

# 注視点と注視対象の動きの類似性に着目した 複数物体への注目度推定法の初期検討

武田 一馬<sup>1</sup> 川西 康友<sup>1</sup> 平山 高嗣<sup>1</sup> 出口 大輔<sup>1</sup> 井手 一郎<sup>1</sup> 村瀬 洋<sup>1</sup> 永野 秀尚<sup>2</sup> 柏野 邦夫<sup>2</sup>

概要：アイドル等の公演では、観客の注目を引き付けるため様々な演出が行なわれており、観客が公演中に注目している対象の定量的分析は演出の構成を考える上で重要な課題である。そこで本研究では、注視点と注視対象の動きの類似性を用いて観客の注視対象を判定し、注目度を推定する手法について述べる。注視対象を判定する手法として、注視点と注視対象の距離で判定する手法が考えられる。しかし、本報告では群衆を対象とするため、個々の顔画像の解像度は低く、視線推定誤差が大きい。このため、注視対象間の距離が小さい場合に判定を誤りやすいと考えられる。この問題に対し、動きの類似性を用いて判定する手法を提案する。注視対象の判定基準に動きを加えることで、注視対象間の距離によらない判定が可能になると考えられる。

## Preliminary Study on Attention Degree Estimation for Multiple Objects Based on Similarity between Gaze-Point and Gaze-Target Movements

### 1. はじめに

アイドルのライブステージなど舞台上の催し物では、観客の注目をより引き付け、楽しんでもらえるよう様々な演出が行なわれている。特に近年のデジタルメディア技術の発展に伴い、これらの演出の種類も多様化している。その際に、観客が公演中の各時刻に注目している対象を定量化できれば、多様な演出を組み合わせることで観客を引き付けるための手がかりとなることが期待される。現在、たとえばアイドルに対する注目度を図るものとして人気投票があるが、これはアイドル個人に対する人気を定量化するものであり、演出の効果を測る指標としては不十分であると考えられる。

ライブ公演では複数人の演者やモニタなど、様々な注視対象が存在する。この注視対象それぞれに対する注目度合を定量化したものを「注目度」とする。この注目度を推定するためには、それぞれの観客の注視対象を知る必要があ

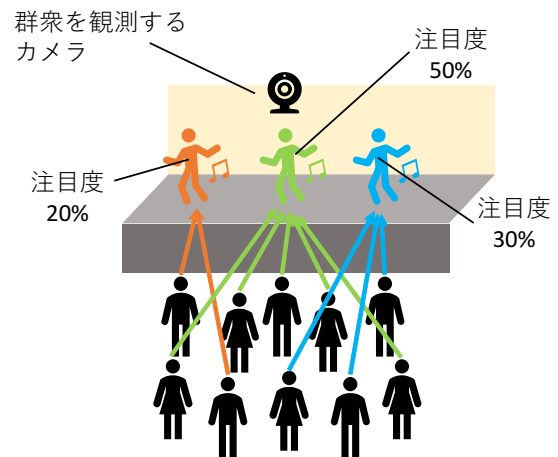


図 1: 注目度推定の概略図

Fig. 1 Overview of attention degree estimation.

る。この際に、観客全体を対象とすることが望ましいため、コストの観点から、カメラで観客全体を一度に撮影した映像から、観客の視線を推定して注視対象を知ることが望ましい。注目度推定の関連研究として、小宮らは全地球カメラによって円卓を囲む複数人の画像を撮影し、頭部姿勢を用いて注目人物を推定することで、円卓を囲む

<sup>1</sup> 名古屋大学  
Nagoya University  
<sup>2</sup> 日本電信電話株式会社 NTT コミュニケーション科学基礎研究所  
NTT Communication Science Laboratories, NTT Corporation

それぞれの人への注目度を推定する手法 [6] を提案している。視線に比べて低解像度でも頑健に推定可能な頭部姿勢を用いているため、群衆のような一度に多くの人を画像フレーム内に捉えることが必要な状況では有効であると考えられるが、注視対象が移動することは想定しておらず、公演などの注視対象の位置が頻繁に入れ替わる状況へのそのままの利用は困難である。

このような注目度推定を実現する際には、視線推定誤差の大きさが問題となる。視線推定誤差は手法によって異なり、例えば角膜反射法を用いた視線追跡装置の1つであるトビー・テクノロジー製 Tobii Pro X3-120 [4] は  $0.24^\circ$  という小さい誤差で視線を計測することができる。しかし、このような視線追跡装置は視線を計測したい人1人につき1台機器を装着する必要があるため、ライブ公演のように多数の観客がいる場合に使用することは困難である。一方、大勢の群衆の注視点を推定する手法として、群衆が映った画像から視線情報を抽出し、それらを統合することで注視点を推定する手法 [2] がある。この手法では群衆の注視点を同時に計測できるが、低解像度の顔画像から視線を推定するため、前述の視線追跡装置を用いた手法に比べて低精度である。注目度を推定するためには各人物の注視対象を判定する必要があるが、複数の注視対象が近接している場合にはこの誤差によって推定を誤りやすくなると考えられる。

