

既設カメラなどの低解像度映像による人物検出システム

辻健太郎^{†1} 鄭明燮^{†1} 中島望夢李^{†2} 松田裕司^{†1} 宮崎信浩^{†1} 皆川明洋^{†1}

あらまし プライバシーの問題に配慮した人流の認識のため、既設のカメラを利用した低解像度映像を入力とする人物検出手法を提案する。低解像度映像では、人物の特定が難しいのと同時に、近接したり重複したりする人物の存在により位置決め精度が低下し、人物ひとりひとりの切り出しが困難になる。提案手法では、まず頭部検出フィルタによって頭部候補を検出し、次にその検出された頭部候補に基づいて、上半身部分の特徴を評価することで、低解像度でも位置決め精度の低下を回避した人物の検出を実現する。延べ428人に対する評価実験により、頭部幅が3画素から8画素の人物の検出率が70%以上であることを確認した。

キーワード 監視カメラ, 低解像度画像, 人物検出, 頭部検出

A method of human detection using multiple detectors from low-resolution video

Kentaro Tsuji^{†1}, Mingxie Zheng^{†1}, Moyuri Nakashima^{†2}, Yuji Matsuda^{†1},
Nobuhiro Miyazaki^{†1} and Akihiro Minagawa^{†1}

Abstract In this paper, we propose a human detection method in which low-resolution video is input from surveillance camera for recognizing human flow considering with privacy protection. Human detection is a difficult problem when low-resolution video is input, and human segmentation is also difficult when some people are captured in near area of low-resolution image. In the proposed method, human detection for low-resolution video is realized by two steps: head detection and body detection. First, head detection is executed, and then features of body are evaluated based on the position of the detected head. From the experimental result for 428 people, we confirmed that the detection rate is over 70% even if the head width is from three pixels to eight pixels.

Key Words surveillance camera, low-resolution video, human detection, head detection

1. はじめに

近年、商業施設の大規模化も手伝って、施設内や街中での人の移動経路である、いわゆる「人流情報」を使って、店舗の配置や店舗内の品ぞろえや人員の配置などを最適化し、顧客サービスの向上を図る傾向にある[1]。たとえば、街中では、イベント時の混雑解消、交通機関の運行計画策定や災害時の避難誘導をおこなうためにも、人流情報が有効に活用されている[2]。

人流情報の獲得には、種々様々な方法が提案されている。ビーコンやRFIDタグを使用する方法、人感センサを利用する方法、スマートホンのWi-Fiの機能で位置を検出する方法もある(たとえば文献[3])。しかしながら、これらの方法は、センサの設置コストの問題、多数の通行者を同時に補足する困難さ、さらにはデバイスやタグを携帯しなければならないなど、不特定多数を対象に人流情報を獲得するには課題が多い。

そこで、セキュリティ用途で設置されたカメラからの映像を処理して人流情報を獲得する方法が注目されている。この方法では、既に設置されているカメラを利用するため、

機材コストやその設置のコストを最小限に抑えることができ、また、人がスマートホンなどのデバイスを携帯することを前提にしないので、不特定多数の人を対象とした人流情報を獲得できる可能性をもっている。

しかしながら、街中や商業施設内に設置されたカメラを活用する際に大きな課題のひとつとして挙げられるのは、プライバシーへの配慮である。特に、最近ではカメラの性能向上は目覚ましく、高解像度化が進み、人物の顔がはっきりと視認できるようになった。そのため、いつどこで何をしていたかなど、撮影された行動自体の情報を利用されたり、第三者に知られたりしたくないというプライバシーに関わる問題が社会問題にもなっている。これらのカメラ映像を活用するには、被写体となる人に用途を明示することが前提であるが、プライバシーに関する情報を利用しない手立てがあれば、被写体となる人の懸念の払拭につながる。この問題を、本稿で扱うプライバシー問題とする。

このプライバシー問題に対して、映像中の顔部分にスクランブルを施す方法や[4]、映像中の物体部分を背景差分に基づいて抽出し、スクランブルを施す方法が提案されている[5]。しかし、いずれの方法も、検出漏れや誤りが避けられないため、課題が残る。映像解析システムとしてプライバシーの問題に対処するには、個人が特定できるような映像を入力することなく、所望の解析ができるシステムを実

^{†1} 株式会社富士通研究所
Fujitsu Laboratories Ltd.

