エネルギー科学研究科

エネルギー社会・環境科学専攻修士論文

題目: **動画像処理とリアルタイムクラスタ分析による身振りの自動分類手法の研究**

指導教官： 吉川 榮和　教授

氏名：　笹井　　寿郎

提出年月日： 平成13年2月7日(水)

論文要旨

題目：動画像処理とリアルタイムクラスタ分析による身振りの自動分類手法の研究

吉川榮和研究室　笹井　寿郎

要旨：

近年の情報技術の発展に伴い、パソコンや携帯電話等の情報端末が次々と開発され、我々の身のまわりに浸透してきている。しかし、現状では情報機器の高機能化・多機能化に重点が置かれ、使う側の操作性の向上はあまり重視されていない。そのため、複雑な情報機器を使いこなせる人と、うまく使いこなせない人との格差、いわゆる「ディジタルデバイド」が広がってきている。使う人間の視点で情報機器を使いやすくし、ディジタルデバイドを解決することもヒューマンインタフェースの研究で重要な課題である。最近のパソコンでは、ディスプレイ、キーボード、マウスなどの入出力機器が人間とコンピュータとの接点となっている。これらのコンピュータ用入出力機器を介したコンピュータと人間とのコミュニケーションは、言語、表情、身振りを用いる人と人とのコミュニケーションとはまったく異なる。そこで人間が本来持っている、このようなコミュニケーション能力を利用すれば、情報機器を使いやすくする新しい形態のヒューマンインタフェースの実現が期待できる。そこで本研究では、人間の意図や感情を表現するノンバーバルメッセージの１つである身振りに着目し研究を進めた。具体的には、機器と対話する個人の身振りを認識して、その動作のタイミングや種類から人間の意図や感情状態を推定し、それに即した応答を返すことによって、コンピュータと人間のスムーズなコミュニケーションを志向する個人適応型インタフェースの実現を目標にした。

本研究では、個人適応型インタフェースの基礎研究として、個人差の大きい人の身振りのリアルタイム認識手法を確立するため、ビデオカメラで撮影した人間の上半身動画像からリアルタイムで身振りを識別し、その動作の特徴から身振りを分類する手法を提案する。従来の身振り認識に関する研究では、主に手話のような規則性ある身振りを認識するものが多く、人と人とのコミュニケーションで自然に生起する身振りを扱うものはほとんどなかった。そこで、人間同士のコミュニケーションで身振り生起のタイミングが重要な意味をもつことに着目して、人間が自然に生起する暗示的情報も包含した身振りの分類手法を導出した。

提案した手法では、上半身動画像から身振りの区間を検出し、その区間にある身振りの特徴をクラスタ分析により分類する。また、過去の身振りの分類結果をもとに、個人の動作の特徴や癖にあわせて分類を再分析し、より個人に応じた身振りの分類へと適応するための分類の再構成を行う。そして提案した手法に基づくリアルタイム身振り分類システム ReD BACS (Real-time and Dynamic Body Action Classi¯cation System) を試作し、試作システムを用いた被験者実験を行って、実験データを分析し、提案した手法を実現した ReD BACS が身振りの自動分類をリアルタイムで実行できること、また、本手法による身振りの自動分類が人間の観察する身振りの検出と分類より、暗示的な身振りを含む身振りを詳細に検出し、分類できることを確認した。

目 次

図 目 次

表 目 次

第 1 章 序論

近年の情報通信技術の発展に伴い、情報通信ネットワークが社会基盤として整備されつつある現在、パソコンや携帯電話のような情報機器が次々と開発され、我々の身のまわりに浸透してきている。現在では、インターネットを用いるとニュース、天気予報、入試情報、就職情報、株価情報などの様々な情報が旧来のメディア、すなわち、新聞、ラジオ、TV、雑誌より簡単に速く手に入れることができるようになってきた。これからの高度情報社会では、情報を速く入手し、うまく活用できるか否かが人生で有利に立つかどうかを左右するようになると考えられる。このため、情報機器を使いこなせることが重要であるが、現状では情報機器は複雑であり、実際に使いこなしているのは一部の人だけである。「機器操作ができない」、「不安に思う」、「そのような機械に反感をもつ」という人も決して少なくない。このような情報機器を利用できるものとできないものの分化を「ディジタルデバイド」というが、情報通信技術が発達した時代にあり、情報化が進む時代の社会問題の 1 つとなっている。これからの情報社会では「誰にでも情報機器が使いこなせる」ことが 1 つの必須条件であろう。

それでは、どのようにすれば人々に使いやすい情報機器ができるのであろうか。人間同士のコミュニケーションにそのヒントが隠されている。人間は、意図的、非意図的にかかわらず、お互いの間で様々な情報を発信し、そして受信している。そこでは、言葉はもちろん重要であるが、ノンバーバル情報と呼ばれる、言葉以外のメッセージ、すなわち、表情、身振り、視線、声の抑揚から、服装や化粧に至るまでの多種多様なメッセージを用いてコミュニケーションが行われる。

このようなノンバーバル情報をコンピュータが人間との間で交わすようになれば、コンピュータや情報機器の人間一般に対する親和性が向上し、ひいてはコンピュータを使いこなせない人間もコンピュータに興味をもって接するようになって、自然に操作性も向上するであろう。

本研究では、以上のような観点で、人間とコンピュータとのノンバーバルコミュニケーションに関する基礎研究を行う。具体的には、ノンバーバル情報の中でも重要な身振りに着目し、個々の人の身振りを認識して、それをヒューマンマシンインタフェースに利用すれば、人間に適応する新しいインタフェース（個人適応型インタフェース）が可能ではないかという観点から研究を行う。

本研究では、身振りをコンピュータに認識させるための基礎的な手法として、ビデオカメラなどにより撮影された上半身の動画像を画像処理して、身振りを分類する手法を提案する[1]。またこの方法に基づいてリアルタイムで身振りを分類するシステムを試作をする。

以下に、本論文の構成を述べる。まず、第2章では、研究の背景として適応型インタフェースやノンバーバル情報について述べ、本研究に関連する動作計測、身振り認識について従来研究をまとめ、それらを背景に本研究の目的を明らかにする。次に、第3章では、人間の動作の特徴を検討し、本研究で提案する身振りの自動分類手法について述べる。そして、第4章では、第3章で提案した身振りの自動分類手法に基づき試作したシステムの詳細を述べる。さらに、第5章では、試作したシステムの評価実験とその結果の考察を述べる。最後に、第6章で本論文の結論と今後の研究課題を展望する。

第 2 章 研究の背景と目的

2.1　研究の背景

2.1.1　適応型インタフェース

高度情報化社会への発展につれ、インターネットや携帯電話という情報機器の多機能化が進み、若者を中心とした情報機器を十分使いこなす人と、複雑な機器をうまく使いこなせない人との格差、いわゆる「ディジタルデバイド」が広がっている。ディジタルデバイドを解消し、すべての人が情報社会のメリットを享受できるためには、情報機器の人間との接点であるヒューマンマシンインタフェースをより人間側の視点に立って、使いやすいものにする必要がある。そこで、この解決方法として、図 2.1 に示すようにコンピュータの側からも人間の心理や生理に合わせるための適応性をもたせる適応型のインタフェースを構築することが考えられている [2]。このような人間の心理や生理に適応するインタフェースを実現するためには、人間に負担をかけず、また意識をさせないで、人間の自然な振る舞いから人間の心理生理状態をモニタする人間情報行動計測手法の開拓が必要である。本研究は、このような新しい人間情報行動計測手法の開発の基礎研究の一環と位置づけられる。

普段の人間同士のコミュニケーションでは、明示的（エクスプリシット）な情報（キー入力、ことば、グラフ、図形、動画など）操作だけでなく、身振り、間合い、視線などの暗示的（インプリシット）な情報を発信している。これまでのインタフェースは、エクスプリシットな情報を入出力として利用しているが、インプリシットな情報も人間の心理状態や生理状態、さらには個性などを表すものとして利用すれば、コンピュータが人間とより自然に接することが可能になるだろう。そこで、次に人間同士のコミュニケーションのあり方を展望する。

2.1.2　人間同士のコミュニケーション

2 つの要素間のコミュニケーションの基本構造は、人間対人間、機械対機械、人間対機械のいずれであれ、図 2.2 のようにまとめられる [3]。何らかの目的や、相手に対する知識や推論を元に、伝えたい情報が生成され、その情報をメッセージの形に符号化し

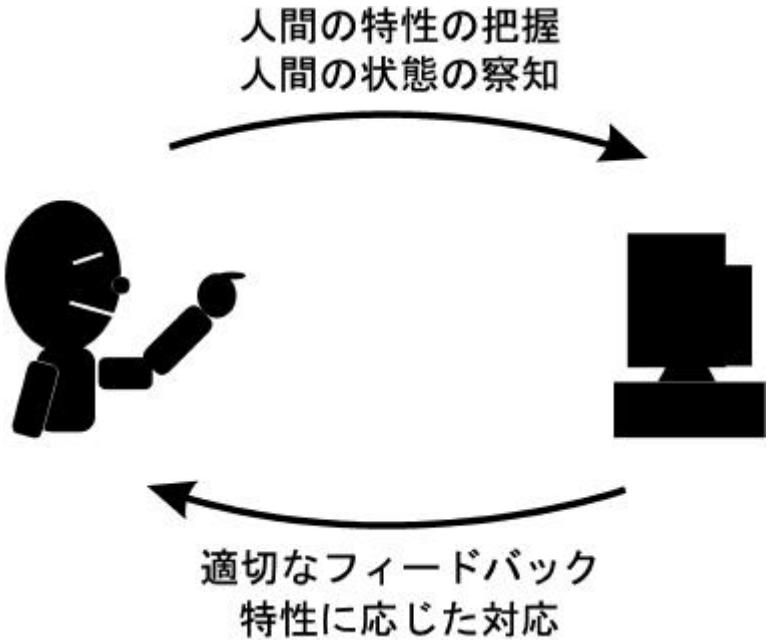

図 2.1: 適応型インタフェース

て、相互にコミュニケーションしている。この際、個々の要素でのメッセージへの符号化のルールは、個々の要素によって異なりうるため、符号化された情報がメッセージとして相手に伝達されて復号化される場合には、かならずしも相手の意図どおりに復号化されるとは限らない。

この符号化－復号化のルールが両者で大きく異なる場合にはコミュニケーションがうまく行えない。これは、2 つの要素が人間同士の場合、あるいは人間と機械の場合の双方の場合に、特に顕著な問題となる。よって、機械に、人間との間で円滑なコミュニケーションを行う能力を持たせるためには、人間特有の符号化－復号化のルールを詳しく調べて、それを機械の仕組みに反映させる新しい取り組みが必要である。

人間と人間とが向かい合ってコミュニケーションを行う場合、「ことば」で表すことのできるバーバル・メッセージ以外にも様々な情報が複数の媒体を通じて伝達されている [4]。これらはノンバーバル・メッセージと呼ばれ、場合によっては「ことば」以上に意味のある意思伝達手段とされる。表 2.1 に代表的なノンバーバル・メッセージを示す。このノンバーバル・メッセージは、手話のようにエクスプリシットな情報を伝達する手段となる場合もあるが、それ以上に、インプリシットな情報を伝達する能力があり、人間同士の日常のコミュニケーションで重要な役割を果たしている。ノンバーバル・メッセージが人間のコミュニケーションで占める割合は、Bindwhistell によると

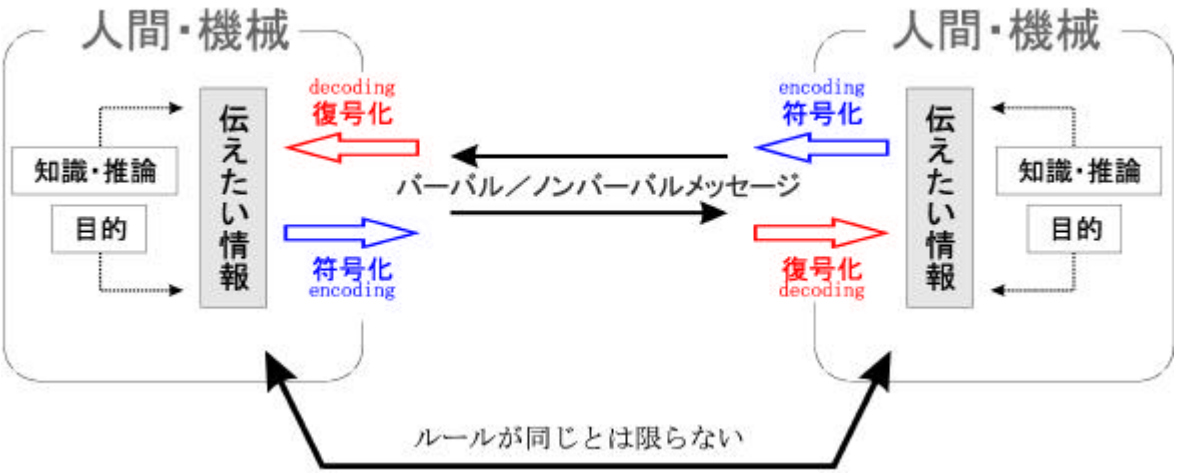

図 2.2: コミュニケーションの基本構造

65～70%、Mehrabian では、93%としている。

ノンバーバル・メッセージは文数、語数のような数量化できる測度があるバーバル・メッセージとは異なり、受け手によりその解釈が異なるものでもあり、これらの数字は主観的なものではあるが、いずれにしてもノンバーバル・メッセージがコミュニケーションにおいて重要であることを示している。

また、情報の次元と伝達モードに関して、バーバルメッセージとノンバーバルメッセージを比較したものを表 2.2 に示す。バーバル・メッセージは単独の言語メッセージを交互に交換することを前提にしているのに対し、ノンバーバル・メッセージは、同時に並列して複数の相互情報交換が可能であり、対話などのコミュニケーションを円滑にするための重要な役割を果たしている [5]。

人間は生来、このようなノンバーバル・メッセージをお互いに送受信することができるため、円滑なコミュニケーションを行うことが可能なのである。このため、人間と機械との間でもノンバーバル・メッセージを伝達することができれば、より自然なインタフェースが実現できると考えられる。

このような特徴をもつノンバーバル・メッセージには、次のような 5 つの機能があるとされている [6]。

1. 明示的情報の伝達

   バーバル・メッセージが利用できない状況やバーバル・メッセージでは表現不可能か、十分正確に表現できない明示的情報を伝達する。

表 2.1: ノンバーバル・メッセージ

ノンバーバル・メッセージ		属性など
身体動作	身振り	
	姿勢（構え）	
	表情	
	視線	視線交差、凝視、無視
	瞳孔	散大、縮小
	口唇の動き	口話を含む
周辺言語	声質	声域、アクセント、発音、ピッチ
	発声法	
	特徴性	笑い、泣き、うめき、叫び、つぶやき
	限定性	強勢、大小、長短
	遊離性	つなぎ、間、沈黙
対人距離		個人空間、距離、位置

表 2.2: バーバル・メッセージとノンバーバル・メッセージの比較

	バーバル	ノンバーバル
情報次元	１次元（直列）	空間次元（並列）
情報伝達	半二重 （一時に片方向）	全二重 （同時に両方向）

2. 発信者の内面状態の伝達

   バーバル・メッセージを生成する心的過程で生じるイメージを伝達する。

3. 発信者のバックグラウンド情報の伝達

   発信者の民族、文化的背景、出身地、家柄、所属社会・所属集団、社会的地位、学歴、職業などを伝達する。

4. 発信者の性格に関する情報の伝達

   発信者の気質や性格を伝達する。

5. メタコミュニケーション機能

   コミュニケーションの内容の規定や、コミュニケーションの進行を調整する。

ヒューマンマシンインタフェース、ノンバーバルコミュニケーションの研究では、これまで上記のうち、主として手話に代表される 1. に関する研究が行われてきたが、ノンバーバル・メッセージが本来持っている 5 つの機能をインタフェース機能へ導入すれば、より自然なインタラクションが可能なヒューマンマシンインタフェースが実現される。

2.1.3　インタフェースへの身振りの利用

表 2.1 に示したノンバーバル・メッセージの中でも、身体動作については表 2.2 に示したように、同時に双方向の情報交換が可能であるため、コミュニケーションにおける意図伝達の補足やタイミングの調整で相互の対話をより円滑に進行させる働きが期待され、特に重要である [7] 。Ekman は人間の身体動作の機能を表 2.3 に示すように、標識、例示子、情感表示、調整子、適応子の 5 つに分類している [8]。

この中で、標識はそれ自体が言語的な意味を持ち、音声に翻訳可能な動作であるが、例えば日本では「人差し指と親指で丸をつくる」サインが「お金」の意味を表すなど、文化ごとに異なる傾向もある。例示子は発話に伴った動作で、発話内容の補完的役割がある。情感表示は感情に伴う動作で、感情の種類や強さを表す。調整子は発話の交換や対話の開始や終了のタイミングの制御を行う動作で、会話の流れを円滑にする。適応子は状況に適応するための動作で、頭をかく、ものをもて遊ぶなど無意識に行われる行為である。

表 2.3: Ekman による身体動作の機能的分類

		内容	例
標識	(emblem)	記号性が強い 言語に変換可能	サイン、手話
例示子	(illustrator)	発話内容と 強い関連がある	指差し
情感表示	(a®ect display)	情動に伴うもの	表情・姿勢
調整子	(regulator)	会話の流れと 関連がある	うなずき
適応子	(adaptor)	状況に適応する ためのもの	頭をかく

人々の対話で重要な役割を果たす身体動作では、標識に分類される記号性の強い手話やサインではなく、相手との対話によって無意識に表出する調整子などの身体動作である [6]。

身体動作は、表情と身振りに分けられるが、表情には情感表示の役割が大きい。表情に関しては、本研究室では人間の顔の動画像からリアルタイムで表情を認識する研究 [9] やコンピュータディスプレイ上の 3 次元顔画像を用いて表情を表出する研究 [10] を行っている。

また、「身振り」という単語は一般的には、相手に意図などを伝えるために行う動作を示すことが多い。しかし、社会心理学や文化人類学などコミュニケーションを研究対象として扱っている分野では、コミュニケーションの対象がある場合は人間の動作すべてを身振りとすることが多い。例えば、Watzlawick は「相手がいる状況では、すべての行動がコミュニケーションとなる」といっている [11]。本研究では、身振りを機械とのコミュニケーションの手段に利用することを視野に入れているため、「身振り」という単語を後者の意味でとらえる。身振りをヒューマンマシンインタフェースに利用する場合、コンピュータ側から見ると、その伝達方向より

- コンピュータから人間への身振り提示による情報発信 [12]
- コンピュータが人の身振りを認識することによる情報受信

の 2 通りが考えられるが、本研究では後者の「コンピュータが人間の身振りを認識す

る」場合に着目する。次節では、この分野で行われてきた研究についてまとめる。

2.2　身振りの認識に関する従来研究

身振りをインタフェースに利用するためには、コンピュータが自動的に人間の動作を計測し、その動作から身振りを検出し、認識する必要がある。ここでは、人間の動作を計測する段階と身振りを認識する段階に分けて、従来研究を展望する。

2.2.1　動作計測

人間の動作計測に関する研究としては、人間に直接計測装置を取り付ける接触型センサによるものと、間接的に計測する非接触型センサによるものに分けられる。接触型センサには、ゴニオメータ、手形状センサ[13]、磁気センサ、光学センサなどが挙げられる。これらの接触型センサは、人間の動作を精度良く計測できる特長がある反面、センサの装着に時間を要したり身体に拘束を与えるため、人間に負担をかけ、その動作も制限される欠点がある。

一方、非接触型センサには、CCD カメラ、レンジセンサ、赤外線センサなどを用いて動画像処理により動作を計測するものが多い。非接触型センサは概して計測精度が良くないものの、人間に身体的拘束を与えず自然な状態で計測できる利点がある。また動画像処理では、センサからの画像情報をそのまま扱うとデータ量が膨大になるため、様々な手法によりその画像の特徴を取り出して動作を計測している。この手法の代表的なものとして以下の 3 つがある。

- 微分処理、エッジ処理などを用いて輪郭を抽出する手法
- オプティカルフローを用いて、特に動きのある成分に着目する手法[14][15]
- 人間の身体に特徴点を設定して、その動きを追跡する手法

またこれらには、単独センサによる 2 次元情報で処理を行うものと、複数センサにより、計測した複数の 2 次元画像から生成した 3 次元画像情報を利用するものがある。

2.2.2　身振り認識

動作計測によって得られた特徴を認識し、身振りと対応づける手法に関する研究も、様々なものが試みられている。これにはパターン認識の研究分野でよく用いられる手

法を利用するものが多く、あらかじめ用意したテンプレート動作との類似性を調べるパターンマッチング[16][17]や、事前にニューラルネットワークやHMM（Hidden Markov Model）を用いて学習しておいた動作を認識するもの、部分空間法、最近傍法、決定木を利用したものがある。

現在行われている研究では、手話などのあらかじめシステムによって固定されたパターンとのマッチングにより動作を認識するものが多く、個人による身振りの違いに着目した研究は、ほとんど報告されていない。

以上をまとめると、現在の身振りの認識に関する研究では、主として手話のような標識にあたる明示的な身振りの認識を対象に行われており、その他の身振りの機能に着目して、人の個性的な癖のような暗示的情報の認識を目的とする研究はほとんどない。

以上の従来研究の展望より、特に、ヒューマンマシンインタフェースとして「身振り」認識の今後の研究方針として、以下の課題が重要と考えた。

- 暗示的な身振りの認識

  人間は明示的情報だけでなく、暗示的情報も伝達することにより、はじめて円滑なコミュニケーションを行うことができる[18]。ヒューマンマシンインタフェースに利用する際に、身振りの計測と認識から、あらかじめ決められた標識的な明示情報の認識だけでなく、標識以外の機能を担う暗示的情報も同時に取り出すことができる身振り認識手法が必要である。

- リアルタイムでの認識

  身振り認識のヒューマンマシンインタフェースへの利用では、リアルタイム性が非常に重要である。身振りの認識処理をリアルタイムに行わなければ、コンピュータが人間の状態変化を即座に認識して、適切に対応することができない。特に、コミュニケーションを行っている際には、発話や身振りの発生タイミングが重要で、それをリアルタイムで認識することが必要である[19]。

2.3　研究の目的

前節で述べたように、明示的な情報だけでなく、暗示的な情報も伝達することにより、コンピュータは、より一層円滑に人間とのコミュニケーションを行うことができる。すなわち、ノンバーバル・メッセージとしての人間の「身振り」をコンピュータが認識して、人間が自然に発する暗示的な情報を読み取ることができれば、人間とコン

ピュータの間の円滑なコミュニケーションに役立つであろう。以上より、本研究では、ノンバーバル・メッセージとしての「身振り」を用いるヒューマンインタフェース実現のための基礎研究として、人間の動画像からリアルタイムで身振りを計測して、その身振りを動的に分類する手法を提案し、それに基づいて計算機でシステムを試作することを目的とする。

なお、本研究では、特に身振りの「認識」ではなく、「分類」を目的とする。そこで、本研究での身振り認識と分類の相異を以下に論じる。「認識」とは、対象そのものの本質や意味するところを正確に判断することである。「身振り」の「認識」であれば、その身振りに正確な意味づけをすることまで対象にしなければならない。しかし、本研究では暗示的な身振りに着目するが、このような身振りの認識は動作だけからその意味を正確に判断することは本来困難であり、「身振り」以外の情報がコミュニケーションの「文脈」を理解しないと「身振り」そのものの正確な意味づけまで至らないと考えている。従来研究では、動作を区別することをもって安易に「身振り認識」と定義しているものが多いが、本研究ではこの段階を「身振りの分類」とし、本研究で考えるところの「身振り認識」が如何にあるべきかは、第6章で今後の展望において論じる。

第 3 章 提案する身振りの分類手法

本章では、まず本研究で提案する身振りの分類手法の概要を述べ、続いて、特徴抽出、特徴分析、分類の 3 段階から構成される本手法の詳細を順に説明する。

3.1　提案する手法の概要

本手法では、人が椅子に座って対話している状況を身振りの分類の対象とし、測定する身振りには、暗示的情報を含むあらゆる身振り動作を対象とする。

また、CCD カメラなどで撮影した人間の上半身動画像を入力とし、その動画像から身振り動作を抽出して分類する。その計算処理はリアルタイムで行われ、出力として以下のものを求める。

- 身振りの分類結果

  まず、ある一定の長さの時間における人間の上半身動画像記録を処理して、身振り動作を検出し、分類する。そして、それらの身振り動作がいくつに分類されるか、また、それぞれの身振り動作がどの時刻に行われたかなどの情報を求める。この手法では、抽出したそれぞれの身振り動作が過去の分類結果で決めたグループのいずれに入るか、またどのグループにも属さない新しいタイプであるかを調べて、その結果をその身振り開始時刻と共に出力する。

- 身振りの開始タイミング

  人間同士のコミュニケーションにおいて、動作や発話の表出タイミングにより、相手との引き込み現象を誘発することが知られている。また、同じ動作や発話においても、その表出タイミングが異なることで、違う意味を伝達することがあるため、身振りの開始タイミングは重要な情報であると考えられる。このため、身振りの分類結果を即座に出力することが望ましい。しかし、身振りの分類は、身振り動作開始からその身振りが終了した時点までのデータを元に初めて可能になるため、身振りを分類した結果は身振り動作の終了後しか、出力できない。これでは、本手法を人間とのインタラクションに応用する際には、身振りの表出タイミ

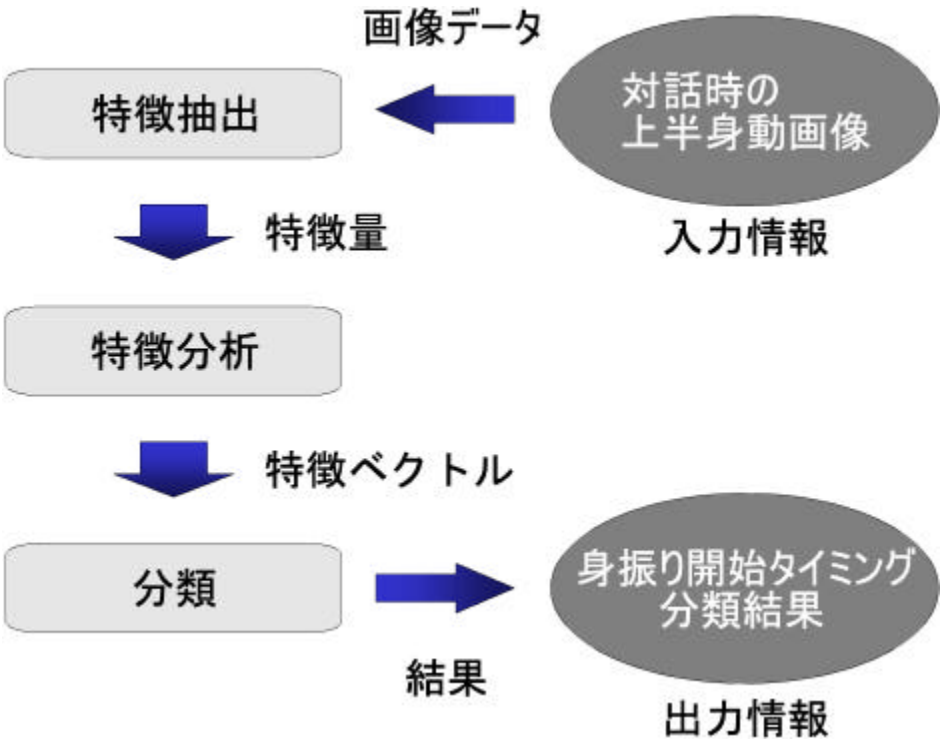

図 3.1: 提案する手法の概要

ングを適切にとることが難しい。そこで、本手法では分類結果とは別に、身振りの開始時点で、まず身振り開始を出力、次に身振りの終了時点で上記の分類結果を出力できるようにする。

図 3.1 に本手法による身振りの分類方法の概要を示す。本手法は特徴抽出、特徴分析および分類の 3 つから構成されいる。まず、特徴抽出部で上半身動画像から身体の部位の動作を表す特徴量を抽出する。そして、特徴分析部で、その特徴量の変化から、身振りを行っている時間帯を抜き出して、その抜き出した部分（身振り動作の部分）の特徴を特徴量を元に分析し、いくつかの特徴成分を取り出す。そして、身振りをそれらの特徴成分を要素とする特徴ベクトルで表現する。最後に身振りを表す特徴ベクトルをクラスタ分析により分類し、区分した身振りが、その時点までに自動的に分類された身振りのタイプのいずれに属するかを求める。

本研究では、人間の身体の動きを「動作」と定義し、その動作から身振りを抽出する。以下では、特徴抽出、特徴分析、および分類の各方法についてその詳細を述べる。

3.2　特徴抽出

特徴抽出部分では、CCD カメラや VTR などによる動的な上半身画像から身体の部位の動作を表す特徴量を抽出する。ここでは、まず動作の特徴について説明し、次にその特徴量の抽出手法について説明する。

3.2.1　動作の特徴と特徴量の役割

本研究では、3.1 で述べたように椅子に座って対話している時に観察される身振り動作の分類を研究対象にする。そのため、ここでは以下のような条件が必要である。

1. 対話時の身振りの特徴を十分に表す特徴量の抽出ができること。

2. リアルタイムで特徴量の抽出が可能な簡便な方法であること。

対象とする人間の身振りは主に上半身の動きとなる。本手法では、上半身の身振りを構成する特徴的な部分として、顔、右手、左手の 3 つの部位に着目し、これらの部位の動作の特徴として、それらの位置情報と形状情報の時間変化を抽出することとする。