この問題への対処として、注視点の位置という誤差の影響を受けやすい基準とは別の基準により注視対象を判定することが必要である。そこで本研究では、注視点と注視対象の動きの類似性を注視対象判定の基準とすることを考える。動きの類似性を用いた注視対象判定の既存研究として、岡本らがユーザの顔画像時系列を用いてその注視対象を判定する手法 [5] を提案しており、注視対象の動きとユーザの視線の時系列データの関連から注視対象を推定するため、複数の注視対象間の距離によらず注視対象を推定できるという利点がある。しかし、常に画面の正面から見ていると仮定しているため、頭部姿勢の変化には対応できず、実用のためには姿勢に関わらず推定可能である必要があると考えられる。しかし、注視対象推定に位置の誤差による影響が少ない動きの類似性を用いることの有効性を示している。

以上のことをふまえて、本研究では複数の移動する注視対象に対する観客全体の注目度を推定することを目的とする。個々の観客の注視点と各注視対象の距離に基づいてその観客の注視対象を判定し、その結果を集計することで注目度を推定する場合、複数の注視対象が移動する状況下では、先に述べたような視線推定誤差により、注視対象が近接している場合に判定を誤りやすいという問題がある。そこで、注視点と注視対象の動きの類似性を注視対象判定の基準に用いる手法を提案する。これにより注視対象の判定精度を向上させ、複数の移動する注視対象へのより高精度

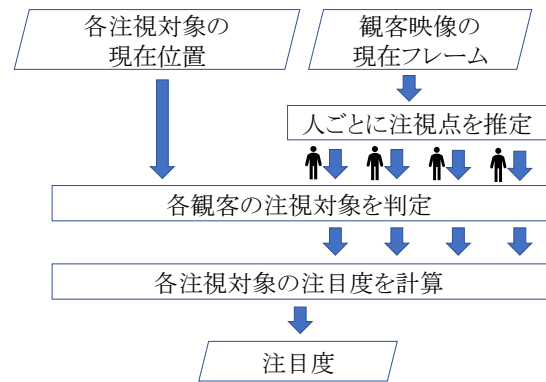


図 2: 注目度推定の基本手法の処理手順

Fig. 2 Process flow of the baseline method for attention degree estimation.

な注目度推定手法の実現を目指す。ただし、ライブ会場は多くの場合平面のステージであるため、注視対象は平面上に存在することを想定する。また、観客は複数の注視対象のいずれかを必ず注視しているものとする。

以後、2節では動きを用いない注目度推定手法を紹介し、3節で提案手法である動きを用いた注目度推定手法を紹介する。4節では実験について、5節では考察について述べ、6節で本研究のまとめを述べる。

## 2. 注視点と注視対象の位置関係に基づいた注目度推定

### 2.1 注目度推定の基本的な処理手順

まず、動きを用いず位置のみにより注視対象を判定し、注目度を推定する手法（以下、基本手法）の処理手順を図2に示す。以下で具体的な処理方法について述べる。

基本手法では、入力として観客全体を撮影した映像の現在フレーム画像及び現在フレーム中の  $M$  個の注視対象の、舞台平面上での現在位置  $\mathbf{t}_i = (x_i, y_i)^T$  ( $i \in \{1, \dots, M\}$ ) が与えられる。この現在フレーム画像に対して頭部検出器及び視線推定器を適用することで、 $j$  番目の観客の頭部座標  $\mathbf{K}^j = (X^j, Y^j, Z^j)^T$  と視線方向ベクトル  $\mathbf{L}^j = (U^j, V^j, W^j)^T$  を得る。このとき、頭部座標はカメラ位置を原点とする世界座標系で表現する。次に、注視対象は平面上に存在するという仮定の下、注視対象が存在する平面  $S$  と得られた頭部位置  $\mathbf{K}^j$  を始点とする視線方向ベクトル  $\mathbf{L}^j$  との交点から、各観客の平面上での注視点の位置  $\mathbf{g}^j = (x^j, y^j)^T$  を計算する。 $\mathbf{g}^j$  と各注視対象の座標  $\mathbf{t}_i$  を用いて、

$$e = \arg \min_i d(\mathbf{g}^j, \mathbf{t}_i) \quad (1)$$

を計算し、距離  $d(\mathbf{g}^j, \mathbf{t}_i)$  が最も近い注視対象をその観客の注視対象  $e$  として判定する。ここで、距離  $d(\mathbf{g}^j, \mathbf{t}_i)$  の計算には Euclidean 距離を用いる。最後に、

