

修 士 論 文

視線を用いた人間の意図推定に基づく  
人間とロボットの柔軟な協調作業

Flexible Cooperation of Human  
and Robot by interpreting Human  
Intention from Gaze Information

指導教官 池内 克史 教授



東京大学大学院情報理工学系研究科  
電子情報学専攻

氏 名 26413 崎田 健二

提 出 日 平成 16 年 1 月 30 日

# 概要

近年、人間と関わりながら活動するロボットに関する研究が盛んに行われており、近い将来、人間と自然なインタラクションを可能とする介護ロボットや家事手伝いロボットなどが実現されると期待できる。

本研究は、ロボットへの教示情報として人間の視線情報を利用し、ロボットが人間の意図や状態を考慮した協調の実現を目的としている。視線情報は人間の意図・興味・関心をダイレクトに表現する生理的情報である上に、その動きから人間の情動を読み取ることも可能であることが知られているため、人間の状態をロボットに把握させるための情報に利用することは非常に有効であると考えた。

従来の人間とロボット間の協調作業に関する研究では、ロボットは人間の手形状の認識、作業物体の認識・トラッキング、音声情報などの情報を基に人間と協調作業を行っている。ロボットはこれらの情報により作業の本質を理解することができ、作業の流れを考慮した適切な協調・補助を行うことはできているが、その行動が人間にとって快適な協調・補助とは言えない。

生理的な情報である視線の動きからは、迷い・探索といった人間の状態および意図に関する情報を汲み取ることが可能である。本研究ではLEGOの組み立て作業を題材とし、直近の視線の履歴から人間の意図や状態を推定する手法と、これに基づくロボットの補助動作決定法を提案する。さらに、実時間3次元視線追跡システムを構築し、人間型ロボットとの協調組立て作業を実現した。

本研究は次の4つから構成される。

1. 視線の解析による人間の意図・状態の抽出
2. 人間の意図・状態に基づくロボットの補助動作の提案
3. 実験プラットフォームである3次元視線追跡システムの構築
4. 提案したロボットの補助動作の実装

本研究では、視線情報をロボットへの教示情報と用いることの有効性を示すこと

ができ、また、ロボットが作業を行っている人間の意図・状態を視線情報から推測し、適切な協調・補助を行うことができた。

# 目次

第1章	序論	2
第2章	関連研究	5
2.1	人間型ロボットとのインタラクションに関する研究	5
2.2	視線情報を利用した研究	6
第3章	人間の視線の解析	9
3.1	視線とは何か	9
3.2	LEGOの組み立て作業における視線情報の取得	11
3.2.1	実験内容	11
3.2.2	実験環境	11
3.2.3	被験者	13
3.2.4	実験手順	17
3.3	視線の解析	18
3.3.1	EMRデータのセグメンテーション	18
3.3.2	注視時間の特徴	22
3.3.3	視線の履歴情報による人間の“迷い”状態の予測	24
3.3.4	組み付け状態における視線の特徴	25
3.4	考察	27
第4章	視線を利用した協調作業の提案	29
4.1	タスクモデルの生成と補助動作の計画	29
4.1.1	タスクモデルの枠組み	29
4.1.2	タスクモデルの自動生成	30
4.1.3	補助動作の計画	30
4.2	視線を利用した補助動作の提案	31
4.2.1	従来の補助動作の問題点	31
4.2.2	探索状態における補助動作の提案	31



4.2.3	組み付け状態における補助動作の提案	34
4.2.4	まとめ	36
<b>第5章</b>	<b>3次元視線追跡システムの構築</b>	<b>38</b>
5.1	ヒューマノイドロボットシステム	38
5.1.1	ビジョン	39
5.1.2	アームとハンド	40
5.1.3	上半身	41
5.2	注視点計測システムの開発	41
5.2.1	EMR-8の概要	41
5.2.2	世界座標系における注視位置の取得手順	43
5.2.3	マーカー座標 $\Rightarrow$ 世界座標	45
5.2.4	アイマークカメラ座標 $\Rightarrow$ マーカー座標	47
5.2.5	注視位置 $\Rightarrow$ アイマークカメラ座標	49
5.2.6	世界座標系における注視位置の取得	54
5.3	3次元視線追跡システム	54
5.3.1	環境物体の認識	55
5.3.2	注視点の可視化	56
5.4	システムの評価	57
<b>第6章</b>	<b>実装および実験結果</b>	<b>61</b>
6.1	協調作業の流れ	61
6.2	実装I	64
6.3	実装II	65
6.4	実装III	66
<b>第7章</b>	<b>結論</b>	<b>69</b>
7.1	まとめ	69
7.2	今後の課題	70
	<b>謝辞</b>	<b>72</b>
	<b>参考文献</b>	<b>73</b>

# 目次

3.1	眼の構造	10
3.2	実験の様子	12
3.3	EMR-8	13
3.4	アイマークカメラ画像	13
3.5	LEGO Parts I	14
3.6	LEGO Parts II	15
3.7	組み付けの種類	16
3.8	ドライバを必要とする組み付け	16
3.9	LEGO の完成図	17
3.10	視線の停留時間分布	20
3.11	視線の停留時間と把持の分布	23
3.12	注視から把持へ行動移行に生じる時間差	24
3.13	組み付け可能なパーツリストの例	26
3.14	組み付けを行う前の注視点の様子	27
4.1	組み立て途中の物体と候補パーツ	34
4.2	機能部に着目した投票式による解析結果	35
4.3	同じパーツによる異なる機能部を使用した組み付け例	35
5.1	実験プラットフォーム	38
5.2	EMR-8	41
5.3	EMR-8 キャリブレーション	43
5.4	EMR-8 の原理	44
5.5	必要な座標変換	45
5.6	9眼ステレオビジョンの画像	46
5.7	画像処理の結果	46
5.8	キャリブレーションの様子	47
5.9	EMR-8 カメラと9眼ステレオビジョンの画像	48

5.10	カメラ座標系における注視点 . . . . .	50
5.11	カメラ座標系における両眼の位置の推定 . . . . .	52
5.12	注視点の獲得過程 . . . . .	55
5.13	3DTM による物体の認識 . . . . .	56
5.14	仮想空間における注視点の可視化 . . . . .	57
6.1	選択状態の補助動作のフローチャート . . . . .	62
6.2	作業状態の補助動作のフローチャート . . . . .	63
6.3	注視物体の把持の実験結果 . . . . .	64
6.4	注視物体の把持の実験結果 . . . . .	65
6.5	Shovel と BlockB の組み付けパターン . . . . .	66
6.6	注視機能部に基づく補助の実験結果 . . . . .	67

# 表 目 次

3.1	セグメンテーションの結果 . . . . .	19
3.2	停留時間の統計結果 . . . . .	21
3.3	停留時間を考慮したセグメンテーションの結果 . . . . .	21
5.1	統計結果 . . . . .	58

# 第1章

## 序論

---

# 第1章 序論

「ロボット」という言葉を耳にした時、主に製造ラインなどの産業用途で使用されているロボットを想像する人はどれくらいいるだろうか。ほとんどの人は、人間と同じような姿をした人間型ロボットを想像するのではないだろうか。このように、人間型ロボットが身近な存在となっている理由には、近年、人間型ロボットが相次いで研究・開発されてきている背景があり、特に、2足歩行の実現に関する研究は精力的に行われ一定の成果を収めてきた。

また一方で、ロボットによる作業の代行や協調作業によるより知的な人間支援の実現の研究が行われている。ここで重要な点は、ロボットの行動をいかに生成するかということである。これまで様々な手法が提案されてきた。これらは大まかには、“人の手によるプログラミング”、“コントロール装置を用いた人による操作”、そして“自動プログラミング”の3つに分類することができる。このうち前の2つの手法は、ロボット行動全般に渡って人間の介入を必要とする。それに引き換え自動プログラミングは、人間の補助を低減し全体の行動を自動的に獲得することを目標としている。この方法は、“Learning from Observation”と呼ばれ、必要な初期知識を最初にロボットに与えることで、ロボットはお手本動作の観察・シュミレーション・学習などによって、視覚から得た膨大な情報を抽象的な情報に置き換え、自身の行動を自動的に獲得していく方法である。

Learning from Observationの研究については2.1で述べるが、従来の研究では視覚、人間の手形状、作業物体の位置・姿勢、音声などを教示情報としてロボットに与えていた。これらの情報では、作業の本質は記述できるが、作業を行っている人間の状況・立場・意図などは記述できない。つまり、人間がロボットと協調作業を行う際には、ロボットは、人間が教示時と全く同じ行動をとることを前提に、設計された静的なモデルに従って作業を進行していく。これは、人間同士の作業のように相手のことを考慮した協調作業には程遠く、人間側にとって自然な協調作業とは言難い。人間とロボットの協調作業において、より自然で柔軟なシステムを構築するためには、作業の情報だけでなく、人間の注視対象・意図・判断等を正しく認識するための情報もロボットに与える必要がある。

我々は、人間の注意や興味をダイレクトに表現する生理的情報である視線情報が教示情報に有効であると考えた。視線は、知的に活動する人間の思考や行動を反映し、これを示唆する有益な情報を含んでいるため、心理学研究の現場においては多く利用され、また工学分野でも研究対象になっている [1]。人間の視線を分析することで、操作対象だけでなく、興味、関心、意図などの情報を抽出し、これをロボットの行動計画に反映させることができれば、人間同士の作業に勝るとも劣らないほど効率的で柔軟な人間とロボットの協調作業を実現できる可能性がある。また、視線を動かすことは無意識の動作なので、人間に余計な動作を強いる必要なしに生成される生理的情報であるという点も望ましい。この点が、音声などと比較して大きな違いである。「物体 A を把持する」という意図的な音声情報は、視線情報においては停留時間や視線の移動速度といった無意識的に含まれる情報に相当し、その情報を解析することで、人間が自分の考えを口に出さなくても、ロボットは人間の考えを把握することが可能になる。このように、視線情報はロボットへの教示情報として非常に有効であるにもかかわらず、なぜこの情報を利用した研究は少ないのであろうか。その理由の 1 つにハードウェアの問題があげられる。頭部装着型の視線追跡装置はユーザの負担が大きいという問題がある。しかし、非装着型はこれまでに存在しなかったが、最近ではカメラとコンピュータビジョンを利用した研究が行われている。もう 1 つはノイズの問題が考えられる。視線を検出するために使用する視線追跡装置の測定誤差だけでなく、人間の実際の視線では停留点がターゲットを外れることが頻繁に起きてしまい、これも視線を解析する上でノイズとなってしまふからである。このような理由から、視線情報は教示情報として採用されていなかったと考えられる。しかし、最新の視線追跡装置の測定精度は非常に高く、視線インタフェースにも使えるようになっている [2]。我々は、ロボットの行動の中でも手を用いた作業を対象として、視線情報を用いた人間とロボットの協調作業の実現を研究目的としており、人間が日常的に行う手作業で用いる物体ほどの大きさであれば、問題なく注視物体を抽出することができる。

本論文ではまず第 2 章では本研究の関連研究について触れる。第 3 章では人間の視線の解析について述べ、第 4 章に、視線を利用した協調作業を提案し、そして、第 5 章では、協調作業を行うシステムの構築について説明する、第 6 章に構築したシステムを用いた協調作業の実装および実験結果を示す。最後に、第 7 章に結論と今後の課題を述べる。

## 第2章

### 関連研究

---



## 第2章 関連研究

本研究に関連の深い研究である“人間型ロボット”と“視線情報”の2つの分野の研究を紹介する。

### 2.1 人間型ロボットとのインタラクションに関する研究

ロボットは，人間行動を獲得するためには，何らかの人間行動モデルを構築する必要がある．この人間行動モデルは，ロボットが人間と同じ行動を再現したり，人間との協調作業を行う際に利用される．このようなタスクモデルに基づくロボットプログラミングの手法の1つである“Learning from Observation”に関する研究について説明する．

これは，環境物体との相互作用を伴うような比較的複雑なタスクにおいて，そのタスクが実行される領域を限定し，その中の全ての事象を構成することが可能な動作プリミティブを人間が設計する方法である．ここで，動作プリミティブとはタスクを記述する上で基本となる動作群のことである．ロボットは観察している人間の行動と動作プリミティブとのマッチングを行い，人間の行動をモデル化している．

従来の研究では，凸物体同士の組み立て作業を対象とした研究 [3][4][5][6]，非接触作業における手と操作物体の役割のモデル化を扱った研究 [7] などがある．また，玩具の組み立て作業において作業手順の因果関係をモデル化しロボットとの協調作業を実現した研究 [8] や剛体ではなく柔軟物であるひもを結ぶという動作に注目した研究 [9] も行われている．その中でも特に本研究と繋がりの深いロボットとの協調作業に関する研究 [8] について述べておく．

木村らは，人間とロボット間での柔軟な協調作業を実現することの可能なタスクモデルを提案した．この手法では，ロボットは視覚によって人間の作業を連続するイベント列として認識し，イベント相互の依存関係 (pre-condition & result) を解析することでタスクモデルを生成する．ロボットは，現在観察されるイベントとタスクモデルより，次にどんなイベントが必要とされるかを知り，作業を行う人間の進行状況に応じて適切な補助動作を行うことができる．この枠組みにより，単一のタ

スクモデルから様々な協調作業パターンを生成することができた。しかし、この研究では、人間の手形状や把持状態を認識し、作業目的・動作文脈に基づいて、必要な補助動作を計画・実行しているが、これでは人間が本当にロボットに助けを求めているのかどうかは不明である。作業中にロボットに介入して欲しくない場合もあるかもしれない。そこで、人間の注意や興味が含まれている視線情報を教示情報に利用することで、さらに柔軟な人間とロボットの協調作業が実現できると考えている。

## 2.2 視線情報を利用した研究

視線情報の利用の仕方は次の 2 つに大別できる。

- 視線による選択作業
- 視線を利用した意図の推定

前者は、視線によるポインティングである。例えば、マウスを用いてアイコンを選択するとき、マウスカーソルより先に視線はアイコンに到達しているため、視線でアイコンを選択できれば、手をマウスに伸ばす必要もなく、選択作業の効率が良くなると考えられる。視線を利用した研究のほとんどがこれに属する。従来の研究では、視線と顎の動きを利用して操作する車いすの研究 [10] や視線の高速性を利用して、階層メニューの選択を高速化する研究 [11] が行われている。これらの研究は、視線の高速性や正確性を利用しており、我々が視線を利用する理由である特性を活かしている研究ではない。

後者は、視線の生理的特性を考慮し、その特徴を捉え、人間の意図を推定するという利用方法である。視線による選択作業と比較して、研究例は少なく手法も確立していないといえる。従来の研究を 3 つ紹介する。

和英翻訳作業の支援環境を題材として、人間の「迷い」状態を自動的に抽出し、適切な支援を行うシステムを提案している研究 [12] がある。人間が翻訳の参考として日本語で単語を入力すると、その単語を含む英文および和文の対訳コーパスが複数画面上に表示される。そして人間の視線がいずれかのコーパス上にあるかを測定し、視線の移動パターンに対して N-Gram 解析を行うことで、視線パターンと人間の心理状態を対応付け、人間の「迷い」を検出している。

隠れマルコフモデル (Hidden Markov Model) を用いて、視線から人間の意図を確率的に確定する手法を提案した研究 [13] がある。画面上に表示されたキーボードを用いて英単語を入力するシステム、方程式を解く過程の推定などにこの手法を利用

している。この手法はあらかじめ人間の作業を正規文法 (regular grammar) による書き換え規則集合として表現する (プロセスモデルと呼ぶ)。また各プロセスにおける視線の対象領域を定める。例えば, “eye” という単語を入力する場合は, “e”, “y”, “e” の順にキーの位置を見るという規則を定める。次に観察された視線データからのプロセスモデルに最も一致しているかを HMM によって算出する。この手法は人間がキーの位置を正確に注視せずに多少ずれた場合でも単語の入力が可能であるという特徴を持つ。

英語の文書を読んでいる人間の視線から, 単語の理解度を推定する研究 [14] がある。この手法は文書中の各単語について, 人間の視線に基づき, 視線の移動方向, 移動距離, 停留時間等の属性から構成された特徴ベクトルを動的に更新する。そしてあらかじめ生成した決定木を用いて特徴ベクトルから人間の理解度を判定する。この手法の応用例として, 自動辞書引きシステムを試作しており, 人間が文書を読みながら, 「この単語の意味が分からない」と思っていると, 自動的に単語の意味を表示することで, 人間の文書理解を促進することを目指している。

他にも, 将棋を題材にして視線と熟練度の関係を研究した研究 [15] などがあるが, これらの研究に共通している点は, 視線のパターンから人間の意図や状態を推定している点である。作業内容や人間の状態により異なると予想される視線の動きなどを解析し, その特徴と捉えることができれば, 視線は人間の情報を推定するには有効な情報であることを示している。しかし, これらの研究では推定方法は統一されておらず, 推定方法はさまざまである。このような手法はまだ十分に研究が進んでおらず, どの程度有用であるかははっきりしていない。有用性を高めるには, 手法の適切性, 推定可能な意図の範囲, 推定精度, 個人差の影響などに関して, さらなる研究を進める必要がある。

## 第3章

### 人間の視線の解析

---

## 第3章 人間の視線の解析

本章では，人間の視線の解析について説明する．視線を利用するためには，その特徴を捉える必要がある．

### 3.1 視線とは何か

まず，眼の構造について簡単に説明する (図 3.1)．外界から入射した光は，角膜および水晶体の表面で屈折し，網膜に到達する．人間がある 1 点を注視しているとき，注視している物体からきた光線は網膜内の最も解像度の高い領域である中心窩に到達する．視線は視軸とも呼ばれ，厳密には注視点 (人間が注視している領域)，節点 (眼をレンズで構成された光学系であると考えたとき，どの角度から入射した光線でも入射角と出射角が等しくなる光軸上の点)，中心窩を結んだ直線として定義される．

視線測定によって得られた視線データを見ると，1 個所への停留と次の個所へのサッケード (跳躍運動) が交互に観察される．人間が眼によって情報を知覚できるのは視線が停留している間であり，サッケード中にはほとんど情報は得られない．サッケードは非常に高速な運動である．また，視線の停留時間は，一般には 100[msec] から 300[msec] 程度であるが，長い場合には 1000[msec] 以上になる場合もある．眼の動きは，サッケードだけでなくさまざまな眼球運動の組み合わせである．代表的な運動である固視微動，追従運動 (smooth pursuit movement)，跳躍運動 (saccadic movement) を説明する．

**固視微動** 眼球は 1 点を固視している場合でも微動している．固視微動はフリック (flick) と呼ばれる 0.03~0.05[sec] 間隔で不規則に生じる視角 20 分程度のステップあるいはパルス状の運動，視角 15[sec] 程度の振幅角を持ち 30~100[Hz] の周波数成分を有する微少振動トレモア (tremor)，視角 5 分以下でフリックの間に存在する非常に低速な変動ドリフト (drift) に分けられる．

