

NAIST-IS-MT1651016

修士論文

生活時間測定のための移動ロボットを用いた 人の注視物体推定

井上 智章

2018 年 3 月 9 日

奈良先端科学技術大学院大学
情報科学研究科 情報科学専攻

本論文は奈良先端科学技術大学院大学情報科学研究科に
修士（工学）授与の要件として提出した修士論文である。

井上 智章

審査委員：

小笠原 司 教授	（主指導教員）
萩田 紀博 教授	（副指導教員）
高松 淳 准教授	（副指導教員）
丁 明 助教	（副指導教員）

生活時間測定のための移動ロボットを用いた 人の注視物体推定*

井上 智章

内容梗概

生活の質を向上のために，生活を記録，評価することは重要である．家政学において注目されている生活の質の評価指標に，人の一日を時間の使い方によって分類した生活時間がある．既存の生活時間記録では，一定時間ごとに単一の行動を手作業で記録しているため測定間隔の粗さや手作業による手間や思い込みによる誤入力
が課題となっている．

そこで，本論文では，より詳細な生活時間測定を支援するため移動ロボットを用いて人の注視物体推定を実現する．生活時間の記録行動の対象である「食事」「テレビ鑑賞」は注視対象がわかれば容易に判断可能であり，人の注視している物体を推定することは，より細かな生活時間の測定につながる．提案手法では，ロボットに搭載したカメラを用いて，周囲に存在する人と物体を認識し，人の顔向き推定結果に基いて，人の注視物体の推定する．また，人の顔向き状態によって注視物体推定が困難になった場合には，ロボットの移動によって，推定精度の改善を図る．固定視点では注視物体の推定が困難な場合でも，ロボットが人の顔向きに応じて視点を変更することで推定精度が改善されることが実験により示唆された．

キーワード

移動ロボット，ライフログ，注視対象推定，視点移動，家政学

*奈良先端科学技術大学院大学 情報科学研究科 情報科学専攻 修士論文, NAIST-IS-MT1651016,
2018 年 3 月 9 日.

Gazed Object Estimation for Obtaining Detailed Time-usage Data using a Mobile Robot*

Tomoaki Inoue

Abstract

To improve quality of life (QOL), it is important to record and analyze our daily life. One of the methods to evaluate and classify the QOL is based on the time-usage in daily-life activities. Time-usage studies are drawing attention in the home economics field. There are several problems of current time-usage studies such as low sampling rate of recording, burden of manual logs, and incorrect records. In this thesis, we propose a method to estimate the gazed object using a mobile robot to acquire detailed time-usage measurements. The gazed objects are closely related to the human activities, for example, eating or watching TV. So, if the gazed objects are recorded, the time-usage records are expected to be more precise. In the proposed method, the human and the surrounding objects are recognized using the camera mounted on the robot, and the object gazed by the human is estimated based on the face orientation. When the estimation of the gazed object becomes difficult due to the face orientation, the estimation accuracy can be improved by changing the viewpoint of the robot. The experiments show that the accuracy of the gazed object estimation is improved by changing the viewpoint of the robot based on the face orientation when it is difficult to perform such estimation from a fixed viewpoint.

Keywords:

Mobile robot, Lifelogging, Gazed target estimation, Changing viewpoint, Home economics

*Master's Thesis, Department of Information Science, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT1651016, March 9, 2018.

目次

図目次	v	
表目次	vii	
第 1 章	はじめに	1
1.1	研究背景と目的	1
1.2	関連研究	2
1.2.1	画像を用いた人の生活の振り返りシステム	2
1.2.2	カメラを用いた人の行動推定とサービスロボットへの応用	3
1.2.3	一人称視点をを用いた行動推定と目的推定	3
1.3	本論文の構成	4
第 2 章	生活時間記録の精査と提案手法	5
2.1	生活時間記録の精査	5
2.2	提案手法の概要	6
第 3 章	人の顔向きに基づく注視物体推定	8
3.1	注視物体推定の概要	8
3.2	環境中に存在する物体の認識	10
3.3	人物検出, 追跡	11
3.4	顔向き推定	18
3.5	人の注視物体の判定	23
第 4 章	ロボットの移動による注視物体推定精度の向上	29
4.1	注視物体推定失敗の場合分け	29
4.2	視点移動時におけるロボットの目標位置の設定	32
4.2.1	注視物体推定失敗における注視物体位置の判定	32
4.2.2	注視物体がロボットの視野内に存在する場合	33
4.2.3	注視物体がロボットの視野外に存在する場合	35

第 5 章	ロボットの視点移動を用いた人の注視物体推定実験	40
5.1	システム構成	40
5.2	顔向きに基づいた注視物体推定実験	43
5.2.1	実験の流れ	43
5.2.2	実験結果	44
5.3	ロボットの視野内に存在する人の注視物体推定実験	45
5.3.1	実験の流れ	45
5.3.2	実験結果	45
5.4	ロボットの視野外に存在する人の注視物体推定実験	48
5.4.1	実験の流れ	48
5.4.2	実験結果	51
5.5	考察	55
第 6 章	おわりに	57
6.1	本論文のまとめ	57
6.2	今後の課題と展望	58
謝辞		59
参考文献		60

図目次

1.1	Archival paper for time-usage investigator	2
2.1	The flowchart of the proposed method	7
3.1	The flowchart of the gazed object estimation	9
3.2	Object recognition	10
3.3	Person detection	12
3.4	Person tracking (Front)	15
3.5	Person tracking (Left)	16
3.6	Person tracking (Right)	17
3.7	Extraction of face feature	19
3.8	3D face model	19
3.9	Result of face orientation estimation (Front)	20
3.10	Result of face orientation estimation (Left)	21
3.11	Result of face orientation estimation (Right)	22
3.12	The coodinate of camera and head pose	25
3.13	Projection to the projected point of face orientation in the 2D image	25
3.14	Distance between objects and the projected point obtained from the face orientation	26
3.15	Result of gazed object estimation in each direction	27
3.16	Comparison of estimation results by changing distance between face and object	28
4.1	Robot failed to estimate the gazed object even when the object is within the robot's view	31
4.2	Robot failed to estimate the gazed object when the object when it is outside the robot's view outside the robot's view	31
4.3	Classification of gazed object position using face orientation . .	32

4.4	The gazed object is in the view of robot	34
4.5	Classification of gazed object is outside view	36
4.6	The target position in the case where the gazed object is the robot	37
4.7	The target position in the case where the gazed object is behind of the robot	38
4.8	The target position in the case where the gaze object is side of the robot	39
5.1	Hardware configuration	41
5.2	Software configuration	42
5.3	Experimental setup	43
5.4	Screenshots of robot moving when the gazed object is in the view of the robot	46
5.5	Experimental setup	49
5.6	Experimental layout	49
5.7	Error of object center from the image center	50
5.8	First-Person-View (FPV) device	50
5.9	Capture of gazed object using by FPV device	52
5.10	Screenshots of robot moving when the gazed object is outside the robot's view	53

表目次

2.1	Classification of time-usage in home activity	6
5.1	Specifications of laptop	42
5.2	Result of the gazed object estimation	44
5.3	Comparison of fixed and changing viewpoint of the robot (P_d = -400[mm])	47
5.4	Comparison of fixed and changing viewpoint of the robot (P_d = 400[mm])	47
5.5	Result of the gazed object estimation for outside the robot view (a),(b),(c),(d) are in Fig. 5.5	54

第 1 章 はじめに

1.1 研究背景と目的

生活の改善や振り返りのために人の日常生活を記録するライフログ [1] が注目されている。人の生活の質を定量化し評価することができれば、収集したライフログを人の生活の振り返りや改善に有効に活用することができる。生活の質を評価する代表的な指標にリハビリテーション分野において高齢者の生活機能を評価するために用いられている日常生活動作 [2] や国際機能分類 [3] がある。日常生活行動は、食事・更衣・移動などの生活を営む上で不可欠な基本的行動の可否を点数化している。国際機能分類は、それらに影響を及ぼす環境や個人の特性との関連性を点数化している。これらの評価項目は、生活の質の評価の視点として、人の「行動の能力」に重きをおいている。

家政学分野において生活の実態を調査するために使用される指標として生活時間がある [4]。生活時間では、人の一日の時間の使い方を食事や入浴など生活に必要不可欠な行動「必需行動」、学校や職場など社会的に拘束される時間「拘束行動」、テレビ鑑賞や会話などの「自由行動」に分類し、各行動に費やす時間の配分から生活の実態を解明しようとしている。生活時間は、それぞれの行動をどこで、誰と、あるいは誰のために行われたかなど、行動の目的、意図に着目し、より多角的に人の生活を把握できるとされている。したがって、生活時間を正確に把握することは、人の生活の質の定量的な評価につながると考えている。現在の生活時間調査では、Fig. 1.1 に示す調査票 [5] に 15 分単位の行動を手で記入しており、測定間隔が荒い、手作業による手間や思い込みによる誤記入の発生、複数の作業を同時に行うながら作業の正確な記録が困難という課題がある。

本研究では、カメラを搭載した移動ロボットを用いて、生活時間収集を支援することを目的とする。そのため、環境中に存在する物体と人の顔向きに基づいて人の注視物体を推定する方法を提案する。人が注視する物体を知ることができれば、生活時間に現れる行動を推定することができる。ロボット自身が移動しながら柔軟に視点を変更し、生活の状況を観測することによって、環境中に多数のカメラを設置しなくとも生活を記録することが可能となる。

食事メニューや食卓周辺での人の移動情報を自動で収集している [7].

1.2.2 カメラを用いた人の行動推定とサービスロボットへの応用

Wu らは、RGB-D カメラから取得される人の姿勢情報と深度画像から得られた特徴量に主成分分析と線形判別分析をそれぞれ適用し、圧縮したデータに対し、Support Vector Machine (SVM) を用いることでカメラを設置した机の上での日常生活行動を認識している [8]. サービスロボットに搭載したカメラを用いた人の行動認識の研究例として、Hu らは、「電子レンジで食べ物を温める」などの家庭内での行動を「ドアを開く」、「手を伸ばす」、「物を移動させる」、「設置する」、「ドアを閉める」など行為単位に分割し、行動と行為の関連性を条件付き確率場を用いて学習することで人の行動を認識している [9]. また Koppula らは、人の行動時における関節角と物体との関係をアフォーダンスとして定義することで認識し、食事後の片付け動作、薬の服薬における水の給仕動作、シリアル食品作成動作におけるロボットを用いた補助動作を実現している [10]. さらに Pyo らは、環境中の任意の位置に存在する人に対して、ロボットによる物品取り寄せタスクを行うために人、物体、ロボットの位置をデータベース上で管理することで状況に応じてロボットの行動を計画可能なプラットフォームを提案している [11]. この研究では、環境中の物体位置の取得に物体が格納されている冷蔵庫に内蔵された RFID とロードセルやロボットに搭載したカメラを用いており、人の行動は、環境中に複数台配置した RGB-D カメラから取得した人の位置、姿勢情報に基いて推定している. いずれの研究例においても、人の行動推定は、深度画像から得られる人の姿勢情報に基づいて行われている.

1.2.3 一人称視点を用いた行動推定と目的推定

人の行動をより詳細に記録するために人の見ている対象を一人称視点カメラを用いて推定し、行動を記録する手法が提案されている. Pirsiavash らは、日常生活での日常生活動作を詳細に測定するために一人称視点画像において物体認識を用いた詳細な行動認識手法を提案している [12]. また、Nakashima らは、サービスロボッ

トの水給仕タスクにおいて、一人称視点カメラを用いて推定された人の注視物体推定結果に基づき、人の要求意図推定を行っている。ユーザが食べ物を見ているときは、飲料水、植木を見ているときは、ジョウロの水を提供するなどユーザの意図に応じた提供物の切り替えを実現している [13]。これらの研究例から人が見ている対象には、人の行動や行動意図に関する情報が多く含まれていると考えられる。

1.3 本論文の構成

以下に本論文の構成を述べる。2章で提案手法の概要について述べ、3章では、人の注視物体の推定方法、4章で、ロボットの移動による注視物体推定精度の改善について述べた後、5章にて提案手法の評価実験について述べる。6章では、本論文のまとめと今後の課題について述べる。

第 2 章 生活時間記録の精査と提案手法

2.1 生活時間記録の精査

生活時間調査では、まず日常生活の行動を「必需行動」、「拘束行動」、「自由行動」の 3 つに大別し、それらを「睡眠」、「学業」、「仕事」などの中分類に分類した後、更に細かく分類した 28 種類の小分類について、記録を行っている [14]。本研究では、より詳細な家庭内での生活時間測定を行うために、まず家庭での生活時間記録において対象となる行動を Table. 2.1 に示すように行動名、注視物体、行動場所への関連の有無で分類した。

これらから生活時間測定のための行動認識手法として、人の注視物体を推定し、人の行動と関連付けることが有効であると考えている。まず Table. 2.1 から生活時間測定において記録する行動は、場所への依存の有無に関わらず、人の注視している物体と関連深いことがわかる。また、注視している物体と関連のない行動の多くは、場所との関連がある行動が多く、場所に依存する行動の生活時間は、家庭内での人の位置を推定することによって測定可能である。さらに記録すべき行動の中には、読書やインターネットなど長時間にわたり姿勢の変化が少ない行動が多く、行動推定における姿勢の特徴が乏しいことが考えられる。

