協働ロボットと魚眼ステレオを統合した安全監視システム

Safety Monitoring System via Integration of Collaborative Robots and Fisheye Stereo

○ 髙村悠真（東京大）　Alessandro Moro（東京大，ライテックス）
中島慎介（東京大）　村上弘樹（不二越）　大井嘉敬（不二越）
中井徹志（不二越）　正　安琪（東京大）　正　山下淳（東京大）

Yuma TAKAMURA, The University of Tokyo, yuma-takamura@robot.t.u-tokyo.ac.jp
Alessandro MORO, The University of Tokyo, RITECS
Shinsuke NAKASHIMA, The University of Tokyo
Hiroki MURAKAMI, NACHI
Yoshitaka OI, NACHI
Tetsushi NAKAI, NACHI
Qi AN, The University of Tokyo
Atsushi YAMASHITA, The University of Tokyo

This study proposes a safety monitoring system that integrates fisheye cameras and a collaborative robot. As collaborative robots do not require safety fences, ensuring the safety of human workers is crucial. Conventional vision-based methods have limitations, as cameras and robots are installed separately, which can restrict robot deployment depending on the camera placement. To address this issue, this study installs multiple fisheye cameras on the robot base to achieve overlapping fields of view in all directions. By utilizing stereo vision, distance measurement is performed to obtain the 3D positions of humans. Additionally, to accurately capture detailed human 3D positions, including human limbs, in real time, this study proposes a method that utilizes pose estimation for human region extraction and applies interpolation to disparity missing areas. In validation experiments, the proposed method successfully demonstrated real-time acquisition of human 3D positions corresponding to the actual environment.

Key Words: Collaborative robots, Safety monitoring, Fisheye stereo, Stereo vision, Image processing

1　序論

1.1　研究背景

現在，日本国内の製造業就業者数の減少に伴い，労働力不足を解決する手段とし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自重が重く，かつ高速で動作するため人と接触した際に重大な危険をもたらす可能性がある．そのため安全柵と呼ばれる物理的な障壁を設けて人とロボットを分離し，人が安全柵内に侵入する場合にはロボットを完全停止させる必要があるなど，設備の専有面積や運用効率の面で問題がある [1].

一方で，協働ロボットは安全柵が必要なく，共同作業者とロボットが同じ空間で作業することが可能である．そのため，人とロボットの長所を組み合わせることで，生産性の向上や労働力不足の解消，人の負荷の軽減が期待されている [2].

しかし，協働ロボットは安全柵がないという性質上，共同作業者の安全性の確保が非常に重要な問題である．文献 [3][4] より，ISO10218 では，協働ロボットの運用においては，共同作業者とロボットの間の距離を監視したり，ロボットのトルクを制限したりすることが求められ，現在様々なセンサシステムの活用が研究されている．本研究では，特に接触前に人とロボットの位置関係を画像を用いて把握する手法について着目する．

1.2　先行研究

画像を用いた手法として，文献 [5] では協働ロボットを見下ろすように RGBD カメラを設置し，シミュレーション空間上で人を円柱近似してロボットとの位置関係を把握する手法が提案されている．しかし，カメラをロボットの動作範囲が俯瞰できる場所に設置する必要があり，ロボットの導入についてカメラの設置場所の制約による限定性が生じる．

そこで，カメラとロボットを統合した手法として，文献 [6] では，ロボットのベース部分に魚眼カメラを設置することで，ステレオビジョンを用いた距離測定による人の 3 次元位置の把握を実現している．しかし，視差が推定できない領域は位置情報が欠損してしまい，人とロボットの位置関係の把握に影響を及ぼすという問題がある．

Table 1 Paramaeters of the system.

