p(\mathbf{h}_t | \mathbf{g}_t)$  は顔向きのゆう度であり、 $p(\mathbf{y}_{t-1}, \mathbf{v} | \mathbf{H}_{t-1})$  は時刻  $t-1$  における事後分布である。

結果として、式 (1)~(5) より、

$$\begin{aligned}
 \hat{\mathbf{y}}_t &\propto \iint \mathbf{y}_t \int p(\mathbf{h}_t | \mathbf{g}_t) \int p(\mathbf{g}_t | \mathbf{g}_{t-1}, \mathbf{y}_t, \mathbf{v}) d\mathbf{g}_{t-1} d\mathbf{g}_t \\
 &\quad \int p(\mathbf{y}_t | \mathbf{y}_{t-1}) p(\mathbf{y}_{t-1}, \mathbf{v} | \mathbf{H}_{t-1}) d\mathbf{y}_{t-1} d\mathbf{v}
 \end{aligned} \quad (6)$$

$$\begin{aligned}
 \hat{\mathbf{v}} &\propto \iint \mathbf{v} \int p(\mathbf{h}_t | \mathbf{g}_t) \int p(\mathbf{g}_t | \mathbf{g}_{t-1}, \mathbf{y}_t, \mathbf{v}) d\mathbf{g}_{t-1} d\mathbf{g}_t \\
 &\quad \int p(\mathbf{y}_t | \mathbf{y}_{t-1}) p(\mathbf{y}_{t-1}, \mathbf{v} | \mathbf{H}_{t-1}) d\mathbf{y}_{t-1} d\mathbf{v}
 \end{aligned} \quad (7)$$

が導出される。

図 4 に本項で提案する手法のグラフィカルモデルを示す。本研究では、パーティクルフィルタ [12] によ

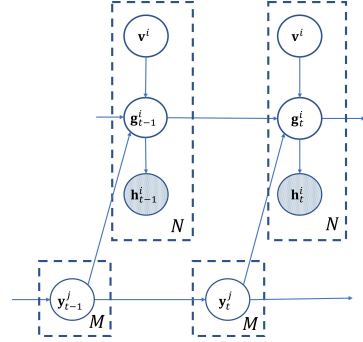


図 4 提案手法 1 のグラフィカルモデルによる表現  
Fig. 4 Graphical model representation of proposed method 1.

り各分布をパーティクルで近似して各潜在変数を推定する。

### 2.1.2 提案手法 2：注目対象の位置を観測できる場合

2.1.1 では、注目対象の位置が未知である場合に、注目対象の位置及び観客の注目対象を推定する手法を提案した。しかし実際のイベントでは、ステージ上を撮影するカメラや、あらかじめ決められた対象の移動経路をもとに、観測誤差を含む形で注目対象の位置が得られる場合もある。そこで本項では、観測誤差を含む注目対象の位置と観測した顔向きに基づいて、注目対象の正確な位置を推定しつつ、観客の注目対象を推定する手法を提案する。

提案手法 1 では、時刻  $t$  における  $N$  人の観客の頭部位置と顔向きの集合  $\mathbf{h}_t = \{\mathbf{h}_t^1, \dots, \mathbf{h}_t^N\}$  を観測値とした。本手法ではこれに加えて、時刻  $t$  における  $M$  個の注目対象の誤差を含む位置の集合  $\mathbf{x}_t = \{\mathbf{x}_t^1, \dots, \mathbf{x}_t^M\}$  を観測値として入力する。

入力された顔向き及び誤差を含む注目対象の位置に基づいて、 $M$  個の注目対象の正確な位置  $\{\mathbf{y}_t^1, \dots, \mathbf{y}_t^M\}$  をまとめた確率分布  $\mathbf{y}_t$  及び、観客それぞれの注目対象の確率分布  $\{\mathbf{v}^1, \dots, \mathbf{v}^N\}$  をまとめた確率変数  $\mathbf{v}$  を推定する。以下で本手法の詳細について述べる。

ある時刻  $t$  における観衆の顔向き集合を  $\mathbf{h}_t$  とすると、時刻  $t$  までの観衆の顔向き集合の列は  $\mathbf{H}_t = (\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_t)$  と表せる。また、時刻  $t$  までの注目対象の誤差を含む位置集合の列を  $\mathbf{X}_t = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_t)$  と表す。この  $\mathbf{H}_t$  及び  $\mathbf{X}_t$  を観測値として時系列フィルタを適用することで、時刻  $t$  における注目対象の正確な位置  $\mathbf{y}_t$  及び各観客がそれぞれ注視している注目対象に関する確率分布  $\mathbf{v}$  を、同時確率分布における条

件付期待値として以下により求める.

$$\hat{\mathbf{y}}_t = E[\mathbf{y}_t | \mathbf{H}_t, \mathbf{X}_t] = \iint \mathbf{y}_t p(\mathbf{y}_t, \mathbf{v} | \mathbf{H}_t, \mathbf{X}_t) d\mathbf{y}_t d\mathbf{v} \quad (8)$$

$$\hat{\mathbf{v}} = E[\mathbf{v} | \mathbf{H}_t, \mathbf{X}_t] = \iint \mathbf{v} p(\mathbf{y}_t, \mathbf{v} | \mathbf{H}_t, \mathbf{X}_t) d\mathbf{y}_t d\mathbf{v} \quad (9)$$