**追従運動** 運動する物体を目で追う場合などの滑らかな眼球運動でその速度は 30[deg/sec] 程度であり，眼が動いている間に知覚が生じることがある．

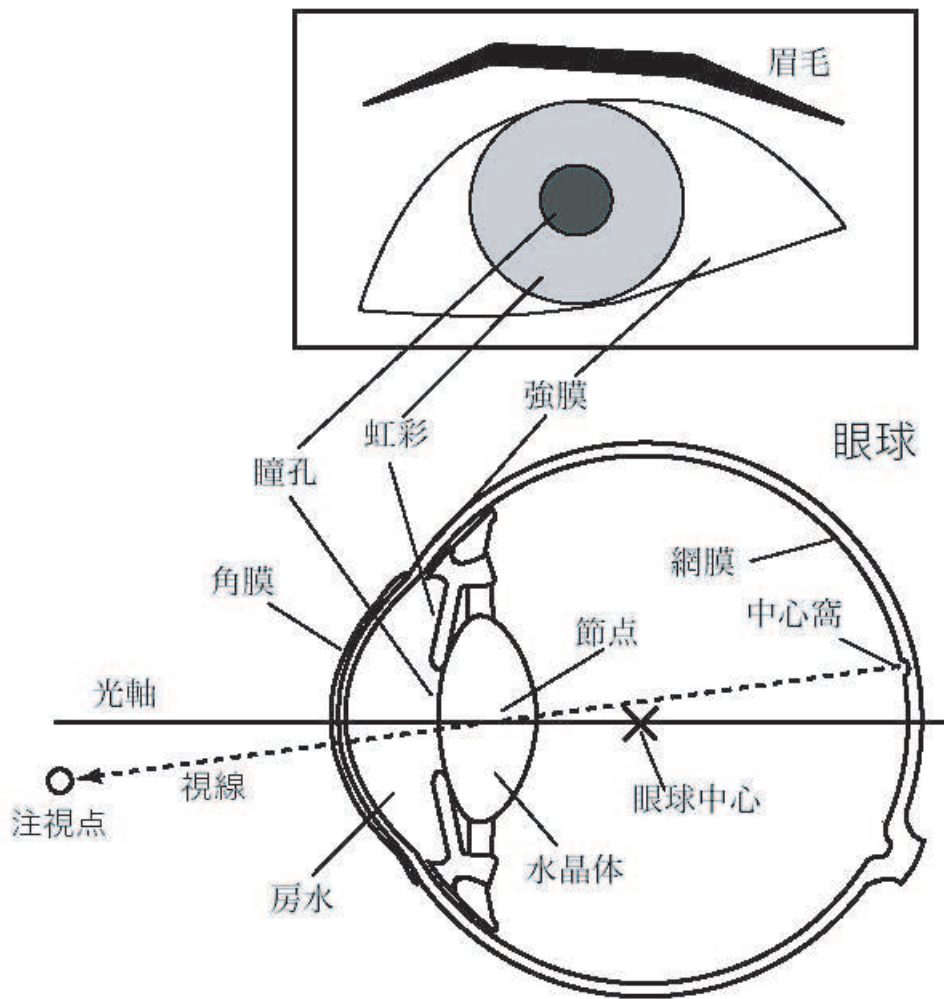


図 3.1: 眼の構造

**跳躍運動** 最高速度が300[deg/sec]にもなる高速な眼球運動で、視線を大きく変化させるときに生じる。眼が動いている間に知覚が生じることはほとんどない。

## 3.2 LEGOの組み立て作業における視線情報の取得

我々の研究の目的は、視線情報を用いて、人間の意図・状態を推定し、それに基づいて人間とロボットの柔軟な協調作業を実現することである。そこで、LEGOの組み立て作業を題材とし、作業中の人間の視線の特性を調べた。LEGOの組み立ては、同じパーツを組み立てた場合でも、複数の組み立てパターンが存在する。また、同じような形状を持つパーツでも機能部レベルでは全く異なるパーツとなる。このように、LEGOの組み立ては想像以上に複雑であり、組み立て作業中に、混乱や迷いが現れると考えられる。それらの状態と特徴的な視線の動きを結び付けることで、視線情報を用いることによる独特な協調作業が可能となる。まずはLEGOの組み立てを行っている人間の視線データを取得する実験を行った。その概要を以下に述べる。

### 3.2.1 実験内容

我々はLEGOの組み立て作業を題材とし、図3.2に示すように、卓上に散乱してあるLEGOの組み立て作業における視線の動きを計測した。作業時間は個人差はあるが、3分程度であり、比較的簡単な作業である。被験者にあらかじめ完成品の写真を覚えてもらい、その後組み立てを行うように指示した。

### 3.2.2 実験環境

#### 3.2.2.1 EMR-8

実験には、(株)ナックイメージテクノロジー社製の眼球運動計測装置EMR-8(図3.3)を用いる。この装置は単眼カメラ(アイマークカメラ)とアイマーク検出ユニットから構成されており、検出された眼球方向からユーザの左右の視軸方向を計測し、図3.4に示すように、アイマークカメラで得られたリアルタイムの視野画像に両眼の視点座標を重畳表示するシステムである。EMR-8の視線サンプリングレートは30[Hz]である。このシステムの視線方向精度と輻輳角による距離計測精度を評価している研究がある[16]。これによると、位置精度は注視点までの距離1mにつき1cm



図 3.2: 実験の様子

程度であることが分かっている。また、注視点距離が50cmから150cmまではほぼ正確に距離推定が可能であるが、それ以上の距離にある点を注視しても、輻輳角に差が生じず距離推定が不可能であることが分かっている。我々の研究対象が手作業であることを考慮すると、人間の比較的近く(150cm以内)を注視することになるので、距離も方向情報も利用できるといえる。視線測定中、被験者は自由に頭部を動かすことが可能であるため、EMR-8の装着による作業の制限はほとんどない状態である。また、EMR-8により計測されるデータは、両眼の水平・垂直方向の視野角データまたはEMR-8カメラ座標系における座標値である。

### 3.2.2.2 LEGO パーツ

実験で使用したパーツを図3.5および図3.6に示す。パーツ名、そのパーツが備えている機能部およびその個数、パーツの色およびその色のパーツの個数を示している。機能部は7つ(Axle, Open-Axle, Bearing, Peg, Hole, Screw, Screw-Hole)に分類され、色は6つ(Red, Blue, Green, Yellow, Black, Grey)に分類される。機能部について詳しく説明する。7つの機能部は、それぞれ対応する機能部とのみ組み付け可能である。組み付け可能な機能部の組み合わせを図3.7に示す。また、“Axle and Bearing”の組み付けと“Screw and Screw-Hole”の組み付けには、組み付け直後に必ず“Driver”を用いてネジ締めを行う必要がある(図3.8)。





図 3.3: EMR-8

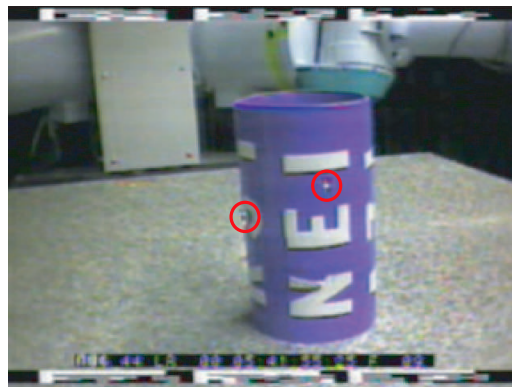


図 3.4: アイマークカメラ画像

被験者に組み立て作業を指示した LEGO の完成図を図 3.9 に示す。作業に必要なパーツ数は 13 つである。卓上にある総パーツ数は、不必要なパーツも合わせて 24 つである。また、本研究では、パーツの機能部の情報に着目して作業の解析を行っており、この妨げとなる色の情報は完成図から取り除いてある。

### 3.2.3 被験者

被験者は大学生の男 5 名である。実験の参加に対して謝礼は支払われていない。

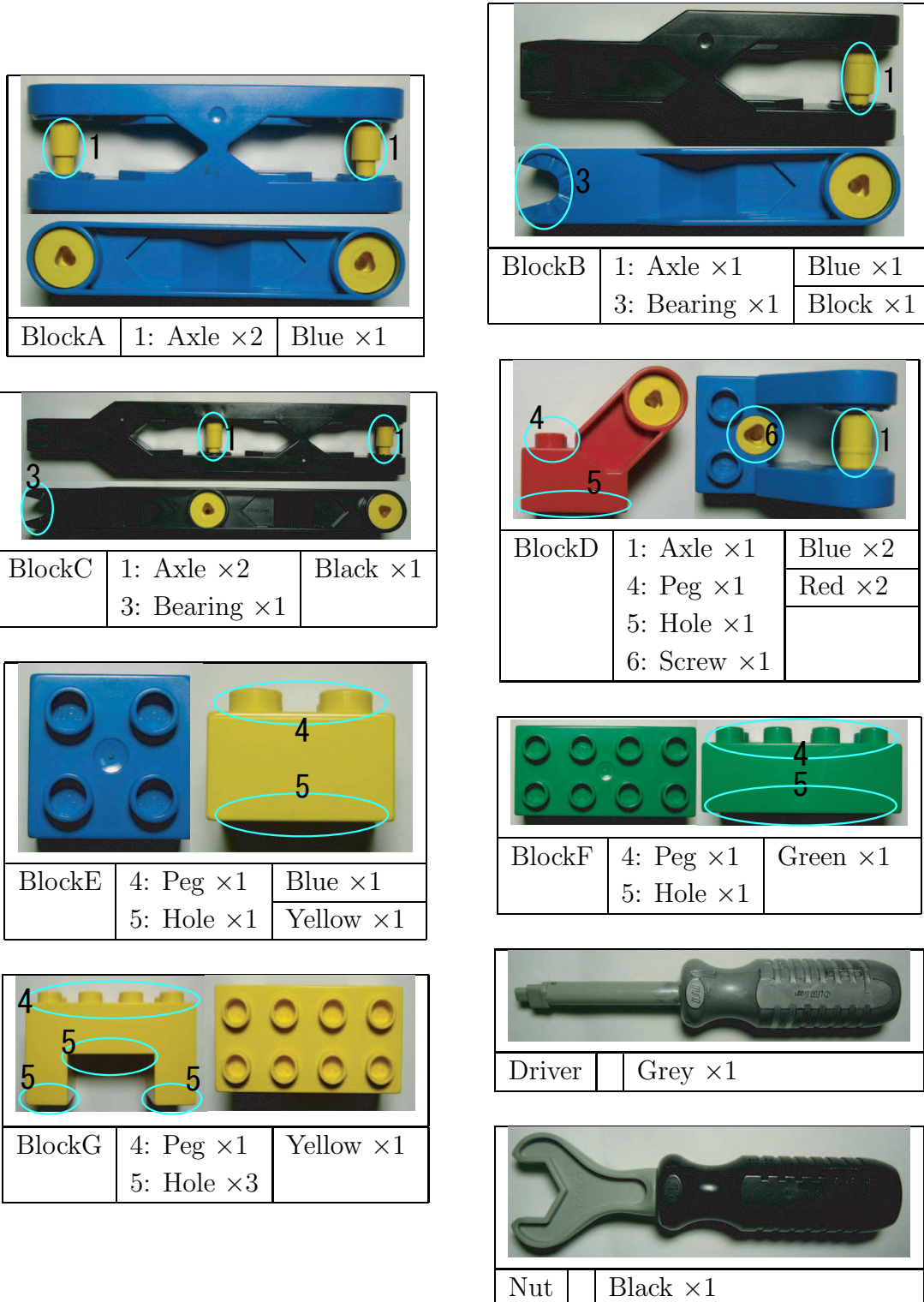


図 3.5: LEGO Parts I

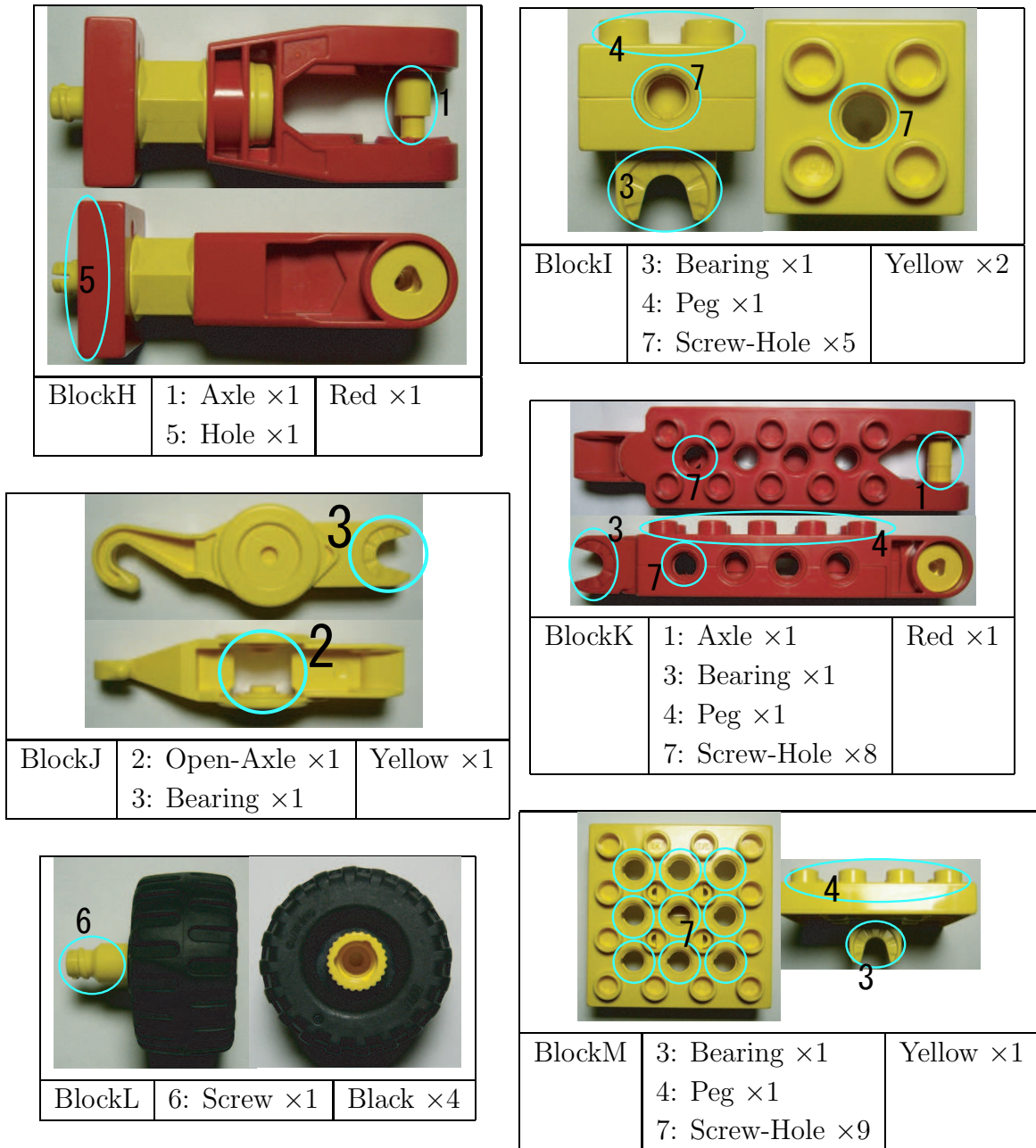
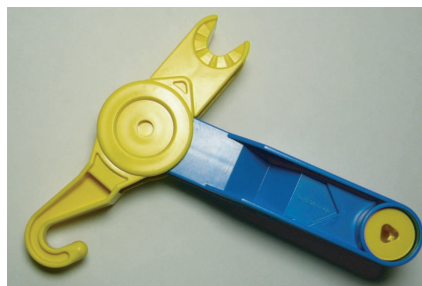


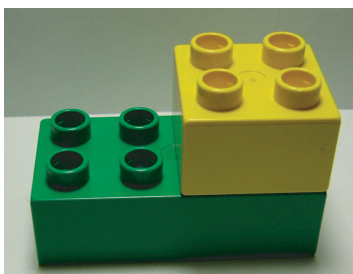
図 3.6: LEGO Parts II



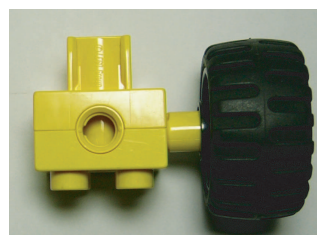
(1) Axle and Bearing



(2) Open-Axle and Bearing



(3) Peg and Hole

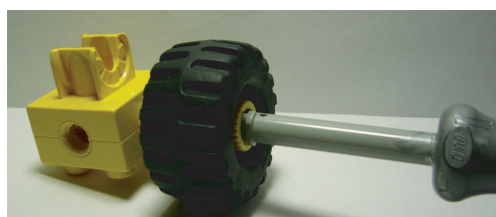


(4) Screw and Screw-Hole

図 3.7: 組み付けの種類



(1) Driver into Bearing



(2) Driver into Screw

図 3.8: ドライバを必要とする組み付け

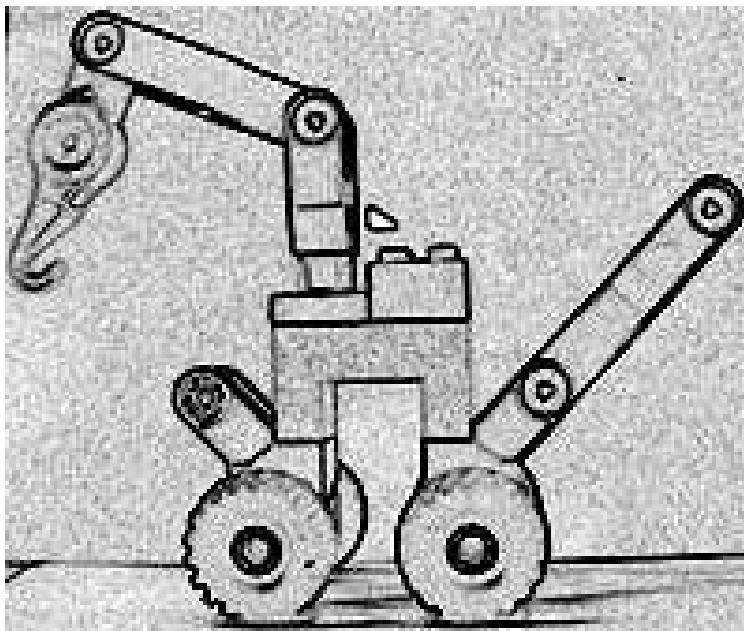


図 3.9: LEGO の完成図

### 3.2.4 実験手順

以下の手順で実験は行われた。

1. 被験者は EMR-8 を装着し，EMR-8 を使用するためのキャリブレーションを行う。
2. 被験者は提示された LEGO の完成図を記憶する。
3. 被験者は記憶を基に LEGO の組み立てを行う。

被験者は当然のことながら，LEGO のパーツの配置については実験を開始するまで知らない。また，組み立て作業中に必要なパーツが分からなくなっても，被験者自身が正しいと思うパーツを使用して組み立てを行うように指示した。さらに，完成図と全く関係の無い組み付けを行った際には，その都度訂正し，進行が完成図から大きく隔たらないようにした。

## 3.3 視線の解析

3.2 の実験で得られたデータを解析することで、視線の特徴を抽出する。

### 3.3.1 EMR データのセグメンテーション

EMR から得られる視線情報は両眼の水平・垂直方向の視野角データの時系列情報である。データ解析において、このデータを注視している物体ごとにセグメンテーションし、意味のある集団にまとめた後、解析を行った方が人間の直感に訴えやすい。そこで、EMR データのセグメンテーションを行い、注視点を算出する。

注視点は、視線の停留時間と跳躍運動 (saccadic movement) の分離により、基準値を設定することにより定義づけることができる。つまり、注視の時間的定義と速度的定義で定めた 2 種類の基準値を用いることで、セグメンテーションを行う。

**注視の時間的定義** 長文読解における眼球運動測定実験で 99[msec] 以上を注視として定義している研究 [17] があるが、150[msec] 以上を注視として定義づけることが最も一般的であるという報告もある [18]。

**注視の速度的定義** 固視微動の速度成分と、ある速度以下の追従運動成分を注視に含めた考え方であり、人間が情報を取り入れうる時間を考慮した定義である。音読における視線速度を 5[deg/sec] 以下と定義した研究 [19] がある。

まず、跳躍運動の基準値、つまり注視の速度的定義に基づくセグメンテーションを行った後、視線の停留時間 (注視の時間的定義) を加えたセグメンテーションを行う。

#### 3.3.1.1 注視の速度的定義に基づくセグメンテーション

以下の解析方法で、EMR-8 から得られた視野角データのセグメンテーションを行う。

**解析方法** 各視線データの重心点を中心とする直径 10[deg] の円を停留範囲とする。停留範囲内に存在する視線データ群を同一グループと判定する。この方法を重心点方法と呼ぶ。

作業は卓上で行われるので、眼球から注視物体までの距離は 50[cm] 程度である。停留範囲の円の直径 10[deg] は、物体周辺では直径 8[cm] 程度の円となる。これは物