パラメータの種類	内容	値
S_a	人とロボットの間の最小分離距離	0.9 m
v_h	共同作業者がロボットに向かう速度	1.6 m s^{-1}
v_r	ロボットが停止命令直前に動く速度	0.5 m s^{-1}
T_s	ロボットが停止命令を受けてから完全停止するまでの時間	400 ms

以上より，本研究の目的を，視差の欠損に頑健な，カメラとロボットを統合した人の位置把握システムの構築と評価とする．

1.3　問題設定

本研究では，協働ロボットに対してロボットの動作範囲外から共同作業者がロボットに向かって接近するという状況を想定する．まず，協働ロボットの周囲の領域の安全性について，文献 [5] では，人とロボットが離れていなければならない最小の距離 S_a を式 (1) で表している．

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

この式に表 1 に示す各種パラメータを代入することで処理時間 T_r の上限を 1 フレームあたり 76.2 ms と定めた．また，共同作業者とロボットの間の最小距離については，処理のリアルタイム性を維持しつつ，表 1 に示した停止時間内の相対移動距離と最小分離距離を考慮した際に，十分な安全マージンが確保されるように真値の 10 %の誤差を許容することとした．

2　提案手法

本研究では，文献 [6] で提案された手法をベースとし，加えて，視差の欠損領域に対しては人の身体の長さと画像上での座標を用

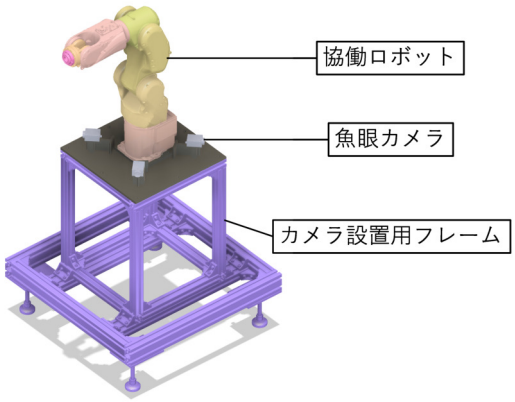

Fig. 1 Camera placement on the collaborative robot.

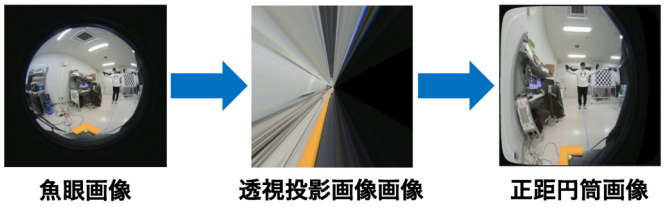

Fig. 2 Conversion of fisheye image to equirectangular image.

いて位置情報を補間する手法を提案する．ベースとなる文献 [6] の手法については 2.1–2.4 節に，補間処理の詳細については 2.5 節で述べる．

2.1 協働ロボットへのカメラ配置

本研究では図 1 に示すように，協働ロボットのベース部分にカメラを設置するためのフレームを作成し，対角線上外側を向くようフレーム上に魚眼カメラを四隅に設置する方法を提案する．視野角が 180° を超えるような広角な魚眼カメラを用いることで，ロボット全周での視野角の重複を実現し，ステレオビジョンを用いた全周での距離測定を行うことができる．

2.2 正距円筒画像への変換を用いた前処理

図 1 のシステムを用いて取得した魚眼画像ペアを用いてステレオビジョンによる距離測定を行う際には，ステレオ平行化を行うために魚眼画像の歪みを除去して透視投影画像に変換する必要がある．しかし，透視投影画像では図 2 中央のように引き伸ばしが大きく，マッチングによる視差推定が正しく行われない．そこで，図 2 右に示すように，文献 [7] で提案されている手法を利用し，平行化後の透視投影画像を正距円筒画像に変換することで，引き伸ばしの少ない画像を生成する．

