協働ロボットのための全方位マルチカメラシステムを用いた人検出

○髙村 悠真(東京大学), Alessandro Moro(東京大学, ライテックス), 中島 慎介(東京大学),

村上 弘樹(不二越),大井 嘉敬(不二越),中井 徹志(不二越),安 琪(東京大学),

山下淳(東京大学)

# Human Detection for Collaborative Robot Using an

# **Omnidirectional Multi Camera System**

○ Yuma TAKAMURA (The University of Tokyo),

Alessandro MORO (The University of Tokyo, RITECS),

Shinsuke NAKASHIMA (The University of Tokyo), Hiroki MURAKAMI (NACHI),

Yoshitaka OOI (NACHI), Tetsushi NAKAI (NACHI), Qi AN (The University of Tokyo),

# Atsushi YAMASHITA (The University of Tokyo)

Abstract: This study discribes a human detection system that integrates cameras and a robot using a new camera arrangement for collaborative robots. In collaborative robots, the safety of collaborative operators is very important due to the nature of sharing workspace. In conventional methods, cameras are installed as a separate system from the robot for safety monitoring, which causes local restrictions. In this study, fisheye cameras are installed on the base of a collaborative robot, thereby removing the aforementioned obstacles. And robot-centered person detection with ROI extraction method using pose estimation is conducted.

# 1. 序論

# 1.1 背景

近年,産業用ロボットの分野において,協働ロボット が多くの関心を集めている.従来の産業用ロボットは人 に重大な危険をもたらす可能性があるため,安全柵で 囲って人とロボットを物理的に分断し,人が安全柵内に 入る場合はロボットを完全停止させる必要がある等,運 用効率の面で問題があった<sup>[1],[2]</sup>.それに対して,協働ロ ボットは安全柵を撤廃することで人とロボットの直接的 なインタラクションを可能にする.これにより,共同作 業者の柔軟性とロボットの正確性,再現性を組み合わせ, 作業の最適化や人にかかる身体的,心理的負荷を軽減す ることが期待されている<sup>[3]</sup>.

しかし,協働ロボットでは人と共に作業するという性 質上,共同作業者の安全性確保が必須の課題であり協働 運転要求事項に定められている「監視された安全停止」 「ハンドガイド」「速度及び隔離監視」「動力及び力の制 限」の4つの規格のうち少なくとも1つを守らなければ ならない<sup>[4]</sup>.そこで,接触前の人検知を実現するために ビジョンセンサを用いて人とロボットの位置関係を把握 する手法が多く用いられている.

## 1.2 先行研究

ビジョンセンサを用いた協働ロボットと人の位置関係 の把握について、協働ロボットを見下ろすように天井に RGBD カメラを設置し、シミュレーション空間上で人を 身長と同じ高さの円柱として近似して 3 次元的な位置関 係を取得する手法が提案されている<sup>[5]</sup>.しかし、この手 法ではカメラの設置作業の負担や、また、カメラ設置場 所の条件によりロボット導入場所についても制約が生じ る. さらに,人を円柱近似すると,人がロボットに対し て腕を伸ばしている場合のように人の姿勢が重要な場合 には対応できない.

一方,ロボット自身にビジョンセンサを設置した研究 も存在する<sup>[6]</sup>.この研究では協働ロボットの肘を挟むよ うにして魚眼カメラを設置して肘部分を境とした全天球 画像を取得し,人が検出された部分を透視投影画像とし て切り出して人の姿勢推定を行っている.しかし,この 研究では人とロボットの位置関係が固定された状況で実 験されている.そのため,人とロボットの位置関係が変 化する場合は想定されておらず,また,人とロボットの 間の距離測定も実施されていない.

## 1.3 研究目的

上記の研究背景及び先行研究の問題点を踏まえて,本 研究では従来研究で実現されていない,「協働ロボット単 体で導入可能な,ロボット周囲の人の姿勢検知と距離測 定を行うビジョンシステムの構築」を目的とする.本シ ステムの構築により,ビジョンセンサ設置場所や設置作 業の負担等協働ロボットの導入の際の障害を取り除くこ とができる.かつ,姿勢検知と距離測定を組み合わせる ことで従来手法よりも人とロボットの実際の状況に近い 位置関係を取得することができ,協働作業者の安全性の 向上が期待される.

## 2. 提案手法

## 2.1 システムの要求仕様

上記の研究目的を達成するために,構築するシステムの要求仕様としては以下の4つが挙げられる.

1. ロボットとビジョンセンサが1つのシステムとして 統合されており,別々に設置する必要がないこと.



