

日本機械学会論文集 Transactions of the JSME (in Japanese)

ステレオカメラを搭載した移動ロボットによる 照明変動に頑健な追跡対象の認識

磯邊 柚香*1, 増山 岳人*2, 梅田 和昇*3

Robust target recognition for a mobile robot using a stereo camera under illumination changes

Yuzuka ISOBE^{*1}, Gakuto MASUYAMA^{*2} and Kazunori UMEDA^{*3}

 *1 Course of Precision Engineering, School of Science and Engineering, Chuo University 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan
*2.*3 Department of Precision Mechanics, Faculty of Science and Engineering, Chuo University 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan

Received 30 June 2015

Abstract

In this paper, we propose a target-tracking system for a mobile robot equipped with a stereo camera. Mobile service robots with the ability to track a specific person in dynamic environments have been required. In such environments, varying illumination and the presence of multiple people are challenges to carrying out target tracking. Color and location information is used for the target's features, that are useful for distinguishing a target from the other people. However, color information is not resilient to illumination changes. On the other hand, location information might be infeasible when non-target people present in the environments. Therefore, it is necessary to combine each information according to the situations, in which either information is infeasible to use as the feature. In order to make the system robust to varying illumination and presence of non-target people, a parameter of illumination changes is introduced in this paper. The parameter is defined using automatically adjusted white balance. The color and location features are weighted based on white-balance changes to determine a target. The effectiveness of the proposed system is verified through target-tracking experiments in outdoor environments. It is demonstrated that the proposed method can successfully recognize a target in the environments where the lighting condition changes extremely and non-target people present.

Key words : Target tracking, Mobile robot, Stereo camera, Illumination changes, Outdoor environment

1. 緒 言

特定の人物の追跡は、自律移動ロボットに必要な能力の一つである. 自律移動ロボットの運用されうる環境は 多岐に渡り、この能力も、オフィスやショッピングセンター(Five Elements Robotics, 2015)、軍事区域(Raibert et al., 2008)、ゴルフコース(Stewart Golf Limited, 2015)にいたるまで想定される運用環境は幅広い(Doog Inc., 2015). これらの動的環境において人物追跡を実現させるためには、ロボットに搭載されたセンサによる周囲の環 境の認識が必要となる(Katz et al., 2008). ロボットに環境認識の機能を与えるために、数々のセンサが用いられ る(Hebert, 2000)(Mattos and Grant, 2004)(Lee et al., 2006).

人物追跡のためには、センサの中でもカメラが広く使用されている.これは、カメラから得られる色情報が、

*2 中央大学理工学部精密機械工学科

No.15-00368 [DOI:10.1299/transjsme.15-00368], J-STAGE Advance Publication date : 10 November, 2015

^{*1} 中央大学大学院理工学研究科精密工学専攻(〒112-8551 東京都文京区春日 1-13-27)

^{*3} 正員,フェロー,中央大学理工学部精密機械工学科

E-mail of corresponding author: isobe@sensor.mech.chuo-u.ac.jp

レーザレンジファインダやレーダ,超音波センサなどから得られる情報に比べ,追跡対象の特徴として有用である ことによる.得られた色情報のみを用い追跡対象領域を抽出し人物追跡を行う手法も多く提案されている(Doi et al., 2001)(Kwon et al., 2005)(Calisi et al., 2007)(Tsalatsanis et al., 2007).しかし,色情報は照明変動の影響を 受けやすいため,これらの手法には照明環境が変化する状況において誤追跡が容易に起こりうる問題がある.色 情報のみを用いた研究に対して,Huら(Hu et al., 2007)は追跡対象の特徴を複数使用した手法を提案している. この研究では,人物形状のエッジ情報と追跡対象の色情報を用いて,追跡を実現している.また,Chakravartyら (Chakravarty and Jarvis, 2006)は、レーザレンジファインダによる人物検出と全方位カメラから得られる色情報を 基にした追跡手法を構築している.これらの研究においては,追跡対象の領域抽出の際に色情報以外も用いてい るが,追跡環境中の照明条件の大きな変化への対応は実現されていない.

Takemura らの手法(Takemura et al., 2009)では、ステレオカメラとレーザレンジファインダを用い、双方から 得られる追跡対象の特徴を組み合わせたアプローチがとられている.これにより、屋外の環境においても人物追 跡が可能となっている.しかし、太陽光の照射方向が多様に変わる環境における追跡には至っていない.複数の センサを用い、様々な特徴量を組み合わせた手法は、Fritschら(Fritsch et al., 2004)によっても提案されている. この研究では、顔の特徴、胴部、声音、および脚部を特徴としているが、追跡というタスクにおいては各特徴を 検出可能な環境が大きく制限されることが課題となっている.また、2種類のカメラを搭載した人物追跡ロボット の開発も行われている(Cielniak et al., 2007).人物検出処理ではサーマルカメラから得られる熱画像を用い、検 出された人物領域においてカラーカメラにより照明変動に強い色情報を抽出する.この手法ではオクルージョン を検出する処理も同時に行うため、複数の人物が存在する混雑した環境でも追跡が行える.しかし、サーマルカ メラによる人物検出は外界の温度に依存するため、屋外環境への適用は困難である.