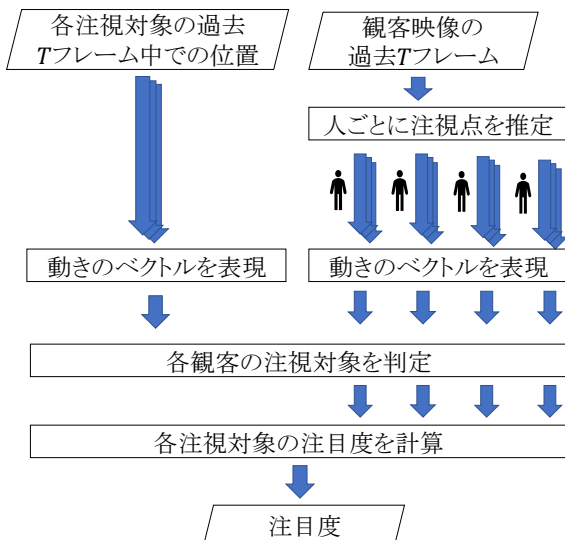


図 3: 注視点と注視対象の動きの類似性に着目した注目度推定手法の処理手順

**Fig. 3** Process flow of the attention degree estimation based on the similarity between gaze-point and gaze-target movement.

$$a_i = \frac{c_i}{C} \quad (2)$$

で表される式を用いて、各フレームから検出された顔の総数  $C$  と、そのうち  $i$  番目の注視対象を注視していると判定された人数  $c_i$  の比率を計算し、これをそのフレームにおける  $i$  番目の注視対象への注目度  $a_i$  とする。注目度の算出において、頭部検出器で誤検出された顔以外の領域除去は人手で行なう。

なお、データセット撮影のため設置したカメラの傾きや、各観客の頭部位置の推定値と実際の値の差異により、推定した視線には誤差が含まれると考えられる。本研究ではこの誤差は予め校正しておくものとする。

## 2.2 基本手法の問題点

基本手法の問題点として、1 節でも述べた通り、視線推定の誤差による影響が大きい点が挙げられる。本研究では、画像から視線を推定する視線推定器を用いることを想定しているため、装着型の視線追跡装置などに比べて誤差は大きい。また、本研究では観客全体を対象としているため、1 人あたりの顔画像の解像度は低くなり、それに伴い視線推定の誤差も拡大する。これらの誤差により、注視対象間の距離が小さい状況下では、注視対象の判定において距離の大小関係が入れ替わりやすくなり、注視対象判定精度の低下に伴って注目度推定精度も低下する。

## 3. 動きの類似性に基づく注目度推定手法

2.2 節では、注目度推定の基本手法の問題点について述べた。本報告では視線推定誤差により、注視対象が近接し

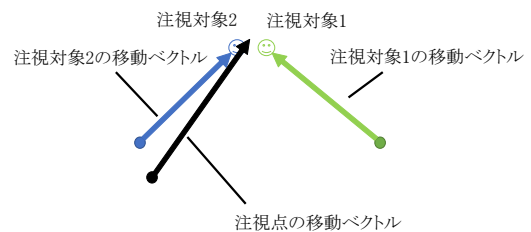


図 4: 提案手法 1: 移動ベクトルを用いた注視対象判定  
**Fig. 4** The proposed method 1: Gaze-target estimation via the similarity of motion vectors.

ている場合に注視対象の判定を誤りやすいという基本手法の問題に対し、注視点と注視対象の動きの類似性に着目し、これを注視対象判定の基準に加えることで判定精度を向上する手法を提案する。本節では入力された各注視対象の位置と各観客の注視点に対して、異なる処理を行なう 2 つの注視対象判定手法を提案する。提案手法の処理手順を図 3 に示す。1 つ目は動きを 2 フレーム間の移動ベクトルで表現した手法であり、もう 1 つは点列からなるベクトルで表現した手法である。

### 3.1 提案手法 1: 2 フレーム間の移動ベクトルによる注目度推定手法

まず、注視対象の判定基準として 2 フレーム間の注視点の位置変化から計算した移動ベクトルを用いる手法を提案する。本手法では、図 4 に示すように現在の注視点と  $\tau$  フレーム過去における注視点の間の変化を求めると移動ベクトルを計算し、注視点と注視対象の絶対位置に加えて移動ベクトルを用いて注視点と注視対象間の距離を計算し、その観客の注視対象を判定する。過去フレームからの移動傾向を判断基準に加えることで、注視対象間の距離が小さくとも区別できるようになると考えられる。