Fig. 1 Camera arrangement.

- 2. ロボットの周囲のどの場所にいても人が検出できる ようにロボットの全方位の映像が取得できること.
- 取得した映像中の人の姿勢検知と距離測定により実際の環境に近い人とロボットの位置関係が取得できること.
- システムの処理時間が協働作業者の安全性を確保す るために許容できるものであること.

4. に関して,許容できる処理時間について定量的に論 じる.協働ロボットと人との最小分離距離 *S<sub>a</sub>* は次の式 で表される<sup>[5]</sup>.

$$S_a = v_h(T_r + T_s) + v_r T_r + \frac{v_r T_s}{2},$$
 (1)

ただし、 $v_h$  は人がロボットに向かって直進する速度で ISO13855 に従い 1,600 mm/s とし、 $v_r$  はロボットが停止 する前の速度として 500 mm/s、 $T_s$  はロボットの停止時間 として 400 ms と仮定する.また、ロボットの最大動作半 径を 600 mm として最小分離距離  $S_a$  を 900 mm と仮定す ると、ビジョンシステムとロボット制御システムの合計 処理時間  $T_r$  は 1 フレームあたり 76.2 ms と求まる.本研 究ではロボット制御システムについては考慮せず、 $T_r$  の 値をビジョンシステムの許容処理時間とする.

### **2.2** 提案手法の概要

上記の要求仕様を受けて、本研究では Fig.1 のように 協働ロボットのベース部分にカメラを設置するフレーム を取り付け、フレームの四隅に対角線上外側を向くよう に魚眼カメラ4台を設置する.魚眼カメラは通常のカメ ラに比べて非常に視野角が広く、四隅に設置することで ロボット全方位の映像が取得可能である.また、Fig.2 のように視野角が全方位で重なるようにし、全方位にお いてステレオビジョンを用いた人検出、ロボットを中心 とした人-ロボット間の距離測定を行う.

#### 2.3 カメラキャリブレーションと座標変換

ステレオビジョンを用いた距離測定ではカメラ同士の 位置関係とそれぞれのカメラの内部パラメータを求める 必要がある.そこで、本研究では Zhang の手法を元にし て、まずそれぞれのカメラでチェッカーボードを様々な 角度で撮影し、内部パラメータを求める.その後、ステ レオビジョンを行う 2 つのカメラでチェッカーボードを 撮影してカメラ間の位置関係を求める<sup>[7]</sup>.

また,カメラで撮影された映像から求まる人の位置は カメラ座標系における人の位置である.そのため,世界 座標系での人の位置を得るにはカメラ座標系と世界座標



Fig. 2 Camera field of view's overlap.



Fig. 3 The relationship between camera coordinate and world coordinate.

系の変換行列をキャリブレーションによって求める必要 がある.本研究では Fig. 3 に示すようにロボットベース 中心を世界座標系原点とし,チェッカーボード左上の交 点の座標をロボットベース中心からの実際の距離を用い て指定してその値をキャリブレーションに用いることで ロボットベース中心を原点とした人の位置を取得する.

#### 2.4 画像処理

以下に,取得した映像に関する具体的な画像処理手法 について述べる.今後,Fig.2で表される視野角の重なり について,ある一面の視野角の重なりを考えた時に,例 えば①と②の視野角共通部分を例に挙げると①を左カメ ラ,②を右カメラと呼称する.また,左カメラによって 撮影された画像を左画像,右カメラによって撮影された 画像を右画像とそれぞれ呼称する.

#### **2.4.1** ステレオ平行化と正距円筒変換

ステレオビジョンで距離測定を行う場合,視差を計算 するために撮影した画像を平行ステレオで撮影した場合 と同じ状態の画像に変換する手法が一般的に用いられる. これはステレオ画像の平行化と呼ばれる.しかし,魚眼 画像をステレオ平行化するために歪みを除去して透視投 影画像に変換すると周辺部分の引き伸ばしが生じ,引き 伸ばしが起きている部分だけでなく中心部分でも人検出 が正しく行われない.そこで,先行研究では平行化した





(a) Fisheye image.

(b) Equirectangular image.

**Fig. 4** Fisheye image before conversion and equirectangular image after conversion.



Fig. 5 Relationship between coordinates to be converted.

後の画像を正距円筒画像に変換することでエピポーラ線 を保ちつつ引き伸ばしや歪みの少ない画像を作成してい る<sup>[8]</sup>.