体の大きさとほぼ同等である。

また，停留点を以下のように定義する。

**停留点** 重心点方法により同一グループと判定された視線データの重心点

重心点方法による視線データのセグメンテーションを行った。得られた結果を評価するため，作業中のアイマークカメラの動画を録画しておき，録画像を用いて手動でセグメンテーションした結果を真値として評価する。結果を表 3.1 に示す。

表 3.1: セグメンテーションの結果

	Error1	Error2	Error3	Truth	Total
Selecting State	6	5	12	59	77
Working State	4	3	8	48	60

表 3.1 の説明をする。

**Selecting State** 組み立て作業を行うための物体を探している状態 (探索択状態).  
視線は卓上にある LEGO のいずれかに停留している (選択状態).

**Working State** LEGO の組み付け作業を行っている状態 (組み付け状態). 従って,  
両手に LEGO を把持しており，ほとんどの場合，視線は作業物体に停留して  
いる。

**Error1** 視線の移動時 (跳躍運動) に停留が検出されるエラー。

**Error2** 停留しているが検出されないエラー。

**Error3** 同一物体に複数回の停留が検出されるエラー。

**Truth** 適切に停留が検出される。

的中率を算出する。なお，上記評価は人間が手動で行った。

- 重心点方法により得られたセグメンテーションの的中率  
 $107 \div 137 = 0.781022$



- “Error3” をエラーに含めない場合の的中率  
 $(107 + 12 + 8) \div 137 = 0.927007$
- 停留が検出されない率  $8 \div 137 = 0.058394$

以上の結果から、本手法では、70% (Error3 を含めない場合は 80%) 以上の的中率を得ることができた。また、検出されない率は 6% 未満という良好な結果を得ることができた。

### 3.3.1.2 注視の時間的定義に基づくセグメンテーション

3.3.1.1 で得られた結果に、注視の時間的定義を考慮する。つまり、重心点方法でセグメンテーションした後に停留時間によるセグメンテーションを行う。図 3.10 に全被験者における停留時間の分布を示す。もっとも発生頻度の高い停留時間は 200[msec] であり、次に発生頻度が高かった停留時間は 167[msec] であった。これらの結果は従来の研究 [15][20] で報告されている結果と同様の結果であり、また人間が注視した物を認識するために必要とされる停留時間の一般的に言われている値とも同じである。停留時間の最小値、最大値、平均値、分散、標準偏差を表 3.2 に示す。

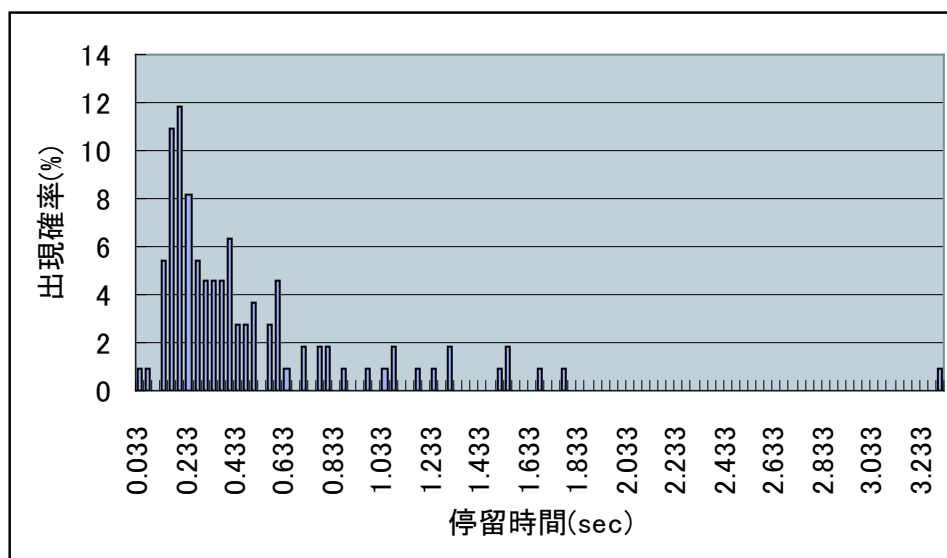


図 3.10: 視線の停留時間分布



表 3.2: 停留時間の統計結果

	$T$	$T'$ (Excluding Working State)
Min	0.033	0.067
Max	3.3	1.3
Ave	0.48063636	0.390205
Var	0.21213034	0.092166
Stdev	0.4605761	0.303589

表 3.2 の  $T'$  は作業中の停留時間を除いた結果である。つまり、物体を選んでいる状態の停留時間を対象とした結果である。表 3.2 より、組み付け状態の停留時間の方が探索状態の停留時間よりも停留時間が長い傾向にあることがわかる。組み立て作業中は、作業物体を長い間注視するためである。

重心点方法に基づくセグメンテーションの結果より、0.167[msec] 以上が全体の 92.7% を占めていた。従って、実験の結果および従来の研究結果より、0.167[msec] を LEGO の組み立て作業における停留時間の基準値とし、この値以上の停留時間の停留を注視とした。停留時間を考慮したセグメンテーションの結果を表 3.3 に示す。

表 3.3: 停留時間を考慮したセグメンテーションの結果

	Error1	Error2	Error3	Truth	Total
Selection State	4	5	8	52	64
Working State	0	3	7	44	51

的中率を算出する。

- 重心点方法 + 停留時間により得られたセグメンテーションの的中率  
 $96 \div 115 = 0.834783$
- “Error3” をエラーに含めない場合の的中率  
 $(96 + 8 + 7) \div 115 = 0.965217$

3.3.1.1 で求めた結果よりも良好な結果が得られた。理由としては、Error1(視線の移動時(跳躍運動)に停留が検出されるエラー)が停留時間の基準値未満である場合が多く、除外可能なエラーであることが考えられる。跳躍運動は非常に高速な運動であるため、停留範囲に留まることは少ないからである。

以上の結果から、視線の停留時間および跳躍運動の基準値を用いたセグメンテーションが有効な手段であることが示せた。つまり、本手法により EMR-8 から得られた視野角データを注視物体および注視時間のデータに変換することが可能となった。

次に、注視物体および注視時間のデータを用いて、LEGO 組み立て作業における視線の特徴を検出する。

### 3.3.2 注視時間の特徴

人間の LEGO 組み立て作業の大まかな流れは以下のようになっている。

1. 必要なパーツの探索
2. 発見したパーツの把持
3. パーツの組み付け
4. 1. へ

ここでは、1. と 2. における特徴を発見する。

#### 3.3.2.1 探索状態における注視時間の特徴

1. の探索状態と 2. の把持状態を分離することができれば、人間が探索したパーツをロボットが把持するという補助が可能になる。注視時間による探索状態と把持状態の分離が可能であれば、人間に無理な動作を強要することなく、ロボット自身が、人間は探索状態であるかどうかを把握することができる。

そこで、LEGO の組み立て作業中における探索状態の停留時間の分布および注視から把持に行動移行した際の停留時間の分布を図 3.11 に示す。停留時間が 0.6 秒以上のときに把持している率は約 70 % である。また、停留時間が 0.6 秒未満で把持を行っているデータの半分が “Driver” を使用するときである。これは、ドライバを何回も使用し、使用後には自分で卓上に置くため、ドライバの場所を記憶しているために短い停留時間にもかかわらず把持が可能であると考えられる。“Driver” の把持をデータから除いたデータにおいて、停留時間が 0.6 秒以上のときに把持している率は約 77 % である。このように、必要物体を探索している時と把持直前の注視時間には異なる特徴があることがわかった。つまり、注視時間による探索状態と把持状態の分離が可能であるといえる。

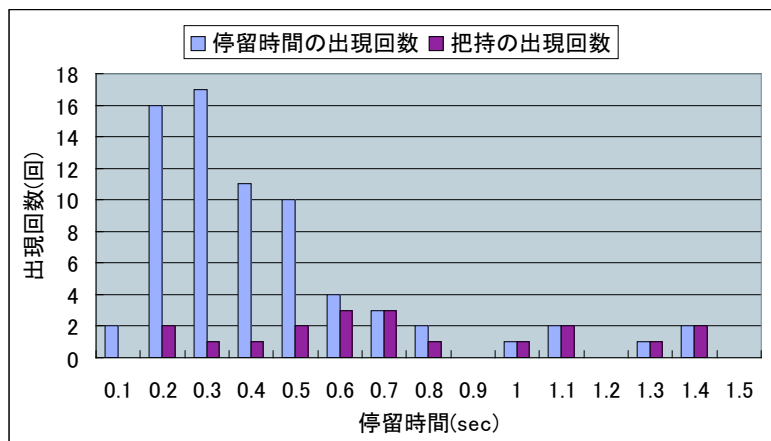


図 3.11: 視線の停留時間と把持の分布

### 3.3.2.2 注視から把持動作開始までの時間差

次に、2.の把持動作における特徴を検出する。必要な物体を探索して発見後、把持動作開始までの時間差があれば、その時間を利用した協調作業を考えることができる。

視線の動きと手の動きは必ずしもパラレルの関係で動いているのではない。しかしながら、互いに何らかの関係があるという報告 [21] もある。また、視線の動きと手の動きの時差に着目した研究 [22] もある。この研究では、注視してから手が動き出すまでの時差が生じることについて述べている。動作開始以前に、長期記憶に蓄積されてある知識から、該当する動作の行動レパートリー (知識や経験の集合) を選択・抽出し、リハーサルを行った後、必要な情報を短期記憶にメンタル・イメージとして保管する。さらに、入力される視覚情報や身体的情報と形成されたメンタル・イメージとを照合させながら動作は遂行されるという一連の処理が行われるため、時差が生じると報告している。このような研究結果を参考に、LEGO の組み立て作業において、注視してから手が動き出すまでに時間差が生じるかどうかを録画しておいた動画を使って調べた。

LEGO の組み立て作業においても同様に、注視から手の始動までの時間差が見られた。結果を図 3.12 に示す。

図 3.12 のデータは作業時間に沿って表示されているが、作業の前半のデータと後半のデータではまったく異なる結果となることがわかる。これは、時間が経つにつれて、作業物体の特徴や位置に関する知識を得ることが原因と考えられる。

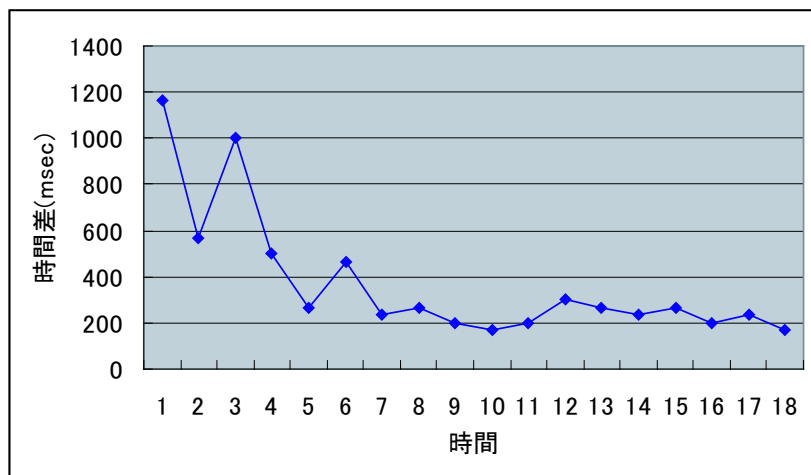


図 3.12: 注視から把持へ行動移行に生じる時間差

### 3.3.2.3 考察

3.3.2.1 と 3.3.2.2 とにおいて、作業中の注視時間に着目し、その特徴を検出した。

3.3.2.1 では、作業において、探索段階と把持段階を注視時間で分離することが可能であることを示すことができた。従って、注視時間を調べることで、人間の状態が探索状態であるかどうかを把握することができる。

3.3.2.2 では、注視から把持に行動移行する際に時間差が生じることを示した。しかし、生じる時間差は非常に短く、この時間差を利用した協調作業は困難であると思われる。

### 3.3.3 視線の履歴情報による人間の“迷い”状態の予測

前述までは、視線情報として「注視時間」に着目した特徴を検出した。次に、視線情報として「視線の履歴情報」に着目する。

探索状態において、必要なパーツを発見するために、複数のパーツを順次注視することが大半であり、場合によっては同じパーツ間での注視の繰り返しが見られるかもしれない。このように、どのパーツをどの順番で注視したかという「視線の履歴情報」を解析する。ここでは、人間の作業時における“迷い”の状態を推測することを目指し、迷い状態における特徴的な視線の動きを発見する。

### 3.3.3.1 注視物体の履歴への登録

LEGO の組み立て作業における視線の履歴は、作業中の人間の EMR-8 データをセグメンテーションし、同一グループに判定された注視物体とそのときの停留時間の情報を逐次登録していくことで作られる。ただし、停留時間が基準値 (0.167[sec]) 未満のデータは登録しない。

### 3.3.3.2 視線の履歴情報の解析

3.3.3.1 で得られる視線の履歴を解析することで、“迷い”における特徴的な視線の動きを発見する。

LEGO のパーツは次の 2 点の情報を保持している。

- 色 (Red · Blue · Green · Yellow · Black · Grey)
- 機能部 (Axle · Open-Axle · Bearing · Peg · hole · Screw · Screw-Hole)

しかし、被験者には、色情報は与えられていないため、機能部情報もしくは形状情報に基づいた探索を行っていると考えられるが、ここでは機能部情報のみを扱う。注視物体の履歴を解析すると以下のような特徴がみられた。

- 組み立て段階の作業物体 (組み立て途中の未完成物体) の機能部と対応する機能部を保持する物体を注視している。

この特徴は、作業を遂行していく上で当然の現象であるといえる。作業中の物体の機能部に対応する機能部を保持する物体以外では作業は進行しない。組み立て段階の作業物体には、組み立て順序の自由度から、組み付け可能な機能部が複数 (図 3.13) あり、その各々の機能部にマッチする候補パーツが卓上には多数存在する。従って、どの候補パーツが適切なパーツなのかかわからず、“迷い”状態に陥る場合があると予想される。人間がこの“迷い”状態に陥った際に、多数の候補パーツの中から、視線を利用して候補パーツを絞り込み、人間が組み付けを行いたいと考えている機能部に対応し、かつタスクに適切なパーツを推測できれば、人間の状態を考慮した柔軟な協調作業が実現できる。視線の履歴解析については、4.2.2.2 で詳しく説明する。

### 3.3.4 組み付け状態における視線の特徴

3.3.2 と 3.3.3 では、物体を選んでいるとき (探索状態) の視線の特徴について述べた。ここでは、物体を把持し、組み付け作業を行っているとき (組み付け状態) の視

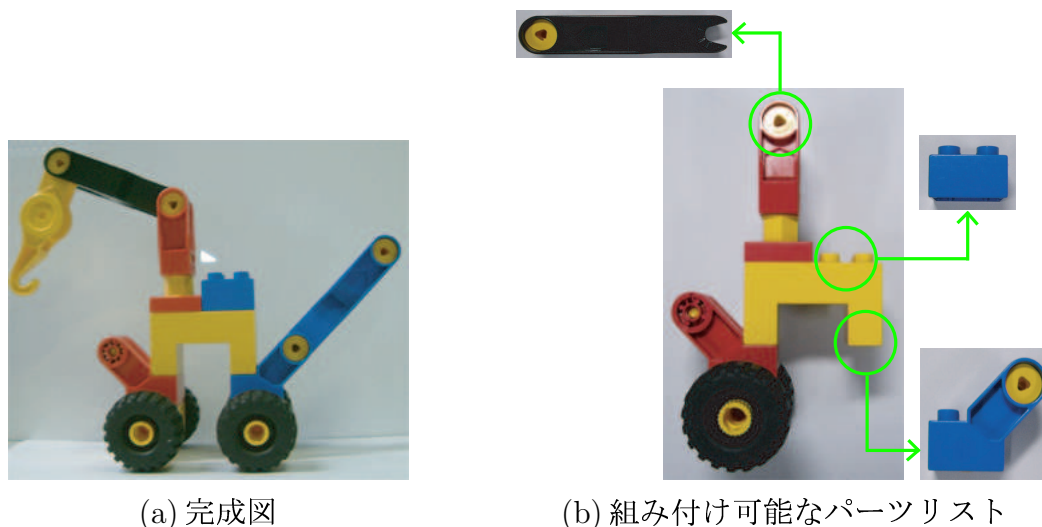


図 3.13: 組み付け可能なパーツリストの例

線の特徴について述べる。作業中の人間の視線には以下のような特徴が見られた。

1. 組み付けを行う前には、両手に把持している作業物体を交互に注視するが、注視点は機能部にほぼ固定されており、物体全体を見ているのではない。
2. “Driver” を使用してネジ締めを行っている間、常にネジ部を注視しているのではなく、卓上の物体を探索していることもある。

1. に関して詳しく説明すると、組み付けを行う前に、2つのパーツの対応する機能部の位置・姿勢を合わせる必要があるので、対応する機能部を注視することになる(図 3.14)。つまり、関係のない機能部を注視することはほとんど考えられない。LEGO のパーツは1パーツに複数の機能部が付いており、同じパーツの組み付けの場合においても複数の組み付けパターンが考えられる。そこで、組み付けを行う直前に注視している機能部から、組み付けパターンが推測できる。つまり、注視機能部による組み付け意図の抽出が可能になるといえる。

2. に関して詳しく説明すると、ネジ締めというような単純作業においては作業状態を常に把握していなくても作業を行うことができる。そのため、ネジ締め作業中に、卓上の物体に注意がいくこともある。このような単調作業における視線の動きから、次の作業に必要な物体を推測できる可能性がある。

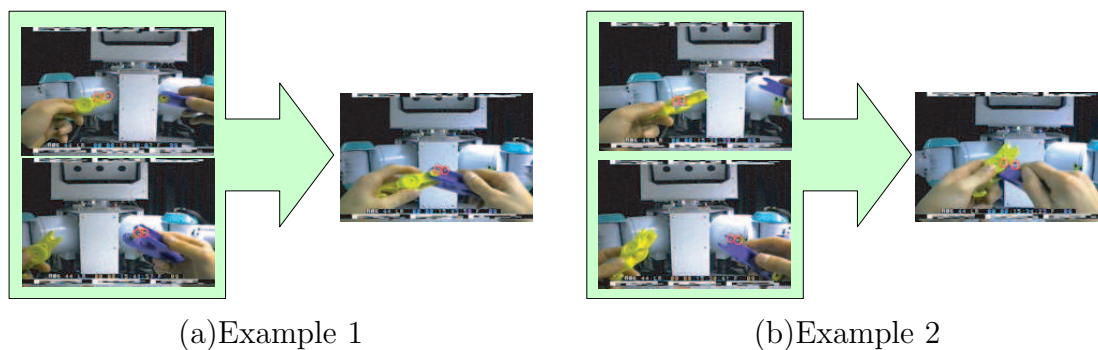


図 3.14: 組み付けを行う前の注視点の様子

### 3.4 考察

本章では、視線の一般的な特徴を述べ、作業中の視線を取得する実験を行い、そのデータを解析した。そして、解析結果より、LEGOの組み立て作業を行っている人間の視線の特徴を検出することができた。その中でも、ロボットの補助動作に利用できると考えられる特徴は次の3つである。