^{†2} 富士通九州ネットワークテクノロジーズ株式会社
Fujitsu Kyushu Network Technologies Limited

現すべきである。この問題に対処する方法として、すべての映像に対して、被写体が誰であるか特定できないぐらいまで解像度を低下させた(ダウン・レゾリューション)映像(以降、低解像度映像と呼ぶ)を用いる方法がある。例えば、人の顔が数画素程度で表され、個人が特定できない映像を用いる。プライバシー問題に対処する一つの方法は、そのような低解像度映像を入力としても、人物が検出でき、移動経路が抽出できるシステムを構築することである。

筆者らは、低解像度映像を処理して、人流情報を獲得するシステムの開発を進めている。本稿は、その第一段階として、低解像度映像を時系列画像として取扱い、それぞれ1枚ずつの静止画像(以降、低解像度画像と呼ぶ)から人物検出する手法について述べる。今後、人物の動きなどの時間軸の情報を使いながら、高精度化を進めてゆく。

本稿では、設置コストの観点から、既設のカメラで実現することを前提に、高所から見下ろして人を俯瞰するように設置したカメラの利用を考える。このような設置では、人物同士の重なり・遮蔽(オクリュージョン)が生じるが、低解像度画像でも、高精度に人物検出が可能な手法を提案する。

2. 人物検出の関連研究

本稿では、低解像度画像1枚を対象に、人物同士のオクリュージョンに対応した人物検出を実現する手法を提案する。通常セキュリティ用途で設置されているカメラは、高所から人が多数撮影できるように俯瞰するように設置されており、人物を横から撮影する角度に近くなる。このような撮影角度に近い、従来から議論されている歩行者検出・人物検出技術は比較対象とするが、筆者らが以前開発した天井から真下を見下ろすカメラを用い、直下の人物の流れを認識する方法[6]は、本論では比較対象外とする。

代表的な歩行者検出・人物検出技術については、文献[7]で16例が述べられている。それらの手法の中で、特に多く用いられている画像特徴は、HOG(Histograms of Oriented Gradients)であり、16手法中14手法で採用されている。

HOG特徴は、近傍画素からセルを構成し、セル内部の画素における局所的な勾配情報をヒストグラム化されることで、人物の形状の変化にも対応可能な柔軟な特徴であるといえる[8]。一方で、HOG特徴は、入力した画像情報をより低いサンプリング間隔で再標本化した特徴であるとみなすことができる。このような再標本化過程を持つ検出方法では、低解像度画像において特徴自体が平坦化され、位置ずれを引き起こしやすくなり、結果として、検出した位置に人物が一致せず、誤検出となる。また、HOG特徴に別の特徴を追加することで、精度の向上が試みられている。例えば、文献[8]のHOG特徴にHaar-like特徴などを併用したり[9]、LBP(local binary pattern)特徴を併用したりすること

が提案されている[10]。これらの方法では、検出対象となる人物のサイズが90画素のように大きい場合に、文献[8]の方法と差異が表れる一方で、45画素のサイズではその効果がほとんど見られない。このことから、学習として与えた画像の解像度を無視すれば、HOGと他の特徴の組み合わせは、高解像度画像から得られる細部の特徴を追加することであると考えられる。同時に、そのような高解像度画像に対する学習が盛んに行われてきている一方で、低解像度画像に関する人物検出は行われてこなかったと捉えることもできる。

