

Title	ロボットの空間状況認識のための物体間のトポロジー 関係を用いた表現記述と関係推論
Author(s)	南, 孝
Citation	
Issue Date	2007-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/3607
Rights	
Description	Supervisor: 丁 洛榮, 情報科学研究科, 修士

修 士 論 文

**ロボットの空間状況認識のための物体間の
トポロジー関係を用いた表現記述と関係推論**

北陸先端科学技術大学院大学
情報科学研究科情報システム学専攻

南 孝

2007 年 3 月

修 士 論 文

ロボットの空間状況認識のための物体間の トポロジー関係を用いた表現記述と関係推論

指導教官 丁 洛榮 助教授

審査委員主査 丁 洛榮 助教授

審査委員 松澤 照男 教授

審査委員 小谷 一孔 助教授

北陸先端科学技術大学院大学
情報科学研究科情報システム学専攻

510098 南 孝

提出年月: 2007 年 2 月

概要

本研究では、まず我々人間が通常位置関係を表すときに用いている実世界の定性的な物理法則に着目し、空間内にある物体間のトポロジー関係を抽出、表現する方法を提案した。外界のデータはステレオ法により空間内の物体の相対的な関係を取得する。得られた画像から接触情報や距離情報などをまとめて関係表現行列で表現した。この行列の特徴を抽出することで物体間のトポロジー関係を抽出することができる。さらに得られた物体間の関係の信頼度を表現する馴化評価関数を定義した。これをシステムに実装し、状況に応じて馴化評価値が変化し、評価値が大きいほど物体間の関係の信頼度が高いことを示した。これにより、このような信頼度の高い物体間の関係を知識に加えていくことは効率的であり、様々な物体の配置に用いられている何らかのルールを導き出すことができる。さらに、これらの配置列を学習することで、世界の配置ルールや常識を知ることができ、部屋の片付けをする、物を探すということも可能になると考えられる。

目次

第1章 序論	1
1.1 背景	1
1.2 目的	2
第2章 関連研究および考慮される課題	3
2.1 状況認識の定義	3
2.1.1 空間の環境認識	3
2.1.2 認知ロボティクス	4
2.2 考慮される課題	5
第3章 空間表現の物理概念	6
3.1 空間における物体間の空間3大要素	6
3.2 物理量からの空間概念の定義と限界	6
3.3 トポロジー関係の定義	8
3.3.1 接触情報	10
3.3.2 距離情報	10
3.3.3 方向情報	10
第4章 システムの概要	13
4.1 概要	13
4.2 ハードウェア	14
4.3 ソフトウェア	16
4.3.1 奥行き画像	16
4.3.2 動きベクトル	22
第5章 物体間の関係表現方法と導出アルゴリズム	24
5.1 関係表現行列	24
5.2 トポロジー関係導出アルゴリズム	25
5.2.1 接触情報	25
5.2.2 距離情報	30
5.2.3 方向情報	30

第 6 章	物体間の関係推論方法	32
6.1	馴化脱馴化パラドクス	32
6.2	馴化評価モデル	34
第 7 章	評価実験および考察	35
7.1	実験内容	35
7.2	トポロジー関係導出アルゴリズムの評価	35
7.2.1	接触情報	35
7.2.2	距離情報	36
7.2.3	方向情報	38
7.3	システム全体の評価	41
第 8 章	結論	59

第1章 序論

1.1 背景

従来の限られた特殊な環境で特定の動作をする産業用ロボットに対し、近年日常的な環境で動作する次世代ロボットの研究が盛んである。1996年、HONDA はかねてより運動性、移動性の追及を目指して研究されていた二足歩行ロボット「P2」を発表、さらに改良された「ASIMO」が発表されてヒューマノイドブームを引き起こすきっかけとなった。また1999年には、SONY が自律型四足歩行エンターテインメントロボット「AIBO」を商品化した。AIBO はロボットを人間の生活の中の商品として初めて事業化されたものである。鉄腕アトムのような夢の実現はそう簡単なものではないが、従来の工場の中で使われている産業用ロボットから、極限環境(宇宙、海洋、原子力)、一般産業(農業、水産業)、医療・福祉、公共施設、さらにはエンターテインメント、家庭といった幅広い一般社会の中で使われるロボットへと発展が予想されている [1] (図 1.1)。特に、これからの社会はますます高齢化していくと予想され、それに伴って生じる人手不足を補う介護の補助役をするといった応用や、人間の心のゆとりを促進するといった効果も期待されている。このようなことから、ロボットと人間が共存する未来への期待がますます高まってきている。しかし、従来のロボットは想定されている環境においては望んだ行動を高精度で実現することはできるが、未知の環境においては状況に応じて行動をとることができない。今後ロボットが人間の日常環境で共存していくためには、現実世界の常に変化していく環境の情報を取得しながら、その場所にあった適切な行動を獲得していくことが必要である。そのため、環境状況や場面的に表現することは重要な課題である。

状況認識とは、「環境に存在する要素の時空間内における知覚とその意味理解、および近い将来におけるそれらの状態の予測」と定義されている [11]。つまり、環境から異常の兆候などを検出することにより、何が起こったのかを知覚し、何がどのような状態で存在しているのかを把握し、この後どうしたらいいのかという将来予測を行うという3段階の過程が意思決定に先立って行われている。人間は優れた五感を通じて外界の情報を脳へ取り込み、認識している。人間は、実世界に存在する物体の大まかな3次元構造を知識として持っているために、単一の画像からでも物体間の3次元的位置関係を推測することができる。さらに、物体の構造に関する知識に加えて、物体は支えがないと下に落ちるといった物理法則の定性的な知識を持っている。2次元画像から3次元世界の構造を認識するために、こうした実世界の定性的な物理法則に基づく3次元推論能力をもつことは、より自然な画像の解釈を実現する上で必要な能力であると考えられる。またヒトとヒトの間の意

思疎通は曖昧表現で位置情報を伝達しており、自位置を伝える際に緯度経度や住所などの絶対的位置表現ではなく、ヒトがおかれている外界の中で視認した周囲のものと自分との空間的關係、「前に何々が見える」、「右に何々が見える」のような相対的位置表現を用いている。

1.2 目的

前節で述べたように、今後、ロボットが人間の日常環境で共存していくためには、実現世界の常に変化していく環境の情報を取得しながら、その場所にあった適切な行動を獲得していくことが必要である。そのため、環境状況や場面を的確に、かつ我々人間が判断に用いている方法に近い表現方法で表現することは重要な課題である。そこで本研究では、実世界の定性的な物理法則に注目して、画像情報から得られる外界のデータから物体間のトポロジー関係を抽出、表現する方法を提案する。さらに得られた情報の信頼度を表現する馴化評価モデルを提案する。そして、環境変化に対応することができる実世界認識システムの適応を検討する。

これにより、物体間の関係を抽出することで、空間の環境状況を把握できるようになるため、様々な分野で利用できる。さらにそれらの知識を蓄えていくと、例えば机と椅子の位置関係、本と本棚の位置関係など様々な物体の配置に用いられている何らかのルールを導き出すことができる。これらの配置例を学習することで世界の配置ルールや常識を知ることができ、部屋の片付けをする、物を探すということも可能になる。

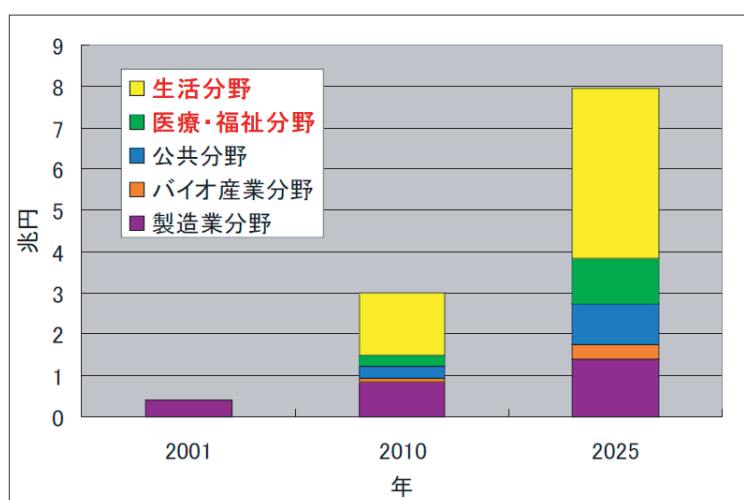


図 1.1: 今後のロボットの市場規模予測 [6]

第2章 関連研究および考慮される課題

前章では，本研究の背景ならびに目的を述べた．本章では，本稿のキーワードである「状況認識」の定義を明示する．2.1節では空間環境とロボティクスの両分野における状況認識に関する研究をいくつか紹介する．そして2.2節ではそれらをふまえた上で，考慮される課題について述べる．

2.1 状況認識の定義

本稿のキーワードである「状況認識」の定義は，M. R. Endsley らがまとめており，「環境に存在する要素の時空間内における知覚とその意味理解，および近い将来におけるそれらの状態の予測」と定義している [11]．つまり，環境から異常の兆候などを検出することにより，何が起こったのかを知覚し，何がどのような状態で存在しているのかを把握し，この後どうしたらいいのかという将来予測を行うという3段階の過程が意思決定に先立って行われているとしている．これらの定義のもと，自動車，無人航空機 [12] [13]，ロボット [14] [15] などの移動可能なシステムや，救急車の管理センタなど移動せずとも緊急事態に的確に状況認識する必要なシステムなど多岐に渡って研究開発がなされている．これらの研究の共通する目的は，知覚情報から得た情報をどのように伝えれば効果があるのかということである．この技術が確立されれば，人間とロボットの共存が可能となることから，今後益々必要になってくると考えられる．以下では，これらの定義および目的をふまえた上で，空間の環境認識，認知ロボティクスについて述べる．

2.1.1 空間の環境認識

空間の環境認識の対象として，遠景の風景画像などを基準にしている研究がほとんどである．MIT [16] では，低次元表現を用いた場所特定システムを開発している．これは，特定の物体やランドマークに依存しないシーン全体で，場所を特定するシステムである．これに対し，空間の環境認識として，室内空間を構成する平面をその連続性など外観的特徴だけでなく，平面自身が持っている機能から平面の向きや接している物体などを認識する手法を提案している研究がある．シーン認識に関する従来の手法の多くは，物体の認識に3次元形状やアスペクトなどの形状モデルを用いているため，認識結果はシーン中の物体の形状やモデルに大きく影響する．出口氏ら [17] は，一般の実世界シーンの単一画像に

対して、そのシーン中に含まれる物体の一般名称と、物体同士の位置関係を定性的に認識する画像理解システムの実現を目指している。これは、物体の3次元構造を記述した定性的モデルをあらかじめ用意されている。画像中に明確に現れている対象に対して、そのモデル当てはめによって物体の3次元構造を推定し、さらに推定された物体の3次元構造を利用して、物体間の支持関係を確かめることにより、物体の候補を検証して、全体として整合性のとれた認識結果を得ている。一方、樋口ら [18] は、物体の形状と機能、性質ならびに人物の動作には関連があると仮定し、人物の介在する映像を人物の行動を解析することにより認識している。物体のモデルを幾何学的な特徴のみではなく、物体の持つ機能・性質を加え、物体を階層構造を用いて抽象的にモデル化する。ここでのシーン認識では、室内のような空間が物体の配置や壁などからどのように構成されているかということや、その空間内で人物がどのように振舞い、行動するかを認識している。またある物体が他の物体によって支えられているといった物体間の力学的物理法則を考慮してシーンを理解する研究 [21] もある。この研究では、物体の認識が目的でなく、物体の物理的作用からの画像の理解に焦点が当てられている。また電子タグを用いた環境認識として柳沢ら [19] は、物体の位相的な特徴に注目し、実世界のイベントを検索、予測する手法を提案した。実世界のセンシングは電子タグで行っており、物体の接触、支持関係を推論している。

2.1.2 認知ロボティクス

エージェントとは本来代理人という意味であるが、人工知能の分野では1990年代にインターネットの発展に伴い人工知能への新しい手法として新しく生まれた概念である。人工知能とは、人間の知的活動をコンピュータにより人工的に実現させることを研究するコンピュータサイエンスの一分野であると定義されている。人工知能の研究分野には人間の知的活動に対応して、自動定理証明、知識表現、自然言語処理、機械学習などの分野が挙げられる。そのうち、エージェントについては、ユーザの意図を解釈して自律的に問題解決を行うプログラムを作成する技術を研究する分野である。エージェントの基本モデルを図 2.1 に示す。

人々のエージェント技術への要求は近年徐々に強くなり、エージェント技術は大きく発展し始めている。例えば、情報検索、情報配信、共同作業、ユーザインターフェース、コミュニティ活動支援など、応用範囲は多岐にわたり、それを実現する要素技術も着実に進歩している。エージェントという名称を使う場合には、要素技術に着目して、様々な形容詞をつけて呼ばれることも多い。例えば、ユーザの代理人として動作するエージェントは自分で自分自身の行動を自律的に決めることが求められる「自律エージェント」がある。その例が認知ロボティクスである。

従来のロボティクスはもっぱら設計者が身体運動や物体操作の物理として構築した理解の体系の帰結のみを明示的にロボットに埋め込むというものが多かった。それに対し、浅田ら [8] は、ロボットの認知プロセスを理解し、具体的メカニズムとして具現することを目指す認知ロボティクスの必要性を唱えている。認知ロボティクスとは [9] [10]、従来、

設計者が明示的にロボットの行動を規定してきたことに対し、環境との相互作用からロボットが自ら行動を学習し、それらを発達させていく過程に内包されるシンボル化を実現するためのロボット設計論である。認知ロボティクスの焦点は、自律エージェントが環境との相互作用を通して、実世界をどのように表現し、行動を獲得していくかといった、ロボットの認知発達過程にある。

2.2 考慮される課題

前章や前節でも述べたように、今後ロボットが人間の日常生活で共存していくためには、その環境状況や場면을的確に表現することは重要な課題である。空間内にある物体間の関係をいかに我々人間が判断に用いている方法に近い表現方法に置き換えることが課題である。そのために、我々が日常用いている空間表現を効率的に形式化することが重要である。さらに、工場で常に同じ環境で同じ作業を行う産業用ロボットと違い、生活環境は各家庭によって異なり、また同じ環境内であっても人間の行動によって環境は変化しうる。このことから、主観によらず一意的に関係を導出することが必要である。

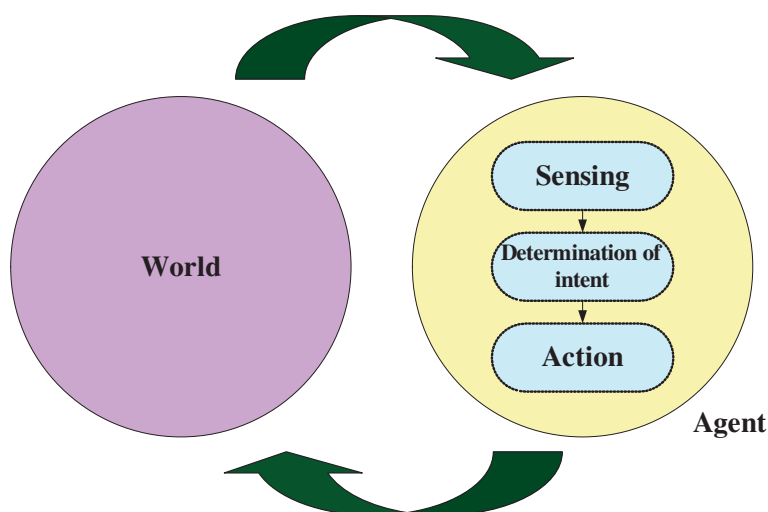


図 2.1: 認知判断と行動のメカニズム

第3章 空間表現の物理概念

前章までに研究の背景や関連研究を述べた。本章では、本研究で用いる「トポロジー関係」の意味および定義を行う。3.1節では空間における物体間の空間表現の3大要素を説明し、各情報の特徴を述べる。3.2節では我々が日常使っている表現方法をまとめる。3.3節では本研究で扱う「トポロジー関係」の定義を示し、各情報の表現方法を明確にする。

3.1 空間における物体間の空間3大要素

空間を表現する語句の物理的解釈については英語の前置詞の研究があるが、英語と日本語における空間概念の現れ方は全く同一ではないので、そこから得られた知見を日本語にそのまま当てはめることができない。しかし、共通する点もあり、位置関係を表現するための言語手段が文化を問わずどの言語にも備わっていると仮定し、英語と日本語に現れる空間表現の共通点、相違点を比較することで、日本語の空間表現の形態的特徴を明らかにしようと試みている [5]。

空間における2つのオブジェクトの関係を表現する空間情報の3大要素として、距離情報、方向情報、接触情報が考えられる。その要素の関係を図 3.1 に示す。

距離情報は「大小」、「遠近」、「長短」などの概念があり、定量的に定まることが多く絶対的表現であるため、強い制約条件が備えられている。またユーザによって認識誤差が生じてしまう。方向情報は「東西南北」などの概念である絶対的方向関係と「前後左右上下」などの概念である相対的方向関係がある。絶対的方向関係は、全ての人で共通の方向を示すが、自分が今どの方向を向いているのか定まっていなければならない。一方、相対的方向関係は、見えるものをそのものの関係で表すことができる。接触情報は、相対的な関係であり、「連続」、「極限」といった概念があり、物体がつながっているか切れているかななどの区別が可能である。

3.2 物理量からの空間概念の定義と限界

距離、方向、位置を示す表現方法に、指示対象となる物体の位置や方向といった物理量を照らし合わせることで物体間の位置関係を表現することができるが、我々人間は日常、物体の物理量からだけではなく、その他の要素によって位置関係を表現する場合がある。純粋に物体の物理量しか知らないということは、その物体の形状が分かっているてもそれ

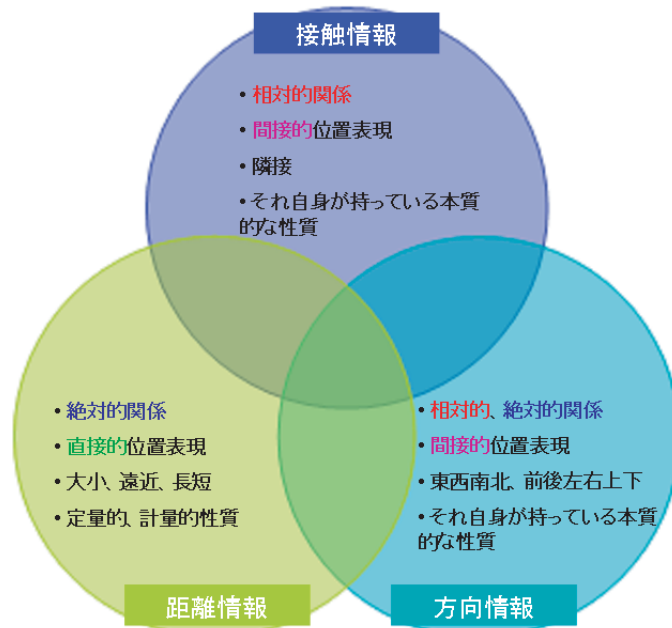


図 3.1: 空間内の物体間の空間関係

が何であるか，物体の用途も知らないため物体自体の方向も知らないということである．例えば机の配置について，足にあたる部分が下になり，広くて平らな部分が上を向くように置くことが正しい置き方だということも分からないことになる．物体の位置の表現は，その物体がテレビである，机であるという事実が大きく左右される．このような物体が固有に持つ意味という属性を無視した物理量のみから絶対的に定義できる理論概念は非常に狭い領域のもので，我々人間を取り囲んでいる複雑で不完全な日常の世界はそれをはるかに上回る広い領域を占めている．その理由の1つは言語が持つ多様性にある．この多義性を引き起こすのが習慣，環境である．この慣習，環境に基づいて言語の理論概念から逸脱した例外を許容することで我々が日常用いる一般的な言語表現が成立する．慣習や環境によって理想的概念から逸脱した例外は，我々が日常例外と意識することは少なく，ある推論規則に基づいて常識として定着してしまっている．このような慣習や環境といった属性を利用する推論規則は物理量からの直接的な定義は困難である．そこで本研究では，物理量から定着できる範囲での位置関係の表現を実現することを目的とする．さらに，距離情報，方向情報，接触情報のそれぞれの相対的表現を合わせたものを「トポロジー関係」と定義する．次節において，トポロジー関係のそれぞれの情報について意味および定義を述べる．

3.3 トポロジー関係の定義

人間は実世界に存在する物体の大まかな3次元構造を知識として持っているために、単一の画像からでも物体間の3次元的な位置関係を推測することができる。さらに、物体の構造に関する知識に加えて、物体は支えがないと下に落ちるといった物理法則の定性的な知識を持っている。2次元画像から3次元世界の構造を認識するために、こうした実世界の定性的な物理法則に基づく3次元推論能力をもつことは、自然な画像の解釈を実現する上で必要な能力だと考えられる。またヒトとヒトの間の意思疎通は曖昧表現で位置情報を伝達しており、自位置を伝える際に緯度経度や住所などの絶対的位置表現ではなく、ヒトがおかれている外界の中で視認した周囲のものと自分との空間的關係、「前に何々が見える」、「右に何々が見える」のような相対的表現を用いている。そこで本研究では、画像情報から得られるトポロジー関係に着目する。前節で空間における2つの物体間を表現する空間的關係の3大要素を述べたように、各要素の相対的特徴を合わせたものをトポロジー関係と定義している。トポロジー表現は、物理的な法則であるので関係を一意に定めることができる。また主に相対的な位置関係を用いているため、表現や推論が容易である。さらに情報が欠落した場合でも他の情報から欠落した情報がある程度補完することができる。これにより、トポロジー関係はより人間と類似した表現方法であり、どんな物体間の関係であっても環境状況を表現することができると考えられる。

我々人間が通常空間関係を表現するときに用いている表現を表3.1にまとめた。図3.2に我々人間が通常見ている空間の風景を示している。人間はこの風景を説明するときに例えば、「マウスパッドの上にマウスがある」「ホワイトボードが近くにあり、奥にテーブルやポットがある」という表現をする。表3.1に示すように、空間関係を表現するものは大きく、「物体」に対して適応されるものと、「場所」に対して適応されるものがある。本研究では、シーンの内容を部屋の中に限定し、物体間の関係を表現することを目的としているため、「場所」に対して適応されるものは取り上げず、「物体」に対して適用されるものだけに着目し、検討することとする。トポロジー関係の接触情報、距離情報、方向情報のそれぞれについて以下に定義する。

表 3.1: 空間表現の分類

トポロジー関係	一例
距離情報	近く, 遠く, そば, 付近, 隣, ~辺, ~沿い, ここ, そこ, あそこ
方向情報	東西南北, 左右, 上下, こちら, そちら, あちら, 手前
接触 (位置) 情報	隣接, 包含, 連続, まわり, 内, 外



(a) 机上空間



(b) 室内空間

図 3.2: 人間が生活している環境風景

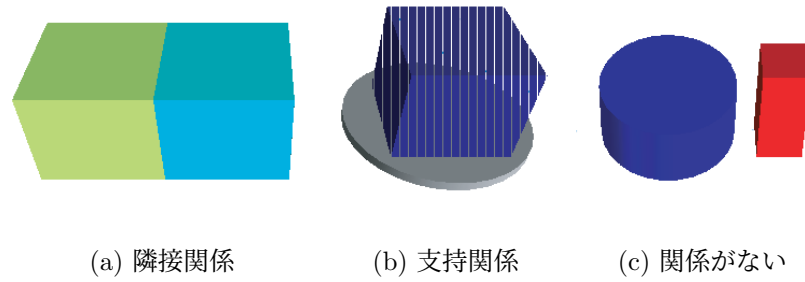


図 3.3: 接触情報の種類

3.3.1 接触情報

基準物体と目的物体との間の接触を用いた位置関係を表現するために接触情報を用いる。接触情報には、近接、交差、包含といったものがある。しかし、多数の類型にしてしまうとそれだけ意味も多様になり、複雑になってしまう。そこで本研究では、図 3.3 に示すように物体の周囲の一部が他と左右の方向情報をもって接している「隣接関係 (CONTACT)」、他方の重さを支えている上下の方向情報をもった「支持関係 (OVERLAP)」、物体間の関係性がない状態 (DISJOINT)」の 3 つの関係で類型化することにする。