- 注視による探索状態と把持状態の分離
- 視線の履歴解析による“迷い”状態における適切な物体の提示
- 注視機能部による組み付け意図の抽出

これらの3つのことは、視線という無意識の動作を解析することにより可能になるので、人間に余計な動作を強いる必要なしに抽出可能である。従って、視線により人間の意図・状態を抽出し、人間の明示的な指示なしに、ロボット自身が適切な補助を考え、実行するような柔軟性に富んだ協調作業の実現ができる。

## 第4章

### 視線を利用した協調作業の提案

---



## 第4章 視線を利用した協調作業の提案

第3章ではLEGOの組み立て作業を題材とし、視線の履歴や注視時間から人間の意図や状態を推定することができた。本章では、人間の意図や状態に基づくロボットの協調作業を提案する。

### 4.1 タスクモデルの生成と補助動作の計画

我々は、木村らの協調行動フレームワーク [8] を拡張する形で、人間とロボット間の協調作業をより柔軟で効率的なものに発展させる。つまり、タスクモデルの生成段階は従来の手法を利用し、ロボットがタスクモデルに基づいて補助動作を行う際に、視線を利用した新しい手法を提案する。まず、木村らが提案したタスクモデルの生成手法について説明する。

#### 4.1.1 タスクモデルの枠組み

タスクはLEGOの組み立て作業を対象とする。従来の手法では、「作業対象物の機能部レベルでのタスクモデル」の枠組みを以下の要素で定義している。

1. 組み付けられる2つの対象物の機能部の状態 (Result)
2. Resultが発生する前提条件 (Preconditions)

具体的には、組み付けられるものは軸合わせや面合わせのオス・メス組み、ネジとドライバの組などであり、1.はその組み合わせにおいて機能部座標が一致した状態を表す。2.も機能部の状態の1つであり、ネジ部を持った軸どうしの組み合わせのあとドライバによるネジ締め動作が行われるなどを表し、動作文脈と呼ばれるものの一部である。

### 4.1.2 タスクモデルの自動生成

ロボットの視覚処理により、LEGOのパーツの3次元認識と把持の有無、把持されているならその把持形状、パーツの機能部間の関係を検出することができる。ここでは、人間によるパーツ把持の発生やパーツ機能部の組み付けの発生などを事象(Event)と呼ぶ。人間が単独で行う一連の作業の観察・解析から、タスクモデルを以下の手順で自動生成することが可能である。

1. 現在のシーンを解析し、検出された事象を“event stack”と呼ばれるスタック型の作業領域に積む。
2. event stackの最上位にパーツやツール間の機能部の組み付け状態“fixed”が検出されたとき、機能部レベルのタスクモデルを生成する。
3. 1に戻る。

2.については具体的には、event stackにおいて、最上位の状態fixedを機能部状態(Result)とし、状態fixedより下に積み上げられている状態を取りだし、前提条件(Preconditions)とする。

この手法に基づき、人間が行う作業を解析することでタスクモデルを生成することができている。

### 4.1.3 補助動作の計画

協調を行うための補助動作は次の3つに分類される。

**タスク目的を実現するための補助** 人間がタスクの目的を実現するための前提条件の一部しか行わなかった場合、残りの前提条件をロボットが実行し、タスクの目的(Result)も実行する。

**作業を効率良く行うための補助** 人間があるタスクモデル中の操作をすべて行ってしまった場合、あらかじめ次のタスクの準備を行うために、次のタスクモデルを呼び出し、前提条件の一部を実行する。

**把持した対象物を人間に手渡す補助** ロボットが対象物を把持した状態で人間がそれを渡して欲しいといった受け取り動作を行った場合、ロボットは把持している対象物を人間に手渡す。

人間が行う動作に対して、4.1.2 と同様に、事象と event stack による人間の動作解析を行い、タスクモデルに基づいて、人間の操作とロボットの操作の分離、機能部の状態からロボットの操作の生成を行うことで、補助動作の計画は行われる。ロボットは現在のシーンを解析、人間の動作を理解することで、適切な補助動作を自動的に選択し、実行することができる。

## 4.2 視線を利用した補助動作の提案

### 4.2.1 従来の補助動作の問題点

従来の手法において、タスクモデルと補助動作の計画に基づき、ロボットは視覚により現在の人間の動作を理解することで、人間が行う様々な動作に対して適切な補助動作を行うことができているが、以下の問題点がある。

- ロボットによる補助動作が行われるには、作業中に新たな事象が一定時間経過しても発生しない条件が必要
- ロボットが把持している対象物を手渡す補助には、人間が受け取り動作を行うことが必要 (明示的な指示が必要)
- タスクが一方方向

明示的な指示なしに、ロボットが人間の意図・状態に基づいて補助動作を開始することができれば、柔軟性および協調性の高い協調作業が実現できる。また、従来のモデルでは、タスクが一方方向であり、一つのタスクに対して、一つのモデルが対応しており、単一の補助動作のみ可能である。従って、作業の流れが一意に決まっており、補助の内容を変化させることは困難であり、柔軟性に欠けている問題がある。そこで、人間の視線を利用することで、従来の研究の問題を改善し、より柔軟かつ効率的な協調を実現する。

### 4.2.2 探索状態における補助動作の提案

#### 4.2.2.1 注視による必要物体の教示

3.3.2.1 で、注視時間によって情報の探索段階と把持段階の分離することが可能なことを示し、3.3.2.2 で、注視から把持に行動移行する際に時間差が生じることを示

したが、後者は3.3.2.2で説明した通り、時間差は生じるが、その時間は非常に短い。従って、この時間差を利用したロボットの補助動作は現実的ではない。ここでは前者の性質を利用する。

情報の探索段階と把持段階においては明らかな注視時間の差があるため、人間が物体を注視している時間がある閾値を越えた場合、探索段階ではないことがわかる。そこで、注視時間を利用した視線によるポインティングの補助動作を提案する。

1. 一定時間以上の注視時間が検出される
2. 一定時間以上、新たな事象が発生しない
3. ロボットは注視物体を把持する
4. 人間がLEGO パーツを把持していれば協調して組み立て作業を行い、把持していなければLEGO パーツを手渡す

この補助は、次のような場合のときに有効である。

- 両手でLEGO パーツを把持しており、人間が把持することができない場合
- LEGO パーツが人間から離れたところにあり、人間が把持するよりもロボットが把持する方が効率的である場合

#### 4.2.2.2 “迷い”状態における適切な物体の提示

本研究では、LEGOの組み立て作業を題材としており、扱うパーツ数は24つであり、必然的に「物体の探索」という状態が考えられる。人間は、必要な物体を探索しようと試みるが、どの物体が作業に適切なものであるか分からない状態、つまり、“迷い”の状態に陥る可能性もある。そこで、人間が“迷い”状態に陥った場合に、タスクモデルと視線情報を利用して、人間の“迷い”を解決する適切な物体を推測し、それを提示する補助を提案する。タスクモデルと視線情報の利用方法について説明する。

**タスクモデル** ロボットはタスクモデルにより、人間により最終的に組み立てられる完成物体に関する情報は獲得している。従って、ある組み立ての段階において、組み立て途中の物体に組み付けを行うべき物体のリストを得ることができる。

**視線情報** 人間が最終的に組み立てる物体は一意に決まっているが、組み付けを行う順番には自由度があり、ある組み立ての段階における候補パーツは複数存在する。そこで、視線情報を利用することで、候補パーツから人間が組み付けを行うことを考えているパーツを絞り込む。

ある組み立て段階において、組み立て順序の自由度によりタスクモデルは複数存在する。つまり、作業の進行に関して正しい解は複数存在する。タスクモデルから正しい解を提示することはできるが、それが人間が組み付けを行おうと考えている機能部にマッチする解であるかは不明である。そこで、人間が注視している物体の機能部に基づく解析から、どのタスクモデルに基づく組み付けを行うかを推定する。解析手法を説明する。

可能な候補パーツの中から、視線情報を利用して候補パーツを絞り込み、その中からタスクモデルに基づき、適切な物体を推測する。この推測された物体が、人間が組み付けを行おうと考えている機能部に対応し、かつタスク遂行に正しい物体であるといえる。

具体的に、タスクモデルと視線情報をどのように利用して適切な物体を推測するかを説明する。図 4.1 に示すように、組み立て途中の物体があるとする。

タスクモデルに基づき、候補パーツは、“BlockB”、“BlockD”、“BlockE” の 3 つである。ここで、人間が注視する物体の機能部に基づく「投票式判定」により、人間が意図している組み立てに適切な物体を推測する。人間が注視する物体は次の 2 つに分けられる。

1. 次のタスクで使用される候補パーツ (“BlockB”、“BlockD”、“BlockE”)
2. 次のタスクでは使用されない候補パーツ (“BlockJ”、“BlockL”)

人間が注視したある物体が保持する機能部の中で、組み立て途中の物体が保持する組み付け可能な機能部と対応している機能部を抽出する。次に、抽出された機能部を保持する 1. の候補パーツに 1 票ずつ投票する。例えば、“BlockD” を注視した場合、“BlockD” の機能部の中で、組み立て途中の物体が保持する組み付け可能な機能部 (Axle, Peg, Hole) に対応する機能部は、“Peg” と “Hole” である。そこで、人間が “BlockD” を注視した場合には、“Peg” と “Hole” のどちらかの機能部を利用したいと推測し、1. の候補パーツの中で、“Peg” または “Hole” を保持する物体である “BlockD”、“BlockE” に 1 票ずつ投票する。人間が物体を注視する度にこの投票を繰り返し、最大得票数を獲得した物体が、人間が組み付けを行う意図がある機能部に組み付け可能で、かつタスクモデルに基づき正しい解であると推定する。

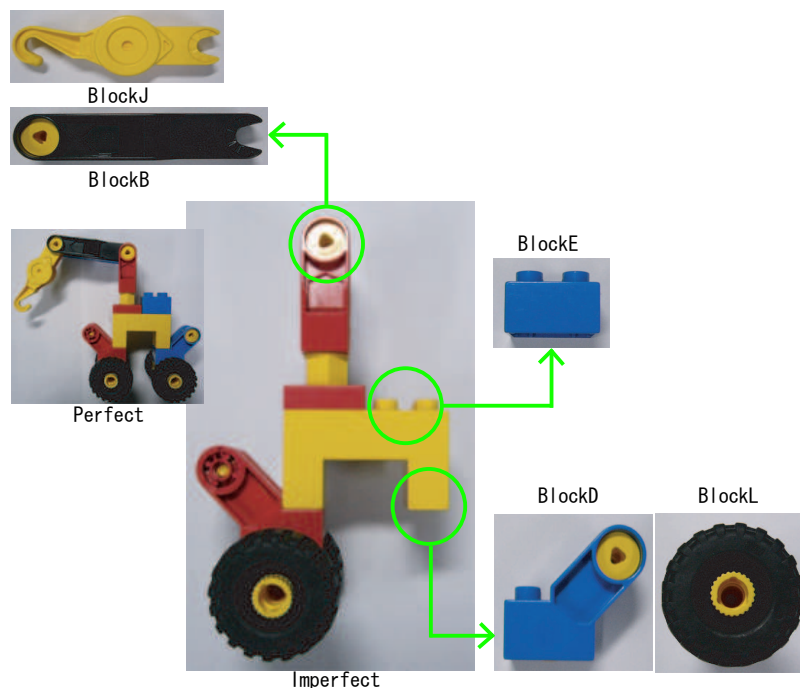


図 4.1: 組み立て途中の物体と候補パーツ

この解析手法により得られた結果を図 4.2 に示す。

極端に探索時間が短くなければ、人間が把持した物体が最大得票数を獲得する結果が得られた。探索時間が極端に短い場合、総投票数が少なくピークが現れにくいいため、推定に失敗する。

この結果から、人間が“迷い”状態に陥った場合に、人間が意図を考慮した正しい解を提示する補助が可能といえる。ここで人間が“迷い”状態に陥っているかどうかの判定は、ある一定以上の時間が経過してもパーツの組み付けが行われない場合、“迷い”状態に陥ったと判定する。そして、ロボットは、その時点での最大得票数を獲得している物体を把持し、人間に手渡すことで、迷いの解消を試みる。

### 4.2.3 組み付け状態における補助動作の提案

タスクモデルにより、組み付けられるパーツの組みは一意に決まっている。しかし、LEGOのパーツには複数の機能部を保持しているパーツもあるため、図 4.3 に示すように、タスクモデルの中には、同じパーツを使用した組み付けではあるが、異

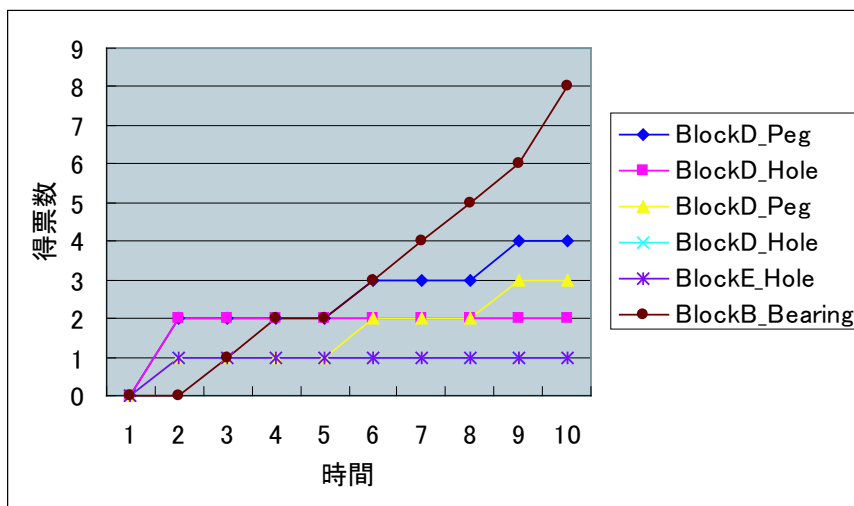


図 4.2: 機能部に着目した投票式による解析結果

なる機能部を使用した組み付けが行われるモデルも存在する場合がある。

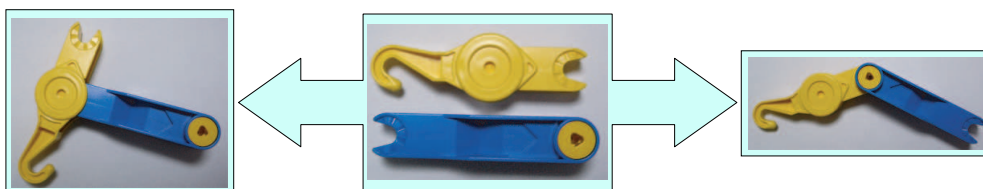


図 4.3: 同じパーツによる異なる機能部を使用した組み付け例

タスクの組み立て順序は一意ではないため、タスクモデルの中に、同じパーツを使用し、かつ異なる機能部に関する組み付けを行うモデルが存在する場合、従来の手法では、どのタスクモデルが実行されるかは、実際に組み付けが行われた後でなければ知ることはできない。そこで、視線を利用することで、実際に組み付けが行われる前に、実行されるタスクモデルを推測し、タスクモデルに基づいて次のタスクの準備をロボットが行う補助を提案する。

3.3.4で述べたように、人間はLEGOの組み付けを行う前には、組み付けを行うと考えている機能部を注視する。従って、どの機能部を注視しているかを調べることで、組み付けを行う前にどのタスクモデルが実行されるかを推測することができる。そして、タスクモデルに基づいて次のタスクの準備をロボットが行うことが可

能となる。ロボットは次のような処理を行い、補助動作を切り替える。

1. 組み付けを行う前に、人間がどの機能部を注視しているかを調べる
2. 注視機能部により、実行されるタスクモデルを推測する
3. タスクモデルに基づき、次のタスクの準備を行う

この補助動作の有効性として、「効率性の向上」があげられる。

**効率性の向上** 人間が組み付けを行う前に、タスクモデルを推測できるため、人間が実際に組み付けを行っている間に、次のタスクの準備を行うことができ、人間が組み付けを行っている時間を有効に使用している。

#### 4.2.4 まとめ

4.2.2 および 4.2.3 で視線を利用した補助動作を提案した。補助内容を以下に示す。

- 注視時間を利用した視線によるポインティング
- “迷い” 状態におけるタスクモデルと機能部情報に基づく適切な物体の提示
- 注視機能部による動作の先回り

以上の補助動作は、従来の研究で提案された補助動作と比較して柔軟性と効率性に長けており、また人間の意図・状態を考慮した補助動作であると言える。



## 第5章

### 3次元視線追跡システムの構築

---

## 第5章 3次元視線追跡システムの構築

人間とロボットの協調作業において、人間の意図・状態に基づくロボットの補助動作を第4章で提案したが、協調作業を行うための実験プラットフォームとして、我々の研究グループで開発したヒューマノイドロボットに、人間の視線を測定するための注視点計測システムを拡張する形で、実験プラットフォームを開発した。

### 5.1 ヒューマノイドロボットシステム

ロボットによる人間行動の学習及び人間とロボットの協調作業の実現のための実験プラットフォームとして、我々の研究グループでは図5.1に示すようなヒューマノイドロボットを開発した。我々の研究では、人間の行動の中でも手作業に焦点を当てており、ロボットによる人間の手作業の獲得及びその再現を実現することが1つの大きな目標となっている。そのため、視覚・双腕・上半身など人間の上半身を模し、ほぼ同等の機能を持つような実験プラットフォームを設計・開発した。

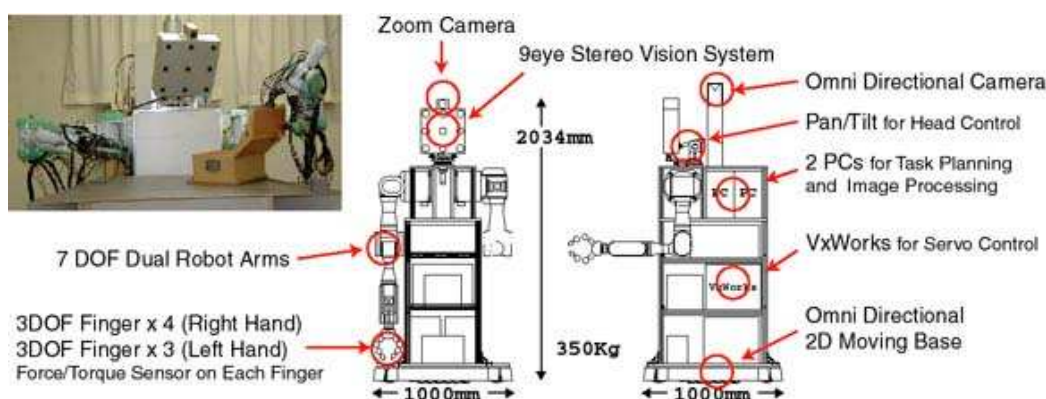


図 5.1: 実験プラットフォーム

本ロボットの主な特徴は以下の通りである。

- ビジョン

視覚として、3次元計測のための9眼ステレオビジョン装置、2次元計測のためのズーム機能付きカメラを搭載。

- アームとハンド

双腕として2つの7自由度 (DOFs) のロボットアームを搭載。右腕、左腕にそれぞれ4本、3本の指を持つ。右手は人間の手のかたちを模倣した人間型ロボットハンドであり、母指を有し、他指と対向できるかたちとなっている。指腹部や手掌に接触センサが配置してある。左手の指は全て3間接3自由度である。右手、左手のすべての指の指先にフォース/トルクセンサを持つ。

- ボディ

ロボットボディは全方向移動可能台車の上に搭載されおり、2次元平面上を自由に並進・回転移動することで視点及び腕の位置を自由に変化させることが可能。

- CORBA

分散オブジェクト技術である CORBA を通信バスとして用いたソフトウェアアーキテクチャによって、ロボット制御のために異なるマシン上の異なるプログラム同士をネットワーク上で簡単に結合でき、また外部機器 (データグローブ etc.) を容易に追加することができる。CORBA の実装系としては、TAO ORB[23] を使用。

