一般社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report
PRMU2021-46(2021-12)

観衆の顔向きの時空間統合による ステージ上の注目対象及び注目度の推定

武田　一馬[†]　川西　康友[†,††] 平山　高嗣[†,†††] 出口　大輔[†]　井手　一郎[†]
村瀬　洋[†]　柏野　邦夫[††††]

† 名古屋大学 〒464-8601 愛知県名古屋市千種区不老町
†† 理化学研究所 〒619-0288 京都府相楽郡精華町光台 2-2-2
††† 人間環境大学 〒444-3505 愛知県岡崎市本宿町上三本松 6-2
†††† 日本電信電話株式会社 〒243-0198 神奈川県厚木市森の里若宮 3-1
E-mail: †takedak@murase.is.i.nagoya-u.ac.jp

あらまし　講演会や音楽のライブ公演では，観衆は，ステージ上の空間に複数存在する演者やスクリーンなどの注目されうる物体（以下，注目対象）を注視する．本研究では，多数の観客で構成される観衆を撮影した映像から，観衆による注目対象に対する注目度合を推定することを目的とする．もし，注目対象の位置が既知で，また観客の視線を正確に計測できるなら，各注目対象への注目度は容易に推定できる．しかし，観衆を撮影した映像中に映った個々の観客の顔の大きさは相対的に小さいため，視線の推定は困難である．また，注目対象の位置は事前に分からない．そこで本報告では，多数の観客の低精度な顔向きの時系列データから，時系列フィルタを用いて注目対象の位置と注目度を同時に推定する手法を検討する．また，本手法を用いてシミュレーションデータに対して注目対象の位置推定及び注目度推定を行ない，精度を評価する．

キーワード　注目度推定，群衆，位置推定，時系列フィルタ

1. はじめに

講演会や音楽のライブ公演などのイベントでは，演者やスクリーン，各種演出など，ステージ上に複数の注目されうる物体（以下，注目対象）が存在することが多い．このようなイベントにおいて，観衆のうち，各々の注目対象を見ている観客の割合を注目度として定量化できれば，特定の注目対象に多くの観客の注意を集めるイベント構成などを検討できると考えられる．また，注目度をイベント中のリアルタイムな人気の指標として用いることで，ライブ公演中に注目を集めているアイドルに特別な演出を行なうなど，新たなエンタテインメントとしても活用できると考えられる．

観衆が各注目対象に注目している割合を推定する場合に最も単純な方法は，各観客の視線と注目対象の位置を対応付けることで，各注目対象を見ている観客の人数を数える方法である．観客の視線方向を推定する手法として，視線推定機器を用いた推定 [1] や，観客を近距離から撮影した顔画像を用いた推定 [2] を利用することができれば，非常に高精度な視線を得ることができる．しかし，これらの手法は撮影機材の設置などに要するコストが高く，多数の観客に対してこれらの手法を適用することは困難である．そのため，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，その領域から視線を推定することが望ましい．しかし，多数の観客を同時に撮影した映像中では顔の大きさが相対的に小さいため，視線推定精度は低くなると考えられる．

図 1: 音楽のライブ公演の例．

また，想定する状況下では，図 1 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と考えられる．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．

これらをふまえ，多数の観客を同時に撮影した映像からでもある程度推定可能な顔向きに着目する．顔向きは視線とある程度相関があると言われている [3]．そこで本報告では，注目対象の位置と各観客の視線及び見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手法を検討する．

なお，本研究では，各観客は常に 1 つの注目対象のみを見続けているものと仮定する．

2. 関 連 研 究

本研究に関連する研究として，複数の観客が映った映像から各時刻における注視領域のヒートマップを生成し，これらを入力としてその映像における観客の注目対象の位置を推定する手法 [4] が挙げられる．この研究ではエンコーダ・デコーダモデルを用いて，映像内に存在しない注目対象の位置を高精度に推定することに成功しているが，美術館の展示物など屋内に存在する静止物体を注目対象として想定しているため，本研究で想定するような注目対象の位置が動的に変化する状況への適用は困難である．