なお,  $\mathbf{y}_t$  と  $\mathbf{v}$  は独立であると仮定する. ここで,  $\mathbf{h}_t$  と  $\mathbf{x}_t$ ,  $\mathbf{y}_t$  と  $\mathbf{H}_{t-1}$ ,  $\mathbf{X}_{t-1}$  もそれぞれ互いに独立であると仮定すると, Bayes の定理を用いて,

$$\begin{aligned} p(\mathbf{y}_t, \mathbf{v} | \mathbf{H}_t, \mathbf{X}_t) &= p(\mathbf{y}_t, \mathbf{v} | \mathbf{h}_t, \mathbf{H}_{t-1}, \mathbf{x}_t, \mathbf{X}_{t-1}) \\ &\propto p(\mathbf{h}_t, \mathbf{x}_t | \mathbf{y}_t, \mathbf{v}) p(\mathbf{y}_t, \mathbf{v} | \mathbf{H}_{t-1}, \mathbf{X}_{t-1}) \end{aligned} \quad (10)$$

が得られる. 更に,  $\mathbf{y}_t$  に対して 1 重 Markov 性を仮定することにより,

$$\begin{aligned} p(\mathbf{y}_t, \mathbf{v} | \mathbf{H}_{t-1}, \mathbf{X}_{t-1}) &= \int p(\mathbf{y}_t, \mathbf{y}_{t-1}, \mathbf{v} | \mathbf{H}_{t-1}, \mathbf{X}_{t-1}) d\mathbf{y}_{t-1} \\ &\propto \int p(\mathbf{y}_t | \mathbf{y}_{t-1}) p(\mathbf{y}_{t-1}, \mathbf{v} | \mathbf{H}_{t-1}, \mathbf{X}_{t-1}) d\mathbf{y}_{t-1} \end{aligned} \quad (11)$$

が導かれる. ここで,  $p(\mathbf{y}_t | \mathbf{y}_{t-1})$  は注目対象の運動モデルであり,  $p(\mathbf{y}_{t-1}, \mathbf{v} | \mathbf{H}_{t-1}, \mathbf{X}_{t-1})$  は時刻  $t-1$  における事後分布である.

また,  $\mathbf{h}_t$  と  $\mathbf{x}_t$  が独立であると仮定し, 提案手法 1 と同様に  $\mathbf{g}_t$  を潜在変数として導入すると,

$$\begin{aligned} p(\mathbf{h}_t, \mathbf{x}_t | \mathbf{y}_t, \mathbf{v}) &= p(\mathbf{x}_t | \mathbf{y}_t, \mathbf{v}) p(\mathbf{h}_t | \mathbf{y}_t, \mathbf{v}) \\ &= p(\mathbf{x}_t | \mathbf{y}_t) \int p(\mathbf{h}_t | \mathbf{g}_t) \int p(\mathbf{g}_t | \mathbf{g}_{t-1}, \mathbf{y}_t, \mathbf{v}) d\mathbf{g}_{t-1} d\mathbf{g}_t \end{aligned} \quad (12)$$

と表せる.

ここで,  $p(\mathbf{g}_t | \mathbf{g}_{t-1}, \mathbf{y}_t, \mathbf{v})$  は視線の運動モデルを表している. また,  $p(\mathbf{x}_t | \mathbf{y}_t)$  は観測誤差を含む注目対象位置のゆう度,  $p(\mathbf{h}_t | \mathbf{g}_t)$  は顔向きゆう度である.

結果として, 式 (8)~(12) より,

$$\begin{aligned} \hat{\mathbf{y}}_t &\propto \iint \mathbf{y}_t p(\mathbf{x}_t | \mathbf{y}_t) \int p(\mathbf{h}_t | \mathbf{g}_t) \\ &\quad \int p(\mathbf{g}_t | \mathbf{g}_{t-1}, \mathbf{y}_t, \mathbf{v}) d\mathbf{g}_{t-1} d\mathbf{g}_t \\ &\quad \int p(\mathbf{y}_t | \mathbf{y}_{t-1}) p(\mathbf{y}_{t-1}, \mathbf{v} | \mathbf{H}_{t-1}, \mathbf{X}_{t-1}) d\mathbf{y}_{t-1} d\mathbf{y}_t d\mathbf{v} \end{aligned} \quad (13)$$

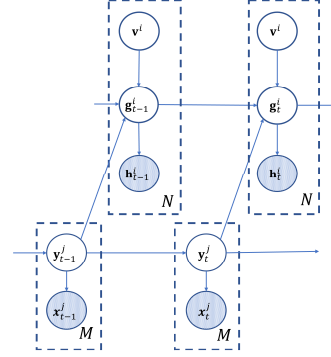


図5 提案手法 2 のグラフィカルモデルによる表現  
Fig. 5 Graphical model representation of proposed method 2.

$$\begin{aligned} \hat{\mathbf{v}} &\propto \iint \mathbf{v} p(\mathbf{x}_t | \mathbf{y}_t) \int p(\mathbf{h}_t | \mathbf{g}_t) \\ &\quad \int p(\mathbf{g}_t | \mathbf{g}_{t-1}, \mathbf{y}_t, \mathbf{v}) d\mathbf{g}_{t-1} d\mathbf{g}_t \\ &\quad \int p(\mathbf{y}_t | \mathbf{y}_{t-1}) p(\mathbf{y}_{t-1}, \mathbf{v} | \mathbf{H}_{t-1}, \mathbf{X}_{t-1}) d\mathbf{y}_{t-1} d\mathbf{y}_t d\mathbf{v} \end{aligned} \quad (14)$$

が導出できる.

図 5 に本項で提案する手法のグラフィカルモデルを示す. 本研究では, パーティクルフィルタ [12] により各分布をパーティクルで近似して各潜在変数を推定する.

