

スポーツフィールドにおける自己校正を使った観客の注視領域推定[†]

阿部和広, 高山侑也, 大坪洋介, 小池哲也

Estimating Attention Area of Spectator using Self-calibration in Sports Fields

Kazuhiro ABE, Yuuya TAKAYAMA, Yosuke OTSUBO and Tetsuya KOIKE

スポーツ映像ソリューションにおいて、観戦者の注視領域を推定することは重要なシーンを抽出する上で有用である。注視領域は観客の視線の先のコート上の領域として特定される。スポーツフィールドの周囲に配置された観客を撮影するカメラを使って注視領域推定をするためには、視線ベクトルをカメラ座標系からコート座標系に変換するためにカメラ校正が必要である。一方で、観客をズームアップで写す観客カメラは校正ターゲットとなるコートなどのマーカーが映っていないため校正が難しい。そこで、本研究では、観客を撮影するカメラと俯瞰撮影するカメラの組み合わせから自己校正して注視領域を推定する手法を提案する。さらに、観客カメラ画像是低解像度であり一人ひとりの正確な視線推定は難しいため、複数の頭部方向を合計することにより最終的な注視領域を算出した。実際に行われた3×3バスケットボールでの撮影映像での検証により、リーズナブルな注視領域推定結果が得られることを示す。

In sports video solutions, estimating the attention area of spectators is useful for extracting important scenes. The attention area is identified as the area on the court corresponding to the gaze direction based on the spectator's gaze. To do this from cameras installed around the sports field that capture the spectators, camera calibration is required to convert the coordinate system of the camera that shoots the spectators and the coordinate system of the court. However, it is difficult to apply the general camera calibration to the camera capturing spectators because the court is not reflected in the camera. Therefore, in this study, we propose a method to estimate the attention area by self-calibrating the combination of the camera capturing the spectators and the camera capturing the overhead view. Since it is difficult to estimate the gaze accurately from low-resolution images of spectators, we approximate the gaze direction as the head direction and reduce the error by aggregating multiple head directions. Verification using the shooting data of an actual 3 x 3 basketball game shows that a reasonable attention area map can be obtained based on three camera inputs.

Key words コンピュータービジョン, スポーツ映像分析, 視線推定
computer vision, sports video analysis, gaze estimation

1 はじめに

近年、スポーツ映像ソリューションにおいて、機械学習や画像認識の応用が進んでいる。特に、スポーツ映像撮影の自動化において有用であり、製品化も進んでいる [10]。バスケットボールやサッカーなどの球技スポーツの映像撮影を自動化する試みにおいて、「スポーツフィールド（コート）上のどの領域を撮影するか？」は大きな課題である。このために、まずボールの位置や選手の位置に基づいて撮影領域を決定することが考えられる。しかし、ボールは動きも速く、大きさやオクルージョンの点で検知が非常に難しく、また、選手の位置検知・動作認識もそれ自体が難し

いため、そこから撮影領域を決定するには課題が多い。そこで、観客の注視領域に基づいて撮影領域を決定するアプローチを考える。これは、映像を見る人が「どこが見たいか？」を直接認識するアプローチであり、フィールド上の選手やボールに依存しない解析であることから、特定のスポーツに特化しない汎用的な方法になり得る。また、観客の注視領域推定は、観客の嗜好性を反映していると考えられ、ハイライトシーン生成 [11]、カメラコントロール [12]、仮想カメラシステム [13] などへの応用や、エンターテインメント（音楽ライブ、舞台演劇）などのスポーツ以外の幅広い適用シーンがある。さらに、注視領域は対象に関する注意や関心の度合いと密接に関係していることか

[†] 本稿は、著者の引用文献 [19] の人工知能学会より許諾を得ている。

ら、注視領域推定それ自体が広告の効果測定 [14]、教育の集中度合い計測や視線操作 [15] などにも応用できる。このような背景から、本研究では観客の注視領域推定の方法を提案する。

