

NAIST-IS-MT9551030

修士論文

プロトコル学習による身振りの実時間画像認識

桐島 俊之

1997年 2月 14日

奈良先端科学技術大学院大学
情報科学研究科 情報処理学専攻

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

桐島 俊之

指導教官： 千原 國宏 教授
鹿野 清宏 教授
佐藤 宏介 助教授

プロトコル学習による身振りの実時間画像認識*

桐島 俊之

内容梗概

VR 環境においてはユーザの位置・姿勢・動作に関する身振り情報が反映されることが重要であり，従来，これらの情報はデータグローブや磁気センサなどの装着型身振りインタフェースにより獲得されてきた．装着型身振りインタフェースの場合，ユーザに負担をかけずに全身の身振り情報を獲得することが困難である．

そこで本研究では，身振り情報を認識するための画像処理アルゴリズムを新たに考案することにより上記問題点を解決し，VR アプリケーションとのインタフェースの透明化を可能とする非接触型身振りインタフェースを提案する．本論文では，身振り認識固有の問題として存在する身振りプロトコルに言及するとともに，認識系を身振りプロトコルに適応させるために多注視点身振り認識法を新たに提案する．

評価実験の結果，多注視点身振り認識法により，身振りプロトコルの学習が可能となることに加えて，専用ハードウェアなしでも実時間での身振り認識処理が可能であり，さらに実時間ベースで身振り情報を獲得できることが分かった．

キーワード

身振りインタフェース，身振りプロトコル，多注視点身振り認識法，活性化マップ，身振り情報

*奈良先端科学技術大学院大学 情報科学研究科 情報処理学専攻 修士論文，NAIST-IS-MT9551030, 1997年 2月 14日.

Realtime Gesture Recognition by Learning of Gesture Protocol*

Toshiyuki Kirishima

Abstract

In virtual environments, it is exceedingly important to reflect the changes of an user's position, posture and body movements. These gestural information have usually been obtained by conventional gesture interfaces such as DataGlove and polhemus sensors. As most of the conventional gesture interfaces are attached onto an user's body, the user's freedom to move is undeniably restricted.

In order to solve above-mentioned problem, novel image processing algorithms and frameworks to realize a vision-based gesture interface are to be explored in this research. In this paper, some important issues which are inherent in gesture recognition problem are discussed and a recognition framework, K-VIPS (K-Visual Interest Point Strategy) is newly introduced.

The proposed system performs recognition in real-time without any special-purpose hardware. The experimental results demonstrate that the proposed methods are effective for developing gesture interface for Virtual Reality.

Keywords:

Gesture Interface, Gesture Protocol, K-VIPS, Activation Map, Gestural Information

*Master's Thesis, Department of Information Processing, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT9551030, February 14, 1997.

目次

1	まえがき	1
2	身振りの画像認識	3
2.1	身体モデルに基づく身振り認識	3
2.2	図形パターン認識技術に基づく身振り認識	3
2.3	従来手法の問題点	4
2.4	本研究の目的と意義	5
3	多注視点身振り認識法	8
3.1	処理の枠組	8
3.2	注視領域の設定	9
3.3	多注視点特徴抽出法	11
3.4	身振り標準パターンの登録	15
3.5	身振り特徴量ベース識別処理	15
3.6	身振りプロトコルベース認識処理	16
3.7	身振り情報の定量化	18
3.7.1	身振り位相値の算出	18
3.7.2	身振り速度と身振り振幅の算出	19
4	評価実験	20
4.1	評価実験用身振りについて	20
4.2	システム構成	34
4.3	身振りの学習方法	34
4.4	身振り認識実験	36
4.4.1	標準身振りサンプルの認識結果	37
4.4.2	派生身振りサンプルの認識結果 (I)	38
4.4.3	派生身振りサンプルの認識結果 (II)	39
4.5	他人身振りの認識実験	40
4.5.1	最も個人に依存する設定で実験を行った場合	40

4.5.2	最も個人に依存しない設定で実験を行った場合	42
4.6	多注視点身振り認識処理の有効性評価実験	44
4.7	プロトコル学習の収束性評価実験	47
4.8	身振り認識処理の実時間性評価実験	48
4.9	身振り情報の認識実験	48
5	考察および検討	52
5.1	身振り認識実験の結果に関する考察	52
5.1.1	標準身振りサンプルの認識結果について	52
5.1.2	派生身振りサンプルの認識結果 (I) について	52
5.1.3	派生身振りサンプルの認識結果 (II) について	53
5.2	他人身振りの認識実験の結果に関する考察	54
5.3	多注視点身振り認識処理の有効性について	55
5.4	プロトコル学習の収束性について	56
5.5	身振り認識処理の実時間性について	56
5.6	身振り情報の認識結果について	57
5.7	問題点の検討と今後の研究課題	58
5.7.1	空間セグメンテーション処理の必要性	58
5.7.2	時間セグメンテーション処理の必要性	59
5.7.3	リアルタイムレンジファインダの必要性	61
5.7.4	階層化身振り認識機構の必要性	62
6	むすび	65
	謝辞	68
	参考文献	69
	付録	I
A	形状特徴抽出法について	I
A.1	ガウス分布との畳み込み演算を行う理由	I
A.2	密度係数を掛け合わせる理由	II

A.3	ガウス分布の勾配係数 a の決定方法	III
-----	--------------------------------	-----

図 目 次

2.1	身振り認識固有の問題：身振りプロトコル	6
3.1	多注視点身振り認識法の処理の枠組	8
3.2	採用した注視領域	10
3.3	背景差分画像と時間差分画像の生成方法	11
3.4	多注視点特徴抽出法における処理の流れ	12
3.5	形状特徴抽出処理の流れ	12
3.6	身振り特徴量ベース識別処理	15
3.7	身振りプロトコルベース認識処理	17
4.1	標識身振り（バイバイ）	22
4.2	標識身振り（グッパ）	22
4.3	例示子身振り（鳥のまね）	23
4.4	情感表示身振り（バンザイ）	23
4.5	調整子身振り（聞き返し）	24
4.6	環境適応子身振り（腕組み）	24
4.7	オブジェクト適応子身振り（マウス操作）	25
4.8	標準身振りサンプルのスナップショット（身振り「バイバイ」）	26
4.9	派生身振りサンプルのスナップショット（身振り「バイバイ」）	26
4.10	標準身振りサンプルのスナップショット（身振り「グッパ」）	27
4.11	派生身振りサンプルのスナップショット（身振り「グッパ」）	27
4.12	被験者 A の身振り動作のスナップショット	28
4.13	被験者 B の身振り動作のスナップショット	28
4.14	被験者 C の身振り動作のスナップショット	29
4.15	被験者 D の身振り動作のスナップショット	29
4.16	被験者 E の身振り動作のスナップショット	30
4.17	被験者 F の身振り動作のスナップショット	30
4.18	被験者 G の身振り動作のスナップショット	31
4.19	被験者 H の身振り動作のスナップショット	31
4.20	被験者 I の身振り動作のスナップショット	32

4.21	被験者 J の身振り動作のスナップショット	32
4.22	被験者 K の身振り動作のスナップショット	33
4.23	実験システム	34
4.24	身振りインタフェース用 GUI	35
4.25	使用した 6 種類の衣服	36
4.26	標準身振りサンプルの認識結果	37
4.27	派生身振りサンプルの認識結果 (I)	38
4.28	派生身振りサンプルの認識結果 (II)	39
4.29	他人身振りサンプルの認識結果 (I)	41
4.30	他人身振りサンプルの認識結果 (II)	43
4.31	身振りプロトコル学習の際の注視点重みの変化	44
4.32	注視点重みの学習結果	45
4.33	生成された身振りプロトコルマップ	46
4.34	プロトコル学習の収束性評価実験結果	47
4.35	システム動作周波数の変化	48
4.36	不規則に身振り動作した際の位相値の変化	49
4.37	不規則に身振り動作した際の身振り速度の変化	49
4.38	不規則に身振り動作した際の身振り振幅の変化	50
4.39	規則的に身振り動作した際の位相値の変化	50
4.40	規則的に身振り動作した際の身振り速度の変化	51
4.41	規則的に身振り動作した際の身振り振幅の変化	51
5.1	ボトムアップとトップダウンアプローチに基づく身振り認識	59
5.2	複数ユーザおよび単一ユーザによる並列身振り動作の認識	60
5.3	CCD カメラと SRF による衣服依存 / 非依存の身振り認識の枠組	62
5.4	階層化身振り認識	63
A.1	ガウシアンオペレータ (左) とその一次導関数 (右)	III
A.2	オペレータの対称性評価関数	IV

表 目 次

3.1	採用した 16 種類の注視点	14
4.1	採用した評価実験用身振り	20
4.2	標準身振りサンプルの認識結果	37
4.3	派生身振りサンプルの認識結果	38
4.4	派生身振りサンプルの認識結果	39
4.5	他人身振りサンプルの認識結果 (I)	40
4.6	他人身振りサンプルの認識結果 (II)	42
4.7	身振り「バイバイ」の認識結果	46
4.8	身振り「グッパ」の認識結果	46
6.1	対応可能な身振り解釈の多様性と対処方法	67

第 1 章 まえがき

新たな情報環境として仮想現実感 (Virtual Reality) が注目されている。VR 技術は、建築設計・医療・教育・娯楽などの広範囲な分野への応用が可能であり、従来の情報環境に大きな革新をもたらす技術である。このため、将来的に VR 技術は、専門家のみならず子供からお年寄りまでの広範囲な人々に利用されるものと考えられ、日常生活の中で人間同士がコミュニケーションを行うために利用しているメディア (音声・顔表情・身振り) を認識する先進的ヒューマンインタフェース [23] への要求が高まることが予想される。VR 環境では、外部環境からの働きかけが仮想環境や仮想物体に反映されるため、ユーザは仮想環境への知覚的な没入感を得ることが可能である。しかしながらこうした没入感は、環境としての CG を自然に経験することができる場合にのみ得られるものであるため、CG 技術のみでなく、ユーザとコンピュータとの間のインタフェース技術が非常に重要な構成要素となっている。

人間対人間のインタラクションにおいては、音声・顔表情に関する情報が身振り情報に対して優位であるが、人間対仮想環境あるいは人間対仮想物体におけるインタラクションにおいては逆に、身振り情報 (位置や姿勢や動作に関する情報) が音声・顔表情に関する情報に対して優位となる。しかしながら、従来、音声・顔表情をコンピュータを利用して認識するための研究は数多くなされ、実用的なシステムも開発されている一方で、身振り情報をコンピュータを利用して認識するための研究は、近年その重要性が認識されたばかり [4] で、まだその途に就いたばかりである。しかし、近年、コンピュータ性能の向上に伴って身振り情報を認識する多種多様なシステムが開発されるようになってきた (これらのシステムは身振りインタフェースと呼ばれる [15])。現在最も一般的に利用されている身振りインタフェースは、データグローブやポヒマスセンサなどの装着型の身振り

インタフェースである．これらの入力装置は，ユーザの体の限定部位の位置や姿勢に関する情報を比較的精度良く検出することが可能であるが，有線であったり装脱着が必要であったりするため，ユーザをシステムに拘束してしまう問題がある．また，ユーザに負担をかけずに全身の身振り情報を獲得することが困難であることに加えて，それらのシステムを不特定多数のユーザが利用する場面，あるいは複数のユーザが同時に利用する場面を想定すると，装置自体の購入コストの増大や衛生面での問題を回避することができない．

以上に述べた問題点は，実用的な身振りインタフェースを開発するためには回避できないものであり，画像処理による非接触身振り認識技術 [27] [28] が不可欠である．そこで，本研究では，画像情報から身振り情報を認識するための画像認識技術を新たに開発することにより，VR アプリケーションとのインタフェースの透明化を実現する非接触型身振りインタフェースを開発することを目的とする [1] [2] [3] ．

本論文では，身振り認識固有の問題として存在する身振りプロトコルに言及するとともに，身振り認識システムを身振りプロトコルに適応させるための枠組として，多注視点身振り認識法を新たに提案する．多注視点身振り認識法では，時々刻々と入力される身振り画像から，複数段・複数種類の身振り特徴抽出フィルタにより各注視点で必要とされる特徴量を抽出し，得られた特徴量を画像ラベルとして表現することによりフレームワイズの身振り認識処理を実現する．さらに，各注視点の画像ラベルを並列的に識別処理することにより各注視点の識別結果を類似度として表現し，多次元時系列パターンである活性化マップを獲得，この活性化マップに基づき身振りプロトコルベースの学習・認識処理を行う手法である．

本論文の第 2 章では，従来提案されている身振りの画像認識手法とその問題点について述べ，第 3 章では，多注視点身振り認識法を新たに提案し，第 4 章では，提案手法の有効性を認識処理の柔軟性と実時間性により評価した際の実験結果を示し，第 5 章では，評価実験で得られた結果に関して考察と検討を加え，さらに，今後の研究課題について述べる．

第 2 章 身振りの画像認識

画像処理技術による身振り認識 [8] [16] には主に二通りのアプローチがある．本章では，主に身体モデルに基づく身振り認識手法と図形パターン認識技術に基づく身振り認識手法について紹介する．続いて従来手法の問題点を明らかにした後で，本研究の目的と意義について述べる．

2.1 身体モデルに基づく身振り認識

身体モデルに基づく身振り認識の目的は，主に実画像中の人物の姿勢を推定することである．推定されるパラメータは，関節角や手のひらの方向などであり，これらの情報に基づき身振りが識別される．推定方法としては，シリンダモデルなどの数学的幾何モデルを入力画像にフィッティングさせる方法 [8] や，複数のマーカーを人体あるいは指などに取り付けてそれらの位置関係から人物姿勢や手形状を推定する方法など [16] が提案されている．これらの手法は，撮影条件・照明条件などを厳密に配慮することが必要であり，また，モデルフィッティングの際に煩雑な計算を必要とするため，実世界で撮影されるユーザの身振り画像を実時間で処理するには不向きである．

2.2 図形パターン認識技術に基づく身振り認識

図形パターン認識技術に基づく身振り認識の目的は，主に実画像中の人物領域などをパターン情報とみなし，入力された実画像の属するカテゴリーを推定することである．従って，身体モデルに基づく手法と比べて撮影条件や照明条件など

を厳密に配慮する必要がなく，またモデルフィッティング処理なども不要であるために，実世界で撮影される身振り画像を実時間で処理するには最も適切なアプローチである．ユーザの身振り動作に対して VR システムが実時間で応答することは，人工現実感を生成する上でも極めて重要である [18] [19] ため，本研究ではこの方式を採用する．

図形パターン認識技術に基づく身振り認識の代表的手法としては，Oka et al. による DP マッチングに基づく手法 [9]，Takagi et al. によるファジー連想記憶に基づく手法 [10]，Ishii et al. による HMM (Hidden Markov Model) に基づく手法 [11]，図形の固有空間内の軌跡の類似性に基づく手法 [14]，さらにニューラルネット的アプローチに基づく手法 [20]，Zernike Moment などの図形モーメント特徴に基づく手法 [21] [22] [24] [25] [6] [29] などがある．従来手法では，主に入力された身振りのカテゴリ推定のみを目的としているが，本研究では，身振りによる指示入力の有効性を高めるために，入力された身振りのカテゴリ推定のみでなく，標準動作と比べた際の身振り動作の方向や相対的な速度（以降身振り速度と呼ぶ），相対的な振幅（以降身振り振幅と呼ぶ）などの身振り情報を獲得するための手法も新たに提案する．

2.3 従来手法の問題点

殆どの従来手法では，身振り認識処理の単純化および高速化を図るために，最適な画像特徴量あるいは関節角などの特徴パラメータをシステム開発者が一つだけ選定した後で認識アルゴリズムが開発されている．画像特徴量の例としては，人物領域の面積率，リダクション画像，背景差分画像（シルエット画像），人物領域のエッジ特徴，時間差分画像，オプティカルフローなどが挙げられる．

認識対象となる画像特徴量や特徴パラメータが固定化されてしまうと，ユーザはシステム開発者の意図に従って身振り動作をしなければならない問題点がある．しかしながら，身振りの定義に関する主導権をユーザに委ねることにより，ユーザは任意に「場（認識空間）」をマッピングすることが可能となるメリットが生じる．従来の身振り認識手法では，システム開発者があらかじめ「場」を固定化し，ユーザが任意の「場」をマッピングすることは不可能であった．体の固定部位

を繰り返し連続的に長時間使用するようなインタフェースは、近年 RSI(Repetitive Stress Injuries) 症候群などを誘発する原因になるとの指摘 [12] もあり、生理学的側面からも柔軟な定義が可能であるインタフェースへの要求は高まっている。身振りによる指示入力 of 柔軟性を高めるためにも、身振り定義に関する主導権をユーザに委ねる [17] ことが不可欠である。

2.4 本研究の目的と意義

認識対象となる身振りの定義をユーザに委ねることは、対象に関して事前に何らかの仮定を設定することが困難になることを意味している。このため、コンピュータ側がユーザの提示する身振りから普遍的な身振り特徴や特定の身振りに共通した特徴を導出することが、柔軟かつ安定した身振り認識処理を実現するためには不可欠となる。この問題は、心理学の立場から焦点選択の問題として古くから研究され、その重要性和難しさは認識されているが、工学的な立場からこの問題を扱っている研究例 [7] はまだまだ少なく、身振りに関しては極稀である。認識対象となる身振りの定義をユーザに委ねた際に生じる問題を、具体的に図 2.1 を用いて説明する。

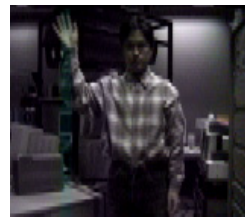
図 2.1 の各写真は、正面を向いて片手を振っている人物のスナップショットである。身振り定義に関する主導権がユーザにある場合、図 2.1 の身振り動作 (A) と (B) は別々の身振り動作としてではなく、同一範疇の身振り動作、すなわち同一の身振りカテゴリに属する動作として定義することが可能である。同様にして、身振り動作 (A) と (B) と (C) と (D) を同一の身振りカテゴリとして定義することも可能であろうし、身振り動作 (A) と (B) は同一であるが、身振り動作 (C) と (D) は同一の身振りカテゴリには属さないといった場合など、何通りものカテゴリ分けが可能である。このような状況に対処する方法の一つとしては、同一の身振りカテゴリに属する身振り画像全てを逐次登録する手法が考えられる。しかしこの方法では、同一カテゴリに属する身振り動作の種類が多いほど、登録すべき身振り画像の枚数が増加する。身振り動作あたりの身振り画像のデータ量が膨大であることを考えると、逐次登録方式では、パターン照合などの計算コストが著しく増加してしまうために、実時間での身振り認識処理は困難となる。



#1



#2

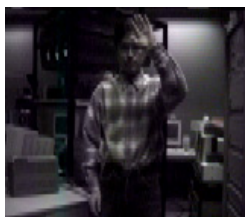


#3

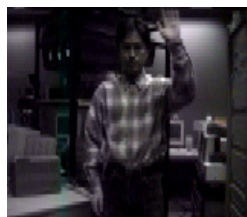


#4

身振り動作 (A)



#1



#2



#3



#4

身振り動作 (B)



#1



#2

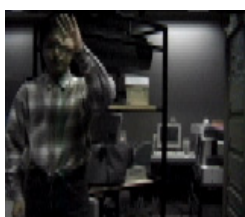


#3

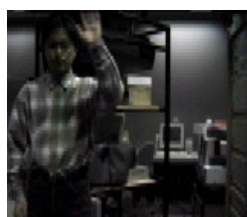


#4

身振り動作 (C)



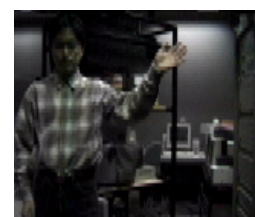
#1



#2



#3



#4

身振り動作 (D)