3.3.2 距離情報

基準物体と目的物体との間の距離を用いた位置関係を表現するために距離情報を用いる。距離情報には、近接度を表す「近くにある」、「隣にある」などと、遠さを表す「遠くにある」、「向こうにある」などがある。特に日本では、「こそあど」言葉を用いてよく表現している。「ここ」、「そこ」、「あそこ」、「どこ」は定量的に表現しているのではなく、曖昧な表現方法である。つまり、我々人間の距離情報に関する定義づけは難しく、ユーザによって異なった判断をしてしまうことがしばしばある。そこで本研究では、近接度を表す「近くにある (NEAR)」の距離情報のみを定式化することで表現する。遠さを表す距離情報は、部屋のシーンのように壁で区切られた部屋の中という領域内で考える場合、距離の最大値が決まってしまう。さらに、「近く」なければ「遠く」にあると言える場合が多く、物体間の位置を表すときには近接度のみを定式化すれば十分であると考えた。

3.3.3 方向情報

目的物体が基準物体に対してどの方向にあたるかを表現するには方向情報を用いる。方向情報を物理的に定義するための基準として、基準物体は自身の中心を原点とする 6 つの

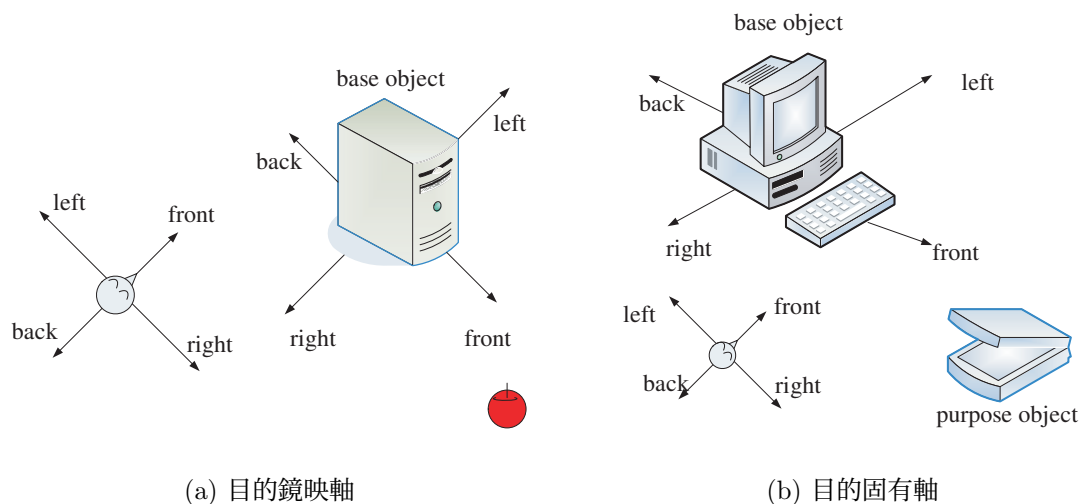


図 3.4: 方向軸の決定

半直線の軸，つまり前後，左右，上下それぞれの方向軸を持つこととする．上下の方向軸の場合，重力の働く方向によって決まるため，鉛直上下の方向に定めた軸を基準に「上」，「下」の方向が決定できる．なお本研究では，“物体は空中に浮いていることはない”としているため上下の方向情報をもつものは前節で述べた「支持関係 (OVERLAP)」に相当する．前後，左右の方向軸の決定は，システムの位置や対象物体の性質に依存するため，いくつかの場合に分けられる．ある目的物体を方向を用いて表すとき目的鏡映軸 (図 3.4(a)) と目的固有軸 (図 3.4(b)) の 2 種類が考えられる．以下にそれぞれの軸について定義する．

- 目的鏡映軸

システムに対面した基準物体を基準とする方向を指す場合，基準物体の中心を原点とし，システムが持つ方向軸を鏡映順にした方向軸を定める．例えば，「サーバの右手前にりんごがある」という位置関係では，サーバの中心を原点としてシステムの方向を前とする方向軸により，「サーバの右手前」と定める．システムの視点が移動すれば軸自体が動いてしまう視点依存の指示である．

- 目的固有軸

システムの位置に関わらず，基準物体がもつ性質から方向軸が決まる場合がある．例えば，「コンピュータの前にスキャナーがある」という位置関係では，コンピュータの中心を原点としてディスプレイがある方向を前とする方向軸を定め，前方向の軸によって「コンピュータの前」を定める．この場合は視点が動いても方向軸は移動しないが，それぞれの物体がもつ方向軸が分かっている必要がある．

目的固有軸は，物体ごとの固有軸を定めるためには，その物体の性質を解析し，前後の決定をしなければならない．しかし，前後が決定できないような物体の場合など前後を決

定することが極めて困難である．なぜなら我々人間は日常生活においてアフォーダンスや習慣によって物体の用途を学び，それによって前後を決定するというプロセスを経ているからである．あらかじめ物体の前後を定義しておく手法も考えられるが，本研究では扱わず，目的鏡映軸による 8 方向の方向情報表現を用いることとする．

第4章 システムの概要

前章では、用いた各手法の一般的解釈について述べた。本章では、本研究のシステムについて述べる。4.1節では全体の概要を述べる。4.2節ではカメラシステムのハードウェアについて述べる。4.3節ではソフトウェアの面からシステム GUI について述べる。

4.1 概要

システムの概要を図 4.1 に示す。動作の流れは大きく、外界からデータを取得、画像処理、関係表現行列の作成、物体間のトポロジー関係抽出、知識蓄積の5つのステップに分けられる。

- Step 1: 外界のデータは PGR 社 Flea を 2 台用いて取得する。なお、カメラのレンズ歪みを補正するために、キャリブレーションを行った。
- Step 2: 入力された画像から奥行き画像を作成する。奥行き画像を使用すれば、面の特性における不連続性を用いなくても異なる物体は 3 次元空間で分離されているという特性によって、シーンの中の物体の分割が可能になる。しかし、他と接触していたり、他の上に乗っている物体はそう簡単には分割できない。そこで、動きベクトルを用いることで変化を抽出し、物体の分割を行う。
- Step 3: 奥行き画像と 2 次元画像を元に、関係表現行列を作成する。作成するタイミングは、動きベクトルがゼロになったときである。この関係表現行列は共起行列を拡張した行列である。なお、関係表現行列の作成手法の詳細は 5 章で述べる。
- Step 4: 関係表現行列から物体間のトポロジー関係を導出する。なお、導出アルゴリズムの詳細は 5 章で述べる。
- Step 5: 得られた情報を馴化評価モデルを元に、知識ベースに効率的に蓄積する。さらに、得られたトポロジー関係と知識を比較することによって、ロボットの行動生成と結びつけることが可能となる。なお、詳細は 6 章で述べる。

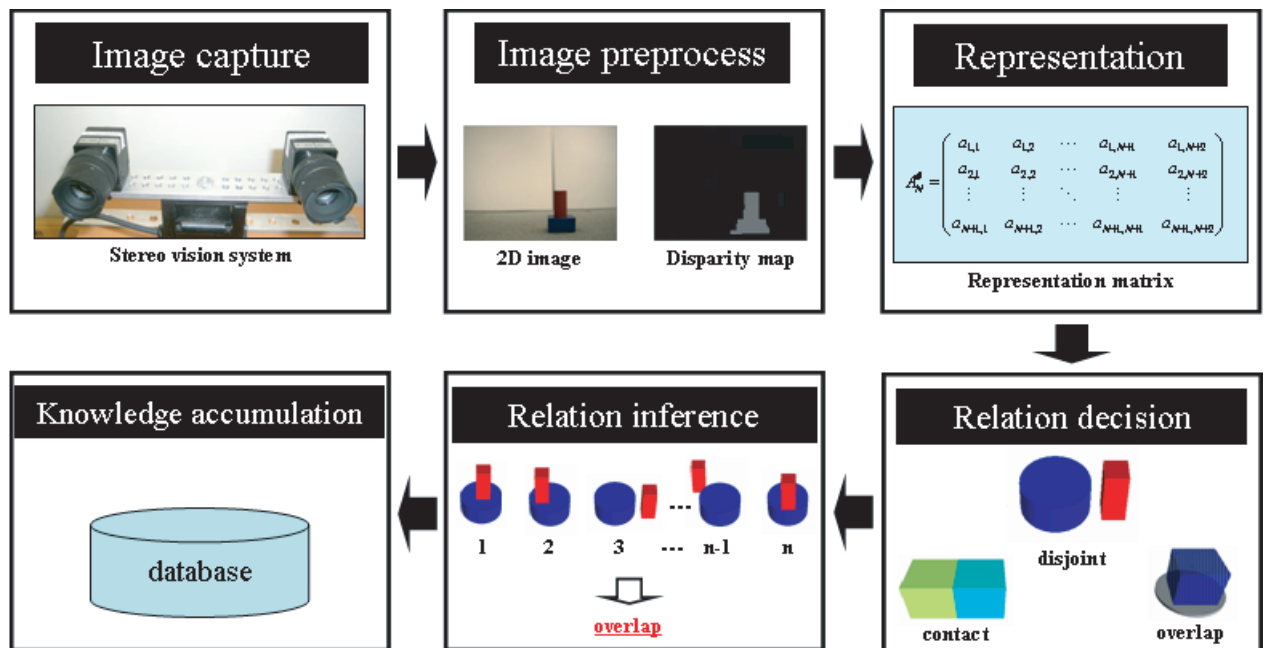
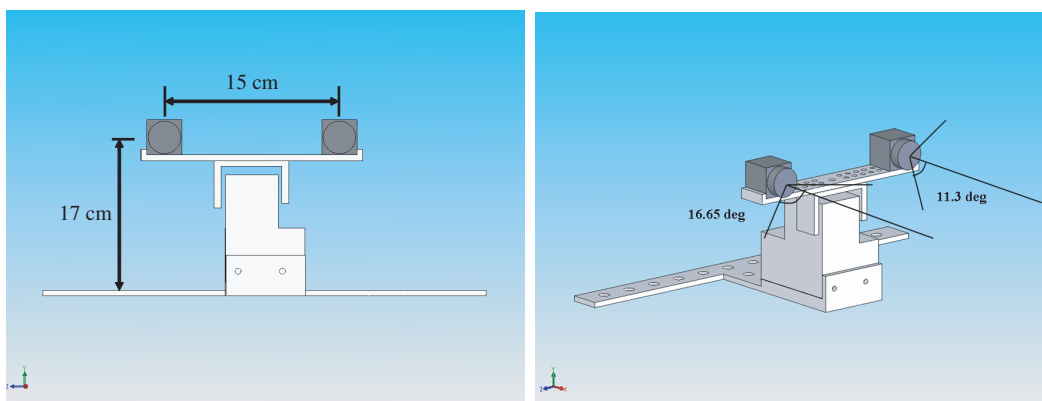


図 4.1: システムの流れ

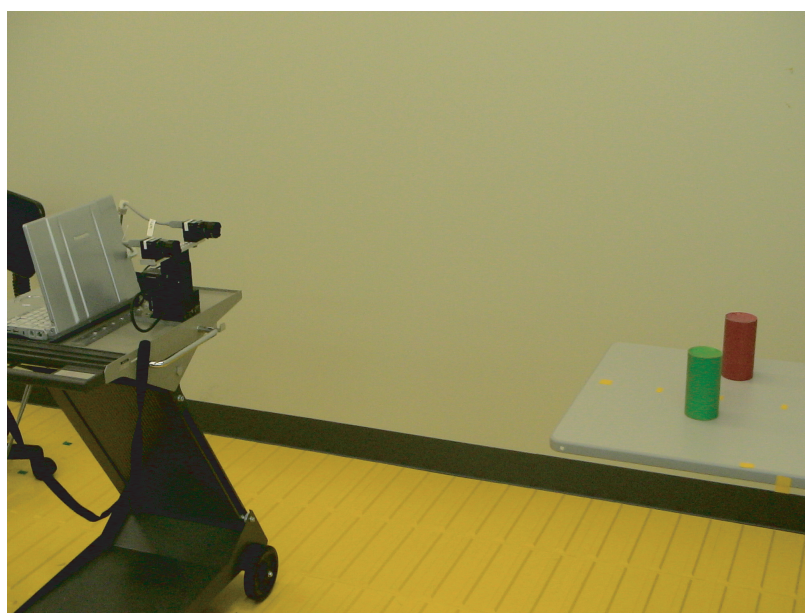
4.2 ハードウェア

外界のデータを取得するために図 4.2 に示すように、カメラシステムを設計した。設計したカメラシステムはステレオ法を使用した。ステレオ法とは、人間が2つの目で外部環境を3次元的に把握するのと同様に2台のカメラを平行に並べることで3次元的に物体の距離を測定する技術である。なおステレオ法に関しては次節および付録 A で詳細に述べる。今回のシステムでは、PGR 社の Flea というカメラを使用した。このカメラはカラー映像を写すことができ、大きさが 30*30*30mm と小型ながら、高解像度を有している。双方のカメラは互いに平行になるよう高さ 17cm、ベース間距離は 15cm、パンチルトユニットの上に設置した。パンチルトユニットとは、ユニット上部に設置したカメラの視点を左右上下に動かす装置である。今回のシステムでは使用していないが、ユニットを制御することによって、より広範囲の環境認識も可能となる。またカメラのパラメータでもある視野角は図 4.2 に示すように左右方向は 16.65deg、上下方向は 11.3deg である。このことから、カメラと物体までの距離が 80cm 以上であれば有効な結果が得られる。80cm に相当する画像上の視差は 100pixel であった。なお、カメラで撮影されたデータは IEEE1394 によってパソコンに送信され、システム GUI に映し出される。



(a) 正面図

(b) 斜めからの図



(c) 実験風景

図 4.2: カメラシステムの概要

4.3 ソフトウェア

ソフトウェアの部分は、システム GUI である。システム GUI は Visual C++ によって作成した。この GUI 上では、外界の画像表示、奥行き画像、動きベクトル、関係表現行列および空間言語表現、推論グラフからなる。この節では、奥行き画像と動きベクトルの部分について述べ、他の部分については後ほど改めて述べることにする。

4.3.1 奥行き画像

空間情報認識の手法の 1 つである距離認識には大きく分けて 2 つの手法が考えられる。1 つ目はレーザ光などを物体に直接当て、その光の状態から物体の場所を判断する方法である。2 つ目はある物体の画像を用いて、その画像を処理することで距離情報を得る方法である。前者の方法は、観測する物体がその場所にあることを前提とする方法であるので、本研究では後者の方法を利用した。画像を使った立体認識には大きく分けて次の 2 つがある。

1. 1 枚の画像を用い、立体の認識をする方法

1 枚の画像では、画像の距離情報が失われてしまっている。よって何らかの方法でそれを補わなくてはならない。その方法としていくつか考えられている。1 つ目は、画像の距離情報が予め既知な点あるいは点列を用いる方法。2 つ目は、画像のセグメンテーションを行い、予め用意したモデルと照合する方法。3 つ目は、画像中の陰影やテクスチャを用いる方法である。

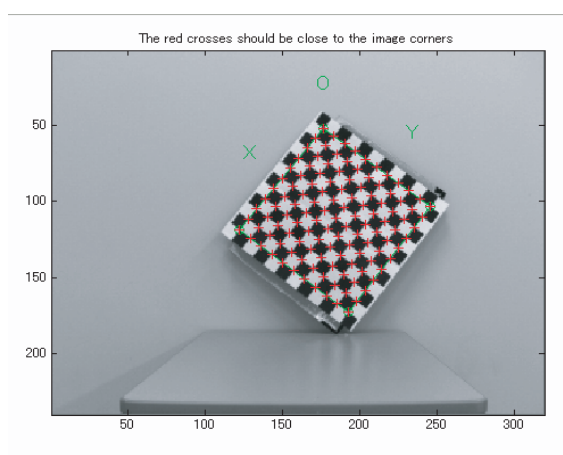
2. ステレオ画像など複数の画像を用いて立体を認識する方法

複数の視点または照明条件で撮影した画像を用いることにより、3 次元情報の算出を行う。その方法を計測原理により大別すると、受動型計測と能動型計測がある。受動型計測の代表的な計測方式としては、ステレオマッチング法が挙げられる。能動型計測の代表的な計測方式としては、照度差ステレオ法などが挙げられる。

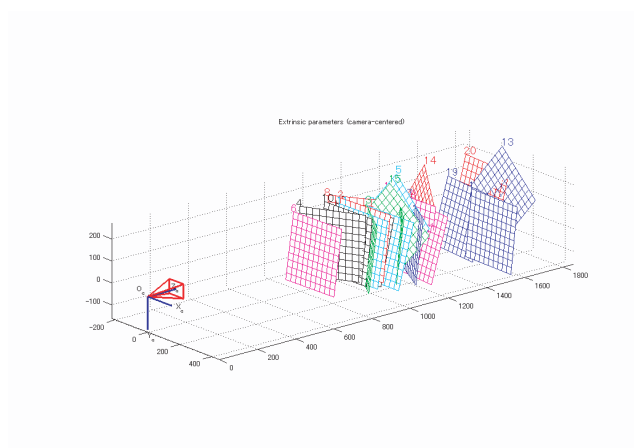
本研究では、2 つのカメラを用いた Pixel-to-Pixel 法 [28] による奥行き画像を作成した。奥行き画像を用いる最大の理由は、物体の面の特性における不連続性を用いなくても異なる物体は 3 次元空間で分離されているという特性によってシーン中の物体の分割が可能になるからである。なお、奥行き画像作成に関する詳細は付録 A で述べる。

奥行き画像を作成する上で、対応点を見つけ出すことが最大の問題である。通常のステレオビジョンシステムでは既にレンズ歪みを考慮された画像が映し出されるが、本研究で用いた Flea はレンズ歪み補正がなされていない。そこでまず、カメラのレンズ歪みによる対応点の誤認識を防ぐため、キャリブレーションを行った。キャリブレーションは、Camera calibration Toolbox [30] を参照した。このライブラリは Z.Zhang の論文などを元に J.Y.Bouguet が実装したものである。図 4.4 に示すように平面パターンの画像列を入力

して、ホモグラフィに基づいたキャリブレーションを行う。次に図 4.3 に示すような GUI が準備されており、画像列を読み込んだ後、チェックボードのコーナーを選択するだけで特徴点を自動抽出して、内部パラメータと外部パラメータを推定する。なお、キャリブレーションに関する詳細は付録 B で述べる。本研究で用いた Flea カメラの各パラメータを求めるため、図 4.4 に示すように左右のカメラでそれぞれ 27 枚のサンプル画像を作成した。そしてこの手法によってキャリブレーションを行った。その結果、用いたカメラの各パラメータは次の通りである。なお、 α は再サンプルによる不確定要素、 f はカメラの焦点距離、 c はレンズ歪みの中心座標、 R は実世界とカメラ間の座標変換による回転要素、 Kc はレンズ歪み係数である。



(a) チェックボードの選択



(b) 特徴点の自動抽出

図 4.3: Camera Calibration Toolbox

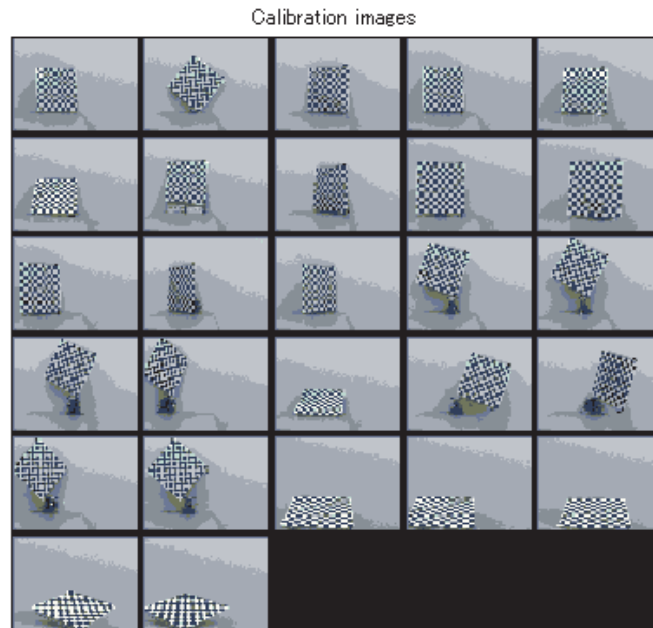


図 4.4: キャリブレーション用サンプル画像

- 左カメラ

$$\alpha = -0.00002$$

$$f = [547.15891, 547.47343]$$

$$c = [159.51754, 132.60466]$$

$$R = [0.9999, -0.001, -0.0105; 0.001, 1, -0.0024; 0.0105, 0.0024, 0.9999]$$

$$Kc = [-0.44619, 0.22353, -0.00319, -0.00111, 0.0]$$

- 右カメラ

$$\alpha = -0.00048$$

$$f = [546.87276, 547.06976]$$

$$c = [156.65204, 136.25426]$$

$$R = [0.9996, -0.0206, -0.0203; 0.0207, 0.9998, 0.0022; 0.0203, -0.0026, 0.9998]$$

$$Kc = [-0.39804, -0.02663, -0.00038, -0.00027, 0.0]$$

これらのパラメータをもとにカメラのレンズ歪み補正画像を作成した。図 4.5 において、左上・右上が補正前の画像，左下・右下が補正後の画像を示している。補正前は左右の画像のエピポーラ線が一致していないが，補正後はエピポーラ線が一致している。それぞれ

の画像で奥行き画像を作成したものを図 4.6 に示す. 図 4.6 に示すように補正前 (右) では文字などの部分でうまく対応点を抽出できないために奥行き画像が正確に得ることができていない. しかし補正後 (左) ではエピポーラ線が一致しているためにきれいな奥行き画像を得ていることがわかる.

またこのシステムの奥行き精度を測定するため, カメラの位置を固定し, 物体を 10cm ずつ移動させることを各 10 回ずつ測定した. その結果を図 4.7 に示す. 図 4.7(a) は画像の視差とカメラと物体までの距離の関係をグラフに示した. 特徴点は累乗曲線に近似できると判断し,

$$y = 13042x^{-1.0953} \quad (4.1)$$

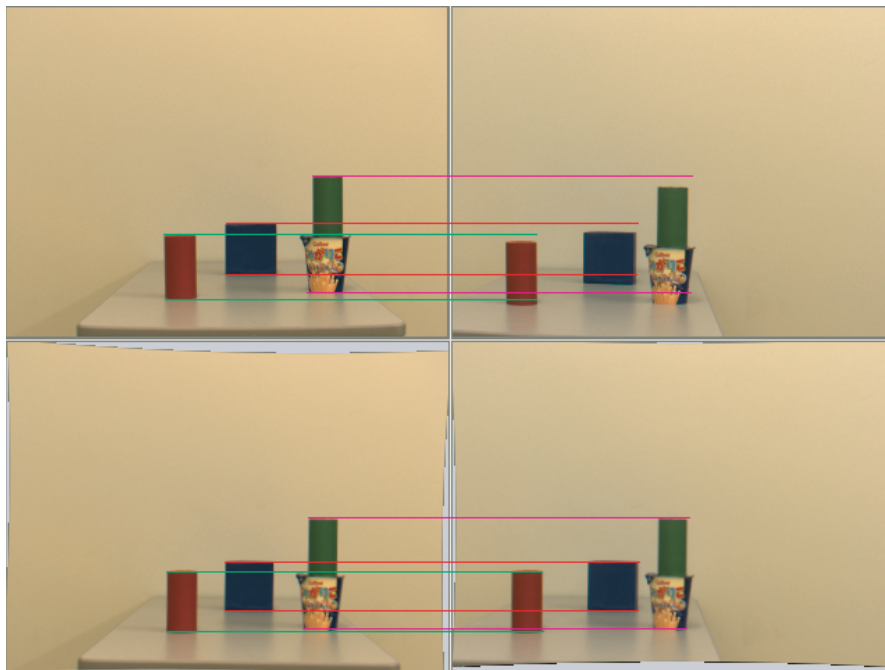
とした. y は物体までの距離, x は画像上の視差である. この距離をモデル距離と定義し, 実際の距離とモデル距離との誤差評価を行った. 図 4.7(b) は実際の距離とモデル距離の誤差が最小, 最大であった結果を示している. 全サンプルの誤差の平均および分散は,

$$\mu = 1.324, \quad \sigma = 1.143 \quad (4.2)$$

であり, 誤差の最大 $\mu_{max} = 4.86$, 最小 $\mu_{min} = 0.01$ であった. よって高精度にカメラから物体までの奥行き情報を取得できているといえる.



