

## ウェアラブル周囲モニタリングシステムと物体追跡方式 Wearable Around View Monitoring System and its Multiple Object Tracking

秦 淑彦<sup>†1</sup> 迫谷 伊織<sup>†1+2</sup> 小竹 涼也<sup>†1+3</sup> 片木 伸哉<sup>†1+4</sup>  
Toshihiko Hata Iori Sakotani Ryoya Kotake Shinya Katagi

### 1. はじめに

人間が周囲の状況を把握する際、視覚は重要な役割を果たすが、人間の安定注視野は水平 60~90 度、垂直 45~70 度と限られた範囲しかない。[1]さらに会話や考え事などにより周囲の認識能力は低下し、スマートフォンやスマートグラスを使うと周囲の視覚情報も低減する。VR グラス装着にいたっては、周囲の視覚情報は全く得られない。十分把握できなければ、不適切な行動や危険な行動をとる恐れがある。

そこで、ウェアラブルカメラを用いて、ユーザが見えない部分も含めて周囲をモニタリングし、適切に状況を知らせるシステムの開発に取り組んでいる。[2][3] 全方位カメラで 360°画像を撮影するとともに、慣性センサを用いてカメラの姿勢情報を取得する。この姿勢情報に基づき、カメラの撮影方向に依らない画像となるよう座標変換し、周囲の物体と位置を検出するとともに、人物や車を追跡する。

本稿では、ウェアラブル周囲モニタリングシステムの基本アイデアについて述べた後、全方位画像とカメラ姿勢情報を用いた周囲物体の追跡方式を説明するとともに、方式検証のための実験について報告する。

### 2. ウェアラブル周囲モニタリングシステム

#### 2.1 ウェアラブルカメラを用いた周囲モニタリング

セキュリティやライフログなどの応用を想定し、スマートグラス等に搭載されたカメラで撮影した画像から、周囲の人物や顔を認識するシステムが提案されている。[4][5] これらのシステムでは、装着者の一人称視点の画像、すなわち顔を向けた方向の限られた視野の画像に写る物体を認識する。また、連続する時間に渡って検出した物体を追跡する機能はない。

文献[6]はウェアラブルカメラを用いて複数の人物を追跡する MOT (Multiple Object Tracking) 方式を提案している。ユーザが視線や場所を移動することにより、ある人物グループの観察から他の人物グループの観察に移行した際、画像情報から視野の変化を検出し、変化が小さい場合は連続する時間フレームにて追跡処理を行い、大きい場合はシーンが切り替わったとして新規に追跡処理を開始する。カメラの撮影方向の変化に対応した時間的に連続する追跡処理ではない。

#### 2.2 提案システム

従来は装着者の視野にある物体を対象としているのに対し、図 1 に示す提案システムでは装着者が見えない/見えない部分も含めた周囲のモニタリングを行う。自動車の

Around View Monitor と同様なコンセプトである。また、人物を対象とする場合は、人物を認識するだけでなく追跡を行い、実空間での位置を推定する。これにより、以下のような応用が考えられる。

- ・ 歩きスマホにおける前方注意に加え、後方から近づく自転車や歩行者との衝突防止。
- ・ 犯罪者や迷子の捜査に加え、人物追跡によるストーカーの検知。
- ・ 仮想空間ゲームやメタバースをしている際の注意喚起。

ユーザの視野範囲外も含めて観察するため、広角レンズや魚眼レンズのウェアラブルカメラを 1 つあるいは複数使用する。カメラの姿勢情報 (方向や位置) を測定するために慣性センサや SLAM (Simultaneous Localization And Mapping) デバイス、画像認識等の処理を行うための小型・高性能なプロセッサ、情報提示のためのディスプレイやスピーカが必要である。これらの小型・高性能化がさらに進み、将来はスマートグラスやスマートキャップなどに搭載されると想定する。処理の流れを以下に示す。