図 2.1 身振り認識固有の問題：身振りプロトコル

同一カテゴリに属するものとして定義される複数の身振り動作の間に共通する特徴があるのであれば、その特徴に着目して身振り動作のグルーピングを行なうことが可能である。身振り動作を少数の特徴量に基づいてグルーピングできれば、不要な身振り画像を逐次登録する必要がなくなるため、パターン照合処理に必要な計算量を最小限に抑えることが可能となると同時に、学習時に使用していない身振り画像サンプル（派生身振りサンプル）に対する認識率の向上も期待できる。このようなグルーピング処理を実現するには、ユーザが同一カテゴリの動作として定義した身振り画像から、時空間的に一貫した特徴量を見い出して、その特徴量に基づいて選択的に身振りを認識する枠組が不可欠である。

ある特徴に注目するためには、身振り認識における「注視機構 (Attention mechanism)」を実現するための問題に取り組む必要がある。本研究では、身振りの注視機構により規定されるものとして「身振りプロトコル」が存在しているものと仮定し、身振りプロトコルを「ある身振りを認識しようとする際に、何（着眼点）をどの程度の広がり（視野）をもって認識するのかを誘導する経験的な取り決め」と定義して、身振りプロトコルに適応するための新たな学習・認識の枠組みを提案する。身振りプロトコルを学習することが可能になれば、ユーザ定義の身振りおよび派生的な身振りを実時間で認識することが可能になるだけでなく、身振り画像に含まれる無用な情報の影響を抑えて、よりロバストな身振り認識処理の実現が期待できる。

第 3 章 多注視点身振り認識法

3.1 処理の枠組

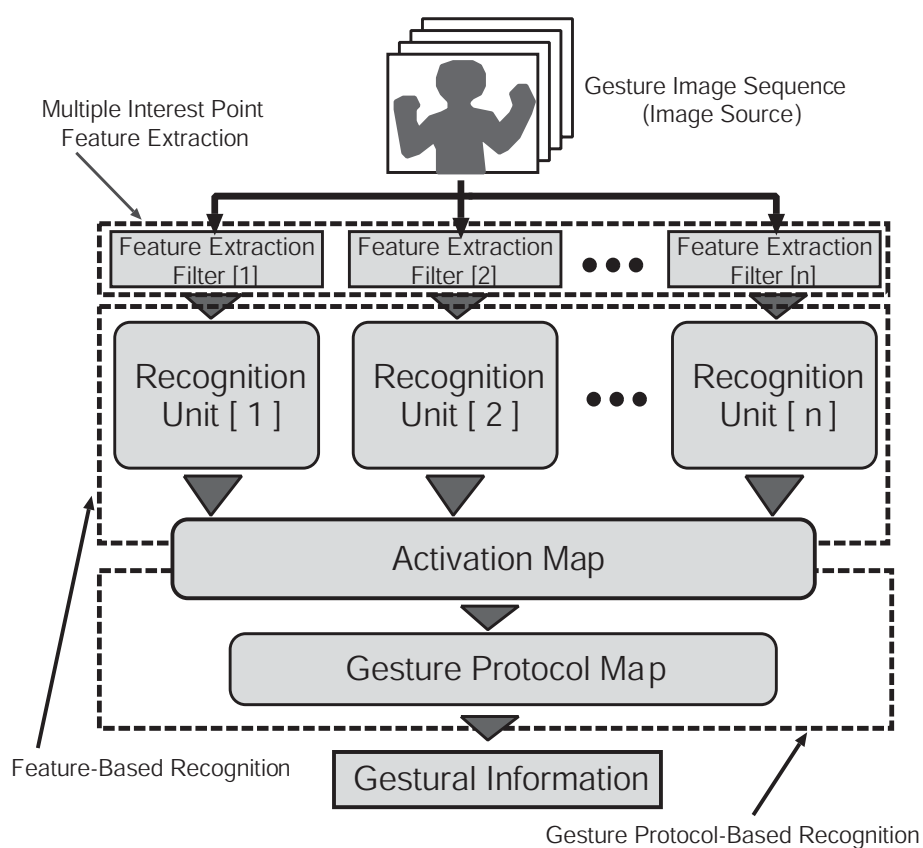


図 3.1 多注視点身振り認識法の処理の枠組

図 3.1 に本研究で提案する多注視点身振り認識法の処理の枠組を示す．多注視

点身振り認識法では、時々刻々と入力される身振り画像が複数の身振り特徴抽出フィルタにかけられ、多注視点特徴抽出法により原画像から各注視点で必要な特徴量が抽出される。抽出された特徴量は画像ラベルとして格納され、身振り学習時には身振り標準パターンとして登録される。一方、身振り認識時には、身振り標準パターンと各注視点の画像ラベルとのマッチング処理が行われ、識別結果は活性化マップとして出力される。この段階での身振り認識処理は特徴量ベース識別処理と呼ばれる。各注視点の重み付け処理を行うために、活性化マップを利用して身振りプロトコル学習が行われる。注視点の重み付け処理を行うことにより、重視すべき注視点や重視すべきでない注視点をグルーピングすることが可能となる。身振りプロトコル学習では、尤度分布を生成することにより、ある身振りを認識する際に時間軸方向に一貫している注視点に、相対的に大きな重みを割り当てる処理が行われる。生成された各注視点における尤度分布は、身振りプロトコルマップとして登録され、以後、この身振りプロトコルマップに基づいて入力された身振りのカテゴリ推定が行われる。この段階での身振り認識処理は身振りプロトコルベース認識処理と呼ばれる。

以上に示したように多注視点身振り認識法は、特徴量ベース識別処理と身振りプロトコルベース認識処理の二段階により構成される階層型認識機構である。入力された身振りのカテゴリ推定が行われた後で、推定された身振りカテゴリの活性化マップに基づき各種身振り情報の算出が行われる。算出される身振り情報は、ユーザの位置・姿勢・動作に関するものであり、身振り標準パターンにおけるフレーム番号に相当する身振り位相値、さらに身振り標準パターンと比べた場合の身振り動作の相対的な速度（身振り速度）や相対的な振幅（身振り振幅）に関する情報である。

3.2 注視領域の設定

身振り特徴抽出フィルタは、特定の画像特徴を抽出する。注視領域の選定は、獲得される身振り情報の有効性を左右するために極めて重要な設計要素である。単純図形における注視領域の選定方法に関しては、Iijima et al. により理論的側面からの研究が進められている [5] が、身振りにおける注視領域に関する工学的

基礎研究はなされていない．身振りを認識するには「経験」が無視できないため，現状では，ヒューリスティックなアプローチで注視領域を定めている．

決定した各注視領域は，注視領域 1（動作部位の相対的位置情報を抽出する），注視領域 2（動作部位の姿勢情報を抽出する），注視領域 3（姿勢情報を抽出する），注視領域 4（位置情報を抽出する），の 4 つである．採用した 4 つの注視領域を図 3.2 に示す．認識対象の運動依存 / 非依存により生じる解釈の多様性に対

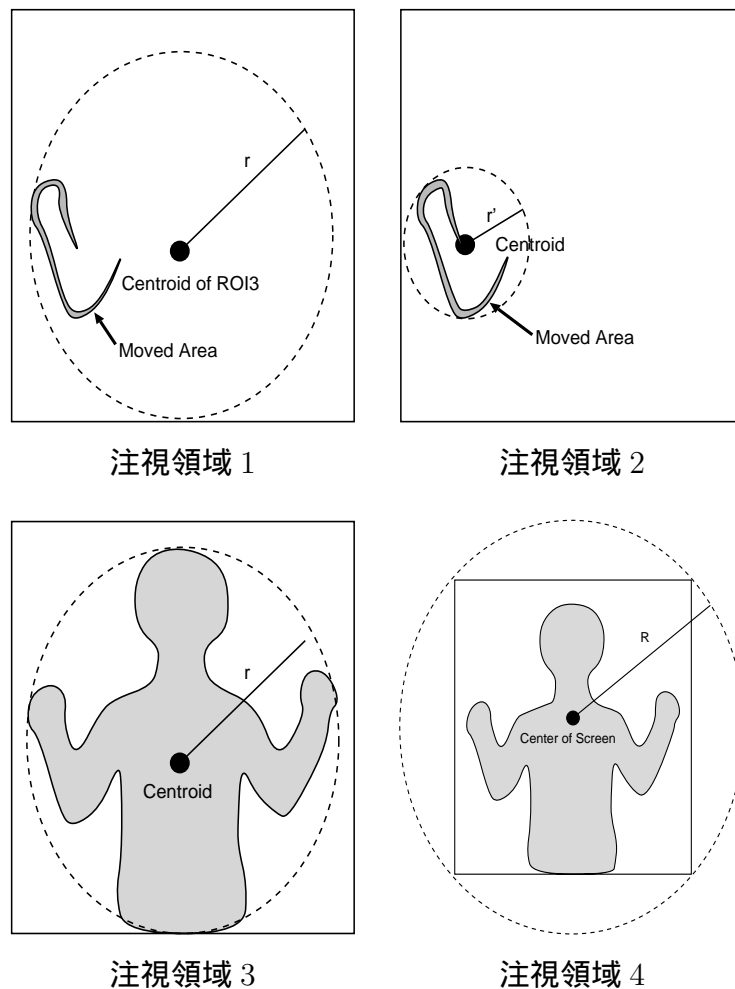


図 3.2 採用した注視領域

応するために，注視領域 1 と注視領域 2 では時間差分画像から，注視領域 3 と注視領域 4 では背景差分画像から特徴量を抽出する．背景差分画像と時間差分画像

の生成方法を図 3.3 に示す．

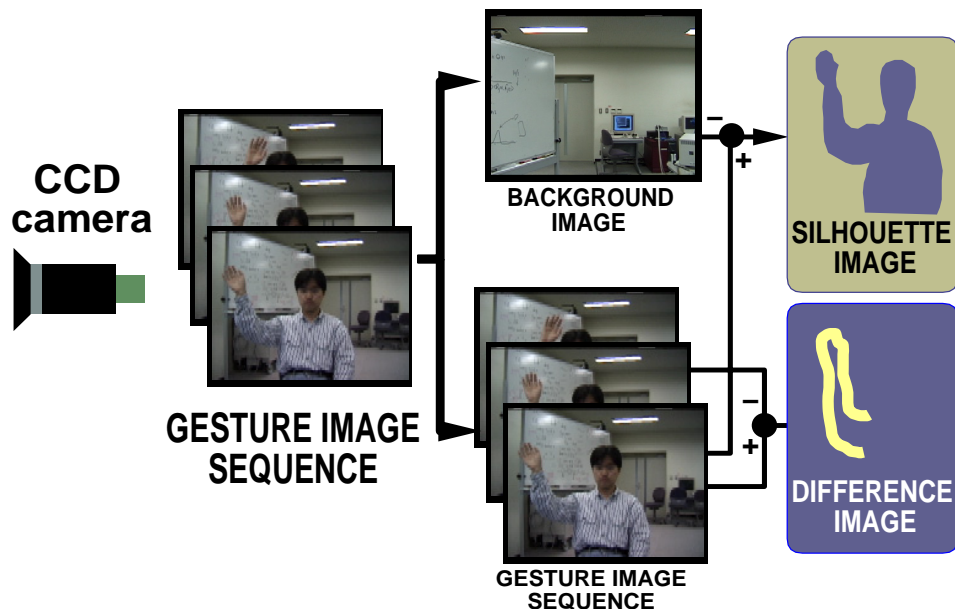


図 3.3 背景差分画像と時間差分画像の生成方法

3.3 多注視点特徴抽出法

本研究で提案する多注視点身振り認識法では，身振りの共通感覚成分である「形」から生じる解釈の多様性を考慮する．注視領域においては「形」の多様性が存在していて，この「形」に関する情報が身振りの共通感覚成分であると同時に，身振り解釈の多様性を保証していると考えられる．そこで設定された注視領域から形状特徴量を抽出する必要性が生じるが，本節では，形状パターンの大きさを考慮すべき場合と考慮すべきでない場合と，形状パターンの回転を考慮すべき場合と考慮すべきでない場合に対応するための特徴量の抽出方法を提案する．新たに形状特徴抽出法を提案する理由としては，回転依存／非依存と大きさ依存／非依存を考慮した特徴パターンを簡便な処理により得ることが挙げられる．図 3.4 に本研究で提案する多注視点特徴抽出法における処理の流れを示し，図 3.5 に形状特徴抽出処理の流れを示す．

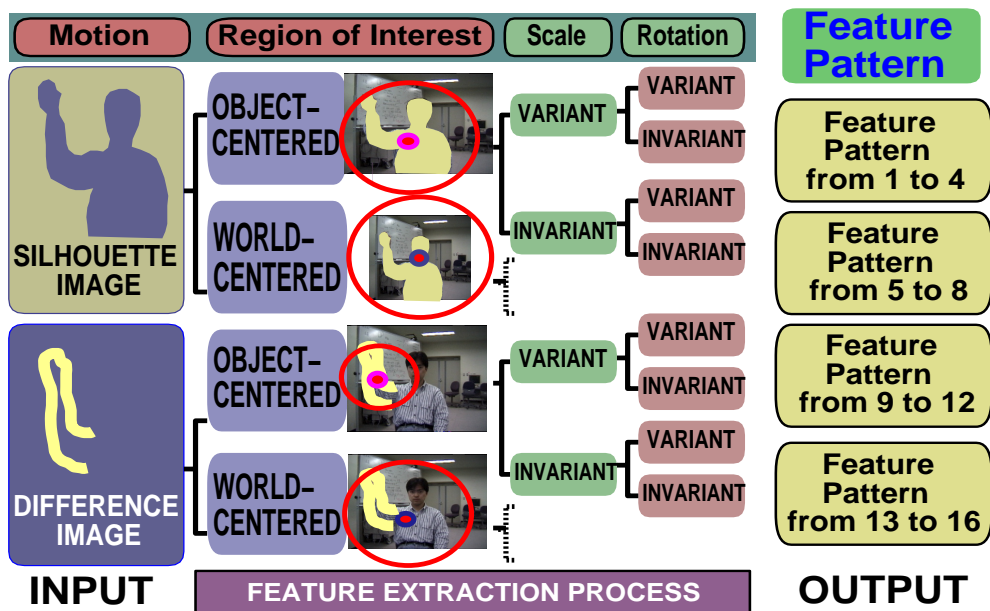


図 3.4 多注視点特徴抽出法における処理の流れ

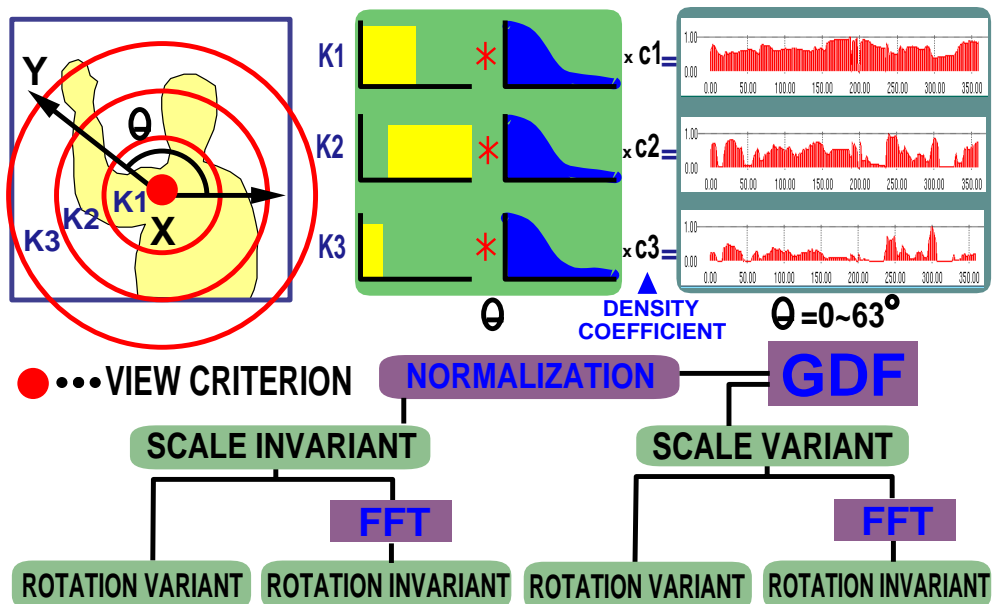


図 3.5 形状特徴抽出処理の流れ

注視領域における身振り画像を $f_i(x, y)$ ($i = 1, 2, \dots, n$) とする．閾値 t により身振り画像 $f_i(x, y)$ を二値化処理し，得られる二値画像を $g_i(x, y)$ とする． $g_i(x, y)$ に対する原点回りのモーメント $m_{p,q}$ は式 3.1 で定義される．

$$m_{p,q} = \sum_x \sum_y g_i(x, y) x^p y^q \quad (3.1)$$

これより，重心 (a_i, b_i) は式 3.2 により求めることができる．

$$(a_i, b_i) = \left(\frac{m_{1,0}}{m_{0,0}}, \frac{m_{0,1}}{m_{0,0}} \right) \quad (3.2)$$

ここで $g_i(x, y)$ を (a_i, b_i) が原点である極座標系で表し， $P_i(r, \theta)$ で表現することにする．角度 θ における半径 ($= R$) 方向のパターン分布を L 分割する．ここで，個々の分割領域をカーネルと呼び，その角度 θ における形状特徴量 $s_l(\theta)$ ($l = 1, 2, \dots, L$) を式 3.3 により定義する．

$$s_l(\theta) = \frac{R \sum_r P_i(r, \theta) \exp\{-\alpha(r - \frac{R(l-1)}{L})^2\}}{\sum_r P_i(r, \theta)} \quad (3.3)$$

式 3.3 を任意の解像度（分割数）で二値画像の全周方向に適用することで，カーネル l における特徴パターンを得る．線形パターンとガウス分布との畳み込みを計算し，その結果に密度係数を掛け合わせるため，この特徴量をガウス密度特徴（Gaussian Density Feature）と以降は呼ぶことにする．式 3.3 の妥当性と勾配係数 α の決定方法を付録 A. に示す．ガウス密度特徴を求めた後，各カーネルにおける特徴パターンに高速フーリエ変換（FFT）を適用する．ここで，変換対象である特徴パターンは回転角 θ に依存する関数とみなすことができるため，FFT の結果得られるパワースペクトルはシフト不変の性質を持つために，ここでは回転不変の特徴パターンとして利用することができる．ガウス密度特徴自体は回転依存の特徴パターンとして利用する．ガウス密度特徴を正規化することにより，形状パターンの回転に依存するが大きさには依存しない特徴パターンが得られる．ガウス密度特徴を正規化しない場合，形状パターンの回転と大きさに依存する特徴パターンが得られる．同様に，パワースペクトル成分を正規化することにより，形状パターンの回転と大きさに依存しない特徴パターンが得られる．パワースペ

クトル成分を正規化しない場合，認識対象の回転には依存しないが大きさには依存する特徴パターンが得られる．以上の処理は，各注視領域について実行される．一つの注視領域につき4通りの注視点（回転依存／非依存と大きさ依存／非依存の組み合わせ）に基づく特徴パターンが出力されるので，4通りの注視領域を設定している場合は合計16通りの注視点に関する特徴パターンが得られることになる．表3.1に採用した16種類の注視点を示す．

表 3.1 採用した 16 種類の注視点

注視点	位置	大きさ	回転	特徴情報源
1	依存	依存	依存	時間差分
2	依存	依存	非依存	時間差分
3	依存	非依存	依存	時間差分
4	依存	非依存	非依存	時間差分
5	非依存	依存	依存	時間差分
6	非依存	依存	非依存	時間差分
7	非依存	非依存	依存	時間差分
8	非依存	非依存	非依存	時間差分
9	非依存	依存	依存	背景差分
10	非依存	依存	非依存	背景差分
11	非依存	非依存	依存	背景差分
12	非依存	非依存	非依存	背景差分
13	依存	依存	依存	背景差分
14	依存	依存	非依存	背景差分
15	依存	非依存	依存	背景差分
16	依存	非依存	非依存	背景差分