## 2.2 注目対象及び視線の運動モデル

注目対象の運動モデル  $p(\mathbf{y}_t | \mathbf{y}_{t-1})$  の計算は, 注目対象  $j$  ごとのパーティクル  $(s_1) \mathbf{y}_t^j (s_1 = 1, \dots, S_1)$  を用いて近似する. ここで,  $(s_1)$  は, 確率分布  $p(\mathbf{y}_t | \mathbf{y}_{t-1})$  を  $S_1$  個のパーティクルで表現するときのパーティクルの番号である. 本手法では注目対象の位置  $\mathbf{y}_t$  及び観客の視線方向  $\mathbf{g}_t$  に 1 重 Markov 性を仮定しており, かつ隣接フレーム間における注目対象位置の変化は大きくないと考えられることから, 注目対象の移動にランダムウォークを仮定し,

$$(s_1) \mathbf{y}_t^j = (s_1) \mathbf{y}_{t-1}^j + \varepsilon_y \quad (15)$$

$$\varepsilon_y \sim \mathcal{N}(0, \sigma_y^2) \quad (16)$$

として時刻  $t$  における注目対象  $j$  の位置  $\mathbf{y}_t^j$  の分布を表す. また, 視線の運動モデル  $p(\mathbf{g}_t | \mathbf{g}_{t-1}, \mathbf{y}_t, \mathbf{v})$  に対しても, 観客  $i$  ごとに, パーティクル  $(s_2) \mathbf{g}_t^i (s_2 = 1, \dots, S_2)$  を用いて視線の確率分布  $p(\mathbf{g}_t | \mathbf{g}_{t-1}, \mathbf{y}_t, \mathbf{v})$  を近似する. まずランダムウォークを仮定し,

$$(s_2) \mathbf{g}_t^i = (s_2) \mathbf{g}_{t-1}^i + \varepsilon_g \quad (17)$$

$$\varepsilon_{\mathbf{g}} \sim \mathcal{N}(0, \sigma_{\mathbf{g}}^2) \quad (18)$$

として、時刻  $t$  における観客  $i$  の視線方向を  $\mathbf{g}_{t-1}^i$  のみから予測したときの  $p(\mathbf{g}_t^i | \mathbf{g}_{t-1}^i)$  を表す。ここで、各観客の視線は、その観客の注目対象の位置に依存する。そこで、観客  $i$  の視線方向予測  $^{(s_2)}\mathbf{g}_t^i$  と、観客  $i$  の注目対象  $j$  の方向  $r_i \left( ^{(s_1)}\mathbf{y}_t^j \right)$  との重み付き和をとる。このとき、観客がより注視している確率が高い注目対象の方向に視線が向くと考えられることから、観客  $i$  が注目対象  $j$  を注視している確率  $v^{ij}$  も重みとして用いる。具体的には、

$$^{(s_2)}\mathbf{g}_t^i = \sum_{j=1}^M v^{ij} \left( (1 - \alpha) ^{(s_2)}\widehat{\mathbf{g}}_t^i + \alpha r_i \left( ^{(s_1)}\mathbf{y}_t^j \right) \right) \quad (19)$$

として、注目対象の位置を考慮した視線方向  $\mathbf{g}_t$  の分布を得る。ただし、 $\alpha$  は注目対象の位置を考慮する度合いの重みであり、本研究では  $\alpha = 0.5$  とした。

### 2.3 観測値のゆう度推定

本研究では、視線方向は顔向きの方向で近似できると仮定し、観測した顔向きのゆう度  $p(\mathbf{h}_t | \mathbf{g}_t)$  を、視線方向との類似度として定義する。観客  $i$  の顔向き  $\mathbf{h}_t^i$  に対して、予測した視線方向  $\mathbf{g}_t^i$  を用いて

$$p(\mathbf{h}_t^i | \mathbf{g}_t^i) \propto l_f(\mathbf{g}_t^i; \mathbf{h}_t^i) = e^{-k \|\mathbf{g}_t^i - \mathbf{h}_t^i\|^2} \quad (20)$$

を計算し、観客  $i$  の顔向き  $\mathbf{h}_t^i$  のゆう度  $l_f(\mathbf{g}_t^i; \mathbf{h}_t^i)$  を得る。ここで、 $k$  は観測値と予測値のずれを許容する程度を決定する係数である。

また、注目対象  $j$  の誤差を含む注目対象の位置を観測できる場合、推定した注目対象の位置  $\mathbf{y}_t^j$  を用いて観測した位置  $\mathbf{x}_t^j$  のゆう度を計算する。具体的には、

$$p(\mathbf{x}_t^j | \mathbf{y}_t^j) = e^{-k \|\mathbf{x}_t^j - \mathbf{y}_t^j\|^2} \quad (21)$$

を計算し、各注目対象の誤差を含む位置に対するゆう度を得る。

### 2.4 パーティクルフィルタを用いた被注目度推定

本研究では、パーティクルフィルタを用いて潜在変数を推定する。また、 $\mathbf{v}$  と  $\mathbf{y}_t$  は片方を固定し、反復的に最適化する。まず  $\mathbf{v}$  を固定し、2.2 で述べた運動モデルを用いて時刻  $t$  での注目対象の位置と観客の視線方向を予測する。次に、観客の視線に対するパーティクル  $^{(s_2)}\mathbf{g}_t^i$  に対して 2.3 で述べたゆう度を計算する。ここで、注目対象の位置はその注目対象を注視している観客の視線が集まっている場所にあると考えら

れる。また、よりゆう度が高い視線ほど推定において参考になると考えられることを考慮し、注目対象の位置のパーティクル  $^{(s_1)}\mathbf{y}_t$  に対して、予測した視線方向  $^{(s_2)}\mathbf{g}_t^i$  及び各観客の注目対象  $\mathbf{v}^i$ 、顔向き  $\mathbf{h}_t^i$  に対するゆう度  $l_f \left( ^{(s_2)}\mathbf{g}_t^i; \mathbf{h}_t^i \right)$  を用いて、

$$l_y \left( ^{(s_1)}\mathbf{y}_t^j; \mathbf{g}_t, \mathbf{v}, \mathbf{h}_t \right) = \frac{\sum_{i=1}^N l_f \left( ^{(s_2)}\mathbf{g}_t^i; \mathbf{h}_t^i \right) v^{ij} l_g \left( ^{(s_2)}\mathbf{g}_t^i; r_i \left( ^{(s_1)}\mathbf{y}_t^j \right) \right)}{\sum_{i=1}^N l_f \left( ^{(s_2)}\mathbf{g}_t^i; \mathbf{h}_t^i \right) v^{ij}} \quad (22)$$