#### ①カメラ画像と姿勢情報の取得

慣性センサの情報を基にカメラの方向を計算する。位置情報が必要な場合は、カメラや慣性センサを用いて自己位置を推定する。

#### ②カメラ姿勢情報を用いた MOT

深層学習を用いた物体認識技術により物体の画像領域を検出し、検出物体の外接矩形の位置や大きさの情報をを用いて、連続する画像フレームにて物体を追跡する。この際、カメラの姿勢情報に基づき撮影画像を基準座標系に変換し、撮影方向に影響されない追跡処理を行う。

#### ③周囲人物の位置推定と提示

検出物体に対し、撮影位置の異なる複数画像における視差情報を用いて、距離を推定する。

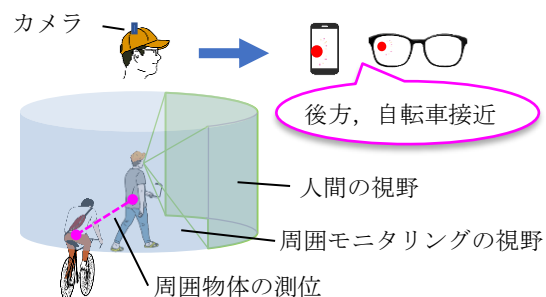


図 1 ウェアラブル周囲モニタリングシステム

### 3. 全方位カメラと姿勢情報を用いた周囲物体追跡

#### 3.1 Multiple Object Tracking

MOT (Multiple Object Tracking) とは、画像から特定の物体を認識し、連続する複数フレームにおいて、同一の物体を

†1 広島工業大学 Hiroshima Institute of Technology

†2 (株) モルテン Molten Corp.

†3 (株) インプレステック Impress-Tech. Co., Ltd.

†4 (株) ハイエレコン Hi-Elecom Co., Ltd.

追跡する手法である。今回使用する Tracking-by-Detection 型の MOT[7]は、各フレームで物体の認識を行った後、前フレームの位置から次フレームの位置を予測し、予測と認識の結果を結び付けることで追跡する。従来方式では、カメラが動かない場合は精度良く追跡できるが、ウェアラブルカメラでは撮影する位置や方向が変化するため、次フレームの位置を正しく予測できず、追跡精度が著しく低下してしまう。

### 3.2 全方位カメラ

魚眼レンズを用いた全方位カメラで採用される等距離射影方式では、レンズの中心を中心とする球面がスクリーンとなり、物体の各点は、点と中心を結ぶ直線が球面と交差する点に写像される (図 2(a))。

全方位カメラから出力されるパノラマ画像は正距円筒図法に基づいている。正距円筒図法では、2次元直交座標の各軸を球面の緯度と経度とするものであり、球面に投影された全方位の空間を矩形の2次元画像として表す (図 2(b))。球面上の点  $p$  の3次元直交座標  $(x, y, z)$  と緯度経度  $(\theta_l, \lambda_l)$  は式(1)で表される。

$$x = \cos\theta_l \times \cos\lambda_l, y = \cos\theta_l \times \sin\lambda_l, z = -\sin\theta_l \quad (1)$$

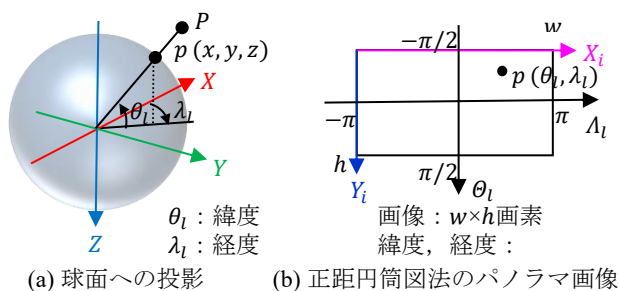


図 2 全方位カメラにおける射影とパノラマ画像

### 3.3 提案方式

撮影方向が変化しても適切に MOT 処理ができるようにするため、カメラ姿勢情報を用いて、カメラ座標系の XY 面が実空間の水平面となるような画像に変換する (図 3)。処理手順を以下に述べる。

