2 台のカメラを用いた 骨格抽出による移動ロボットの 3 次元人物追跡*

— 追跡復帰機能による人物追跡のロバスト化 —

松 原 慎 也** 池 勇 勳† 梅 田 和 昇††

Three-dimensional Human Tracking of a Mobile Robot
by Skeleton Extraction Using Two Cameras
- Robust Human Tracking with Tracking Return Function -
Shinya MATSUBARA, Yonghoon JI and Kazunori UMEDA

In this paper, we propose a new human tracking method equipped with a function to return tracking a person. Under the stereo framework of fusing the processing results of two cameras, there is a three-dimensional tracking system for a specific person by a mobile robot using OpenPose that can acquire skeletal information using deep learning. We propose a new human tracking system equipped with the function to return tracking a person by template matching that limits the search range by epipolar constraint. The function is to return tracking a person who failed to be tracked by using the image information that was successfully tracked. With this function, even if one camera fails to track a person, the robot can continue to track the person. The effectiveness of the proposed human tracking system is verified by experiments.

Key words: human tracking, stereo camera, autonomous mobile robot, three-dimensional measurement, robust tracking

1. 結 言

近年、自律移動ロボットは工場現場での運搬作業や高齢者向けの介護支援といった、様々な場面での活用が期待されている。これらの自律移動ロボットにおける重要なタスクの一つに、特定人物の追跡が挙げられ、該当人物を安定して追跡する性能の向上に向けた研究が数多く行われている^{1)~7)}。この追跡機能を実現するための研究には、ステレオカメラを用いることが多く、取得した三次元情報を用いることで追跡ロボットの追跡を実現している⁸⁾。

我々は新たなステレオ視の枠組みとして、シーン全体の距離画像は取得せず、単眼カメラ 2 台がそれぞれでタスクに応じた処理を行った後にその処理結果を融合することで必要な対象のみに関しての 3 次元情報を取得するという枠組みを提案した⁹⁾。画像全体で距離を求める必要がないことから、誤マッチングを抑制し、必要なタスクにとって最適な 3 次元情報を得ることが可能である。具体的なタスクとして特定の人物の 3 次元追跡を取り上げる。この枠組みの下で、人物検出に OpenPose¹⁰⁾を用いた 3 次元追跡を提案した¹¹⁾。しかし、ここで提案した追跡手法は、2 台のカメラのうちどちらか片方で特定の人物の追跡に失敗してしまうと、距離計測が行えず移動ロボットによる 3 次元人物追跡が失敗してしまうという課題があった。

そこで本論文では、2 台のカメラのうちどちらかが追跡に成功していれば移動ロボットによる 3 次元人物追跡が行えるよう追

跡手法を改善し、2 台のカメラを用いて追跡を復帰させる機能を備えた新たな追跡システムを提案する¹²⁾。

2. 追跡手法の概要

人物検出、人物識別、距離計測、ロボット制御を毎フレーム繰り返すことで、人物追跡を実現する。図 1 に示すように、本追跡手法では、人物検出に単一画像から複数の人物の骨格情報をリアルタイムに検出することができる OpenPose を用いる。OpenPose とは、CNN (Convolution Neural Network) を用いて単一画像から複数の人間の体や顔などの複数の骨格点をリアルタイムに検出することができるライブラリである。この章では、OpenPose を用いた人物追跡手法¹³⁾について説明する。人物追跡の全体の流れを図 2 に示す。

Fig.1 Detection of human using OpenPose

* 原稿受付 令和 3 年 5 月 14 日

掲載決定 令和 3 年 11 月 15 日

** 中央大学大学院 (東京都文京区春日 1-13-27)

† 北陸先端科学技術大学院大学 (石川県能美市旭台 1-1)

†† 正 会 員 中央大学 (東京都文京区春日 1-13-27)

Fig.2 Flow chart of human tracking

2.1 初期テンプレートの作成

まず初めに、追跡対象者には挙手をしてもらう。OpenPoseで手を挙げている人物を検出し(図 3(a))、その人物の領域画像を初期テンプレート画像とする(図 3(b))。このテンプレート画像を HSV 画像に変換し、HSV のうち明るさや照明変動の影響を受けにくいとされる色相 (H) のみで表された色ヒストグラム(図 3(c))を本手法では用いる。このテンプレート画像は後の人物識別を行う際に用いる。

