# 博士論文

時系列画像に基づく動的事象の推測に関する研究

# 2007年1月

名古屋大学大学院 情報科学研究科

# 大塚和弘

# 内容梗概

コンピュータビジョンは、人間などの生物がもつ視覚機能を人工的に実現することを目指す学 問分野である.コンピュータビジョンの中でも特に動的な事象を観測対象とする領域のことを時 系列画像解析と呼ぶ.時系列画像解析では,時々刻々と観測される2次元画像の集合を入力とし, その画像集合から,そこに映っている3次元世界中の対象の状態やその時間変化に関する推測が 行われる.しかし、観測情報である2次元画像は,3次元世界を2次元平面へ射影して得られる 縮退された情報であり,その観測情報から対象とする事象の状態を一意に決定することは必ずし も容易ではない.本論文では,この時系列画像に基づく推測の困難性を,観測の部分性と間接性 という2つの側面から捉え,それらに対処可能な時系列画像解析の新しいアプローチを提案する. ここで観測の部分性とは,複数の対象間のオクルージョン(隠れとも呼ぶ)や観測範囲の制限に よって生じる観測情報の欠落のことを指し,観測の間接性とは,観測情報と推測対象の状態との 間の質的な乖離のことを指す.本論文では,観測の部分性,間接性の度合いの組み合わせに応じ て,以下に述べる3つの問題を取り上げる.

まず、観測の部分性が顕著な例として、複数の物体の間に生じるオクルージョンに着目する、オ クルージョンとは,カメラに対して,手前の物体が後方の物体を遮蔽する現象であり,時系列画 像解析を妨げる主要因として知られている.オクルージョンが生じる場合,遮蔽された領域に関 する観測情報が得られないという問題が生じる.また,画像上で複数の物体のオーバーラップが 生じ、観測情報の解釈が困難になるという問題も生じる、本論文では、多視点観測による複数物 体の追跡(トラッキングとも呼ぶ)を課題として取り上げる.この課題に対して本研究では,物 体間のオクルージョンに対して頑健な追跡を実現するために,物体間のオクルージョンを含めた 観測過程を陽にモデル化し,オクルージョンの構造と物体の状態(位置や姿勢)を同時に推定す るという方法を提案する.この方法では,各々の視点からどの物体が見えていて,どの物体が隠 れているかというオクルージョンの空間構造が陽に記述され、物体追跡の問題は、このオクルー ジョンの空間構造についての多重仮説の生成・検証,及び,物体の状態の事後確率分布の推定と いう2つの問題からなる再帰的ベイズ推定法として定式化される.この方法は,オクルージョン に起因する物体の配置や状態の不確定性を推定することができるため,多様にオクルージョンが 変化する状況においても , 安定した物体追跡を行うことができる . 人工データ , 及び , 実データを 用いた実験を行い、その結果、一時的に全ての視点においてオクルージョンが発生している状況 においても,頑健に追跡が継続できることを確認する.

次に、観測の間接性が顕著である例として、事象を構成する複数の要素間の相互作用に着目す る.その中でも本論文では,視覚的に直接観測はできないが,相互作用が事象のダイナミクスに 対して決定的な影響をもつ事例として、複数人物の対面会話シーンを取り上げる、対面会話シー ンにおいては、各人物の行動は時系列画像として比較的容易に観測できるものの、それら行動を 支配するダイナミクスは,人物間の相互作用(インタラクションとも呼ぶ)に依存し,非常に複雑 である.本論文では,会話中の人物間の相互作用について陽にモデル化を行い,各人物の行動か ら人物間の相互作用の状態を推定する新しい方法を提案する.具体的には,相互作用として,話 し掛け,及び,傾聴に着目し,誰が誰に話し掛けているか?誰が誰の話を聞いているか?という人 物間でのメッセージの流れのパターンを推測の対象とする.本研究では,これを会話構造と呼ぶ. また、この推測の手掛かりとして、会話参加者の非言語行動に着目し、会話構造と非言語行動と の関連性を動的ベイジアンネットワークを用いてモデル化する.ここでは非言語行動として,発 話の状態(発話の有無)と,対面会話において重要な役割をもつことが知られている視線を用い る.ただし,会話中の視線方向を直接的に計測することは困難であるため,代わりに頭部方向を 計測対象とし , センサや画像上での頭部追跡によって計測された頭部方向から視線方向の推測を 行う.本論文では,頭部方向と発話の有無の時系列データから,会話構造,視線方向,及び,モデ ルパラメータを同時に推定する方法として、ギブスサンプリングと呼ばれる一種のマルコフ連鎖 モンテカルロ法を用いる方法を提案する.最後に,4人会話を対象とした実験により,会話構造と 視線方向の推定精度を評価し,提案法の有効性を確認する.

さらに,観測の部分性と間接性がともに顕著な事例として,時系列画像パターンとして観測さ れる自然現象の予測の問題を取り上げる、本研究では、自然現象の中でも気象現象、その中でも 特に降水現象に着目し,その観測手段として気象レーダを用いる.気象レーダは,降水強度の空 間分布を画像パターンとして観測する装置である.これを用いることで,降水の空間分布の予測 を行う問題は,気象レーダ画像上のパターンの時間変化を予測する問題として捉えることができ る.しかし,このパターンは,非剛体の不定形パターンであり,様々な気象的要因が重なり合っ た複雑なダイナミクスを持つため,その予測は容易でない.気象レーダによる観測は,パターン の元となる降水現象の空間的な広がりに対して、気象レーダから一定範囲内、かつ、ある空間断 面上に限定された観測であり,降水現象の全容を捉えることができない.その観点から観測には 部分性が伴うといえる.また,気象レーダでは,気温,気圧,風速といった気象現象の内部状態 を直接的に観測することができず、さらに、現時点の観測情報から未来のパターンを予測すると いう問題の性質からも,観測には間接性が伴うと考えられる.これらの観点より,気象レーダ画 像上のパターンの予測の問題は,観測の部分性,間接性がともに顕著な事例といえる.本論文で は、このような複雑な時系列画像パターンの予測の方法として、過去の画像パターンの記憶に基 づく予測法を提案する.これは,現在の画像系列に類似する過去の画像系列をデータベースより 検索し,検索された画像中のパターンの時間発展に基づいて予測画像を生成するという方法であ

る.この方法の要素技術として,本論文では,時系列の気象レーダ画像より抽出された画像特徴 に基づいて類似画像系列を検索する方法,及び,検索された画像系列の未来の画像を合成するこ とで予測画像を生成する方法を提案する.北海道の約6冬期分のデータセットを用いた実験を行 い,予測先行時間,データ量の観点から予測精度を評価し,提案方法の有効性を確認する.

以上のように,本論文では,従来の時系列画像解析において未だ解決法が確立されていない観 測の部分性,間接性の問題に対する新しいアプローチを提案する.具体的には,未解決課題の中で も特に重要な事例として,オクルージョンに頑健な物体追跡,複数人物の対面会話シーンの分析, 気象レーダ画像パターンの予測の3つを対象とし,それぞれ,観測過程のモデル化,相互作用のモ デル化,記憶に基づくモデル化というアプローチを提案する.これら3つの事例について提案さ れる具体的な方法は,個別の問題設定に特化したものであるが,その根幹となるアプローチは幅 広い事象に適用できるものであり,今後,時系列画像解析分野の発展に資することが期待される.

# 目 次

内容梗概				
第1章	序論		1	
1.1	まえがき			
1.2	1.2 研究の背景と従来研究		2	
	1.2.1	観測の部分性 ―オクルージョン― .............................	5	
	1.2.2	観測の間接性 ― 複数要素間の相互作用― ........................	7	
	1.2.3	観測の部分性,間接性 ――非剛体・不定形パターン―	9	
1.3	本研究	の目的と位置付け	11	
	1.3.1	多視点オクルージョン解析による複数物体の追跡	12	
	1.3.2	非言語行動に基づく会話構造の推定.........................	12	
	1.3.3	記憶に基づく気象レーダ画像パターンの予測...............	14	
	1.3.4	本論文で提案するアプローチ間の位置づけ	15	
1.4	本論文	での構成	17	
第2章	多視点	マイクルージョン解析による複数物体の追跡	19	
2.1	緒言		19	
2.2	.2 観測過程とオクルージョンの構造			
	観測過	超程とオクルージョンの構造	21	
	<b>観測</b> 遁 2.2.1	過程とオクルージョンの構造 観測過程の定義	21 21	
	観測過 2.2.1 2.2.2	2程とオクルージョンの構造 観測過程の定義	<ul><li>21</li><li>21</li><li>23</li></ul>	
	観測過 2.2.1 2.2.2 2.2.3	2程とオクルージョンの構造 観測過程の定義 視角表現に関する幾何的制約 部分オクルージョンと完全オクルージョン	<ul><li>21</li><li>21</li><li>23</li><li>25</li></ul>	
	観測過 2.2.1 2.2.2 2.2.3 2.2.4	<ul> <li>         ・・・・・・・・・・・・・・・・・・・・・・・・・・・・・</li></ul>	<ul> <li>21</li> <li>21</li> <li>23</li> <li>25</li> <li>26</li> </ul>	
	観測通 2.2.1 2.2.2 2.2.3 2.2.4 2.2.5	<ul> <li>         程とオクルージョンの構造         <ul> <li>             ・ ・・・・・・・・・・・・・・・・・・・・・・・</li></ul></li></ul>	<ul> <li>21</li> <li>21</li> <li>23</li> <li>25</li> <li>26</li> <li>26</li> </ul>	
	観測通 2.2.1 2.2.2 2.2.3 2.2.4 2.2.5 2.2.6	程とオクルージョンの構造	<ul> <li>21</li> <li>21</li> <li>23</li> <li>25</li> <li>26</li> <li>26</li> <li>29</li> </ul>	
2.3	観測通 2.2.1 2.2.2 2.2.3 2.2.4 2.2.5 2.2.6 オクル	程とオクルージョンの構造	<ul> <li>21</li> <li>21</li> <li>23</li> <li>25</li> <li>26</li> <li>26</li> <li>29</li> <li>31</li> </ul>	
2.3	観測通 2.2.1 2.2.2 2.2.3 2.2.4 2.2.5 2.2.6 オクル 2.3.1	程とオクルージョンの構造	<ul> <li>21</li> <li>21</li> <li>21</li> <li>23</li> <li>25</li> <li>26</li> <li>26</li> <li>29</li> <li>31</li> <li>31</li> </ul>	
2.3	観測通 2.2.1 2.2.2 2.2.3 2.2.4 2.2.5 2.2.6 オクル 2.3.1 2.3.2	程とオクルージョンの構造	<ul> <li>21</li> <li>21</li> <li>23</li> <li>25</li> <li>26</li> <li>26</li> <li>29</li> <li>31</li> <li>31</li> <li>32</li> </ul>	
2.3	観測通 2.2.1 2.2.2 2.2.3 2.2.4 2.2.5 2.2.6 オクル 2.3.1 2.3.2 2.3.3	福祉 とオクルージョンの構造   観測過程の定義   視角表現に関する幾何的制約   部分オクルージョンと完全オクルージョン   視角表現と物体状態の不確定性   オクルージョン構造の記述    イクルージョン構造のダイナミクス   インリージョン構造の推定   インリージョン構造の推定   インリージョン構造の推定   インリージョン低談と物体状態の尤度の定義   イクルージョン仮説の生成	21 21 23 25 26 26 29 31 31 32 33	
2.3	観測通 2.2.1 2.2.2 2.2.3 2.2.4 2.2.5 2.2.6 オクル 2.3.1 2.3.2 2.3.3 2.3.4	程とオクルージョンの構造   観測過程の定義   視角表現に関する幾何的制約   部分オクルージョンと完全オクルージョン    視角表現と物体状態の不確定性   オクルージョン構造の記述    イクルージョン構造のダイナミクス   インレージョン構造のが生定    イクルージョン仮説の生成   オクルージョン仮説の検証	21 21 23 25 26 26 29 31 31 32 33 35	

	2.3.5 物体状態の更新	36
	2.3.6 物体状態の予測	36
	2.3.7 初期化	37
2.4	実験	37
	2.4.1 パーティクル・フィルタを用いた実装	38
	2.4.2 評価データ	38
	2.4.3 オクルージョンに対する振る舞い	39
	2.4.4 統計的評価	41
	2.4.5 <b>実データへの適用</b>	43
2.5	議論	46
2.6	結言	48
第3章	非言語行動に基づく会話構造の推定	49
3.1	緒言	49
3.2	会話レジームの設定	53
3.3	会話モデルと推定アルゴリズム..................................	56
	3.3.1 会話モデルの構造	57
	3.3.2 ギブスサンプリングによるベイズ推定	59
3.4	頭部方向の推定方法と会話構造推定への適用.................	63
	3.4.1 会話シーン分析における要求条件	63
	3.4.2 頭部方向推定の従来研究	63
	3.4.3 疎テンプレート Condensation 追跡法 (STC 法)	65
	3.4.4 STC 法の会話構造推定への適用	67
3.5	実験1 ―磁気センサによる頭部方向計測に基づく推定―	68
	3.5.1 会話データの収集と準備	69
	3.5.2 ハイパーパラメータの設定	70
	3.5.3 ギブスサンプリング	72
	3.5.4 評価 1: 視線方向の推定精度	72
	3.5.5 評価 2: 会話レジームについての評価	75
	3.5.6 評価のまとめ	80
3.6	実験2 ―トラッキングによる頭部方向計測に基づく推定―	80
	3.6.1 頭部方向の推定精度	81
	3.6.2 推定された視線方向の評価	84
	3.6.3 推定された視線パターン,及び,会話レジームの様子.........	84

	3.6.4 会話レジームの評価	85	
3.7	議論	85	
3.8	結言	88	
第4章	記憶に基づく気象レーダ画像パターンの予測	91	
4.1	緒言	91	
4.2	記憶に基づく予測の枠組みと実現手法	94	
	4.2.1 特徴抽出	94	
	4.2.2 特徴の時系列の表現法	99	
	4.2.3 時系列パターンの検索	100	
	4.2.4 予測画像の生成	103	
4.3	実験	103	
	4.3.1 データセット	104	
	4.3.2 類似パターン検索に関する実験と評価	104	
	4.3.3 予測に関する実験と評価	108	
4.4	結言	113	
第5章	結論 1	15	
謝辞	1	.19	
参考文南	考文献 12		
研究発表	研究発表一覧 13		

## vii

# 第1章 序論

## 1.1 まえがき

本論文は,時系列画像に基づく動的事象の推測に関する著者の研究成果をまとめたものである. 本研究は,人間などの生物がもつ視覚機能を人工的に実現することを目指す学問であるコンピュー タビジョンの中でも,特に動的な事象を観測対象とする時系列画像解析と呼ばれるに属する.時 系列画像解析では,カメラなどの手段を用いて観測される2次元画像の時系列から,そこに映って いる3次元世界中の対象の状態やその時間変化に関する推測が行われる.ここで観測情報である 2次元画像は,3次元世界を2次元平面へ射影して得られる縮退された情報であり,その観測情報 から対象の状態を一意に決定することは必ずしも容易ではない.本論文では,この時系列画像に 基づく推測の困難性を,観測の部分性と間接性という2つの側面から捉え,それらに対処可能な 時系列画像解析のアプローチを提案する.ここで観測の部分性とは,複数の対象間の隠れ(オク ルージョンとも呼ぶ)や観測範囲の制限による観測情報の欠落のことを指し,観測の間接性とは, 観測情報と推測対象の状態との間の質的な乖離のことを指す.本論文では,観測の部分性,間接 性の度合いの組み合わせに応じて,以下に述べる3つの問題を取り上げる.

まず,観測の部分性が顕著な例として,複数の物体の間に生じるオクルージョンに着目する.オ クルージョンとは,カメラに対して,手前の物体が後方の物体を遮蔽する現象であり,時系列画 像解析を妨げる要因の一つとして知られている.オクルージョンが生じる場合,遮蔽された領域 に関する観測情報が得られないという問題が生じる.また,画像上で複数の対象のオーバーラッ プが生じ,どの物体が見えていて,どの物体が隠れているかという観測情報の解釈が困難になる という問題も生じる.本論文では,複数物体の追跡(トラッキングとも呼ぶ)を課題として取り 上げ,対象間のオクルージョンを含めた観測過程を陽にモデル化し,オクルージョンの空間的な 構造と対象の状態(位置や姿勢)を同時に推定することで,オクルージョンに対して頑健に追跡 を行う方法を提案する.

次に,観測の間接性が顕著である例として,事象を構成する複数の要素間の相互作用に着目す る.要素間の相互作用には,物理的な衝突のような視覚的に観測可能なものの他,遠隔的に作用 し視覚的に直接観測できない事象も数多く存在する.本研究では,視覚的に観測できないが,事 象のダイナミクスに対して相互作用が決定的な影響をもつ事例として,複数人物の対面会話シー ンを取り上げる.この会話シーンにおいては,各人物の動作は時系列画像として比較的容易に観 測できるものの,それら動作を支配するダイナミクスは,人物間の相互作用(インタラクション とも呼ぶ)に依存し,非常に複雑である.ここでは相互作用として,話し掛け,及び,傾聴の動作 に着目し,人物間でのメッセージの流れのパターンを推測の対象とする.本研究では,これを会 話構造と呼び,この会話構造と参加者の行動との関連性を動的ベイジアンネットワークを用いて モデル化し,観測された参加者の行動から会話構造の推定を行う方法を提案する.

さらに,観測の部分性と間接性がともに顕著な事例として,画像パターンとして観測される自 然現象の予測の問題を取り上げる.本研究では,自然現象の中でも気象現象,その中でも特に降 水現象に着目し,その観測手段として気象レーダを用いる.気象レーダは,降水強度の空間分布 を画像パターンとして観測する装置である.そのため,これを用いることで降水の空間分布の予 測を行う問題は,気象レーダ画像上のパターンの時間変化を予測する問題として捉えることがで きる.本論文では,このパターンのことを気象レーダ画像パターンと呼ぶ.しかし,このパター ンは,非剛体の不定形パターンであり,様々な気象的な要因が重なり合った複雑なダイナミクスを 持つため,その予測は容易でない.気象レーダの観測は,パターンの元となる降水現象の空間的 な広がりに対して,レーダから一定範囲内,かつ,ある空間断面上に限定された観測であり,降 水現象の全容を捉えられない.その観点から観測には部分性が伴うといえる.また,気象レーダ では,気温,気圧,風速といった気象現象の内部状態が直接的に観測できず,さらに,現時点の観 測情報から未来のパターンを予測するという問題の性質からも,観測の間接性が顕著である.こ れらの観点より,非剛体不定形パターンとして観測される気象レーダ画像パターンの予測の問題 は,観測の部分性,間接性ともに顕著な事例といえる.本研究では,このような時系列パターン の予測の方法として,過去の画像パターンの記憶に基づく予測法を提案する.

以上のように,本論文は,従来の時系列画像解析において未だ解決法が確立されていない観測 の部分性,間接性の問題に着目し,新しいアプローチを提案するものである.具体的な提案方法 は,各々の事象に特化したものであるが,その根底にある方法論は幅広い事象に適用できるもの であり,今後の時系列画像解析分野の発展に資することが期待される.

本章では,まず,1.2節で研究の背景と従来研究について概観し,次に,1.3節では本研究の目 標と位置づけを述べる.最後に1.4節で本論文の構成を述べる.

## 1.2 研究の背景と従来研究

視覚は,人間を含めた生物にとって外界に関する知識を得るための重要な感覚である[1,2,3]. このような視覚の機能を人工的に実現することを目指して,過去,四半世紀以上に渡って,コン ピュータビジョンの研究が盛んに行われている[4,5,6,7].コンピュータビジョンにおいては,外 界の観測手段として生物の目に代わって主にカメラを用い,視覚情報の処理を行う脳に代わって 計算機が用いられる.コンピュータビジョン分野の中でも,特に動的な事象を観測対象とする領



図 1.1: コンピュータビジョンの観測過程.

域のことを,時系列画像解析や動画像解析などと呼んでいる(以後,時系列画像解析と統一して 呼ぶ).時系列画像解析においては,時系列的に観測される2次元画像の集合を入力とし,その画 像集合から元の3次元世界中の事象の推測が行われる.一般に事象とは,自然環境や社会環境内 で生起する変化全般を指す用語であり,物体の移動などの運動学的変化,形状の変形などの力学 的変化,酸化などの化学結合の変化など様々なものが事象として存在する.コンピュータビジョ ンにおいて推測の対象となる事象には,物体の位置や姿勢,形状などの物理的な状態とその変化, それら変化の種別や要因などが含まれる.前者の物理的な状態の時間変化を推測する問題は,対 象追跡,トラッキングなどと呼ばれており[8],後者の変化の種別や要因を推測する問題は,一般 的に時系列パターン認識と呼ばれている[9].代表的な追跡の対象としては,人物や車両などの物 体の位置やその変化,人物の姿勢などがあげられる.また,代表的な時系列パターン認識の対象 としては人物のジェスチャなどがあげられる.これらのタスクは,遠隔監視やマン・マシン・イ ンタフェース,ロボットビジョンなど様々な産業への応用が期待されている[4,6,7].

コンピュータビジョンにおける3次元世界の観測は,図1.1のように,3次元世界を2次元の画 像に投影するプロセスとして捉えられる.実際の観測には,カメラやレーダなどが用いられる.例 えば,カメラを用いる場合,観測される画像上の濃淡分布は,物体表面の物理的・光学的特性や 照明の状態,物体の形状,カメラと物体の位置関係,カメラの特性,物体表面からカメラに至る 経路上の大気の性質などに依存し,同一対象を観測する場合においても多様に変化する.このよ うに観測される2次元の画像は,一種の縮退された視覚情報であり,そこから元の3次元世界中の 事象を推測する場合,必ずしも対象とする事象の状態が一意に決定できるとは限らない.本論文 では,このような観測情報から事象の状態が一意に決定できないことの要因として,観測の部分 性と間接性に着目する.観測の部分性は,図1.2のようなオクルージョンと呼ばれる光学的な遮蔽 現象や,図1.3のようなカメラなどの観測範囲の制限により生じる観測情報の欠落のことを指す. また,観測の間接性とは,観測情報と推測対象の状態との間の質的な乖離のことを指す.観測の



図 1.2: 観測の部分性(オクルージョンの例).



図 1.3: 観測の部分性(観測範囲の制限の例).

間接性が顕在化する例としては,2次元画像上の時間変化として観測される人物の行動からその行動の意図を推測する場合などがあげられる.

これらの観測の部分性と間接性は、時系列画像解析の障壁となっており、従来は、これらが少 ない撮影環境や事象が解析の対象とされてきた.本研究では、時系列画像解析の適用範囲の拡大 を目指し、これまで十分に対処法が確立されていない範囲の観測の部分性・間接性に対する新し いアプローチを提案する.具体的には、観測の部分性の要因として、物体間のオクルージョンに 着目し、具体的なタスクとしては、複数物体の追跡の問題を取り上げる.また、観測の間接性の 要因としては、複数の対象間の相互作用に着目する.その具体的なタスクとしては、複数人物の 会話シーンにおける人物間の相互作用の推測の問題を取り上げる.さらに、観測の部分性、間接 性がともに顕著な例として、非剛体不定形パターンとして観測される自然現象に着目する.具体 的には、気象レーダ画像パターンの予測の問題に取り組む.

以下では,それぞれについて研究の背景や従来研究を概説する.また,その参考として,図1.4 には,この3つの事例について,従来研究によって対処可能な観測の部分性,間接性の度合い,及 び,本研究で対象とする領域を図示する.



図 1.4: 従来研究に対する本研究の位置づけ.

#### 1.2.1 観測の部分性 — オクルージョン—

観測の部分性の主要な要因の一つとして,オクルージョンと呼ばれる光学的な遮蔽現象がある. オクルージョンとは,図1.2に示すように,カメラに対して手前の物体表面が後方の物体を遮蔽 する現象であり,視覚を用いて外界を観察する際の本質的な障壁として知られている.オクルー ジョンが生じる場合,遮蔽された領域に関する観測情報が得られないという問題が生じる.また, カメラや物体の位置関係に依存して,どの物体が見えていて,どの物体が隠されているかという オクルージョンの構造が変化する.このオクルージョンの構造に依存して,画像上では複数の物 体がオーバーラップして映るため,観測された画像情報の解釈が困難になるという問題も生じる. ここでオクルージョンの種類として,自己オクルージョンと相互オクルージョンがあげられる.自 己オクルージョンとは,同一物体の一部の表面により他の部分の表面が遮蔽される場合のことを 指し,相互オクルージョンとは,異なる物体間で生じるオクルージョンのことを指す.従来の時 系列画像解析において,オクルージョンは忌避されている現象であり,事前にオクルージョンが 生じないような撮影状況を設定することが望まれている.しかし,時系列画像解析の適用範囲を より拡大するため,近年ではオクルージョンが発生している状況においても頑健に動作する手法 が強く求められている.

このようなオクルージョンの発生が重大な障害となる時系列画像解析のタスクとして,対象の

追跡がある.追跡はトラッキングとも呼ばれ,狭義には対象の位置を時間的に追跡することを指す が,本論文では,より広義に対象の位置や姿勢などの状態を逐次的に推定することと捉える.一般 に追跡の問題は,現時点までに観測された情報に基づいて,対象の現時点での状態を推定するとい う時系列フィルタリングの枠組みで定式化される.時系列フィルタリングでは,状態空間モデル に基づいて対象の状態と観測情報との関係が記述される.状態空間モデルは,状態変数,観測値, 観測過程のモデル,及び,ダイナミクスモデルから構成される.状態変数は隠れ変数であり直接観 測できないが,その代わりに,その状態に依存した観測値が入手できるものと仮定される.観測過 程のモデルは,観測モデル,出力モデルなどと呼ばれ,状態変数と観測値との関係を記述するモデ ルである.ダイナミクスモデルは,状態遷移モデルや運動モデルなどと呼ばれ,系の状態の時間変 化を記述するためのモデルである.時系列フィルタリングの実行には,仮定される状態空間モデ ルの性質に応じて,カルマンフィルタ[10,11,12]やパーティクルフィルタ[13,14,15,16,17,18] などが用いられる.

時系列フィルタを用いた追跡の典型的な流れは以下のようである.まず,各時刻において,画 像上から画像特徴が抽出される.画像特徴の例としては,濃淡分布の局所的極大・極小点,コー ナー,対象領域の重心位置,大きさ,輪郭,色,低解像度の濃淡分布などがあげられる.次に,前 時刻から予測された状態の確率分布に基づいて,対象と画像特徴との対応付けが行われる.この 過程は data association とも呼ばれる [19, 20, 21, 22].この対応付けは,画像上での特徴の位置 と対象の予測位置との距離に加えて,色,形状,アピアランスなどの類似性などに基づいて行わ れることが一般的である.対応付けが完了した後,対象の状態の更新が行われる.最後に次時刻 における対象の状態の確率分布が,ダイナミクスモデルに従って予測される.以後は,同様の処 理を各時刻において繰り返し行うことで,対象状態の時間変化が推測される.これまで,着目す る画像特徴の違いなどにより様々な追跡の方法が提案されている.例えば,古典的な追跡として, 画像上で対象の領域を抽出し,その重心の追跡を行う方法があるが,この方法では,追跡の対象 は画像上の質点として捉えられ,オクルージョンは考慮されない.よって,対処可能な観測の部 分性の程度は低いと想定され,図1.4の中では a1 として位置づけられる.

また,従来,対象追跡においてオクルージョンに対処するためのアプローチが幾つか考えられ ている.その一つとして,物体を輪郭で表現し,2物体間のオーバーラップ(重なり)の状態を許 容しつつ追跡を行う方法が提案されている[23].この方法では,平面上を運動する物体を斜め上 方向から撮影することを仮定しており,画像座標上で,上方に位置する物体が常に重なりの背後 に位置するというような物体の重なり方に制限が設けられている.また,これらの制限を設けず, 2物体間での重なりの順序の反転を許容する方法も提案されている[24].この方法では,カメラに 対する複数の物体の奥行き順序(depth order)を推定し,オクルージョンによって生じる画像上 の物体の重なりを解釈する機構が導入されている.しかし,単一視点の観測の性質上,奥行きを 推定するための手掛かりに乏しく,適用可能な範囲が2物体程度の部分的なオクルージョンと限 定されたものに留まっている.また,これらの方法では,ある物体が他の物体の影に完全に隠れてしまうという状態は想定していない.図1.4では,これらの方法は a2 付近に位置づけられる.

また,画像上での追跡において,一時的に対象が完全に隠れてしまう現象(完全オクルージョ ンと呼ぶ)に対するアプローチとして,オクルージョンが発生している時間の前後において,追 跡対象の対応付けを行うアプローチが知られている[25,26].このアプローチでは,対象の消失か ら再出現までの区間において,対象の運動の連続性・継続性や,色や形状の不変性を仮定する必 要があるため,適用可能なシーンが限定される.また,このアプローチでは,オクルージョンが 発生している区間において物体の位置を推定する仕組みを持たない.これらの方法は,図1.4の中 では,a3として位置づけられる.

一方,近年,3次元空間中における複数の対象の位置を,オクルージョンに対して頑健に追跡する アプローチとして,多視点観測に基づくアプローチに注目が集まっている[27,28,29,30,31,32,33]. このアプローチは,空間中の複数の位置にカメラを設置し,観測できない領域を減らすことで,観 測の部分性の解消を狙ったものである.このアプローチでは,各視点で得られる対象の画像特徴 を統合し,追跡が行われる.これにより,一部の視点においてオクルージョンが存在する場合に おいても,他の視点の情報によって欠落した観測情報を補うことができ,各物体の空間中におけ る位置を頑健に推定することができる.この多視点観測に基づく手法では,単一視点の手法と比 較し,飛躍的な耐オクルージョン性の向上が実現されている.しかしながら,オクルージョンの 影響の少ない視点が十分に存在することを暗黙のうちに仮定しており,物体が全ての視点からみ て隠蔽されているような重度のオクルージョン(完全オクルージョン)には十分に対処できない. また,これらの従来法では,複数視点間の情報の統合方法がヒューリスティクスに基づくものが 多く,オクルージョンの現象を正確に取り扱うことは,その複雑性から避けられてきた.これら の方法は,図1.4の中では,a4として位置づけられる.

1.2.2 観測の間接性 — 複数要素間の相互作用—

実世界の事象は,複数の対象の相互作用により構成されることが一般的であり,各対象それぞ れの状態だけではなく,それらの間の相互作用の推測も時系列画像解析のタスクとして重要であ る.このような要素間の相互作用には,物体の衝突のような物理的な近接や接触といった視覚的 に観測可能なものの他,物理的な接近や接触などを伴わず,遠隔で作用するために,視覚的に観 測されにくいものも存在する.そのような場合,相互作用に伴って生じる各対象の変化などを手 掛かりとした間接的な推測が必要とされる.本論文では,このような相互作用の中でも特に重要 な事象として,複数の人物間の相互作用に着目する.

従来,時系列画像解析の分野において,人物の行動の認識や理解は重要なテーマとして研究が 進められている.その代表例としては,前節でみた人間の歩行動作の追跡があげられる[25,31, 32, 27, 28, 33].また,頭部や肢体の姿勢・動作の認識に関しては,1980年代にNTTの末永・間 瀬らのグループが世界に先駆けて研究を開始し,ハンドリーダ,ヘッドリーダ,ヒューマンリーダ [34, 35, 36, 37] などの成果を残している.その後,この分野には多数の研究者が参入し,現在まで に様々な方法が提案されている[38, 39, 40, 41].これら身体の物理的な位置や姿勢を推定する問題 は,2次元の画像上での位置から3次元空間中での状態を推測するという観点から,一種の観測の 間接性を伴う問題であるといえる.しかし,適切なカメラモデルやダイナミクスモデル,対象の 形状モデルなどが与えられた場合には,推定される状態の不確定性は低く抑えることが可能であ る.そのため,これらの手法において対処すべき観測の間接性は比較的小さいと位置づけられる.

また,人間のジェスチャを対象として,その認識を時系列パターン認識の枠組みで行う方法も提 案されている[42,43,44,9].時系列パターン認識では,まず,認識対象とする時系列パターンの カテゴリを複数個設定し,各カテゴリを代表するサンプル集合を収集し,モデルを作成する.そ のモデルへの入力情報としては,時系列画像の各フレームにおいて抽出される対象の位置・姿勢 や画像特徴などがあげられる.モデルの構築後,新たに入力される時系列データに対して,最も 適合したモデルのカテゴリが認識結果として出力される.時系列パターン認識のための人物動作 のモデルとしては,隠れマルコフモデル(Hidden Markov Model;以後,HMM)が多く用いられ ている[45,46].HMMを用いた時系列パターン認識は,音声認識における標準的な方法として古 くより用いられているが[45,46],コンピュータビジョン分野においては,大和らのテニスのスイ ング動作の認識[47]に端を発し,その後,手話認識[43]やマン・マシン・インタフェースなど様々 な事象の認識に適用がなされている.ジェスチャを人間の意図の表出と捉えると,推定対象の状 態(意図)と観測された画像上のジェスチャとの間に質的な乖離があるといえる.しかし,一般 にジェスチャ認識においては,ジェスチャの行動としての認識が主に対象とされており,観測の 間接性は比較的少ないといえる.そのため,図1.4の中では,b1として位置づけられる.

また,近年では,時系列画像から人間の心的状態や意図を推測するという課題に注目が集まり つつあり,現在,顔の表情の認識や,顔表情から心理状態を推測する技術の開発が進んでいる.顔 の表情の認識は,顔面のオプティカルフローとして表情変化を捉えて,認識を行う方法や,顔部 品の位置関係や濃淡分布など様々な特徴量を用いた方法など様々な方法がこれまでに提案されて いる[48,49,50,51,52,53].また,人物の表情を心理状態の表れであると考え,心理状態と表情 との間の関係をベイジアンネットによってモデル化して,感情認識を行う方法も提案されている [54].感情認識においては,人物の心理状態と画像として観測される顔の表情の間に質的な乖離が 存在し,その観点から観測の間接性が含まれる事例といえる.図1.4の中では,b2として位置づ けられる.

以上で述べた方法は,いずれも一人物の動作・行動を対象としたもので,複数の人物間の相互 作用に関する研究は,いまだ初歩的な段階にある.例えば,複数人物の歩行シーンを対象に人物 の出会い,別れ,連れ添って歩くなどの相互作用を認識対象とし,これらを画像上での人物間の 距離の変化から認識する方法 [55] などが提案されるに留まっている.

本研究では,人物間の心理的な相互作用が重要な働きを持つ現象として,対面会話シーンに着 目する.複数人物による対面会話は,情報の伝達・共有,他者の意図・感情の理解,集団の意志 決定などにおいて欠かすことのできないコミュニケーションの形態である、対面会話おいて伝達・ 交換されるメッセージには, 音声による言語的なメッセージのほか, 顔の表情や視線, 手振り・身 振り,姿勢,声の調子・韻律など非言語的なメッセージも多く含まれ,それらは重要な役割をもつ ことが知られている[56,57,58,59,60].過去の心理学研究によると,会話中に他者に与える印象 のうち,非言語的なメッセージによるものが7割から9割以上占めるという報告もあり[57,61], 非言語的なメッセージの重要性が伺える.対面会話において人間は,受信したメッセージから他 者の意図や心的状態を推測し,また,自らの意図を他者に伝達すべく,非言語的な行動によりメッ セージを発する.このような会話参加者の相互作用によって会話という現象が進展する.ここで, 顔の表情,視線,手振り・身振り,姿勢などの非言語的なメッセージの多くは,視覚のチャネルを 通して伝達される.したがって,会話参加者は,視覚情報として他者の行動を観測し,そこから 他者の意図や心的状態を推測し、会話の状況を理解するというタスクを実行していると考えられ る.対面状況においては,会話参加者は,一般的に互いの顔や身体が十分に観察できる位置関係 にあり,視覚的な観測のプロセスは比較的単純なものと仮定できる.しかしながら,会話現象の進 展は,各人のメッセージの表出・解釈というプロセスを介した相互作用に依存し,複雑なダイナ ミクスを有する.このような複数人物による会話をカメラやマイクロフォンなどを通じて観測し, そこから会話中の相互作用として,例えば,誰が誰に向かって話をしているのか?誰が誰の話を 聞いているのか?という会話の状態を自動的に推定する問題を考えた場合,この人物間の相互作 用は、それ自体直接的に観測されるものではなく、観測される人物の行動から間接的に推測する ことが求められる.そのため、このような複数人物による対面会話シーンの認識・理解は、観測 の間接性が大きい課題であり、従来の時系列画像解析法にとって難しい課題であると考えられる.

観測の部分性,間接性がともに顕著である事例として,非剛体・不定形の時系列画像パターンとして観測される気象現象に着目する.そのなかでも,気象レーダ画像として観測される降水パターンの予測の問題を取り上げる.気象現象は,微視的には大気中の分子の相互作用によって引き起こされる現象であるが,その空間スケールは地球規模まで広い幅をもち,また,数時間の局所的な雷雨から,気候変動まで極めて大きい時間スケールの幅を有する.また,スケール間での相互作用 [62,63] や,カオス的な振る舞いのため,予測が難しい対象としても知られている[64,65].本研究では,気象現象の中でも,降雨や降雪などの降水現象に焦点を当て,その局所的(数 km~1000km)で短時間(30 分から 3 時間先),かつ,高解像度(1km~3km 単位)の降水分布の予測に着目する. この局所・短時間の降水予測は,防災や農業,治水など産業・行政において重要な情報として求められている[66,67].

現在,降水現象を高解像度かつリアルタイムに観測する一つの手段として,気象レーダと呼ば れる装置が一般的に用いられている.気象レーダは,降水強度の空間分布を画像として得ることが できる装置である[68,69].その原理としては,気象レーダからマイクロ波を細いビーム状のパル スとして発信し,それが雲の中の雨粒や氷粒によって散乱される現象を利用したものである.発 信した電磁波の方向,及び,散乱されて返ってくるまでの時間から,降水粒子の空間分布がわか り,また,散乱されて返ってくる電磁波(レーダエコーと呼ぶ)の強度から,降水強度が推定でき る.従って,降水分布の予測の問題は,気象レーダ画像上のパターン(レーダエコーパターンと も呼ぶ)の時間変化の予測の問題として捉えることができる.

気象レーダによる観測は,空間的にみて,地球上のごく狭い領域(半径~数100km)のみの観測 であり,さらに,3次元空間中のある断面を切り取ったものである.また,降水が存在する領域以 外の情報は得られない.これらの観点からレーダによる降水現象の観測は,部分性が顕著である と言える.また,気象レーダでは,気温,気圧,風速といった気象現象のダイナミクスを記述す るための内部状態を直接的に観測できない.それにより気象現象の結果として生じる降水の空間 分布に対応した気象レーダ画像パターンのダイナミクスも直接的に知ることができない.その観 点から観測の間接性も顕著であるといえる.さらに,現時点の観測情報から未来のパターンを予 測するという問題の性質からも,本質的に観測の間接性は避けられないと言える.このように気 象レーダ画像上のパターンの予測の問題は,観測の部分性,間接性の双方が顕著な事例といえる.

従来,気象レーダ画像は,気象予報士により目視で観察され,予報を作成する際の参考情報とし て用いられてきたが,近年では,予報の自動化や予報士の支援のために,自動的に気象レーダ画像 上のパターンの時間変化を予測する技術の研究開発が行われている.気象分野における時系列画像 解析の例としては,古くより気象衛星画像中の雲の動きを計測する方法が知られている[70,71,72]. 気象衛星により撮影された画像は,雲を可視光や赤外光で捉えたものであり,その動きは,物質の 移動現象であり短時間の間では,ほぼ,そのパターンの形状が保存される.そのため画像フレー ム間での相互相関法などにより動きベクトルを精度良く計算でき,得られた雲の動きベクトルは, 雲頂高度における高層風の近似とみなせ,数値予報の初期化などに利用されている[73].

一方,この方法を参考にして,気象レーダ画像上での降水パターンの動きを計測して,予測に 活用する方法も提案されている.例えば,計測された動きに従ってパターンを平行移動させると いう線形外挿法が知られている[74,75].この方法は,現在の画像上のパターンを時間的な変形の ない一つの塊とみなし,また,その動きを等速直線運動と仮定するなど極めて単純な方法であり, その性能は低いものに留まっている.この手法は,時系列パターンの予測の問題を,観測の部分 性,間接性が小さい問題へと近似して解いていると位置づけられる(図1.4 中の c1).

また,より現象を考慮した方法として移流拡散モデルを用いた方法が提案されている[76,77].

これは,気象レーダ画像パターンのダイナミクスを一種の流体と仮定して予測を行う方法である. しかしながら,気象衛星画像上の雲と異なり,気象レーダ画像上のパターンの動きは,物質(水粒 子)の移動現象を直接捉えたものではない.このパターンの動きは,おおよそ大気の流れにそって 移動する傾向はあるものの,パターンを構成する要素の動きの方向とパターン全体の動きの方向 は必ずしも一致しないことが知られている[78].また,個々の要素は生成消滅を繰り返す一方で, パターン全体の形状は長時間維持される場合もある[79].このようにレーダ画像上のパターンは, 極めて複雑に変化し,従来の方法では,特にパターンの時間変化が大きい場面等に対応できず,そ の適用範囲は限定的なものに留まっているのが現状である.なお,この方法では,観測された画 像系列から,パターンが存在しない領域を含めた画像全域に渡る速度ベクトル場(及び,その時 間変化)を推定する必要がある.この速度ベクトル場の推定において,観測の部分性,間接性の 双方が存在すると位置づけられる(図1.4 中の c2).

## 1.3 本研究の目的と位置付け

前節で見たように,従来の時系列画像解析においては,観測の部分性,間接性が大きい事象に 対処できる方法がいまだ確立されていないのが現状である.本研究では,時系列画像解析の適用 範囲をより広範囲の実世界事象へと広げるために,図1.4中の三角形の領域で示される,従来では 対処が困難である領域を対象とし、新しいアプローチを提案する.まず、観測の間接性は比較的 小さいものの,観測の部分性が大きい事象として,複数物体間の相互オクルージョンに着目する. また,具体的な推測のタスクとして,複数物体の追跡の問題を取り上げ,多視点観測におけるオ クルージョンの空間的な構造と、物体の位置・姿勢とを同時に推定することで、耐オクルージョン 性を向上させるというアプローチを提案する.このテーマをここでは「多視点オクルージョン解 析による複数物体の追跡」と呼ぶ.なお,図1.4中ではa5に位置づけられる.次に,観測の部分 性は比較的小さいものの,観測の間接性が大きい事象として,複数人物間の相互作用に着目する. 特に,代表的な相互作用の例である複数人物の対面会話シーンを取り上げ,会話参加者の行動の 視覚観測から,参加者間の相互作用を規定する会話の構造を推定するというアプローチを提案す る.このテーマをここでは、「非言語行動に基づく会話構造の推定」と呼ぶ.なお,図1.4中では b3 に位置づけられる.さらに,観測の部分性,間接性がともに大きい事例として,非剛体・不定 形の画像パターンとして観測される自然現象に着目する.本論文では,特に,気象レーダ画像上 のパターンとして観測される降水現象に着目し、そのパターンの変化を予測するというタスクに |取り組む.本論文では,そのための新しいアプローチとして,過去の記憶に基づいて予測すると いうアプローチを提案する.このテーマを「記憶に基づく気象レーダ画像パターンの予測」と呼 ぶ.なお,図1.4中ではc3に位置づけられる.