#### ① 画像&姿勢情報取得

全方位カメラで周囲を撮影するとともに、カメラに取り付けた慣性センサで測定した角速度等からカメラの撮影方向を計算する。図 3 に示すようにカメラ座標系において、XY 面が実空間の水平面、X 軸は任意の方向となるよう基準座標を設定し、この基準座標に対するカメラの撮影方向を、基準座標の 3 軸周りの回転角  $(\psi, \theta, \varphi)$  として求める。

#### ② 姿勢情報に基づき画像を基準座標に変換

パノラマ画像における各画素のカメラ座標  $(x, y, z)$  を式(1)から求め、カメラ撮影方向に基づき基準座標  $(x_r, y_r, z_r)$  を式(2)により計算する。ここで  $R$  はカメラ撮影方向を表す回転角  $(\psi, \theta, \varphi)$  による回転行列である。

$$(x_r, y_r, z_r)^T = R(x, y, z)^T \quad (2)$$

さらに式(3)により、基準座標に対する緯度経度  $(\theta_{lr}, \lambda_{lr})$  を計算し、元のパノラマ画像の各画素をこの緯度経度の位置に移動する。

$$\theta_{lr} = -\sin^{-1}z_r, \lambda_{lr} = \tan^{-1}(y_r/x_r) \quad (3)$$

#### ③ Tracking-by-Detection 型 MOT により人物追跡

基準座標でのパノラマ画像から物体認識アルゴリズム YOLO[8]により人物領域を検出し、motpy[7]により同一人物を認識し追跡する。

パノラマ画像の左右両端へフレームアウトした人物は反対側から再度出現する。このままでは別人物として追跡を行うため、連続するフレームにおいて、左右両端に存在する人物領域の縦方向の位置が近い場合、同一人物として追跡する。

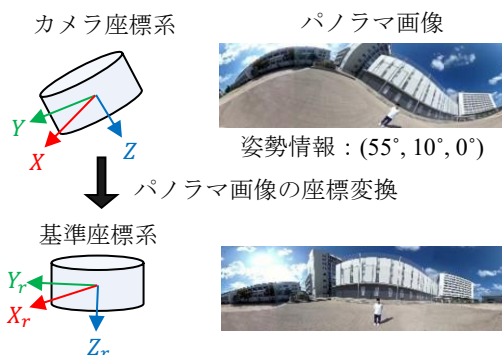


図 3 パノラマ画像の基準座標系への変換

### 3.4 画像変換処理の高速化

3.3 の提案方式は、高解像度パノラマ画像の全画素に対する処理を含むため、実時間処理を目指した高速化の検討が必要である。高速化として画像の解像度とフレームレートの低減、並列計算等のプログラム実装効率化が考えられ、ここでは、3.3②の画像変換における解像度とフレームレートの低減について検討する。

画像変換は、各フレームに対する回転マトリクス計算、各画素に対する座標変換 (3D 直交座標系にて回転し、正距円筒座標に変換) と正距円筒画像への写像から成り、全方位パノラマ画像は高解像度であるため、全画素に対して行う座標変換と写像の処理に時間を要する。要求される追跡性能 (誤追跡が少ない) の範囲で、どこまで画像の解像度とフレームレートを低減できるか、実験により明らかにする必要がある。解像度の低減は YOLO による人物認識に影響し、遠くにいる人物の画像は画素数が少なくなり、認識されにくくなる。また、1 画素当たりの実空間での距離が大きくなるため、motpy の追跡にも影響し、適切にパラメータ設定したとしても誤追跡が増加する可能性がある。フレームレートの低減は、フレーム間での人物の移動距離が大きくなるため、motpy での追跡性能が低下する可能性がある。解像度とフレームレートに応じた適切なパラメータを設定した上で、追跡性能を評価しなければならない。

画像変換では、回転後の各画素の位置を計算して出力画像を作成するが、近傍の画素は互いに近い位置に変換される。そこで、座標変換は低解像度で計算し、近傍画素の補完により高解像度化して写像することを考える。