空間的位置情報を意味する「位置情報」については、本来、3 次元座標による位置情報として算出することが望ましい。しかし、3 次元位置情報を取得する場合、例えば、複数カメラで撮影された画像から奥行き情報を算出するような複雑な処理が必要であり、またデータ量が 2 次元位置情報と比較して大きくなり、特徴抽出に要する時間が長くなる。そこで本手法では、奥行き情報は画像上の対象領域面積の増減にある程度反映されると考え、3 次元位置情報の代わりに 2 次元位置情報を用いる。具体的には、1 台のカメラによる撮影画像を画像処理して 3 部位を抽出し、それぞれの重心点と領域面積を求める。

一方、形状情報とは、上記の 3 部位がどのような形状かを示す情報である。本来、形状情報を正確に表すには、複雑で多くのパラメータが必要になる。本手法では、あらかじめ対象とする部位を限定しているため、前述の領域面積とアスペクト比（図形の縦、横の比）の 2 つの情報で十分に形状を表すことができると考え、計算処理の簡略化を図ることにした。

本手法では、以上の理由により、図 3.2 に示すように、顔、右手、左手の重心点、領域面積および領域のアスペクト比を特徴量として抽出する。

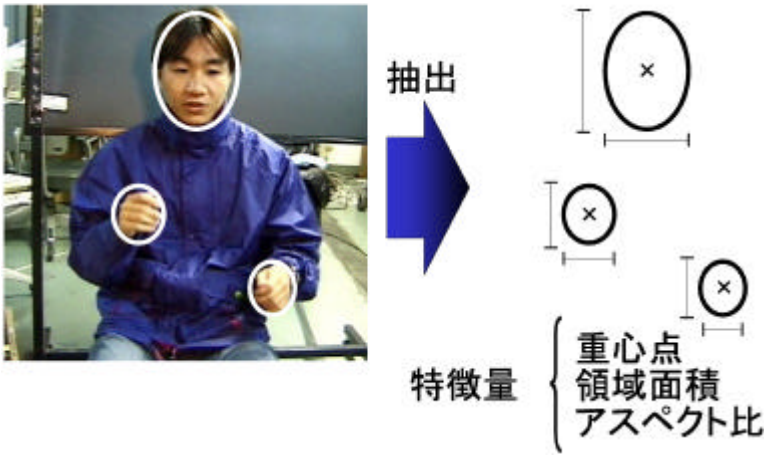

図 3.2: 抽出する特徴量

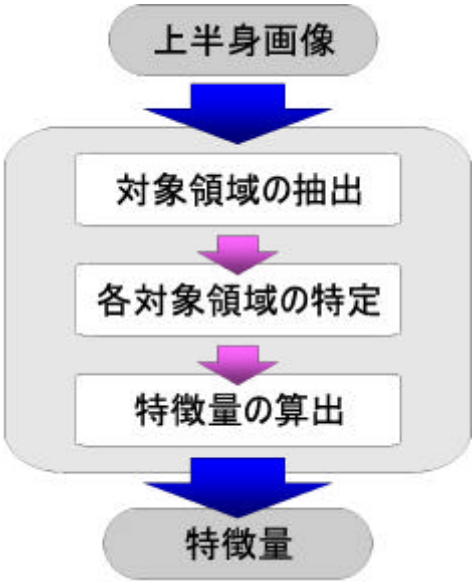

図 3.3: 特徴量抽出手法の流れ

3.2.2 特徴抽出手法の流れ

図 3.3 に 3.2.1 で述べた特徴量を抽出する手法の流れを示す。この手法は、次の 3 つの部分から構成される。

1. 顔、右手、左手領域の抽出
2. 顔、右手、左手領域の特定
3. 特徴量の算出

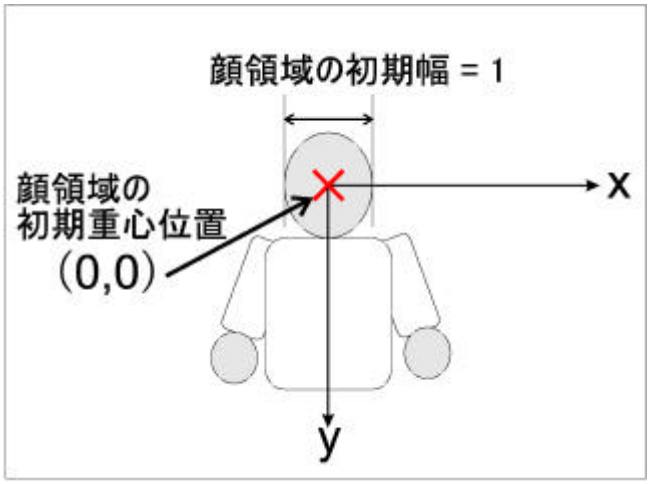

図 3.4: 入力画像の座標系

まず入力された画像から、顔、右手、左手の領域を抽出し、次に、それらの領域が顔、右手、左手のどの部位に相当するかを特定し、最後に各部位の特徴量を算出する。

なお、入力画像では、撮影に用いたカメラの種類や人間とカメラとの距離により人物の大きさが必ずしも一定であるとは限らない。そこで、特徴を抽出する際には、図 3.4 に示すように、顔領域の初期重心位置の座標を原点とし顔領域の初期幅を 1 とする x; y 座標系を一貫して用いることによって、撮影時の人間とカメラとの位置関係が変化した場合の影響を取り除く。

以下では、対象領域の抽出、対象領域の特定、特徴量の算出の各方法について説明する。

3.2.3 対象領域の抽出手法

対象領域の抽出では、入力画像から対象とする部位である顔、右手、左手の部分の領域を抽出する。本手法では、入力される画像の色成分を用い、これらの部位が肌色領域であること利用して対象領域を検出する。色情報を人体部位の認識に利用する場合、一般に使われている RGB 表色系ではなく、HSV 表色系、YIQ 表色系が広く使われる [20]。これらの表色系が、人間の皮膚の色を含む肌色領域の色調を表色の軸とする H 成分または I 成分をもつためである。本手法では、このうち輝度による影響の少ない YIQ 表色系の I 成分を利用する。入力画像の RGB 表色系から YIQ 表色系への変換は次の式（3.1）により行われる。

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.275 & -0.321 \\ 0.212 & -0.523 & 0.311 \end{bmatrix} \times \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{3.1}$$

しかし、肌色領域を対象領域の抽出に用いる場合には、以下に述べる 3 つの問題がある。

1. 肌色領域である「腕」を抽出してしまう。

2. 肌色領域が身体や物の影になる場合に抽出できない。

3. 対象領域が重なった場合、1 つの領域として抽出される。

本研究では、リアルタイム性の維持のため、この簡便な領域抽出方法を用いることとし、上記の問題は以下のように取り扱う。1. の問題については、人間が長袖の衣服を着用し、「腕」の部分を隠すことで解決するが、将来的に克服すべき問題の 1 つであると考える。また、後の 2 つの問題は、対象領域の特定の際に処理方法を工夫し解決する。その詳細については、後述の 3.2.4 で説明する。

前述のように各対象領域の抽出は、肌色領域の抽出により行う。具体的には以下の 6 段階に分けられる。

i. 解像度 640×480［pixel］の入力画像の画素を間引くことにより、解像度を 80×60［pixel］にする。これにより、以降の画像処理を大幅に高速化できるとともに、画像上の細かい雑音成分を除去できる。

ii. 解像度を 80×60［pixel］にした画像から、肌色画素で大きな値をとる YIQ 表色系の I 成分画像を求める。I 成分値は入力画像の各画素の RGB 値から式（3.1）により算出する。

iii. I 成分画像に平滑化を行い、雑音成分をさらに削減する。平滑化には 3×3 メディアンフィルタを用いる。3×3 メディアンフィルタは各画素を中心とする近傍の 3×3 画素の領域の中央値をその画素値とするものである。

iv. iii. の画像を二値化し、I 成分の大きい部分を抽出する。二値化の閾値は、画像の照明環境や肌の色に依存する。本手法では二値化の閾値はあらかじめ設定する。

v. 二値化した画像に膨張・収縮処理を施すことにより、微小な島状の領域を除去し、必要な領域の周囲の凹凸を除去する。この処理は 8 近傍に対し、収縮→膨張→膨脹→収縮の順に行う。

vi. v. の画像に対し領域分割を行う。ここでは、残された領域に対してラベリングを行う。

図 3.5 に、この方法による対象領域の抽出例を示す。

背景画像除去

本手法では、色情報を領域抽出の手がかりにしているため、背景画像に段ボール箱のような肌色の物体がある場合、対象とする領域以外の領域も誤って抽出してしまう。そのため、抽出の信頼性を向上させるために背景画像の除去を行う。背景画像除去の概要を図 3.6 に示す。この方法では、事前に人のいない状態で背景の画像（初期背景画像）を取り込んでおき、これを入力画像と比較し、人物のみの領域を特定する。この処理により、背景画像に肌色の物体がある場合でも、例えば図 3.7 に示すように、正しく人体部分の対象領域を抽出することができ、さらに、偶然、初期背景画像の肌色領域と対象領域の肌色領域が重なった場合でも、その影響を低減することができる。

しかし、この手法は最初に背景画像を取り込む必要がありやや煩雑である。なお、撮影する人物の背景に肌色の物体がない場合にはこの手法を用いる必要はない。

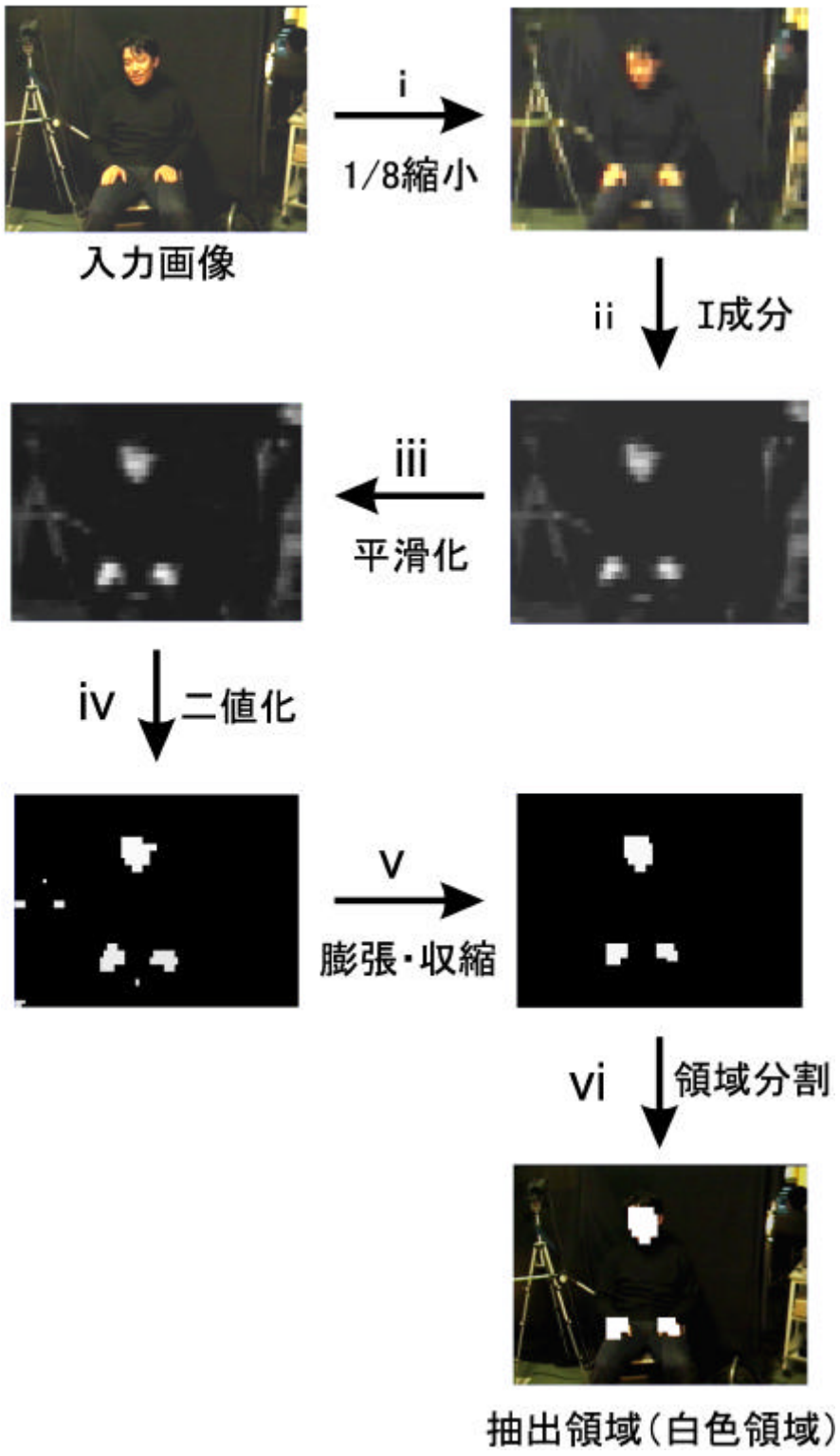

図 3.5: 対象領域の抽出例

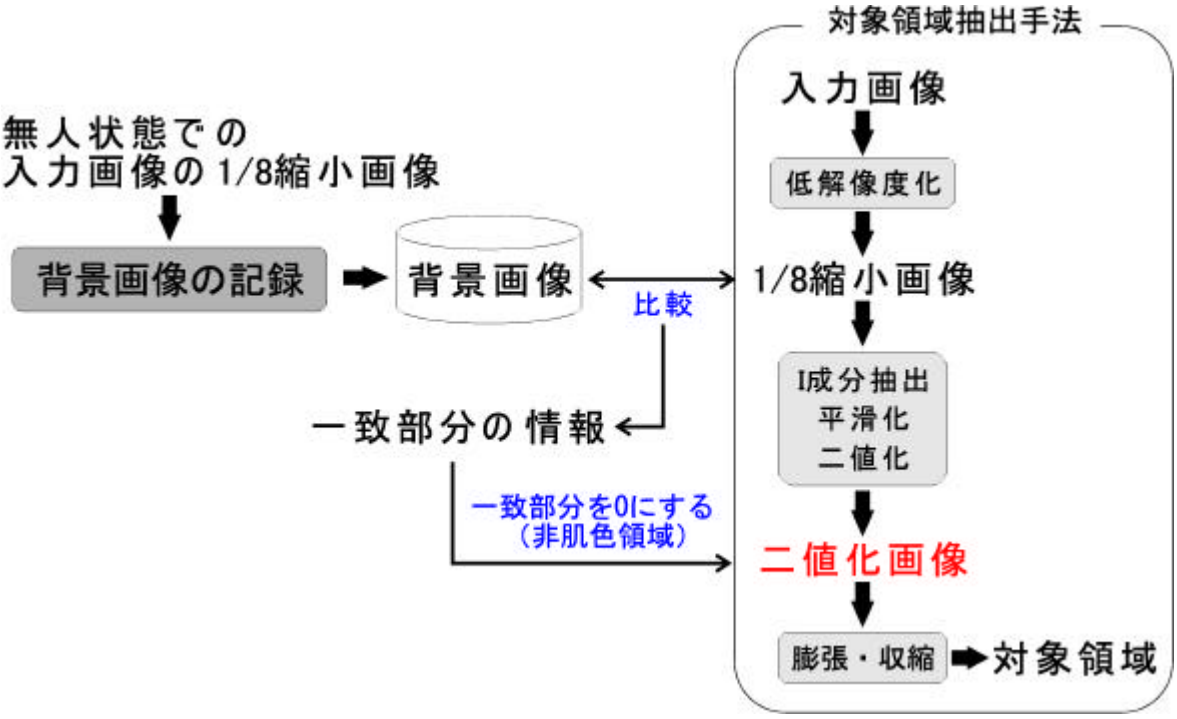

図 3.6: 背景画像除去手法

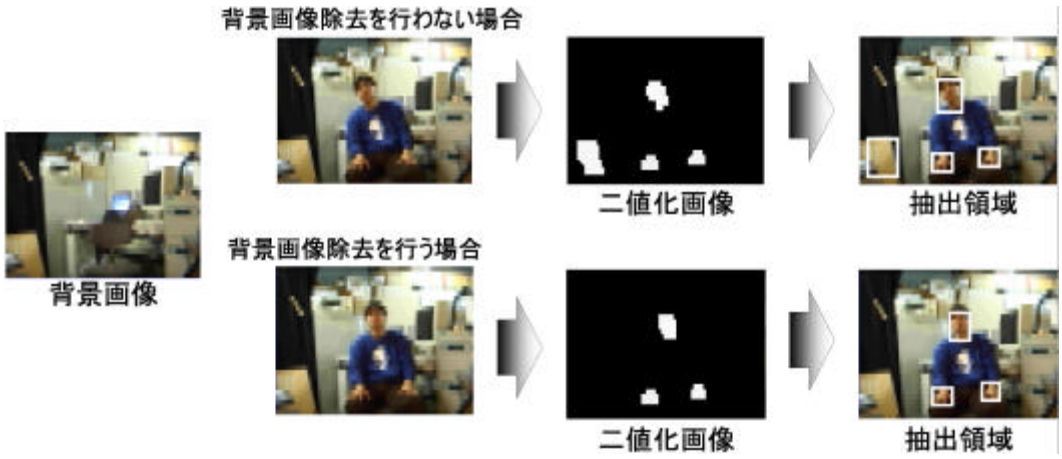

図 3.7: 背景画像除去手法の使用例

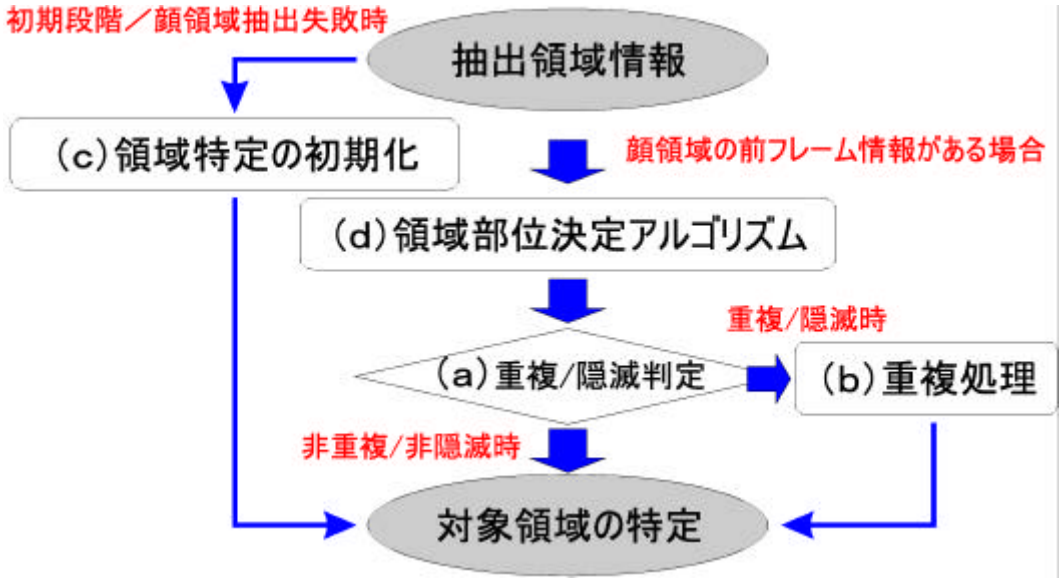

図 3.8: 対象領域の特定手法

3.2.4 対象領域の特定手法

3.2.3 で抽出した領域は、色情報を利用して抽出した肌色領域であり、そのうちのどこが顔、右手、左手に該当するかは特定されていない。ここでは、抽出した領域が顔、右手、左手のどの部分に対応するかを特定する。特定手法の概要を図 3.8 に示す。

この手法の基本は、入力画像の過去 2 フレームの情報を利用し、顔、右手、左手のそれぞれに対して画面上での探索範囲を設定し、その探索範囲内に抽出した領域が存在するかどうかを調べ、対象領域の特定を行うことである。しかし、3.2.3 で述べたよう

な「対象領域が身体や物の影になる場合に抽出できない」（以下、隠滅と呼ぶ）、「対象領域が重なった場合、1つの領域として抽出される」（以下、重複と呼ぶ）という問題があるため、上記のような方法だけでは正しく顔、右手、左手領域の特定ができないことがある。これを考慮し、対象領域の特定では図3.8に示す手法を用いる。以下では、この手法の基本的な考え方を説明する。なお、具体的な手法の詳細は付録Aに譲る。

a. 重複、隠滅の判定

対象同士が重なる場合や、対象が物体の後ろに入る場合には、抽出される肌色領域が減り正しく対象領域を抽出できない。このような場合、対象領域が重複したか隠滅したかを判断する必要がある。ここでは条件を設定し、重複したか隠滅したかを判断する。

b. 重複、隠滅時の処理

前述のように、肌色領域から対象領域を特定するために、過去2フレームの情報を用いる。しかし、領域の重複や隠滅が起こった場合には、そのフレームの領域情報が失われることになり、現フレームでの領域特定だけでなく、次フレームや次々フレームの領域特定ができなくなる。そこで、領域が重複や隠滅した場合に適切な代用データを当てはめることでこの問題を解決する。すなわち、顔領域が隠滅した場合には、顔領域の抽出が失敗したとして前フレームの情報をそのまま用い、次フレームの処理で、次に述べる領域特定の初期化を行う。

また、2つ以上の部位が重複して1つの領域として認識されている状態から、再び各部位に分離する場合には、あらかじめ設定した条件を元に判定する。

c. 領域特定の初期化

この手法では、基本的に過去2フレームの領域特定情報を利用するが、当然、処理の開始時では、過去フレームの領域特定情報は存在しない。この場合、入力画像のうち、あらかじめ設定した範囲から、顔、右手、左手領域を特定する。
また、前述したように顔領域の抽出に失敗した場合にも、この領域特定の初期化を行う。

d. 領域部位特定アルゴリズム

部位を特定する基本手法である本アルゴリズムでは、顔や手の動きに関する特性を考慮に入れた領域特定方法を用いる。基本的な方法は、過去フレームの領域特

定情報を元に部位ごとに、領域を探索する範囲を設定し、その領域内に抽出した領域が存在するかどうかを判定し、最終的に部位を領域に割り当てる。

以上の手法を通して、顔、右手、左手領域を特定する。

3.2.5 特徴量の算出

ここでは、前項で述べた方法により特定した顔、右手、左手の領域から特徴量を算出する。ただし、顔の動きは手と比較して小さいため、本手法で用いる解像度を落とした画像では、顔の動きの特徴量を高い精度で検出できない可能性がある。このため、顔領域の特徴量を算出する際には解像度の高い画像を併用し、そこから顔領域を再検出し、再検出された顔領域から特徴量を算出する。

以下に、顔領域の再検出と抽出する特徴量について詳しく述べる。

顔領域の再検出

前述にように、顔の移動は、左右の手の移動に比べ速度も遅く距離も小さい。例えば、本手法で領域探索や特徴量抽出に用いる 80 £ 60［pixel］の画像解像度では、画面上の顔領域の幅が $w_{face} = 10$［pixel］程度であり、顔の動きを十分に特徴量に反映させることは難しい。このままでは、「うなずき」などの顔による身振りを十分に検出できない可能性がある。

このため、顔領域の特徴量を抽出する際、両手の特徴量抽出時の縦横 2 倍にあたる 160 £ 120［pixel］の解像度の画像を用いる。このため図 3.9 で示すように、顔領域探索範囲において縦横 2 倍の解像度の画像を用いて顔領域を再検出する。すなわち、顔領域探索範囲の縦横 2 倍の解像度の画像に対して、図 3.5 に示した手法で領域抽出を行い、この領域で最大の面積をもつ領域を顔領域とする。図 3.10 に顔領域の再検出例を示す。

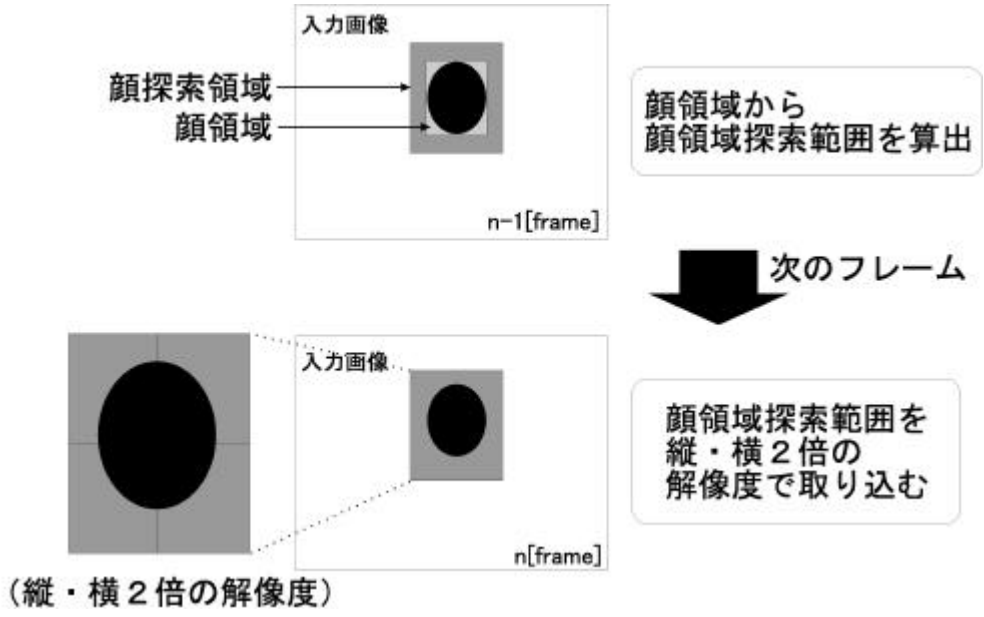

図 3.9: 顔領域の再検出手法

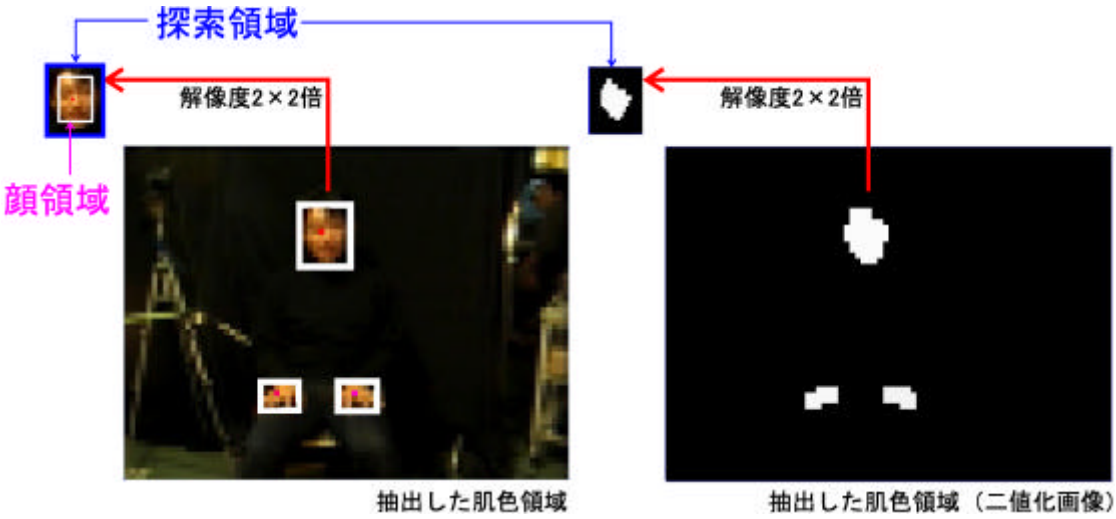

図 3.10: 顔領域の再検出例

抽出する特徴量

抽出する特徴量は、顔領域については再検出された高解像度の顔領域、右手と左手の領域については、領域特定部分で得られた領域から算出する。算出する特徴量は表 3.1 に示す通りである。

面積や重心点の座標は図 3.4 に示した x; y 座標系を用いて表す。アスペクト比は縦／横の対数をとる。すなわち、領域の高さが h_{face}、幅が w_{face} の場合、アスペクト比は $\log_{10}(h_{face}=w_{face})$ となる。

これらの特徴量を毎フレームごとに算出し、さらに、雑音成分除去のため過去 3 フレームでの移動平均を算出して特徴量とする。

表 3.1: 算出する特徴量

記号	抽出する特徴量	基準
ef_{face_x}	顔の重心点の x 座標	初期顔幅を 1 とする（$w_{face\_ini} = 1$）
ef_{face_y}	顔の重心点の y 座標	初期顔幅を 1 とする（$w_{face\_ini} = 1$）
ef_{face_area}	顔の領域面積	初期顔面積を 1 とする（$ef_{face\_area_{ini}} = 1$）
ef_{face_asp}	顔のアスペクト比	$\log_{10}$（縦幅／横幅）= $\log_{10}(h_{face}=w_{face})$
ef_{rhd_x}	右手の重心点の x 座標	初期顔幅を 1 とする（$w_{face\_ini} = 1$）
ef_{rhd_y}	右手の重心点の y 座標	初期顔幅を 1 とする（$w_{face\_ini} = 1$）
ef_{rhd_area}	右手の領域面積	初期顔面積を 1 とする（$ef_{face\_area_{ini}} = 1$）
ef_{lhd_x}	左手の重心点の x 座標	初期顔幅を 1 とする（$w_{face\_ini} = 1$）
ef_{lhd_y}	左手の重心点の y 座標	初期顔幅を 1 とする（$w_{face\_ini} = 1$）
ef_{lhd_area}	左手の領域面積	初期顔面積を 1 とする（$ef_{face\_area_{ini}} = 1$）