一方で、人物の形状を利用せず、低解像度の映像を対象に、人物を抽出する手法も提案されている。文献[11]の手法は、1台のカメラで数十メートル四方を監視する広域監視技術であり、全身の大きさが10画素四方で撮影される人物の検出を可能にする。この方法では、画像中で変化した部分の輝度値の周波数から人物かノイズかを判定することで、極めて小さなサイズで撮影された人物でも検出できる。その一方で、この方法では、人物以外の物体の検出も生じること、また人が重なった場合に、静止し、かつ、重なりのある人は抽出できないなどの問題がある。

3. 低解像度人物検出手法

3.1 概要

本手法では、低解像度画像からの人物検出を実現するため、低解像度画像に現れる人物特有の情報に基づいて検出を実現する。

具体的には、人は頭部をもった形状である、と定義し、最初に頭部形状(らしい領域)に合致するフィルタを設計し、発見する。もちろん、頭部が隠れた人は検出できないが、頭部が隠れる場合、体も隠れるので、静止画像1枚の抽出では漏れを許容した。頭部形状(らしい領域)が発見できると、続いて、その頭部を含む上半身部分の輝度勾配の分布を確認し、発見した領域が頭部であることを検証する。このように、頭部とその上半身の形状と輝度分布を使った二段階の手法で低解像度でも人を検出できる手法を考案した。次節以降では、それぞれの特徴と検出手法の得失について述べたのち、その統合方法について述べる。

3.2 頭部の検出

図1に今回の低解像度映像から抽出される、頭部画像の例(頭部径が約8画素)を示す。これらの画像では、頭部より細かい部位の検出、例えば顔の部位の抽出は困難である一方で、解像度が低いわりに、顔には陰影によるパターンが残存し、髪との境界も多様であることがわかる。同時に、顔の向きに関するパリエーションも存在している。例えば、文献[4]では、身体の向きに対して7つのパターンを与えている。しかし、髪と肌の境界のパターンの多様性ま

で考慮すると、相当な数の辞書が必要になる。これこそが低解像度で生じる問題であり、特徴自体が少ないにも関わらず、得られる情報に共通性が見いだせず、照合が困難となり、一括した認識手法が適用できない。結果として、情報の多様性に合わせて辞書数を増加すると、検出の安定性が低下するだけでなく、辞書数に応じた処理時間も必要となる、という弊害が現れる。

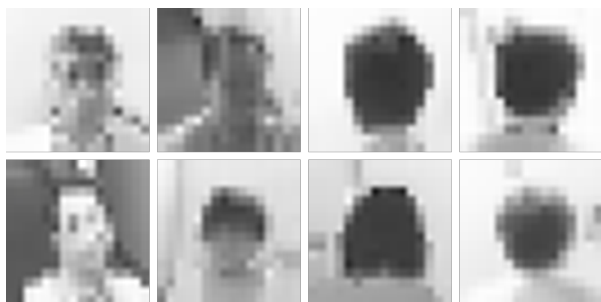


図 1 低解像度映像で撮影された頭部の例。

一方で、図 1 の右半分にある後頭部からの撮影では、顔が撮影されないことから、顔の中に見えるパターンが存在しない。そのため、これらの画像から共通性を見出すことができる。これらの観測に基づき、本手法では、後頭部を検出する辞書を前向き頭部にも適用できるように、画像を加工することで、安定かつ高速な検出を実現する。

具体的には、前向き頭部画像を後ろ向きにそれになるように加工した後、照合する。前向き頭部画像には顔領域が存在する。顔の領域を肌色に基づいて抽出し、顔の領域のそれぞれの画素の色を、頭髪色に変換する。これにより、複雑なパターンを消去した、低解像度でも利用可能な単純かつ検出が容易なテクスチャとなるように頭部画像を生成する。