また，観衆の注目領域の推定手法として，観衆を撮影した画像から推定した各観客の視線情報から，注目対象の位置を推定する手法 [5] がある．この手法では注視対象が特定の平面上に存在することを仮定してその平面上での注視点を計算し，その分布からただ 1 つ存在すると仮定した注目対象の位置を推定している．また，低精度な視線を集約することで推定誤差による影響を軽減し，注視位置の推定精度を向上できることを示しており，多数の視線を統合することの有効性を示している．各時刻において推定を行なうことで移動する注目対象の位置を推定することが可能であるが，本研究で想定しているステージ上の空間のような，平面を仮定しない状況，複数の注目対象がある状況への直接の適用は困難である．

また，本研究の目的に最も近い手法として，観衆の頭部方向の観測値と対象物の位置情報の時系列データを用いて実際の視線方向を予測する手法 [6] がある．この研究では注目対象の位置を既知として実際の視線方向を推定することを目的としている．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る推定を行なうことで高精度な推定を可能にしているが，本研究では注目対象の位置を未知としているため直接の適用は困難である．

3. 多数の時系列顔向きデータを用いた注目度推定

3.1 提案手法の枠組み

本研究では，N 人の観客からなる観衆のうち，ある注目対象を見ている人数が n 人の時の割合 n/N を，その注目対象の注目度と定義する．これを計算するためには，各観客が見ている注目対象を知る必要があるが，本研究では複数ある注目対象の位置を未知としているため，この注目対象の位置を推定しつつ，観客が見ている注目対象を同時に推定する．

また，観衆を一度に撮影した映像を用いるため，画像中の個々の観客の顔の大きさは相対的に小さく，正確な視線を推定することは困難である．そこで，低解像度な顔画像からでも比較的推定が容易な顔向きの推定結果を入力として用いる．観衆映像に対し頭部検出器及び顔向き推定器を適用し，カメラ座標系における各観客 i の頭部位置と顔向き $\mathbf{h}_t^i$ を推定する．

本研究では時刻 t における N 人の観客の頭部位置と顔向きの集合 $\mathbf{h}_t = \{\mathbf{h}_t^1, ..., \mathbf{h}_t^N\}$ を観測値とし，M 個の注目対象の位置 $\mathbf{y}_t^j$ の集合 $\mathbf{y}_t = \{\mathbf{y}_t^1, ..., \mathbf{y}_t^M\}$ 及び，各観客がそれぞれ見ている注目対象の確率分布を表す j 次元ベクトル $\mathbf{v}^i$ の集合 $\mathbf{v} = \{\mathbf{v}^1, ..., \mathbf{v}^N\}$ を推定することを目的としたモデルを提案する．ここで，各観客はそれぞれ 1 つの注目対象を見続けていると仮定し，$\mathbf{v}$ は t によらず一定とする．以下で本モデルの枠組みについて述べる．

ある時刻 t における観衆の顔向きを $\mathbf{h}_t$ とすると，時刻 t までの観衆の顔向きの集合は $\mathbf{H}_t = \{\mathbf{h}_1, \mathbf{h}_2, ...\mathbf{h}_t\}$ と表せる．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 $\mathbf{y}_t$ 及び各観客がそれぞれどの注目対象を見ているかの確率分布 $\mathbf{v}$ を，同時確率分布における条件付期待値を，

$$\widehat{\mathbf{y}}_t = E[\mathbf{y}_t|\mathbf{H}_t] = \iint \mathbf{y}_t p(\mathbf{y}_t, \mathbf{v}|\mathbf{H}_t) \mathrm{d}\mathbf{y}_\mathrm{t} \mathrm{d}\mathbf{v} \tag{1}$$

$$\widehat{\mathbf{v}} = E[\mathbf{v}|\mathbf{H}_t] = \iint \mathbf{v} p(\mathbf{y}_t, \mathbf{v}|\mathbf{H}_t) \mathrm{d}\mathbf{y}_\mathrm{t} \mathrm{d}\mathbf{v} \tag{2}$$

として求める．なお，$\mathbf{y}_t$ と $\mathbf{v}$ は独立であるとする．ここで，Bayes の定理を用いて，

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

と置き換える．さらに，$\mathbf{y}_t$ に対して 1 重 Markov 性を仮定し，$\mathbf{y}_t$ と $\mathbf{v}$ が独立であることから，

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

となる．また，各時刻 t での注目対象の位置 $\mathbf{y}_t$ と，各観客がそれぞれどの注目対象を見ているかの確率分布 $\mathbf{v}$ から決まる真の視線方向 $\mathbf{g}_t$ を潜在変数として導入すると，