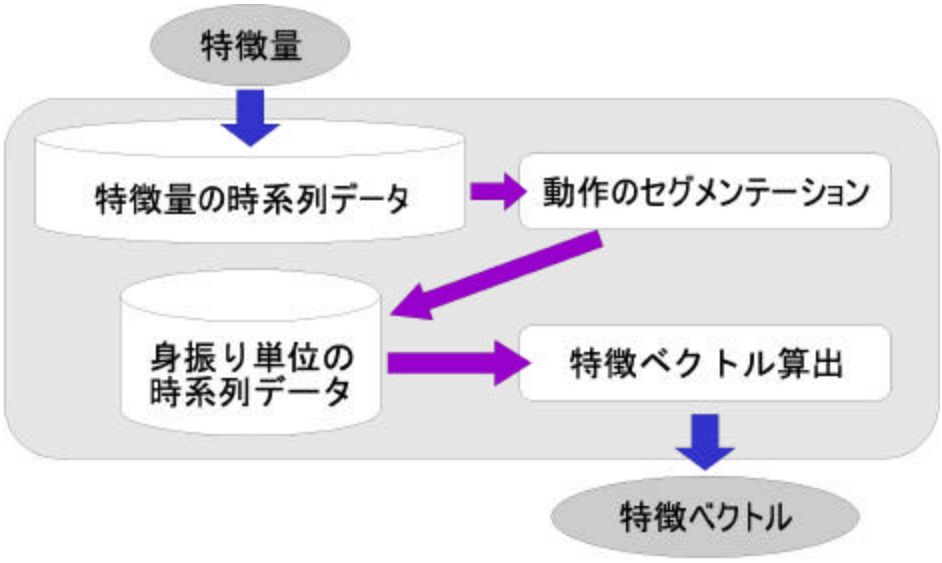

図 3.11: 特徴分析の手順

3.3 特徴分析

特徴分析の手順を図 3.11 に示す。ここでは、3.2 で抽出した動作の特徴量の時系列データを分析し、身振りとする部分を検出し、その身振りの特徴を表す特徴ベクトルを作成する。まず、人の動作を身振りとして意味のある部分で区切り、1 つの身振りとして扱う。これを動作のセグメンテーションと呼ぶ[21]。そして、区切った部分の特徴量の時系列データを分析し、特徴的な情報を取り出す。そして取り出した情報を成分としたベクトルを作成し、これを身振りの特徴を表した特徴ベクトルとする。以下では、この方法について詳しく述べる。

3.3.1 動作のセグメンテーション手法

2.3 で述べたように、本研究ではすべての動作を身振りとして扱うが、「身振り」と見なす部分を区分するために、「動作をしていない」、すなわち「身振りをしていない」状態を定める必要がある。本研究では、この状態を以下のように定める。

[2] 身振りをしていない状態

ある一定時間、顔、右手、左手のすべての特徴量が、微小な身体の揺らぎから生じると考えられる変動値を超えない場合、身体が動いていないとし、身振りをしていない状態とする。

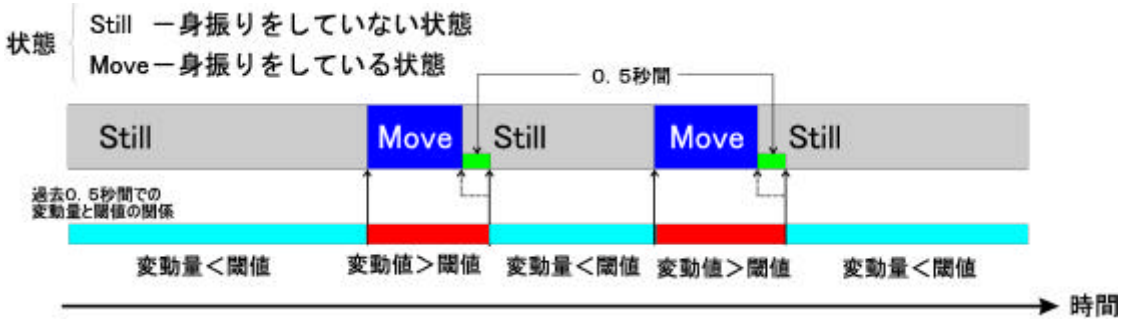

図 3.12: 動作のセグメンテーション手法

本手法では、身振りをしていない状態（Still）の条件を満たさない状態を身振りをしている状態（Move）とする。これは本研究では、身体の揺らぎ以上の動きをすべて「身振り」と定義しているためである。

これにより、身振りの区間は図 3.12 に示すように、身振りをしていない状態の条件を満たさなくなった時点を身振りの開始時間とし、再度、身振りをしていない状態と判断された時点から、判断に用いた時間を逆上った時点を身振りの終了時間として表す。この区間の動作を 1 つの身振りとして扱う。以下、この定義による動作を太字の「身振り」と記述する。

本研究では、上記に示した時間や変動値を決定するため、実際の人間同士の対話を観察して、変動を検知する時間窓を 0.5 秒間（15frame）とし、一方、身体の揺らぎから生じると考えられる閾値を表 3.2 とした。

すなわち、前述の特徴抽出方法で算出されたすべての特徴量 ef_k が表 3.2 の閾値を越えた時を身振りの開始とし、その変動が 0.5 秒間、表 3.2 の閾値より小さくなった時点から 0.5 秒間さかのぼった時点を身振りの終了とする。

ただし、この手法では、1 フレームでの変動が表 3.2 の閾値より小さくなる緩やかな身振りを検出することができないが、このような動作は極度に緩やかな動作であるため、無視できるものと考える。

以上の手法を検討するために、実験を行った。

3.3.2　実験によるセグメンテーション手法の再検討

上記の手法で実際に身振りとする区間を決定できるかを検討するために、動作をセグメンテーションしたものと、人間が身振りを区切ったものとを比較する実験を行った。次に、この実験について述べる。

表 3.2: 身体の揺らぎとする値

セグメンテーションに用いる特徴量	身体の揺らぎとする値
ef_{face_x}（顔の重心点の x 座標）	w_{face_ini}（初期顔幅）× 0.3
ef_{face_y}（顔の重心点の y 座標）	w_{face_ini}（初期顔幅）× 0.3
ef_{face_area}（顔の領域面積）	$ef_{face\_area_{ini}}$（初期顔面積）× 0.3
ef_{face_asp}（顔のアスペクト比）	$\log_{10}2$
ef_{rhd_x}（右手の重心点の x 座標）	w_{face_ini}（初期顔幅）× 0.5
ef_{rhd_y}（右手の重心点の y 座標）	w_{face_ini}（初期顔幅）× 0.5
ef_{rhd_area}（右手の領域面積）	$ef_{face\_area_{ini}}$（初期顔面積）× 0.5
ef_{lhd_x}（左手の重心点の x 座標）	w_{face_ini}（初期顔幅）× 0.5
ef_{lhd_y}（左手の重心点の y 座標）	w_{face_ini}（初期顔幅）× 0.5
ef_{lhd_area}（左手の領域面積）	$ef_{face\_area_{ini}}$（初期顔面積）× 0.5

セグメンテーション実験

［目的］

上記で検討した手法を用いて身振りを適切に区切ることができるかどうかを検証し、さらに、人間が何を手がかりに身振りを認識しているかを調べることを目的とする。

［実験方法］

概要

本実験は、被験者に対話時の人の上半身映像を提示し、被験者が対話している人の動作を身振りと認識した時点を指摘してもらう。また、被験者が何を手がかりに身振りと考えて決めていたかを知るために、実験後にアンケートに答えてもらう。

用意した映像

まず、この実験で用いる対話時の人の上半身動画像データを取得するために、対話時の人の上半身を撮影した。撮影時の状況を図 3.13 に示す。二人に向かい合って話をしてもらい、片方の対話者の上半身画像をディジタルビデオカメラにより撮影した。対話者は男子学生 2 名、時間は 5 分間程度とし、このときの話題は「自分の研究について」とした。このとき撮影した映像の例を図 3.14 に示す。

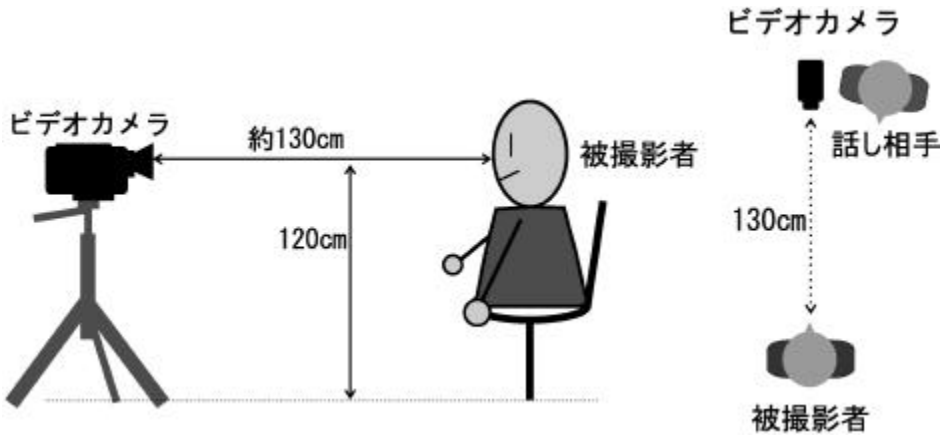

図 3.13: 対話時の上半身動画像撮影状況

図 3.14: 対話時の上半身動画像例

実験手順

セグメンテーション実験の手順を図 3.15 に示す。事前に撮影した対話時の上半身動画像をビデオデッキで再生し、被験者にその画像をテレビモニタで見てもらう。この時の画像はすべての被験者で同じものとし、提示する映像の時間は 3 分間である。ただし、事前に実験者から被験者に約 1 分間、本実験とは別の上半身動画像をビデオを見せながら、実験手順を説明する。被験者に身振りと判断できる動作を見つけるごとに発話により指摘してもらい、その発話の時刻を隣にいる実験者が記録用紙に記載する。

人間は対話時に相手の動作の開始と終了を正確に区切って考えるのではなく、一連の動作をまとめて身振りとして認識するため、実験では身振りの開始と終了の時点を正確に指摘する形式ではなく、身振りであると認識した時点を指摘する形式をとる。また、被験者に対して、「人の癖や無意識の動作も身振りとする」ことを指示した。ここ

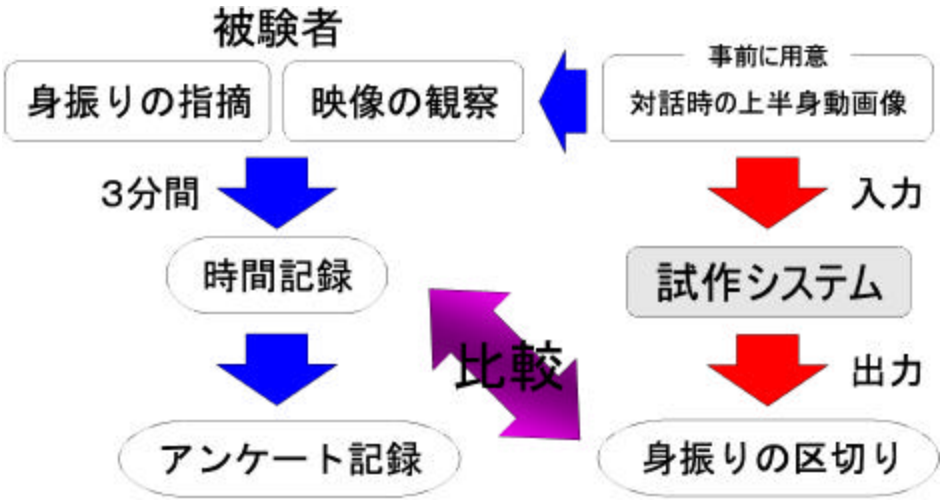

図 3.15: セグメンテーション実験の流れ

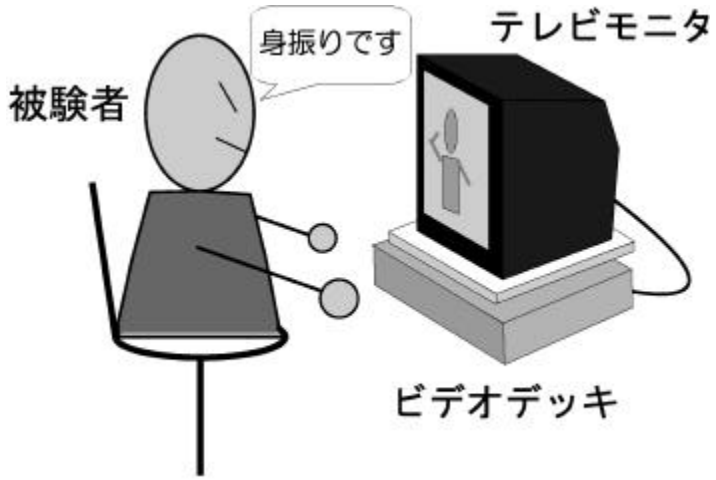

図 3.16: セグメンテーション実験の実験システム

で用いた実験システムを図 3.16 に示す。

また、第 4 章で述べる試作システムのうちの特徴抽出を行う部分を用いて、被験者に提示したものと同じ上半身動画像を入力し、特徴量の変化を調べ、図 3.12 に示した手法で、動作のセグメンテーションを行いその結果を出力する。このとき用いたシステムは、4.1 で述べるハードウェアおよび 4.2 で述べるソフトウェアで構成され、その詳細は第 4 章で説明する。

被験者

被験者は男子学生 YK、KM、EK の 3 名である。

［実験結果と考察］

この実験の結果を表 3.4 に示す。これは、実験で用意した動画像を試作システムに取り込んだときの冒頭からのフレーム番号を左に示し、その時点で被験者および上述の手法を用いた試作システムが身振りと区切ったものを○で示した。この時の 1 フレームの長さは、約 40［msec］である。この結果をまとめたものを表 3.3 に示す。

この結果によると、提案手法が身振りと区切った部分が被験者の区切った数よりも多く、また 3 名の被験者の区切りと提案手法の区切りとは、明らかに異なっている。例えば、3000〜3500 フレーム付近では、被験者の区切りは提案手法での区切りの半分程度しかない。この部分の特徴量の変化を図 3.17〜3.19 に示す。これらのグラフでは、横軸がフレーム数（33msec/frame)、縦軸が特徴量である。この部分では、人が指摘した身振りは、動作をしてから元の位置に戻るまでとしていることに対し、提案手法により区切った身振りは、位置に関わらず、動作して止まるまでと判断していることがわかる。他にも、例に挙げたような、提案手法が 2 つの別々の身振りが連続しているとして判断する動作に対して、人間は 1 つの身振りと判断する動作が多くあった。

また、人間が何を手がかりに身振りと判断しているかのアンケートでは、「動き出してから止まるまで」とする意見が多かった。また、どのような身振りが多く見られたかの問いには、全員が「左手で顔を触る」と回答した。これは、用意した映像の人物の癖が目立ったためであるが、これを本手法では、2 つの身振りと検出することが多かった。よって、本手法では、このような身振りも考慮する必要がある。

表 3.3: セグメンテーション実験の結果のまとめ

被験者	人による区分数	提案手法による区分数	人は指摘するが提案手法が区分しない数	人は指摘しないが提案手法が区分した数
YK	31	46	2	17
KM	25	46	1	22
EK	32	46	2	16

表 3.4: セグメンテーション実験の結果

フレーム数	被験者 YK	KM	EK	提案手法
1670	○	○	○	○
1692				○
1842	○	○	○	○
1870				○
1991	○	○	○	○
2067	○	○	○	○
2183				○
2244	○	○	○	○
2352	○	○	○	○
2413	○		○	○
2548	○		○	○
2576				○
2842	○	○	○	○
2877				○
3083	○	○	○	○
3121				○
3223	○		○	○
3248			○	○
3354	○	○	○	○
3388				○
3555	○	○	○	○
3629	○	○	○	○
3656				○
3764	○	○	○	○
3813				○
3851	○		○	
3981	○	○	○	○
4018				○
4171	○	○	○	○
4197				○
4605	○	○	○	○
4814	○	○	○	○
4850				○
5013	○	○	○	○
5226	○	○	○	○
5260				○
5303	○	○	○	
5477	○	○	○	○
5533	○	○	○	○
5599				○
5767	○	○	○	○
5921	○	○	○	○
6011	○		○	○
6069	○	○	○	○
6142		○	○	○
6162	○		○	○
6200				○
6225	○			○

図3.17 | 図3.19

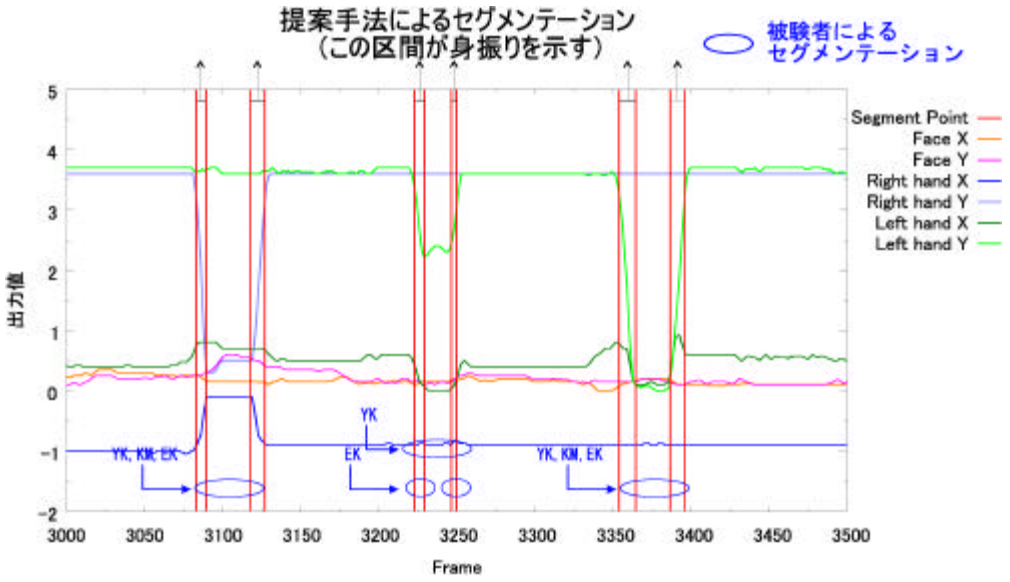

図 3.17: セグメンテーション実験の結果（重心点座標）

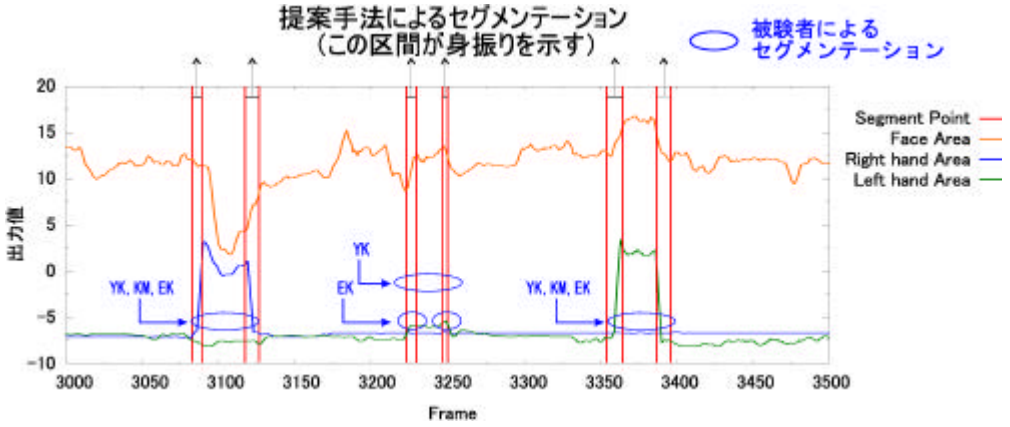

図 3.18: セグメンテーション実験の結果（領域面積）

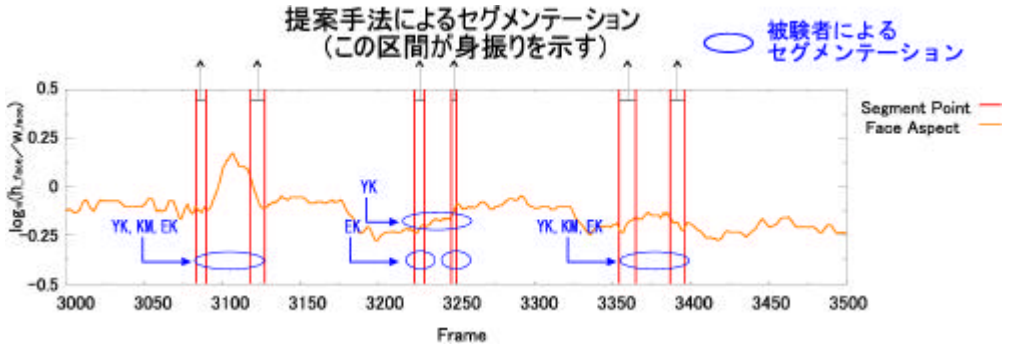

図 3.19: セグメンテーション実験の結果（顔のアスペクト比）

［セグメンテーション手法の再検討］

前述のように、人間は顔や手が動作開始位置に戻るまでに他の位置で止まっていたとしても、元の位置に戻るまでの一連の動作を１つの身振りとして認識しているが、前述のセグメンテーション手法では、2つの身振りとして認識してしまう。このような人間の特性を手法に盛り込むため、セグメンテーション手法を再検討した。具体的には、前述のセグメンテーション手法で得られる2つの身振りを次の式（3.2）に示す連結アルゴリズムにより連結し、人間と同等の身振りの認識を実現する。

$$\begin{aligned} |ef_{k_{st}}(j) - ef_{k_{end}}(j+1)| &< ef_{k_{thrd}} \\ \text{かつ} & \\ i_{st}(j+1) - i_{end}(j) &< 30\text{［frame］（1秒）} \end{aligned} \tag{3.2}$$

の場合、
j と j + 1 の身振りを連結する。

$ef_{k_{st}}$:身振り開始時の特徴量成分 k の値
$ef_{k_{end}}$:身振り終了時の特徴量成分 k の値
i_{st} :身振り開始時のフレーム番号
i_{end} :身振り終了時のフレーム番号

ただし、添字 k は、ff ace_x; f ace_y; f ace_area; rhd_x; rhd_y; lhd_x; lhd_yg である。

これは、1秒間の間に、連続する2つの身振り j と身振り j + 1 を検出した場合、その1つ目の身振りの開始時点の特徴量 $ef_{k_{st}}(j)$ と、2つ目の身振りの終了時点での特徴量 $ef_{k_{end}}(j+1)$ とを比較し、その差がある閾値 $ef_{k_{thrd}}$ 以内である場合は、同じ身振りとみなして連結することを示している。ただし、特徴量のうち顔のアスペクト比と左右の手の領域面積の変化は用いない。ここでは、セグメンテーション実験の結果を分析して、閾値 $ef_{k_{thrd}}$ を表 3.5 に示す値とした。

表 3.5: 身振りを連結する際の特徴量の差

特徴量	身振りを連結する時の特徴量の差
ef_{face_x}（顔の重心点のｘ座標）	w_{face_ini}（初期顔幅）× 0.2
ef_{face_y}（顔の重心点のｙ座標）	w_{face_ini}（初期顔幅）× 0.2
ef_{face_area}（顔の領域面積）	$ef_{face\_area_{ini}}$（初期顔面積）× 1.0
ef_{rhd_x}（右手の重心点のｘ座標）	w_{face_ini}（初期顔幅）× 0.2
ef_{rhd_y}（右手の重心点のｙ座標）	w_{face_ini}（初期顔幅）× 0.2
ef_{lhd_x}（左手の重心点のｘ座標）	w_{face_ini}（初期顔幅）× 0.2
ef_{lhd_y}（左手の重心点のｙ座標）	w_{face_ini}（初期顔幅）× 0.2

表 3.6: セグメンテーション実験の結果のまとめ（連結アルゴリズム導入時）

被験者	人による区分数	提案手法による区分数	人は指摘するが提案手法が区分しない数	人は指摘しないが提案手法が区分した数
YK	31	30	3	2
KM	25	30	1	6
EK	32	30	4	2

この連結アルゴリズムを組み込み、前述の実験と同じ上半身動画像を入力した場合、図 3.17〜3.19 の同一時間帯の結果が図 3.20〜3.22 のようになった。これによって、人がセグメンテーションした場合と同様の結果が本手法により得られた。連結アルゴリズムを導入した場合のセグメンテーション実験の結果を表 3.6 および表 3.7 に示す。

表 3.6 から、連結アルゴリズムの導入により適切に動作をセグメンテーションできているのがわかる。一方、表 3.7 を見ると、少数ではあるが、人が身振りを区切っているにもかかわらず、提案手法が区切っていないケースが増加している。これは連結アルゴリズムの導入により、人が区別している部分を逆にシステムが連結してしまったものと考えられる。

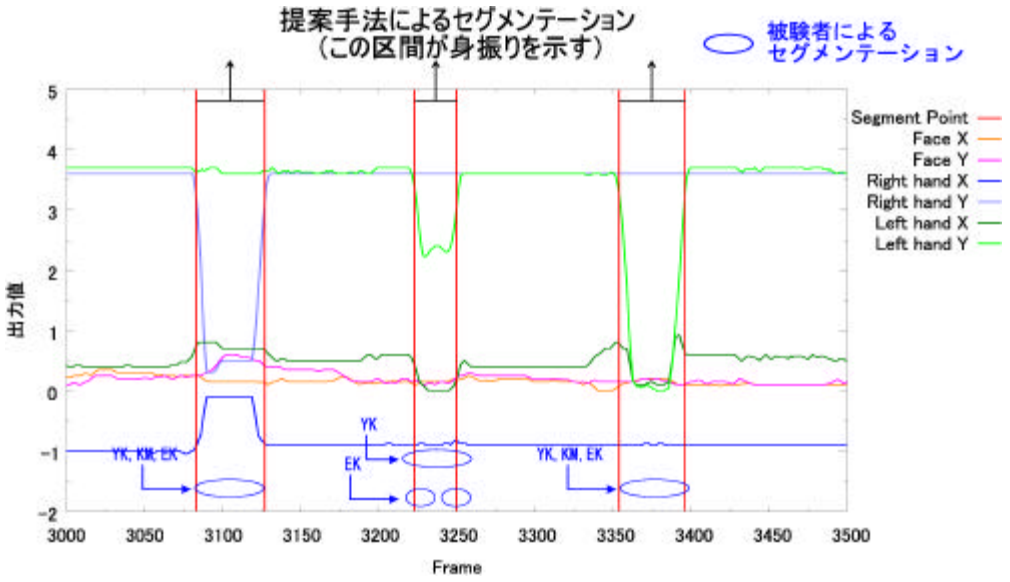

図 3.20: セグメンテーション実験の修正結果（重心点座標）

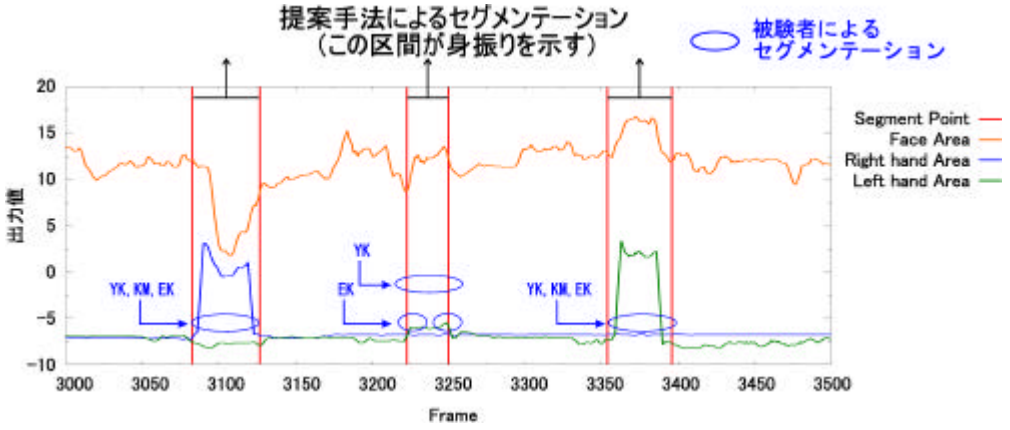

図 3.21: セグメンテーション実験の修正結果（領域面積）

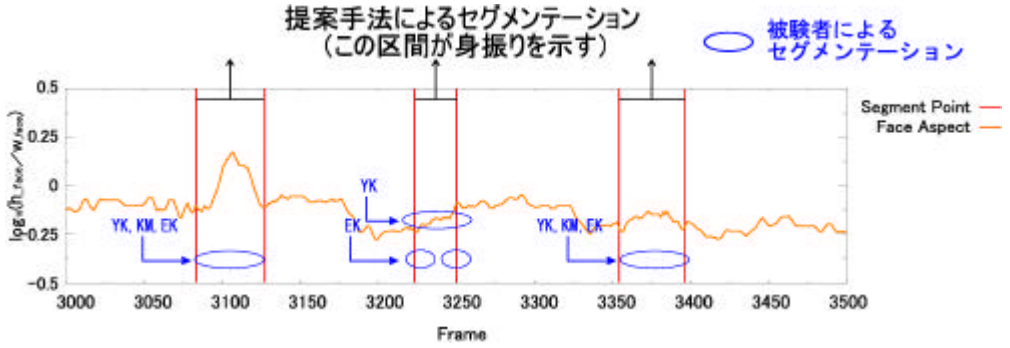

図 3.22: セグメンテーション実験の修正結果（顔のアスペクト比）

表 3.7: セグメンテーション実験の結果（連結アルゴリズム導入時）

フレーム数	提案手法（連結アルゴリズム未導入）	提案手法（連結アルゴリズム導入）	YK	KM	EK
1670	○	○	○	○	○
1692	○				
1842	○	○	○	○	○
1870	○				
1991	○	○	○	○	○
2067	○	○	○	○	○
2183	○	○			
2244	○	○	○	○	○
2352	○	○	○	○	○
2413	○		○		○
2548	○	○	○		○
2576	○				
2842	○	○	○	○	○
2877	○				
3083	○	○	○	○	○
3121	○				
3223	○	○	○		○
3248	○				○
3354	○	○	○	○	○
3388	○				
3555	○	○	○	○	○
3629	○	○	○	○	○
3656	○				
3764	○	○	○	○	○
3813	○				
3851			○		○
3981	○	○	○	○	○
4018	○				
4171	○	○	○	○	○
4197	○				
4605	○	○	○	○	○
4814	○	○	○	○	○
4850	○				
5013	○	○	○	○	○
5226	○	○	○	○	○
5260	○				
5303			○	○	○
5477	○	○	○	○	○
5533	○	○	○	○	○
5599	○				
5767	○	○	○	○	○
5921	○	○	○	○	○
6011	○	○	○		○
6069	○	○	○	○	○
6142	○	○		○	○
6162	○	○	○		○
6200	○				
6225	○	○	○		