以下では,この3つのテーマについて,そのアプローチの概略と位置づけを述べる.

#### 1.3.1 多視点オクルージョン解析による複数物体の追跡

本研究では、1.2.1 節で紹介した背景を踏まえ、多視点観測による複数物体の追跡において、物体間の相互オクルージョンに対するロバスト性を向上させるためには、多視点の観測情報を単に ヒューリスティックに統合するのではなく、オクルージョンという現象を陽に分析し、各視点から どの物体が見えていて、どの物体が隠れているかというオクルージョンの構造を正しく推定する 必要があると考える.そこで、本研究では、従来、あまり省みられていなかったオクルージョン の構造を陽に定義して、推定を行うアプローチを新たに提案する.具体的には、2次元平面上のパ ラメトリックな形状をもつ物体と、同平面上の固定視点からなる単純化された世界を考え、オク ルージョンの空間的な構造を推定しつつ、複数対象の状態推定を行うことで、オクルージョンに 頑健に複数物体の追跡を行う方法を提案する.本提案方法は、すべての視点から物体が隠れるよ うな場合にも対処が可能であると期待でき、また、対象の画像上での位置ではなく、空間中の位 置・姿勢を推測対象としている点より、図1.4 では a5 に位置づけられると考える.本研究は、多 視点観測におけるオクルージョンの空間的な構造を陽に推定するというアイディアを提案した点 において新規性が認められ、多視点オクルージョン解析という分野を新たに提案するものと位置 づけることができ、今後、様々な追跡技術の基礎となるものと期待される.

#### 1.3.2 非言語行動に基づく会話構造の推定

本研究では,複数の人物間の相互作用が重要な場として,複数人物の対面会話シーンに着目す る.対面で行われる会話は,情報通信技術の発達した今日にあっても,我々人間の社会的活動に おいて欠かすことができないものである.しかし,対面の会話は,会話参加者全員がその場に居 合わせることが必要であり,また,そこで取り交わされる情報の内容も各人の記憶のみに留まり 通常,形として残ることがない.現在,このようなその場限りという会話の限界を超え,より人 間同士のコミュニケーションが円滑になるようサポートする技術として,遠隔会議[80],映像議事 録[81],会議映像の自動編集[82],会話エージェント[83,84]や会話ロボット[85,86]などへ大き な期待が寄せられている.その上で,これらのアプリケーションの実現のために,人間同士の会 話が理解できる計算機の実現が強く求められている.

現在まで,会話シーンを対象とした工学的な研究は,音声分野において先行している.その代 表的な課題としては話者検出があげられる.また,近年では,複数人物の音声ストリームから,会 話をしている人物の組を検出する研究や,2つの音声ストリーム間において発話・応答ペアを検出 し,そこから複数人物間の関係性を見いだす方法などの検討が進められている[87,88,89,90].ま た,近年では,音声情報に画像情報も加えたマルチモーダルな会話シーンの認識にも注目が集ま りつつある.例えば,会話中の動作,プレゼンテーション,ディスカッション,モノローグなどの シーンを,HMM などを用いて認識する手法が提案されている[91,92].この手法では,画像上で の人物の位置・動作と認識対象のカテゴリとの対応関係が明確であり,かつ,対象となる部屋や 会話参加者ごとの事前の学習が必要であるという制約条件があり,初歩的な段階に留まっている. また,人物の行動の生成要因となっている人物の内部状態や相互作用のモデル化・推定も検討の 対象とはしていない.これらの研究に加えて,近年では,実際の会話の場面を,複数のカメラや センサーなどを用いて撮影・データ収集するシステムの開発 [93,94,95] や,得られた会話データ の分析によって会話参加者の行動分析を行う研究 [96,97] が盛んになりつつあり,今後,この種の 研究分野は,大いに発展することが見込まれている.

本研究では、以上の背景を踏まえ、複数人物の会話シーンを理解するためには、人物の行動を 生じせしめる人物の内部状態や相互作用を含む場の状態,そのダイナミクスまで踏み込んだモデ ル化と推測が不可欠であると考える.そこで,その初歩的な段階として,本論文では,会話の場に おいて,人物間でやりとりされるメッセージの流れを明らかにすることを目指す.このメッセージ の流れは,各参加者の参与役割(発話者や聞き手)[98,99]や,誰が誰に話し掛けているか,誰が 誰の話を聞いているかという会話行動によって示唆される.本論文では,このメッセージの流れ のことを会話構造と呼ぶ.また,その推測の手掛かりとして,各参加者の視線行動(誰が誰を見 ているか)と会話構造との関係に着目する.対面会話において,視線は,数ある非言語行動の中 でもとりわけ重要な役割を果たしていることが知られている [100, 101, 102]. その役割としては, 他者のモニタリングの他,自らの態度や意図の表出,ターン交替などのタイミング制御などがあ るといわれている [103].また,話し手と聞き手の間の相互の注目は,会話を構成する上で不可欠 な要素であるという指摘もある [104]. さらに,視線の方向は,注意(アテンション)の方向性の キューの一つであり,人間の意図を示唆するものとしてコミュニケーションにおいて重要であるこ とが知られている [102, 105]. しかし, このような視線の方向は, 現状, 画像から直接的に計測す ることは困難であるため,本研究では,その推測の手掛かりとして頭部方向[106,107]に着目し, 観測された各人の顔画像から頭部方向を推測する方法を導入する.このような考えに基づき本研 究では,会話のモデルとして,頭部方向と視線方向の関係,及び,視線方向・発話状態と会話の 構造との関係を,単一の動的ベイジアンネットを用いてモデル化する.このモデルでは,聞き手 は話し手の方を見る傾向がある」というような、会話の構造と参加者の行動との関係が表現され る.本研究は,時系列画像解析を会話シーン分析へ適用した先駆的な例であり,今後,会話を対 象とした研究のための足掛かりを提供するものと位置づけることができる.

対面会話シーンにおける人物間の相互作用は,画像として直接的には観測できず,観測可能な人物の行動から間接的に推測を行う必要がある.そのため,観測の間接性は大きいといえる.また, 人物行動の観測において,カメラと人物の位置関係によって自己オクルージョンが生じるケース も多く(例えば,頭部を正面から横顔に回転させるなど),ある程度の観測の部分性にも対処する ことが求められる.これらの理由により,提案法は,図1.4中においてb3として位置づけられる.

#### 1.3.3 記憶に基づく気象レーダ画像パターンの予測

本研究では,観測において部分性,間接性がともに大きい事例として,画像パターンとして観 測される降水現象の予測という問題に着目する.特に,観測手段として気象レーダを用い,観測 される気象レーダ画像上のパターンの予測を目標とする.1.2.3節で述べたように,気象レーダ画 像上のパターンは,降水の空間分布に対応しているため,このパターンの時間変化を予測するこ とで,降水分布の予測が可能とされる.しかし,このレーダ画像上のパターンは,非剛体,かつ, 不定形であり,そのダイナミクスは極めて複雑である.そのため,従来法では,特にパターンが 大きく変化するような状況へ対処ができなかった.

本研究では、このような時系列画像パターンに対処するため、記憶に基づく予測と呼ぶアプロー チを導入する.これは、類似する気象現象が不定期的に再現されるという性質に基づいて、長期 の観測により蓄積したデータを用いて、現時点の観測情報に類似する過去の事例を検索し、現時 点の観測情報との類似性が認められた場合、その観測の元となっている現象そのものも類似して いるという仮定のもと、現時点の現象も過去と同じような変化を辿るであろうと予測するもので ある.本研究では、この考えのもと、現時点の気象レーダ画像を入力とし、それと類似する過去の 気象レーダ画像を検索し、検索された過去の画像の時間変化に基づき、未来のレーダ画像パター ンを予測するという方法を提案する.この方式は、観測される画像の生成要因となっている現象 そのものに関するモデルや知識が不要であり、事象の観測過程、及び、ダイナミクスのモデル化 が困難な本研究の事例に適合すると考えられる.このような記憶に基づく予測法によって、気象 レーダの観測範囲内の直前の数フレーム内のパターンの動きだけではなく、より広範囲・長時間に わたる現象の変化を反映した予測が可能になると期待できる.これにより、例えば、画像パター ンの平行移動や移流・拡散としてモデル化できる範囲を超えて大きくパターンが変化するような 状況にも対処が可能となると期待できる.これらの観点から、記憶に基づく予測法は、図1.4 にお いて c3 の位置づけにあるものと考える.

このような予測の原理は,人間の経験則に基づく予測のアナロジーとなっている.人間は,現 象の仕組みや原理が分からない場合でも,長期的な観測と記憶に基づいて未来の現象について経 験則を立て,それに基づいてある程度の予測を行うことができる.このような,いわば経験則に 基づく予測のことを,気象分野では,観天望気と呼んでいる[67].これは,空の状況を観察するこ とで天気の変化を予測する,古来より行われてきた方法である.例えば,夕焼けは晴 というの は,西空が晴れているから翌日は晴れると予測することである.

本論文で提案する記憶に基づく予測(Memory-Based Forecasting; MBF)の方法は,過去の類 似画像系列の検索と,検索された類似画像系列の集合を用いて予測画像を合成するという2つの 要素技術から構成される.前者の時系列画像の検索の方法としては,時系列画像を低次元の固有空 間上の軌跡として表現し,軌跡の部分区間同士の類似性の観点から,入力された現在の時系列と 類似する過去の時系列を検索するという方式を採用する.このような固有空間を用いた画像の表 現方法は,アピアランスベースと呼ばれるコンピュータビジョンの一技法であり,Turkらによる 固有顔[108]や,村瀬らによるパラメトリック固有空間法[109,110]に端を発し,その後,歩行動 作,読唇,ジェスチャなどの様々な時系列パターンの認識に応用されている[111].このアプロー チは,画像の見かけの変動を低次元の部分空間上で統計的にモデル化することで,対象の構造や ダイナミクスを陽にモデル化する必要がない点が利点である.本研究では,類似の技法を時系列 画像の予測に適用した点に新規性がある.

また,本論文の提案法は,非線形力学系の分野で提案されている非線形予測法[112,113,114, 115,116,117,118]を画像系列へと応用したものとも捉えることができる.この非線形予測法も, 過去の類似時系列に基づく予測の方法であり,これまでに為替や売上高,電力消費量などカオス 性をもつ低次元時系列を対象とした予測法として提案がなされている[114,119,120],本研究で は,これを高次元の画像パターンの予測に適用した点に新規性が認められる.

また,過去の記憶に基づいて推論を行うという方法は,人工知能分野においては,事例ベース推 論 (Case-Based Reasoning; CBR)[121, 122, 123] や記憶に基づく推論 (Memory-Based Reasoning; MBR)[124, 125] として,シンボルに関する推論のアプローチとして知られている.これらのアプ ローチを気象分野に適用した例としては,過去の気温や気圧などの物理観測量のデータベースの 検索に基づく方法が提案されている[126, 127, 128, 129].これらの方法では,比較的スムーズに変 化する大規模 (1000km<sup>2</sup>-10000km<sup>2</sup>) な現象を対象とし,数値化された物理量やシンボル化された 高次の特徴を使用しているのに対して,本研究では,レーダ画像情報を直接用い,より小さなス ケール (10km<sup>2</sup>-100km<sup>2</sup>) で生じるより激しい現象の予測を対象としている.この点において,従 来研究と本研究とでは,対象とする事象の性質が異なり,また,より難易度の高い問題に取り組 んでいるとも位置づけられる.

#### 1.3.4 本論文で提案するアプローチ間の位置づけ

以上のように本論文では,従来の時系列画像解析において未だ解決法が確立されていない観測 の部分性,間接性の問題に着目し,それらの度合いに応じた3つの事例を取り上げ,それぞれ新し いアプローチを提案する.それぞれのアプローチにおいて,未解決課題の中でも特に重要なテー マとして,多視点オクルージョン解析による複数物体の追跡,非言語行動に基づく会話構造の推 定,記憶に基づく気象レーダ画像パターンの予測の3つのテーマを対象とする.各テーマにおい て提案する手法は,対応する問題設定に特化したものであるが,その手法の背後にあるアプロー チは,幅広い事象に適用できるものと考えられる.その観点から以下では,この3つのアプロー チの間の位置づけについて述べる.

本論文で提案するアプローチは,i) 観測過程のモデル化,ii) 相互作用のモデル化,iii) 記憶に基



図 1.5: 本論文で提案するアプローチ間の位置づけ,及び,対象とする事象の性質.

づくモデル化の3種に分類ができる.図1.5にこれらの関係を図示する.観測過程のモデル化の アプローチは、物体間の相互オクルージョンへの対処を主眼として、オクルージョンを含む観測 過程を陽にモデル化し、オクルージョンの構造と物体の状態とを同時に推測するものである.ま た,相互作用のモデル化のアプローチは,複数の人物間の相互作用の推測を主眼として,観測さ れる個々の人物の行動と相互作用との関係についてモデルを立てて,推測を行うものである.さ らに,記憶に基づくモデル化のアプローチは,観測過程や相互作用を陽にモデル化する代わりに, 観測された画像と類似する過去の画像に基づいて事象のモデル化を行うものである.一般に実世 界の事象は、複数の要素から構成され、それらの間の相互作用により事象が進展する、時系列画 像解析では,その様子をカメラなどにより視覚的に観測し,事象の状態の推測を行う.その際,要 素の個数が多くなるに従い,要素間のオクルージョンが発生する可能性もより高まる.また,要 素数の増加に伴い,要素間の相互作用もより複雑なものとなる.こうした要素が比較的少数の場 合,それらの間のオクルージョンや相互作用を陽に記述し,事象の推測を行うアプローチは有効 であると考えられる.本論文では,アプローチi)とii)をそれぞれ単独に用いているが,それらを 同時に用いることで,複数要素のオクルージョンと相互作用の双方を含む事象へ対処ができるも のと期待される.一方,事象を構成する要素数が多くなる場合には,その組み合わせ数の爆発と も関連して,オクルージョンの構造や相互作用の状態を観測情報から推測することが困難となる. このような事象に対して、そこで本論文では、事象の観測過程や相互作用を陽にモデル化するこ とを断念し、その代わりに観測された画像と類似する過去の画像に基づいて事象のモデル化を行 うというアプローチを提案する.このアプローチは,過去のデータが大量に存在するという前提 を要するが,対象に関する事前知識が少ない場合にも適用できるという利点ではある.1.3.3節で 述べた気象レーダ画像パターンは,このアプローチが有効に機能する一例である.このように本 論文で提案する3つのアプローチを,事象の複雑さや要素数に応じて切り替えて適用することで, 広域な事象に対して有効な時系列画像解析法を構成することが可能となる.本論文は,こうした 時系列画像解析法の基本的な構成要素として,これらアプローチを提案するものと位置付けるこ とができる.

## 1.4 本論文の構成

本論文は5つの章から構成される.第1章は序論であり,本研究の目的や背景,及び,各章の位 置付けについて述べた.

第2章では、観測の部分性が顕著な例として、複数の物体の間に生じるオクルージョンに着目 する.オクルージョンとは,カメラに対して,手前の物体が後方の物体を遮蔽する現象であり,時 系列画像解析を妨げる主要因として知られている.オクルージョンが生じる場合,遮蔽された領 |域に関する観測情報が得られないという問題が生じる.また,画像上で複数の物体のオーバーラッ プが生じ,観測情報の解釈が困難になるという問題も生じる.本論文では,多視点観測による複 数物体の追跡(トラッキングとも呼ぶ)を課題として取り上げる.この課題に対して本研究では. 物体間のオクルージョンに対して頑健な追跡を実現するために,物体間のオクルージョンを含め た観測過程を陽にモデル化し,オクルージョンの構造と物体の状態(位置や姿勢)を同時に推定す るという方法を提案する.この方法では,各々の視点からどの物体が見えていて,どの物体が隠 れているかというオクルージョンの空間構造が陽に記述され,物体追跡の問題は,このオクルー ジョンの空間構造についての多重仮説の生成・検証,及び,物体の状態の事後確率分布の推定と いう2つの問題からなる再帰的ベイズ推定法として定式化される.この方法は,オクルージョン に起因する物体の配置や状態の不確定性を推定することができるため , 多様にオクルージョンが 変化する状況においても,安定した物体追跡を行うことができる.人工データ,及び,実データを 用いた実験を行い,その結果,一時的に全ての視点においてオクルージョンが発生している状況 においても,頑健に追跡が継続できることを確認する.

次に,第3章では,観測の間接性が顕著である例として,事象を構成する複数の要素間の相互作 用に着目する.その中でも本論文では,視覚的に直接観測はできないが,相互作用が事象のダイ ナミクスに対して決定的な影響をもつ事例として,複数人物の対面会話シーンを取り上げる.対 面会話シーンにおいては,各人物の行動は時系列画像として比較的容易に観測できるものの,そ れら行動を支配するダイナミクスは,人物間の相互作用(インタラクションとも呼ぶ)に依存し, 非常に複雑である.本論文では,会話中の人物間の相互作用について陽にモデル化を行い,各人 物の行動から人物間の相互作用の状態を推定する新しい方法を提案する.具体的には,相互作用 として,話し掛け,及び,傾聴に着目し,誰が誰に話し掛けているか?誰が誰の話を聞いている か?という人物間でのメッセージの流れのパターンを推測の対象とする.本研究では,これを会 話構造と呼ぶ.また,この推測の手掛かりとして,会話参加者の非言語行動に着目し,会話構造 と非言語行動との関連性を動的ベイジアンネットワークを用いてモデル化する.ここでは非言語 行動として,発話の状態(発話の有無)と,対面会話において重要な役割をもつことが知られて いる視線を用いる.ただし,会話中の視線方向を直接的に計測することは困難であるため,代わ りに頭部方向を計測対象とし,センサや画像上での頭部追跡によって計測された頭部方向から視 線方向の推測を行う.本論文では,頭部方向と発話の有無の時系列データから,会話構造,視線 方向,及び,モデルパラメータを同時に推定する方法として,ギブスサンプリングと呼ばれる一 種のマルコフ連鎖モンテカルロ法を用いる方法を提案する.最後に,4人会話を対象とした実験に より,会話構造と視線方向の推定精度を評価し,提案法の有効性を確認する.

さらに,第4章では,観測の部分性と間接性がともに顕著な事例として,時系列画像パターン として観測される自然現象の予測の問題を取り上げる.本研究では,自然現象の中でも気象現象, その中でも特に降水現象に着目し、その観測手段として気象レーダを用いる、気象レーダは、降 水強度の空間分布を画像パターンとして観測する装置である.これを用いることで,降水の空間 分布の予測を行う問題は、気象レーダ画像上のパターンの時間変化を予測する問題として捉える ことができる.しかし,このパターンは,非剛体の不定形パターンであり,様々な気象的要因が 重なり合った複雑なダイナミクスを持つため,その予測は容易でない.気象レーダによる観測は, パターンの元となる降水現象の空間的な広がりに対して、気象レーダから一定範囲内、かつ、あ る空間断面上に限定された観測であり,降水現象の全容を捉えることができない.その観点から 観測には部分性が伴うといえる.また,気象レーダでは,気温,気圧,風速といった気象現象の内 部状態を直接的に観測することができず,さらに,現時点の観測情報から未来のパターンを予測 するという問題の性質からも,観測には間接性が伴うと考えられる.これらの観点より,気象レー ダ画像上のパターンの予測の問題は,観測の部分性,間接性がともに顕著な事例といえる.本論 文では、このような複雑な時系列画像パターンの予測の方法として、過去の画像パターンの記憶 に基づく予測法を提案する.これは,現在の画像系列に類似する過去の画像系列をデータベース より検索し、検索された画像中のパターンの時間発展に基づいて予測画像を生成するという方法 である.この方法の要素技術として,本論文では,時系列の気象レーダ画像より抽出された画像 特徴に基づいて類似画像系列を検索する方法,及び,検索された画像系列の未来の画像を合成す ることで予測画像を生成する方法を提案する.北海道の約6冬期分のデータセットを用いた実験 を行い,予測先行時間,データ量の観点から予測精度を評価し,提案方法の有効性を確認する.

最後に,第5章において,本論文のまとめと今後の課題・展望を述べる.

# 第2章 多視点オクルージョン解析による複数物体 の追跡

## 2.1 緒言

オクルージョンとは,物体の表面から反射または発せられた光が,他の物体表面によって妨げ られる現象である.ある視点からみて複数の物体間にオクルージョンが生じる場合,観測画像上 において物体の重なりや観測情報の欠落が発生する.この影響により,画像情報に基づいて複数 物体の追跡(トラッキングとも呼ぶ)を行う場合,追跡中の物体を見失ったり,また,異なる物体 と入れ替わったりという深刻な問題が生じることが知られている.そのため,オクルージョンへ の対処は,実世界の動的な事象の推測において極めて重要な課題であると考えられている.本章 では,時系列画像解析における観測の部分性の要因として,このオクルージョンに着目し,オク ルージョンに対してロバストに複数の対象物体を追跡する方法を提案する.具体的には,多視点 観測に基づいた物体間の相互オクルージョンの解析と,物体の位置・姿勢等の状態の逐次推定を 行う方法を提案する.ここでは相互オクルージョンとは,ある視点から見て前方の物体が後方の 物体の一部,または,全部を隠蔽する現象と定義する.なお,本論文においては,関節物体に生 じるセルフ・オクルージョンは対象外とする.

従来,単一視点観測に基づく物体追跡法におけるオクルージョン対策として,オクルージョン が発生している時間の前後において,追跡対象の対応付けを行うアプローチが一般的に用いられ ている [25, 26].このアプローチにおいて,対象間の対応付けは,色や形状の一貫性,運動の連続 性などの対象の性質に基づき行われる.また,オクルージョンの発生・解消は,blobの合併・分 裂 [25] や,特徴点の消滅・再出現 [26] などの画像上のイベントとして解釈・検出される.しかし, このアプローチは,オクルージョンが発生している区間において,観測情報に基づいて物体の位 置を推定する仕組みを持たないため,個々の物体の位置関係を正確に推定することは困難とされ る.また,オクルージョンが生じている状況においても,個々の物体の位置を特定し,追跡を継 続することを狙った方法も提案されている [23, 24, 130].これらの方法では,物体の輪郭 [23, 24] や見かけ [130] に関するモデルを用いることで,カメラに対する複数の物体の奥行き順序(depth order)を推定し,オクルージョンによって生じる画像上での物体の重なりを解釈する機構が導入 されている.しかし,単一視点の観測の性質上,奥行きを推定するための手掛かりに乏しく,適 用可能な範囲が2物体程度の部分的なオクルージョンと限定されたものに留まっている.

19

一方,単一視点観測の限界を克服するためのアプローチとして,複数の異なる視点において観 測を行うアプローチに近年,注目が集まっている.これは,各々の視点で得られる画像から,そ れぞれ物体領域の重心 [27,28] やシルエット像 [29,30] を観測情報として抽出し,複数の視点の観 測情報を統合し,物体追跡を行うものである.これにより,一部の視点においてオクルージョン が存在する場合においても,他の視点の情報により補うことができ,各物体の空間中における位 置を推定することができる.また,さらにオクルージョンの悪影響を軽減するために,オクルー ジョンのない視点の選択的利用 [28,33] や,オクルージョンの度合いに応じた観測情報の重み付け 統合 [31,32] などの方式が提案されている.

こらら複数の視点を用いた手法では,単一視点の手法と比較し,飛躍的なオクルージョン耐性 の向上が実現されている.しかしながら,これまでに提案されている手法は,オクルージョンの 影響の少ない視点が十分に存在することを暗黙のうちに仮定しており,ある物体が全ての視点か らみて隠蔽されているような重度のオクルージョンには対処できない.また,これら手法が対処 できるオクルージョンの種類や程度についても十分な議論がなされていないのが現状である.今 後,様々なオクルージョンに対処できるロバストな追跡手法の実現と,その適用可能範囲の理論 的な検証のためには,物体間にオクルージョンが生じるプロセスを適切に理解し,それを観測情 報から推測する手法を構築する必要があると考えられる.

本研究では、様々なオクルージョンの状況下において頑健に物体追跡を行うために、多視点観 測において,どのカメラからどの物体が見えているか?どの物体によってどの物体が遮蔽されて いるか?といったオクルージョンの空間的な構造を陽にモデル化するというアプローチに着目し ている.本論文では,その初歩的な試みとして,2次元平面上の既知の幾何形状をもつ物体と,同 平面上の固定視点からなる単純化された世界を対象とし,オクルージョンの空間的な構造を推定 することで,オクルージョンに頑健に複数物体の追跡を行う方法を提案する.本論文では,その ために観測過程を2次元平面上の透視投影の幾何的な表現法である「視角」によりモデル化する. ここで視角とは,視点から物体と背景の境界に伸ばした半直線で囲まれた領域であると定義する. 次に,物体の空間的な配置や観測情報と物体との対応関係を表現するために,観測された視角とそ の交差の性質に基づいてオクルージョンの空間的な構造を記述する方法を提案する.また,本研 究ではこのオクルージョンの空間的な構造の記述法に基づき , 複数物体のトラッキングの問題を , オクルージョンの空間的な構造の多重仮説の生成・検証と、各仮説のもとにおける物体状態の事後 確率分布の推定という2つの問題に分解して考え,これを2重構造をもつ再帰的ベイズ推定の問 題として定式化する.この推定の枠組みは MHT(Multiple Hypothesis Tracking)[22] と呼ばれる data association の解法の拡張であると位置付けることができる. Data association[19] とは, 多体 軌跡追跡における追跡対象と観測値の対応付け問題のことを指す.近年,MHT や JPDAF(Joint Probabilistic Data Association Filter) [19] などの data association の解法は, コンピュータ・ビ ジョン分野における複数物体の追跡へも適用され始めている[131,132].これら従来の方法では, すべて対象を質点と捉えて,追跡を行っているのに対して,本研究では大きさをもつ物体を対象 として捉え,物体間に生じるオクルージョンの空間的な構造の動的な推定を行っている点に新規 性がある.本論文の提案法は,各時刻においてオクルージョンに起因する物体の状態や配置の不 確定性を推定することができるため,重度のオクルージョンが生じる状況においても,安定した 物体追跡が実現される.

また,提案法では,観測過程の記述法として,2次元平面上の視角,及び,視角の交差を用い ているが,これはいわゆる視体積交差[133]の2次元版と位置付けることができる.視体積交差と は,各視点のシルエット像を元の空間に逆投影してできる交差の領域のことを指す.視体積交差 は,元来,物体の3次元形状を推定するために導入されたものであるが,近年,複数物体の追跡 においても用いられている[86,30].しかし,これらの方法では,視体積交差の領域が,直接,対 象物体の形状の近似となると考えて追跡を行っており,オクルージョンによって生じる視体積交 差と物体形状の相違や,物体の存在しない交差領域を単なるノイズとみなしている.このように これらの方法では,オクルージョンの生じるプロセスを陽に考慮しておらず,オクルージョン問 題の本質的解決は期待できない.

本章は以下のように構成されている . 2.2 節において, 観測過程とオクルージョンの空間的な構造の記述法を定義し, 2.3 節において, オクルージョンの空間的な構造と物体状態の推定の枠組みを提案する . 2.4 節において,提案した推定法の枠組みの有効性を実験により検証する . 2.5 節において,研究の課題と方向性について考察を行い, 2.6 節でまとめを行う.

### 2.2 観測過程とオクルージョンの構造

はじめに本研究が対象とする世界とその観測過程を定義する.次にオクルージョンの空間的な 構造の記述法を述べ,さらにオクルージョンの空間的な構造のダイナミクス・モデルを導入する. なお,表2.1に本節で導入するオクルージョンに関する用語の一覧を記載する.

#### 2.2.1 観測過程の定義

本論文では 2 次元平面上を運動する 2 次元物体,及び,物体と同一平面上の複数の固定視点からなる世界を考える.これら物体の追跡の問題を,複数の視点からの観測情報に基づいて,物体の位置や姿勢等の状態を逐次的に推定する問題として捉える.なお,ここでは物体数 N は未知とし,その推定も問題に含まれる.物体は 2 次元平面  $R^2$  上の滑らかな輪郭をもつ凸閉領域  $O \subset R^2$ であり,剛体性を仮定する.本論文では図 2.1(a)のように,床面から一定の高さにある水平断面における人体断面の近似形状として想定できる楕円を物体形状モデルとする.物体  $O_i$ の状態変数は,図 2.1(b)のように楕円の中心座標  $(x_i, y_i)$ と x座標軸に対する回転角  $\phi_i$ ,及び,それらの時

表 2.1: オクルージョンに関する用語.

視角	視点から物体の両端に引いた2本の半直線で囲まれる領域.
視角辺	視角の境界線.
セル	視角の交差領域.
視角分離不可能性	ある一視点からみて , オクルージョンの関係にある複数の
	物体が一つの視角として観測されるという性質.
排他的接触性	各視角辺はただ一つの物体と接触するという性質.
物体包含性	各物体はいずれか一つのセルに含まれるという性質.
空間的排他性	物体は互いに交差しないという性質 .
オクルージョン構造	各物体の属するセル,及び,各物体と視角辺との接触関係
	の記述.
オクルージョン仮説	オクルージョン構造についての仮説.

間変化分  $\dot{x}_i, \dot{y}_i, \dot{\phi}_i$  からなるベクトル  $x_i = [x_i, \dot{x}_i, y_i, \dot{y}_i, \phi_i, \dot{\phi}_i]^T$  として定義する.ここでは回転角  $\phi_i$ を物体の姿勢と呼ぶ.各時刻 k において,各々の物体の状態  $x_{i,k}$ を推定することが目標である. 物体の状態のことを単に物体状態とも呼ぶ.本論文では説明の簡便性のため,楕円の形状とサイ ズは既知とするが,これらを物体状態のベクトルに含めて推定することも可能である(2.4.5節参 照).物体は空間的に互いに交差をしない限り独立に運動するものと仮定する.

視点  $v \in V$ は,物体と同一の平面上に存在し,物体が運動する領域を取り囲むように配置される.各視点は既知の位置に固定されている.視点数は  $M \ge 2$ とする.物体は全ての視点から観測可能な領域内に存在すると考え,この領域からの退出,及び,新しい物体の進入は考えない.各視点において,物体は背景から分離されたシルエット像として得られることを想定し,図 2.2(a)のような視角 (visual angle)の集合が各視点において観測されるものとする.本研究において,視角とは,ある視点から物体を見たときの,視点から物体の両端へ引いた2本の半直線で囲まれた領域であると定義する.視角は,2次元平面上における透視投影の幾何的な表現法の一つである.時刻 kにおいて観測される視点 v の w 番目の視角を  $\Omega_{v,w,k}$  と記す.視角の境界線を視角辺と呼び,  $l_{v,w,k}^+$ ,  $l_{v,w,k}^-$  と記す.視角辺は,各視点から物体と背景の境界を見込んだときの視線に相当する.各時刻 kにおける観測情報は,視角辺の集合として,

$$Z_k = \{l_{1,1,k}^-, l_{1,1,k}^+, l_{1,2,k}^-, \cdots, l_{1,W_1,k}^+, l_{2,1,k}^-, \cdots, l_{M,W_M,k}^+\}$$
(2.1)

のように得られるものとする.ここで  $W_v$  は視点 v で観測される視角の個数を指す.時刻 k までの観測情報の集合を  $Z^k = \{Z_1, Z_2, \cdots, Z_k\}$  と記す.本論文では,この観測過程において視角の誤



図 2.1: 物体形状のモデル.

検出や未検出はないものと仮定する<sup>1</sup>.また,各視角辺は,物体の状態が与えられた条件のもとでは,空間的,時間的に互いに独立に観測されると仮定する.

本論文で想定する世界は,屋内の平坦な床面上を歩行する複数の人物を対象とし,壁面に設置 された複数台の固定カメラにより側面から人物を撮影し,追跡を行うという状況を単純化したも のである.この場合,本論文の2次元平面の世界は,図2.1(a)のように,3次元空間中において 床面と平行であり,人物の胸部辺りと交わるような平面であると想定される.この平面を基準平 面と呼ぶ.本論文では,基準平面と人物の交わりの形状を楕円として近似できると想定している. 楕円の回転角が人物の胴体の方向に対応している(ただし,前後の区別はない).また,物体のシ ルエット像は,撮影された画像に対して一般的な画像処理手法である背景差分と2値化の処理を 施すことで得られると考える.観測値である視角は,このシルエット像から,校正済みカメラの カメラモデルにより定まる世界座標と画像座標との関係に基づき計算できると想定している.な お,2.5節において,提案法を現実世界へ適用する際の各種仮定・条件の妥当性や問題点,及び, 問題の解決に向けた提案法の拡張の方向性について論じる.

#### 2.2.2 視角表現に関する幾何的制約

観測された視角から,物体の状態を推定するために,物体と視角の間に存在するいくつかの幾何的制約を導入する.

まず,図2.2(a)のように,ある視点に対してある物体が後方の物体の一部を隠蔽する場合など, 複数の物体が互いに隠蔽・被隠蔽の関係にある場合,これらの物体の視角は,一つの視角として 統合され観測される.この性質を「視角分離不可能性」と呼ぶことにする.また,一つの視角内 に存在する物体の集合のことをオクルージョンの関係にある物体と呼ぶ.この性質は2次元平面

<sup>&</sup>lt;sup>1</sup>本論文では,議論の単純化のためこの仮定をおいたが,提案する方法の枠組み自体は,誤検出・未検出を含む状況 へ拡張可能である(2.5節を参照).



図 2.2: 観測過程.(a)物体と視角,(b)セル内における取り得る物体配置の2つの例.



図 2.3: 完全オクルージョンが生じる物体の位置関係の例.

上の物体のシルエット像の観測という前提に由来する.

また,図 2.2(a) のように視角の辺は物体の接線となっている.本研究では,各視角辺はただ-つの物体と接触するという仮定をおく.これを「排他的接触性」と呼ぶ.なお,物体に接触しな い視角辺は存在しないとする.排他的接触性の制約は,視角の観測において誤検出・未検出がな いという仮定に由来する.なお,現実には2つ以上の物体が一つの視角辺に接するような状況も 考えられるが,本論文ではそのような状況を,その視角辺にそれぞれ一つの物体が接するという 複数の仮説により表現する(2.2.5節参照).

さらに,全ての視点からみて,物体はいずれかの視角内に存在するため,物体は各視点の視角 の交差がつくる領域内に存在することがわかる.本論文では,この視角の交差の領域を「セル」と 呼ぶ.各物体は必ずいずれか一つのセルに含まれるが,一つのセルに複数の物体が含まれること もあり,また,物体を一つも含まないセルも存在し得る.これらの性質のことを「物体包含性」と 呼ぶことにする.時刻 k におけるセルの集合 *C*k は,

$$\boldsymbol{C}_{k} = \{\Omega_{1}^{\prime} \cap \Omega_{2}^{\prime} \cap \dots \cap \Omega_{M}^{\prime} | \Omega_{v}^{\prime} \in \boldsymbol{\Omega}_{v,k}\} \setminus \{\phi\}$$

$$(2.2)$$

のように得られる.ただし,  $\Omega_{v,k}$  は時刻 k において視点 v から観測される視角の集合を指す.また,式 (2.2) 中の\は,差集合を表す記号であり,例えば, $X \setminus Y$  は,集合 X から集合 Y の要素を除いた差集合を表す.各セル C は多角形領域であり,その辺は視角辺の部分集合から構成される.これを E(C) と記す.M 視点からの観測においては,セルは最多で  $2 \times M$  本の辺から構成される.なお,物体包含性の条件は,視角の観測において未検出がないことに由来する.

また,複数の物体が一つのセルに包含される場合,セル内における物体同士の交差はないもの とする.これを「空間的排他性」と呼ぶ.この条件は物体の剛体性の仮定に由来している.

#### 2.2.3 部分オクルージョンと完全オクルージョン

2.2.1 節で導入した観測過程,及び,2.2.2 節の視角に関する性質や制約を踏まえ,物体間のオク ルージョンと観測される視角との関係について述べる.ある一つの視点からみて,2つの物体が オクルージョンの関係にある場合,図2.2(b)のような各物体がそれぞれ一方の視角辺に接する2 つの場合と,図2.3のような両方の視角辺に同一の物体が接する2つの場合とが考えられる.図 2.2(b)の場合,視点 $v_1$ から見て手前の物体 $O_1$ が後方の物体 $O_2$ の一部分を隠している.このよう な状況を「部分オクルージョン」と呼ぶことにする.また,図2.3(a)の場合,前方の物体 $O_1$ が両 方の視角辺に接しており,視点 $v_1$ から見て後方の物体 $O_2$ が前方の物体 $O_1$ によって完全に遮蔽さ れている.このような状況を「完全オクルージョン」と呼ぶ.一方,図2.3(b)の場合は,後方の 物体 $O_2$ が両方の視角辺に接しており,本来,視点 $v_1$ から前方の物体 $O_1$ の表面は見える位置関係 にあるが,シルエット観測の前提のもとでは,この物体 $O_1$ の位置を特定できる観測情報がこの視 点 $v_1$ の視角に含まれない.この観点から本論文では図2.3(b)の状況も「完全オクルージョン」と みなすこととする.このようにシルエット観測においては,物体の前後関係(遮蔽・非遮蔽の関係)を決めるための直接的な手掛かりが観測された単一の視角には含まれないという性質がある.

#### 2.2.4 視角表現と物体状態の不確定性

2.2.2 節の幾何的な制約は, 観測された視角の集合に対し, 物体の存在し得る領域や, 物体のと り得る状態を規定する制約となっている.しかし, これらは比較的緩い制約であるため, 制約条 件を満たす物体の配置や状態が多数存在し, 一意に定めることができない状況が存在する.

例えば,ある視点から見て,二つの物体が部分オクルージョンの関係にある場合,オクルージョ ンがない場合と比較して,それぞれの物体に接触している視角辺の本数が減少し,セルの面積が 増加する.これにより,セル内の物体の配置に不確定性が生じ,与えられた視角のみからは一意 に決定できない状況が生じる.具体的には図2.2(b)の例の場合,一つの視角辺が2つのセルの辺 として共有されるため,物体と視角辺との接触関係に不確定性が生じ,物体の形状・サイズが与 えられた場合,図2.2(b)の実線の物体配置と破線の物体配置の2通りの可能性が存在する<sup>2</sup>.

また,仮に物体と視角辺の接触関係が与えられた場合にも,各物体の位置・姿勢には自由度が 残る可能性がある.例えば,図 2.4(a)のように,物体  $O_1$ が視角辺  $l_{1,1,k}^+$ , $l_{2,1,k}^-$  に接し,物体  $O_2$  が 視角辺  $l_{1,1,k}^-$ , $l_{2,1,k}^+$  に接すると仮定した場合,観測された視角から解釈し得る物体の配置には,図 中の破線,一点鎖線のような配置を許容する自由度が存在し,与えられた物体と視角辺の接触関 係のみからは一意に決定できない.また,ある視点からみてオクルージョンの関係にある物体の 状態には「視角分離不可能性」を満足するような相互依存性が存在する.例えば,図 2.4 の場合に は,視点  $v_1$  からみて,図 2.4(b)の破線部のように,物体間に隙間が生じるような物体の配置は, 観測値と矛盾するためにあり得ないといえる.

また,図 2.5 のような場合,観測された視角の集合を満足すまた,物体数も一意に決定できない.このような場合,複数の物体状態の間には大局的な相互依存性が存在し,各物体の状態や存在はセル毎に単独に決定することができない.なお,ここでは物体を含まないセルのことをファントム・セルと呼ぶ.

#### 2.2.5 オクルージョン構造の記述

前節 2.2.4 で述べたように,物体の状態や配置には不確定性や相互依存性が存在し,観測された 視角の集合から,直接,各物体の状態や配置を一意に推定することは困難な問題である.そこで, まず本研究では,物体の配置や状態を一意に決定することを直接の目標とするのではなく,それら の不確定性の推定を目標とする.その上で物体の配置・状態の不確定性を「排他的接触性」と「物

 $<sup>^2</sup>$ 図 2.2(b) の場合,物体サイズの関係上,図 2.3 で示した配置はとり得ないことに注意されたい.


図 2.4: 物体配置の不確定性と拘束.(a) 物体と視角辺の接触関係が与えられた後に残る物体配置の不確定性,(b) 観測情報と矛盾する物体配置の例.



図 2.5: 観測から解釈し得る物体-セル配置の2つの例.

体包含性」に関連する部分と「視角分離不可能性」と「空間的排他性」に関連する部分に分解し て考える.そして,まず前者に関する物体の大局的な配置の不確定性を推定した後に,後者に関 する局所的な物体の状態を不確定性を含め推定するという2段階の推定のアプローチをとる。こ こでは前者2つの制約を大局的制約と呼び,後者2つを局所的制約と呼ぶことにする.

まず,物体の空間的な配置に関する大局的制約である「排他的接触性」と「物体包含性」に基 づいた物体の配置の記述法を提案する.この記述法は,各物体  $O_i$  に接触する視角辺の集合  $L^{(i)}$ , 及び,各物体が包含されるセル $C^{(i)}$ のペアから構成される.これにより複数の物体の空間的な配 置と,物体と視角辺との関連が表現できる.本論文では,この記述のことを「オクルージョン構 造」と呼ぶ.また,観測された視角に対して大局的制約を満足するオクルージョン構造を,複数 の仮説として数え上げることで、物体配置の大局的な不確定性が表現できる.ここではこの仮説 のことをオクルージョン仮説,または、単に仮説と呼ぶ.あるシーンのオクルージョン構造の一 仮説 H は , 物体毎の部分仮説  $H^{(i)}$  の集合からなり , 部分仮説  $H^{(i)}$  は物体  $O_i$  に接触する視角辺の 集合  $L^{(i)}$  と属するセル $C^{(i)}$ の対として,

$$H = \{H^{(1)}, H^{(2)}, \cdots, H^{(N)}\},$$
(2.3)

$$H = \{H^{(1)}, H^{(2)}, \cdots, H^{(N)}\},$$

$$H^{(i)} = (L^{(i)}, C^{(i)})$$
(2.3)
(2.4)

のように表される.なお,Nは仮説Hにおいて,その存在が支持される物体の個数である.た だし,各物体に接触する視角辺の集合(以下では接触辺集合と呼ぶ)L<sup>(i)</sup>は,「排他的接触性」  $(L^{(i)} \cap L^{(j)} = \emptyset, i \neq j$ かつ $\sum_i L^{(i)} = Z_k$ )を満たす.本論文では,観測される視角集合からオク ルージョン構造を推定することをオクルージョン解析と定義する.この記述法を用いると,例え ば,図2.2(b)の場合,オクルージョン構造の仮説の集合は,

$$\begin{split} \boldsymbol{H} &= \{H_1, H_2\}, \\ H_1 &= \{H_1^{(1)}, H_1^{(2)}\} \\ &= \{(\{l_{1,1,k}^+, l_{2,1,k}^-, l_{2,1,k}^+\}, C_{1,k}), (\{l_{1,1,k}^-, l_{2,2,k}^-, l_{2,2,k}^+\}, C_{2,k})\}, \\ H_2 &= \{H_2^{(1)}, H_2^{(2)}\} \\ &= \{(\{l_{1,1,k}^-, l_{2,1,k}^-, l_{2,1,k}^+\}, C_{1,k}), (\{l_{1,1,k}^+, l_{2,2,k}^-, l_{2,2,k}^+\}, C_{2,k})\} \end{split}$$

のように表現することができる.ただし, $H_1, H_2$ は図2.2(b)の実線と破線の物体の配置にそれぞ れ対応する仮説であり, $H_1^{(1)}, H_1^{(2)}$ は,仮説 $H_1$ における各物体 $O_1, O_2$ の部分仮説を表す.また,  $H_2^{(1)}, H_2^{(2)}$ も同様である.このようにオクルージョン構造の多重仮説によって,大局的な物体配置 の不確定性に対応した複数の異なる物体-セル配置を表現することができる.

