

# 顔向き変化に着目した競技観戦者中の不審人物候補検出

## Detection of Candidates for Suspected Persons in Spectators Focusing on Face Direction Change

岩井 佑樹†      竹長 慎太郎‡      石川 武典†      波部 斉†  
Yuki Iwai      Shintaro Takenaga      Takenori Ishikawa      Hitoshi Habe

### 1. はじめに

近年家庭への防犯カメラの普及や、公共空間での犯罪抑止のため防犯カメラの設置によって、防犯カメラで記録される映像データ量が急速に増加している。また、東京オリンピック・パラリンピックや大阪万博などの幅広い国・地域から多くの人が集まるため、犯罪のターゲットにされることが多々ある。そのため、膨大な防犯カメラ映像から不審な人物を見つけ出し、犯罪を抑止することが必要になるが、人の眼だけでは困難である。

本研究では、防犯カメラ映像から不審人物を割り出す補助となることを目指し、不審人物候補を検出する手法を提案する。不審な行動は多岐にわたるが、その中でも競技観戦中に周囲の人物と異なる動きをしている人物の検出を行う。観戦中の人物は競技の流れに沿って同じような方向を向く傾向があると考えられるが、悪意をもつ人物は競技に関心を示さずに、異なった動きをしていると考えることができる。そこで、競技観戦中に競技を観戦せず他の方向を向いたり、周りを見回している人物を不審人物候補とみなし、顔の向きを手掛かりにそれを検出する手法を提案する。

### 2. 関連研究

人物群衆映像を解析しその中から不審人物候補を検出するためのシステムを作り上げるためには、そのシステムの評価を行うためのデータベースが必要である。黒沢らは群衆及び不審行動人物の映像 DB を作成している [1]。データベースは外国人や子供や高齢者を含む多様な年齢層、人種のエキストラを用意し、駅等の待ち合わせ場所やスポーツ競技場の観客席など全 5 種類のシーンにおいて不審行動を行なっている人物が含まれるものとなっている。今回はこのデータベースの最新版のうち、スポーツ競技場の観客席におけるシーンを使用した。

不審行動人物の検出の先行例としては、寺田らのインターネットカメラの映像を用いた研究がある [2]。この研究では、窃盗行為の際の不審行動検出を行なった。窃盗行為を行う際には物を探す動作を行うと考え、時々ゆっくり移動したり立ち止まったりする人の検出を行なったが、本研究では着席している群衆の映像を取り扱うため不審人物の移動に関する特徴に着目せず、顔向きに着目した不審人物候補の検出を行う。

また、他の不審行動をとっている人物の検出として齋藤らは不審行動をとっている人物ではなく、周りの人物に着目して行なっている [3]。そこでは、カルマンフィルタによって回避行動と思われる行動の検出を行い、隠れマルコフモデルを用いて、行動変化が不審人物に対しての回避行動であるかどうかを判断することによって不審人物を回避した行動を検出し、不審人物の検出を行なったが本研究

では、周りの人が気付くほど明らかに不審である行動へ移る前に不審人物の検出を行う。

### 3. 提案手法の概要

本研究では、競技場内において不審人物候補を検出することを目指し、科学警察研究所によって撮影された不審人物行動データベース [1] を用いて顔の向きに着目することによって、不審人物候補を検出することを目的とする。最近では深層学習、特に畳み込みニューラルネットワーク (CNN) を用いて映像中の顔の検出や、その向きの推定を行う研究が多く行われている。本研究では、それらを用いて顔の検出と顔向き推定を行って不審人物候補を検出する。そのため本研究では、以下の手順を用いる。

- ① 畳み込みニューラルネットワーク (CNN) を用いた Finding Tiny Faces [4] を用いて 1 フレーム内に映っている人物の顔の検出を使用動画全編のフレームに対して行う。
- ② 時系列画像で検出された顔の対応づけを行う。
- ③ 同じく CNN を用いた Head pose estimation in the wild using convolutional Neural Networks and adaptive gradient methods (Deepgaze) [5] を用いて①で検出した顔に対して顔向きの推定を行う。
- ④ ③の顔向き座標データを用いて②で識別したそれぞれの人物毎の顔に対して二つの指標を用意し、比較を行う。

上記の中で、Finding Tiny Faces と Deepgaze が先に述べた顔の検出や顔向き推定の先行事例であるので、それらについて述べた後で、本研究の独自の処理である②、④について述べる。