また、実際の家庭内において生活時間を記録するためには、測定範囲に制限がなく、装着による負荷がない非接触な測定手法が理想的である。Faulhammer らは環境中に存在する物体の 3D モデルを復元するために移動ロボットを用いている [15]。カメラを搭載したロボットは、移動することにより様々な視点における物体の画像を取得でき、ロボットの自己位置推定により、視点位置の取得も容易である。移動ロボットの自由に視点を変更できる利点は、注視物体の推定時においてもより推定が容易な視点位置に自ら移動することで推定精度の改善にもつながると考えられる。以上のことから本研究では、生活時間測定を支援するため、自由に視点を変更可能な移動ロボットを用いて人の注視物体を推定することを目標とした。

Table 2.1: Classification of time-usage in home activity

行動分類	行動名	注視物体との関連	行動場所との関連
必需行動	食事	あり	なし
	睡眠	なし	あり
	洗顔	なし	あり
	入浴	なし	あり
	トイレ	なし	あり
	着替え	あり	なし
	化粧	あり	なし
拘束行動	炊事	あり	あり
	洗濯	あり	あり
	掃除	なし	なし
拘束行動	テレビ鑑賞	あり	あり
	読書	あり	なし
	新聞	あり	なし
	ラジオ	なし	なし
	インターネット	あり	なし
	休息	あり	なし
	会話	あり	なし

2.2 提案手法の概要

Fig. 2.1 に提案する移動ロボットを用いた人の注視物体推定の流れを示す．ここで、本研究における注視物体とは、人の顔向き方向に存在する物体を示す．提案手法では、移動ロボットが物体検出と人の顔向き推定に基づいて人の注視物体を推定する．まず、一般物体認識によって環境中に存在する物体を認識し、同時に人物位置を推定する．そして、検出された人に対し、人の顔向き推定する．人の顔向きと環境中に存在する物体の位置関係を把握し、人の注視物体推定を実現する．また、ロボットは、現在の位置において注視物体の推定が困難であった場合は、人の顔向

きに応じて視点を変更することによって，注視物体推定精度の改善を図る．

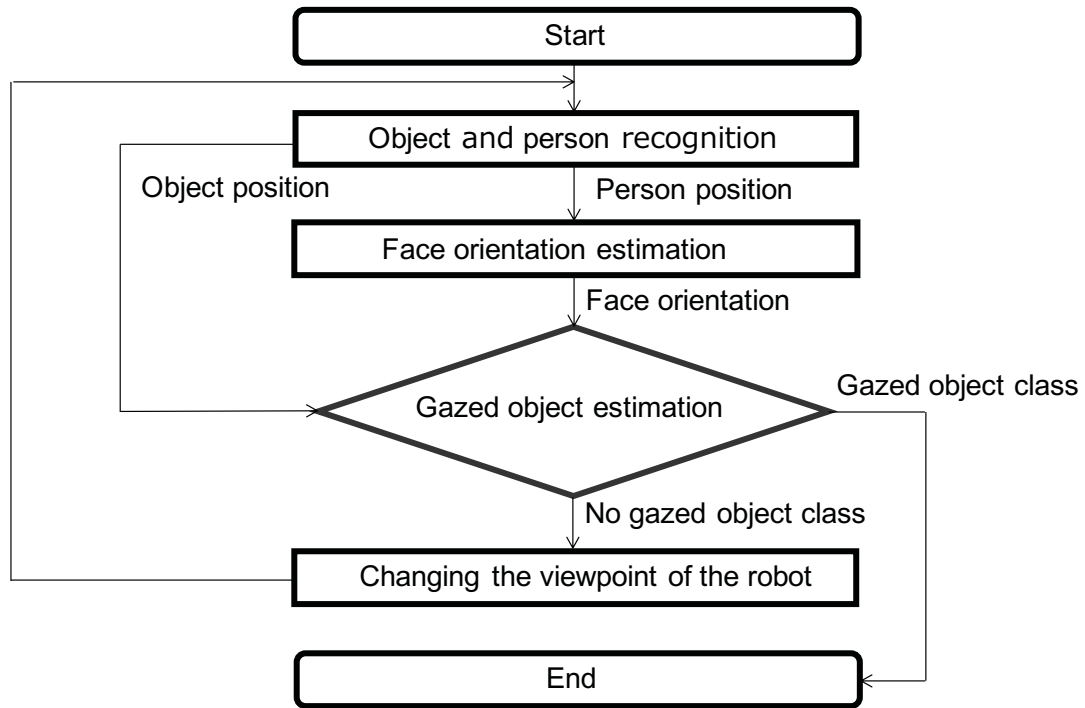


Fig. 2.1: The flowchart of the proposed method

第 3 章 人の顔向きに基づく注視物体推定

3.1 注視物体推定の概要

本研究では，物体認識と人の顔向き推定に基づいて人の注視物体を推定する．Fig. 3.1 に注視物体推定の流れを示す．まず，取得した RGB 画像から人の顔の特徴点を抽出し，頭部の三次元姿勢を求める．次に推定した頭部姿勢における注視位置を画像中に投影することで，画像中において顔向きの方向と顔向きからの奥行きを表現した顔向きベクトルを求める．そして，得られた顔向きベクトルの終点座標と YOLO によって取得した画像中における各物体の重心座標と位置関係に基づいて人の注視物体を推定する．推定された顔向きが画像中における物体位置に対応付けできず，注視物体が推定できなかった場合においては，人の顔向きと画像中の各物体との位置関係に基づいてロボットの視点を移動する．

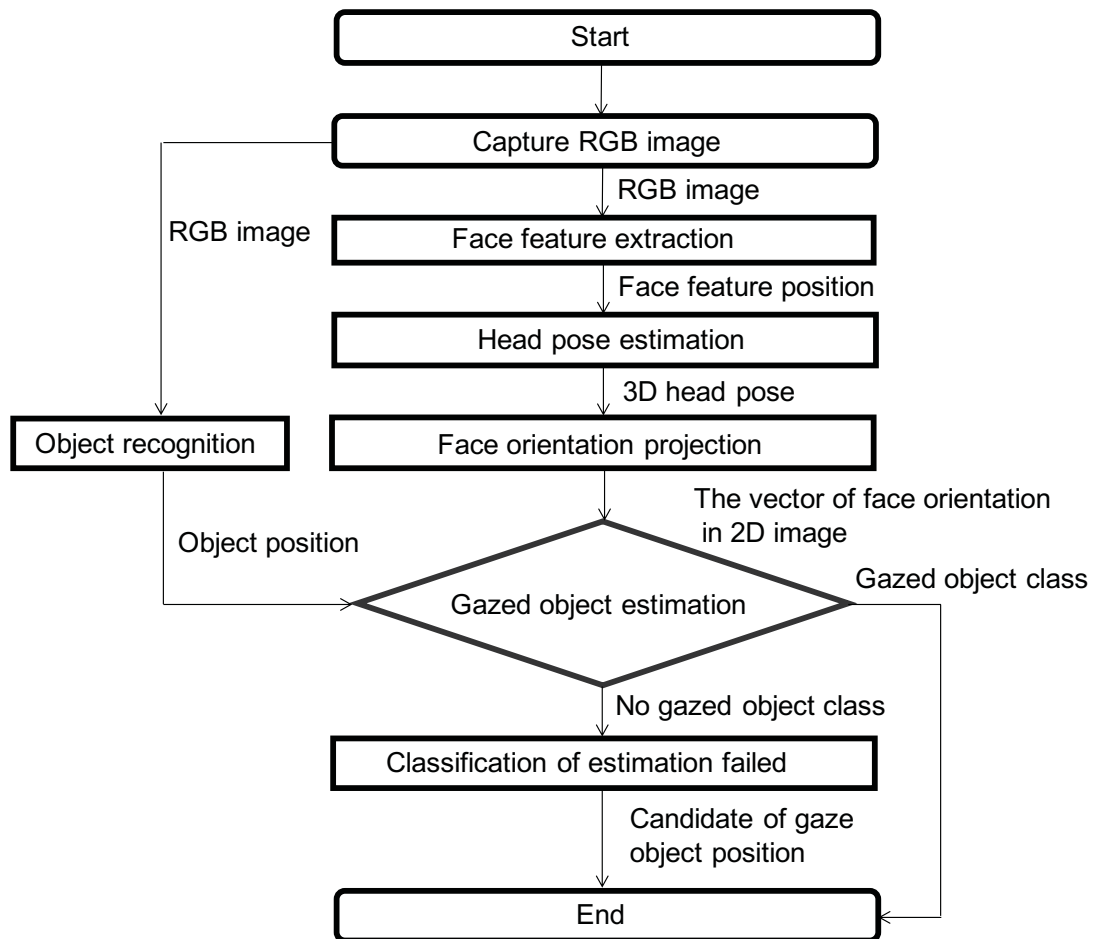


Fig. 3.1: The flowchart of the gazed object estimation

3.2 環境中に存在する物体の認識

人が注視している物体を推定するためには，環境中に存在する物体をリアルタイムに認識する必要がある．本研究では，物体認識に Deep learning を用いた一般物体認識手法の一つである YOLO（You Look Only Once）[16] を用いる．YOLO は，画像全体をグリッド分割し，各領域ごとに物体のクラスとバウンディングボックスを算出することで高速な物体検出を実現している．物体認識のモデルには，Microsoft 社の COCO データセット [17] を用いた学習済みのモデルを使用しており，80 クラスの物体を分類可能である．

Fig. 3.2 に物体認識の結果を示す．この物体認識結果から画像中に存在する Person ラベル以外の n 個の物体の重心座標 (x_{c_n}, y_{c_n}) を算出する．

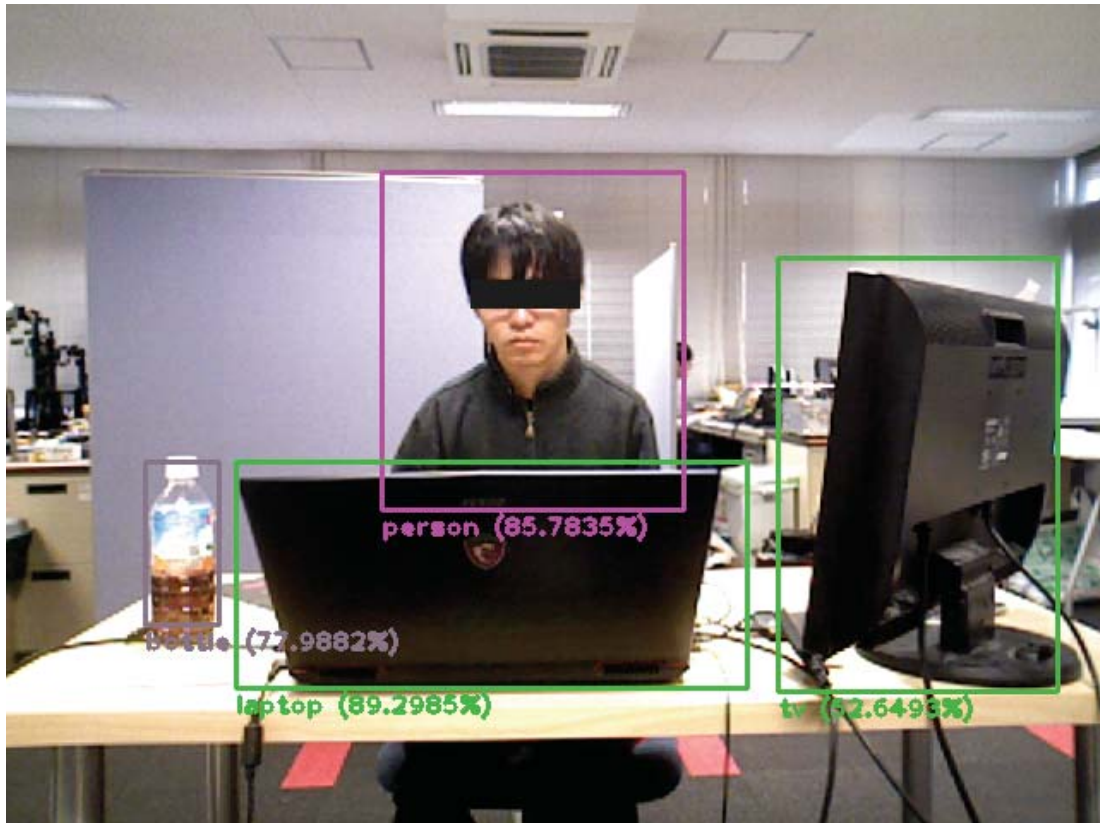


Fig. 3.2: Object recognition

3.3 人物検出，追跡

環境中におけるロボットと人の位置関係を把握するために，人の検出，追跡を利用する．移動ロボットに用いられている人追跡の手法には，主に Laser Range Finder (LRF) を用いた手法 [18] とカメラを用いた手法 [19] がある．LRF を用いた手法は人の脚の点群を認識することで人までの距離を正確に測定できる利点があるが，家庭内のような複雑な環境下では，机や椅子の脚など人の脚の形状と似たものを間違えて検出する可能性がある．

そこで，本研究では，人の検出に Fig. 3.3(a) に示すように YOLO の認識によって取得される RGB 画像中の人物領域を用いる．ロボットの正面を基準とした人物方向 θ_p は，人物領域の重心座標 (x_{cp}, y_{cp}) ，取得される画像の幅 W_{img} ，カメラの画角 W_a を用いて式 (3.3.1) で求められる．また，人物とロボット間の距離 d_p には，Fig. 3.3(b) に示すように深度画像中の人物領域の重心座標における深度値を使用している．

$$\theta_p = \frac{W_a}{2} - \frac{x_{cp} \cdot W_a}{W_{img}} \quad (3.3.1)$$

そして，自己位置推定から得られる世界座標系におけるロボットの自己位置を (x_r, y_r, θ_r) とすると世界座標系における人物位置を表現する同次変換行列は，式 (3.3.2) で表現される座標変換で示すことができる．