注視領域の推定は、観客の視線方向を推定し、観客の視点から視線方向に投影したコート上の領域を推定することで達成できる。そのために、まずは、観客の視線推定について考える。スポーツを観戦している観客は複数人が集合しており、多くの観客が1画角に捉えられているため、一人ひとりの顔が低解像度で撮影されている状況で視線推定できることが望ましい。従来の視線推定は、ターゲットに装着する専用のデバイスを必要とし、スポーツを観戦する観客に装着させるのは大きな負担となり現実的ではない。他方、近年、画像から視線推定を行う手法も研究されている [16] が、高解像度の顔画像で瞳領域を必要とするため、本研究で対象とする観客画像には適していない。そこで本研究では、低解像度顔画像からでも推定が可能な頭部方向推定を用いることで、視線方向を頭部方向と近似することとした。一方で、観客一人ひとりの頭部方向は、観客の集中度合いなどによってばらつくため、複数人の観客の合計をとることで頑健な注視領域推定を実現した。

次に、視線方向のフィールド上への投影について考える。観客を撮影するカメラ (= 観客カメラ) から推定した視線方向は観客カメラの座標系で推定されたものであるため、コート上に投影して注視領域を得るためには、視線ベクトルをコート座標系に変換する必要がある。この変換は、観客カメラのカメラ校正で実現できる。カメラ校正なしで注視領域を直接学習する手法も提案されている [17] が、全員が同じ対象を注視しているという限定的な仮定に基づいており、カメラ校正による手法の重要性は依然として高い。コートを俯瞰する画角を持つカメラ (= 俯瞰撮影カメラ) を用いる場合は、コートを参照物体とし、画像内のコートと参照コートとの対応からホモグラフィーを推定し、カメラ校正を行うことができる。一方、観客カメラではコートが映っていないため、この手法を使うことができない。そこで、コート外の物体の対応を用いて、観客カメラと俯瞰撮影カメラとの自己校正を行った。

以上から、本研究では、スポーツフィールドにおいて、1台の観客カメラと2台の俯瞰撮影カメラの組み合わせから自己校正を行い、複数観客の注視領域の合計をとることによって、スポーツフィールド上の注視領域を推定する手法を提案する。そして、実際の3×3バスケットボール試合の撮影データを用いた検証によって、3台のカメラ入力をもとに合理的な時空間の注視領域マップを得ることができることを示した。

2 手法

本研究で提案する注視領域推定は、カメラ校正、視線推定、視線投影からなる。本手法では、観客をズームで撮影する「観客カメラ」1台と、コート全体を撮影する「俯瞰カメラ」2台を用いる。この構成を、Fig. 1 に示す。また、全体のアルゴリズムフローを Fig. 2 に示す。

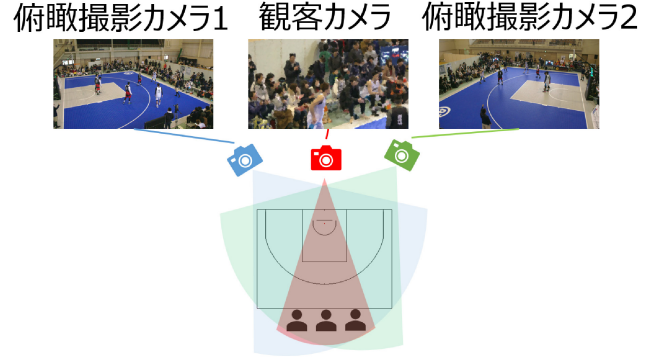


Fig. 1 カメラの構成

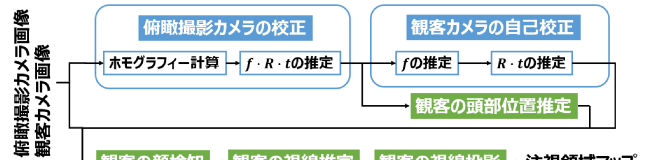


Fig. 2 注視領域推定のアルゴリズムフロー

2.1. カメラ校正

カメラ校正では、コート座標系における、内部パラメーターである校正行列 K と外部パラメーターである回転 R 、位置 t を推定する。ここで、アスペクト比を1、スキュー歪を0、レンズ歪みが無く、principal point を画像中心 (x_0, y_0) と仮定すると、内部パラメーターのうち推定すべきパラメーターは焦点距離 f のみとなり、校正行列 K は

$$K = \begin{bmatrix} f & 0 & x_0 \\ 0 & f & y_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (1)$$