(a)



(b)

図 4.5: 補正画像 (左上・右上：補正前画像，左下・右下：補正後画像)

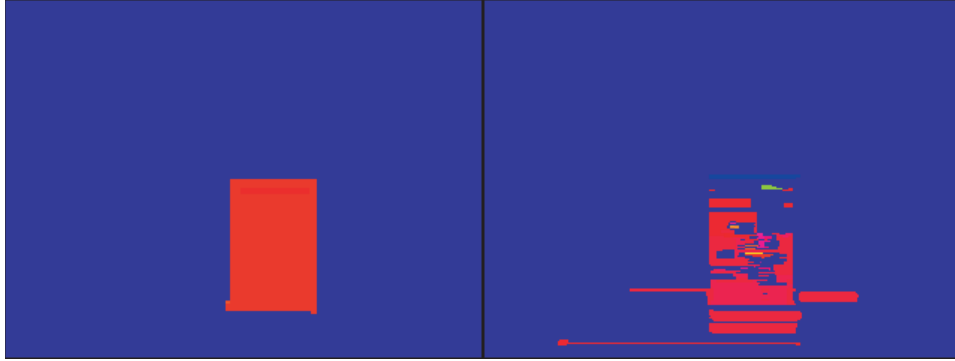
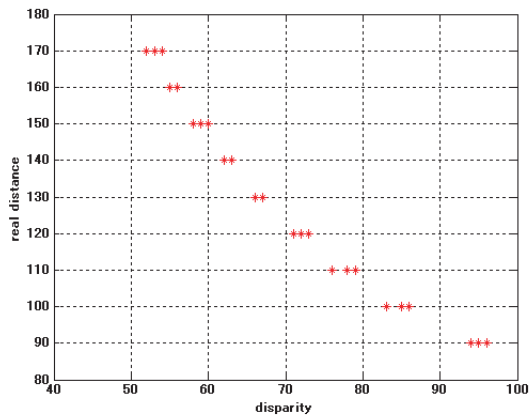
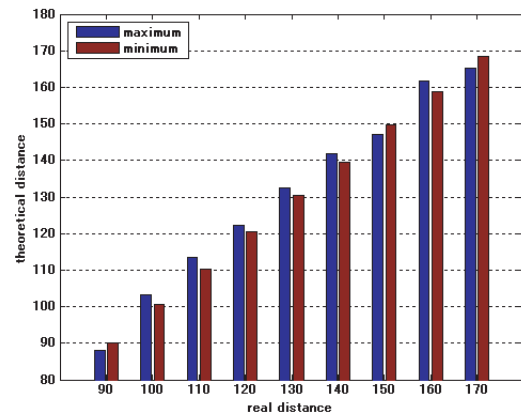


図 4.6: 補正前と補正後の奥行き画像 (左：補正後, 右：補正前)



(a) 視差と奥行きの関係



(b) 実際の距離とモデル距離の誤差

図 4.7: 奥行き情報の評価

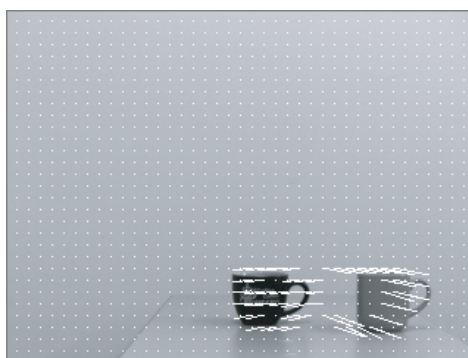
4.3.2 動きベクトル

動きベクトルとはカメラと移動物体との相対的な動きによって生じる，画面上に発生する速度ベクトルを求めたものである．速度ベクトルとは画像中のある1点において示される画素情報が次の画面上で，どちらの方向へ，どの程度の距離を移動するかを示したものである．これは図4.8で示すように，平面上や空間内の矢印として幾何学的に表現されることが多い．図4.8は，カップが移動したときのそれぞれの動きベクトルを表示したものである．ここで，白い点は，その点の速度ベクトルの大きさが0であることを示しており，長さを持った線分の場合は，線分の向きがその点の速度の方向，線分の長さが動きの量を表している．この動きベクトルを求める方法にはブロックマッチング法と勾配法の2つがある．各手法の詳細は付録Cで述べるため，ここでは簡潔に特徴を述べる．ブロックマッチング法は精度よく動きを検出できるが，処理量が多い．一方，勾配法は処理量がブロックマッチング法に比べて少なく処理できるが，検出の精度が低い．



(a) t フレーム目の画像

(b) $t + 1$ フレーム目の画像



(c) 動きベクトルの表示

図 4.8: 動きベクトル

そこで，本研究では動きベクトルをブロックマッチング法により抽出している．ブロックマッチングのブロックの大きさは処理速度と検出数を考慮して定めた．まず処理速度に関しては，数回の実験を行ったところ，ブロックサイズが2，4，8，16の場合，それぞれ12s，3s，1s，0.5sであった．実環境で動作することを考えると処理に長時間かかってはいけない．また検出数が少なすぎると対応する点を見つけることができないときもある．よって，本研究では実験の結果からブロックサイズは，8が適切であると判断した(図4.9)．

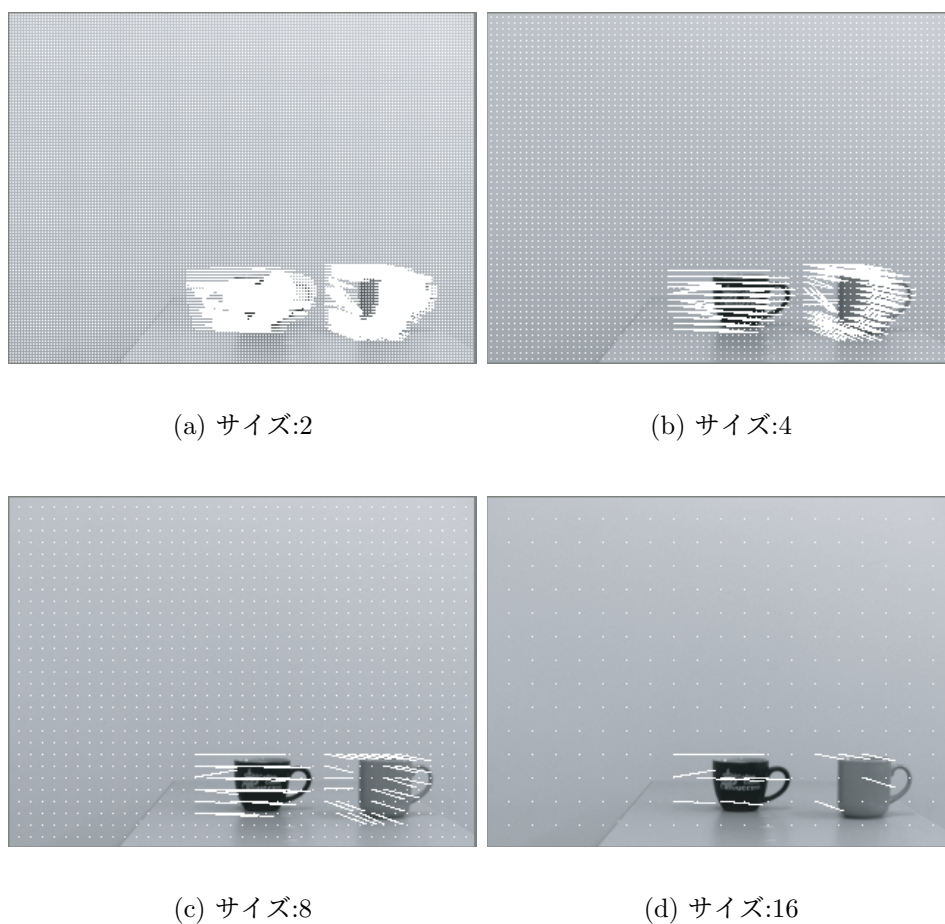


図 4.9: ブロックマッチングのブロックサイズ

第5章 物体間の関係表現方法と導出アルゴリズム

前章では、本システムの概要について述べた。本章では、システムにおける物体間の関係表現方法とその導出アルゴリズムについて述べる。6.1節では空間内にある物体の関係を表現する関係表現行列の定義と意味について述べる。6.2節では本研究で定義したトポロジー関係の導出アルゴリズムについて、それぞれの情報にわけて述べる。

5.1 関係表現行列

共起行列とは、出現するイベントのそれぞれの二項間の関連性の強さをある距離の間に現れるイベントの出現頻度により表し、全ての二項間のイベントの関連性を表現した行列である。共起行列の詳細は付録Dで述べる。本研究の表現方法は、この共起行列を拡張したものをを用いる。つまり、2つの物体間の関連性に着目し、シーン全体に現れている全ての物体間の関連性を共起行列として表現する。そして、そのデータからどのトポロジー関係のカテゴリに属するものであるか否かを判断する。

本研究の提案する関係表現行列を示す。 $a_{A,B}$ は物体Aと物体Bとの関連性の割合、 N は物体の数(1toN)、 θ は画素間の方向性(0, 45, 90, 135deg)である。 A_{11}^θ は物体間の関連性を表す行列であり、接触情報があるか否かの指標となる。 A_{12} は奥行き情報を表す行列であり、カメラから物体までの距離のデータが表示され、方向情報である前後関係の指標となる。 A_{13} は物体間の相対的な方向情報を表す行列であり、シーンの中の一番左側にある物体を基準物体としたときのそれぞれの目的物体との方向情報の指標となる。つまり、 A_N^θ は、それぞれの情報を表している3つの行列を一つにまとめた主行列となっている。

$$A_N^\theta = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,N+1} & b_1 & c_1 \\ a_{2,1} & a_{2,2} & \cdots & a_{2,N+1} & b_2 & c_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ a_{N+1,1} & a_{N+1,2} & \cdots & a_{N+1,N+1} & b_{N+1} & c_{N+1} \end{pmatrix} \quad (5.1)$$

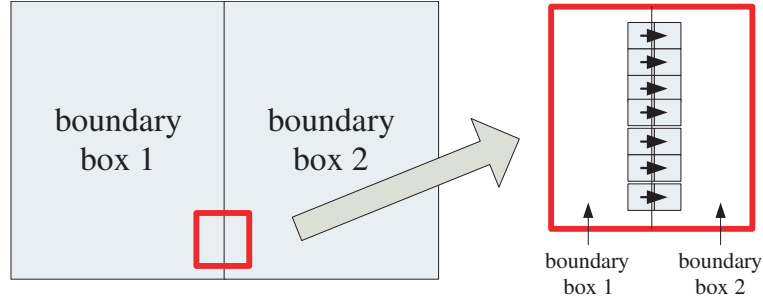


図 5.1: 接触判定

$$A_{11}^{\theta} = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,N+1} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,N+1} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N+1,1} & a_{N+1,2} & \cdots & a_{N+1,N+1} \end{pmatrix} \quad (5.2)$$

$$A_{12} = \begin{pmatrix} b_1 & b_2 & \cdots & b_{N+1} \end{pmatrix}^T \quad (5.3)$$

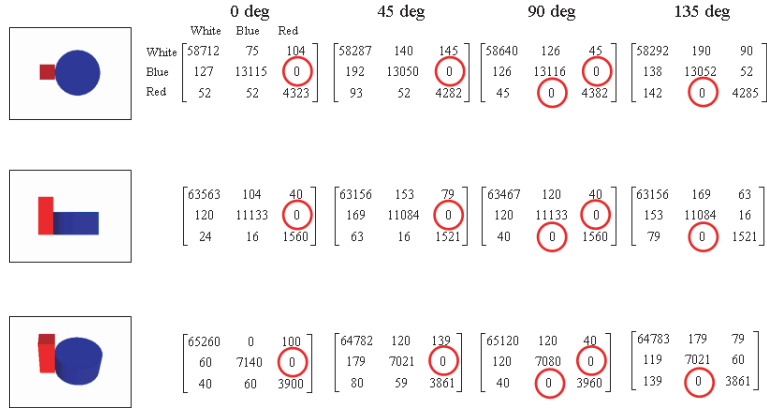
$$A_{13} = \begin{pmatrix} c_1 & c_2 & \cdots & c_{N+1} \end{pmatrix}^T \quad (5.4)$$

5.2 トポロジー関係導出アルゴリズム

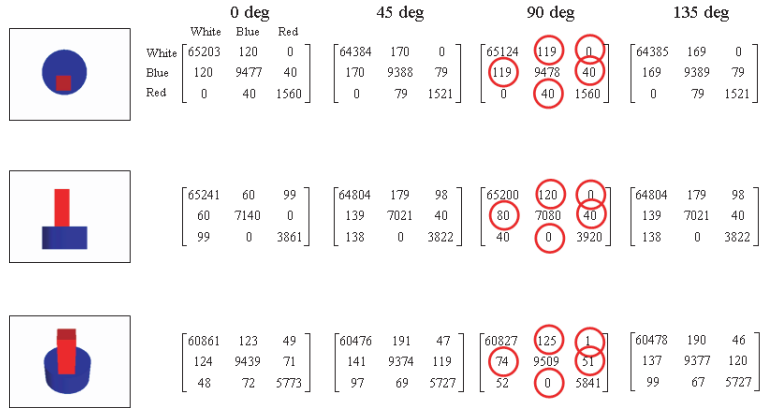
本研究では、「接触情報」, 「方向情報」, 「距離情報」の3つで表現される相対的な表現方法を用いた物体間の関係をトポロジー関係と定義した。以下では、それぞれの情報の抽出手法をまとめる。

5.2.1 接触情報

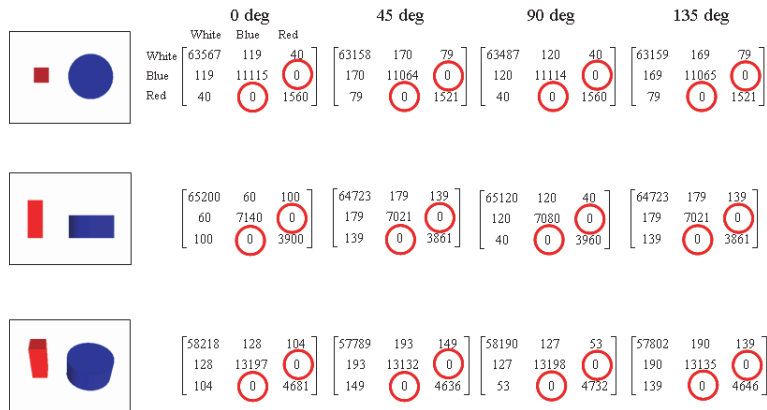
接触情報は、左右および上下の方向情報も同時に得ることができる。それぞれの接触の判断は基準物体と目的物体の boundary box の接触している割合を利用する。この割合は、前節で述べた関係表現行列 A_{11}^{θ} の各要素 $a_{i,j}$ に相当する。例えば、「隣接関係 (CONTACT)」は図 5.1 に示すように、左側にある物体は右側にある物体に対して、 $0deg$ 方向に要素の特徴がある。同様に、「支持関係 (OVERLAP)」は $90deg$ 方向に要素の特徴がある。このことを確かめるため、いくつかのサンプル画像を作成し、前節で述べた関係表現行列 A_{11}^{θ} に対して、パターン特有の性質を担う特徴抽出を行った。図 5.2, 図 5.3 はその一例である。その結果、それぞれの接触情報に対し、次のような規則性を見出すことができた。



(a) 隣接関係 (CONTACT)

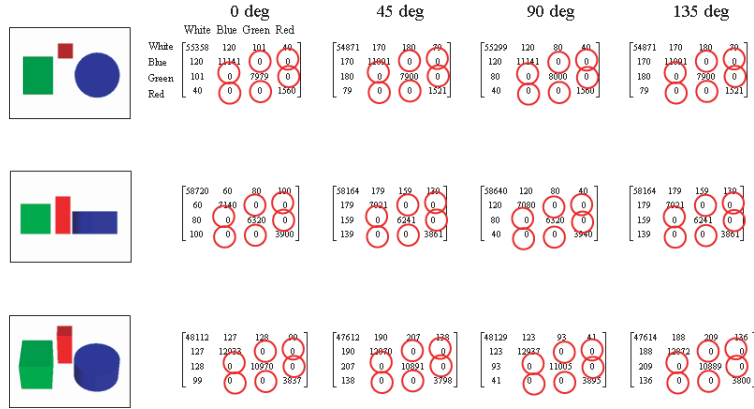


(b) 支持関係 (OVERLAP)

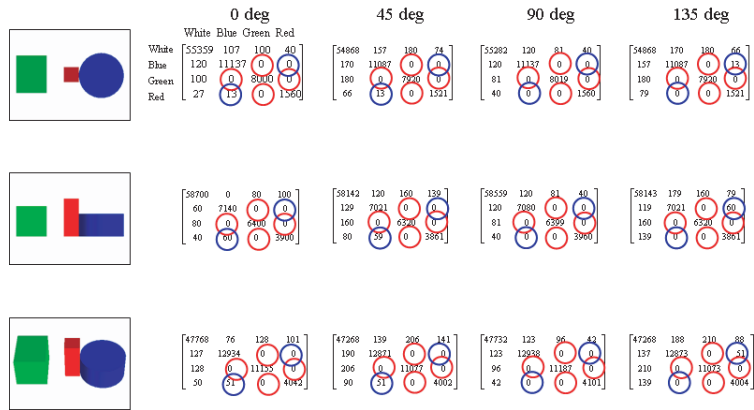


(c) 物体間の関係性がない (DISJOINT)

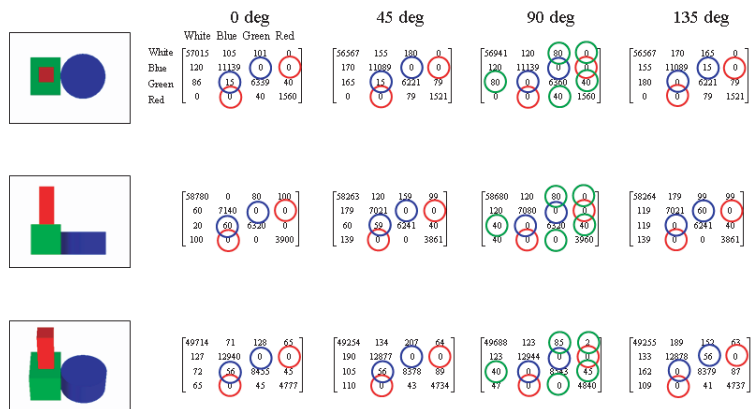
図 5.2: 接触関係



(a) 全く関係がない場合



(b) 隣接関係と関係がない場合



(c) 全ての関係がある場合

図 5.3: 3つの物体間の関係

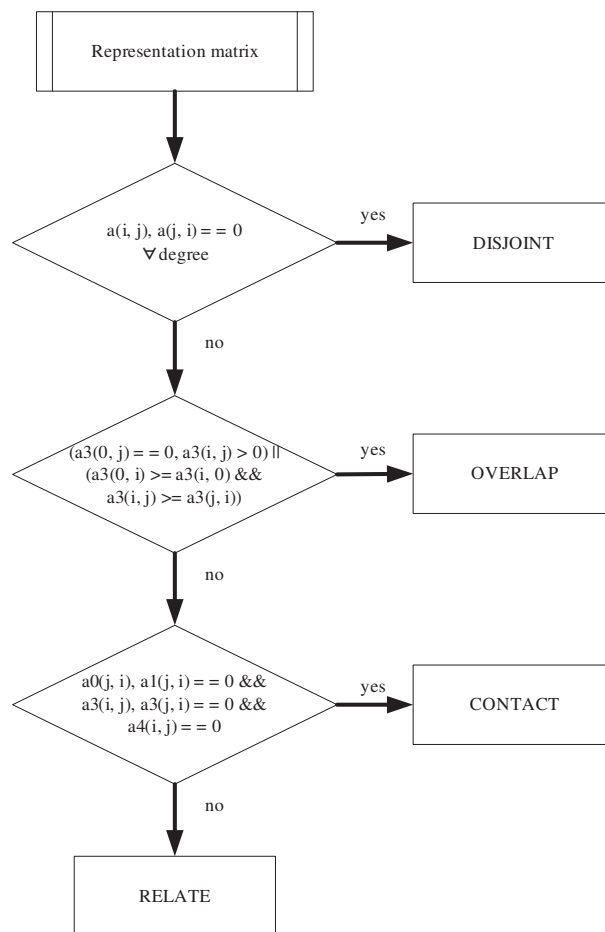


図 5.4: 接触情報導出フロー

— 接触情報の規則性 —

* 物体間の関係がない状態 (DISJOINT)

$$a_{i,j}^{\theta} == a_{j,i}^{\theta} == 0, \forall \theta$$

* 隣接関係 (CONTACT)

$$a_{j,i}^0 == a_{j,i}^{45} == 0 \wedge a_{i,j}^{90} == a_{j,i}^{90} == 0 \wedge a_{i,j}^{135} == 0$$

* 支持関係 (OVERLAP)

$$(a_{0,j}^{90} == 0 \wedge a_{i,j}^{90} > 0) \parallel (a_{i,j}^{90} > a_{j,i}^{90} \wedge a_{0,i}^{90} > a_{i,0}^{90})$$

* 何らかの関係あり (RELATE)

other

なお、「何らかの関係あり (RELATE)」はリジェクト領域のことである。クラスタ問題を考えるとき、どのクラスタに属するかという問題と、どのクラスタに属さないかという問題が存在する。認識精度の面を考慮すると、正確なクラスタに分類されなくてもリジェクト領域があることである程度、認識精度が補償される。

導出アルゴリズムの流れを図 5.4 に示す。まずはじめに、奥行き情報を用いて大まかな3次元位置での関係性があるか否かを判断する。本研究で類型化する3つの関係のうち「隣接関係 (CONTACT)」と「支持関係 (OVERLAP)」は、距離情報で説明する「近くにある (NEAR)」に属している必要がある。つまり「近くにある (NEAR)」と判断されたものののみ A_{11}^{θ} によって判断すればよい。

なお A_{11}^{θ} の各要素 $a_{i,j}^{\theta}$ は、接触情報があればあるほど大きな値となる。仮想的な環境であれば誤差などを考慮する必要はないが、実環境での動作を考えた場合、ノイズによる誤差を考慮する必要がある。さらに、個々の物体の大きさによって閾値の評価が異なってしまうので、次のように各要素を標準化することにした。

$$a_{i,j}^* = \max\left(\frac{a_{i,j}}{\sum_i a_{i,j}}, \frac{a_{i,j}}{\sum_j a_{i,j}}\right) \quad (5.5)$$

これにより、実環境の閾値を実験の結果、1.85 と定めた。

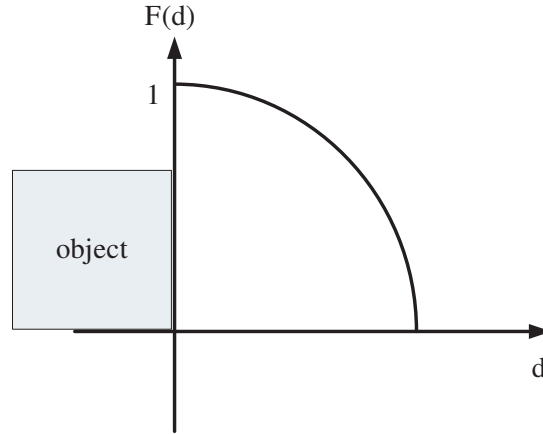


図 5.5: 「近くにある」の定義

5.2.2 距離情報

「近くにある (NEAR)」の確実性は、目的物体の boundary box からの距離の 2 乗に比例し、距離が 0 のときに最も確実性が高く、距離が大きいほど確実性が低くなる関数を利用して決定する (図 5.5). つまり、1 に近ければ近いほど前節で述べた「隣接関係 (CONTACT)」となる可能性が高くなる. この関数の比例定数、つまり距離に応じてどれだけ確実性が変化するかという値は物体の高さや横幅に依存すると考えられる. 通常、物体の高さや横幅が大きければ大きいほど確実性の高い範囲が広くなると考えられるので、比例係数は小さくなる. よって「近くにある (NEAR)」の確実性を示す関数を次のように定義する. ただし、 R_h は物体の高さ、 d は基準物体から目的物体までの距離を示している.

$$F(d) = -\alpha \left(\frac{d}{R_h} \right)^2 + 1 \quad (5.6)$$

5.2.3 方向情報

基準物体に対する目的物体の方向情報は、目的鏡映軸で表現する. 基準物体と目的物体の位置はステレオ法により測定できる. 図 5.6 のように各パラメータを求めると、

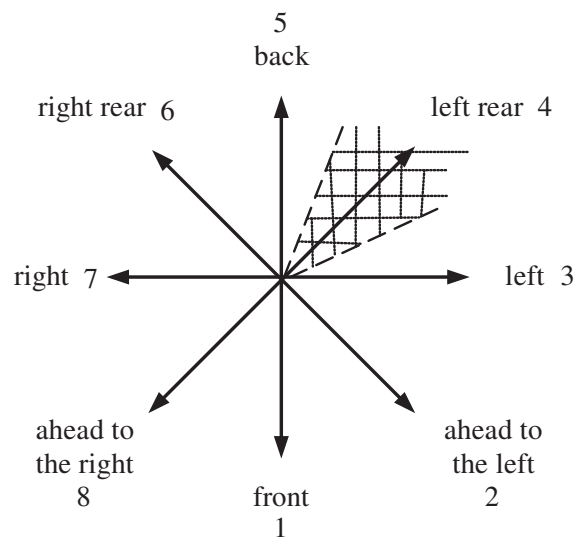
$$X = X_{\text{目的物体}} - X_{\text{基準物体}} \quad (5.7)$$

$$D = D_{\text{目的物体}} - D_{\text{基準物体}} \quad (5.8)$$

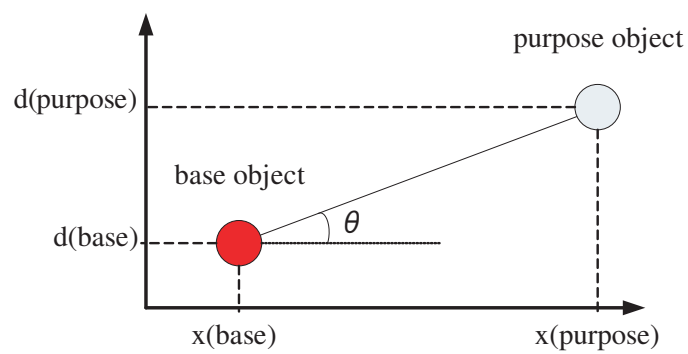
$$\theta = \arctan \frac{|X|}{|D|} \quad (5.9)$$

となる. よく用いられている 4 方向に対し、本研究では 8 方向で方向情報を表現する. そのため、 θ を 8 分割し、各方向に対し、 $\pm 22.5\text{deg}$ の幅を持たせた. これによって、相

対的な方向情報「前」,「後」,「左」,「右」,「右手前」,「左手前」,「右後」,「左後」の8方向を決定できる.