基本手法と同様に推定した視線から算出した  $j$  番目の注視点座標  $\mathbf{g}^j$  を、現在より  $\tau$  フレーム過去における注視点座標  $\mathbf{g}^{\tau j}$  と比較することにより、注視点の移動ベクトル  $\mathbf{v}^j = (u^j, v^j)^T = \mathbf{g}^j - \mathbf{g}^{\tau j}$  を算出する。比較するためには現在のフレームと同一人物を過去のフレームにおいても検出する必要があるが、本研究ではこれらを顔追跡により対応付ける。

同様に、 $i$  番目の注視対象に対しても現在のフレーム中の座標  $\mathbf{t}_i = (x_i, y_i)^T$  と  $\tau$  フレーム過去のフレーム中の座標  $\mathbf{t}_i' = (x_i', y_i')^T$  を比較することにより、注視対象の移動ベクトル  $\mathbf{v}_i = \mathbf{t}_i - \mathbf{t}_i' = (u_i, v_i)^T$  を求める。現在のフレームにおける注視点または注視対象の座標  $(x, y)^T$  と、このようにして求めた移動ベクトル  $(u, v)^T$  を統合したものを、注視点および注視対象の特徴ベクトル  $\mathbf{f} = (x, y, u, v)^T$  とする。基本手法においては  $j$  番目の人物の注視点  $\mathbf{g}^j$  と

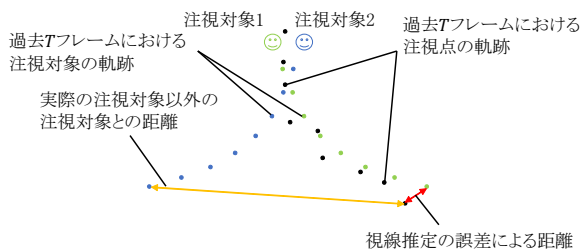


図 5: 提案手法 2 : 軌跡を用いた注視対象判定

Fig. 5 The proposal method 2 : Gaze-target estimation via the similarity of trajectories.

$i$  番目の注視対象の位置  $\mathbf{t}_i$  との距離  $d(\mathbf{g}^j, \mathbf{t}_i)$  を用いて注視対象を判定するのに対し, 提案手法では注視点の特徴ベクトル  $\mathbf{f}^j = (x^j, y^j, u^j, v^j)^T$  と注視対象の特徴ベクトル  $\mathbf{f}_i = (x_i, y_i, u_i, v_i)^T$  の距離  $d_v(\mathbf{f}^j, \mathbf{f}_i)$  を用いてその観客の注視対象を判定する. 最後に, 基本手法と同様の手法で, 各注視対象の注目度を算出する.

### 3.2 提案手法 2 : 点列からなる軌跡ベクトルによる注目度推定手法

注視対象の判定基準として注視点と注視対象の移動の軌跡の類似性に着目する手法を提案する. 2.2 節では視線推定の誤差により注視対象が近接している場合に注視対象判定を誤りやすいと述べたが, 注視対象間の距離が誤差に対して十分大きければ, 誤差を含んだ推定値を用いても注視対象を正しく判定できると考えられる. そこで図 5 に示すように, 過去  $\tau$  フレーム間の注視点と各注視対象の軌跡を用いて各観客の注視対象を判定することで, 現在フレームのみでは注視対象が近接していて区別が困難であっても, 過去  $\tau$  フレーム中において距離が大きければ, 注視対象の区別が可能になると考えられる. これにより視線推定誤差の影響を少なくし, 注視対象の判定精度向上を図る.

基本手法と同様に推定した視線から算出した  $j$  番目の観客の注視点座標を  $\mathbf{g}^j = (x_0, y_0)^T$  とし, これに加えて過去  $\tau$  フレーム間における同一人物の注視点  $(x_t, y_t)^T$  ( $t \in \{1, \dots, \tau\}$ ) を用いて, 注視点の軌跡を表す軌跡ベクトル  $\ell^j = (x_0, y_0, x_1, y_1, x_2, y_2, \dots, x_\tau, y_\tau)^T$  を生成する. 同一人物は, 3.1 節で述べた手法と同様に顔追跡により対応付ける.  $i$  番目の注視対象に対しても同様に行なって注視対象の軌跡を表す軌跡ベクトル  $\ell_i$  を生成し,  $\ell^j$  と  $\ell_i$  の平均 2 乗誤差を用いてその観客の注視対象を判定する. 最後に, これまでの手法と同様に, 各注視対象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, 大勢の観客の映像を撮影し, 各観客の注視対象を真値として付与した公開データセットがないこと



図 6: 撮影した映像中の観客画像の例

Fig. 6 Example of audience image in the video.