これらの色情報を特徴とした手法に対し,Satake ら(Satake et al., 2013)は,照明変動の問題を考慮し色情報以 外を特徴量として人物検出,追跡対象判別を行う手法を提案している.視差画像中の物体の形状を基に人物検出 を行い,人物の着ている服のデザインから得られる SIFT(Scale-Invariant Feature Transform)特徴量により追跡対 象を判別する.照明変動に対応した手法と言えるが,特徴量抽出やレジストレーションにおける処理時間が課題 である.また,Petrovic ら(Petrovic et al., 2013)により,ステレオカメラから得られる3次元情報のみを用いた 手法も提案されている.照明環境に依存しない3次元情報により,屋内外の照明変動のある環境で人物追跡実験 が行われている.しかし,3次元情報のみを追跡対象の特徴としているため,複数の人物が存在する環境での運用 は実現されていない.

本論文では,照明変動があり複数の人物が存在する環境における,移動ロボットによる人物追跡手法を提案する.ステレオカメラを用い,取得したカラー画像と視差画像を利用して人物追跡を行う.照明変動に応じて色情報と位置情報に重みづけをした評価値を算出し,検出した人物から追跡対象を判別する.屋外環境における実験により,提案手法を用いた追跡対象検出の性能を評価する.

2. 提案 手法

2.1 手法の流れ

人物追跡の流れを図1に示す.手法にはステレオカメラを用いる.これは、ステレオカメラが、同一環境下に おいて追跡対象の特徴として有用なカラー情報と照明変動の影響を受けにくい3次元情報とを同時に取得するこ とが可能であり、またカメラと距離センサを組み合わせる場合よりも同期等を考慮する必要がなく処理が簡便な ためである.まず、ステレオカメラより得られる視差画像を利用して人物検出を行う.検出された各人物領域に 対して、色情報と図2に示す3次元のロボット座標系での位置情報を抽出する.予め取得した追跡対象の色情報、 カルマンフィルタを用いて予測した位置情報と、それぞれの人物の各情報とを照合し、追跡対象と各人物の相違 度を算出する.照明環境の変化に対応するために、照明変動に応じて各情報の相違度への重みを変化させる.検 出された追跡対象人物とロボットの距離と方位に基づいて、ロボット制御を行う.以上の処理を毎フレーム繰り返 すことで、移動ロボットによる人物追跡を実現する.





Fig. 2 Coordinate system of the proposed method

Fig. 1 Flow chart of the proposed system

2.2 従来手法における問題点

我々はこれまで,様々な色情報の中で照明環境の変化に強い色相と彩度を追跡対象の特徴として人物追跡を行う手法を提案してきた(Isobe et al., 2014).しかし,照明変動の小さな環境における追跡は可能となったが,急激な照明条件の変化には対応できていない.また,図3のように短時間にホワイトバランスが大きく変化する環境においても,フレーム間での追跡対象の対応付けや他の人物との区別は困難となる.同図において,図3(a)の次フレームが図3(b)であり,フレーム間の時間差は約0.06sである.

また,ホワイトバランスの変化による追跡対象人物の色情報の変化を図4に示す.図中で青線は,従来手法にお ける各画像中の正しい追跡対象の色情報と予め保存した追跡対象の色情報の相違度を表す.赤線は,各画像での 追跡開始時からのホワイトバランスの変化量を示す.使用したステレオカメラ(Bumblebee2, Point Grey Research) において,ホワイトバランスは,それぞれ 1024 段階の赤ゲインと青ゲインにより調節される.ホワイトバランス の変化量は,追跡開始時と各フレームでのホワイトバランスの,それぞれのゲインの変化量の和により算出され る.同図とカラー画像の出力例(図5)より,ホワイトバランスの変化が大きいときに照明環境も変化する傾向が あると考えられる.

以上より,提案手法においては,ホワイトバランスの変化により照明環境の変化が示されると仮定する.そこで,ホワイトバランスの変化を照明変動を表すパラメータとして用いることで,様々な照明環境における追跡の 実現を図る.

2.3 人物検出

提案手法において,人物の検出は2つのフェーズからなる.3次元情報を利用して人物候補領域を抽出し,視差 画像を用いて候補の中から人物を検出する.

2.3.1 人物候補領域の抽出

画像中に存在する各物体の領域を抽出するために,取得した視差画像に対して,距離画像セグメンテーション (Ubukata et al., 2010)を適用する.まず,視差画像中の各画素の持つ3次元情報を俯瞰平面にプロットする.この 俯瞰平面を投影平面,プロットされた各点を投影点とする.提案手法において,投影平面をX方向は各距離にお いてステレオ計測可能な範囲,Y方向はYmin~Ymax mの範囲として与える.図6(a)の入力画像に対して,投影結



(a) The previous frame

(b) The next frame

Fig. 3 Example of extreme white-balance changes. The interval of these frames is about 0.06 s. However, the white balance extremely changes. It becomes difficult to detect a target based on color infomation.



Fig. 4 The influence of white-balance changes on color information. The blue line shows degrees of dissimilarity between the color information of a calculated target on each frame and that of a predetermined target, and the changes in white balance are given as a red line. It is seen that, as the white-balance changes severely, the dissimilarity tends to change as well.