本研究でもこの手法を用いて Fig. 4 のようにエピポー ラ線を保ったまま歪みの少ない画像を生成する.

以下の式及び Fig. 5 に変換を行う座標系間の関係式と その概略を示す.

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} c_x + x f_x \\ c_y - y f_y \\ 1 \end{bmatrix},$$
 (2)

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{X}{Z} \\ \frac{Y}{Z} \\ 1 \end{bmatrix}, \qquad (3)$$

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} \sin \phi \\ \cos \phi \sin \lambda \\ \cos \phi \cos \lambda \end{bmatrix},$$
 (4)

$$\begin{bmatrix} \phi \\ \lambda \\ 1 \end{bmatrix} = \begin{bmatrix} \left(\frac{2u'}{w} - 1\right)\frac{\pi}{2} \\ \left(1 - \frac{2v'}{h}\right)\frac{\pi}{2} \\ 1 \end{bmatrix},$$
 (5)

ただし,(2)式は透視投影画像座標と正規化画像座標の関係式であり, $c_x$ , $f_x$ , $c_y$ , $f_y$ はカメラの内部パラメータ を,(3)式は正規化画像座標と球面座標の関係式を,(4) 式は球面座標と円筒モデル座標の関係式を,(5)式は円筒 モデル座標と正距円筒座標の関係式であり,w,hはそれ ぞれ画像の幅と高さを表す.

(2)~(5) 式を用いて平行化した透視投影画像と正距円 筒画像の対応関係を求め目的の変換を行う.

# 2.4.2 セミグローバルマッチングを用いた人の視差 推定

生成した正距円筒画像を用いた人の視差推定に関して, セミグローバルマッチング(SGM)を用いた視差推定を 行う. SGM とは,2つの画像ペアを用いて画素ごとに独 立に対応点を探索するのではなく,周辺の画素との関係 も考慮しながら全画素に対する最適な視差を求めるステ レオマッチングアルゴリズムである<sup>[9]</sup>.

SGM では、画素の数に応じて処理の複雑さが増加する ため元画像のサイズのまま対応点探索を行うと計算に多 くの時間がかかる.また、視差推定を行いたい領域以外 の画素が存在すると、対応点の探索においてミスマッチ ングを引き起こしやすくなる.そのため、視差を求めた い関心領域 (ROI)のみを抽出することが重要である.

Ziran らはステレオマッチングの際の人領域抽出について,差分画像を取得し動いている人が含まれる矩形領域 を関心領域として抽出したが,背景によるミスマッチン グが生じた<sup>[10]</sup>.

そこで、本研究では、YOLO-Pose<sup>[11]</sup>を用いた姿勢推定 により次の2段階でさらに限定的な人領域の抽出を行う.

- 左右のペア画像から検出された人のバウンディング ボックスの共通部分を抽出し画像サイズの小さい新 たな画像ペアを作成する.
- Fig.6に示すように,推定された各関節,部位(キー ポイント)を中心として,ウィンドウを設定し,そ の部分の画素のみの値を取得する.

この2段階の前処理を行ったのちに SGM による視差推 定を行うことで、ミスマッチングを抑えつつ、人領域の 視差を求めることが可能である.また、有効な視差範囲 を限定することで3次元座標の決定に必要な計算時間を 抑えることもできる.

## 2.4.3 距離測定

生成した視差画像を用いた各画像座標の3次元座標の 決定について,概略を Fig.7 に示す.生成された視差を 正距円筒画像における左右のカメラ座標の方位角の差に 変換すると,次の式のようになる.

$$\Delta \phi = \phi_l - \phi_r$$

$$= \frac{d\pi}{w},$$
(6)

ただし、左カメラ座標系での方位角を  $\phi_l$ 、右カメラ座標 系での方位角を  $\phi_r$ 、視差を d、画像の横幅 (pixel) を d と している.

Fig.7より, 左カメラ座標系から画像中のある点 A ま





(a) Extraction of human regions based on pose estimation.

(b) Actual extracted human area.





**Fig. 7** Relationship between camera coordinate and 3D coordinates.

での距離 D1 は次のように計算できる.

$$D_l = \frac{b\sin(\frac{\pi}{2} - \phi_r)}{\sin(\Delta\phi)}.$$
(7)

また, 左カメラ座標系からみた点 A の 3 次元座標は次 の式で表される.

$$X_{l} = \begin{bmatrix} X_{l} \\ Y_{l} \\ Z_{l} \end{bmatrix} = \begin{bmatrix} D_{l} \sin \phi_{l} \\ -D_{l} \cos \phi_{l} \sin \lambda_{l} \\ D_{l} \cos \phi_{l} \cos \lambda_{l} \end{bmatrix},$$
(8)

ただし、左カメラ座標系での仰角を $\lambda_l$ 、右カメラ座標系 での仰角 $\lambda_r$ 、基線長の長さをbとしている.また、左カ メラ座標系と右カメラ座標系はどちらもステレオ平行化 されたものであることに留意する.