### 5.1.1 ビジョン

ビジョンシステムは人間の視覚の代わりに用いられる。ロボットには3次元の認識を行う9眼のステレオビジョン (詳細は後述) が搭載されている。ビジョンシステムはロボットの頭部にセットしてあり、パン/チルトの回転駆動系によって、ロボットは目の前にある任意の位置に焦点を合わせることができる。

### 5.1.1.1 9眼ステレオビジョンシステム

9眼ステレオビジョンシステムによりロボットは周囲の環境をリアルタイムに認識できる。このシステムはコマツ社製 [24] であり、マルチベースライン法 [25] を使っている。システムは以下のような特徴を持っている。

- ロバストなステレオマッチング  
同時に8つのステレオペア（中心のカメラと周囲8つのカメラのペア）から視差画像を計算し、その中でもっとも信頼性の高い距離データを採用する。
- ハードウェアチップを用いたリアルタイム処理  
ステレオ計算はハードウェアチップにより行い、 $280 \times 200$  の距離画像をリアルタイムに生成する。（15[fps] から 30[fps] まで）
- 簡単なカメラ設定  
ユーザは自分の要求に合うように簡単にカメラの設定をすることができる。またステレオシステムが短い時間で高解像度の距離画像を生成できるように、焦点距離を伸ばし、外側の8つのカメラを内側に向けるようにした。測定範囲は可変である。

### 5.1.2 アームとハンド

このロボットは双腕として2つの7自由度 (DOFs) の PA10 ロボットアーム (三菱重工社製) を持つ。この自由度はロボットハンドが3次元空間の領域内を (位置・姿勢の面で) 広範囲に動くのに十分なものである。人間の手のかわりに、アームの先端に指を有するロボットハンドが取りつけてある。右腕, 左腕にそれぞれ4本, 3本の指を持つ。右手は人間の手を模倣した人間型ロボットハンドである。

#### 5.1.2.1 人間型ロボットハンド (右手)

人間の把持スキルの模倣をスムーズに行うために、右手については人間型のハンドを開発した。全体とのバランスの関係から指は環指と小指をまとめて1本とし、指の本数は全部で4本としてある。指の各関節の駆動には安川電機の指関節用アクチュエータを使用し、指先にはBLオートテック社の6軸力覚センサ (NANO センサ) を取りつけてある。

### 5.1.3 上半身

上半身は3次元空間中でロボットアームの稼動範囲およびビジョンのカバーする領域を増やすため作られている。人間が手作業をするときに、物体を様々な角度から見るために頭を動かす。それと同時に、腕が届かないところで作業をするときには上半身を曲げたりひねったりする。ロボットに同様な能力を持たせるために我々は前方向移動可能台車 [26] を搭載した。この台車によりロボットは体をひねったりフロア上を移動したりできるため、視点や作業スペースを動的に変更することができる。

## 5.2 注視点計測システムの開発

注視点計測システムでは、視線検出装置により計測された視線情報とロボットの9眼ステレオビジョンを利用することで、注視している物体や世界座標系における注視位置を抽出することができる。

### 5.2.1 EMR-8の概要

#### 5.2.1.1 EMR-8の構成

本システムでは視線検出装置として、(株)ナックイメージテクノロジー社製の眼球運動計測装置アイマークレコーダ:EMR-8(図 5.2)を用いる。EMR-8はコントローラ部とヘッドユニット部から構成されている。ヘッドユニット部にはイメージセンサと反射鏡がついており、反射鏡が眼球周辺をイメージセンサに反射し観測させた上で、角膜反射方式によってイメージセンサ観測画像における眼球の位置を観測する。

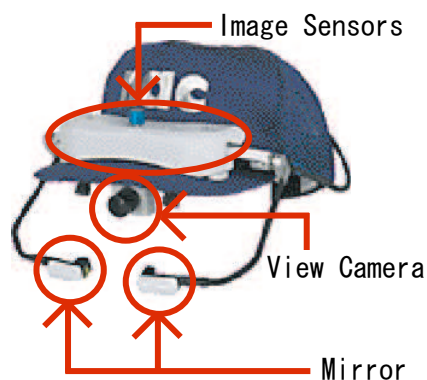


図 5.2: EMR-8

またヘッドユニットの中央には、視野カメラ (アイマークカメラ) が装備されており、リアルタイムにユーザの視野の画像を観測している。コントローラ部は、これ

らの入力を受け取り，後述のキャリブレーションによって入力情報の対応関係を求める．この対応関係に基づいて視線方向を推定し，アイマークカメラ画像にその方向が重畳表示される．つまり，EMR-8 から取得できる情報は，EMR-8 に装着されているアイマークカメラで撮影された画像座標系における 2次元の座標値である．

### 5.2.1.2 EMR-8 のキャリブレーション

イメージセンサによって眼球を観測する際，眼球・イメージセンサ・アイマークカメラの空間的な相対位置関係は未知である．そのため，キャリブレーションによって，イメージセンサ画像における  $x, y$  方向への眼球の変位量として定義される視野角とアイマークカメラの画像座標系における注視点位置の対応関係を調べ，対応式を得る．キャリブレーション後は，得られた対応式をもとに眼球観測結果から視線方向を計算することが可能となる．このキャリブレーションは，(株)ナックイメージテクノロジー社によって提供されるオペレーションである．

具体的なキャリブレーション作業について述べる．まず実環境中に壁などの 1 平面を用意し，EMR-8 ヘッドユニットのアイマークカメラの光軸が壁面に垂直になるようにユーザの頭部位置を定める．このときのアイマークカメラと壁面間の距離をキャリブレーション距離と呼び，その壁面をキャリブレーション面と呼ぶことにする．図 5.3 に示すように，アイマークカメラの画像座標系上で座標値が既知の 9 つの注視点状ターゲットを実環境中のキャリブレーション面に投影し，投影された点をユーザが注視した際の眼球の視野角を観測する．これらの座標値対から対応式を決定する．

アイマークカメラの画像座標系における座標値と注視点の関係を説明する．図 5.4 に示すように，ユーザは点  $P$  を注視したとする．キャリブレーション面においては，右眼の注視位置は  $P_R$  に，左眼の注視位置は  $P_L$  に現れる． $P_R, P_L$  とカメラ中心  $C$  を結ぶ直線とアイマークカメラ画像面との交点  $C_R, C_L$  がアイマークカメラの画像座標系における座標値である．このデータが EMR-8 から出力される 2次元のデータであり，RS232C 経由で 30Hz のサンプリング周期で PC に取り込まれる．



図 5.3: EMR-8 キャリブレーション

### 5.2.2 世界座標系における注視位置の取得手順

人間とロボットが協調作業を行うには、世界座標系における注視位置を抽出する必要がある。しかし、EMR-8 から得られるデータは、搭載されているアイマークカメラの画像座標系における座標値である。アイマークカメラをキャリブレーションすることで、アイマークカメラ座標系における3次元の注視位置を取得できるが、作業中には人間の頭部が動いてしまうため、アイマークカメラ自体の位置・姿勢が変化する。従って、アイマークカメラの位置・姿勢を知る必要がある。そこで、ロボットの9眼ステレオビジョンを利用する。EMR-8 にマーカーを装着し、ロボットがマーカーを観察することで、マーカーの位置・姿勢を獲得する。また、あらかじめマーカーとアイマークカメラの位置関係を求めておくことで、同時にアイマークカメラの位置・姿勢を獲得することができる。このようにして、EMR-8 から得られる2次元データから、世界座標系における注視位置を求めることができる。以上の処理をまとめると、図 5.5 に示すような座標変換が必要である。

1. マーカー座標 (Marker Coordinates) から世界座標 (World Coordinates) への座標変換

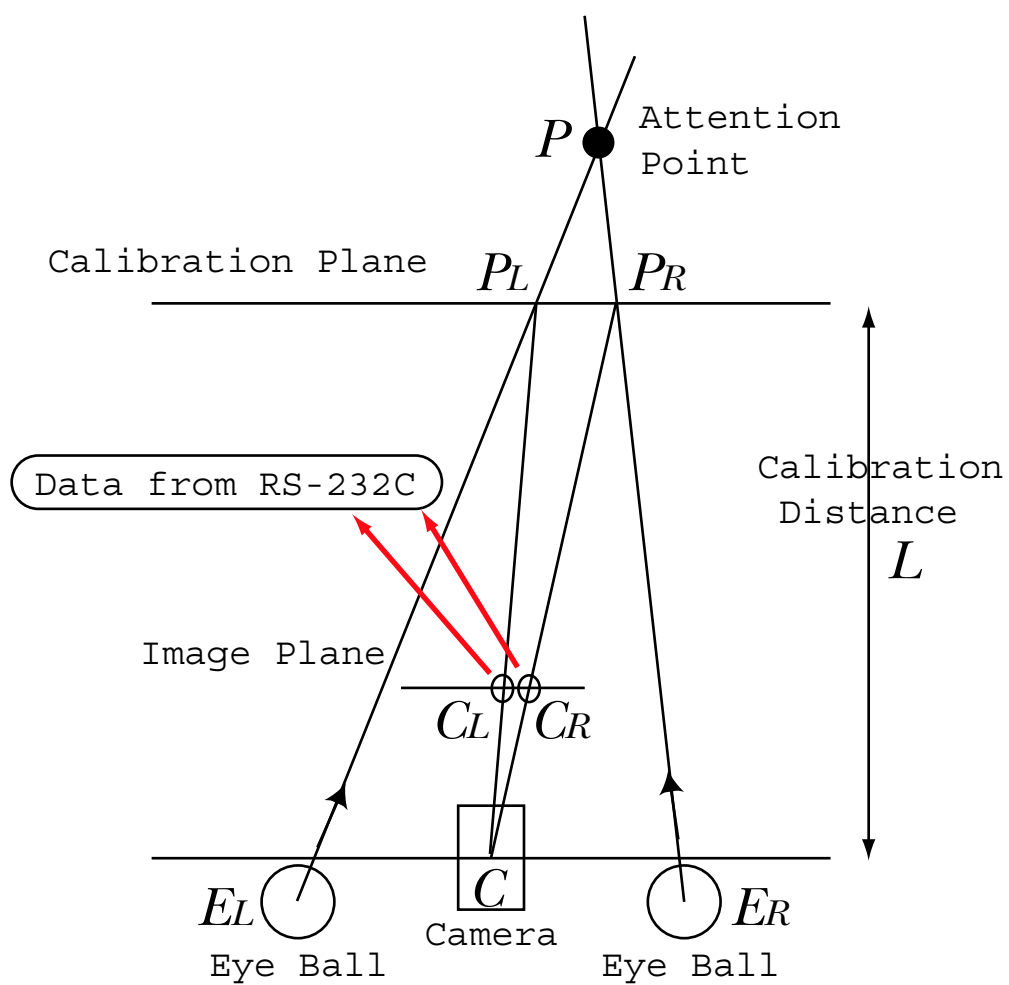


図 5.4: EMR-8 の原理



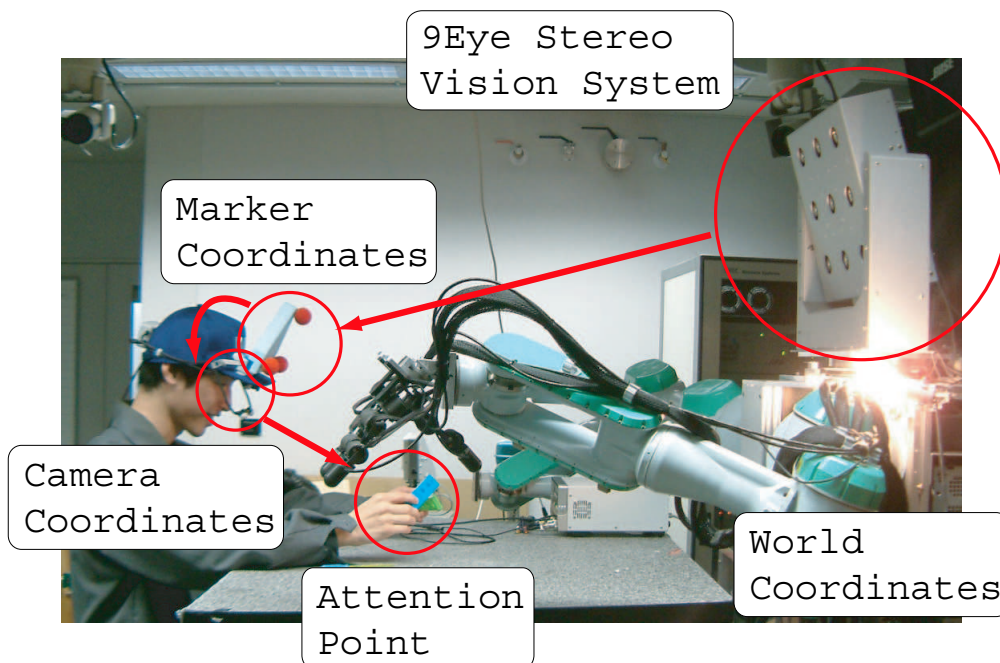


図 5.5: 必要な座標変換

2. アイマークカメラ座標 (Camera Coordinates) からマーカース座標 (Marker Coordinates) への座標変換
3. アイマークカメラ画像座標系における注視点位置 (Attention Point) からアイマークカメラ座標 (Camera Coordinates) への座標変換

これらの座標変換により，世界座標系における注視点位置  $P_W$  は次の式で表される．

$$P_W = \begin{bmatrix} \mathbf{R}_M & \mathbf{T}_M \\ \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{R}_C & \mathbf{T}_C \\ \mathbf{0} & 1 \end{bmatrix} f(\vec{V}_{C_R}, \vec{V}_{C_L}, M_C) \quad (5.1)$$

ただし， $\mathbf{0}$  は  $1 \times 3$  のゼロベクトルである．

次に，式 (5.1) の各座標変換について詳しく説明する．

### 5.2.3 マーカース座標 $\implies$ 世界座標

図 5.6 にロボット搭載の 9 眼ステレオビジョンから得られる画像を示す．

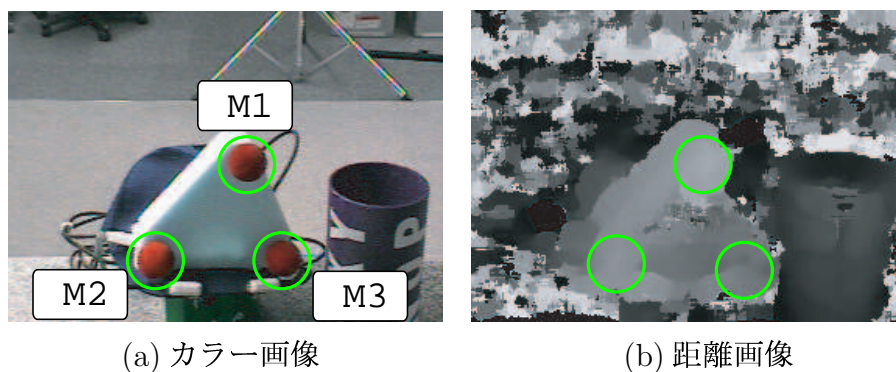


図 5.6: 9眼ステレオビジョンの画像

まず、あらかじめマーカの色を登録しておく。カラー画像 (図 5.7(a)) から、登録してあるマーカの色ヒストグラムと相関の高い領域を切り出し、これを登録してあるマーカの参考面積の大きさ範囲と対応する部分のみを切り出し、物体領域の候補画像 (図 5.7(b)) を得る。次に、ラベリング・面積ソートして登録してある面積に適合するものであれば、マーカを見つけたもの (図 5.7(c)) とし、その画像をマスクとして9眼ステレオビジョンの距離画像をマスキングすることで、世界座標系におけるのマーカの位置と姿勢を得ることができる。

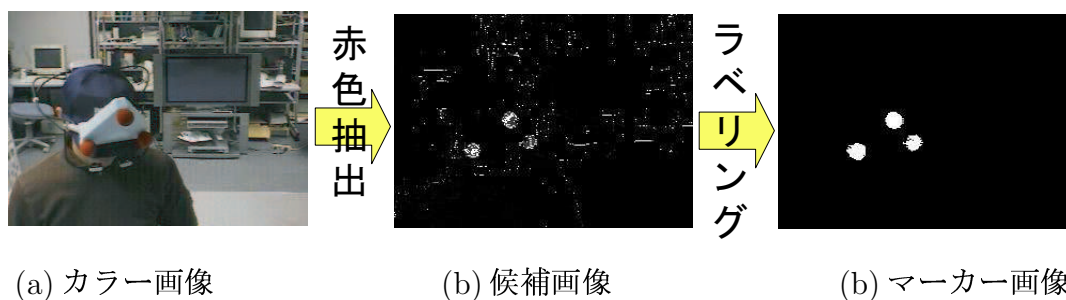


図 5.7: 画像処理の結果

マーカ座標系を次のように定義する。マーカ座標系は3つのマーカの重心  $M_G$  を原点とし、マーカが作る平面の法線ベクトルを  $Z_M$  軸とし、 $\overline{M_G M_1}$  を  $Y_M$  軸とする。そして、 $X_M$  軸は、 $Y_M$  軸と  $Z_M$  軸の両方の軸と直交する軸とする。

マーカ座標の世界座標への変換行列は回転行列  $R_M$  と平行移動ベクトル  $T_M$  からなる。世界座標系におけるマーカの座標値は既知であるので、 $R_M$  と  $T_M$  は求

まる.

マーカース座標系で表現された点  $P_M$  は, 回転と平行移動で世界座標系では点  $P_W$  で表現されるとすると, 座標変換の関係式は,

$$\begin{aligned} P_W &= R_M P_M + T_M \\ &= \begin{bmatrix} \mathbf{R}_M & \mathbf{T}_M \\ \mathbf{0} & 1 \end{bmatrix} P_M \end{aligned} \quad (5.2)$$

で表される.

#### 5.2.4 アイマークカメラ座標 $\Rightarrow$ マーカース座標

アイマークカメラ座標からマーカース座標への変換行列を求めるために, 図 5.8 に示すようなキャリブレーションを行う.