図3.20
|
図3.22

3.3.3　特徴ベクトル作成手法

1つの身振りを表す特徴量の時系列データは、多くの情報を含んでおり、そのままではリアルタイムに分類するのは難しい。このため、特徴量の時系列データから分類に適した情報だけを選択し、適切に分類できる指標を作成することが必要である。

本研究では、人間が身振りの分類をする際に動作のどこを手がかりとして着目しているかを調べ、その結果に基づいて分類に適した指標を決定する。

そのため、人間が実際に対話時の身振りを分類するときの手がかりを調べる実験を行い、分類に適した指標を検討した。

分類実験

［目的］

実際に人間が身振りの分類を行う際の手がかりを調べ、人の身振りを分類するための指標を検討することを目的とする。

［実験方法］

概要

本実験は被験者に対話時の映像を提示し、被験者が身振りと認識した時点を指摘してもらい、さらに、指摘した身振りを分類してもらう。また、実験後に何を手がかりに身振りを分類したかを被験者にアンケート用紙に記入してもらう。

実験手順

実験の手順を図 3.23 に示す。実験では、セグメンテーション実験と同じように図 3.13 に示した撮影環境で事前に記録した対話時の身振りの上半身動画像をビデオデッキで再生し、被験者にその映像をテレビモニタで見てもらう。用意した映像は 5 分間である。被験者には身振りと認識した時点で、その時刻をビデオデッキのカウンタで調べて、それを記録用紙に記入してもらう。次に、繰り返し映像を見てもらい、身振り動作の特徴が似ていると判断した動作ごとに分類してもらう。また実験後、被験者に、何を手がかりに身振りを分類したかをアンケート用紙に記入してもらう。

被験者

被験者は男子学生 SH の 1 名である。

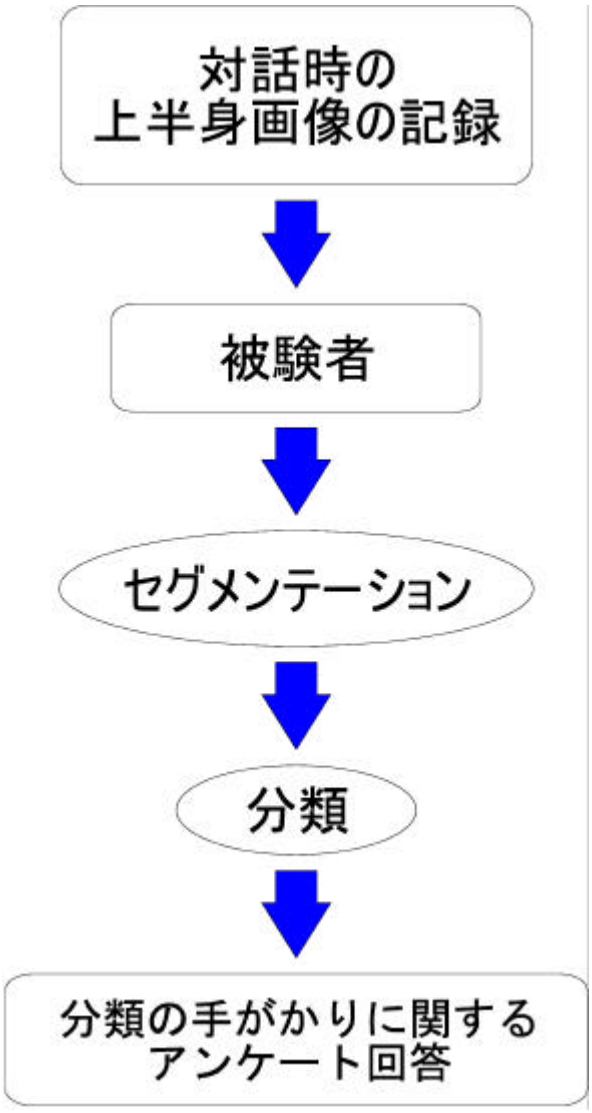

図 3.23: 分類実験の流れ

表 3.8: 分類実験の結果

身振りの分類番号	身振り数	動作内容
1	16	左手で顔周辺を触る身振り
2	6	右手を顔の高さまで上げて右を指さすような身振り
3	2	右手で正面を指さすような身振り
4	2	右手を縦に振る身振り
5	2	右手で地面を指さすような身振り
6	4	両手を広げる身振り
7	3	右手で顔周辺を触る身振り
8	4	左手で胸を掻く身振り
9	4	右手で腹を触る身振り
その他 10〜17	1（計 8）	それぞれ 1 つずつ別の身振り
総数	51	

［実験結果］

被験者による身振りの分類結果を、表 3.8 に示す。

この表では、例えば身振りの分類番号 1 に分類された身振りは、全部で 51 回の身振りのうち 16 回であり、その動作は「左手で顔周辺を触る身振り」であることを示している。被験者が分類した身振りの例を図 3.24 と図 3.25 に示す。図 3.24 は同じ身振りとして分類されたもの（身振りの分類番号 1）である。一方、図 3.25 は、別の身振りとして分類されたもの（左:身振りの番号 7、右:身振りの分類番号 8）である。

また、被験者が何を手がかりに分類しているのかを尋ねるアンケートでは、以下のような点に着目したとの回答を得た。

- 主にどの部位がどの位置に動いたか
- 顔や手がどれぐらい動いたのか
- 動作の長さ（時間）はどれぐらいであったか

これらの回答を考慮すると、各部位の動作開始時点の位置と動作終了時点の位置の情報が特に重要と考えられ、さらに動作の時間や大きさも考慮しているようであった。そこで、自動的に身振りを分類するための指標として、次の特徴ベクトル成分を考えた。

1. 身振りの持続時間
2. 各部位の最大移動距離
3. 動作開始時点の各部位の位置・形状、動作終了時点の位置・形状
4. 動作した部位

特徴ベクトルの成分は、その数が大きすぎると後述のように、その分類手法が複雑になり、さらに計算量が増大して処理に時間がかかる。そのため、特徴ベクトル成分は、身振りを分類するために必要かつ十分な成分の数があり、なるべく少ない方が望ましい。よって、必要最低限の特徴ベクトル成分として重心点は顔、右手、左手のすべての部分の身振り開始位置、終了位置、最大変動量を、x; y 座標ごとに成分とし、領域面積は顔、右手、左手のすべての部分の最大変動量とする。アスペクト比は、手の変動が激しくあまり重要な意味をもたないと思われるため、顔のみの最大変動量を用いる。また、動作の時間は身振り全体の緩急を表しており、重要な意味があると考え、身振りの持続時間を用いる。具体的には、これらの情報を成分とする特徴ベクトルとして、表 3.9 に示す 23 の成分をもつベクトルを身振りの特徴ベクトルＦＶ とする。例えば、1 秒間に相当する 30 フレームの身振りが検出された場面を想定する。この場合、その身振りに関する特徴量は、10 種類の特徴量が 30 フレーム分あるため、300 となる。一方、この 30 フレーム分の身振りの特徴ベクトル成分は 23 であるため、考慮すべきデータ量は大幅に低減できる。

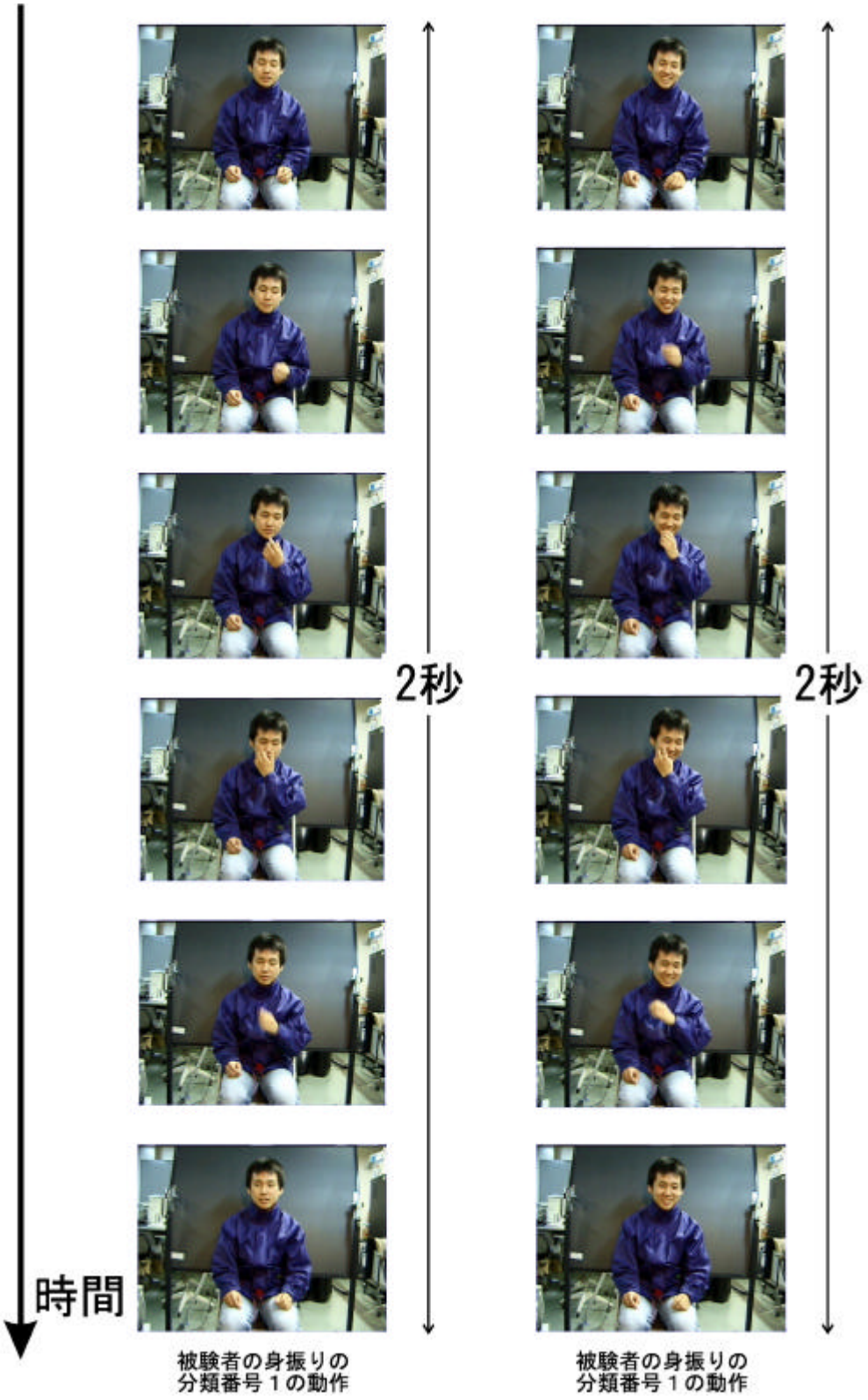

図 3.24: 分類実験の結果（同じ分類の動作例）

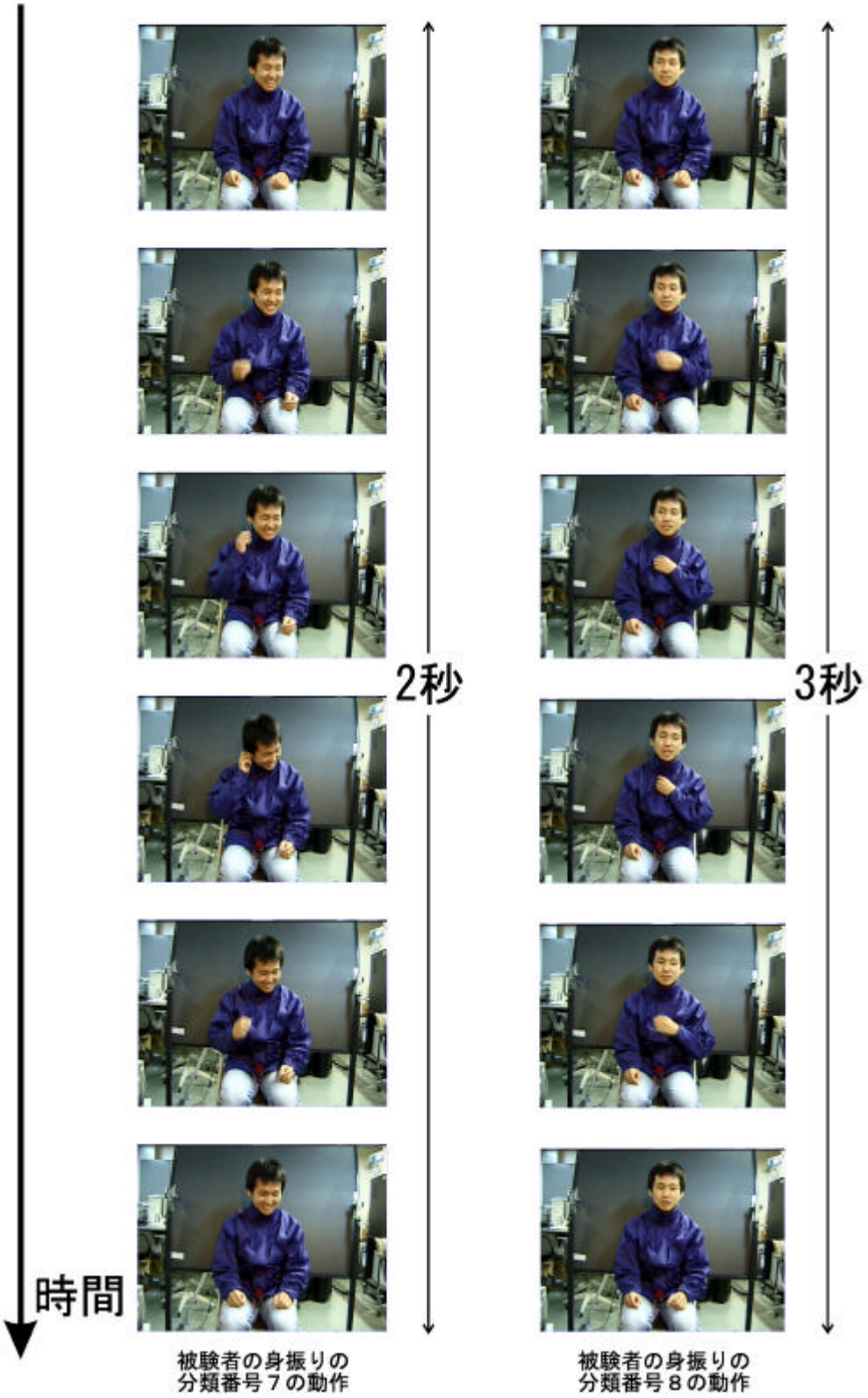

図 3.25: 分類実験の結果（異なる分類の動作例）

表 3.9: 特徴ベクトル成分

番号	記号	特徴ベクトル成分
1	fv_{time}	身振りの継続時間（frame）
2	$fv_{face\_x\_st}$	身振り開始時の顔の重心点 x 座標
3	$fv_{face\_x\_end}$	身振り終了時の顔の重心点 x 座標
4	$fv_{face\_x\_max}$	身振り中での顔の重心点 x 座標の最大変動量
5	$fv_{face\_y\_st}$	身振り開始時の顔の重心点 y 座標
6	$fv_{face\_y\_end}$	身振り終了時の顔の重心点 y 座標
7	$fv_{face\_y\_max}$	身振り中での顔の重心点 y 座標の最大変動量
8	$fv_{face\_area\_max}$	身振り中での顔の領域面積の最大変動量
9	$fv_{face\_asp\_max}$	身振り中での顔のアスペクト比の最大変動量
10	$fv_{rhd\_x\_st}$	身振り開始時の右手の重心点 x 座標
11	$fv_{rhd\_x\_end}$	身振り終了時の右手の重心点 x 座標
12	$fv_{rhd\_x\_max}$	身振り中での右手の重心点 x 座標の最大変動量
13	$fv_{rhd\_y\_st}$	身振り開始時の右手の重心点 y 座標
14	$fv_{rhd\_y\_end}$	身振り終了時の右手の重心点 y 座標
15	$fv_{rhd\_y\_max}$	身振り中での右手の重心点 y 座標の最大変動量
16	$fv_{rhd\_area\_max}$	身振り中での右手の領域面積の最大変動量
17	$fv_{lhd\_x\_st}$	身振り開始時の左手の重心点 x 座標
18	$fv_{lhd\_x\_end}$	身振り終了時の左手の重心点 x 座標
19	$fv_{lhd\_x\_max}$	身振り中での左手の重心点 x 座標の最大変動量
20	$fv_{lhd\_y\_st}$	身振り開始時の左手の重心点 y 座標
21	$fv_{lhd\_y\_end}$	身振り終了時の左手の重心点 y 座標
22	$fv_{lhd\_y\_max}$	身振り中での左手の重心点 y 座標の最大変動量
23	$fv_{lhd\_area\_max}$	身振り中での左手の領域面積の最大変動量

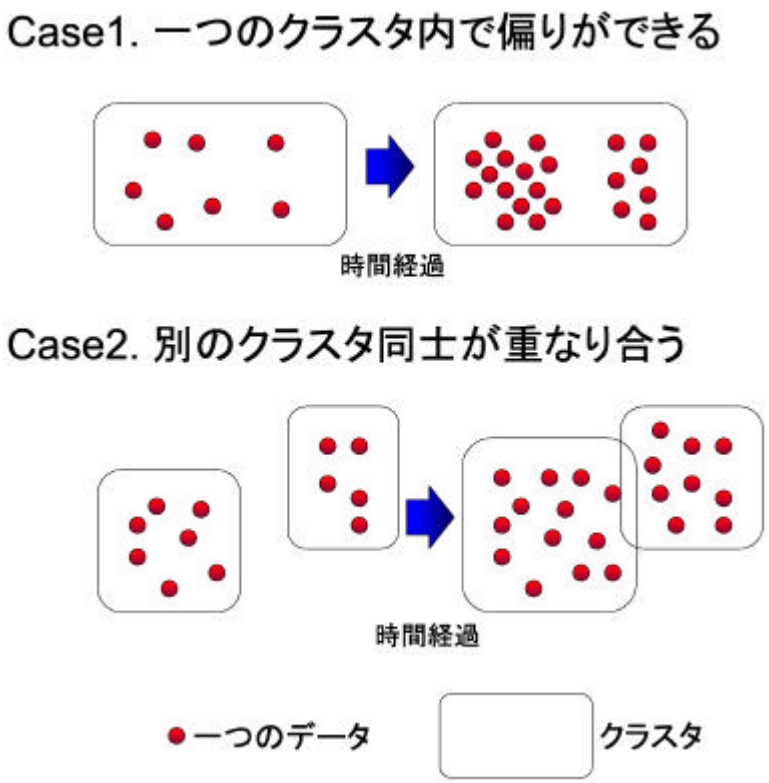

図 3.26: 不適切な分類の例

3.4 分類

分類部分では、 3.3.3 に述べたように身振りの時系列データを符号化した形式である特徴ベクトルを元に、入力された身振りを順次分類していく。そして、リアルタイム性を確保するためには、次々とセグメンテーションされた身振りをそれまでの身振りと比較し、似た身振りの集合であるクラスタに分類していく必要がある。しかし、身振りの検出ごとに既存のクラスタ同士やクラスタ内の身振りのデータを考慮してクラスタを変化させるためには、膨大な計算が必要であり、リアルタイムで処理することは困難である。このため、クラスタの適切な分離や結合ができず、図 3.26 に示すように、時間経過と共に偏った分類になる可能性がある。この問題を解決するため、本手法では適当な時点で分類結果を修正する機構を組み込む。

以下に分類手法の詳細を述べる。

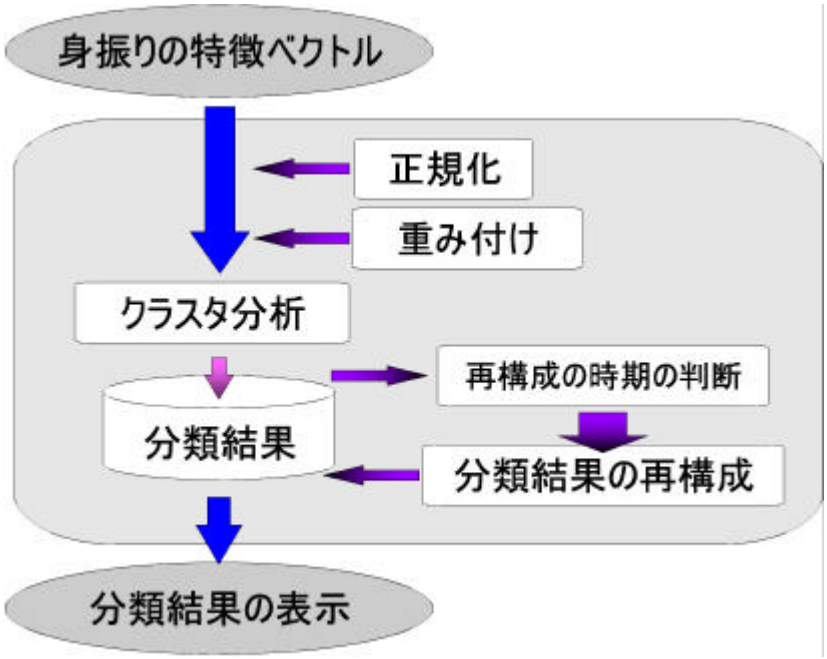

図 3.27: 分類手法の流れ

3.4.1 分類手法の概要

提案する分類手法の流れを図 3.27 に示す。まず、身振りの特徴ベクトルの成分間のばらつきを正規化し、特徴ベクトル成分ごとに、分類に適した重み付けを行う。次に正規化、重み付けをした特徴ベクトル用い、1 つずつ逐次的にクラスタ分析を行い、分類結果を得る。ここで、特徴ベクトルを 1 つのデータ（オブジェクト）として扱い、分類結果を特徴ベクトルの集合（クラスタ）として表現する。

さらに、図 3.26 に示したように、クラスタの中で偏りが生じたり、クラスタが別のクラスタと重なる場合の問題点を解決するために、適当な時点でクラスタ分析の際の閾値などのパラメータを変更し、分類結果を再構成する機構を導入する。

また、クラスタ分析を行う際には、同一クラスタ内のオブジェクト間やクラスタ間の類似度を表す指標として、同一クラスタ内の異なったオブジェクト間および異なったクラスタ間の特徴ベクトル空間上の距離測度を定義する。本手法では、この距離測度としてミンコフスキー距離を用いる。ミンコフスキー距離を用いる理由とその詳細は後述の 3.4.4 のクラスタ分析の説明の際に合わせて説明する。

3.4.2　特徴ベクトルの正規化

ベクトルで表現されたデータ間の距離測度を求める際に、ベクトル中の各成分の変動範囲が異なるため、それぞれのベクトル成分をそのまま用いると、変動範囲の差が距離測度に影響を与える。このため、事前に各ベクトル成分を正規化する必要がある。

正規化する際には、クラスタ分析において各成分の差が距離測度へ与える影響を避けるために、すべてのオブジェクトの集団について、各ベクトル成分の標準偏差を 1 にするように正規化することが多い。しかし、本手法では、逐次的に身振りの特徴ベクトルが追加されていく構成であるため、その都度、蓄積した過去の特徴ベクトルをすべて正規化し直すと処理時間がかかり、リアルタイムに分類することができなくなる。

このため、本手法における特徴ベクトルの正規化では、実際の対話中の上半身動画像を元に事前に算出した特徴ベクトルの各成分の標準偏差を利用する。

ただし、個人の特性に強く影響されないように、複数人による複数の身振りの動画像を用意し、それらの特徴ベクトルの標準偏差を用いる。なお、この値は後述する分類の再構成の際には、抽出した過去の身振りを手がかりに導出したものに置き換えられる。すなわち、分類の再構成では、実際に動作を計測する人の特性に合わせた。

具体的には特徴ベクトルの正規化には次の式（3.3）を用いる。

$$fv_{std_l} = \frac{fv_l}{\sigma_{fv_l}} \tag{3.3}$$

ただし、fv_{std_l}　：正規化後の fv_l 成分の値

σ_{fv_l}　：事前に用意した成分 fv_l の標準偏差

3.4.3　重み付け

一方、身振りの分類を適切に行うには、特徴ベクトルのどの成分が分類によく寄与するかを判断し、その成分を重点的に用いることが必要である。また、分類の精度を上げるには、生成されるクラスタに関して次の 2 つの条件を満たすことが望ましい。

条件 1. 類似した身振りがクラスタ内で凝集すること。

条件 2. 別々のクラスタ同士ができるだけ分散すること。

条件 1. を満たすにはクラスタ内の特徴ベクトル間の距離の平均ができるだけ小さくなることが必要であり、一方、条件 2. では各クラスタの代表値として、クラスタ内の

すべてのオブジェクトのベクトル成分の平均値を用いた代表ベクトル間の距離の平均が大きくなることが必要である。よって、これらの条件を満たすように、特徴ベクトルの成分に重み付けを行う。

実際に重み付けに用いる値として、3.3.3 で行った分類実験の結果を参考にこの条件を満たす値を設定する。分類実験の結果を用いることで、上記の条件を満たすだけでなく、分類に大きく寄与する特徴ベクトルの成分に大きな重みをつけることができる。その手法は図 3.28 に示すように、まず、分類実験と同じ画像から特徴量を算出し、次に、この実験で得られた身振りの特徴ベクトルの標準偏差を用いて正規化を行う。これにより得られた正規化された特徴ベクトルを分類実験で被験者が分類したクラスタに当てはめる。このとき、セグメンテーションの位置が分類実験で行った人間と前述のセグメンテーション手法とで異なるものは除去する。

そして、条件 1.、条件 2. を次のような制約条件下の目的関数の最適化問題に置き換える。

・目的関数　クラスタ間距離の平均
・制約条件　クラスタ内のベクトル間距離平均 ・ 定数 1
　　ただし クラスタ間距離の平均 ¸ 定数 2

ここで、クラスタ間距離を表す目的関数を表す変数を P(WGT) は、次式（3.4）で与えられる。

$$P(WGT) = \frac{1}{n} \sum \overline{FV_{rep_l} FV_{rep_m}} \tag{3.4}$$

$$FV_{rep_l} = \frac{1}{k} \sum_{i=1}^{k} FV_{std_i} \cdot WGT^T \tag{3.5}$$

FV_{std_i} :正規化された特徴ベクトル i
k :クラスタに含まれる特徴ベクトルの個数
WGT :重みベクトル
FV_{rep_l}、FV_{rep_m} :クラスタ l、m の代表ベクトル
$\overline{FV_{rep_l} FV_{rep_m}}$: FV_{rep_l} と FV_{rep_m} の距離
n :分類された全クラスタのうちの 2 つのクラスタの組み合わせ数