3.4 身振り標準パターンの登録

身振り動作の知覚は，位置や姿勢の状態が連続的に変化することにより生成されと考え，各特徴パターンを画像ラベルとして登録する．画像ラベルには，身振りカテゴリ番号・位相値・注視点番号・特徴パターンなどが格納される．位相値は，ユーザが登録時に身振り動作を開始した時点から終了するまでの身振り画像のフレーム番号である．このように，一枚の身振り画像を複数の画像ラベルによって分解表現することによって，高次認識処理部における注視点の重み付け処理が可能となる．この際に登録される画像ラベル（群）を身振り標準パターンとする．身振り標準パターンの一貫性を保証するために，各身振りにつき一通りの標準パターンが登録される．

3.5 身振り特徴量ベース識別処理

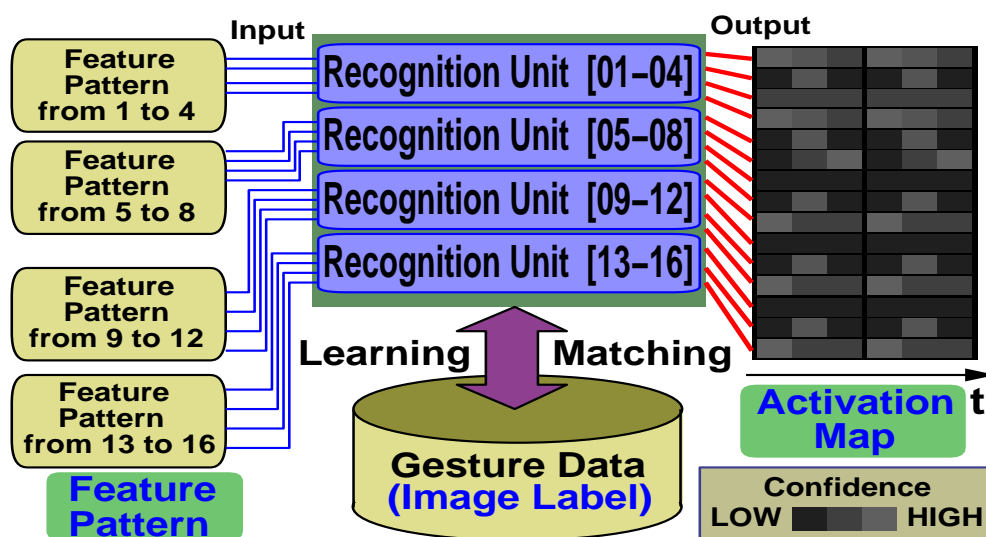


図 3.6 身振り特徴量ベース識別処理

図 3.6 に身振り特徴量ベース識別処理における処理の流れを示す．任意の注視点 l において得られる特徴パターンを $T_l = (T_1, T_2, \dots, T_n)^t$ ，各画像ラベルに格納されている身振り標準パターンを $K_l = (K_1, K_2, \dots, K_n)^t$ とする．両パ

ターン間の距離を式 3.4 により定義する．

$$d_l^{(g)} = \sum_n ||Tn - Kn^{(g)}|| \quad (3.4)$$

ここで n は解像度， g は画像ラベル番号である．身振りカテゴリ i において $d^{(g)}$ を最小にする画像ラベルの番号を k_i とする．身振りカテゴリ i における類似度 Si を式 3.5 により定義する．

$$Si = 1 - \frac{d^{(k_i)}}{\text{Max}(d^{(g)})} \quad (3.5)$$

こうして得られる類似度 $Si(0 \leq Si \leq 1)$ と画像ラベル k_i に格納されている位相値を特徴量ベース識別処理の結果として出力する．類似度を識別結果とする理由としては，各注視点は大きさや回転などの異なる次元量であるため，識別結果を類似度という次元に統一することにより各注視点間の重み付けおよび評価値の計算を可能にするということが挙げられる．以上より識別結果は，各注視点において得られる類似度の多次元時系列パターンとなる．これを活性化マップと呼ぶ．特徴量ベースの学習処理のみでは身振りプロトコルの学習が出来ないため，特徴量ベース識別処理で得られる活性化マップを次章で説明する身振りプロトコルベース学習・認識処理のための入力データとする．

3.6 身振りプロトコルベース認識処理

図 3.7 に身振りプロトコルベース認識処理における処理の流れを示す．特徴量ベース識別処理では，各注視点における形状パターンの空間的な類似度を求めることを目的としており，類似度の時間的な変動には対応していない．時間的に一貫している注視点を見い出すには，類似度の時間軸方向の変動を考慮する必要がある．提案手法では，特徴量ベース識別処理からフレーム毎に得られる時系列パターンである活性化マップに基づき身振りプロトコルの推定を行う．身振りカテゴリ c 注視点 l における活性化マップ $X_c^{(l)} = (X_{cl1}, X_{cl2}, \dots, X_{clk})^t$ が与えられている場合，注視点 l における平均 $\hat{\mu}_l$ ，分散 $\hat{\sigma}_l$ は，

$$\hat{\mu}_l = \frac{1}{K} \sum_{k=1}^K X_{clk} \quad (3.6)$$

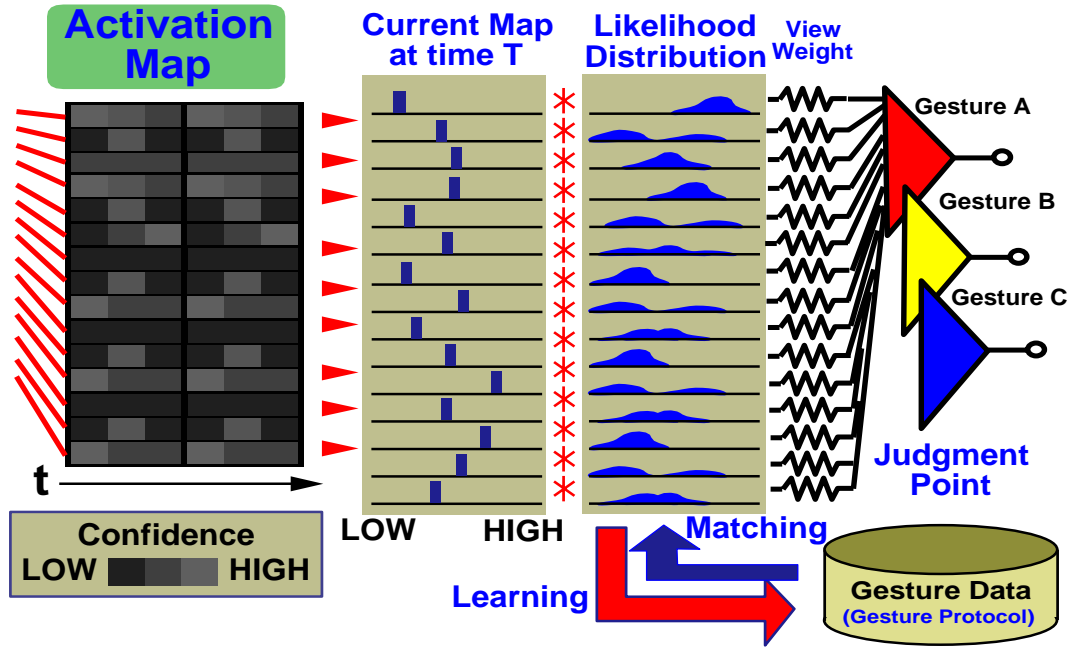


図 3.7 身振りプロトコルベース認識処理

$$\hat{\sigma}_l = \frac{1}{K} \sum_{k=1}^K (\mathbf{X}_{clk} - \hat{\mu}_l)(\mathbf{X}_{clk} - \hat{\mu}_l)^t \quad (3.7)$$

により与えられる．ここで注視点重み ω_l を式 3.8 により定義する．

$$\omega_l = \frac{\hat{\mu}_l}{\hat{\sigma}_l + \alpha} \quad (3.8)$$

ここで α は注視点強調係数であり，小さい程分散 $\hat{\sigma}_l$ が注視点重み ω_l に大きな影響を与えるようになる．また式 3.8 より分かるように，一貫した類似度が得られる場合，その注視点の重みはより大きく設定されることになる．学習時には，注視点毎に重みを算出するとともに $\mathbf{X}_c^{(l)}$ を身振りプロトコルデータとして登録しておく．この際，登録するデータの属するカテゴリはあらかじめユーザが指定しておくものとする．一方，識別時には $\mathbf{X}_c^{(l)}$ の各要素の分布型を正規分布と仮定し，フレーム番号 i の活性化マップ（類似度） $\mathbf{X}_s^{(l)}$ が入力された場合の評価値 $Ei^{(c)}$ を

$$Ei^{(c)} = \sum_{l=1}^L \omega_l \sum_{k=1}^K \exp\{-\beta(\mathbf{X}_s^{(l)} - \mathbf{X}_{clk})^2\} \quad (3.9)$$

により定義する． β は分離係数であり，大きい程要素間の相互作用が小さくなる．ここで，評価値 $E_i^{(c)}$ をすべての類似度区間 $[0, 1.0]$ において計算し生成した尤度分布を注視点 c における身振りプロトコルマップと呼ぶことにする．

一方， N フレームの画像が入力されている場合の評価値 $E^{(c)}$ を，

$$E^{(c)} = W_c \sum_{n=1}^N E_n^{(c)} \quad (3.10)$$

により表す．ここで W_c は各身振りカテゴリに与える重みであり，全注視点数を M とした場合，

$$W_c = \frac{M}{\sum_{m=1}^M \omega_m} \quad (3.11)$$

により定義する．式 3.11 より，身振りプロトコル学習の結果，注目すべき注視点が決まっている程，カテゴリ重み W_c は大きく設定されることになる．本システムの認識対象である身振りは，基本的に動作の開始と終了が時間差分量により検出できるものである．これは，身振り動作が「構え（低速）」 「ストローク（高速）」 「戻し（高速）」の三要素から成り立つとの A.Kendon et al.[26], F.Quek et al.[27] の主張に基づくものである．以上の理由から，動作開始からの評価値 $E_i^{(c)}$ を加算し始めて，現在までの評価値 $E^{(c)}$ が最大になる身振りカテゴリに入力された身振りが属するものと判定する．

3.7 身振り情報の定量化

入力された身振りのカテゴリ推定を行った後，推定された身振りカテゴリの活性化マップに格納されている位相値の時系列変化から，身振り標準パターンと比べた場合の身振り位相値や身振り速度，身振り振幅などの身振り情報を算出する．

3.7.1 身振り位相値の算出

推定された身振りカテゴリ \hat{c} の注視点 l における活性化マップの位相値の時系列パターンを $P_{\hat{c}}^{(l)} = (P_{\hat{c}l1}, P_{\hat{c}l2}, \dots, P_{\hat{c}lk})^t$ とする．今，フレーム番号 k における入力身振り画像の位相値 \hat{P}_{0k} を，

$$\hat{P}_{0k} = \frac{\sum_{l=1}^L P_{\hat{c}lk} \times Z_{lk}}{\sum_{l=1}^L Z_{lk}} \quad (3.12)$$

により算出する．ここで Z_{lk} はフレーム番号 k , 注視点 l における評価値重みであり ,

$$Z_{lk} = \sum_{i=1}^K \sum_{n=1}^N \exp\{-\beta(S_i^{(l)} - \mathbf{X}_{cln})^2\} \quad (3.13)$$

により与えられる． β は前節で述べた分離係数である．位相値 \hat{P}_{0k} の急激な変動を抑えるために , 移動平均 (フィルタ長 U) を位相値 \hat{P}_{0k} に適用した結果得られる \hat{P}_k を推定身振り位相値とする．

$$\hat{P}_k = \frac{\sum_{u=0}^{U-1} \hat{P}_{0k-u}}{U} \quad (3.14)$$

一方 , フレーム番号 k における各注視点の位相値の平均 $\hat{\mu}$, 分散 $\hat{\sigma}$ は ,

$$\hat{\mu} = \frac{1}{L} \sum_{l=1}^L \mathbf{P}_{cll} \quad (3.15)$$

$$\hat{\sigma} = \frac{1}{L} \sum_{l=1}^L (\mathbf{P}_{cll} - \hat{\mu})(\mathbf{P}_{cll} - \hat{\mu})^t \quad (3.16)$$

により与えられる．分散 $\hat{\sigma}$ は , 位相値 \hat{P}_k の信頼性の指標にもなり , 認識結果をリジェクトする際に利用できる．単一特徴量に基づく身振り認識手法では , このような指標を利用できない．認識結果をリジェクトする際の指標を容易に獲得できるということも多注視点身振り認識法の長所の一つであると考えられる．

3.7.2 身振り速度と身振り振幅の算出

身振り速度 \hat{v} と身振り振幅 \hat{w} を式 3.17 と式 3.18 により算出する．

$$\hat{v} = \hat{P}_k - P_{k-\Delta} \quad (3.17)$$

$$\hat{w} = \text{Max}(\hat{P}_k) - \text{Min}(\hat{P}_k) \quad (3.18)$$

ここで式 3.17 の Δ は , 微小時間幅に対応するフレーム数である．式 3.17 より分かるように , 登録時の身振り動作の方向を順方向とすれば , 順方向の身振り動作の際 v は正 , 停止の際には 0 , 逆方向身振り動作の際には負の値をとる．位相値 \hat{P}_k は画像ラベル (身振り標準パターン) に基づくために , 推定可能な身振り動作の速度や振幅に関する情報は物理的なものではなく相対的な量である．

第4章 評価実験

第3章では，身振り認識の基本問題として存在する身振りプロトコルに認識系を適応させるための枠組み，およびその枠組みを実現するための各種手法を提案した．本章では，提案手法の有効性を確認するために行った評価実験の結果を示す．

4.1 評価実験用身振りについて

現在，画像処理アルゴリズムの客観的評価を行うためのS I D B Aの様な標準データベースは，身振りに関しては存在していない．従って，現状では任意で実験に利用する身振りを定義せざるを得ない．そこで，本論文では表4.1に示す身振りを採用した．各身振りクラスの簡単な説明を以下に示すが，詳細は[15]を参

表 4.1 採用した評価実験用身振り

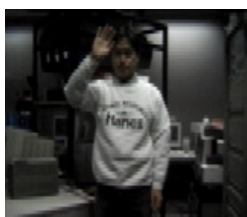
身振りクラス	身振り動作名	略記表現	標準身振り図番号
標識（コード）	「バイバイ」	（G-A）	図 4.1
標識（コード）	「グッパ」	（G-B）	図 4.2
例示子（模倣）	「鳥のまね」	（G-C）	図 4.3
情感表示	「バンザイ」	（G-D）	図 4.4
調整子	「聞き返し」	（G-E）	図 4.5
環境適応子	「腕組み」	（G-F）	図 4.6
オブジェクト適応子	「マウス操作」	（G-G）	図 4.7

照されたい．標識身振りとは，明示的情報の伝達を目的とした身振りであり特定

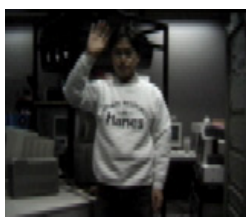
の集団ごとに発達し、その成員すべてがそのコードを知っていることを期待される身振りである。例示子身振りとは、対象の形や大きさを空間に描いたり、出来事のリズムを示す動作などに相当するもので会話内容の強調や補足を行うものである。情感表示身振りとは、情感表示の基本的な場は表情であるが、表情に出る感情を補足する場合にとられる二次的な身体動作である。調整子身振りとは、話す順番を決定したり、発話権のやりとりを制御したり、会話の流れを円滑にする機能をもつ身体動作である。環境適応子身振りとは、身体的要求を満たしたり、情緒を管理したりといった、状況や環境に適応するための身体動作である。オブジェクト適応子とは、道具や機械を使用するために学習されるものであるが、関連する情緒や構えが刺激されると、会話中であっても生じることがある身体動作である。

各身振りに対応して定義した標準身振りを図 4.1 から図 4.7 に示す。これらの標準身振りに加えて、提案手法の柔軟性を評価するために、位置に関する派生身振り（本実験では、標準身振りは画面中央位置での動作を指し、派生身振りは画面左寄りと画面右寄りの位置での動作を指す）を定義した。また、多注視点身振り認識処理の有効性評価実験のために作成した身振り「バイバイ」と「グッパ」の身振り画像のスナップショットを、標準身振りサンプルに関しては図 4.8 と図 4.10 に、派生身振りサンプルに関しては図 4.9 と図 4.11 にそれぞれ示す。さらに、11 名の被験者の身振り画像のスナップショットを図 4.12 から図 4.22 に示す。なお、被験者の身振りサンプルを収集する際には、例えば、「バイバイの身振りをしてください」などといったキーワードにより身振り動作を求めた。この条件設定により、被験者それぞれが有している身振り概念に対する特徴的動作を獲得することができると考えられる。

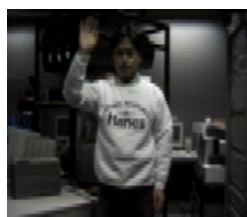
以上に定義した評価実験用身振り動画は、デジタルビデオにより撮影されたものを利用し、テープを再生することで繰り返しオンラインでの認識実験を行う。リジェクト処理のための閾値の設定については、実験前に任意の値に設定した。各種評価実験の間には、閾値を固定し認識結果の客観的な比較ができるようにした。なお、提案手法が背景や照明に関してロバストであることを確認するために、身振り画像の撮影時には特別な背景や照明は利用していない。



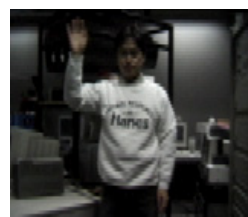
バイバイ#1



バイバイ#2



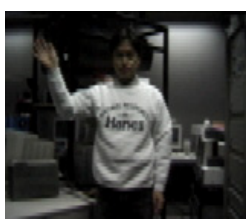
バイバイ#3



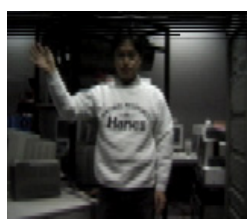
バイバイ#4



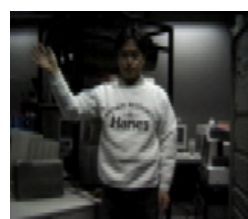
バイバイ#5



バイバイ#6



バイバイ#7



バイバイ#8

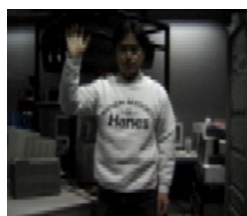
図 4.1 標識身振り (バイバイ)



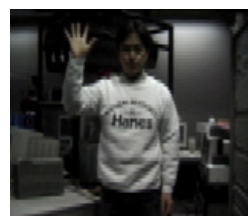
グッパ#1



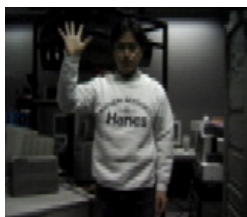
グッパ#2



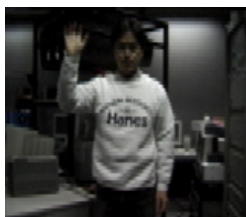
グッパ#3



グッパ#4



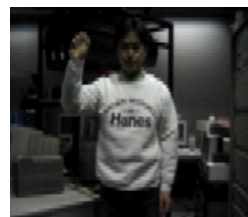
グッパ#5



グッパ#6

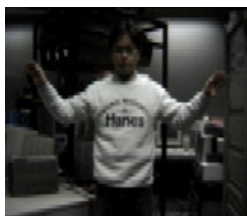


グッパ#7

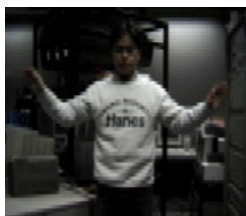


グッパ#8

図 4.2 標識身振り (グッパ)