#### 3.1 Finding Tiny Faces

Finding Tiny Faces とは CNN を用いて様々なサイズの顔を精度よく検出する技術である [4]。ImageNet を ResNet101 へ学習させたモデルを用いている。また、ImageNet 内データの 80%以上が 40 から 140 ピクセルであったため 40 から 140 ピクセル内に収まらない顔の検出精度が上がらなかったが、40 ピクセル以下の画像を 2 倍の解像度に、140 ピクセル以上の画像を半分の解像度にリサイズし、その間のサイズの画像はそのままの解像度で学習させることにより、様々なサイズの顔を非常に高い精度で検出をすることができるようにしたものである。

†近畿大学, Kindai University

‡近畿大学大学院, Graduate School of Kindai University

### 3.2 Deepgaze

Deepgaze は最適化手法に RMSProp を使用し、CNN を用いた頭部姿勢推定技術のことである [5]。入力された顔画像に対して CNN を用いて頭部の向きを推定し、roll, pitch, yaw 三方向で正面を基準の 0 度として向きを出力する。

ここで、roll は頭部を左右に傾ける方向を、pitch は頭部を上下に振る方向つまり頷く動作の際に動かす方向を、yaw は頭部を左右に振る方向を表している。

この中で、頭部を左右に動かす動作を示す yaw が不審者候補の推定に有用であると予想される。以下の実験ではそれを確かめる。

### 3.3 検出した顔の時系列対応付け

Finding Tiny Faces では画像フレームごとで顔の位置を検出する。顔の動きを得るためには、隣接するフレーム間で同一人物に相当する顔を対応づけなくてはならない。今回対象とする観客席にいる人物の場合は、速い動作をすることは少ないため、検出された顔領域の中心座標を計算し、隣接するフレームで最も近い座標をもつものを同一人物に対応すると見なして対応づけることにした。

### 3.4 不審人物候補推定の指標

今回不審人物候補の検出では、二つの指標を提案し、それぞれの有効性を後の実験で確かめる。一つ目の指標 (1) は、それぞれの人物がどの程度顔向きを変化させたかを定量化したものである。二つ目の指標 (2) は、他の人物との顔向きの違いを定量化したものである。以下では、それぞれについて詳細に述べる。

**指標 (1) :** 数式①, ②, ③のように定義する。ここで、 $R_t, P_t, Y_t$  はある人物の第  $t$  フレームでの顔の向きの roll, pitch, yaw をそれぞれ示している。初期フレームから隣接する時刻での角度の変化の絶対値を計算し、それを足し合わせたものを  $R_{sum}, P_{sum}, Y_{sum}$  としている。不審者は周囲の様子を探るために顔向きを頻繁に変えられ、その場合、これらの値は極端に大きくなると考えられる。

**指標 (2) :** あるフレーム  $t$  において Finding Tiny Faces で検出した  $n$  人分の roll, pitch, yaw それぞれの中央値をもとめ、 $R_{med_t}, P_{med_t}, Y_{med_t}$  とする。次に、数式④, ⑤, ⑥のようにフレーム毎それぞれの人物の顔向きと中央値の差分  $|R_t - R_{med_t}|, |P_t - P_{med_t}|, |Y_t - Y_{med_t}|$  を人物毎に全フレームで足し合わせたものを  $\Delta_R, \Delta_P, \Delta_Y$  として指標 (2) とする。先に述べたように、競技観戦中はゲームの進行にあわせて同じような方向を向くのが自然であるので、それからの逸脱度合いを定量化したものである。

$$R_{sum} = \sum_{t=2}^m |R_t - R_{t-1}| \quad \dots \textcircled{1}$$

$$P_{sum} = \sum_{t=2}^m |P_t - P_{t-1}| \quad \dots \textcircled{2}$$

$$Y_{sum} = \sum_{t=2}^m |Y_t - Y_{t-1}| \quad \dots \textcircled{3}$$

$$\Delta_R = \sum_{t=1}^n |R_t - R_{med_t}| \quad \dots \textcircled{4}$$

$$\Delta_P = \sum_{t=1}^n |P_t - P_{med_t}| \quad \dots \textcircled{5}$$

$$\Delta_Y = \sum_{t=1}^n |Y_t - Y_{med_t}| \quad \dots \textcircled{6}$$

これらの指標は、それぞれ大きな値になるほど不審人物候補としての尤度が高くなると言える。そこで、次節の実験では、指標 (1), (2) を検出した全ての人物に対して計算して降順に並べ、不審人物候補として検出すべき人物が何番目に位置するかを評価することにした。