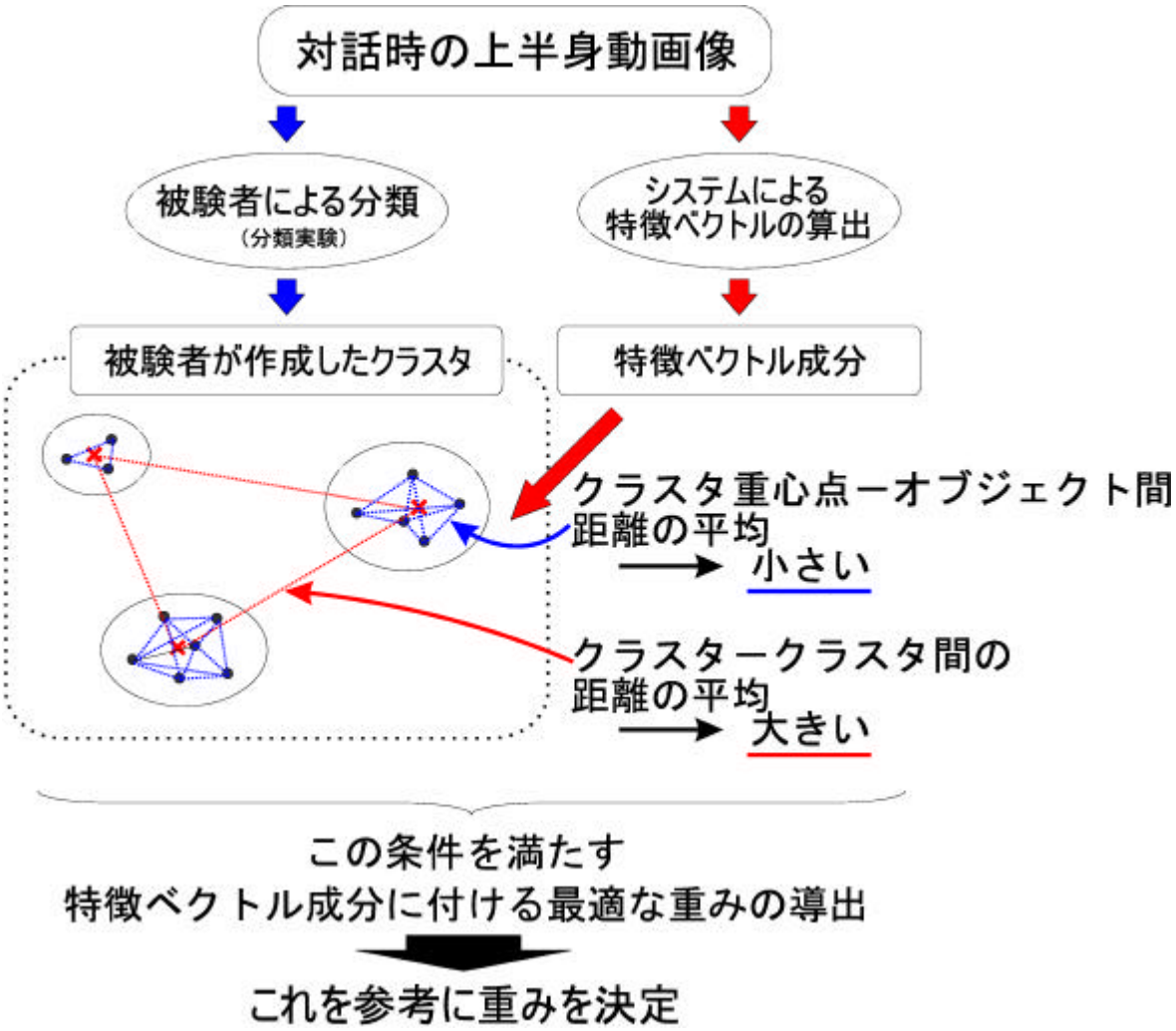

図 3.28: 重み付け決定手順

そして、準ニュートン法[22]により、目的関数を最大にする重み成分の最適解を求めた。これにより求めた重み成分を表3.10に示す。

ただし、求めた最適値は、3.3.3で述べた分類実験の際に使用した動画像に写っている身振りの種類に関係しているため、この重み成分を特徴ベクトルへの重み付けに単純にこのまま用いることは適切ではない。例えば、身振り開始時の右手のｙ座標成分の重みは$wgt_{rhd_y_st}$=1.6、左手については$wgt_{rhd_y_st}$=0.49である。右手の方の重み成分が大きいのは、分類実験で用いた上半身動画像に写っている動作の中で右手を上下に動かす身振りが多かったためである。元来、人体は左右対称のため、左右で重み成分が異なるのは好ましくない。そこで、分類実験から求めた重み成分を左右で同じ値になるように再検討し、表3.10の右列に示した値を用いることにした。ここでは、簡単化のため各値の最小値を0.25にして、それ以上は0.25ごとに離散化した値にしている。

次項で述べるクラスタ分析の際には、正規化した身振りの特徴ベクトルFV_{std}の各成分に、次の式（3.6）に示す式で重みを付加した特徴ベクトルFV_{wgt}を用いる。

$$FV_{wgt} = \begin{pmatrix} fv_{wgt_1} \\ fv_{wgt_2} \\ \vdots \\ fv_{wgt_{23}} \end{pmatrix} = \begin{pmatrix} fv_{std_1} \\ fv_{std_2} \\ \vdots \\ fv_{std_{23}} \end{pmatrix} \pounds \begin{pmatrix} wgt_1 \\ wgt_2 \\ \vdots \\ wgt_{23} \end{pmatrix}^T \tag{3.6}$$

ただし、重み付けする前の正規化した特徴ベクトル成分をfv_{std_l}、重み成分をwgt_l、重みを付加した特徴ベクトル成分をfv_{wgt_l}とする（ただし、l=1～23）。

表 3.10: 重み成分

番号	記号	重み成分 ［分類実験結果より導出］	重み成分 ［本手法で使用］
1	wgt_{time}	$2.7 \pounds 10^{i5}$	0.25
2	$wgt_{face\_x\_st}$	1.0	1.0
3	$wgt_{face\_x\_end}$	1.1	1.0
4	$wgt_{face\_x\_max}$	0.55	0.50
5	$wgt_{face\_y\_st}$	0.68	1.0
6	$wgt_{face\_y\_end}$	1.2	1.0
7	$wgt_{face\_y\_max}$	0.71	0.75
8	$wgt_{face\_area\_max}$	0.013	0.25
9	$wgt_{face\_asp\_max}$	0.59	0.50
10	$wgt_{rhd\_x\_st}$	0.38	1.0
11	$wgt_{rhd\_x\_end}$	0.059	1.0
12	$wgt_{rhd\_x\_max}$	1.3	1.5
13	$wgt_{rhd\_y\_st}$	1.6	1.0
14	$wgt_{rhd\_y\_end}$	1.6	1.0
15	$wgt_{rhd\_y\_max}$	1.7	1.5
16	$wgt_{rhd\_area\_max}$	1.23	1.0
17	$wgt_{lhd\_x\_st}$	1.9	1.0
18	$wgt_{lhd\_x\_end}$	1.3	1.0
19	$wgt_{lhd\_x\_max}$	1.7	1.5
20	$wgt_{lhd\_y\_st}$	0.49	1.0
21	$wgt_{lhd\_y\_end}$	0.61	1.0
22	$wgt_{lhd\_y\_max}$	1.6	1.5
23	$wgt_{lhd\_area\_max}$	0.33	1.0

3.4.4　クラスタ分析

定量的なデータ群の分類を扱う多変量解析手法の1つとしてクラスタ分析がある。クラスタ分析には、大別して「階層的手法」、「分割最適型」の2つがあり、「階層的手法」の代表的なものは「凝集法」であり、「分割最適型」の代表的なものは「K-means 法」である[23][24][25]。以下にそれぞれを説明する。

- 凝集法
  最初は、すべてのデータを1つのクラスタとして扱い、順次、距離（非類似度）の小さいクラスタ同士を融合していく手法である。

- K-means 法
  あらかじめクラスタ数を決めておき、その数のクラスタを作成していく手法である。最初はランダムなクラスタから開始し、クラスタ内での変動が小さくなりクラスタ間の変動が大きくなるように、クラスタ間でオブジェクトを移動させていく手法である。

本研究では、次々に発生する人間の身振りをリアルタイムで逐次的に分類することを前提とするため、事前に適当なクラスタ数を与えておくことはできない。このため類似性を利用した凝集法が最も適していると考えられる。凝集法を用いる際の類似性を表す距離測度として様々なものが提案されているが、ここではミンコフスキー距離を用いる。また、凝集のためのクラスタ間の距離の定義にも様々なものがあるが、本研究ではクラスタ重心間の距離をクラスタ間の距離とする重心法を利用する。

すなわち、式（3.7）に示すように、クラスタ C_m に属する n 個の特徴ベクトル FV_{wgt_i} の各成分の平均をそのクラスタの代表ベクトルとし、クラスタ間の距離やクラスタと特徴ベクトル間の距離を計算する際に用いる。そして、クラスタ m の代表ベクトルを表す変数 FV_{rep_m} は次式（3.7）で与えられる。

$$FV_{rep_m} = \frac{\sum_{i=1}^{n} FV_{wgt_i}}{n} \tag{3.7}$$

ただし、FV_{wgt_i} は重み付け後の特徴ベクトル i である。

なお、これにより、後述の式（3.12）を用いて、クラスタ間距離や身振りの特徴ベクトルとクラスタとの距離を算出する。

本研究でのクラスタ分析の流れを図 3.29 に示し、以下にその計算の手順を説明する。

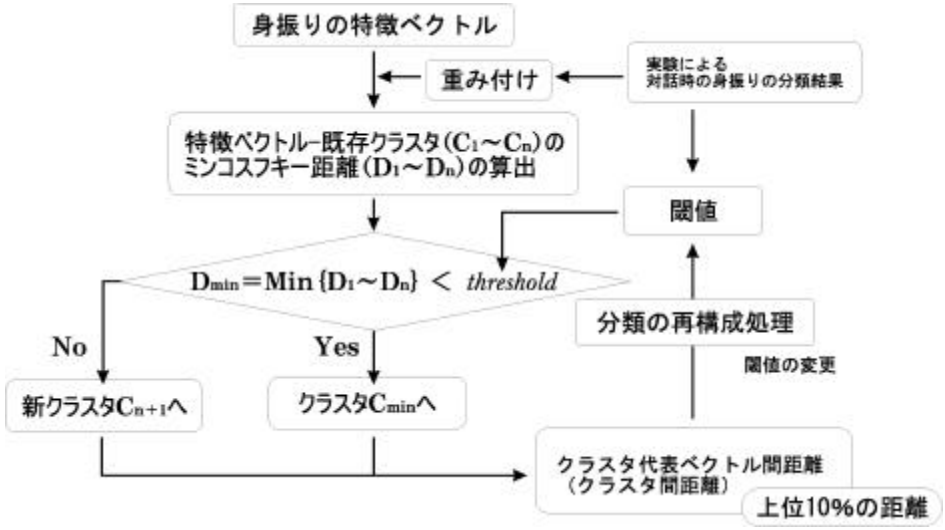

図 3.29: クラスタ分析の流れ

(1) **身振りが検出されると、その身振りの重み付け後の特徴ベクトルと既存の各クラスタ** C_m **との各距離** D_m **を次式（3.8）で計算する（**m = 1～n; n **は既存のクラスタ数)。**

$$D_m = \overline{FV_{wgt} FV_{rep_m}} \tag{3.8}$$

FV_{wgt} :**特徴ベクトル**

FV_{rep_m} :**クラスタ** m **の代表ベクトル**

(2) **次に、各距離** D_m **のうち次の条件式（3.9）を満たす最小値をとるクラスタ** C_k **を探す。**

$$8m; D_k \cdot Dm \tag{3.9}$$

(3) **そして、** D_m **が次の条件式（3.10）を満たせば、その特徴ベクトル** FV_{wgt} **をクラスタ** C_k **に分類する。**

$$D_m \cdot D_{thd} \tag{3.10}$$

ただし、 D_{thd} **は閾値である。**

(4) 一方、D_m が条件式（3.10）を満たさないときは、FV_{wgt} を要素とする新しいクラスタ C_{n+1} を作成する。

以下には、本手法でのクラスタ分析における、ミンコフスキー距離と、分類の初期閾値について詳しく述べる。

ミンコフスキー距離

距離測度として一般的に広く用いられているものは、ユークリッド距離である。2つのデータ A、B（$A(a_1; a_2; :::; a_n); B(b_1; b_2; :::; b_n)$）の間のユークリッド距離 $\overline{AB}$ は次の式（3.11）によって表される。

$$\overline{AB} = \sqrt{\sum_{l=1} (a_l - b_l)^2} \tag{3.11}$$

しかし、この距離測度では各成分値の差が距離に与える影響が大きい。例えば、2つのベクトル間において、ある特定の成分の値が大きく異なり、他の成分の値はほぼ同じような値の場合でも距離が大きくなる。

一方、ミンコフスキー距離と呼ばれる距離測度での AB 間の距離 $\overline{AB}$ は式（3.12）で表される。

$$\overline{AB} = \left(\sum_{l=1} (a_l - b_l)^p \right)^{\frac{1}{r}} \tag{3.12}$$

ただし、この r と p はあらかじめ指定する値であり、p は個々の成分の違いに与える重みを調整し、一方 r はオブジェクト間の大きな差を与える重みを調整する。

ここでは、$r = p = \frac{1}{2}$ の場合を考える。この場合、2つのベクトル間で、ある特定の成分の値だけが大きく異っても、他の成分の値がほぼ同じなら、ミンコフスキー距離は大きくならない。逆に、各成分の値が少しずつ違う 2 つのベクトル間のミンコフスキー距離は大きくなる。本手法では、特徴ベクトルの成分が 23 あり、同じ身振りにもかかわらず、ノイズ等の影響によりそのうちの特定の成分だけが著しく違う特徴ベクトルが得られる可能性があるため、本研究では、その影響を省き分類時のロバスト性を向上させるために、$r = p = \frac{1}{2}$ としたときのミンコフスキー距離を距離測度として用いる。

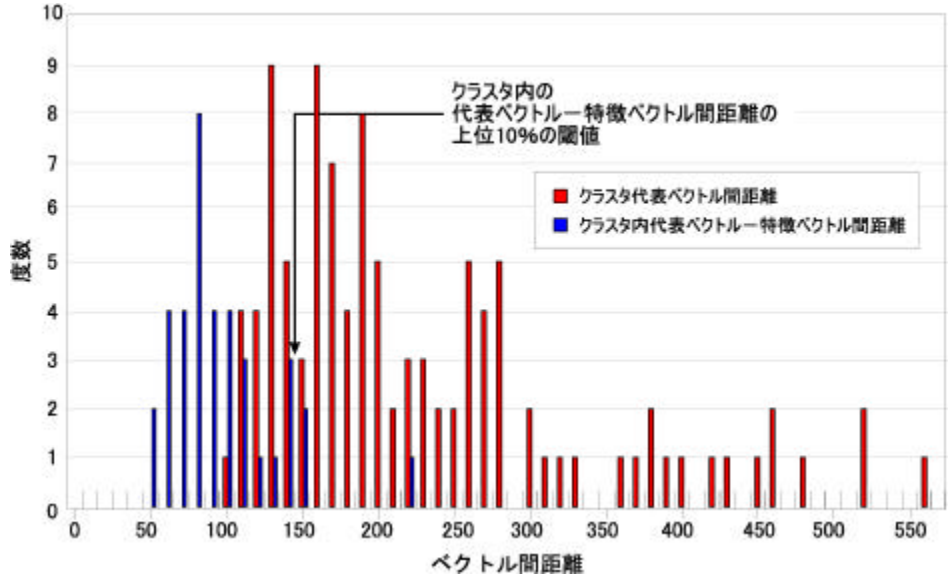

図 3.30: クラスタ間距離・クラスタ内ベクトル間距離の分布

分類の初期閾値

凝集法では、クラスタ間距離の閾値 D_{thd} の程度によってクラスタの分類の良し悪しが決定される。本研究でのクラスタ分析では、初期段階において、3.3.3 で述べた分類実験の結果を元に分類の閾値を算出し、その閾値により分類を行う。すなわち、新しく身振りが抽出されたときには、その特徴ベクトル FV_{wgt} と既存のクラスタ C_m との距離 D_m を調べ、その距離が閾値 D_{thd} 以下のクラスタに分類するものとする。すべての既存のクラスタとの距離 D_m が閾値 D_{thd} より大きければ、その特徴ベクトル FV_{wgt} を新しいクラスタ C_{n+1} を生成する。

閾値 D_{thd} は、前述の分類実験で被験者が身振りを分類した結果を用い、同じクラスタに分類された特徴ベクトル FV_{wgt} とそのクラスタの代表ベクトル FV_{rep} との距離 D_m の分布を参照して適切な値に設定する。

図 3.30 に、分類実験で同じクラスタ内に分類された特徴ベクトル FV_{wgt} とそのクラスタの代表ベクトル FV_{rep} との距離 D_m の分布を示す。グラフには各クラスタの代表ベクトル間の距離も示す。ただし、分類実験で作成された複数のオブジェクトをもつクラスタは 43 あった。

クラスタの生成を適切に行うためには、各クラスタ間の距離と、クラスタ内の代表ベクトルとの距離が適切に区分されるように閾値 D_{thd} を決める必要がある。

図 3.30 より、この閾値 D_{thd} として、その度数の変化から、140 前後の値で区切るべ

きであると判断した。具体的には、特徴ベクトル FV_{wgt} とそのクラスタの代表ベクトル FV_{rep} との距離 D_m の分布で、上位 10%に当たる距離を閾値 D_{thd} とする。

ただし、重み付けの際と同様、分類した人間や身振りに深い関わりがあるため、分類の初期閾値として、D_{thd} = 140 として用い、後述する分類の再構成時には実際に抽出した身振りを再分析することにより閾値を変更する。

3.4.5　分類結果の再構成手法

3.4.1 で述べたように、上記の分類手法により順次入力される身振りを逐次分類すると、クラスタに偏りが生じたり、2 つのクラスタが重なったりして適切に分類できなくなる可能性がある。そこで、本研究では適当な時期に今までの分類を再分析し、分類結果を再構成する。分類結果の再構成処理の流れを図 3.31 に示す。

再構成の手法は以下の 3 つからなる。

1. 過去のすべての特徴ベクトル成分 fv_l から、各成分の標準偏差 $¾_l$ を算出し、正規化に用いる値を変更する。

2. 分類の閾値となる距離 D_{thd} を変更する

3. 過去の身振りに対して、変更した正規化値 $¾_l$ と分類の閾値 D_{thd} を用い、改めてクラスタ分析を行う。

1. では 3.4.2 で述べたように、初期段階において特徴ベクトル成分 fv_l の正規化に用いる標準偏差 $¾_l$ は事前に求めた一定値であるが、身振りの特徴ベクトル成分の変動には個人差があると考えられるため、過去の身振りの特徴ベクトル成分の標準偏差を正規化に利用するように変更する。

2. では前項で述べたように、初期段階では分類の閾値 D_{thd} として事前に求めた一定値を用いているが、1. と同様に身振りには個人差があると考えられるため、分類に用いる閾値として、抽出した過去の身振りの分類結果を参考にし、各クラスタ内のベクトル間距離の分布から、前項の手法により閾値 D_{thd} を決める。

3. では、分類を行う際に、身振りの特徴ベクトルとの距離を測定する対象となるクラスタを適切なものにするため、過去の分類結果を破棄し、1.、2. で決定した正規化に用いる標準偏差 $¾_l$、分類の閾値 D_{thd} を用い、再びクラスタ分析を行う。このクラスタ分析の手法としては、はじめにすべての特徴ベクトル FV_{wgt} を個別のクラスタと考え、

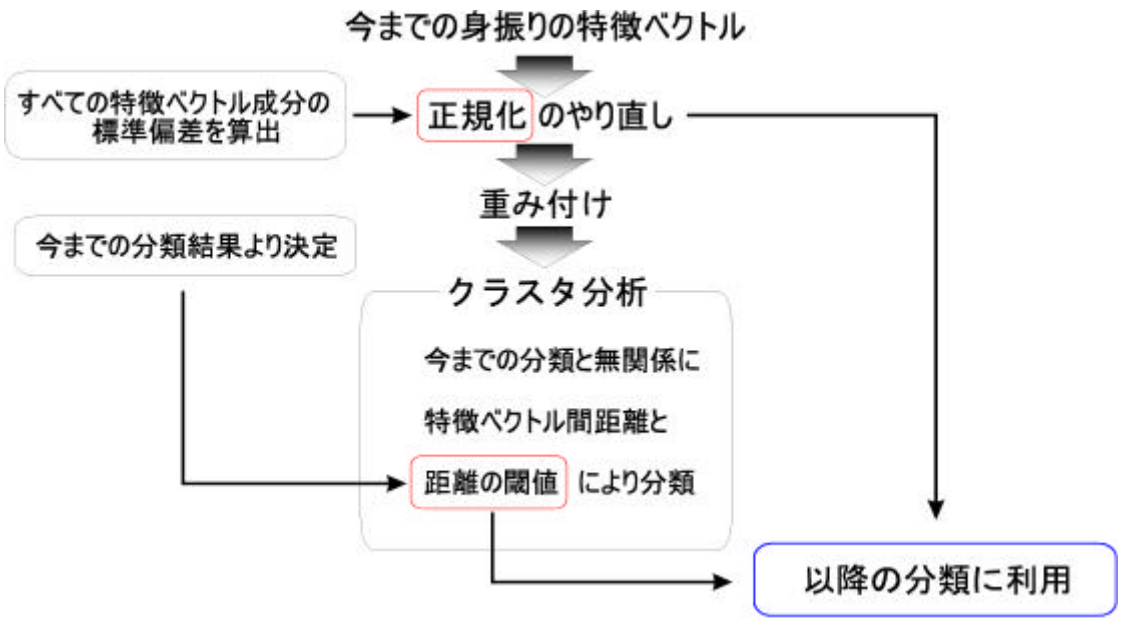

図 3.31: 分類結果の再構成処理の流れ

クラスタ C_i とクラスタ C_j の距離が 2. で算出した閾値 D_{thd} より小さくなっている場合に C_i と C_j を合併する。この操作を繰り返し、特徴ベクトル FV_{wgt} を再分類する。

このように、正規化に用いる標準偏差 ¾l と分類のための閾値 D_{thd} を更新して、分類結果の再構成を行うことにより、過去の身振りを適切に再分類する。

第 4 章 試作システム ReD BACS の構成

第 3 章に述べた身振りの分類手法に基づいてリアルタイム身振り分類システム ReD BACS（Real-time and Dynamic Body Action Classi¯cation System ）を試作した。本章では、そのハードウェア構成、ソフトウェア構成について述べる。

4.1 ハードウェア構成

図 4.1 に本研究室で試作したシステムのハードウェア構成を示す。本システムは以下のハードウェアにより構成されている。

- グラフィックワークステーション O2 R10000/195MHz
  （SGI 社製、以下、O2）
- グラフィックワークステーション OCTANE R12000/400MHz£2
  （SGI 社製、以下、OCTANE）
- CCD カメラ EVI-D30 （ソニー社製）
- デジタルビデオカセットレコーダ WV-D9000 （ソニー社製、以下、VTR）
- ディスプレイ （SGI 社製）

CCD カメラまたは VTR から出力される画像は NTSC 方式のアナログ・ビデオ信号であり、グラフィックワークステーション O2 に内蔵されているビデオシステム MVP（Multiport Video Processor for the O2 system）に入力される。MVP ではその画像をディジタル信号に変換しビデオメモリに蓄えることで O2 で処理できるようにする。そして、O2 で身振りの自動分類処理を行い、その処理過程と結果をディスプレイに出力する。OCTANE はビデオ入力のためのハードウェアを備えていないが、O2 に比べ処理速度が速いため、後述する評価実験での処理時間の計測に用いる。

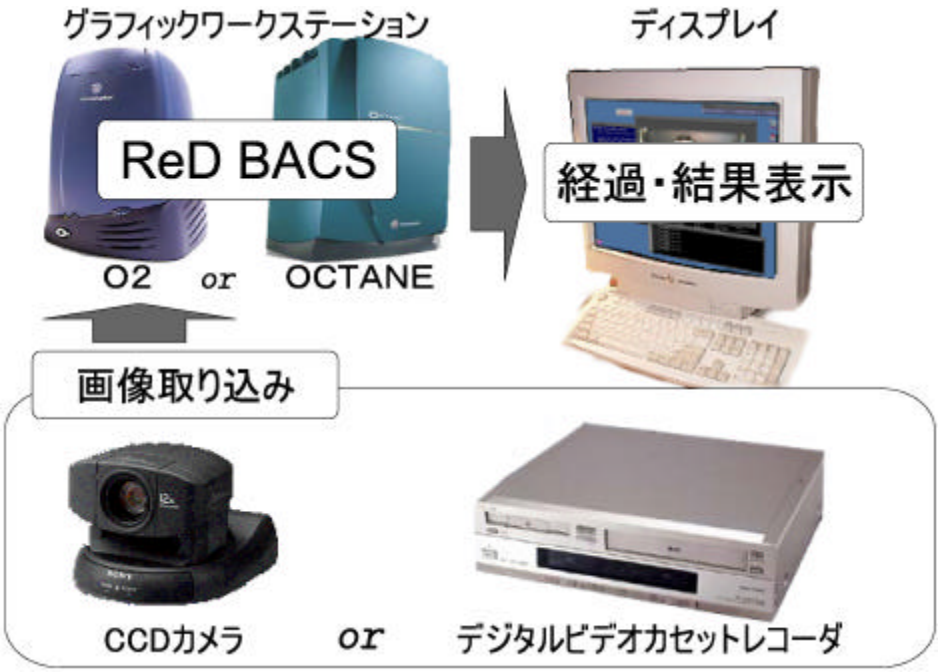

図 4.1: ハードウェア構成

4.2 ソフトウェア構成

試作したシステムは以下のソフトウェアから構成されている。

- SGI ビデオライブラリ（以下、VL）：ビデオデバイス用 C 言語ライブラリ
- Open GL：画像表示用グラフィックスライブラリ
- 本研究で作成した身振り自動分類プログラム

ソフトウェア構成を図 4.2 に示す。ビデオカメラまたは VTR から MVP に入力された画像を VL を用いて R、G、B 各 8 ビットの RGB データに変換し、メモリ上に格納する。その RGB データに対して本研究で作成した身振り自動分類プログラムで特徴抽出、特徴分析、分類の処理を行い、Open GL を用いてその過程と結果をディスプレイに表示する。

身振り自動分類プログラムは、3.1 で述べた身振り分類手法の各部に対応した次の 3 つのサブシステムから構成される。

1. 特徴抽出サブシステム

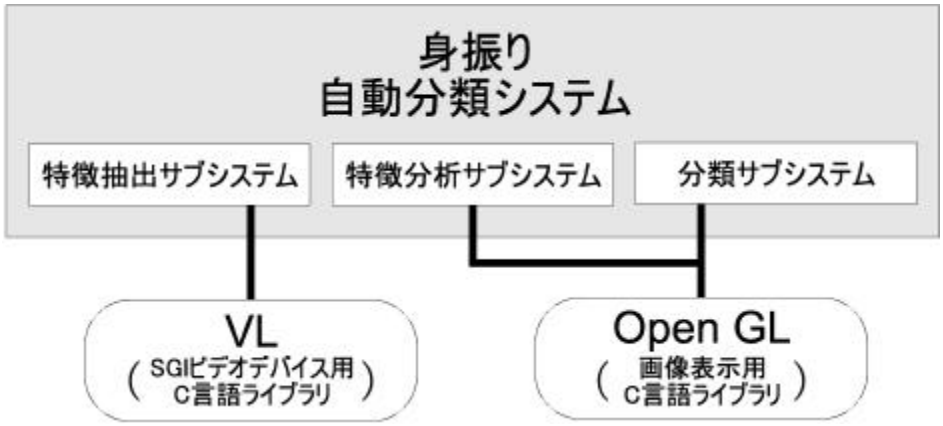

図 4.2: ソフトウェア構成

2. **特徴分析サブシステム**

3. **分類サブシステム**

身振りの自動分類システムの処理の流れを図 4.3 に示す。まず、入力インタフェースで上半身動画像の入力を行う。特徴抽出サブシステムでは上半身画像の RGB データから特徴量を抽出し、特徴分析サブシステムでその特徴量の変化から身振りを分析し、ベクトルで表す。そして分類サブシステムで身振りを分類する。最後に出力インタフェースでは、各サブシステムにおける処理過程の画像や分類結果を出力する。以下の各項では、入出力インタフェースと、自動分類プログラム中の 3 つのサブシステムについて述べる。

4.2.1 入力インタフェース

入力インタフェースでは、画像の取り込みを行う。この部分の実装のため、O2 に付属するビデオシステムを操作する VL インタフェースの作成を行い、入力画像のフォーマットは、コンポジット形式及び、S-Video（YC-358）形式の画像に対応するように作成した。

4.2.2 特徴抽出サブシステム

ここでは、3.2 に述べた特徴抽出手法を実装している。肌色領域抽出部分は、画像処理アルゴリズムにより抽出した、顔、右手、左手の重心点座標・領域面積・アスペクト比の時系列データを特徴分析サブシステムに送る。