2.3 姿勢推定を用いた人領域抽出

ステレオマッチングでは片方の画像のある注目画素に対して，もう片方の画像中で対応点を探索する．そのため，画像のサイズに応じて計算時間が増加し，また，視差推定を行いたい領域以外が存在すると，探索の際にミスマッチングを引き起こしやすくなる．そこで，本研究では文献 [8] で提案されている深層学習ベースの姿勢推定である，YOLO-pose を用いて以下の手順で人領域抽出を行い，人領域部分のみでマッチングが行われるようにし，視差情報を求める．また，その概要を図 3 に示す．

1. 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，画像サイズの小さな新たな画像ペアを作成する．この処理により，マッチングで対応点を求める際に走査する領域を削減できる．
2. 推定されたキーポイントを中心としたウィンドウを設置し，その領域内の画素のみを有効化する．この時，有効化された領域以外の画素値を左画像では 0，右画像では 255 にすることで，人領域において背景部分とのミスマッチングを防ぎつつ，背景部分同士でのマッチングを防ぎ，人領域のみに視差の生成を限定できる．

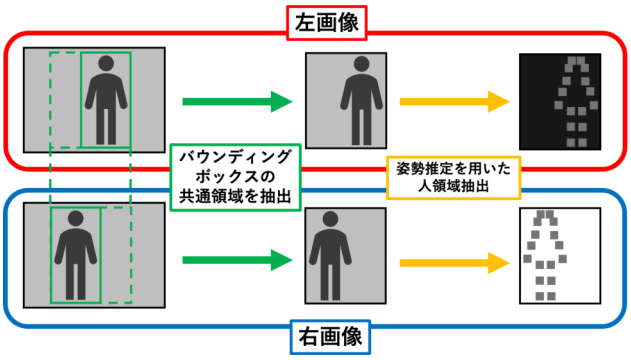

Fig. 3 Extraction of human region.

2.4 マッチングによる視差推定と 3 次元位置決定

2.3 節で作成した人領域抽出後の画像ペアを元に，マッチング処理を行い視差を推定する．視差推定のためのマッチング手法として，文献 [9] で提案されているセミグローバルマッチング（SGM）を用いる．SGM では，対応する画素同士だけではなく，周辺の画素との関係も考慮しながら最適な視差を求めるため，滑らかさと一貫性を持ちながら，不連続な場所も適切に表現することができるという利点が存在する．さらに，テクスチャの乏しい領域においても周囲の画素の情報を活用することで視差推定の精度が向上する．

こうして推定された視差情報を元に，式 (2) を用いて，平行化後の左カメラ座標系における，ある点の 3 次元位置を決定する．

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} \frac{b\cos\phi_r}{\sin(\phi_l-\phi_r)}\sin\phi_l \\ -\frac{b\cos\phi_r}{\sin(\phi_l-\phi_r)}\cos\phi_l\sin\lambda_l \\ \frac{b\sin\phi_r}{\sin(\phi_l-\phi_r)}\cos\phi_l\cos\lambda_l \end{bmatrix}, \tag{2}$$

ここで，b は左右カメラ間の基線長の長さ，ϕ_l, ϕ_r はそれぞれ左右カメラ座標系における方位角，$\phi_l - \phi_r$ が視差情報を方位角の差に変換したもの，λ_l は左カメラ座標系における仰角である．最終的に，平行化後の左カメラ座標系における 3 次元位置を世界座標系に変換することで，世界座標系から見たある点の 3 次元位置を求める．

2.5 人情報と画像情報を用いた位置情報補間

ステレオビジョンを用いた距離測定では，視差がうまく推定できない領域は距離の推定も難しく，その結果人とロボットの間の最小距離が実際よりも大きく推定される可能性がある．そのため，視差が求まらない領域に対して位置情報を補間する処理が必要である．まず，補間処理の判定として，2.4 節で設置したウィンドウごとに有効な視差値を持つ画素数の割合を計算し，閾値以下の場合には補間処理を行うように定める．今回，閾値の決定は手動で行った．