## 4. 実験

提案方式の実現可能性と課題を抽出するための実験を行った。図 4 に実験システムの構成、表 1 に全方位カメラ、

表 2 に慣性センサの仕様を示す。カメラとセンサをノートパソコンに USB 接続し、データ取得プログラムにより周期的に全方位画像と姿勢情報を取得しファイルに保存する。画像変換プログラムが姿勢情報に基づき画像をカメラ座標系から基準座標系へ変換し、MOT 処理プログラムが人物領域抽出および追跡を行う。プログラムは Python で記述している。(1)~(4)の実験では 1920×960 画素、24fps の画像データを使用した。

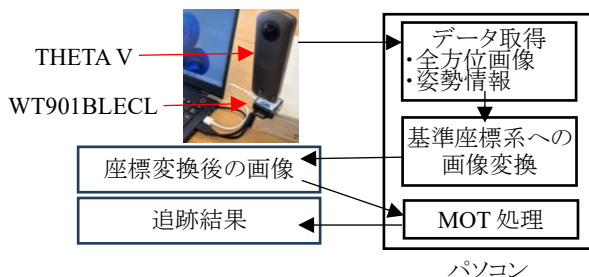


図 4 実験システムの構成

表 1 THETA V (Ricoh Co., Ltd) の仕様[9]

画像	360°全天球画像 魚眼レンズ(6群7枚, F2.0)×2 1/2CMOS(1200万画素)×2 3840×1920/29.97fps, 1920×960/29.97fps JPEG/MP4
他	USB2.0, 内蔵バッテリー 45.2×130.6×22.9mm <sup>3</sup> , 121g

表 2 WT901BLECL (WitMotion Shenzhen Co.) の仕様[10]

データ	角度 XYZ(X&Z:±180°, Y:±90°) 加速度 XYZ(±16g, 0.005g) 角速度 XYZ(±2000°/s, 0.61°/s) 磁界 XYZ(±4900μT, 16 bits) 周期 0.2~200Hz
他	USB, BLE, 内蔵バッテリー 51.3×36×15mm <sup>3</sup> , 20g

(1) 基準座標系への座標変換

三脚に取り付けたカメラの撮影方向 (XYZ 軸の回転) を変化させて撮影した画像を基準座標系に変換し、目視で確認した。撮影方向によってはある程度の歪は存在するが、追跡に利用する水平部分は想定通りに変換できている。

(2) 2名の周囲人物の追跡

装着者と2名の周囲人物が、各々、歩く/走る運動状態と立ち止まる静止状態を組み合わせた動作 (図 5) を行い、装着者がカメラを手に持って各軸回転させた画像を撮影した。追跡処理の結果を表 3 に示す。MOT の追跡誤りには、人物が写っていないのに存在するとした場合 (False Positive), 写っているのに追跡しなかった場合 (False Negative), 近くの人物で ID が入れ替わる、もしくは同一人物に新しい ID がつく場合 (ID Switch) がある。これらの誤りの合計数を追跡判定回数 (正解と誤りの合計) で割った値を MOTA とする。表 3 には MOTA と誤り合計数/追跡処理を行った画像のフレーム数を示す。