(a) 相対的な方向情報



(b) 方向情報抽出

図 5.6: 方向情報の定義

第6章 物体間の関係推論方法

前章では、画像に映し出されている物体間の関係を導出する方法について述べた。本章では、自律的に知識を蓄積する関係推論方法を説明する。6.1節では、人間の慣れや動物の逃避行動を例に挙げ、馴化脱馴化について述べる。6.2節では、馴化評価モデルの定義と意味について述べる。

6.1 馴化脱馴化パラドクス

人間は生まれながらすべての知識を持っているわけではなく、学習(強化学習, 試行錯誤学習, 洞察学習)して知識を得ていく。これは先見的知識に基づいて外見を予測し、感覚系を通して取得した観察との一致度を高めるように知識を更新していく。このプロセスは、外界が間違いのない情報を観測系に提示することを”知って”おり、更新途中であるにも関わらず、予測が有効になされていることを”確信”している。すなわち、知覚循環系は、外界、先見的知識ならびに予測方法に対する信頼に基づいて、認識と行動を並行化する能動的システムである。

人間、特に乳幼児は新しい刺激、珍しい刺激を与えると注意を払う。しかし、同じ刺激が与えられ続けると人間は、徐々にその刺激に慣れ、そのうちその対象物に飽きて、別の刺激に注意を払うようになる。この慣れの現象を馴化という。これは、動物の日常生活において意味のない反応を排除しようとする生物の環境への適応の好例であり、繰り返された刺激への応答という経験を積んだことで行動に変化が生じる。つまり最も単純な学習の一形態といえる。また、一度馴化が成立すると刺激停止後も比較的長期にわたって保持されている場合が多く、記憶の一側面といった面も合わせ持っている。例えば、動物の逃避行動はその一例である。北海道大学の長山氏 [35] [36] は、アメリカザリガニの逃避行動(図 6.1)について研究しており、その逃避行動の馴化曲線を図 6.2 に示す。



図 6.1: アメリカザリガニの逃避行動

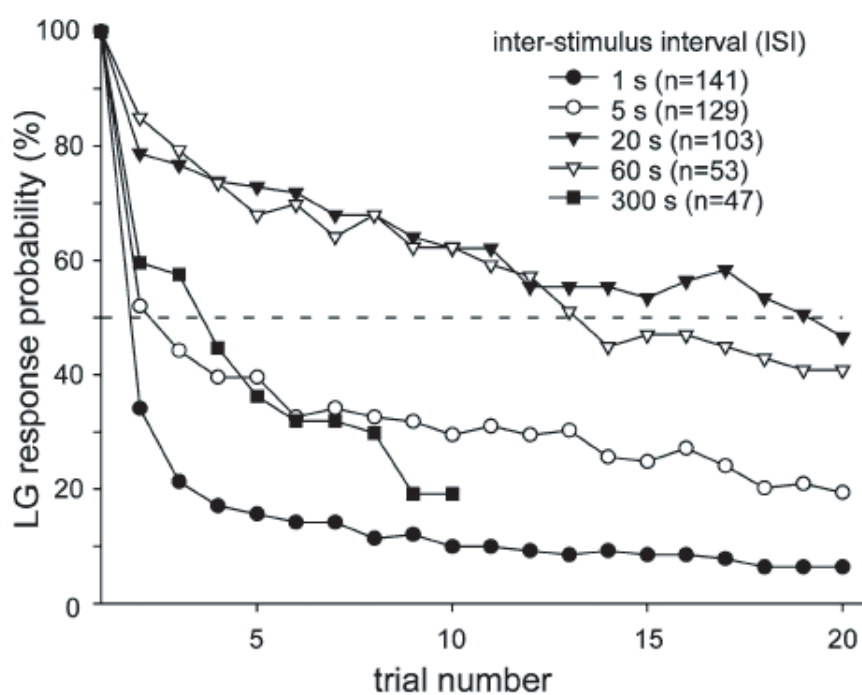
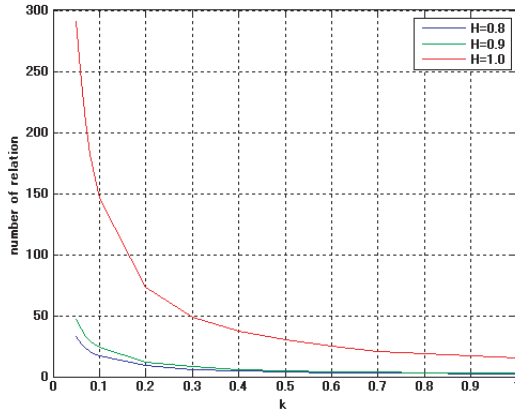
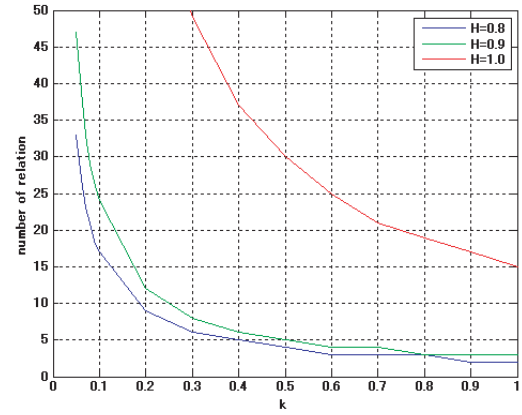


図 6.2: 逃避行動の馴化曲線



(a)



(b)

図 6.3: κ と成立数の関係

6.2 馴化評価モデル

ロボットが環境中の未知の物体を扱うためには自ら物体と相互作用を行い、自律的に物体間の関係を獲得していく必要がある。そこで、本研究ではロボットが時間変化に応じて自律的に学習を進めていくために、前節で述べた動物の逃避曲線に基づいて馴化評価モデルを定義する。

$$H = 1 - \alpha \exp(-\kappa x_{j,A,B}) \quad (6.1)$$

ここで、 α, κ は定数、 j はトポロジー関係の種類、 A, B は物体を示す。つまり $x_{j,A,B}$ は、「 A と B の関係は j である」という意味を持っている。この関数 H は、0から1の値を取り、その情報が正しいほど1に近い値となる。さらに、入力が少ないときは大きく変化するが、大きくなるにつれて変化は小さくなる。これは、馴化曲線をよく表した関数であり、これによって、物体間の関係、さらにロボットが行うべき行動の取捨選択ができるようになると思う。

図 6.3 に κ と関係の数の関係を示した。馴化評価モデル H が 1, 0.9, 0.8 となる場合の結果を示した。 κ の設定は使用する状況に応じて変更する必要がある。物体間の関係が頻繁に変化する場合、多くの変動から物体間の関係を推論するため、 κ を小さい値にする必要がある。これによってより信頼性のある関係推論が実現できる。逆に、物体間の関係に変化があまりない場合、これは得た情報に信頼性があるということである。ただし、推論完了するまでの速度はトレードオフの関係がある。なお、本研究では机上空間および物体間の関係の変動は少ないことを想定しているため、 $\kappa = 0.5$ と設定した。

第7章 評価実験および考察

これまでの章では、本研究で提案した手法などについて述べてきた。本章では、提案した手法の評価を行う。8.1節では実験の概要を述べる。8.2節ではトポロジー関係導出アルゴリズムの有用性を示す。8.3節ではシステム全体の評価として、馴化評価モデルによる配置ルール導出について述べる。

7.1 実験内容

定式化したトポロジー関係を実装し、システムが実環境内の物体の関係を表現する実験を行った。実験は大きく、(1) トポロジー関係の判定アルゴリズム、(2) システム全体の2つの評価実験に分けられる。(1)については、トポロジー関係である「接触情報」、「距離情報」、「方向情報」の3通りについてそれぞれ別々に実験を行った。実験は静止画キャプチャによって行った。(2)については、本研究が提案した手法をすべてまとめた評価である。動画キャプチャから得られた画像をもとに奥行き画像や動きベクトルなどを作成し、各トポロジー関係を抽出できているか、さらに知識となる物体間の配置ルールが導出されているかを評価した。

7.2 トポロジー関係導出アルゴリズムの評価

7.2.1 接触情報

本研究が提案した手法では、1枚の画像に n 個の物体が存在すれば、シーン中には $n*(n-1)$ つの接触情報が得られる。そこで、接触情報の評価として個別評価と総合評価の2種類によって評価した。まず個別評価では、1枚の画像から得られる $n*(n-1)$ つの接触情報を1つ1つの情報に分割し、それぞれが正しく抽出されているかどうかを評価した。総合評価では、1枚の画像から得られる複数の接触情報を1つの情報とし、それが正しく抽出されているかどうかを評価した。よって個別評価におけるOVERLAP, CONTACT, DISJOINTの各サンプル数は、123, 41, 268であり、総合評価のサンプル数は182である。

個別評価の結果を表7.1に示す。単位は[%]である。DISJOINTの場合、100%の接触情報を得られているのに対し、OVERLAPやCONTACTの場合、約80%である。その理由として、実画像を用いることでノイズが発生し、個々の物体のboundary boxを正確

表 7.1: (a) 個別評価: 単位 [%]

	correct	incorrect	reject
OVERLAP	81.3	6.5	12.2
CONTACT	75.6	22.0	2.4
DISJOINT	100	0	0

に定めることができないときがある。DISJOINT の場合, boundary box は明らかに個々の物体は離れたところにあるためノイズが発生しても誤差を吸収することができる。しかし, CONTACT や OVERLAP は個々の物体が接触しているために boundary box を定めるときにノイズが大きく影響していると思われる。本研究では, 個々の物体の認識には注目していないが, より正確な情報を得るためには認識手法に関して検討する必要がある。総合評価は, サンプル数 182 に対し, 139 枚の画像においてシーンの中にある全ての物体間の関係を抽出できていた。よって 76.4 % の認識率であった。2 つの評価から, ノイズが影響しているものの 80 % 前後の信頼性があり, 空間内の物体配置 (接触情報) を認識できているため, 有用性があると言える。

7.2.2 距離情報

距離情報の評価方法として, 図 7.1 に示すように, 基準物体を固定し, 目的物体を $d=5 \sim 15\text{cm}$ と移動させたときのそれぞれの距離情報を表示させ, 実際の配置と一致しているかどうか評価した。「理論」は実際に配置した距離情報, 「実測」はシステムの出力である。

表 7.2, 7.3 が示すように距離情報の表現方法は, 理論値と実測値に誤差は含まれているものの, 想定した距離範囲内であれば「近くにある」と判断できていることがわかる。なお, 評価実験において $\alpha = 0.3$ と設定した。理由として我々人間が「近くにある」と認識している範囲を考えると, 自分の身長約 1.3 倍の距離に相当する。本研究で定義した「近くにある」の定式化では, 物体の大きさに依存するため, $\alpha = 0.3$ とすることで物体間の距離情報を表現できると考えたためである。

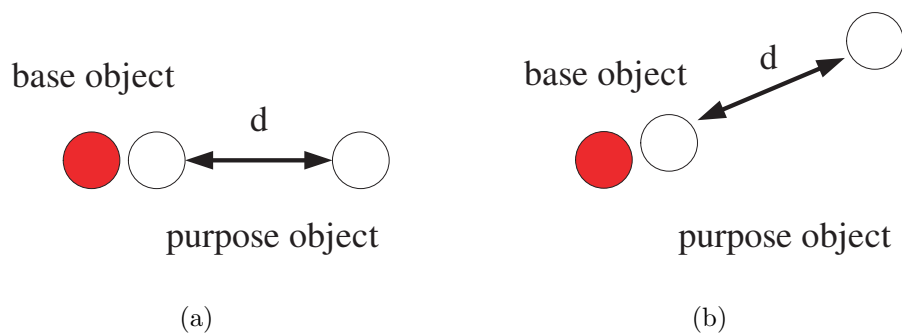


図 7.1: 距離情報の評価方法

表 7.2: 図 7.2.2(a) の場合

距離	確実性 (理論)	確実性 (実測)	関係	判定
5	0.925	0.900	contact	○
6	0.892	0.853	disjoint	○
7	0.853	0.853	disjoint	○
8	0.808	0.853	disjoint	○
9	0.757	0.757	disjoint	○
10	0.700	0.700	disjoint	○
11	0.637	0.568	disjoint	○
12	0.568	0.568	disjoint	○
13	0.493	0.490	disjoint_{th}	○
14	0.412	0.412	disjoint_{th}	○
15	0.325	0.325	disjoint_{th}	○

表 7.3: 図 7.2.2(b) の場合

距離	確実性 (理論)	確実性 (実測)	関係	判定
5	0.925	0.853	contact	○
6	0.892	0.850	disjoint	○
7	0.853	0.850	disjoint	○
8	0.808	0.757	disjoint	○
9	0.757	0.730	disjoint	○
10	0.700	0.688	disjoint	○
11	0.637	0.634	disjoint	○
12	0.568	0.556	disjoint	○
13	0.493	0.556	disjoint, disjoint _{th}	△
14	0.412	0.481	disjoint _{th}	○
15	0.325	0.400	disjoint _{th}	○

7.2.3 方向情報

方向情報の評価方法として、図 7.2 に示すように、基準物体を固定し、目的物体を $x=10 \sim 20\text{cm}$, $y=0 \sim 60\text{cm}$ と移動させたときのそれぞれの方向情報を表示させ、実際の配置と一致しているかどうか評価した。表 7.4, 7.5 で表示されている「理論」は実際に配置した距離関係から求めた方向情報であり、「実測」はシステムの出力で表示された方向情報である。

図 5.6(a) の 6(右後) $\rightarrow$ 5(後), 4(左後) $\rightarrow$ 5(後), 2(右手前) $\rightarrow$ 1(前), 8(左手前) $\rightarrow$ 1(前)に関わるところで誤認識している割合が高かった。原因として考えられることは、基準物体と目的物体との距離が離れているために誤差が生じてしまっているものと考えられる。さらに本手法の場合、例えば 4(左後) $\rightarrow$ 5(後) の場合を考えると、4(左後) と判断される θ の範囲は $22.5 < \theta \leq 67.5$ であり、67.5 以上になれば 5(後) と判断されてしまう。よってこの基準の選定によって生じた誤差による誤認識であると考えられる。しかし、全く違う方向情報を示すことはなかったため、問題はないといえる。表 7.6 に 2 種類の評価方法の精度をまとめた。情報は 80 % の信頼性があるため、空間内の物体配置 (方向情報) を認識できているといえる。

さらに、物体を机上空間にランダムに配置し、物体間の方向情報から 3D 画像を作成した結果を図 7.3 に示す。これは、画像から得られた関係表現行列および言語をもとに個々の局所関係を集めて、机上空間全体の関係を推測した結果である。実際に机上空間に物体を配置したように 3D 画像が得られている。これによって個々の局所関係を集めて、全体の関係を推測することで、大局的な状況を明らかにできることが分かる。

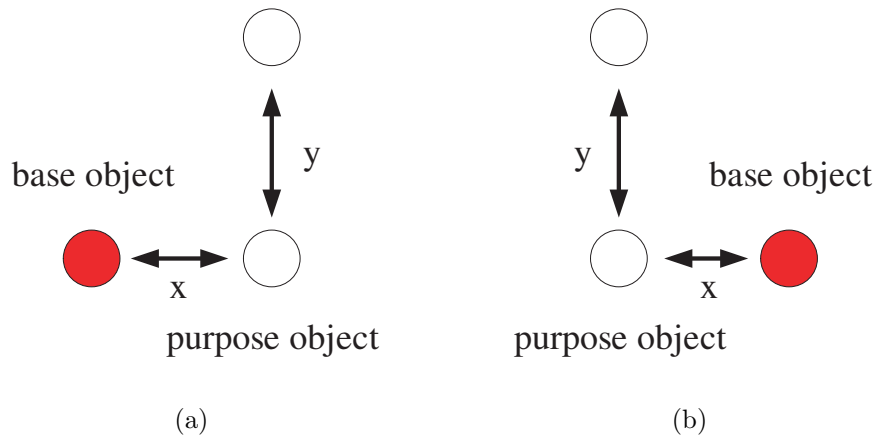


図 7.2: 方向情報の評価方法

表 7.4: 図 7.2.3(a) の場合

奥行き距離	理論 10	実測 10	判定	理論 20	実測 20	判定
0	(3,7)	(3,7)	○	(3,7)	(3,7)	○
10	(4,8)	(4,8)	○	(4,8)	(4,8)	○
20	(4,8)	(5,8)	△	(4,8)	(4,8)	○
30	(5,1)	(5,1)	○	(4,8)	(4,8)	○
40	(5,1)	(5,1)	○	(4,8)	(5,8)	△
50	(5,1)	-	-	(5,1)	(5,1)	○
60	(5,1)	-	-	(5,1)	(5,1)	○

表 7.5: 図 7.2.3(b) の場合

奥行き距離	理論 10	実測 10	判定	理論 20	実測 20	判定
0	(3,7)	(3,7)	○	(3,7)	(3,7)	○
10	(2,6)	(2,6)	○	(2,6)	(2,6)	○
20	(2,6)	(2,6)	○	(2,6)	(2,6)	○
30	(1,5)	(2,5)	△	(2,6)	(2,6)	○
40	(1,5)	(1,5)	○	(2,6)	(2,6)	○
50	(1,5)	(1,5)	○	(1,5)	(2,5)	△
60	(1,5)	-	-	(1,5)	(2,5)	△

表 7.6: 方向情報の評価まとめ: 単位 [%]