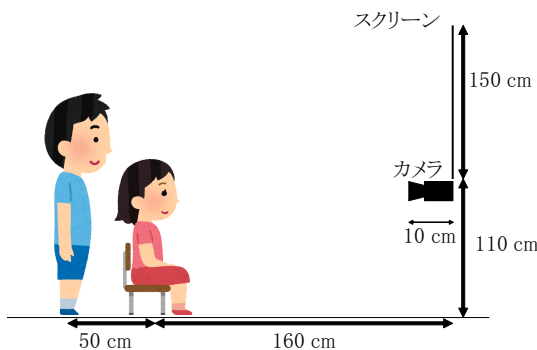


図 7: 撮影環境の側面図

Fig. 7 Side view of the shooting environment.

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を構築した際の撮影条件及びその内容について述べる.

#### 4.1.1 撮影条件

本研究では, 個々の観客が複数の視標のうちいずれかを注視している状況を想定する. ここでは, 問題を簡単にするために, 複数の実験参加者に鉛直平面なスクリーン上の 2 つの視標のうちいずれかを注視させ, その様子を撮影した. 撮影した映像から切り出した画像の例を図 6 に示す. また, 図 7 に撮影環境の側面図を示す. スクリーンの幅は 200 cm, 高さは 150 cm であり, 実験参加者は 8 名 (男性 7 名, 女性 1 名) であった. 各実験参加者は前列と後列に 4 名ずつ, 前列では椅子に着席し, 後列では立った状態で視標を注視した. また, 実験参加者は眼鏡を着用せず, 裸眼もしくはコンタクトレンズを着用して注視を行なった.

#### 4.1.2 注視対象物体の詳細及び移動パターン

本節では注視対象物体として選んだ視標について説明する. 1 節で述べたように, 本研究では公演中のアイドルなどへの注目度推定を目的としている. そのため, 本実験では CelebA データセット [3] から画素数  $72 \times 86$  の 2 枚の顔画像を選び, 図 8 のように背景領域を除去したものを視標として用いた.

本研究では移動する視標への注目度を推定することを目的とするため, 図 9 に示す 5 つの移動パターンを用意した. 表 1 に移動パターンの概要を示す. なお, どのパターンに



図 8: 視標とした顔画像 (CelebA データセット [3] の画像を加工)

Fig. 8 The target face images (From CelebA dataset[3]).

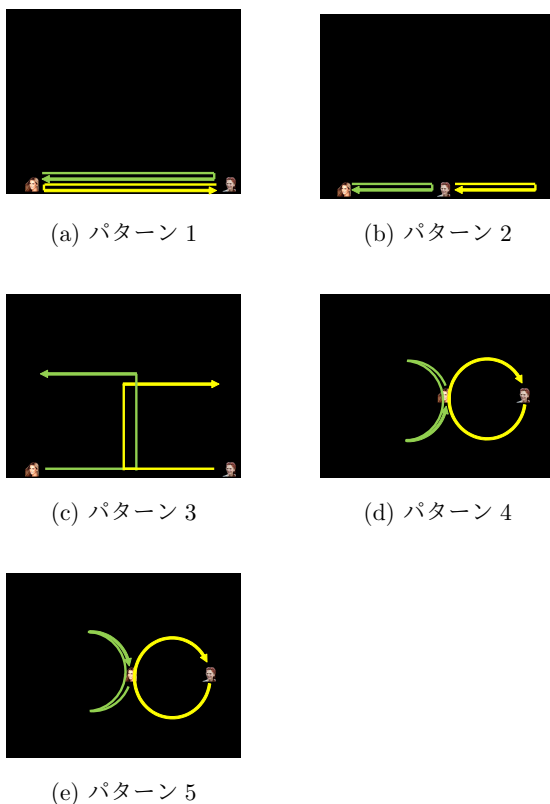


図 9: 視標の移動パターン

Fig. 9 Movement patterns of the target images.

表 1: 移動パターンと概要

Table 1 Descriptions of the movement patterns.