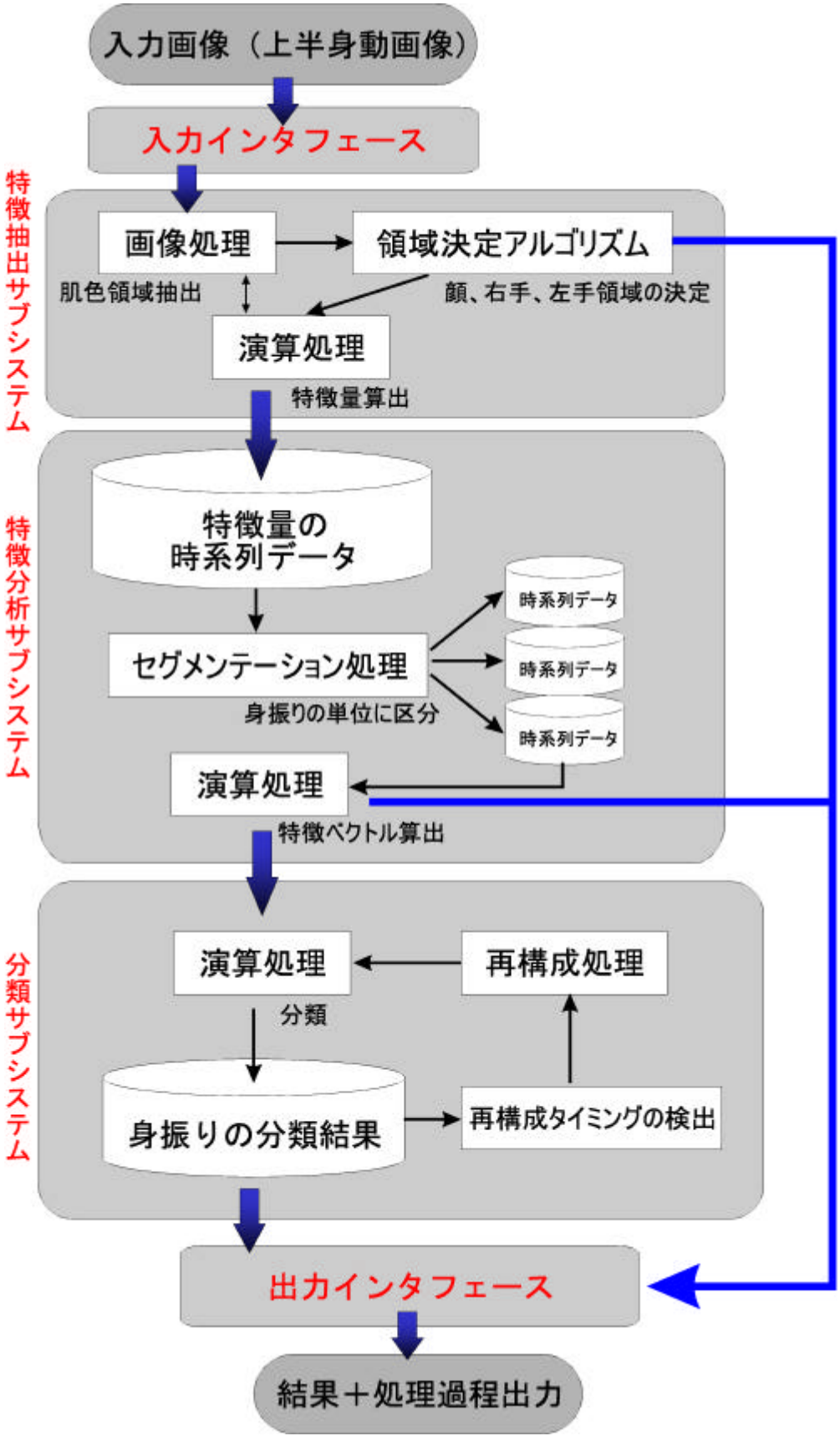

図 4.3: 身振り自動分類システムの処理の流れ

4.2.3　特徴分析サブシステム

ここでは、3.3.3 に述べた特徴分析手法を実装している。各特徴量を用いて様々な演算を行い、特徴量に関する時系列データを別々に保存し、このデータをまとめて、特徴ベクトルとしている。そして、この特徴ベクトルに重みを付け、分類サブシステムに送る。

4.2.4　分類サブシステム

ここでは、3.3.3 に述べた分類手法を実装している。特徴ベクトルを元に、距離演算を行い、その結果に応じてクラスタ分析を行う。再構築処理は、重み付けを変更し、今までのクラスタ分析結果を再分類する。この部分は別プロセスで動作することにより、リアルタイム処理を損なわずに、分類結果を変更することが可能である。

4.2.5　出力インタフェース

出力インタフェースでは、各サブシステムからの処理経過の状況を表示し、分類結果の表示を行う。

この部分の実装には、画像処理の経過を表示するための Open GL インタフェースを作成し、マウス操作により 3.3 に述べた手法による画像処理の過程を表示できるようにしている。また実行状況の表示や分類結果の表示には、X ウィンドウインタフェースを作成し、独立した別のウィンドウにその結果が表示できるようにした。実際に作成した出力インタフェース例を図 4.4 に示す。

図 4.4: 出力インタフェース例

第 5 章 試作システムの評価実験

本章では第 4 章で述べた試作システムの評価実験について述べる。まず、上半身動画像を身振りの単位に区切るセグメンテーション機能の評価実験として、システムによるセグメンテーションと人間による身振りの区分を比較する実験を行った。次に、身振りの分類機能の評価実験として、システムの分類結果と人間の分類結果とを比較する実験を行った。そして、分類の再構成機能については、システムの分類結果を再構成し、再構成前の分類と再構成後の分類とを比較した。

以下では、それぞれの実験についてその目的、方法、結果および考察を述べる。

5.1 動作のセグメンテーション機能評価実験

5.1.1 実験目的

動作のセグメンテーション機能は、対話中の人の上半身動画像からその人の動作を身振りと判断できる単位に区切るものである。ここでは、人間が対話している VTR 画像から、試作システムによるセグメンテーションと、実際の人間が行う区切りとを比較することにより、人間の判断基準と比較してどの程度試作システムが人間の動作を身振りとして抽出することができるかを評価する。ただし、同じ入力情報だけで身振りの区分を比較するため、音声を除去し、画像情報だけを用いる。

また同時に、動作のセグメンテーションの前段階に行われる動作の特徴量の抽出について、領域の重複や隠滅の判定が正しく行われているかどうかを確認する。

5.1.2 実験方法

概要

本実験は、あらかじめ用意した対話中の人の上半身動画像を被験者に提示し、対話している人の動作を被験者が身振りと認識できた時点を指摘してもらう。また、試作したシステムに同じ動画像を入力し、特徴量の抽出、および身振りのセグメンテーションを行う。

用意した対話時の上半身動画像

まず、この実験で用いる対話時の人の上半身動画像を用意するために、対話時の人の上半身を撮影し、ビデオテープに記録した。撮影時の状況を図 5.1 に示す。ここでは、2 名の対話者に向かい合って自由に対話をしてもらい、両方の対話者の上半身をディジタルビデオカメラにより撮影した。撮影に際しては、カメラの設置位置や照明などを次の条件を満たすように整えた。

1. カメラに対して、身体が正面を向いている。

2. 撮影中に身体に向けられる照明が変わらない。

3. 大きな動きをしても、撮影範囲から外れない。

1. は、試作システムに入力される動画像として身体が正面を向いていることを前提にしているためである。2. は、照明が大きく変化する場合には、対象領域の色情報が変化し、正しく特徴量を抽出できないためである。3. は、対話者の上半身がビデオカメラの撮影範囲から外れると特徴量の抽出ができなくなるためである。ただし、過度に動作範囲を制限すると、身振りがぎこちないものになってしまうため、椅子に着座する姿勢を保つ限り、どのような動作でも画面内におさまるようにビデオカメラを設置した。使用したビデオカメラは、ソニー製 VX-1000 とソニー製 TRV900 である。

これらの条件を考慮して、図 5.2 に示すように、相対する 2 名の対話者の正面から約 155cm 離れた位置にそれぞれビデオカメラを設置した。また、対話者間の距離は約 125cm とし、照明環境は図 5.3 に示すように、上方には対話者との位置関係が対称となるように 1 本の蛍光灯を配置し、下前方にそれぞれ 1 つの白熱灯を設置した。なお、この照明環境では、対話者の顔面上の照度は 280［lux］（上方の蛍光灯による照度 200［lux］、下前方の白熱灯による照度 80［lux］）であった。

対話者は、本研究室の男子学生 OT と女性秘書 FM である。対話は 15 分以上で、話題が終了するまで続けてもらった。

2 名の対話者には、事前に表 5.1 に示す指示書を読んでもらい、実験者から実験内容の簡単な説明を行った。

撮影時の時間経過と話題の推移を表 5.2 に示す。なお、表中の経過時間とそのフレームはおおよその目安である。

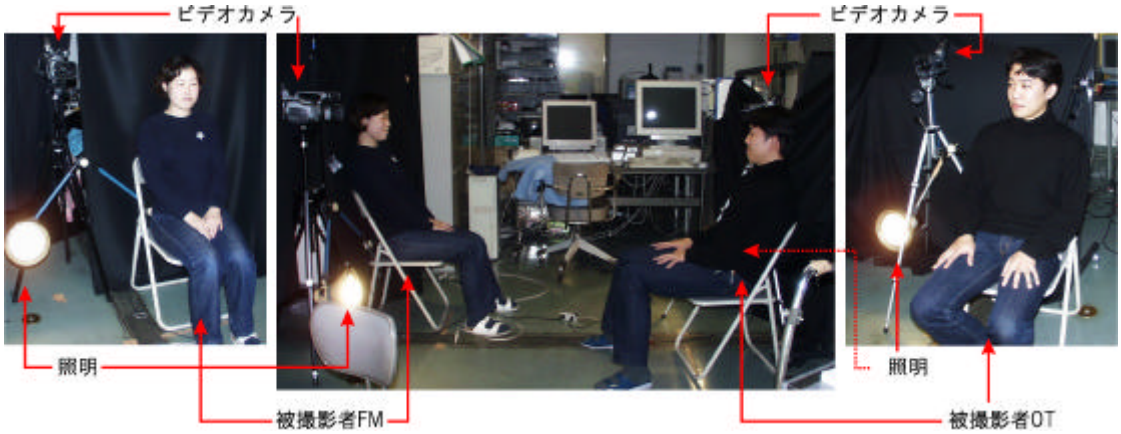

図 5.1: 対話時の上半身動画像記録実験風景

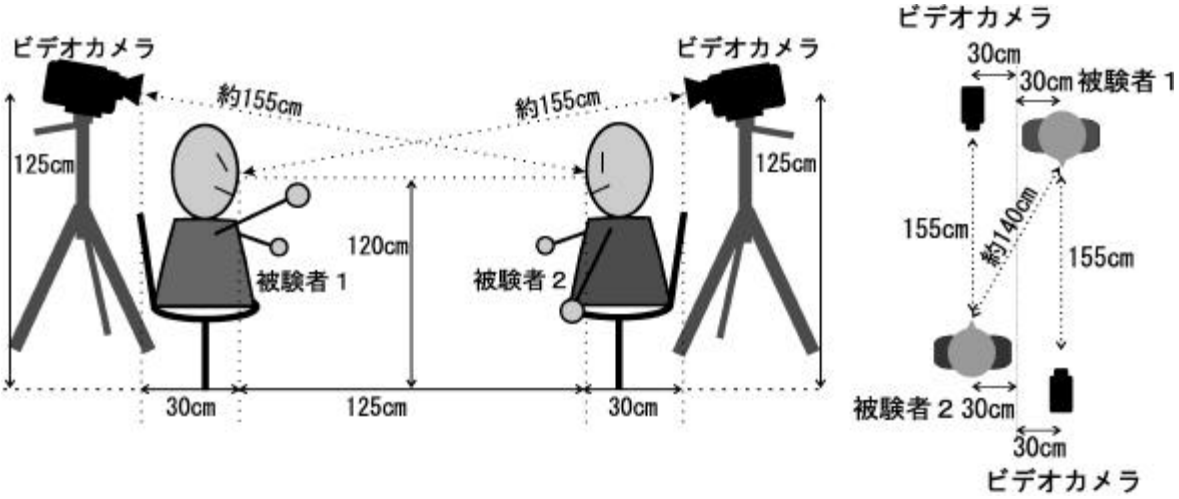

図 5.2: 対話時の上半身動画像の記録実験システム

表 5.1: 対話者に与えた撮影時の指示書

実験目的	本実験では、対話中の自然な行動（身振り） の観察を目的とする。
指示事項	カメラを意識せず相手との対話に集中すること。 対話時間は 15 分以上とし、15 分経過時に合図をするが そのまま対話を続けても良い。
禁止事項	立ち上がらない。物をもたない。 実験器具（ビデオカメラ、照明）に触らない。

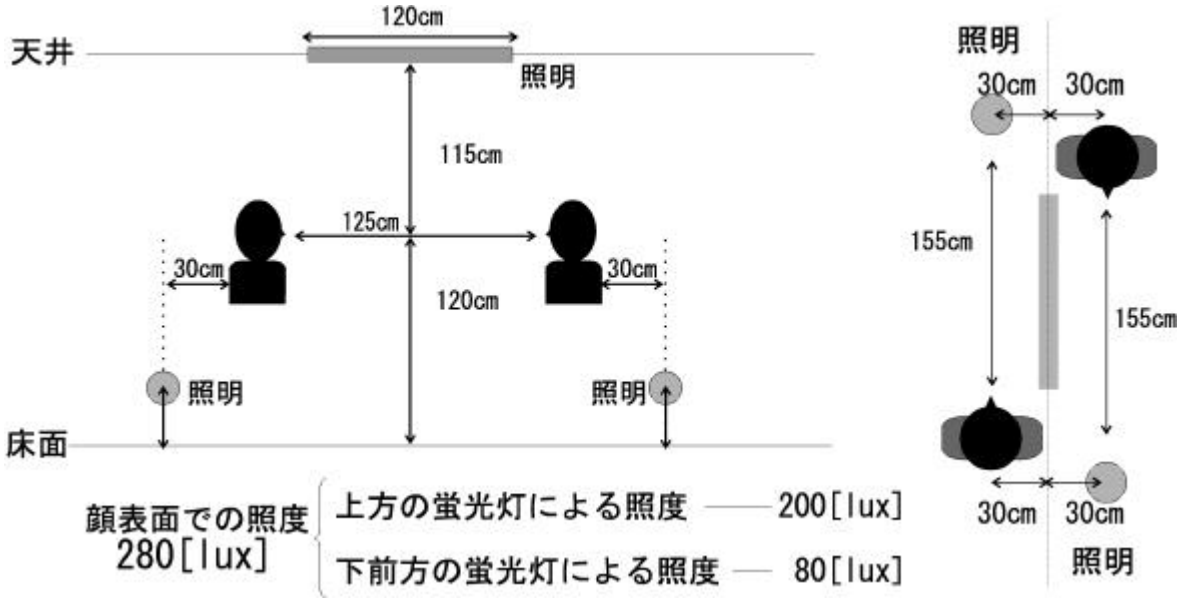

図 5.3: 実験時の照明環境

表 5.2: 時間経過と話題

経過時間	(フレーム)	対話内容
↓	↓	友人夫婦の話題
30 秒	(900)	↓
↓	↓	この実験の話題
2 分 30 秒	(4500)	↓
↓	↓	対話者 OT の個人的な話題
3 分	(5400)	↓
↓	↓	お酒の話題
4 分	(7200)	↓
↓	↓	この実験の話題
4 分 30 秒	(8100)	↓
↓	↓	対話者 FM の個人的な話題
5 分	(9000)	↓
↓	↓	この実験の話題
7 分	(12600)	↓
↓	↓	対話者 FM の個人的な話題
8 分 30 秒	(15300)	↓
↓	↓	学会に関する話題
9 分	(16200)	↓
↓	↓	電話に関する話題
9 分 30 秒	(17100)	↓
↓	↓	友人の話題
10 分	(18000)	↓
↓	↓	アルバイトに関する話題
12 分	(21600)	↓
↓	↓	収入に関する話題
13 分	(23400)	↓
↓	↓	後輩学生に関する話題
15 分	(27000)	↓

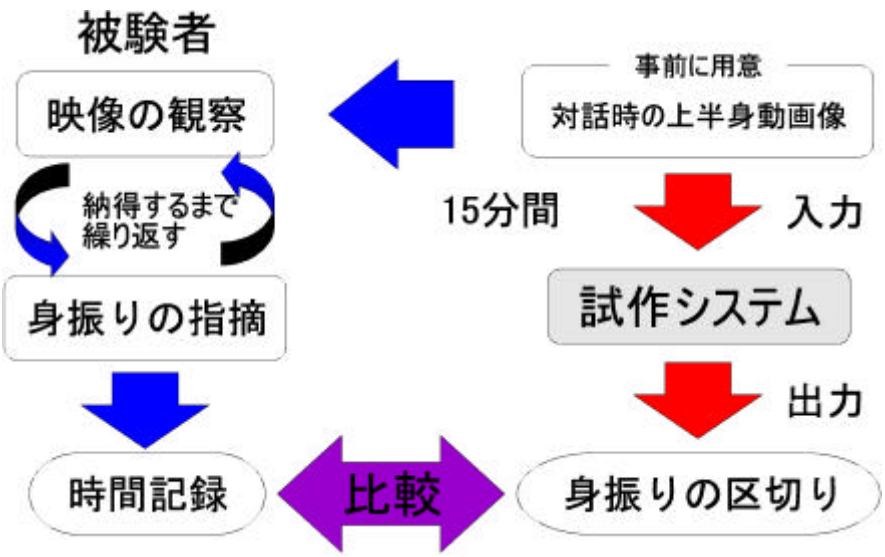

図 5.4: セグメンテーション機能評価実験の流れ

実験手順

実験時の手順を図 5.4 に示す。事前に撮影した対話時の身振りの上半身動画像をビデオデッキで再生し、被験者にその画像をテレビモニタで見てもらう。ここで用いる画像はすべての被験者で同じものとし、提示する映像の時間は 1 つの動画像あたり 15 分間である。ただし、事前に実験に用いるものとは別の 1 分間の上半身動画像を提示し、実験の手順を説明する。

実験は図 5.5 に示すように、被験者にテレビモニタの前に着座してもらい、ビデオテープに記録されている映像を見てもらう。その際、被験者には身振りと認識した動作を指摘してもらい、ビデオデッキのテープカウンタを参照して、その時刻を記録用紙に記入してもらう。ただし、被験者にはビデオデッキを自由に操作し、繰り返し画像を確認してもらう。

また、本実験でも第 3 章に述べたセグメンテーション実験と同様に、身振りの開始と終了の時点を正確に指摘する形式ではなく、身振りであると認識した時点を指摘してもらう。また、被験者への指示として、「人の癖や無意識の動作も身振りとする」ことを教示する。

また、これとは別に被験者に提示する上半身動画像と同じものを試作システムに入力し身振りを抽出する。なお、実験で使用する映像は、対話者 OT のもの（映像 1）と対話者 FM のもの（映像 2）のそれぞれ 15 分間の上半身動画像である。

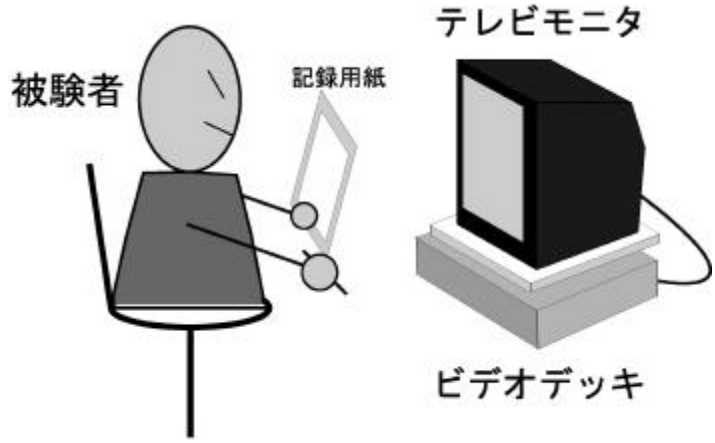

図 5.5: セグメンテーション機能評価実験の実験システム

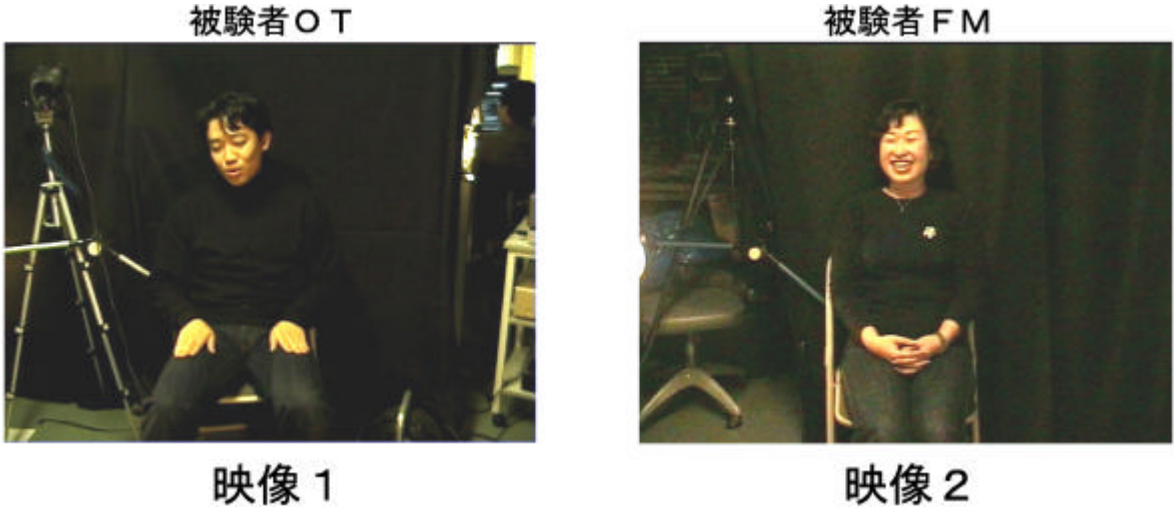

図 5.6: 上半身動画像の例

被験者

被験者は男子学生 SK、OY、MT の 3 名である。

5.1.3 実験の結果と考察

実験に用いた上半身動画像の例を図 5.6 に示す。

また、この実験の結果を表 5.3、表 5.4 に示す。表中では、実験で用意した動画像のフレーム番号を左列に示し、その時点でシステム、および各被験者が身振りと区切った場合を○で示した。映像のフレームレートは、30frame/sec である。また、この結果

表 5.3: セグメンテーション評価実験結果（映像 1）

フレーム	システム	被験者SK	被験者OY	被験者MT
182	○			○
650	○			○
723	○			
1050	○			
1793	○	○	○	○
1866	○			
2125		○	○	○
2792	○			○
2821	○	○	○	
2909	○			
3515	○	○	○	○
3938				○
4178				○
4456		○	○	○
4535	○	○	○	○
4590	○	○	○	○
4657			○	○
4763	○			
4818	○	○	○	○
4866		○	○	
5006	○	○	○	○
5053	○		○	○
5185	○	○	○	○
5408		○		○
5512	○	○		
5554	○			
5664	○		○	○
7150	○	○	○	○
7654	○	○	○	○
8004	○	○	○	○
8057	○	○	○	○
8254	○	○		
8485	○		○	○
8524	○	○		
8656	○		○	○
8707	○	○	○	○
8837	○			○
8904	○	○	○	○
9018	○		○	○
9047	○	○	○	○
9171	○	○	○	○
9277	○			
9403	○	○	○	○
9810			○	○
9829	○	○	○	○
10105	○	○		○
10388		○	○	○
10719			○	
11121	○	○	○	○
11605	○	○	○	○

フレーム	システム	被験者SK	被験者OY	被験者MT
12616	○	○	○	○
12661	○			○
14033				○
14803	○	○	○	○
15005	○	○	○	○
15370	○	○	○	○
15425		○	○	○
15770				○
15946	○		○	○
16000	○	○		
16450		○		○
16628	○	○	○	○
16988	○	○	○	○
17071	○	○	○	○
17158	○	○	○	○
17236	○			
17271	○	○	○	○
17591	○	○	○	○
17730	○	○	○	○
17927	○			○
18344	○	○	○	○
18587		○		
18739	○	○		○
18763	○			○
19030		○		○
19995				○
20367	○	○	○	○
21040	○	○	○	○
21113	○			○
21972		○	○	
22136	○	○	○	○
22211	○	○	○	○
22833		○		○
23230				○
23380				○
23577	○	○		○
23604	○			
23669	○	○		
23779	○	○		○
23871	○			
23949	○			
24144	○	○		○
24393				○
25311	○	○		○
26136				○
26459		○		
26563	○			○
26829	○	○	○	○
26978	○	○	○	○
27004		○		

をまとめたものを表 5.5、表 5.6 に示す。

表 5.4: セグメンテーション評価実験結果（映像 2）

フレーム	システム	被験者SK	被験者OY	被験者MT
440	○	○	○	○
508	○	○		
882	○			
1268	○	○	○	○
1320	○			
1766	○			
2194				○
2342	○			
2413	○	○	○	○
3274				○
3582	○			○
4082	○	○	○	○
4143	○	○	○	○
4223	○		○	○
4289	○	○		
4559	○		○	○
4607		○		
4670	○	○	○	○
5184	○	○	○	○
5446	○			
5597	○	○	○	○
6749	○	○	○	○
6875	○	○	○	○
7000	○		○	
7196		○		
7613	○	○	○	○
7712	○	○	○	
7738	○	○	○	○
7761	○			
8371	○	○	○	○
8403	○			
8438	○			
8487	○		○	
8549	○	○	○	
8604			○	
8681	○	○	○	○
8717	○	○		
8880	○	○	○	○
8944	○	○	○	
8996	○			
9549	○	○	○	○
9706	○	○	○	○
9771	○	○	○	
9798	○			
9912	○			
9978	○	○	○	
10222	○	○	○	○
10476	○			
10531	○	○	○	
10586	○	○	○	○
10621	○	○		
10653	○			
10735	○	○		
10802	○			○
10831	○	○	○	
10895	○			
10924	○			
11118	○	○	○	○
11264	○	○	○	○
11390	○			

フレーム	システム	被験者SK	被験者OY	被験者MT
11616	○	○		
11655	○			
12133	○	○	○	○
12185	○			
12622	○	○		
12800	○			
12922	○	○	○	○
12948	○		○	
13024	○			
13078	○	○	○	○
13298	○	○	○	○
13349	○			
13524	○	○	○	○
13661	○		○	○
13696	○	○		
14662	○	○	○	○
14962	○	○	○	○
15109	○	○	○	○
15163		○	○	
15454	○			
15541	○			
15592	○			
16461	○			
16575	○	○	○	○
16702	○			
16898	○			
16923	○			
17149	○	○	○	○
17270	○			
17813	○	○	○	○
17828			○	
17945	○	○	○	○
18041	○	○	○	○
18436	○	○	○	○
18464	○			
19620	○	○	○	○
19669	○			
19914	○	○	○	○
21255	○	○	○	○
22882	○	○	○	○
24997	○	○	○	○
25055	○			
25173	○			
25320	○	○	○	○
25405	○	○	○	○
25443	○	○		
25584	○			
25631	○			
25854			○	○
25893	○	○	○	
25935	○	○		○
25967	○			
26420	○	○	○	○
26543	○	○	○	
26597	○	○	○	
26700	○			
26736	○			
26764	○	○	○	○
26874	○	○	○	○
26973	○	○	○	○

表 5.5: セグメンテーション評価実験結果合計（映像 1）

被験者	人による区分数	システムによる区分数	人は身振りと判断するがシステムが区分しない数	人は身振りと判断しないがシステムが区分する数
SK	62	75	12	26
OY	53	75	9	27
MT	77	75	18	16

表 5.6: セグメンテーション評価実験結果合計（映像 2）

被験者	人による区分数	システムによる区分数	人は身振りと判断するがシステムが区分しない数	人は身振りと判断しないがシステムが区分する数
SK	69	112	3	46
OY	66	112	2	50
MT	55	112	1	60

表 5.3 と表 5.4 から、3 人の被験者によって身振りと区分する動作に若干ばらつきがあることがわかる。また、表 5.5 と表 5.6 から、映像 1 ではシステムによる区分数（75）と 3 人の被験者 SK、OY、MT による区分数（62、53、77）がほぼ同じであるが、映像 2 ではシステムによる区分数（112）は 3 人の被験者による区分数（69、66、55）より多いことがわかる。映像 1 では、人間が区分してもシステムが区分しないもの（12、9、18）、あるいはシステムが区分しても人間が区分しないもの（26、27、16）が多かった。一方、映像 2 では、人間が区分したものをシステムが区分しないもの（3、2、1）は、ほとんどないが、システムが区分して人間が区分しないもの（46、50、60）は非常に多かった。

ここで、システムが適切に身振りを区分していないものは、被験者全員が身振りと区分しているが、システムは区分していないものとする。このような身振りは、映像 1 の場合で 4 つの動作だけであり、映像 2 の場合にはなかった。これより、システムは全被験者が 身振りと判断する動作をほぼ適切に抽出していることがわかる。

逆に、システムが身振りとして区分しているが、被験者の一人も身振りとして区分していないものもあった。このような身振りは、映像 1 の場合で 8、映像 2 の場合で 34

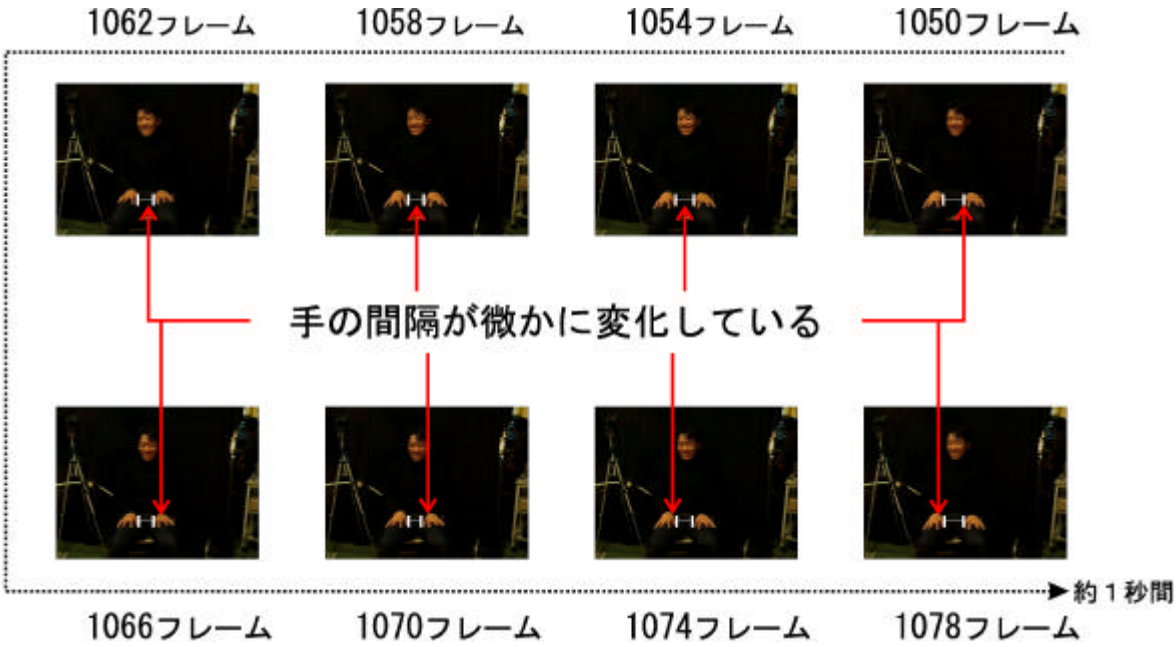

図 5.7: 小さな動作の例

であった。これらの動作をビデオテープで確認すると、そのほとんどは短時間、手を少しだけ動かす個人の癖のような動作であり、暗示的な情報を含む適応子に対応するものである。この一例を図 5.7 に示す。また、特に、システムの区分数が被験者による区分数より多い映像 2 については、「うなずき」などの頭を動かす動作が多くみられた。つまり、被験者が身振りとは判断しない小さな動作でも、システムは抽出することができた。このことは、この手法が被験者が見逃すような微かな暗示的情報を含む身振りも検出できることを示している。