となる。

2.1.1. 俯瞰カメラの校正

俯瞰カメラは、コートが映っているため、4点の対応から参照コートとのホモグラフィー H が計算できる。このとき、ホモグラフィー $H = [h_1 \ h_2 \ h_3]$ 、校正行列 K 、回転 $R = [r_1 \ r_2 \ r_3]$ と位置 t の関係は、

$$[h_1 \ h_2 \ h_3] = \lambda K [r_1 \ r_2 \ t] \quad (2)$$

である [1]。また、 r_1, r_2 が正規直交であることから、

$$h_1^\top K^{-\top} K^{-1} h_2 = 0 \quad (3)$$

$$\mathbf{h}_1^\top \mathbf{K}^{-\top} \mathbf{K}^{-1} \mathbf{h}_1 = \lambda^2 \quad (4)$$

$$\mathbf{h}_2^\top \mathbf{K}^{-\top} \mathbf{K}^{-1} \mathbf{h}_2 = \lambda^2 \quad (5)$$

が導かれる。式 (2) より、焦点距離 f は、

$$f = \sqrt{\frac{(\mathbf{h}_{11} - x_0 \mathbf{h}_{13})(\mathbf{h}_{21} - x_0 \mathbf{h}_{23}) + (\mathbf{h}_{12} - y_0 \mathbf{h}_{13})(\mathbf{h}_{22} - y_0 \mathbf{h}_{23})}{\mathbf{h}_{23} \mathbf{h}_{13}}} \quad (6)$$

と求めることができる。ここで、 $[\mathbf{h}_{11} \ \mathbf{h}_{12} \ \mathbf{h}_{13}]^\top = \mathbf{h}_1$ 、 $[\mathbf{h}_{21} \ \mathbf{h}_{22} \ \mathbf{h}_{23}]^\top = \mathbf{h}_2$ 、である。回転 $R = [\mathbf{r}_1 \ \mathbf{r}_2 \ \mathbf{r}_3]$ と位置 \mathbf{t} は、式 (2) - (5) より、

$$\mathbf{r}_1 = \lambda^{-1} \mathbf{K}^{-1} \mathbf{h}_1 \quad (7)$$

$$\mathbf{r}_2 = \lambda^{-1} \mathbf{K}^{-1} \mathbf{h}_2 \quad (8)$$

$$\mathbf{r}_3 = \mathbf{r}_1 \times \mathbf{r}_2 \quad (9)$$

$$\mathbf{t} = \lambda^{-1} \mathbf{K}^{-1} \mathbf{h}_3 \quad (10)$$

$$\lambda = \|\mathbf{K}^{-1} \mathbf{h}_1\| = \|\mathbf{K}^{-1} \mathbf{h}_2\| \quad (11)$$

と求めることができる。

2.1.2. 観客カメラの自己校正

観客カメラは、コートが映っていないため、参照コートとの対応点を特定できず、直接校正することができない。そこで、観客カメラと一部画角がオーバーラップしている校正済み俯瞰撮影カメラを使い、俯瞰撮影カメラ画像と観客カメラ画像との対応点から、自己校正する。まず、5点以上の対応点から、5点法 [2] により、Essential matrix を求める。次に、特異値分解により R と \mathbf{t} を復元する。このとき、可能な解が4通りあるが、chirality constraint によって1通りに決定する [2]。ここで、 \mathbf{t} はスケール倍の不定性がある。

焦点距離 f は、撮影時の機材情報から得ることもできるが、映像編集過程でのメタデータの消失、撮影中の調整による変更、などの理由から必ずしも利用できない。そこで、本研究では、画像情報から校正する手法を4通り試した。画像対応点のみから自己校正する手法として、2カメラ（俯瞰撮影カメラ1、2のどちらかと観客カメラ）を用いる① Fundamental matrix から求める手法と、3カメラ（俯瞰撮影カメラ1、2と観客カメラ）を用い②角度を用いる手法を試した。一方で、これらの方法は実用的なシナリオでは適用が難しいことが知られているため、③シーンの知識を利用する方法と④頭部の大きさを使ったヒューリスティックによる手法を試した。