	図 7.2.3(a)	図 7.2.3(b)
10	80	83
20	86	71.4

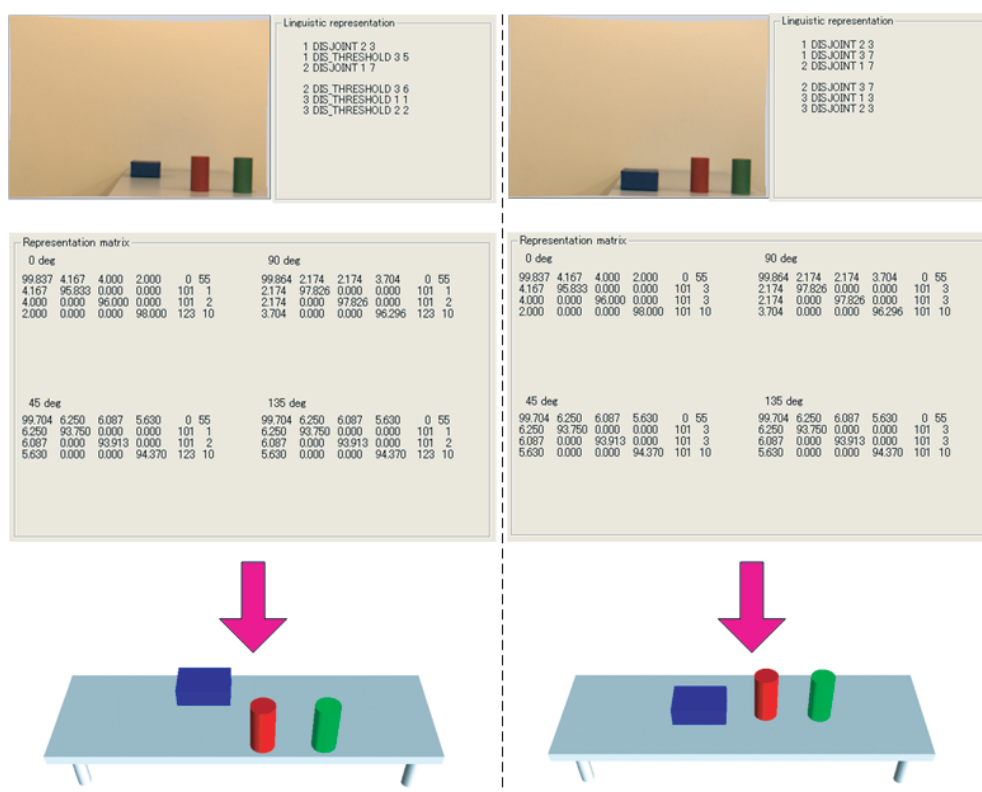


図 7.3: 机上空間での方向情報と 3D 画像

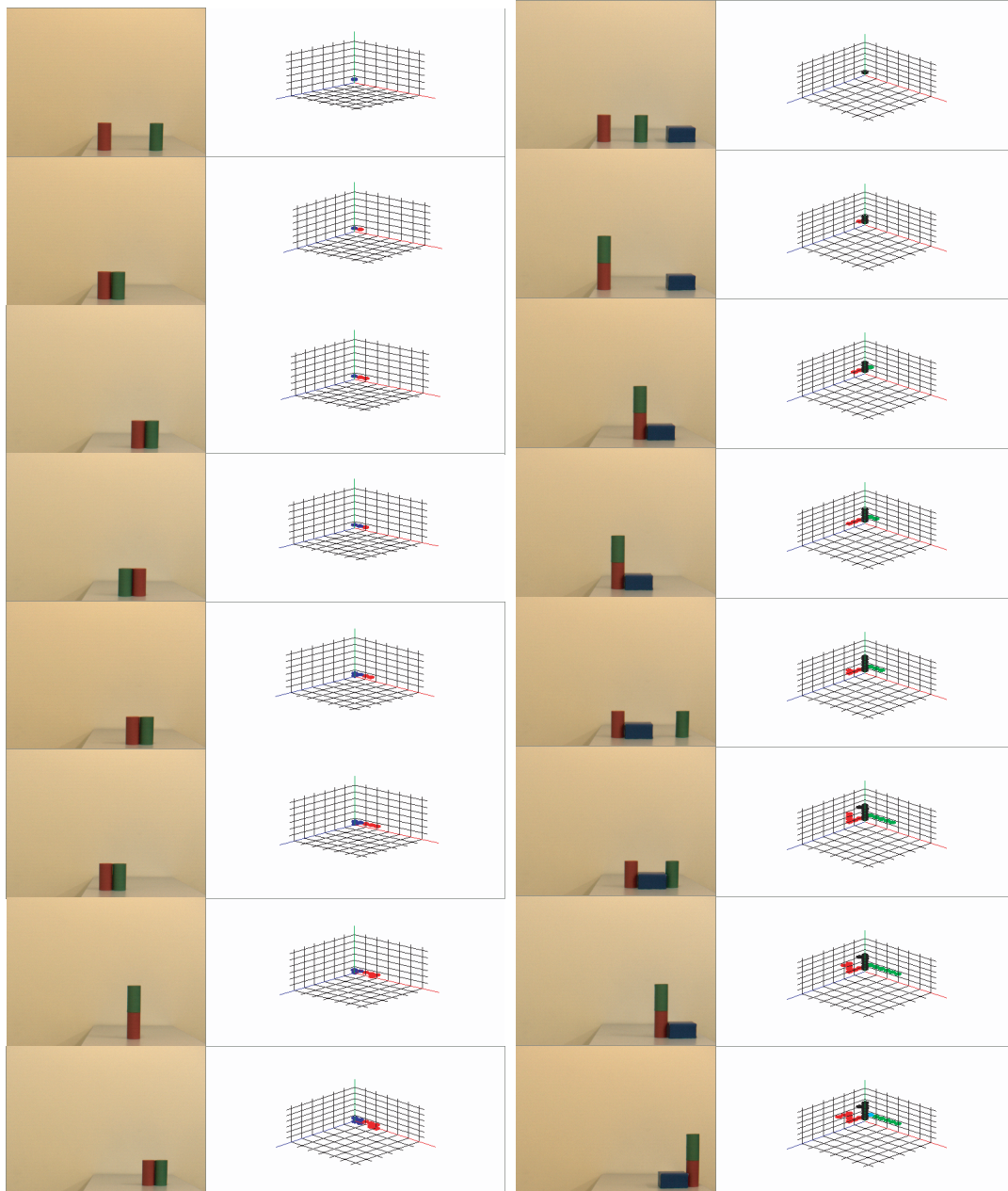
7.3 システム全体の評価

これまでの前節で各トポロジー関係が静止画で正確に抽出されることを示した。この節では、システム全体の評価として動画にてリアルタイムに実用できるかどうか、さらに効率的に知識蓄積されているかどうかを検証する。実験環境として、机上空間において人間が物体を移動させる風景をシステムで撮影する。eventが発生すると物体間の各トポロジー関係を自動的に導出し、馴化評価モデルにより関係の信頼性を導出する。ここで、eventとは動きベクトルがゼロになったことを示している。

まず動画にて評価を行う前に、物体を動かした後の画像、つまり event が発生したと仮定した画像を数枚用意し、順にシステムに入力し、動作を確認した。図 7.4(a) の左側は入力した画像である。画像を入力するたびに推論グラフが変化している。推論グラフは個々の物体を基準物体としたとき、目的物体との関係をグラフィック化したものである。赤、青、緑の各軸はそれぞれ接触情報の隣接関係、支持関係、関係なしを示している。さらに、図 7.5(a)(b) は馴化評価モデルの event ごとの変化を表している。これらのことから、個々の局所関係を集めて全体の関係を推定することで、図 7.5(c) のような大局的な状況を明らかにすることができている。同様にして図 7.4(b)、図 7.6 に物体の数を 3 つにした場合を示す。

静止画によって物体間の配置ルールを導出できたので、次にリアルタイムに実用できるか、動画によって評価する。図 7.7～図 7.11 はそのときの動画シーケンスである。右から順に、左画像、動きベクトル、奥行き画像、推論グラフである。図を見て明らかなように、左画像と動きベクトルは常に取得しており、eventが発生すると奥行き画像と推論グラフが更新されている。図 7.12(a)(b) は馴化評価モデルの event ごとの変化を示している。これらのことから、個々の局所関係を集めて全体の関係を推定することで、図 7.12(c) のような大局的な状況を明らかにすることができている。同様にして図 7.13～図 7.21 に物体の数を 3 つにした場合を示す。

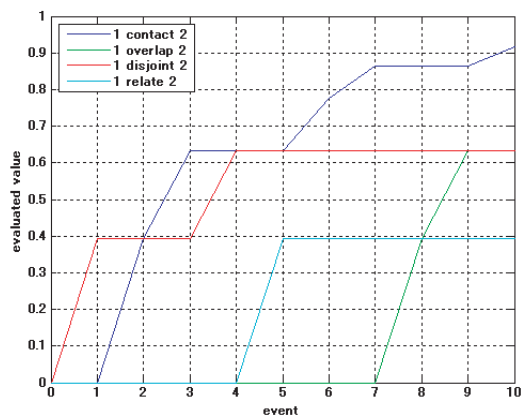
以上のことから、リアルタイムでも実用できること、さらに馴化評価モデルによって物体間の配置ルールを推論できていることを明らかにした。