Fig. 5 Varying illumination scenes. The frame numbers are corresponding to the values of the horizontal axis of Fig. 4

果は図 6(b) として示される.次に,投影平面を 0.1 m× 0.1 m セルに分割し,各セルに含まれる投影点の点数を 算出する.なお,このとき,各セルに含まれる投影点数に距離の2乗に比例する重みづけを行い,距離に応じた1 セルに含まれる点数の減少を補正する.各セルにおける投影点数のヒストグラムを図 6(c) に示す.赤いセルほど 投影点数が多く,青いほど点数が少ないことを表す.投影点数の多いセルを抽出するために,投影平面上の各セ ルに対して投影点数に関する閾値処理を行う.投影平面上で補正後の投影点数が閾値以上のセルを抽出し,それ らのセルに対して8近傍のラベリングを行う.ここで,複数の候補領域同士の距離が近いときにセル群が結合し, 一つの領域とみなされてしまうのを防ぐために Mean-shift クラスタリングを施す. Mean-shift ベクトルを式(1)の



(c) Histogram of projected pointsin each cell

(d) Labeled cells

Fig. 6 Process of setting search regions. 3D infomation is given, simultaneously with a color image. The infomation is project onto an overlooked plane (*projected plane*). By dividing the plane and calculating projected points in each cell, the histogram of the points is executed. In (c), the more points a cell includes, the closer to red the color becomes. In order to detect candidate regions of human, the cells which include many points are extracted and labeled.

ように表し、点群の密度分布の重心位置を求めていく.

$$\boldsymbol{m}(\boldsymbol{\nu}) = \frac{\sum_{c \in rectangle} \boldsymbol{P}_c p'(X_c, Y_c)}{\sum_{c \in rectangle} p'(X_c, Y_c)} - \boldsymbol{\nu}$$
(1)

式中で*v*は重心位置を,カーネル(重心計算を行う範囲)として用いる*v*を中心位置とする矩形を rectangle と表 す.また,任意のセル*c*における位置ベクトルを*P_c*とし,セル中の投影点数を $p'(X_c,Y_c)$ とする.*v*の初期値を各 セル群内に一様に配置し, $v'^{t+1} = v' + m(v')$ を逐次計算することで,重心位置を算出していく.得られた重心位置 を中心とする rectangle により,セル群の統合・分割をした結果が図 6(d) である.同図より,ラベル付けされたセ ル群のうち,人物領域に対応するものは一定の大きさを持つことがわかる.各セル群の大きさ,及び各セル群に 対応する物体の高さ情報にも閾値処理を施す.投影平面上で X 方向を *C_x*セル,Y 方向を *C_y*セル,高さに関して は 1.3~1.9 m を閾値として,人物の候補となる領域を抽出する.この抽出された領域を人物候補領域とする.

2.3.2 人物領域の抽出

抽出された各候補領域に対応する視差画像を2値化することで,各物体のシルエットを取得する.シルエット の取得例を図7に示す.図7(a)中で矩形で表される候補領域に対して,視差情報より物体の存在する領域のみを 白画素として表したものが図7(b)である.次にこの2値画像を低解像度化し,低解像度画像の各行,各列に含ま れる白画像のピクセル数を示すヒストグラムをそれぞれ作成する(図8).なお,ヒストグラムの作成には,人物 の足から肩までの範囲を用いる.人物の肩の位置は,セル群の右左端のセルの持つ高さ情報より算出する.この 範囲でヒストグラム作成を行うと,人物の背中の広さや上半身と比較した下半身の細さから,行方向に関しては, 低解像度画像の上方で頻度が高くなり,下方で頻度は低くなる.また,列方向に関しては,両足のある列ではヒス トグラムの頻度は高く,それ以外では頻度が低く示される傾向がある.これらの傾向に人物の特徴が表れていると 考え,頭部を除いた範囲を使用する.事前に複数の人物画像から作成したヒストグラムの平均値を算出し,追跡 処理中に検出した人物候補から得られるヒストグラムと比較することで,候補領域から人物領域を抽出する.各 ヒストグラムの比較には,類似度式(2)を用いる.

$$R_{\text{SAD}} = \sum_{j=1}^{N} |I_r(j) - T_r(j)| + \sum_{i=1}^{M} |I_c(i) - T_c(i)|$$

(2)



(a) Input image



(b) Binary image

Fig. 7 Example of extraction of person's silhouette. A region which is enclosed by a red rectangle shows a candidate region. In a disparity image, the pixels corresponding to a candidate region are showed as white, and the other pixels are showed as black. Then, the executed binary image shows human silhouette.



Fig. 8 Histograms extracted from low resolution binary image

ここで, N, M は行, 列方向のヒストグラムの要素数, $I_r(j)$, $I_c(i)$ は取得したヒストグラムの各方向の j, i 番目の要素の値, 同様に $T_r(j)$, $T_c(i)$ はテンプレートヒストグラムでの各方向での各要素の値である.式(2)の値が閾値よりも小さい探索領域を人物領域とする.