パターン	概要
パターン 1	左右方向に等速で移動し中央で 2 回交差
パターン 2	左右方向に等速で平行移動
パターン 3	左右上下方向に移動し加減速や停止を含む
パターン 4	等速で移動し中央で逆方向に移動しながら交差
パターン 5	等速で移動し中央で同方向に移動しながら交差

においても交差時には図 8 の視標 2 が前面に配置される。

### 4.1.3 撮影手順

撮影の際は、4.1.2 節で紹介した 5 つのパターン各 1 回の撮影を 1 セットとし、これを 5 セット録画した。以下にその手順を示す。

進行役は 2 つの視標のうちどちらを注視するかを記したカードを各実験参加者に配り、実験参加者は指示された視標を一連の撮影で注視し続けるものとした。この時、進行役は注視すべき視標のみを指示し、その他の指示はせず、実験参加者は各自で自由に注視した。一連の撮影において各パターン間に休憩を挟み、各パターンの撮影開始時には進行役により合図を行なった。各セットで各実験参加者が注視する視標を無作為に変えて撮影を行なった。また、実験参加者の並び順は 2 セット目の撮影と 3 セット目の撮影の間に無作為に変更した。それぞれの視標に対する注視人数の割合は、1 セット目から順に 4 : 4, 6 : 2, 2 : 6, 3 : 5, 5 : 3 とした。

以上のタスクを 8 名の実験参加者に課して撮影を行ない、データセットを構築した。なお、全実験参加者が本撮影の実施及び撮影データの研究目的の使用及び公開に同意した。

## 4.2 実験の概要

### 4.2.1 実験方法

本実験では頭部姿勢推定器及び視線推定器として OpenFace [1] を用い、これにより世界座標系での頭部位置と視線方向を推定した。

評価実験の評価指標として、注視対象判定精度に対しては式 (3) に示す正解率を、 $i$  番目の視標に対する注目度推定に対しては式 (4) に示す平均絶対誤差 (Mean Absolute Error ; MAE) を用いる。精度の計算は各映像に対して行ない、パターンごとに 5 つの映像に対する結果の平均をとったものを最終的な実験結果とする。

$$\text{Acc} = \frac{c_c}{c_T} \quad (3)$$

$$\text{MAE} = \frac{1}{T} \sum_{k=1}^T |a_i^k - \hat{a}_i^k| \quad (4)$$

なお、正解率において  $c_T$  は全フレームから検出された観客の数のべ数を表し、 $c_c$  はそのうち正しく注視対象を判定できた観客の数のべ数を表す。また、平均絶対誤差において  $T$  は映像の総フレーム数を表し、 $a_i^k$  は  $k$  フレーム目における観客の総数と実際に  $i$  番目の視標に注視していた観客数の比率から求めた注目度の真値を表し、 $\hat{a}_i^k$  は  $k$  フレーム目における  $i$  番目の視標への注目度推定の結果を表す。また、注目度を推定する際に比較フレームが存在しない場合がある冒頭部のデータは無視し、60 フレーム目以降について評価する。

視線推定は、設置したカメラの微妙な傾きや、観客との

位置関係の差により影響を受けるため、推定値には偏りを含んだ誤差が生じる。これを補正するため、本実験では求めた注視点に対して平面射影変換を行なう。補正では、パターン1, 2, 4, 5に対してはパターン3の映像を、パターン3に対してはパターン4の映像を用いた。

実験では位置のみで注視対象を判定する手法と相関係数を用いる手法[5]を比較手法とし、本報告で提案する手法のうち、3.1節で提案した動きを2フレーム間の移動ベクトルとして表現する手法を提案手法1、3.2節で提案した軌跡中の点列のベクトルとして表現する手法を提案手法2とする。補正した注視点に対して比較手法及び提案手法を適用し、注視対象判定の正解率及び注目度推定の平均絶対誤差を求める。

#### 4.2.2 実験結果

表2に注視対象判定手法ごとの正解率を、表3に注目度推定手法ごとの平均絶対誤差を示す。提案手法1では比較対象とするフレームを15フレーム前及び60フレーム前と変化させたときの結果を、提案手法2では用いる軌跡の区間を過去15フレーム、及び60フレームと変化させたときの結果を示した。実験結果から、動きの類似性に着目した提案手法を用いることで、どのパターンにおいても注視対象判定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2に示した注視対象判定精度において、位置情報のみを用いる手法に比べて、提案手法1及び提案手法2のどちらにおいても全体的に精度が向上した。これは動きの類似性による注視対象判定が有効であることを示していると考えられる。また、表3に示した注目度推定誤差と表2に示す注視対象判定精度を比較すると、注視対象判定精度の向上に伴って注目度推定誤差が小さくなっているのが確認できる。このことから、注視対象判定精度の向上が注目度推定精度の向上にとって重要であると考えられる。ただし、移動パターンごとに結果を確認すると、移動ベクトルを用いる場合はパターン2において精度の向上がないことがわかる。これは、パターン2においては視標の動きが完全に同一であり、距離計算において移動ベクトルに関する距離が同一になり、結果として位置のみで判定する場合と同等になったためであると考えられる。これらの結果から、類似した動きが多い状況では提案手法2を、そうでなければ提案手法1を用いるなど、視標の動き方に応じて判定方法を変更することで、さらなる精度向上が期待できる。