上記で得られた画像に対して、検出では、Haar-like 特徴を拡張した図 2 に示すような検出フィルタを利用して検出する。図 2 のフィルタにおいて、領域 A, B, C は、頭部、背景、肩をそれぞれ表す。本方法では、このフィルタで対象画像をスキャンすることで、頭部領域を検出する。なお、領域 C については、服の色の多様性を考慮し、別の値を検出する複数のマスクを用意し、用意したマスクの中で対象の座標に対して実行した際に、最も高い値を対象座標における頭部の評価値とする。

なお、剃髪している人などのように、頭髪に相当する画素がない場合には、背景との差分が大きくなるような値で変換した。

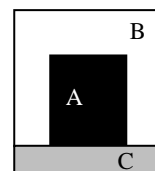


図 2 頭部検出用フィルタの一例。

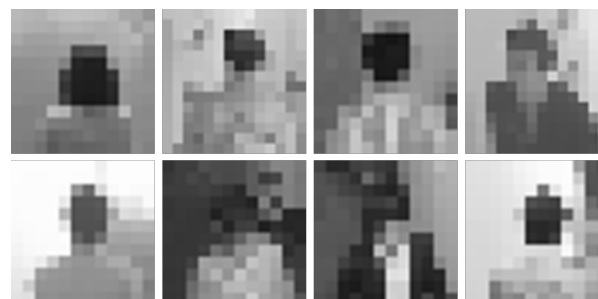


図 3 低解像度映像から抽出した上半身画像。

3.3 上半身の検出

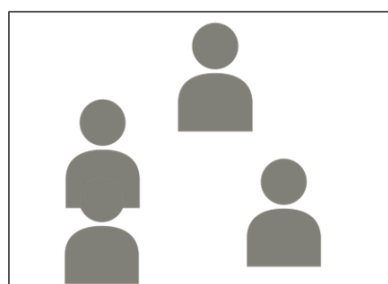
図 3 に低解像度映像から抽出した上半身画像を示す。この図より、上半身の検出では、前述の頭部検出ほど形状が限定できず、変形の自由度が高いことがわかる。さらに、服に関する色の自由度を考慮すると、頭部検出で用いた Haar-like 特徴では膨大なテンプレート数が必要になるため、適用が難しい。そこで、上半身の検出として、より自由度の高い HOG 特徴を用いて概形を照合することとする。

3.4 検出手法の統合

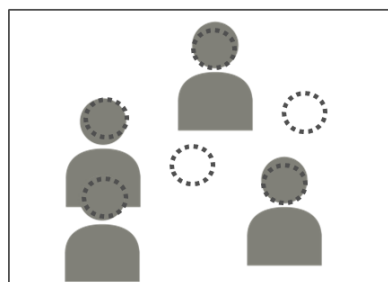
今回の頭部検出手法では、頭部のシンプルな形状と色とを抽出することを目的としており、頭部は抽出できるものの、円形ないしは中心部が比較的暗い色である物体すべてを検出するという問題がある。

一方、上半身検出手法では、HOG を用い、上半身に相当する箇所を検出する。しかし、第 2 章で述べたように、人物検出に関する位置感度が低下していることから、近傍に人物がいても比較的高い評価値を出力するという問題がある。

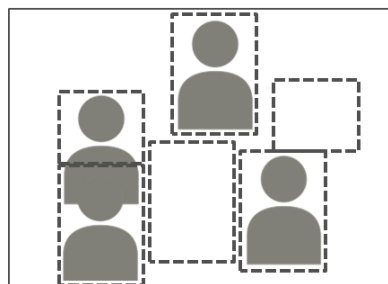
このような二つの手法を相補的に利用することを考えると、最初に、頭部検出で頭部となる候補を画像中から検出し、次に、それが人物であるかどうかを上半身で検証する枠組みが実現可能である。すなわち、頭部検出によって離散的に得られた人物候補から、上半身に対して概形を照合することで人物に相当する領域にあるかどうかを特定する。これにより、頭部検出における誤検出を回避しながら、HOG を用いた上半身検出の位置決め精度の低下も回避する。この手順を図 4 に示す。この手順では、画像全体に HOG を実行する必要もなくなり、さらなる処理の高速化が可能となる。この手順は、文献[12]で行っている手順に等しい。



(a) 入力画像



(b) 頭部の検出



(c) 上半身の検出 (検証)

図 4 本手法の流れ.