次にあるオクルージョン構造の仮説が与えられた後における,局所的な物体状態の不確定性を 表現する方法を述べる.ここではオクルージョン構造によって物体毎に割り当てられた視角辺が,



図 2.6: オクルージョン構造の変化.(a)オクルージョンの発生・解消,(b)部分オクルージョン-完全オクルージョン間の遷移(接触性の変化).

物体の接線に対応するという性質を用いて,与えられた接線に対して残留する物体状態の自由度 を確率密度分布として表現する.これを物体の状態分布と呼ぶ.また,その際に局所的制約(「視 角分離不可能性」と「空間的排他性」)を反映するために,全物体の結合状態分布として物体状態 を表現する.この分布は,物体毎に割り当てられた接触辺の本数や配置,物体の形状モデルに応 じて多様な形状をもつ.

また,オクルージョン構造によって与えられる物体の接触辺集合は,その物体の可視性を表現 していると考えることができる.つまり,ある物体 $O_i$ に視角辺 $l \in L^{(i)}$ が接している場合,この 視角辺lの通る視点からこの物体の輪郭が見えていることを意味する.従って,接触辺の本数は一 種のオクルージョンの度合いを表す尺度とみなせる.例えば,ある物体の接触辺数が $2 \times M$ の場 合,この物体は全視点から見えていることを意味し,接触辺数が1の場合,ただ一つのみの視点 から物体の片方の輪郭が見えることを意味する.また,ある物体の接触辺数が0の場合,全ての 視点からみて物体は完全に隠蔽されていることを意味する<sup>3</sup>.

# 2.2.6 オクルージョン構造のダイナミクス

対象物体の状態や配置が変化する場合,それに対応してオクルージョン構造にも変化が生じる. したがって,物体追跡を実行するためには,オクルージョン構造の動的な推定が必要となる.こ こでは,そのためにオクルージョン構造の時間変化の性質を表現するダイナミクス・モデルを導 入する.本研究では,オクルージョン構造の時間変化を,オクルージョンの発生・解消に伴うも

<sup>&</sup>lt;sup>3</sup>2.2.3 節でも議論したように,ある物体の接触辺数が0の場合でも,実際には視点から物体表面が見えている場合 もあり得る.しかし,シルエット観測の前提の元では,この場合,この物体の状態を直接的に決定する情報が得られな いという観点から,完全に隠蔽されているとみなす.

のと,部分オクルージョンと完全オクルージョンの間の遷移に伴うものの2種類に分解し,これ らをそれぞれ決定論的な遷移モデルと確率論的な遷移モデルとして表現する.

オクルージョンの発生とは,図2.6(a)の左 右のようにある視点から見て,物体の一部が他の 物体の影に入る事象を指し,また,オクルージョンの解消とは,図2.6(a)の右 左のように,逆に ある物体が他の物体の影から完全に脱却する事象を指す.図2.6(a)からわかるように,オクルー ジョンの発生は複数の視角の合併に対応し、また、その解消は複数の視角への分裂に対応してい る.したがって,これらの事象が生じる場合,各物体の属するセルは,視角の合併・分裂によっ て生じる新しいセルへと変化することがわかる.オクルージョンが発生する場合,複数の視角が 一つの視角へと合併するため,図2.6(a)のように,合併する複数の視角の最も外側にある視角辺  $l^+, l^-$  が,合併後の視角の両辺 $l^{+'}, l^{-'}$ に対応することがわかる.そのため,これらオクルージョ ン発生後の視角辺 l+', l-' に接する物体は, 合併前の視角辺 l+, l- に接する物体と同一であると仮 定できる.一方,オクルージョンが解消する場合,図2.6(a)の右 左のように,分裂前の視角の 両辺 $l^{+\prime}, l^{-\prime}$ が,分裂後の視角の集合の最も外側にある視角辺 $l^+, l^-$ に対応することがわかる.そ のため,オクルージョン解消後の視角の視角辺  $l^+, l^-$  に接する物体は,分裂前の視角辺  $l^{+\prime}, l^{-\prime}$ に 接する物体と同一であると仮定できる.このような変化をオクルージョンの発生・解消に伴うオ クルージョン構造の変化と呼ぶ.なお,本研究では単位時間内における物体の移動量は,物体の サイズと比較して十分小さく,隣接時刻間における視角の対応付けは不確定性なく決定できると 仮定する.

一方,部分オクルージョンと完全オクルージョンの間の遷移は,図 2.6(b)のように,視角辺に 接触する物体の入れ替わりに対応する.このことを接触性の変化と呼ぶ.この変化は前述の視角 の合併・分裂のように明確に観測できないため,これを確率的に発生する現象であるとみなすこ とにする.ここではその確率モデルとして,図 2.6(b)の視角辺 l,l'に対する物体  $O_1, O_2$ のように, ある視角辺に接触し得る物体が複数ある場合,単位時刻間で同一の物体が接触し続ける確率は,あ る確率  $\eta$  で与えられるというモデルを仮定する.この  $\eta$  を停留確率と呼ぶ.停留確率  $\eta$  は,オク ルージョン構造の変化の起こりにくさ(起こりやすいさ)を表すダイナミクスモデルのパラメー タである.この遷移モデルは,例えば,図 2.6(b)の場合,単位時刻間において,右の配置のまま 留まる確率が  $\eta$  で与えられ,右から左へ変化する確率が  $1 - \eta$  で与えられることを仮定するもので ある.また本研究では,視点は十分に疎に配置されていると仮定し,各辺の接触性変化は独立事 象であると仮定する.



図 2.7: 推定の枠組みのブロック図.

# 2.3 オクルージョン構造と物体状態の推定

# 2.3.1 概要

本論文では,複数物体の状態の逐次的な推定の問題を,オクルージョン構造の推定と,推定さ れたオクルージョン構造のもとでの物体の状態推定の問題に分解して捉える.そして,これらを オクルージョン仮説の生成・検証と,物体状態の事後分布の推定という2つのプロセスからなる 再帰的ベイズ推定として定式化する.図2.7に提案する推定法のブロック図を示す.これは相互作 用を伴う2重の再帰的ベイズ推定からなり,外側のループは物体状態の推定に対応し,内側のルー プはオクルージョン構造の推定に対応する.この枠組みでは,オクルージョン構造の推定と物体 状態の推定が交互に実行される.なお,一般の再帰的ベイズ推定法は,この外側のループに対応 するステップのみから構成される[14,13].

提案する推定法の流れを以下で概説する.まず,初期時刻k = 1の観測情報 $Z_1$ よりオクルージョン構造の初期仮説と物体状態の初期分布が計算される.初期時刻以降k > 1においては,直前の時刻k - 1で得られたオクルージョン構造の仮説集合 $H_{k-1}$ と,その時刻の物体の状態分布から予測された予測状態分布に基づいて,時刻間におけるオクルージョン構造の変化を反映した事後仮説集合 $H_k$ への更新が行われる.次に各仮説毎に物体状態の事後確率分布が推定され,第一位の仮説のもとでの物体状態の代表値が推定結果として出力される.また,物体状態の事後確率分布と事後仮説集合が次時刻に伝播される.なお,これらの各過程において前述の幾何的制約条件に合致しない仮説は棄却され,仮説の合計が0個となった時点で処理は停止される.ここで事前仮説とは,各時刻の観測情報による更新前の仮説のことであり,事後仮説は,その時刻の観測情報に基づいて更新された後の仮説のことである.また,物体状態の事前分布,事後分布も同

様の意味をもつ.

この推定法により,オクルージョンに起因する物体の状態や配置の不確定性を陽に推定するこ とができ,さらに複数の時刻にわたる観測情報の累積の効果により,この不確定性を減少させる ことができる.そのため,一時的に全ての視点から見て物体が隠れるような重度のオクルージョ ンが生じる状況においても,安定に物体の追跡を継続することができる.

# 2.3.2 オクルージョン仮説と物体状態の尤度の定義

説明の準備として,オクルージョン仮説と,物体状態の尤度をそれぞれ定義する.ある時刻kにおける観測値 $Z_k$ に対する仮説Hの尤度 $P(Z_k|H)$ を,物体間の状態の独立性,及び,各視角辺の観測の独立性の仮定より,物体毎の部分仮説 $H^{(i)} = (L^{(i)}, C^{(i)}) \in H$ の尤度 $P(Z_k|H^{(i)})$ の積として,

$$P(Z_k|H) = \prod_{i=1}^{N} P(Z_k|H^{(i)}), \qquad (2.5)$$

$$P(Z_k|H^{(i)}) = \int p(Z_k|\boldsymbol{x}_{i,k}, H^{(i)}) \cdot p(\boldsymbol{x}_{i,k}|H^{(i)}, Z^{k-1}) d\boldsymbol{x}_{i,k}$$
(2.6)

のように定義する.式(2.6)では,物体状態の事前分布 $p(\boldsymbol{x}_{i,k}|H^{(i)},Z^{k-1})$ に対して,観測データ $Z_k$ が合致する度合いを評価している.

式 (2.6) 中の  $p(Z_k | \boldsymbol{x}_{i,k}, H^{(i)})$  は,仮説  $H^{(i)}$  のもとにおける物体  $O_i$  の状態  $\boldsymbol{x}_{i,k}$  の観測値に対す る尤度である.これをこの物体が属するセルの各辺への接触・非接触の尤度の積として,

$$p(Z_k | \boldsymbol{x}_{i,k}, H^{(i)}) = \prod_{l \in L^{(i)}} t(l, \boldsymbol{x}_{i,k}) \cdot \prod_{l \notin L^{(i)}, l \in E(C^{(i)})} \bar{t}(l, \boldsymbol{x}_{i,k})$$
(2.7)

のように定義する.ただし,t(l,x), $\bar{t}(l,x)$ は,状態xを持つ物体が視角辺lに接触する尤度,及び,接触せず視角内に存在する尤度をそれぞれ表す関数である.これらの関数は,例えば,視角の左の辺l<sup>+</sup>に関しては,観測された視角辺l<sup>+</sup>の角度 ang l<sup>+</sup>と推定された物体状態xから得られる視角 $\hat{l}^+(x)$ の角度 ang  $\hat{l}^+(x)$ の差  $\Delta = ang l^+ - ang \hat{l}^+(x)$ の関数として,

$$t(l^{+}, \boldsymbol{x}) = \begin{cases} (1 - (\Delta/h)^{2})^{2} & \text{if } |\Delta/h| \leq 1, \\ 0 & \text{otherwise,} \end{cases}$$

$$\bar{t}(l^{+}, \boldsymbol{x}) = \begin{cases} 1 & \text{if } \Delta/h > 1, \\ 1 - t(l^{+}, \boldsymbol{x}) & \text{if } 0 \leq \Delta/h \leq 1, \\ 0 & \text{if } \Delta/h < 0 \end{cases}$$
(2.8)
$$(2.8)$$

の様に定義したものを用いることができる.図 2.8 にこれらの関数の様子を示す.接触の尤度 t は 角度差  $\Delta$  が 0 になる場合に最大値をもつ単峰性の分布をもつ.また,非接触の尤度  $\overline{t}$  は,角度差  $\Delta$  が 0 になる場合に 0 となり,物体がセルの内側の方向に位置するにつれ大きくなるような分布



図 2.8: 観測された視角辺に対する物体状態の尤度関数 (実線:接触の尤度,破線:非接触の尤度).

を持つ.また,hはこれらの分布の幅を定めるパラメータである.これらの分布は視角辺の観測の際に想定される観測ノイズをモデル化したものと考えることができる.また,視角の右側の辺 l<sup>-</sup> に関しても同様の尤度が定義できる.

2.3.3 オクルージョン仮説の生成

各時刻 k において,前時刻 k-1 との間で生じたオクルージョン構造の変化を反映させるため, 事前仮説集合  $H_{k-1}$  から事後仮説の候補集合  $\Theta_k$  が生成される.この過程の概略を図 2.9 に示す. まず,オクルージョンの発生・解消に関する構造変化を反映して,個々の事前仮説から中間仮説  $H'_{n,k}$  が生成される.続いて,個々の中間仮説から視角辺の接触性変化を反映した事後仮説の部分 候補集合  $H''_{n,k}$  が生成される.最後に異なる部分候補集合に含まれる同一の仮説を併合し,事後仮 説の候補集合  $\Theta_k = \bigcup_n H''_{n,k}$  を得る.

オクルージョンの発生・解消

各々の事前仮説  $H_{n,k-1} \in H_{k-1}$ に対して,2.2.6 節で述べたオクルージョンの発生・解消についての遷移モデルを適用し,中間仮説  $H'_{n,k} = \{(L_{n,k}^{'(i)}, C_{n,k}^{'(i)})\}_i^{N'_{n,k}}$ を得る.物体数  $N'_{n,k}$ は事前仮説から継承される  $(N'_{n,k} = N_{n,k-1})$ .物体  $O_i$  が属するセル  $C_{n,k}^{'(i)}$ を,物体が各セル  $C \in C_k$ 内に含まれる確率  $P(O_i \subset C|H_{n,k-1})$ に基づいて,

$$C_{n,k}^{'(i)} = \arg \max_{C \in \mathbf{C}_k} P(O_i \subset C | H_{n,k-1}),$$
(2.10)

$$P(O_i \subset C | H_{n,k-1}) = \int_{O(\boldsymbol{x}_{i,k}) \subset C} p(\boldsymbol{x}_{i,k} | H_{n,k-1}, Z^{k-1}) d\boldsymbol{x}_{i,k}$$
(2.11)

のように決定する.ただし, $p(x_{i,k}|H_{n,k-1}, Z^{k-1})$ は,2.3.6節において述べる物体 $O_i$ の状態の予測分布を表す.事前仮説の中で $P(O_i \subset C|H) = 0, \forall C \in C_k$ となるような物体 $O_i$ を含む仮説Hは「物体包含性」の条件を満たさないため棄却される.



図 2.9: 仮説の生成・検証の過程における仮説集合の遷移.

また, 2.2.6 節で述べた遷移モデルより, 視角辺の接触性は視角の合併・分裂により消滅や出現 する辺を除き保存されるため, 各物体の接触辺集合  $L_{nk-1}^{'(i)}$ は,

$$L_{n,k}^{'(i)} = \{f_{k-1}^{k}(l) | l \in L_{n,k-1}^{(i)}, f_{k-1}^{k}(l) \in E(C_{n,k}^{'(i)})\}$$
(2.12)

のように定められる.ただし,  $f_{k-1}^k(l)$ は,前時刻の視角辺lに対応する現時刻の視角辺を指す. なお,図 2.6(a)のようなオクルージョン解消の際に,新たに出現した視角辺の接触性は,この段 階においては未定とする.

# 接触性変化

次に,個々の中間仮説 H'<sub>n,k</sub> について,部分オクルージョン・完全オクルージョン間の遷移によって生じる可能性のある視角辺の接触性の変化を推測し,事後仮説の部分候補集合 H''<sub>n,k</sub> を生成する.そのため,「排他的接触性」の条件を満たすような,各物体が属するセルと,各物体が接するセルの辺の組み合わせの対を

$$\left\{ \left\{ (L^{(i)}, C_{n,k}^{'(i)}) \right\}_{i=1}^{N'_{n,k}} \middle| L^{(i)} \in \mathcal{B}(E(C_{n,k}^{'(i)})), L^{(i)} \cap L^{(j)} = \emptyset, i \neq j, \cup_i L^{(i)} = Z_k \right\}$$
(2.13)

のように列挙する.ただし, $\mathcal{B}(\cdot)$ はべき集合をあらわす.この中から可能性の低い遷移を除外するために,式(2.5)で定義される尤度の高いものを上位 $K_S$ 個選択し,事後仮説の部分候補集合 $H''_{n,k}$ とする.

2.3.4 オクルージョン仮説の検証

事後仮説の候補集合  $\Theta_k$  の各要素について,現時刻の観測情報が与えられたときの事後確率を計算し,その値が閾値  $\epsilon$  を超えた仮説の中から上位  $K_P$  個を選択し,事後確率が高い順番に並び替えたものを事後仮説集合  $H_k$  とする.この仮説の閾値処理と個数の制限は,可能性の低い仮説の存在を抑制する効果がある.候補集合  $\Theta_k$  の各要素  $H \in \Theta_k$  の事後確率  $P(H|Z^k)$  は,視角辺の観測の時間的独立性を仮定し,ベイズ則を用いて,

$$P(H|Z^{k}) = \alpha_{H} \cdot P(Z_{k}|H) \cdot \sum_{n} P(H|H_{n,k-1}) \cdot P(H_{n,k-1}|Z^{k-1})$$
(2.14)

のように定義できる.ただし, $\alpha_H$ は正規化の係数である.また, $P(H_{n,k-1}|Z^{k-1})$ は,前時刻の 仮説  $H_{n,k-1}$ の事後確率を表す.式 (2.14)において  $P(H|H_{n,k-1})$ は,事前仮説からの事後仮説へ の遷移確率を指し,これは 2.2.6 節で述べた 2 種のオクルージョン構造の変化についての遷移確率 の積として,

$$P(H|H_{n,k-1}) = P(H|H'_{n,k}) \cdot P(H'_{n,k}|H_{n,k-1})$$
(2.15)

のように記すことができる.ただし, $P(H'_{n,k}|H_{n,k-1})$ はオクルージョンの発生・解消に関する遷移確率に対応し,ここでは1とする.また,部分・完全オクルージョン間の変化に対応する中間仮説  $H'_{n,k}$ から事後仮説への遷移確率 $P(H|H'_{n,k})$ は,各視角辺の接触性変化の独立性を仮定し,各視角辺の遷移確率の積として,

$$P(H|H'_{n,k}) = \prod_{l \in Z_k} P_l(H|H'_{n,k}), \qquad (2.16)$$

$$P_l(H|H'_{n,k}) = \begin{cases} 1 & \text{if } N_l = 1, \\ 1/N_l & \text{else if } l \text{ is a new edge}, \\ \eta & \text{else if tangency is unchanged}, \\ \frac{1-\eta}{N_l-1} & \text{else if tangency is changed} \end{cases}$$

のように定義する.ただし, $P_l(H|H'_{n,k})$ は視角辺lに生じる接触性変化の確率を表す.また, $N_l$ は視角辺lに接し得る物体の個数を指す.これを本論文では,視角辺lを辺として共有するセルの内部に含まれる物体の個数の合計として計算する.式(2.16)において,ある辺lに接し得る物体がただ一つの場合には,確率1を与える.また,オクルージョン解消により出現した新しい辺の場合には,均等の確率を与える.さらに,ある視角辺lが,中間仮説 $H'_{n,k}$ で割り当てられた物体と同一の物体に接触する場合には,停留確率 $\eta$ を与える.また,この視角辺lに接触する物体が他の一つの物体へと変化する場合には,停留確率 $\eta$ と視角辺lに接し得る物体の個数 $N_l$ を用いて, $(1-\eta)/(N_l-1)$ と与える.

#### **2.3.5** 物体状態の更新

各事後仮説  $H \in H_k$  のもとにおいて,現時刻の観測情報が与えられたときの物体状態の事後分布を計算する.ここでは複数の物体状態の間に存在する相互依存性の制約である「視角分離不可能性」及び「空間的排他性」を考慮するため,全ての物体の結合状態  $X_k = (x_{1,k}, \cdots, x_{N,k})$ に対する事後分布を計算する.これは,視角辺の観測の時間的な独立性を仮定し,ベイズ則を用いることで,

$$p(\boldsymbol{X}_k|H, Z^k) = \alpha_S \cdot p(Z_k|\boldsymbol{X}_k, H) \cdot \prod_{i=1}^{N} p(\boldsymbol{x}_{i,k}|H, Z^{k-1})$$
(2.17)

のように得ることができる.ただし, $\alpha_S$ は正規化係数である.式(2.17)において, $p(\mathbf{x}_{i,k}|H, Z^{k-1})$ は仮説 *H*における物体  $O_i$ の状態の事前分布であり,

$$p(\boldsymbol{x}_{i,k}|H, Z^{k-1}) = \sum_{n} P(H|H_{n,k-1}) \cdot P(H_{n,k-1}|Z^{k-1}) \cdot p(\boldsymbol{x}_{i,k}|H_{n,k-1}, Z^{k-1})$$
(2.18)

のように定義できる.ただし, $P(H|H_{n,k-1})$ は式 (2.15)で定義された事前仮説  $H_{n,k-1} \in H_{k-1}$ から事後仮説  $H \in H_k$ への遷移確率である.また, $p(x_{i,k}|H_{n,k-1}, Z^{k-1})$ は,2.3.6 節で述べる物体状態の予測分布である.式 (2.17) において,観測値  $Z_k$ に対する物体の結合状態の尤度  $p(Z_k|X_k, H)$ は,式 (2.7)で定義された各物体毎の尤度  $p(Z_k|x_{i,k}, H^{(i)})$ の積を用いて,

$$p(Z_k | \boldsymbol{X}_k, H) = \begin{cases} \prod_{i}^{N} p(Z_k | \boldsymbol{x}_{i,k}, H^{(i)}) & \text{if } \mathcal{G}_I \text{ and } \mathcal{G}_S \text{ are satisfied,} \\ 0 & \text{otherwise} \end{cases}$$

のように定義できる.ただし, $\mathcal{G}_I$ , $\mathcal{G}_S$ は,それぞれ「視角分離不可能性」と「空間的排他性」の条件を指し,どちらか一つでも満足しない場合には,尤度として0を与える.事後分布が $p(\mathbf{X}_k|H, Z^k) \equiv 0$ となるような仮説 H は事後仮説集合  $\mathbf{H}_k$ から除外される.

各物体状態の代表値 $\hat{x}_{i,k}$ は,第一位の仮説 $H_{1,k} \in H_k$ のもとでの結合状態の事後分布 $p(X_k|H_{1,k},Z^k)$ から,分布の重心など必要に応じて計算することができる.また,各物体毎の状態の事後分布 $p(x_{i,k}|H,Z^k)$ は,結合状態の事後分布 $p(X_k|H,Z^k)$ の周辺分布として得る.

#### **2.3.6** 物体状態の予測

前時刻 k-1の事後仮説  $H \in H_{k-1}$ のもとでの物体状態の事後分布  $p(\mathbf{x}_{i,k-1}|H, Z^{k-1})$ より,現時刻における予測分布  $p(\mathbf{x}_{i,k}|H, Z^{k-1})$ を求める.この予測の過程には,一般的に再帰的ベイズ推定法 [14,13] で用いられている予測モデルが利用できる.例えば,物体状態のマルコフ性,及び,物体間の状態の独立性を仮定することで,

$$p(\boldsymbol{x}_{i,k}|H, Z^{k-1}) = \int p(\boldsymbol{x}_{i,k}|\boldsymbol{x}_{i,k-1}) \cdot p(\boldsymbol{x}_{i,k-1}|H, Z^{k-1}) \, d\boldsymbol{x}_{i,k-1}$$
(2.19)

のように記述することができる.ただし,式(2.19)の $p(x_{i,k}|x_{i,k-1})$ は物体状態の遷移確率を表す. 本研究では,これは等速運動成分とシステムノイズと呼ばれる運動の不確定性成分からなるモデ ルを用いる.このシステムノイズは,等速運動では説明できない加速・減速などの物体の運動を 説明するためのモデルである.この遷移確率は,

$$p(\boldsymbol{x}_{i,k}|\boldsymbol{x}_{i,k-1}) = \int \delta(\boldsymbol{x}_{i,k} - F(\boldsymbol{x}_{i,k-1}, \boldsymbol{w}_{i,k-1})) \cdot p(\boldsymbol{w}_{i,k-1}) \, d\boldsymbol{w}_{i,k-1}$$
(2.20)

のように表すことができる.ただし, $p(w_{i,k-1})$ はシステムノイズの分布を表し, $w_{i,k-1}$ はノイズ 成分のベクトルを表す. $\delta$ はディラックのデルタ関数を表す.Fは遷移関数であり,物体の中心位 置のx成分に関しては,

$$x_{i,k} = x_{i,k-1} + \Delta k \cdot \dot{x}_{i,k-1}, \qquad (2.21)$$

$$\dot{x}_{i,k} = \dot{x}_{i,k-1} + w_{i,k-1}^{(x)} \tag{2.22}$$

のような演算を含むものである.ただし, $w_{i,k-1}^{(x)}$ はノイズ成分を表す. $\Delta k$ は単位時間ステップである.また,物体の中心位置のy成分や姿勢 $\phi$ についても同様の演算が行われる.

#### 2.3.7 初期化

初期時刻k = 1において,観測情報 $Z_1$ よりオクルージョン構造の初期仮説集合 $H_1$ と物体状態 の初期分布 $p(x_{i,1}|H, Z_1), H \in H_1$ が求められる.ここでは,初期時刻の各セル $C \in C_1$ に,それ ぞれ最大1つの物体が含まれるという仮定をおき,2.2.2節で述べた制約条件を満たすようなオク ルージョン構造の仮説を数え上げ,その中から尤度が高いものを最大で $K_I$ 個選択し,初期仮説集 合 $H_1$ とする.その後,各々の初期仮説について,物体状態の初期分布を2.3.5節と同様の処理に より得る.この初期化のプロセスは,一つのセルに2物体以上が存在する場合などにおいて,必 ずしも成功するとは限らない.その場合には,次の時刻で再び初期化を試みるという方策が考え られる.また,この初期仮説集合には,いくつかの異なる物体-セル配置や物体数を支持する仮説 が含まれることがあり,誤った物体-セル配置を示す仮説も含まれ得る.しかし,これら誤った仮 説は時間経過とともに淘汰され,結果的に正解の物体-セル配置をもつ仮説のみが生き残っていく と期待できる.

# 2.4 実験

提案した物体追跡法の有効性を確認するために実施した実験について述べる.まず,実行の手段として用いたパーティクル・フィルタについて述べ,続いて,実験に用いた評価データについて概説する.次に,実験結果の一例に基づき提案法の振る舞いを説明し,その後,統計的な実験結果を示し,オクルージョン構造や物体状態の推定精度を検証する.

#### 2.4.1 パーティクル・フィルタを用いた実装

実験において,提案した推定法を実行する手段としてパーティクル・フィルタを用いた.パー ティクル・フィルタ [14, 13] は,ベイジアン・フィルタ,逐次モンテカルロフィルタ, Condensation とも呼ばれ,系の状態の確率分布をパーティクル集合として近似的に表現し,逐次的に推定する方 法として知られている.特に本研究では,複数の仮説のもとでの物体の状態分布を単一のパーティ クル集合によって表現し,推定を行うために,一種の拡張を施したパーティクル・フィルタを導入 している.このパーティクル・フィルタでは,各パーティクルの属性として,各仮説を支持する割 合が追加されている.これを仮説支持率と呼び, $\Lambda_{i,k}^{(s)} = {\lambda_{i,m,k}^{(s)}}_m$ と記す.ただし, $\sum_m \lambda_{i,m,k}^{(s)} = 1$ である.ここで*m* は各仮説を指すインデックスである.これにより各物体 *O<sub>i</sub>*のパーティクル集 合は

$$\widetilde{X}_{i,k} = \left\{ \left( \widetilde{\boldsymbol{x}}_{i,k}^{(s)}, q_{i,k}^{(s)}, \Lambda_{i,k}^{(s)} \right) \right\}_{s=1}^{\mathcal{N}}$$
(2.23)

のように表記できる.ただし, $\tilde{x}_{i,k}^{(s)}$ は,物体状態 $x_{i,k}$ のs番目の実現値である.また, $q_{i,k}^{(s)}$ はパーティクルの重みを表し, $\sum_{s}^{\mathcal{N}} q_{i,k}^{(s)} = 1$ である. $\mathcal{N}$ はパーティクル数を表す.仮説支持率 $\Lambda_{i,k}^{(s)}$ を用いると,ある仮説 $H_{m,k}$ のもとでの物体状態分布は,重み $q_{i,k}^{(s)}$ を $q_{i,k}^{(s)} \times \lambda_{i,m,k}^{(s)}$ に置き換えたパーティクル集合により表現でき,通常のパーティクルフィルタの枠組みを適用することができる.

また,各時刻における推定の結果は,物体の中心位置 $\hat{x},\hat{y}$ ,及び,姿勢 $\hat{\phi}$ とした.ここでは,第一 位の仮説のもとでの物体状態の事後分布 $p(x_{i,k}|H_{1,k},Z^k)$ から各成分についての重心を計算し,推 定値とした.また,2.3.6 節で述べた物体の状態遷移モデルにおけるシステムノイズの分布として, 各状態の要素に対して独立なガウス分布(平均0)を用いた.その標準偏差は,物体の中心位置, 姿勢について,それぞれ,経験的に1.0,4.5[deg]とした.なお,単位時間ステップは $\Delta k = 0.02$ とした.その他のパラメータは,経験的にパーティクル数 $\mathcal{N} = 5000$ 個,停留確率 $\eta = 0.6$ ,仮 説の事後確率の閾値  $\epsilon = 10^{-3}$ ,仮説数の条件 $K_I = K_P = 6, K_S = 10$ ,観測の尤度関数の分布幅 h = 2.0[deg]とした.

## 2.4.2 評価データ

評価データとして,図 2.10(a) のような複数の楕円形の物体が閉領域内で運動する様子を観測し たデータを作成した.楕円は短半径 0.25,長半径 0.5 とし,物体数は 8,6,4 の 3 通りとした.物体 の運動モデルは,等速運動成分にランダムな加速度成分を加え,さらに,物体間,及び,物体と 閉領域の壁面との間の斥力を加えたものを用いた.この斥力は,物体同士の衝突回避と物体の領 域内拘束の役割をもち,そのモデルとして,分子動力学法におけるソフトコアモデル [134] を用い た.なお,閉領域の壁面は観測に影響を与えないものとした.この運動モデルにより,物体に複 雑な運動が生じ,その結果として,絶え間なく変化するオクルージョンを含むデータが生成され た.初期時刻において,各物体を互いに重なりのないようにランダムに配置した.また,初期時



図 2.10: 評価データの例 (8 物体, 6 視点). (a) 視角の集合, (b) 各視点で観測された視角の区間 [deg].

刻における物体の速度は0とした.

物体を観測する視点を,図 2.10(a)のように半径7の円周上に等間隔に配置した.なお,視点数 は8,6,4個の3種類とし,観測の誤差はないものとした.図 2.10(b)に各視点における観測像を 示す.一つの矩形が一つの視角に対応する.一つの視角内に含まれる物体数の平均は8,6,4物体の 場合,それぞれ2.70,1.97,1.48となり,ほぼ全ての視点において何らかのオクルージョンが生じ ている.また,各物体当りの平均の接触辺数は,最少2.97本(4視点8物体の場合),最多10.8 本(8視点4物体の場合)となった.

2.4.3 オクルージョンに対する振る舞い

提案法の振る舞いを図 2.11 に示す一例(6 視点 8 物体)を用いて説明する.図 2.11(a)-(d)は, 当初別個のセルに含まれていた4つの物体が段階的に一つのセルに包含され(k = 75),物体 $O_1$ に ついて全ての視点からみて完全オクルージョンが発生し(k = 81),その後,部分的にオクルージョ ンが解消され(k = 85),再び個別のセルへと復帰する状況を表したものである.図中の実線の楕 円は推定値,破線の楕円は真値をそれぞれ表す.また,影のついた領域はセルを表す.さらに,図 中のパーティクル集合は,推定された物体状態の中心位置に関する分布を表す.なお,これは全仮 説における混合分布となっている.パーティクルの重みは一様である.また,図 2.12 は,物体毎 の状態分布(パーティクル集合)の分散の時間変化を表したグラフである.この分散の値は,物体 の中心座標の推定値と真値との間の距離の自乗の平均値として計算した.図 2.11 において,物体  $O_1$ は,時間とともに接触辺数が減少し,時刻k = 81においては,いずれの視角辺とも接しない,



図 2.11: オクルージョンに対する振る舞い.(太線の楕円:推定結果,細線の楕円:真の楕円,直線:視角辺).



図 2.12: 推定された物体状態分布 (パーティクル集合)の位置に関する分散の時間遷移.

つまり、いずれの視点からも隠れていて観測できないという状況になる.この場合、図 2.11(c) の ようにパーティクルが広範囲に分布し、物体状態には大きな不確定性が存在することがわかる.こ の様子は図 2.12 のプロットからも読み取れる.しかし、その後、部分的にオクルージョンが解消 し、物体  $O_1$  の接触辺が増えることにより、k = 85 に示すように、パーティクル分布の分散が急 激に減少し、再び確実性の高い推定値へと復帰した.一方、物体  $O_2$  のように、オクルージョン状 況下においても、その接触辺が安定に保持される場合には、確実な状態の推定が継続されること がわかった.

このように提案した推定法では,オクルージョンが生じる状況において,それに起因する物体 状態の不確定性を時間的に伝播し,オクルージョン解消後の観測情報と統合させることにより,物 体の追跡が安定に継続できることがわかった.

## 2.4.4 統計的評価

物体数 8, 6, 4, 視点数 8, 6, 4 の 9 通りの組合せの条件のもと,初期時刻から 100 時刻分の運動 データをそれぞれ 50 試行分作成し,実験を行った.その結果に対して,物体-セル配置,平均仮説 数,接触性の正解率の観点からオクルージョン構造の推定の精度について検証した.また,物体 の状態に関しては,位置と姿勢の推定誤差の観点から評価を行った.

## 初期化,及び,物体-セル配置の誤り

まず,初期化の成功率は,6視点8物体,4視点8物体,4視点6物体の場合,各々96%,90%, 98%となり,その他の場合では100%となった.今回のデータに対して1セル1物体の仮定は概ね 有効であることがわかった.表2.2(a)(b)(c)には,初期化に成功した試行のなかで,誤った物体-セル配置の仮説を一つでも含む試行の割合の時間的推移(k = 1,5,20)を,8物体,6物体,4物 体の場合についてそれぞれ記す.ここで誤った物体-セル配置とは,推定された物体数が真値と異 なる場合や,真の物体が含まれるセルに推定結果の物体が存在しない場合,また,真の物体が含 まれないセルに推定結果の物体が存在する場合のことを指す.表2.2(a)(b)より,誤配置の仮説を 含む試行の割合は,時間とともに淘汰され,減少していくことが確認できた.しかし,視点数に 対して物体数が多い4視点8物体の場合,初期時刻以降においても誤配置の仮説への遷移が頻繁 に存在するため,その誤配置の仮説を含む試行の割合は0%へは収束しないことが確認された.

#### オクルージョン構造の不確定性

表 2.2(d) には,全試行の各時刻において得られた仮説数の平均値を示す.仮説の個数は,オク ルージョン構造の不確定性の大きさに対応すると考えられる.表 2.2(d) より平均仮説数は物体数

#### 第2章 多視点オクルージョン解析による複数物体の追跡

表 2.2: 仮説に関する評価結果.(a)(b)(c) 誤配置の仮説を含む試行の割合 [%]((a)8 物体,(b)6 物体,(c)4 物体),(d) 平均仮説数.

	(a)8 objects			(b)6 objects			(b)4 objects			(d)number of hyps.			
views	k=1	k=5	k = 20	k=1	k=5	k = 20	k=1	k=5	k = 20	8 obj.	6 obj.	4 obj.	
8	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.93	2.95	1.53	
6	14.6	6.98	0.0	2.0	2.0	0.0	0.0	0.0	0.0	5.28	3.79	1.78	
4	71.1	51.2	27.3	26.5	17.0	9.09	0.0	0.0	0.0	5.76	4.72	2.28	

表 2.3: 推定精度の評価結果 , (a) 接触性正解率 , (b) 平均持続長 , (c) 中心位置の平均誤差 , (d) 姿 勢  $\phi$ の平均絶対値誤差 .

	(a)tangency			(b)continuity				(d)posture				
	num. objects			num. objects			num	num. objects				
views	8	6	4	8	6	4	8	6	4	8	6	4
8	95.1	97.7	99.1	80.6	97.7	98.4	2.87e-3	6.00e-4	1.65e-4	1.21	0.50	0.32
6	93.5	96.9	98.8	69.9	95.0	97.8	1.71e-2	2.37e-3	2.60e-4	2.55	0.88	0.44
4	84.9	94.7	98.2	34.6	73.5	100.0	1.72e-1	1.13e-2	1.38e-3	11.8	4.33	1.51

が多いほど大きい値をもつことがわかる.これは物体数が多いほど,より複雑なオクルージョン構造が生じていることに対応すると考えられる.また,表2.2(d)より同一の物体数においては,視点数が多いほど平均仮説数が減少していることがわかる.これは視点数が多いほど確実にオクルージョン構造が推定できることを示唆するものと考えられる.

#### オクルージョン構造の推定精度

表 2.3(a) には,接触性正解率を示す.これは第一位の仮説によって割り当てられた視角辺と物体との組合せが正解していた割合を,全試行の各時刻について平均したものである.これはオクルージョン構造の推定精度の尺度とみなすことができる.表 2.3(a)より,接触性正解率は,視点数が同一の場合,物体数が少ないほど高い値を示し,また,同一の物体数の場合では,視点数が多いほど高い値を示すことがわかった.このように,視点数と物体数の条件により接触性正解率は左右されるが,物体数に対し十分な視点数を与えることで,オクルージョン構造が正確に推定できることがわかった.また,第一位の仮説の接触性正解率は,第2位以下のものよりも平均的に高い値をもつことが確認されている.

#### 追跡が継続した時間の長さ

表 2.3(b) には,各試行の平均持続長を示す.これは各試行において追跡が持続した時間の長さ の平均値であり,この値が長いほど,頑健に追跡が実行されたことを意味する.よって,これは 物体追跡法としての性能を表す一尺度であるといえる.表 2.3(b)より,物体数が少ないほど,ま た,視点数が多いほど,長い時間にわたって追跡が持続したことがわかり,そのような場合にお いて,提案法は物体追跡法として有効に機能することがわかった.また,4視点8物体の例のよう に物体数に対して視点数が大幅に少ない場合,平均持続長は著しく小さくなり,有効な追跡が困 難であることがわかった.これらより,オクルージョンに頑健な追跡を行うためには,想定され る物体数に対して適切な視点数の設定が不可欠であることが示唆された.

この平均持続長の傾向は,表2.2(d),表2.3(a)の平均仮説数,接触性正解率の傾向と符合し,オ クルージョン構造の不確定性が大きい場合に,追跡が中断される確率が高くなるという特性があ るといえる.今後,これらの関係をさらに調査し,追跡が継続可能な視点数の条件等を理論的に導 出することが課題としてあげられる.さらに,実際にこの推定法を応用する際には,仮説数が0に なる段階において再初期化を導入することで,追跡の中断を回避することができると期待される.

#### 物体の位置・姿勢の推定誤差

表 2.3(c)(d) に,第一位の仮説のもとで推定された各物体の中心位置と姿勢の推定誤差をそれぞ れ示す.物体の中心位置の誤差は,推定値と真値との距離の平均であり,姿勢の推定誤差は推定値 と真値との平均絶対値誤差である.これらは全試行の各時刻,各物体に関する平均値である.中 心位置に関しては,表 2.3(c)より,物体の大きさ(長半径 0.5)と比較して,全般的に精度の高い 推定値が得られたことがわかった.また,姿勢に関しては,表 2.3(d)より,物体数に対して視点 数が比較的少ない場合(特に4視点8物体)において,姿勢の誤差が大きいことがわかった.こ れは,物体当りの接触辺数が減少し,物体を含むセルの面積が大きくなることで,物体の回転を 許容するような物体状態の自由度が生じるためであると考えられる.また,表 2.3(c)(d)から,視 点数の増加,及び,物体数の減少に伴い推定精度の向上の傾向が確認できた.このような推定誤 差の傾向は,平均持続長とあわせて,必要な視点数を決定する際の参考情報として利用できる.

# 2.4.5 実データへの適用

提案手法の実データへの適用例として,図 2.13(a)のように 5 人の人物が約  $3 \times 3$ [m<sup>2</sup>]の領域をランダムに歩行するシーンを対象とした実験結果を示す.ここでは,6 台の固定カメラを図 2.13(a)中の $v_1 \sim v_6$ で指し示される壁面の位置に設置し,撮影を行った(VGA サイズ,30[fps]).各画像において,背景差分法により人物領域の抽出を行い,抽出された領域を基準平面(床面上 1.3[m])へ投影



図 2.13: 実データの様子 , (a) 上部から見たシーンの全体像 (k=53), (b) 推定結果 (k=53).



図 2.14: 各カメラからの画像 (k=53).



図 2.15: 中心位置の軌跡. 300[frame](=10[sec]) に渡って描画.

することで視角の集合を形成し,観測データとした.人工データとは異なり,観測された視角は分解 能が低く (=0.5[deg]),多くのノイズや外れ値を含む.これらに対処するために,本実験では,形状 のパラメータとして,楕円を短径  $b \in [0.08, 0.12]$ [m],楕円の短径と長経の比 (長経/短径) $a \in [1.8, 2.7]$ のように可変のパラメータにより表し,それらを状態ベクトル  $x_i = [x_i, \dot{x}_i, y_i, \dot{y}_i, \phi_i, \dot{\phi}_i, a_i, b_i]^T$ に含めて推定を行った.これらパラメータの初期分布は,パラメータの範囲上の一様分布とした.なお,人工データを対象とした実験と比較し,状態ベクトルの次元数が増えたため,パーティクル数を $\mathcal{N}=7000$ と増加させた.

図 2.14 には, 典型的なオクルージョンの状況において観測された画像を示す.この例では, どの視点からも人物  $O_4$  と人物  $O_5$  を分離して観測することができない.図 2.13(b) には, この時点における推定結果を示す.図 2.13(b) 中のパーティクル分布から,同一のセル内に位置する人物  $O_4$  と人物  $O_5$  の推定結果に大きな不確定性が生じていることが読み取れるが,図 2.13(a) に示した実際の人物配置と比較して,この推定結果はほぼ妥当なものといえる.

また,図2.15は,10秒間にわたって各人物の中心位置の軌跡を描画したものである.図2.15に おいて,軌跡の太さは不確定性の大きさを表す.一時的に不確定性が増大した箇所や局所的な軌 跡の不安定性はみられるものの,人物の移動に伴って多様にオクルージョンの状況が変化する状 況に対して,提案手法は追跡を継続することができたことが分かる.