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

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，$p(\mathbf{g}_t|\mathbf{g}_{t-1}, \mathbf{y}_t, \mathbf{v})$ は視線の運動モデルを表している．また，$p(\mathbf{h}_t|\mathbf{g}_t)$ は顔向きの尤度であり，$p(\mathbf{y}_{t-1}, \mathbf{v}|\mathbf{H}_{t-1})$ は時刻 $t-1$ における事後分布である．

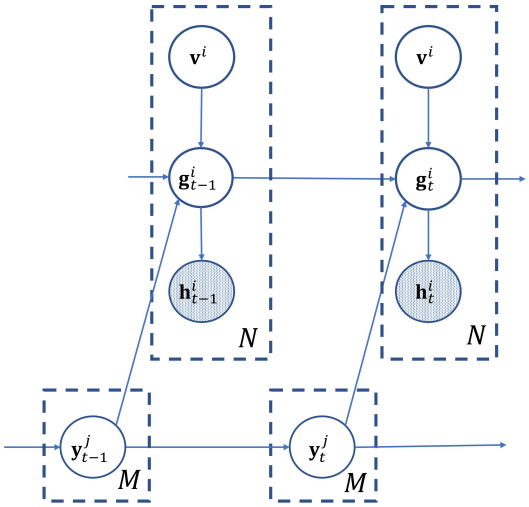


図 2: 提案手法のグラフィカルモデルによる表現.

結果として，式 (1)–(5) より，

$$\widehat{\mathbf{y}}_t = \iint \mathbf{y}_t \int p(\mathbf{h}_t|\mathbf{g}_t)p(\mathbf{g}_t|\mathbf{g}_{t-1},\mathbf{y}_t,\mathbf{v})\mathrm{d}\mathbf{g}_\mathrm{t} \int p(\mathbf{y}_t|\mathbf{y}_{t-1})p(\mathbf{y}_{t-1},\mathbf{v}|\mathbf{H}_{t-1})\mathrm{d}\mathbf{y}_{\mathrm{t}-1}\mathrm{d}\mathbf{y}_\mathrm{t}\mathrm{d}\mathbf{v} \tag{6}$$

$$\widehat{\mathbf{v}} = \iint \mathbf{v} \int p(\mathbf{h}_t|\mathbf{g}_t)p(\mathbf{g}_t|\mathbf{g}_{t-1},\mathbf{y}_t,\mathbf{v})\mathrm{d}\mathbf{g}_\mathrm{t} \int p(\mathbf{y}_t|\mathbf{y}_{t-1})p(\mathbf{y}_{t-1},\mathbf{v}|\mathbf{H}_{t-1})\mathrm{d}\mathbf{y}_{\mathrm{t}-1}\mathrm{d}\mathbf{y}_\mathrm{t}\mathrm{d}\mathbf{v} \tag{7}$$

が導出できる.

図 2 に本研究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により各分布を近似して推定する．以下で，注目対象及び視線の運動モデル，尤度，具体的な推定手法についてそれぞれ述べる.

3.2 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 ${}^{(s_1)}\mathbf{y}^j_t\,(s_1 = 1,2,...,S_1)$ を用いて近似する．注目対象の移動にランダムウォークを仮定し，

$${}^{(s_1)}\mathbf{y}^j_t = {}^{(s_1)}\mathbf{y}^j_{t-1} + \varepsilon_\mathbf{y} \tag{8}$$

$$\varepsilon_\mathbf{y} \sim \mathcal{N}(0,\sigma^2_\mathbf{y}) \tag{9}$$

として時刻 t における注目対象 j の位置 $\mathbf{y}^j_t$ 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 ごとのパーティクル ${}^{(s_2)}\mathbf{g}^i_t\,(s_2 = 1,2,...,S_2)$ を用いて近似する．まずランダムウォークを仮定し，

$${}^{(s_2)}\mathbf{g}'^i_t = {}^{(s_2)}\mathbf{g}^i_{t-1} + \varepsilon_\mathbf{g} \tag{10}$$