世界座標系からみた点 A の 3 次元座標を  $X_w$ ,世界座 標系からステレオ平行化した左カメラ座標系への変換行 列を M とすると,  $X_w$  は以下の式で表される.

$$X_w = M X_l. \tag{9}$$



Fig. 8 Actual system for experiment and its coordinate system.



Fig. 9 Schematic of the experiment.

## 3. 評価実験

提案するシステムの評価実験として, Fig.8 に示すように, ロボットベースの隣り合う2つの頂点に対角線上外側を向くように魚眼カメラを設置したシステムを用いて提案手法の実装, 評価実験を行った.また, Fig.8 に示すようにカメラの左右および各座標系を定めている.

評価実験では **Fig. 9** に示すように世界座標系原点から 見た人の位置の Y 座標が 2.0 m から 1.0 m と移動する映 像を取得した.取得した動画は 15FPS で総フレーム数は 135,フレームごとの画素数は 1024×1024 であった.

実験について, プロセッサはインテル (R) Core (TM) i7-14700KFを, メモリは 64GB (DDR5-4800)を, GPU は NVIDIA GeForce RTX 4070 SUPER を搭載した デバイスを使用した.また,実験に使用したカメラ は FLIR Grasshopper 3 GS3-U3-41C6C-C, 魚眼レンズは FE185C086HA-1 である.

また、本実験のプログラムは OpenCV ライブラリを用いて python 3.10 で構成されている.

## 4. 結果と考察

4.1 キャリブレーション

実際の左カメラ座標系と世界座標系の位置関係を Fig. 10 に測定した距離および角度とともに示す. 左カメ ラ座標系から世界座標系までの X, Y, Z 軸方向の平行移動 量を  $X_l, Y_l, Z_l$  とし, 順に X 軸に  $\alpha$ , Y 軸に  $\beta$ , Z 軸に  $\gamma$  だ



**Fig. 10** Relationship between left camera coordinate and world coordinate in the experiment.

**Table 1**Comparison of experimental and true values forcoordinate transformations.

		Parallel	l	Angle of		
	transformation (mm)			rotation (deg)		
	$X_l$	$Y_l$	$Z_l$	α	β	γ
Results	-225	-20.9	751	-85.7	-26.7	5.10
Ground Truth	-152	152	770	-90	-45	0

け回転するとする.この時,実験値及び測定値に基づく 真値は Table 1 のようになる.

Table 1 より,実験値と真値を比較すると特に Y 軸方向 の平行移動量及び回転角についての誤差が大きく,座標 変換が正しく行われていないことがわかる.

# 4.2 距離測定

**Fig. 11** に 2.0 m, 1.0 m 地点での距離測定結果および元 映像において対応するフレームを示す.ただし, Fig. 11 中の緑点が人を,黒の点が世界座標系原点を示している. また, **Fig. 12** に各フレームの人の点群の XY 座標の平均 値及び真値の推移を示す.

Fig. 11 に示すように,実際の映像と比較して,距離測 定結果では実際の人の姿勢が正しく復元されていること がわかる.一方,本来 Fig. 9 に示されるように,世界座 標系原点の Y 軸上に人が検出され,原点方向に向かって 軸上を人が直進してくるはずである.しかし,Fig. 11 及 び Fig. 12 に示された結果では,Y 軸上に人が検出され ず,かつ人の進行方向が Y 軸から傾いた方向となり,検 出された距離も実際の世界座標系からの距離とは異なっ ている.これは,前項のキャリブレーションで左カメラ 座標系から世界座標系への変換が正しく行われていない ことが原因である.実際 Table 1 より,今回の距離測定結 果は実際の人の位置を Y 軸方向に-172.9 mm 移動させ, Y 軸から 18.3 度 X 軸方向に傾けた位置となっていること が窺える.