3.5 監視カメラ固有の配置に基づく重複への対応

上記の方法では、全身での特徴が照合されるため、人物が重複している場合には、遮蔽部分に対して妥当な照合結果が得られなくなる。その結果、頭部が検出できても検証処理で棄却されることになる。この問題を解消するため、複数の検証部位を事前に学習しておき、遮蔽されることがわかった場合に、遮蔽領域が含まれない検証部位を用いて照合を行ない、検証を行なう。ところが、この枠組みを実現するためには、事前に遮蔽される領域が既知である必要がある。

一般に、監視カメラの設置は、人間の頭部より高いところに、下向きに設置される。このとき、カメラに近いものは画像中の下側に、遠くのものには画像上部に撮影される。したがって、手前の人物から決定していけば、遮蔽の有無を認識しながら奥の人物の領域を決定できる。

そこで、本手法では、図 5 に示すように、画像の下に撮影された頭部から人物を検証することで、人物の領域を決定していく。これにより、人物と人物とで重複があっても、既に決定している手前の人物領域による遮蔽を回避した検

証部位によって照合を行うことが可能になり、より安定した検証が可能となる。今回は、図 6 に示す 3 通りの異なる照合部位を事前に学習することで、重複のない照合を可能にする。

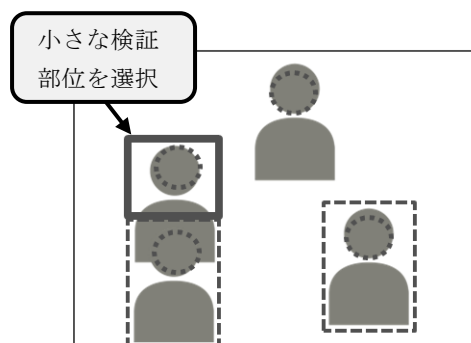


図 5 遮蔽を考慮した検出手順。頭部が下に存在する手前の人物から検証を進め、上部の候補では、手前の人物との遮蔽を考慮して検証部位を決定する。

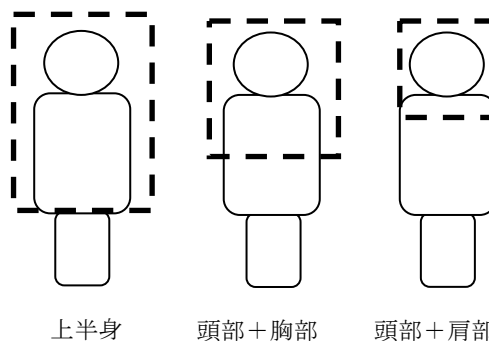


図 6 今回使用した照合部位.

4. 実験

4.1 実験概要

本手法の有効性を示すために、以下について評価した。

- ・従来手法との検出率の比較
- ・異なる解像度間での検出率の比較
- ・遮蔽の有無に対する検出率の比較

評価用のデータセットとして、建物屋内に設置された監視カメラ映像を利用した。このカメラは、高さ約 2.5m の天井から斜め下向きに設置されており、撮影された映像は 160×120 画素、10fps のフレームレートである。今回の実験では、この映像の中からランダムに 180 枚の画像を抽出し、80×60 画素の画像にダウンコンバートすることで低解像度画像を作成した。また、今回の検出対象は、頭部と上半身が撮影画像中に映る位置にいる人物で、かつ、特定の大きさの人物とした。具体的には、人物の顔幅がおよそ 3 画素から 8 画素となる大きさの人物を検出対象とした。この顔幅を人物の高さに換算すると、約 20 画素から約 50 画素にそれぞれ相当する。今回の検出対象である 428 人の顔