2.4 追跡対象人物の判別

色情報と位置情報を追跡対象の特徴とし、検出されたそれぞれの人物の各情報との相違度を算出し、追跡対象 人物を判別する.相違度算出の際に、照明変動に応じて各情報への重みを変化させることで、照明条件が変化す る環境への対応を図る.色情報には、抽出された人物シルエット(図7(b))に対応するカラー画像中の画素を用 いる.従来手法と同様に、人物領域中の画素値を RGB 色空間から HSV 色空間の値に変換し、色相と彩度に関す るヒストグラムを作成する.検出された人物の色ヒストグラムと予め取得した追跡対象の色ヒストグラムとの相 違度 *R_{color}* は、以下の式により与えられる.

$$R_{color} = \sqrt{1 - \sum_{h} \sum_{s} \sqrt{H_{input}(h, s) H_{template}(h, s)}}$$
(3)

 $H_{input}(h,s)$, $H_{template}(h,s)$ はそれぞれ入力画像, テンプレート画像における色相 h, 彩度 s での正規化された頻度 を表す. なお, 色相 h, 彩度 s はそれぞれ h_{max} 段階, s_{max} 段階の値に分割した値を使用する.

また,位置情報も追跡対象の特徴として用いる.位置情報として,図2に示すロボット座標系のX,Yについて評価を行う.各人物領域の位置は,投影平面(図6(d))における各領域に対応するセル群の重心位置により表される.追跡対象の位置と速度を状態変数としてカルマンフィルタを適用し,前フレームの追跡対象の位置から現

在のフレームでの位置を予測する.なお,カルマンフィルタのモデルには等速直線運動モデルを用いる.検出さ れた各人物の位置と,予測した追跡対象の位置の相違度 *E* を次式で定義する.

$$E = k \sqrt{(X_s - X_e)^2 + (Y_s - Y_e)^2}$$
(4)
(X_s,Y_s) は人物の位置, (X_e,Y_e) は追跡対象人物の予測位置を示す. ここで, k は E を無次元化するための定数で

R_{color} と*E* をそれぞれ色情報と位置情報に関する評価値とする.これらの評価値を照明変動に応じて重みづけした相違度 *D* により、検出された人物と追跡対象人物の比較を行う.

$$D = \begin{cases} (1 - \alpha)R_{color} + \alpha E & (\alpha < \alpha_{th}) \\ E & (otherwise) \end{cases}$$
(5)

Dの計算結果に閾値処理を施し,領域の絞り込みを行う. 閾値よりもDの値が小さい人物領域が存在するとき,その中で最も値の小さい領域を追跡対象人物の領域とする. ただし, α は照明変動を表すパラメータでホワイトバランスの変化量 |W| を用いて $\alpha = p|W|$ と定める. また, p は定数で, $0 \le \alpha \le 1$ となるように定められる. 式(5) 中で, α_{th} は照明変動による色情報への信頼度の閾値である. ホワイトバランスの変化が大きく色情報も大きく変わるとき, α は α_{th} 以上の値となる. $\alpha \ge \alpha_{th}$, およびDの算出式の関係を以下に詳述する. (1) $\alpha < \alpha_{th}$ の場合

ホワイトバランスの変化に応じて,各評価値に重みづけを行う.これは,照明条件の変化により,色情報への 信頼度が変化することによる.照明変動が小さいときは,色情報による追跡対象の判別は有効であるが,照明条 件が様々に変わる環境においては,色情報はその影響を受けやすい.位置情報は視差情報により与えられるため, 照明の影響は受けにくいが,単独で用いると追跡対象の実際の運動と予測モデルが異なったときに追跡が困難と なる.よって,色情報と位置情報の両方を組み合わせて用いることで,各情報を単独で用いる場合よりも実環境 における追跡能力を向上させることが可能となる.

(2) $\alpha \ge \alpha_{th}$ の場合

ある.

ホワイトバランスの変化が極端に大きいとき,色情報への信頼度は著しく低下する.そのため,相違度を位置 情報のみにより与える.

追跡中は動的な環境における色情報の変化に対応するために, D の値が小さいときにテンプレートの色情報を 更新する.なお,テンプレートの色情報が更新されるとき,同時にそのフレームでのホワイトバランスも登録す る.登録後は,現フレームでのホワイトバランスと登録されたホワイトバランスの各ゲインの差の絶対和により |W|を算出する.また,何も検出されない状態が1s以上連続したときに追跡対象を見失ったとみなす.見失う前 に最後に追跡対象が検出された位置を (*X_e*, *Y_e*) として,相違度を算出し再検出を行う.

提案した手法を用いて,図5の環境で追跡対象の検出を行った際に得られた相違度 D の値の変化を図9に示す. 図の表し方は図4と同様である.これより,相違度 D の値がホワイトバランスの変化に依存せず,照明変動のあ る環境でも低い値を取りつづけられていることがわかる.