$$\varepsilon_\mathbf{g} \sim \mathcal{N}(0,\sigma^2_\mathbf{g}) \tag{11}$$

として，時刻 t における観客 i の視線方向 $\mathbf{g}'^i_t$ の分布を表す．ここで，各観客の視線は，注目対象の位置と，どの注目対象を見ているかにも依存する．そこで，観客 i の視線方向予測 ${}^{(s_2)}\mathbf{g}'^i_t$ と，観客 i から見た注目対象 j の方向 $r_i({}^{(s_1)}\mathbf{y}^j_t)$ との重み付き和をとり，さらに注目対象 j を見ている確率 v_{ij} を重みとして足し合わせる．具体的には，

$${}^{(s_2)}\mathbf{g}_t = \sum_{j=1}^{M} v_{ij}\left((1-\alpha)^{(s_2)}\mathbf{g}'^i_t + \alpha r_i({}^{(s_1)}\mathbf{y}^j_t)\right) \tag{12}$$

として，注目対象位置を考慮した視線方向 $\mathbf{g}_t$ の分布を得る．ただし，α は注目対象位置を考慮する度合いの重みであり，本研究では $\alpha = 0.1$ とした.

3.3 尤　　度

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，観測した顔向きの尤度 $p(\mathbf{h}_t|\mathbf{g}_t)$ を，視線方向との類似度として定義する．観客 i の顔向き $\mathbf{h}^i_t$ に対して，予測した視線方向 $\mathbf{g}^i_t$ を用いて

$$p(\mathbf{h}^i_t|\mathbf{g}^i_t) \propto l_f(\mathbf{g}^i_t;\mathbf{h}^i_t) = e^{-k\|\mathbf{g}^i_t-\mathbf{h}^i_t\|} \tag{13}$$

を計算し，各観客の顔向き $\mathbf{h}_t$ の尤度 $l_f(\mathbf{h}_t,\mathbf{g}_t)$ 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.

3.4 パーティクルフィルタを用いた注目度推定

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，3.2 節で述べた運動モデルを用いて時刻 t での値を予測する．次に，観客の視線に対するパーティクル ${}^{(s_2)}\mathbf{g}^i_t$ に対して 3.3 節で述べた尤度を計算する．さらに，注目対象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 ${}^{(s_2)}\mathbf{g}^i_t$ 及び各観客が注目対象を見ている確信度を表す $\mathbf{v}$，顔向き $\mathbf{h}^i_t$ に対する尤度 $l_f({}^{(s_2)}\mathbf{g}^i_t;\mathbf{h}^i_t)$ を用いて，

$$l_\mathbf{y}({}^{(s_1)}\mathbf{y}^j_t;\mathbf{g}_t,\mathbf{v},\mathbf{h}_t) = \frac{\sum_{i=1}^{N} l_f({}^{(s_2)}\mathbf{g}^i_t;\mathbf{h}^i_t)\cdot v_{ij}\cdot l_\mathbf{g}\left({}^{(s_2)}\mathbf{g}^i_t;r\left({}^{(s_1)}\mathbf{y}^j_t\right)\right)}{\sum_{i=1}^{N} l_f({}^{(s_2)}\mathbf{g}^i_t;\mathbf{h}^i_t)\cdot v_{ij}} \tag{14}$$

を計算することで，注視対象位置の尤度 $l_\mathbf{y}({}^{(s_1)}\mathbf{y}^j_t;\mathbf{g}_t,\mathbf{v},\mathbf{h}_t)$ を得る．ここで，$l_\mathbf{g}\left({}^{(s_2)}\mathbf{g}^i_t;r\left({}^{(s_1)}\mathbf{y}^j_t\right)\right)$ は，観客 i の視線方向 ${}^{(s_2)}\mathbf{g}^i_t$ と，観客 i がある注目対象を見ていた場合の視線方向 $r({}^{(s_1)}\mathbf{y}^j_t)$ の類似度である．各観客がその注目対象を見ているであろう確率分布 $\mathbf{v}$ と，予測した視線に対する観測値の尤度 $l_f({}^{(s_2)}\mathbf{g}^i_t;\mathbf{h}^i_t)$ 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，これらの尤度を各パーティクルの重みとしてリサンプリングを行ない，$\mathbf{y}^j_t$ 及び $\mathbf{g}^i_t$ 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 が各注目対象を見ている確率の分布 $\mathbf{v}_i$ を更新する．更新前の確率分布を $\mathbf{v}'_i$ として，観客 i に対して予測した視線の時刻 T までの時系列データ $\{\mathbf{g}^i_1,...,\mathbf{g}^i_T\}$ と，観客 i が注目対象 j を見ていた場合の時刻 T までの視線方向の時系列データ $\{r(\mathbf{y}^j_1),...,r(\mathbf{y}^j_T)\}$ を用いて，

$$v_{ij} = v'_{ij}\cdot\frac{1}{T}\sum_{t=1}^{T}\max\left(0,\eta-\left\|\mathbf{g}^\mathrm{i}_\mathrm{t}-\mathrm{r}\left(\mathbf{y}^\mathrm{j}_\mathrm{t}\right)\right\|\right) \tag{15}$$