図 4 に補間処理の概略を示す．キーポイント A（以下 A）とキーポイント B（以下 B）は隣接するキーポイントとし，A は補間処理が不要であり，B について補間処理を行う場合を考える．AB 間の距離を L_{joint}，カメラ座標系から A までの視線ベクトルを $\boldsymbol{P}_{\text{known}}$，B までの視線ベクトルを $\boldsymbol{P}_{\text{unknown}}$ とし，その単位ベクトルを $\boldsymbol{u}_{\text{unknown}}$ とする．このとき，$\boldsymbol{P}_{\text{unknown}}$ は求める値であり未知だが，画像座標からその単位ベクトル $\boldsymbol{u}_{\text{unknown}}$ は既知であり，以下のベクトル方程式が成り立つ．

$$\|\boldsymbol{P}_{\text{known}} - t\boldsymbol{u}_{\text{unknown}}\| = L_{\text{joint}}, \tag{3}$$

この方程式を解いて未知数である t を求めることで B の 3 次元位置である $\boldsymbol{P}_{\text{unknown}}$，すなわち $t\boldsymbol{u}_{\text{unknown}}$ を推定する．

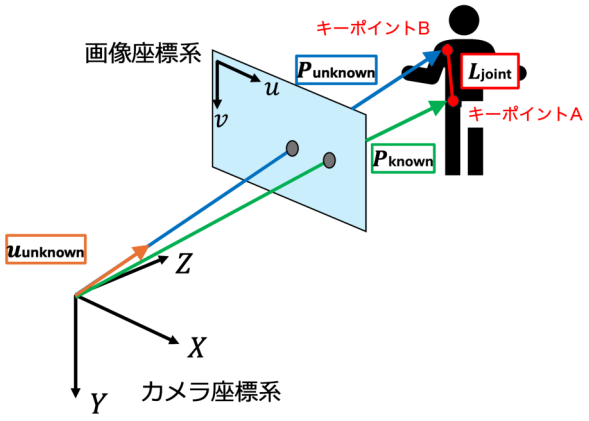

Fig. 4 Interpolation method of 3D position.

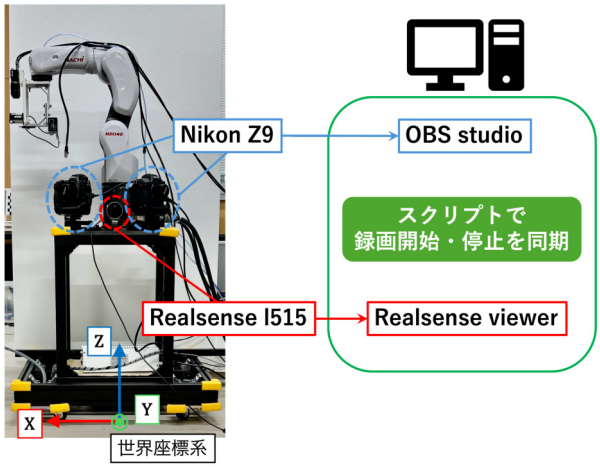

Fig. 5 Experimental setup.

3 検証実験

実験では，カメラは Nikon Z9（画像サイズ 1024×1024）を，レンズは Laowa 4mm F2.8 fisheye（FoV 210° ）を使用し，RGB 画像を取得した．また，人の 3 次元位置の真値を取得するために Realsense l515 を使用して深度データの取得を行った．それぞれのカメラを図 5 のように設置した．

RGB 画像の取得には Nikon Z9 を PC に接続し，OBS studio を用いて 2 台のカメラを同時に録画した．また，深度データの取得には Realsense l515 を同様に PC に接続し，Realsense viewer を用いて点群データを取得した．ステレオビジョンで人の 3 次元点群を取得し，かつそれに対応する真値を取得するには，上記で挙げた 3 台のカメラのフレーム間のずれを最小限に抑える必要がある．そこで，録画開始・停止の同期を確保するため，スクリプトを用いて両ソフトの操作を統一し，フレーム間のずれを最小限に抑えた．その概略を図 5 に示す．