3. 実 験

3.1 オフライン実験

提案手法の有用性を検証するために、屋外の照明変動がある環境において実験を行った.まず、提案手法(I)、 色情報のみを用いた相違度(II, III)、位置情報をのみを用いた相違度(IV)、色情報と位置情報に一定の重みを つけた相違度(V, VI, VII)のそれぞれによる追跡性能の比較を行う.表1に各条件を示す.すべての条件にお いて、各プロセスにおける閾値は同じ値とする.また、提案手法において用いる各種閾値やパラメータの値を表2 に示す.種々の条件での実験環境を等しくさせるため、本実験では予め取得した画像に対し各手法を適用する.実 験環境における追跡経路とシーンを表したものが図11である.図中、Sは追跡開始位置、Gは追跡終了位置を示 す.実験に用いる画像は、歩行中の特定の人物を追従するように移動ロボットを操作した際にロボットに搭載の ステレオカメラより取得される.オフラインでの追跡対象検出実験後、実験環境を照明条件に応じて6つのシー ンに分け、各シーンに対して追跡性能を評価する.各シーンでの画像の取得例を図12に、詳細な条件を表3に示



Fig. 9 The influence of white-balance changes on the proposed method. The environments where the dissimilarity is given are the same environments as shown in Fig. 4.

1	Table 1 Compared settings of features and dissimilarity equations								
	Feature	Dissimilarity equation							
Ι	Color and Location Information	(5)							
II	Color Information (frequently updated)	R _{color}							
III	Color Information (no updated)	R _{color}							
IV	Location Information	Ε							
v	Color and Location Information	$0.7R_{color} + 0.3E$							
VI	Color and Location Information	$0.5R_{color} + 0.5E$							
VII	Color and Location Information	$0.3R_{color} + 0.7E$							

Table 2 Thresholds and parameters for experiments

The distance of a projection plane: Y_{min} and Y_{max}	<i>Y_{min}</i> : 1.5 m, <i>Y_{max}</i> : 4.0 m
The threshold value of projected points in one cell which is extracted to be labeled	1000
The threshold of the width and depth of search regions: C_x and C_y	C_x : 8 cells, C_y : 5 cells
The numbers of the histogram of Fig. 8: N and M	N: 20, M: 10
The numbers of steps of a Hue-Saturation histogram: h_{max} and s_{max}	h_{max} : 20 steps, s_{max} : 16 steps
The constant k in (4)	$1.0 \ {\rm m}^{-1}$
The threshold α_{th} in (5)	0.25
The proportionality constant <i>p</i> between α and $ W $	0.8

す. また,図 13 は各フレームでの追跡対象以外の人物の人数(青線),オクルージョンの発生(黄色),及びホ ワイトバランスの変化(赤線)を示す.なお,画像は 6.9 fps で取得され,1728 枚を実験に用いる.

本研究では,照明環境によらず追跡対象を正しく,網羅的に検出することを目的としている.そこで,追跡対象 検出の正確性と網羅性に対応する評価値として,適合率と再現率,及びその調和平均であるF値を用い実験の評 価を行う.各評価値は以下で定義される.

適合率 P=
$$\frac{A}{A+B}$$
, 再現率 R= $\frac{A}{A+C}$, F 値 = $\frac{2PR}{P+R}$

[DOI: 10.1299/transjsme.15-00368]



Fig. 10 Details of each frame that the evaluation values consist of. In A frames, a target is identified correctly. B shows the frames that non-target is detected as a target. If no object is identified as a target despite target's presence, the frames are classified to C frames.



Fig. 11 Path of the offline experiment. S and G show the starting and goal points, respectively.

A:追跡対象人物が正しく検出されたフレーム数(図10(a))

B:追跡対象人物以外が検出されたフレーム数(図10(b))

C:追跡対象人物が存在していたが何も検出されなかったフレーム数(図 10(c))

表4に評価値の算出結果を示す.なお,表中の*は評価値算出式の分母の値が0になったものを表している.適 合率は、シーン4とシーン6以外では提案手法を用いた追跡結果が最も高い値である.また,再現率及びF値は、 全てのシーンにおいて提案手法による値が一番高い.適合率,再現率及びF値の全てのシーンを通しての平均と 標準偏差より,総合的に提案手法での追跡対象の検出性能が最も良いことがわかる.

まず,適合率に関して考察する.シーン4では,領域分割に失敗したフレームで色更新が行われ,その後のフレームにおいて誤検出が起こっていた.図14に領域分割の失敗(図14(a))と,その後の誤検出の様子(図14(b))を示す.図中の矩形は追跡対象領域を表す.左側の追跡対象と右側の人物の領域の分割が正しく行われず,右の人物の領域も追跡対象領域に含まれてしまっている.このフレームで色情報の更新が行われてしまったため,右の人物の色情報も追跡対象の色情報として保存され,図14(b)のように誤検出が起こった.このとき,提案手法のみではなく,色情報を使っている条件 V, VI, VII においても同じフレーム数だけ誤検出となっていた.しかし,色情報を使用しない条件 IV においては,領域分割が失敗した影響が小さく,誤検出フレーム数も少なくなっていた.そのため,シーン4において条件 IV の適合率が最も高くなっている.また,シーン6ではオクルージョンの影響が提案手法の適合率の低下に起因していた.誤検出が起こった際に,ホワイトバランスの変化量が閾値よりも大きくなっており,位置情報のみに基づいて追跡を行っていた.よって,色情報にも重みを置かれている条件 V, VI, VII ではオクルージョン発生中や発生後に未検出となっていたフレームにおいて,提案手法では誤検出が発生

Table 3Details of offline experimental scenes. Condition means the lighting condition (e.g. back means back
lighting); Number of people shows how many the other people were present at a frame on average;
Occlusion presents the number of occlusion, the average number of occluded frames, and the maximum
number of occluded frames; Shadow indicates how often the shadow appeared.