鳥のまね#1



鳥のまね#2



鳥のまね#3



鳥のまね#4



鳥のまね#5



鳥のまね#6



鳥のまね#7



鳥のまね#8

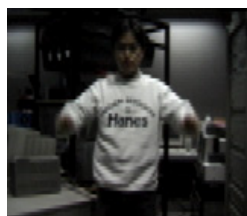
図 4.3 例示子身振り（鳥のまね）



バンザイ#1



バンザイ#2



バンザイ#3



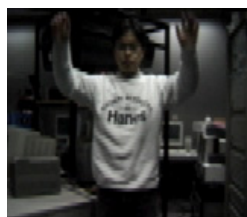
バンザイ#4



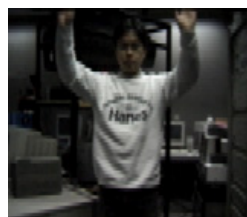
バンザイ#5



バンザイ#6

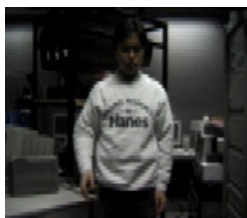


バンザイ#7

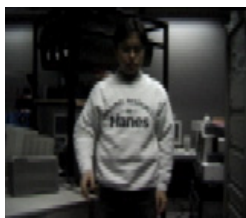


バンザイ#8

図 4.4 情感表示身振り（バンザイ）



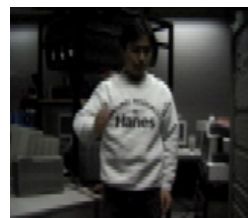
聞き返し#1



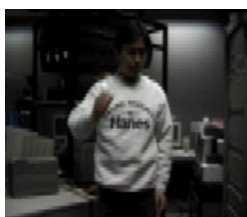
聞き返し#2



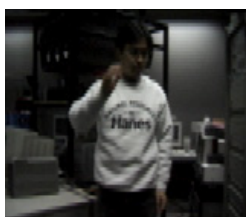
聞き返し#3



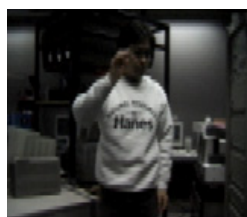
聞き返し#4



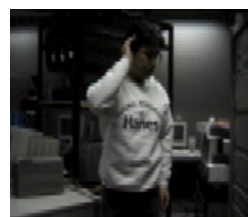
聞き返し#5



聞き返し#6



聞き返し#7



聞き返し#8

図 4.5 調整子身振り（聞き返し）



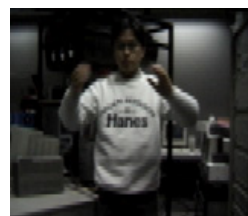
腕組み#1



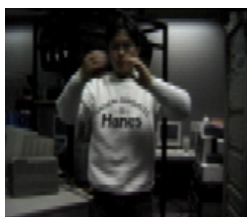
腕組み#2



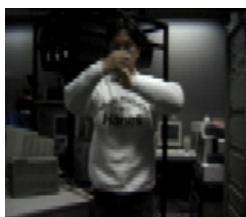
腕組み#3



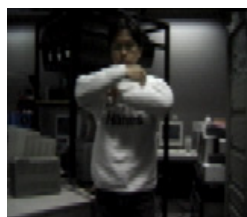
腕組み#4



腕組み#5



腕組み#6



腕組み#7



腕組み#8

図 4.6 環境適応子身振り（腕組み）



マウス操作#1



マウス操作#2



マウス操作#3



マウス操作#4



マウス操作#5



マウス操作#6

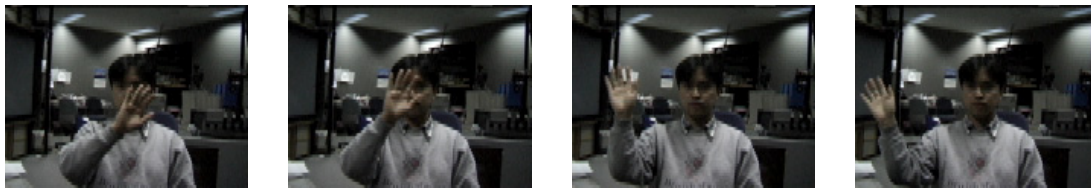


マウス操作#7



マウス操作#8

図 4.7 オブジェクト適応子身振り（マウス操作）



バイバイ #1

バイバイ #2

バイバイ #3

バイバイ #4

図 4.8 標準身振りサンプルのスナップショット（身振り「バイバイ」）



バイバイ A



バイバイ B



バイバイ C



バイバイ D



バイバイ E



バイバイ F



バイバイ G



バイバイ H



バイバイ I



バイバイ J

図 4.9 派生身振りサンプルのスナップショット（身振り「バイバイ」）



グッパ#1

グッパ#2

グッパ#3

グッパ#4

図 4.10 標準身振りサンプルのスナップショット（身振り「グッパ」）



グッパ A



グッパ B



グッパ C



グッパ D



グッパ E



グッパ F



グッパ G



グッパ H



グッパ I



グッパ J

図 4.11 派生身振りサンプルのスナップショット（身振り「グッパ」）



「バイバイ」



「グッパ」



「鳥のまね」



「バンザイ」



「聞き返し」



「腕組」



「マウス操作」

図 4.12 被験者 A の身振り動作のスナップショット



「バイバイ」



「グッパ」



「鳥のまね」



「バンザイ」



「聞き返し」

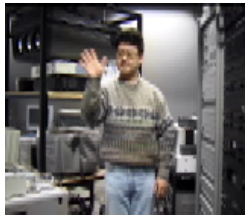


「腕組」



「マウス操作」

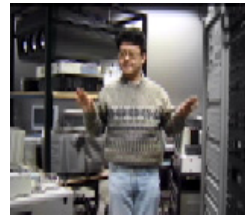
図 4.13 被験者 B の身振り動作のスナップショット



「バイバイ」



「グッパ」



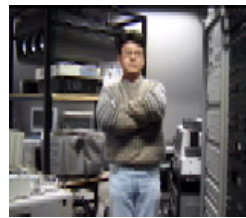
「鳥のまね」



「バンザイ」



「聞き返し」



「腕組」

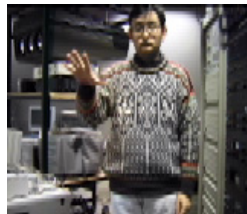


「マウス操作」

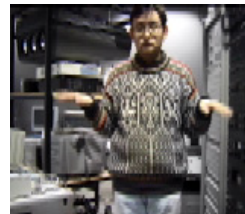
図 4.14 被験者 C の身振り動作のスナップショット



「バイバイ」



「グッパ」



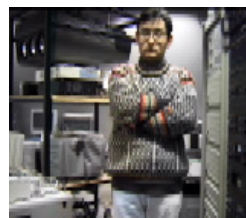
「鳥のまね」



「バンザイ」



「聞き返し」



「腕組」



「マウス操作」

図 4.15 被験者 D の身振り動作のスナップショット



「バイバイ」



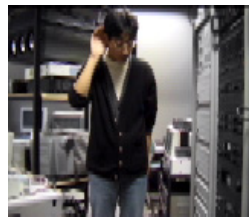
「グッパ」



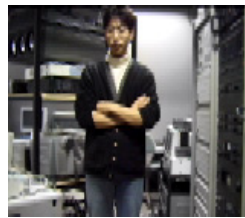
「鳥のまね」



「バンザイ」



「聞き返し」



「腕組」



「マウス操作」

図 4.16 被験者 E の身振り動作のスナップショット



「バイバイ」



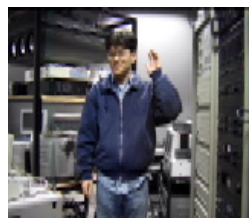
「グッパ」



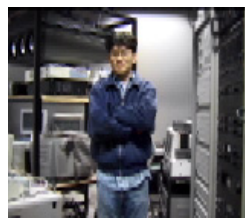
「鳥のまね」



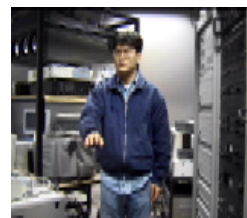
「バンザイ」



「聞き返し」

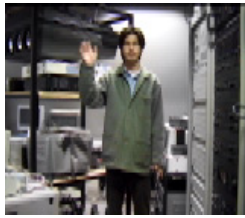


「腕組」



「マウス操作」

図 4.17 被験者 F の身振り動作のスナップショット



「バイバイ」



「グッパ」



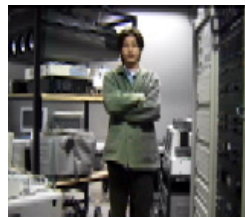
「鳥のまね」



「バンザイ」



「聞き返し」



「腕組」



「マウス操作」

図 4.18 被験者 G の身振り動作のスナップショット



「バイバイ」



「グッパ」



「鳥のまね」



「バンザイ」



「聞き返し」



「腕組」



「マウス操作」

図 4.19 被験者 H の身振り動作のスナップショット



「バイバイ」



「グッパ」



「鳥のまね」



「バンザイ」



「聞き返し」



「腕組」



「マウス操作」

図 4.20 被験者 I の身振り動作のスナップショット



「バイバイ」



「グッパ」



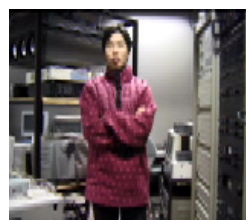
「鳥のまね」



「バンザイ」



「聞き返し」



「腕組」



「マウス操作」

図 4.21 被験者 J の身振り動作のスナップショット



「バイバイ」



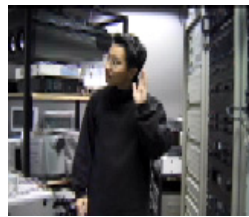
「グッパ」



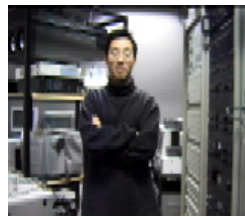
「鳥のまね」



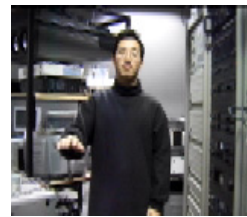
「バンザイ」



「聞き返し」



「腕組」



「マウス操作」

図 4.22 被験者 K の身振り動作のスナップショット

4.2 システム構成

評価実験を行うために本論文で提案した手法をWSに実装した．実験システムの構成を図 4.23 に示す．CCD カメラにより入力されるユーザのモノクロの身振り画像 (160dot \times 120dot) は，WSに取り込まれ提案手法により認識される．なお本実験では，シリコングラフィクス社の INDIGO2TM と画像入力ボード Galileo video を使用した．提案手法の高速化のための特殊なハードウェアは一切使用していない．



図 4.23 実験システム

4.3 身振りの学習方法

4.1節において定義した身振りの登録・学習作業は，身振りインタフェース用の GUI を通して行う．図 4.24 に作成した身振りインタフェース用 GUI の画面を示す．図 4.24 に示すように，身振りインタフェース用 GUI は，内部情報表示のためのウィンドウと現在のビデオ入力画像表示用ウィンドウおよび，定義済み身振

り標準パターン表示用ウィンドウから構成される．身振り登録の際には，各身振り標準パターン表示用ウィンドウ内において，マウスの左ボタンのクリックで身振り標準パターンの登録（特徴量ベースの学習），中央ボタンのクリックで派生身振りの学習（身振りプロトコルベースの学習），右ボタンのクリックで身振り標準パターンの再生（replay）を行うことが可能である．学習の際には，数秒の準備時間の後，ユーザはカメラの前で身振り動作を行う（本論文の実験では，認識結果の相互比較を可能とするためにデジタルビデオを使用して身振り画像を入力した）．なお，認識結果（トラッキング情報）は，各身振り標準パターン表示用ウィンドウ内にフレームワイズで表示される．作成した身振りインタフェース



図 4.24 身振りインタフェース用 GUI

用 GUI においては，以上に述べた機能のみでなく，各種身振り情報（認識結果）のグラフ表示機能，処理画像表示機能，身振りデータの保存および読出し機能，リジェクト閾値などの各種閾値の設定機能を利用することが可能である．また，身振りプロトコルマップの表示機能も利用でき，実時間で内部状態を確認することも可能である．

4.4 身振り認識実験

提案手法による身振り認識処理が，ユーザの衣服に依存するかしないかを調べるために，図 4.25 に示す 6 種類の衣服 A～F を利用して身振り認識実験を行った．



衣服 A (モザイク模様のシャツ)



衣服 B (白のトレーナー)



衣服 C (グレーのトレーナー)



衣服 D (縦縞模様のシャツ)



衣服 E (横縞模様のシャツ)



衣服 F (黒のジャージ)

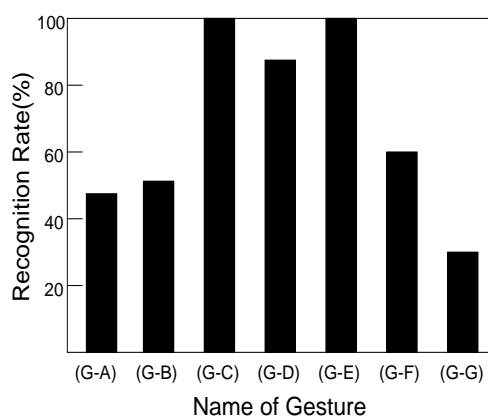
図 4.25 使用した 6 種類の衣服

4.4.1 標準身振りサンプルの認識結果

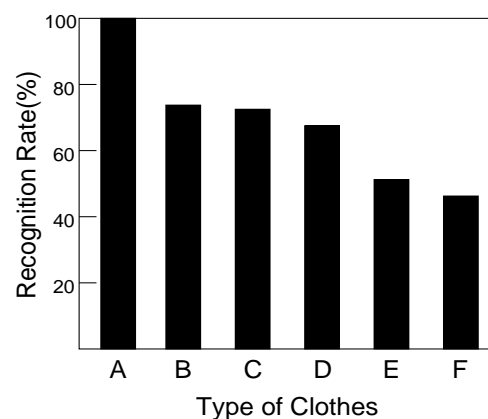
衣服 A の標準身振りサンプルを用いて身振り標準パターンを登録し，衣服 A の標準身振りサンプルにより身振りプロトコル学習を行った．学習処理の結果として得られた身振りデータに基づき，全ての標準身振りサンプル（各 10 パターン）の認識実験を行った際に得られた認識率を表 4.2 に示す．認識率は，正答サンプル数をテストサンプル数により割ることで求めた．衣服 A に関する身振りデータを用いた際に得られた各身振り・衣服における平均認識率を図 4.26 に示す．

表 4.2 標準身振りサンプルの認識結果

	(G-A)	(G-B)	(G-C)	(G-D)	(G-E)	(G-F)	(G-G)	認識率 (%)
衣服 A	100	100	100	100	100	100	100	100
衣服 B	0	80	100	100	100	50	90	74
衣服 C	60	80	100	100	100	70	0	73
衣服 D	30	50	100	80	100	90	0	64
衣服 E	20	10	100	90	100	50	0	53
衣服 F	60	0	100	60	100	0	0	46
認識率 (%)	45	53	100	88	100	60	32	68



各身振りにおける平均認識率



各衣服における平均認識率

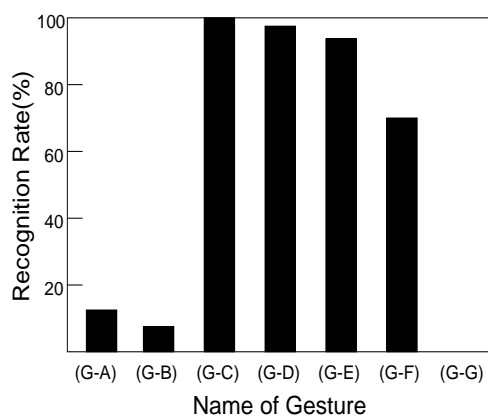
図 4.26 標準身振りサンプルの認識結果

4.4.2 派生身振りサンプルの認識結果 (I)

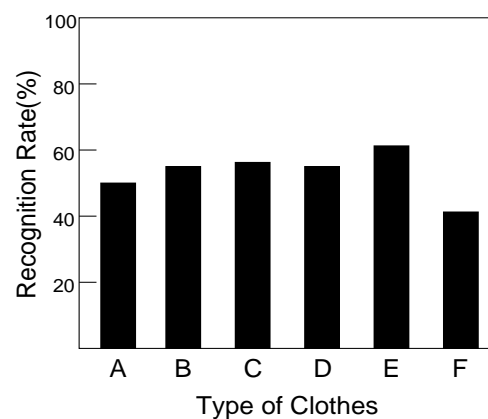
衣服 A の標準身振りサンプルを用いて身振り標準パターンを登録し，衣服 A の標準身振りサンプルにより身振りプロトコル学習を行った．学習処理の結果として得られた身振りデータに基づき，全ての派生身振りサンプル（各 10 パターン）の認識実験を行った際に得られた認識率を表 4.3 に示す．また，衣服 A に関する身振りデータを用いた際に得られた各身振り・衣服における平均認識率を図 4.27 に示す．

表 4.3 派生身振りサンプルの認識結果

	(G-A)	(G-B)	(G-C)	(G-D)	(G-E)	(G-F)	(G-G)	認識率 (%)
衣服 A	0	0	100	100	90	70	0	51
衣服 B	10	10	100	100	100	80	0	57
衣服 C	20	0	100	100	100	90	0	58
衣服 D	10	0	100	90	100	100	0	57
衣服 E	40	10	100	100	100	80	0	61
衣服 F	0	20	100	100	70	0	0	41
認識率 (%)	13	7	100	98	93	70	0	54



各身振りにおける平均認識率



各衣服における平均認識率

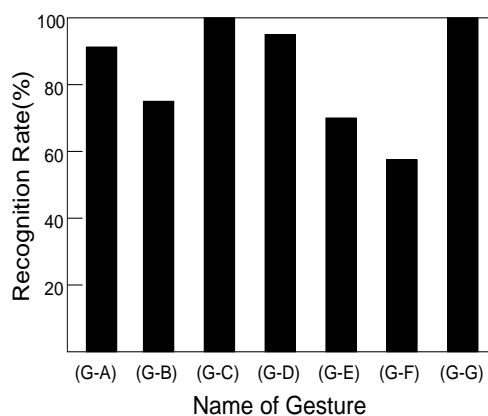
図 4.27 派生身振りサンプルの認識結果 (I)

4.4.3 派生身振りサンプルの認識結果 (II)

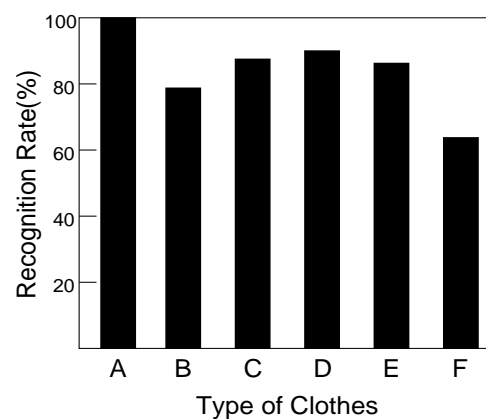
衣服 A の標準身振りサンプルを用いて身振り標準パターンを登録し，衣服 A の派生身振りサンプルにより身振りプロトコル学習を行った．学習処理の結果として得られた身振りデータに基づき，全ての派生身振りサンプル（各 10 パターン）の認識実験を行った際に得られた認識率を表 4.4 に示す．また，衣服 A に関する身振りデータを用いた際に得られた各身振り・衣服における平均認識率を図 4.28 に示す．