映像として，定量評価のために人が Realsense l515 の視野角内で自由に動く様子を撮影し，提案手法の評価を行った．取得した映像は 30 fps で合計フレームは 652 フレームであった．また，今回の実験では図 5 のように世界座標系を設定し，ロボットの先端が $(0, 0, 1.50)$(単位は m) にあると仮定して人とロボットの間の最小距離を求めた．

4 結果

図 6 に，提案手法による人の 3 次元位置の測定結果と真値を，実験に用いた映像中から 300, 400, 500 フレーム目の結果を抜粋して示す．図 6 より，実環境と対応した人の 3 次元位置が取得できていることがわかるが，その一方で距離の誤差が一部顕著にみられる．

Table 2 Comparison of the error with and without interpolation (m).

真値	補間処理あり			補間処理なし		
	平均値	MAE	RMSE	平均値	MAE	RMSE
0.9–1.2	0.912	0.147	0.184	0.985	0.0807	0.0945
1.2–1.5	1.10	0.235	0.256	1.13	0.209	0.225
1.5–1.8	1.31	0.368	0.397	1.37	0.311	0.334
1.8–2.1	1.55	0.391	0.414	1.61	0.327	0.344
2.1–2.4	1.62	0.614	0.644	1.73	0.504	0.512

そこで，提案手法の距離測定の精度について定量的に評価するために，提案手法による点群と真値点群とで今回設定したロボットの手先位置からの最小距離の比較を行った．また，その際に補間処理の有無による最小距離の変化についても比較するため，補間処理を行った場合と行わなかった場合の結果を比較した．真値点群と提案手法の点群における最小距離の差分をプロットしたものを図 7 に示す．また，真値点群における最小距離を距離順に 5 つの区間に分け，真値点群と提案手法の点群における最小距離の平均絶対誤差（MAE）および二乗平均平方根誤差（RMSE）をそれぞれ計算したものを表 2 に示す．

図 7，表 2 より，補間処理を行わなかった場合について，最小距離の真値が小さい領域では要求仕様を満たし得るが，真値が大きくなるほど真値との誤差が大きくなり，要求仕様が満たされないことがわかる．また，MAE と RMSE の値が大きく変わらないことから，これは誤差が外れ値によるものではなく，全体的な誤差が増大していることを意味する．

この原因として，今回用いた正距円筒画像への変換を活用したステレオビジョンによる距離測定の原理的な限界が挙げられる．2.4 節より，正距円筒画像への変換を用いたステレオビジョンでは，平行化後の左カメラ座標系から見たある点の 3 次元座標は式 (2) で表される．ここで，視差 $\phi_l - \phi_r$ を D と再定義し，視差の微小変化量 ΔD による 3 次元座標の微小変化量 $[\Delta X, \Delta Y, \Delta Z]$ を考える．式 (2) を D で微分することにより，式 (4) のように求めることができる．

$$\begin{bmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{bmatrix} = \begin{bmatrix} \frac{\sin(\phi_l+\phi_r)}{b(\sin\phi_l\cos\phi_r)^2} X^2 \Delta D \\ \frac{(\cos\phi_l^2+\cos\phi_r^2)}{b\sin\lambda_l(\cos\phi_l\cos\phi_r)^2} Y^2 \Delta D \\ \frac{(\cos\phi_l^2+\cos\phi_r^2)}{b\cos\lambda_l(\cos\phi_l\cos\phi_r)^2} Z^2 \Delta D \end{bmatrix}. \tag{4}$$

式 (4) より，X, Y, Z 座標の変化量はそれぞれの座標の値の 2 乗に比例していることがわかる．このため，人とロボットの間の最小距離が大きくなるほど，各座標の値も大きくなるため，表 2 のような誤差が生じたと考えられる．