## 4. 実験

### 4.1. データセット

本研究で用いる映像データは、科学警察研究所によって撮影された不審行動データベースのうち競技場で競技観戦を行なっている 93 人の群衆を正面から撮影した映像である。当該映像には一名の不審人物役の男性が映っている。その男性は最初周りと同じように競技を観戦していたが、途中から左右をしきりに確認した後、下を向いて自分のカバンの中を物色して携帯電話を取り出し、競技観戦とは関係のない方向を向きながら携帯電話で電話をかけ始める。

また、使用する映像は競技観戦者が途中で立ち上がることや、映像内で人物の数が増減しない場合のみとする。

さらに映像中では、対象の男性が不審な行動を行っている部分と、そうでない部分がある。そこで、(A) 不審行動と不審ではない行動の両方を含む映像と、(B) 不審行動を行っている部分だけを含む映像の二種類の映像で実験を行った。以上のような映像を映像のフレームレートである 59.94fps でフレーム毎に分割し、実験に使用した。



図 1 今回実験で使用する動画のあるフレーム

### 4.2. 実験手法

本研究では、3.1 章で用意した画像に対して 2 章で述べた手法を適用した。以下の図 2 は Finding Tiny Faces 適用後画像と、その画像から犯人役の男性の顔を切り抜いたものである。また、左上から右下へと顔向きの様子を時系列順に示したものであり、特に顔向きの変化が大きかった部分を抽出して図示した。4.1 章で述べた様に使用した動画のフレームレートは 59.94fps であるため、図 2 中の矢印記号の下に記述したフレーム数は、59.94 フレームで 1 秒を表している。さらに、図 2 中の画像のうち最初と最後に使用している画像が、それぞれ実験に使用した動画の第 1 フレームと最終フレームを抽出してきたものではない。