# 2.5 議論

本章では,多視点観測環境におけるオクルージョン解析の枠組みを提案し,実験によりその基本的な原理の有効性を確認した.本章では単純化された世界を対象としているものの,2次元の 視角という極めて限定された観測情報のみにより,一時的に完全オクルージョンが生じるような 状況に対しても頑健な追跡が実行できることを確認した.本研究では,屋内の壁面にカメラを設 置し,歩行する人間を追跡するという人間行動監視において一般的な状況を念頭に置いているが, 本論文では,議論の焦点をオクルージョン問題に絞るため,様々な仮定や制約条件を含む単純化 された世界を対象とした.そのため,本論文の提案法を直接,現実世界へ適用することは難しい が,その枠組み自身は,現実世界へ適用可能な拡張性を有する.本節では,提案法を現実世界へ 適用する際の各種の仮定・条件の妥当性や問題点,及び,問題の解決に向けた提案法の拡張の方 向性について論じる.

まず,本論文では,視角の観測に際し,誤検出や未検出がない状況を仮定していたが,現実の観 測においてこれらの発生は避けられない問題である.視角の誤検出としては,観測画像上にノイ ズが重畳し,物体に由来しない視角が観測される状況が考えられる.また,未検出としては,本 来,一つの視角として観測されるべきに視角が複数の視角に分断されて観測される状況が考えら れる.前者の場合,誤って検出された視角に対しては「排他的接触性」が満足されず,また,後者 の場合,セルが分断されるため「物体包含性」,「排他的接触性」ともに成立しない.そのため,現 状の方法ではこれらの状況には対処できない.これら状況に対する対処策としては,例えば,観 測された個々の視角辺について,それらが誤検出・未検出に由来するものである可能性を考慮し, オクルージョン構造の多重仮説を生成するという方法が考えられる.この場合,前時刻から予測 された物体状態とオクルージョン構造に対して,大幅に食い違う視角の辺を外れ値とみなすよう な仮説を生成し,この外れ値を物体状態の推定に使用しないことで,誤検出や未検出の悪影響が 抑制されると期待できる.

また,本論文では視角辺の角度の観測誤差のモデルとして,式(2.8),(2.9)のような分布を仮定 することで,微小な視角辺の角度のずれに対処している.しかしながら,現実世界においては,こ の誤差モデルでは吸収しきれないような大きな観測誤差が生じることもあり得る.例えば,視角 辺が視角の外側に大きくずれる場合,この視角辺がセルから外れてしまい,排他的接触性が満足 されない状況が生じる.また,視角辺が視角の内側に大きくずれる場合,この視角辺が構成する セルの面積が狭まり,物体包含性が満足されなくなる.これらの問題に対処するためには,上述 の誤検出・未検出への対処と同様に,観測された視角辺に対して,それらが外れ値である可能性 を含めた仮説生成を行うという拡張が必要とされる.また,このような大きな観測誤差に対処す るため,誤差モデルのパラメータhを大きくする場合,式(2.8),(2.9)の尤度分布の幅が広がり, セルに対して当てはまりの悪い物体の配置が解として許容されるため,推定される物体の状態分 布の分散が大きくなる.従って,相対的に正しい解が時間伝播される確率が低下し,結果的にオ クルージョンに対する頑健性が低下するという現象がみられる.この悪影響を回避するためには, より現実の状況を反映した観測誤差のモデルが必要であると考えられる.

また,本論文では,オクルージョンによって重なった物体像から,個々の物体を識別,分離する ことはできないという仮定(視角分離不可能性)を導入している.しかし,物体表面の色や見か け等の各物体に固有の特徴を利用することで,重なり合った物体像から個々の物体を分離するこ とも可能だと考えられる.また,視点間のphoto-consistency[135]も物体の位置に対する制約とし て有効であると考えられる.このような,色・見かけの特徴や制約を,オクルージョンの空間構 造に関する制約と組み合わせて利用することで,オクルージョンに対するロバスト性のさらなる 向上が期待できる.

また,環境やカメラの配置に関する制約の緩和も課題である.本論文では,全カメラの視野の 共通領域内に物体が存在し,この領域内外への物体の進入・退出がないことを仮定している.さ らに,物体間のオクルージョンのみを考察の対象としている.しかし,現実の環境では,観測可 能な領域内外への物体の進入・退出や,カメラから見て物体が柱や壁など障害物の影に隠れるよ うな環境によるオクルージョンも存在する.提案法を現実の環境へと適用する際には,このよう な状況への対処が不可欠である.例えば,観測可能領域への物体の進入・退出は,視角の出現・消 滅に対応するため,予め設定された各カメラの観測範囲の境界における視角の変化を物体の退出・ 進入として解釈し,オクルージョン仮説を生成することで対処できると考えられる.また,環境 によるオクルージョンに対しても,各視点から障害物までの距離が既知であれば,該当位置にお ける視角の消滅や出現を,環境によるオクルージョンとして解釈し,仮説を生成することが可能 であると考えられる.

本論文で想定しているような屋内の壁面にカメラを設置し,歩行する人間を側面から観測すると いう状況は,コンピュータビジョン分野の人物追跡においては一般的な状況である [25, 31, 27, 28]. しかしながら,このような側面からの観測は,最も隠れが発生し易い状況であるといえる.その ため,実際の人物行動の追跡・監視においては,斜め上方からの観測の方が,物体間の隠れを削 減できるという観点から望ましいと考えられる.このようなカメラ配置に対応するための提案法 の拡張として,床面上の人物のように平面上を運動する物体を対象とする場合には,3次元的な観 測情報を2次元の基準平面上で統合し,その上で本論文の2次元的な追跡法を適用するという方 法が考えられる.例えば,物体形状モデルを楕円からの楕円柱に拡張し,これを斜め上方に設置 されたカメラから観測する場合,観測画像上に現れる楕円柱の垂直輪郭線を基準平面上に逆投影 することで視角を得ることができる.この場合得られる視角辺は,物体の高さとカメラ位置に応 じた有限の長さを持つ.本論文の2次元世界の観測で得られる視角辺は無限長であるため,1本の 視角辺からは物体の位置が限定できなかったのに対し,このような3次元的な観測により得られ る視角辺からは,物体の存在し得る位置が限定できるという性質がある.この性質によりカメラ 配置の3次元化によってオクルージョンに対するロバスト性の向上が期待できる.

提案した推定の枠組みは,仮説生成における組み合わせ最適化の問題などを含み,その実行に は大きな計算量が要求される.その計算量の削減のためには,仮説生成における近似アルゴリズ ムの利用が有効であると考えられる.例えば,同時に接触性が変化する視角辺の本数を制限する ことにより,仮説の探索の範囲を限定することができる.また,仮説の生成や物体状態の事後分 布の推定など,仮説毎に独立した演算を多く含むため,並列計算も有効なアプローチであると考 えられる.

# 2.6 結言

本章では,時系列画像解析における観測の部分性の一要因として,複数の物体間のオクルージョ ンに着目した.具体的には,多視点観測による複数物体の追跡の問題を取り上げ,オクルージョン に対して頑健な追跡を目指して,物体間の相互オクルージョンの解析と物体状態の逐次推定を行 う方法を提案した.この方法では,2次元平面上の2次元物体を対象とし,2次元平面上の視角と 物体との接触性に基づいて,オクルージョンを含む観測過程が陽に記述される.また,物体追跡 の問題を,オクルージョンの構造の仮説生成・検証と,物体状態の事後確率分布の推定からなる 再帰的ベイズ推定法として定式化した.人工データを用いた実験により,複数の物体が密集し,か つ,多様にオクルージョンが変化する状況においても,安定した物体追跡が可能であることを示 し,提案法の基本原理の有効性を確認した.また,実データに対しても試行的な実験を行い,今 後の本格的な実環境適用に向けた研究の課題と方向性を議論した.

本章で提案したオクルージョンの観測過程を陽にモデル化するというアプローチは,観測の部 分性の問題に対する有効なアプローチの一つとして位置づけることができ,人物の移動シーンの 他に様々な実世界の動的事象へと適用が期待できる.なお,本章では対象間の相互作用は,衝突 回避など最小限に留まっていると仮定したが,現実の実世界事象は,対象間の相互作用が事象の ダイナミクスに対して決定的な影響を持つ場合も少なくない.このような場合には,オクルージョ ンによる観測の部分性の他に,対象間の相互作用などの観測の間接性への対処が必要不可欠とな る.次章では,そのような事例の一つとして,複数人物による対面会話シーンを取り上げ,各人 物の行動の観測から人物間の相互作用の状態を推測する方法を検討する.

# 第3章 非言語行動に基づく会話構造の推定

# 3.1 緒言

複数人物による対面会話は,情報の伝達・共有,他者の意図・感情の理解,グループの意思決定 などにおいて,欠かすことができないコミュニケーションの形態の一つであり,情報通信技術の 発達した今日にあっても,我々人間の社会的活動において欠かすことができないものである.し かし,対面の会話は,会話参加者全員がその場に居合わせることが必要であり,また,そこで取 り交わされる情報の内容も各人の記憶のみに留まり通常,形として残ることがない.近年,この ようなその場限りという会話の限界を超え,より人間同士のコミュニケーションが円滑になるよ うサポートする技術として,遠隔会議[80],映像議事録[81],会議映像の自動編集[82],会話エー ジェント [83,84] や会話ロボット [85,86] などへ大きな期待が寄せられている.そこで現在,これ らのアプリケーション実現のための基本技術の一つとして,会話シーンを自動的に認識・理解す る技術が求められている.

複数人物による会話は,参加者間でのメッセージ交換の過程として捉えることが可能である.こ のメッセージには,言語的なメッセージに加えて,顔の表情,視線,手振り・身振り,姿勢などの 非言語的なメッセージが含まれる.対面会話において人間は,受信したメッセージから他者の意 図や心的状態を推測し,また,自らの意見・意図を他者に伝達すべく言語・非言語行動を通じて メッセージを発する.このような会話参加者の相互作用によって会話という現象が進展する.そ こで,対面会話シーンの自動的な認識・理解には,会話参加者間でどのような相互作用が生じて いるか推測することが不可欠であると考えられるが,その相互作用自体は,視覚的に直接観測は できず,観測可能な人物の行動から間接的に推測を行う必要がある.そのため,対面会話シーン の認識・理解の問題は,観測の間接性が顕著であるといえる.

人間の行動認識は,古くよりコンピュータビジョン分野などにおいて研究が進められており,近年,その対象は,単一の人物の姿勢やジェスチャから,複数の人物間のインタラクションまで拡大している.近年,中でも人物間のインタラクションが生じる場として,会議や会話シーンに注目が集まりつつあり,隠れマルコフモデルやその拡張である HMMs[91], Layered-HMM[92], Coupled-HMM[87],動的ベイジアンネットワーク[90] などを用いて,会議・会話中の各参加者の行動やグループ行動を認識する方法が提案されている.例えば,McCowanらは,会議中の行動として,プロジェクタ利用の発表,ホワイトボードの使用,メモ書き,ディスカッションなどの認識を目標

として,HMMsを用いた手法を提案している[91].これらの方法は,従来の時系列パターン認識 の枠組み[9]に基づいており,認識対象となる人物行動のカテゴリーと画像上での人物の位置・動 作との対応関係が明確であり,かつ,事前に十分な学習データが入手可能であることが前提とさ れる.また,人物の行動の生成要因となっている人物の内部状態や相互作用のモデル化・推測は 検討の対象とはしていない.このような観点から,会話シーンの認識・理解技術は,いまだ研究 の初期段階にあるといえる.

本研究では、複数人物の会話シーンを理解するためには、人物の行動を生じせしめる人物の内 部状態や相互作用を含む場の状態,そのダイナミクスまで踏み込んだモデル化と推測が不可欠で あると考える.そこで前述したメッセージ交換の観点から,本研究では,人物間でやりとりされ るメッセージの流れを明らかにすることを目標の一つとして定めている.このメッセージの流れ は,誰が誰に話し掛けているか?,誰が誰の話を聞いているか?という会話行動として言い換える ことができる.また,会話の各時点において,誰が話し手で誰が聞き手であったか?という各参 加者の役割として記述することができると考える、会話における参加者の役割は、社会心理学の 分野では「参与役割」と呼ばれている [98, 99, 96, 95, 97] . Goffman [98] によると会話参加者は , 承認された参加者と承認されていない参加者(立ち聞きしている者)に分類され,さらに,承認 された参加者は,話し手,受け手,及び,傍参与者に分類される.ここで,受け手は,話し手に話 し掛けられている人物を指し,傍参与者は,承認された参加者のうち,話し手でも受け手でもな い人物を指す.話し手は,発話権(ターン)を保持している者として位置付けられる.また,受け 手と傍参与者は聞き手とも呼ばれる.会話の進行とともに参加者間で発話権が移動し,それに伴 い,各参加者の参与役割も動的に変化する.このような参与役割を自動的に推定する技術は,様々 なアプリケーションにおいて重要な要素技術であると考えられる.例えば,会議映像の自動編集 システムにおいては,参与役割に応じた映像の切り替え表示などにより,「誰が誰に話し掛けてい るか」といった情報を視聴者にわかりやすく伝達できるようになると期待される.また、会話ロ ボットにとってもグループ会話に参加するうえで参与役割の推定は不可欠であることが指摘され ている [86, 85].

本研究では,参加者間のメッセージの伝達パターンを基本的な会話の構造として捉え,その記述法として参加者とその参与役割との組み合わせを採用する.また,参与役割の推定の手掛かりとして,以下の観点から参加者の非言語的な行動に着目する.従来,参与役割の内,話し手については,各参加者の音声信号から各人の発話の有無を判定することで,おおよそ同定が可能であることが知られている.しかし,三者以上の会話においては,聞き手である受け手と傍参与者を音声信号のみから区別することは困難とされている.また,発話者が,ある参加者に話し掛ける場合,常に,その相手の名前を発話に含めるとは限らないため,各人の発話に含まれる言語的な情報も,受け手と傍参与者を区別するための手掛かりとしては不十分であると考えられる.

その一方,対面状況においては参加者の非言語的な行動も,コミュニケーションを遂行するた

めの重要な要素であることが知られている [59, 56, 58].このような非言語行動には,視線や顔の 表情,頷き,手振り・身振り,姿勢などがあるが,中でも視線の役割の重要性は古くより指摘され ている [103, 104, 100, 97].例えば,Kendonは,視線には,他者の行動をモニタリングする機能, 自らの態度や意図を表出する機能,会話の流れを調整する機能が具わっていることを示唆してい る [103].また,Goodwinは,話し手は視線を使って,誰に話し掛けているのかを他の参加者に提 示し,また,受け手は,その視線を話し手に向けることで,話し手の話を聞いていることを話し 手に合図しており,このような話し手と受け手の間の相互の注目が会話の成立の上で不可欠であ ると主張している [104].また,Vertegaalらは,会話参加者は参与役割を理解するための手掛か りとして視線を用いていることを実験的に明らかにしている [136].これらの知見に基づき,近年 では,会話参加者の視線行動を分析することにより,自動的に参与役割を求めることができるも のと期待が高まっている [137].

このように会話中の視線と参与役割との関連性は古くより示唆されているが,参与役割の自動 的な推定を実現するためには,両者の関係を定量化する必要があり,そのためには,まず会話中 の視線行動を実際に計測する必要があると考えられる.人間の視線方向を自動的に計測する方法 は,これまでにもいくつか提案がなされているが[138,139],複数人の自然な会話を妨げないよう に視線方向を正確,かつ,安定に計測することは,いまだ困難なタスクとされている.そのため, これまで視線の代用として,より容易に計測が可能な頭部の姿勢・方向を用いるというアプロー チが試みられている[106,83,85,93].このアプローチは,人間には興味の対象をその視野の中央 で捉えようとする性質があり,それにより,視線を向けた先の人物との位置関係に応じて,頭部 や胴体の姿勢が変化する,という性質に立脚するものである.このアプローチの先駆的な例とし て,Stiefelhagenらは,4人会話において,頭部方向に基づいて,誰が誰を見ているか判定可能で あることを示している[106,107].また,簡単な会話の状況を対象に,話し手の頭部方向から,そ の発話が向けられている相手(受け手)を判定する試みも,会話ロボット開発の一環として始まっ ている[86,85].

このような背景を踏まえ,本研究では,N人物 ( $N \ge 3$ )による対面会話を対象とし,会話中 の各時刻における,会話参加者とその参与役割の組み合わせを会話構造として捉え,その会話構 造の自動的な推定を目標とする.その一環として,本論文では,従来研究において採られてきた, 人物行動の直接的な認識・解釈に基づくアプローチとは異なる新しいアプローチを提案する.こ のアプローチとは,会話の各時点における参加者の行動は,会話の構造(以下,会話構造とも呼 ぶ)によって規定されるという仮説を立て,会話という現象を,会話の構造に相当する一種の上 位プロセスと参加者の行動という下位のプロセスとの間の相互作用によって時間発展する系と見 立てて,確率的なモデルを構築し,このモデルに基づいて会話構造を推定するというものである. ここでは,このような参加者の行動と会話の構造との関係,及び,そのダイナミクスを表現する 会話現象のモデル(以後,会話モデルと呼ぶ)として,マルコフ切替モデル(Markov-Switching Model)[140,141]と呼ばれる一種の動的ベイジアンネットを用いる.マルコフ切替モデルは,レ ジームと呼ばれる上位にあるプロセスの状態がマルコフ過程に従って変化しつつ,その状態に依 存して下位にあるプロセスのダイナミクスが決定されるという階層的な構造を有する.本論文で は,このモデルのレジームを「会話レジーム」と呼び<sup>1</sup>,本研究で推論の対象とする会話構造とみ なす.なお,現実の会話は多様な構造を取り得るが,本論文で扱う会話レジームは,その部分集合 であり,会話中に頻繁に出現する代表的な会話構造に対応すると考える.また,下位のプロセス の状態が参加者の行動に対応すると仮定し,会話レジーム(以下,単にレジームとも呼ぶ)の状 態に依存して確率的に参加者の行動が生成されるというモデルの構成を試みる.ここでは,参加 者の行動として,会話構造との関連性が示唆されている視線行動(視線の方向)と発話状態(発 話の有無)を導入する.また,視線方向は直接的に計測が困難であるため,これもレジームと同 様に推定すべき隠れ変数とみなし,その推定の手掛かりとして頭部方向を計測の対象とし,その 観測過程のモデルを会話モデルに組み込む.

本研究では,このような会話モデルを具体的に構成するため,会話中に出現する視線パターン (各参加者の視線方向の集合)の頻度を分析し,頻出する視線パターンに含まれる特徴的な構造に 基づいて,会話モデル中の会話レジームと現実の会話構造との関係,及び,会話レジームと参加 者の行動との関係を仮説的に設定する.また,観測された頭部方向と発話状態に基づいて,会話 中の各時刻におけるレジーム状態と視線パターン,及び,モデルパラメータのベイズ推定を行う 方法として,ギブスサンプリングと呼ばれる一種のマルコフ連鎖モンテカルロ法 [142,143]を用 いた方法を提案する.本論文の提案方法は,頭部方向から視線方法を近似的に推定するという問 題と,視線方向と発話状態から会話構造を推測するという2つの問題を同時に解く方法とも捉え られる.視線方向に対する頭部方向の自由度より,頭部方向からの視線方向の推定には曖昧性が 伴うが,会話モデルを用いることで,会話の構造に応じて出現しやすい視線方向が想定できるこ とから,精度の高い視線方向の推測が可能となると期待される.

また,本論文では4人会話を対象として実験を行った結果を示す.まず,会話レジームと人物 行動の関係が適切にモデル化されていることを確認するために,視線方向の推定精度を評価する. 次に,推定された会話レジームが実際の会話構造を反映していることを確認するため,発話の種 別・方向性のラベルを用いた評価を行う.これら結果に基づいて,提案した会話モデル,及び,会 話構造の推定方法の有効性を確認する.

なお,本研究では,会話参加者の頭部方向の計測手段として,画像中の人物頭部のトラッキン グによる方法と,装着式の磁気式センサによる方法を採用し,両者それぞれを用いて評価をおこ なう.トラッキングによる方法と比較して,磁気式センサによる計測は安定かつ正確である.よっ て,まず,磁気式センサによる頭部方向の計測値を用いて,頭部方向が正しく計測された条件に

<sup>&</sup>lt;sup>1</sup>レジーム(regime)という用語は,一般的には政治体制,気候変動の型などを指す場合に使われる.本論文では会話構造によって参加者の行動が規定されるという会話モデルの構造を含意させるためにこの用語を用いることとした.

おいての会話構造推定法の有効性を検証する.その後,これと頭部トラッキングによる結果を比較することで,頭部トラッキングの有効性を検証する.また,トラッキングの方法としては,疎 テンプレート Condensation 追跡法 [144] と呼ばれる方法を導入する.

本章は以下のように構成される.3.2節において,視線パターンの分析に基づいて会話レジーム の設定をおこなう.3.3節においては,会話モデルの定義を行い,会話レジームのベイズ推定を行 うアルゴリズムを提案する.続いて,3.4節では,頭部方向の推定方法を述べる.その後,3.5節, 及び,3.6節では実験結果を示し,提案したモデル,及び,推定法の有効性を検証する.さらに, 3.7節において議論を行い,3.8節では本章のまとめを示す.

# 3.2 会話レジームの設定

本節では,会話モデルに含まれる会話レジームの状態を具体的に設定し,会話レジームと会話 構造との関係,及び,会話レジームと参加者の行動との関係について仮説を提案する.3.1節で述 べたように,本論文の会話モデルにおいて,会話レジームは会話構造に相当するものであり,ま た,会話レジームの状態に依存して,参加者各人の行動や参加者間のインタラクションが確率的 に生成されるものと仮定される.また,発話権の交替に伴う参与役割の変化など,会話のダイナ ミクスは,会話レジームの状態遷移として表わされるものと考える.

このような会話レジームの設定を行うため,ここではまず,参加者の行動として,会話構造との 関連性が示唆されている視線行動に着目する.また,会話構造によって参加者行動が規定される という仮説より,同じ会話構造が継続する区間においては,同種の視線行動が観測されると考え, 参加者の視線方向(視線パターンと呼ぶ)の出現頻度を分析する.続いて,その結果から,出現 頻度が高く,かつ,長時間継続するような視線パターンに対応する会話構造を考察することで,会 話ジレームの状態の設定を試みる.ここで視線パターンとは,ある時刻における全参加者の視線 方向の集合を指す.本研究では,参加者がN人の場合,各参加者の視線方向は,他の参加者のう ちいずれか一人の顔に向けているか,あるいは,誰の顔からも視線を逸らしているかというN個 の離散的かつ排他的な状態のいずれかにあるものと考える<sup>2</sup>.従って,視線パターンは合計でN<sup>N</sup> 通り存在し得る.なお,本論文では,参加者は着席しており,会話中の参加者の増減や移動がな いことを前提とする.

なお,本研究では,話し手の話し掛けや,聞き手の傾聴に際して,視線を向けるという行為が 伴うことを仮定している.ただし,実際の会話では,参加者は対象への凝視と凝視回避を繰り返 すことから,同一の会話構造が継続する状況においても,一つの視線パターンが継続することは 期待できない.そこで,このような個々の人物の視線方向の変化に影響を受けずに,会話構造を 示唆するような安定した視線パターンの特徴を抽出するため,次のような分析を行った.

<sup>&</sup>lt;sup>2</sup>各参加者が注意を向ける先を限定するため、ノートや黒板などの道具は使用しないことを想定している.



図 3.1:4 人会話における視線パターンの出現頻度(会話データ G1-C1 より算出.3.5.1 節参照). 視線パターンは有向グラフにて表現(頂点=人物,辺=視線方向,出力辺がない頂点は,視線を逸 らしている人物を表す).同型グラフのクラスに対して相対頻度の大きい順番に1から19まで番 号を付与.各グラフの下部の3行の数値は,上から(チャンスレベルに対する相対頻度,この同 型グラフに分類される視線パターン数),全時刻中の(絶対)出現頻度[%],平均継続時間長[秒].

図 3.1 には,ある 4 人会話(3.5.1 節のデータ G1-C1)を対象として,人手により検出された視 線方向に基づいて,視線パターンの出現頻度を分析した結果を示す.図 3.1 において,視線パター ンは有向グラフとして表されている.ただし,グラフの頂点は各参加者,辺は視線方向を表わす. 図 3.1 に示した 19 種類のグラフは,人物の入れ替えに不変な視線パターンの構造に着目するため, 同型グラフの関係にある視線パターンを一つのグラフとしてまとめたものである.また,これら は,チャンスレベルに対する相対出現頻度の大きい順番に並べられている.ここで,ある二つの グラフが同型であるとは,一方のグラフを他方のグラフに変換するような頂点集合間の一対一写 像が存在する場合のことをいう[145].図 3.1 には,各グラフについて含まれる異なる視線パター ンの個数,全体の時間中で占める割合,一つの視線パターンが持続する平均時間長を併せて示す. 図 3.1 より,一人物への視線の集中,二人物間の相互凝視のような特徴的な構造を持つ視線パター ンが高い頻度で出現し,また,それらは比較的長時間継続する傾向があることがわかった.

また,視線パターンの時間変化の性質を探るために,これら特徴的な構造を含む視線パターン が継続する時間の長さを調べた.会話中においては,各参加者の視線方向が変化するに従って,視 線パターンも一つのパターンから別のパターンへと変化していくが,視線パターンが前述の特徴 的な構造を有する場合,その構造は時間的に維持される傾向があることが分かった.つまり,あ る人物への二人以上の視線の集中が生じる場合,その視線パターンの変化は,図3.1の1,2,5,6 のような同一人物への視線集中を示すパターン間の遷移に留まる傾向があり,同様に,二者間で の相互凝視の場合,2,4,5,8のように同じ二者間での相互凝視をもつパターン間の遷移が起こり やすいことが分かった.これら2種類の特徴をもつ構造が継続する時間の長さ(平均継続時間長) を計算したところ,前者については1.76[sec]となり,後者は1.33[sec]となった.これは,一つの 視線パターンが継続する平均的な時間(=0.52[sec])と比較しても長く,これら視線パターンの特 徴的な構造は時間的に安定しているといえる.本論文では,このような視線パターンの構造を作 り出すメカニズムと,会話の構造との間には強い関連性があるものと予想し,視線パターンの構 造に基づいて,以下のように一者集中,二者結合,分散と呼ぶ3つの会話レジームのクラスを仮 説的に設定する.

最初に「一者集中」(convergence)と呼ぶレジームは,図3.1の1,6のように,参加者の視線が 一人の人物に集中する視線パターンをもたらすような会話の構造に対応する.このレジームにおい ては,この最も多くの視線を受けている人物(中心人物と呼ぶ)が話し手となり,他の人物が受け 手になると想定される.そのため,中心人物が,他の参加者に向かって主に発話を行い,受け手は その発話を聞いており,受け手の発話は,相槌などに限定されると考えられる.また,このレジー ムにおける情報伝達のパターンは,中心人物から他の人物への一対多の一方向性であり,いわゆ る,モノローグと呼ばれる会話の型に対応すると考えられる.ここで,このレジームを $R_i^C$ と記す. ただし,*i*は中心人物を指す.N人会話の場合,N通りの一者集中レジームの状態  $\mathbf{R}^C = \{R_i^C\}_{i=1}^N$ が存在し得る.なお,話し手に対して全く視線を向けないような人物が存在する状況(考え事を している.聞き耳を立てている等)も考えられるが,そのような状況においても,他の参加者が 話し手へ注目している場合には,その会話の場を代表する会話構造として,一者集中レジームが 該当するものと考える.

次に「二者結合」(dyad-link)と呼ぶレジームは,図3.1の4,8のように,二人の人物が互いを見ている,つまり,相互凝視を含む視線パターンが生じるような会話の構造に対応する.このレジームにおいては,この相互凝視の関係にある二者間に限定された情報交換が行われ,この二者が話し手または受け手となり,他の参加者は傍参与者となるものと想定される.そのため,この二者は高い確率で発話を行い,他の参加者の発話は限定的であると仮定する.このレジームは,ダイアローグと呼ばれる会話の型に対応すると考えられる.一方,二者間の相互凝視は,発話交替の合図としても機能することより[103,146,97],発話交替時にも瞬間的にこのレジームが出現すると考えられる.このレジームを $R_{(i,j)}^{DL}$ と表記する.ただし,(i,j)は相互凝視の関係にある二者を表す.N人会話においては, $NC_2$ 通りの二者結合レジームの状態  $R^{DL} = \{R_{(i,j)}^{DL}|i=1,\dots,N; j=i+1,\dots,N\}$ が存在し得る.

最後に「分散」(divergence)と呼ぶレジームは,上の二つのレジームに該当しない視線パター ンが生じるような会話の構造に対応する.つまり,図3.1の3,9,13のように,各参加者が,別々 の人物を見ていたり,視線を逸らしているような場合に該当する.このレジームにおいては,グ ループとしての会話は生じていないと想定される.そのため,各参加者が発話を行う確率も低い と考えられる.また,このレジームは,会話の開始前や,会話中の話題の切れ目などにしばし現 れると考えられる.このレジームを *R*<sup>0</sup> と記す.

以上のように,提案する会話モデルにおいては,各レジームについて,対応する会話の構造,及び,出現する参加者行動(視線パターン,発話状態)の傾向が仮定される.これらの傾向は,確率 分布として会話モデルのパラメータとして表現され,推定の対象となる(3.3節参照).また,各 レジームが想定する性質を持つように予め事前確率分布がモデルパラメータに対して設定される (3.5.2節参照).

# 3.3 会話モデルと推定アルゴリズム

本節では,会話中の各時刻における会話レジームの状態を推定するために,マルコフ切替モデ ル[140,141]<sup>3</sup>に基づく会話モデルを提案し,続いて,このモデルに基づく推定のアルゴリズムを 提案する.

<sup>&</sup>lt;sup>3</sup>マルコフ切替モデルは, Interactive Multiple Model, Switching Dynamical System, Jump-Markov Model など とも呼ばれている.また,一般的には,下位の隠れ変数は連続状態をとるが,本論文では,離散状態である視線パター ンに拡張して考え,この名称を採用した.

#### 3.3.1 会話モデルの構造

図 3.2 は,提案する会話モデルの構造をグラフとして図示したものである.図 3.2 において,各 頂点は状態変数を表し,各辺は状態変数間の依存関係を表す.このモデルには,隠れ状態変数と して,会話レジームの状態変数の時系列  $S_{1:T} = \{S_1, S_2, \cdots, S_T\}$ ,及び,視線パターンの時系列  $X_{1:T} = \{X_1, X_2, \cdots, X_T\}$ が含まれる.ただし,本モデルでは,t = 1からt = Tまでの離散時 間区間をモデル化の対象とし,この区間内でN人物による会話が行われているものとする.ある 時刻  $t \in \{1, \cdots, T\}$ (時間ステップとも呼ぶ)のレジームの状態変数  $S_t$ は,3.2 節で定義された  $M(= N + NC_2 + 1)$ 個のレジームの内,いずれか一つの状態  $S_t = R \in \mathbf{R} = \mathbf{R}^C \cup \mathbf{R}^{DL} \cup R^0$ を とる.また,時刻 tの視線パターン  $X_t$ は,各人物  $i \in \{1, \cdots, N\}$ の視線方向  $X_{i,t}$ の集合  $X_t =$  $\{X_{1,t}, X_{2,t}, \cdots, X_{N,t}\}$ として定義される.ここで,人物 iが人物  $j(\neq i)$ の顔を見る場合を $X_{i,t} = j$ と表し,人物 i がいずれの人物の顔からも視線を逸らしている場合を $X_{i,t} = i$ と表す.

また,このモデルには,観測可能なデータとして,頭部方向の時系列 $H_{1:T} = \{H_t\}_{t=1}^T$ ,及び, 発話状態の時系列 $U_{1:T} = \{U_t\}_{t=1}^T$ が含まれる.時刻tの頭部方向 $H_t$ は,各参加者の頭部方向の 集合 $H_t = \{h_{i,t}\}_{i=1}^N$ として定義される.ただし, $h_{i,t}$ は,時刻tにおける人物iの頭部方向を表し, 図 3.6(a)のように世界座標軸Xと顔面正面の方向との成す方位角として計測できるものと考える<sup>4</sup>. また,時刻tの発話状態 $U_t$ は,各参加者の発話状態の集合 $U_t = \{u_{i,t}\}_{i=1}^N$ として定義される.ただ し, $u_{i,t}$ は,人物iの発話状態を表わし,時刻tにおいて人物iが発話をしている場合は $u_{i,t} = 1$ , 沈黙している場合には $u_{i,t} = 0$ の値をとるものとする.

このモデルの同時確率分布は,

$$p(\boldsymbol{X}_{1:T}, \boldsymbol{S}_{1:T}, \boldsymbol{Z}_{1:T}, \varphi) \propto f(\boldsymbol{Z}_{1:T} | \boldsymbol{X}_{1:T}, \boldsymbol{S}_{1:T}, \varphi) \cdot p(\boldsymbol{X}_{1:T} | \boldsymbol{S}_{1:T}, \varphi) \cdot p(\boldsymbol{S}_{1:T} | \varphi) \cdot p(\varphi)$$
(3.1)

のように観測データ  $Z_{1:T} = \{H_{1:T}, U_{1:T}\}$ についての尤度関数  $f(\cdot)$  と事前分布(右辺第2項以降) の積として表される.ここで, $\varphi$ はモデルパラメータの集合を表し(詳細は後述),以降,記述の 明瞭性のため必要がない限りこれを省略する.

式 (3.1) において,観測データ  $Z_{1:T}$  に基づくレジームの状態の系列  $S_{1:T}$  と視線パターンの系列  $X_{1:T}$  の尤度関数  $f(\cdot)$  は,時刻間の観測データの独立性,及び,レジーム状態が与えられたときの 頭部方向と発話状態の条件付独立性を仮定して,

$$f(\mathbf{Z}_{1:T}|\mathbf{X}_{1:T}, \mathbf{S}_{1:T}) = \prod_{t=1}^{T} f_{H}(H_{t}|\mathbf{X}_{t}) \cdot f_{U}(U_{t}|S_{t})$$
$$= \prod_{t=1}^{T} \prod_{i=1}^{N} f_{h}(h_{i,t}|X_{i,t}) \cdot f_{u}(u_{i,t}|S_{t})$$
(3.2)

<sup>&</sup>lt;sup>4</sup>各参加者の頭部は,床面に平行な面上に位置すると考え,各人の視線方向に応じて変化する頭部方向は,この面上 の成分のみで十分に表現できるものと考えた.なお,本論文で提案する会話モデルは,この1次元の頭部方向に限定さ れるものではなく,3次元頭部方向にも容易に適用可能である.



図 3.2: 提案した会話モデルの構造.

のように定義される.ただし,ここでは,視線パターンが与えられたときの各人の頭部方向の条件付独立性を仮定し,頭部方向の観測の尤度  $f_H(H_t|X_t)$  を各人の頭部方向の尤度  $f_h(h_{i,t}|X_{i,t})$  の 積として定義している.また,発話の尤度  $f_U(U_t|S_t)$  についても,与えられたレジーム状態に対し て各人の発話状態の条件付独立性を仮定して,各人の発話の尤度  $f_u(u_{i,t}|S_t)$  の積として定義して いる.本論文においては,各人の頭部方向の尤度分布は,眼球運動の自由度によって生じる視線 方向に対する頭部方向の不確定性を反映するために,ガウス分布を用いて,

$$f_h(h_i|X_i = j) = \frac{1}{\sqrt{2\pi\sigma_{i,j}^2}} \exp\left[-\frac{(\mu_{i,j} - h_i)^2}{2\sigma_{i,j}^2}\right]$$
(3.3)

のように定義される.ここで  $\mu_{i,j}$ ,  $\sigma_{i,j}^2$  は, 人物 *i* が方向 *j* を見るときの頭部方向の尤度分布の 平均と分散を表すパラメータである.これらパラメータは会話中,参加者の座席移動はないと の前提により,時間的に不変であると仮定する.また,各人の発話はレジームの状態に依存した ベルヌーイ過程に従って生成されると仮定し,式(3.2)中の各人物の発話の尤度  $f_u(u_{i,t}|S_t)$ は,  $f_u(u_{i,t} = 1|S_t = R) = \eta_{R,i}, f_u(u_{i,t} = 0|S_t = R) = 1 - \eta_{R,i}$ のように定義される.ただし,パラ メータ  $\eta_{R,i}$ は,レジーム  $S_t = R$ において人物 *i* が発話を行う確率を表わす.

式 (3.1) 中の右辺第 2 項  $p(X_{1:T}|S_{1:T}, \varphi)$ は,レジームの状態に依存して視線パターンが生成される条件付確率分布を表しており,与えられたレジーム状態のもと各人の視線方向の条件付独立性を仮定し,

$$p(\mathbf{X}_{1:T}|\mathbf{S}_{1:T}) = p(\mathbf{X}_1|S_1) \cdot \prod_{t=2}^T p(\mathbf{X}_t|\mathbf{X}_{t-1}, S_t, S_{t-1})$$
  
= 
$$\prod_{i=1}^N \left[ g(X_{i,1}|S_1) \cdot \prod_{t=2}^T p(X_{i,t}|X_{i,t-1}, S_t, S_{t-1}) \right]$$

$$\propto \prod_{R \in \boldsymbol{R}} \prod_{i=1}^{N} \prod_{j=1}^{N} \left[ \theta_{R,i,0,j}^{m_{R,i,0,j}} \times \prod_{k=1}^{N} \theta_{R,i,k,j}^{m_{R,i,k,j}} \right]$$
(3.4)

のように定義する.式 (3.4) において,  $p(X_{i,t}|X_{i,t-1}, S_t, S_{t-1})$  は人物 i の視線方向が前時刻 t-1 において  $X_{i,t-1}$  であった場合,現時刻 t における視線方向が  $X_{i,t}$  となる確率分布を表わす.本論 文では,これを

$$p(X_{i,t}|X_{i,t-1}, S_t, S_{t-1}) \propto g(X_{i,t}|S_t) \cdot w(X_{i,t}|X_{i,t-1}, S_{t-1})$$
(3.5)

のように生成重み  $g(X_{i,t} = j | S_t = R) = \theta_{R,i,0,j}$  と遷移重み  $w(X_{i,t} = j | X_{i,t-1} = k, S_{t-1} = R) = \theta_{R,i,k,j}$  の積に比例するものとして定義する.ここで,生成重み  $\theta_{R,i,0,j}$ は,あるレジーム R において,人物 i が方向 j を見る傾向の強さを表わし,遷移重み  $\theta_{R,i,k,j}$ は,人物 i がその視線を方向 k から j に変化させる傾向の強さを表わすパラメータである.ただし, $\sum_{j=1}^{N} \theta_{R,i,k,j} = 1, \forall k \in \{0, 1, \dots, N\}$ とする.また,式 (3.4) 中の  $m_{R,i,k,j}$  ( $k \neq 0$ )は,レジーム状態 R において,人物 i が視線を方向 k から j に変化させた回数を表わし, $m_{R,i,0,j}$ は,レジーム状態が R の場合に,人物 i が方向 j を見た時間ステップの総数を表わす.以上の視線パターンに関するパラメータをまとめて, $\Theta = \{\theta_{R,i,k} | R \in \mathbf{R}; i = 1, \dots, N; k = 0, \dots, N\}$ , $\theta_{R,i,k} = \{\theta_{R,i,k,j}\}_{j=1}^{N}$ と記す.

また,式 (3.1) 中のレジーム系列の事前分布  $p(S_{1:T}|\varphi)$ は,レジームのダイナミクスが1次のマルコフ過程であると仮定することにより,

$$p(\mathbf{S}_{1:T}) = p(S_1) \cdot \prod_{t=2}^{T} p(S_t | S_{t-1}) = \prod_{R \in \mathbf{R}} \left[ \pi_{0,R}^{\delta_R(S_1)} \times \prod_{R' \in \mathbf{R}} \pi_{R,R'}^{n_{R,R'}} \right]$$
(3.6)

のように定義することができる.式 (3.6) において,レジーム状態の初期確率は, $p(S_1 = R) = \pi_{0,R}, R \in \mathbf{R}$ として表わされる.また,時刻t - 1からtにおいて,レジーム状態がRからR'に変化する遷移確率は $p(S_t = R'|S_{t-1} = R) = \pi_{R,R'}$ として表わされる.ただし, $\sum_{R' \in \mathbf{R}} \pi_{R,R'} = 1, \forall R \in 0 \cup \mathbf{R}$ とする.式(3.6) において, $\delta_R(S)$ は,S = Rのときに1をとり,その他の場合0をとる関数である.また, $n_{R,R'}$ は,レジームがRからR'に変化する回数を表す.これらのレジームに関するモデルパラメータをまとめて, $\Pi = \pi_0 \cup \{\pi_R | R \in \mathbf{R}\}$ ,  $\pi_R = \{\pi_{R,R'} | R' \in \mathbf{R}\}$ のように記す.

以上で説明したモデルパラメータをまとめて,  $\varphi = \{\Pi, \Theta, \{\mu_{i,j}\}_{i,j}, \{\sigma_{i,j}^2\}_{i,j}, \{\eta_{R,i}\}_{R,i}\}$ と表わ す.本論文では,これらのパラメータは時間的に変化しないものと仮定する.また,式(3.1)中の パラメータの事前分布  $p(\varphi)$ は,それぞれのパラメータの独立性を仮定し,各々のパラメータの事 前分布の積として定義する(3.3.2節参照).

### 3.3.2 ギブスサンプリングによるベイズ推定

前節において定義された会話モデルに基づき,観測データ $Z_{1:T}$ が与えられたときの,レジームの状態系列 $S_{1:T}$ ,視線パターンの系列 $X_{1:T}$ ,及び,モデルパラメータ $\varphi$ の推定を行う.本論文で

はベイズ流のアプローチ [147, 148, 149] を採用し,全ての未知変数についての同時事後確率分布  $p(S_{1:T}, X_{1:T}, \varphi | Z_{1:T})$ の推定を行う.しかしながら,この事後分布を厳密に計算することは,モデ ルの複雑性の観点から困難であるため,本論文では,ギブスサンプリング [142, 140, 150, 141, 143] と呼ばれる一種のマルコフ連鎖モンテカルロ法を用いる.ギブスサンプリングとは,各々の未知 変数について全条件付事後分布からのランダムサンプリングを繰り返し行い,その過程で生成さ れるサンプル値の系列を用いて同時事後分布を近似的に求める手法である.これまで,ギブスサ ンプリングを隠れマルコフモデルのベイズ推定へ適用する方法が提案されており [150, 141],本論 文ではそれにならい以下のように推定を行う.

#### 事前分布

まず,モデルに関する事前知識を,モデルパラメータの事前分布として設定し,各々のパラメー タについて,それぞれの事前分布からのサンプリングにより初期値を設定する.本論文では,そ の事前分布として自然共役事前分布を採用する.その理由としては,i)事前分布と事後分布とが同 じ分布族に属するため,事後分布の導出が容易である,ii)多様な分布が表現できる,iii)ハイパー パラメータの解釈が容易である,といった点 [147,148,149] があげられる.