追跡精度は、装着者と周囲人物の動作状態の違いに大きな差はなく、全ての項目において 99.6%以上正しく追跡で

きている。誤追跡のあった画像を分析すると、頭を左右に早く振るなど、カメラが速く回転するとき、さらに人物の距離が遠い場合に誤っている。

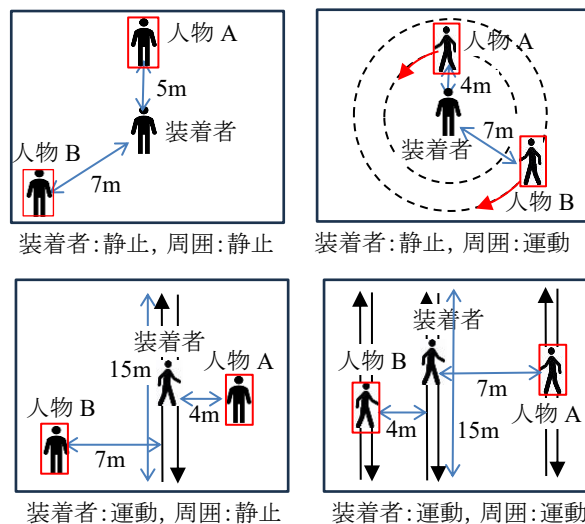


図 5 実験 (2) における装着者と周囲人物の動作

表 3 実験(2)における追跡結果

装着者		周囲人物			
		静止		運動	
		誤追跡数 / フレーム数	MOTA (%)	誤追跡数 / フレーム数	MOTA (%)
装着者	静止	0 / 1841	100.0	3 / 1856	99.8
	運動	8 / 2830	99.7	12 / 2819	99.6

(3) 人物距離とカメラ回転速度

実験(2)における誤追跡の原因を確認する実験を行った。三脚に取り付けたカメラで人物とカメラの距離を変化させ、横に 5m 歩行して追跡処理を行った。距離 23m までは 90%以上の追跡率であるが、25m を超えると 50%を下回る。これは人物領域の画素数が少なくなり、YOLO の認識率が低下することが要因である。

装着者が方向転換する、振り向く、屈むなどによる急激な撮影方向の変化による追跡性能への影響を調べるため、三脚に取り付けたカメラを早く回転させて追跡処理を行った。10m 離れた人物であれば、およそ 180°/s の回転速度までは 100%追跡できるが、360°/s で 74%、720°/s で 25%と追跡性能が低下する。慣性センサが高速回転に追従できず、正しい姿勢情報が得られていないことが要因と思われる。

(4) 多数の周囲人物の追跡

キャンパス建物内を計 12 名の人物が行き来する空間にて、カメラを回転させて撮影し追跡処理を行った (図 6)。装着者はカメラを X 軸 ±50°, Y 軸 ±40°, Z 軸 ±180° の範囲で回転させながら歩いた。表 4 に追跡結果を示す。画像変換無しの MOTA が 79.9%, 有りの MOTA が 95.7%であり、画像変換により誤りが大きく低減している。誤追跡となる原因は実験(2)の理由に加え、複数人の人物領域が重なる、人物と背景の区別がつかない、人でない物体を人と認識するなど、YOLO における認識誤りと、オクルージョンにより同一人物として識別されない motpy における追跡誤りがある。



画像変換無



画像変換有

図 6 実験(4)におけるパノラマ画像と人物追跡の例

表 4 実験(4)における追跡結果

画像 変換	False Positive	False Negative	ID Switch	Total	MOTA(%)
無	2	862	458	1322	79.9
有	5	250	27	282	95.7

### (5) 画像変換処理の高速化

3.4 で述べた画像変換処理の高速化について、保存した画像データと姿勢情報を読み込み、画像の解像度とフレームレートを変化させた場合の処理時間と、低解像度で座標変換した場合の処理時間を測定するとともに、追跡処理を行い誤追跡数を求めた。装着者と 2 名の周囲人物が歩く／走る動作を 1920×960 画素、24fps で撮影し、解像度については 1920×960、960×480、480×240 画素、フレームレートについては 24、12、6fps について測定した。座標変換は入力画像の解像度と縦横 1/2 に間引いた画像で行った。ノートパソコン (Core i5-10210U, 8GB, Widonws10) 上で画像変換処理を実行し、座標変換と写像処理の時間を計測し、フレーム毎の平均値を計算した。表 5 に処理時間と誤追跡数／画像フレーム数を示す。

フレーム毎の処理時間はほぼ解像度に比例している。誤追跡数は、解像度が低いほど、フレームレートが小さいほど増加するが、今回の約 90 秒の画像コンテンツでは、960×480、12fps であれば誤追跡は 20 回程度であり、比較的良く追跡できている。解像度が 480×240、あるいはフレームレートが 6fps になると誤追跡数は 100 を超えてしまう。