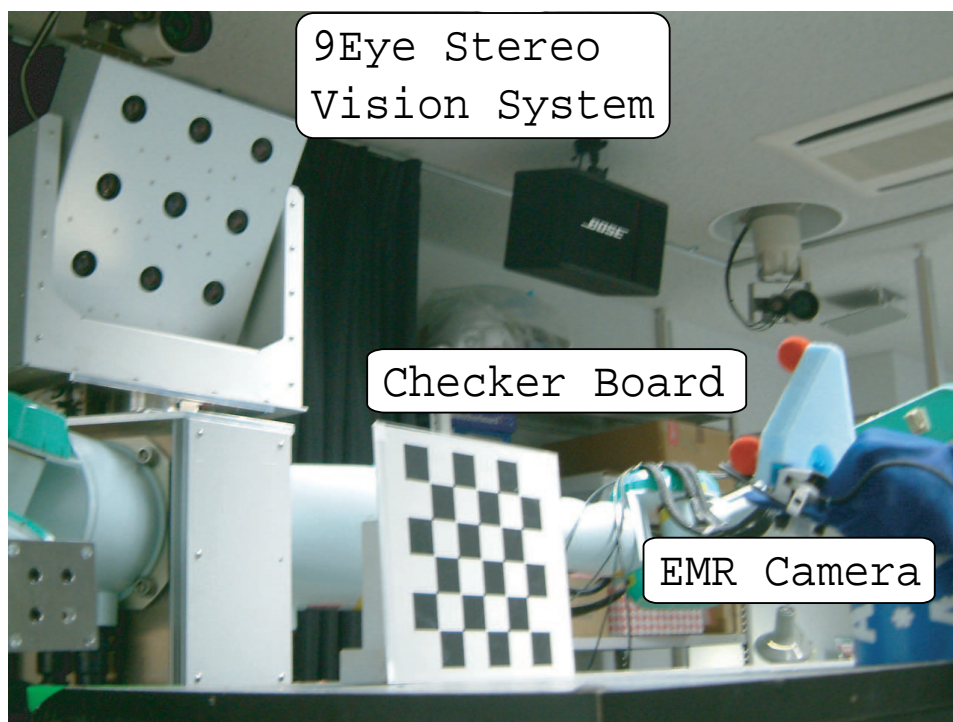
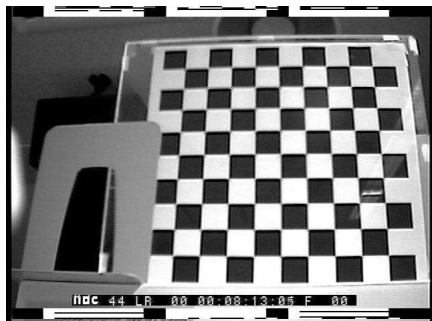
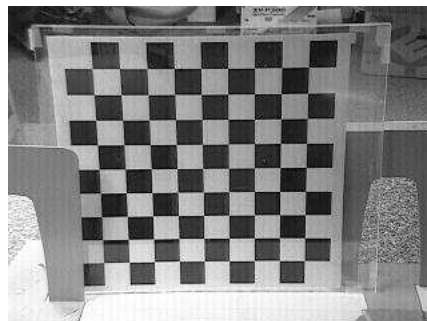


図 5.8: キャリブレーションの様子



(1) EMR-8



(2) 9眼ステレオビジョン

図 5.9: EMR-8 カメラと 9 眼ステレオビジョンの画像

ロボット搭載の 9 眼ステレオビジョンとアイマークカメラの間に、表裏に同じマーカがついているチェッカーボードを置くことで、両方のカメラを同時にキャリブレーションすることができる。図 5.9 に示すように、両方のカメラから同じチェッカーボードをキャプチャーすることができる。このキャリブレーションによって得られるパラメータと 9 眼ステレオビジョンから得られるマーカの座標値から以下の座標変換行列が得られることになる。

- マーカー座標系から世界座標系への変換行列  $M_{M \rightarrow W}$
- チェッカーボード座標系からアイマークカメラ座標系への変換行列  $M_{CB \rightarrow C}$
- 世界座標系からチェッカーボード座標系への変換行列  $M_{W \rightarrow CB}$

従って、アイマークカメラ座標系からマーカ座標系への変換行列を  $M_{C \rightarrow M}$  とすると次式が成立する。

$$\begin{aligned} I &= M_{W \rightarrow CB} M_{CB \rightarrow C} M_{C \rightarrow M} M_{M \rightarrow W} \\ M_{C \rightarrow M} &= M_{CB \rightarrow C}^{-1} M_{W \rightarrow CB}^{-1} M_{M \rightarrow W}^{-1} \end{aligned} \quad (5.3)$$

ただし、 $I$  は単位行列である。

式 (5.3) より、アイマークカメラ座標系からマーカ座標系への変換行列が求まる。アイマークカメラ座標のマーカ座標への変換行列は回転行列  $R_C$  と平行移動ベクトル  $T_C$  で表現できるので、アイマークカメラ座標系で表現された点  $P_C$  は、回

転と平行移動でマーカー座標系では点  $P_M$  で表現されるとすると、座標変換の関係式は、

$$\begin{aligned} P_M &= R_C P_C + T_C \\ &= \begin{bmatrix} \mathbf{R}_C & \mathbf{T}_C \\ \mathbf{0} & 1 \end{bmatrix} P_C \end{aligned} \quad (5.4)$$

で表される。

### 5.2.5 注視位置 $\Rightarrow$ アイマークカメラ座標

EMR-8 から得られる 2次元データからアイマークカメラ座標系における座標値を取得する。

#### 5.2.5.1 アイマークカメラのキャリブレーション

2次元画像から3次元情報を取り出すにはカメラキャリブレーションが必要である。カメラキャリブレーションに関する研究は盛んに行われており、数々の手法が提案されている。アイマークカメラのキャリブレーションには、比較的新しい手法である Zhang[27] の手法を用いる。Zhang の手法は、複数の平面を使用するが各平面ごとに世界座標系を設定するため、ターゲットを自由に設定でき、比較的容易にキャリブレーションを行うことができるというメリットがある。

このキャリブレーションによって得られた内部パラメータ  $\alpha (= s_x \frac{f}{d_x})$ ,  $\beta (= \frac{f}{d_y})$ ,  $u_0$ ,  $v_0$  ( $f$ : 焦点距離,  $d_x, d_y$ : 画素サイズ,  $s_x$ : 画素の縦横比) を用いて、アイマークカメラ画像座標系における 2次元座標値  $C_R(X_R, Y_R)$ ,  $C_L(X_L, Y_L)$  はアイマークカメラ座標系では次のように 3次元ベクトルとして表現される。

$$\vec{V}_{C_R} = \left( \frac{X_R - u_0}{\alpha}, \frac{Y_R - v_0}{\beta}, 1 \right) \quad (5.5)$$

$$\vec{V}_{C_L} = \left( \frac{X_L - u_0}{\alpha}, \frac{Y_L - v_0}{\beta}, 1 \right) \quad (5.6)$$

## 5.2.5.2 アイマークカメラ座標系における視線ベクトルの表現

アイマークカメラ座標系における視線ベクトルは注視点の定義によって異なる。注視点の定義は以下の2つが考えられる。

**定義 1** 両眼の視線ベクトルの交点を注視点とする。

アイマークカメラ座標系における注視点位置は、両眼の画像座標の位置により、図 5.10 に示すような関係によって求まる。この関係について詳しく説明する。

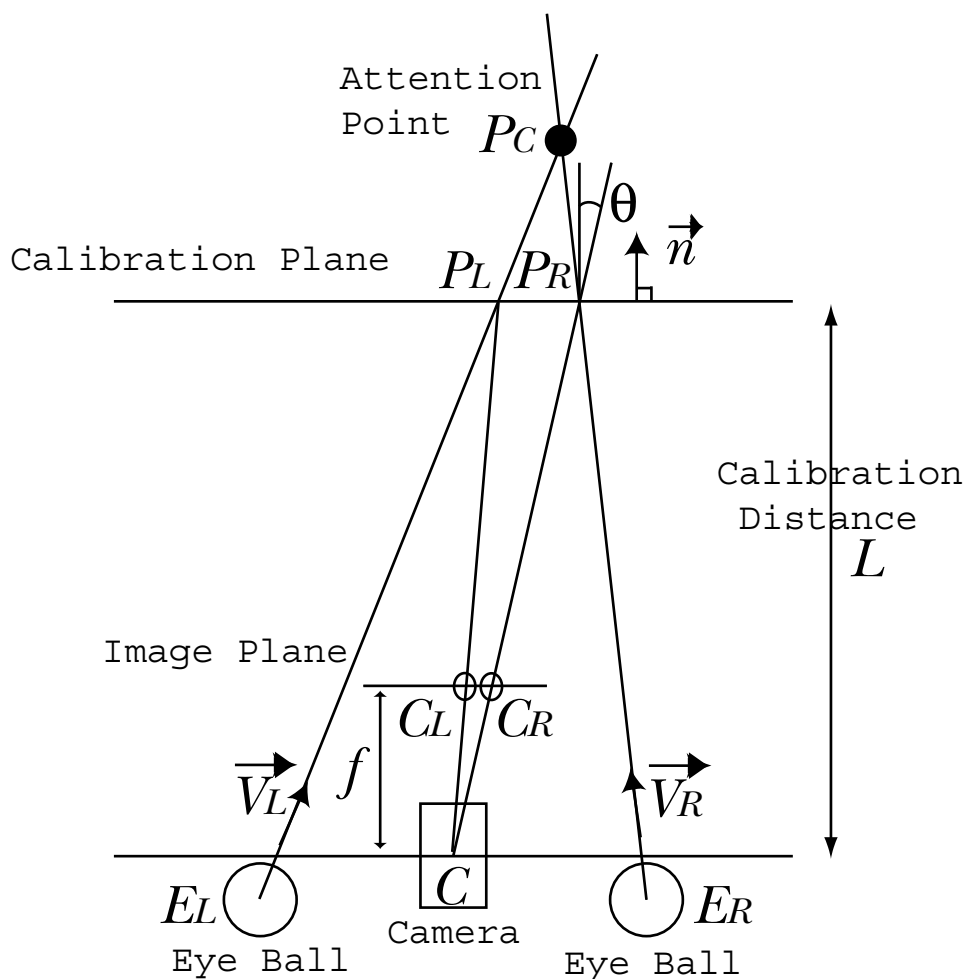


図 5.10: カメラ座標系における注視点

アイマークレコーダのキャリブレーションの距離を  $L$ 、アイマークカメラ座

標系における両眼の位置を  $E_R, E_L$ , アイマークレコーダから得られる情報を  $C_R, C_L$ , 両眼の視線単位ベクトルを  $\vec{V}_R, \vec{V}_L$  とおく. 定義 1 より, 注視点  $P_C$  は  $\vec{V}_R$  と  $\vec{V}_L$  の交点である.

従って, 既知である  $L, E_R, E_L, C_R, C_L$  から  $\vec{V}_R, \vec{V}_L$  を表すことができれば注視点が求まることになる (ここで,  $E_R, E_L$  は既知の情報として扱っているが,  $E_R, E_L$  の推定方法については後述).

$\vec{V}_R, \vec{V}_L$  の算出方法は同じなので, ここでは  $\vec{V}_R$  についてのみ説明する.

$\vec{V}_R$  は単位ベクトルなので,

$$\vec{V}_R = \frac{\overrightarrow{E_R P_R}}{|\overrightarrow{E_R P_R}|} \quad (5.7)$$

と表される. また,  $\overrightarrow{E_R P_R}$  は図 5.10 から,

$$\overrightarrow{E_R P_R} = \overrightarrow{C P_R} - \overrightarrow{C E_R} = k \overrightarrow{C C_R} - \overrightarrow{C E_R} \quad (5.8)$$

と表現できる.  $\overrightarrow{C E_R}$  は既知である.  $\overrightarrow{C C_R}$  を単位ベクトルとすると,

$$k = |\overrightarrow{C P_R}| \quad (5.9)$$

となる. ここで, アイマークレコーダのキャリブレーション平面の単位法線ベクトルを  $\vec{n}$  とする.  $\vec{n}$  はアイマークカメラの光軸方向と一致するので既知である. ここで,  $\overrightarrow{C P_R}$  と  $\vec{n}$  のなす角を  $\theta$  とおくと,  $|\overrightarrow{C P_R}| \cos \theta = L$  の関係式と式 (5.9) より,

$$k = \frac{L}{\cos \theta} \quad (5.10)$$

$$\text{ただし, } \theta = \cos^{-1} \frac{\vec{n} \cdot \overrightarrow{C C_R}}{|\vec{n}| |\overrightarrow{C C_R}|}$$

以上のことから,  $\vec{V}_R$  が求まる.  $\vec{V}_L$  も同様である. 従って,  $\vec{V}_R$  と  $\vec{V}_L$  の交点を求めることでアイマークカメラ座標系における注視点位置  $P_C$  が求まる.

以上のことから,  $P_C$  は次の関数式によって表すことができる.

$$P_C = f(C_R, C_L, E_R, E_L) \quad (5.11)$$

次に、既知の情報としていたアイマークカメラ座標系における両眼の座標値  $E_R$ ,  $E_L$  の推定方法を説明する。

両眼の位置はアイマークレコーダのキャリブレーションが完了した後、世界座標系において既知の座標である数点を注視することで求める。

説明を簡単にするために、図 5.11 に示すように世界座標系における既知の座標点  $P_1$ ,  $P_2$  の 2 点を注視したとする。また、右眼に関してのみ説明を行う。左眼に関しても同じ手法で求めることができる。

図 5.11 から、 $\overrightarrow{P_1P_{R1}}$ ,  $\overrightarrow{P_2P_{R2}}$  の交点が眼球の位置であると仮定できる。  $P_1$ ,  $P_2$  に関しては座標変換を行うことにより、アイマークカメラ座標系での座標が求まる。また、キャリブレーションの距離  $L$  とアイマークレコーダから得られる情報  $C_{R1}$ ,  $C_{R2}$  から、 $P_{R1}$ ,  $P_{R2}$  が求まるので、 $\overrightarrow{P_1P_{R1}}$ ,  $\overrightarrow{P_2P_{R2}}$  が求まることになる。このベクトルの交点を求めることで、アイマークカメラ座標系における眼球の位置を求めることができる。

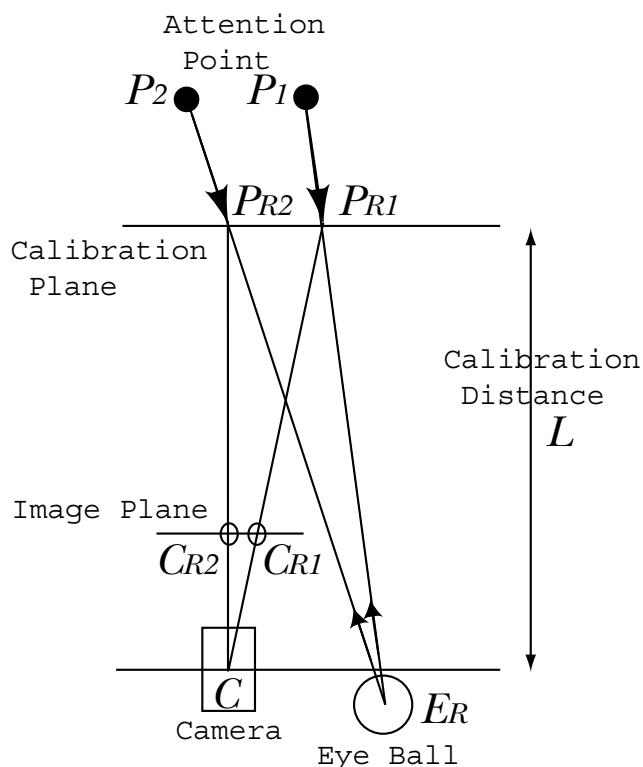


図 5.11: カメラ座標系における両眼の位置の推定



**定義 2** 視線ベクトルと環境物体との交点を注視点とする.

人間は環境中の何らかの物体を見ている. 人間は像が存在していない空中などには焦点を合わせることはできない. 従って, 定義2のように定義することができる. ここでの視線ベクトルは図 5.10 より, 以下のように定義できる.

$$\begin{aligned}
 \overrightarrow{CP_C} &= \overrightarrow{CE_L} + \overrightarrow{E_LP_C} \\
 &= \overrightarrow{CE_L} + k_L \overrightarrow{E_LP_L} \\
 &= \overrightarrow{CE_L} + k_L (\overrightarrow{E_L C} + \overrightarrow{CP_L}) \\
 &= (1 - k_L) \overrightarrow{CE_L} + k_L \overrightarrow{CP_L} \\
 &= (1 - k_L) \overrightarrow{CE_L} + k'_L \overrightarrow{CC_L}
 \end{aligned} \tag{5.12}$$

同様に,

$$\overrightarrow{CP_C} = (1 - k_R) \overrightarrow{CE_R} + k'_R \overrightarrow{CC_R} \tag{5.13}$$

式 (5.12), (5.13) から,

$$\overrightarrow{CP_C} = \frac{(1 - k_L) \overrightarrow{CE_L} + (1 - k_R) \overrightarrow{CE_R} + k'_L \overrightarrow{CC_L} + k'_R \overrightarrow{CC_R}}{2} \tag{5.14}$$

作業状況を考えると, キャリブレーション距離と作業中の視距離 (眼球と注視物体までの距離) はほぼ等しいので,  $k_L, k_R \simeq 1$  と近似できる. また  $\overrightarrow{CE_L}$  と  $\overrightarrow{CE_R}$  のベクトルの大きさ・向きを考慮すると, 式 (5.14) の第 1 項と第 2 項は除くことができる. さらに,  $k'_L \simeq k'_R$  と近似できる. 従って, 式 (5.14) は次のように近似できる.

$$\overrightarrow{CP_C} \simeq k (\overrightarrow{CC_L} + \overrightarrow{CC_R}) \tag{5.15}$$

このように, 視線ベクトルを  $\overrightarrow{CC_R}, \overrightarrow{CC_L}$  の和で表現し, この視線ベクトルと環境物体との交点を注視点とする. 環境物体の位置・姿勢はロボットビジョンによって取得できる. 環境物体の位置・姿勢の取得については 5.3.1 で説明する.

本研究では, 定義 2 を採用した. その理由を以下に示す.

1. 定義1において、両眼の視線ベクトルは3次元空間中の2直線となるため、一般には交差せず、2直線の交点として解を得ることはできず、両眼の視線直線の方程式について最小二乗法的に解を求める必要があり、計算誤差が生じてしまう。
2. 定義1において、両眼の座標値を推定する際に生じる誤差が大きいと考えられる。1.と同様の理由のため生じる計算誤差と固視微動による誤差により正確に両眼の位置を推定するのは困難である。
3. ヒューマノイドロボットには、9眼ステレオビジョンシステムが搭載されており、環境物体の位置・姿勢を容易に取得できる。

以上のような理由のため、本システムでは、視線ベクトルと環境物体の交点を注視点と定義する。アイマークカメラ座標系における視線ベクトルを $\overrightarrow{CC_R}$ と $\overrightarrow{CC_L}$ で表現しているが、これらのベクトルは、式(5.5)と式(5.6)における $\overrightarrow{V_{C_R}}$ と $\overrightarrow{V_{C_L}}$ と同等である。ここで、アイマークカメラ座標系における環境物体の位置・姿勢を $M_C$ とすると、注視位置 $P_C$ は次の関数式で表現できる。

$$P_C = f(\overrightarrow{V_{C_R}}, \overrightarrow{V_{C_L}}, M_C) \quad (5.16)$$

### 5.2.6 世界座標系における注視位置の取得

5.2.3, 5.2.4, 5.2.5 で説明した座標変換を行うことで、式5.1により世界座標系における注視位置を取得することができる。EMR-8のデータとロボットビジョンを用いた処理の流れを図5.12に示す。人間の注視点をロボットに教示できるようになり、視線を利用した人間とロボットの協調作業を行うことができる。

## 5.3 3次元視線追跡システム

ロボットシステムと注視点計測システムにより、仮想空間で人間の注視点を可視化できる実時間3次元視線追跡システムを構築した。

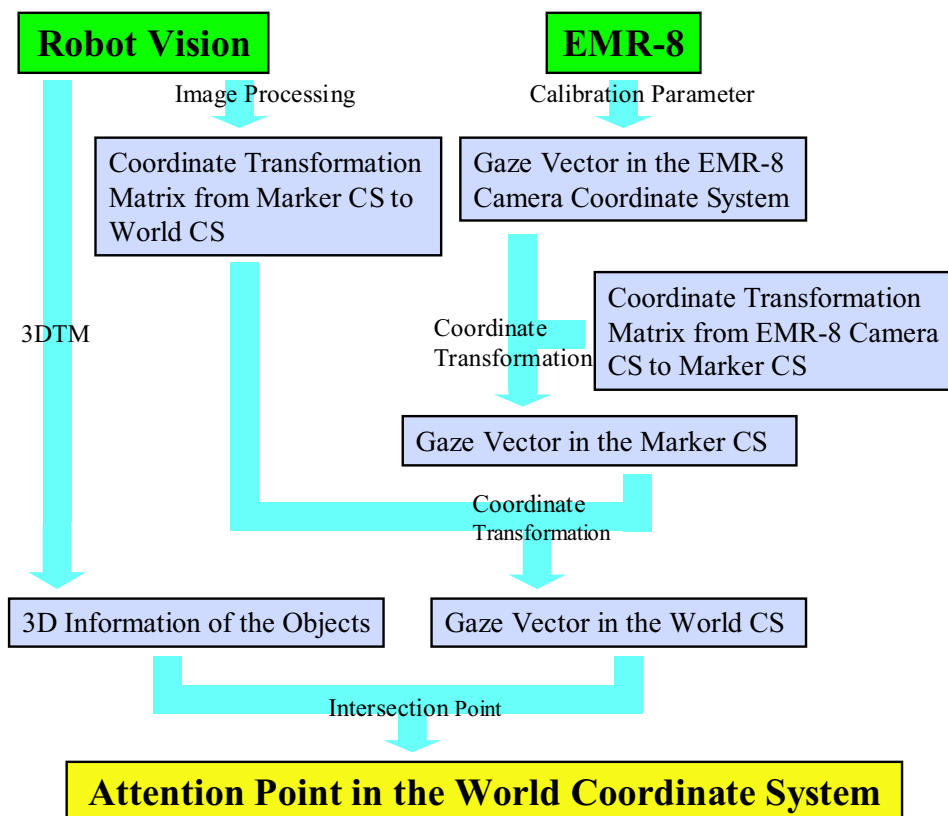


図 5.12: 注視点の獲得過程

### 5.3.1 環境物体の認識

環境物体の認識には3次元テンプレートマッチング (3D Template Matching:3DTM)[28]法を使用している。3DTMとは、3次元距離データ中において、あらかじめ対象のおおまかな位置と3次元幾何モデルが既知であるという前提の下で、対象物の正確な位置と姿勢を推定する手法である。