また、図10に、提案手法1における比較フレーム及び提案手法2における軌跡長を変化させたときの正解率の変化を示す。横軸 $\tau$ は提案手法1においては比較フレームが $\tau$ フレーム過去であることを示し、提案手法2においては過去 $\tau$ フレーム間の軌跡を用いることを示している。提案

手法2では $\tau$ を大きくするほど、どの移動パターンに対しても精度が向上している。提案手法2では軌跡の途中で視標間の距離が大きくなる区間では、距離計算を行なう際に2乗誤差が大きくなり正しい判定が行なわれやすい。用いる軌跡を長くするとこの区間が含まれやすくなるため、精度が向上すると考えられる。一方、移動ベクトルを用いる提案手法1では比較フレームから現在にかけての視標間の移動ベクトルの差が大きいかほど正しい判定が行なわれやすい。そのため、図10に示すように移動ベクトルの計算における比較対象フレームを現在フレームから離していくと、経路の重複がなく比較フレームにおける位置からの移動量が平均して大きいパターン3では精度が向上した。反対に、パターン1では一度正解率が向上したのちに低下した。これは、移動ベクトルは中間の経路を考慮せず現在フレームと比較フレームにおける注視点位置の差のみを表しているため、折り返し等経路の重複がある動きについては、現在フレームと比較フレームで注視点位置が同じになってしまい動きを表現できなくなるためと考えられる。

### 6. おわりに

本報告では、移動する複数の注視対象のいずれかを観客全体が注視している状況において、注視点と注視対象の動きの類似性を用いて各観客の注視対象を判定し、これらを集計することで注視対象ごとの注目度を推定する手法を提案した。観客全体の視線を一度に推定するためには、コストや手間の面から、カメラで観客全体をまとめて撮影した映像から推定するのが望ましい。この場合、広範囲を一度に撮影するために各観客の顔画像の解像度は低くなり、視線推定誤差が増大する。注視対象を注視点と各注視対象の絶対位置のみで判定する場合、注視対象間の距離が小さいと、視線推定誤差により各観客の注視対象の判定を誤りやすくなる。

一方、注視点と各注視対象の動きの類似性を注視対象の判定に用いることで、注視対象が近接し、絶対位置では区別が難しい状況下でも、動きの差異により注視対象を区別できる可能性がある。そこで本研究では、動きを過去のあるフレームから現在のフレームへの移動ベクトルとして、もしくは過去のあるフレームから現在のフレームまでの軌跡を示す点列として表現し、これを注視対象判定に用いることで、高精度な注目度推定を行なう手法を提案した。

実験結果から、動きの類似性を用いることで判定精度及び推定精度が向上することを確認した。これにより、提案手法の有効性を確認した。今後の課題として、注視をしていない観客への対応や、注視対象の切り替わりへの対応が挙げられる。

謝辞 本研究の一部は科研費(17H00745)による。

表 2: 注視対象判定の正解率 [%]  
**Table 2** Accuracy[%] of the gaze-target estimation.

手法	パターン 1	パターン 2	パターン 3	パターン 4	パターン 5	平均	
比較手法	位置のみ	86.00	94.72	83.20	81.47	83.66	85.81
	15 フレーム間の相関係数 [5]	63.06	48.86	57.52	60.18	51.00	56.12
	60 フレーム間の相関係数 [5]	79.16	49.18	67.32	76.65	56.25	65.71
提案手法 1	15 フレーム前からの移動ベクトル	92.09	94.72	88.61	83.76	84.13	88.66
	60 フレーム前からの移動ベクトル	<b>99.27</b>	94.72	<b>95.38</b>	<b>93.48</b>	<b>90.25</b>	<b>94.62</b>
提案手法 2	過去 15 フレーム	87.06	95.71	84.15	81.99	83.91	86.57
	過去 60 フレーム	94.41	<b>97.66</b>	90.19	86.63	85.02	90.78

表 3: 推定した注目度 [%] の平均絶対誤差  
**Table 3** Mean Absolute Error[%] of the estimated attention degrees.