表 4.4 派生身振りサンプルの認識結果

	(G-A)	(G-B)	(G-C)	(G-D)	(G-E)	(G-F)	(G-G)	認識率 (%)
衣服 A	100	100	100	100	100	100	100	100
衣服 B	70	60	100	100	60	60	100	79
衣服 C	90	90	100	100	80	50	100	87
衣服 D	100	60	100	100	90	80	100	90
衣服 E	100	70	100	90	80	60	100	86
衣服 F	90	70	100	80	0	0	100	63
認識率 (%)	92	75	100	95	68	58	100	84



各身振りにおける平均認識率



各衣服における平均認識率

図 4.28 派生身振りサンプルの認識結果 (II)

4.5 他人身振りの認識実験

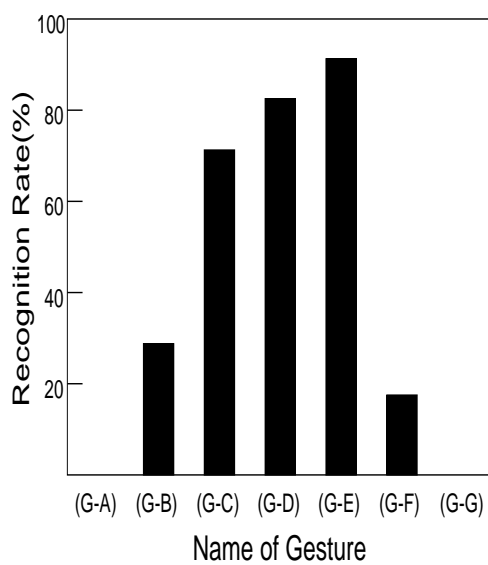
本節では，単一ユーザにより作成された身振りデータを用いて，他のユーザの身振り（他人身振り）の認識実験を行った際の実験結果を報告する．

4.5.1 最も個人に依存する設定で実験を行った場合

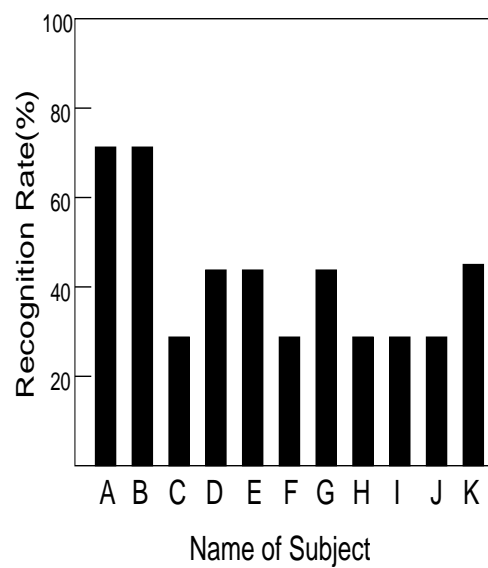
衣服 A の標準身振りサンプルを用いて身振り標準パターンを登録し，衣服 A の標準身振りサンプルにより身振りプロトコル学習を行った．学習処理の結果として得られた身振りデータに基づき，11 名の被験者の身振りの認識実験を行った際に得られた結果を表 4.5 に示す．表 4.5 においては，認識結果が正しい場合○，間違った場合× にて認識結果を示した．また，各身振りおよび各被験者における平均認識率を図 4.29 に示す．

表 4.5 他人身振りサンプルの認識結果 (I)

	(G-A)	(G-B)	(G-C)	(G-D)	(G-E)	(G-F)	(G-G)	認識率 (%)
被験者 A	×	○	○	○	○	○	×	71
被験者 B	×	○	○	○	○	○	×	71
被験者 C	×	×	×	○	○	×	×	29
被験者 D	×	×	○	○	○	×	×	43
被験者 E	×	×	○	○	○	×	×	43
被験者 F	×	×	○	×	○	×	×	29
被験者 G	×	○	○	×	○	×	×	43
被験者 H	×	×	×	○	○	×	×	29
被験者 I	×	×	×	○	○	×	×	29
被験者 J	×	×	○	○	×	×	×	29
被験者 K	×	×	○	○	○	×	×	43
認識率 (%)	0	27	72	82	91	18	0	41



各身振りにおける平均認識率



各被験者における平均認識率

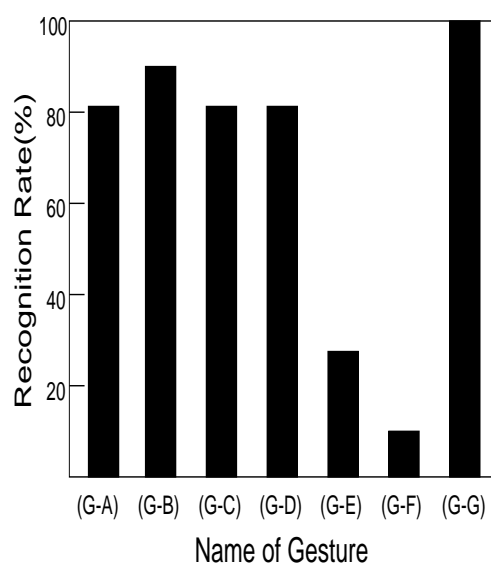
図 4.29 他人身振りサンプルの認識結果 (I)

4.5.2 最も個人に依存しない設定で実験を行った場合

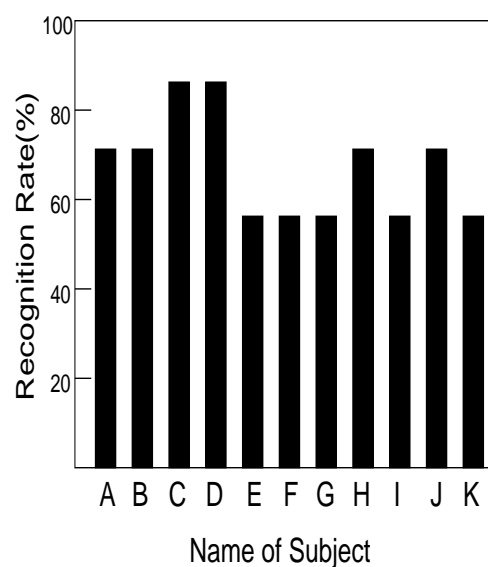
衣服 A の標準身振りサンプルを用いて身振り標準パターンを登録し，衣服 A ～ F の全ての派生身振りサンプルにより身振りプロトコル学習を行った．学習処理の結果として得られた身振りデータに基づき，11 名の被験者の身振りの認識実験を行った際に得られた結果を表 4.6 に示す．表 4.6 においては，認識結果が正しい場合○，間違った場合× にて認識結果を示した．また，各身振りおよび各被験者における平均認識率を図 4.30 に示す．

表 4.6 他人身振りサンプルの認識結果 (II)

	(G-A)	(G-B)	(G-C)	(G-D)	(G-E)	(G-F)	(G-G)	認識率 (%)
被験者 A	×	○	○	○	×	○	○	71
被験者 B	○	○	○	○	×	×	○	71
被験者 C	○	○	○	○	○	×	○	86
被験者 D	○	○	○	○	○	×	○	86
被験者 E	○	○	×	○	×	×	○	57
被験者 F	○	○	○	×	×	×	○	57
被験者 G	○	○	○	×	×	×	○	57
被験者 H	○	○	○	○	×	×	○	71
被験者 I	○	○	×	○	×	×	○	57
被験者 J	○	○	○	○	×	×	○	71
被験者 K	×	×	○	○	○	×	○	57
認識率 (%)	82	91	82	82	27	9	100	68



各身振りにおける平均認識率



各被験者における平均認識率

図 4.30 他人身振りサンプルの認識結果 (II)

4.6 多注視点身振り認識処理の有効性評価実験

本節では，多注視点戦略により身振りを安定して認識できるかどうかを調べるための実験を行った際の結果を報告する．本実験では，認識対象として身振り「バイバイ」と「グッパ」を採用した．本実験で使用する身振り画像のスナップショットを身振り「バイバイ」に関しては図 4.8，身振り「グッパ」に関しては図 4.10 に示す．実験手順としては，まず，身振り標準パターンの登録を行う．次に身振りプロトコルの学習を派生身振りを学習させることにより行う．その後，図 4.9 と図 4.11 に示したテストサンプルを使って，認識実験を行う．以上の手順に従って身振りプロトコルを学習させた際の注視点重みの変化を，図 4.31 に示す．なお，図 4.31 においては全ての注視点重みではなく，重みが最大となる注視点と最小となる注視点に関する重みの変化を示した．

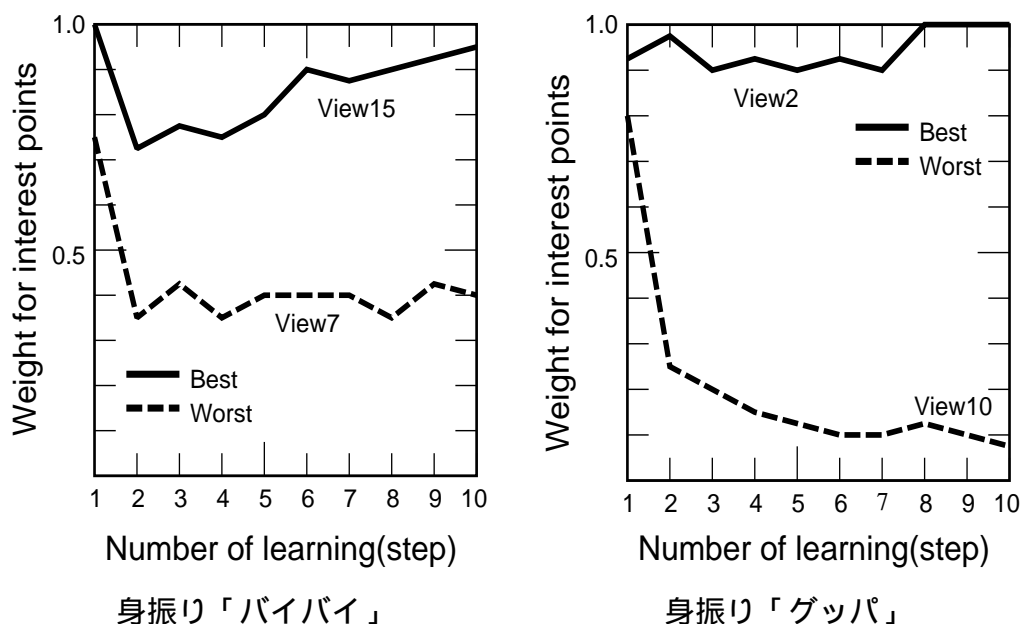


図 4.31 身振りプロトコル学習の際の注視点重みの変化

身振りプロトコル学習の後 (1) 時間差分画像で最も有効な注視点のみ (2) 背景差分画像で最も有効な注視点のみ (3) 全ての注視点で最も有効な 4 つの注視点 (4) 時間差分画像に関する全ての注視点 (5) 背景差分画像に関する

全ての注視点 (6) 全ての注視点 , の 6 種類の注視点利用条件の下で身振り認識実験を行った . ここで 3 番目の「全注視点において最も有効な 4 つの注視点」とは , 身振りプロトコルが学習の結果として得られる各注視点の重みの大きい順に 4 つの注視点を選択することを意味する . 従って , どの注視点が選ばれるかは事前には予測できず認識対象に依存する . 今回の実験では , 身振り「バイバイ」の場合 , 注視点 (15,2,10,14) が選択され , 身振り「グッパ」の場合は同様にして注視点 (2,1,8,3) が選択された . 身振りプロトコル学習の結果 , 身振り「バイバイ」と身振り「グッパ」に関して得られた注視点重みを図 4.32 に示す .

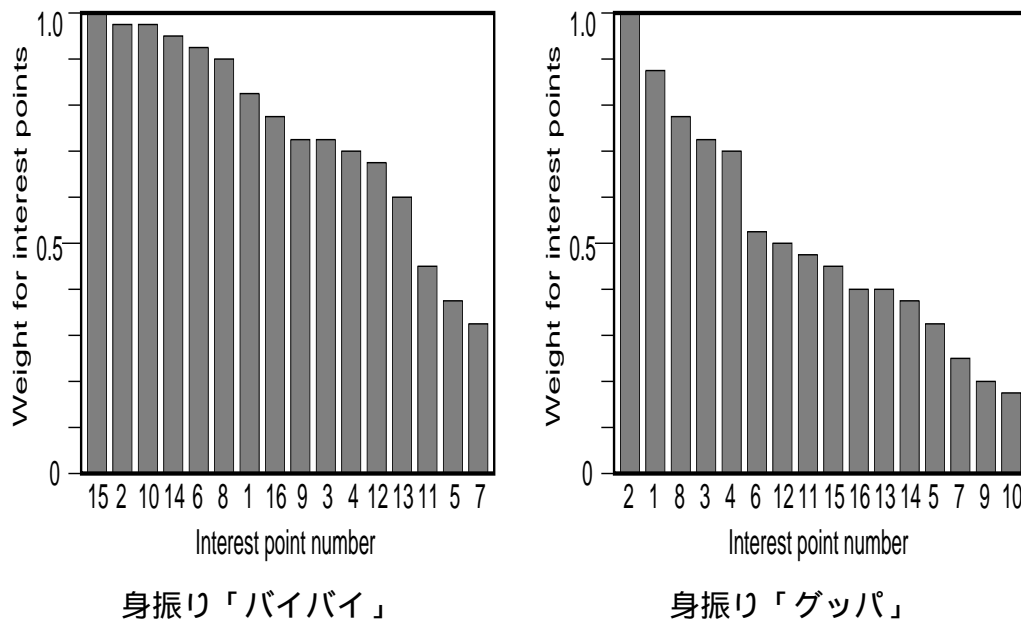


図 4.32 注視点重みの学習結果

また , 身振り「バイバイ」において注視点重みが最大である注視点 15 の身振りプロトコルマップと , 注視点重みが最小である注視点 7 の身振りプロトコルマップを図 4.33 に示す . 以上で述べた 6 種類の注視点利用条件の下で身振り認識実験を行った際の認識実験の結果を表 4.7 と表 4.8 に示す .

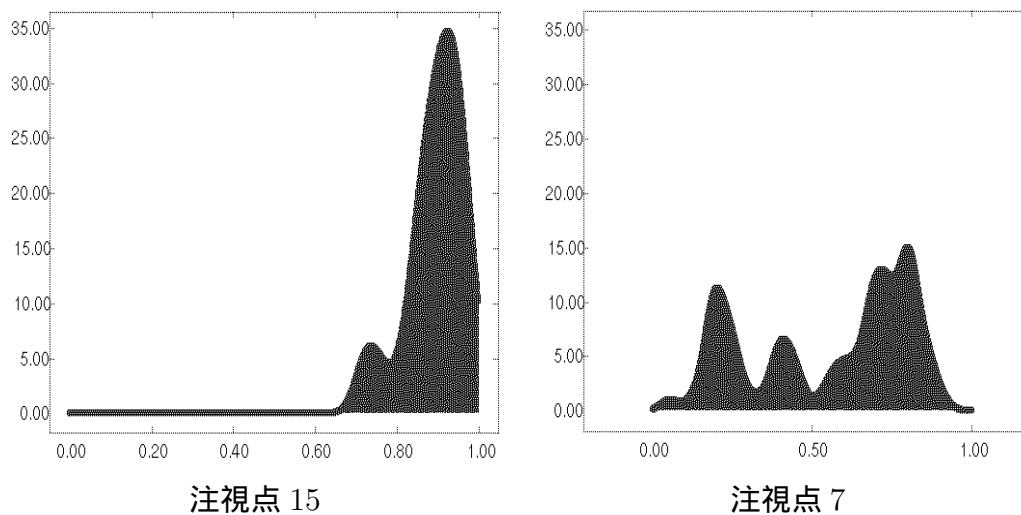


図 4.33 生成された身振りプロトコルマップ

表 4.7 身振り「バイバイ」の認識結果

実験条件	(1)	(2)	(3)	(4)	(5)	(6)
認識率 (%)	70	10	100	90	30	100

表 4.8 身振り「グッパ」の認識結果

実験条件	(1)	(2)	(3)	(4)	(5)	(6)
認識率 (%)	90	10	100	100	20	100

4.7 プロトコル学習の収束性評価実験

本節では、身振り「バイバイ」の身振りプロトコルを「右手を振っても左手を振っても同じ意味を持つ」と設定した場合の学習効率の評価を行う。具体的には、右手を振った時と左手を振った際の評価値の差とカテゴリ重みの変化を測定することにより学習効率の評価を行う。

まずカテゴリ A の標準パターンを登録するために右手を一回振る。次に右手と左手を交互に振ることで「右手を振っても左手を振っても同じ意味を持つ」という身振りプロトコルを学習させる。身振りプロトコル学習の際、右手と左手を振った際の評価値の差を測定する。カテゴリ重みは式 3.11 に従って算出されたものを測定する。カテゴリ A の標準パターンを登録する際に左手を一回振った場合でも同様にして右手と左手を振った際の評価値の差とカテゴリ重みを測定する。以上の手順に基づいて実験を行った際に得られた結果を評価値の差の変化とカテゴリ重みの変化を図 4.34 に示す。

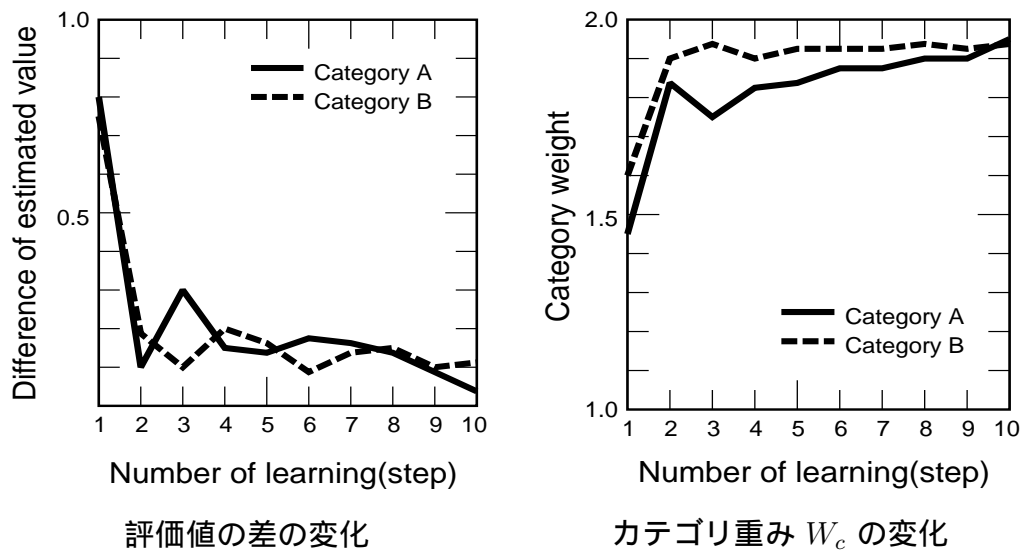


図 4.34 プロトコル学習の収束性評価実験結果

4.8 身振り認識処理の実時間性評価実験

提案手法により身振り認識処理を実時間で実行できることが可能であることを評価するために、身振り標準パターンを増加させて行った際の認識処理時間の変化を調べた。図 4.35 に身振り標準パターンを増加させた際のシステム動作周波数の変化の様子を示す。なお、図 4.35 は特別なハードウェアを一切使用しないで得られた結果である。