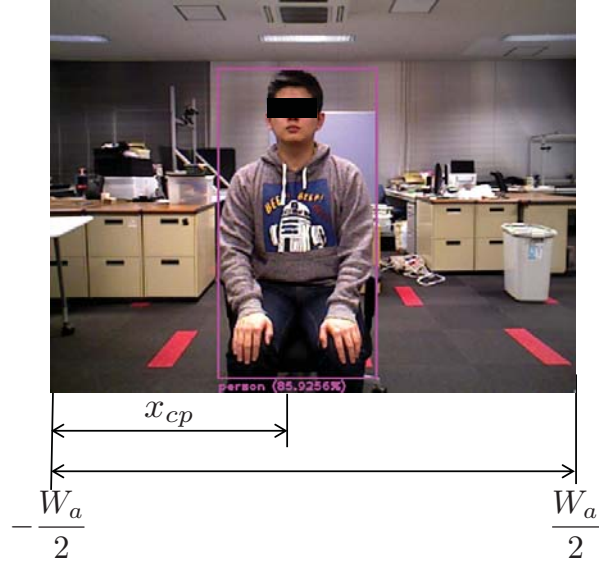
$$\mathbf{M}_{\mathbf{T}_P} = \mathbf{M}_{\mathbf{T}_R} \mathbf{R}_{\mathbf{T}_P} \quad (3.3.2)$$

$$\mathbf{M}_{\mathbf{T}_R} = \begin{pmatrix} \cos \theta_r & \sin \theta_r & x_r \\ \sin \theta_r & \cos \theta_r & y_r \\ 1 & 0 & 1 \end{pmatrix}$$

$$\mathbf{R}_{\mathbf{T}_P} = \begin{pmatrix} \cos \theta_p & \sin \theta_p & d_p \cos \theta_p \\ \sin \theta_p & \cos \theta_p & d_p \sin \theta_p \\ 1 & 0 & 1 \end{pmatrix}$$

$\mathbf{M}_{\mathbf{T}_R}$ は，世界座標系におけるロボット位置を表現する同次変換行列である．また， $\mathbf{R}_{\mathbf{T}_P}$ は，ロボット座標系における人物位置を表現する同次変換行列である．最終的な世界座標系における人物位置 (x_p, y_p) は， $\mathbf{M}_{\mathbf{T}_P}$ から式 (3.3.3) で得られる．

$$\begin{cases} x_p = d_p \cos \theta_r \cos \theta_p + d_p \sin \theta_r \sin \theta_p + x_r \\ y_p = d_p \sin \theta_r \cos \theta_p + d_p \cos \theta_r \sin \theta_p + y_r \end{cases} \quad (3.3.3)$$



(a) Angle of person in the image



(b) Depth image around the person

Fig. 3.3: Person detection

検出された人物の位置に対してカルマンフィルタを適用することで人物の追跡を実現している．状態 $\mathbf{x}$ は式 (3.3.4)，式 (3.3.5) 式で表現される．

$$\mathbf{x}_t = A\mathbf{x}_{t-1} + \epsilon_t \quad (3.3.4)$$

$$A = \begin{pmatrix} 1 & 0 & \Delta t & 0 & \frac{1}{2}\Delta t^2 & 0 \\ 0 & 1 & 0 & \Delta t & 0 & \frac{1}{2}\Delta t^2 \\ 0 & 0 & 1 & 0 & \Delta t & 0 \\ 0 & 0 & 0 & 1 & 0 & \Delta t \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\mathbf{x}_t = (x_p \quad y_p \quad v_{xp} \quad v_{yp} \quad a_{xp} \quad a_{yp})$$

$$z_t = C\mathbf{x}_t + \delta_t \quad (3.3.5)$$

$$C = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

A は，システム行列を示している，状態 $\mathbf{x}$ は，(3.3.2) 式で算出されたロボットのマップ上での人の位置 (x_p, y_p) ，に対して等加速度モデルを適用している． (v_{xp}, v_{yp}) は人の速度， (a_{xp}, a_{yp}) は人の加速度である． ϵ_t はシステムノイズ， C は観測行列， δ_t は観測ノイズである．予測ステップにおける事前推定値 $\hat{\mathbf{x}}_t$ ，事前誤差共分散 $\hat{P}_t$ はそれぞれ式 (3.3.6)，式 (3.3.7) により与えられる．

$$\hat{\mathbf{x}}_t = A\hat{\mathbf{x}}_{t-1} \quad (3.3.6)$$

$$\hat{P}_t = A\hat{P}_{t-1}A^T + Q \quad (3.3.7)$$

ここで， P は誤差共分散， Q はシステムノイズの分散であり正規分布に従う．そして，フィルタリングステップにおけるカルマンゲイン K_t ，事後誤差共分散 $\hat{P}$ ，人の

位置の推定値である事後推定値 $\hat{\mathbf{x}}_{t+1}$ はそれぞれ式 (3.3.8), 式 (3.3.9), 式 (3.3.10) により与えられる.

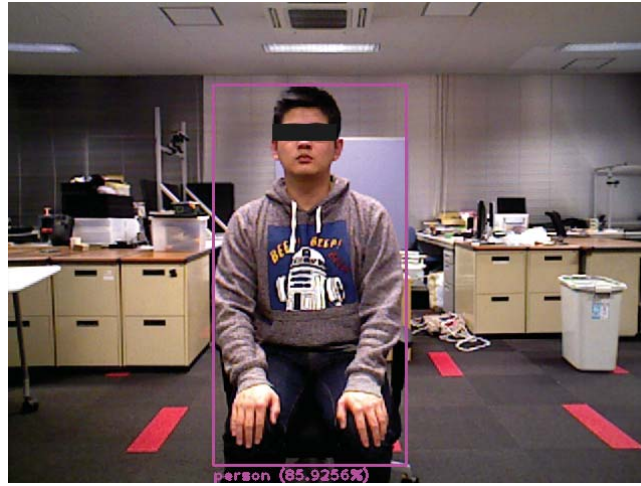
$$K_t = \hat{P}_t C^T (C \hat{P}_t C^T + R)^{-1} \quad (3.3.8)$$

R は, 観測ノイズの分散であり正規分布に従う.

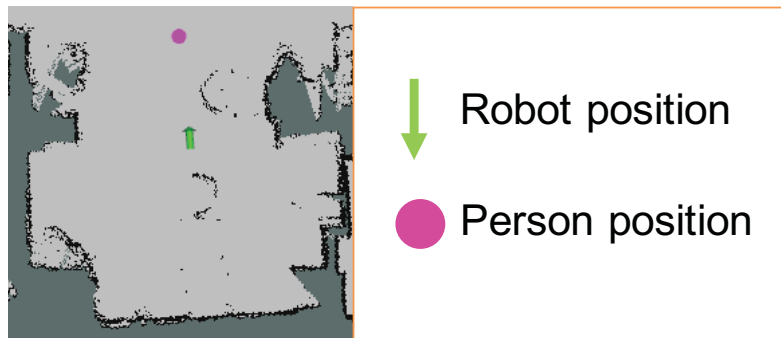
$$\hat{P} = (I - K_t C) P \quad (3.3.9)$$

$$\hat{\mathbf{x}}_{t+1} = \hat{\mathbf{x}}_t + K_t (Z_t - C \hat{\mathbf{x}}_t) \quad (3.3.10)$$

Fig. 3.4, Fig. 3.5, Fig. 3.6 に人がロボットの正面, 左側, 右側に立った場合におけるカルマンフィルタによる人物追跡結果を示す. Fig. 3.4(b), Fig. 3.5(b), Fig. 3.6(b) 中の緑色の矢印はロボットのマップ中における位置姿勢を示しており, 赤い丸は, 人物位置の推定結果を示している. これらより, 人の動きに対して人物位置の推定値が追従していることからカルマンフィルタにより人を追跡できていることがわかる.

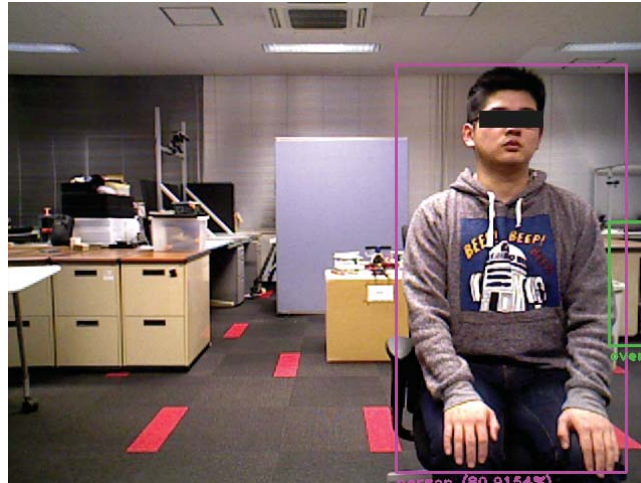


(a) Person detection

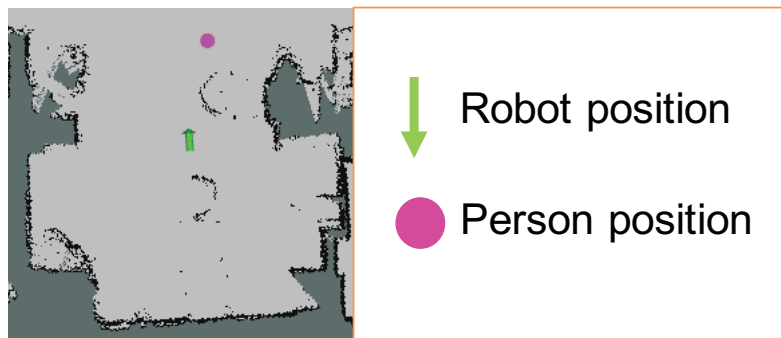


(b) Person position on the robot's map

Fig. 3.4: Person tracking (Front)

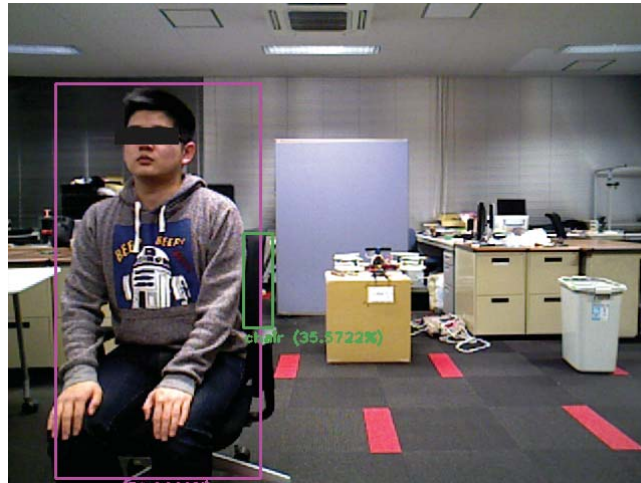


(a) Person detection

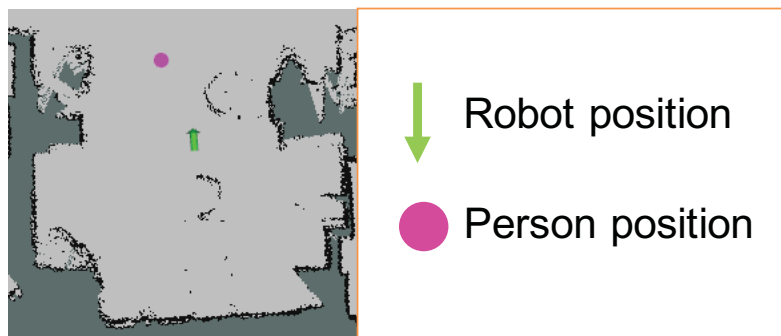


(b) Person position on the robot's map

Fig. 3.5: Person tracking (Left)



(a) Person detection



(b) Person position on the robot's map

Fig. 3.6: Person tracking (Right)

3.4 顔向き推定

人の顔の特徴点を利用して顔向きを推定する．顔の特徴点の抽出には，顔認識ライブラリである OpenFace [20] を使用している．Fig. 3.7 に OpenFace を用いて抽出した顔の左目端，右目端，鼻，左口端，右口端の特徴点を示す．この各特徴点の位置座標と Fig. 3.8 に示す標準的な人の顔の三次元モデルとの位置合わせによって，カメラ座標系における頭部の三次元姿勢を推定する．この三次元姿勢は，既知の点である顔の三次元モデルと観測した顔の特徴量位置との複数の対応を扱った Perspective-n-Point problem (PnP 問題) を解くことで推定可能である [21]．最終的な人の顔向き方向は，カメラ座標系における頭部姿勢をオイラー角で変換することで求めることが可能である．Fig. 3.9, Fig. 3.10, Fig. 3.11 に人がロボットに対して正面を向いた場合，左側を向いた場合，右側を向いた場合における顔向き推定結果を示す．Fig. 3.9(b), Fig. 3.10(b), Fig. 3.11(b) 中の青矢印は，ロボットの地図上での人の顔向きを二次元平面で表現しており，ロボットの地図上でのロボットの自己位置，人物位置，人の顔向きの位置関係を表現できていることがわかる．