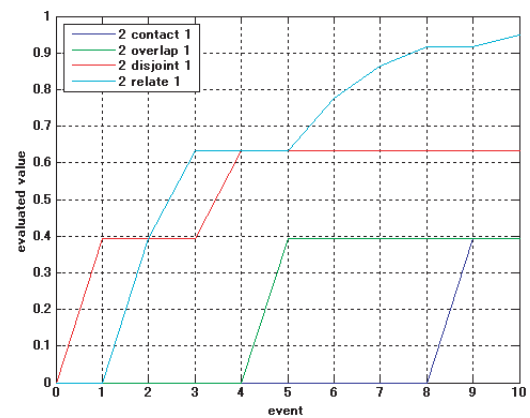
(a) 物体が 2 つ

(b) 物体が 3 つ

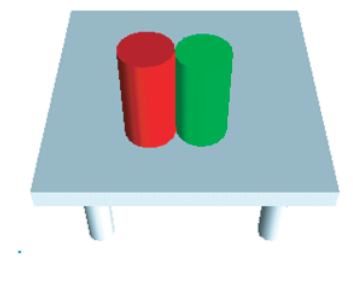
図 7.4: event が発生したときの主な動画シーケンス



(a) 物体 (赤) を基準とした場合

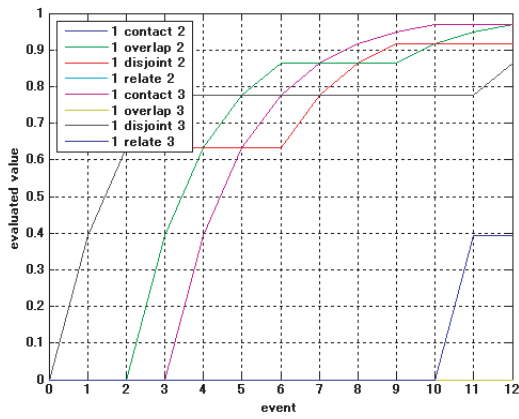


(b) 物体 (緑) を基準とした場合

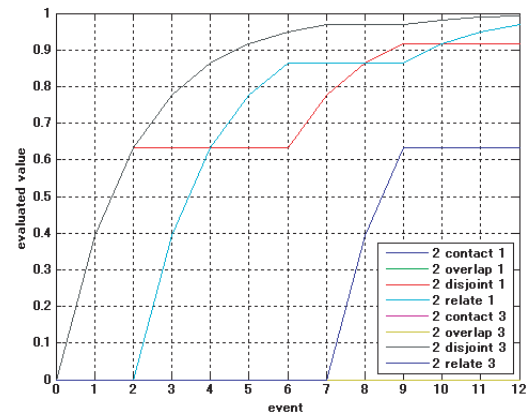


(c) 導出された物体の配置
ルール

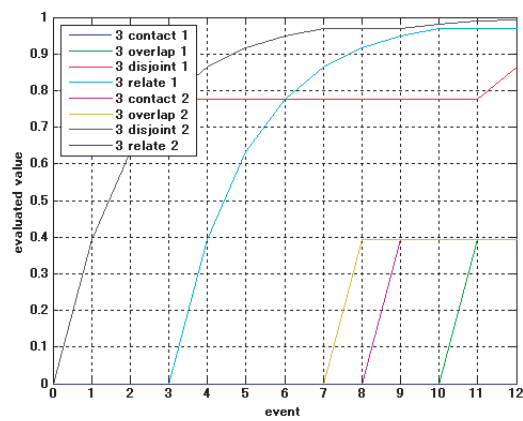
図 7.5: 馴化評価モデルの event ごとの変動と関係推論



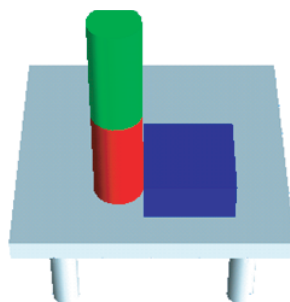
(a) 物体 (赤) を基準とした場合



(b) 物体 (緑) を基準とした場合



(c) 物体 (青) を基準とした場合



(d) 導出された物体の
配置ルール

図 7.6: 馴化評価モデルの event ごとの変動と関係推論

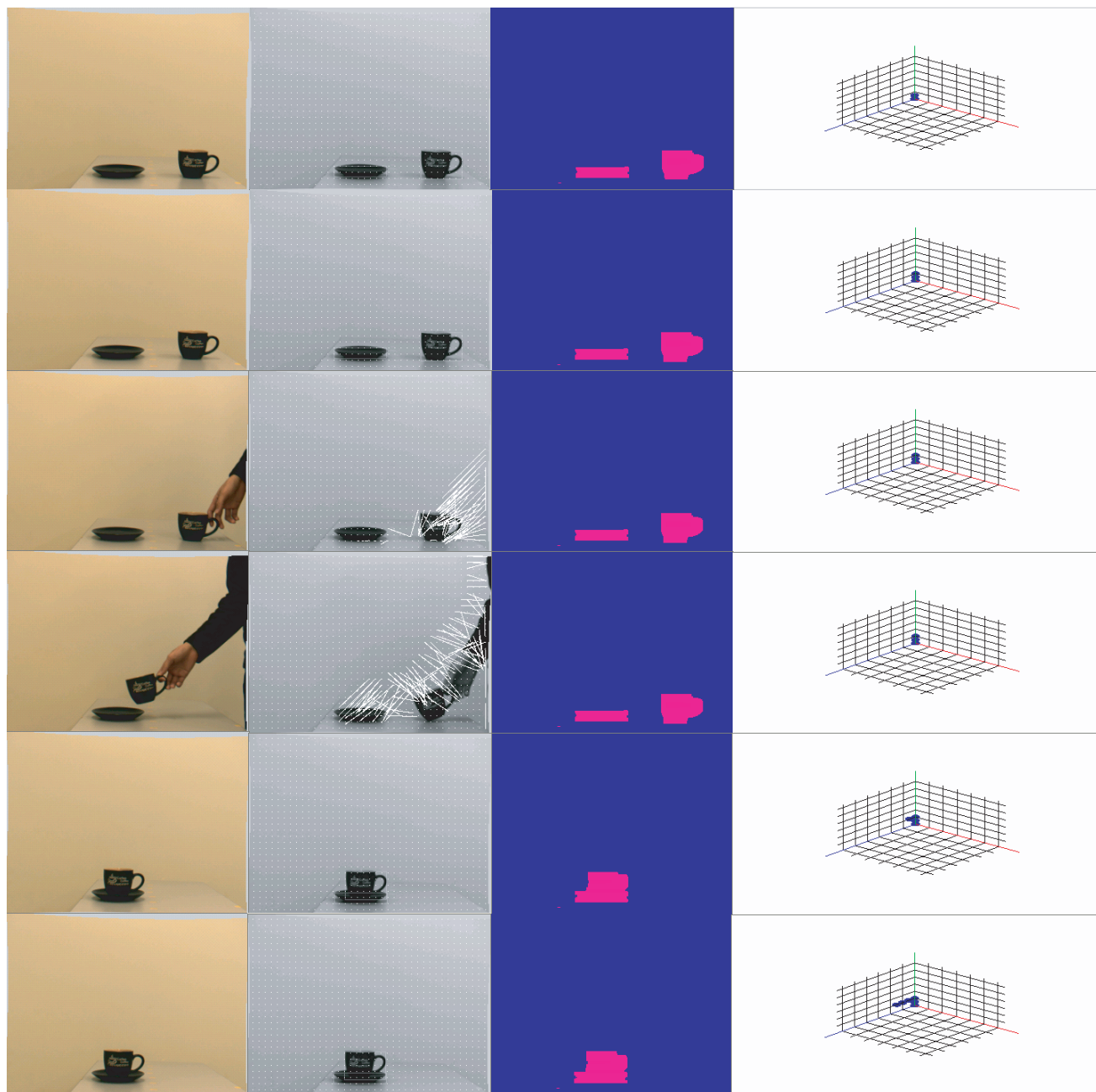


図 7.7: 動画シーケンス (1/5): 上から 0, 5, 8, 10, 19, 25 sec

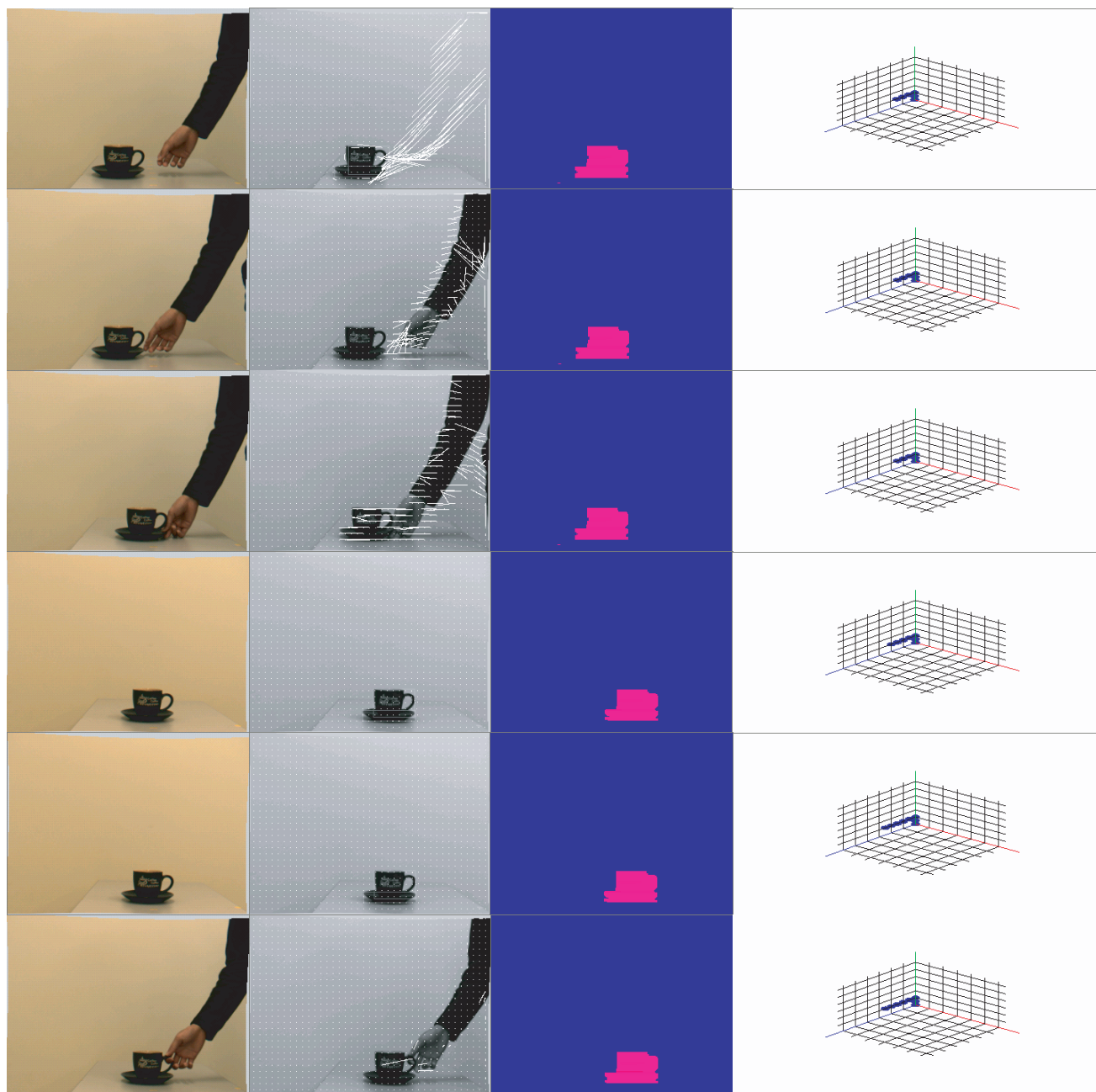


図 7.8: 動画シーケンス (2/5): 上から 34, 36, 38, 43, 50, 54 sec

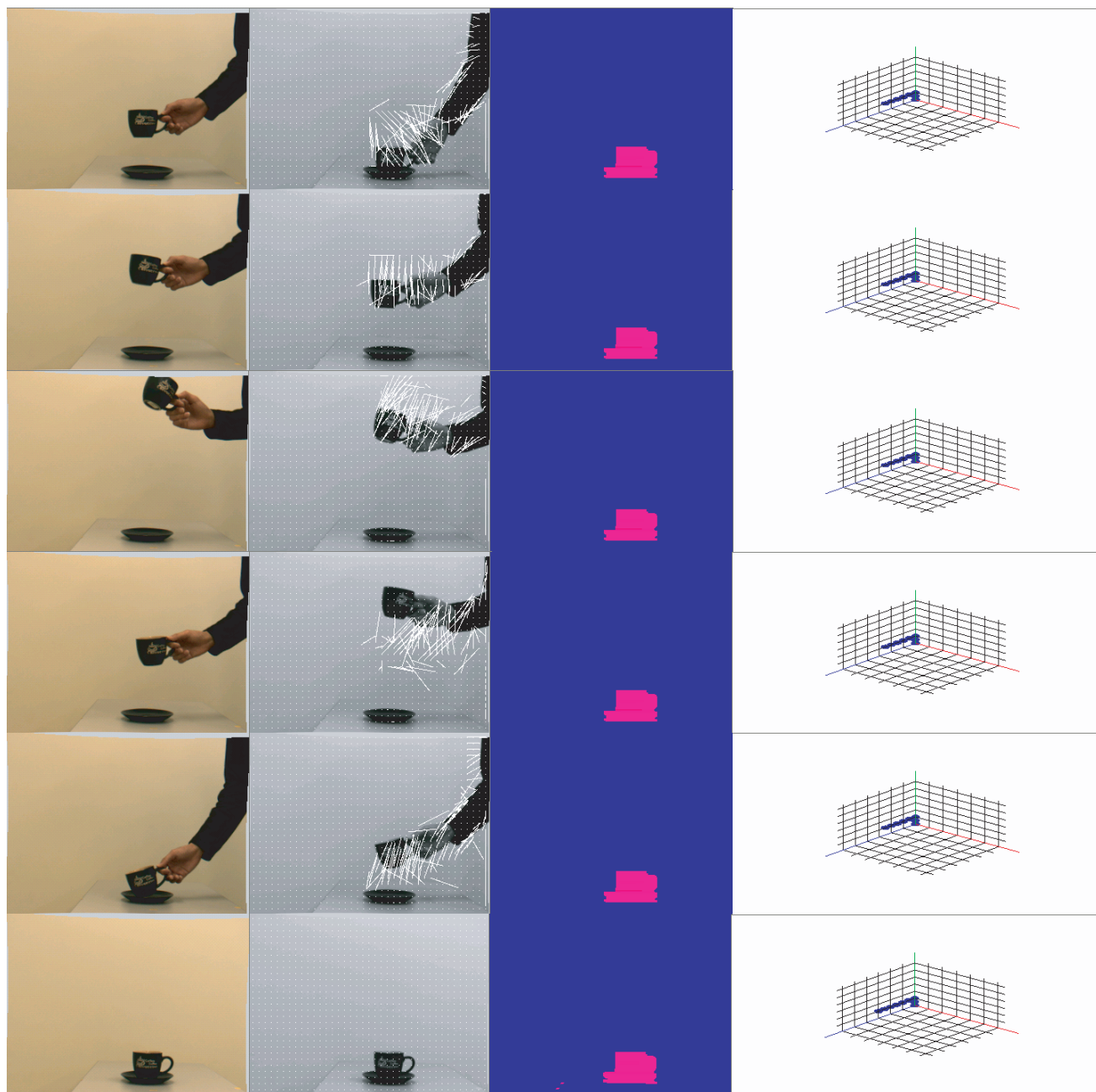


図 7.9: 動画シーケンス (3/5): 上から 58, 61, 64, 70, 73, 80 sec

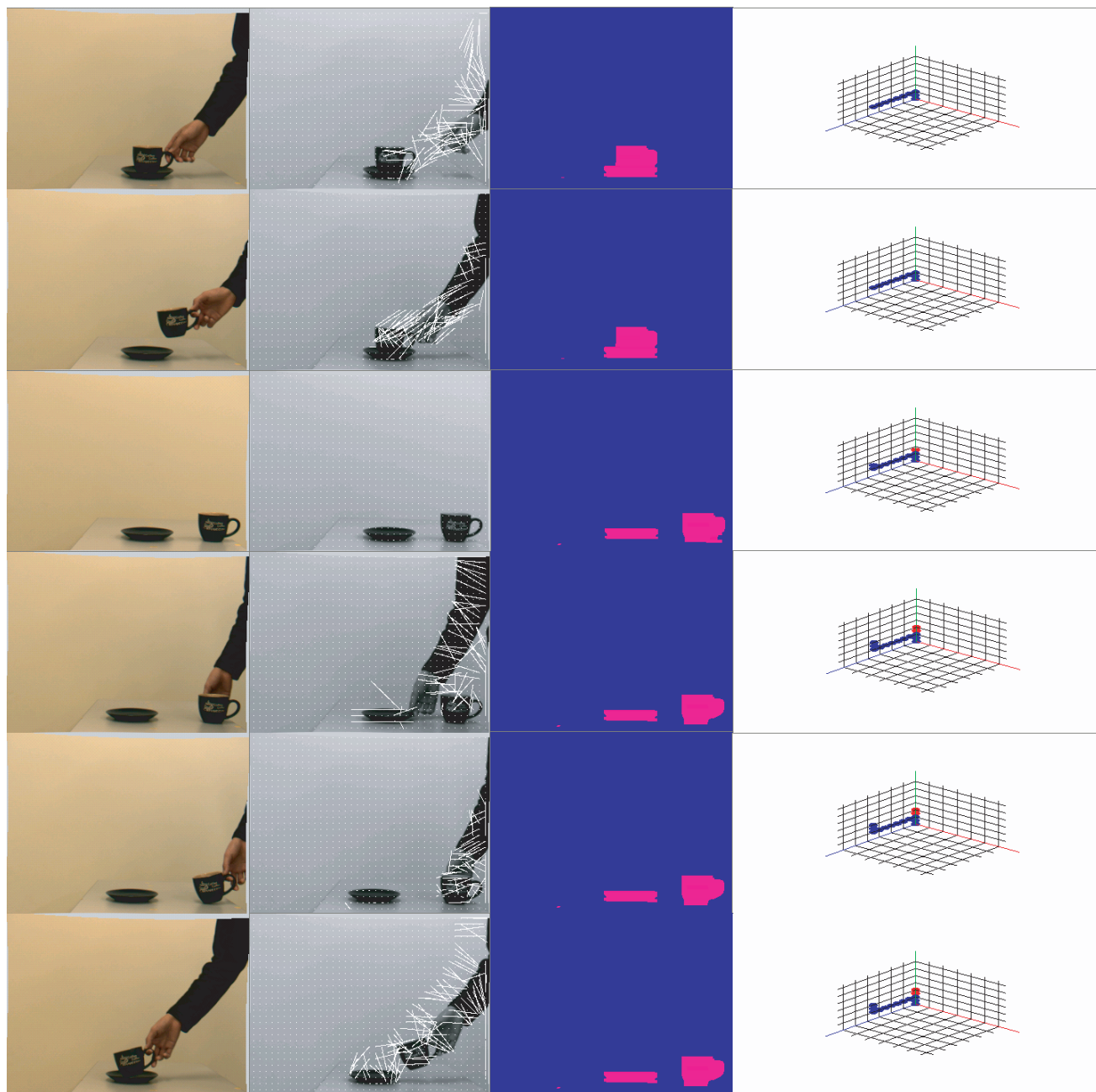


図 7.10: 動画シーケンス (4/5): 上から 93, 95, 97, 107, 108, 112 sec

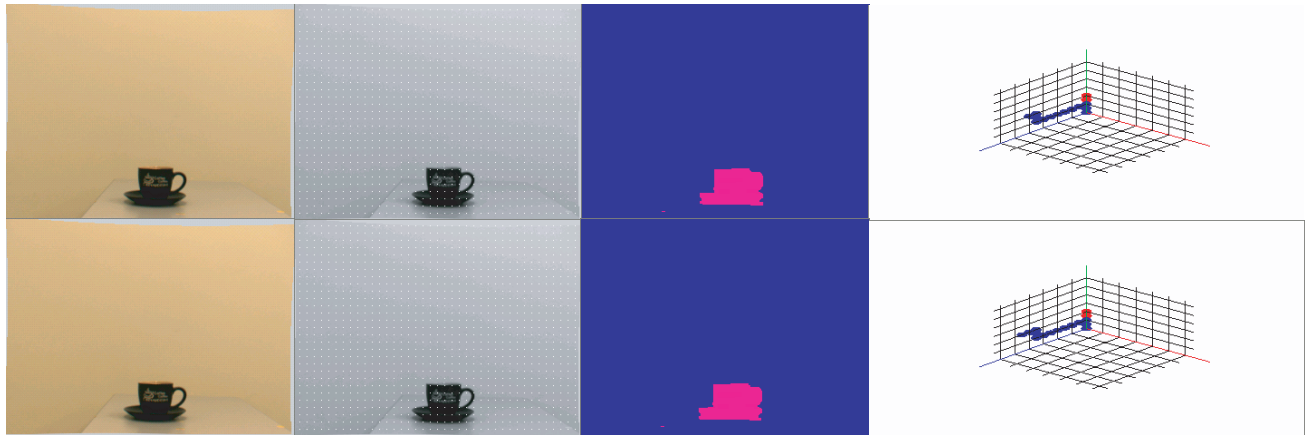
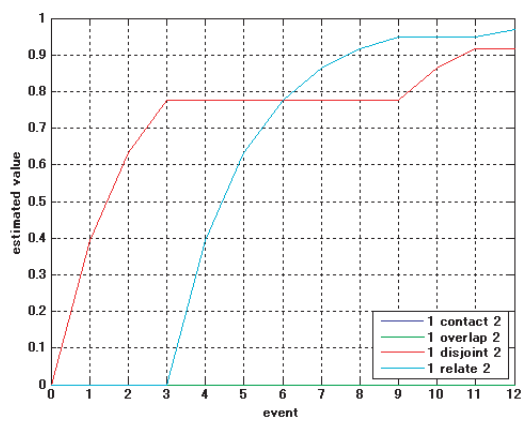
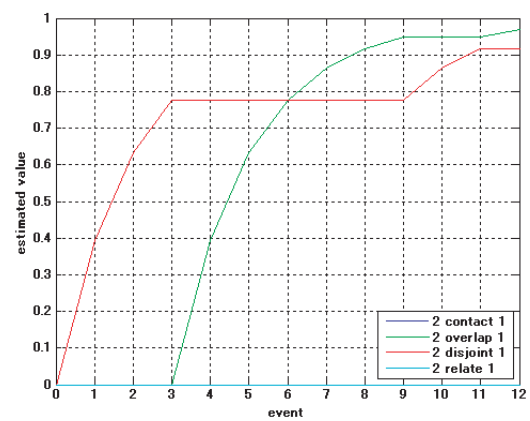