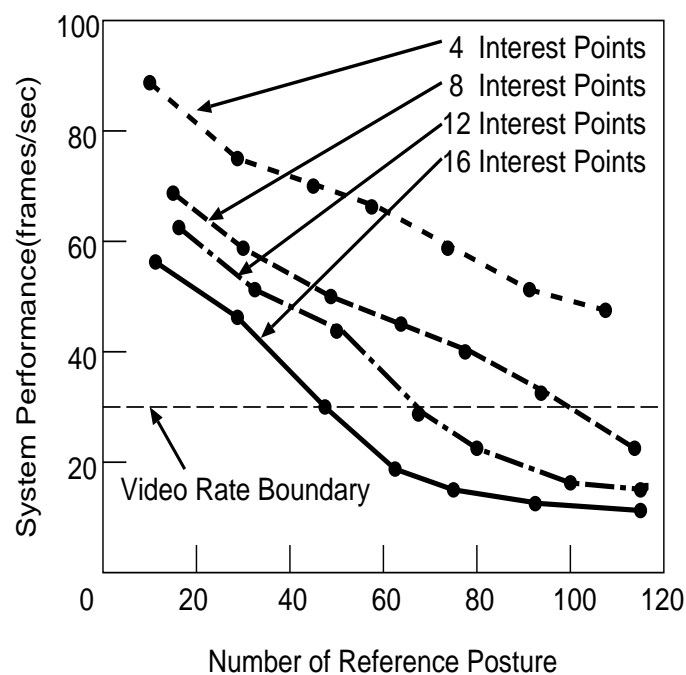


図 4.35 システム動作周波数の変化

4.9 身振り情報の認識実験

身振り「バイバイ」の動作を反復した際に得られた身振り情報の算出結果を不規則に動作した際の位相値の変化を図 4.36，身振り速度の変化を図 4.37，身振り振幅の変化を図 4.38 に示す。また，規則的に動作した際の位相値の変化を図 4.39，身振り速度の変化を図 4.40，身振り振幅の変化を図 4.41 に示す。

Phase Fluctuation

Phase

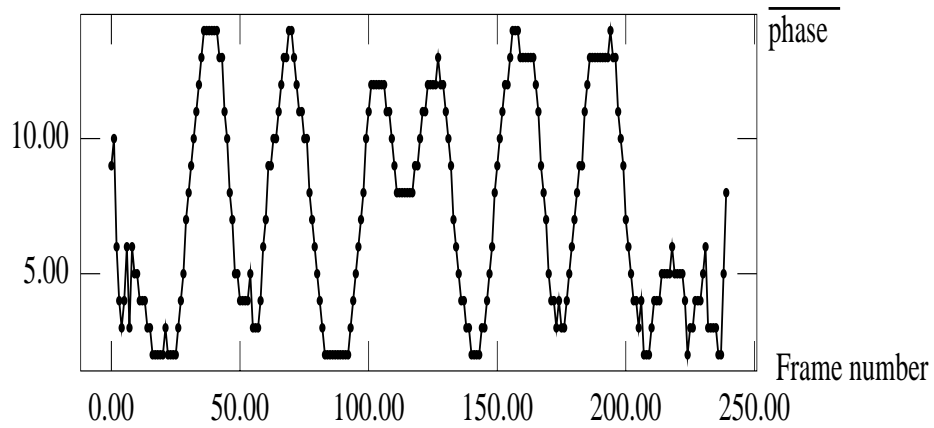


図 4.36 不規則に身振り動作した際の位相値の変化

Speed Fluctuation

speed

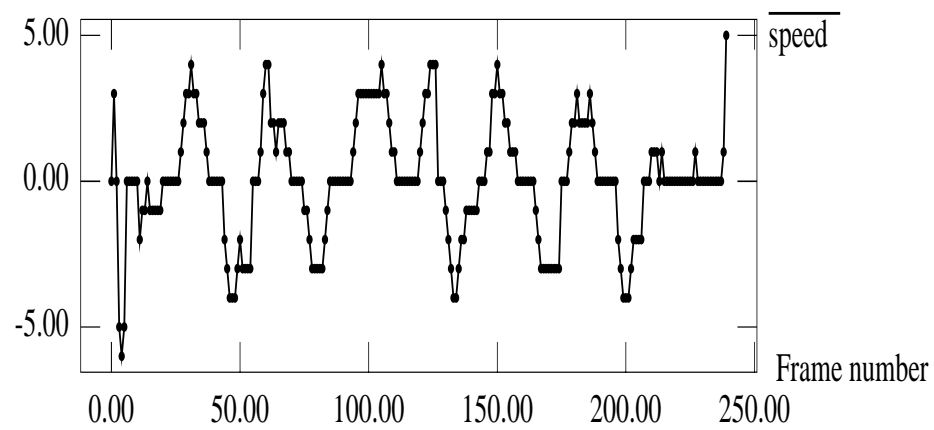


図 4.37 不規則に身振り動作した際の身振り速度の変化

Width Fluctuation

width

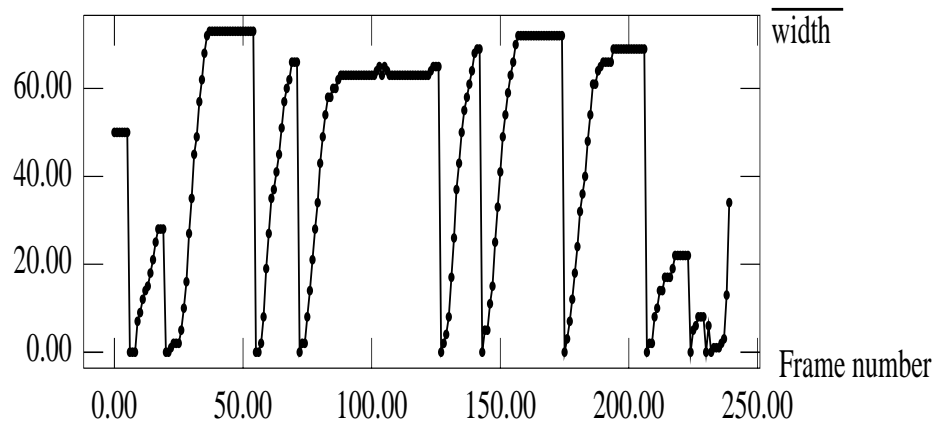


図 4.38 不規則に身振り動作した際の身振り振幅の変化

Phase Fluctuation

phase

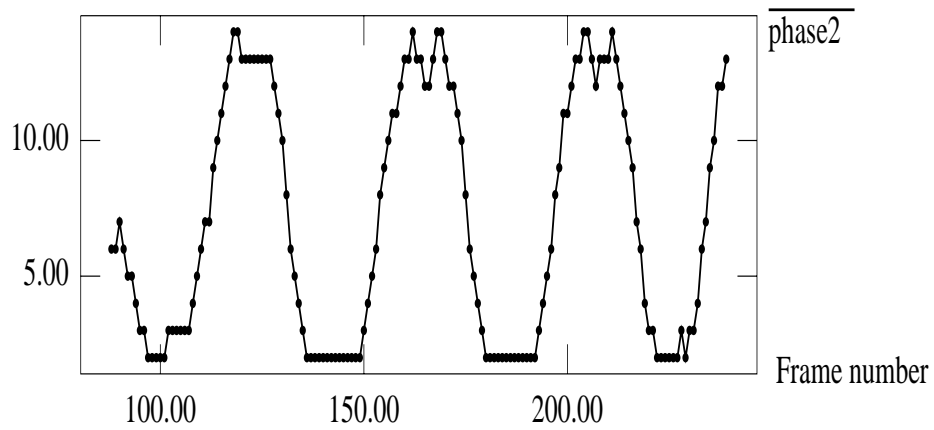


図 4.39 規則的に身振り動作した際の位相値の変化

Speed Fluctuation

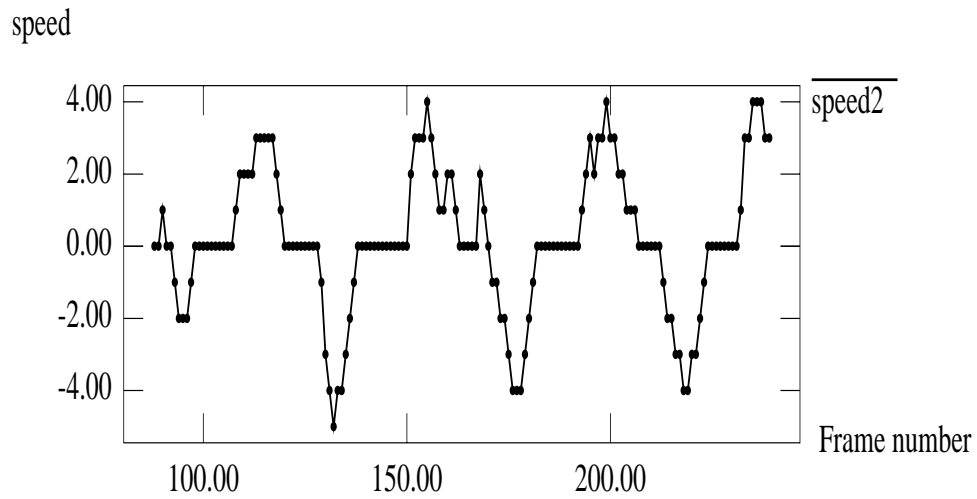


図 4.40 規則的に身振り動作した際の身振り速度の変化

Width Fluctuation

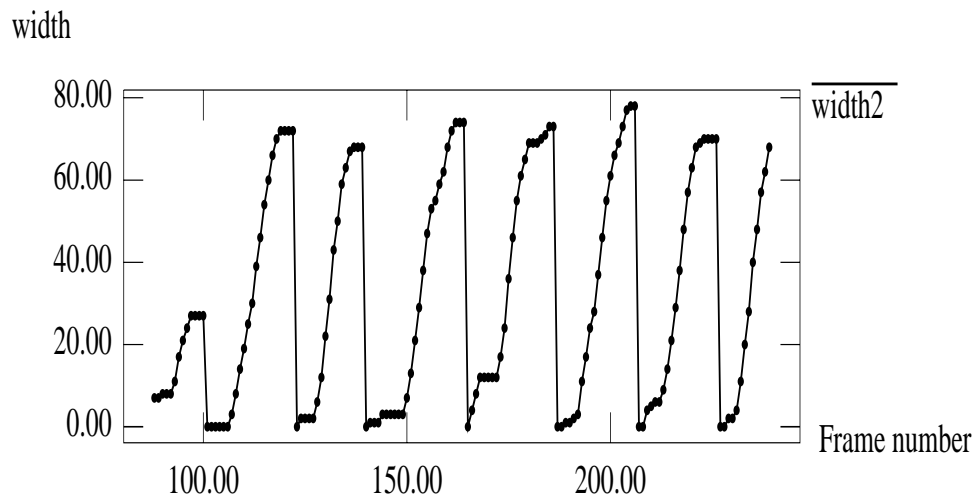


図 4.41 規則的に身振り動作した際の身振り振幅の変化

第 5 章 考察および検討

5.1 身振り認識実験の結果に関する考察

5.1.1 標準身振りサンプルの認識結果について

表 4.2 に示すように，衣服 A の標準身振りサンプルを用いて身振り標準パターンを登録し，衣服 A の標準身振りサンプルにより身振りプロトコル学習を行った場合，衣服 A に関する標準身振りサンプル全てに関して正答を得ることができた．しかしながら，図 4.26 に示すように，衣服 A 以外の衣服 B～F に関しては 80(%) 以下の平均認識率しか得られていない．この結果は，単一の衣服による訓練サンプルのみを用いたために，衣服選択性が生じたことを示していると考えられる．一方，各身振りにおいても平均認識率が著しく異なる結果が得られた．この結果は，テストサンプルが標準身振りサンプルであっても衣服が異なる場合，身振りによっては平均認識率に強い影響を受けることを示していると同時に，身振り認識の難しさは認識対象の表面的差異の吸収にあることを示している．

5.1.2 派生身振りサンプルの認識結果 (I) について

表 4.3 に示すように，衣服 A の標準身振りサンプルを用いて身振り標準パターンを登録し，衣服 A の標準身振りサンプルにより身振りプロトコル学習を行い，全ての派生身振りサンプルに関して認識実験を行ったところ，全体で 54(%) と低い認識率しか得られないことが分かった．また，図 4.26 に示した各衣服における平均認識率の標準偏差が $\sigma = 17.37$ であり図 4.27 の場合は， $\sigma = 6.59$ であることから分かるように，テストサンプルが標準身振りサンプルであった際に見られ

た衣服選択性が、テストサンプルを派生身振りサンプルにした場合、減少していることが分かる。これは、学習時に派生身振りサンプルに関する情報を与えなかったために、派生身振りサンプルの特定の衣服への選択性が付与されなかったことを示していると考えられる。また、図 4.27 に示すように、学習時に派生身振りサンプルに関する情報を与えなかった場合、図 4.26 と比べて身振り (G-A),(G-B),(G-G) の平均認識率が著しく低下しているが、身振り (G-C),(G-D),(G-E),(G-F) の平均認識率に関しては大きな変化がない。これは、後者の身振りは両手を使う動作や、左右の対称性が大きい動作であるために派生身振りサンプルでもバリエーションが前者（片手動作の身振り）に比べて乏しいということがその要因として挙げられる。

5.1.3 派生身振りサンプルの認識結果 (II) について

表 4.4 に示すように、衣服 A の標準身振りサンプルを用いて身振り標準パターンを登録し、衣服 A の派生身振りサンプルにより身振りプロトコル学習を行い、全ての派生身振りサンプルに関して認識実験を行ったところ、全体で 84(%) と、身振りプロトコル学習の際に派生身振りサンプルに関する情報を与えていない場合の結果である表 4.3 と比べて 30(%) の著しい認識率の向上が見られた。ここで重要なことは、身振りプロトコル学習の際には衣服 A の派生身振りサンプルしか用いていないにも関わらず、全体の派生身振りサンプルに対する認識率が著しく向上したことである。図 4.28 から分かるように、身振りプロトコル学習の際に、衣服 A の派生身振りサンプルのみを用いたために、衣服 A への衣服選択性（各衣服における平均認識率の標準偏差 $\sigma = 11.33$ ）が生じている。一方、身振りプロトコル学習の際に、派生身振りサンプルに関する情報を与えたために、特に片手動作の身振りである (G-A),(G-B),(G-G) の平均認識率が著しく向上していることが分かる。また、各身振りにおける平均認識率の標準偏差は $\sigma = 15.63$ となり、標準身振りサンプルの認識結果における各身振りの平均認識率の標準偏差 $\sigma = 25.53$ と派生身振りサンプルの認識結果 (I) における各身振りの平均認識率の標準偏差 $\sigma = 42.48$ と比較して、最も身振りに依存しない認識処理が達成できていることが分かった。この結果から、身振りプロトコル学習は、衣服に依存しない身振り認識処理のみでなく、身振りの種類の依存しない身振り認識処理に対し

でも有効であることが分かった．さらに，身振りプロトコル学習の際に，標準身振りサンプルのみでなく派生身振りサンプルに関する情報を与えておけば，両手動作の身振りのみでなく，片手動作の身振りに関しても著しい認識率の向上が期待できることも分かった．

5.2 他人身振りの認識実験の結果に関する考察

提案手法が，個人に依存する身振り認識と個人に依存しない身振り認識のどちらにも対応できるかどうかを調べるために，単一ユーザにより作成された身振りデータを用いて，他のユーザの身振り（他人身振り）の認識実験を行った．

最も個人に依存する設定で実験を行うために，衣服 A の標準身振りサンプルを用いて身振り標準パターンを登録し，衣服 A の標準身振りサンプルにより身振りプロトコル学習を行った．学習処理の結果として得られた身振りデータに基づき，11 名の被験者の身振りの認識実験を行った結果，表 4.5 に示すように全サンプルの平均認識率は 41(%) であった．また，各身振りにおける平均認識率の標準偏差は $\sigma = 36.31$ となり，各被験者における平均認識率の標準偏差は $\sigma = 15.18$ となった．

最も個人に依存しない設定で実験を行うために，衣服 A の標準身振りサンプルを用いて身振り標準パターンを登録し，衣服 A ~ F の全ての派生身振りサンプルにより身振りプロトコル学習を行った．学習処理の結果として得られた身振りデータに基づき，11 名の被験者の身振りの認識実験を行った結果，表 4.6 に示すように全サンプルの平均認識率は 68(%) であった．また，各身振りにおける平均認識率の標準偏差は $\sigma = 32.29$ となり，各被験者における平均認識率の標準偏差は $\sigma = 10.83$ となった．

以上より，身振りプロトコル学習の際に，個人性を強く反映するような身振りサンプルを使用することにより個人に強く依存した身振り認識が行われ，個人性をあまり反映せずに一般的な性質を持つような身振りサンプル（派生身振りサンプルなど）を使用することにより個人にあまり依存しない身振り認識が行われることが明らかとなった．また，被験者に身振り動作を要求する際に「バイバイの身振りをしてください」などといったキーワードにより行ったにも関わらず，全

サンプルの平均認識率が 68(%) に達したことは、あるキーワードに対して各個人が取る動作にはある程度の共通性があることを示していると言える。逆説的に言えば、各個人の動作に共通性があるということは、身振りプロトコルが存在していることを暗黙に示しているとも考えられ、興味深い。しかしながら、細かく観察すれば、各個人の動作（運動）特性はそれぞれ異なっており、将来的にはさらなる認識率の改善のために、各個人の動作特性の多様性を吸収することが可能なアルゴリズムの検討が必要になるものと考えられる。

5.3 多注視点身振り認識処理の有効性について

表 4.7 と表 4.8 から分かるように、全ての注視点を利用する場合と、全ての注視点で最も有効な 4 つの注視点を利用する場合、全てのテストサンプルを正しく認識することが出来ている。全体的に、時間差分画像の注視点における認識結果が背景差分画像よりも良くなっている。これは、身振りプロトコル学習の結果として、身振り「バイバイ」と身振り「グッパ」については「時間差分画像特徴に注目すれば良い」ということが提案手法により導かれたことを示している。また、表 4.7 と表 4.8 より、単一の画像特徴のみでは認識できない身振りが、多注視点アプローチに基づくプロトコル学習により認識できるようになったことが分かる。

しかしながら、現在のシステムでは n 種類の注視点を使用する場合、パターン照合処理において、単一注視点による身振り認識処理の n 倍の処理時間が必要となる問題がある。これは現在のシステムではパターン照合処理をすべて逐次方式により行っていることに原因がある。 n 種類の注視点に関する特徴抽出処理およびパターン照合処理では、注視点間の関係が独立しているので、比較的容易に並列化できるものと考えられる。

一方、特徴抽出処理およびパターン照合処理の並列化を行わなくても、選択的に注視点を使用する機構を開発することにより、逐次処理方式におけるパフォーマンスの向上を図ることも可能であると考えられる。すべての注視点を使用する方式を、全注視点方式と呼ぶとすると、限定的に注視点を使用する方式は、限定注視点方式と呼ぶことができる。限定注視点方式が採用できるのは、限定された種類の身振りしか入力されないことが保証されている場合のみである。これはす

なわち，使用できる注視点の数が減少すればするほど，「無視すべき身振り動作」のリジェクト処理が困難になることを意味している．多種多様な視点から身振り動作を観測すれば，リジェクト処理のための情報は得られやすいが，単一的な視点からの情報では，リジェクト処理のための情報は相対的に欠乏する．現在のシステムでは 16 種類の注視点を使用しているが，考えられる注視点の数が実際には無限通り存在することを考慮すると，16 種類の注視点でも相当に限定された注視点に過ぎないことは明らかである．

無限通りの注視点に基づく認識処理が不可能である以上，実際には全注視点方式は実現不可能であり，多かれ少なかれあらゆる身振り認識システムは，限定注視点方式であると言える．問題は，限定されていたとしても利用可能な注視点全体から，無意味動作をリジェクトすることが可能であると同時に，認識対象である身振りを十分に認識することを可能とするような注視点を選定することであると考えられ，今後の研究課題としたい．