3DTMでは、対象物体の3次元モデルを、レンジデータなどによって得られた対象物体の3次元データ中に投影し、モデルの原点とそれに最も近い周囲の3次元点との重みつき距離の和の平均を求める。この和を指標として、これが最小となる状態に収束するようにモデルの位置・姿勢を変化させていく。収束地点でのモデルの位置・姿勢が推定値となる。

ロボットは、次の手順で環境物体を認識する。

1. 物体表面色をモデルとして持っておき、色画像上でその領域をセグメンテーションする。
2. 物体は安定姿勢にあるとの仮定をおくと、セグメンテーションにより物体のおおまかな位置・姿勢が分かるため、それを初期位置とし、対応する距離画像上でセグメンテーションされた物体領域と物体の3次元幾何モデルを使用して3DTMをかけ、位置と姿勢の正確な推定を行う(図5.13)。

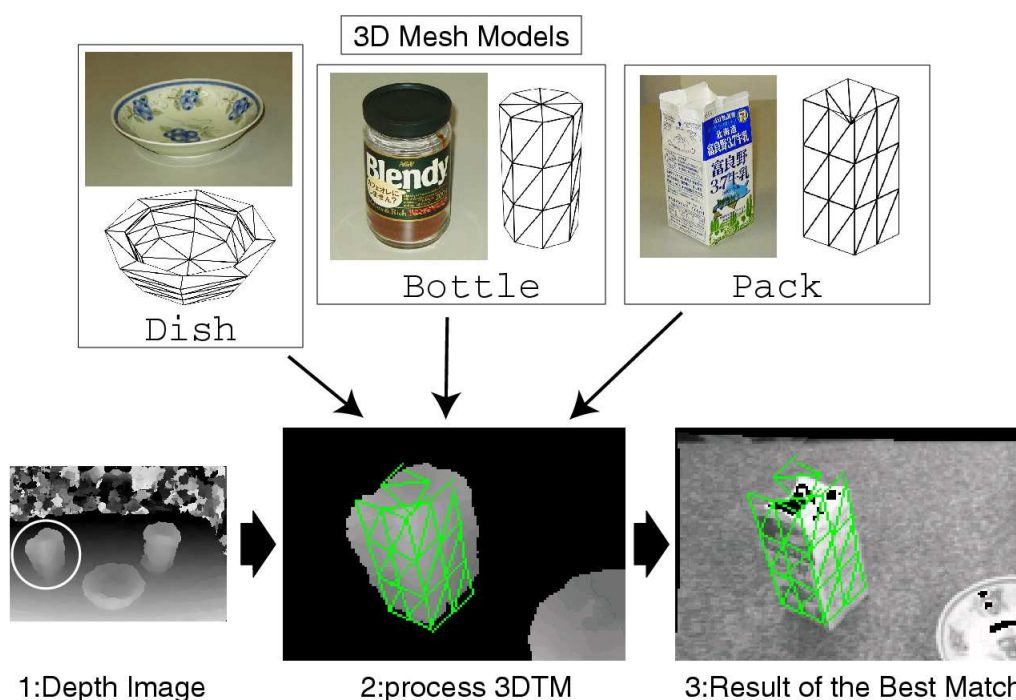


図 5.13: 3DTM による物体の認識

### 5.3.2 注視点の可視化

5.3.1で説明した手法により、環境物体の位置・姿勢を認識することができる。従って、環境物体と視線ベクトルの交点を求めることで注視点の推定が可能となる。人間が知覚困難である注視点を実空間内で可視化することができる3次元視線追跡システムを構築することにより、人間が対象物体をどの位置・角度から注視しているかなど、人間の作業状況を詳しく知ることができる。

図 5.14 に示すように，人間の注視点だけでなく，人間がどの位置からどのような姿勢で物体を注視しているかを取得することができる．

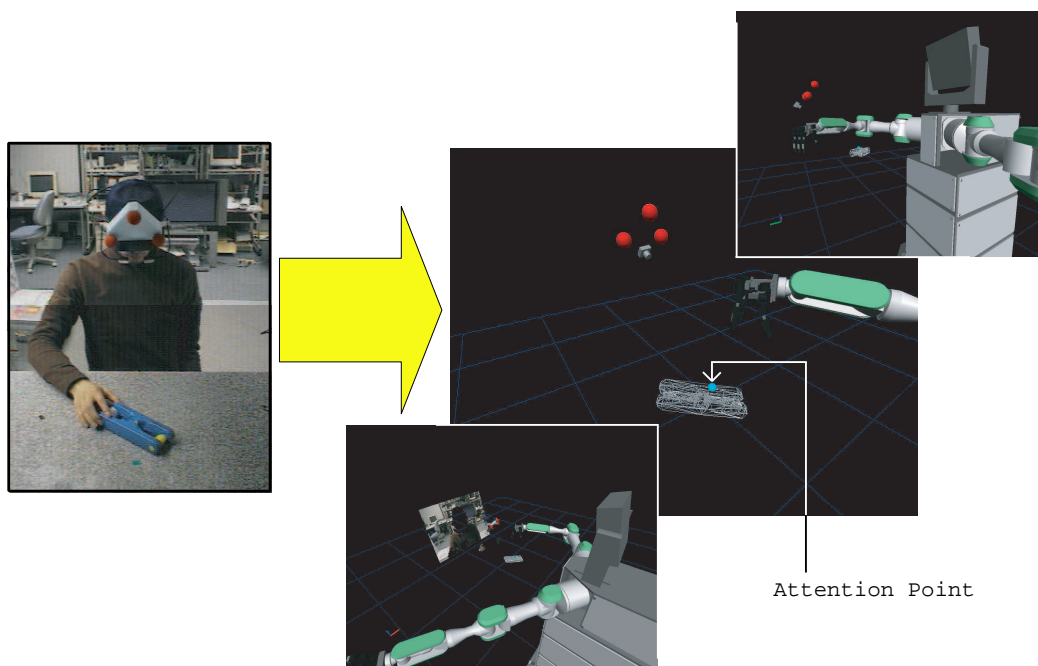


図 5.14: 仮想空間における注視点の可視化

## 5.4 システムの評価

構築した 3 次元視線追跡システムでは，約 13[fps] で仮想空間において，実時間で人間の注視点を可視化することができる．測定誤差は，約 50[cm] の距離にある物体を注視する場合に，最大 1[cm] 程度である．誤差が生じる様々な要因の中で，最大誤差になると考えられる要因は，マーカー抽出部の処理における誤差である．マーカーの位置・姿勢により人間の視線のベクトルの方向が決定されるためである．

マーカーの位置を固定した状況で，EMR-8 から得られるデータを一定値にした際のマーカーと注視点の統計を求めた．結果を表 5.1 に示す．なお，データ数は 50 個であり，視線ベクトルと卓上の交点を注視点とした解析した結果である．

表 5.1: 統計結果

Marker0	x	y	z
Average	-34.382328	1168.698608	1529.547363
Variance	0.257680	2.310269	1.054992
Deviation	0.507622	1.519957	1.027128
Marker1	x	y	z
Average	-85.325859	1239.429565	1436.716431
Variance	2.013706	11.860677	4.677338
Deviation	1.419051	3.443933	2.162715
Marker2	x	y	z
Average	43.102810	1222.900146	1425.289062
Variance	1.244638	9.972414	3.386955
Deviation	1.115633	3.157913	1.840368
Attention Point	x	y	z
Average	-14.608809	993.031494	1230.000000
Variance	34.296329	87.034462	0.000000
Deviation	5.856307	9.329226	0.000000

表 5.1 より、いずれについても  $y$  の分散が大きいことがわかる。これにより、奥行き座標の誤差が生じやすい傾向にあることがわかる。奥行き情報は、9眼ステレオビジョンにより得られる情報であり、ステレオ計算時の誤差が最終的に得られる注視点の位置に影響する。マーカの抽出段階の 3[mm] 程度の誤差が、注視点においては 1[cm] 程度の誤差となってしまう。

本システムでは、搭載カメラの視野角の制限のため、卓上の環境認識と人間の観察(マーカの抽出)を同時に行うことはできない。そこで、ロボットは、パン/チルトの回転駆動系により駆動するビジョンシステムを回転させることで、卓上の環境認識と人間の観察を交互に行うことで、人間の注視点を獲得する。処理の手順は以下のようになる。

1. ロボットは卓上の環境を認識する。つまり、卓上の物体の位置・姿勢を把握する。
2. 環境認識後、人間観察に処理を移す。
  - (a) カラー画像からマーカ色を抽出し、ラベリングを行うことで、マーカ画像を得る。
  - (b) マーカ画像に対応する距離データから、マーカの座標値を取得する。
  - (c) 取得した座標値から世界座標系とマーカ座標系の座標変換行列を求める。
  - (d) EMR データと座標変換行列から、視線ベクトルを求める。
3. 環境情報と視線ベクトルから、人間の注視点を求める。

このような処理を繰り返すことで、ロボットは人間の視線情報を取得することができ、視線に基づく補助を行うことが可能となる。

## 第6章

### 実装および実験結果

---



## 第6章 実装および実験結果

本章では，第5章で説明したシステムを用いて，第4章で提案した人間とロボットの協調作業を実装した．その結果について述べる．

### 6.1 協調作業の流れ

協調作業を行うためのロボットの処理の流れは以下のようにになっている．

1. 環境物体の位置・姿勢の認識
2. 人間の注視点の獲得
3. 視線の解析による人間の意図・状態の推定
4. 人間の意図・状態に基づく補助動作
5. 1.へ

3.の解析の結果によって，ロボットが行う補助動作は決定される．我々が提案した補助動作は以下の3つである．

- 注視時間を利用した視線によるポインティング
- “迷い”状態におけるタスクモデルと機能部情報に基づく適切な物体の提示
- 注視機能部による動作の先回り

視線を利用した協調作業のフローチャートを図6.1と図6.2に示す．視線情報を用いることで，従来の研究で提案された補助の問題点を解決し，人間の意図・状態を考慮したロボットによる補助が可能となった．

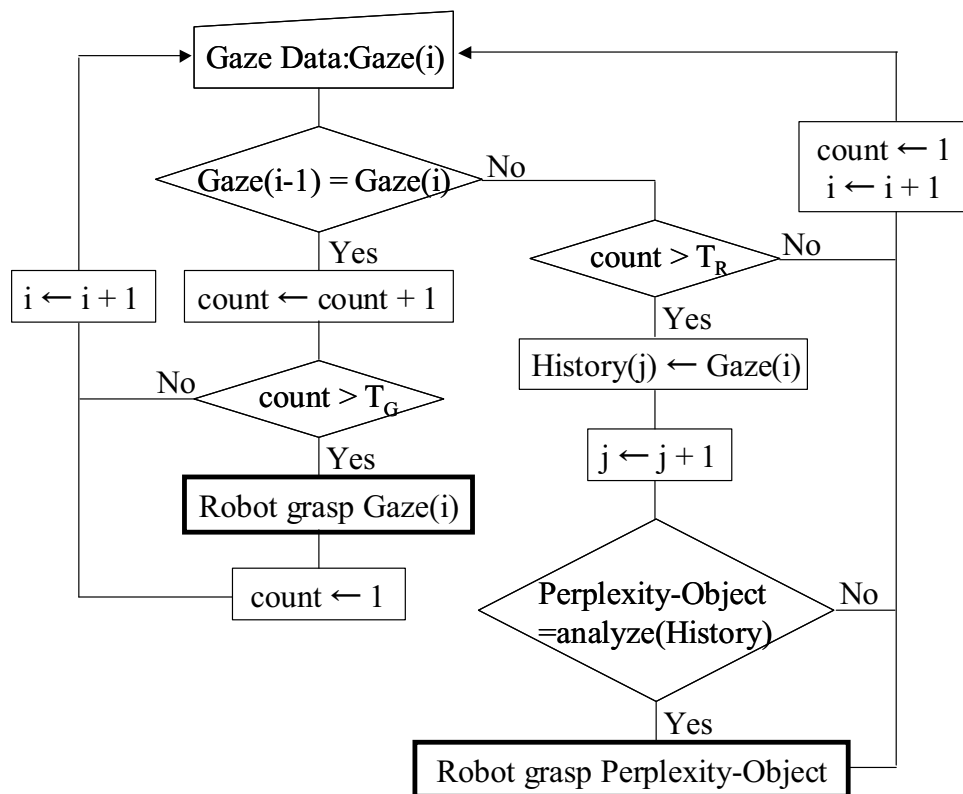


図 6.1: 選択状態の補助動作のフローチャート

**Gaze(i)**  $i$  番目の処理で得られる注視物体

$T_G$  探索段階と把持決定段階を分離する停留時間の閾値

$T_R$  人間が注視物体を認識するために必要な停留時間の閾値

**History(j)**  $j$  番目に登録されている視線履歴情報

**analyze()** 視線履歴を解析し, “迷い” 状態が検出されれば, 次の作業に適切な物体 “Perplexity-Object” を返す

**Robot Grasp Gaze(i)** ロボットの補助動作. 注視物体を把持

**Robot Grasp Perplexity-Object** ロボットの補助動作. “Perplexity-Object” を把持

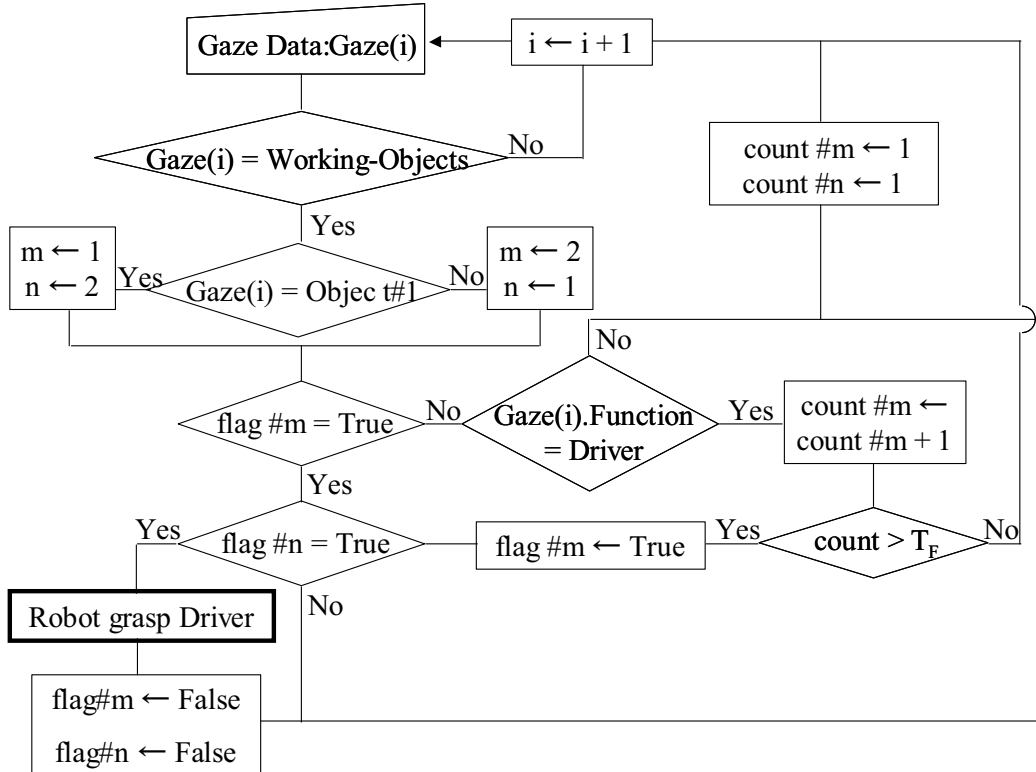


図 6.2: 作業状態の補助動作のフローチャート

**Working-Objects** 人間が組み付けを行う際に把持している 2 つの物体

**Object #1** “Working-Objects” のうち、右手が把持している物体

**flag #m** “Object #m” の “Driver” を必要とする機能部への注視フラグ

**Gaze(i).Function** i 番目の処理で得られる注視物体の機能部のうち注視している機能部

$T_F$  人間が組み付ける機能部を決定するのに要する停留時間の閾値

**Robot Grasp Driver** ロボットの補助動作. “Driver” を把持

それぞれの補助動作を実装した。また、ロボットが把持した作業物体は、把持した後、組み付け作業まで行うことが理想であるが、ここでは、把持した物体を人間に手渡すことを補助動作としている。

使用した作業物体は、市販の LEGO を 2 倍の大きさに拡大したものである。ロボットは、市販の LEGO を把持することは困難であり、把持可能な大きさに拡大する必要があったためである。また、総物体数は 5 つであり、物体数は十分であるとは言えないが、今後、物体数が増えても同様の補助が行える。

## 6.2 実装 I

「視線によるポインティング」の補助を実装した。人間が一定時間以上の注視を行うとロボットが注視物体を把持し、人間に手渡す。図 6.3 に結果を示す。

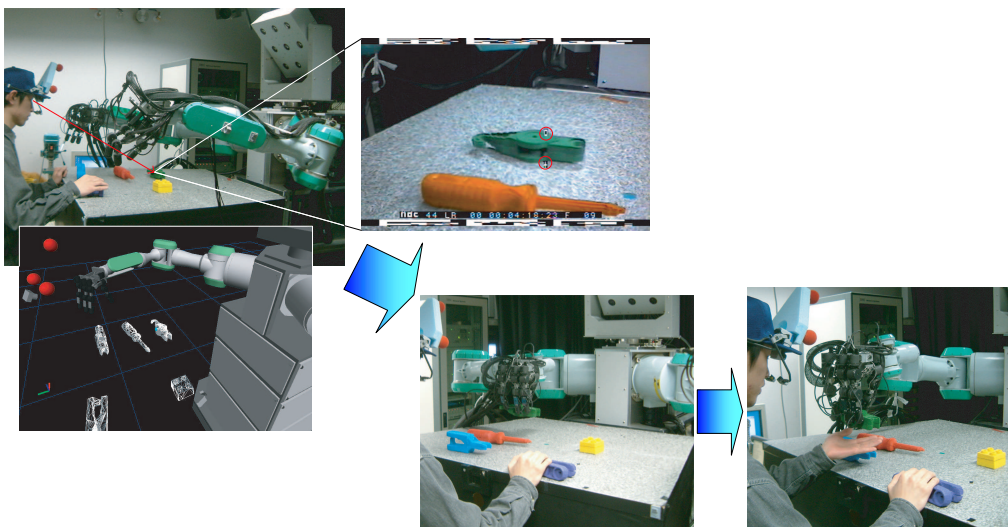


図 6.3: 注視物体の把持の実験結果

## 6.3 実装 II

「迷い状態における補助」を実装した。図 6.4 に結果を示す。「Gaze Record」に示すような履歴情報を取得した場合に、履歴に登録されている“BlockA(Green)”, “BlockB(Light Blue)”, “BlockC(Yellow)”, “BlockD(Green)”の機能部の中で、組み立て途中の物体“BlockA(Purple)”の機能部が保持している機能部“Axle”と対応する機能部を保持する物体に投票を行う。このように、投票を繰り返し、一定時間以上経過しても組み付けが行われていないので、その時点で最大得票数を獲得している“BlockB”を把持し、人間に手渡すことができる。