を計算し、注目対象位置のゆう度  $l_y \left( ^{(s_1)}\mathbf{y}_t^j; \mathbf{g}_t, \mathbf{v}, \mathbf{h}_t \right)$  を得る。ここで、 $l_g \left( ^{(s_2)}\mathbf{g}_t^i; r_i \left( ^{(s_1)}\mathbf{y}_t^j \right) \right)$  は、観客  $i$  の視線方向  $^{(s_2)}\mathbf{g}_t^i$  と、観客  $i$  がいる注目対象を注視していた場合の視線方向  $r_i \left( ^{(s_1)}\mathbf{y}_t^j \right)$  の類似度である。観客  $i$  がその注目対象を注視しているであろう確率分布  $\mathbf{v}_i$  と、予測した視線に対する観測値のゆう度  $l_f \left( ^{(s_2)}\mathbf{g}_t^i; \mathbf{h}_t^i \right)$  を用いることで、それぞれの注目対象にとって重要かつ精度が高いと思われる視線に高い重みを付けている。そして、これらのゆう度と 2.3 で計算したゆう度を各パーティクルの重みとしてリサンプリングを行い、 $\mathbf{y}_t^j$  及び  $\mathbf{g}_t^i$  を更新する。

次に、推定した注目対象位置及び視線方向をもとに、観客  $i$  が各注目対象を注視している確率の分布  $\mathbf{v}^i$  を更新する。更新前の確率分布を  $\mathbf{v}^i$  とし、観客  $i$  に対して予測した時刻  $T$  までの視線の列  $(\mathbf{g}_1^i, \dots, \mathbf{g}_T^i)$  と、観客  $i$  が注目対象  $j$  を注視していた場合の時刻  $T$  までの視線方向の列  $(r(\mathbf{y}_1^j), \dots, r(\mathbf{y}_T^j))$  を用いて、

$$v^{ij} = v'^{ij} \frac{1}{T} \sum_{t=1}^T \max \left( 0, \eta - \left\| \mathbf{g}_t^i - r \left( \mathbf{y}_t^j \right) \right\| \right) \quad (23)$$

を計算し、最後に  $v^{ij} (j = 1, \dots, M)$  の総和が 1 になるよう正規化をすることで、確率分布  $\mathbf{v}$  を更新する。ここで、 $\eta$  は計算時に考慮する角度誤差の度合を決める定数である。一連の推定と確信度の更新を被注目度の変動がしきい値以下に達するまで反復した上で、最後に全ての観客の  $v^{ij}$  の平均値を

$$A^j = \frac{1}{N} \sum_{i=1}^N v^{ij} \quad (24)$$

として算出することで、注目対象  $j$  の被注目度  $A^j$  を算出する。推定の不確かさを扱うため、各観客の注目

対象に関する確率分布の平均値を計算している。

### 3. 実 験

#### 3.1 データセット

本研究では、ステージ上を移動している複数の注目対象を多数の観客が注視している状況を想定している。しかしこのような条件を満たす公開データセットは存在しないことから、実験用のデータセットを独自に構築した。以下では、データセット構築における撮影条件及び設定について述べる。

本研究で想定する状況を再現するために、幅 8 m、奥行き 5 m のステージ上に複数の注目対象を、ステージ前方の観客席に 24 人の観客（実験参加者）を各々配置した。図 6 に撮影の様子を示す。注目対象には合図とともにあらかじめ決められた移動経路に沿って歩くよう指示し、各観客にはあらかじめ指定した注目対象を注視し続けるよう指示した。図 7 に撮影した映像から切り出した画像の例を示す。実験参加者は 48 名の男女であり、24 名ずつの二つのグループに分けて、あらかじめ指定した席に着席した状態で注視を行わせた。

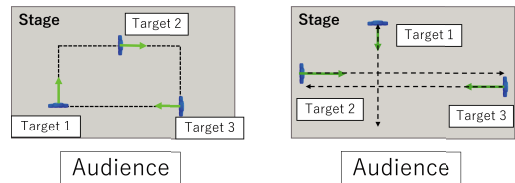
本研究では、ステージ上を移動する複数の注目対象の被注目度推定を目的とする。そこで、移動経路による差異を分析するため、注目対象の移動パターンを 2 種類用意し、それぞれの移動パターンに対して、各観客が注視する注目対象や各注目対象を注視する観客の人数の比率を変更しながら撮影を行った。移動パターンを図 8 に示す。ステージは平面であるため、例えばパターン 1 において注目対象がステージ奥を移動している際には、観客から見て手前にはほかの注目対象が重なって見える区間がある。パターン 1 は 18 m を約 24 秒、パターン 2 は 6 m を約 8 秒で移動しており、移動速度は毎秒約 0.75 m である。なお、注目対象の数は三つに固定した。また、移動経路は平面上の 2 次元の移動であるが、注目対象の位置は高さ方向を加えた 3 次元のデータとして扱った。三つの注目対象を注視する人数の比率として、8:8:8、12:8:4、18:6:0 の 3 種類の設定を用意した。比率が 18:6:0 の場合は、どの観客にも注視されていない注目対象が存在する。一つの比率に対して 2 種類の移動パターンを撮影して 1 セットとし、このセットを 3 種類の比率に対してそれぞれ 2 回ずつ撮影することで、12 本の映像を撮影した。また、各セットの合間に、各観客が注視する注目対象を無作為に変更した。このように各注目対象に注目する観客の配置を無作為にすることで、被注目度の推定値に偏



図 6 撮影の様子  
Fig. 6 Recording scene.