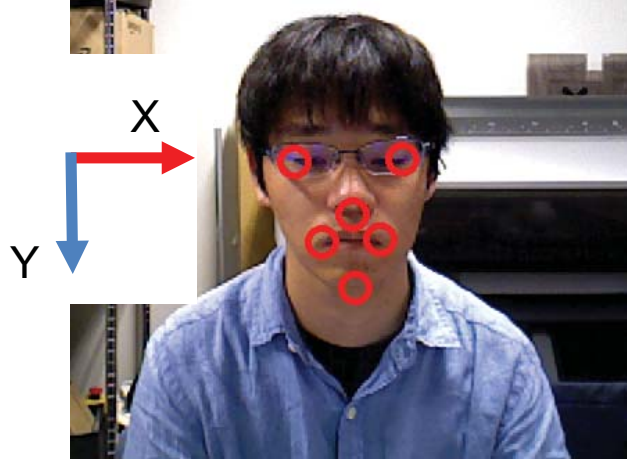


Fig. 3.7: Extraction of face feature

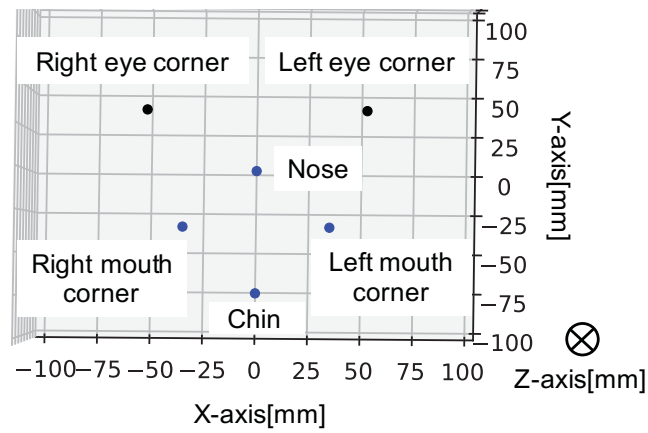
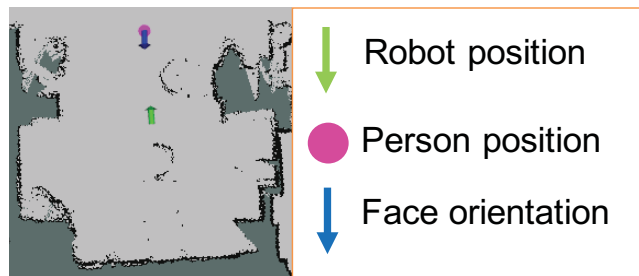


Fig. 3.8: 3D face model



(a) RGB image

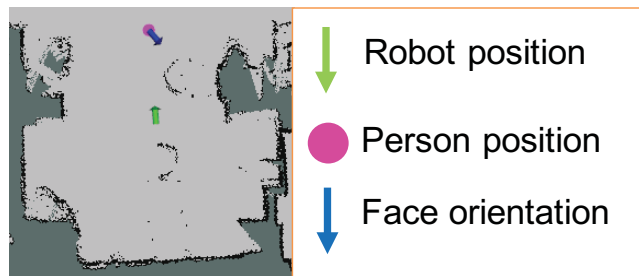


(b) Face orientation on the robot's map

Fig. 3.9: Result of face orientation estimation (Front)



(a) RGB image

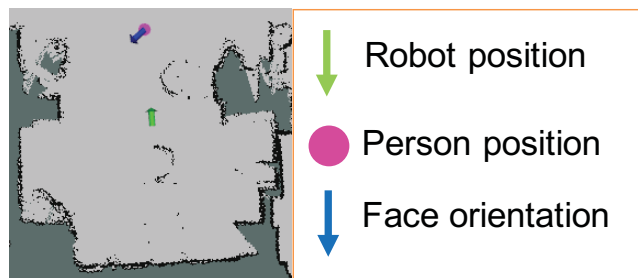


(b) Face orientation on the robot's map

Fig. 3.10: Result of face orientation estimation (Left)



(a) RGB image



(b) Face orientation on the robot's map

Fig. 3.11: Result of face orientation estimation (Right)

3.5 人の注視物体の判定

顔向きに基づいて注視物体を判定するために Fig. 3.12 に示すようにカメラ座標系 Σ_c における注視位置と物体との位置関係を算出する．そこで，推定した頭部姿勢座標系 Σ_h において，顔向き方向である Z 軸上の点 $(0, 0, d_{proj})$ を二次元画像中に投影することで，Fig. 3.13 に示すカメラ座標系での顔向きの方向と奥行きを画像中で表現した顔向きベクトルを求める．ここで，顔向きベクトルの始点は，画像中における顔の鼻位置座標，終点は，画像中に頭部姿勢座標系上の点 $(0, 0, d_{proj})$ を投影した座標 (x_f, y_f) であり，式 (3.5.1) によって求められる．

$$\begin{pmatrix} x_f \\ y_f \\ 1 \end{pmatrix} = A_c R|t \begin{pmatrix} 0 \\ 0 \\ d_{proj} \\ 1 \end{pmatrix} \quad (3.5.1)$$

$$A_c = \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix}$$

$$R|t = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{pmatrix}$$

A_c は，カメラキャリブレーションによって求めたカメラの内部パラメータで (f_x, f_y) は焦点距離， (c_x, c_y) は光学中心である．また， $R|t$ はカメラ座標系 Σ_c における頭部三次元姿勢を表現した剛体変換である． d_{proj} の大きさは，画像中において顔向きベクトルのノルムの大きさに反映される．したがって，画像中における物体位置とのユークリッド距離が最小になる顔向きベクトルの終点座標 (x_f, y_f) が得られる d_{proj} は，カメラ座標系 Σ_c における顔と注視物体との距離となる．

注視物体推定では，画像中で注視位置に最も近い物体を注視物体として判定するため，Fig. 3.14 に示すように点 $(0, 0, d_{proj})$ を変更しながら顔向きベクトルの終点座標 (x_f, y_f) と各物体の重心座標 (x_{c_n}, y_{c_n}) とのユークリッド距離 r_{ob_n} が最小となる顔向きベクトルの終点座標 $(x_{f_{min}}, y_{f_{min}})$ を求める．認識された全ての物

体と顔向きとの終点座標 $(x_{f_{min}}, y_{f_{min}})$ のユークリッド距離の最小値 $r_{ob_{min}}$ は、式 (3.5.2) で求められる。

$$r_{ob_{min}} = \min(\sqrt{(x_{c_n} - x_{f_{min}})^2 + (y_{c_n} - y_{f_{min}})^2}) \quad (3.5.2)$$

$$s.t. \quad 0 < d_{proj} \leq d_{max}$$

ここで、 d_{proj} の最大値 d_{max} は人の腕のおおよその長さである 650 [mm] とした。算出された $r_{ob_{min}}$ が閾値 T_d より小さい場合には、 $r_{ob_{min}}$ となる物体を注視物体と判定する。Fig. 3.15 に注視物体推定結果の一例を示す。Fig. 3.15 は、顔向きに応じてペットボトル、ノートパソコン、モニターを注視した場合における推定結果を示している。Fig. 3.15 中の赤点は、 $r_{ob_{min}}$ となる顔向きベクトルの終点座標、青点は、注視判定された物体の重心座標、青線は、顔の鼻位置座標から注視物体重心までのベクトル、緑線は、 $d_{proj}=650$ [mm] における顔向きベクトルを示している、Fig. 3.15 の各場面において $r_{ob_{min}}$ となる顔向きベクトルの終点座標が顔向きベクトルの方向において物体の重心位置に最も近い顔向きベクトルの終点座標となっていることから人の顔向きとの推定結果に基づいて注視物体を推定できていることがわかる。

また、Fig. 3.16 に示すスマートフォンを注視している例では、Fig. 3.16(a) の状態を中立とすると Fig. 3.16(b)、Fig. 3.16(c) における顔向きベクトルの終点座標は、Fig. 3.16(b) に示すように物体を顔に近づけた場合は、顔に近づき、Fig. 3.16(c) に示すように物体を遠ざけた場合は、顔から遠ざかっていることから、カメラ座標系において顔と物体との距離を考慮した注視物体推定を実現できていることがわかる。

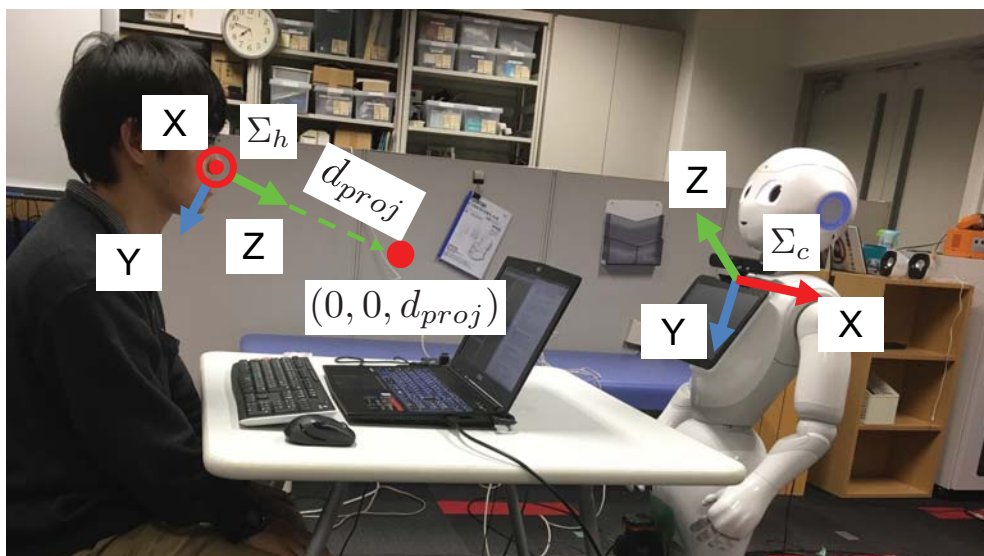


Fig. 3.12: The coodinate of camera and head pose



Fig. 3.13: Projection to the projected point of face orientation in the 2D image

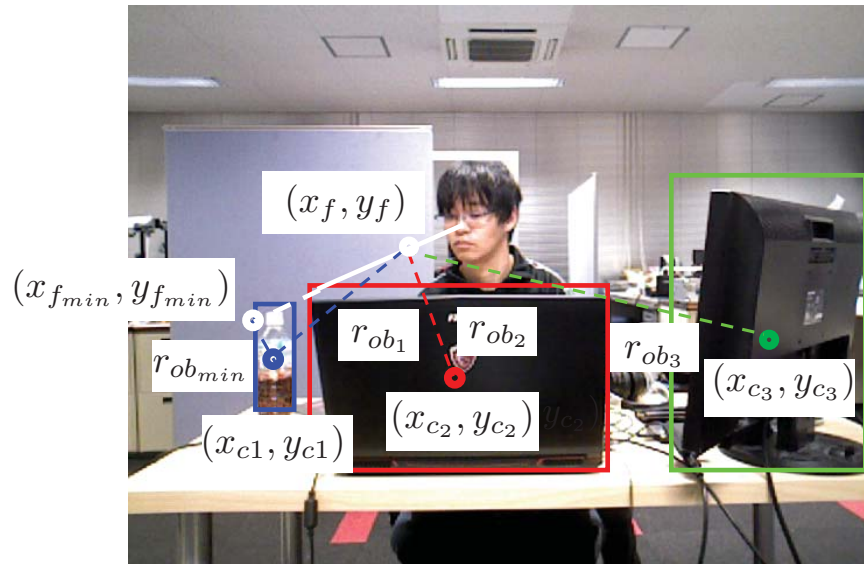


Fig. 3.14: Distance between objects and the projected point obtained from the face orientation



(a) Front (laptop)

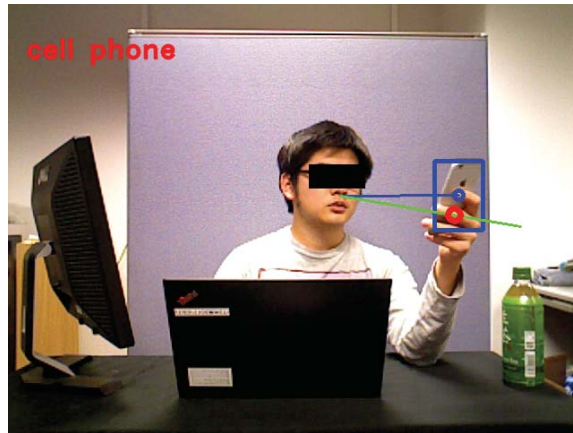


(b) Left (monitor)

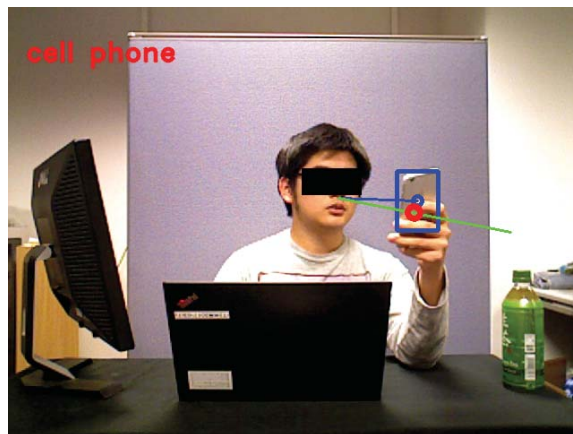


(c) Right (bottle)

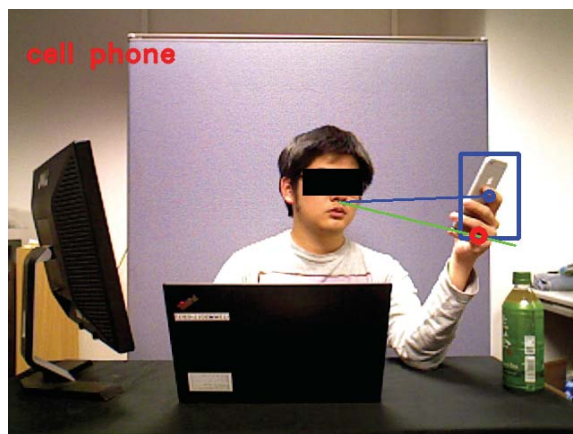
Fig. 3.15: Result of gazed object estimation in each direction