2.2 OpenPose を用いた人物候補領域の検出

人物候補領域の検出にも OpenPose を用いる。図 4 に示すように、OpenPose により得られた骨格情報(図 4(a))のうち、左右の肩と足首の座標を用いて人物領域画像(図 4(b))を切り取る。人物候補領域を赤紫色の矩形で表す。

2.3 人物候補領域の分割

図 5 に示すように、OpenPose によって検出した人物候補領域を分割する。取得した骨格座標のうち腰の座標を用いて上半身と下半身を分割することで、上着と下着それぞれで色の比較が可能となる。

2.4 色比較による人物特定

服の色の違いから人を識別し、追跡対象者を 1 人決定する。検出した人物候補領域画像の色ヒストグラムと作成したテンプレート画像の色ヒストグラムを比較し、類似度を計算する。類似度計算には正規化相互相関を使用する。現フレームの色相ヒストグラムとテンプレート画像の色相ヒストグラムの類似度 d を以下の式で算出する。

$$d(H_1, H_2) = \frac{\sum_l (H_1(l) - \bar{H}_1)(H_2(l) - \bar{H}_2)}{\sqrt{\sum_l (H_1(l) - \bar{H}_1)^2 \sum_l (H_2(l) - \bar{H}_2)^2}} \quad (1)$$

H_1 はテンプレート画像のヒストグラム、 H_2 は人物候補領域のヒストグラムである。類似度が一定の値を超えている場合、その人物候補領域に対応する人物を追跡対象者と決定する。

2.5 距離情報を用いた追跡

ここまでの処理を 2 台のカメラで行い、それぞれで追跡対象者を特定する。その後、図 6 に示すように追跡対象人物の領域の中心点(赤点)に着目し、2 台のカメラの視差(横方向のずれ)により三角測量の原理を用いて追跡対象までの距離 Z を以下の式より求める。

$$Z = \frac{b \cdot f}{\delta(u_l - u_r)} \quad (2)$$

ここで、 u_l , u_r はそれぞれ左右カメラの中心点の横座標であり、 $u_l - u_r$ は左右カメラでの視差である。 b は左右カメラの基線長、 f はカメラの焦点距離、 δ は画素間隔である。以上で求めた人物と移動ロボット間の距離と位置の情報を用いてロボットを制御する。これを毎フレーム繰り返すことで追跡を可能にしている。追跡の概念図を図 7 に示す。

Fig.3 Histogram generation for template image: (a) Raised hands detected (b) template image and (c) color histogram

Fig.4 Detection of human region: (a) detection of human body key points by OpenPose and (b) examples of region used to calculate histogram similarity

Fig.5 Separation of human region

Fig.6 Examples of points used for distance calculation: (a) left camera image and (b) right camera image

Fig.7 Conceptual diagram of target tracking with a mobile robot

3. 追跡復帰機能

3.1 追跡失敗の原因

図 8 に示すように、色ヒストグラム同士の比較には、前フレームの追跡対象者の人物領域画像をテンプレート画像として用いる。テンプレート画像は毎フレーム更新する。このとき、現フレームで決定した人物領域画像を次フレームでのテンプレート画像とするため、フレーム間で照明変動が起こると、色ヒストグラムが変化し、追跡対象者を見失う場合がある。この課題を解決するために、提案手法では追跡復帰機能を搭載する。

3.2 追跡復帰手法

エビポーラ拘束により探索範囲を限定したテンプレートマッチングによって追跡復帰を実現する。2 台のカメラのうち片方のみが追跡に失敗した場合、三角測量で距離を測ることができず、移動ロボットでの追跡は失敗してしまうという課題があった。そこで、この場合にも、テンプレートマッチングを用いて、追跡に成功しているカメラの人物領域画像からもう一方の追跡に失敗しているカメラの人物領域画像を探索することで追跡の復帰を試みる。このとき、誤対応を減らし探索時間を短縮させるために、テンプレートマッチングを行う際にエビポーラ拘束を用いる。図 9 にこの手法の流れを示す。

Fig.8 Similarity calculation between adjacent frames

Fig.9 tracking return function using template matching

4. 実験

4.1 実験概要

提案手法の有効性を検証するために、2 種類の実験を行った。図 10 に示すように、実験は比較的明るく照明変動がある屋内環境で行った。1 つ目の実験は、追跡復帰機能を搭載する手法と追跡復帰機能を搭載しない手法を比較した。2 つの手法での実験環境はどちらもオクルージョンが 5 回発生する条件の元、事前に撮影した動画を用いて（オフラインで）実験を行った。2 つ目の実験は、オンライン環境かつより長い距離で、移動ロボットによる 3 次元人物追跡を行った。両実験とも被追跡者は一定の速度でまっすぐに歩行動作を行った。実験で使用した装置を図 11 に示す。カメラには ELECOM 製 UCAM-DLI500TN、移動ロボットには AdeptMobileRobots 社製の Pioneer3-AT を使用した。カメラの解像度は 640×480 pixel である。図 12 に示すように、移動ロボット上に 2 つのカメラを平行に並べることで平行ステレオを実現した。基線長は 95mm とした。