図 7 撮影した観衆映像の例  
Fig. 7 Examples of recorded audience video.



(a) パターン 1 (b) パターン 2

図 8 注目対象の移動パターン  
Fig. 8 Moving pattern of the gaze targets.

りが生じないようにした。これを 2 グループに対して 2 回ずつ行い、計 48 本の映像を撮影した。なお、撮影した映像の解像度は  $4,096 \times 2,160$  画素であり、顔領域の解像度は  $55 \times 72$  画素から  $136 \times 169$  画素程度であった。

一方、観客に対しては、より自然な環境下での様子を撮影するために、注視すべき注目対象のみ指示し、注目対象のどこを注視するかや、顔や目の動かし方などは指示しなかった。なお、1 回の撮影中は注目対象を切り替えず、開始から終了まで同じ注目対象を注視し続けるように指示した。

#### 3.2 実験方法

本研究では顔向き推定器及び顔検出器として OpenFace [13] を用い、世界座標系での各観客の顔向きと頭



部位置を推定した。この推定結果を入力顔向きとし、注目対象の位置が未知であることを想定した提案手法1及び、注目対象の誤差を含む位置が観測できることを想定した提案手法2を用いて、注目対象の位置及び被注目度を推定した。また、顔向きの推定精度が手法に与える影響を検証するために、顔向きの正解値に対して正規分布から無作為に抽出した誤差を上下及び左右方向にそれぞれ加えたデータを生成して検証した。この際の顔向きの正解値として、本実験では撮影した映像から推定した頭部位置と、観客ごとにあらかじめ指示した注目対象の位置を結んだ方向を用いた。なお、公演開始時における注目対象の初期位置は決まることが多いことから、注目対象の移動パターンは未知としつつも、初期位置は既知とし、この位置と最初の時刻における各観客の顔向きを用いて  $v^i$  の初期値を設定し、実験を行った。

実験結果の評価指標として、注目対象の位置推定と被注目度推定のそれぞれに対して平均絶対誤差を用いる。評価はまず各映像に対して行い、それを平均したものを最終的な実験結果とする。また、提案手法1に対しては注目対象の位置推定誤差を左右・上下・奥行き方向それぞれに対して評価したが、提案手法2に対しては観測誤差を含む注目対象の位置を入力として与えることから、位置推定精度を評価せず、被注目度のみを評価した。

比較手法として、ステージ上の空間に各時刻における視線を重ね合わせることで、空間中の各座標に集まる視線のヒートマップを作成し、その値から注目対象の位置と被注目度をそれぞれ推定する手法を用意した。提案手法1との比較としては、各座標における値の大きさに応じた個数生成した点について、混合 Gaussian モデルを用いてクラスタリングし、各クラスに割り当てられた要素の座標の平均位置を注目対象の位置、その周辺ヒートマップの値を平均したものを被注目度とする（比較手法1）。提案手法2との比較としては、入力された観測誤差を含む注目対象位置周辺ヒートマップの平均値を被注目度とする（比較手法2）。この提案手法1及び提案手法2並びに、比較手法1及び比較手法2に対して、

- 正解値の顔向きに対して誤差を加えない場合
- $\sigma = 5^\circ$  及び  $\sigma = 10^\circ$  の正規分布から抽出した誤差を加えた場合
- OpenFace で推定した顔向きの場合による結果を比較した。

### 3.3 実験結果

表1及び表2に実験結果を示す。表1の実験結果から、提案手法1において、注目対象の位置及び被注目度推定の結果が比較手法1より優れていることが分かる。しかし、OpenFaceにより推定した顔向きに対する結果を見ると、顔向きの正解値に誤差を加えたデータの結果に比べて、注目対象の位置推定及び被注目度推定の誤差が非常に大きい。

次に、表2に示す提案手法2における結果を提案手法1の結果と比較すると、顔向きに加えて注目対象位置の観測値を入力したことで、全体的に被注目度推定誤差が減少していることが分かる。

また、OpenFaceで推定した顔向きを入力とした場合に、比較的正しく推定できていると考えられる提案手法2の被注目度の推定結果と、各注目対象を注視する人数比率ごとに分けた際の各注目対象の被注目度の真値を、図9にバイオリンプロット[14]で示す。図中の赤い十字は推定結果の中央値を示し、プロットの形状は推定結果の確率分布を示している。また、注目対象は三つ存在するが、観衆はそのうちの二つしか見えない実験条件（比率18:6:0）が存在するため、被注目度が0%の場合が存在する。図9から、比率が偏るほど真値と中央値のずれが大きくなることが分かる。

## 4. 考察

表1及び表2の結果を見ると、提案手法と比較手法において被注目度推定の精度に大きな差が見られる。これは、比較手法では時系列情報を用いた追跡を行っていないため、観客から見て複数の注目対象が重なって見えるような時刻においては、両方の注目対象に注目しているように計算されてしまい、実際には注目対象ごとに被注目度が偏っていたとしても平均化されてしまうためだと考えられる。一方、提案手法においては時系列情報を用いて注目対象の追跡と各観客の注目対象の推定を同時に行うため、注目対象が一時的に重なって見えるような場合においても、より注視しているであろう注目対象を推定して計算ができたため、誤差が小さかったと考えられる。この誤差低下による効果を示す例として、注視人数の比率に偏りがあるデータに対して、注視人数に応じた被注目度の順位付けの正解率を表3に示す。なお、それぞれの比率に対する映像は16本ずつあり、三つの注目対象の被注目度推定結果の大小関係が注視人数比率の大小関係と同じ場合を正解、それ以外を不正解として正解率を計算した。

表 1 注目対象の位置を未知とした場合の位置推定及び被注目度推定の誤差

Table 1 Errors in location and Attention Degrees estimation when the locations of the targets are unknown.