① Fundamental matrix から求める方法

8点以上の対応点から8点法により Fundamental matrix を求め、それを使って f を求める方法である。本研究では、Bougnoux の手法 [9] と Hartley の手法 [3] を試した。

Bougnoux の手法によれば、俯瞰撮影カメラの焦点距離を f' とし、観客カメラの焦点距離 f は

$$f = \sqrt{\frac{\mathbf{p}'^\top [\mathbf{e}']_x \tilde{\mathbf{I}} \mathbf{F} \mathbf{p} \mathbf{p}^\top \mathbf{F}^\top \mathbf{p}'}{\mathbf{p}'^\top [\mathbf{e}']_x \tilde{\mathbf{I}} \tilde{\mathbf{I}} \mathbf{F}^\top \mathbf{p}'}} \quad (12)$$

と求められる。ここで、 \mathbf{p}, \mathbf{p}' はそれぞれ観客カメラ、俯瞰撮影カメラの principal point、 \mathbf{e}' は俯瞰撮影カメラのエピポール、 $\tilde{\mathbf{I}} = \text{diag}(1, 1, 0)$ である。Hartley の手法は、Kruppa 方程式を解くことにより f を決定する。ただし、これらの方法は実用的なシナリオでは適用が難しいことは注意が必要である [4]。

②角度を利用する方法

コート座標における2点 $\mathbf{X}_1, \mathbf{X}_2$ が観客カメラに投影される点が $\mathbf{x}_1, \mathbf{x}_2$ とすると、 $\mathbf{x}_1, \mathbf{x}_2$ を逆投影したときに作る線がなす角度 α は、カメラ幾何から、

$$\cos \alpha = \frac{\mathbf{x}_1^\top \boldsymbol{\omega}^* \mathbf{x}_2}{\sqrt{\mathbf{x}_1^\top \boldsymbol{\omega}^* \mathbf{x}_1} \sqrt{\mathbf{x}_2^\top \boldsymbol{\omega}^* \mathbf{x}_2}} \quad (13)$$

と導ける [4]。これを f について解くことにより f を求めることができ、 α は、観客カメラのカメラ中心 \mathbf{C} と、 $\mathbf{X}_1, \mathbf{X}_2$ から求めることができる。コート座標における \mathbf{C} は、エピポールがカメラ中心の像であることを利用して、俯瞰撮影カメラ1に対する観客カメラのエピポール \mathbf{e}_1 と俯瞰撮影カメラ2に対する観客カメラのエピポール \mathbf{e}_2 の三角測量により求めた。2点 $\mathbf{X}_1, \mathbf{X}_2$ は、俯瞰撮影カメラ1、2から三角測量によって求めた任意の点のコート座標を利用することができるが、本研究では、俯瞰撮影カメラ1画像の最も右側と左側の2人の観客の頭部位置を利用した。

③シーンの知識を利用する方法

シーンについての事前知識がある場合、その知識を制約として校正に利用できる。本研究では、観客席がコートと同一平面上にあることを利用して、観客カメラ-俯瞰カメラのホモグラフィ $H_{\text{spec-over}}$ を求め、2.1.1. で得られた俯瞰カメラ-コートのホモグラフィ $H_{\text{over-court}}$ より、観客カメラ-コートのホモグラフィ $H_{\text{spec-court}} = H_{\text{spec-over}} \cdot H_{\text{over-court}}$ を求め、2.1.1. と同様にホモグラフィから焦点距離を得る。

④頭部の大きさを使ったヒューリスティックによる方法

カメラからの距離が s である大きさ y_0 の被写体をカメラで撮影したとき、撮影した像の大きさ y_i は、 $s \gg f$ とすると、

$$y_i = y_0 \cdot f / (s - f) \approx y_0 \cdot f / s \quad (14)$$

と表せる。被写体を観客の頭部として y_0 を頭部の大きさとし、同じ観客頭部を観客カメラと俯瞰撮影カメラから撮影したときの観客頭部の大きさをそれぞれ y_i, y_i' 、距離を s, s' 、俯瞰撮影カメラの焦点距離を f' とすれば、

$$y_i / y_i' = f s' / f' s \quad (15)$$

となる。ヒューリスティックに、 $s = s'$ とすれば、 $f = y_i f' / y_i'$ と求めることができる。

2.2. 視線推定

観客カメラの画像では視線推定が困難であるため、頭部方向を視線方向と近似することとした。頭部方向推定は、顔検知によって得られた顔画像に対して行った。顔検知データセット [5] で訓練済みの顔検知モデル [6] を用い、画像内の観客の顔領域を検知した。検知した顔画像に対し、facial landmark データセット [7] で訓練済みの頭部方向推定モデル [8] を用いて3次元頭部方向 \mathbf{h} を出力した。