Fig.10 Experiment environment

Fig.11 Experimental equipment: (a) web camera and (b) mobile robot

Fig.12 Parallel stereo constructed by arranging two inexpensive web cameras

4.2 実験結果

1つ目の追跡実験の結果を表1に、2つ目の追跡実験の結果を表2に示す。

追跡復帰機能の有効性を検証するための1つ目の実験の様子を図13に示す。図14~16のように追跡対象者と移動ロボットの間を非追跡対象者が横切った場合にも追跡を継続して行うことができた。しかし、表1からわかるように、追跡復帰機能未搭載の手法では11フレームで移動ロボットによる追跡に失敗していた。これは全て2台のカメラのうちの片方が追跡に失敗してしまい距離計測が行えないことが原因であった。図17に追跡失敗の例を示す。左カメラでは問題なく追跡に成功しているが、右カメラでは照明の変化によって画像中の服の色の見え方が変わってしまい、追跡に失敗していることがわかる。ここでは追跡対象者の人物領域をピンク色、追跡対象者以外の人物領域を黄緑色で表している。つまり、右カメラでは追跡対象者以外の人物と判定してしまっている。これに対し、追跡復帰機能を備えた提案手法を用いることで、追跡に失敗していた11フレーム全てで追跡復帰に成功した。図18には、提案手法を用いて追跡の復帰に成功したフレームの例を示している。ここでは右カメラで一度追跡に失敗したが、左カメラで追跡に成功した人物領域画像をテンプレート画像としたことで、追跡復帰が成功した。また、移動ロボットの動作を含めた評価を図19に示す。追跡対象者と移動ロボット間の距離は2.0~3.5mの範囲内で保つことができおり、リアルタイムでの追従が可能である。赤く矩形で囲った部分は非追跡対象者が横切ったフレームを示しており、未検出の際はロボットが停止してしまうため、距離が離れてしまうことがわかる。

追跡距離を長めにとった2つ目の実験の様子を図20に示す。追跡復帰機能未搭載の手法では左右のカメラ合計32フレームで追跡に失敗していた。追跡失敗のうち、3フレームがカメラ2台ともでの失敗であった。しかし、29フレームはいずれか1台のカメラで追跡が行われていたため、追跡復帰処理を行うことでそれら全てのフレームで追跡復帰に成功した。追跡復帰に成功したフレームの様子を図21に示す。よって、3フレームを除いた378フレームで移動ロボットによる追跡が成功し、追跡成功率99.2%という結果を得た。なお、ここでの追跡成功率とは、左右のカメラのうち追跡成功数の少ない方を全フレーム数で割ったものを表す。また、本実験では左右のカメラの画角に多少のずれがあっても追跡を継続して行うことができることも示すことができた。

本追跡システムを用いることにより、2台のカメラ両方で追跡対象者を見失った場合以外では追跡を復帰させること成功した。図22に2台のカメラ両方で追跡対象者を見失ったフレームの左右のカメラ画像を示す。このフレームで追跡に失敗してしまったのは、追跡対象人物の真上に照明がきたため人物領域画

像の色ヒストグラムが大きく変化してしまったことが原因であるとする。

Fig.13 Scene of human tracking in experiment 1: (a) left camera and (b) right camera

Fig.14 Before a person crosses: (a) left camera and (b) right camera

Fig.15 While a person is crossing: (a) left camera and (b) right camera

Fig.16 After a person crosses: (a) left camera and (b) right camera

Fig.17 A frame that failed to track a person using a method without a return function: (a) tracking success in left camera and (b) tracking failure in right camera

Fig.18 A frame that succeeded in returning to track a person with the proposed method: (a) tracking success in left camera and (b) tracking return in right camera

Table 1 Target tracking success rate in experiment 1

	Without return function	With return function
Total number of frames	164	164
Number of tracking failure frames	11	11
Number of tracking return frames	-	11
Number of successful tracking frames	153	164
tracking success rate %	93.3	100