レジーム状態の初期確率  $\pi_0$ ,及び,遷移確率  $\pi_R$ の事前分布としては,隠れマルコフモデルの推定において一般的に用いられているディリクレ分布  $\mathcal{D}(\pi_R|\alpha_R)$ をそれぞれ採用する.ここで,ディリクレ分布は, $\mathcal{D}(\pi_R|\alpha_R) = c \cdot \prod_{R' \in \mathbf{R}} \pi_{R,R'}^{\alpha_{R,R'}-1}$ のように定義される.ただし, $\alpha_R = \{\alpha_{R,R'}\}_{R' \in \mathbf{R}}$ はディリクレ分布のパラメータであり,条件  $\alpha_{R,R'} > 0$ を満たす.また,cは,正規化のための係数を表す.また,視線パターンの生成重み/遷移重み  $\theta_{R,i,k}$ の事前分布についても,それぞれ独立なディリクレ分布  $\mathcal{D}(\theta_{R,i,k}|\beta_{R,i,k})$ を用いる.ただし, $\beta_{R,i,k} = \{\beta_{R,i,k,j}\}_{j=1}^{N}$ はそのパラメータを表す.また,発話確率  $\eta_{R,i}$ は,ベルヌーイ過程の仮定により,その事前分布として知られている[147,148,149]ベータ分布  $\mathcal{B}e(\gamma_{R,i,0},\gamma_{R,i,1})$ を用いる.ただし, $\gamma_{R,i,0},\gamma_{R,i,1}$ はベータ分布のパラメータを表わす.

頭部方向の尤度分布のパラメータの事前分布は,一般的なガウス分布のベイズ推定[147,148,149] にならい,平均値  $\mu_{i,j}$ についてはガウス分布  $N(\phi_{i,j},\xi_{i,j}^2)$  を仮定し,分散  $\sigma_{i,j}^2$ については逆カイニ 乗分布  $\chi^{-2}(\nu_{i,j},\lambda_{i,j})$  を採用する.ここで, $\phi_{i,j}$ , $\xi_{i,j}^2$ は,ガウス分布の平均と分散をそれぞれ表し,  $\nu_{i,j}$  と  $\lambda_{i,j}$ は,逆カイ二乗分布の自由度と尺度パラメータをそれぞれ表す.

これらのモデルパラメータの初期値は,それぞれの事前分布からの乱数発生により設定される. また,レジームの状態系列,及び,視線パターンの系列の初期値は,各モデルパラメータの初期 値を用いて,式(3.6),式(3.4)で表わされるモデル構造に従い生成される. 全条件付事後分布

ギブスサンプリングでは,各未知変数について,各々の全条件付事後分布からのサンプリングを 行い,逐次的に各変数の値を新しいサンプル値に置き換えるという処理を反復的に実行する.各 変数についての全条件付事後分布は,その変数以外の全ての変数の値が与えられたという条件の もとでの,その変数についての事後分布であり,ベイズ則に基づき各変数に関連した事前分布と 尤度の積から以下のように導出できる.

まず,レジーム状態の遷移確率  $\pi_R$ の全条件付事後分布は,式 (3.1)に表わされるモデルの構造を用いて,

$$p(\boldsymbol{\pi}_{R}|\boldsymbol{S}_{1:T}, \boldsymbol{X}_{1:T}, \boldsymbol{\varphi} \setminus \boldsymbol{\pi}_{R}, \boldsymbol{Z}_{1:T}) \propto p(\boldsymbol{S}_{1:T}|\boldsymbol{\varphi}) \cdot p(\boldsymbol{\varphi})$$

$$\propto \prod_{R' \in \boldsymbol{R}} (\pi_{R,R'})^{n_{R,R'}} \cdot \mathcal{D}(\boldsymbol{\pi}_{R}|\boldsymbol{\alpha}_{R})$$

$$\propto \prod_{R' \in \boldsymbol{R}} (\pi_{R,R'})^{n_{R,R'}+\alpha_{R,R'}-1}$$

$$\propto \mathcal{D}(\boldsymbol{\pi}_{R}|\boldsymbol{\alpha}_{R}^{*}) \qquad (3.7)$$

のように導出される.ただし, $\alpha_R^* = \{\alpha_{R,R'}^*\}_{R'\in \mathbf{R}}$ は, $\alpha_{R,R'}^* = \alpha_{R,R'} + n_{R,R'}$ により与えられる.また,レジーム状態の初期確率に関する全条件付事後分布としては, $\mathcal{D}(\pi_0|\alpha_0^*)$ , $\alpha_{0,R'}^* = \alpha_{0,R'} + n_{0,R'}$ を用いる.ただし, $n_{0,R'}$ は,レジーム状態がR'であった総時間ステップ数を表わす.

また,視線パターンの生成重み・遷移重み $\theta_{R,i,k}$ の全条件付事後分布は,レジームの場合と同様 に導出でき,

$$p(\boldsymbol{\theta}_{R,i,k}|\boldsymbol{S}_{1:T}, \boldsymbol{X}_{1:T}, \varphi \setminus \boldsymbol{\theta}_{R,i,k}, \boldsymbol{Z}_{1:T}) \propto \mathcal{D}(\boldsymbol{\theta}_{R,i,k}|\boldsymbol{\beta}_{R,i,k}^*)$$
(3.8)

のように定義できる.ただし,パラメータ $\beta_{R,i,k}^* = \{\beta_{R,i,k,j}^*\}_{j=1}^N$ は, $\beta_{R,i,k,j}^* = \beta_{R,i,k,j} + m_{R,i,k,j}$ として与えられる.

発話確率  $\eta_{R,i}$  の全条件付事後分布は,その事前分布であるベータ分布が,2変数の場合のディリクレ分布と等価であることから,レジームの場合と同様にして, $p(\eta_{R,i}|S_{1:T}, X_{1:T}, \varphi \setminus \eta_{R,i}, Z_{1:T}) \sim \mathcal{B}e(\gamma_{R,i,0}^*, \gamma_{R,i,1}^*)$ のように導出できる.ただし, $\gamma_{R,i,0}^* = \gamma_{R,i,0} + y_{R,i}, \gamma_{R,i,1}^* = \gamma_{R,i,1} + n_{0,R} - y_{R,i}$ である.ここで  $y_{R,i}$ は,レジーム状態がRのときに,人物iが発話をしていた時間ステップの総数を指す.

また,頭部方向の尤度分布の平均値  $\mu_{i,j}$  の全条件付事後分布は,一般的な分散既知の場合のガウス分布のベイズ推定 [147, 148, 149] に倣い,

$$p(\mu_{i,j}|\boldsymbol{S}_{1:T}, \boldsymbol{X}_{1:T}, \varphi \setminus \mu_{i,j}, \boldsymbol{Z}_{1:T}) = N\left(\phi_{i,j}^{*}, \xi_{i,j}^{*2}\right)$$
  

$$\phi_{i,j}^{*} = \xi_{i,j}^{*2} \cdot (l_{i,j} \cdot \overline{h}_{i,j}/c_{i,j}^{2} + \phi_{i,j}/\xi_{i,j}^{2})$$
  

$$\xi_{i,j}^{*2} = (l_{i,j}/c_{i,j}^{2} + 1/\xi_{i,j}^{2})^{-1}$$
(3.9)

のように平均  $\phi_{i,j}^*$ , 分散  $\xi_{i,j}^{*2}$  をもつガウス分布として与えられる.ただし,  $l_{i,j}$  は, 人物 i が方向 j を見た総時間ステップ数を表わし,  $\overline{h}_{i,j}$ ,  $c_{i,j}^2$  は, そのときの頭部方向の標本平均, 標本分散をそ れぞれ表わす.ただし, この平均値  $\mu_{i,j}$ のサンプリングにおいては,参加者の位置関係が既知で あるという仮定のもと,ある人物から他の人物を見るときの頭部方向について制約条件を導入す る.例えば,図 3.6(a) のように参加者が位置する場合,人物 1 の頭部方向の尤度分布について,  $\mu_{1,2} > \mu_{1,3} > \mu_{1,4}$ を満たすような平均値の組が得られるまで繰り返しサンプリングを行う.また, 頭部方向尤度分布の分散  $\sigma_{i,j}^2$ の全条件付事後分布は,一般的な平均既知の場合のガウス分布のベ イズ推定 [147, 148, 149] に倣い,

$$p(\sigma_{i,j}^2 | \boldsymbol{S}_{1:T}, \boldsymbol{X}_{1:T}, \varphi \setminus \sigma_{i,j}^2, \boldsymbol{Z}_{1:T}) = \chi^{-2}(\nu_{i,j}^*, \lambda_{i,j}^*)$$
$$\nu_{i,j}^* = \nu_{i,j} + l_{i,j}$$
$$\lambda_{i,j}^* = \lambda_{i,j} + l_{i,j} \cdot c_{i,j}^2$$
(3.10)

のように,自由度 $\nu_{i,j}^*$ ,尺度パラメータ $\lambda_{i,j}^*$ の逆カイ二乗分布として与えられる.

さらに,各時刻 $t \in \{1, \dots, T\}$ におけるレジーム状態 $S_t$ の全条件付事後分布は,式(3.1),式(3.2)から,

$$p(S_t | \boldsymbol{S}_{1:T} \setminus S_t, \boldsymbol{X}_{1:T}, \varphi, \boldsymbol{Z}_{1:T}) \propto p(S_t | S_{t-1}) \cdot p(S_{t+1} | S_t)$$
$$\cdot p(\boldsymbol{X}_t | \boldsymbol{X}_{t-1}, S_t, S_{t-1}) \cdot p(\boldsymbol{X}_{t+1} | \boldsymbol{X}_t, S_{t+1}, S_t) \cdot f_U(U_t | S_t)$$
(3.11)

のように求めることができる.また,各時刻 t における視線パターン Xt の全条件付事後分布は,

$$p(\boldsymbol{X}_t | \boldsymbol{S}_{1:T}, \boldsymbol{X}_{1:T} \setminus \boldsymbol{X}_t, \varphi, \boldsymbol{Z}_{1:T})$$

$$\propto p(\boldsymbol{X}_t | \boldsymbol{X}_{t-1}, S_t, S_{t-1}) \cdot p(\boldsymbol{X}_{t+1} | \boldsymbol{X}_t, S_{t+1}, S_t) \cdot f_H(H_t | \boldsymbol{X}_t)$$
(3.12)

として与えられる.

#### 事後分布からの統計量の計算

全ての未知変数について一通りサンプリングと値の更新を行う処理を一単位とし,それを  $\mathcal{N}$ 回,繰り返し実行する.反復終了後,反復回  $q = \mathcal{N}'$ から  $\mathcal{N}$ において得られたサンプル集合  $\{S_{1:T}^{(q)}, X_{1:T}^{(q)}, \varphi^{(q)}\}_{q=\mathcal{N}'}^{\mathcal{N}}$ から未知変数についての統計量が計算される.ただし, $\mathcal{N}'$ はサンプル値 の系列が十分に収束したとみなせる反復回とする.また,上付き添え字 (q)は,q回目の反復回で 得られた各変数の値を表す.レジーム系列,及び,視線パターン系列については,最大事後確率 推定値を  $\hat{S}_t = \arg \max_{R \in \mathbf{R}} \sum_{q=\mathcal{N}'}^{\mathcal{N}} \delta_R(S_t^{(q)})$ のように計算する.他の変数については,推定値を  $\hat{\mu} = (\mathcal{N} - \mathcal{N}' + 1)^{-1} \sum_{q=\mathcal{N}'}^{\mathcal{N}} \mu^{(q)}$ のように計算する.
## 3.4 頭部方向の推定方法と会話構造推定への適用

本節では,まず,本研究の会話シーン分析における頭部方向(頭部姿勢,顔向きとも呼ぶ)の追跡・推定に対する要求条件について述べた後,それに照らし合わせて従来研究を概観し,要求条件を満たす方法として疎テンプレート Condensation 追跡法 (Sparse Template Condensation Tracker; 以後,STC法と呼ぶ)を選択する(3.4.2節).続いて,STC法の概要を説明し(3.4.3節),会話構造推定に適用する方法を述べる(3.4.4節).

### 3.4.1 会話シーン分析における要求条件

複数人物の対面会話シーンを対象として,頭部方向推定を行う場合,以下の条件・状況に対処 することが求められる.

- 既存システム等の映像を流用するため,カメラパラメータなどの撮影環境に関する詳細情報が入手困難である.
- ステレオカメラ等,特殊な撮影手段は利用できず,一人物につき最大1台のカメラの利用を 前提とする.
- 参加者の顔について事前にモデルを獲得することが困難である.
- 不特定多数の人物に対して適用可能であること.
- 顔向きは正面顔から横顔まで多様に変化する (図 3.14(b)).
- 会話に伴う口の動きや表情の変化が大きい (図 3.14(c)).
- 顔を手で覆うなどオクルージョンが発生する (図 3.14(d)).

従来,この種の技術の主たるターゲットとして想定されていた HCI (Human-Computer Interaction) とは異なり,対面会話シーンの分析では,人間の自然な動作を対象とし,それを妨げないように 撮影を行う必要があるため,より要求条件は厳しいものとなる.ただし,本論文ではオフライン 処理を対象とする.

## 3.4.2 頭部方向推定の従来研究

従来,画像から人物の頭部方向を推定する代表的なアプローチとして,特徴点ベースの手法[39] と,アピアランスベースの手法[40]が知られている.特徴点ベースの手法では,目や鼻孔などの 顔部品の特徴点の位置関係に基づいて,頭部方向の推定を行う.そのため,姿勢変動やオクルー ジョン(隠れ,遮蔽とも呼ぶ)などにより特徴点を安定に抽出できない状況には適用が困難である.



図 3.3: 疎テンプレートの例 (グループ G1 の人物 3 (3.5.1 節参照)). 領域,特徴点を示す. (a) 基 本テンプレート T<sub>0</sub>, (b) 部分テンプレート T<sub>1</sub>, (c) ゼロ交差境界と境界ダイポール (線分で表す).



図 3.4: 疎テンプレート設定の例 (グループG1), (a)P1, (b)P2, (c)P4.設定時に与えた頭部姿勢 を併記 (単位=度).人物 i を Pi と記す.



図 3.5: STC 法の動作の様子 . (a) 抽出された頭部領域 (灰色で示す),及び,頭部範囲を示す円 (黒 い線), (b) 追跡中のテンプレート (白枠);黒点の集合:粒子の分布,大きい円:頭部中心,小さ い円:テンプレート表面の中心.

また,アピアランスベースの手法は,事前に姿勢既知の条件で撮影された大量の顔画像データを 用いた学習のプロセスを要し,また,遮蔽や顔領域の切り出しの精度に敏感であるという欠点を 有する.また,Active Appearance Model (AAM)を用いるアプローチも近年,着目されている [38].AAMは,メッシュ状の構造で顔形状を記述するモデルであり,顔面の細かな形状変化まで 表現できる点が利点であるが,その構築には,メッシュの頂点を手動で与えるなど煩雑なプロセ スが必要である.また,人物非依存を狙った Generic AAMは,一人物に特化した AAM と比較し て,その性能が大きく劣ることが指摘されている[41].そのため,これら従来法は,本研究には向 かないと考えられる.

一方,近年,疎テンプレート Condesantion 追跡法と呼ばれる手法が提案されている[144].こ の手法は,汎用的な物体追跡法であるが,人物の顔領域をテンプレートとして設定し,追跡を行 うことで,人物頭部の方向を推定することが可能である.テンプレートは,一枚の画像のみから 得られるため,事前の学習プロセスが不要である.また,ロバスト照合,Condensation による多 重仮説の生成・検証,複数テンプレートの切替などにより,遮蔽や表情変化に対する高いロバス ト性を実現している.これらの性質から STC 法は 3.4.1 節で述べた状況にも適応可能であると考 え,本研究ではこれを採用する.

## 3.4.3 疎テンプレート Condensation 追跡法 (STC 法)

疎テンプレートは,画像上の方形領域中の離散的な画素の集合から構成されるテンプレートであり,図 3.3(a)のように顔部品を含むように設定される.STC法では,単一の画像から切り出された複数のテンプレートを切り替えて追跡を行うことで,部分的な遮蔽に対するロバスト性を高める機構が導入されている.ここではテンプレートの集合 *T* は,顔面全体をカバーする基本テンプレート *T*<sub>0</sub> と, *N*<sub>T</sub> 個の部分テンプレート *T*<sub>n</sub> (*n* = 1,...,*N*<sub>T</sub>; *N*<sub>T</sub> ≥ 0)の集合 *T* = {*T*<sub>n</sub>}<sup>*N*<sub>T</sub></sup><sub>*n*=0</sub> から構成される.部分テンプレートは,基本テンプレートの部分領域からなり,顔の一部が遮蔽された場合を想定し,残る可視領域に対応するものとして設定される.各テンプレート *T*<sub>n</sub>  $\in$  *T* は, テンプレート画像 *J* 上での各特徴点の位置 *p*<sub>n,m</sub> = [*X*<sub>n,m</sub>,*Y*<sub>n,m</sub>]<sup>T</sup> とその輝度値 *J*(*p*<sub>n,m</sub>)の組の集合 *T*<sub>n</sub> = {(*p*<sub>n,m</sub>, *J*(*p*<sub>n,m</sub>))}<sup>*N*<sub>T</sub>=1</sup> から構成される (*N*<sub>T</sub>n は特徴点数).ただし,*T*<sub>n</sub>  $\subset$  *T*(*n* > 0)である.特徴点は,画像の輝度分布の局所的な極大・極小点,及び,図 3.3(c)に示すようなゼロ交差境界に跨る境界ダイポールの両端点から選択される(詳細は [144]).特徴点数は,テンプレート領域内の全画素数の 1.5%程度と少なく,通常のテンプレートより格段に高速な照合が可能である.

STC 法では,各時刻においてテンプレートの状態を推定する問題を,Condensationの枠組み で捉え,観測情報に基づいてテンプレートの状態の確率分布を更新するステップと,その次時刻 での分布を予測するという2つのステップを交互に実行する.テンプレートの状態の確率分布は, パーティクル (粒子) と呼ばれるサンプル集合  $\{s_t^{(l)}, \omega_t^{(l)}\}_{l=1}^{N_s}$ によって近似的に表現される.ここで  $\mathcal{N}_s$ は粒子の個数を表し, $s_t^{(l)}$ ,及び, $\omega_t^{(l)}$ は,それぞれ時刻tにおけるl番目の粒子の状態,及び,重みを表す.各粒子 $s_t^{(l)}$ の状態は,7次元ベクトル $s_t^{(l)} = [x_t^{(l)}, y_t^{(l)}, r_{x,t}^{(l)}, r_{y,t}^{(l)}, r_{z,t}^{(l)}, s_t^{(l)}, n_t^{(l)}]$ として定義される.ここで, $(x_t^{(l)}, y_t^{(l)})$ は,入力画像上でのテンプレートの位置を表す.また, $r_{x,t}^{(l)}, r_{y,t}^{(l)}, r_{z,t}^{(l)}$ は,それぞれ頷き,首振り,傾げに対応する頭部姿勢(回転角)を表し,カメラに正対する顔向きを原点とする.さらに, $s_t^{(l)}$ はスケール, $n_t^{(l)}$ は,テンプレートの番号をそれぞれ表す.

更新のステップでは,各粒子の状態 $s_t^{(l)}$ に対応するテンプレート $\mathcal{T}_{n'}(n' = n_t^{(l)}$ とおく)と入力画像Iとの照合が行われ,その照合誤差 $\varepsilon_t^{(l)}$ に基づいて,粒子の重み $\omega_t^{(l)}$ が更新される.照合誤差 $\varepsilon_t^{(l)}$ は,テンプレート $\mathcal{T}_{n'}$ 内の各特徴点 $p_{n',m}(m = 1, \cdots, \mathcal{N}_{Tn'})$ の輝度値と,その特徴点が入力画像I上に投影される点 $q_{n',m}$ の輝度値との相対残差 $\kappa(J(p_{n',m}), I(q_{n',m}))$ に基づき計算される.ここで, $q_{n',m}$ は,カメラモデルとして弱透視投影を仮定することで, $q_{n',m} = s_t^{(l)} \cdot \mathbf{R} \, \mathbf{p}_{n',m} + [x_t^{(l)}, y_t^{(l)}]^T$ のように定義される.ただし, $\mathbf{R}$ は, $r_{x,t}^{(l)}, r_{y,t}^{(l)}, r_{z,t}^{(l)}$ によって定まる回転行列を表す.相対残差 $\kappa(J,I)$ ,及び,照合誤差 $\varepsilon_t^{(l)}$ は,

$$\begin{split} \kappa(J,I) &= \frac{\alpha \cdot I - J}{J}, \\ \varepsilon_t^{(l)} &= \frac{1}{\mathcal{N}_{Tn'}} \sum_{m=1}^{\mathcal{N}_{Tn'}} \rho\left(\kappa(J(\boldsymbol{p}_{n',m}), I(\boldsymbol{q}_{n',m}))\right) \end{split}$$

のように計算される.ここで $\rho(\cdot)$ は, $\rho(\kappa) = \kappa^2/(\kappa^2 + 1)$ と定義される一種のロバスト関数であ リ,これにより外れ値の影響を軽減した照合誤差が計算できる.また, $\alpha$ は,画像輝度の変化を吸 収するための係数であり,照合誤差を最小とする値が選択される(詳細は [144]).各粒子の重みは 照合誤差から $\omega_t^{(l)} \propto 1/\varepsilon_t^{(l)}, \sum_l \omega_t^{(l)} = 1$ のように計算される.その後,全粒子の中から,重みの値 が大きい順番に上位 $\mathcal{N}_U(=10)$ 個が選択され,この選択された粒子の重み付き平均として,各時刻 のテンプレート状態の推定値 $\overline{s}_t^{(l)}$ が計算される.また,推定された頭部姿勢の内,首振り角に対 応する $\overline{r}_{y,t}$ が水平方位角 $h_{i,t}$ として会話構造の推定のために出力される.また,選択された粒子 の集合に基づいてリサンプリングが行われる.

次の予測のステップにおいては,現時刻 t の各粒子の状態  $s_t^{(l)}$  から,次時刻 t + 1 における状態  $\hat{s}_{t+1}^{(l)}$  が予測される.状態のうち位置と姿勢については,各々独立なガウス分布 (平均=0) に従う システムノイズが加えられる.また,スケールの値は,確率  $\xi_1(=0.955)$  にて持続し  $(s_{t+1}^{(l)} = s_t^{(l)})$ , 残りの場合,均等な確率で増大または減少する  $(s_{t+1}^{(l)} = s_t^{(l)} \times (1.02)^{+1}$  or  $^{-1}$ ) という規則を適用す る.テンプレートに関しては,確率  $\xi_2(=0.3)$  にて,基本テンプレートが選択され  $(n_{t+1}^{(l)} = 0)$ ,そ れ以外の場合には,確率  $\xi_3(=0.4)$  にて持続し  $(n_{t+1}^{(l)} = n_t^{(l)})$ ,さらに,それ以外の場合に他のテン プレートに均等の確率で遷移するような規則を適用する.この予測規則は,フィッティングの安定 性の観点から基本テンプレートの適用を常に試行する一方,現時刻の遮蔽の状態が次時刻でも継 続するという予想から,同じテンプレートの継続使用を試みるという考えに基づいている.

以上において, 文献 [144] の STC 法と比較して, 本研究の STC 法は, テンプレートの照合誤差

の計算が簡素化され,また,テンプレートの予測規則が新たに導入された点が主に異なる.

### 3.4.4 STC 法の会話構造推定への適用

STC 法を会話構造推定へ適用するために試みた本論文独自の工夫について述べる.

テンプレートの設定方法 各人についてカメラに正対した平静時の顔が写っている画像フレーム を手動で探し、その顔面領域をマウスで選択することテンプレートを設定する.なお、カメラに 正対することのない人物については、図3.4(a)(c)のように、傾いた状態のテンプレートを目視で 設定する.また、図3.14(d)のように手で口を覆うような動作が頻発している人物については、図 3.3(b)のように口を除く領域を部分テンプレートとして設定する.

頭部回転角の符号不定性の解消 カメラモデルとして弱透視投影を採用している関係上, テンプレートの姿勢の回転角に符号の不定性の問題が存在する.これは回転角の符号を変えても画像上に投影される特徴点分布は同じになり,顔が左右どちらを向いているか判断できないという問題である.ここでは,以下の方法により解決を図る.まず,テンプレートに奥行きのオフセットを導入し,テンプレート面を顔面に一致させ,位置(x,y)を頭部の中心(頭の回転中心を画像上に投影した位置)に変更する.この場合,特徴点の座標は $p = [X,Y,Z]^T$ のように経験的に決めたオフセット成分Zを加えたものとする.次に,入力画像上で頭部領域の抽出を行い(図 3.5(a)),頭部中心(x,y)を中心とする円(頭の画像上での見かけの大きさに対応)が頭部領域内に包含されない場合,その度合いに応じて,対応する粒子の重みにペナルティを課す.具体的には,この円の円周が頭部領域内に含まれる長さの割合を粒子の重みに乗じることで実施する.なお,頭部領域の抽出は,単純にフレーム間差分の累積により行う.以上の処理により,正しい頭部回転角の符号をもつ粒子のみが生き残ることで,不定性の問題が解消される.図 3.5(b)に,追跡中のテンプレート,及び,粒子の分布を図示する.

追跡の初期化・再初期化 初期時刻において,テンプレート状態の各要素について,その想定され る範囲内の一様分布からの乱数生成により粒子集合の初期化を行う.また,追跡の失敗を,照合誤 差の最小値が閾値以上になった場合として検出し,その場合には,更新ステップ中の上位 N<sub>U</sub> 個 の粒子の選択のプロセスを省く.これにより次時刻で再び対象を捉えるべく,より広い範囲に渡っ てテンプレート照合を試みることができる.

未校正カメラへの対応本研究では,各参加者の画像を別々の未校正カメラにより撮影する.そのため,計測される各人の頭部の方位角は,それぞれのカメラの座標系に依存したものとなる.3.3節の提案法では,各人物に共通した世界座標系上での方位角を用いていたが,これは本質的な必要条件ではない.この方法では,各人物*i*が他者 $j(\neq i)$ に視線を向ける場合の頭部方向の尤度分布 $f(h_i|X_i = j) = N(\mu_{i,j}, \sigma_{i,j}^2)$ のパラメータ(平均 $\mu_{i,j}$ ,分散 $\sigma_{i,j}^2$ )を推定しており,これは人物毎の別個の座標系にもそのまま適用できる.ただし,この推定においては,これらパラメータの事



図 3.6: 会話シーンの概略.(a)参加者の配置(平面図),(b)参加者の全体ショット(グループG1).

前分布を設定する必要があるが,本稿では,STC法の計測データを利用して, $\mu_{i,j}$ の事前分布(これもまたガウス分布)の平均値 $\phi_{i,j}$ を設定するというアプローチを採る.具体的には,人物1(P1とも記す)を例にとると,この人物から見た他者の位置関係(P3はP2の右にいる等)の事前知識を用いて,頭部方向尤度分布の平均値の間に $\mu_{1,4} < \mu_{1,3} < \mu_{1,2}$ の関係(図 3.6(a)参照)を仮定し,これに基づき, $\phi_{1,4},\phi_{1,2}$ には,計測された頭部方位角の最小値,及び,最大値をそれぞれ与え,また, $\phi_{1,1},\phi_{1,3}$ には,平均値を割り当てることとした.なお,この設定は結果に対して敏感に影響しないことを実験により確認している.また,画像の水平軸が床面に平行であることは前提とする.

# 3.5 実験1 —磁気センサによる頭部方向計測に基づく推定—

提案した会話モデル,及び,会話構造推定法の有効性を検証するため,4人会話を対象とした実 験を行った.この実験は2部から構成されており,第1部(本節)と,第2部(3.6節)に分けて 記載する.まず,第1部の実験1では,頭部方向計測の方法として,磁気式センサを用いる.こ れは,画像のトラッキングによる方法と比較して,安定かつ正確な計測が可能である.そのため, 第1部では頭部方向が正確に計測された条件においての会話構造推定法の有効性を検証する.そ の後,第2部においては,頭部計測の方法として,3.4節において述べたトラッキングによる方法 を用いた結果を示し,本節の結果と比較して頭部トラッキングの有効性を検証する.

本節の以後の構成は以下のようになっている.まず,実験に使用したデータについて述べ,続いて,モデルのハイパーパラメータの設定について説明する.その後,会話レジームと人物行動の関係が適切にモデル化されていることを確認するために,視線方向の推定精度を評価する(評価1と呼ぶ).次に,推定された会話レジームが実際の会話の構造を反映したものであることを確認するために,発話の種別・方向性のラベルに基づいた評価法を提案し,他手法との比較を交え

表 3.1: 会話の議題.

グループ	会話を指す記号	議題
G1	G1-C1	恋愛と結婚は一緒か別か?
G1	G1-C2	安楽死は法的に認めるべきか?
G2	G2-C1	恋愛と結婚は一緒か別か?
G2	G2-C2	専業主婦に対して法的優遇措置をとるべきか?



図 3.7: 観測されたデータ(会話 G1-C1), (a) 頭部方向, (b) 発話区間(Pi は人物 i を示す).

て,推定されたレジームの評価を行う(評価2と呼ぶ).最後にこれらの結果をあわせて,提案した会話モデル,及び,会話構造推定法の有効性を確認する.

## 3.5.1 会話データの収集と準備

本実験では,参加者数4名によるグループ会話を対象とした.参加者は女性8名とし,4人ずつの2グループG1,G2について,それぞれ会話を収録した.各グループの年齢構成は,グループG1は,23歳~28歳(平均25.8歳),グループG2は,28歳~29歳(平均28.5歳)であった.参加者に対しては,与えられた議題に対して議論を行い,5分を目安にグループとして一つの結論を出すよう指示が与えれた.表3.1には,各グループに与えた議題を示す.なお,これらの議題の選択に際し,会話収録前にいくつかの議題に対してアンケートを行い,同一グループの参加者間で意見の異なる議題を選択した.

頭部方向の計測は,磁気式の6自由度センサ(POLHEMUS Fastrak<sup>TM</sup>)を採用し,各参加者の 頭部にヘアバンドによりセンサを装着し,30[Hz] にて計測を行った.図3.7(a)には,会話G1-C1 において計測された頭部の方向(水平方向の方位角)の時系列の一部(3600ステップ = 2[min]) を示す.また,音声データを,各参加者に装着したピンマイクによって収録し,各人物の発話区間 を,音声波形編集ソフトを用いて人手により抽出した.ここでは一つの発話区間を 300[ms] 以上 の無音区間に挟まれた区間として定義した.図3.7(b)には,図3.7(a)と同一の時間区間について 得られた発話区間の様子を示す.さらに,会話の様子を映像として記録するため,全参加者を捉 えた全体ショット(図3.6(b)),各参加者のバストショット(図3.11(a)),及び,各参加者の顔領 域を捉えたアップショットを撮影した.フレームレートは30[frame/sec] とした.これらのデータ は,単位時間ステップ1/30[sec] にあわせて同期された.分析に用いる会話データG1-C1,G1-C2, G2-C1,G2-C2の時間長は,それぞれ10000,9300,9100,10000[frame] とした(5.1分から5.6分に 相当).

3.5.2 ハイパーパラメータの設定

会話モデルの各パラメータの事前分布について,その分布の形状を規定するハイパーパラメー タの設定を行った.この設定は,会話モデルを特徴付ける上で重要であり,ここでは,次のよう な方針に従って経験的にハイパーパラメータの値を決めた.なお,全ての会話データに対して同 一の値を用いた.

以下では,事前分布としてディリクレ分布を用いるパラメータについては,その分布のパラメー タ  $\alpha_R$ (レジームの初期・遷移確率の場合)を, $\alpha_R = \{\alpha_{R,R'}\}_{R'\in \mathbf{R}} = \{\alpha'_R \cdot \alpha'_{R,R'}\}_{R'\in \mathbf{R}}$ のよう に,事前平均  $\alpha'_{R,R'} = E[\pi_{R,R'}]$ と,仮設的な事前分布の標本サイズ  $\alpha'_R$ の積に分解し,それぞれを 設定することとした.ただし, $\sum_{R'\in \mathbf{R}} \alpha'_{R,R'} = 1$ である.なお,視線パターンの生成重み/遷移 重み,及び,発話確率についても同様にパラメータを設定した.

レジーム

レジームの初期確率  $\pi_0$ の事前平均は,一者集中レジームの各々,及び,分散レジームに対し て均等な値  $\alpha'_{0,R} = 0.19, R \in R^0 \cup \mathbf{R}^C$ を与え,残りの二者結合レジームの各々についても均等な 値を設定した.また,レジーム間の遷移確率の事前平均については,分散レジームから一者集中 レジーム ( $R^0 \to R \in \mathbf{R}^C$ )については均等な値を設定し,分散レジームから二者結合レジーム ( $R^0 \to R \in \mathbf{R}^{DL}$ )への遷移はないものと仮定した.一者集中からの遷移, $R_i^C \to R_j^C$  ( $j \neq i$ ),  $R_i^C \to R^0$ ,については均等な値を設定した.また,一者集中  $R_i^C$ から二者結合への遷移は,一者 集中の中心人物 i が二者結合のペアに含まれるもの ( $R_i^C \to R_i^{DL}$  (i = k or i = j))のみ生じる ものと仮定した.二者結合  $R_{(i,j)}^{DL}$ からの遷移は, $R_{(i,j)}^{DL} \to R_i^C, R_j^O, R^0$ に限り均等な事前平均を与



図 3.8: ギブスサンプリングにおける頭部方向尤度分布の平均値  $\mu_{1,1}, \mu_{1,2}, \mu_{1,3}, \mu_{1,4}$ の遷移 (G1-C1の場合).

え,それ以外の二者結合からの遷移はないものと仮定した.また,仮設的な事前分布の標本サイズは  $\alpha'_R = 5000, \forall R \in \mathbf{R}$  とした.

視線パターンと発話

ー者集中レジーム  $R = R_i^C$  においては,中心人物 *i* の視線方向分布は一様であると仮定し,視線パターンの生成重みの事前平均  $\beta'_{R,i,0,j}$  は均等とした.一方,他の人物  $j(\neq i)$  は,高い確率で話し手である中心人物 *i* の方を向くと仮定した( $\beta'_{R,j,0,i} = 0.7$ ).なお,視線パターンの遷移重みの事前平均は,聞き手の視線が中心人物を向いていない場合,中心人物へと遷移しやすく,また,既に中心人物を見ている場合,見続ける傾向が強くなるように設定した.さらに,発話確率の事前平均は,中心人物 *i* が主に発話を行い( $\gamma'_{R,i,0} = 0.97$ ),その他の人物の発話は相槌などに限定されるよう設定した( $\gamma'_{R,i,0} = 0.03, j \neq i$ ).

二者結合レジーム  $R = R_{(i,j)}^{DL}$ においては,ペアを成す人物 i, j は高い確率で互いを見るものとの 仮定をおき,生成重みの事前平均を  $\beta'_{R,i,0,j} = \beta'_{R,j,0,i} = 0.95$ と設定した.一方,他の人物について は各視線方向に均等な事前平均を設定した.遷移重みについては,ペアを成す人物は,互いの方向を 見つづけるとし( $\beta'_{R,i,j,j} = \beta'_{R,j,i,i} = 1.0$ ),他の人物については,ランダムな方向に視線が変化する よう均等な事前平均を設定した.また,ペアをなす人物が主に発話を行い( $\gamma'_{R,i,0} = \gamma'_{R,j,0} = 0.7$ ), 他の人物の発話はほとんどないものと仮定した( $\gamma'_{R,k,0} = 0.03$ ).

分散レジーム  $R = R^0$  においては,各参加者  $i \in \{1, \dots, N\}$ の視線方向はランダムであると仮定し,各々の視線方向の生成重みの事前平均は均等とした( $\beta'_{R,i,0,j} = 1/N, j = 1, \dots, N$ ).また, 視線方向の変化もランダムであると仮定し,均等な事前平均を設定した.また,このレジームにおいては,発話は少ないものと仮定した( $\gamma'_{R,i,0} = 0.01$ ).

なお,仮設的な事前の標本サイズについては,視線パターンに関しては, $\beta'_{R,i,0} = 10000$ , $\beta'_{R,i,k} = 500 \ (k \neq 0)$ とし,発話確率については $\gamma'_{R,i} = 500$ とした.

group	conversation	Average	P1	P2	P3	P4
G1	G1-C1	71.1	80.9	65.7	71.5	66.4
G1	G1-C2	59.3	70.4	48.2	67.4	51.4
G2	G2-C1	72.4	57.4	83.8	78.2	70.1
G2	G2-C2	75.9	49.0	90.5	84.1	80.1

表 3.2: 視線方向の推定精度 [%].

頭部方向の尤度分布

頭部方向の尤度分布の平均値の事前分布  $p(\mu_{i,j})$  については,その平均値  $\phi_{i,j}$ を,図 3.6(a) のように,ある人物 *i* から他の人物 *j* を見込んだ方位角  $\Delta \phi_{i,j}$  として与えた( $i \neq j$ の場合).また,視線を逸らしている状態に関しては, $\phi_{i,j}$ は,他の人物に関する  $\phi_{i,j}$  ( $j \neq i$ )の平均値を与えた.また,その他の頭部方向に関するパラメータは,経験的に $\xi_{i,j}^2 = 0.2$ ,  $\nu_{i,j} = 1000$  ( $i \neq j$ ),  $\nu_{i,i} = 3000$ ,  $\lambda_{i,j} = 100$ のように与えた.

## 3.5.3 ギブスサンプリング

ギブスサンプリングの反復回数は,反復回に従ってパラメータの値がどのように遷移するかを示したグラフを作成し,それを目視することで,収束に至る反復回数を把握し, $\mathcal{N} = 700$ , $\mathcal{N}' = 500$ のように設定した.図 3.8 にはそのようなグラフの一例として,会話データ G1-C1 について得られた,頭部方向の尤度分布の平均値 { $\mu_{1,j}$ } $_{j=1}^4$ の遷移を示す.

## 3.5.4 評価 1: 視線方向の推定精度

提案した会話モデルによって,会話レジームと人物行動の関係が適切にモデル化されているこ とを確認するために,視線方向の正解ラベルを用いて,視線方向の推定精度を検証した.なお,視 線方向の正解ラベルは,撮影された映像を詳細に観察することにより人手で付与した.視線方向 の推定精度として,表3.2には,各会話データ,各参加者について,推定された視線方向と正解ラ ベルが一致した時間ステップ数の割合,及び,それらの全参加者についての平均値を示す.また, 図 3.9(a)には,推定結果の一例として,会話G1-C1の一部(最初の2分間)における各人の視線 方向の時系列を正解ラベルと併せて図示する.推定結果と正解ラベルが一致しない状況を全ての 会話データについて調べた結果,視線を逸らした状態(凝視回避と呼ぶ)を凝視として誤判定し たケースが,全ての誤りの内,約60%を占め,また,凝視の状態を凝視回避と誤ったケースが同 約13%を占め,推定誤りの多くは視線を逸らした状態に関連することがわかった.このような誤



図 3.9: 会話データ G1-C1 に対する推定結果: (a) 各人物 P1, P2, P3, P4 の視線方向  $\{X_{1,t}, X_{2,t}, X_{3,t}, X_{4,t}\}$  (実線:推定値,破線:正解ラベルの方向), (b) 推定されたレジーム;各 時刻 t において,一者集中レジーム  $S_t = R_i^C$  の場合,iの位置のみにバンド,二者結合レジーム  $R_{(i,j)}^{DL}$  の場合,i,jの位置にバンド,分散レジーム  $R^0$ の場合は空白により示す.



図 3.10: 会話データ G1-C1 より推定された頭部方向の尤度分布  $p(h_1|X_1 = i)$ (曲線),及び,頭 部方向のヒストグラム(折れ線).  $p(h_1|X_1 = i), i \neq 1$ :人物1が人物 iを見ている.  $p(h_1|X_1 = 1)$ : 人物1が視線を逸らしている.

りの要因を探るため,図3.10に,推定された頭部方向の尤度分布,及び,正解ラベルを用いて作 成された各視線方向毎の頭部方向のヒストグラムを示す.図3.10からは,凝視回避時の分布と凝 視時の分布が,尤度分布,ヒストグラムともに大幅にオーバーラップしていることがわかる.こ のことは,頭部方向を固定したまま,視線のみを動かすことができる,つまり,流し目を使う,と いう人間の性質が反映された結果であり,視線方向の推定誤りの要因になっていると考えられる.

また,比較のために,視線方向の正解ラベルを用いて,頭部方向尤度分布(ガウス分布)を求め,視線方向を最大事後確率推定により求めた結果,各会話データについての平均推定精度は,それぞれ,G1-C1:68.6%,G1-C2:65.4%,G2-C1:74.7%,G2-C2:71.1%となった.これらの精度は,床面に水平な頭部方向成分をデータとして用い,その尤度分布にガウス分布を採用する際の一種の限界を示唆するものと考えられる.表3.2記載の提案法の推定精度は,これらの推定精度より若干,劣る程度,あるいは,場合によっては上回ることがわかった.

本節の結果からは次のような考察が可能である.提案方法では,各人の相対的な位置関係に関 する知識は使用しているものの,頭部方向の尤度分布のパラメータは未知という条件のもと,そ のパラメータや視線方向の推定を行っている.この問題設定に対処するため,提案方法では会話 モデルを導入し,それにより会話構造に応じて出現しやすい視線パターンが想定できることから, 曖昧な頭部方向のデータからも精度良く視線方向が推定できることを期待している.ここで会話 モデルが適切ではない場合,各人物の頭部方向の尤度分布が精度良く求まらず,結果的に視線方 向の精度も低いものになると考えられ,また逆に,視線方向の推定精度が高いことは,会話モデ ルが適切なものであった証拠であると考えられる.本実験において,提案方法の推定精度が,上 記,比較対象の手法の精度に匹敵し得るものであったことから,会話レジームと参加者行動の関 係のモデル化という観点から会話モデルの妥当性が示唆されたといえる.

### 3.5.5 評価 2: 会話レジームについての評価

#### 定性的な側面

図 3.9(b) に,会話データ G1-C1 より推定されたレジームの時系列の一部を示す.また,図 3.11 には,この会話から特徴的な個所を取り出し,その会話の流れに沿った 3 時刻(t = 310, t = 485, t = 578)における各参加者の様子と発言(図 3.11(a)),及び,推定されたレジームと視線パターン(図 3.11(b))を示す.この個所において,最初(t = 310),人物4は,他の全ての参加者に対して意見を述べており,他の参加者は人物4の発話を聞いている.このときの会話レジームは,人物4への一者集中  $R_4^C$  と推定されており,実際の会話の構造に合致していると考えられる.次に(t = 485),人物2が人物4に対して頷きを伴う発話により強い反応を表出し,人物4も人物2に視線を向け,その反応を確認している.ここでは人物2と4の間で相互凝視が生じている.この個所での会話レジームは,この2人物間の二者結合  $R_{(2,4)}^{DL}$ と推定され,この状況を適切に反映したものになっていると考えられる.さらに(t = 578),人物2は発話を継続し,それに対して人物4は相槌を打った後,発話を中断し,人物2に発話権を譲渡した.また,同時に,人物3は,新しい話し手である人物2の発言に注意を向けるため,その視線を人物4から人物2へと移した.このような現象を反映して,この個所では会話レジームは人物2への一者集中  $R_2^C$ と推定された.このような観測により,推定されたレジームの系列は,実際の会話の流れを反映したものであることが確認された.