幅の分布を図7に示す。また、検出の可否判定には、検出した人物の顔の領域と目視で抽出した真の領域との重複率を採用し、50%以上の重複率がある場合を合格とした。

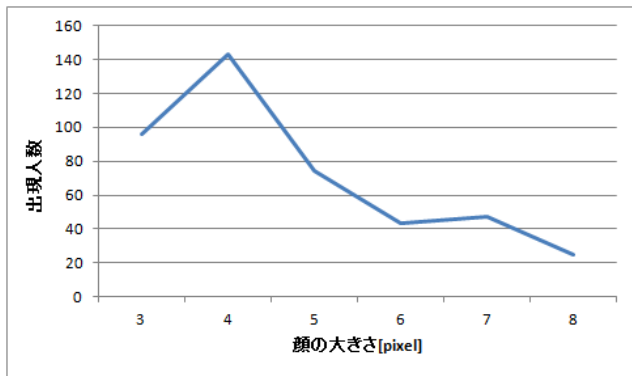


図7 評価画像中の顔の大きさの分布。

4.2 従来手法との比較

最初に、従来手法として HOG 特徴を用いた人物検出と検出率に関する比較を行った。従来手法の検出器には、識別器として Real-boosting を用いた。また、学習における正事例は、同じカメラの別の映像の画像を切り出すことで作成した。

図8に提案手法による検出結果の例を、図9に提案手法と従来手法の検出率をそれぞれ示す。図8では、頭部候補を多数検出し、頭部候補から得られる人物領域候補を上半身の形状特徴で評価することで、画像中の人物を正しく検出していることがわかる。

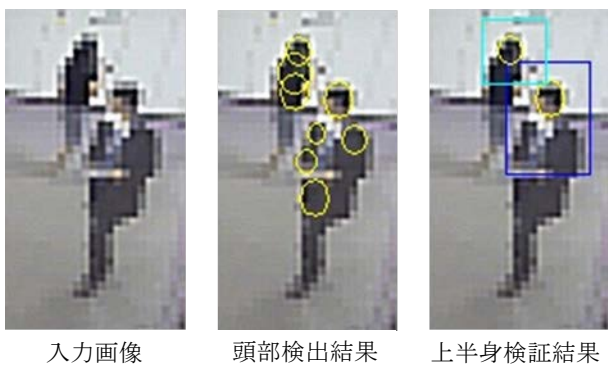


図8 人物検出結果例。

また、図9の結果より、提案手法が顔の大きさによらず、おおよそ70%以上の検出率を達成していることがわかる。一方、誤認識率を同程度として比較した従来手法の検出率は、顔の大きさが大きくなるほど検出率が向上しているが、最大でも50%に満たないことから、提案手法が低解像度の画像に対して高精度な検出を実現していることがわかる。この原因として、今回のカメラのように、斜め上から撮影された映像では、人物の位置によって見え方が大きく異なる

るものの、単一の照合辞書しか与えなかった従来手法では、多くの位置でずれが生じ、性能が低下したと考えられる。一方、提案手法では、見えの変化の少ない頭部と上半身の形状で人物を検出することから、見えの変化に大きな影響を受けることなく検出できたと考えられる。

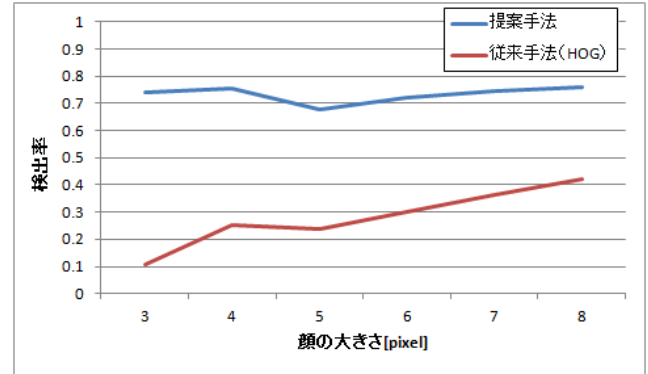


図9 従来手法との比較。

4.3 解像度間での比較

次に、解像度だけが異なる画像を利用して、本手法の有効性を検証した。高解像度画像については、入力画像である160×120画素の画像を直接用いて比較した。

図10に比較結果を示す。この結果より、低解像度画像でも高解像度画像でも同程度の検出結果を示しており、解像度の違いによって検出性能に大きな差が生じないことがわかる。