5.4 プロトコル学習の収束性について

図 4.34 より数回の学習だけで，右手と左手を振った際の評価値の差は激減した後後に安定している．この傾向はカテゴリ A とカテゴリ B それぞれの実験結果において明らかである．また，図 4.34 より学習の度合の指標となるカテゴリ重み W_c に関しても数回の学習のみで急激に増加した後後に安定している．この傾向はカテゴリ A とカテゴリ B それぞれの実験結果において明らかである．ここで，評価値の差およびカテゴリ重み W_c の変化の様子が，カテゴリ A とカテゴリ B の場合において著しく類似していることは興味深く，右手と左手の対応関係が逆である場合でも同様の学習結果が得られたことを意味している．以上の実験結果より，提案手法は訓練サンプル数が少ない場合でも高い学習効率を得られることが分かった．ユーザが対話的にコンピュータに身振りを学習させる際に，このような性質は非常に重要であると言える．

5.5 身振り認識処理の実時間性について

図 4.35 より，本研究で開発したシステムでは，注視点数が 16 個であってもシステム動作周波数（（画像の入力時間 + 認識処理時間）の逆数により定義され

る) 約 30(frames/sec) 以上を実現できることが分かった。処理時間については、画像入力時間に 8(ms) 程度、特徴抽出処理に 10(ms) 程度、パターン照合処理に 10(ms) 程度 (画像ラベル数が 1600 個程度である場合)、身振りプロトコルベース認識処理に 3(ms) 程度必要であることを確認している。処理時間の内訳からも分かるように、特徴抽出処理とパターン照合処理に多くの時間が必要とされることが分かる。特徴抽出処理にかかる時間は、注視領域および注視点の数に依存しており、特に注視領域数には強く依存する。また、時間差分画像と背景差分画像の生成にかかる時間についても、処理対象とする画像の解像度次第で多くの処理時間を要する。このように画像領域を直接サーチして処理する場合には処理時間が長くなる傾向がある (実時間システムにおいては、数百 (ms) の処理時間でさえも許容されない)。

システム動作周波数約 30(frames/sec) 以上のパフォーマンスを維持した上で、さらに注視領域および注視点の数を増やす場合には、時間差分画像と背景差分画像の生成処理および特徴抽出処理の並列化あるいはハードウェア化が必要になると考えられる。パターン照合処理についても、認識する身振りシーケンスの長さにも依存するが、身振り数が数百を越える場合は、並列化あるいはハードウェア化が不可欠なものとなると考えられる。提案手法は、並列的に構成されているために処理の並列化には容易に対応できると考えられる。一方、システム動作周波数は使用するマシン能力に依存するために、将来的にはさらに改善されることが期待できる。

5.6 身振り情報の認識結果について

図 4.36 から図 4.41 に示すように、身振り「バイバイ」の反復動作を行った際に、その動作の位相・方向・速度・振幅などの各種身振り情報を実時間で認識できていることが分かる。本論文で提案した手法により得られる各種身振り情報は、最も基礎的な情報であるために、これらの情報を利用してさらに高次の認識処理を行うことができるものと考えられる。具体的には、身振り情報の時間方向の波形解析による身振り動作区間のセグメンテーション処理や記号化処理などであり、提案手法により得られた身振り情報により、タスク依存の認識処理が可能になることが期待される。

5.7 問題点の検討と今後の研究課題

評価実験により提案手法の有効性を確認することができたが、一方で、提案手法では対応できない多くの問題点がある。本節では以下に列挙する問題点を検討するとともに改善方法を提案する。

問題点 1 複数人物および単一ユーザによる並列身振り動作を認識できない（並列身振り動作とは、複数種類の身振りを同時に入力する動作を指す）。

問題点 2 身振り動作の区切りを画像差分量により検出しているため、動作途中での身振り動作の切り替わりを瞬時に検出することが出来ない。

問題点 3 CCD カメラ画像では、完全に衣服に依存しない身振り認識処理の実現は困難である。

問題点 4 状況依存の身振り認識に対応していない。

5.7.1 空間セグメンテーション処理の必要性

本論文で提案した多注視点身振り認識法は、複数種類の画像特徴量に基づき、入力された身振りのカテゴリ推定をボトムアップに行うものである。ボトムアップアプローチのみでは、カテゴリ名・動作方向・身振り速度・身振り振幅などの定量的な身振り情報を獲得することが可能であるが、状況依存の身振り認識や身振り画像領域のセグメンテーション処理など、トップダウンアプローチを必要とする処理には対応できない。

このために提案手法では、人物領域および動作領域を単一のセグメントとみなして認識処理を行っている。従って、複数ユーザおよび単一ユーザによる並列身振り動作を認識することは、領域分割処理を行っていない現時点では出来ない（動作に同期が取れていれば複数ユーザでも認識可能）。画像処理方式による身振りインタフェースの他方式に対する明確な優位性を主張するためには、人物領域および動作領域のセグメンテーション処理を行うことにより、複数ユーザおよび単一ユーザによる並列身振り動作の認識処理を可能とする必要がある。図 5.1 にボトムアップとトップダウンアプローチに基づく身振り認識処理の構図を示す。

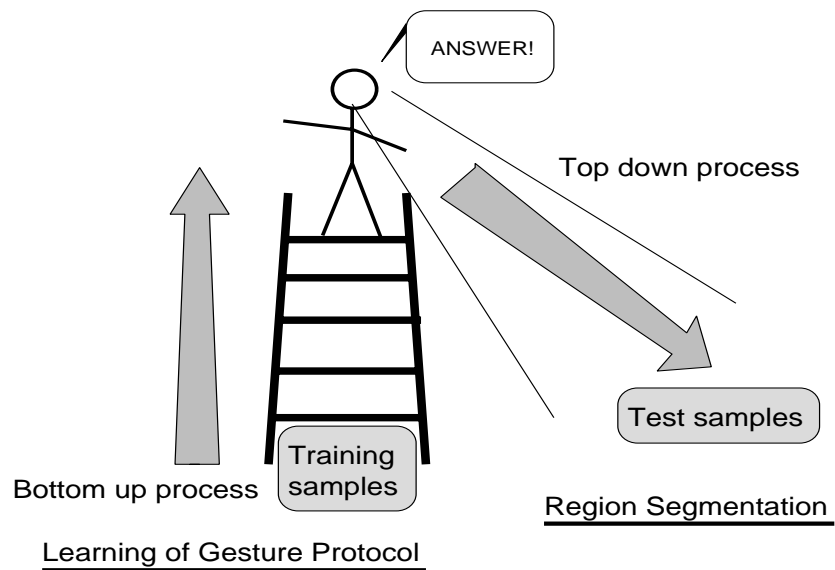


図 5.1 ボトムアップとトップダウンアプローチに基づく身振り認識

図 5.1 に示したように，学習時にはセグメンテーション処理は不要であるが，身振りプロトコル学習は必要である．一方，認識時にはセグメンテーション処理は必要であるが身振りプロトコル学習は不要である．具体的な画像のセグメンテーションアルゴリズムについては，身振り画像に特化してさらに実時間（数 msec 以内）での処理が可能なアルゴリズムの開発が必要になると考えられる．現在，複数ユーザおよび単一ユーザによる並列身振り動作の認識は，図 5.2 に示す方式で可能になると考えている．

5.7.2 時間セグメンテーション処理の必要性

提案手法では，動作の開始と終了が時間差分量により検出できる身振りを対象としている．これは，身振り動作が「構え（低速）」 「ストローク（高速）」 「戻し（高速）」の三要素から成り立つとの A.Kendon et al.[26], F.Quek et al. [27] の主張に基づくものである．このため，本手法では，動作開始からの評価値 $E_i^{(c)}$ を加算し始めて，現在までの評価値 $E^{(c)}$ が最大になる身振りカテゴリに，入力された身振りが属するものと判定している．しかしながら，この方式では，動作

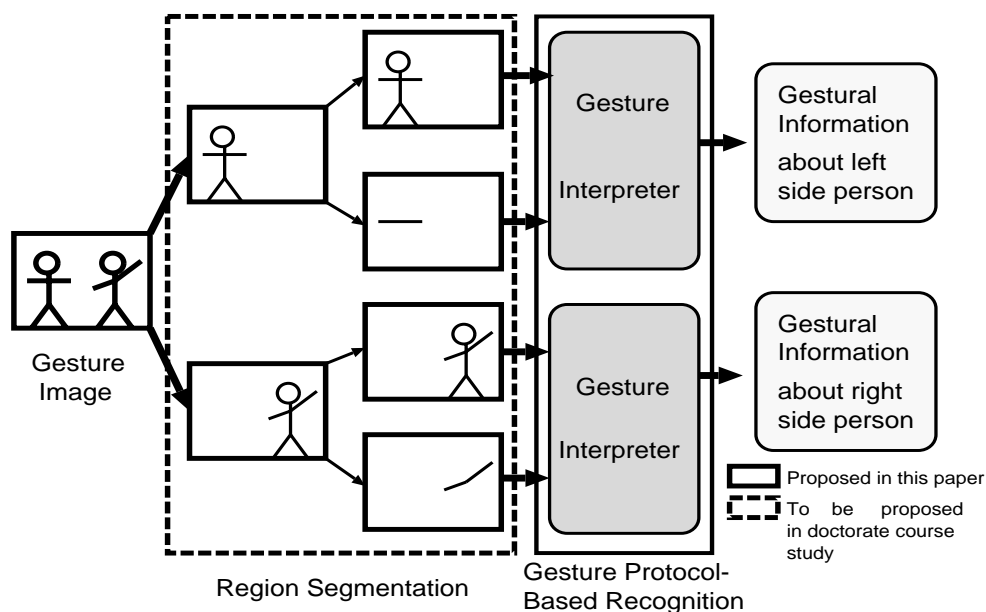


図 5.2 複数ユーザおよび単一ユーザによる並列身振り動作の認識

途中での身振りの種類の切り替わりを瞬時に検出することが出来ない．頻繁に身振りを入力する場合、「構え（低速）」 「ストローク（高速）」 「戻し（高速）」を繰り返していたのでは甚だ不効率である．従って，身振り入力の時間的な効率を高めるためには，動作途中でも身振りの種類の切り替わりを検出するための手法を考案する必要がある．

改善方法としては，累積評価値におけるカテゴリ順位と瞬時評価値におけるカテゴリ順位の相関に基づく時間セグメンテーション手法を考えている．これは，順位変化の起こりにくい累積評価値の性質と，順位変化が頻繁に発生する瞬時評価値の性質を利用した方式である．両評価値の順位相関係数を利用することで，身振り動作の切り替わりに敏感にする場合や，逆に鈍感にする場合にも対応できるものと考えている．

5.7.3 リアルタイムレンジファインダの必要性

評価実験の結果より、完全に衣服に依存しない身振り認識処理は、CCD カメラ画像を使用している限り極めて困難であることが分かった。従来の研究の中には、衣服に依存しない身振り認識処理を実現するために、極端に粗いリダクション画像（例えば 10x10）を使用する例 [13] もある。しかしながら、極端に低い解像度の画像を利用する場合、細かい身振り動作（手振りなど）を認識することが困難になる。また、解像度の低い画像を使用する場合、単位時間当たりで把握できる姿勢変化の状態数が減少する。これは、身振りインタフェースを VR に適用する際には大きな問題となる。なぜなら、ユーザが仮想環境あるいは仮想オブジェクトに対してとる身振り動作は、応答性を高めるために時間解像度を高める必要があるからである。VR への応用を視野に入れているために、提案手法はリダクション画像の使用は前提としていない。

衣服に依存する身振り認識が、完全に無用な訳ではない。衣服に依存するということは衣服の違いに基づくセキュリティシステム開発の可能性を暗に示している。重要なことは、衣服に依存する身振り認識処理と、衣服に依存しない身振り認識処理を如何にして同一の認識の枠組の中で実現することができるかということである。衣服依存 / 非依存の身振り認識処理を、画像の解像度のみの変更で対処することは本質的に困難であると考え、本研究では、衣服依存 / 非依存の身振り認識処理を実現するために、画像センサ系の並列化を図りたいと考えている。例えば、CCD カメラとレンジファインダ（距離画像カメラ）が同時に利用できる場合、提案手法を図 5.3 に示すように拡張することにより、同一解像度の下で容易に衣服依存 / 非依存の身振り認識が可能になると考えている。図 5.3 の枠組により、衣服依存の場合と衣服非依存の場合を完全に分離することなく、認識対象に応じて衣服依存識別器と衣服非依存識別器への重み付け処理の自動化が可能になると考えられる。

一方、実世界における身振りを認識する際に無視できない問題が「図と地」の識別である。レンジファインダを利用することが可能であれば、人物と背景の区別を容易に行うことが可能となり背景画像が不要になる。また、影の影響を受けない性質も、衣服に依存しない身振り認識処理を実現する際には、極めて有効であると考えられる。この結果、リアルタイムレンジファインダの使用により、照

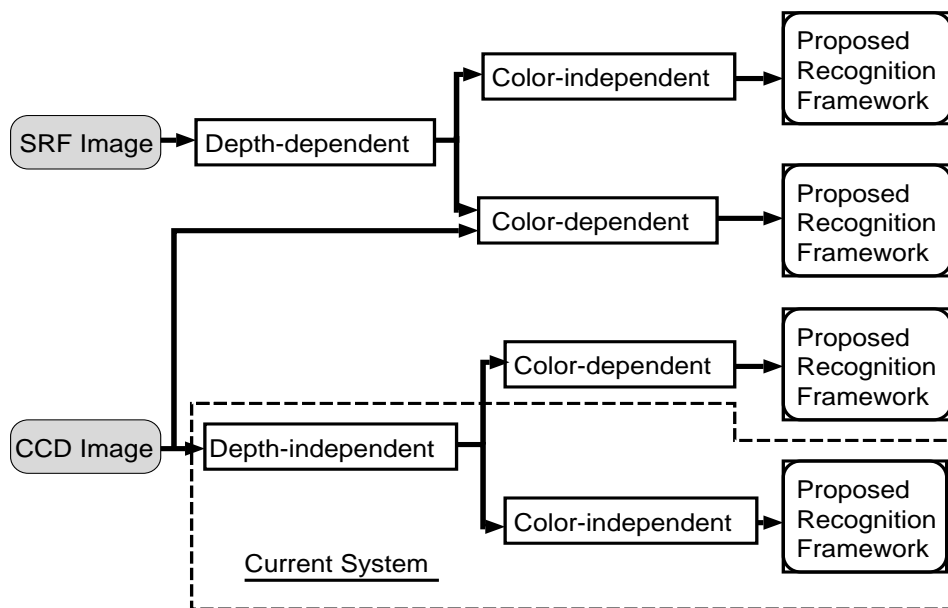


図 5.3 CCD カメラと SRF による衣服依存 / 非依存の身振り認識の枠組

明などの環境変動要因に対してより一層ロバストな身振り認識処理が可能となる
ことが期待できる。

5.7.4 階層化身振り認識機構の必要性

提案手法は、画像特徴量に基づいて、入力された身振りのカテゴリ推定をボトムアップに行うものであり、状況依存の身振り認識問題に対処することは出来ない。これは、インタフェースとしての側面を重視すれば、状況依存の身振り認識は特別に要求されるものではないという信念に基づくものである。しかしながら、より広範な身振り認識問題、あるいはアプリケーションの開発に対処するには、トップダウンアプローチに基づく認識機構が不可欠であり、本研究では今後、図 5.4 に示す階層化身振り認識法を提案して行きたいと考えている。図 5.4 に示すように、階層化身振り認識機構においては、本論文で提案した多注視点身振り認識法は最下層である第一層の役割を担うものと位置付けることができる。第一層では、主に状況非依存の身振り（Context-Free Gesture）の認識を対象としてお

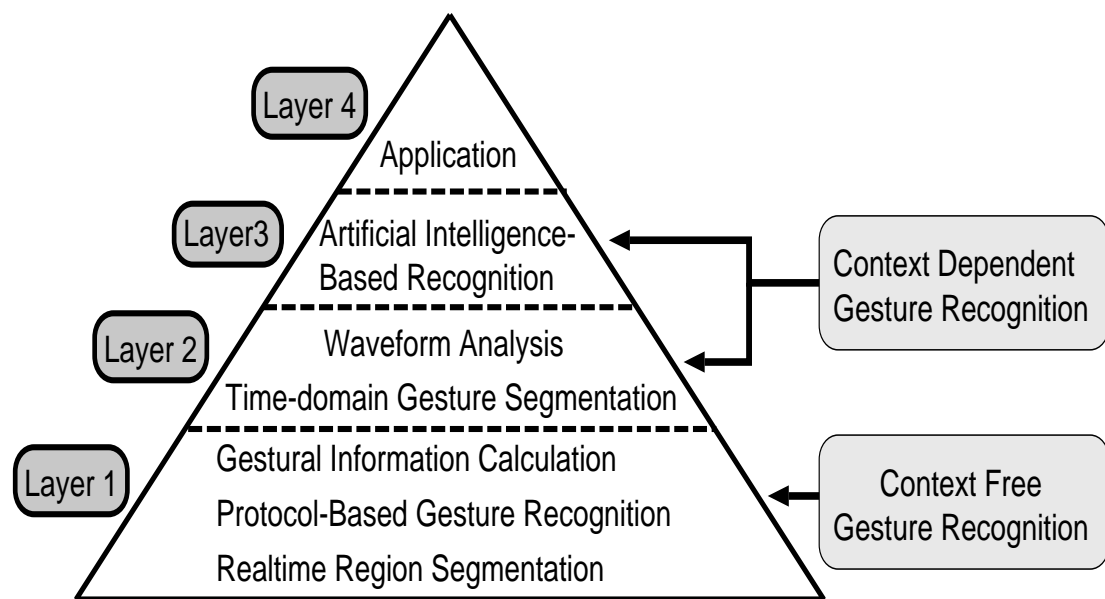


図 5.4 階層化身振り認識

り，種々の身振り情報を獲得する役割を果たす．獲得される身振り情報は，ユーザの位置・姿勢・動作に関するものであり，身振り標準パターンにおけるフレーム番号に相当する身振り位相値，さらに身振り標準パターンと比べた場合の身振り動作の相対的な速度（身振り速度）や相対的な振幅（身振り振幅）といった情報である．第一層における研究トピックスは，身振りプロトコル学習と身振りプロトコルに基づく認識処理，さらに身振り画像の空間セグメンテーション処理などが挙げられる．

第二層の役割は，身振り情報の時間方向の波形解析による身振り動作区間のセグメンテーション処理や記号化処理などであり，研究トピックスとしては，デジタル信号処理や DP マッチング処理などが挙げられる．

第三層の役割は，状況依存の身振り（Context-Dependent Gesture）を対象とした状況依存のトップダウン身振り認識であり，研究トピックスとしては，オートマトンや HMM（Hidden Markov Model）による構文解析処理が挙げられる．また，第三層において人工知能的アプローチに基づき第一層において利用される

身振りデータに関する知識ベースを構築することにより，上位層から下位層へのフィードバック処理を行い，Interactive Cinema, Video などの状況依存のトップダウン身振り認識を必要とするアプリケーションにも対応できると考えられる．

最上層である第四層のアプリケーション層における研究トピックスとしては，VR・HCI などの人間-機械系のコミュニケーションを支援するようなアプリケーションの開発が挙げられる．

ここで注意すべき点としては，各層の認識結果はそれぞれ独立して利用することが可能であるということである．すなわち，第一層の役割を果たす多注視点身振り認識法により得られる認識結果は，既に VR・HCI などへの応用が可能である．このように，身振り認識の階層化を図る理由は，状況依存 / 非依存の認識処理をより目的に適合したレベルで合理的に行うことにある．図 5.4 で示した身振り認識機構は，並列処理と逐次処理，さらにボトムアップとトップダウン処理を結合させる問題を扱うものであり，センサヒュージョン問題に対処するための基本的構造を備えていると考えられ，図 5.4 で示した階層化身振り認識機構によってより広範な身振り認識問題を扱うことが出来るようになることが期待される．