## 推定された会話レジームの評価法

次に,推定されたレジームの精度を定量的に評価するための方法について述べる.現在のところ,会話構造を定量的に評価する一般的な方法が存在しないため,本論文では独自の評価方法を考案した.この方法は,会話中の各発話区間について,人手によって付与されたラベルと推定されたレジームとの照合を行うことで実行される.発話区間のラベルは,発話の種別とその発話が誰に向けられたものか,あるいは,誰の発話を受けて発せられたものか,という方向性の情報を含む.このラベルは,具体的には表 3.3 に示すような「発話種別+方向性」の形式をとり,各発話区間について,その区間に該当する項目のラベルをすべて付与した.発話種別には,意見の表明 $\{a\}$ ,他者への問い掛け(質問 $\{r\}$ ,同意の要求 $\{R\}$ ),問い掛けに対する応答(質問に対する応答  $\{r\}$ ,要求された同意に対する応答 $\{R\}$ ),その他,他者の発話に対する反応(相槌 $\{n\}$ ,笑い $\{s\}$ ,驚き $\{!\}$ ),その他の発声 $\{e\}$ (溜息,唸り声など)が含まれる.加えてラベルには,聞き手が記号 h に続いて記述される.意見の表明は,一者集中レジームにおける話し手から受け手へ



図 3.11: 推定されたレジームの時間遷移の様子(会話データ C1-G1 中の  $t_1 = 310, t_2 = 485, t_3 = 578$ ). (a) 各参加者のバストショット, (b) 推定されたレジーム,及び,視線パターン(細い矢印: 推定された視線方向,幅の広い矢印:正解ラベル).

Label	Definition
$ad_1d_2,\cdots$	express opinion toward persons $d_1 d_2, \cdots$
$\mathbf{q}d_1d_2,\cdots$	open question toward persons $d_1d_2, \cdots$
$Qd_1d_2,\cdots$	tag question toward $d_1 d_2, \cdots$
$\mathbf{r}d$	response for question from person $d$
$\mathbf{R}d$	response for others' utterances (other than 'r')
$\mathbf{n}d$	back-channel utterance for person $d$
$\mathbf{s}d$	laugh caused by what person $d$ said or did
ed	expressions with sound such as hum or sigh
!d	exclamation caused by what person $d$ said or acted
$hd_1d_2,\cdots$	persons $d_1 d_2, \cdots$ listening to the utterance

表 3.3: 発話区間に付与されるラベルの定義.

の情報伝達に関連し,また,問い掛けやそれに対する応答は,二者結合レジームにおける二者間の双方向的な情報交換に関連する.また,分散レジームに関連したラベルとしては,方向性のラベルが無い,つまり,他者に向けられていない独り言や,聞き手のいない発話などがあげられる.

例えば, ラベル  $\{a234,h234,q2\}$  が人物1の発話区間について与えられた場合, 人物1は, 人物2,3,4に向かって自分の意見を表明しており, また, 人物2,3,4は人物1の発話に注意を傾けて聞いていることが表わされる.さらに, 人物1は, その発話区間の終端において, 人物2に対して問い掛け(質問)を発している.また, この発話区間に後続する人物2の発話区間にラベル  $\{r1,a1,h134\}$  が与えられた場合, 人物2は, 人物1の問い掛けに対して応答し, 自らの意見を人物1のみに向かって表明しており, その発話を人物1,3,4 が聞いているという状況が表わされる.このようなラベル付けによって, 各人の発話区間の間の因果関係を記述することができる.

次に,レジームの推定精度として,各時刻 t において,推定されたレジーム  $\hat{R}_t$  に対応する会話 構造が,発話ラベルによって示唆された現象中に含まれるかどうか判定し,含まれていると判定さ れた時間ステップの割合を,レジームの正答率として計算する<sup>5</sup>.そのため,まず,各時刻 t につ いて,その時刻を範囲に含む発話区間のラベル集合 L(t)を抽出する.ここでは,レジームの時間 変化のスケールが単位時間ステップ(=1/30[sec])よりも大きいことを考慮し,推定されたジレー ム状態  $\hat{R}_t$  と同じ推定値をもつ連続時間区間  $[t_0, t_1](t_0 \le t \le t_1, \hat{R}_t = \hat{R}_{t'}, \forall t' \in [t_0, t_1])$ との交差を

<sup>&</sup>lt;sup>5</sup>発話ラベルは,物理的な発話区間を単位としているため,一つの連続した発話区間内に,複数の異なる意図をもった発話が含まれる可能性がある.また,複数の人物がそれぞれ話者として同時に発話を試みるような状況も存在しうる. よって,一時刻に対応する発話ラベルには,複数の現象が含まれる場合がある.そのため,レジームとラベルとの一致 率という観点からではなく,「含まれるかどうか」という観点からの評価を行うこととした.

表 3.4: レジームの正答率 [%], (a) 提案手法 *M*, (b) 視線既知 *M<sub>GG</sub>*, (c) 視線のみ *M<sub>GO</sub>*, (d) 発 話のみ *M<sub>UO</sub>* (Conv.: 一者集中レジーム, DL.: 二者結合レジーム, Div.: 分散レジーム, Total: 全レジーム).

	(a)Our method $\mathcal{M}$			(b)Gaze Given $\mathcal{M}_{GG}$				
	Total	Conv.	DL.	Div.	Total	Conv.	DL.	Div.
G1-C1	81.8	85.0	77.7	53.2	78.1	85.8	70.1	68.6
G1-C2	92.1	94.9	85.9	69.0	81.0	91.3	76.7	42.9
G2-C1	91.4	95.7	72.6	100	92.6	93.5	91.3	88.5
G2-C2	96.3	98.8	83.5	100	96.0	97.6	91.3	88.5

(c)Gaze Only $\mathcal{M}_{GO}$				(d)Utterance Only $\mathcal{M}_{UO}$			
Total	Conv.	DL.	Div.	Total	Conv.	DL.	Div.
69.7	86.3	52.7	54.1	78.5	80.2	64.6	100
79.0	92.1	67.5	56.9	83.4	85.9	73.4	100
84.0	95.4	64.6	52.5	89.6	88.4	78.5	100
92.0	98.6	41.7	49.2	92.4	93.2	83.4	98.7

もつ発話区間の集合に含まれるすべてのラベルを抽出し,照合に用いるラベル集合 L(t) とした. 次に各時間ステップ毎にレジーム毎に異なる条件を用いて評価を行った.下記には,実際の会話の状況と照らし合わせてレジームが正解である条件をレジーム種別毎に記す.

- i) 一者集中 Â<sub>t</sub> = R<sub>i</sub><sup>C</sup> が正解の条件: ラベル集合 L(t) が下記の条件 i-i, または, 条件 i-ii の事象を
   含む場合:
  - 条件 i-i) 人物 i が他の全ての参加者に向けて意見を表明していた.
  - 条件 i-ii) 人物 i は一人の人物に向けて意見を表明していたが,全ての人物がその話に注意 を傾けて聞いていた<sup>6</sup>.
- ii) 二者結合  $\hat{R}_t = R_{(i,j)}^{DL}$  が正解の条件: ラベル集合 L(t) が下記の条件 ii-i, ii-ii, ii-iii のいずれかの 事象を含む場合:

条件 ii-i) 人物 i は人物 j のみに向けて問い掛け,若しくは,応答・反応を行った.

<sup>&</sup>lt;sup>6</sup>この場合,人物*i*としては,特定の一者への話し掛けという意図を持っているのであるが,結果的にこの場で生じているコミュニケーションは,一対多の一方向性の情報伝達であると考えられるため,3.2節で定義した一者集中レジームに該当するものと考えた.

条件 ii-ii) 人物 *i* は人物 *j* のみに向けて意見を表明し,人物 *j* のみがその話を注意を傾けて 聞いていた.

条件 **ii-iii**) 条件 ii-i), ii-ii) において, 人物 *i* と *j* を入れ替えた条件.

iii) 分散  $\hat{R}_t = R^0$  が正解の条件: ラベル集合が空集合  $L(t) = \emptyset$  である,若しくは,次の2つの条件 (iii-i, iii-ii) を同時に満たす場合:

条件 iii-i) 発話は独り言であった.

条件 iii-ii) 他者の発話に対して誰も応答をしなかった.

レジームの推定結果の評価結果

表 3.4 に,各会話データについて得られたレジーム推定値の正答率を,各レジームのクラス毎, 及び,全体について示す.この表には提案方法(*M*と記す)の結果に加えて,比較対象として3つ の異なる方法による結果も併記した.この比較対象の方法は,提案した会話モデルにおいて,観測 可能な変数の条件を変えたものであり,i)視線既知*M<sub>GG</sub>*,ii)視線のみ*M<sub>GO</sub>*,iii)発話のみ*M<sub>UO</sub>* とした.視線のみ*M<sub>GO</sub>*と発話のみ*M<sub>UO</sub>*の手法については,それぞれのみを観測値として与え た条件のもと,ギブスサンプリングによりレジームの系列を推定した.また,視線既知*M<sub>GG</sub>*と 視線のみ*M<sub>GO</sub>*においては,視線パターン系列*X*<sub>1:T</sub>について正解ラベルを観測値として与えた. なお,これら方法において,新たに既知とした変数以外の条件やパラメータは,提案方法*M*と同 ーとした.これらの方法との比較は,正答率の大小という観点を補完し,より多角的に提案方法 を検証することを狙ったものである.具体的には,視線のみ*M<sub>GO</sub>*,発話のみ*M<sub>UO</sub>*の各方法との 比較により,着目した人物行動の観点から会話モデルの妥当性を検証する.また,提案方法*M*と 視線既知*M<sub>GG</sub>*の方法を比較することで,頭部方向を視線方向の代用として計測することの妥当 性を議論する.

表 3.4 より, 提案方法 *M* により推定されたレジームの正答率は, 81.8% ~96.3%と平均的に高い値を示していることがわかる.また,総合的にみて,提案方法 *M* の結果は,視線のみ *M<sub>GO</sub>*, 発話のみ *M<sub>UO</sub>* の方法の正答率を上回り,また,視線既知 *M<sub>GG</sub>* に匹敵する正答率を実現していることがわかる.

また,表 3.4 からは,視線,発話を単独に用いた方法( $\mathcal{M}_{GO},\mathcal{M}_{UO}$ )よりも,それら双方を用いた方法( $\mathcal{M},\mathcal{M}_{GG}$ )の方が全体的に高い正答率を達成していることも読み取れる.なお,数値上は,発話のみの方法 $\mathcal{M}_{UO}$ でも十分な精度が得られているように見えるが,この方法で得られたレジームの系列は,二者以上の同時発話や瞬時的な沈黙など発話の状態に敏感なものとなっており,会話の流れを適切に反映しているとは言い難いことが観察により判明した.それに対して,視線パターンと発話の双方を利用した方法では,発話の状態に敏感に左右されることもなく,発

話情報だけからは誰が話し手であるか特定困難な同時発話などの状況においても,発話権を保持 している話し手を安定に同定できることが確認された.これらの結果により,視線パターンと発 話状態の双方をモデルに組み込むことの有効性が示唆された.

さらに表 3.4 より,提案方法 *M* は,視線既知の方法 *M<sub>GG</sub>* と比較して,同程度かそれを上回る レジームの正答率を達成していることがわかり,視線方向の代わりに頭部方向を計測することの 妥当性が示唆された.本方法 *M* で推定された視線方向には,3.5.4 節で述べたように誤差が含ま れている.それにも関わらず,このような正答率が得られた原因として,以下のような考察が成り 立つ.まず,話し手が受け手に対して話し掛ける場合や,聞き手が話し手の発話を傾聴する場合, 対象へのアテンションは継続するものの,その視線は対象への凝視と凝視回避を繰り返すという 特性があり [103, 59, 100],図 3.9(a)からもそのような特性を反映した同一対象への凝視・凝視回 避が読み取れる.一方,このような場面では,視線方向の変化に伴う頭部方向の変化が比較的小 さく,この性質により 3.5.4 節で述べたように凝視回避を凝視として誤判定するケースが多く生じ ると考えられる.このように提案方法によって推定された視線方向は,実際の視線方向よりも頭 部方向の影響を強く受けたアテンションの方向を示唆するものと考えられ,これが反って,話し 掛け・傾聴時のアテンションの方向と符合した結果,推定された会話レジームは実際の会話現象 を十分に反映したものとなり,高いレジームの正答率が得られたものと考察できる.

### 3.5.6 評価のまとめ

3.5.4 節で述べた視線方向の推定精度の評価から,会話レジームと参加者行動の関係が適切にモ デル化されていることが確認され,また,3.5.5 節において,推定された会話レジームが実際の会 話構造を反映していることが確認された.これらの結果より,本論文で提案した会話モデル,並 びに,会話モデルに基づく会話構造の推定方法の有効性が示唆された.

# 3.6 実験2 ―トラッキングによる頭部方向計測に基づく推定―

本節では,参加者の頭部方向の計測法として,3.4節において述べたトラッキングによる方法を 用いた結果を示し,3.5節の磁気センサによる結果と比較して,頭部トラッキングの有効性を検証 する.実験データや,その他の条件は,3.5.1節記載のものと同一のものを用いた.なお,STC法 のSTC法の粒子数は $N_s$ =2000とした.G1の人物3以外については,基本テンプレートのみを設 定した.



図 3.12: 会話シーンの撮影環境 . (a) 参加者 , 及び , カメラの配置 , (b) カメラ  $C_W$  から撮影され た参加者の全体ショット (グループ G1) .

## 3.6.1 頭部方向の推定精度

図 3.13 には,会話データ G1-C1 について得られた各参加者の頭部方向(水平方位角)の時系列 (最初の 3600 フレーム=120 秒を抜粋)を示す.また,図 3.13 には,各参加者の頭部にヘアバンド を用いて装着した磁気式センサ(POLHEMUS Fastrak<sup>TM</sup>)の計測データも併せて示す.図 3.13 では,比較のため,STC 追跡法の計測値に対して,その平均値がセンサ出力の平均値に等しくな るようなバイアスが加えられている.図 3.13 中の STC 法による計測値には,絶え間ない微小振動 がみられるものの,全体の傾向はセンサ出力と類似している.特に人物が視線をある人物から別 の人物に移す場合に見られる頭部方向の急速な変化に対して,良好に追従していることが分かる. 両者の計測値の平均偏差(差の絶対値の平均値)は,会話 G1-C1 の各人  $P1 \sim P4$  についてそれ ぞれ,5.1,6.9,10.0,3.7[deg]であった.また,全データについての平均偏差は5.9[deg]であった.

図 3.14 には, G1-C1 における頭部追跡の様子を示す.図 3.14(b) には,横顔による自己遮蔽が 生じる場合,図 3.14(c) には,登録されたテンプレートとは異なる表情が生じる場合,図 3.14(d) には,部分的に顔領域が遮蔽された場合の各々において頑健に追跡が継続された例を示す.なお, 各人物の追跡は,オフラインで実施され,各データの初期フレームから終端フレームまで自動的 に追跡が行われた.また,データ中には,顔の全域を両手で覆うという完全オクルージョンが生 じる区間が存在し,その区間における追跡精度は極端に劣化するものの,その状態から脱した時 点で再初期化に成功し,追跡が続行された.処理速度は,動作周波数 3.2GHz の PC において,約 0.20[sec/frame] であった.以下の実験では,STC 法の計測値に対して平滑化などの前処理は施さ ないデータを使用した.なお,単純な時系列平滑化を導入した場合も,ほぼ同じ結果が得られる ことを予備実験により確認している.



図 3.13: 計測された頭部方向 (水平方位角)の時系列 (G1-C1の会話開始から2分間を抜粋).実線:STC 追跡法による計測値,グレーの線:磁気式センサの計測値.



図 3.14: 会話中の各参加者 (グループG1) と追跡の様子 . (a) 全参加者 , (b) 横顔 (P1) , (c) 笑い (P2) , (d) 手で口を覆う動作 (P3) .



図 3.15: 推定された系列(図 3.13 と同じデータ・時間区間). (a) 視線方向  $\{X_{i,t}\}_{i=1}^{4}$  (実線:推定値,破線:正解データ), (b) 会話レジームの状態;各時刻 t において,一者集中レジーム  $S_t = R_i^C$ の場合, iの位置のみにバンド,二者結合レジーム  $R_{(i,j)}^{DL}$ の場合,i,jの位置にバンド,分散レジーム  $R^0$ の場合は空白により示す.

表 3.5: 推定された視線方向の推定精度 [%]. Proposed: 提案法の結果, Reference: 磁気式センサ で計測を行った結果.

	G1-C1	G1-C2	G2-C1	G2-C2
Proposed	64.8	57.3	67.5	69.3
Reference	71.1	59.3	72.4	75.9

表 3.6: 会話レジームの平均正答率 [%] . Proposed: 提案法の結果, Reference: 磁気式センサで計 測を行った結果.

	G1-C1	G1-C2	G2-C1	G2-C2
Proposed	78.7	88.9	86.1	90.7
Reference	81.8	92.1	91.4	96.3

## 3.6.2 推定された視線方向の評価

図 3.15(a) には,提案法によって推定された視線方向と,対応する正解データを示す.視線方向 の正解データは,人手により映像を観察することで作成された.表3.5 には,推定方向と正解が一 致したフレームの割合を推定精度として示す.比較のため,表3.5 には,磁気式センサによる結果 も併記した.3.6.1 節で述べたように,STC 法の頭部方向データには,多くのノイズや誤差が含ま れる.しかし,表3.5 から分かるように,提案法の結果は,磁気センサの結果とそれほど大差がな いことから,頭部方向の計測手段として STC 法は有望な手段であることが示唆される.また,推 定誤りの内訳をみると,視線を逸らした状態に関する誤りが多く(平均70.4%)を占めている.こ れは,人間は頭部方向を変えずに視線方向を変化させることができるという性質に起因するもの であり,センサを用いた場合と同様に,頭部方向から視線方向を推定する際の本質的な限界とい える.

## 3.6.3 推定された視線パターン,及び,会話レジームの様子

図 3.16 には,会話 G1-C1 から推定された視線パターン,及び,会話レジームの様子を示す.最 初  $(t = t_1)$ ,人物4が他の参加者に対して意見を述べている場面において,人物4への視線の集中, 及び,人物4への一者集中レジーム  $R_4^C$ が推定結果として得られた.次に  $(t = t_2)$ ,人物2が人物 4 に対して応答を示す場面において,人物2,4の間の相互凝視,及び,この二者間の二者結合レ ジーム  $R_{(2,4)}^{DL}$ が推定された.その後  $(t = t_3)$ ,人物4が人物2に発話権を譲渡した後,人物2が人物 1 へ話し掛け,人物1がそれに応える場面においては,人物1,2間の二者結合レジーム  $R_{(1,2)}^{DL}$ が推定された.センサによる結果と比較して,提案法の結果は,推定された視線方向が若干,正 確性に欠けるものの,会話レジームの遷移は,実際の会話の流れを反映したものであることが確



図 3.16: 推定結果の様子 (G1-C1,  $t_1 = 300, t_2 = 558, t_3 = 673$ ). 含む, 各参加者の画像, 視線方向(細い矢印:推定結果, グレーの矢印:正解データ), 及び, 推定されたレジーム  $\hat{S}_t$ .

認された.

## 3.6.4 会話レジームの評価

図 3.15(b) には,提案法によって推定された会話レジームの状態系列の一部を示す.また,表3.6 には,各会話データについて,レジームの推定値の平均正答率を評価した結果を示す.レジーム の正答率は,3.5.5 節記載の方法を用いて,会話中の各フレームにおいて,推定されたレジームの 状態と発話区間に付与されたアノテーションとの比較により計算される.表3.6 から,提案法の正 答率は,比較対象とした磁気センサを用いた方法よりは劣るもの,それに十分に類するものであ ることが分かった.

以上の実験より提案法の有効性が確認され,また,STC法による画像中の人物頭部追跡は,会 話構造推定のための計測手段として有効であることが分かった.

# 3.7 議論

本節では以下の各項目について議論を行う.

会話レジームとその評価について

本論文では,会話を構成する代表的な会話構造を定義するために,視線パターンの構造的な特徴 に着目し,3クラスの会話レジームを仮説的に設定した.しかしながら,実際の会話においては, 本論文の会話レジームでは想定していない状況,例えば,話し手が一人の受け手のみに向かって 一方的に話し掛ける場面や,参加者が複数のサブグループに分かれて会話が進行するような場面 もあり得る.今後は,これらの場面へ対処するために,データに基づいて適応的に会話レジーム を設定することが検討課題としてあげられる.また,参加者数の異なる会話や,会話中の参加者 数の増減,様々な人物配置,ノートや資料,黒板等の使用など,より多様な会話の状況に対処で きるよう提案法を拡張し,評価を行うことが望まれる.

また,本論文では,推定されたレジームの評価のために,発話区間を単位として,その人物間の 問い掛け・応答に関するラベルを用いて評価を実施した.しかしながら,会話中には,発話とし て現れない非言語的なメッセージ交換,例えば,話し手の問い掛けに対して,表情のみで無言に て反応を返す場面など,も含まれており,今後は,それらも含めた評価法を検討する必要がある.

#### 関連研究に対する本研究の位置付け

本研究は,Stiefelhagenらの研究[106,107]の発展形であると位置付けることができる.彼らは, グループ会話において,参加者の頭部方向,及び,発話の状態から各人物の視線方向(注意の焦 点と呼んでいる)をベイズ推定により求める手法を提案している.本論文の方法との類似点とし て,頭部方向の尤度関数にガウス分布を仮定している点があげられ,また,相違点として以下の 項目があげられる.まず,彼らは,個々の参加者の視線方向を独立に求めており,参加者間のイン タラクションや会話の構造について陽にモデル化を行うというアプローチをとっていない.また, 視線方向と発話の共起関係を用いて推定を行っている点において,部分的に会話の性質を利用し ていると考えられるが,彼らの提案手法は,会話の構造の推定を目的とするものではない.さら に,視線を逸らした状態を推定の対象から除外している点でも本研究とは立場を異にする.

また,坊農らは,ポスター発表会場における説明者と来訪者との間の会話を対象とした研究を 進めている [96,95].この研究では,来訪者の参与役割が,非参与者から傍観者を経て傍参与者, 受け手,話し手へと動的に遷移する過程に着目し,参加者の音声,及び,赤外線 ID システムから 得られる位置・頭部方向などの情報を用いて,来訪者の参与役割の同定を試みている [96].また, レクチャーモードとインタラクションモードと呼ばれる2つの会話のモードを定義し,説明者の 無音区間持続長の長短によりこれら2つのモードが判別できることを示唆している [95].この研 究で対象とする会話は,参加者の役割が説明者と来訪者に明確に分離されており,また,その場 において生じる会話も説明者からのポスター説明,及び,来訪者・説明者間の質疑応答と限定さ れている.また,来訪者の移動により会話参加者の人数や位置が時間的に変化する状況を対象と している.それに対して,本研究では,対等な立場の参加者による自由な会話を対象としており, 現在のところ参加者の増減・移動がないことを前提としている.これらの点において,本研究と 坊農らの研究とでは,提案方法の適用範囲が異なるといえる.さらに,本研究は,会話の構造の モデル化を志向している点においても,坊農らの研究とは方向性が異なる.

# 確率的なアプローチの利点・発展性

本論文で提案した確率的なアプローチの利点としては,まず,複数の人物から表出されるマル チモーダルな情報を統一的な枠組みのもと統合することができるという点があげられる.そのた め,頭部方向や発話状態に加えて,他の種類の人物行動を組み込んだ会話モデルにも容易に拡張 できると考えられる.例えば,頷きや相槌,胴体姿勢,表情,ジェスチャなどの非言語的な行動 も会話においては重要な役割を果たしていることが知られているが[59,56,58],それらを会話モ デルに組み込むことで,より精緻に会話の状態を推測することが可能になると思われる.

また,提案方法では,この確率的な枠組みのもとで,視線パターンと発話状態から会話レジー ムを推定するという問題と,頭部方向から視線方向を推定するという2つの問題を同時に解いて いると考えることができる.これにより,それぞれの問題を単独に解く場合よりも,より精度の高 い推定が実現されていると考えられる.その理由としては,聞き手は話し手の方を良く見るとい うような会話の性質に関する情報を利用することで,視線方向の推定精度が高まり,また,より 高精度の視線方向を入力とすることで,会話レジームの推定の精度も向上するというような,互 いの推定情報が相補的に作用する点があげられる.このような利点をさらに活用することにより, 人物が密集して着席する場合など,頭部方向による視線方向の判別が困難な状況にも対処できる 可能性がある.例えば,一者集中の場合において,一部の聞き手の視線方向が判別できない場合 でも,他の聞き手を含めた全体での視線方向の多数決により視線の焦点を求め,さらにその結果 を発話状態の情報と統合することで,ロバストに会話構造を同定できるものと期待される.

## 画像からの頭部方向の推定に関して

3.6節の実験結果より,STC法は会話構造推定のための計測手段として有効であることが分かった.今後の改良点としては,顔面の平面近似によるフィッティングの不安定性を解消し,計測精度 を高めるため,より顔面形状に近い立体的なテンプレートへの拡張が考えられる.また,テンプ レート設定を自動化するため,顔検出技術[151]の導入も検討課題としてあげられる.さらに,本 論文では,一人一台のカメラを使用したが,現実の応用では,全方位カメラ等を利用し,より少 数のカメラでより多くの参加者を捉えることが導入コストや利便性の観点から望ましい.そのた め,画像中の各人の顔領域が小さく,また,複数の人物が混在する状況においても,安定に各人 物の頭部を追跡することが課題となる.さらに,現在,低解像度画像から直接,視線方向の推定 を行う研究も進展しつつあり[152],その結果との統合も興味深い課題である.

## アプリケーションと課題

提案方法の応用先としては,会議映像の自動編集・アーカイブや多地点間の複数人対複数人の 遠隔会議,会話エージェント・ロボットなどがあげられ,これらの応用において,会話構造の推 定結果は有力な入力情報になると期待される.例えば,会議映像の自動編集システムにおいては, 話し手の映像と受け手の映像を時間的に切り替えて表示する等,参加者の役割に応じた様々な映 像表現が可能となり,誰が誰に向かって話をしているか等,視聴者にとって会話の内容がより分 かりやすい映像を自動的に編集することが可能になると期待される.また,3.3.2節で示した方法 は,対象となる会話の時間区間についてバッチ処理を行うものであったが,実時間性が要求され るアプリケーションに提案方法を供する場合,オンライン推定法の利用が必須となる.

また,提案方法は,ハイパーパラメータの設定が煩雑であり,ヒューリスティクスが介在する 点が欠点としてあげられる.しかし,これらのパラメータやモデル構造の設定は,会話現象の性 質と密接に関連するため,それらの吟味・検討のプロセスは,会話の様々な側面を分析・定量化す るツールの開発へと発展する可能性がある.例えば,推定されたパラメータに基づいて,会話に 対する積極性や他者への影響力などといった参加者個人の特性や,会話の円滑さなどのグループ としての特性などを定量化することも興味深い検討課題の一つである.このように,従来,定量 的な分析が困難であった複数人物の会話という現象に対して,本論文は,一つの新しいアプロー チを提供するものと位置付けられる.

## 3.8 結言

本章では,時系列画像解析における観測の間接性の要因として,複数対象間の相互作用に着目 し,その事例として複数人物による対面会話を対象として,会話の構造を推定するための方法を 提案した.会話の構造として,会話中の各参加者とその参与役割の組み合わせに着目した.また, 会話の構造に依存して参加者の行動が規定されるとの仮説を立て,マルコフ切替モデルに基づく 会話モデルを提案した.さらに,参加者の頭部方向,及び,発話状態を観測データとし,会話構 造に対応すると想定される会話レジームの状態,視線パターン,及び,モデルパラメータを推定 するため,ギブスサンプリングを用いた方法を提案した.4人会話を対象とした実験を行い,視線 方向の推定精度,及び,会話レジームの正答率を評価し,提案した会話モデル,及び,会話構造 推定方法の有効性を確認した.

このように本章で提案した相互作用を陽にモデル化するというアプローチは,時系列画像解析

における観測の間接性の問題に対する有効なアプローチであると考えられる.なお,本章では,対 象間の相互作用に焦点を当てるため,第2章で取り上げた対象間の相互オクルージョンがない環 境を対象とした.しかしながら,現実の会話の状況では,オクルージョンが生じるような人物・カ メラ配置も想定でき,その場合にはオクルージョンに頑健な手法が必要となる.その際には,本 章の相互作用のモデルと第2章で述べたオクルージョンを含む観測過程のモデルを統合したモデ ルの構築が求められる.

また,第2章,及び,本章で取り上げた事象は,要素数が少数であり,比較的,観測過程や相 互作用のモデル化が容易な事象であったともいえる.しかしながら,実世界中の事象には,視覚 的な観測情報と事象の状態との関係が明確にモデル化できない事象も数多く存在する.次章では, そのような事例として,気象レーダ画像パターンの予測の問題を取り上げる.

# 第4章 記憶に基づく気象レーダ画像パターンの予測

## 4.1 緒言

気象現象は,実世界の動的事象のなかでもとりわけ産業や生活に及ぼす影響が大きい事象であ る.本研究では,気象現象の中でも特に降雨や降雪などの降水現象に的を絞り,その局所・短時間 の予測を支援する技術の確立を目指している.局所・短時間の降水予測とは,概ね30分から3時 間先の降水量の空間分布を高解像度(1km~3km単位),かつ,リアルタイムに予測することを指 す.この局所・短時間の降水予測情報は,防災や農業,治水など産業・行政において,その重要 性が知られている[66,67].本研究では,観測の部分性,間接性がともに大きい事例として,画像 パターンとして観測される降水現象の予測という問題に着目する.ここでは観測手段として気象 レーダを用いる.気象レーダは,降水強度の空間分布を画像として得ることができる装置である [68,69].気象レーダは,降水強度の空間分布を画像として得ることができる装置である [68,69].気象レーダは,図4.1のようにマイクロ波を細いビーム状のパルスとして発信し,それ が雲の中の雨粒や氷粒によって散乱される現象を利用している.発信した電磁波の方向,及び,散 乱されて返ってくるまでの時間から,降水粒子の空間分布がわかり,また,散乱されて返ってく る電磁波(レーダエコーと呼ぶ)の強度から降水強度が推定できる.従って,降水分布の予測の 問題は,気象レーダ画像上のパターン(レーダエコーパターンとも呼ぶ)の時間変化の予測の問 題として捉えることができる.

従来,気象レーダ画像上でのパターンの動きを計測して,予測に活用する方法が提案されてい る.例えば,計測された動きに従ってパターンを平行移動させるという線形外挿法[74,75]や,降 水パターンのダイナミクスを移流拡散方程式を用いて近似的に表現し,予測を行う方法が知られ ている[76,77].しかしながら,気象衛星画像上の雲と異なり,気象レーダ画像上の降水パターン の動きは,物質(水粒子)の移動現象を直接捉えたものではない.降水パターンの動きは,大気 の流れにそって移動する傾向はあるものの,パターンを構成する要素の動きの方向とパターン全 体の動きの方向とが必ずしも一致しない[78].また,個々の要素は生成消滅を繰り返す一方で,パ ターン全体の形状は長時間維持される場合もある[79].このようにレーダ画像上のパターンは,極 めて複雑に変化し,従来の方法では,特にパターンの時間変化が大きい場面等に対応できず,そ の適用範囲は限定的なものに留まっているのが現状である.

気象レーダによる観測は,空間的にみて,地球上のごく狭い領域(半径~数100km)のみの観測

91



図 4.1: 気象レーダによる降水の観測.

であり,さらに,その3次元空間中のある断面を切り取ったものである.また,降水が存在する領 域以外の情報は得られない.さらに,気象レーダにより得られる観測量は,降水のパターンとい う気象現象の結果であり、気象現象の内部状態を記述する物理量(気圧、気温など)とは異なる. このような気象現象の内部状態に依存して変化する降水パターンのダイナミクスは,気象レーダ画 像によって観察される瞬時的なパターンの動きとは異なるため、降水パターンの予測は困難な問題 とされる.これらの観点から降水パターンの予測の問題は,観測の部分性,間接性の双方が顕著な 問題といえる. 本研究では, このような事象に対処するため, 記憶に基づく予測(Memory-Based Forecasting; MBF) と呼ぶアプローチを導入する.これは,類似する気象現象が不定期的に再現 されるという性質に基づいて,長期の観測により蓄積したデータを用いて,現時点の観測情報に 類似する過去の事例を想起し,現時点の現象も過去と同じような変化を辿るであろうと予測する ものである.この予測の方式は,観測情報が類似していれば,その観測の元になっている現象そ のものも類似しているという仮定に立脚している、本研究では、この考えのもと、現時点の気象 レーダ画像を入力とし,それと類似する過去の画像の検索を行い,検索された画像系列の時間発展 に基づき,未来の画像パターンを予測するという方法を提案する.このようなプロセスは,人間 の経験則に基づく予測のアナロジーとなっている.人間の気象予報者は,あるパターン A の後に 別のパターン B が出現するといった規則性を経験により獲得し, 蓄積された過去の経験を用いて, 将来起りうる現象を予想するというプロセスを日常行っている.記憶に基づく予測(MBF)は,こ うした人間の思考プロセスを模倣したものであるともいえる.そのため,現況に類似する過去の レーダ画像を検索し,予報者に提示することで予報者の記憶能力を補い,過去事例の想起を支援 することが可能になると考える.また,検索された画像から一定時間未来の画像を用いて予測画 像を合成することで,より付加価値の高い参考情報が提供できるものと考えている.

このような過去の事例に基づくアプローチは,これまで人工知能分野において,事例に基づく 推論(Case-Based Reasoning),記憶に基づく推論と呼ばれる枠組みの元,機械翻訳や医療診断な どへの応用が検討されている[123][124].これらを気象分野に適用した研究として,毛利らは,地 上観測により得られた雨量や風向などの離散化された気象データを用い,12時間先までの降水の 有無の予測を行う手法を提案している[129].また,Tothは,北半球の気圧の尾根の位置や本数に 関するデータを用いて,5日間平均気温や10日間合計降水量といった長時間予測の可能性を検討 している[126].これらの研究では,比較的スムーズに変化する大規模(1000km<sup>2</sup> ~ 10000km<sup>2</sup>)な 現象を対象とし,シンボル化された高次の特徴を用ているのに対して,我々は,レーダ画像情報 を直接用い,より小さなスケール(10km<sup>2</sup> ~ 100km<sup>2</sup>)で生じるより激しい現象の予測を対象として いる点が大きく異なる.また,非線形力学系の分野でも,過去の事例に基づく予測の方法が提案 されており[112,113,114,115,116,117,118],これまでに,為替や売上高,電力消費量などカオ ス性をもつ低次元時系列を対象とした予測法として検討されている[114,119,120],本研究では, これを高次元の画像パターンの予測に適用した点に新規性が認められる.

以上のような,記憶に基づくアプローチは,対象とする事象の観測過程やダイナミクスモデル を陽にモデル化する代わりに,過去の事例に基づいてモデル化を行っていると捉えることができ, 未知の複雑な現象に対処する一つの方法論として有望であると期待できる.本論文では,従来,シ ンボルや数値を対象としてきた記憶に基づくアプローチを,時系列画像パターンの予測の問題へ 拡張した方法を提案する.具体的には,過去の類似画像系列の検索と,検索された類似画像系列の 集合を用いて予測画像を合成するという2つの要素技術から構成される予測方法を提案する.ま ず,時系列画像の検索の方法としては,時系列画像を低次元の固有空間上の軌跡として表現し,軌 跡の部分区間同士の類似性の観点から、入力された現在の時系列と類似する過去の時系列を検索 するという方法を導入する.このような固有空間を用いた画像の表現方法は,アピアランスベー スと呼ばれるコンピュータビジョン分野の一技法であり, Turk らによる固有顔 [108] や, 村瀬ら によるパラメトリック固有空間法 [109, 110] に端を発し,顔や3次元物体の認識や姿勢推定,ジェ スチャー認識や,読唇など多数の応用が考案されている[153][154].また,固有空間上の軌跡の検 索,照合は,時系列画像検索やロボットの自己位置推定などにも用いられている[155].このアプ ローチは,画像の見かけの変動を低次元の部分空間上で統計的にモデル化することで,画像の生 成・観測過程に関するモデルや知識を不要とする点が利点であり,事象の観測過程及びダイナミク スのモデル化が困難な本研究の事例に適合すると考えられる.また,検索された類似画像系列の集 合を用いて予測画像を合成するため,前述の非線形時系列の予測法を参考にした方法を提案する.

本章は以下のように構成されている.4.2節では,記憶に基づく予測の枠組みと,その具体的な 実現方法を示し,4.3節では,実データを用いた検索,及び,予測の精度を評価した結果を示し, 提案方法の有効性を論じる.最後に4.4節において,まとめと今後の研究課題を示す.



図 4.2: 記憶に基づく画像予測の枠組み.

# 4.2 記憶に基づく予測の枠組みと実現手法

図 4.2 に示すように,記憶に基づく予測の枠組みは,特徴抽出,検索,予測の3つの段階より 構成される.始めに,データセットとなる画像集合が収集され,画像データベースへ蓄積される. 次に,その中の画像系列について,一定間隔毎に画像パターンの性質を表現する画像特徴量が計 算され,画像特徴データベースに蓄積される.次に,最新の画像系列(または,任意の画像系列) が入力されたとき,それらの画像特徴とデータベース中の画像特徴とが比較され,特徴量が類似 する画像系列が検索される.最後に,検索された系列に続くその未来の部分を用いて,入力系列 の画像特徴量の未来の予測点が求められ,その点に対応する予測画像がデータベース中の過去の 画像を用いて合成される.以下では,以上の過程の一実現手法を記述する.

4.2.1 特徴抽出

気象レーダ画像中のレーダエコーパターンは,降水強度の空間分布を表現したものであり,そ の構成要素であるエコーセルが移動,生成,消滅を繰り返しつつ,全体としてあるまとまった領 域(降水域と呼ぶ)を形成するパターンである.エコーパターンは時間とともに変化する非剛体 パターンであるが,その時間的な変化の性質は,観測するスケールの違いにより異なることが知 られている.その性質の一端を,図4.3に示したエコーパターンを例にとって概説する.図4.3に 含まれるエコーパターンは,筋状パターンと呼ばれ,エコーセルと呼ばれる小さい領域の集合が 幾つかの平行な列を構成するように配置されるパターンである.この例では,図4.4のように,あ る領域において出現したエコーセルは,おおよそ大気の流れに沿って移動しながら成長し,別の



図 4.3: レーダエコーパターンの系列.(a)5 分間隔で観測された 3 フレームの画像(サイズ =340×340[画素]=[km<sup>2</sup>]), (b) 白線枠内(60×60 km<sup>2</sup>)の拡大図.



図 4.4: エコーセルの生成,移動,消滅の様子.



図 4.5: 大局的特徴の様子 . (a) メッシュ特徴 , (b) 速度ベクトル場 . 一つのメッシュ領域のサイズ は 55 画素 .



図 4.6: エコーパターンの種類. (a) 筋状, (b) 散在状, (c) 層状. (表示の範囲:210×210[画素]).



図 4.7: テンポラルテクスチャ特徴の抽出(図 4.3 の部分領域内). (a) 局所的な時空間領域内の運動軌跡の様子, (b) ノーマルフローの分布( $\theta$ :極座標表示した方向, $\phi$ : ノーマルフローの大きさ).

ある領域において消滅する.このような個々のエコーセルの生成,移動,消滅が繰り返されるこ とでパターン全体の形状は,個々のエコーセルの寿命と比較して,長時間維持されるという性質 がある[79].また,エコーセルの動きの方向とパターン全体の動きの方向とが必ずしも一致しない ことも知られている[78].このようにエコーパターンは,着目する空間スケールに依存して固有 の空間的な広がり,速度,寿命を持つという性質がある.そのため,本研究では,エコーパター ンの性質を,大小2つのスケールの画像特徴量に分解して表現するという方策をとる.大きいス ケール,言い換えれば,エコーパターンの大局的な性質を表現するために,メッシュ特徴,速度 ベクトル場と呼ぶ画像特徴量を導入する.また,エコーパターンの局所的な性質を表現するため に,テンポラルテクスチャ特徴と呼ばれる特徴量を導入する.

#### 大局的な画像特徴

大局的な画像特徴として,本研究では,メッシュ特徴,速度ベクトル場と呼ぶ画像特徴量を用 いる.メッシュ特徴は,エコーパターンの大局的な空間分布,すなわち,大局的なパターンの位 置と形状を表すために導入する.具体的には,図4.5(a)のような画像を粗いメッシュ状の領域へ と分割した一種の低解像度の画像濃淡値情報であり,パターン表面の細かなテクスチャの変化の 影響に左右されずに,大局的なパターンの位置と形状を表現することができる.具体的には,メッ シュ特徴  $x_1(t)$ は,各メッシュ内の濃淡値を,複数の画像フレームについて時間平均した値からな るベクトル

$$\boldsymbol{x}_{1}(t) = [x_{1,1}, x_{1,2}, \cdots, x_{1,M}, \cdots x_{M,M-1}, x_{M,M}]^{t}$$

$$(4.1)$$

として表現される.ただし, $M \times M$ はメッシュ分割の数を表し, $x_{i,j}$ は,メッシュの位置(i,j)において,時刻tを終端とする直近の複数フレームについて平均された濃淡値を表す.メッシュ特徴

の時間変化は,降水域全体の移動や成長,減衰などの現象の長期的な変化を表現できる.

速度ベクトル場は,図4.5(b)のように各メッシュ内のエコーセルの移動速度と方向を表現した ものである.具体的には,速度ベクトル場  $x_2(t)$ は,各メッシュ領域についての速度ベクトルの成 分を

$$\boldsymbol{x}_{2}(t) = [v_{x;1,1}, v_{y;1,1}, \cdots , v_{x;1,M}, v_{y;1,M}, \cdots , v_{x;M,1}, v_{y;M,1}, \cdots , v_{x;M,M}, v_{y;M,M}]^{t}$$
(4.2)

のように並べたベクトルとして構成される.ただし, $v_{x;i,j} \ge v_{y;i,j}$ は,メッシュ(i,j)において,そ れぞれ速度ベクトルのx成分とy成分(図4.3上では,それぞれ水平方向と垂直方向に対応)を表 す.ただし,この速度ベクトルは,時刻tを終端とする直前の複数フレームにわたって計算された 速度ベクトルの平均ベクトルである.ここで,各々の速度ベクトルは,一般的なプロックマッチ ング法や著者らが提案した方法 [156] などのオプティカルフロー計算により求められる.速度ベク トルの時間平均操作は,個々のエコーセルの動きに伴う変動によらず安定した動きベクトルを得 るために行うものである.このようにして得られた速度ベクトル場は,レーダでスキャンした高 度の大気の流れを近似したものであると考えられる.さらに,速度ベクトル場の時間的な変化は, 降水系に影響を与える前線や低気圧の移動などの大きなスケールでの気象現象の変化を表すと期 待される.