(a) Neutral



(b) Close



(c) Far

Fig. 3.16: Comparison of estimation results by changing distance between face and object

第 4 章 ロボットの移動による注視物体推定精度の向上

4.1 注視物体推定失敗の場合分け

提案手法では、注視物体を推定ができなかった場合において移動によってロボットが適切に視点を変更することで、推定精度の改善を図る。ここで、注視物体推定における失敗とは、実際に人が環境中に存在する物体を注視している場面において、ロボットがその物体を注視物体と認識できないことを示す。注視物体推定は、環境中の物体認識と人の顔向き推定結果に基いて実現されている、そのため、注視物体が失敗の原因は、それぞれの結果が影響していると考えられる。そこで、以下のように注視物体推定が失敗した場面における場合分けを行う。

(I) 物体認識に起因する場合

- (a) ロボットの視野内に存在する物体を正しく認識できていない場合

(II) 顔向き推定に起因する場合

- (a) 顔向き推定に失敗している場合
- (b) 顔向き推定には成功しているが、顔向きと物体の位置関係のよってロボットの視野内に存在する注視物体を正しく注視物体判定できていない場合
- (c) 人の注視物体がロボットの視野外に存在する場合

まず、(I)(a)、(II)(a) は、物体・人検出に失敗した場合、つまり存在するものをしてしないと判定した場合である。この場合において注視物体推定精度を向上させるためには、物体認識と顔向き推定のそれぞれの性能を向上させる必要がある。(II)(b) は、Fig. 4.1 に示すように物体の重心座標 (x_{c_n}, y_{c_n}) との顔向きの終点座標 $(x_{f_{min}}, y_{f_{min}})$ とのユークリッド距離 $r_{ob_{min}}$ が閾値 T_d より大きい場合である。この場面では、顔位置と注視物体の位置が離れているため、 d_{proj} の最大値における r_{ob_n} が閾値 T_d より小さくならず、注視物体として推定されていない。(II)(c) は、Fig. 4.2 に示すように人がロボットの視野外に存在している物体を注視している場合である。この場面では、人の注視物体はロボットに視野内に存在しないため、現在の画像だけからは、注視物体推定は不可能である。したがって、(II)(b)、(II)(c) における注視物体推定が失敗する原因は、環境中のロボットの視点位置、人の顔向き、

注視物体位置の位置関係であり，これらはロボットが視点を移動することで解決できる．そのため本研究では，この2つの場面における推定性能の改善を目指す．本研究で使用している顔向き推定は，3.4章で述べたように顔の特徴点位置と三次元モデルの位置合わせにより求めた頭部三次元姿勢に基いており，精度良く顔向きを推定するためには，ロボットが人と対峙し，人の顔向きの正面から顔の特徴点位置を取得することが望ましい．そのため，ロボットは常に人の顔向きの正面において，注視物体を推定するように視点移動する．



Fig. 4.1: Robot failed to estimate the gazed object even when the object is within the robot's view



Fig. 4.2: Robot failed to estimate the gazed object when the object when it is outside the robot's view outside the robot's view

4.2 視点移動時におけるロボットの目標位置の設定

4.2.1 注視物体推定失敗における注視物体位置の判定

まず，注視物体推定を失敗した場面において，人の顔向き方向から人が注視している物体がロボットの視野内外のどちらに存在しているかを識別する．注視物体が視野内外に存在するかを判定するために顔向きの端点座標と鼻位置座標がなす角度 θ_{fimg} と $r_{ob_{min}}$ となった物体重心座標と鼻位置座標がなす角度 θ_{ob} との差分を用いる．式 (4.2.1) に示すように $\theta_{fimg} - \theta_{ob}$ が角度の閾値 T_θ より小さい場合には，推定できなかった注視物体は，ロボットの視野内，大きい場合にはロボットの視野外に存在すると判定し，それぞれの場合においてロボットが視点移動するための目標位置を算出する．

$$\begin{cases} \theta_{fimg} - \theta_{ob} \leq T_\theta & \text{Object is in view of the robot} \\ \theta_{fimg} - \theta_{ob} > T_\theta & \text{Object is outside view of the robot} \end{cases} \quad (4.2.1)$$

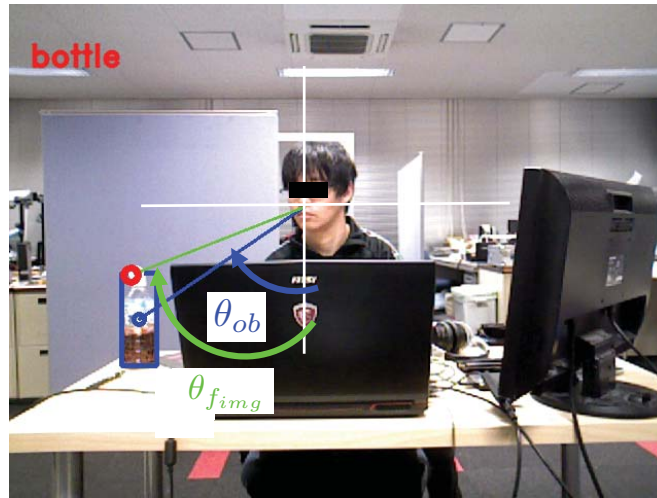


Fig. 4.3: Classification of gazed object position using face orientation

4.2.2 注視物体がロボットの視野内に存在する場合

この場面において，注視物体推定が失敗する原因は，4.1 章において述べたように物体の重心座標 (x_{c_n}, y_{c_n}) との顔向きの終点座標 $(x_{f_{min}}, y_{f_{min}})$ とのユークリッド距離 $r_{ob_{min}}$ が閾値 T_d より大きいことである．これは，画像中の顔と対象物との距離が遠くなるほど顔向きの推定精度の影響が大きくなるためである．これを解決する方法の一つとして，視点を顔向きの正面に移動し，顔を正面から観測することが考えられる．これによって，顔向き推定精度を向上すると同時に顔と物体の画像中の見た目の距離が近くなることが期待される．つまり，視点移動時のロボットの目標位置 $(x_{rT}, y_{rT}, \theta_{rT})$ は，式 (4.2.2) により人を中心とした円周上に算出される． d_p は，3.3 章において算出した人とロボットとの距離である．Fig. 4.4 に注視物体がロボットの視野内に存在する場合における目標位置の設定の一例を示す．Fig. 4.4(b) は，Fig. 4.4(a) におけるロボットの目標位置を示している．Fig. 4.4(b) 中の赤い矢印は算出されたロボットの目標値を示しており，ロボットの視野内に人の注視物体が存在する場合において，人の顔向きの正面になるようにロボットの目標値を設定できていることがわかる．

$$\begin{cases} x_{rT} = d_p \cos \theta_f + x_p \\ y_{rT} = d_p \sin \theta_f + y_p \\ \theta_{rT} = \theta_r + \pi \end{cases} \quad (4.2.2)$$



(a) Face orientation



(b) Target position of the robot

Fig. 4.4: The gazed object is in the view of robot

4.2.3 注視物体がロボットの視野外に存在する場合

ロボットが顔向き推定から得られる情報は、注視物体と顔向きとの方向関係のみであり、顔から注視物体まで正確な距離を得ることはできない。注視物体が人やロボットの近くに存在すると仮定する。注視物体がロボットの視野外に存在する場合には、ロボットは現在の位置において人の顔向きに基づき、その場で回転し、人の視線を追うことで推定できる可能性がある。人とロボットが正面に対峙する場面において、人が注視している物体位置は、次の3つが考えられる。

- (a) ロボットを注視している場合
- (b) ロボットの背後方向に存在する物体を注視している場合
- (c) ロボットの左右に存在している物体を注視している場合

したがって、Fig. 4.5 に示すように人がロボットの視野外の物体を見ている状況の場合分けを3つに場合分けする。Fig. 4.5(a) は人がロボットを見ている場合、(b) は、注視物体がロボットの背後方向に存在している場合、(c), (d) は、注視物体がロボットの左右方向に存在している場合を示している。これらは、4.2.3 式により算出される世界座標系におけるロボットの姿勢角 θ_r と人の顔向き θ_f の角度の差分 θ_{rf} に応じて場合分けしている。ここで θ_f は、ロボットの方を見たときに0となるようにしている。

$$\theta_{rf} = \theta_r - (\theta_f + \pi) \quad (4.2.3)$$

この場合分けに基づいて、式 (4.2.4) により、ロボットの目標姿勢 θ_{rT} を算出する。ここで、 θ_{th1} , θ_{th2} , θ_{th3} は、それぞれの場合分けにおける θ_{rf} の境界を示す角度の閾値である、 $0 \leq \theta_{rf} < \theta_{th1}$, $-\theta_{th1} < \theta_{rf} \leq 0$ の場合においては、すでに注視物体がわかっているため、視点移動を行わない。次に $\theta_{th1} \leq \theta_{rf} < \theta_{th2}$, $-\theta_{th2} < \theta_{rf} \leq -\theta_{th1}$ の場合においては、人とロボットとの距離が人と注視物体まえの距離と比べて無視できるほどに小さいと考え、人の顔向き方向をそのまま目標姿勢としている。最後に $\theta_{th2} \leq \theta_{rf} \leq \theta_{th3}$, $-\theta_{th3} \leq \theta_{rf} \leq -\theta_{th2}$ の場合においては、注視物体が人やロボットの近くに存在すると仮定し、視線上でロボットと距離が最も近くなる位置、つまり、人、ロボット、注視物体が直角三角形を構成して

いると仮定して目標姿勢を設定している．Fig. 4.6, Fig. 4.7, Fig. 4.8 に注視物体がロボットの視野外に存在する場合における目標位置の設定の一例を示す．Fig. 4.6(b), Fig. 4.7(b), Fig. 4.8(b) におけるロボットの目標位置を示している．それぞれの図中の赤い矢印は算出されたロボットの目標値を示しており，ロボットの視野外に人の注視物体が存在する場合において，その物体が存在する方向にロボットの目標値を設定できていることがわかる．

$$\theta_{rT} = \begin{cases} \theta_r & (0 \leq \theta_{rf} < \theta_{th1} \quad -\theta_{th1} < \theta_{rf} \leq 0) \\ \theta_f & (\theta_{th1} \leq \theta_{rf} < \theta_{th2} \quad -\theta_{th2} < \theta_{rf} \leq -\theta_{th1}) \\ \theta_{rf} + \frac{3}{4}\pi & (\theta_{th2} \leq \theta_{rf} \leq \theta_{th3}) \\ \theta_{rf} - \frac{3}{4}\pi & (-\theta_{th3} \leq \theta_{rf} \leq -\theta_{th2}) \end{cases} \quad (4.2.4)$$

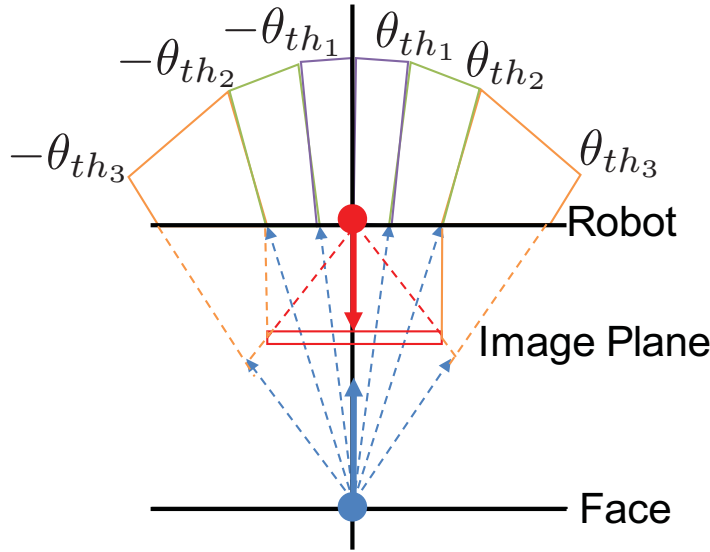
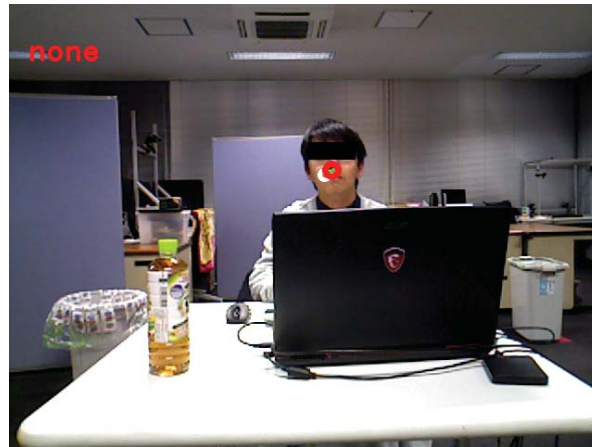