第 6 章 むすび

本論文では、身振りインタフェースを開発する際には、基本問題としての身振りプロトコルが存在し、ユーザ定義の身振りの認識処理および派生身振りの認識処理を実現するには、身振りプロトコルに適応できる認識系を設計することが不可欠であることを示し、身振りプロトコルに適応するための基本的枠組として、多注視点身振り認識法を新たに提案した。

多注視点身振り認識法は、特徴量ベース識別処理と身振りプロトコルベース認識処理の二段階により構成される階層型認識機構である。多注視点身振り認識法により、入力された身振りのカテゴリ推定のみでなく、身振り情報の算出が可能である。算出される身振り情報は、ユーザの位置・姿勢・動作に関するものであり、身振り標準パターンにおけるフレーム番号に相当する身振り位相値、さらに身振り標準パターンと比べた場合の身振り動作の相対的な速度（身振り速度）や相対的な振幅（身振り振幅）といった情報である。

多注視点身振り認識法では注視点の数に比例して計算コストが増加するために、処理時間の大部分を占めている形状特徴抽出処理とパターン照合処理の高速化を図る必要がある。このために本研究では形状特徴抽出法を新たに考案した。新たに考案した形状特徴抽出法では、設定された注視領域における形状パターンの大きさを考慮すべき場合と考慮すべきでない場合と形状パターンの回転を考慮すべき場合と考慮すべきでない場合に対応するための特徴量を抽出するものである。

評価実験の結果、多注視点身振り認識法により以下の事項が可能となることが分かった。

- 提案手法を実際にWSに実装することにより、30(Hz)以上のシステム動作周波数が得られ実時間での身振り認識処理が可能になるとともに、実時間

ベースで身振り情報を獲得することが可能となった．この結果は，提案した身振りインタフェースが VR アプリケーションに十分適用可能であることを示している．

- 単一の注視点のみでは認識できない身振りが，複数種類の注視点を利用することにより認識できるようになることが分かった．この結果は，多注視点身振り認識法により柔軟かつ安定した身振り認識処理が可能となることを示すものである．
- 身振りプロトコルを学習させる際には数回の身振り動作で学習結果が安定することが分かった．この結果は提案手法においては，少ない訓練サンプル数でも高い学習効率を得られることを示している．
- 身振り認識の難しさは認識対象に強く依存しているにもかかわらず，身振りプロトコル学習により，従来は困難であった認識対象である派生身振りを認識することが可能となった．
- 背景や照明に関して，実験遂行上問題が生じない程度にロバストな身振り認識が可能となった．
- 統合された枠組みの中でユーザの位置・姿勢・動作に関する情報を認識することが可能となった．

以上に加えて提案手法では，限定的ではあるが身振り解釈の多様性に対応することが可能である．表 6.1 に対応可能な身振り解釈の多様性と考案した対処方法を示す．

現在，身振りインタフェースへの入力として CCD カメラにより得られた濃淡画像を使用しているが，本論文で提案した身振り認識の枠組は，あらゆる身振り情報入力用センサ（CCD カメラ，レンジファインダ，データグローブ，ポヒマスセンサなど）から得られるデータに適用できる．また，多注視点身振り認識法は大規模な並列化が可能であり，複数種類のセンサ情報を選択的に利用する身振り認識システムの構築が将来的には可能であると考えられる．

本研究で提案した非接触型身振りインタフェースは，インタフェースの透明化を実現することが可能であり，このため，ユーザは興味の対象である仮想世界あ

表 6.1 対応可能な身振り解釈の多様性と対処方法

多様性生起要素	対処方法
形状	形状に関する閾値を（自動）調整する．
位置関係	必要な注視領域を任意に設定する．
大きさ	形状特徴パターンを正規化処理する／しない．
回転	回転不変特徴量を利用する／しない．
移動速度	時間差分画像と背景差分画像を利用する．

るいは仮想物体に直接働きかけ，それを直接眺めているという感じをもつことができる．身振りは，多くの人が既にマスターしたコミュニケーションメディアであり，身振りインタフェースによるダイレクトマニピュレーションによりユーザの能力を最大限に発揮させることが将来的には可能になると考えられる．

本研究の成果は，VR 以外にも以下に示す分野に応用できると考えられる．

- Interactive Art, Video, Cinema
- マルチメディア動画像検索（非言語論理に基づく画像検索）
- 教育（CAI）
- 医療・福祉（診断画像表示 I/F・リハビリテーション・介護ロボット）
- 感性コミュニケーション（リラクゼーション）
- CAD,CAM
- プラント監視，プラント制御
- 発想・想像・発見支援
- 個人識別

以上に示したように，今後，身振りインタフェース研究の成果が応用される分野は極めて広範囲であり，マンマシンコミュニケーションを中心とした人間—機械系に新たな Breakthrough をもたらすものと期待している．

謝辞

本研究を行うにあたり，適切な御指導と多大な御助言を頂きました千原國宏教授に深く感謝致します．

ゼミ発表において，御助言を頂きました鹿野清宏教授に感謝の意を表します．

学会論文発表や研究室進捗報告会などにおいて，研究の方針や手法などに常に適切なアドバイスを頂きました佐藤宏介助教授に深く感謝致します．

また本研究の実験を行うにあたり，様々な場面でお世話頂きました大城理助教授，眞溪歩助手，陳謙助手に感謝致します．

最後に，日頃から御協力や御支援を頂きました宇井信子秘書，並びに像情報処理学講座の諸氏に心から感謝申し上げます．

参考文献

- [1] Toshiyuki Kirishima, Kosuke Sato, Kunihiro Chihara: “A novel approach on gesture recognition: the gesture protocol- based gesture interface” Proceedings of International Conference on Virtual Systems and Multimedia(VSMM'96),pp.433-438, September 18-20,1996
- [2] Toshiyuki Kirishima, Kosuke Sato, Kunihiro Chihara: “Realtime Gesture Recognition by Learning of Gesture Protocol” Proceedings of the Meeting on Image Recognition and Understanding (MIRU'96),vol.II, pp.163-168, 17-19 July, 1996.
- [3] Toshiyuki Kirishima, Kosuke Sato, Kunihiro Chihara: “Gesture Recognition from Sequence of Difference Images” IEICE technical report, PRU 95-190, pp.1-6, January, 1996.
- [4] 大須賀節雄: “ヒューマンインタフェース” 知識工学講座 10, オーム社, pp.218-229, 1992
- [5] 飯島泰蔵: “パターン認識” 電気・電子工学大系 43, コロナ社, pp.1-25, 1973
- [6] Trevor Darrell and Alex P.Pentland: “Attention-driven Expression and Gesture Analysis in an Interactive Environment” International Workshop on Automatic Face- and Gesture-Recognition, Zurich, pp.135-140, 1995
- [7] 國吉康夫: “実世界エージェントにおける注意と視点-情報の分節・統合・共有-” 人工知能学会誌 Vol.10 No.4, pp.507-514, July, 1995

- [8] Paul Ekman: “Essential Behavioral Science of the Face and Gesture that Computer Scientists Need to Know” International Workshop on Automatic Face- and Gesture-Recognition, Zurich, pp.7-11, 1995
- [9] 高橋勝彦, 関進, 小島浩, 岡隆一: “ジェスチャー動画像のスポッティング認識” 電子情報通信学会論文誌, vol.J77-D-II, No.8, pp.1552-1561, August, 1994
- [10] 牛田博英, 山口亨, 高木友博: “ファジー連想記憶システムを用いた動作認識” 電子情報通信学会論文誌, vol.J77-D-II, No.8, pp.1571-1581, August, 1994
- [11] 大和, 大谷, 石井: “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, vol.J76-D-II, No.12, pp.2556-2563, December, 1993
- [12] 山田尚勇: “アメリカにおけるキーボード障害裁判の技術的評価” bit, Vol.28, No.9, pp.10-17, September, 1996
- [13] 西村拓一, 向井理朗, 岡隆一: “複数人物によるジェスチャーの単一画像からのスポッティング認識” 電子情報通信学会技術研究報告, PRMU96-90, pp.77-84, Nov., 1996
- [14] 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研究報告, PRMU96-110, pp.49-56, Dec., 1996
- [15] 黒川隆夫: “ノンバーバルインタフェース” オーム社, 1994
- [16] Thomas S.Huang and Vladimir I.Pavlovic: “Gesture Modeling, Analysis, and Synthesis” International Workshop on Automatic Face- and Gesture-Recognition, Zurich, pp.73-79, 1995
- [17] C.Dennis Allen, Don Ballman, Vivienne Begg, Harold H.Miller Jacobs, Michael Muller, Jakob Nielsen, Jared Spool: “User Involvement In The Design Process: Why, When and How?” Proceedings of the Human Factors

in Computing Systems, INTERCHI'93, pp.251-254, Amsterdam, 24-29 April, 1993.

- [18] I.Scott MacKenzie, Colin Ware: "Lag as a Determinant of Human Performance in Interactive Systems" Proceedings of the Human Factors in Computing Systems, INTERCHI'93, pp.488-493, Amsterdam, 24-29 April, 1993.
- [19] Colin Ware, Kevin Arthur, Kellogg S.Booth: "FISH TANK VIRTUAL REALITY" Proceedings of the Human Factors in Computing Systems, INTERCHI'93, pp.37-42, Amsterdam, 24-29 April, 1993.
- [20] John J.Weng, N.Ahuja and T.S. Huang: "Learning Recognition and Segmentation of 3-D Objects from 2-D Images" Proceedings of the Fourth International Conference on Computer Vision, Berlin, Germany, pp.121-128, 11-14 May, 1993.
- [21] E.Hunter, J.Schlenzig and R.Jain: "Posture Estimation in Reduced-Model Gesture Input Systems" Proceedings of the International Workshop on Automatic Face- and Gesture-Recognition, pp.290-295, Zurich, 1995.
- [22] William T.Freeman and Michael Roth: "Orientation Histograms for Hand Gesture Recognition" Proceedings of the International Workshop on Automatic Face- and Gesture-Recognition, pp.296-301, Zurich, 1995.
- [23] Christoph Maggioni: "Gesture Computer—New Ways of Operating a Computer" Proceedings of the International Workshop on Automatic Face- and Gesture-Recognition, pp.166-171, Zurich, 1995.
- [24] Andrew D.Wilson, Aaron F.Bobick: "Configuration States for the Representation and Recognition of Gesture" Proceedings of the International Workshop on Automatic Face- and Gesture-Recognition, pp.129-134, Zurich, 1995.

- [25] Yuntao Cui and John Weng: “Learning-Based Hand Sign Recognition” Proceedings of the International Workshop on Automatic Face- and Gesture-Recognition, pp.201-206, Zurich, 1995.
- [26] A.Kendon: “Current Issues in the study of gesture” Biological Foundations of Gestures: Motor and Semiotic Aspects, pp.23-47, Lawrence Erlbaum Association, 1996.
- [27] F.Quek: “Toward a vision-based hand gesture interface” Virtual Reality Software and Technology Conference, Aug., 1994.
- [28] Robert Aston, Joyce Schwarz: “Multimedia-Gateway to the Next Millennium-” AP professional, pp.214-217, 1994
- [29] Trevor J. Darrell, Irfan A.Essa, Alex P. Pentland: “Task-Specific Gesture Analysis in Real-time Using Interpolated Views” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 18, No. 12, Dec., 1996.

付録

A. 形状特徴抽出法について

本付録では，本研究で新たに提案した式 A.1 で表される形状特徴抽出オペレータの妥当性と，勾配係数 a の決定方法について検討する．

$$q = \frac{R}{L} \frac{\sum_r p(r) \exp\{-a(r - \phi)^2\}}{\sum_r p(r)} \quad (\text{A.1})$$

A.1 ガウス分布との畳み込み演算を行う理由

式 A.1 の目的は，線形パターン $p(r)$ からユニークな特徴量 q を算出することである．従って，畳み込み演算の際に使用するオペレータは，任意の位置 r においてユニークである必要がある．このような分布としては，

$$f(r) = ar \quad (\text{A.2})$$

$$f(r) = \exp(-ar) \quad (\text{A.3})$$

$$f(r) = \exp(-ar^2) \quad (\text{A.4})$$

などが考えられる．ここで，使用するオペレータが対称性を持つ場合，左右対称なパターン分布からユニークな特徴量を抽出することが出来ないことに注意する必要がある．各オペレータの対称性は式 A.5 により判定できる．

$$\sum_{r=0}^R \frac{d}{dr} \{f(r) + f(R-r)\} \begin{cases} = 0 & (\text{symmetry}) \\ \neq 0 & (\text{asymmetry}) \end{cases} \quad (\text{A.5})$$

式 A.2 のオペレータに式 A.5 を適用すると式 A.6 の結果が得られる．

$$\sum_{r=0}^R \frac{d}{dr} \{ar + a(R-r)\} = \sum_{r=0}^R \frac{d}{dr} aR = 0 \quad (\text{A.6})$$

式 A.5 に示した判定基準より，式 A.2 のオペレータでは左右対称なパターン分布からユニークな特徴量を抽出できないことが分かった．一方，式 A.3 のオペレータについては式 A.7 の結果が得られ，式 A.4 のオペレータについては式 A.8 の結果が得られる．

$$\sum_{r=0}^R \frac{d}{dr} \{\exp(-ar) + \exp(-a(R-r))\} = \sum_{r=0}^R a \exp(-ar)(\exp(-aR) - 1) \neq 0 \quad (\text{A.7})$$

$$\sum_{r=0}^R \frac{d}{dr} \{\exp(-ar^2) + \exp(-a(R-r)^2)\} = \sum_{r=0}^R 2a \exp(-ar^2) \{\exp(aR(2r-R))(R-r) - r\} \neq 0 \quad (\text{A.8})$$

以上より，式 A.2 は分布に対称性を有しているために形状特徴抽出オペレータとしては不適切であるが，式 A.7 と式 A.8 の計算結果より，式 A.3 と式 A.4 は，オペレータとして使用できることが分かった．本研究では，低域通過フィルタ特性を持たせることが可能なガウシアンオペレータを使用することにする．

A.2 密度係数を掛け合わせる理由

式 A.1 において式 A.9 により表される部分を密度係数と呼んでいる．

$$d(\theta) = \frac{R}{\sum_{r=0}^R p(r)} \quad (\text{A.9})$$

密度係数を掛け合わせる理由としては，パターン分布区間 $[0, R]$ において，畳み込み演算の際の有効（オンドット）加算回数を R に統一することが挙げられる．加算回数を統一することにより，パターン分布における有効要素数（オンドット数）に依存しない特徴量を抽出することが可能となる．

A.3 ガウス分布の勾配係数 a の決定方法

A.1節において，ガウシアンオペレータにより対称性を持つようなパターン分布でもユニークな特徴量を抽出することが可能であることが分かった．しかしながら，得られる特徴量のユニークさの程度については，勾配係数 a に依存するために，目的に応じて勾配係数を決定する必要がある．図 A.1 に勾配係数 a を 1～9 まで変化させた際に得られるガウシアンオペレータとその一次導関数を示す．なお，オペレータとして実際に用いるガウス分布は，対象図形の外側輪郭の特徴に関して一層敏感となるように分布の左右を反転させて用いることにする．式 A.10 により表される対称性評価関数において r を 0～0.5 まで変化させた場合のグラフを図 A.2 に示す．

$$\frac{d}{dr} \left\{ \exp(-ar^2) + \exp(-a(R-r)^2) \right\} = 2a \exp(-ar^2) \left\{ \exp(aR(2r-R))(R-r) - r \right\} \quad (\text{A.10})$$

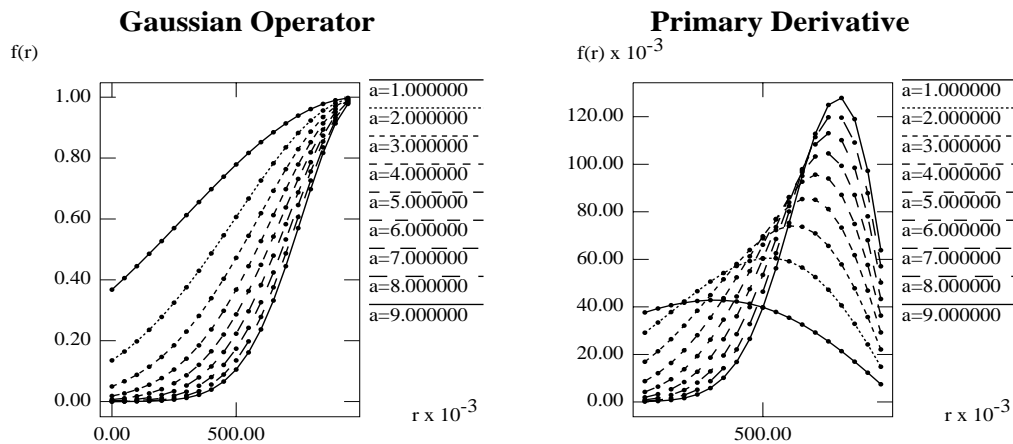


図 A.1 ガウシアンオペレータ（左）とその一次導関数（右）

図 A.2 に示したように，形状特徴量の外側輪郭依存性と uniqueness は勾配係数 a に依存しそれぞれがトレードオフの関係にあることが分かる．以上のことを考慮し，本論文の評価実験においては，勾配係数 a を 5.0 に設定して実験を行った．

Symmetry Evaluation

$f(r)$

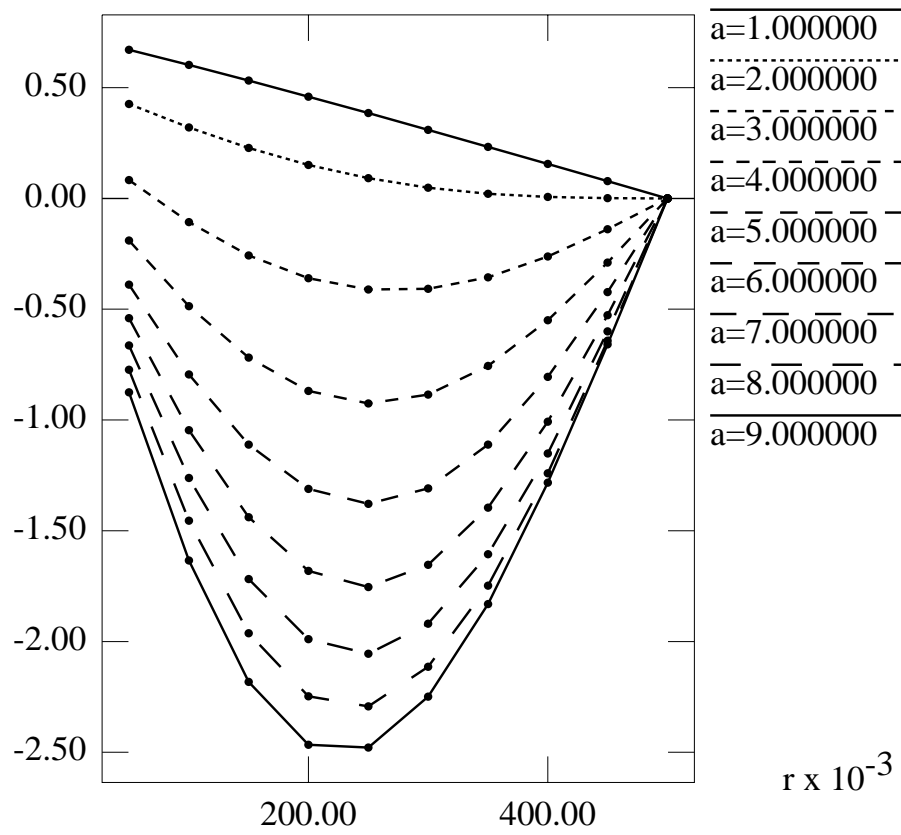


図 A.2 オペレータの対称性評価関数