

迷子対策のための親子連れ検出

堤 陽菜[†] 中島 克人[‡]

東京電機大学未来科学部情報メディア学科^{†‡}

1 はじめに

自分の子どもが迷子になる経験をする親が多いと言われる事から、モール等の施設において防犯用の監視カメラを活用した迷子発見を支援するシステムは大変有用であろう。そこで我々は、迷子になる前に子供とその傍らにいる親を特定して記録しておき、迷子が発生した際には、はぐれた子からその親を、あるいは、親から迷子を発見する事を容易化する事を提案している。親子を識別する一つの手がかりとして、今回は手を繋いだ親子連れの検出を目標とする。まず物体検出器で全身と頭の検出を行う。次に、検出した全身と頭の比率を元に子供と大人に分類する。そして、子供とその足元から一番近くにいる大人とが手を繋いでいるかを骨格推定技術を用いて判定する。

2 関連研究

Ince らは CCTV カメラを用いて歩行者を子供と大人の 2 つのカテゴリーに分類した[1]。まず、Haar カスケード分類器を用いて全身と頭の検出を行う。次に検出した全身と頭の高さの比率を算出している。移動平均アルゴリズムを用いて閾値の比率を取得し、比率によりを分類している。本研究でもこの分類手法を用いる。

3 提案手法

モール等の施設の防犯カメラの映像から手を繋いでいる親子を認識するためには、まず人物を子供と大人に分類する必要がある。そこで、図 1 のような子供と大人の分類、親子ペアの検出、手繋ぎの検出の三段階の手法によるシステムとした。

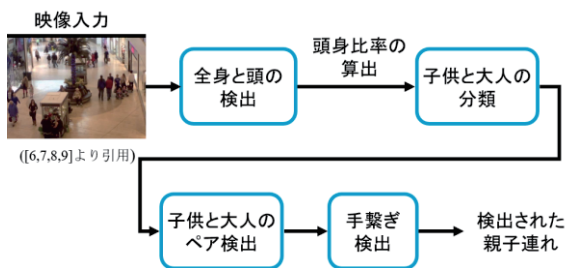


図 1 提案手法の流れ

3.1 子供と大人の分類

映像から、生体情報の一つである頭身比率を用いて子供と大人に分類するためには、まず全身と頭を検出する必要がある。そこで、映像を物体検出器 YOLOv5 [2]に投入し全身と頭を検出し、その両者の高さ比率とある閾値とを比較する事により子供と大人の分類を行う。一般的に防犯カメラの映像は俯瞰方向から撮影されたものとなる。そのため、同じ人が遠くからカメラの真下近くに近づくにつれて、物体検出器が検出する頭身の比率は頭部の厚みによって大きくなる筈である。この比率変化は物体検出器の特性によって変わるため、予備実験として撮影の高さと距離が既知の環境で計測し、子供を分類する閾値を求めることにする。

3.2 頭身比率の計測

モールの防犯カメラの位置を想定し、高さ 7.3m から歩行者の距離 10~29m で撮影し、身長 180cm の男性、160cm の女性、子供を模すために 110cm 程度にしゃがんで歩いた女性に対する物体検出器が出力した頭身比率を計測した(図 2)。子供を分類する閾値については後述する。

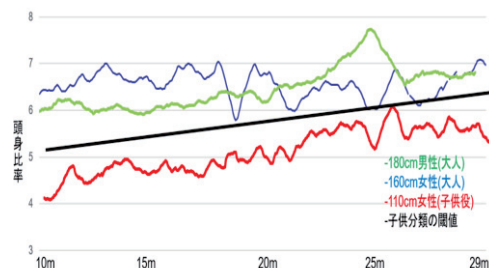


図 2 頭身比率の変化

3.3 親子ペア検出と手繋ぎ判定

子供の分類後、その子供の足元からの一定以内の距離で一番近くに映る大人を子供の保護者と推定し、親子ペアとして検出する。その最大距離は大人の身長¹の 1/2 以下とした。

次に、検出したペアの親子それぞれに対して姿勢推定器 OpenPose [3]を適用し、手を繋いでいるかの判定を行う。なお、遠方からの俯瞰画像のため、多くの場合は手の検出が困難なため、親子それぞれの手首のキーポイントを取得し、手首同士の距離が手を繋げるに十分近いペアを、手を繋いでいる親子として検出する。

Parent-child pair detection to prepare to find lost children

[†] Haruna Tsutsumi · Tokyo Denki University

[‡] Katsuto Nakajima · Tokyo Denki University

4 評価

4.1 全身と頭の検出精度

COCO データセットで学習済の YOLOv5m に対して群衆の中の間人を検出するベンチマークである CrowdHuman [4]データセットで追加学習されたモデルである crowdhuman_yolov5m [5]を全身と頭の検出に使用した。