#### テンポラルテクスチャ特徴

エコーパターンは,その元になる降水現象の種類により大きくその性質が異なる.降水現象の 構造の違いは,局所的なエコーパターンの形状や配置,生成・消滅などの動きの特徴として観測 できる.ここでは,その降水現象の違いを画像特徴量として捉えるために,著者らが以前に提案 したテンポラルテクスチャ特徴[157]を導入する.テンポラルテクスチャ特徴は,一般に知られる テクスチャ特徴とは異なり,複数のフレームに渡って分布する画像要素の動的な性質を記述する ことができる画像特徴である.文献 [157]で提案されている7種のテンポラルテクスチャ特徴量 は,エコーセル配置の規則性や,その運動の不確定性,生成・消滅の頻度などを記述する目的で 提案されたものである.これら特徴により,図4.6で示した筋状,散在状,層状といった降水現象 の違いに対応したエコーパターンのタイプの種類が見分けられることが確認されている[157].具 体的には,この特徴量は,動きの集中度(Motion intensity)  $f_1$ , ノーマルフローのエントロピー (entropy)  $f_2$ ,高速成分の割合(High-speed ratio)  $f_3$ ,優勢な速度の大きさ(Speed)  $f_4$ ,輪郭の方 向性の強さ(Directionality strength)  $f_5$ ,輪郭の密度(Density)  $f_6$ ,運動軌跡の隠蔽率(Occlusion ratio)  $f_7$  から構成されている. $f_1$ , $f_2$ , $f_3$ は,局所的な領域に含まれるエコーセルの動きの一様 性や多様性に関係し, $f_5$ と $f_6$ は,エコーセルの空間的な分布に関する特徴量である.また, $f_7$ は, エコーセルの出現,消滅の頻度に関係する特徴量である.
これらのテンポラルテクスチャ特徴は,図4.7(a)で示すように,局所的な時空間領域において, パターンの輪郭の移動によって描かれる運動軌跡を解析することで得られる.この解析は,ある 時空間領域内に含まれ得るノーマルフローの分布を,その方向と速度の大きさをパラメータとす る分布として計算するものであり,図4.7(b)にその一例を示す.なお,ノーマルフローとは,輪 郭の法線方向にそった運動成分の事である.

特徴量の値は,図4.7(b)で示すような運動軌跡の分布から計算されるノーマルフローの確率 分布に基づいて定義される.テンポラルテクスチャ特徴は,図4.5で示したようなメッシュの各 位置について計算され,その値の平均値が時刻tにおけるテンポラルテクスチャ特徴ベクトル  $x_3(t) = \{f_1, \dots, f_7\}^t$ となる.ここで,エコーパターンの局所的性質は画像内においてほぼ一様で あると仮定している.また,このテンポラルテクスチャ特徴は,エコーパターン全体の位置や大 きさ,形状には不変であるという性質をもつ.

### 4.2.2 特徴の時系列の表現法

記憶に基づく予測に向けた類似パターンの検索のためには,入力されたパターン(クエリと呼ぶ)と検索されたパターンがある一時点において類似しているだけではく,その時間変化の傾向も類似している必要がある.このような時間変化の傾向は,特徴ベクトルの時系列として表現できるが,ここでは,検索をより効果的に行うために,各時刻tにおける各特徴ベクトル $x_k(t), k = 1, 2, 3$ を主成分分析により,低次元の固有区間上の新しいベクトルへと変換する.この変換には,次元圧縮や相関のある成分の削除などの効果がある.また,この変換は,メッシュ特徴,速度ベクトル場,テンポラルテクスチャ特徴の各々について独立に実施するものであり,この固有空間へ変換されたメッシュ特徴,速度場,テンポラルテクスチャ特徴を,それぞれ $y_1(t), y_2(t), y_3(t)$ と表す.

以下では,特徴 k についての計算法を示す.まず,過去のレーダ画像データベース中のデータ の各時刻に対する特徴ベクトルの集合について,共分散行列 Σ<sub>k</sub>

$$\Sigma_k = \frac{1}{N_S} \sum_{t=1}^{N_S} (\boldsymbol{x}_k(t) - \overline{\boldsymbol{x}}_k) (\boldsymbol{x}_k(t) - \overline{\boldsymbol{x}}_k)^t$$
(4.3)

が計算される.ここで, $N_S$ は,特徴ベクトルの時間ステップ数の合計,つまり,データの個数を 表す.また, $\overline{x}_k$ は, $x_k(t)$ の平均ベクトルを表す.次に,共分散行列 $\Sigma_k$ について,その固有ベクトル $e_{k,i}$ と固有値 $\lambda_{k,i}$ を

$$\Sigma_k \boldsymbol{e}_{k,i} = \lambda \boldsymbol{e}_{k,i} \tag{4.4}$$

の解として計算する.その後,特徴ベクトルは,大きい固有値に対応した固有ベクトルの集合を 基底とする固有空間内の特徴ベクトル yk へと,

$$\boldsymbol{y}_{k}(t) = [\boldsymbol{e}_{k,1}, \boldsymbol{e}_{k,2}, \cdots, \boldsymbol{e}_{k,M_{k}}]^{t} (\boldsymbol{x}_{k}(t) - \overline{\boldsymbol{x}}_{k})$$

$$(4.5)$$

のように変換される.ここで, $M_k$ は固有ベクトルの個数を表す.これは,固有値の累積寄与率が閾値より大きくなる最小の固有ベクトルの個数として決定される.3種類の特徴ベクトル $y_1(t), y_2(t), y_3(t)$ が,各時刻tにおいて計算され,データベースに蓄積される.以下ではこの固有空間を特徴空間と呼ぶことにする.エコーパターンの時間変化は,図4.8のように特徴空間上の軌跡として表現される.類似するエコーパターンの系列は,この空間上でも類似する軌跡となることから,2つの軌跡の近さによって,系列間の類似性を定めることが可能である.

### 4.2.3 時系列パターンの検索

予測に向けた類似パターンの検索のためには,現時点の最新パターンである入力パターン(ク エリ)と検索されたパターンがある一時点において類似しているだけではく,その時間変化の傾 向も類似する必要がある.そのような検索を実現するため,パターンの時系列を図 4.9 のように, 特徴空間中における特徴点の軌跡として表現し,パターンの時間的な変化を,軌跡上での特徴点 の移動とみなす.これにより,二つの時系列パターン間の類似性は,特徴空間中の二つの軌跡の 類似性によって定義することができる.

ここでは,検索の問い合わせ(クエリ)として入力される時刻  $\{T - L + 1, \dots, T - 1, T\}$ の画 像系列を入力系列と呼び,また,検索の対象となる時刻  $\{t - L + 1, \dots, t - 1, t\}$ の画像系列をマッ チング系列と呼ぶ.以下では,入力系列およびマッチング系列を,それぞれその終端の時刻を用 いてT,tと記す.なお,Lは部分系列の長さを表す.検索の処理は,検索候補の取得と,その順 位付けという二つの手順よりなる.

まず,特徴量kに関する二つの時刻m, nのパターン間の相違の度合い(非類似度とも呼ぶ)  $d_k(m, n)$ を,各々の特徴空間中における特徴点間の距離を用いて,

$$d_k(m,n) = \frac{|\boldsymbol{y}_k(m) - \boldsymbol{y}_k(n)|}{V_k(T)}$$

$$(4.6)$$

ように定義する.ただし, $|\cdot|$ はユークリッドノルムを表し, $V_k(T)$ は入力部分系列内において特徴点が単位時間ステップ当りに進む距離に対応する遷移速度を表す.例えば,この遷移速度 $V_k(T)$ は,

$$V_k(T) = |\mathbf{y}_k(T) - \mathbf{y}_k(T-1)|$$
(4.7)

のように計算することができる.この遷移速度  $V_k(T)$  による正規化は,エコーパターンの種類や 面積による影響を軽減したパターン間の相違の度合いを計算するために行うものである.これに より,後述の検索候補選別における閾値の決定が容易になるという利点が生じる.具体的に,ここ では,入力系列の時間変化の大きさを基準にして,パターン間の相違の度合い $d_k$ を定めている.

次に,与えられた入力系列 T について,それに類似する検索候補の部分系列の集合  $S_T$  を抽出 する.ここでは,そのために,データセット中の全ての時刻 t について,順次,各特徴量 k 毎の非



図 4.8: 特徴点の軌跡 ( 時刻  $t_1 < t_2 < t_3$  ). (a) 固有空間中のメッシュ特徴, (b) 固有空間中の速度 場, (c) テンポラルテクスチャ特徴 [軸=エントロピー  $f_2$ ,速度の大きさ  $f_4$ ,輪郭の密度  $f_6$ ].各 点の間隔は1時間.



図 4.9: 特徴空間中の入力 (問い合わせ) 系列と,マッチング対象の過去の系列の様子.

類似度  $d_k$ を計算し,区間 Lに渡って閾値を満たす部分系列を検索候補として,

$$S_T = \bigcap_{\substack{k \in \{1,2,3\}\\i \in \{0,\cdots,L-1\}}} \{t | d_k(T-i,t-i) < \xi, t = 1,\cdots,N_S\}$$
(4.8)

のように選択する.ただし, $\xi$ は許容する非類似性の上限を定める閾値である.このようにして 得られた集合  $S_T$ は,入力系列と一定時間 Lの間にわたって類似した過去の部分系列から構成される.

次に,検索候補の集合 S<sub>T</sub> 中から,最も類似する幾つかの検索候補を選択し,順位をつけて出力 するために,部分系列間の非類似度を定義する.ここでは,まず,DP マッチングを用い,問い合 わせの系列と検索候補の系列間の時間的な伸縮を許容した非類似度を次のように計算する.

$$D_{k}(T,t) = \min \left\{ \begin{array}{l} D_{k}(T-1,t) + d_{k}(T,t)/L \\ D_{k}(T-1,t-1) + d_{k}(T,t)/L \\ D_{k}(T,t-1) + d_{k}(T,t)/L \end{array} \right\}$$
$$D_{k}(m,n) = d_{k}(m,n)/L$$
$$\text{if } m = T - L + 1 \text{ or } n = t - L + 1.$$

(4.9)

DP マッチングとは,二つの時系列間の照合において,系列の時間伸縮を許容して照合を行う方法 として一般に知られてる.DP マッチング用いることで,類似した気象現象間に存在する時間発展 の速度のばらつきを吸収することができ,検索の見逃しを少なく抑えた尺度を得ることができる.

続いて,各特徴量について得られた非類似度を統合して,パターン全体の非類似度E(T,t)を

$$E(T,t) = \frac{1}{R} \sum_{k=1}^{R} D_k(T,t)$$
(4.10)

のように求める.ただし,ここでは特徴ベクトルの個数 R は 3 である.この非類似度を検索誤差 とも呼ぶことにする.

最後に,検索誤差 E(T,t) の小さい順に上位いくつかの検索候補を選択する.ここでは,利用者 から指定される検索結果の個数の上限 K を用い,次のように検索結果の集合  $C_T$  を得る.

$$C_{T} = \{q_{1}, q_{2}, \cdots, q_{K'}\}$$
(4.11)  
where
$$E(T, q_{i}) \leq E(T, q_{i+1})$$

$$K' = \max\{i | i \leq K, E(T, q_{i}) \leq E_{TH}, i \leq N\}$$

$$|q_{i} - q_{j}| > r, i \neq j$$

$$q_{1} = \arg\min_{t \in S_{T}} E(T, t).$$

ただし,Nは検索候補の集合 $S_T$ 内の要素数であり, $E_{TH}$ は検索誤差の閾値である.また,rは検 索結果の系列の間で時間的に近いものが出力されることを抑制するための定数である.

### 4.2.4 予測画像の生成

最後に,検索された画像系列集合  $C_T$ 中の各部分系列の未来に位置する画像を用いて予測画像 を合成する方法を述べる.ここでは,入力系列 T の $\eta$ 時間ステップ後の予測を目標とする.その ために,まず,入力画像の濃淡分布に関する特徴量であるメッシュ特徴  $x_1$ について,予測対象時 刻  $T + \eta$ における特徴点の位置を予測する.ここでは,非線形予測手法の中でもノイズにロバス トであるという性質がある修正ローレンツ法 [115]を用いる.この手法では,時刻 Tの入力系列に 対する検索結果の時刻  $q_1, q_2, \dots, q_{K'}$ の $\eta$ ステップ後, $q_1 + \eta, q_2 + \eta, \dots, q_{K'} + \eta$ ,における特徴 を用い,その線形結合によって予測点  $\hat{y}_1(T + \eta)$ を得るという方法である.その線形結合の重み 付け係数として,検索誤差の逆数に比例した量を用いることで,本手法では,予測点  $\hat{y}_1(T + \eta)$ を

$$\widehat{\boldsymbol{y}}_{1}(T+\eta) = \sum_{i=1}^{K'} w(q_{i}) \, \boldsymbol{y}_{1}(q_{i}+\eta)$$
(4.12)

$$w(q_i) = \frac{1}{E(T,q_i)} \cdot \frac{1}{\sum_{j=1}^{K'} (1/E(T,q_j))}$$
(4.13)

のように計算する.

次に,検索された過去の画像を用いて予測特徴点 $\hat{y}_1(T + \eta)$ に対応する予測画像の再構成を行う.ここでは,時刻Tに観測された画像をI(T),その $\eta$ ステップ後の予測画像を $\hat{I}(T + \eta)$ と記す.入力画像からメッシュ特徴を求める変換Gは一意であるが,その逆に画像を特徴点 $y_1(q_i + \eta)$ から一意に求めることはできない.そこで,生成した予測画像から再度求められるメッシュ特徴が,予測した特徴点と等しくなるという拘束条件 $|\hat{y}_1(T + \eta) - G(\hat{I}(T + \eta))| = 0$ を仮定する.変換Gは,画像輝度の平均演算と固有空間への変換からなる線形変換であるので,拘束条件を満たす予測画像 $\hat{I}(T + \eta)$ の一つとして,

$$\widehat{I}(T+\eta) = \sum_{i=1}^{K'} w(q_i) I(q_i + \eta)$$
(4.14)

のように,検索された系列の未来の画像  $I(q_i + \eta)$ の加重和として予測画像を合成する方式をここでは採用する.

### 4.3 実験

本節では,提案した方法の有効性を検証するために行った実験について述べる.まず,4.3.1節に おいて,実験に用いたデータセット,及び,実行のためのパラメータについて述べる.次に,4.3.2 節において,類似パターン検索に関する実験例,並びに,評価結果を示す.その後,4.3.3節にお いて,予測を行った結果について評価を行う.

### 4.3.1 データセット

本研究では,札幌市を中心とする範囲 340km<sup>2</sup>の気象レーダ画像を対象とし,約6冬分の期間 (92~93年,93~94年,95~96年,96~97年,97~98年,98~99年)より,降水のあった期間 のみを切り出した計約9000時間,約120000フレームからなるデータセットを構築した.画像サ イズは340×340 画素であり,1 画素が1km<sup>2</sup> に対応している.観測間隔は5分である.

実験にあたり,特徴量計算の単位時間ステップは1時間とした.メッシュ特徴のサイズは 6×6 であり,各々のメッシュについて 20 フレームの時間平均された濃淡値を特徴量とした.特徴空間の次元は,固有空間構成の際の寄与率が 90%となるようにメッシュ特徴,速度ベクトル場,テンポラルテクスチャ特徴をそれぞれ 8,30,4 とした.以後,特に断りの無い場合,部分系列の長さは L=3とし,式(4.8)中の閾値は  $\xi=L$ ,式(4.11)中における r=3, $E_{TH}=\infty$ とした.なお,これらのパラメータは経験的に決定した.予測精度の評価は約 9000時間ステップのデータから 800時間ステップをランダムに選択し,一つ抜き法(Leave-One-Out 法)で評価を行った.なお,その際には,入力系列の直前 3 時間,及び,目標予測時刻から 24 時間未満にかかる部分系列は検索されないように制限を加えた.

### 4.3.2 類似パターン検索に関する実験と評価

記憶に基づく予測法が有効である前提条件として,十分に類似するパターンが検索できること が求められる.本節では,検索に関して実験を行った結果を示す.

#### 検索の様子

まず,検索の一例として,図4.10,及び,図4.11に,クエリの系列,及び,検索された系列の 上位3個をそれぞれ示す.図4.10のクエリ系列では,太い線状のパターンが南東方向に移動し, 陸地に差し掛かろうとしている.それに対して,図4.11にて示す検索結果の系列でも,同じよう な分布範囲,及び,移動の傾向をもったパターンが検索された.これらの検索結果は,視覚的に 類似しているばかりでなく,気象学的にみても類似した現象であることが確認されている.この ような検索結果は,人間の予報士に対しても,現象の理解や将来の気象パターンを予測する上で, 有用な参考情報となりうると考える.

#### 検索誤差と検索拒否率

類似パターン検索においては,検索誤差が小さく,かつ,検索に失敗する割合が小さいことが 望まれる.本研究では,与えたパラメータ設定において,検索結果の個数がゼロ個の場合を検索 の失敗と見なし,複数回の試行における失敗の割合を検索拒否率と呼ぶことにする.一般に検索



図 4.10: クエリ系列 . (a) 開始フレームの画像 , (b) 終了フレームの画像 .



図 4.11: 検索された系列 . (a) 第 1 位 , (b) 第 2 位 , (c) 第 3 位 . [(左) 開始フレーム , (右) 終了フレーム] .



図 4.12: 平均検索誤差と拒否率の関係を表す図.横軸:平均検索誤差,縦軸:検索拒否率.様々に 検索ランク K(=K'),及び,閾値  $E_{TH}$ の設定を変えた場合の変化を図示.

	A	В	С	D
イベントの数	5	4	5	5

43

15

19

34

合計の長さ「時間」

表 4.1: 検索されたパターンの類似性を検証するためのデータセット.

誤差の閾値(式(4.11)の $E_{TH}$ )を小さくすると検索誤差の平均値も小さくなるが,その反面,閾値を満たす類似パターンが存在せず,検索に失敗する割合が増加する.このように検索誤差と拒否率はトレードオフの関係にある.図4.12は,この関係を実験結果に基づきプロットしたものである.具体的には,検索のパラメータとして,上位何位までの検索結果を要求するかという個数K(=K'),及び,検索誤差の閾値 $E_{TH}$ の2つのパラメータを変化させた場合の,平均検索誤差と検索拒否率の値の関係を示す.図4.12より,平均検索誤差と検索拒否率の間に一種のトレードオフがあることが確認できる.また,図4.12を用いると,許容できる拒否率と必要とする検索結果の個数を与えた際に,検索誤差の大きさを見積もることができ,検索システムを運用する場合の参考となると考えられる.

検索されたパターンの現象としての類似性

画像特徴の選択の妥当性,及び,検索結果の精度を検証するために,画像パターン系列の元に なっている現象の観点から一つの評価を試みる.ここでは,気象予報士によって予め4つの気象 現象のクラスに分類されたエコーパターンのデータセットを用いる.気象現象のクラスとしては, 図4.13のような,A)NW-SE 筋状,B)W-E 筋状,C)SW-NE 層状,D)NW-SE 散在の4つのクラ



図 4.13: 気象予報士によって分類されたエコーパターンの例 . (a)NW-SE 筋状 , (b)W-E 筋状, (c)SW-NE 層状 , (d)NW-SE 散在 .

スが設定された.ここで,NW-SE などの記号は,パターンの流れる方角を表す.なお,画像の上 方が北に対応している.A)NW-SE 筋状は,北西から南東に向かって複数の平行な筋をなすよう なパターンである.B)W-E 筋状は,西から東の方向に向かう筋状のパターンである.C)SW-NE 層状は,大きな拡散状の表面をもつパターンであり,南西から北東の方向に向かって流れつつ,急 速に出現,消滅を繰り返す.D)NW-SE は,複数のエコーセルがランダムに分布し,変形を繰り 返すものであり,その基本的な流れは北西から南東の方向であるようなパターンである.表4.1 に は,本実験で用いるデータセットに含まれる,A,B,C,Dの各クラスに属するイベントの個数, 及び,各クラスに含まれる系列の長さの合計時間を示す.ここで,イベントとは,同一の現象に 属するエコーパターンが出現してから消滅するまで一連の過程のことを指す.一つのイベントの 持続時間は,およそ5時間から15時間であった.

上記のデータセットの各時刻をクエリとして,クエリ系列を除くデータセット全体から検索を 行い,検索された系列がクエリと同じカテゴリーに属するかどうか検証を行った.検索された上 位 K の中で,クエリと同じカテゴリーに属するものの割合をヒット率と呼び計算を行った.ここ では,K=5とし,また,閾値 E<sub>TH</sub> は使用しなかった.表4.2には,検索ヒット率の平均値を各 パターンのクラス毎,及び,使用する特徴の組み合わせについて示す.ここで,特徴の組み合わせ は,i)メッシュ特徴のみ,ii)メッシュ特徴と速度ベクトル場のみ,iii)全部(メッシュ特徴,速度

Features	1	1,2	$1,\!2,\!3$
All Classes	78.0	83.5	86.0
Class A)	81.6	82.6	81.6
Class B)	56.2	63.9	71.4
Class C)	75.8	100.0	100.0
Class D)	86.2	87.6	93.4

表 4.2: 平均のヒット率 [%].1:メッシュ特徴,2:速度場特徴,3:テンポラルテクスチャ特徴.



図 4.14: メッシュ特徴に関する平均予測誤差  $\overline{\epsilon_T}$ ,及び,その標準偏差を検索の個数 K(=K') に対してプロットした図.ただし,閾値  $E_{TH}$  は未使用.

ベクトル場, テンポラルテクスチャ特徴)とした.なお,ここでは,テンポラルテクスチャ特徴として, $f_1 \sim f_7$ までを使用したが,その選択の妥当性については,著者らの以前の論文で確認されている [157].表4.2からは,クラスA)とB)において比較的,ヒット率が低いものの,これを除くと良好な検索が行われていることがわかる.このクラスA)とB)は,ともに筋状のタイプに属しており,動きの方向による区分される.そのため,両者の境での曖昧性が誤りの一因であると考えられる.また,表4.2からは,3種すべての特徴を使用することが有効であると示唆される.

## 4.3.3 予測に関する実験と評価

本節では,予測に関する実験として,まず,特徴空間における予測誤差を評価した結果を示し, 次に予測画像の合成の様子と予測精度の評価結果を記載する.



図 4.15: CSI 尺度.

特徴空間における予測誤差

メッシュ特徴空間における予測精度を評価した.ここで,メッシュ特徴を用いる理由は,この 特徴が大局的な降水分布を表すと考えられるからである.ある時刻Tを基準とした検索・予測の 試行において,一定時間 $T + \eta$ 後のメッシュ特徴 $y_1(T + \eta)$ と予測されたメッシュ特徴 $\hat{y}_1(T + \eta)$ に基づいて,予測誤差を

$$\epsilon_T = \frac{|\widehat{\boldsymbol{y}}_1(T+\eta) - \boldsymbol{y}_1(T+\eta)|}{|\boldsymbol{y}_1(T) - \boldsymbol{y}_1(T+\eta)|}$$
(4.15)

のように計算した.ここで分母は,持続予報(クエリのパターンをそのまま予測とする)の場合 の予測誤差を表し,これにより正規化することで分布の絶対的な大きさの影響が除外される.図 4.14 には,検索された系列の個数K(=K')の関数として,平均の予測誤差 $\epsilon_T$  とその標準偏差を 示す.なお,検索の閾値 $E_{TH}$ は使用していない.図4.14 から,検索のランク数が増えるに従い 平均予測誤差が減少し,K=10程度で飽和するという傾向が伺える.また,グラフの縦軸に対し て,標準偏差は相当大きく,各試行ごとに検索・予測の振る舞いが大きく違うことが示唆される. また,このような傾向は,類似のパターンが過去に何度も生起している場合と,そうでない場合 の両方のケースが混在していることを示唆するものとも捉えられる.なお,予測誤差が1を上回 る場合,これは,持続予測の誤差を上回る事から,予測情報として価値は持たないといえる.図 4.14 のK(=K')=25の場合,全体の試行のうち,60.8%は,平均予測誤差が1.0を下回っており, この場合には有効な予測情報を提供できる可能性があるといえる.

予測降水分布の評価

予測画像として,η-1時間後からη時間後までの累積降水量の分布を合成した.ここでは,対 象となる時間区間に含まれるレーダ画像の各フレームの濃淡値をレーダ方程式[69]を用いて降水量 へ変換し,その総和をとった分布を予測画像とした.予測画像の精度の評価尺度には CSI(Critical Success Index) と呼ばれる尺度を用いた.これは降水の有無に関して予測が的中した面積の比率 に関する尺度であり,

## $CSI = \frac{$ 予測的中面積 予測的中面積 + 見逃し面積 + 空振り面積 × 100 (4.16)

のように定義されている.ここで,予測的中面積,見逃し面積,空振り面積は,図4.15の各領域 の面積に相当する.ここでは降水の有無の閾値を0.1 [mm/h] とした.なお,精度の比較のために 持続予報と線形外挿手法についても同一の試行ステップについて評価を行った.持続予報とは現 況の分布を予測分布とする手法であり,予測情報としての価値はないが,この CSI 値を上回るこ とが,実用化上の一つの目安となっている.また,線形外挿手法は,入力画像系列上において動 きベクトルを計算し,現況の分布をその動きベクトルにより平行移動させ予測画像を生成する手 法であり,代表的な従来手法の一つである.

予測パターンの例 図 4.16 には,本手法で予測された降水分布の例 (a) を実際の分布 (b),及び, 持続予報 (c) の分布と併せて示す.予測先行時間は 2~3 時間後 (η=3) とし, K=25 とした.CSI は図 4.16 の見出しに示した通りである.線形外挿手法の CSI は Case 1) より 24.2, 24.0, 1.0[%] と なった.Case 1) は,図 4.16(1-c) のような太いバンド状のパターンが,同 (1-b) のような筋状を伴 う広がったパターンに変化する事例であるが,本手法では詳細な筋の模様までは再現できていな いものの,パターンの広がりは予測できていることがわかる.また,Case 2) および Case 3) は, 右上の方向へ向って流れる層状パターンの例であり,観測時点では存在しない領域が新たに出現 している例である.従来の線形外挿手法など直前の情報のみに基づく手法では,このような新た に出現する領域を予測できないが,本手法では過去のパターンの履歴情報を活用することで,こ のようなパターンに対しても,類似パターンが存在する場合には適切な予測画像が生成できる可 能性を有することがわかる.

検索系列数  $K \ge 3$ 測精度の関係 図4.17には,指定された検索画像系列数の上限  $K \ge 1,3,5,7,10,15,20,25$ と変化させたときの,800 ステップの試行結果の CSI の平均値を示す.予測先行時間は 2~3 時間 後 ( $\eta$ =3) とした.類似系列が存在しないと見なされた試行の割合である検索拒否率は 1.75%であ り,また,K=25 のときの実際に検索できた系列の数 K'の期待値は 21.2 と良好な結果を示した. 図 4.17 より K の増加とともに CSI 値が向上し,K=15 付近において持続予報を若干だが上回る ようになり,K=20 付近で飽和する傾向が観測できる.また,この Kの増加に対する CSI 値の増 加は,特徴空間中における予測誤差の傾向とほぼ対応する.検索された画像の類似性は,検索順 位が後退するものほど低くなるが,同時に検索誤差も大きくなるため,その画像への重み付けが 減少し,上記の飽和傾向が現れると考えられる.これらの結果より,複数の画像の合成を用いた 予測画像生成方式の有効性が示される.また,図 4.17 中に示した線形外挿手法の結果と比較して も本方式の優位性が確認できる.



図 4.16: 累積降水分布 (0.1[mm/h] 以上), (a) 本手法による 2~3 時間後の予測分布 (上より CSI=44.5, 72.9, 41.9[%]), (b) 2~3 時間後の実際の分布, (c) 持続予報の分布 (上より CSI=16.5, 35.9, 3.0[%]).



図 4.17: 指定された検索系列数 K に対する平均 CSI 値 .  $\eta = 3$  .



図 4.18:予測先行時間に対する平均 CSI 値 ( $\eta$ -1 $\sim \eta$ 時間後の累積降水量分布).検索系列数 K=25.



図 4.19: 記憶データ量 (全データ量に対する比率) と平均 CSI 値の関係.

予測先行時間と予測精度の関係 図4.18 には,予測先行時間を変化させたときの平均 CSI 値の変 化を持続予報の結果と照し合せて示す.なお,K=25 とした.図4.18 より,予測先行時間の増加 とともに,本手法,持続予報とともに減少する傾向を示すが,本手法の CSI の低下の傾向は緩や かであり,η=3 において精度が逆転している.なお,-1~0 時間後の結果は,検索された画像系列 そのもので合成したパターンと実際のパターンとを比較したもので,この精度が本手法における 予測精度の上限を意味する.持続予報とは異なり,本手法はこの時点においても現況の画像との 間に100%の一致はみられない.この点は本方式の欠点であるが,逆に,より長時間先の予測につ いては本手法の優位性がうかがえる.

データ量の増減による予測精度の変化 記憶に基づいた予測手法において,入力系列に対して類 似する過去のデータが存在しない場合には,検索結果が得られない,若しくは,予測精度は著し く低下するという本質的な欠点がある.ここでは,記憶データ量が予測精度に及ぼす影響を調査 するため,約9000ステップの全データについて,時間的に古いものから1/8,1/4,1/2,3/4,1/1 の区間を抜き取り,それぞれについて,先述の800ステップを入力としたときのCSI値の評価を 行った.なお,予測先行時間は2~3時間後(η=3)とし,K=25とした.図4.19には,各データ量 における平均CSI値,および,持続予報の平均CSI値を示す.記憶データ量の増大とともにCSI 値が増加しているが,これは記憶データ量が増大するに従い,より類似するパターンが予測画像 生成に利用できるようになったためであると考えられる.また,記憶データ量1/8を除き,CSI値 はほぼ直線的に増加していることから,今後さらに記憶データ量を増やすことにより,より一層 の予測精度の向上が期待できる.また,検索拒否率はデータ量が増加するに従い,低下する傾向 が確認できた.

## 4.4 結言

本章では,観測の部分性,間接性がともに顕著な事例として,非剛体・不定形の時系列画像パ ターンとして観測される気象現象の予測の問題を取り上げた.具体的には,気象レーダ画像上の パターンの時間変化を予測することを目標とし,過去の画像の記憶に基づく画像パターンの予測 方法を提案した.これは,現在の画像系列に類似する過去の画像系列をデータベースより検索し, 検索された画像中のパターンの時間発展に基づいて予測画像を生成する方法である.北海道の約6 冬期分のレーダ画像を用いた実験を行い,検索精度,及び,予測精度の評価結果から提案方法の 有効性を確認した.

複数の検索された画像の合成による予測画像生成方式の有効性を確認した.

今後の課題として,まず,本論文では検討しなかった最適な K の選択や,与えられた予測先行時間に対して最適な部分系列の長さ L を決定する方法の検討などがあげられる.また,記憶に基

づく予測法において,特徴量の選択,及び,統合時の重み付けの方法は,性能に対する影響が大 きい.本論文では,気象レーダ画像パターンの性質を鑑みて,メッシュ特徴,速度ベクトル場,テ ンポラルテクスチャ特徴の3種類の特徴量を導入し,それらの均等な重み付けに基づいて検索を 行った.今後は,検索・予測性能の更なる向上のために,本論文で用いた特徴量の他,他の画像特 徴や,それらの重み付けの方法の検討が課題となる.さらに,本提案手法では,観測情報を気象 レーダ画像に限定しており,画像の類似性のみに基づいた予測を行っている.しかしながら,目 的は気象現象の予測にあるため,大気の内部状態をより直接的に捉えることのできる気温,気圧, 湿度,風速・風向などの物理データを用いた記憶に基づく予測や,物理データと画像特徴との併 用も重要な検討課題としてあげられる.さらに,提案手法を実際の予報現場に導入する際のユー ザインタフェースや,ユーザからのフィードバックによる精度向上など様々な研究課題が考えら れる.

## 第5章 結論

本論文は,時系列画像に基づく動的事象の推測に関する著者の研究成果をまとめたものである. 以下では,各章で述べた成果を要約した後,それらを総括して研究の今後の課題・展望について 論じる.

まず,第1章では,時系列画像に基づく動的事象の推測に関する研究の背景を述べ,観測の部 分性,間接性の問題に対して十分な対処法が確立されていない現状を指摘した.次に,それを踏 まえて,本論文では,時系列画像解析の適用範囲を拡大することを目指して,観測の部分性,間 接性のそれぞれの度合いに応じた3つの事例を取り上げ,各々新しいアプローチを提示し,論文 の構成を述べた.

第2章では、観測の部分性が顕著な例として、複数の物体の間に生じるオクルージョンに着目 した.ここでは、物体間のオクルージョンに対してロバストに複数の運動物体の位置・姿勢を追 跡する方法として、対象間のオクルージョンを含めた観測過程を陽にモデル化し、オクルージョ ンの構造と対象の状態(位置や姿勢)を同時に推定する方法を提案した.この方法では、どの物 体が見えていて、どの物体が隠されているかというオクルージョンの空間構造が2次元世界上の 複数視点における視角と物体との接触関係により陽に記述される.また、追跡の問題をオクルー ジョンの空間構造についての多重仮説の生成・検証、及び、物体状態の事後確率分布の推定とい う2つの問題からなる再帰的ベイズ推定法として定式化した.この方法は、オクルージョンに起 因する物体の配置や状態の不確定性を推定することができるため、多様にオクルージョンが変化 する状況においても、安定した物体追跡を行うことができる.人工データ、及び、実データを用い た実験を行い、その結果、一時的に全ての視点においてオクルージョンが発生している状況にお いても、頑健に追跡が継続できることを確認した.

次に,第3章では,観測の間接性が顕著である例として,事象を構成する複数の要素間の相互 作用に着目した.その中でも本論文では,直接観測はできないが,相互作用が事象のダイナミク スに対して決定的な影響をもつ事例として,複数人物の対面会話シーンを取り上げた.対面会話 シーンにおいては,各人物の動作は時系列画像として比較的容易に観測できるものの,それら動 作を支配するダイナミクスは,人物間の相互作用(インタラクションとも呼ぶ)に依存し,非常 に複雑である.本研究では,人物間の相互作用について陽にモデル化を行い,各人物の動作から 人物間の相互作用の状態を推定する新しい方法を提案した.具体的には,相互作用として,話し 掛け,及び,傾聴の動作に着目し,誰が誰に話し掛けているか?誰が誰の話を聞いているか?と

115

いう人物間でのメッセージの流れのパターンを推定対象とした.本研究では,これを会話構造と 呼び,会話構造と各参加者の行動との関連を動的ベイジアンネットワークを用いてモデル化した. このモデルは,会話構造に依存して確率的に参加者の行動(視線行動,及び,発話)が生成され, さらに各人の視線方向に応じて頭部方向が観測されるという階層構造を有する.また,ギブスサ ンプリングと呼ばれる一種のマルコフ連鎖モンテカルロ法を用いて,観測された頭部方向と発話 の有無の時系列データより,会話構造,視線パターン,及び,モデルパラメータのベイズ推定を 行う方法を提案した.最後に,4人会話を対象とした実験により,視線方向と会話構造の推定精度 を評価し,提案した方法の有効性を確認した.

さらに,第4章では,観測の部分性と間接性がともに顕著な事例として,画像パターンとして観 測される自然現象の予測の問題を取り上げた.本研究では,自然現象の中でも気象現象,その中 でも特に降水現象に着目し、その観測手段として気象レーダを用いた、気象レーダは、降水強度 の空間分布を画像パターンとして観測する装置である.そのため,これを用いることで降水の空 間分布の予測を行う問題は、気象レーダ画像上のパターンの時間変化を予測する問題として捉え ることができる.しかし,このパターンは,非剛体の不定形パターンであり,様々な気象的要因が 重なり合った複雑なダイナミクスを持つため,その予測は容易でない.気象レーダによる観測は, パターンの元となる降水現象の空間的な広がりに対して,レーダから一定範囲内,かつ,ある空 間断面上に限定された観測であり、降水現象の全容を捉えることができない、その観点から観測 には部分性が伴うといえる.また,気象レーダでは,気温,気圧,風速といった気象現象の内部 状態が直接的に観測できず,さらに,現時点の観測情報から未来のパターンを予測するという問 題の性質からも,観測には間接性が伴うと考えられる.これらの観点より,気象レーダ画像上の パターンの予測の問題は,観測の部分性,間接性がともに顕著な事例といえる.本論文では,こ のような複雑な時系列パターンの予測の方法として、過去の画像パターンの記憶に基づく予測法 を提案した.これは,現在の画像系列に類似する過去の画像系列をデータベースより検索し,検 索された画像中のパターンの時間発展に基づいて予測画像を生成するという方法である.この方 法の要素技術として,本論文では,時系列の気象レーダ画像より抽出された画像特徴に基づいて 類似画像系列を検索する手法,及び,検索された画像系列の未来の画像を合成することで予測画 像を生成する方法を提案した.北海道の約6冬期分のデータセットを用いた実験を行い,予測先 行時間,データ量の観点から予測精度を評価し,提案方法の有効性を確認した.

以上のように本論文では,従来の時系列画像解析において未だ解決法が確立されていない観測 の部分性,間接性の問題に着目し,それらの度合いに応じた3つの事例を取り上げ,それぞれ新 しいアプローチを提案した.具体的には,未解決課題の中でも特に重要なものとして,オクルー ジョンに頑健な物体追跡,複数人物の対面会話シーンの分析,気象レーダ画像パターンの予測の3 つの問題を対象とした.これら3つの事例それぞれについて提案した方法は,個別の問題設定に 特化したものであるが,その根底にあるアプローチは幅広い事象に適用できるものと考えられる. 以下では,その観点から各章で提案した方法を振り返る.

まず、第2章では、対象間の相互オクルージョンへの対処を主眼として、オクルージョンを含む 観測過程を陽にモデル化し,オクルージョンの空間構造を物体の状態と同時に推測するというア プローチを提案した.第3章では,対象間の相互作用の推測を主眼として,観測される人物行動と 人物間の相互作用の間にある関係をモデル化するというアプローチを提案した.さらに第4章で は, 第2章, 第3章で行ったような観測過程や相互作用を陽にモデル化する代わりに, 過去の記 憶に基づく事象のモデル化を提案した.このように本論文では,i) 観測過程のモデル化,ii) 相互 (作用のモデル化, iii) 記憶に基づくモデル化の3つのアプローチを提案したが, これらのアプロー チは、第2章~第4章の各事象に限定されるものではなく、それらのアプローチを適切に選択し、 または、組み合わることで、様々な事象に対して有効な時系列画像解析法を構成することが可能 である.例えば,アプローチi,iiは,事象に関する事前知識が得られ,モデル構築が可能な場合に 適する.一方,アプローチ iii は,事象に関する事前知識が少ない状況で,かつ,過去のデータが 大量に得られる場合に適する、今後は、対象事象やその観測手段に応じて最適なアプローチを選 択するための方法論の確立が望まれる.また,複数のアプローチの併用・組み合わせも有効な手 立てであると考えられる.例えば,本論文の第2章,及び,第3章で対象とした,複数人物の移動 シーンや会話シーンは多様な人間行動の一部分に過ぎず,現実社会ではそれらが混在しているの が普通である.例えば,人物の移動と会話だけを取り上げた場合でも,立ち話のように人物の移 動や集合・離散と会話の状態とが密接に関連する場面は多い.そのような状況においては,人物 間のオクルージョンが頻発し、かつ、人物間の相互作用も複雑に変化することが想定される、こ のような場合、本論文で提案したアプローチiとiiを統合して推測を行うことが有効であると思わ れる.また,事象に関する事前知識が少なく,さらに,過去のデータも少ないような状況に対し ては,アプローチ i, ii とアプローチ iii を融合した方法が求められる.これに加えて,現象に応じ て様々なセンサー情報と画像情報との統合を行うことも、観測の部分性、間接性を減じるために は重要な検討課題であると思われる。

研究分野としてコンピュータビジョンは,既に四半世紀を超える歴史をもつが,人間のもつような高度な視覚機能を人工的に実現するという観点からは,未だに未成熟な分野であるといえる.しかしながら,21世紀の社会情報基盤として,高度道路交通システム(ITS),知能ロボット,マン・マシン・インタフェース,マルチメディアなど様々な場面で利用が期待されている.こうした社会の要請により,従来の制御された環境から,実世界の環境へとコンピュータビジョンの適用範囲を拡大することが急務となっている.また,歴史的にコンピュータビジョンの主眼となっている物理的な対象の構造や運動の推定のみならず,複数の対象の相互作用により構成される実世界の事象の推測へと対象領域の拡大を推し進める時期に差し掛かっているといえる.こうした背景から本研究では,実世界事象のうち有益なアプリケーションに結びつくと考えられる,人間の歩行シーン,人間同士のコミュニケーション,気象現象という3つの事象に的を絞り,従来の時

系列画像解析技術では対応が困難であった観測の部分性,間接性に対処する新しいアプローチを 提案した.しかし,これらは,実世界の多岐にわたる事象のうち,ほんの一部分に過ぎず,異な る事象ごとに試行錯誤で解法を構築するという域を脱していないともいえる.その観点から,研 究は入り口に差し掛かったばかりであり,幅広い実世界事象の推測に対する統一的な方法論を構 築するには機が熟していないのが現状である.しかし,本研究で提案した幾つかのアプローチは, 既存の時系列画像解析の問題点に対する突破口となり得る方法論であり,今後,多様な場面に展 開できるものと期待される.このように,本研究は,今後,実世界事象の推測に関する研究を促 進する上での一つの足掛かりを提供したといえ,その点で,貴重な成果であると考える.

## 謝辞

本論文をまとめるにあたり,名古屋大学大学院 情報科学研究科 村瀬 洋 教授には,終始懇切な 御指導と御教示を賜りました.ここに謹んで深謝いたします.また,論文をまとめるに際し,同 研究科 末永康仁 教授,大西 昇 教授,井手一郎 助教授,三輪和久 教授,中京大学 目加田 慶人 教授には,有益な御討論,御助言を賜りました.ここに記し,深く感謝いたします.

本研究は,著者が日本電信電話株式会社 NTT ヒューマンインタフェース研究所に在籍中に取 り組み始めたものであります.それ以降,同じく NTT 横須賀 R&D センタを拠点とする NTT サ イバースペース研究所,NTT サイバーソリューション研究所の在籍期間にわたり,本研究の機会 を与えて頂き,また,多大なる御助言・御指導を賜りました NTT ヒューマンインタフェース研究 所 元所長 加藤邦紘氏(現在,独立行政法人 情報通信研究機構 理事),同研究所 元所長 岡田和 |比古氏(現在 , エヌ・ティ・ティ アイティ株式会社 代表取締役社長 ) , NTT サイバースペース研 究所 元所長 木暮賢司 博士 ( 現在 , NTT アドバンステクノロジ株式会社 取締役 メディア事業本 部長), NTT サイバーソリューション研究所 元所長 安原隆一氏(現在,NTT アドバンステクノ ロジ株式会社 取締役 システムソリューション事業本部長), NTT ヒューマンインタフェース研 究所 映像処理研究部 元部長 徳永幸生 博士 (現在,芝浦工業大学 教授),同研究部 映像インタ ラクティブ研究グループリーダー , NTT サイバーソリューション研究所 コンテンツ流通プロジェ クトマネージャーを歴任された曽根原 登 博士 (現在,国立情報学研究所教授),NTT ヒューマ ンインタフェース研究所 映像処理研究部 映像学習研究グループ 元リーダー 鈴木 智 博士 (現在, NTT システムインテグレーション基盤研究所 ビジネスプロモーション SE プロジェクト プロジェ クトマネージャ),同グループ 元主任研究員 堀越 力 博士(現在,株式会社 NTT ドコモ マルチ メディア研究所 映像信号処理研究室 室長), NTT サイバーソリューション研究所 コンテンツ流 通プロジェクト デジタルアーカイブグループ 元リーダー児島治彦 氏 (現在 , NTT ユニバーサ ル IT デザインセンタ・センタ長),同グループ元 主任研究員 森本 正志 博士(現在,NTT サイ バーソリューション研究所 メディアコンピューティングプロジェクト 主幹研究員)に深く感謝い たします.