誤追跡の要因の一つは、解像度が低下することにより、YOLO での人物認識率が低下することである。表 5 の結果は、YOLO の人物確率パラメータを 0.5 としている。この値を低くすることにより誤追跡を低減できるが、人以外の物体を人と認識する副作用があり、解像度に応じた人物確率パラメータの検討が必要である。もう一つの要因は、フレームレートが低下することにより motpy での追跡精度が低下することである。フレームレートと解像度に応じて適切な motpy のパラメータが存在するはずであるが、まだ検討できていない。

間引きをした座標変換の場合、ほぼ同様の追跡精度で、処理時間は、座標変換で 40%程度、トータルで 35%程度、低減されている。

表 5 実験(5)における画像変換の処理時間

画像 解像度	処理 fps	処理 間引	処理時間(ms)			
			誤追跡数 /フレーム数	座標変換	写像	計
1920x960	24	無	3/2128	293.2	33.4	326.6
	12		10/1063	290.8	33.0	323.8
	6		122/532	295.3	29.6	324.9
960x480	24		6/2128	80.0	9.1	89.1
	12		19/1063	80.4	9.6	90.0
	6		131/532	86.6	10.4	97.0
480x240	24		281/2128	20.0	3.7	23.7
	12		200/1063	15.2	3.8	19.0
	6		222/532	18.8	3.5	22.3
1920x960	24	有	3/2128	185.4	39.7	225.1
	12		12/1063	183.3	39.4	222.7
	6		119/532	182.8	39.4	222.2
960x480	24		7/2128	43.3	12.5	55.8
	12		20/1063	46.7	11.7	58.4
	6		132/532	44.0	11.8	55.8
480x240	24		297/2128	5.9	7.3	13.2
	12		218/1063	8.4	6.7	15.1
	6		233/532	9.6	5.2	14.8

## 5. まとめ

ウェアラブルな Around View Monitoring を基本コンセプトとするシステムにおける、全方位画像とカメラ姿勢情報を用いた周囲物体の追跡方式を提案した。アルゴリズム検証のための実験を行い、その実現可能性を確認した。今後は実環境での実験を継続して課題を抽出し改良するとともに、ウェアラブル環境での実時間処理の検討、周囲物体の距離測定方式の検討を行う。

### 参考文献

- [1] 清川清, "知能と情報\_バーチャルリアリティにおける視覚提示技術," 知能と情報, Vol.19, No.4, pp.318-325, 2007
- [2] 半田悠貴, 瀬野広夢, 廣山将太郎, 村上兼一, 秦淑彦, "ウェアラブルカメラと SLAM を用いた周囲モニタリングに関する検討," 計測自動制御学会中国支部学術講演会, November 2022
- [3] 迫谷伊織, 片木伸哉, 小竹涼也, 秦淑彦, "全方位カメラを用いたウェアラブル周囲モニタリングシステムに関する検討," 計測自動制御学会中国支部学術講演会, November 2023
- [4] Lu Han, Zushu Li, Heng Zhang, Dongyi Chen, "Wearable Observation Supporting System for Face Identification Based on Wearable Camera," ICCSIT, July 2010
- [5] Hyun-Woo Kim, Min Young Kim, Seung-Ho Yang, Kyo-Yeol Kim, Hyung-Min Son, Yun-Jung Lee, "Smart Wearable Robot Glasses for Human Visual Augmentation Based on Human Intention and Scene Understanding," ISOT2010, October 2010
- [6] Sibio Wang, Ruize Han, Wei Feng, Song Wang, "Multiple Human Tracking in Non-Specific Coverage with Wearable Cameras," ICASSP, 2021
- [7] Wiktor Muron, Maciej Budyś, "motpy - Simple Multi Object Tracking Library," <https://github.com/wmuron/motpy>
- [8] Ultralytics Inc., "YOLOv5," <https://github.com/ultralytics/yolov5>
- [9] Ricoh Co., Ltd, "Ricoh THETA Technology," <https://theta360.com/ja/etc/technology.html>
- [10] WitMotion Shenzhen Co.,Ltd, "WT901BLECL5.0," <https://www.wit-motion.com/BLE/52.html>