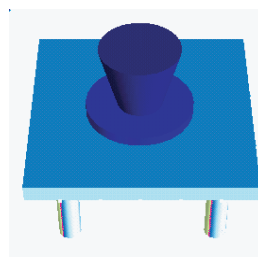
図 7.11: 動画シーケンス (5/5): 上から 120, 123 sec



(a) カップを基準とした場合



(b) コースタを基準とした場合



(c) 導出された物
体配置ルール

図 7.12: 馴化評価モデルの event ごとの変動と関係推論

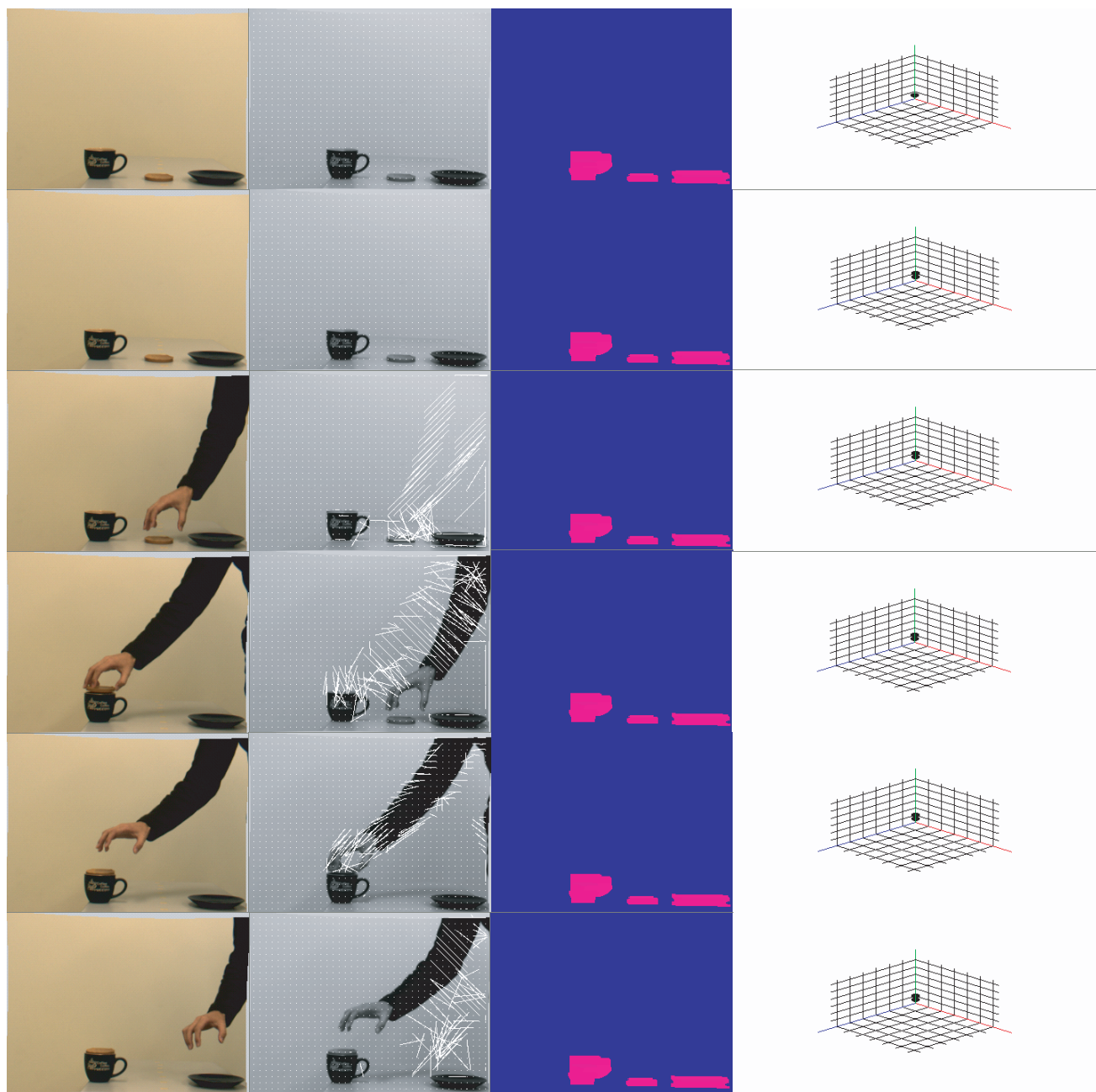


図 7.13: 動画シーケンス (1/8): 上から 0, 5, 7, 10, 14, 16 sec

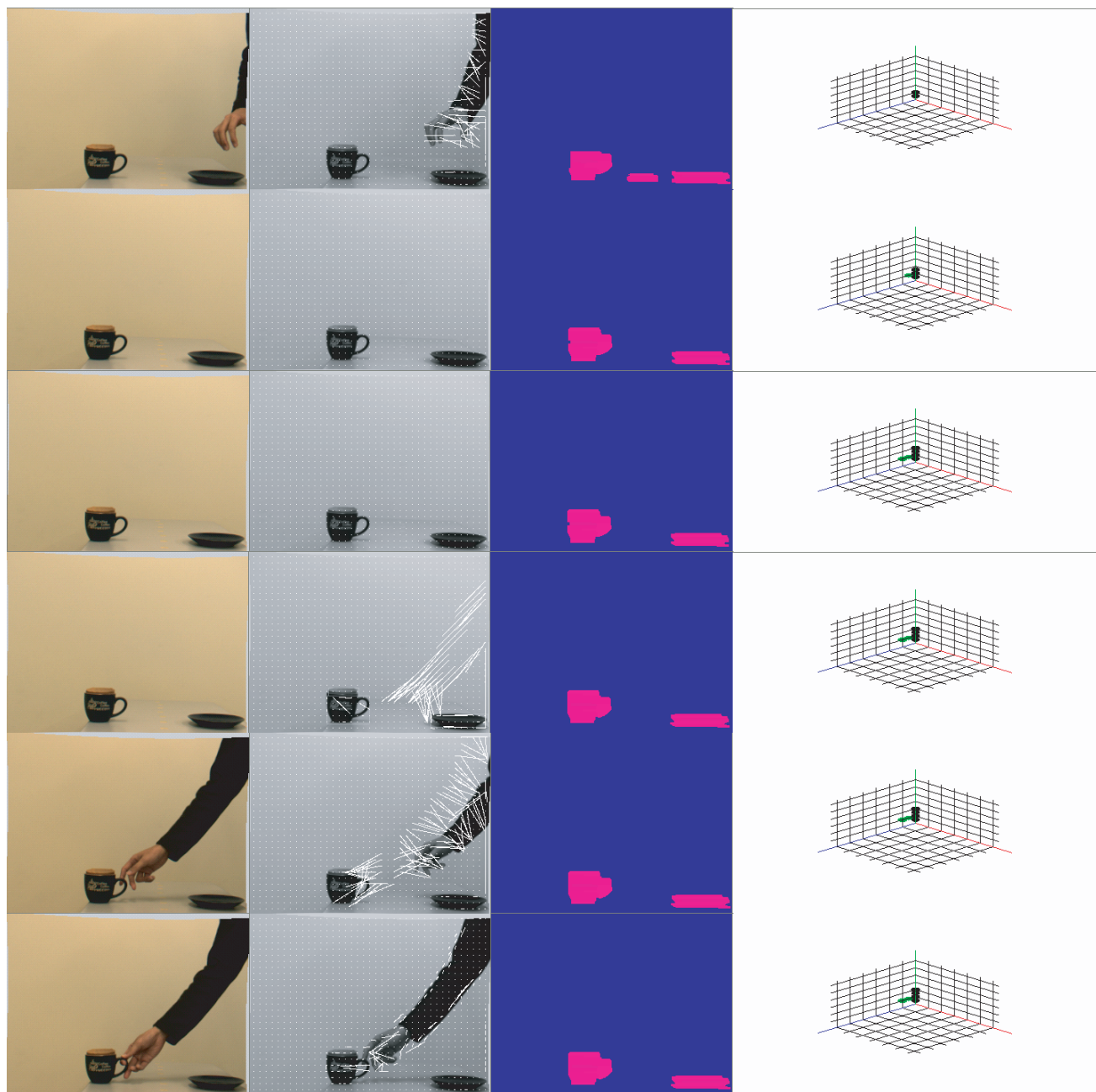


図 7.14: 動画シーケンス (2/8): 上から 19, 27, 32, 34, 36, 40 sec

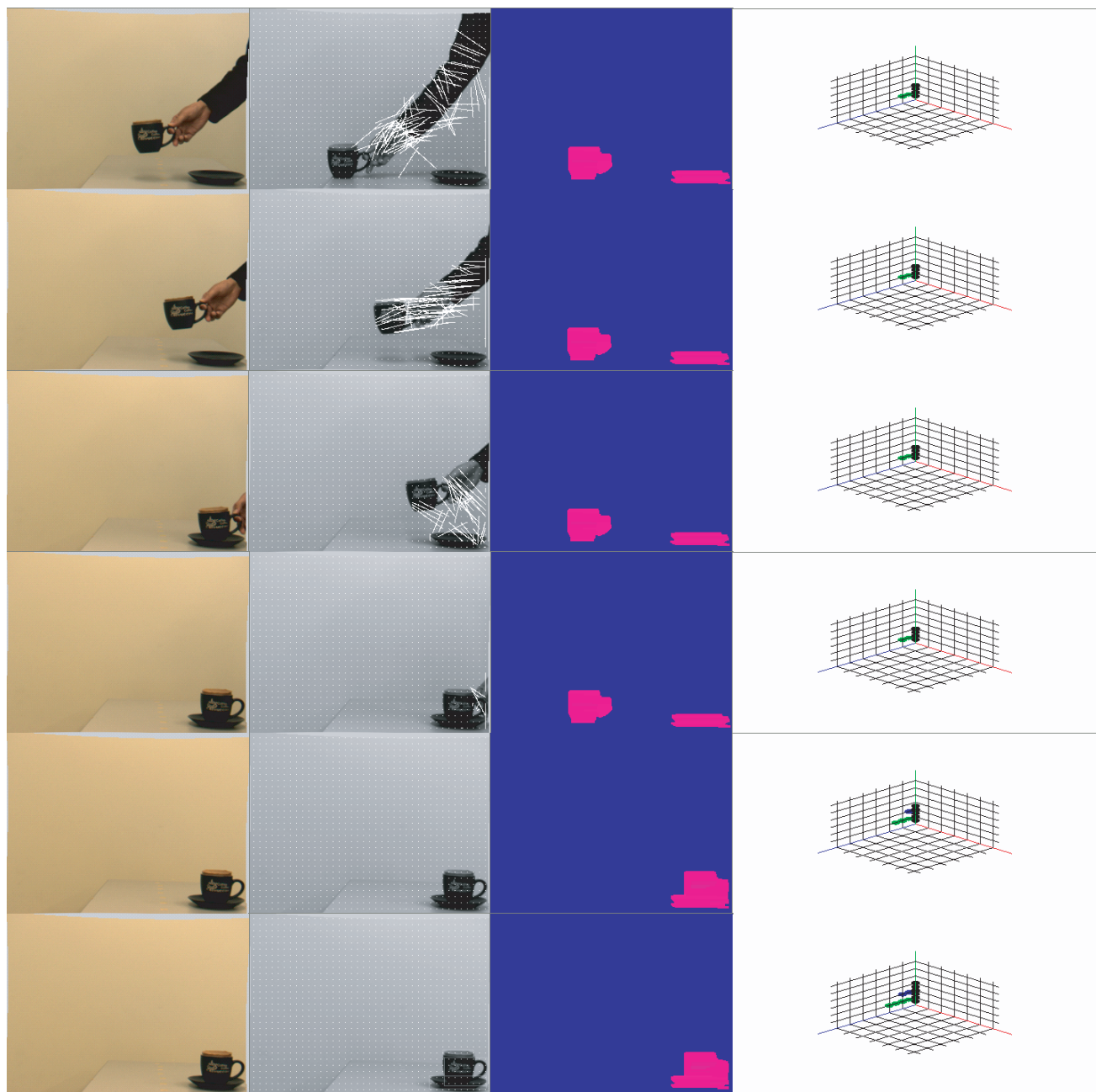


図 7.15: 動画シーケンス (3/8): 上から 43, 44, 46, 49, 56, 60 sec

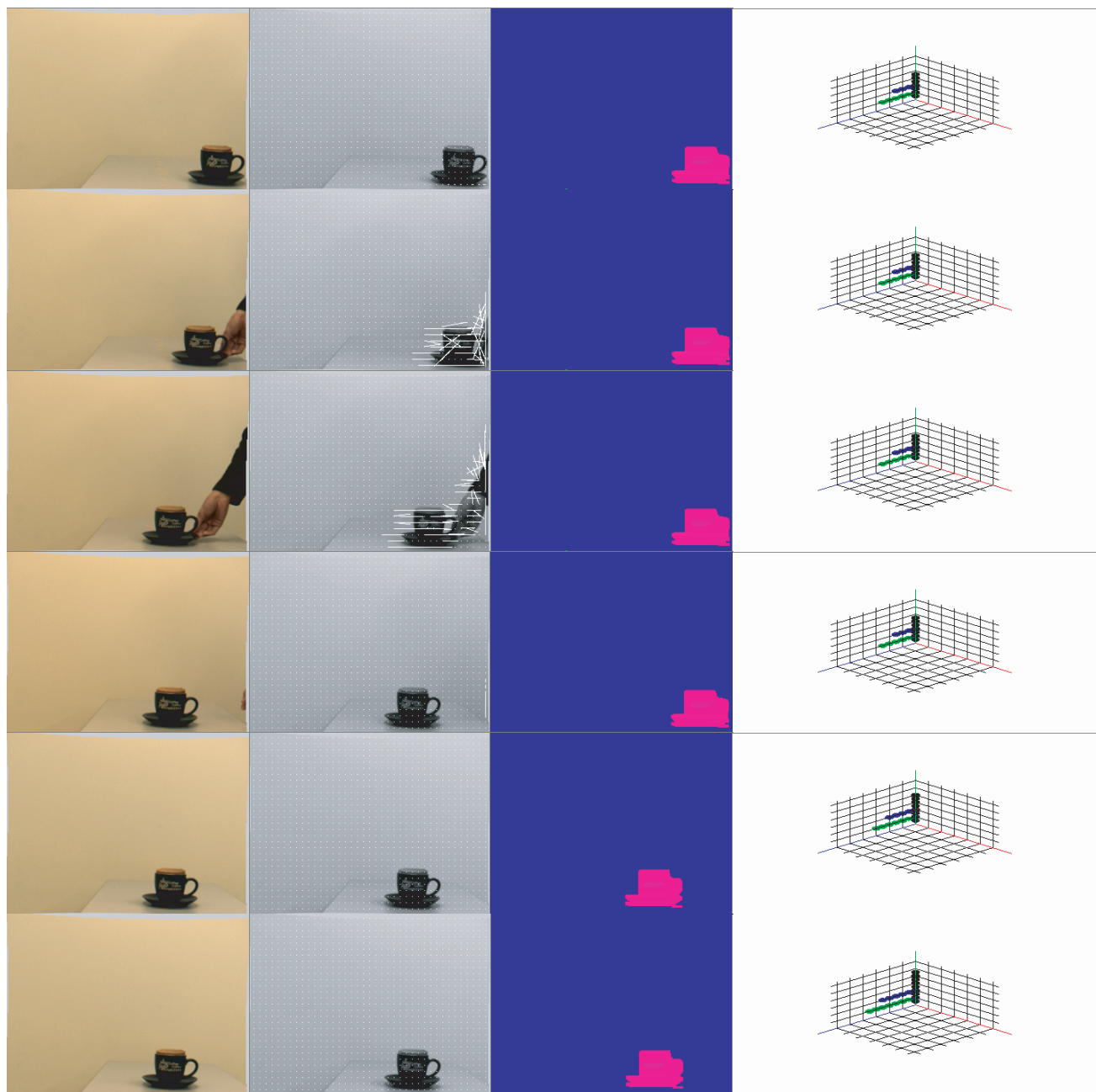


図 7.16: 動画シーケンス (4/8): 上から 65, 67, 70, 72, 76, 77 sec

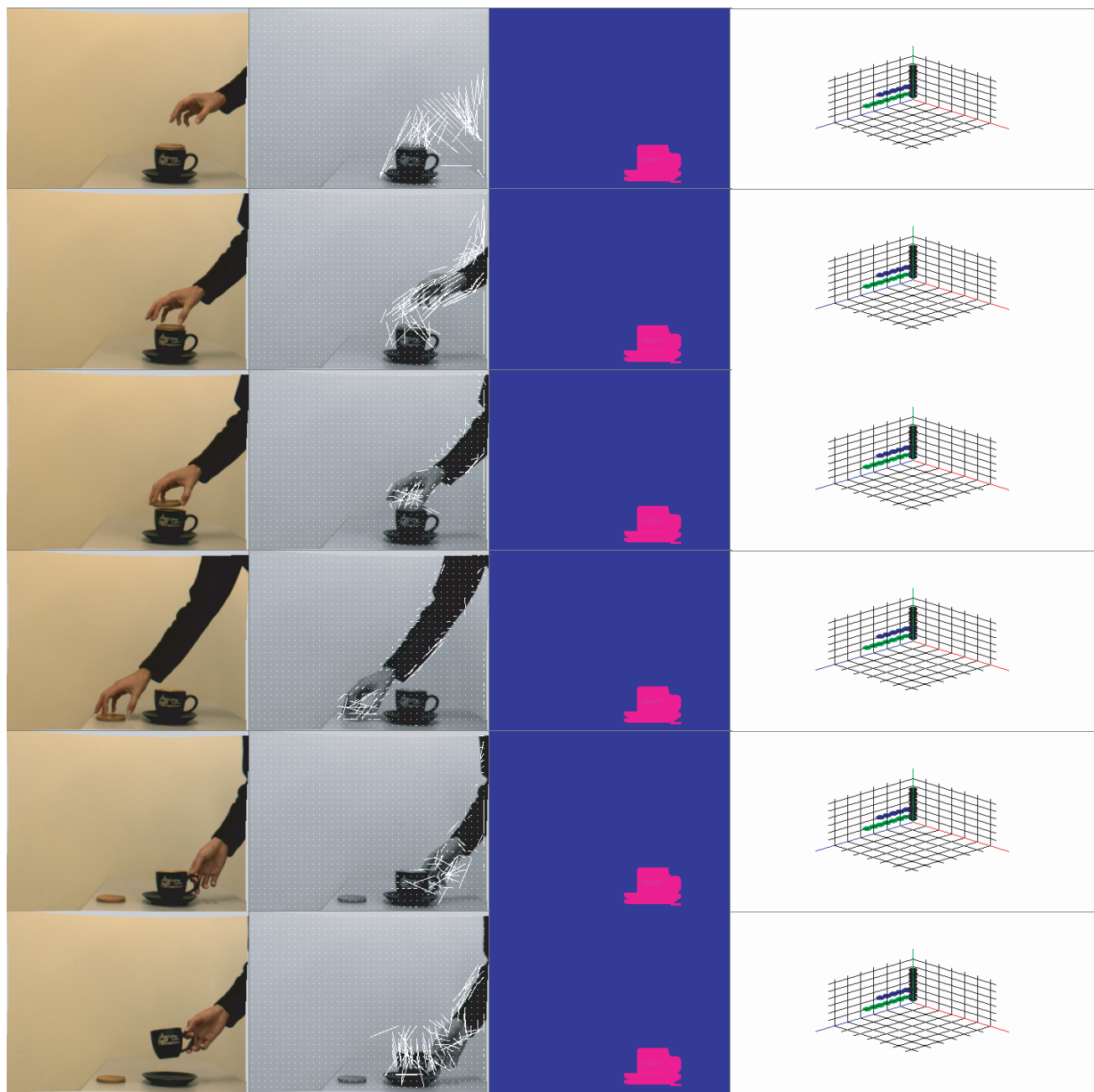


図 7.17: 動画シーケンス (5/8): 上から 78, 79, 81, 83, 84, 86 sec

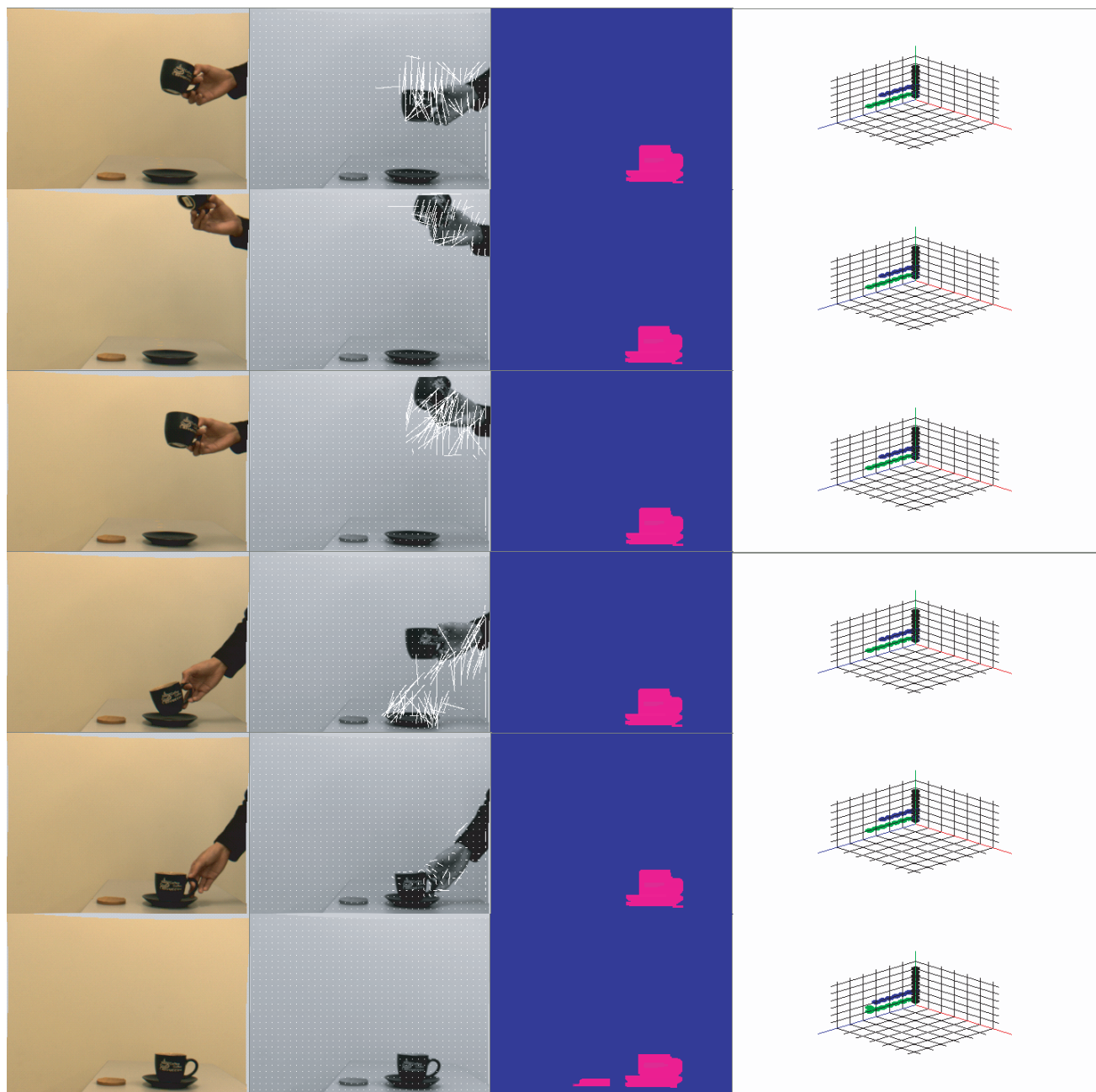


図 7.18: 動画シーケンス (6/8): 上から 87, 90, 94, 96, 98, 103 sec

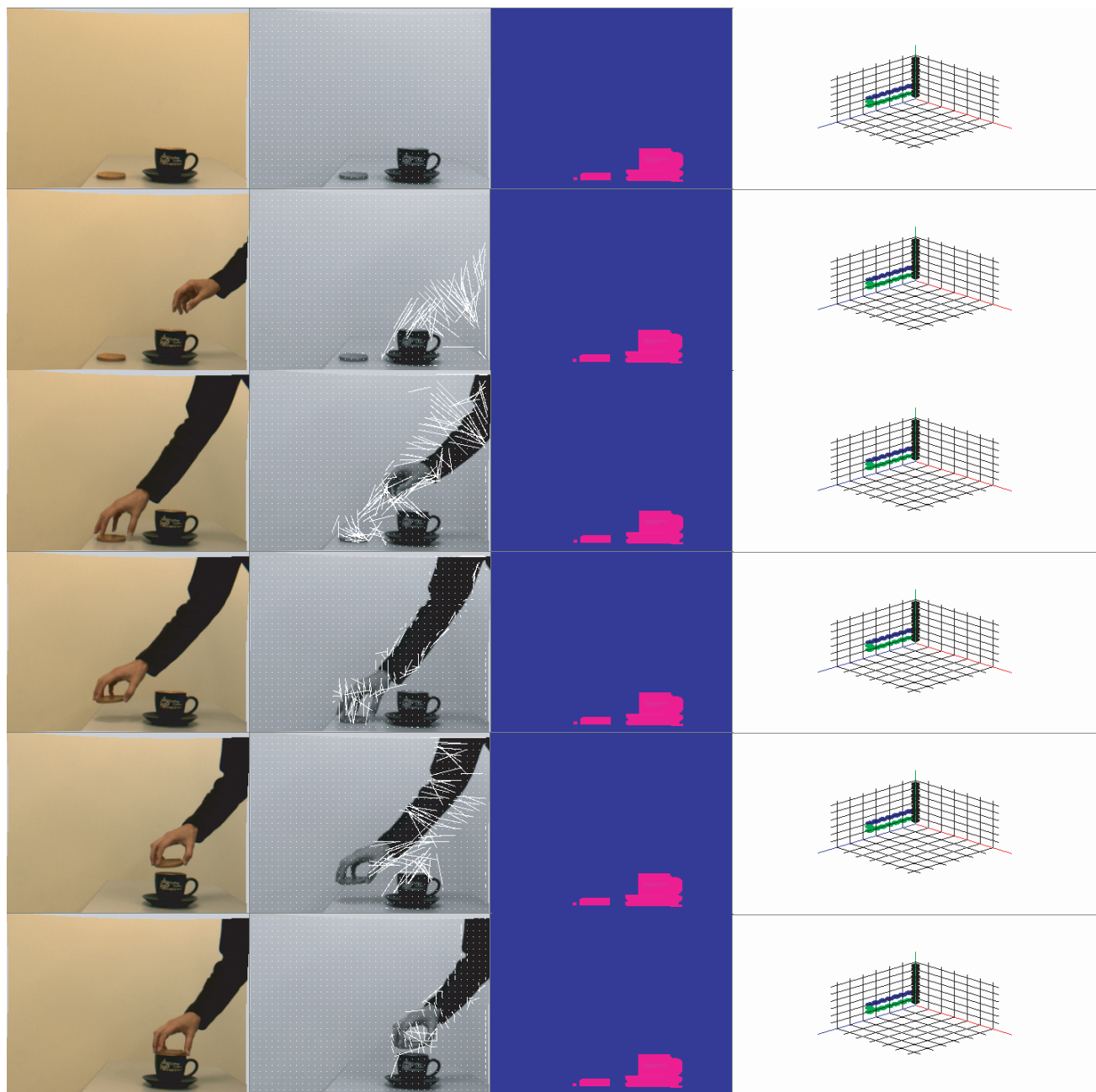


図 7.19: 動画シーケンス (7/8): 上から 105, 107, 110, 112, 113, 114 sec

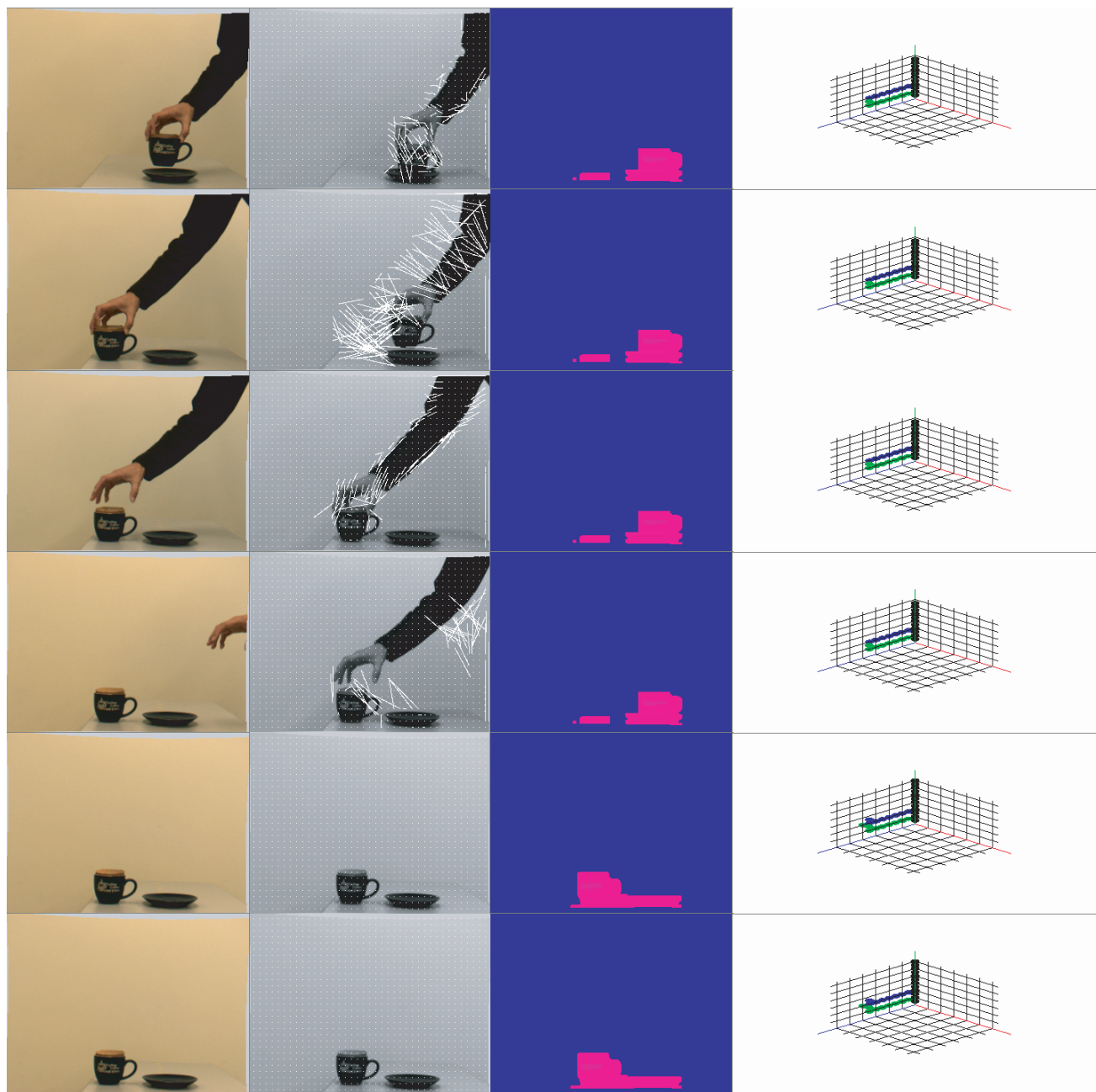
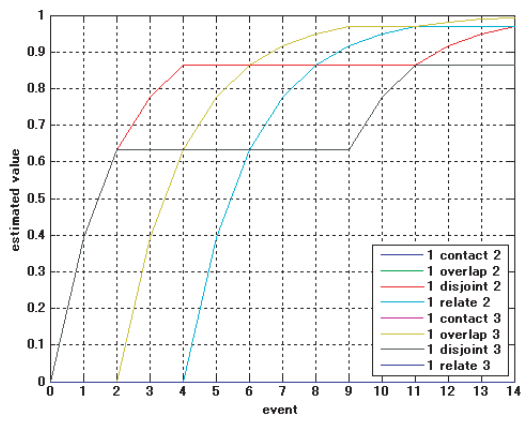
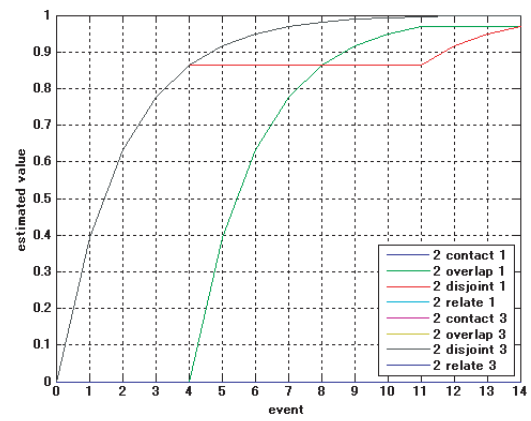


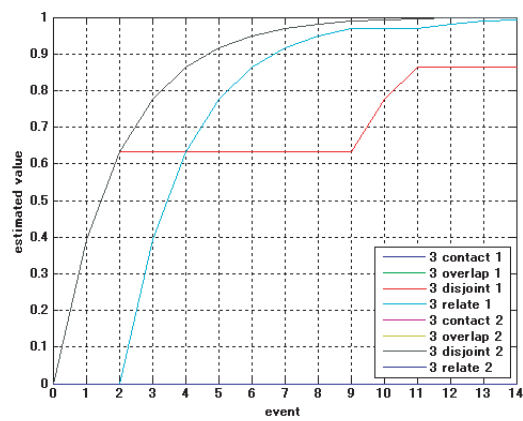
図 7.20: 動画シーケンス (8/8): 上から 115, 117, 118, 122, 124, 128 sec



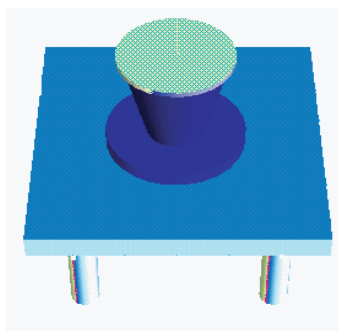
(a) カップを基準とした場合



(b) コースタを基準とした場合



(c) 蓋を基準とした場合



(d) 導出された物体配置
ルール

図 7.21: 馴化評価モデルの event ごとの変動と関係推論

第8章 結論

本研究では、まず我々人間が通常位置関係を表すときに用いている表現方法について、相対的な表現を用いていることを示した。そしてその相対的な表現方法として「距離情報」、「方向情報」、「接触情報」の3つを組み合わせてトポロジー関係と定義した。そして、空間内の物体間の関係を表現、抽出する手法として関係表現行列を提案した。その行列の特徴抽出することで、関係を抽出でき、静止物体や移動物体の個々の局所関係を集めて全体の関係を推定することで、大局的な状況を明らかになることを示した。さらに、得られた物体間の関係の信頼度を表す馴化評価モデルを定義した。これをシステムに実装し、状況に応じて馴化評価モデルが変化し、評価値が大きいほど物体間の関係の信頼度が高いことを示した。これにより、このような信頼度の高い物体間の関係を知識に加えていくことは効率的であり、様々な物体の配置に用いられている何らかのルールを導き出すことができる。さらに、これらの配置列を学習することで世界の配置ルールや常識を知ることができ、部屋の片付けをする、物を探すということも可能になる。今後の展望として、次のようなことが考えられる。

1. 自律エージェントとの統合環境の整備

本研究では、静止物体の位置を表現することに重点を置いた。また簡単な評価モデルを定義し、空間内の物体の配置ルールを導出することができる。このシステムをロボットなどエージェントに搭載していくことが必要である。

2. よりふさわしい表現方法の検討

本研究では、複雑な環境で1つの物体につき、複数出力される位置関係で表現している。複数出力される情報の中からよりふさわしい表現方法を特定できれば、処理速度、今後の動作計画などに役立つと考える。

謝辞

本研究を進めるにあたり，ご多忙中にも関わらず，常に熱心にご指導ご鞭撻を頂いた北陸先端科学技術大学院大学 情報科学研究科 丁 洛榮助教授に心から深く御礼申し上げます。また，日々の様々な面で助言やご協力を下さった Cognitive Robotics Team のリーダーである博士後期過程 柳 在官氏をはじめ，博士後期課程，前期課程の皆様に御礼申し上げます。

付録 A: 奥行き画像

従来の空間情報認識の手法の1つである距離認識には大きく分けて2つの方法が考えられる。1つ目はレーザ光などを物体に直接当て、その光の状態から物体の場所を判断する方法である。2つ目はある物体の画像を用いて、その画像を処理することで距離情報を得る方法である。

画像を使った立体認識には大きく分けて次の2つがある。

1. 1枚の画像を用い、立体の認識をする方法

1枚の画像では、画像の距離情報が失われてしまっている。よって何らかの方法でそれを補わなくてはならない。その方法としていくつか考えられている。1つ目は、画像の距離情報が予め既知な点あるいは点列を用いる方法。2つ目は、画像のセグメンテーションを行い、予め用意したモデルと照合する方法。3つ目は、画像中の陰影やテクスチャを用いる方法である。

2. ステレオ画像など複数の画像を用いて立体を認識する方法

複数の視点または照明条件で撮影した画像を用いることにより、3次元情報の算出を行う。その方法を計測原理により大別すると、受動型計測と能動型計測がある。受動型計測の代表的な計測方式としては、ステレオマッチング法、レンズ焦点法などが挙げられる。能動型計測の代表的な計測方式としては、光レーザ法、照度差ステレオ法などが挙げられる。以下では、ステレオマッチング法と照度差ステレオ法について詳しく説明する。

ステレオマッチング法

ステレオマッチング法とは、人間の両眼のように左右にカメラを配置し、左カメラと右カメラで撮影される2枚の画像よりカメラ間での視差を求め、視差より奥行き画像を計測する方式である。ステレオマッチング法の原理を図 8.1 に示す。図 8.1 は、3次元空間に対し XYZ 座標を考えている。左カメラ、右カメラのレンズの中心がそれぞれ O_L , O_R になるようにカメラ間のベース距離を d だけ離して平行に並んでいる。また空間中の計測点を P 、カメラの焦点距離を f とする。

物体を撮影したとき、左右のそれぞれの画像は x_L, y_L 平面, x_R, y_R 平面に投影される。ただし、左右の画像中の点の座標はそれぞれの画像の中心を原点とする相対座標である。こ

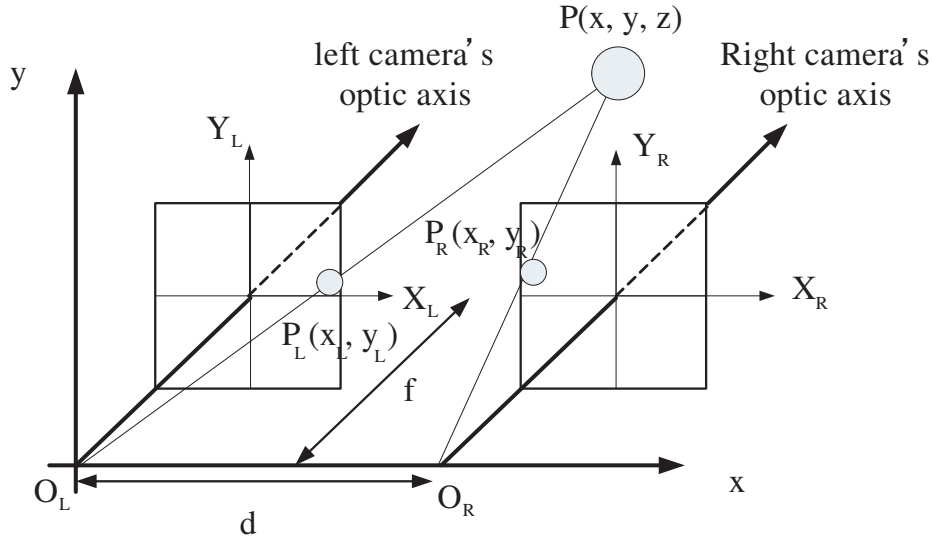


図 8.1: ステレオマッチング法の原理

のとき，左画像中の点 P_L と右画像中の点 P_R は対象物体上の同一の点 P にそれぞれが対応している．この点 P の3次元座標は三角計測の原理で求めることができる．これを図8.2で説明する．図8.2は Y 軸方向の正方向から見たときの様子を示している．図の相似な三角形 PO_LO_R , PP_LP_R のそれぞれの底辺の長さ d , $d - (x_L - x_R)$ と，それぞれの三角形の高さの方向の長さ z , $z - f$ の間には比例関係がある．これにより

$$z = \frac{d \cdot f}{x_L - x_R} \quad (8.1)$$

同様にして， X ， Y に関しても同様にして求めることができる．

$$x = \frac{d \cdot x_L}{x_L - x_R}, y = \frac{d \cdot y_L}{x_L - x_R} = \frac{d \cdot y_R}{x_L - x_R} \quad (8.2)$$

これらの式から，左右の画像の対応する点を決定することにより，画像中の各点の3次元座標を求めることができる．

このようにステレオマッチングを行うにあたっては，処理の簡単化のためにいくつかの必要条件が挙げられる．まず左右のカメラのレンズをつなぐ線分が水平であること．これは左右の画像平面が x 軸に平行であれば常に成立することを意味している．上記の式(3.2)に示されるように $y_L = y_R$ であるので左右の画像のある水平線上でその間の対応関係を決定すればよい．つまり，図8.2で示すように左右の画像の同じ高さにある水平線 l_L , l_R を取り出し，それぞれの線上の何らかの特徴点とその対象関係をみて，決定すればよい．さらに照明条件に変化すると左右のカメラで撮影される画像も変化するが，両方のカメラで同じように変化するため照明条件の変化には影響されない．しかし，このような計測方式では，左右の画像の対応点決定が難しいといった問題点がある．

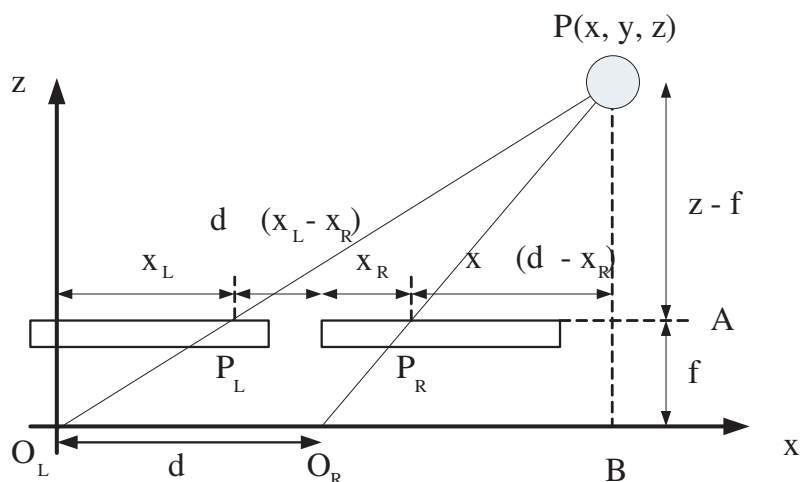


図 8.2: 三角測量の原理

照度差ステレオ法

照度差ステレオ法とは、複数の照明光源を切り替えて写した複数枚の画像から面の方向を求めようとするものである。反射率分布が分かっている被写体を既知の方向から照射すると、ある画素の傾きに対して拘束条件を与える。その後、別の光源から被写体を照明し、もう1つの拘束条件を求め、この2つの拘束条件の好転として画素の方向を求める。このとき、カメラは1台だけを用い、その照明条件を変えた画像を複数使い、3次元形状を求める。この場合、カメラは移動しないことから、対応付けの問題は省くことができる。空間中のある点 P を図 8.3 に示すように光源 L_i のいずれか1つだけを転倒して複数の画像を撮影する場合について考える。このときの点 P が属する面の法線を n 、点 P からの光源 L_i に向かう方向を n_i とし、 n と n_i のなす角度を θ_i とする。

点 P における反射が完全拡散反射であると仮定すると、点 P に対応する画像上の画素の階調値 $f(x, y)$ は $\cos \theta_i$ に比例する。物体表面 $z = z(x, y)$ とすると、法線ベクトル n は次式で表される。

$$n = (\frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}, -1) = (p, q, -1) \quad (8.3)$$

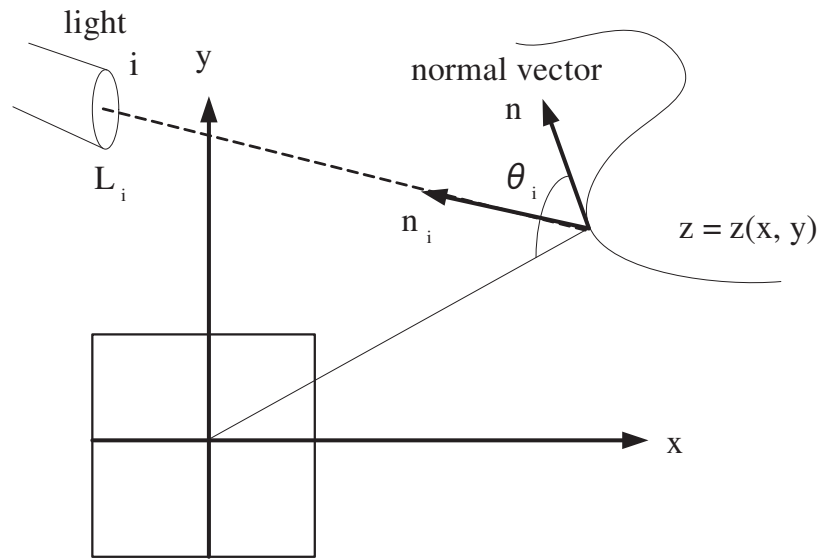


図 8.3: 照度差ステレオ法の原理

ただし, p_i, q_i を定数, すなわち光源が十分遠方にあると仮定すると, 式 (8.5) は pq 平面上の曲線を作る. 光源を複数設定してそれぞれ式 (8.5) を算出し, それらを連立させることで, 物体の表面の反射率が不明であっても, その法線ベクトルを求めることができる.