2.3. 視線投影

コート座標系における観客の視線方向 \mathbf{g} は、コート座標系での観客カメラの回転 R 、観客カメラから推定した観客の視線方向を \mathbf{h} 、 z 軸方向の単位ベクトルを \mathbf{u} とすると、幾何的な関係から、

$$\mathbf{g} = R^{-1} \cdot \text{Rot}(\mathbf{h}) \cdot \mathbf{u} \quad (16)$$

と書ける。ただし、 $\text{Rot}(\mathbf{h})$ は、 \mathbf{h} と \mathbf{u} とのなす角の yaw, pitch, roll 角をそれぞれ $\theta_y, \theta_p, \theta_r$ としたときの回転行列

$$\text{Rot}(\mathbf{h}) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta_p & -\sin \theta_p \\ 0 & \sin \theta_p & \cos \theta_p \end{bmatrix} \begin{bmatrix} \cos \theta_y & 0 & \sin \theta_y \\ 0 & 1 & 0 \\ -\sin \theta_y & 0 & \cos \theta_y \end{bmatrix} \begin{bmatrix} \cos \theta_r & -\sin \theta_r & 0 \\ \sin \theta_r & \cos \theta_r & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (17)$$

である。

注視領域は、観客の視点 (= 頭部位置) から視線方向へコート上に投影した領域である。視線投影は、コート座標系の z 方向は考慮せず、視線方向 \mathbf{g} を x - y 平面上に投影した2次元視線方向 $\bar{\mathbf{g}}$ を用いた。人間の視野は点ではなく一定の大きさを持つため、注視領域を視線上にピークを持ち角度方向に広がりがある分布を持っていると仮定するのが自然であろう。具体的には、角度方向の分布を Gauss 分布と仮定し、コート上の点 \mathbf{q} での注視領域スコア $a(\mathbf{q})$ を、

$$a(\mathbf{q}) = \exp\left(-\frac{\theta^2}{2\sigma^2}\right) \quad (18)$$

とした。ここで、 θ は、観客の頭部位置を \mathbf{p} としたときの、ベクトル $\bar{\mathbf{g}}$ とベクトル $\mathbf{q} - \mathbf{p}$ とのなす角である。 σ は広がり具合を表すパラメーターであり、実験的に $\sigma = 0.2 \text{ rad}$ (11.46°) と設定した。観客の頭部位置 \mathbf{p} は、観客を撮影する異なる2台の俯瞰カメラから検知した顔位置から、三角測量によって推定した。

最後に、複数の観客の注視領域スコアを算出し、それらを合計し、最終的な注視領域スコアとした。

3 実験

3.1. データセットと精度評価

本実験では、茨城県つくば市に本拠地を置く 3×3 バスケットボールチーム「アルボラーダ」主催の 3×3 バスケットボールのリーグにおいて、自然に観客を入れた状態で、コート周辺に設置した俯瞰撮影カメラ2台、観客カメラ1台から撮影した1試合分の映像 [18] を用いて検証を行った。対象とする観客は、対象とする試合時間内に離席せず始終観戦している観客を7人抽出した。対象とする観客のみを抽出するために、対象の観客の顔検知の結果に対してトラッキングを行い、トラッキングがロストした場合は手動で補正を行った。

定量的な精度評価は、ボールを保持しているプレイヤーに注目が集まりやすいことから、注視領域にボールを保持しているプレイヤーが含まれるかどうかを基準として評価した。具体的には、試合全体からランダムにフレームを100フレームサンプリングし、ボールを保持しているプレイヤーの足のコート座標における位置を目視でアノテーションを行った。そして、コート内の領域のうち、注視領域スコア上位50%の領域を抽出し、その領域にアノテーション済みのボール保持プレイヤー座標が含まれるフレームを正解と評価した。なお、試合の展開が停止したフレームおよびパスやシュート動作後にボールが空中に存在しているフレームは評価する100フレームに含めないこととした。