Scene	Condition	Number of people	Occlusion the number of occlusion/ the average frames/ the maximum frames	Shadow
			5.0	
1	back	1.8	5	no appearance
			11	
			4	
2	direct	1.9	5	no appearance
			7	
			8	
3	side	1.1	4	no appearance
			7	
			15	
4	4 direct		7	no appearance
			14	
			7	
5	side	1.2	7	continuous appearance by buildings
			12	
			14	
6	direct	1.3	7	no appearance
			17	

し適合率が下がってしまった.

また,提案手法の再現率が最も低いシーン6では,スミア現象により視差情報が得られず未検出となることが あった.同様にスミア現象が起こっている際,ホワイトバランスの変化と画像中の追跡対象の色情報の変化が対応 していない状況もあった.その例を図 15 に示す.図中では追跡対象のみがフレーム内に存在している.また,2 フレーム間では服の色に変化があることがわかる.しかし,このとき画像中のホワイトバランスの値は変化せず, この場面では未検出フレームが連続して起こっていた.これより,ホワイトバランスの変化は,照明環境の変化と 完全には対応していないことがいえる.

3.2 オンライン実験

提案手法の有用性を検証するために,屋外と屋内の実環境において移動ロボットによる人物追跡実験を行った. 移動ロボットには Blackship(セグウェイジャパン)を,ステレオカメラには Bumblebee2(Point Grey Research)



Fig. 12 Offline experimental scenes. The scenes are classified according to illumination environments.



Fig. 13 Details of the number of people excluding a target, occlusion and white-balance changes in offline experimental environments. The blue line shows the number of people, and the period with occlusion is shown as the orange line. The red line corresponds to the changes in white balance.





(a) Failure of the segmentation

(b) Mis-detection caused by the failure

Fig. 14 Example of mis-detection of the target region

を使用した. PID 制御を用い,追跡対象人物と移動ロボットとの距離と方位に応じて,移動ロボットの運動を制御 した.また,提案手法における各種閾値やパラメータは表2に示す値とした.手法の評価には3·1項で定義した適 合率と再現率,F値を用いる.

Isobe, Masuyama and Umeda, Transactions of the JSME (in Japanese), Vol.82, No.834 (2016)

Table 4Results of precision and recall in offline experiments. I is the proposed method. II and III use only color
information. IV is depend on only location information. V, VI and VII are based on both color and location
information, with fixed weights. The bold letters show the results, which well indicate the usefulness of
the proposed method.

	Ι	II	III	IV	v	VI	VII	Ι	II	III	IV	v	VI	VII
	Scene 1 (91 frames)							Scene 4 (660 frames)						
Precision [%]	100	80.0	81.6	100	100	100	100	99.6	0.0	0.0	99.8	99.6	99.6	99.6
Recall [%]	93.8	74.1	67.8	93.8	93.8	93.8	93.8	87.9	0.0	0.0	77.8	87.9	87.9	82.0
F – measure [%]	96.8	76.9	74.1	96.8	96.8	96.8	96.8	93.4	0.0	0.0	87.4	93.4	93.4	89.9
	Scene 2 (78 frames)						Scene 5 (347 frames)							
Precision [%]	100	60.0	66.7	83.3	100	100	100	100	0	*	100	100	100	100
Recall [%]	88.3	16.1	6.8	53.6	73.3	73.3	73.3	73.0	0.0	0.0	72.3	73.0	72.7	73.0
F – measure [%]	93.8	25.4	12.3	65.2	84.6	84.6	84.6	84.4	0.0	0.0	83.9	84.4	84.2	84.4
	Scene 3 (196 frames)						Scene 6 (356 frames)							
Precision [%]	100	16.7	75.8	89.6	100	100	100	98.9	*	*	93.1	100	100	100
Recall [%]	93.8	6.6	63.9	78.4	93.8	93.8	93.8	67.0	0.0	0.0	58.5	59.8	59.4	59.4
F – measure [%]	96.8	9.4	69.3	83.6	96.8	96.8	96.8	79.9	0.0	0.0	71.8	74.8	74.5	74.5

	Mean and standard deviation(SD)						
	Ι	II	III	IV	V	VI	VII
Mean of Precision	99.7	31.3	56.0	94.3	99.8	99.8	99.8
SD of Precision	0.4	32.8	32.8	6.3	0.2	0.2	0.2
Mean of <i>Recall</i>	84.0	16.1	23.1	72.4	80.3	80.1	79.2
SD of Recall	10.0	26.6	30.4	13.3	12.5	12.7	12.2
$\begin{array}{c} \text{Mean of} \\ F-measure \end{array}$	90.8	18.6	26.0	81.5	88.5	88.4	87.8
SD of F-measure	6.4	27.6	32.7	10.3	8.0	8.1	7.8



(a) High saturation value of target's clothes (b) Low saturation value of target's clothes



3·2·1 屋外実験

屋外の照明変動があり複数の人物が存在する環境で 151s の追跡を行った. なお,実験を通して計6回のオクルージョンが発生し,オクルージョンの継続時間は平均して約0.6s,最大約1sであった. また,追跡経路は図16の通りで,照明条件によりシーン分けを行った.シーンは以下の4つとする.