以上のように, 特殊環境での奥行き画像計測では, 認識対象や計測方式によっては奥行き画像を計測できないこともあり得る. しかし, 多種の計測原理があり, 様々な環境で奥行き画像計測を実現するために開発されている. そのために, 環境に応じた計測原理を用いることにより, 環境中に存在する認識対象の特性を計測することが可能であり, 安定した状況認識などを容易に実現することができる.

作成したプログラムの概要

Pixel-to-Pixel Stereo

```
/*Matching scanlines independently*/
for  $\delta \leftarrow 0$  to  $\Delta$ 
   $\psi[\delta, 0] \leftarrow d(\delta, 0)$ 
  for  $y \leftarrow 1$  to  $n - 1$ 
    for  $\delta \leftarrow 0$  to  $\Delta$ 
       $\psi[\delta, y] \leftarrow \infty$ 
    for  $x \leftarrow 0$  to  $n - 1$ 
       $m_x[x] \leftarrow \infty$ 
    for  $y \leftarrow 0$  to  $n - 2$ 
       $m_y \leftarrow \min(\psi[0, y_p], \psi[1, y_p], \dots, \psi[\Delta, y_p])$ 
      for  $\delta_p \leftarrow 0$  to  $\Delta$ 
        update( $\delta_p, y_p, \delta_p, y_p + 1$ )
        if  $\psi[\delta_p, y_p] \leq m_y$  then
           $y \leftarrow y_p + 1$ 
          for  $\delta \leftarrow y_p + 1$  to  $\Delta$ 
            if  $v_L[y + \delta - 1]$  then update( $\delta_p, y_p, \delta, y$ )
          if  $\psi[\delta_p, y_p] \leq m_x[\delta_p + y_p]$  and  $v_R[y_p + 1]$  then
            for  $\delta \leftarrow 0$  to  $\delta_p + 1$ 
               $y \leftarrow y_p + \delta_p - \delta + 1$ 
              update( $\delta_p, y_p, \delta, y$ )
/*update function*/
 $\psi' \leftarrow \psi[\delta_p, y_p] + d(y + \delta, y) - \kappa_r + \kappa_{occ} * (\delta \neq \delta_p)$ 
if  $\psi' < \psi[\delta, y]$  then
   $\psi[\delta, y] \leftarrow \psi'$ 
   $\pi[\delta, y] \leftarrow [\delta_p, y_p]$ 
   $m_x[y + \delta] \leftarrow \min(m_x[y + \delta], \psi')$ 
```

付録B:カメラキャリブレーション

Camera Calibration Toolbox で用いられているパラメータの定義について述べる。パラメータは内部パラメータと外部パラメータがある。

- 内部パラメータ

fc: 焦点距離 cc: 中心座標 alpha: ゆがみ係数 kc: 歪曲係数

- 外部パラメータ

Rc: 回転要素行列 Tc: 平行要素行列

カメラ基準座標上 $XX_c = (X_c, Y_c, Z_c)$ で点 P を定義する。そして固有内部パラメータを用いて点 P を画像平面上に投射する。ここで x_n を標準化された画像とすると

$$x_n = \begin{pmatrix} \frac{X_c}{Z_c} \\ \frac{Y_c}{Z_c} \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} \quad (8.6)$$

ここで $r^2 = x^2 + y^2$ とすると、レンズ歪み補正後の正常化された座標は以下の通り定義できる。

$$x_d = \begin{pmatrix} x_d(1) \\ x_d(2) \end{pmatrix} = (1 + kc(1)r^2 + kc(2)r^4 + kc(5)r^6)x_n + d_x \quad (8.7)$$

ただし、 d_x は接線の歪みベクトルであり

$$d_x = \begin{pmatrix} 2kc(3)xy + kc(4)(r^2 + 2x^2) \\ kc(3)(r^2 + 2y^2) + 2kc(4)xy \end{pmatrix} \quad (8.8)$$

歪みがいったん適用されると、点 P の投射の最終的なピクセル座標は

$$\begin{aligned} x_p &= fc(1)(x_d(1) + alpha * x_d(2)) + cc(1) \\ y_p &= fc(2)x_d(2) + cc(2) \end{aligned} \quad (8.9)$$

したがってピクセル座標ベクトルと正常化された座標は 1 次の関連性がある。

$$\begin{pmatrix} x_p \\ y_p \\ z_p \end{pmatrix} = KK \begin{pmatrix} x_d(1) \\ x_d(2) \\ 1 \end{pmatrix} \quad (8.10)$$

ここで KK はカメラ行列と定義し、以下に示す.

$$KK = \begin{pmatrix} fc(1) & alpha * fc(1) & cc(1) \\ 0 & fc(2) & cc(2) \\ 0 & 0 & 1 \end{pmatrix} \quad (8.11)$$

さらに, XX_c はカメラ基準座標, XX はグリッド参照フレーム上での座標とすると

$$XX_c = Rc * XX + Tc \quad (8.12)$$

これによってレンズ歪みを考慮に入れた座標変換ができるようになる.

付録C:動きベクトル

動きベクトルとはカメラと移動物体との相対的な動きによって生じる、画面上に発生する速度ベクトルを求めたものである。速度ベクトルとは画像中のある1点において示される画素情報が次の画面上で、どちらの方向へ、どの程度の距離を移動するかを示したものである。

この動きベクトルを求める方法にはブロックマッチング法と勾配法の2つがある。ブロックマッチング法は精度よく動きを検出できるが、処理量が多い。一方、勾配法は処理量がブロックマッチング法に比べて少なく処理できるが、検出の精度が低い。以下では、ブロックマッチング法と勾配法について詳しく説明する。

勾配法

勾配法とは画像の時空間微分の拘束方程式による条件から速度ベクトルを推定する手法である。「物体上の明るさは移動後も変化しない」と仮定し、時空間微分と動きベクトルとの関係式を利用して対象の動きを推定する。

画面上の点 (x, y) の時刻 t での明るさを $E(x, y, t)$ とし、時刻 t から $t + dt$ の間に対象が (x, y) から $(x + dx, y + dy)$ に移動し、明るさが変化しないと仮定すると次式が成り立つ。

$$E(x, y, t) = E(x + dx, y + dy, t + dt) \quad (8.13)$$

上記の式をテーラー展開すると、

$$E(x, y, t) = E(x, y, t) + E_x(x, y, t)dx + E_y(x, y, t)dy + E_t(x, y, t)dt + e \quad (8.14)$$

ここで、 $E_x = \frac{\partial E}{\partial x}$, $E_y = \frac{\partial E}{\partial y}$, $E_t = \frac{\partial E}{\partial t}$ であり、 e は dx, dy, dt の高次項である。 E は微小であると仮定して切捨て、両辺を dt で割り、 $dt \rightarrow 0$ とすると、

$$E_x(x, y, t)u + E_y(x, y, t)v + E_t(x, y, t) = 0 \quad (8.15)$$

と表せる。ここで、 u , v はそれぞれ x 方向の速度成分、 y 方向の速度成分である。これが勾配法における拘束方程式である。

ブロックマッチング法

ブロックマッチング法とは、図 8.4 に示すように連続するフレーム間において濃淡パターンによるパターン付けをし、第 1 フレームの画素に第 2 フレームの中の対応点をブロック間のパターンマッチングに基づいて決定する方法である。

原理的には 1 枚目の読み込んだ画像から、正方形の部分画像をブロックとして切り出す。次に、2 枚目に読み込んだ画像からもっともマッチング率が高いブロックを探し出す。そして動きの量として $(\Delta x, \Delta y)$ の速度ベクトルを出力する。これをブロックごとに繰り返す方法である。

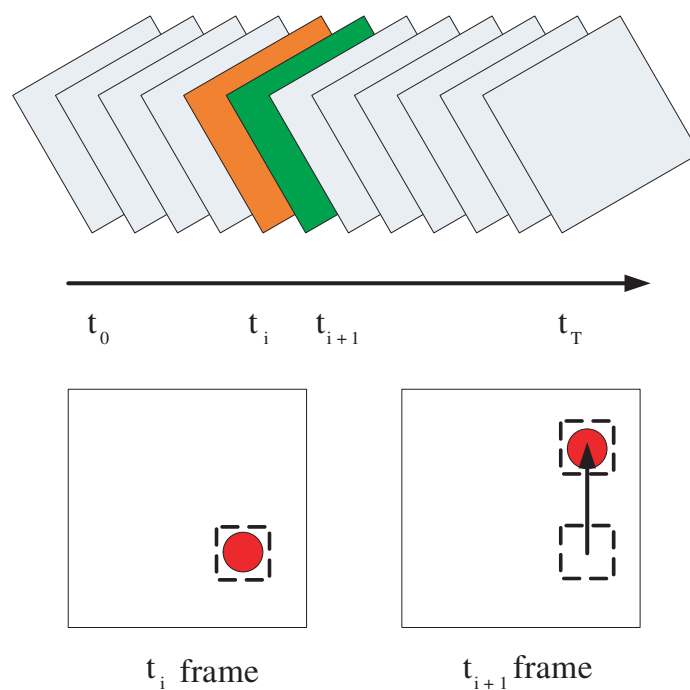


図 8.4: ブロックマッチング法

作成したプログラムの概要

Calculation motion vector

```
/*Calc_motion_vector_function*/
if  $x - x\_tmp < 0 \parallel x + x\_tmp \geq x\_size1 \parallel$ 
     $y - y\_tmp < 0 \parallel y - y\_tmp \geq y\_size1$  then
     $*vx \leftarrow 0$ ;  $*vy \leftarrow 0$ 
else
    /*cutout of templete: tmp_image*/
     $num\_tmp \leftarrow (2 * x\_tmp + 1) * (2 * y\_tmp + 1)$ 
     $min\_gray \leftarrow 255$ ;  $max\_gray \leftarrow 0$ 
    for  $i \leftarrow -y\_tmp$  to  $y\_tmp$ 
        for  $j \leftarrow x\_tmp$  to  $x\_tmp$ 
             $xp \leftarrow x + j$ ;  $yp \leftarrow y + i$ 
             $tmp\_image = image1$ 
            if  $image1 < min\_gray$  then
                 $min\_gray = image1$ 
            else if  $image1 > max\_gray$  then
                 $max\_gray = image1$ 
        if  $(max\_gray - min\_gray) \leq threshold\_gray$  then
             $*vx \leftarrow 0$ ;  $*vy \leftarrow 0$ 
        else
            /*Searching*/
             $min\_error \leftarrow 10000.0$ 
            for  $m \leftarrow -y\_offset$  to  $y\_offset$ 
                for  $n \leftarrow -x\_offset$  to  $x\_offset$ 
                     $xs \leftarrow x + n$ ;  $ys \leftarrow y + m$ 
                    /*calculation (error)*/
                    if  $xs - x\_tmp \geq 0, xs + x\_tmp < x\_size1,$ 
                         $ys - y\_tmp \geq 0, ys + y\_tmp < y\_size1$ 
                         $sum \leftarrow 0.0$ 
                        for  $i \leftarrow -y\_tmp$  to  $y\_tmp$ 
                            for  $j \leftarrow -x\_tmp$  to  $x\_tmp$ 
                                 $xp \leftarrow xs + j$ ;  $yp \leftarrow ys + i$ 
                                 $sum \leftarrow sum + abs(tmp\_image - image2)$ 
                         $error\_ratio \leftarrow sum / num\_tmp$ 
                        if  $error\_ratio \leq min\_error$  then
                             $min\_error \leftarrow error\_ratio$ 
                             $*vx \leftarrow xs - x$ 
                             $*vy \leftarrow ys - y$ 
                        end
                    end
                end
            end
        end
    end
end
```

付録D:共起行列

共起行列とは，出現するイベントのそれぞれの二項間の関連性の強さをある距離の間に現れるイベントの出現頻度により表し，全ての二項間のイベントの関連性を表現した行列である．共起行列はネットワーク，Web 検索やデータベース，コンピュータビジョンなど多岐にわたって利用されている．

コンピュータビジョンの分野では，濃度共起行列として用いられている．濃度共起行列とは，クラスタ画像 $f(x, y)$ において図 8.5 に示すように，ある特定の相対的な位置関係 (d, θ) にある画像対 (x_1, y_1) と (x_2, y_2) においてその濃度が (i, j) になる，すなわち $f(x_1, y_1) = i$ ， $f(x_2, y_2) = j$ となる頻度を表したものである．画像の濃度階調度を n とすると， $n * n$ の正方行列となる．主にテンプレートマッチングなどの物体認識の一手法として用いられている．この濃度共起行列を用いて，特徴量を求めることができる．得られる特徴量としては次のようなものが挙げられる．ここで， v_x, v_y は濃度の平均であり， σ_x, σ_y は分散である．

- エネルギー

$$E\{S_\theta(d)\} = \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} s_\theta(i, j|d)^2 \quad (8.16)$$

- エントロピー

$$H\{S_\theta(d)\} = - \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} s_\theta(i, j|d) \log s_\theta(i, j|d) \quad (8.17)$$

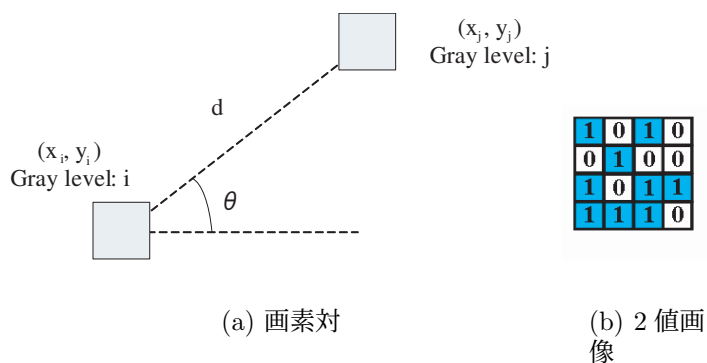


図 8.5: 濃度共起行列の求め方

- 相関

$$C\{S_\theta(d)\} = \frac{\sum_{i=0}^{n-1} \sum_{j=0}^{n-1} (i - v_x)(j - v_y) s_\theta(i, j|d)}{(\sigma_x \sigma_y)} \quad (8.18)$$

- 局所一様性

$$L\{S_\theta(d)\} = \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} \frac{s_\theta(i, j|d)}{1 + (i - j)^2} \quad (8.19)$$

- 慣性

$$I\{S_\theta(d)\} = \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} (i - j)^2 s_\theta(i, j|d) \quad (8.20)$$

例えば図 8(b) に示す 2 値画像の場合，共起行列は次のようになる．

$$\begin{pmatrix} 2 & 8 \\ 8 & 6 \end{pmatrix} \quad \begin{pmatrix} 6 & 2 \\ 2 & 8 \end{pmatrix} \quad \begin{pmatrix} 2 & 9 \\ 9 & 4 \end{pmatrix} \quad \begin{pmatrix} 4 & 4 \\ 4 & 6 \end{pmatrix}$$

この共起行列は，右から方向 0, 45, 90, 135, 距離 d が 1 のときの結果である．これによりテクスチャ画像の特徴を抽出することができる．

参考文献

- [1] 中山 眞, ”ロボットが日本を救う” 東洋経済新報社, 2006.
- [2] J. C. Gomez, 長谷川 眞理子 (訳), ”霊長類のこころ -適応戦略としての認知発達と進化-”, 新曜社, 2005.
- [3] けいはんな社会的知能発生学研究会, ”知能の謎 -認知発達ロボティクスの挑戦-”, 講談社, 2004.
- [4] 田村 秀行, ”コンピュータ画像処理入門”, 総研出版, 2003.
- [5] 田中 茂範, 松本 曜, ”日英語比較選書 6 空間と移動の表現”, 研究社出版, 1997.
- [6] (社) 日本ロボット工業会, 21 世紀におけるロボット社会創造のための技術戦略報告書.
- [7] General ヒーリング, <http://naonyan.ath.cx/General/General-top.html>
- [8] 浅田 稔, 石黒 浩, 国吉 康夫, ”認知ロボティクスの目指すもの”, 日本ロボット学会誌, Vol.17, No.1, pp2-6, 1999.
- [9] 中村 恭之, 佐藤 知正, 国吉 康夫, 開 一夫, 柴田 智広, 浅田 稔, K. F. MacDorman, 谷 淳, ”なぜ認知ロボティクスは有望なのか?”, 日本ロボット学会誌, Vol.17, No.1, pp38-43, 1999.
- [10] 浅田 稔, ”認知発達ロボティクスによる赤ちゃん学の試み”, 日本赤ちゃん学会, vol.4, pp2-7, 2004.
- [11] M. R. Endsly, ”Toward a theory of situation awareness in dynamic system”, Human Factors, 37(1), pp32-64, 1995.
- [12] H. A. Yanco and J. Drury, ””Where Am I ?” Acquiring Situation Awareness Using a Remote Robot Platform”, Proceeding of the IEEE Conference on Systems, Man and Cybernetic, October 2004.
- [13] W. Zhang and R. W. Hill, ”A Template-Based and Pattern-Driven Approach to Situation Awareness and Assessment in Virtual Humans”, Proceeding of the Fourth International Conference on Autonomous Agents, Barcelona, Cataluma, Spain, pp116-123, 2000.

- [14] S. Hughes, M. Lewis, "Robotic camera control for remote exploration", Proceeding of the Conference on Human Factors in Computing System, Vienna, Austria, pp511–517, 2004.
- [15] W. E. Green, P. Y. Oh, "An aerial robot prototype for situation awareness in closed quarters"
- [16] A. Torralba, K. P. Murphy, W. T. Freeman, M. A. Rubin, "Context-Based Vision System for Place and Object Recognition", Proceeding of the IEEE Conference on Computer Vision, Nice. France, pp273–, 2003.
- [17] 出口 光一郎, 柳井 啓司, "物体間の支持関係を利用した室内画像の認識", 電子情報通信学会論文誌 D, Vol.J84-D2, No.8, pp1741–1752.
- [18] 樋口 未来, 小島 篤博, 福永 邦雄, "物体の機能と人物の動作解析によるシーン認識", 信学技報.
- [19] 柳沢 豊, 服部 正嗣, 平松 薫, 岡留 剛, "物体間の関係記述に基づく実世界中のイベント検出手法", 情報処理学会全国大会, 講演番号 5k-4-4, 2005.
- [20] W. Chou, J. Xiao, "Real-time and Accurate Multiple Contact Detection between General Curved Objects", Proceeding of International Conference on Intelligent Robots and Systems, Beijing, China, pp.556–561, 2006.
- [21] M. Brand, "Physical-based visual understanding".
- [22] 藤森 史生, 有川 正俊, 森田 喬, "SVG 地図を対象とした空間関係抽出とその応用", DEWS, 1-C-02, 2003.
- [23] 中村 嘉志, 並松 祐子, 宮崎 伸夫, 松尾 豊, 西村 拓一, "実世界での局所的位相関係とトポロジーを用いた情報支援に向けて", 第 20 回人工知能学会全国大会, 講演番号 1F2-03, 2006.
- [24] X. Lin, Q. Lin, Y. Yuan, and X. Zhou, "Multiscale Histograms: Summarizing Topological Relations in Large Spatial Datasets", Manuscript, 2003.
- [25] C. Sun, D. Agrawal, A. E. Abbadi, "Exploring Spatial Datasets with Histograms", Proceeding of the International Conference on Data Engineering, 2004.
- [26] J. M. Tenenbaum, H. G. Barrow, "Experiments in interpretation guided segmentation", Artificial Intelligence, Vol.8, pp241–274, 1977.
- [27] Y. Ohta, "Knowledge-Based Interpretation of Outdoor Natural Color Scenes", Research Notes in Artificial Intelligence, Vol.4.

- [28] S. Birchfield, C. Tomasi, "Depth Discontinuities by Pixel-to-Pixel Stereo", *Internal Journal of Computer Vision*, 35(3), pp269–293, 1999.
- [29] M. Z. Prown, D. Burschka, G. D. Hager, "Advances in Computational Stereo", *IEEE, Transactions on Pattern Analysis and Machine Intelligence*, Vol.25, No.8, pp993–1008, 2003.
- [30] Camera Calibration Toolbox, "<http://www.vision.caltech.edu/bouguetj/index.html>"
- [31] 別所 克人, 古瀬 茂, 片岡 良治, "単語と意味属性との共起に基づく概念ベクトル生成手法", 第20回人工知能学会全国大会, 講演番号 3C3-1, 2006.
- [32] 石尾 豊, 石塚 満, "語の共起の統計情報に基づく文章からのキーワード抽出アルゴリズム", *人工知能学会論文誌*, 17-3D, pp217–223, 2002.
- [33] S. Ekvall, D. Kragic, "Receptive Field Cooccurrence Histograms for Object Detection", *IEEE, International Conference on Intelligent Robots and Systems*, pp3969–3974, 2005.
- [34] P. Chang, J. Krumm, "Object Recognition with Color Cooccurrence Histograms", *IEEE Conference on Computer Vision and Pattern Recognition*, Fort Collins, CO, 1999.
- [35] 北海道大学大学院理学研究院 生命理学部門 生命機能学分野 行動知能学 Ia 長山 俊樹 研究室 ホームページ, <http://nag3.sci.hokudai.ac.jp/welcome.html>
- [36] M. Araki and T. Nagayama, "Decrease in excitability of LG following habituation of the crayfish escape reaction", *J Comp Physical A* 191, pp481–489, 2005.