を計算し，最後に $v_{ij}\,(j = 1,...,M)$ の総和が 1 になるよう調整を行なうことで，確率分布 $\mathbf{v}$ を更新する．一連の推定と確信度

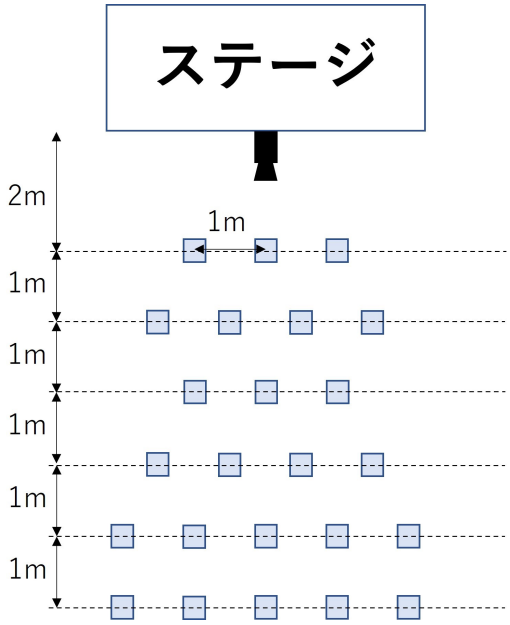


図 3: 撮影環境の模式図.

図 4: 撮影した映像の例.

の更新を規定回数に達するまで繰り返した上で，最後に全ての観客の $\mathbf{v}_i$ の平均値を算出することで，観衆全体の注目度を算出する．

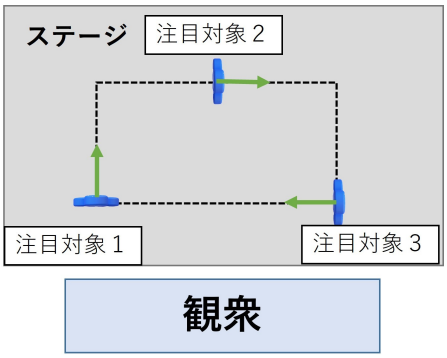
4. 実　　験

4.1 データセット

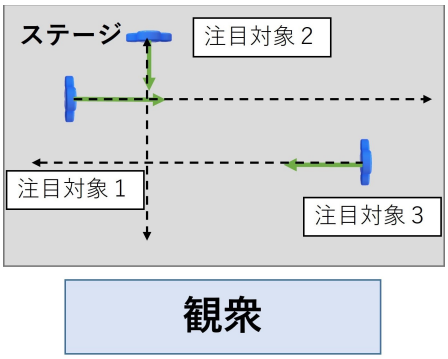
本研究では，ステージ上を移動している複数の注目対象を多数の観衆が注視している状況を想定しているが，条件を満たす公開データセットがなかったことから，実験用のデータセットを独自に撮影した．本節では，データセット構築における撮影条件及びその内容について述べる．

4.1.1 撮 影 条 件

本研究で想定する状況を再現するために，ステージ上に注目対象となる複数人の演者を，ステージ前方の観客席に 24 人の観客を配置した．演者には合図とともにあらかじめ決めた移動経路に沿って歩いてもらい，各観客にはあらかじめ指定した演者を注視し続けるよう指示した．図 3 に撮影環境の模式図を，図 4 に撮影した映像から切り出した画像の例を示す．ステージの幅は 8 m，奥行きは 5 m である．実験参加者は 48 名の男女



(a) パターン 1



(b) パターン 2

図 5: 注目対象の移動経路.