Fig. 4.5: Classification of gazed object is outside view

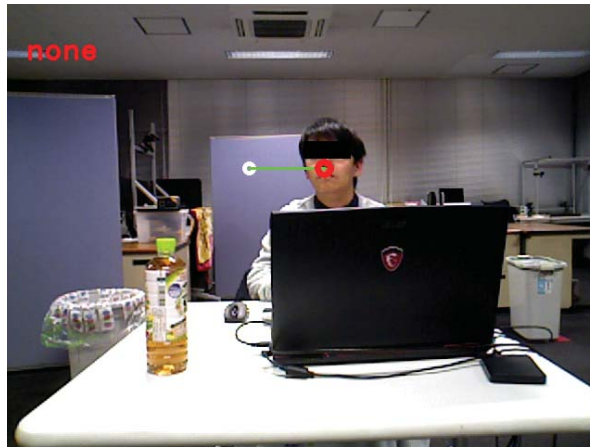


(a) Face orientation



(b) Target position of the robot

Fig. 4.6: The target position in the case where the gazed object is the robot

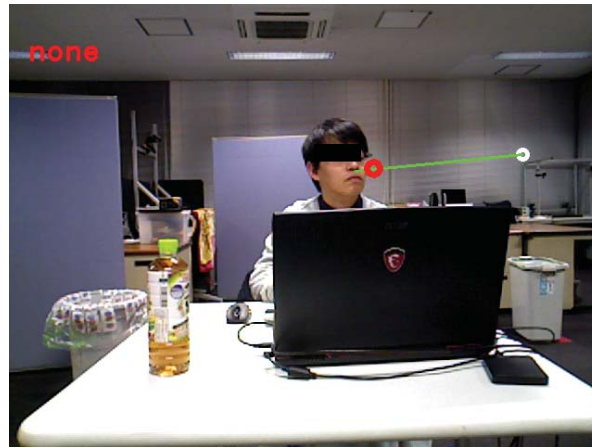


(a) Face orientaiton



(b) Target position of the robot

Fig. 4.7: The target position in the case where the gazed object is behind of the robot



(a) Face orientation



(b) Target position of the robot

Fig. 4.8: The target position in the case where the gaze object is side of the robot

第 5 章 ロボットの視点移動を用いた人の注視物体推定実験

本章では，提案した注視物体推定の精度評価実験について述べる．まず，実験に使用するロボットシステムについて述べる．そして，固定視点における注視物体の場所に応じた注視物体推定の推定実験について述べた後，固定視点において推定精度の低い場面におけるロボットの視点移動による影響の検証実験について述べる．最後にロボットの視野外に存在する注視物体推定実験について述べる．

5.1 システム構成

提案する人の注視物体推定手法では，移動ロボットが物体認識結果に基づき環境中における物体や人を検出，追跡し，顔向きを推定することで人の注視物体を推定する．したがって，構築するロボットシステムには，以下の機能が必要となる．

- 環境中に存在する物体の認識
- 人の検出，追跡
- 人の顔向き推定
- ロボットの自己位置推定

Fig. 5.1 にこれらを実現するために本研究で構築したロボットシステムを示す．移動ロボットには，ソフトバンクロボティクス社の Pepper [22] を使用している．Pepper は全方向移動機構を有しているため，環境中の任意の位置，姿勢に制御可能である．そして，ロボットの周囲の人や物体を把握するために RGB-D カメラである ASUS 社の Xtion PRO LIVE を搭載した．さらに環境内でロボットが自己位置を把握するための LRF として北陽電機の UTM-30LX を搭載している．各センサやロボットの制御は，Table. 5.1 に示す外部に接続されている PC 上で行う．さらに各センサとロボットシステムの統合には，ロボットプラットフォームの一種である Robot Operating System (ROS) [23] を用いることで，システム全体を機能ごとのノードに分割している．

Fig. 5.2 にシステムのソフトウェア構成を示す．ロボットシステムは，各セン

サからのデータを取得するセンサー部，人の注視物体推定をする注視物体推定部，ロボット自己位置取得や移動を行うロボット制御部から構成されている，まず，センサー部では，RGB-D カメラから RGB 画像，深度画像，LRF から距離情報をそれぞれ取得している．次に注視物体推定部では，物体認識ノード，顔認識ノードにおいて RGB-D カメラから得られた RGB 画像を用いて物体認識結果と顔の特徴位置を取得する．次に環境中の人物位置を推定するトラッキングノードでは，3.3 章で述べたように物体認識結果と深度情報から世界座標系での人物位置を推定している．これらに基づいて，注視物体推定ノードにおいて注視物体の推定している．最後にロボット制御部では，自己位置推定ノードにおいて事前に gmapping [24] を用いて作成した環境地図上で LRF から得られた距離データとロボットのオドメトリ情報を用いて amcl [25] を用いてロボットの自己位置を推定している．そして，ロボットの移動が必要な場合には，ロボットに速度指令値を入力する．

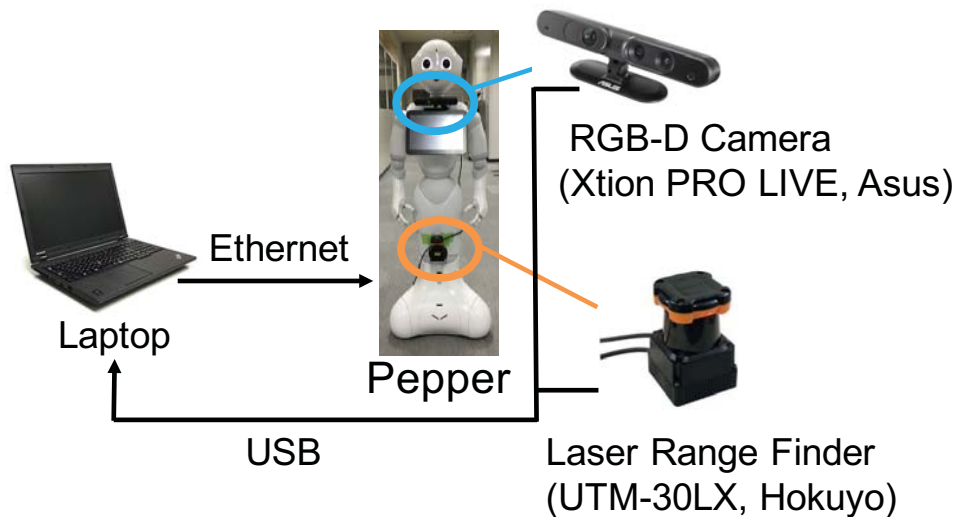


Fig. 5.1: Hardware configuration

Table 5.1: Specifications of laptop

OS	Ubuntu 16.04 LTS (64bit)
CPU	Intel Core i7-770HQ (2.80GHz)
Memory	16GB
GPU	NVIDIA GeForce GTX 1060/PCIe/SSE2

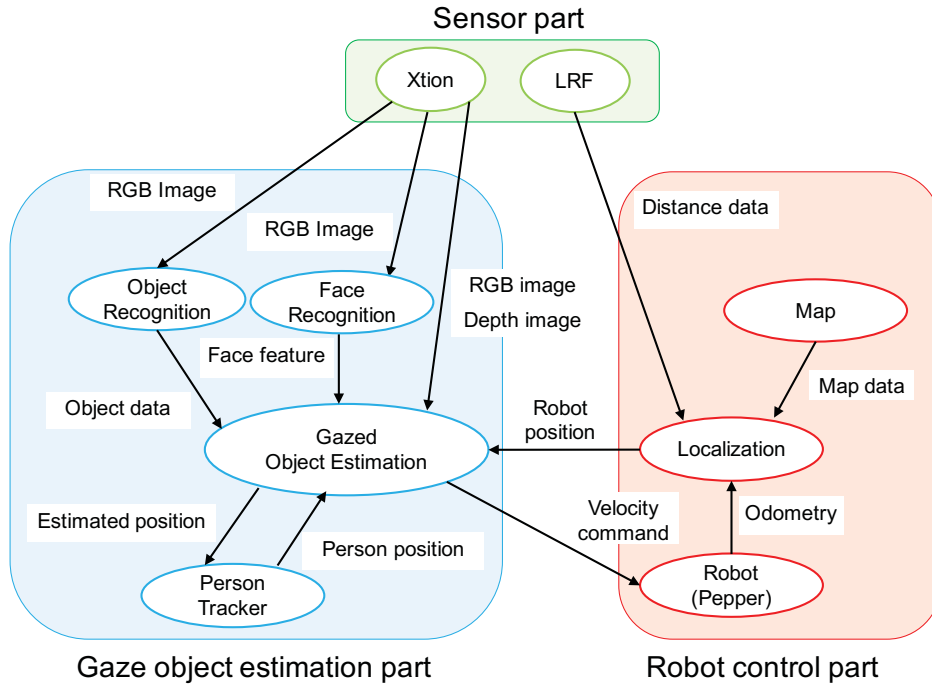


Fig. 5.2: Software configuration

5.2 顔向きに基づいた注視物体推定実験

5.2.1 実験の流れ

本実験では，固定視点における提案する注視物体推定の性能を評価する．実験の様子を Fig. 5.3 に示す．実験は，健康な成人男性 3 人を対象に実施し，被験者は，ロボットと正面に対峙した状態で着席する．被験者は，Fig. 5.3 に示すように机の上に置いたペットボトルを注視し，ロボットは，被験者の顔向きに基いて，被験者の注視物体を 30 秒間にわたり推定する．そして，Fig. 5.3 に示すようにペットボトルの位置 P_b を $-400 \sim 400$ [mm] までペットボトルを 200 [mm] ずつ移動させて実験を繰り返す．

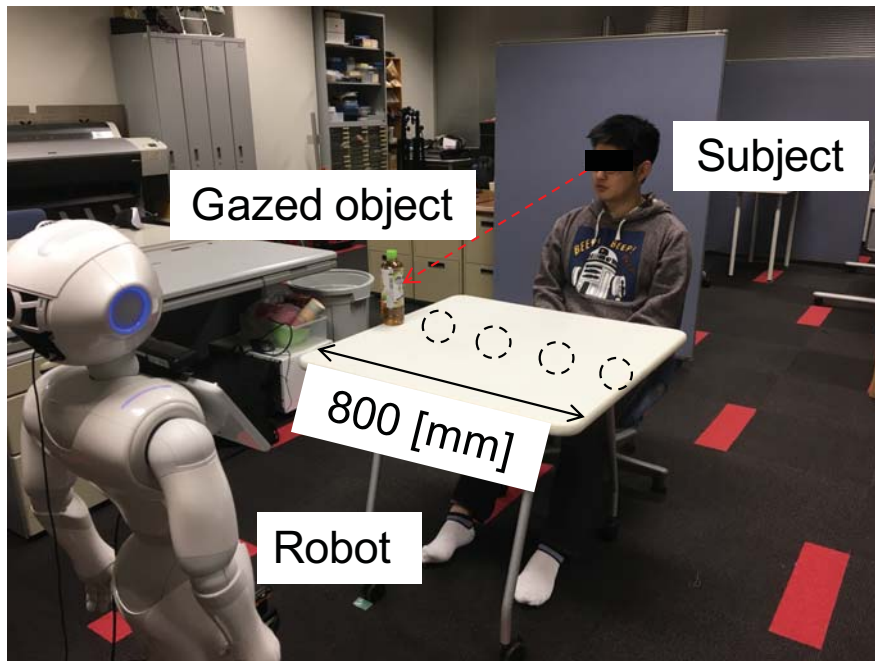


Fig. 5.3: Experimental setup

5.2.2 実験結果

Table. 5.2 に実験結果を示す．推定精度は，処理された全フレーム数に対して，ペットボトルを注視していると推定されたフレーム数の割合を示している．Table. 5.2 より，被験者の正面付近 ($P_b=-200, 0, 200$ [mm]) における注視物体精度は，被験者から見て斜め方向 ($P_b=-400, 400$ [mm]) における注視物体の推定精度より高いことから，提案する注視物体推定手法では，推定精度を改善するために顔向きの正面付近から注視物体を推定することが有効であることがわかる．

Table 5.2: Result of the gazed object estimation

	Estimation rate [%]				
$P_b[mm]$	-400	-200	0	200	400
Subject1	17.4	99.1	96.7	98.6	55.9
Subject2	36.1	99.5	97.7	99.5	57.6
Subject3	68.7	99.1	97.0	99.1	54.6
Average	40.5	99.2	97.1	99.1	54.6

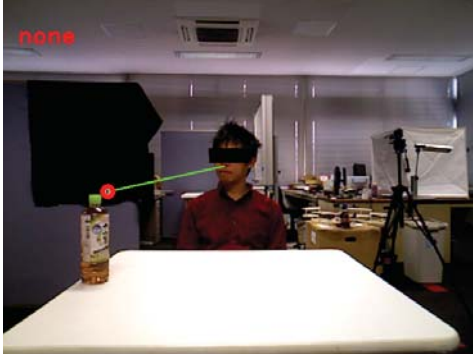
5.3 ロボットの視野内に存在する人の注視物体推定実験

5.3.1 実験の流れ

本実験では，人の注視物体がロボットの視野内に存在する場合におけるロボットの視点移動による注視物体推定精度への影響を評価する．5.2 章において推定精度が低いペットボトルの位置 $P_b = -400, 400$ [mm] の場合において，ロボットが人の顔向き正面に移動し，移動前後の注視物体精度を比較することで，ロボットの視点移動によって注視物体推定精度が改善されることを確認する．

5.3.2 実験結果

Fig. 5.4 にロボットが視点移動中に取得した画像を時系列で示す．Fig. 5.4 からロボットは， $t = 0$ [s] では，人の顔向きと注視物体との角度が大きく，注視物体を推定できていないが，ロボットが人の顔向きの正面に移動した $t = 15$ [s] においては，注視物体を推定できていることがわかる．Table. 5.3, Table. 5.4 に ($P_b = -400, 400$ [mm]) におけるロボットに移動前と移動後の推定精度の比較を示す．Table. 5.3, Table. 5.4 から移動前における注視物体の推定精度は，50 [%] 前後であり，移動後の推定精度は，85 [%] 前後と改善されていることからロボットの移動によって注視物体の推定精度の改善が可能である．



(a) $t=0$ [s]



(b) $t=3$ [s]