今回の我々の評価では Mall Dataset [6,7,8,9]の中から手を繋いでいる親子が映っている画像 24 枚に対して 640x480 px から 1280x960 px にリサイズを行った上で検出精度を求めた (表 1)。

表 1 全身と頭の適合率, 再現率(%)

	適合率	再現率
全身	88.8	89.1
頭	85.7	85.4

適合率は全身と頭のどちらに対しても 90%未満にとどまるものの, 誤検出の多くはマネキンや鏡に反射した人影であり, この検出では子供と大人のペアを検出することを目的としているため, 単独の誤検出は問題ないと考えられる。また, 遠くに小さく映る人影は検出ミスとなり, 再現率低下の原因となるが, カメラの近くまで移動してくれば検出可能であると考えられる。

4.2 親子ペア検出の精度

図 2 の計測結果から, 子供と大人の分類の閾値 r は下式(1)のような 1 次式で表して良からう。

$$r = a + b \times x \quad (1)$$

ここで, 距離 x (m), 傾き b , 定数 a (m)であり, 図 2 からおよそ $b = 2/30 = 0.07$, $a = 4.5$ が求まる。

Mall Dataset の 24 枚も図 2 の計測時と同じ俯角および距離で撮影されたものと仮定すると, 遠方で 6.5 頭身, 一番手前で 5.2 頭身が閾値となる。

この閾値による子供の分類精度と, その子供とペアになる大人の検出精度を表 2 に示す。なお, ここでは座っている大人が子供に誤分類された場合等は除外している。

表 2 子供分類とペア検出の適合率, 再現率(%)

	適合率	再現率
子供分類	31.9	71.9
ペア検出	52.9	64.3

子供分類の精度が低い原因として, 閾値決定のための計測時と 24 枚の評価画像でのカメラの設置位置の違いと, 服装や髪型による頭の検出サイズの揺れが考えられる。

4.3 手繋ぎ判定による親子連れ検出の精度

前節で検出された全ペアから大人同士の手繋ぎペアを除外するために, 今回はペアの身長差が 1/3

以上のペアだけを対象とした。なお, これらの画像サイズは小さく, ノイズが手首位置の検出に影響を与えかねないため, OpenPose の入力前にカーネルサイズ 3 のメディアンフィルタを掛けた。

ペアそれぞれの手首位置が求まればその距離を求め, それが大人の背骨の長さの 2/5 以下であれば手繋ぎであると判定した。

手繋ぎ親子, 即ち, 親子連れの検出精度を表 3 に示す。再現率が低い原因の 1 つは, 前段のペア検出率がそもそも低い事, もう 1 つは, 解像度の低さや服装と背景色の類似等によって手首の関節点が検出できない事であった。しかしながら, 適合率は高く, 親子連れの検出に手繋ぎ判定を用いる事は有効と言える。

表 3 親子連れ検出の適合率, 再現率(%)

	適合率	再現率
親子連れ	88.9	33.3

5 考察

全身と頭の検出は, 適合率, 再現率ともに 80%以上であり俯瞰からでも検出可能であることが分かる。人混みのシーンでは人同士が重なってしまうこと等で未検出や誤検出が多発したが, これは俯瞰角度の群衆画像を学習時に増やすことで改善できると考えられる。

頭身比率による子供分類は撮影条件により閾値が変わるため適合率を上げるのは困難が予想されるが, 再現率を重視し, その後のペア検出, 身長差によるふるい落とし, 手繋ぎ判定で親子連れか否かを絞っていけると考えられる。

6 今後の予定

今回は頭身比率や関節推定による手繋ぎ判定を使用した親子連れの検出を行ったが, 手を繋ぐ事なく寄り添っている場合を含めた親子のペア自体を物体検出器で検出する方法も今後検討していく。

参考文献

- [1] O.F. Ince, et al., "Child and adult classification using biometric features based on video analytics," ICIC, 2017
- [2] G. Jocher, "Yolov5," <https://github.com/ultralytics/yolov5>, 2020, 2021/9 参照
- [3] Z. Cao, et al., "Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields," CVPR, 2017
- [4] S. Shao, et al., "CrowdHuman: A Benchmark for Detecting Human in a Crowd," arXiv, 2018
- [5] S. Akkash, "yolov5-crowdhuman," <https://github.com/SibiAkkash/yolov5-crowdhuman>, 2021/9 参照
- [6] C.C. Loy, et al., "From Semi-Supervised to Transfer Counting of Crowds," ICCV, 2013
- [7] K. Chen, et al., "Cumulative Attribute Space for Age and Crowd Density Estimation," CVPR, 2013
- [8] C.C. Loy, et al., "Crowd Counting and Profiling: Methodology and Evaluation," Springer, 2013
- [9] K. Chen, et al., "Feature Mining for Localised Crowd Counting," BMVC, 2012