であり，24 名ずつの 2 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なった．また，実験参加者は眼鏡もしくはコンタクトレンズを着用するか，裸眼の状態であった．

4.1.2 注目対象の移動パターン及び注視方法

本研究では，ステージ上を移動する複数の注目対象への注目度推定を目的とする．そこで，移動経路による差異などを分析するため，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，各観客が注目する対象や各注目対象を見ている観客の人数の比率を変更しながら撮影を行なった．なお，注目対象の数は 3 つで固定とする．2 種類の移動経路を図 5 に示す．3 つの注目対象を見る人数の比率は 8:8:8，12:8:4，18:6:0 の 3 種類を用意した．1 つの比率に対して 2 種類の移動パターンの撮影を行なうことを 1 セットとし，3 種類の比率に対してそれぞれ 2 セットの撮影を行なうことで，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．これを 2 グループに対して 2 回ずつ行ない，計 48 本のデータを作成した．

また，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，見るべき注目対象の指示のみを与え，注目対象のどの部分を見るのかや，注目対象を見る際の顔や目の動かし方などの指示は行わなかった．更に，本研究では 1 回の撮影中においては注目対象を切り替えるなどの動作は行なわず，注目対象の移動開始から終了まで同じ注目対象を見続けるように指示した．

表 1: シミュレーションデータに対する実験結果

手法	条件	左右誤差 [m]	上下誤差 [m]	前後誤差 [m]	注目度誤差 [%]
比較手法	角度誤差加算なし	0.340	0.072	0.475	12.94
	角度誤差加算あり（$\sigma = 5^\circ$）	0.786	0.272	0.739	14.39
	角度誤差加算あり（$\sigma = 10^\circ$）	1.256	0.451	0.891	14.06
提案手法	角度誤差加算なし	0.213	0.187	0.234	1.04
	角度誤差加算あり（$\sigma = 5^\circ$）	0.346	0.193	0.357	2.18
	角度誤差加算あり（$\sigma = 10^\circ$）	1.132	0.219	0.667	8.56

4.2 実験方法

本研究では，頭部検出器及び顔向き推定器として OpenFace [7] を用い，カメラ座標系での各観客の頭部位置と顔向き方向を推定した．

実験の評価指標として，注目対象位置の推定と注目度推定のそれぞれに対して平均絶対誤差を用いる．評価はそれぞれの映像に対して行ない，各映像での結果を平均したものを最終的な実験結果とする．また，注目対象位置の推定に対しては左右・上下・奥行き方向それぞれに対して誤差を計算し，考察を行なう．

また，提案手法を様々な条件の入力データに対して検証するために，OpenFace によって推定した頭部位置と注目対象を結んだ方向を真の顔向き方向と仮定し，その方向に対して正規分布から無作為に抽出した誤差を加えたシミュレーションデータを用いて検証を行なう．なお，公演開始時における最初の立ち位置などは決まっていることが多いことから，移動経路は未知であるが，注目対象の初期位置のみ既知とし，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 $\mathbf{v}_i$ の初期値を設定し，実験を行なう．

時系列情報を用いない比較手法として，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，視線の集中度合を値としてもつボクセル空間を作成し，その値の大きさに応じて生成した点に対して混合 Gaussian モデルを用いてクラスタリングする手法を用意する．各クラスタ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，その周辺のボクセルの値を平均したものを注目度とする．

4.3 実験結果

表 1 にシミュレーションデータに対する結果を示す．入力の顔向きに対して誤差を加えない場合と，$\sigma = 5^\circ$ 及び　$\sigma = 10^\circ$ とした正規分布から無作為に抽出した誤差を加えた場合で結果を比較した．

実験結果から，シミュレーションデータに対する結果において，注視対象位置及び注目度推定結果が比較手法より優れていることが分かる．

5. 考　　察

表 1 において，注目対象位置推定の方向ごとの誤差を比較すると，入力である顔向きに加える誤差が大きくなるほど，前後方向に比べて左右方向の位置の誤差が大きいことがわかる．これは，顔向きの誤差が大きくなり注目対象位置の推定が困難になるにつれて，2 種類の移動パターンにおける，方向ごとの注目対象の移動量の差が表れるようになっているものだと考えられる．一方で，誤差が小さいときには前後方向のほうが誤差が大きいのは，観客とステージの位置関係上視線はおおむね前後方向に伸びており，その延長線上の位置推定が難しいからだと考えられる．