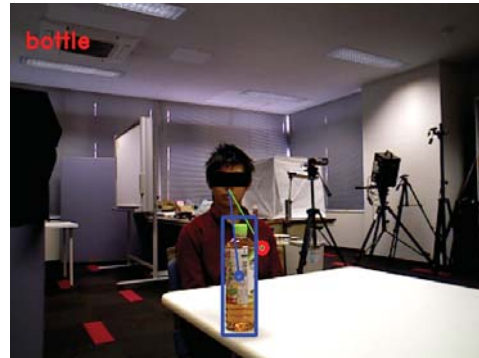
(c) $t=6$ [s]



(d) $t=9$ [s]



(e) $t=12$ [s]



(f) $t=15$ [s]

Fig. 5.4: Screenshots of robot moving when the gazed object is in the view of the robot

Table 5.3: Comparison of fixed and changing viewpoint of the robot ($P_d = -400[\text{mm}]$)

	Estimation rate [%]	
	before	after
Subject1	17.4	89.1
Subject2	53.1	87.4
Subject3	68.7	80.6
Average	40.5	85.9

Table 5.4: Comparison of fixed and changing viewpoint of the robot ($P_d = 400[\text{mm}]$)

	Estimation rate [%]	
	before	after
Subject1	61.9	84.3
Subject2	49.1	91.2
Subject3	50.4	92.5
Average	54.6	89.2

5.4 ロボットの視野外に存在する人の注視物体推定実験

5.4.1 実験の流れ

本実験では、4.2.3 章において述べたロボットの視野外に存在する人の注視物体推定精度を検証する．実験の様子を Fig. 5.5 に示す．5.2 章，5.3 章と同様に被験者は，ロボットと向かいあった状態で着席する．実験開始時に被験者は，机の上にあるノートパソコンを注視する．そして，指示されたタイミングで Fig. 5.6 中の (a)～(d) の位置に配置したペットボトルを注視する．そして，ロボットは顔向きに応じて，目標位置を設定し，姿勢角を変更することで注視物体を推定する．また，4.2.3 章において述べた人がロボットの視野外の物体を見ている状況の分類に用いる式 (4.2.4) 中の θ_{th_1} ， θ_{th_2} ， θ_{th_3} は，式 (5.4.1) に示すように設定している．

$$\begin{cases} \theta_{th_1} = \frac{\pi}{18} \\ \theta_{th_2} = \frac{\pi}{6} \\ \theta_{th_3} = \frac{\pi}{2} \end{cases} \quad (5.4.1)$$

なお，ロボットが目標位置に到達後は，人が視野内に存在しなくなるため人の顔向きは，推定不能となる．そこで，注視物体推定の精度評価には，Fig. 5.7 に示すように移動後に検出されたペットボトルの重心座標 (x_c, y_c) と画像の中心座標 $(x_{c_{img}}, y_{c_{img}})$ との差分 e_x を用いる．この e_x が式 (5.4.2) により求められ，小さいほど，ロボットの視野の中心で物体に注視物体を捉えているため注視物体推定の精度が高い．実験で取得する画像のサイズは， 480×640 である．実験は，Fig. 5.6 中の (a)～(d) のペットボトルに対して 5 回ずつ繰り返す．また，実験中は，Fig. 5.8 に示すヘルメットにアクションカメラ (GoPro HERO3) 装着したデバイスを用いて，人の注視対象を確認する．

$$e_x = x_c - x_{c_{img}} \quad (5.4.2)$$

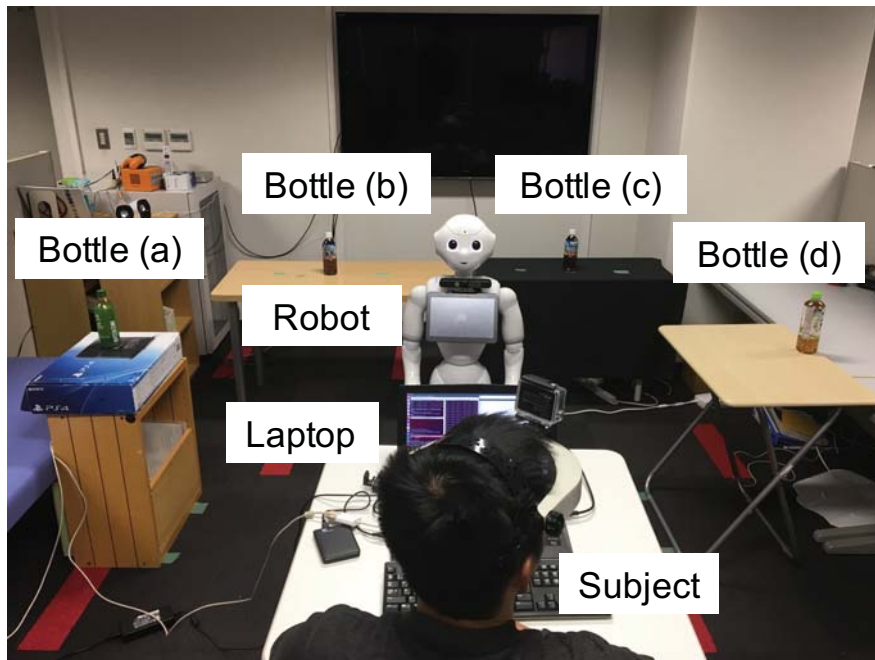


Fig. 5.5: Experimental setup

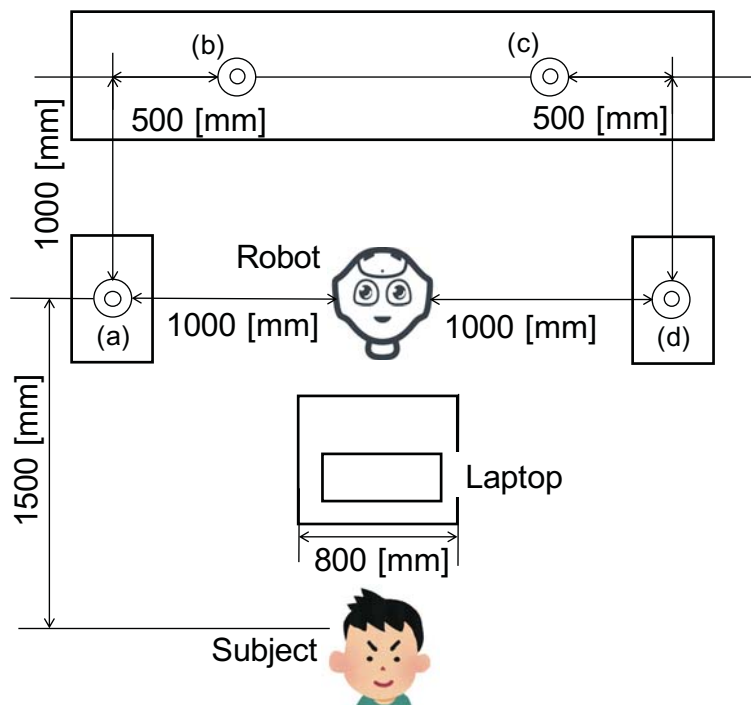


Fig. 5.6: Experimental layout

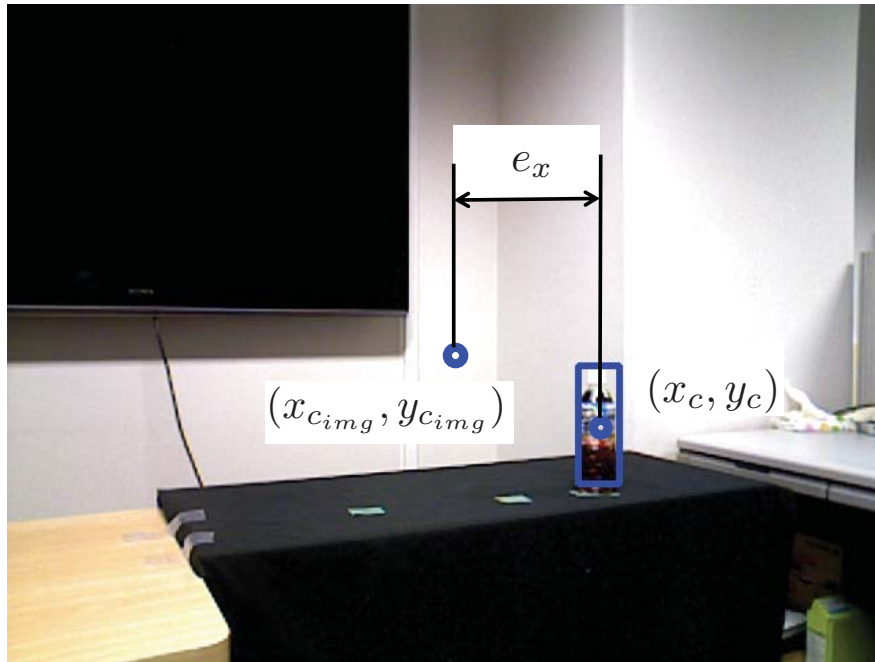


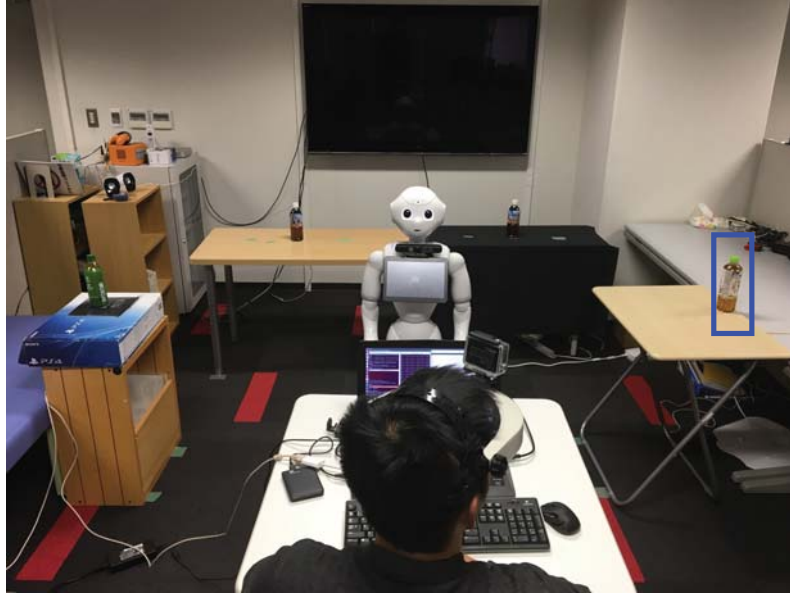
Fig. 5.7: Error of object center from the image center



Fig. 5.8: First-Person-View (FPV) device

5.4.2 実験結果

Fig. 5.9(b) に示す FPV デバイスから取得された画像の中央にペットボトルを捉えていることから実際に人がペットボトルを注視していることがわかる．そして，Fig. 5.10 に示した視点移動中にロボットが取得した画像からロボットは，その物体がロボットの視野外に存在する場合において，顔向きに基づいて視点を注視物体が存在する物体の方向に移動していることがわかる．Table. 5.5 に視点移動後のロボットの視野内の注視物体位置と視野の中央との誤差を示す．Table. 5.5 は，各方向のペットボトルにおいて 5 回の試行で得られた誤差 e_x の絶対値の平均値である．Table. 5.5(a), (d) から，ロボットの左右方向に存在する物体に関しては，視点移動後にロボット視野の中央付近に人の注視物体を捉えることができていることから，人の注視物体を正しく推定できていることがわかる．一方で，Table. 5.5(b), (c) からロボットの奥側に存在する注視物体に関しては，視点移動後に，ロボットの移動後に視野の中央付近に人の注視物体視野を捉えることができていないことがわかる．しかしながら，注視物体が存在する方向にロボットが視点移動を行っていることから Fig. 4.5 中における人の注視物体がロボットの視野外に存在する場合の場合分けができていることがわかる．



(a) Face orientation



(b) Captured image using FPV device

Fig. 5.9: Capture of gazed object using by FPV device



(a) $t=0$ [s]



(b) $t=3$ [s]



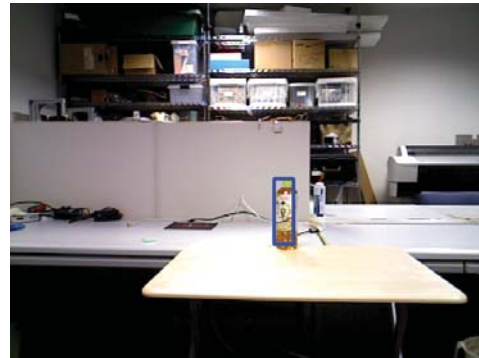
(c) $t=6$ [s]



(d) $t=9$ [s]



(e) $t=12$ [s]



(f) $t=15$ [s]

Fig. 5.10: Screenshots of robot moving when the gazed object is outside the robot's view

Table 5.5: Result of the gazed object estimation for outside the robot view
(a),(b),(c),(d) are in Fig. 5.5

	e_x [pixel]			
Gazed object	(a)	(b)	(c)	(d)
Subject1	99	168	212	57
Subject2	49	144	227	49
Subject3	36	87	220	50
Average	62	133	222	51

5.5 考察

5.2 章の顔向きに基づいた注視物体推定実験から、画像中の顔向きと物体位置によって推定精度が異なることが明らかになった。提案する注視物体推定では、頭部姿勢をカメラ画像中に投影した顔向きの終点座標と物体認識によって得られた物体重心座標のユークリッド距離に基づいて注視物体を判定しているため、 $P_b = -200, 0, 200$ [mm] の場合のように顔の正面付近に存在する物体と比較して $P_b = -400, 400$ [mm] の場合のように顔と物体との角度が大きい物体の推定精度は低くなると考えられる。これは、顔向きが正面から離れるにつれて、顔の特徴量の取得が不安定になるため、顔向きの推定精度が低くなる傾向があるためであると考えられる。