Fig. 16 Path of the online experiment



Fig. 17 Details of the number of people excluding a target, occlusion and white-balance changes in online experimental environments. What each line shows is detailed in Fig. 13.

シーン1: 順光

シーン 2: 逆光

シーン3:ビルの陰(順光)

シーン4:ビルの陰(逆光)

また,実験を通して,フレーム毎の追跡対象以外の人物の数,オクルージョンの発生,開始フレームからのホワイトバランスの変化を表したものが図17である.実験の様子を図18に,追跡対象の検出例を図19に示す.図19中で,赤い矩形は追跡対象領域を示し,赤い点は領域と対応する2値画像中の白画素の重心を表す(図7(b)参照). 実験結果に対する評価値を表6に示す.結果より,全ての実験環境において,91%以上の適合率,88%以上の再現率,90%以上のF値が得られたことがわかる.

各評価値が他のシーンに比べて低くなったシーン3について,その原因を考察する.シーン3においては,ホ ワイトバランスの変化が頻繁に起こっていた.その間,追跡は位置情報のみに依存し,追跡対象の運動は仮定し た運動モデルに沿っていなかった.よって,環境中に追跡対象が存在しているにも関わらず未検出となる状況が あった.この未検出のフレームが増加することで,更に追跡対象の位置予測が不安定になり,誤検出が起こったと 考えられる.また,図20のように誤検出となった人物を誤追跡してしまう状況もあった.しかし,追跡対象の色 情報保存時からのホワイトバランスの変化αが0.5であり,色情報にも信頼を置いた追跡となっていた.そのため 誤検出も連続的ではなく,結果1秒以上連続した未検出の状態となり見失いが発生したとみなされた.見失った 後は,位置情報の比較に用いる追跡対象の位置(*X_e*,*Y_e*)は見失う前に検出された位置にリセットされた.その付近 で保存された色情報に似た服の人物を追跡対象として正しく検出することができ,追跡が再度開始された.



Fig. 18 Outdoor experimental scenes. The scenes are classified according to illumination environments.

Scene	Number of frames	Precision [%]	Recall [%]	F-measure [%]
1	559	99.4	97.0	98.2
2	472	100	94.8	97.4
3	462	91.6	89.2	90.4
4	716	100	98.4	99.2

Table 5 Evaluation of results in outdoor scenes

Table 6 Evaluation of results in indoor scenes

Number of frames	Precision [%]	Recall [%]	<i>F-measure</i> [%]		
791	96.7	81.0	88.1		

3·2·2 屋内実験

次に,屋内環境で49sの追跡実験を行った.屋外環境との差別化を図るために,夜間,天井や部屋からの光が 処々で点灯している環境を実験環境とした.実験を通して,天井や部屋の光が強く当たる場所では照度が高くな り,そうでない場所では照度が低くなることで,照明変動が発生していた.フレーム毎の追跡対象以外の人数,オ クルージョンの発生,及び追跡開始からのホワイトバランスの変動を図21に表す.

実験環境における追跡の様子を図 22 に,追跡対象の検出例を図 23 に示す. 照度の強弱が実験全体で交互に現 れていたため,本実験ではシーン分けは行わず,実験全体での評価値を算出する.評価値の算出結果を表6に示 す.実験全体で全ての評価値は80%を超えている.また,図 24 のように短時間(約 1.3~1.4 s)で追跡対象の服の



Isobe, Masuyama and Umeda, Transactions of the JSME (in Japanese), Vol.82, No.834 (2016)



Fig. 19 The results of the outdoor experiment. The regions enclosed by red rectangles show the target's regions. The red dots indicate the centroids of the white pixels of the binary images.



(a) 1467th (The first frame of mistracking)



(b) 1476th



(c) 1489th (No-detection frame)

(d) 1503rd (Re-detection frame)

Fig. 20 Failure of target tracking



Fig. 21 Details of the number of people excluding, occlusion and white-balance changes in indoor experimental environments. What each line shows is detailed in Fig. 13.















(e) 453rd

(f) 508th

(g) 635th

(h) 730th

Fig. 22 Indoor experimental environment. In the environment, the luminance changed frequently by ceiling lights.





色が変化するときにおいても、照明の変化に合わせて相違度に重みを置くことや、色情報の更新を行うことで照明 変動にロバストに追跡が行われていたことがわかる.しかし,図 25 のように照明具の間にいるときには,進行方 向の照明が常に点灯している状態となり、追跡対象の存在する付近の照度が低くてもホワイトバランスが変化し ないことがあり、未検出が増えていた. 今後はこのような環境にも対応するために、光の当たる方向や光源の種類 に適応させるホワイトバランスのみではなく、明暗にも対応したパラメータを導入する必要があると考えられる.

16



Fig. 24 Though the color of the target's clothes was changed frequently, the target was identified correctly. It was due to the weighting based on the changes in white balance and the updating.