3.2. 結果

焦点距離の推定について、各方法で算出した結果を Table 1 に示す。手法①については、実数解を得ることができなかった。

各焦点距離推定手法による注視領域スコアのマップの例を Fig. 3 に示す。また、各焦点距離推定手法における、ボール保持プレイヤーが含まれるかどうかの精度評価結果を Table 2 に示す。なお、比較のため、ランダムに領域を抽出した場合の期待値も Table 2 に示す。

Table 1 各手法で算出した焦点距離 (①-1: FundMat (Bougnoux), ①-2: FundMat (Hartley): Fundamental matrix から求める方法 (それぞれ Bougnoux, Hartley の方法), ②Angle: 角度を利用する方法, ③Scene: シーンの知識を利用する方法, ④Heuristic: 頭部の大きさを使ったヒューリスティックによる方法)

手法	焦点距離 (pixel)
①-1 FundMat (Bougnoux)	NaN
①-2 FundMat (Hartley)	NaN
② Angle	8668.86
③ Scene	15494.6
④ Heuristic	18761.4

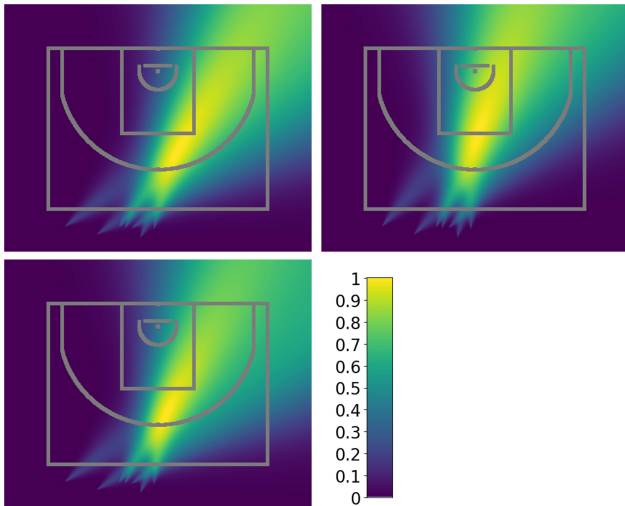


Fig. 3 各手法における注視領域スコアのマップ（左上が② Angle, 右上が③ Scene, 左下が④ Heuristic）. コートラインをオーバーレイ表示している.

Table 2 各焦点距離推定手法による精度評価結果

手法	精度
② Angle	73%
③ Scene	90%
④ Heuristic	78%
ランダム	50%

どの手法においても、ランダムな領域抽出による精度を上回っていることから、ボール保持プレイヤーを含む領域を有意に抽出できていることが分かる。ゲームフィールドにおいて、注目が集まるのはボールを保持しているプレイヤーの周辺領域であることから、この結果は、本研究で提案する手法によって、注目が集まるフィールド領域を抽出できていることを示す。また、③ Scene の手法の精度が高いことは、焦点距離の推定精度も高いことを示唆している。

4 まとめ

スポーツフィールドにおける観客の注視領域を、1台の観客カメラと2台の俯瞰撮影カメラの組み合わせから自己校正を行い、観客全体の注視領域を推定する手法を提案した。そして、実際の3×3バスケットボール試合の撮影データを用いた検証によって、3台のカメラ入力をもとに合理的なコート上の注視領域マップを得ることができることを示した。ボール保持プレイヤーを含む領域を有意に抽出できていることから、推定された注視領域を撮影領域の決定に用いることが期待できる。