5.3 章のロボットの視野内に存在する人の注視物体推定実験の結果は、注視物体推定の精度が低く、注視物体がロボットの視野内に存在している場合において、ロボットが顔向きの正面に移動することによって、その推定精度が改善されることを示している。これは、ロボットの移動によって視点が変更され、 $P_b = -400, 400$ [mm] それぞれの場合において、顔と物体との角度が小さくなったためであると考えられる。また、視点の移動後は、顔を正面から捉えることで、顔向き推定に使用している顔の特徴量をすべて安定して取得することが可能となり、顔向き推定の精度が向上することも推定精度改善の一因であると考えられる。しかしながら、視点移動後の推定精度は、Table. 5.2 に示した推定精度と比較して低くなっている。これは、顔向きの推定誤差や、ロボットの移動後の目標位置に対する誤差から誤差が原因で人の顔向きの正面に移動することができていないためである。これらの結果は、注視物体推定の精度の一つの改善策としてロボットの移動による視点変更が有用であることを示唆している。

5.4 章のロボットの視野外に存在する人の注視物体推定実験結果から注視物体がロボットの左右方向に存在する場合において、移動後のロボットの画像の中心座標と注視対象物体の重心座標の誤差が小さく、注視物体を顔向き情報のみで推定可能であることが確かめられた。しかしながら、ロボットの奥側に存在する物体の推定結果においては、移動後のロボットの画像の中心座標と注視対象物体の重心座標に大きな誤差が確認された。これは、現状の手法では、顔向きに基づいて注視物体が存在する方向にロボットが姿勢を変化させており、その方向の誤差は、ロボットと物体の距離に比例しているため、物体とロボットとの距離が離れるほど画像中での

誤差が大きくなるためである。したがって、注視物体とロボットの距離が大きく顔向き情報のみでは、視点移動によってロボットの後ろ側に存在する人の注視物体を一意に推定することは困難であり、ロボットの視野外に存在する注視物体推定においては、注視物体は、人やロボットの近くに存在することを前提にする必要があると考えられる。また、Table. 5.5 のすべての結果から、ロボットは、移動後の視野内に人の注視物体を捉えているため、注視物体の候補となりうる物体の推定は可能であると考えられる。そのため、環境中の物体位置とロボットが持つマップとの関連付けや会話などの人とのインタラクションによって顔向き以外の情報を考慮して視点移動することで推定精度は改善すると考えられる。

第 6 章 おわりに

6.1 本論文のまとめ

本論文では，日常生活における生活の質を評価する上で重要な指標である生活時間収集を支援することを目的として，カメラを搭載した移動ロボットを用いた注視物体推定を行った．人の注視物体を推定する手法として，本論文では，カメラ画像を用いた環境中の物体や人の顔向き推定結果に基いた人の注視物体を推定手法を提案した．人の頭部姿勢推定によって得られた人の顔向きを画像中に投影することで画像中における物体認識結果と人の顔向きとの対応関係を求め，人の注視物体推定を実現した．また，注視物体推定の失敗時における場合分けを行うことで人の顔向き方向から人が注視している物体がカメラの視野内外のどちらに存在しているかを識別し，それぞれに応じてロボットの移動によって視点を変更することで推定精度の向上を図った．

提案する注視物体推定の有用性を確かめるために人とロボットが対峙した場面において顔向きに基づいた注視物体推定の実験を行った．その実験結果から，提案する注視物体推定では，顔向きの正面付近の推定精度が高く，顔と物体の三次元距離が遠くなり，顔向きと物体との角度が大きくなることで推定精度が低下する傾向があることが確かめられた．次にロボットの移動による注視物体推定の精度向上の検証実験を行い，人の注視物体がロボットの視野内に存在し，推定精度が低い場面において顔向きの正面にロボットが移動することで，推定精度が改善することが示された．また，精度は十分とは言えないものの人がロボットの視野外の物体を注視した場合においてもロボットの視点移動によってその注視物体やその候補となる物体を推定できることが確かめられた．このことから提案する人の注視物体推定手法によって生活時間測定を支援できる可能性が示唆された．

6.2 今後の課題と展望

課題として、まず、注視物体推定の精度向上が挙げられる。現状の注視物体推定手法は、5.2 章における実験で明らかになったように、注視対象と人間の顔との距離が離れている場合における推定精度に課題がある。ロボットの視点変更により、推定精度は向上することは確認できたが、実際の生活時間測定においては、人に与える違和感を考慮してロボットの移動頻度は少ないことが望ましい。具体的な推定精度の改善方法として考えられるのが、時系列データを用いた注視物体推定である。本研究では、フレーム毎に取得される物体認識と顔向き推定結果に基づいて注視物体推定を実現したが、過去の推定結果を考慮した推定を行うことによって注視物体推定の頑健性を高めることができると考えられる。

次に人がロボットの視野外に存在する物体を注視した場合における推定精度の向上が挙げられる。現在の手法では、人の顔向き情報に基づいて注視対象の判定を行っている。そのため、ロボットが視点移動によって人の顔向きを推定困難な位置に移動した場合において、移動後のロボットの視野内に物体が複数個存在する際に注視物体を一意に推定することは困難である。さらに注視物体の推定精度は、顔向き推定、ロボットの自己位置推定、ナビゲーションなど様々な要素技術の精度に依存している。そこで、ロボットならではの改善方法として、注視物体の推定が困難な場面においては、人に対して質問し、回答を得るなどインタラクションによる推定手法の導入が考えられる。この手法は、その他の注視物体推定の失敗事例についても有効な手法であると考えられる。また、あらかじめロボットの持つ地図上に物体の配置情報を追加することで、顔向き方向と注視物体の関連付けがより容易になると考えられる。

また、本研究では、生活時間を測定するために人の行動目的と関連深い注視物体に着目したが、実際の生活時間を取得するためには、食事の箸などの人が操作している物体も同時に推定することが必要となる。そのために物体認識によって得られる物体位置のフレーム間の移動量を算出することによって操作物体を推定し、注視物体推定結果と統合することでより詳細な生活時間記録に繋げることが可能になると考えられる。

謝辞

本論文は、奈良先端科学技術大学院大学情報科学研究科情報科学専攻ロボティクス研究室で博士前期課程在籍中の研究成果をまとめたものである。

同専攻領域の小笠原司教授には本研究を遂行するにあたり様々な御指導，御助言を賜りました。ここに深く御礼申し上げます。

同専攻領域の萩田紀博教授には論文執筆にあたり丁寧な御検討，御教示を賜りました，ここに深く御礼申し上げます。

同専攻領域の高松淳准教授には本研究を遂行するにあたり，アイデアや論文執筆など多くの点について，様々な御指導，御助言を賜りました。ここに深く御礼申し上げます。

同専攻領域の丁明助教には本研究についての様々な議論や学生生活に関して多くの点において御指導，御助言を賜りました。ここに深く御礼申し上げます。

本研究室博士研究員の GARCIA RICARDEZ Gustavo Alfonso 氏には本研究についての様々な議論，英文添削など多くの点において御指導，御助言を賜りました。ここに深く御礼申し上げます。

本研究の遂行にあたり，様々な議論やアドバイスを頂いたサービスロボットグループの築地原里樹氏，袴田有哉氏，湯口彰重氏に感謝致します。

本研究室秘書の大脇美千代氏，池田美輝氏には大学生活や事務手続きなど様々なサポートをして頂きました，ここに深く御礼申し上げます。

本研究の遂行にあたり，様々な意見や御助言，頂いた先輩，同輩，後輩の皆様に深く感謝致します。最後に，長年の学生生活を支えてくださった両親，友人に心から感謝致します。

参考文献

- [1] 相澤清晴: “ライフログ“, 映像情報メディア学会誌, vol. 63, no. 4, pp. 445–448, 2009.
- [2] 厚生労働省: “総論参考資料“, http://www1.mhlw.go.jp/topics/kenko21_11/s1.html, (2018 年 1 月 30 日 参照)
- [3] 厚生労働省: “「国際生活機能分類－国際障害分類改訂版－」(日本語版)“, <http://www.mhlw.go.jp/houdou/2002/08/h0805-1.html>, (2018 年 1 月 30 日 参照)
- [4] 筑紫恒男: 家族生活の支援 理論と実践, pp. 136-140, 建帛社, 2014.
- [5] 総務省統計局: “統計局ホームページ/平成 23 年社会生活基本調査“, <http://www.stat.go.jp/data/shakai/2011/index.html>, (2018 年 1 月 30 日 参照)
- [6] Ryota Matura, Satoki Tsuichihara, Garcia Ricardez Gustavo Alfonso, Ming Ding, Jun Takamatsu, Tsukasa Ogasawara: “Reviewing Eating Habits Using Video Summarization of Eating Scenes“, Proc. of 19th ARAHE Biennial International Congress, MPF01, 2017.
- [7] 佐藤琢磨, 前川卓也: “天井照明に設置したセンサデバイスを用いた食事画像自動記録システムの提案“, 情報処理学会研究報告, Vol.2014-UBI-42, no. 11, pp. 1–8, 2014.
- [8] Min-Yu Wu, Tzu-Yang Chen, Kuan-Yu Chen and Li-Chen Fu: “Daily Activity Recognition Using the Informative Features from Skeletal and Depth Data,” Proc. of the 2016 IEEE Int. Conf. on Robotics and Automation (ICRA 2016), pp. 1628-1633, 2016.
- [9] Ninghang Hu, Gwenn Englebienne, Zhongyu Lou and Ben Kröse: “Latent Hierarchical Model for Activity Recognition,” IEEE Trans. on Robotics, vol. 31, no. 6, pp.1471-1472, 2015.
- [10] Hema Swetha Koppula, Rudhir Gupta and Ashutosh Saxena: “Learning human activities and object affordances from RGB-D videos,” The International Journal of Robotics Research, vol. 32, no. 8, pp.952-970, 2013.

- [11] Yoonseok Pyo, Kouhei Nakashima, Shunya Kuwahata, Ryo Kurazume, Tokuo Tsuji, Ken'ichi Morooka and Tsutomu Hasegawa: "Service robot system with an informationally structured environment," *The International Journal of Robotics and Autonomous Systems*, vol. 74, Part A, pp.148-165, 2015.
- [12] Hamed Pirsiavash and Deva Ramanan: "Detecting activities of daily living in first-person camera views," *Proc. of the 2012 IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR 2012)*, pp. 2847-2854, 2012.
- [13] Kazuto Nakashima, Yumi Iwashita, Pyo Yoonseok, Asamichi Takamine and Ryo Kurazume : "Fourth-person sensing for a service robot," *Proc. of the 2015 IEEE Int. Conf. on Sensors*, pp. 1110-1113, 2015.
- [14] 関根智江, 渡辺洋子, 林田将来: "日本人の生活時間・2015 ～睡眠の減少が止まり, 必需時間が増加～", vol. 66, no. 5, *放送研究と調査*, pp. 2-27, 2016.
- [15] Thomas F  ulhammer, Rareş Ambruş, Chris Burbridge, Michael Zillich, J Folkesson, Nick Hawes, Patric Jensfelt and Markus Vincze: "Autonomous Learning of Object Models on a Mobile Robot," *IEEE Trans. on Robotics and Automation Letters*, vol. 2, no. 1, pp.26-33, 2017.
- [16] Joseph Redmon, Santosh Divvala, Ross Girshick and Ali Farhadi: "You Only Look Once: Unified, Real-Time Object Detection," *Proc. of the 2016 IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR 2016)*, pp. 779-788, 2016.
- [17] Microsoft: "Microsoft COCO" ,<http://cocodataset.org>, (2017 年 1 月 30 日 参照)
- [18] Woojin Chung, Hoyeon Kim, Yoonkyu. Yoo, Chang-Bae Moon, and Jooyoung Park: "The Detection and Following of Human Legs Through Inductive Approaches for a Mobile Robot With a Single Laser Range Finder," *IEEE Trans. on Industrial Electronics*, vol. 59, no. 8, pp.3156-3166, 2012.
- [19] Meenakshi Gupta, Laxmidhar Behera, Venkatesh K. Subramanian and Mo M. Jamshidi: "A Robust Visual Human Detection Approach With

- UKF-Based Motion Tracking for a Mobile Robot,” IEEE Systems Journal, vol. 9, no. 4, pp.1363-1375, 2015.
- [20] Tadas Baltrusaitis, Peter Robinson and Louis-Philippe Morency: “Open-Face: An open source facial behavior analysis toolkit,” Proc. of the 2016 IEEE Winter Conf. on Applications of Computer Vision (WACV 2016), pp. 1-10, 2016.
- [21] 出口光一郎: “射影幾何学による pnp カメラ補正問題の統一的解法 “ , コンピュータビジョン,’90 シンポジウム, pp. 41-50, 1990.
- [22] Pepper, <https://www.softbank.jp/robot/consumer/products/>, (2018 年 1 月 30 日 参照)
- [23] Morgan Quigley, Ken Conley, Brian P. Gerkey, Josh Faust, Tully Foote, Jeremy Leibs , Rob Wheeler, and Andrew Y. Ng, “Ros: an open-source robot operating system,” Proc. of the IEEE Int. Conf. on Robotics and Automation Workshop on Open Source Software, 2009.
- [24] gmapping, <http://wiki.ros.org/gmapping>, (2018 年 1 月 30 日 参照)
- [25] amcl, <http://wiki.ros.org/amcl>, (2018 年 1 月 30 日 参照)