#### 4.3 処理時間

処理時間評価として, (a) 提案手法による ROI 抽出, (b) 差分画像を用いた ROI 抽出<sup>[10]</sup>, (c)ROI 抽出を行わない, の 3 つの場合において 2 章で論じた (i) ステレオ平行化 と正距円筒変換, (ii)ROI 抽出, (iii)SGM を用いた視差推 定, (iv) 距離測定の 4 つそれぞれの 1 フレームあたりの 平均処理時間およびその合計時間の比較を行った. 結果 を Table 2 に示す. ただし, (i) に関しては共通の処理と して処理時間は (a), (b), (c) の平均値を示す.

Table 2 より,提案手法による距離測定は他の 2 つの手法に比べて 4~9 倍ほど処理時間が短く,また,2.1 項で





(a) Distance measurement result at 2.0 m.



(b) Distance measurement result at 1.0 m.

**Fig. 11** Corresponding between distance measurement result and original movie.

**Table 2**Average processing time of the system (ms).

	i	ii	iii	iv	total time
(a) Proposed		9.13	23.9	1.65	37.9
(b) Subtraction	3.25	4.47	74.3	64.6	147
(c) Original		0	239	115	358

論じたシステムの要求処理時間が 76.2 ms であることを 考えると,提案手法が本研究で設定した要求事項を満た す手法として適していることがわかる.

# 5. 結論

本研究では、協働ロボットとカメラを統合した新たな システム及び協働ロボット運用の際の共同作業者の安全 性確保のための姿勢検知および人とロボットの3次元的 位置関係把握手法を提案した.今回の評価実験ではキャ リブレーションの際にカメラ座標系と世界座標系の変換 行列が理論値とは異なり、それに準じて距離測定結果も 実際の実験状況とは異なるものとなったものの、従来手 法よりも詳細な人とロボットの3次元位置関係の把握お よび処理時間の要求を達成することができた.

今後の課題として,カメラ4台を用いた提案手法全体 の実現および検証や,ロボットや他の障害物による人の 遮蔽がある場合への対応が挙げられる.

#### 謝辞

本研究の一部は公益財団法人里見奨学会の援助を受けた.

#### 参考文献

- [1] V. Ales and N. Petr. Human-robot Collaboration in Industry. MM Science Journal, pp. 903–906, (2016).
- [2] H. Roni-Jussi, L. Minna, K. Joni, P. Roel, L. Jyrki, and H. Antti. Review of Vision-based Safety Systems for Human-robot Collaboration. Procedia CIRP, pp. 111– 116, (2018).
- [3] M. Melchiorre, L. S. Scimmi, S. P. Pastorelli, and S. Mauro. Collision Avoidance using Point Cloud Data Fu-



**Fig. 12** Changes in average xy-coordinate values of distance measurement results.(The red dashed line is ground truth and the green scatterplot is results.)

sion from Multiple Depth Sensors: a Practical Approach. 2019 23rd International Conference on Mechatronics Technology (ICMT), pp. 1–6, (2019).

- [4] 榊原伸介. 知能ロボットによる工場自動化と IoT, AI 活用について. システム/制御/情報, pp. 101–106, (2017).
- [5] A.-M. L. María, D.-S. Nicolás, J. R. Daniel, and M. Carol. Vision-Based Safety System for Barrierless Human-Robot Collaboration. Proceedings of the 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 7331–7336, (2022).
- [6] 永谷智貴,湯口彰重,高松淳,小笠原司.小売店での 安全な人間共同作業のための全天球画像視点による ロボットアーム周囲の人間行動観測システム.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 2020, 2A1-G10, (2020).
- [7] Z. Zhang. A Flexible New Technique for Camera Calibration. IEEE Transitions On Pattern Analysis And Machine Intelligence, pp. 1330–1334, (2000).
- [8] 飯田浩貴,池勇勲,梅田和昇,大橋明,福田大 輔,金子修造,村山純哉,内田吉孝.魚眼ステレ オカメラの2眼ステレオとモーションステレオ の融合による距離画像計測.日本機械学会論文集, pp. 19-00069-1-19-00069-14, (2019).
- [9] H. Hirschmüller. Stereo Processing by Semi-Global Matcing and Mutual Information. IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 328–341, (2008).
- [10] Z. He and N. Dahnon. A Contactless Solution for Monitoring Social Distancing: A Stereo Vision Enabled Real-Time Human Distance Measuring System. 2021 10th Mediterranean Conference on Embedded Computing (MECO), pp. 19-00069-1-19-00069-14, (2021).

[11] M. Debapriya, S. Nagori, M. Mathew, D. Poddar, and T. I. Inc. YOLO-Pose:Enhancing YOLO for Multi Person Pose Estimation Using Object Keypoint Similarity Loss. Proceeding of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2637–2646, (2022).