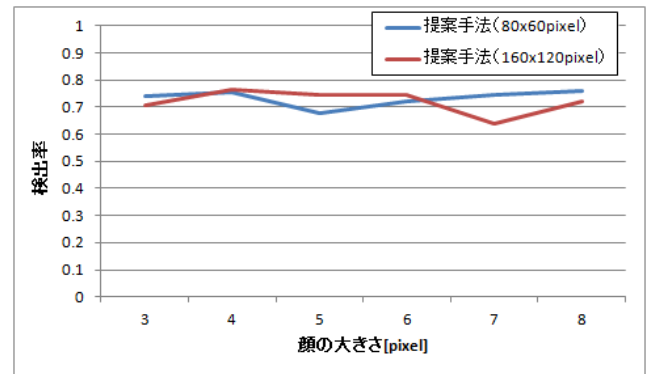


図10 入力画像の解像度による検出率の違い。

4.4 遮蔽の有無に対する比較

最後に、画像上で人同士が重なることによる遮蔽の有無について比較結果を表1に示す。評価画像中には、他の人を遮蔽した人、もしくは他の人に遮蔽された人が195人存在しているが、それら遮蔽に関連した人も67.7%の精度で検出できたことを確認した。これは本手法の複数の検出器による検出が機能することによって、検出できていると考えられる。

表 1 遮蔽状況と検出率.

人物の状態	検出率
遮蔽に関連しない人物	78.1% (182/233)
遮蔽に関連した人物	67.7% (132/195)

4.5 考察

本提案手法では、従来手法[7]で検出できない大きさで撮影された人物高さが約 20 画素の場合でも約 70%以上の検出を実現し、その有効性が確認できた。一方、検出できなかった理由の一つとして、同一の服の色で重なった人の検出ミスがあることが分かった。図 11 に検出に失敗した例を示す。画像上の上下方向に並んでいる 3 人のうち、最も手前にいる人の頭部検出に誤っており、最終的に人物として検出できていないことがわかる。これは、今回利用した頭部検出手法が、背景との差異を評価しているため、背景の服の色と頭部が同化した際に、頭部候補が検出できなかったことに起因する。このようなコントラストの低い背景からの検出精度の向上は今後の課題とする。

一方で、人流の認識における実際の運用では、時間的に連続する画像系列が利用できることから、一度検出できれば追跡により検出することが可能となる。そのため、重なりが無い時刻の画像から人物を検出することができれば、検出率向上が見込める。また、もう一つの未検出の原因として、頭部検出処理で検出された人物候補に対して上半身の検出時に、人物を人物でないと誤判定することがあった。これは、今回評価に使用した画像が、斜め上から撮影したものであるため、撮影位置によって人物の見えの形状が大きく異なり、学習したパターンと一致しないものが存在したためと考える。この問題に対しては、画像上の位置に応じたパターンを複数学習させることで解決できると考える。

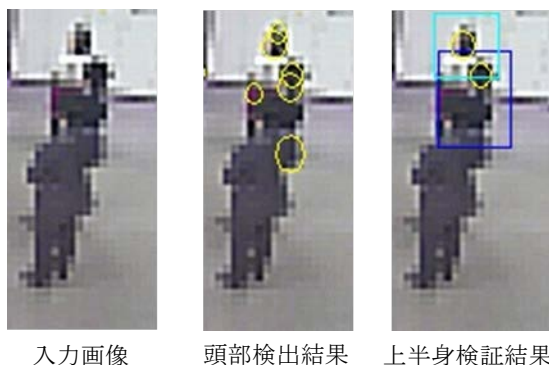


図 11 人物検出の失敗例.