手法	顔向き	左右誤差 [m]	上下誤差 [m]	前後誤差 [m]	被注目度誤差 [%]
比較手法 1	正解値	0.340	0.072	0.475	12.94
	正解値 + 誤差 ( $\sigma = 5^\circ$ )	0.786	0.272	0.739	14.39
	正解値 + 誤差 ( $\sigma = 10^\circ$ )	1.256	0.451	0.891	14.06
	OpenFace [13] で推定	1.134	0.338	1.023	13.88
提案手法 1	正解値	0.213	0.187	0.234	1.04
	正解値 + 誤差 ( $\sigma = 5^\circ$ )	0.346	0.193	0.357	2.18
	正解値 + 誤差 ( $\sigma = 10^\circ$ )	1.132	0.219	0.667	8.56
	OpenFace [13] で推定	2.025	0.119	1.289	15.03

表 2 注目対象の位置を既知とした場合の被注目度推定の誤差

Table 2 Errors in Attention Degrees estimation when the locations of the targets are known.

手法	顔向き	被注目度誤差 [%]
比較手法 2	正解値	5.79
	正解値 + 誤差 ( $\sigma = 5^\circ$ )	6.76
	正解値 + 誤差 ( $\sigma = 10^\circ$ )	8.98
	OpenFace [13] で推定	12.57
提案手法 2	正解値	0.30
	正解値 + 誤差 ( $\sigma = 5^\circ$ )	0.33
	正解値 + 誤差 ( $\sigma = 10^\circ$ )	0.91
	OpenFace [13] で推定	<b>8.72</b>
提案手法 1 (再掲)	正解値	(1.04)
	正解値 + 誤差 ( $\sigma = 5^\circ$ )	(2.18)
	正解値 + 誤差 ( $\sigma = 10^\circ$ )	(8.56)
	OpenFace [13] で推定	(15.03)

また、比率が 8:8:8 のデータでは被注目度が等しくなり順位付けができないため、計算を行っていない。表 3 から、提案手法 2 では比較手法 2 に比べて順位付け精度が向上しており、特に偏りが大きいデータでは高い精度で順位付けできていることを確認した。

また、被注目度の真値と推定値の関係として図 9 を見ると、三つの注目対象間の被注目度の差は、真値と比較して小さくなる傾向がある。例えば比率 18:6:0 のデータに対する結果では、真値では被注目度が 75.0%:25.0%:0.0% であるが、推定の中央値はそれぞれ 52.6%:33.8%:12.5% である。本手法では、観客の顔向きと注目対象がある方向の一致度合に基づいて各観客の注目対象を推定しており、注目対象が左に動いた際に顔向きも同時に左方向に変化するなど、両者の同期が取れていると一致しやすくなる。しかし、観客がステージから離れている場合などには、顔向きの変化量が小さく同期が取れなくなり、結果として複数の注目対象をそれぞれ均等な確率で注視しているように推定してしまうためだと考えられる。注視している観客の数が多き注目対象ほどこのような観客が含まれやす

表 3 注目対象の注視人数の偏りに応じた被注目度の順位付けに対する正解率 [%]

Table 3 Percentage of correct responses to ranking of Attention Degrees according to the bias of the number of people gazing at the target.

手法	比率 12:8:4	比率 18:6:0
比較手法 2	37.5	68.8
提案手法 2	<b>68.8</b>	<b>93.8</b>

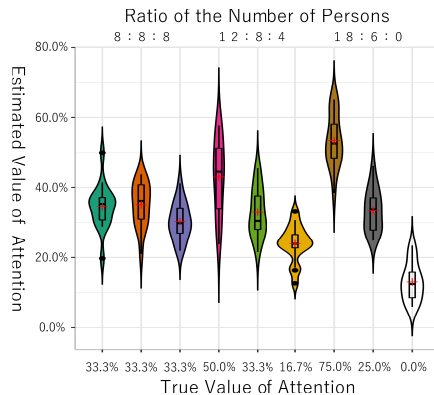


図 9 各注視人数比率における各注目対象の真の被注目度と推定値の関係

Fig. 9 Relation between true attention and estimated values of each target at each ratio of number of people gazing at it.

くなるため、図 9 に表される傾向が生じたと考えられる。また、図 9 において比率 12:8:4 のデータに対する結果を見ると、ばらつきが大きく出ていることが分かる。これは、無作為に観客の位置を入れ替えた結果、前方や後方の観客が固まってしまう、推定に影響を与えたことなどが考えられる。

本研究は、注目対象の位置が未知である場合においても、観客がどこに注目しているかを推定する課題として取り組んでいる。この課題をふまえて実験結果を見ると、提案手法 1 のように位置が未知の場合には、顔向きの推定精度が低下すると比較手法に比べて被注

目度推定精度が低下している一方、顔向きの推定精度が高い場合には比較手法に比べて被注目度推定精度が向上していることが分かる。また、提案手法2のように注目対象の位置を観測できる場合には、比較手法と比べて精度が向上しており、提案手法の有効性が示されている。これらの結果は、顔向き推定精度の向上や、周辺情報の活用により、提案手法を用いた性能改善ができることを示していると考えられる。

## 5. む す び

本論文では、観衆映像を用いて、ステージ上に存在する複数の注目対象に対し、各注目対象の位置及び被注目度を推定する手法を提案した。複数人の顔向きや視線を用いた研究には、注目対象となる静止物の位置推定[3]や円卓における被注目度推定[4]などがあるが、移動する注目対象に対して推定ができないことや、前提として注目対象の位置が既知である必要があることなどの課題があった。そこで本研究では、多数の観客の顔向きを時空間統合することで、位置が未知であり、移動する複数の注目対象に対して、それらの位置及び被注目度を同時に推定する手法を提案した。