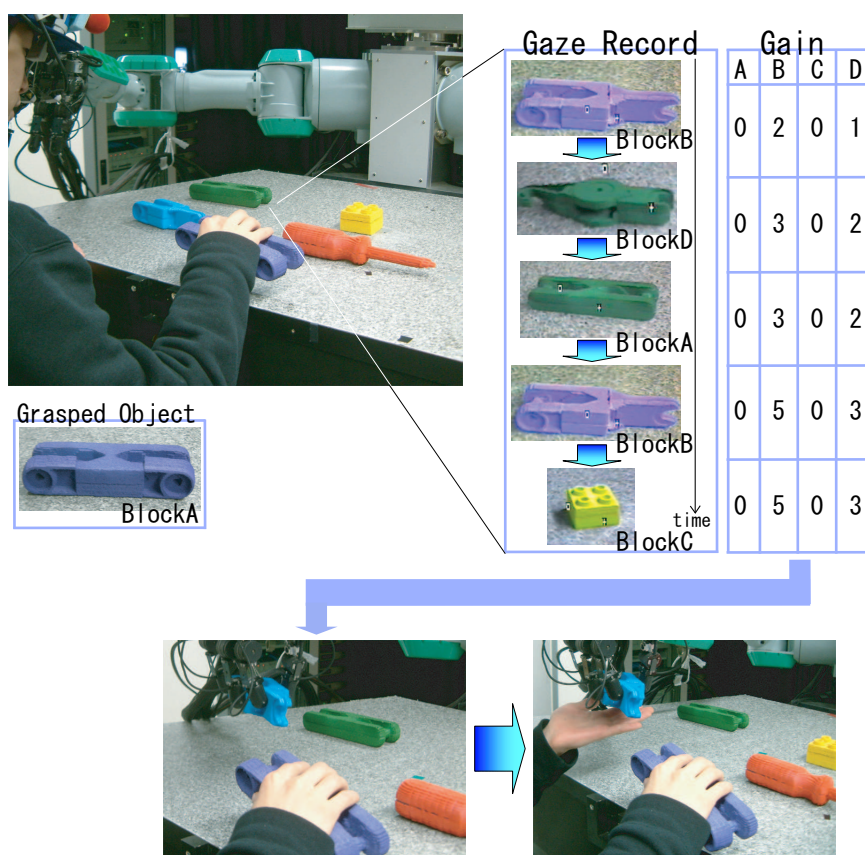


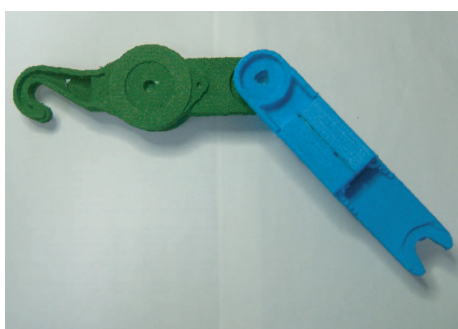
図 6.4: 注視物体の把持の実験結果

## 6.4 実装 III

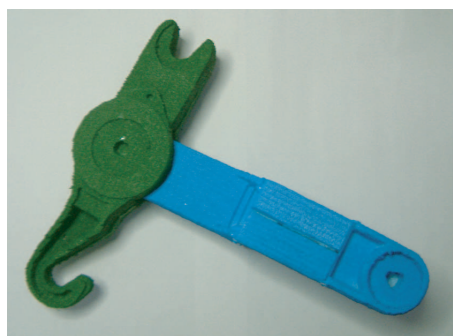
「注視機能部による動作の先回り」を実装した。作業を行っている人間が、作業物体のどの機能部を注視しているかということを出し、注視機能部より実行されるタスクモデルを推測し、そのタスクモデルの次のタスクモデルが一意に決定している場合、人間が現在のタスクモデルを実行している間に、ロボットが次のタスクモデルの準備を行う。

実験には、“Shovel”と“BlockB”の組み付けを行った。“Shovel”は“Bearing”と“Open-Axle”の機能部を持ち、“BlockB”は“Bearing”と“Axle”の機能部を持つ。この2物体を使用した組み付け作業は、次の2パターンが考えられる。

1. Shovel:Bearing  $\iff$  BlockB:Axle (図 6.5(a))
2. Shovel:Open-Axle  $\iff$  BlockB:Bearing (図 6.5(b))



(a) Bearing and Axle



(b) Open-Axle and Bearing

図 6.5: Shovel と BlockB の組み付けパターン

1. の組み付け作業は次の作業に“Driver”を必要とし、2. の組み付け作業は必要としない。このように、同じ物体の組み付け作業においても次の作業は異なった作業となる。そこで、人間がどの機能部を注視しているかを調べることで、効率的で柔軟な補助が可能となる。

図 6.6 に結果を示す。上段は、“Bearing”と“Open-Axle”の組み付けのため、次の作業として“Driver”は不要であり、ロボットは何の補助も行わない。下段は、“Bearing”と“Axle”の組み付けであり、次の作業に“Driver”が必要なので、人間が組み付けている間に、ロボットは“Driver”を把持し、人間に手渡している。このように、ロボットは、人間が注視している機能部に基づいた動作の先回りを行うことができる。

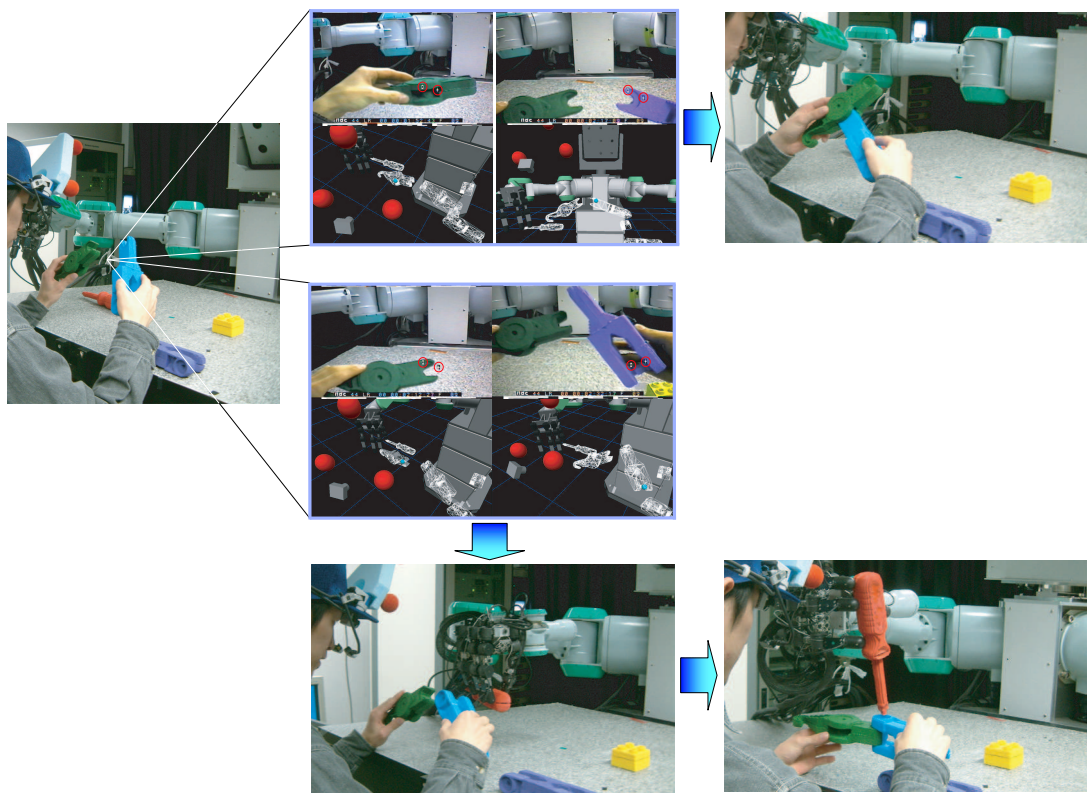


図 6.6: 注視機能部に基づく補助の実験結果

## 第7章

### 結論

---



## 第7章 結論

近年になり研究・開発の著しい人間型ロボットは，従来の産業用ロボットとは異なり，我々一般人と共存し，人間を直接支援することを目的としている．適切な支援を行うには，環境の状況や人間の状態を正確に把握する必要がある．

本研究では，人間の意図を内包した生理的情報である視線の運動を利用して，人間の意図や状態を反映したロボットによる協調動作を実現した．具体的には，LEGOの組み立て作業を題材とし，直近の視線の履歴から人間の意図や状態を推定する手法と，これに基づくロボットの補助動作決定法を提案した．さらに，実時間3次元視線追跡システムを構築し，人間型ロボットとの協調組み立て作業を実現した．

### 7.1 まとめ

本研究で行ったことは次の通りである．

1. 視線の解析による人間の意図・状態の抽出
2. 人間の意図・状態に基づくロボットの補助動作の提案
3. 実験プラットフォームである3次元視線追跡システムの構築
4. 提案したロボットの補助動作の実装

**視線の解析による人間の意図・状態の抽出** LEGOの組み立て作業を行っている人間の視線を解析した．解析の結果，以下のようなロボットの補助動作の生成に有効な特徴を発見することができた．

- 注視による探索状態と把持状態の分離
- 視線の履歴による“迷い”状態の予測
- 注視機能部による組み付け意図の抽出

**人間の意図・状態に基づくロボットの補助動作の提案** 視線の解析結果から、人間とロボットの協調作業において、視線をどのように利用するかを提案した。提案したロボットの補助動作は以下の3つである

- 注視時間を利用した視線によるポインティング
- “迷い”状態におけるタスクモデルと機能部情報に基づく適切な物体の提示
- 注視機能部による動作の先回り

**3次元視線追跡システムの構築** 提案した補助動作を行うために、実時間3次元視線追跡システムを構築した。本システムは、EMR-8のデータとロボットビジョンを利用することで、ロボットが人間の視線情報を取得できる。また、仮想空間における注視点の可視化ができる。

**提案したロボットの補助動作の実装** 構築したシステムを用いて、提案した補助動作を実装し、ロボットとの協調組み立て作業を実現した。ロボットは、作業中の人間の視線情報を取得し、解析することで適切な補助を行うことができる。

## 7.2 今後の課題

- 構築したシステムで獲得する注視点の精度を向上させる必要がある。マーカー抽出するときの誤差が最終的に得られる注視点の位置の誤差に大きく影響する。照明環境を工夫することで改善されると期待できる。
- 停留時間や物体の注視順番などには、当然のことながら個人差がある。今回の実装ではそれらのパラメータは固定値として扱っているが、より柔軟性を高めるためには、個人ごとにパラメータを変化させる必要がある。ロボットが作業している人間をしばらくの間観察し、リアルタイムで解析を行った上で、パラメータを決定し、それに基づいた協調作業を実行することができれば、作業者に応じた適切な補助が可能となる。
- ロボットは人間が行っている作業を補助するように設計しているが、人間とロボットが共に作業を行いながら、ロボットが人間の補助を行うように設計することで、作業効率が大幅に改善されると期待できる。
- 構築したシステムにおいて、注視点取得のサンプリングレートは約 13[Hz] である。これでは視線の特性を解析するには不十分である。視線特性の研究に

において十分なサンプリング速度は 100~1000[Hz] であるが，サンプリングレート 30[Hz] でも，視線移動における潜時は少なくとも 100[msec] 程度あるので，30[Hz] のサンプリングでも，視線が動いているか，あるいは止まっているのかを測定することは可能であると報告されている [29]．従って，並列処理を行うなど工夫し，十分なサンプリングレートで測定可能なシステムに発展させる必要がある．

- EMR-8 から取得できるデータに“瞬き (自発性瞬目)”がある．瞬きの率を快適性の評価のための心理生理指標とする試み [30] があるように，瞬きと心理状態とは密接な関係がある．従って，瞬きから人間の心理状態を推定し，その状態に適した補助動作の生成が可能となると考えられる．
- 本研究では，ロボットは補助動作として適切な物体を把持し，人間への手渡しを行う．物体のモデル情報から組み付け作業まで行うことは可能であり，実装する必要がある．
- 視線の履歴情報に登録されている色と機能部の情報から人間の心理状態である“迷い”を推定したが，履歴情報と実験後のインタビューを関連づけることで迷い状態における履歴に現れる特徴的なパターンを発見した．このパターンの妥当性を定量的に評価していないため，評価手法を考える必要がある．
- 本研究では，機能部の情報に基づいた解析を行った．しかし，LEGO は色の情報や形状情報を持つ．これらの情報も解析し，人間の意図推定に利用する必要がある．

## 謝辞

本研究を進めるにあたり，素晴らしい研究環境を整えていただき，忙しい御身にも関わらず，適切なお指導とお鞭撻を頂きました池内克史教授に心からお礼を申し上げます。また，終始変わらぬ励ましを頂きました電気通信大学の木村浩助教授に深い感謝の意を表します。

更に，様々な相談にのっていただき，数々のご助言をくださった小川原光一さんには大変お世話になりました。この場を借りて深い感謝および敬意を表します。

また，高松淳さん，佐藤啓宏さんにも的確なアドバイスを頂きました。本当にありがとうございました。英文の修正をして頂きました Keni Bernardin さん，後輩の池田直人君にも感謝します。そして，大阪大学の中澤篤志講師，踊りグループの工藤俊亮さん，中岡慎一郎さん，白鳥貴亮君，Miti Ruchanurucks 君には，ミーティングのときに様々なアイデアを出して頂き，ありがとうございました。

学会出席や事務手続きで大変お世話になった池内研究室の秘書のみなさま，研究に対するお指導の他にも研究生活を進めるにあたりご尽力して頂いた池内研究室の職員みなさまに感謝いたします。

もちろん，池内研究室のその他全ての先輩，後輩，同期の方々にも感謝いたします。研究に従事できたのは皆さんの存在が大きかったと実感しています。

最後に学生生活をいろんな面から支えてくれた家族と友人たちに最大限の感謝の気持ちを表します。

平成 16 年 1 月 30 日

## 参考文献

- [1] 吉川, 大野 : “視線を読む -ユーザにやさしい視線測定環境-,” NTT R&D, **48**, 4, pp. 399–408 (1999).
- [2] 大野 : “視線を用いたインタフェース,” 情報処理学会, **44**, 7, pp. 726–737 (2003).
- [3] K. Ikeuchi and T. Suehiro: “Toward an assembly plan from observation part i,” IEEE Transactions on Robotics and Automation, **3**, 3, pp. 368–385 (1994).
- [4] Y. Kuniyoshi, M. Inaba and K. Ikeuchi: “Learning by watching,” IEEE Transactions on Robotics and Automation, **10**, 6, pp. 799–822 (1994).
- [5] J. Miura and K. Ikeuchi: “Task-oriented generation of visual sensing strategies in assembly tasks,” IEEE Transactions on Pattern Analysis and Machine Intelligence, **20**, 2, pp. 126–138 (1998).
- [6] J. Takamatsu, H. Tominaga, K. Ogawara, H. Kimura and K. Ikeuchi: “Symbolic representation of trajectories for skill generation,” Inernational Conference on Robotics and Automation, **4**, pp. 4077–4082 (2000).
- [7] K. Ogawara, S. Iba, T. Tanuki, H. Kimura and K. Ikeuchi: “Acquiring hand-action models by attention point analysis,” Inernational Conference Robotics and Automation, **4**, pp. 465–470 (2001).
- [8] H. Kimura, T. Horiuchi and K. Ikeuchi: “Task-model based human robot co-operation using vision,” IROS'99, **2**, pp. 701–706 (1999).
- [9] 森田 : “観察によるひも結び動作の学習,” 日本ロボット学会 (2002).
- [10] C. K.H.Law, M. Y. Leung, Y. Xu and S. Tso: “A cap as interface for wheelchair control,” IEEE/RSJ Conference on Intelligent Robots and Systems, pp. 1439–1444 (2002).

- [11] 大野: “階層メニュー選択における視線の利用,” 情処研報1997-HI, **97**, 24 (1997).
- [12] 高木: “視線の移動パターンに基づくユーザの迷いの検出-効果的な作業支援を目指して,” 情報処理学会論文誌, **41**, pp. 1317-1327 (2000).
- [13] D. Salvucci: “Inferring intent in eye-based interfaces: Tracing eye movements with process models,” CHI’99 Conference, pp. 254-261 (1999).
- [14] 東中, 大野: “視線に基づくユーザ理解度測定法とその応用,” 情処研報2003-HI, **102**, pp. 31-38 (2003).
- [15] 伊藤, 松原, R. Grimbergen: “将棋の認知科学的研究 (1) -記憶実験からの考察,” 情報処理学会誌, **43**, 10, pp. 2998-3010 (2002).
- [16] 満上, 浮田, 河野, 木戸出: “両眼視線情報を利用した対象の位置推定とその応用,” The 16th Annual Conference of JSAI (2002).
- [17] 鴻巣, 重松, 鈴木, 福田: “第二言語としての日本語読解における視覚情報処理に関する人間工学的研究,” 人間工学, **35**, 3, pp. 135-144 (1999).
- [18] M. G. Kwok, 石崎, 福田: “音読困難者の眼球運動特性に基づく新しい読字モデル,” KEIO SFC JOURNAL, **2**, 1, pp. 92-113 (2003).
- [19] 山田, 福田: “画像における注視点の定義と画像分析への応用,” 電子通信学会論文誌, **J69-D**, 9, pp. 1335-1342 (1986).
- [20] 大野: “Web画面における情報選択行動と視線の関係,” 信学技報HIP2000-11, pp. 31-36 (2000).
- [21] 宮代, 横溝: “眼球運動の時間構造からみた両手動作の眼と手の協調性について,” 人間工学, **25**, 2, pp. 101-107 (1989).
- [22] 鳥居塚: “保修作業時の分解・組立作業における人間行動の特性に関する考察-視線および手の動きから類推したメンタル・イメージの観点から-,” 日本プラント・ヒューマンファクター学会誌, **3**, 1, pp. 34-46 (1998).
- [23] “<http://www.cs.wustl.edu/schmidt/tao.html>,”
- [24] “<http://www.komtsu.co.jp/research/study56.htm>,”

- [25] S. Kimura, T. Kanade, H. Kano, A. Yoshida, E. Kawamura and K. Oda: “Cmu video-rate stereo machine,” Proc. of Mobile Mapping Symposium (1995).
- [26] S. Hirose and S. Amano: “The vuton: High payload high efficiency holonomic omni-directional vehicle,” Proc. Int. Symp. on Robotics Research, pp. 253–260 (1993).
- [27] Z. Zhang: “A flexible new technique for camera calibration,” Microsoft Research Techniqal Report (1998).
- [28] M. D. Wheeler and K. Ikeuchi: “Sensor modeling, probabilistic hypothesis generation, and robust localization for object recognition,” IEEE Transactions on Pattern Analysis and Machine Intelligence, **17**, 3, pp. 252–265 (1995).
- [29] 大野: “視線インタフェースから視線コミュニケーションへ-視線のある環境を目指して-,” 情処研報 2001-HI, **95**, pp. 171–178 (2001).
- [30] 山田: “瞬目による感性の評価—驚愕性瞬目反射と自発性瞬目による感情評価,” 心理学評論, **45**, pp. 20–32 (2002).