手法	パターン 1	パターン 2	パターン 3	パターン 4	パターン 5	平均	
比較手法	位置のみ	5.87	5.00	11.70	16.41	10.79	9.96
	15 フレーム間の相関係数 [5]	21.12	29.41	22.76	21.69	23.04	23.60
	60 フレーム間の相関係数 [5]	16.77	30.10	20.16	17.82	22.52	21.47
提案手法 1	15 フレーム前からの移動ベクトル	4.51	5.00	9.44	14.75	10.23	8.79
	60 フレーム前からの移動ベクトル	<b>0.73</b>	5.00	<b>4.25</b>	<b>6.28</b>	<b>7.27</b>	<b>4.71</b>
提案手法 2	過去 15 フレーム間の軌跡	5.72	3.88	11.50	16.19	10.76	9.61
	過去 60 フレーム間の軌跡	3.39	<b>2.34</b>	8.36	12.79	10.20	7.41

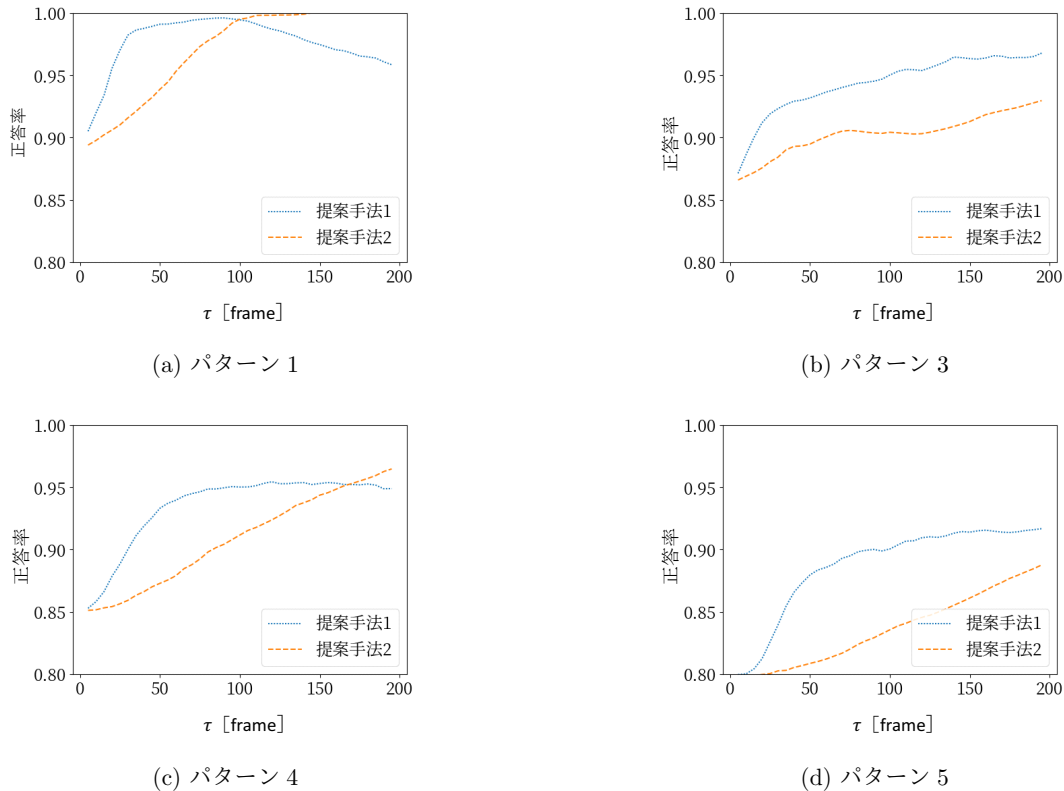


図 10: 過去フレーム区間  $\tau$  を変化させたときの精度の推移

**Fig. 10** Transitions of accuracy against the length of the past frame interval  $\tau$ .

参考文献

- [1] Baltrusaitis, T., Zadeh, A., Lim, Y. C. and Morency, L.-P.: Openface 2.0: Facial behavior analysis toolkit, *Proc. 2018 IEEE Int. Conf. on Automatic Face & Gesture Recognition*, pp. 59–66 (2018).
- [2] Kodama, Y., Kawanishi, Y., Hirayama, T., Deguchi, D., Ide, I., Murase, H., Nagano, H. and Kashino, K.: 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 (2018).
- [3] Liu, Z., Luo, P., Wang, X. and Tang, X.: Deep learning face attributes in the wild, *Proc. 2015 IEEE Int. Conf. on Computer Vision*, pp. 3730–3738 (2015).
- [4] トビー・テクノロジー株式会社: トビー・テクノロジー, Tobii Pro X3-120, 入手先 (<https://www.tobii.com/ja/product-listing/tobii-pro-x3-120/>).
- [5] 岡本宇正, 藤江真也, 小林哲則: 視線方向とオブジェクトの動きの関係に基づいた注目オブジェクト推定, 第9回画像の認識・理解シンポジウム, no. OS1B-1 (2006).
- [6] 小宮凜子, 齊藤剛史, 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, no. J-007 (2017).