一般に、観客が注視している注目対象を推定するためには、観客の視線方向及び注目対象の位置を正確に知り、これらに対応付けることが必要である。その際には、コストや手間の問題から、観衆全体を単一のカメラで撮影した映像から個々の観客の視線を推定するのが望ましい。しかし、このように観衆全体を撮影した映像から切り出した顔画像は、各観客を個々に撮影した場合と比べて低解像度であり、視線推定の精度は低い。そこで本研究では、視線に代わる入力として、視線方向と相関があり、低解像度でも比較的推定がしやすい顔向きを用いた。

また、実際のライブ公演などのイベントでは、ステージ上を撮影するカメラなどを用いることで注目対象の位置を観測できる場合も考えられる。そこで本研究では、顔向きに加えて誤差を含む注目対象の位置の観測値を入力するように変更したモデルも提案した。観測誤差を含む注目対象の位置のゆが度を評価するとともに、注目対象の位置を推定し、被注目度を推定した。

提案手法の有効性を確認するために、実際のアイドルのライブ公演を模した観衆映像を撮影し、OpenFace [13]により推定した顔向きを用いてデータセットを構築した。また、顔向きの推定精度が与える影響を検証するため、顔向きの正解値に誤差を加えたデー

タによるデータセットを構築した。そして、これらのデータセットを用いて注目対象の位置及び被注目度を推定する実験を行った。実験結果から、正解値に誤差を加えたデータにおいて、注目対象の位置が未知であっても、高精度な注目対象位置及び被注目度推定ができることを確認した。また、注目対象が未知である場合には推定した顔向きを入力とした推定は困難であったが、注目対象の位置の観測値を入力として与えることで大きな精度向上が得られた。

今後の課題としては、現在の手法における注目対象の数や位置、観客の注視方法に対する制約を減らしていくことが必要であると考えられる。

謝辞 本研究の一部は科研費(17H00745)による。

## 文 献

- [1] 岡本早苗, “視線追従能力の発達—比較認知発達の視点からの実験的検証—,” 心理学評論, vol.52, no.2, pp.111–124, March 2009. DOI:10.24602/sjpr.52.2\_165.  
S. Okamoto, “Development of gaze following from a comparative perspective,” Jpn. Psychol. Rev. (Japanese Edition), vol.52, no.2, pp.111–124, March 2009. DOI:10.24602/sjpr.52.2\_165.
- [2] Y. Sugano, X. Zhang, and A. Bulling, “AggreGaze: Collective estimation of audience attention on public displays,” Proc. 29th ACM Annual Symposium on User Interface Software and Technology (UIST2016), pp.821–831, Tokyo, Japan, Oct. 2016. DOI:10.1145/2984511.2984536.
- [3] B. Massé, S. Lathuilière, P. Mesejo, and R. Horaud, “Extended gaze following: Detecting objects in videos beyond the camera field of view,” Proc. 14th IEEE Int. Conf. Automatic Face & Gesture Recognition (FG2019), pp.1–8, Lille, France, May 2019. DOI:10.1109/FG.2019.8756555.
- [4] 小宮凜子, 齋藤剛史, 嶋田和孝, “画像処理技術を用いた対話シーンにおける注目度推定,” 第16回情報科学技術フォーラム (FIT2017), J-007, Sept. 2017.  
R. Komiya, T. Saitoh, and K. Shimada, “Image based attention estimation for interaction scene,” 16th Forum on Information Technology (FIT2017) (Japanese Edition), J-007, Sept. 2017.
- [5] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling, “Appearance-based gaze estimation in the wild,” Proc. 2015 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR2015), pp.4511–4520, Boston, MA, USA, June 2015. DOI:10.1109/CVPR.2015.7299081.
- [6] N. Funatsu, T. Takahashi, D. Deguchi, I. Ide, and H. Murase, “A study on gaze estimation using head and body pose information,” Proc. 2013 Int. Workshop on Advanced Image Technology (IWAIT2013), pp.231–235, Nagoya, Japan, Jan. 2013.
- [7] R. Stiefelhagen and J. Zhu, “Head orientation and gaze direction in meetings,” Extended Abstracts on Human Factors in Computing Systems, pp.858–859, April 2002. DOI:10.1145/506443.506634.
- [8] M. Slaney, A. Stolcke, and D. Hakkani-Tür, “The relation of eye gaze and face pose: Potential impact on speech recognition,” Proc. 16th Int. Conf. Multimodal Interaction, pp.144–147, Nov. 2014. DOI:10.1145/2663204.2663251.

- [9] Y. Kodama, Y. Kawanishi, T. Hirayama, D. Deguchi, I. Ide, H. Murase, H. Nagano, and K. Kashino, "Localizing the gaze target of a crowd of people," *Computer Vision—ACCV 2018 Workshops*, Perth, Australia, Dec. 2018, Revised selected papers, *Lect. Notes Comput. Sci.*, vol.11367, pp.15–30, June 2019. DOI:10.1007/978-3-030-21074-8\_2.
- [10] 小嶋翔太, 浮田宗伯, 萩田紀博, M.-H. Yang, "歩行者群の注視方向とシーン中の注視点の同時時系列最適化," *情処学研報*, 2015-CVIM-199(21), Oct. 2015.  
S. Kojima, N. Ukita, N. Hagita, and M.-H. Yang, "Simultaneous temporal optimization of pedestrians' gaze orientation and gaze point in a scene," *IPSI SIG Tech. Rep. (Japanese Edition)*, 2015-CVIM-199(21), Oct. 2015.
- [11] B. Massé, S. Ba, and R. Horaud, "Tracking gaze and visual focus of attention of people involved in social interaction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.40, no.11, pp.2711–2724, Nov. 2018. DOI:10.1109/TPAMI.2017.2782819.
- [12] 北川源四郎, "モンテカルロ・フィルタおよび平滑化について," *統計数理*, vol.44, no.1, pp.31–48, June 1996.  
G. Kitagawa, "On Monte Carlo filter and smoother," *Proc. Inst. Stat. Math. (Japanese Edition)*, vol.44, no.1, pp.31–48, 1996.
- [13] T. Baltrusaitis, A. Zadeh, Y.C. Lim, and L.-P. Morency, "Openface 2.0: Facial behavior analysis toolkit," *Proc. 13th IEEE Int. Conf. Automatic Face & Gesture Recognition (FG2018)*, pp.59–66, Xi'an, China, May 2018. DOI:10.1109/FG.2018.00019.
- [14] J.L. Hintze and R.D. Nelson, "Violin plots: A box plot-density trace synergism," *Am. Stat.*, vol.52, no.2, pp.181–184, May 1998. DOI:10.1080/00031305.1998.10480559.

