

令和 2 年 5 月 22 日現在

機関番号：34419

研究種目：基盤研究(C)（一般）

研究期間：2017～2019

課題番号：17K00256

研究課題名（和文）個の意図・状態を考慮した群衆フローの詳細解析

研究課題名（英文）Crowd flow analysis based on individual intentions and states

研究代表者

波部 斉 (Habe, Hitoshi)

近畿大学・理工学部・准教授

研究者番号：80346072

交付決定額（研究期間全体）：（直接経費） 3,600,000円

研究成果の概要（和文）：本研究では顔向き変化に着目し、主に、不審者検出への応用を中心に画像処理技術の開発とその検証を実施した。顔検出および顔向き変化の推定は、それぞれの分野において現在最高水準の精度を持つとされている、TinyFaceとDeepgazeを用い、そこから得られた時系列の顔向きデータを利用して、人物の意図、特に不審行動をとっている人物の推定を行った。時系列の顔向きデータの中から、他と比べて特異な動きをしている人物や、不自然に大きな動きをしている人物を特定し、それが不審行動をとっている人物とすることにした。この手法の有効性を検証するために、スタジアムでの観客席を想定した映像データセットで実験を行った。

研究成果の学術的意義や社会的意義

顔検出および顔向き変化の推定といった「低レベル」な情報の推定は深層学習によって大いに発展してきている。それに比べて、意図や状態といった「高レベル」な情報の推定は大きな課題となっていて様々なグループが研究を行っている。本研究も、そのような課題にチャレンジしたものであり、学術的な意義が大きい。また、不審人物候補の検出はセキュリティ、防犯などの社会的な要求が大きい応用分野であるので、本研究の社会的な意義も大きいと言える。

研究成果の概要（英文）：In this study, we developed and verified an image processing technique focusing on the change in facial orientation and mainly applying it to the detection of suspicious persons. For face detection and face change estimation, we used TinyFace and Deepgaze, both of which are considered to have the highest level of accuracy in their respective fields, to estimate the intentions of the person, especially the person acting suspiciously, using the time series of face orientation data obtained from TinyFace and Deepgaze. From the time series data, we chose to identify people who moved strangely or unnaturally large in comparison with other people, and identified them as people who were acting suspiciously. In order to verify the effectiveness of this method, we conducted experiments with a video dataset assuming spectator seats in a stadium.

研究分野：コンピュータビジョン，映像解析

キーワード：群衆行動 不審行動 人物検出 畳み込みニューラルネットワーク

様式 C-19、F-19-1、Z-19 (共通)

1. 研究開始当初の背景

映像を用いた人物の行動理解は、防犯などのセキュリティ用途、適切な避難誘導など通行者への情報提供、人物行動の統計情報の施設管理者への提供など、様々な用途が考えられる。我々人間は他者の行動をみてその意図や目的を容易に推測できることが多く、それと同等な機能を計算機で実現できれば、人間の知能の解明にもつながり学術的な意義も大きい。そのため、古くから多くの研究者が着目して様々な研究を展開している。

一方、特に日本国内では、プライバシー意識の高まりから街中に取り付けられたカメラでの個人を識別できる映像の取得と解析には抵抗感を感じる人々が多く、プライバシー情報を除去する試みも行われている。しかし、最終的には公共の利益と個人々のプライバシーとのトレードオフになり、社会的なコンセンサスが得られているとはいえない。

このような状況で有効な手段と考えられるのは、個人々が判別できないような解像度で捉えた映像を用い、行き交う人々を群として扱って様々な情報を得るアプローチである（以下、群衆映像解析と呼ぶ）。群衆映像解析も以前より注目を集めており、多くの研究事例がある。しかし、多くの手法はまさしく「群」としての振る舞いにしか注目しておらず、災害発生時などの異常行動の検出や、群衆フローの分割などを行うにとどまっている。人が観たときに理解できる豊富な情報の理解には至っていない。

2. 研究の目的

このような問題意識のもとに意図・状態を考慮した群衆フローの詳細解析を行うための、人間の意図を推定する技術を開発する。本研究ではその中でも特に、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した。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、周りを見回したり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。

3. 研究の方法

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、その向きを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。具体的には、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 [2] 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) [3] 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 [2]。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。

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 [3]。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。今回は頭部を左右に動かす動作を元に不審者候補の推定を行うため、特に yaw に着目した。

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。

$$R_{sum} = R_1 + \sum_{t=1}^m |R_t - R_{t-1}| \quad \dots \textcircled{1}$$

$$P_{sum} = P_1 + \sum_{t=1}^m |P_t - P_{t-1}| \quad \dots \textcircled{2}$$

$$Y_{sum} = Y_1 + \sum_{t=1}^m |Y_t - Y_{t-1}| \quad \dots \textcircled{3}$$

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。

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

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

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

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. 研究成果

3. で述べた提案手法の妥当性を以下の実験で評価した。

4.1 データセットと手法

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を利用した(右図)。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。

以下の図は 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き、左上から右下へと顔向き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ている。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。



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。

4.2 実験の結果

不審人物役の男性が指標 (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、4.1 節で述べた (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。

なお、4.1 節で使用映像の中に 93 人映っていると述べているが、先に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。

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

さらに 4.1 節の最後に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 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

4.3 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た。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。

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] Peiyun Hu, Deva Ramanan, Finding Tiny Faces, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017).
- [3] Massimiliano Patacchiola, Angelo Cangelosi (2017), Head pose estimation in the wild using Convolutional Neural Networks and adaptive gradient methods, Pattern Recognition.

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件/うち国際共著 0件/うちオープンアクセス 1件）

1. 著者名 波部 斉, 橋本 知典, 満上 育久, 鷲見 和彦, 八木 康史	4. 巻 29
2. 論文標題 人物のジェスチャーを加味した歩行者グループ検出	5. 発行年 2017年
3. 雑誌名 知能と情報	6. 最初と最後の頁 605 ~ 610
掲載論文のDOI (デジタルオブジェクト識別子) 10.3156/jssoft.29.3_605	査読の有無 有
オープンアクセス オープンアクセスとしている(また、その予定である)	国際共著 -

〔学会発表〕 計4件（うち招待講演 1件/うち国際学会 1件）

1. 発表者名 村田一星, 波部 斉, 阿部孝司, 井口信和
2. 発表標題 実用的な大規模魚群中の個体検出手法の検討
3. 学会等名 第81回情報処理学会全国大会
4. 発表年 2019年

1. 発表者名 村田一星, 波部 斉, 阿部孝司, 井口信和
2. 発表標題 機械学習を用いた密度マップ推定による大規模魚群中の個体検出
3. 学会等名 情報処理学会関西支部支部大会
4. 発表年 2018年

1. 発表者名 Hitoshi Habe
2. 発表標題 Visual Surveillance for Collective Behavior Analysis: From Human to Fish
3. 学会等名 ICMA-SURE 2018 (招待講演) (国際学会)
4. 発表年 2018年

1. 発表者名 岩井佑樹, 竹長慎太郎, 石川武典, 波部斉
2. 発表標題 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
3. 学会等名 情報処理学会関西支部支部大会
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----