今後の展開としては、コートのz軸方向も考慮した注視領域の推定、自己校正に利用する対応点の自動検出、撮影領域への応用、スポーツ以外の領域への展開が考えられる。

謝辞

注視領域に関して、技術的な助言を頂いた株式会社コンピュータマインドの青木溪氏、菅野柊紀氏、山崎貴史氏に感謝を申し上げます。

引用文献

- [1] Z. Zhang, "A Flexible New Technique for Camera Calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330-1334, 2000.
- [2] D. Nistér, "An efficient solution to the five-point relative pose problem," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 6, pp. 756-770, 2004.
- [3] R. Hartley, "Kruppa's equations derived from the fundamental matrix," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 19, no. 2, pp. 133-135, 1997.
- [4] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, 2004.
- [5] S. Yang, P. Luo, C. C. Loy, and X. Tang, "Wider face: A face detection benchmark," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5525-5533, 2016.
- [6] S. Zhang, X. Zhu, Z. Lei, H. Shi, X. Wang, and S. Z. Li, "S3fd: Single shot scale-invariant face detector," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 192-201, 2017.
- [7] X. Zhu, Z. Lei, X. Liu, H. Shi, and S. Z. Li, "Face alignment across large poses: A 3d solution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 146-155, 2016.
- [8] N. Ruiz, E. Chong, and J. M. Rehg, "Fine-grained head pose estimation without keypoints," in *Proceedings of the The IEEE International Conference on Computer Vision and Pattern Recognition Workshops*, pp. 2155-215509, 2018.
- [9] S. Bougnoux, "From Projective to Euclidean Space under any practical situation, a criticism of self-calibration," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 790-796, 1998.
- [10] "AI-Automated Sports Camera, Streaming & Analytics | Pixellot." Pixellot: AI-Automated Sports Camera, Streaming & Analytics, <https://www.pixellot.tv>.
- [11] M. Merler, D. Joshi, Q. B. Nguyen, S. Hammer, J. Kent, J. R. Smith, and R. S. Feris, "Automatic Curation of Golf Highlights using Multimodal Excitement Features," *IEEE Transactions on Multimedia*, vol. 21, no. 5, pp. 1147-1160, 2018.
- [12] J. Chen and J. J. Little, "Where should cameras look at soccer games, Improving smoothness using the overlapped hidden Markov model," *Computer Vision and Image Understanding*, vol. 159, pp. 59-73, 2017.

- [13] K. K. Rachavarapu, M. Kumar, V. Gandhi, and R. Subramanian, "Watch to edit: Video retargeting using gaze," In *Computer Graphics Forum*, vol. 37, no. 2, pp. 205-215, 2018.
- [14] Y. Sugano, X. Zhang, and A. Bulling, "AggreGaze: Collective Estimation of Audience Attention on Public Displays," in *Proc. 29 th ACM Symposium on User Interface Software and Technology*, pp. 821-831, 2016.
- [15] R. J. K. Jacob, "What you look at is what you get: eye movement-based interaction techniques," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 11-18, 1990.
- [16] X. Zhang, Y. Sugano, M. Fritz, and B. Andreas, "Appearance-Based Gaze Estimation in the Wild," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4511-4520, 2015.
- [17] 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," *Asian Conference on Computer Vision workshop*, pp. 15-30, 2018.
- [18] K. Abe, C. Nakamura, Y. Otsubo, T. Koike, and N. Yokoya, "Spectator Excitement Detection in Small-scale Sports Events," in *Proceedings of the 2nd International Workshop on Multimedia Content Analysis in Sports*, pp. 100-107, 2019.
- [19] 阿部和広, 高山侑也, 大坪洋介, 小池哲也, "スポーツフィールドにおける自己校正を使った観客の注視領域推定," *人工知能学会全国大会論文集*, 第36回, 4Yin2-15, 2022.

阿部和広 Kazuhiro ABE
 先進技術開発本部 数理技術研究所
 Mathematical Sciences Research Laboratory
 Advanced Technology Research & Development Division

高山侑也 Yuuya TAKAYAMA
 先進技術開発本部 数理技術研究所
 Mathematical Sciences Research Laboratory
 Advanced Technology Research & Development Division

大坪洋介 Yosuke OTSUBO
 先進技術開発本部 数理技術研究所
 Mathematical Sciences Research Laboratory
 Advanced Technology Research & Development Division

小池哲也 Tetsuya KOIKE
 先進技術開発本部 数理技術研究所
 Mathematical Sciences Research Laboratory
 Advanced Technology Research & Development Division



阿部和広
Kazuhiro ABE



高山侑也
Yuuya TAKAYAMA



大坪洋介
Yosuke OTSUBO



小池哲也
Tetsuya KOIKE