Fig. 25 The C frames caused by no change in white balance. When there was the target between lights from the rooms, the luminance changed though white balance did not change.

4. 結 論

本論文では、ステレオカメラを用いた、移動ロボットによる人物追跡手法を提案した。色情報と位置情報の評価値への照明変動に応じた重みづけにより、照明変動があり複数の人物が存在する環境における追跡の実現を図った。実環境における追跡実験により、91%以上の適合率、88%以上の再現率、90%以上のF値を得た。

今後は,照明変動を示すパラメータを改良を検討している.また,追跡対象の特徴量の追加等による誤検出や 未検出の改善も図っていく.より混雑した環境や,照明変動がより複雑な環境での追跡の実現を目指す.

References

- Calisi, D., Iocchi, L. and Leone, R., Person following through appearance models and stereo vision using a mobile robot, Proceedings of VISAPP (Workshop on Robot Vision) (2007), pp. 46–56.
- Chakravarty, P. and Jarvis, R., Panoramic vision and laser range finder fusion for multiple person tracking, Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robotics and Systems (2006), pp. 2949–2954.
- Cielniak, G., Duckett, T. and Lilienthal, A. J., Improved data association and occlusion handling for vision-based people tracking by mobile robots, Proceedings of the 2007 IEEE/RSJ International Conference on Intelligent Robotics and Systems (2007), pp. 3436–3441.
- Doi, M., Nakakita, M., Aoki, Y. and Hashimoto, S., Real-time vision system for autonomous mobile robot, Proceedings of the 10th IEEE International Workshop on Robot and Human Interactive Communication (2001), pp. 442–449.
- Doog Inc. (online), available from 〈http://en.doog-inc.com/product.html〉, (accessed on 30 June, 2015).
- Five Elements Robotics (online), available from http://5elementsrobotics.com/budgee-specification-page/ , (accessed on 30 June, 2015).
- Fritsch, J., Kleinehagenbrock, M., Lang, S., Fink, G. A. and Sagerer, G., Audiovisual person tracking with a mobile robot, Proceedings of the 2004 International Conference on Intelligent Autonomous Systems (2004), pp. 898–906.
- Hebert, M., Active and passive range sensing for robotics, Proceedings of the 2000 IEEE International Conference on Robotics and Automation (2000), pp. 102–110.
- Hu, C., Ma, X. and Dai, X., A robust person tracking and following approach for mobile robot, Proceedings of the 2007 IEEE International Conference on Mechatronics and Automation (2007), pp. 3571–3576.
- Isobe, Y., Masuyama, G. and Umeda, K., Human following with a mobile robot based on combination of disparity and color images, Proceedings of 10th Japan-France Congress and 8th Asia-Europe Congress on Mecatronics (2014),

pp. 84–88.

- Katz, D., Kenney, J. and Brock, O., How can robots succeed in unstructured environments?, RSS Workshop on Robot Manipulation: Intelligence in Human Environments (2008).
- Kwon, H., Yoon, Y., Park, J. B. and Kak, A. C., Person tracking with a mobile robot using two uncalibrated independently moving cameras, Proceedings of the 2005 IEEE International Conference on Robotics and Automation (2005), pp. 2877–2883.
- Lee, J. H., Tsubouchi, T., Yamamoto, K. and Egawa, S., People tracking using a robot in motion with laser range finder, Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems (2006), pp. 2936– 2942.
- Mattos, L. and Grant, E., Passive sonar applications: target tracking and navigation of an autonomous robot, Proceedings of the 2004 IEEE International Conference on Robotics and Automation (2004), pp. 4265–4270.
- Petrovic, E., Leu, A., Ristic-Durrant, D. and Nikolic, V., Stereo vision-based human tracking for robotic follower, International Journal of Advanced Robotic Systems, Vol. 10 (2013), pp. 1–10.
- Raibert, M., Blankespoor, K., Nelson, G., Playter, R. and the BigDog Team, BigDog, the rough-terrain quadruped robot, Proceedings of the 17th World Congress of the International Federation of Automatic Control (2008), pp. 10822– 10825.
- Satake, J., Chiba, M. and Miura, J., Visual person identification using a distance-dependent appearance model for a person following robot, International Journal of Automation and Computing, Vol. 10, Issue 5 (2013), pp. 438–446.
- Stewart Golf Limited (online), available from 〈http://www.stewartgolf.com/X9Follow/features.html〉, (accessed on 30 June, 2015).
- Takemura, H., Nemoto, Z. and Mizoguchi, H., Development of vision based person following module for mobile robots in/out door environment, Proceedings of the 2009 IEEE International Conference on Robotics and Biomimetics (2009), pp. 1675–1680.
- Tsalatsanis, A., Valavanis, K. and Yalcin, A., Vision based target tracking and collision avoidance for mobile robots, Journal of Intelligent and Robotics Systems, Vol. 48, Issue 2 (2007), pp. 285–304.
- Ubukata, T., Terabayashi, K., Moro, A. and Umeda, K., Multi-object segmentation in a projection plane using subtraction stereo, Proceedings of the 20th IEEE International Conference on Pattern and Recognition (2010), pp. 3296–3299.