さらに、システムが動作の特徴量を抽出する際に、領域の重複や隠滅の判定が正しく行われているかどうかを確認するため、用意した上半身動画像の中から、顔や手が互いに重なっているフレームを探し出し、システムがそのフレームの領域をどのように判定したかを調べた。図 5.8 にその結果を示す。これより複数の領域が重なった場合や手が身体の後ろに入って隠れた場合でも、正しく領域が特定できていることがわかる。

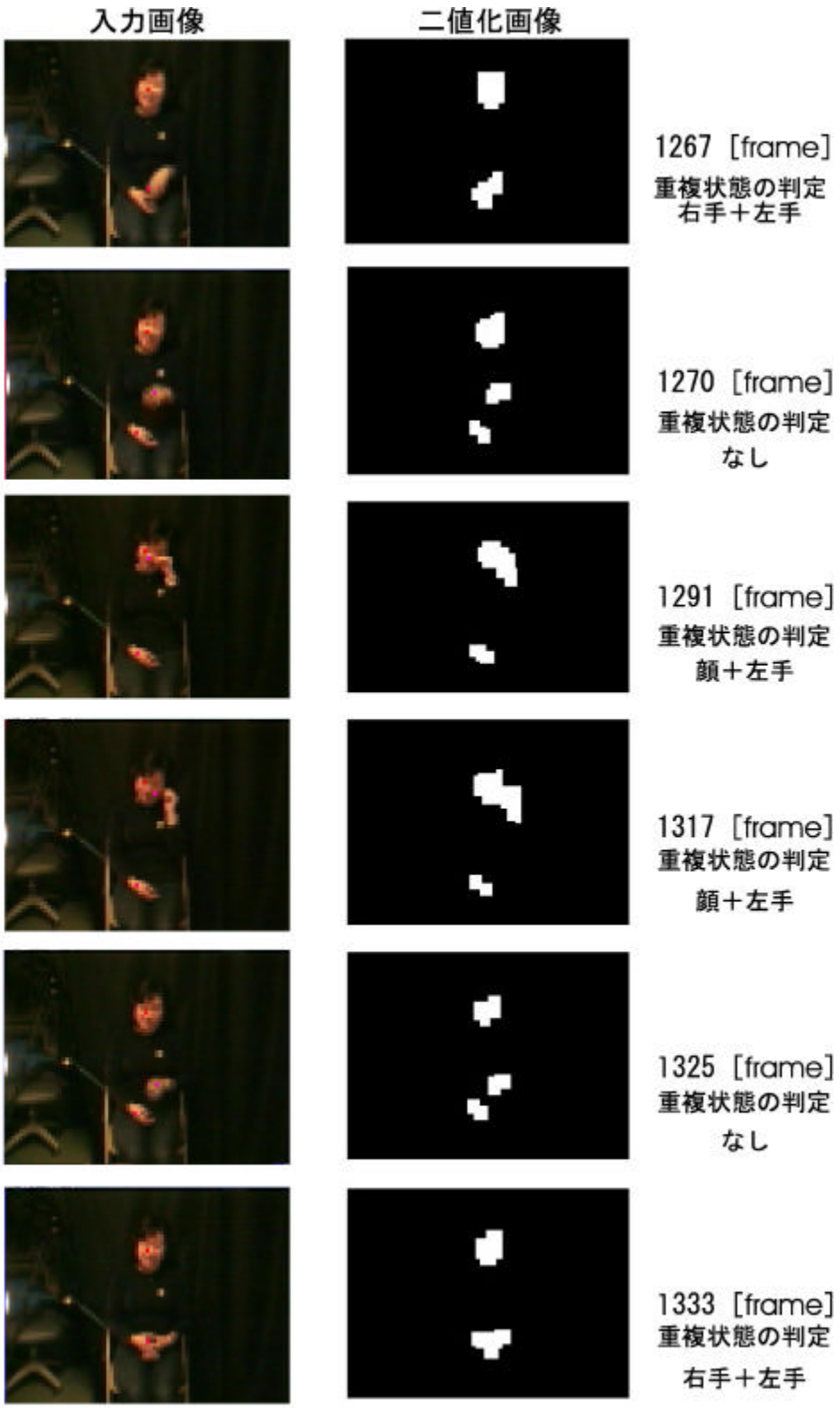
入力画像
二値化画像
1267 [frame]
重複状態の判定
右手＋左手
1270 [frame]
重複状態の判定
なし
1291 [frame]
重複状態の判定
顔＋左手
1317 [frame]
重複状態の判定
顔＋左手
1325 [frame]
重複状態の判定
なし
1333 [frame]
重複状態の判定
右手＋左手

図 5.8: 重複時の例

5.2　身振りの分類機能評価実験

5.2.1　実験目的

身振りの分類機能は、入力された上半身動画像から身振りと定義した単位を類似なものに分類するものである。ここでは、前節のセグメンテーション機能の評価実験で使用した対話時の上半身動画像を用いて、試作システムによる身振りの分類結果と、人間による分類結果とを比較し、どの程度試作システムが身振りを分類することができるかを評価する。ただし、同じ入力情報に対する身振りの分類を比較するため、音声を除去し、画像情報だけを提示する。

また、分類処理がリアルタイムに行われるかどうかを調べるため、4.2 に述べた身振りの自動分類プログラムの各サブシステムの処理時間を計測する。

5.2.2　実験方法

概要

本実験は、あらかじめ用意した対話中の人の上半身動画像を被験者に見せ、被験者が身振りと認識した動作を分類してもらう。上半身動画像は前節で述べたセグメンテーション機能評価実験と同じものを用いる。また、被験者も前節の実験と同じとし、身振りの区分はセグメンテーション評価実験でそれぞれの被験者が区分したものを用いる。また、試作したシステムに同じ動画像を入力し、システムが行った分類結果を調べるとともに、特徴抽出サブシステム、特徴分析サブシステム、特徴分類サブシステムにおける処理時間を計測する。

実験手順

実験時の手順を図 5.9 に示す。前節の実験で用いた身振りの上半身動画像をビデオデッキで再生し、被験者にその画像をテレビモニタで見てもらう。実験で使用する映像は、前節の実験と同じもの、すなわち、対話者 OT のもの（映像 1）と対話者 FM のもの（映像 2）でそれぞれ 15 分間の上半身動画像である。

実験の前には、実験に用いるものとは別の 1 分間の上半身動画像を提示し、実験の手順を説明する。

被験者には、前節の実験でその被験者が動作を身振りと判断した時刻を記入してもらった記録用紙を提示する。実験では、その記録用紙と上半身動画像から、前節の実

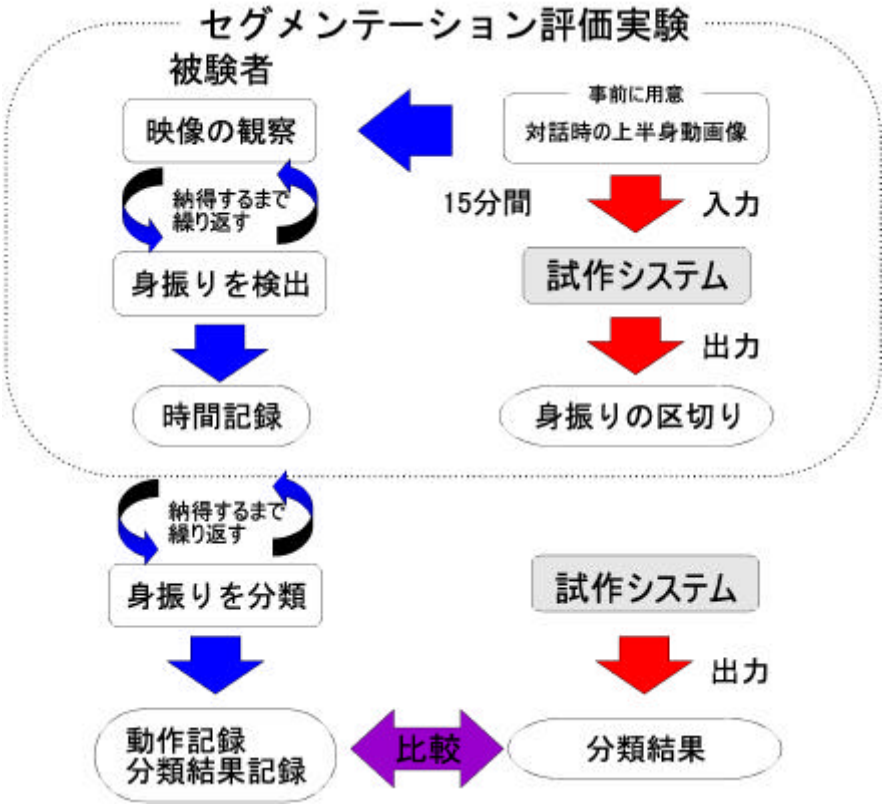

図 5.9: 身振りの分類機能評価実験の流れ

験で区切った身振りを分類してもらう。その際、被験者にはビデオデッキを自由に操作して、繰り返し映像を確認してもらう。

被験者には、(1) 個々の身振りの動作内容、(2) その身振りの分類、(3) 分類した身振りを代表する動作内容を別の記録用紙に記入してもらう。被験者へは、「動作が同じであればすべて同じ分類の身振りとする」と指示する。

また、これとは別に被験者に提示する上半身動画像を試作システムに入力し、身振りを分類する。なお、システムには、1 フレームあたりの各サブシステムの処理時間を計測するため、あらかじめ時間計測プログラムを組み込んでおく。

被験者

前節の実験と同じ被験者 SK、OY、MT の 3 名である。

5.2.3　実験の結果と考察

この実験の結果を表 5.7、表 5.8 に示す。表では、実験で用意した動画像のフレーム番号を左列に示し、その右側には、その時点の身振りをシステムが分類した結果、および 3 名の被験者が分類した結果を示す。表中のアルファベットで示した被験者の分類は、それぞれ各被験者ごとの分類であり、アルファベットが同じでも被験者間で関連はない。また、\XX" と示したものは、被験者がその身振りをどこにも分類できなかったことを示す。この実験で、システム、および被験者が区分した身振り数と分類した動作の種類（クラスタ）数を表 5.9、表 5.10 に示す。また、システムに映像 2 を入力したときの 1 フレームあたりの処理時間を表 5.11 に示す。表に示した処理時間は、映像 2 の全フレームの処理時間の平均である。OCTANE は、画像取り込み用のハードウェアを持たないため、O2 で取り込んだ画像を一旦ハードディスクのファイルとして保存し、それを OCTANE で処理したときの処理時間を計測した。

まず、分類の結果について考察する。

表 5.9 と表 5.10 から、映像 1 では、システムが分類したクラスタ数（19）は、被験者のクラスタ数（19、20、25）とほぼ同じか若干少ないことがわかる。一方、映像 2 では、システムが分類したクラスタ数（46）は、被験者のクラスタ数（22、36、22）よりも多かった。前節の考察で述べたように、映像 2 では、システムが対話者の癖のような小さな動作までもを身振りとして抽出しているため、分類対象となる身振り数が多くなり、クラスタ数も多くなったものと思われる。

また、表 5.7 および表 5.8 から、被験者間で、同じように分類された身振りもあるが、別のクラスタに分類されたものもあり、各被験者ごとに分類のばらつきがみられることがわかる。実験後の被験者の報告によると、「映像 1 および映像 2 の各 15 分間の映像を見て動作を分類するのにそれぞれ約 2 時間もの時間がかかり、その間の分類基準が必ず一定ではなく、分類開始直後と終了直前では異なっていた」となっている。これらの分類結果と報告から、実験で課した身振りの分類作業は被験者にとって非常に難しいものであることがわかる。

ここで、分類に関する実験結果を詳細に検討するため、被験者 MT の分類結果に着目する。まず、表 5.12 および表 5.13 に、被験者 MT の分類による各クラスタについて、そのクラスタに分類した身振り数と、そのクラスタの動作を代表する動作内容を示す。被験者 SK、OY の結果は付録 B に譲る。被験者 MT の分類では、例えば、表 5.12 に示すように、「右手で相手を指す」クラスタ F の動作と「右手で相手を指で指す」クラスタ K の動作が別の身振りとして分類されている。しかし、これらをビデオテープで確

認すると、クラスタＦに分類されている動作にも、「指で指す」という動作が確認できた。前述のように、人間が意識的に動作を分類する場合、その判断基準に揺らぎが生じ、同じ動作でも異なる動作に分類することがあることがわかる。

次に、システムの分類結果と被験者 MT の分類結果を比較する。表 5.14 および表 5.15 にシステムの分類結果を元にした被験者 MT の分類結果と個々の身振りの動作内容（以下、動作メモと呼ぶ）を示す。表 5.14 中で被験者 MT が「すりすり」と記した動作メモの動作は、「両手を膝の上で擦る」動作である。システム分類ごとに被験者 MT の動作メモを見ると、各分類で被験者の動作メモは、ほぼ同じ動作内容であった。

ここで、システムが適切な分類を行っておらず、分類に失敗したものを、被験者の動作メモの「動作部位が異なる」場合と考え、被験者 MT の実験結果を調べた。すると、映像 1 では、システムが分類したクラスタ番号（1）に関しては 11 の身振りのうち 9 が両手に関係する動作であり、（2）に関しては 18 の身振りのうち 13 が右手、（3）では 16 の身振りのうち 13 が左手と、失敗しているものは少なかった。映像 2 では、システムが分類したクラスタ番号（4）では 4 の身振りのうち 3 が右手、（5）では 5 の身振りのうち 5 つとも右手であるとしており、失敗しているものは少なかった。これらにより、動作部位が異なるものはほぼ分類できていることがわかった。

また、表 5.15 の動作メモを見ると「複雑でわからない」としている動作がいくつかあることがわかる。しかし、このような動作に関してもシステムは何らかの分類を行っており、このような動作を分類する機能としては、人間よりも優れている可能性がある。この分類結果の例を図 5.10、図 5.11 に示す。

次に処理速度について考察する。システムに入力される動画像のフレームレートは 30frame/sec である。そのため、すべてのフレームに対して遅滞なく処理を行うためには、毎フレームの処理時間が 33.3msec 以下である必要がある。表 5.11 に示すように、全体の処理時間は、O2 で 26.62msec、OCTANE で 13.30msec（画像取り込み時間を除く）であり、各フレームを遅滞なく処理できることを確認した。サブシステムごとに処理時間をみると、特徴抽出サブシステムの処理時間が最も長く、全体の処理時間の 95%以上を占める。これは、特徴抽出サブシステムが 2 次元の画像データそのものを扱っているため、大量のデータを処理する必要があり、処理時間がかかるものと思われる。一方、特徴分析サブシステムと分類サブシステムでは、動作の特徴量や特徴ベクトル等のデータを処理するため、ほとんど処理時間はかからない。

表 5.7: 分類評価実験結果（映像 1）

フレーム	システム	被験者SK	被験者OY	被験者MT
182	1			A
650	1			A
723	1			
1050	1			
1793	2	A	A	B
1866	2			
2125		B	B	C
2792	1			D
2821	3	C	C	
2909	3			
3515	2	B	B	C
3938				A
4178				E
4456		D	D	F
4535	2	E	D	F
4590	3	B	E	G
4657			B	H
4763	2			
4818	3	E	F	I
4866		E	F	
5006	2	D	D	F
5053	2		D	F
5185	3	E	H	I
5408		F		A
5512	4	G		
5554	5			
5664	6		I	J
7150	2	H	G	K
7654	3	I	J	L
8004	3	I	J	L
8057	6	J	XX	XX
8254	7	J		
8485	3		J	L
8524	3	K		
8656	3		K	XX
8707	8	L	L	XX
8837	1			E
8904	2	M	L	M
9018	2		D	N
9047	2	A	D	N
9171	9	C	M	XX
9277	1			
9403	3	E	H	O
9810			G	N
9829	10	D	XX	J
10105	1	N		E
10388		J	N	XX
10719			N	
11121	3	E	H	O
11605	3	E	F	P

フレーム	システム	被験者SK	被験者OY	被験者MT
12616	2	G	O	Q
12661	1			A
14033				A
14803	3	I	J	L
15005	11	G	O	Q
15370	2	A	A	N
15425		A	A	N
15770				A
15946	12		XX	XX
16000	2	G		
16450		F		A
16628	2	D	D	F
16988	3	J	M	XX
17071	13	H	P	R
17158	14	J	Q	S
17236	15			
17271	1	O	N	T
17591	2	A	R	XX
17730	2	M	R	B
17927	1			A
18344	2	D	D	F
18587		P		
18739	1	P		A
18763	16			A
19030		P		A
19995				A
20367	10	Q	R	U
21040	3	J	S	V
21113	3			W
21972		B	T	
22136	17	G	I	J
22211	2	G	R	J
22833		P		A
23230				X
23380				A
23577	1	R		A
23604	16			
23669	18	P		
23779	2	S		XX
23871	2			
23949	1			
24144	1	P		A
24393				A
25311	19	P		A
26136				A
26459		P		
26563	2			A
26829	3	E	H	G
26978	3	E	J	G
27004		E		

表 5.8: 分類評価実験結果（映像 2）

フレーム	システム	被験者SK	被験者OY	被験者MT
440	1	A	A	A
508	2	B		
882	2			
1268	2	C	B	B
1320	2			
1766	2			
2194				C
2342	2			
2413	3	A	C	A
3274				C
3582	2			C
4082	4	D	D	D
4143	5	D	D	D
4223	6		B	B
4289	6	C		
4559	5		E	E
4607		E		
4670	7	E	F	F
5184	6	F	G	G
5446	2			
5597	8	G	H	B
6749	4	H	I	A
6875	9	I	J	XX
7000	7		K	
7196		J		
7613	5	K	F	I
7712	7	H	F	
7738	10	L	L	XX
7761	11			
8371	12	H	I	A
8403	13			
8438	2			
8487	2		B	
8549	14	M	M	
8604			I	
8681	15	H	K	J
8717	16	J		
8880	2	N	N	B
8944	2	N	O	
8996	2			
9549	7	A	P	F
9706	3	H	I	XX
9771	17	O	AD	
9798	12			
9912	18	M		
9978	19		C	
10222	20	P	XX	XX
10476	21			
10531	18	H	I	
10586	22	Q	Q	XX
10621	23	R		
10653	24			
10735	25	S		
10802	5			K
10831	26	K	F	
10895	26			
10924	26			
11118	5	D	R	K
11264	27	C	B	B
11390	5			

フレーム	システム	被験者SK	被験者OY	被験者MT
11616	28	S		
11655	28			
12133	29	T	XX	XX
12185	30			
12622	31	S		
12800	26			
12922	31	P	S	XX
12948	32		T	
13024	7			
13078	7	O	U	L
13298	33	O	V	M
13349	12			
13524	12	E	W	M
13661	21		E	N
13696	34	W		
14662	26	D	R	O
14962	35	T	X	P
15109	7	K	Y	K
15163		U	Z	
15454	26			
15541	26			
15592	26			
16461	26			
16575	26	V	AA	Q
16702	36			
16898	26			
16923	26			
17149	37	N	AB	B
17270	26			
17813	4	D	AC	XX
17828			D	
17945	7	K	F	I
18041	33	W	AD	R
18436	38	T	XX	S
18464	39			
19620	40	T	XX	XX
19669	41			
19914	40	E	Q	F
21255	41	T	XX	XX
22882	4	D	Q	XX
24997	7	X	A	A
25055	42			
25173	2			
25320	26	D	D	D
25405	40	D	D	D
25443	43	L		
25584	2			
25631	2			
25854			B	B
25893	8	C	AE	
25935	44	D		XX
25967	40			
26420	6	G	H	XX
26543	10	L	L	
26597	12	O	AF	
26700	12			
26736	45			
26764	46	Y	AG	T
26874	28	Z	AH	U
26973	26	L	L	V

表 5.9: 分類評価実験結果（映像 1）

		被験者		
	システム	SK	OY	MT
身振り区分数	75	62	53	77
クラスタ数	19	19	20	25

表 5.10: 分類評価実験結果（映像 2）

		被験者		
	システム	SK	OY	MT
身振り区分数	112	69	66	55
クラスタ数	46	22	36	22

表 5.11: 1 フレームあたりのシステムの処理時間［msec］

画像処理ワークステーション	O2	OCTANE
画像取り込み（HW 処理）	0.35	—
特徴抽出サブシステム	25.99	13.20
特徴分析サブシステム	0.08	0.02
分類サブシステム (*1)	0.20	0.08
処理時間合計	26.62	13.30(*2)

(*1) 分類の再構成処理を除く

(*2) 画像取り込み処理を除く

表 5.12: 被験者による分類（映像 1：被験者 MT）

クラスタ	分類数	分類した動作
A	21	両手を膝の上ですりすり
B	2	右手を肩の高さまで挙げる
C	2	両手を下に向けて20cm上に
D	1	左手で耳を掻く
E	3	両手を膝の上で合わせる
F	6	右手で相手を指す
G	3	左手を振る
H	1	両手を20cm挙げて置くようなしぐさ
I	2	左手を振り降ろす
J	4	両手でそれはこっち（左）においといてのしぐさ
K	1	右手で相手を指で指す
L	4	左手で相手を指す
M	1	右手を振る
N	5	左手を挙げる
O	2	右手をちょっと挙げる
P	1	左手を挙げて手首を隠す
Q	2	両手でそれはこっち（右）においといてのしぐさ
R	1	右手で下を指さし
S	1	左手で涙を拭う
T	1	左手で首に触る
U	1	右手を右から左に動かす
V	1	左手を後ろを指さす
W	1	両手を上に向ける
X	1	指をわらわら動かす
XX	9	複雑でわからない
	77	

表 5.13: 被験者による分類（映像 2：被験者 MT）

クラスタ	分類数	分類した動作
A	5	右手を顔のあたりにやる
B	7	左手を顔のあたりにやる
C	3	両手を膝の間に
D	4	右手を相手を指さす
E	1	右手を自分を指す
F	3	右手を相手に向かって振り下ろす
G	1	両手で何かをもつしぐさ
I	2	右手を相手を指す
J	1	右手を振る
K	3	右手を少し挙げる
L	1	両手を少し挙げる
M	2	両手を顔のあたりにやる
N	1	右手を胸にやる
O	1	右手で上を指す
P	1	両手を右から左に
Q	1	右手で下を指す
R	1	両手で自分を抱きしめる
S	1	両手でおなかをさする
T	1	左手で右側を指す
U	1	左手で足をたたく
V	1	両手をたたく
XX	12	複雑でわからない
	54	

表 5.14: システムによる分類結果（映像 1：被験者 MT）

システム	フレーム	被験者MT	実験時の動作メモ
1	182	A	すりすり
	650	A	両手を足の間に
	723		
	1050		
	2792	D	頭を掻く
	8837	E	両手を足の間に
	9277		
	10105	E	両手を足の間に
	12661	A	すりすり
	17271	T	袖を直す
	17927	A	すりすり
	18739	A	すりすり
	23577	A	すりすり
	23949		
	24144	A	すりすり
2	1793	B	右手を挙げる
	1866		
	3515	C	両手を少し挙げる
	4535	F	右手で相手を指す
	4763		
	5006	F	右手で相手を指す
	5053	F	右手で相手を指す
	7150	K	右手で相手を指差す
	8904	M	右手を振る
	9018	N	右手で相手を指す
	9047	N	右手を挙げる
	12616	Q	両手で何かを持って右に
	15370	N	右手を挙げる
	16000		
	16628	F	右手で相手を指す
	17591	XX	右手で電話するしぐさ
	17730	B	右手で電話するしぐさ
	18344	F	右手で相手を指す
	22211	J	それはこっち(左)に置いといて
	23779	XX	ごちゃごちゃ
	23871		
	26563	A	すりすり

システム	フレーム	被験者MT	実験時の動作メモ
3	2821		
	2909		
	4590	G	左手を左右に振る
	4818	I	右手を振り下ろす
	5185	I	左手を挙げる
	7654	L	左手で相手を指差す
	8004	L	左手で相手を指差す
	8485	L	左手で相手を指差す
	8524		
	8656	XX	左手をあごに当てる
	9403	O	左手を挙げる
	11121	O	左手挙げる
	11605	P	左手挙げる
	14803	L	左手で相手を指差す
	16988	XX	左手で相手を指す
	21040	V	右手で後ろを指差す
	21113	W	両手を裏返す
	26829	G	左手を振る
	26978	G	左手を振る
4	5512		
5	5554		
6	5664	J	それはこっち(左)に置いといて
	8057	XX	両手で数字を数える
7	8254		
8	8707	XX	右手をあごに当てる
9	9171	XX	右手を頭の付近でほわほわ
10	9829	J	両手を右左
	20367	U	右手で釘打ち
11	15005	Q	それはこっち(右)に置いといて
12	15946	XX	両手で何かを持つ
13	17071	R	右手でしたを指差す
14	17158	S	左手で涙を拭くう
15	17236		
16	18763	A	すりすり
	23604		
17	22136	J	それはこっち(左)に置いといて
18	23669		
19	25311	A	すりすり

表 5.15: システムによる分類結果（映像 2：被験者 MT）

システム	フレーム	被験者MT	実験時の動作メモ
1	440	A	右手を顔のあたりにやる
2	508		
	882		
	1268	B	左手を顔のあたりにやる
	1320		
	1766		
	2342		
	3582	C	両手を膝の間に
	5446		
	8438		
	8487		
	8880	B	左手を顔のあたりにやる
	8944		
	8996		
	25173		
	25584		
	25631		
3	2413	A	右手を顔のあたりにやる
	9706	XX	（複雑でわからない）
4	4082	D	右手で相手を指さす
	6749	A	右手を顔のあたりにやる
	17813	XX	（複雑でわからない）
	22882	A	右手を顔のあたりにやる
5	4143	D	右手で相手を指さす
	4559	E	右手で自分を指す
	7613	I	右手で相手を指す
	10802	K	右手を少し上げる
	11118	K	右手を少し上げる
	11390		
6	4223	B	左手を顔のあたりにやる
	4289		
	5184	G	両手で何かを持つしぐさ
	26420	XX	（複雑でわからない）
7	4670	F	右手を相手に振り下ろす
	7000		
	7712		
	9549	F	右手を相手に振り下ろす
	13024		
	13078	L	両手を少し上げる
	15109	K	右手を少し上げる
	17945	I	右手で相手を指す
	24997		
8	5597	B	左手を顔のあたりにやる
	25893		
9	6875	XX	（複雑でわからない）
10	7738	XX	（複雑でわからない）
	26543		
11	7761		
12	8371	A	右手を顔のあたりにやる
	9798		
	13349		
	13524	M	両手を顔のあたりに
	26597		
	26700		
13	8403		
14	8549		

システム	フレーム	被験者MT	実験時の動作メモ
15	8681	J	右手を振る
16	8717		
17	9771		
18	9912		
	10531		
19	9978		
20	10222	XX	（複雑でわからない）
21	10476		
	13661	N	右手を胸に
22	10586	XX	（複雑でわからない）
23	10621		
24	10653		
25	10735		
26	10831		
	10895		
	10924		
	12800		
	14662	O	右手で上を指さす
	15454		
	15541		
	15592		
	16461		
	16575	Q	右手で下を指さす
	16898		
	16923		
	17270		
	25320	D	右手で相手を指さす
	26973	V	両手をたたく
27	11264	B	左手を顔のあたりにやる
28	11616		
	28		
	26874	U	左手で足をたたく
29	12133	XX	（複雑でわからない）
30	12185		
31	12622		
	12922	XX	（複雑でわからない）
32	12948		
33	13298	M	両手を顔のあたりに
	18041	R	両手で自分を抱きしめる
34	13696		
35	14962	P	両手を右から左に
36	16702		
37	17149	B	左手を顔のあたりにやる
38	18436	S	両手でおなかをさする
39	18464		
40	19620	XX	（複雑でわからない）
	19914	F	右手を相手に振り下ろす
	25405	D	右手で相手を指さす
	25967		
41	19669		
	21255	XX	（複雑でわからない）
42	25055		
43	25443		
44	25935	XX	（その他）
45	26736		
46	26764	T	左手で右を指さす