(2022年4月20日受付, 9月16日再受付,  
2023年1月10日早期公開)



武田 一馬

令2名大・工・電気電子情報卒。令4同大大学院情報学研究科博士前期課程了。修士(情報学)。同年株式会社日立製作所入社。令2年度本会学術奨励賞受賞。



川西 康友 (正員:シニア会員)

平18京大・工・情報卒。平24同大大学院情報学研究科博士後期課程了。博士(情報学)。平24同大学術情報メディアセンター特定研究員。平26名大未来社会創造機構特任助教。平27同大情報科学研究科助教。令2同大情報学研究科講師。令3理化学研究所 ガーディアンロボットプロジェクト 感覚データ認識研究チーム チームリーダー。現在に至る。ロボットによる周囲環境認識及び、人物追跡・属性認識・行動認識などの人物画像処理に関する研究に従事。平23年度 PRMU 研究奨励賞受賞。IEEE ITS Society Nagoya Chapter Young Researcher Award 受賞。IEEE, 画像電子学会各会員。



平山 高嗣 (正員)

平10奈良高専・情報卒。平12金沢大・工・電気情報卒。平17阪大大学院基礎工学研究科博士後期課程了。博士(工学)。同年より京大大学院情報学研究科特任助教。平23より名大大学院情報科学研究科特任助教。平24より同助教。平26より同特任准教授。令3より人間環境大学人間環境学部教授。現在に至る。顔画像認識, 視行動分析, インタラクションに関する研究に従事。平21ヒューマンインタフェース学会論文賞, ACM ETRA2014 Best Short Paper Award, IEEE ISM2014 Best Paper Award, 平26電子情報通信学会情報・システムソサイエティ論文賞受賞。情報処理学会, IEEE, ACM 各会員。



出口 大輔 (正員)

平13名大・工・情報卒。平18同大大学院情報科学研究科博士後期課程了。博士(情報科学)。平16~18日本学術振興会特別研究員。平18名大大学院情報科学研究科研究員, 平18名大大学院工学研究科研究員, 平20~24同大大学院情報科学研究科助教, 平24より同大情報連携統括本部情報戦略室准教授。令2より同大大学院情報学研究科准教授。現在に至る。主に画像処理・パターン認識技術の開発とその ITS 及び医用応用に関する研究に従事。CARS2004 Poster Award, CADM2004 大会賞, 平18日本医用画像工学会奨励賞, 平18日本コンピュータ外科学会講演論文賞。情報処理学会, IEEE 各会員。



**井手 一郎** (正員：シニア会員)

平 6 東大・工・電子卒. 平 8 同大大学院工学系研究科情報工学専攻修士課程了. 平 12 同研究科電気工学専攻博士課程了. 博士 (工学). 同年国立情報学研究所助手. 平 16 名古屋大学大学院情報科学研究科助教授. 平 19 より准教授. 平 29 同大大学院情報学研究科准教授. 令 2 より同大数理・データ科学教育研究センター教授. 令 4 より同大大学院情報学研究科教授. 現在に至る. この間, 平 14~16 総合研究大学院大数物科学研究科助手併任. 平 16~22 情報・システム研究機構国立情報学研究所客員助教授・准教授兼任. 平 17, 18, 19 フランス情報学・統計システム研究所 (IRISA) 招聘教授. 平 22~23 オランダアムステルダム大情報学研究科上級訪問研究員. パターン認識技術の実応用や映像メディア処理全般に興味をもっている. 情報処理学会, IEEE 各シニア会員, 映像情報メディア学会, 人工知能学会, ACM 各会員.



**村瀬 洋** (正員：フェロー)

昭 53 名大・工・電気卒. 昭 55 同大大学院修士課程了. 同年日本電信電話公社 (現 NTT) 入社. 平 4 から 1 年間米国コロンビア大客員研究員. 平 15 から名古屋大学大学院情報科学研究科教授. 令 3 より特任教授. 現在に至る. 文字・図形認識, コンピュータビジョン, マルチメディア認識の研究に従事. 工博. 昭 60 本会学術奨励賞, 平 6 IEEE-CVPR 最優秀論文賞, 平 7 情報処理学会山下記念研究賞, 平 8 IEEE-ICRA 最優秀ビデオ賞, 平 13 高柳記念奨励賞, 平 13 本会ソサエティ論文賞, 平 14 本会業績賞, 平 15 文部科学大臣賞, 平 16 IEEE Trans. MM 論文賞, 平 22 前島密賞, 平 24 紫綬褒章, 他受賞. IEEE, 情報処理学会各フェロー.



**柏野 邦夫** (正員：フェロー)

平 2 東大工電子卒. 平 7 同大大学院電気工学専攻博士課程了. 工博. 同年 NTT に入社. 現在, 同社上席特別研究員. 国立情報学研究所客員教授を兼務. メディア情報・生体情報の探索・認識・変換・推論の研究に従事. IEEE, ACM, 情報処理学会, 日本音響学会, 人工知能学会, 各会員. 本会業績賞, 高柳健次郎業績賞, 前島賞, 科学技術分野の文部科学大臣表彰, IEEE Trans. MM 論文賞, などを受賞.