Fig.19 Distance between tracked person and mobile robot

Fig.20 Scene of human tracking in experiment 2: (a) left camera and (b) right camera

Fig.21 A frame that succeeded in returning to track a person with the proposed method: (a) tracking success in left camera and (b) tracking return in right camera

Fig.22 A frame that failed to track a person: (a) left camera and (b) right camera

Table 2 Target tracking success rate in experiment 2

	Proposed method
Total number of frames	381
Number of tracking failure frames	32
Number of tracking return frames	29
Number of successful tracking frames	378
tracking success rate %	99.2

5. 結 言

本論文では、2 台の単眼カメラの処理結果を融合した移動ロボットによる特定人物の 3 次元追跡を提案した。そしてエビポータ拘束により探索範囲を限定したテンプレートマッチングによる追跡復帰機能を搭載することで、ロバストな人物追跡を実現した。この機能により、2 台のカメラのうちどちらかのカメラで追跡に成功していれば移動ロボットによる 3 次元人物追跡が実現できることを実験で示し、2 台のカメラを用いることによる新たな利点を生み出すことができた。しかし、カメラ 2 台両方で追跡に失敗した場合、追跡の復帰は不可能である。今後は、この時の対処法としてカルマンフィルタによる位置予測を手法に取り入れることを考えている。また、曲がり角など今の手法では追跡が困難な場所での追跡の実現を目指す。

謝 辞

本論文は、故 本多明彦氏の研究を発展させたものです。先駆的な優れた研究を行い、本論文の礎を築いた故 本多氏に心より感謝します。

参 考 文 献

- 1) K. Koide, J. Miura, "Identification of a specific person using color, height, and gait features for a person following robot," *Robotics and Autonomous Systems*, **84**, (2016) 76.
- 2) C. Hu, X. Ma, and X. Dai, "A Robust Person Tracking and Following Approach for Mobile Robot," *Proc. of the 2007 IEEE Int. Conf. on Mechatronics and Automation*, (2007) 3571.
- 3) J. H. Lee, T. Tsubouchi, K. Yamamoto, and S. Egawa, "People Tracking Using a Robot in Motion with Laser Range Finder," *Proc. of the 2006 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, (2006) 2936.
- 4) J. Fritsch, M. Kleinhagenbrock, S. Lang, G. A. Fink, and G. Sagerer, "Audiovisual Person Tracking with a Mobile Robot," *Proc. of the 2004 Int. Conf. on Intelligent Autonomous Systems*, (2004) 898.
- 5) M. Hebert, "Active and Passive Range Sensing for Robotics", *Proc. of the 2000 IEEE Int. Conf. on Robotics and Automation*, (2000) 102.
- 6) B. X. Chen, R. Sahdev, and J. K. Tsotsos, "Person following robot using selected online ada-boosting with stereo camera," *2017 14th conference on computer and robot vision (CRV)*, IEEE, (2017) 48.
- 7) M. Danelljan, et al, "Accurate scale estimation for robust visual tracking," *British Machine Vision Conference, BMVA Press*, (2014).
- 8) Y. Isobe, G. Masuyama, K. Umeda, "Occlusion handling for a target-tracking robot with a stereo camera", *Robomech Journal*, **5**, 4, (2018) 1.
- 9) 本多 明彦, 梅田 和昇, "二台のカメラのトラッキングの融合による

- 移動ロボットの 3 次元人物追跡”,ロボティクス・メカトロニクス講演会, 1A1-M09, (2018).
- 10) Z. Cao, T. Simon, S. Wei, Y. Sheikh, “Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields,” Proc.of 2017 IEEE conference on computer vision and pattern recognition, (2017) 7291.
 - 11) S. Matsubara, A. Honda, Y. Ji and, K. Umeda, “Three-dimensional Human Tracking of a Mobile Robot by Fusion of Tracking Results of Two Cameras,” Proc. of 2020 21st International Conference on Research and Education in Mechatronics (REM), (2020).
 - 12) 松原 慎也, 梅田 和昇, “2 台のカメラを用いた移動ロボットの 3 次元人物追跡における追跡復帰機能の実現”,動的画像処理実利用化ワークショップ DIA2021 講演論文集, IS4-1, (2021).
 - 13) 松原 慎也, 池 勇勲, 梅田 和昇, “二台のカメラを用いた OpenPose の骨格抽出による 移動ロボットの 3 次元人物追跡” 日本機械学会ロボティクス・メカトロニクス講演会講演論文集, 1P2-104, (2020).