また，表 1 の結果を見ると，提案手法と比較手法において注目度推定の結果に大きな差異が見られる．これは，比較手法では時系列情報を用いた追跡を行なっていないため，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，実際の注目度が偏ってたとしても平均化されてしまうためだと考えられる．一方で，提案手法においては時系列データをもとに注目対象の追跡と各観客の注目対象の推定を同時に行なう．これにより，注目対象が一時的に重なっているような場合においても，より注目しているであろう注目対象に高い重みをつけて計算ができ，精度が向上すると考えられる．

なお、比較手法及び提案手法において，OpenFace により推定した顔向きを用いた実環境データに対する推定は困難であった．これは，実環境データにおける誤差が，真の顔向きを中心に正規分布に従って生成されていないことが要因であると考えられる．実環境データは，注視を行なう際の視線と顔向きの動かし方が観客ごとに異なることや，顔向き検出器の出力傾向が観客の位置や顔向きなどにより異なることで，偏った分布から生成されている．本研究で用いた較正方法では，この偏りを軽減しきれていないと考えられる．このことは，実環境データにおける誤差の標準偏差が左右方向で $\sigma = 11.8^\circ$，上下方向で $\sigma = 0.59^\circ$ であるにもかかわらず，シミュレーションデータにおいて両方向に対して $\sigma = 10^\circ$ のより大きい誤差を加えて実験した際と異なり，推定に失敗していることからも読み取ることができる．一方で，シミュレーションデータに対する結果では誤差を加えた場合でも比較的推定が可能なことから，顔向き検出器の精度向上や較正方法の改良により実環境データにおける偏りを現在よりも軽減することができれば，より高い精度で推定が可能になると考えられる．また，カメラから近い観客ほど顔向き推定の精度が向上することや，ステージを見渡すために必要な顔向きの移動範囲が広がることから，これらを考慮して推定することで，改善できる可能性があると考えられる．

6. おわりに

本報告では，観衆全体を撮影した映像を用いて，ステージ上に存在する複数の位置が未知の注目対象に対し，各注目対象の位置及び注目度の推定を行なうことを目的とした手法を検討した．観衆全体の視線を推定する際には，コストの面から観衆全体を撮影した画像から個々の観客を推定することが望ましいが，低解像度の顔画像から推定した顔向きは低精度なものとなる．また，ライブ公演などでは，常に注目対象の位置を知ることができるとは限らない．

そこで，低精度の顔向きでもそれを多数集めることで精度が高い推定が可能になるという考えと，過去の視線情報と注目対象の位置を用いることでより精度が高い視線推定が可能になるという考えに基づいて，注目対象の位置と観客の視線を同時に推定し，これらの関係を用いて注目度を推定する手法を検討した．

実験結果から，シミュレーションデータにおいて，誤差がある視線に対しても，高精度な注目対象位置及び注目度推定ができることを確認した．一方で，実環境データに対しては推定が困難であったことから，観客の位置を考慮した評価や，較正手法の見直しなど，より実環境に適した手法の検討が今後の課題である．

謝辞　本研究の一部は科研費（17H00745）による．

文　献

[1] トビー・テクノロジー株式会社, “トビー・テクノロジー, tobii pro グラス 3,” https://www.tobiipro.com/ja/product-listing/tobii-pro-glasses3/. (Accessed on Nov. 22, 2021).

[2] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling, “Appearance-based gaze estimation in the wild,” Proc. 2015 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp.4511–4520, Jun. 2015.

[3] 船津暢宏，高橋友和，出口大輔，井手一郎，村瀬洋，“体に対する顔の向きと視線方向の関係に関する予備的調査，” 2012 年電子情報通信学会総合大会講演論文集，第 2 巻，p.157，Mar. 2012.

[4] 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 Intl. Conf. on Automatic Face & Gesture Recognition (FG 2019), pp.1–8, May 2019.

[5] 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,” Proc. 2018 Asian Conf. on Computer Vision Workshops, Lecture Notes on Computer Science, vol.11367, pp.15–30, Dec. 2018.

[6] B.S. Massé Benoit and H. Radu, “Tracking gaze and visual focus of attention of people involved in social interaction,” IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.40, no.11, pp.2711–2724, Nov. 2018.

[7] T. Baltrusaitis, A. Zadeh, Y.C. Lim, and L.-P. Morency, “Openface 2.0: Facial behavior analysis toolkit,” Proc. 13th IEEE Intl. Conf. on Automatic Face & Gesture Recognition (FG2018), pp.59–66, May 2018.