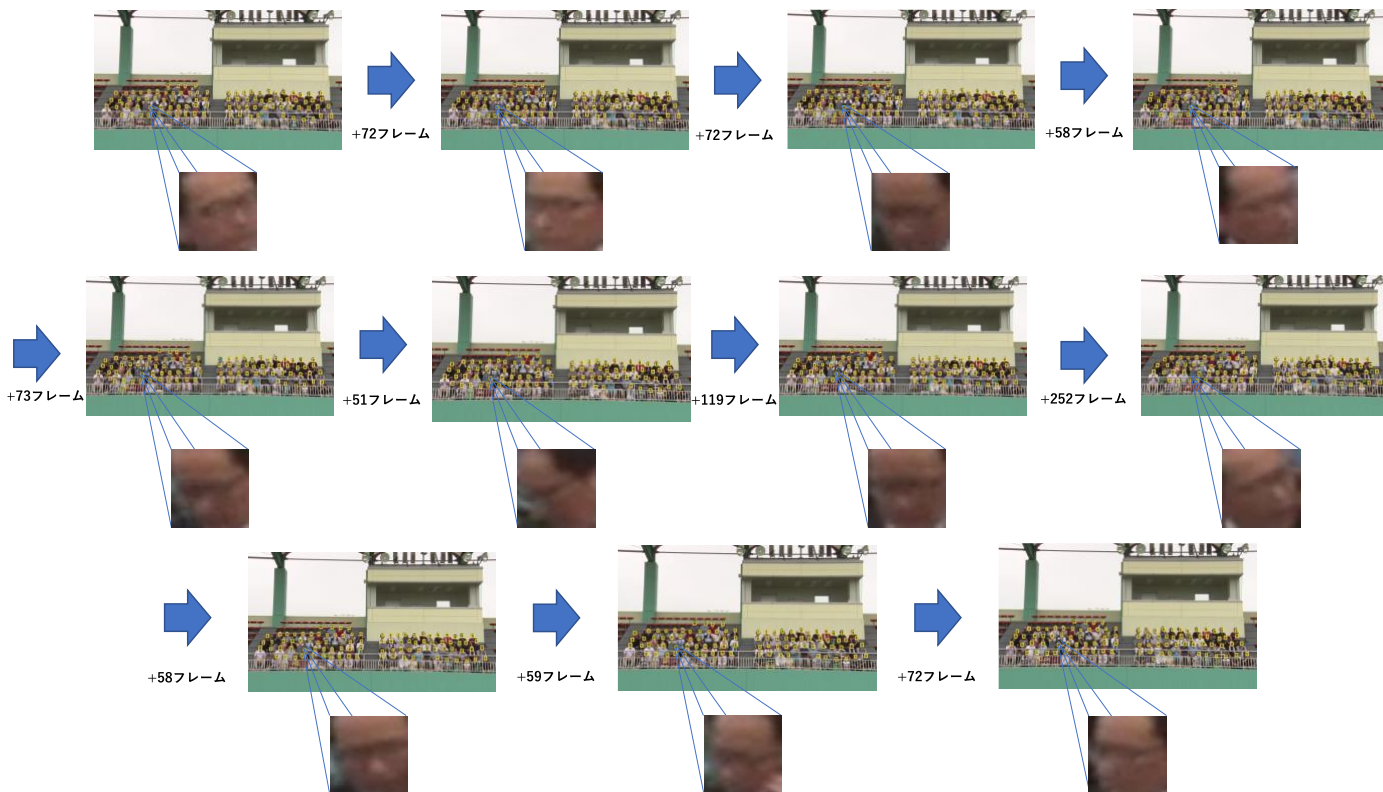


図2 Finding Tiny Faces 適用後の犯人役男性の様子

はじめに、今回最前列に座っている人物の顔がフェンスと被っているため、最前列に座っている 20 人の Finding Tiny Faces による顔検出、Deepgaze による顔向き推定の精度がフェンスのない場所に座っている人物に比べて著しく下がってしまい、不審度リストを生成する際に悪影響を与えてしまうことが考えられるため、最前列において Finding Tiny Faces によって検出されたものを、不審度リスト生成時にリストに加えないようにして実験を行なった。その後、追加実験としてどの程度悪影響を与えているかどうかを調査するために、最前列において Finding Tiny Faces によって検出されたものも含めて不審度リストを生成する実験を行なった。

### 4.3 実験の結果

4.2 章で述べた実験を行った後の結果のうち、不審人物役の男性が 3.4 章で述べた指標 (1)、指標 (2) で何番目に多くの不審行動を行なっているかどうかについて、3.1 章で述べた (A) の映像を用いた結果を表 1 に、(B) の映像を用いた結果を表 2 に示す。表に示した数値は、全体で対象とした人物の数のうち、不審な行動をとる人物が何番目に位置していたかを示している。たとえば、66/73 は全体で 73 人いるうちで、不審な行動をとる人物が 66 番目のスコアをとっており、望ましくない結果と言える。

また、4.1 章で使用映像の中に 93 人映っていると述べているが、4.2 章で述べたように最前列の検出結果は不審度

リスト生成時に除外しているため、最前列に座る 20 人分少ない 73 人で不審度リストを生成している。

まず、表 1 から指標 (1) での結果は映像 (A)、(B) 共に yaw の結果のみ少し良好だが、全体の半分程度の位置であり、不審人物候補と言えるほどの結果ではないことがわかる。

次に、表 2 から指標 (2) では全体的に表 1 の結果よりも優れており、yaw の結果が特に優れていることがわかる。その中でも映像 (B) における yaw が不審人物役の男性がとった行動を正しく捉えていることがわかる。

さらに 4.2 章で述べた追加実験を行ない、先ほどと同様に指標 (1) の結果を表 3 に、指標 (2) の結果を表 4 に示した。結果は先ほど行なったフェンスの部分を除いた実験と傾向は似たものとなった。ただ、表 4 に示す指標 (2) の yaw の結果が、表 2 の yaw の結果に比べて悪くなり、不審行動役の男性を上から三番目に不審な行動をとったと検出していることがわかる。

表 1 指標 (1) の結果

	roll	pitch	yaw
映像 (A) 「不審+通常」	66/73	58/73	28/73
映像 (B) 「不審」	64/73	72/73	36/73

表2 指標(2)の結果

	roll	pitch	yaw
映像(A) 「不審+通常」	34/73	27/73	3/73
映像(B) 「不審」	26/73	23/73	1/73

表3 指標(1)の結果  
(フェンス部分の人物を含めた場合)

	roll	pitch	yaw
映像(A) 「不審+通常」	92/95	68/95	32/95
映像(B) 「不審」	90/95	92/95	43/95

表4 指標(2)の結果  
(フェンス部分の人物を含めた場合)

	roll	pitch	yaw
映像(A) 「不審+通常」	50/95	31/95	7/95
映像(B) 「不審」	44/95	31/95	3/95

## 5 考察

本研究では、競技場において競技観戦を行なっている群衆の中から不審者候補の検出を行い、人の目で不審者を探す際の補助になりうるような手法を提案し、実験を行なった。4.3章で示した実験結果のうち、不審人物役の男性が不審行動を行なっている部分のみが映っている映像(B)を使った実験において、指標(2)の群衆が顔を向けている方向とは別の方向に頻繁に顔を向けているかどうかの指標で、yawの値が不審人物役の男性を最も不審行動をとっていると示した。これは不審人物役の男性の行動がyaw方向の動きである、しきりに左右を確認したり、競技とは関係のない方向を向いて電話をかけたりにしていたためであると言える。それを補強する結果として、表2のrollとpitchでは上位に来てはいるものの不審人物候補と言えるほど不審人物役の男性の不審度が高くなっていないことがわかる。