また，補間処理を行った場合では，補間処理を行わなかった場合と比較して MAE および RMSE 両方で全ての区間で誤差が増大していることがわかる．これは図 7 からわかるように，補間によって実際の位置よりもカメラ座標系に対して近い点に点が付加され，その結果全ての区間で最小距離が小さくなることが原因である．実環境での動作を考慮すると，実際よりも最小距離が小さくなることでロボットの動作効率に影響を及ぼす可能性があるため，実環境での動作に向けてはロボットの動作も含めて本手法の有効性を再考する必要がある．しかし，本研究で考慮する安全性の観点においては最小距離が小さく算出されることは問題なく，現時点ではその有効性が示された．

また，提案手法の処理時間についての有効性を検証するために，提案手法と比較手法の間で処理時間の評価を行った．比較手法として，文献 [10] で提案されている差分画像を用いた手法と，関心領域抽出を行わない手法を用いた．比較の際には，(i) 正距円筒画像変換，(ii) 人領域抽出，(iii) マッチングによる視差推定と 3 次元位置決定，(iv) 位置情報補間の 4 つの項目に分け，それぞれ 1 フレームあたりの平均処理時間およびその合計処理時間を

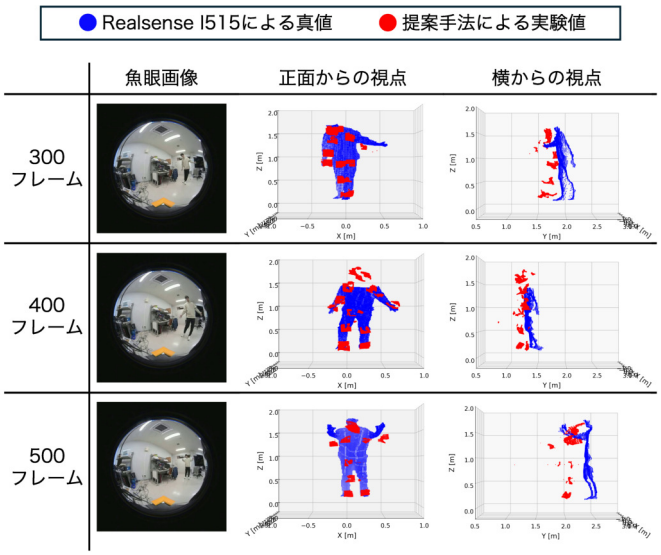

Fig. 6 Compare of the result and the ground truth. The blue scatter means the ground truth and the red scatter means the result of the proposed method.

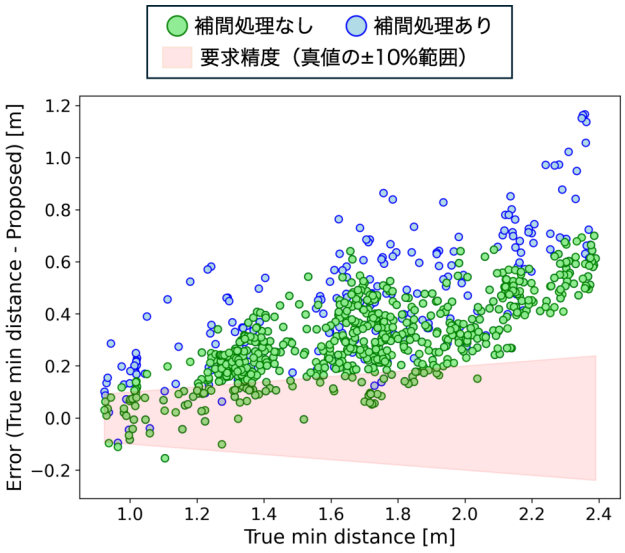

Fig. 7 Compare of the minimun distance error. Blue scatter means the error with interpolation method, green scatter means the error without interpolation method, and pink area means the required accuracy.

Table 3 Compare of the average processing time (ms).