また,その後,NTT コミュニケーション科学基礎研究所においても,本研究を継続する機会を 与えて頂き,多大なる御助言・御指導を賜りました,NTT コミュニケーション科学基礎研究所元 所長 石井健一郎 博士(現在,名古屋大学教授),同研究所元所長 管村 昇 博士(現在,工学院 大学教授),同研究所前所長 片桐 滋 博士(現在,NTT R&D フェロー,同志社大学教授),同 研究所 現所長 外村佳伸 博士,同研究所 メディア情報研究部 元部長 萩田紀博 博士(現在,ATR 知能ロボティクス研究所 所長),同研究部 部長 牧野昭二 博士,同研究部 環境理解研究グループ 元リーダ 武川直樹 博士(現在,東京電機大学 教授),同研究部 メディア認識研究グループリー ダ 大和淳司 博士に深く感謝いたします.

また,本研究の途上において種々の御討論,ご協力を頂きました NTT ヒューマンインタフェー ス研究所 映像処理研究部 諸氏, NTT サイバースペース研究所 コンテンツ生成プロジェクト 諸 氏,NTT サイバーソリューション研究所 コンテンツ流通プロジェクト 諸氏,NTT コミュニケー ション科学基礎研究所 メディア情報研究部 元研究主任 大野健彦 氏(現在,日本電信電話株式会 社 第2部門 担当課長),竹前嘉修 氏(現 NTT サイバーソリューション研究所 ヒューマンアプ ライアンスプロジェクト),他 メディア情報研究部 諸氏,名古屋大学大学院 情報科学研究科 村 瀬研究室 諸氏に厚く御礼申し上げます.

また,学会や研究会において多くの先生,研究者の方々に様々なご助言,ご指導を頂きました ことも心から感謝申し上げます.

# 参考文献

- [1] R. L. グレゴリー (近藤倫明, 訳): "脳と視覚 グレゴリーの視覚心理学 –", ブレーン出版 (2001).
- [2] J. J. ギブソン: "生態学的視覚論 ヒトの知覚世界を探る –", サイエンス社 (1986).
- [3] D. マー:"ビジョン -視覚の計算理論と脳内表現 -", 産業図書 (1987).
- [4] L. G. Shapiro: "Computer Vision", Prentice Hall (2001).
- [5] 高木幹雄, 鳥脇純一郎, 田村秀行 ( 編 ): "画像処理アルゴリズムの最新動向", 新技術コミュ ニケーションズ (1986).
- [6] 松山隆司, 久野善徳, 井宮淳(編): "コンピュータビジョン: 技術評論と将来展望", 新技術 コミュニケーションズ (1998).
- [7] 金出武雄: "ロボット視覚", 情報処理, 44, 11, pp. 1130–1137 (2003).
- [8] 加藤丈和, 深尾隆則, 羽下哲司: "対象追跡 フレーム間の類似性に着目した手法から動きの モデルに着目した手法まで –", 情報処理学会 研究報告 コンピュータビジョンとイメージメ ディア CVIM-150, pp. 185–198 (2005).
- [9] 川嶋宏彰, 西村拓一: "コンピュータビジョンにおける時系列パターン認識", 情報処理学会 研究報告 コンピュータビジョンとイメージメディア CVIM-154, pp. 197-209 (2006).
- [10] R. E. Kalman: "A new approach to linear filtering and prediction problems", Journal of Basic Engineering, Transactions ASMA, Series D, 82, pp. 35–45 (1960).
- [11] 有本 卓: "カルマン・フィルター", 産業図書 (1977).
- [12] 片山 徹: "新版応用カルマンフィルタ", 朝倉書店 (2000).
- [13] G. Kitagawa: "Monte Carlo filter and smoother for non-Gaussian nonlinear state space models", Journal of Computational and Graphical Statistics, 5, 1, pp. 1–25 (1996).

- [14] N. J. Gordon, D. J. Salmond and A. F. M. Smith: "Novel approach to nonlinear/non-Gaussian Bayesian state estimation", IEE proceedings F: Communications, Radar and Signal Processing, 140, 2, pp. 107–113 (1993).
- [15] Arnaud Doucet (Ed.): "Sequential Monte Carlo Methods in Practice", Springer-Verlag (2001).
- [16] M. Arulampalam, S. Maskell, N. Gordon and T. Clapp: "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking", IEEE Transactions on Signal Processing, 50, 2, pp. 174–188 (2002).
- [17] B. Ristic: "Beyond the Kalman Filter: Particle Filters for Tracking Applications", Artech House (2004).
- [18] 樋口知之: "粒子フィルタ", 電子情報通信学会誌, 88, 12, pp. 989-994 (2005).
- [19] Y. Bar-Shalom and T. E. Fortmann: "Tracking and Data Association", Academic Press (1987).
- [20] I. J. Cox: "A review of statistical data association techniques for motion correspondence", International Journal of Computer Vision, 10, 1, pp. 52–66 (1993).
- [21] C. Rasmussen and G. D. Hager: "Probabilistic data association methods for tracking complex visual objects", IEEE Transactions on Pattern Analysis and Machine Intelligence, 23, 6 (2001).
- [22] D. B. Reid: "An algorithm for tracking multiple target", IEEE Transactions on Automatic Control, AC-24, 6, pp. 843–854 (1979).
- [23] D. Koller, J. Weber and J. Malik: "Robust multiple car tracking with occlusion reasoning", Proceedings of European Conference on Computer Vision (ECCV'94), pp. 189–196 (1994).
- [24] J. MacCormick and A. Blake: "A probabilistic exclusion principle for tracking multiple objects", International Journal of Computer Vision, 39, 1, pp. 57–71 (2000).
- [25] I. Haritaoglu, D. Harwood and L. S. Davis: "W<sup>4</sup>:real-time surveillance of people and their activities", IEEE Transactions on Pattern Analysis and Machine Intelligence, 22, 8, pp. 809–830 (2000).
- [26] I. K. Sethi and R. Jain: "Finding trajectories of feature points in a monocular image sequence", IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-9, 1, pp. 56–73 (1987).

- [27] 十河 卓, 石黒 浩, モーハン M. トリベディ: "複数の全方位視覚センサによる実時間人物追跡", 電子情報通信学会論文誌 D-II, J83-D-II, 12, pp. 2567–2577 (2000).
- [28] 森 大樹, 内海 章, 大谷 淳, 谷内田正彦, 中津良平: "非同期多視点画像による人物追跡システムの構築", 電子情報通信学会論文誌 D-II, J84-D-II, 1, pp. 102–110 (2001).
- [29] 早坂光晴, 富永英義, 小宮一三: "逆投影法とカルマンフィルタを用いた複数移動物体位置認
   識とその追跡", 電子情報通信学会技術報告, PRMU2001-132, pp. 133–138 (2001).
- [30] 中島 平, 浜崎浩二, 岡谷貴之, 出口光一郎: "Condensation を用いた多視点画像の融合による 複数人物の追跡", 画像の認識・理解シンポジウム (MIRU'02) 論文集, pp. II-317-322 (2002).
- [31] S. L. Dockstader and A. M. Tekalp: "Multiple camera tracking of interacting and occluded human motion", Proceedings of the IEEE, 89, 10, pp. 1441–145 (2001).
- [32] A. Mittal and L. S. Davis: "M2tracker: A multi-view approach to segmenting and tracking people in a cluttered scene using region-based stereo", Proceedings of European Conference on Computer Vision (ECCV'02), pp. 18–32 (2002).
- [33] 中澤篤志,日浦慎作,加藤博一,井口征士: "分散視覚エージェントを用いた複数人物追跡シ ステム",情報処理学会論文誌,42,11,pp. 2699-2710 (2001).
- [34] 間瀬健二, 末永康仁: "顔画像の動き検出の一手法", 情報処理学会第 30 回全国大会講演論文 集, pp. 1247–1248 (1985).
- [35] 末永康仁,間瀬健二: "手の動きの映像処理による情報入力法の検討",情報処理学会第 30 回 全国大会講演論文集, pp. 1249–1250 (1985).
- [36] 間瀬健二,渡部保日児,末永康仁: "ヘッドリーダ:画像による頭部動作の実時間検出",電子 情報通信学会論文誌 D, **J74-D2**, 3 (1991).
- [37] 末永康仁, 間瀬健二, 福本雅朗, 渡部保日児: "Human Reader: 人物像と音声による知的インタフェース",電子情報通信学会論文誌 D, **J75-D2**, 2 (1992).
- [38] J. Xiao, S. Baker, I. Matthews and T. Kanade: "Real-time combined 2D+3D active appearance models", Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'04) (2004).
- [39] A. Nikolaidis and I. Pitas: "Facial feature extraction and pose determination", Pattern Recognition, 33, pp. 1783–1791 (2000).

- [40] J. Sherrah, S. Gong and E. J. Ong: "Face distributions in similarity space under varying head pose", Image and Vision Computing, 19, pp. 807–819 (2001).
- [41] R. Gross, I. Matthews and S. Baker: "Generic vs. person specific active appearance models", Image and Vision Computing, 23, pp. 1080–1093 (2005).
- [42] J. K. Aggarwal and Q. Cai: "Human motion analysis: A review", IEEE Nonrigid and Articulated Motion Workshop (1997).
- [43] V. I. Pavlovic, R. Sharma and T. S. Huang: "Visual interpretation of hand gestures for human-computer interaction: A review", IEEE Transactions on Pattern Analysis and Machine Intelligence, 19, 7, pp. 677–695 (1997).
- [44] D. M. Gavrila: "The visual analysis of human movement: A survey", Computer Vision and Image Understanding, 73, 1, pp. 82–98 (1999).
- [45] L. R. Rabiner: "A tutorial on hidden Markov models and selected applications in speech recognition", Proceedings of the IEEE, 77, 2, pp. 257–286 (1989).
- [46] L. R. Rabiner, B. Juang, (古井貞熙訳): "音声認識の基礎 (上)(下)", NTT アドバンステ クノロジ (1995).
- [47] 大和淳司,大谷 淳,石井健一郎: "隠れマルコフモデルを用いた動画像からの人物の行動認 識",電子情報通信学会論文誌 D-II, J76-D2, 12, pp. 2556-2563 (1993).
- [48] 間瀬健二: "動画像処理を用いた新しいマンマシンインタフェースの研究",名古屋大学博士 学位論文 (1992).
- [49] Y. Zhang and Q. Ji: "Active and dynamic information fusion for facial expression understanding from image sequences", IEEE Transactions on Pattern Analysis and Machine Intelligence, 27, 5, pp. 699–714 (2005).
- [50] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman and T. J. Sejnowski: "Classifying facial actions", IEEE Transactions on Pattern Analysis and Machine Intelligence, 21, 10, pp. 974–989 (1999).
- [51] Y. Tian, T. Kanade and J. F. Cohn: "Recognizing action units for facial expression analysis", IEEE Transactions on Pattern Analysis and Machine Intelligence, 23, 2, pp. 97–115 (2001).

- [52] M. Pantic and L. J. M. Rothkrantz: "Automatic analysis of facial expressions: The state of the art", IEEE Transactions on Pattern Analysis and Machine Intelligence, 22, 12, pp. 1424–1445 (2000).
- [53] B. Fasel and J. Luettin: "Automatic facial expression analysis: A survey", Pattern Recognition, 36, pp. 259–275 (2003).
- [54] R. El Kaliouby and P. Robinson: "Real-time inference of complex mental states from facial expressions and head gestures", Proceedings of Computer Vision and Pattern Recognition Workshop (CVPRW'04), pp. 154–154 (2004).
- [55] N. M. Oliver, B. Rosario and A. Pentland: "A Bayesian computer vision system for modeling human interactions", IEEE Transactions on Pattern Analysis and Machine Intelligence, 22, 8, pp. 831–843 (2000).
- [56] M.F. ヴァーガス: "非言語コミュニケーション", 新潮社 (1987).
- [57] 黒川隆夫: "ノンバーバルインタフェース", オーム社 (1994).
- [58] M.L. パターソン: "非言語コミュニケーションの基礎理論", 誠信書房 (1995).
- [59] M. Argyle: "Bodily Communication 2nd ed.", Routledge, London and New York (1988).
- [60] ヴァージニア・P. リッチモンド: "非言語行動の心理学 対人関係とコミュニケーション理 解のために –", 北大路書房 (2006).
- [61] 松尾太加志: "コミュニケーションの心理学", ナカニシヤ出版 (1999).
- [62] 小倉義光: "メソ気象の基礎理論", 東京大学出版会 (1997).
- [63] 二宮洸三: "大気現象の階層構造とメソスケール擾乱", 科学, 64, 12, pp. 763-764 (1994).
- [64] E. N. Lorenz: "Deterministic nonperiodic flow", Journal of Atomospheric Sciences, 20, pp. 130–141 (1963).
- [65] 合原一幸: "応用カオス カオスそして複雑系へ挑む", サイエンス社 (1994).
- [66] 立平良三: "天気予報とメソスケール現象", 科学, 64, 12, pp. 799-803 (1994).
- [67] 新田 尚, 住明 正, 伊藤朋之, 野瀬純一 ( 編 ): "気象ハンドブック", 朝倉書店 (2005).
- [68] 小倉義光: "一般気象学", 東京大学出版会 (1984).
- [69] H. Sauvageot: "Radar meteorology", Artech House, Inc. (1992).

- [70] J. Leese, C. Novak and V. Taylor: "The determination of cloud pattern motions from geosynchronous satellite image data", Pattern Recognition, 2, 4, pp. 279–280 (1970).
- [71] E. Smith and D. Phillips: "Automated cloud tracking using precisely aligned digital ATS pictures", IEEE Transactions on Computers, 21, 7, pp. 715–729 (1972).
- [72] T. M. Hamill and T. Nehrkorn: "A short-term cloud forecast scheme using cross correlations", Weather and Forecasting, 8, 4, pp. 401–411 (1993).
- [73] 平木 哲, 伊藤朋之, 佐伯理朗: "予報を支える観測システム", 日経サイエンス, pp. 40-47 (1995).
- [74] 立平良三:"新しい天気予報 確率予報とナウキャスト",東京堂出版 (1986).
- [75] 遊馬芳雄, 菊池勝弘, 今 久: "簡易気象レーダーによるエコーの移動速度について", 北海道 大学地球物理学研究報告, 44, pp. 23-34 (1984).
- [76] 境野英朋, 堀越 力, 安達文夫, 鈴木智: "流体モデルによる時系列レーダパターン変化の予測 方法", 電子情報通信学会論文誌 A, J82-A, 4, pp. 550–562 (1999).
- [77] 椎葉充晴, 高棹琢馬, 中北英一: "移流モデルによる短時間降雨予測手法の検討", 第28回水
   理講演会論文集, pp. 423-428 (1984).
- [78] 浅井富雄, 武田喬男, 木村龍治: "雲や降水を伴う大気", 東京大学出版会 (1981).
- [79] 小倉義光: "対流細胞の増殖と組織化", 科学, 64, 12, pp. 765-778 (1994).
- [80] R. Cutler, Y. Rui, A. Gupta, J. Cadiz, I. Tashev, L. He, A. Colburn, Z. Zhang, Z. Liu and S. Silverberg: "Distributed meetings: A meeting capture and broadcasting system", Proceedings of ACM Multimedia '02, pp. 503–512 (2002).
- [81] M. Bett, H. Y. R. Gross, X. Zhu, Y. Pan, J. Yang and A. Waibel: "Multimodal meeting tracker", Proceedings of RIAO 2000: Content-Based Multimedia Information Access (2000).
- [82] 竹前嘉修,大塚和弘,武川直樹:"対面の複数人対話を撮影対象とした対話参加者の視線に基づく映像切り替え方法とその効果",情報処理学会論文誌,46,7, pp. 1752–1767 (2005).
- [83] D. Reidsma, R. Akker, R. Rienks, R. Poppe, A. Nijholt, D. Heylen and J. Zwiers: "Virtual meeting rooms: From observation to simulation", Proceedings of Social Intelligence Design '05 (2005).

- [84] D. Heylen, I. V. Es, A. Nijholt and B. V. Dijk: "Experimenting with the gaze of a conversational agent", Proceedings of International CLASS Workshop on Natural Intelligent and Effective Interaction in Multimodal Dialogue Systems, pp. 93–100 (2002).
- [85] M. Katzenmaier, R. Stiefelhagen and T. Schultz: "Identifying the addressee in humanhuman-robot interactions based on head pose and speech", Proceedings of ACM International Conference on Multimodal Interfaces (ICMI'04), pp. 13–15 (2004).
- [86] 松坂要佐,東條剛史,小林哲則: "グループ会話に参与する対話ロボットの構築",電子情報通信学会論文誌 D-II, J84-D-II, 6, pp. 898–908 (2001).
- [87] S. Basu, T. Choudhury and B. Clarkson: "Learning human interactions with the influence model", MIT Media Lab. TR#539 (2001).
- [88] S. Basu: "Conversational Scene Analysis", Ph.D thesis, Massachusetts Institute of Technology (2002).
- [89] T. K. Choudhury: "Sensing and Modeling Human Networks", Ph.D thesis, Massachusetts Institute of Technology (2004).
- [90] A. Dielmann and S. Renals: "Dynamics Bayesian networks for meeting structuring", Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'04), pp. V- 629-32 (2004).
- [91] I. McCowan, D. Perez, S. Bengio, G. Lathoud, M. Barnard and D. Zhang: "Automatic analysis of multimodal group actions in meetings,", IEEE Transactions on Pattern Analysis and Machine Intelligence, 27, 3, pp. 305–317 (2005).
- [92] D. Zhang, D. G. Perez, S. Bengio, I. McCowan and G. Lathoud: "Modeling individual and group actions in meetings: A two-layer HMM framework", Proceedings of 2nd. IEEE Workshop on Event Mining (2004).
- [93] 伊藤禎宣, 岩澤昭一郎, 土川 仁, 角 康之, 間瀬健二, 片桐恭弘, 小暮 潔, 萩田紀博: "装着型体験記録装置による対話インタラクションの判別機能実装と評価", ヒューマンインタフェース 学会論文誌, 7, 1, pp. 167–178 (2005).
- [94] 角 康之, 伊藤禎宣, 松口哲也, Sidney Fels, 間瀬健二: "協調的なインタラクションの記録と 解釈", 情報処理学会論文誌, 44, 11, pp. 2628–2637 (2003).
- [95] 坊農真弓, 鈴木紀子, 片桐恭弘: "多人数会話を対象としたデータ収集と分析-参与構造分析 を例として---",国際文化学, 11, pp. 81-94 (2004).

- [96] 坊農真弓,鈴木紀子,片桐恭弘: "ユビキタスセンサを用いた会話参与手続きの認識",第41
   回人工知能学会 音声・言語理解と対話処理研究会 (SIG-SLUD), pp. 27-32 (2004).
- [97] 榎本美香,伝康晴: "3 人会話における参与役割の交代に関わる非言語行動の分析",第 38 回人工知能学会 音声・言語理解と対話処理研究会 (SIG-SLUD-A301), pp. 25-30 (2003).
- [98] E. Goffman: "Forms of Talks", University of Pennsylvania Press, Philadelphia (1981).
- [99] H. H. Clark and T. B. Carlson: "Hearers and speech acts", Language, 58, pp. 332–373 (1982).
- [100] 福井康之:"まなざしの心理学", 創元社 (1984).
- [101] C. L. Kleinke: "Gaze and eye contact: A research review", Psychological Bulletin, 100, 1, pp. 78–100 (1986).
- [102] 遠藤利彦: "読む目・読まれる目 視線理解の進化と発達の心理学 –", 東京大学出版会 (2005).
- [103] A. Kendon: "Some functions of gaze-direction in social interaction", Acta Psychologica, 26, pp. 22–63 (1967).
- [104] C. Goodwin: "Conversational Organization: Interaction between Speakers and Hearers", Academic Press, New York (1981).
- [105] S. Baron-Cohen: "Mindblindness: An Essay on Autism and Theory of Mind", Bradford Books (1997).
- [106] R. Stiefelhagen, J. Yang and A. Waibel: "Modeling focus of attention for meeting index based on multiple cues", IEEE Transactions on Neural Networks, 13, 4, pp. 928–938 (2002).
- [107] R. Stiefelhagen and J. Zhu: "Head orientation and gaze direction in meetings", Proceedings of ACM Conference on Human Factors in Computing Systems (CHI'02) (2002).
- [108] M. A. Turk and A. P. Pentland: "Face recognition using eigenfaces", Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'91), pp. 586–591 (1991).
- [109] H. Murase and S. K. Nayar: "Visual learning and recognition of 3-d objects from appearance", International Journal of Computer Vision, 14, 1, pp. 5–24 (1995).
- [110] 村瀬 洋, シュリー・ナイヤー: "2次元照合による3次元物体認識-パラメトリック固有空間法-",電子情報通信学会論文誌 D-II, J77-D2, 11, pp. 2179-2187 (1994).

- [111] H. Murase and R. Sakai: "Moving object recognition in eigenspace representation", Pattern Recognition Letters, 17, pp. 155–162 (1996).
- [112] 池口 徹, 合原一幸: "非線形現象の解析手法 [vi・完] –カオスと時系列解析–", 電子情報通信
   学会誌, 79, 8, pp. 814–819 (1996).
- [113] E. N. Lorenz: "Atomospheric predictability as revealed by naturally occuring analogues", Journal of Atomospheric Sciences, 26, pp. 636–646 (1969).
- [114] F. J. Mulhern and R. J. Caprara: "A nearest neighbor model for forecasting market response", International Journal of Forecasting, 10, pp. 191–207 (1994).
- [115] T. Ikeguchi and K. Aihara: "Prediction of chaotic time series with noise", IEICE Transactions on Fundamentals, E78-A, 10, pp. 1291–1298 (1995).
- [116] J. D. Farmer and J. J. Sidorowich: "Predicting chaotic time series", Physical Letter Review, 59, 8, pp. 845–848 (1987).
- [117] J. Jiménez, J. A. Moreno and G. J. Ruggeri: "Forecasting on chaotic time series: A local optimal linear-reconstruction method", Physical Review A, 45, 6, pp. 3553–3558 (1992).
- [118] 松葉育雄:"非線形時系列解析", 朝倉書店 (2000).
- [119] 水上雄一,西森壽郎,岡本順子,合原一幸: "グラムシュミット直交化法を用いた決定論的予 測手法",電気学会論文誌 C, 115, 6, pp. 792–797 (1995).
- [120] 寺崎健,池口徹,合原一幸,田中智: "経済時系列データの決定論的非線形ダイナミカル特性 に関する解析",電子情報通信学会論文誌A, J78-A, 12, pp. 1601–1617 (1995).
- [121] 小林重信: "事例ベース推論の現状と展望", 人工知能学会誌, 7, 4, pp. 559-566 (1992).
- [122] K. M. Cupta and A. R. Montazemi: "A connectioninst approach for similarity assessment in case-based reasoning systems", Decision Support Systems, 19, pp. 237–253 (1997).
- [123] J. L. Kolodner: "Improving human decision making through case-based decision aiding", AI Magazine, pp. 52–68 (1991).
- [124] C. Stanfill and D. Waltz: "Toward memory-based reasoning", Communications of the ACM, 29, 12, pp. 1213–1228 (1986).
- [125] J. Rachlin, S. Kasif and D. W. Aha: "Towards a better understanding of memory-based reasoning systems", Proceedings of 11th International Conference on Machine Learning, pp. 242–250 (1991).

- [126] Z. Toth: "Long-range weather forecasting using an analog approach", Journal of Climate, 2, pp. 594–607 (1989).
- [127] H. M. Dool: "A new look at weather forecasting through analogues", Monthly Weather Review, 117, pp. 2230–2247 (1989).
- [128] T. Mohri, M. Nakamura and H. Tanaka: "Weather forecasting using memory-based reasoning", Proceedings of 2nd. International Workshop on Parallel Processing for Artificial Intelligence (PPAI'93), pp. 40–45 (1993).
- [129] 毛利隆夫,田中英彦: "記憶に基づく推論による天気予測",人工知能学会論文誌,10,5,pp.
   798-805 (1995).
- [130] M. Isard and J. MacCormick: "Bramble: A Bayesian multiple-blob tracker", Proceedings of International Conference on Computer Vision (ICCV'01), pp. 34–41 (2001).
- [131] I. J. Cox and S. L. Hingorani: "An efficient implementation of Reid's multiple hypothesis tracking algorithm and its evaluation for the purpose of visual tracking", IEEE Transactions on Pattern Analysis and Machine Intelligence, 18, 2, pp. 138–150 (1996).
- [132] D. F. D. Schulz, W. Burgard and A. B. Cremers: "Tracking multiple moving objects with a mobile robot", Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'01), pp. 371–377 (2001).
- [133] W. N. Martin and J. K. Aggarwal: "Volumetric descriptions of objects from multiple views", IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-5, 2, pp. 150–158 (1983).
- [134] 田中 實, 山本良一: "計算物理学と計算化学", 海文堂出版 (1988).
- [135] K. N. Kutulakos and S. M. Seitz: "A theory of shape by space carving", International Journal of Computer Vision, 38, 3, pp. 199–218 (2000).
- [136] R. Vertegaal, R. Slagter, G. Veer and A. Nijholt: "Eye gaze patterns in conversations: There is more to conversational agents than meets the eyes", Proceedings of ACM Conference on Human Factors in Computing Systems (CHI'01), pp. 301–308 (2001).
- [137] N. Jovanovic and R. Akker: "Towards automatic addressee identification in multi-party dialogues", Proceedings of SIGdial'04, pp. 89–92 (2004).

- [138] T. Ohno and N. Mukawa: "A free-head, simple calibration, gaze tracking system that enables gaze-based interaction", Proceedings of Eye Tracking Research & Application Symposium (ETRA'04), pp. 115–122 (2004).
- [139] Y. Matsumoto and A. Zelinsky: "An algorithm for real-time stereo vision implementation of head pose and gaze direction measurement", Proceedings of International Conference on Automatic Face and Gesture Recognition '04, pp. 499–504 (2000).
- [140] C. J. Kim and C. R. Nelson: "State-Space Models with Regime Switching", MIT Press (1999).
- [141] S. Frühwirth-Schnatter: "Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models", Journal of American Statistical Association, 96, 453, pp. 194–209 (2001).
- [142] W. R. Gilks, S. Richardson and D. J. Spiegelhalter: "Markov Chain Monte Carlo in Practice", Chapman & Hall/CRC (1996).
- [143] 中妻照雄: "ファイナンスのための MCMC 法によるベイズ分析",(財) 三菱経済研究所 (2003).
- [144] 松原康晴, 尺長 健: "疎テンプレートマッチングとその実時間物体追跡への応用", 情報処
   理学会論文誌:コンピュータビジョンとイメージメディア, 46, SIG9(CVIM11), pp. 60-71
   (2005).
- [145] R. ディーステル: "グラフ理論", シュプリンガー・フェアラーク東京 (2000).
- [146] D. G. Novic, B. Hansen and K. Ward: "Coordinating turn-taking with gaze", Proceedings of International Conference on Spoken Language '96, pp. 1888–1891 (1996).
- [147] J. M. Bernardo and A. F. M. Smith: "Bayesian Theory", John Wiley & Sons, Ltd. (1994).
- [148] 繁桝算男: "ベイズ統計入門", 東京大学出版会 (1985).
- [149] 渡部 洋:"ベイズ統計学入門", 福村出版(株) (1999).
- [150] R. Chen and T. Li: "Blind restoration of linearly degraded discrete signals by Gibbs sampling", IEEE Transactions on Pattern Analysis and Machine Intelligence, 43, 10, pp. 2410–2413 (1995).
- [151] P. Viola and M. J. Jones: "Robust real-time face detection", International Journal of Computer Vision, 57, 2, pp. 137–154 (2004).

- [152] 小野泰弘, 岡部孝弘, 佐藤洋一: "目領域の切り出しの不定性を考慮した低解像度画像からの
   視線方向推定", 画像の認識・理解シンポジウム (MIRU'05) 論文集, pp. 96–103 (2005).
- [153] Y. Yacoob and M. J. Black: "Parameterized modeling and recognition of activities", Computer Vision and Image Understanding, 73, 2, pp. 232–247 (1999).
- [154] A. F. Bobick and A. D. Wilson: "A state-based technique for the summarization and recognition of gesture", Proceedings of International Conference on Computer Vision (ICCV'95), pp. 382–388 (1995).
- [155] 藤本泰史,岩佐英彦,横矢直和,竹村治雄: "固有空間内の軌跡の類似性に基づく動画像検索", 電子情報通信学会技術報告 PRMU96-110, pp. 49-56 (1996).
- [156] 大塚和弘, 堀越 力, 安達文夫, 落合慶広, 鈴木智: "時空間中の運動軌跡に基づくオプティカルフローの推定", 電子情報通信学会論文誌 D-II, J81-D-II, 9, pp. 2061–2073 (1998).
- [157] 大塚和弘, 堀越力, 安達文夫, 曽根原登, 鈴木智: "時空間中の運動軌跡に基づくテンポラル テクスチャー解析", 情報処理学会論文誌, 40, 8, pp. 3155-3168 (1999).

# 本論文に関する研究発表

## [学術論文]

- [1] 大塚和弘, 堀越力, 安達文夫, 落合慶広, 鈴木 智: "時空間中の運動軌跡に基づくオプティ カルフローの推定", 電子情報通信学会論文誌 D-II, Vol. J81-D-II, No. 9, pp. 2061–2073, September (1998)
- [2] 大塚和弘, 堀越力, 安達文夫, 曽根原登, 鈴木 智: "時空間中の運動軌跡に基づくテンポラ
   ルテクスチャー解析", 情報処理学会論文誌, Vol. 40, No. 8, pp. 3155–3168, August (1999)
- [3] Kazuhiro Otsuka, Tsutomu Horikoshi, Haruhiko Kojima, and Satoshi Suzuki: "Image Sequence Retrieval for Forecasting Weather Radar Echo Pattern", IEICE Transactions on Information and Systems, Vol. E83-D, No. 7, pp. 1458–1465, July (2000)
- [4] 大塚和弘,武川直樹: "多視点観測に基づく複数物体の相互オクルージョン解析と逐次状態 推定",情報処理学会論文誌, Vol. 44, No. SIG17(CVIM8), pp. 109–125, December (2003)
- [5] 大塚和弘,竹前嘉修,大和淳司,村瀬 洋: "複数人物の対面会話を対象としたマルコフ切 替えモデルに基づく会話構造の確率的推論",情報処理学会論文誌, Vol. 47, No. 7, pp. 2317–2334, July (2006)
- [6] 大塚和弘, 落合慶広, 鈴木 智, 堀越力, 小田寿則, 稲垣充廣: "局地・短時間気象予報シス テム: MyWeather", 情報処理学会論文誌, Vol. 40, No. 3, pp. 1117–1126, March (1999)

## [Book Chapter]

 Kazuhiro Otsuka, Yoshinao Takemae, Junji Yamato, and Hiroshi Murase: "Probabilistic Inference of Gaze Patterns and Structure of Multiparty Conversations from Head Directions and Utterances", Lecture Notes in Artificial Intelligence (LNAI) New Frontiers in Artificial Intelligence, Joint JSAI 2005 Workshop Post-Proceedings, Vol. 4012, pp. 353– 364, Springer (2006)

## [国際会議 (査読あり)]

- Kazuhiro Otsuka, Tsutomu Horikoshi, and Satoshi Suzuki: "Image Velocity Estimation from Trajectory Surfaces in Spatiotemporal Space", Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'97) (Puerto Rico), pp. 200–205, June (1997)
- [2] Kazuhiro Otsuka, Tsutomu Horikoshi, Keihiro Ochiai, and Satoshi Suzuki: "Precipitation Forecast Based on Segmentation of Radar Echo Motion Field", Proceedings of 28th International Conference on Radar Meteorology (Austin), pp. 218–219, September (1997)
- [3] Kazuhiro Otsuka, Tsutomu Horikoshi, Satoshi Suzuki, and Masaharu Fujii: "Feature Extraction of Temporal Texture based on Spatiotemporal Motion Trajectory", Proceedings of IAPR 14th International Conference on Pattern Recognition (ICPR'00) (Brisbane), pp. 1047–1051, August (1998)
- [4] Kazuhiro Otsuka, Tsutomu Horikoshi, and Satoshi Suzuki: "Image Sequence Retrieval for Forecasting Weather Radar Echo Pattern", Proceedings of IAPR Workshop on Machine Vision Applications (MVA'98) (Makuhari), pp. 238–241, November (1998)
- [5] Kazuhiro Otsuka, Tsutomu Horikoshi, Satoshi Suzuki, and Masaharu Fujii: "Local Precipitation Forecast based on Retrieval of Similar Echo Pattern in Radar Image Sequences", Proceedings of American Meteorological Society 15th International Conference Interactive Information and Processing Systems for Meteorology, Oceanography, and Hydrology (Dallas), pp. 99–102, January (1999)
- [6] Kazuhiro Otsuka, Tsutomu Horikoshi, Satoshi Suzuki, and Haruhiko Kojima: "Memorybased Forecasting of Complex Natural Patterns by Retrieving Similar Image Sequences," Proceedings of IAPR 10th. International Conference Image Analysis and Processing (Venice), pp. 874–879, September (1999)
- [7] Kazuhiro Otsuka, Tsutomu Horikoshi, Satoshi Suzuki, and Haruhiko Kojima: "Memorybased Forecasting for Weather Image Patterns", Proceedings of 17th National Conference on Artificial Intelligence (AAAI'00) (Austin), pp. 330–336, July (2000)
- [8] Kazuhiro Otsuka and Naoki Mukawa: "Multiview Occlusion Analysis for Tracking Densely Populated Objects based on 2-D Visual Angles", Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'04) (Washington, DC), Vol. 1, pp. 90–97, June (2004)
- [9] Kazuhiro Otsuka and Naoki Mukawa: "A Particle Filter for Tracking Densely Populated Objects based on Explicit Multiview Occlusion Analysis", Proceedings of IAPR International Conference on Pattern Recognition (ICPR'04) (Cambridge), Vol. 4, pp. 23–26, August (2004)
- [10] Kazuhiro Otsuka, Yoshinao Takemae, Junji Yamato, and Hiroshi Murase: "Probabilistic Inference of Gaze Patterns and Structure of Multiparty Conversations from Head Directions and Utterances", Proceedings of 1st. International Workshop on Conversational Informatics (Kitakyushu), pp. 7–12, June (2005)
- [11] Kazuhiro Otsuka, Yoshinao Takemae, Junji Yamato, and Hiroshi Murase: "A Probabilistic Inference of Multiparty-Conversation Structure Based on Markov-Switching Models of Gaze Patterns, Head Directions, and Utterances", Proceedings of ACM International Conference on Multimodal Interfaces (ICMI'05) (Trento), pp. 191–198, October (2005)
- [12] Kazuhiro Otsuka, Junji Yamato, Yoshinao Takemae, and Hiroshi Murase: "Quantifying Interpersonal Influence in Face-to-Face Conversations based on Visual Attention Patterns", Proceedings of ACM Conference on Human Factors in Computing Systems (CHI'06) Extended Abstract (Montréal), pp. 1175–1180, April (2006)
- [13] Kazuhiro Otsuka, Junji Yamato, Yoshinao Takemae, and Hiroshi Murase: "Conversation Scene Analysis with Dynamic Bayesian Network based on Visual Head Tracking", Proceedings of IEEE International Conference on Multimedia & Expo (ICME'06) (Toronto), July (2006)

## [国内会議・研究会]

- [1] 大塚和弘, 堀越力, 鈴木 智: "時空間画像中の運動軌跡の統合に基づくオプティカルフロー 推定法", 情報処理学会研究報告 コンピュータビジョンとイメージメディア (CVIM), No. 103–10, pp. 75–81, January (1997)
- [2] 大塚和弘, 堀越力, 鈴木 智: "時空間中の運動軌跡に基づくテンポラルテクスチャー解析", 画像の認識・理解シンポジウム, pp. I-407-412, July (1998)
- [3] 大塚和弘, 堀越力, 鈴木智, 児島治彦: "気象レーダ画像系列を対象とした記憶に基づく予 測画像の生成", 画像の認識・理解シンポジウム, pp. II-265-270, July (2000)

- [4] 大塚和弘,武川直樹: "多視点観測に基づく複数物体の相互オクルージョン解析と逐次状態推定",情報処理学会研究報告 コンピュータビジョンとイメージメディア (CVIM), Vol. 2003-CVIM-136, No. 17, pp. 123–130, January (2003)
- [5] 大塚和弘,大和淳司,村瀬 洋: "複数人物の対面会話シーンを対象とした画像中の人物頭部 追跡に基づく会話構造のモデル化と確率的推論",画像の認識・理解シンポジウム(MIRU2006) ダイジェスト冊子, No. OS3A-1, July (2006)
- [6] 大塚和弘: "非言語行動の観測に基づく複数人物の会話シーン分析",人工知能学会研究会資料第48回言語・音声理解と対話処理研究会 (SIG-SLUD), No. SIG-SLUD-A602-01, pp. 1–6, November (2006)

[国内全国大会]

- [1] 大塚和弘, 堀越力, 鈴木 智: "時空間中の運動軌跡に基づく画像速度の推定", 1997 年電子 情報通信学会 総合大会講演論文集, 情報・システム 2, p. 413, March (1997)
- [2] 大塚和弘,堀越力,鈴木智: "時空間速度場の分割に基づく降水予測",情報処理学会第54
  回全国大会講演論文集, Vol. 2, pp. 409-410, March (1997)
- [3] 大塚和弘, 堀越力, 鈴木 智: "時空間中の運動軌跡に基づく密なオプティカルフローの推定", 情報処理学会 第 55 回全国大会講演論文集, Vol. 2, pp. 266-267, September (1997)
- [4] 大塚和弘, 堀越力, 鈴木智, 藤井雅晴: "時空間画像解析によるレーダエコーパターンの特 徴抽出と分類", 日本気象学会 1997 年度秋季大会講演予稿集, A307, October (1997)
- [5] 大塚和弘, 堀越力, 鈴木 智: "時空間中の運動軌跡に基づくテンポラルテクスチャーの特 (1998)
   [5] 大塚和弘, 堀越力, 鈴木 智: "時空間中の運動軌跡に基づくテンポラルテクスチャーの特 (1998)
- [6] 大塚和弘,小田寿則,堀越力,曽根原登: "局所降水予測に向けた類似気象パターン検索", 電子情報通信学会 1998年 ITS に関する情報通信技術シンポジウム講演論文集,SAD-6-11, pp. 131–132, September (1998)
- [7] 大塚和弘, 堀越力, 曽根原登: "映像予測のための検索類似尺度の検討", 1999年電子情報
  通信学会総合大会講演論文集, D-11-140, March (1999)
- [8] 大塚和弘, 堀越力, 児島治彦: "気象レーダ画像を対象とした検索に基づく映像予測の検討",
  2000 年電子情報通信学会 総合大会講演論文集, D-11-160, March (2000)

## [その他, 共著による発表(学術論文)]

- [1] 瀬戸篤志,大塚和弘,石井六哉: "ニューラルネットを用いた劣化画像の復元法",電子情報
  通信学会論文誌 D-II, Vol. J80-D-II, No. 5, pp. 1276–1285 (1997)
- [2] Atsushi Seto, Kazuhiro Ohtsuka, and Rokuya Ishii: "Restoration of Degraded Images with Neural Networks", System and Computers in Japan, Vol. 29, No. 8, pp. 57–67 (1998)
- [3] 竹前嘉修,大塚和弘,武川直樹: "対面の複数人対話を撮影対象とした対話参加者の視線に基づく映像切り替え方法とその効果",情報処理学会論文誌, Vol. 46, No. 7, pp. 1752–1767 (2005)
- [4] Yoshinao Takemae, Kazuhiro Otsuka, Junji Yamato, and Sinji Ozawa: "The Subjective Evaluation Experiments on Automatic Video Editing System Using Vision-based Head Tracking for Multiparty Conversations", IEEJ Transactions on Electronics, Information and Systems, Vol. 126-C, No. 4, pp. 435–442 (2006)

## [その他, 共著による発表(国際会議)]

- Yoshinao Takemae, Kazuhiro Otsuka, and Naoki Mukawa: "Video Cut Editing Rule Based on Participants' Gaze in Multiparty Conversation", Proceedings of ACM Multimedia 2003, pp. 303–306 (2003)
- [2] Yoshinao Takemae, Kazuhiro Otsuka, and Naoki Mukawa: "An Analysis of Speakers' Gaze Behavior for Automatic Addressee Identification in Multiparty Conversation and Its Application to Video Editing", Proceedings of IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN 2004), pp. 581–586 (2004)
- [3] Yoshinao Takemae, Kazuhiro Otsuka and Naoki Mukawa: "Impact of Video Editing Based on Participants' Gaze in Multiparty Conversation", Proceedings of ACM Conference on Human Factors in Computing Systems (CHI2004) Extended Abstract, pp. 1333–1336 (2004)
- [4] Yoshinao Takemae, Kazuhiro Otsuka, and Junji Yamato: "Automatic Video Editing System Using Stereo-Based Head Tracking for Multiparty Conversation", Proceedings of ACM Conference on Human Factors in Computing Systems (CHI2005) Extended Abstract, pp. 1817–1820, 2005.

- [5] Yoshinao Takemae, Kazuhiro Otsuka, Junji Yamato: "Effects of Automatic Video Editing System Using Stereo-Based Head Tracking for Archiving Meetings", Proceedings of IEEE International Conference on Multimedia & Expo (IEEE/ICME 2005), (2005).
- [6] Yoshinao Takemae, Kazuhiro Otsuka, and Junji Yamato: "Development of Automatic Video Editing System Based on Stereo-Based Head Tracking for Archiving Meetings", Proceedings of The Third International Conference on Active Media Technology (AMT2005), p. 269 (2005)

## [その他, 共著による発表(国内研究会・全国大会)]

- [1] 落合慶広,鈴木 智,大塚和弘,小田寿則,境野英明,安達文夫: "局地・短時間気象予測シ ステム: My Weather",電子情報通信学会技術報告 OFS97-25 (1997)
- [2] 小田寿則,落合慶広,大塚和弘,曽根原 登: "ITS のための局地気象予測",電子情報通信学
  会 1998 年 ITS に関する情報通信技術シンポジウム講演論文集,SAD-6-12, September (1998)
- [3] 竹前嘉修,大塚和弘,武川直樹: "視線の機能に着目した映像切り替え 規則とその効果-人 に分かりやすい映像生成を目指して-",電子情報通信学会技術報告 HCS2002-52, pp. 49-54 (2003)
- [4] 竹前嘉修,大塚和弘,武川直樹: "複数人対話シーンの映像編集のための対話参加者の 視線
  に基づく映像切り替え方法とその効果",電子情報通信学会技術報告 HIP2003-137, pp. 1–6 (2004)
- [5] 木村昭悟,川西隆仁,大塚和弘,柏野邦夫: "重み付き特徴点照合に基づく高速画像検索",電 子情報通信学会技術報告, Vol. 105, No. 118, PRMU2005-23, pp. 7–12 (2005)