また、表2の映像(A)の結果においては、表2の映像(B)の結果と比べると不審度が低くなっているが、映像(A)は不審人物役の男性が不審行動を行なっていない部分も含んだ映像であるため、表2の映像(A)と比較して不審度が下がったことは正しく検出できていると言える。

しかしながら、追加で行なったフェンス部分を含める実験の結果のうち表4で示したように、映像(B)において不審度は上から三番目、映像(A)において不審度は上から七番目となっており、フェンスを除外した時の結果と比較すると結果が悪くなっているのがわかる。これは図3に示したように、フェンスと腕が重なっている部分を Finding Tiny Faces が人間の顔であると誤検出したことや、図4で示したように、表4の映像(B)の不審度リストで一番不審とされた人物と二番目に不審とされた人物の顔の一部がフェンスで隠れているためでもある。

さらに表4の映像(B)の不審度リストの上から3人の人物のyawの値を表5に示した。表5では映像(B)全編である1186フレーム分の値を足し合わせたものであるが、yawにおけるフレーム毎の群衆が向いている方向との差分の平均を計算すると、上から順番に50.510度、25.741度、23.854度となる。競技を観戦しているだけである1番目の人物と2番目の人物の値が極端に大きくなっていることがわかる。これらの結果から群衆の中に顔の大部分が隠れている人物が複数人いる場合、本研究で提案する手法を用いて行う不審人物候補の検出の精度が落ちてしまうと言える。

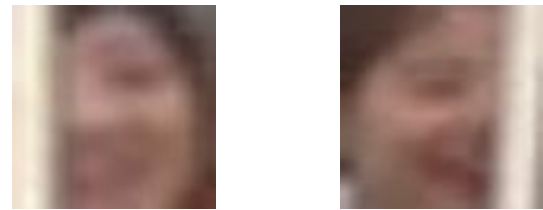
表5 表4映像(B)のyawの値

順位	yawの数値(度×フレーム数)
1	59905.529
2	30529.641
3	28291.106



## 誤検出部分

図3 Finding Tiny Facesの誤検出



一番目

二番目

図4 表5の一番目と二番目

## 5 終わりに

Finding Tiny FacesとDeepgazeによる顔向きから不審人物候補を検出する手法を提案し、科学警察研究所の作成した不審行動データベースのうち競技場内で、競技観戦者の中から、顔を他の人と別の方向に向けていることが多くないかどうかを算出することにより、不審人物役の男性を検出することができた。そのため不審人物候補の検出に一定の成果があったと考えられる。



しかしながら今回の研究では、座っている人物しかいない状況に限定し、誤検出を減らし精度を向上させるためにフェンスと顔が被っている人を除外した。このように条件を絞った状況での結果であるので、顔向き以外の指標を用意することによって、今回除外せざるをえなかった状況においても不審人物候補の検出を行えるようにすることが今後の課題である。

## 謝辞

本研究の一部は、JSPS 科研費 JP19H04939, JP17K00256 の助成を受けて行った。

## 参考文献

- [1] 黒沢健至, 秋葉教充, 土屋兼一, 角田英俊, 井元大輔, 平林学人, 五十嵐直明, 黒木健郎, 映像解析を用いた安全安心技術開発のための評価用映像データベースの構築, 画像センシングシンポジウム (SSII) 2017.
- [2] 寺田賢治, 鴨頭大輔, インターネットカメラを用いた不審行動人物の検出, 電気学会論文誌 C 127 巻 11 号, pp1871-1879, 2007.
- [3] 齋藤伶奈, 大沢英一, 周辺歩行者の挙動変化に基づく不審者検出, 2018 年度人工知能学会全国大会 (32 回) .
- [4] Peiyun Hu, Deva Ramanan, Finding Tiny Faces, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017).
- [5] Massimiliano Patacchiola, Angelo Cangelosi (2017) , Head pose estimation in the wild using Convolutional Neural Networks and adaptive gradient methods, Pattern Recognition.