	(i)	(ii)	(iii)	(iv)	合計時間
(a) 提案手法		8.88	32.7	1.22	46.0
(b) 差分画像を用いた手法	3.25	4.47	139	–	147
(c) 関心領域抽出を行わない手法		0	354	–	358

比較した．その結果を表 3 に示す．

表 3 より，提案手法による処理時間は，他の手法と比較した際に 3–8 倍ほど処理時間が短く，処理時間の要求が 76.2 ms であることを考慮すると，提案手法が本研究で設定した要求を満たす手法として適切であることが確認された．

5 結論

本研究では，協働ロボットのための，複数の魚眼カメラとロボットを統合した，視差欠損に頑健な人の位置把握システムを提案した．実験では，提案手法を用いて人の 3 次元位置を取得し，設定したロボットの手先位置との最小距離を求め，真値と比較することで提案手法の有効性を検証した．その結果，最小距離の真値が大きくなるほど誤差が増大するものの，距離が小さい領域では設定した要求を満たし得ることを確認した．また，視差が欠損した領域に対して補間処理を行うことで，安全性の向上が図れることを確認した．最後に，提案手法の処理時間についての評価を行った結果，提案手法が本研究で設定した要求を満たす手法であることが確認された．

今後の展望としては，ステレオビジョンの原理的な限界による距離測定の誤差を改善するため，距離測定については他のセンサを用いることが挙げられる．また，実環境への適用を考え，ロボットと人が共に動作している場合に最小距離を動的に把握し，ロボットの行動を制御するシステムへの応用も展望として挙げられる．

参考文献

[1] A. Vysocky and P. Novak: “Human – Robot Collaboration in Industry,” MM Science Journal, Vol. 2016, No. 02, pp. 903–906, 2016.

[2] M. Melchiorre, L. S. Scimmi, S. P. Pastorelli and S. Mauro: “Collison Avoidance Using Point Cloud Data Fusion from Multiple Depth Sensors: A Practical Approach,” In Proceedings of the 2019 23rd International Conference on Mechatronics Technology (ICMT2019), pp. 1–6, 2019.

[3] ISO: “ISO 10218-1:2011(en), Robots and robotic devices — Safety requirements for industrial robots — Part 1: Robots,” https://www.iso.org/obp/ui/#iso:std:iso:10218:-1:ed-2:v1:en, (accessed on March 3, 2025).

[4] ISO: “ISO 10218-2:2011(en), Robots and robotic devices — Safety requirements for industrial robots — Part 2: Robot systems and integration,” https://www.iso.org/obp/ui/#iso:std:iso:10218:-2:ed-1:v1:en, (accessed on March 3, 2025).

[5] L. M. Amaya-Mejía, N. Duque-Suárez, D. Jaramillo-Ramírez and C. Martinez: “Vision- Based Safety System for Barrierless Human-Robot Collaboration,” In Proceedings of the 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2022), pp. 7331–7336, 2022.

[6] 髙村 悠真, Alessandro Moro, 中島 慎介, 村上 弘樹, 大井 嘉敬, 中井 徹志, 安 琪, 山下 淳: “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,” 第 25 回計測自動制御学会システムインテグレーション部門講演会論文集, pp. 813–818, 2024.

[7] 飯田 浩貴, 池 勇勳, 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, Vol. 85, No. 875, pp.19–00069, 2019.

[8] D. Maji, S. Nagori, M. Mathew and D. Poddar: “YOLO-Pose: Enhancing YOLO for Multi Person Pose Estimation Using Object Keypoint Similarity Loss,” In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW2022), pp. 2636–2645, 2022.

[9] H. Hirschmuller: “Stereo Processing by Semiglobal Matching and Mutual Information,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 30, No. 2, pp. 328–341, 2008.

[10] H. Ziran and N. Dahnoun: “A Contactless Solution for Monitoring Social Distancing: A Stereo Vision Enabled Real-Time Human Distance Measuring System,” In Proceedings of the 2021 10th Mediterranean Conference on Embedded Computing (MECO2021), pp. 1–6, 2021.