5. おわりに

本稿では、プライバシーの問題に配慮した人流認識のため、既設のカメラを利用した低解像度の静止画像 1 枚から

人物を検出する手法を提案した。低解像度画像では、人物の特定が難しいと同時に、近接したり重複したりする人物の存在により位置決め精度が低下し、人物ひとりひとりの切り出しが困難になる。提案手法では、まず頭部検出フィルタによって頭部候補を検出し、その検出された頭部候補に基づいて、上半身部分の特徴を評価することで、低解像度でも位置決め精度の低下を回避した人物の検出を実現した。延べ 428 人に対する評価実験により、頭部幅が 3 画素から 8 画素の人物の検出率が 70%以上であることを確認した。また、人物同士の重なりに対しても対応できる手法であることを確認した。今後、低解像度映像の時間変化の情報を利用することで、少ない情報量のもとでさらなる高精度化を実現する。

参考文献

- 1) 島崎 康信, 関本 義秀, 柴崎 亮介, 秋山 祐樹, “人の流れによる時間帯別人口と店舗数との相関関係についての研究,” 都市計画学会都市計画論文集, vol. 44, no. 3, pp.781-786, 2009.
- 2) 山下 倫央, 大西 正輝, “オリンピックにおける人の流れの解析,” 情報処理, vol. 55, no.11, pp.1189-1195, 2014.
- 3) 山下 倫央, 副田 俊介, 野田 五十樹, “人流計測による避難誘導効果の実証的検証,” 情報研報, vol. 2009-UBI-24, no. 25, pp. 1-8, 2009.
- 4) H. Sohn, W. D. Neve, and Y. M. Ro, “Privacy Protection in Video Surveillance Systems: Analysis of Subband-Adaptive Scrambling in JPEG XR,” *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 21, no. 2, pp. 170-177, 2011.
- 5) S.-C.S. Cheung, M.V. Venkatesh, J.K. Paruchuri, J. Zhao, and T. Nguyen, “Protecting Privacy in Video Surveillance,” in *Protecting and Managing Privacy Information in Video Surveillance Systems*, Springer, pp. 11-13, 2009.
- 6) K. Tsuji, M. Zheng, E. Segawa, M. Shiohara and T. Morihara, “Robust Human Tracking Using Occlusion-free Images from Multiple Video Cameras,” *Proc. 12th IAPR Conf. on Machine Vision Applications*, pp.532-535, 2011.
- 7) P. Dollar, C. Wojek, B. Schiele, and P. Perona, “Pedestrian Detection: An Evaluation of the State of the Art,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 34, no. 4, pp. 743-761, 2012.
- 8) N. Dalal and B. Triggs, “Histograms of Oriented Gradients,” *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition*, vol.2, pp.886-893, 2005.
- 9) C. Wojek and B. Schiele, “A performance evaluation of single and multi-feature people detection,” *Proc. DAGM Symposium on Pattern Recognition*, pp.82-91, 2008.
- 10) X. Wang, T. X. Han, and S. Yan, “An hog-lbp human detector with partial occlusion handling,” *Proc. IEEE Int. Conf. on Computer Vision*, pp. 32-39, 2009.
- 11) D. Abe, E. Segawa, O. Nakayama, M. Shiohara, S. Sasaki, N. Sugano and H. Kanno, “Robust Small-Object Detection for Outdoor Wide-Area Surveillance,” *IEICE Trans. on Information and Systems*, pp.1922-1928, 2008.
- 12) L. Wang and N. H. C. Yung, “Crowd counting and segmentation in visual surveillance,” *Proc. 16th IEEE Int. Conf. on Image Processing*, pp. 2573-2576, 2009.