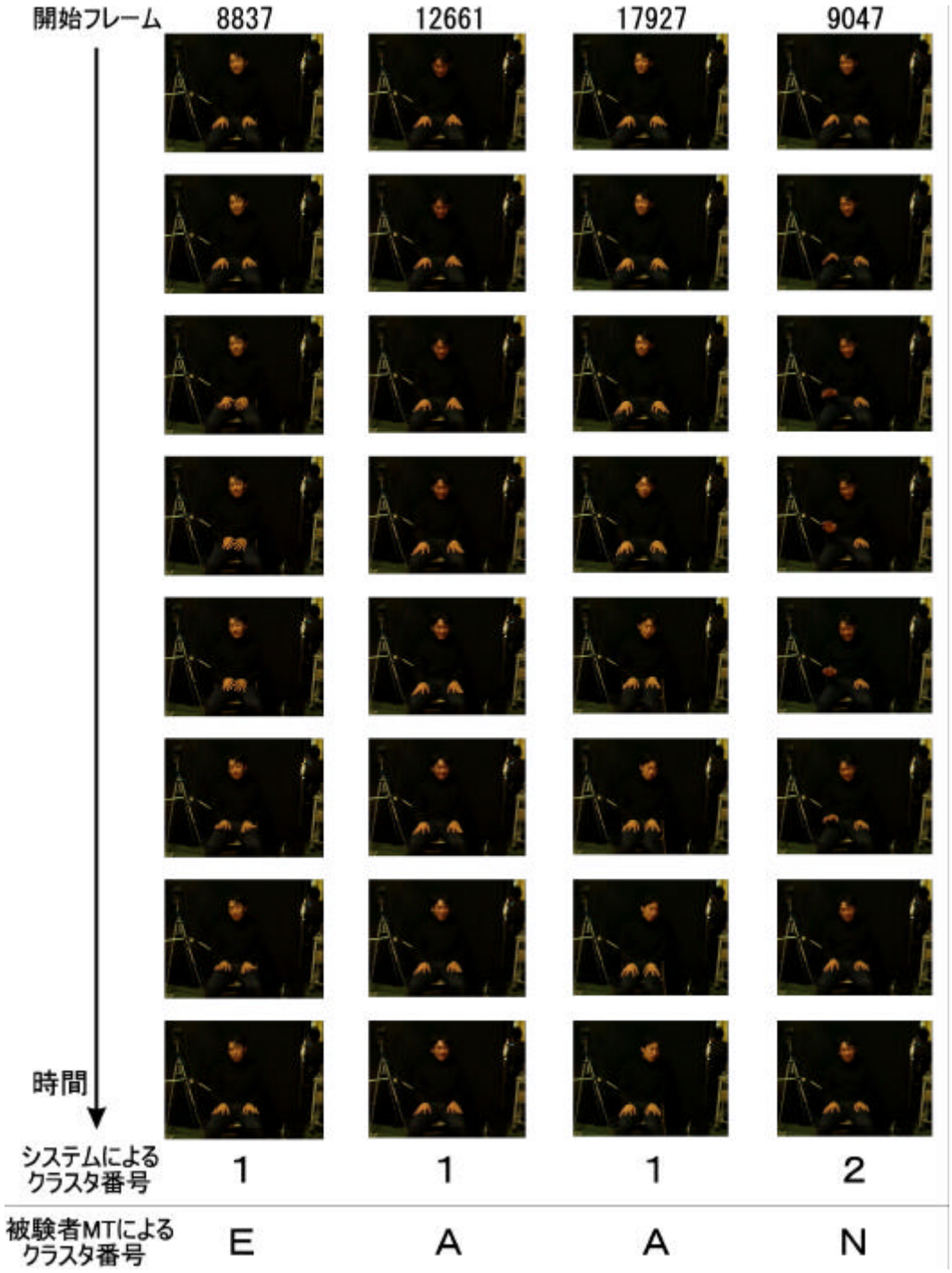

図 5.10: 分類結果の画像例（映像 1）

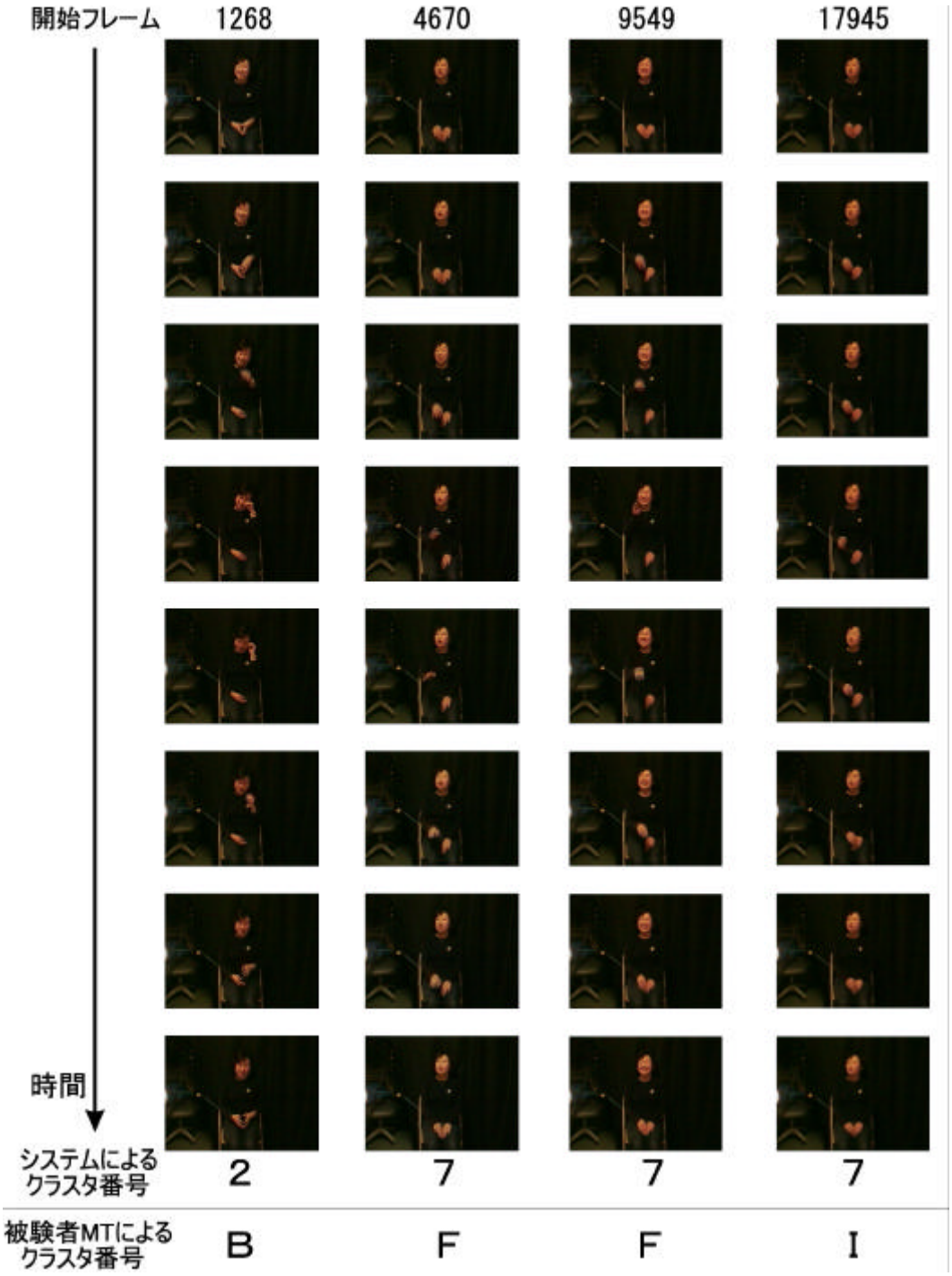

図 5.11: 分類結果の画像例（映像 2）

5.3　分類の再構成機能確認実験

5.3.1　実験目的

分類の再構成機能は、3.4.1 で述べたように、長時間の動作を分類した場合に生じる分類結果の偏りを適切に修正し、さらに以降の分類を個人の動作特徴に応じて適応させるためのものである。ここでは、試作システムに人間が対話している時の上半身動画像を入力し、分類の再構成を行った場合と行わなかった場合とを比較して、再構成機能の動作確認を行う。

5.3.2　実験方法

概要

本実験は、5.1.2 で述べた実験と同様に、あらかじめ用意した対話者の上半身動画像をシステムに入力し、身振りを分類する。さらに、この分類結果を再構成する。上半身動画像は 5.1.2 で述べた撮影環境で新たに 1 時間以上のものを記録する。

用意した対話時の上半身画像

5.1.2 で述べた方法と同様の対話を 4 回行う。撮影される対話者は 5.1.2 と同じ OT であり、対話相手は男子学生 4 名が順に交代して行う。各対話者はそれぞれ 15 分を目安に対話してもらう。これにより、対話者 OT の 1 時間以上の上半身動画像を撮影する。

実験手順

まず、用意した対話時の上半身動画像を試作システムに入力し、身振りを分類する。さらに、この分類結果を再構成機能により再分類する。そして、再構成前後の分類を比較し、再構成機能の動作を確認する。

5.3.3　実験の結果と考察

用意した上半身動画像の時間は 72 分間であった。実験結果を表 5.16 に示す。表には、再構成前後のクラスタ数、そのクラスタの中で 1 つの身振りだけで構成されているクラスタの数、および、システムが抽出した 身振り数を示している。

身振りが 1 つのクラスタとは、他のどのクラスタにも分類することができない身振りのことである。表からもわかるように、このような身振りが再構成前、再構成後と

表 5.16: 分類後・再構成後のクラスタ数

	再構成前	再構成後
クラスタ数	151	184
身振りが 1 つのクラスタ数	102	147
抽出した身振り数	545	

もに多くあることがわかる。これらの身振りは、対話者が同種の動作をあまりしない特殊なものであり、身振り自体の種類よりもその発生タイミングが重要である。

また、分類したクラスタを構成する身振りの数の分布を図 5.12 および図 5.13 に示す。ただし、クラスタの身振り数が 1 であるクラスタは分類の対象外として示していない。

この結果では、再構築後に 1 つのクラスタの身振り数が大きくなっている。身振り数が最大のクラスタの動作は「右手が膝の上にあり、左手を顔付近まであげる」動作であり、再構成前には別の動作として別クラスタに分類されていたものが 1 つのクラスタに再構成されたものである。

このように、分類の再構成により、主に別々の身振りとして分類されていたクラスタを併合して 1 つのクラスタに再構築していることがわかる。例えば、身振り数が 5 以上のクラスタ数は、再構成前が 22 であったのに対し、再構成後は 13 になっている。

また、分類の再構成に要する時間は OCTANE 上で 69.49[sec] であった。この処理時間は、身振りの数が増えるとその数の 2 乗に比例して増大し、身振り数があまりにも多くなると分類の再構築を行うことは困難である。本システムでは、再構成のために最大過去 1000 個の身振りの特徴ベクトルと分類結果を保持している。実験で 545 個の身振りを再構成したときの処理時間から推測すると、1000 個の身振りを再構成するのに要する処理時間 t_{1000} は、

$$t_{1000} = 69{:}49 \pounds \frac{1000^2}{545^2} = 235{:}67[\mathrm{sec}] \tag{5.1}$$

となり、4 分程度と実用時間内で再構成できることがわかる。

5.4 まとめ

ここでは、本章で行った実験により得られた結論について述べる。

まず、上半身動画像を身振りの単位に区切るセグメンテーション機能の評価実験と

して、システムによるセグメンテーションと人間による身振りの区分を比較する実験を行った。セグメンテーション機能評価実験では、

- システムは人間が 身振りと判断する動作をほぼ適切に抽出できる。
- 人間が身振りとは判断しない小さな動作でも、システムは抽出することができ、無意識に起こる暗示的な身振りをも十分とらえることができる。

ということがわかった。

次に、身振りの分類機能の評価実験として、システムの分類結果と人間の分類結果とを比較する実験を行った。分類機能評価実験では、

- 人間は長期間身振りを分類することが不得意である。
- システムは身振りを動作の種類ごとにほぼ適切に分類することができる。
- システムは十分にリアルタイムに動作できる。

ということがわかった。

そして最後に、システムの分類結果を再構成し、再構成前の分類と再構成後の分類とを比較した。分類の再構成の動作確認実験では、

- 再構成では、主に別の動作として分類されているクラスタを併合して適切な分類に再構成される。
- システムが最大限保持できる 1000 個の身振りを 4 分程度で再構成できる。
- リアルタイムに分類結果を再構成することは困難である。

ということがわかった。

以上により、試作システムが人間の身振りをリアルタイムに分類する機能を有することが確認できた。

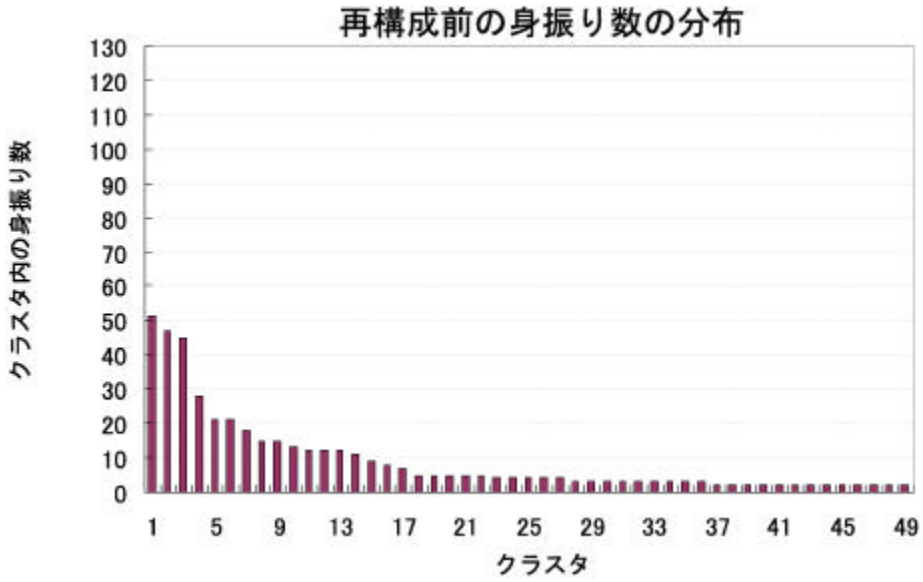

図 5.12: 分類後のクラスタ分析結果

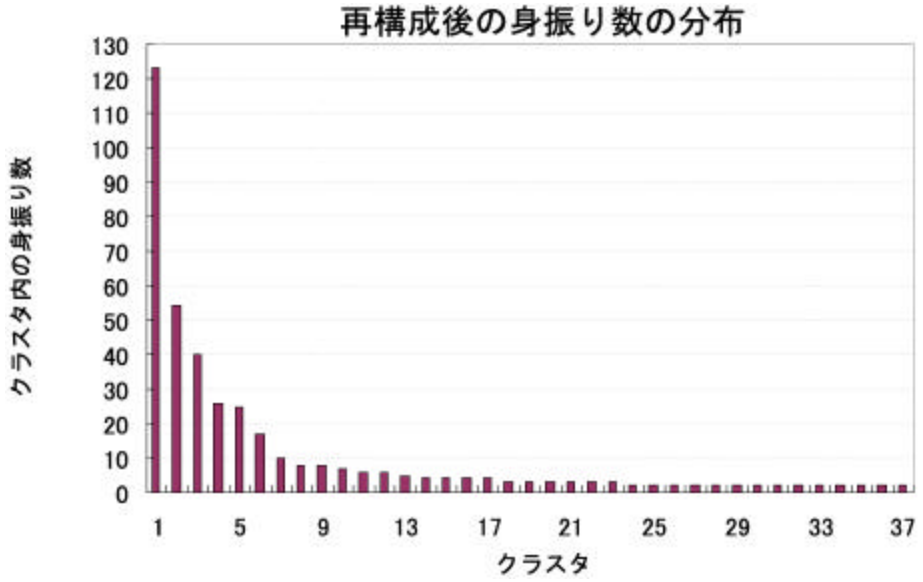

図 5.13: 再構成後のクラスタ分析結果

第 6 章 結論と今後の展望

本研究では、より良い操作性を実現する個人適応型のインタフェースの実現を目標とし、その基礎研究としてノンバーバルメッセージとしての身振りの利用に着目し、対話中の人間の上半身画像からリアルタイムに身振りとする動作を区分し、その動作を分類する手法を提案した。そして、その手法に基づく実時間動的人体動作分類システム ReD BACS （Real-time and Dynamic Body Action Classi¯cation System ）を試作し、身振りの分類実験によりその機能を評価した。

以上、本研究では、個人適応型のインタフェース実現のための基礎研究として、動的な身振りの分類手法の提案と、そのシステムの試作を行った。この研究を通して以下のことがわかった。

- 提案した身振りの自動分類手法は、個人適応型インタフェースの入力機構として利用可能である。
- 動作のセグメンテーションは、人間は感覚的にとらえているため、確固たる基準はない。
- 動作の分類にも、人間に基準はない。
- 動的な分類を長時間行う際には、その間に分類結果を再構成することがよい。

なお、本研究で試作した現在のシステムには次のような問題点がある。顔や手の領域抽出に肌色部分を利用しているため、身振りの分類対象の人は長袖の服を着ていること、照明環境が一定であることなど、入力画像の条件に制約がある。この問題を解決するには、新たに画像処理アルゴリズムを考案する必要があるが、その分処理が複雑になるので、リアルタイム性との兼ね合いが難しい。また、現在のシステムでは分類の再構成を行う判断基準を設定できない。この問題では、長時間使用する場合を考慮し、リアルタイムでの身振り分類と並行して分類の再構成を行い、その結果を反映させる機構が必要である。

次いで、本研究の将来展望を述べる。

本研究の将来展望

本研究の将来展望として、図 6.1 に示すように、身振りや表情だけでなく、音声などの複数のモダリティをヒューマンマシンインタフェースに利用した個人適応型インタフェースを構築することを視野に入れている。

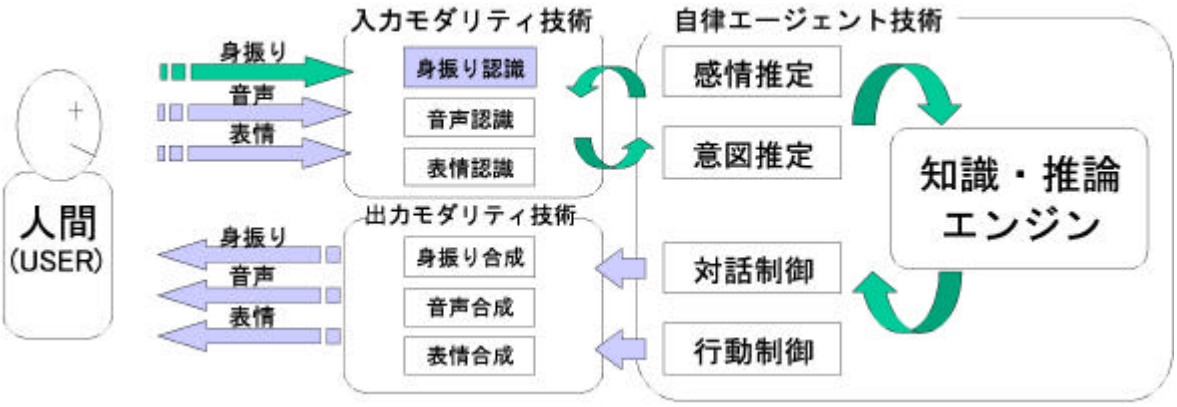

図 6.1: マルチモーダル・インタフェースの概要

近年では、携帯電話や携帯型コンピュータなど、その機器の使用者が特定の個人に限定される情報機器が多くなってきた。このような機器では、その機器の所有者である特定のユーザが使いやすければよい。これまでのは不特定多数のユーザ向け情報機器に個人が合わせていたが、これからは、個人個人に「カスタム化」できる商品が一層求められるものと考えられる。このような個々のユーザにカスタム化できる情報機器のマスプロダクションの方法として上記のようなマルチモーダルによる適応型インタフェースの機能を導入することが考えられる。この効果を以下に述べる。

- 同一人物が長期間にわたり使用すると、複数のモダリティを通して、その個人の性格や好みを把握し、その個人の好みにあったコンピュータの操作環境を実現することが可能となる。

- コンピュータとのインタラクションに人間同士が自然に行っているコミュニケーションを組み入れるため、これまでコンピュータに心理的な障壁や距離を感じていた「情報貧民」もコンピュータに親しみをもつ。

- 高齢者・障害者には、その失われたモダリティを補償するコミュニケーション機能を提供できる。

個人用の情報機器に前述のようなマルチモーダルな適応型インタフェースを適用し、同一人物が長期間に渡り使用すれば、複数のモダリティからその個人の性格や癖が推定され、個人に使いやすい環境を実現することが可能となる。また、機器とのインタラクションに人間同士で自然に行っているコミュニケーション手段も利用できるため、これまで機器操作に不安を感じていた人も、気軽に使用することが可能になると考える。

謝 辞

本研究を進めるにあたり、研究の全般にわたって熱意溢れるご指導頂きました吉川榮和教授に深く感謝致します。

本研究を進めるにあたり、研究の全般にわたって直接ご指導を頂き、幾度となく適切なご助言を頂きました下田宏助教授に心より感謝致します。

本研究を進めるにあたり、ご助言を頂きました石井裕剛助手に感謝致します。

本研究を進めるにあたり、貴重な時間を実験に費やし、顎が疲労するほど対話をして頂いた被験者の皆様、目が充血するほど人の身振りを観察して頂いた被験者の皆様に深く感謝致します。

本研究室で、卒業研究にてご支援頂き、修士論文では原稿の校正をして頂きました博士課程の小澤尚久氏に大いに感謝致します。

本研究をまとめるにあたり、データ整理などにご協力頂きました修士課程の松崎剛士君をはじめとする学生の皆様に感謝致します。

また、共によく遊びよく学んだ修士課程二回生のみなさまに大いに感謝します。

最後に、研究を進めるにあたり、研究室での快適な研究生活を送るために、大変お世話頂いた谷友美秘書、藤岡美紀秘書、森知寿秘書、吉川万里子秘書ならびに、昼夜問わず共に過ごして下さった吉川研究室の学生の皆様に深く感謝致します。

参 考 文 献

[1] 笹井寿郎：身振りの自動分類法に関する研究, ヒューマンインタフェースシンポジウム 2000 論文集, pp.479-482, (2000).

[2] Hartmut Dieterich, Uwe Malinowski, Thomas Kuhme, Matthias Schneider-Hufschmidt：State of the Art in Adaptive User Interfaces, Adaptive User Interfaces Principles and Practice, NORTH-HOLLAND, pp. 13-48 (1993).

[3] 池田謙一, 村田光二：こころと社会 認知社会心理学への招待, 東京大学出版, 第 6 章 1991).

[4] 大坊郁夫：しぐさのコミュニケーション 人は親しみをどう伝えあうか, サイエンス社 (1998).

[5] 田村博 編：ヒューマンインタフェース, オーム社, 第 12 章 (1998).

[6] 黒川隆夫：ノンバーバルインタフェース, オーム社, 第 3 章 (1994).

[7] 藤井美保子：コミュニケーションにおける身振りの役割, 教育心理学研究, 47-1, pp.87-96 (1999).

[8] Paul Ekman and Wallace V. Friesen：The Repertoire of Nonverbal Behavior:Categories, Origins, Usage, and Coding, Semiotica, 1, pp.49-98 (1969).

[9] 下田宏, 國弘威, 吉川榮和: 動的顔画像からのリアルタイム表情認識システムの試作, ヒューマンインタフェース学会論文誌, Vol.1, No.2, pp.25-32 (1999).

[10] Hiroshi Shimoda, Dazhao Yang, Hidekazu Yoshikawa: Dynamic Facial Expression Generation by using Facial Muscle Model, Proceedings of World Multiconference on Systemics, Cybernetics and Informatics, Vol.IX, pp.56-61 (2000).

[11] Paul Watzlawick, Janet Beavin Bavelas, Don D. Jackson 著：山本和郎 監訳, 尾川丈一 訳, 人間コミュニケーションの語用論, 二瓶社 (1998).

[12] 河野純大, 但田育直, 黒川隆夫：日本語文節列からの手話アニメーションの規則合成, 第４回SIGNOI研究会・第１回ヒューマンメディア研究会合同ヒューマンメディア研究会講演資料集, pp. 9-14 (2000).

[13] 毛利工：手指ジェスチャ認識に基づくウェアラブル型操作入力インタフェース, ヒューマンインタフェース学会論文誌 Vol. 2, No. 4, pp. 283-292(2000).

[14] 西川 敦, 大西 映生, 西村 正典, 平野 敦士, 小荒 健吾, 宮崎 文夫：局所相関演算に基づくオプティカルフローを用いた身振り動作の認識手法, 情報処理学会論文誌, Vol. 40, No. 8, pp. 3118-3133(1999).

[15] 岡田隆三, 白井良明, 三浦純, 久野義徳：オプティカルフローと距離情報に基づく動物体追跡, 電子情報通信学会論文誌 A, Vol.J79-A, No.10, pp.1-8 (1996).

[16] 佐々木大輔, 目加田慶人, 春日正男, 植田信夫：パーツの特徴に基づく動きパターンを用いた表情分類法に関する検討, 電子情報通信学会技術研究報告, Vol. 100, No. 376, pp. 33-38(2000).

[17] 渡辺孝弘, 李七雨, 谷内田正彦：インタラクティブシステム構築のための動画像からの実時間ジェスチャ認識手法－仮想指揮システムへの応用－, 電気情報通信学会論文誌 D-II, Vol.J80-D-II, No.6, pp.1571-1580 (1997).

[18] Albert Mehrabian：Silent messages : implicit communication of emotions and attitudes , Wadsworth Pub. Co. (1981).

[19] 渡辺富夫：コミュニケーションにおける身体性, ヒューマンインタフェース学会誌, Vol.1, No.2, pp.14-18 (1999).

[20] 阿部友一, 萩原将文：単眼視動画像からの人物頭部動作の解析と認識, 電気情報通信学会論文誌, D-II, bf Vol.J83-D-II, No.2, pp.601-609 (2000).

[21] 高田雄二, 長嶋祐二, 関宜正, 武藤大至, 呂山, 猪木誠二, 松尾英明：手話文認識のためのセグメント要素の解析, ヒューマンインタフェース学会論文誌, Vol. 2, No. 3, pp. 239-246(2000).

[22] 矢部博, 八巻直一：応用数値計算ライブラリ 非線形計画法, 朝倉書店 (1999).

[23] 西田春彦, 吉田光雄, 平松闊, 田中邦夫：クラスター分析, マイクロソフトウェア (1983).

[24] Michael R. Anderberg 著, 西田英郎 監訳, 佐藤嗣二 他 訳：クラスタ分析とその応用, 内田老鶴圃 (1988).

[25] 奥野忠一, 片山善三郎, 上郡長昭, 伊東哲二, 入倉則夫, 藤原信夫：工業における多変量データの解析, 日科技連出版社 (1986).

付 録 目 次

付 録 図 目 次

付 録 表 目 次

付録 A 領域特定手法の詳細

この手法は、基本的に入力画像の過去 2 フレームの情報を利用し、顔、右手、左手のそれぞれに対して画面上での探索範囲を設定し、その探索範囲内に抽出した領域が存在するかどうかを調べることで、対象領域の特定を行う。しかし、3.2.4 で述べたような「対象領域が身体や物の影になる場合に抽出できない」、「対象領域が重なった場合、1 つの領域として抽出される」という問題が存在するため、上記のような方法だけでは正しく顔、右手、左手領域の特定ができないことがある。

ここでは、まず部位が重複しているかどうかの判定部分の説明をし、次に、重複している場合の領域情報の取り扱いについて述べる。そして、対象領域の特定手法について詳しく説明する。

重複時の判定

対象が影に入る、もしくは対象同士が重なる場合には、抽出される対象領域数が減る。このような場合、重複したかどうか、影に入ったかどうかの判定は、次のルールに基づいて行う。

－顔と右手が重複
右手の探索範囲内に、顔領域と決定した領域しか存在しない状態

－顔と左手が重複
左手の探索範囲内に、顔領域と決定した領域しか存在しない状態

－右手と左手が重複
右手と左手のそれぞれの探索範囲内の領域が 1 つであり、かつ同じ領域である状態

－すべての部位が重複
右手と左手が重複状態であり、かつその領域が顔領域と決定した領域である状態

－右手もしくは左手が影に入る
右手もしくは左手の探索範囲内に抽出した領域が存在しない状態

このような状態であるとき、重複状態もしくは、影に入った状態とし、この場合の領域情報は次の重複時、消滅時の処理ルールに基づいて、取り扱う。

表 A.1: 重複時の処理ルール

状態	重複部位	ルール
顔＋右手	顔	最後に単独で領域特定できた領域
	右手	重複している領域
顔＋左手	顔	最後に単独で領域特定できた領域
	左手	重複している領域
右手＋左手	右手	重複している領域
	左手	重複している領域
顔＋右手＋左手	顔	最後に単独で領域特定できた領域
	右手	重複している領域
	左手	重複している領域

表 A.2: 消滅時の処理ルール

状態	ルール
顔領域なし	最後に単独で領域決定できた領域を代用する ただし、基本的に抽出失敗としてとらえ、 領域特定の初期化を行う
手領域なし	最後に単独で領域決定できた領域を代用する

（1）重複時、消滅時の処理ルール

顔は、右手、左手と比較して、移動量が小さい。よって、この重複時の処理ルールとしては、顔領域は、ほとんど移動しないと仮定し、重複時の顔領域は、元の領域と同じ領域情報で代用することとする。この処理ルールを表 A.1 に示す。

また、消滅時の処理ルールを表 A.2 に示す。ここでは、身体の影に入り領域抽出できない場合には、影に入った場所から出現すると仮定し、消滅する直前の領域と同じ領域情報で代用することとする。ただし、顔領域は影には入らないと考え、顔領域が消滅した場合は、すべての領域抽出に失敗したものとみなし、後述する領域特定の初期化を行う。

（2）領域特定の初期化

最初もしくは顔領域の抽出に失敗した場合には、過去 2 フレームの情報を利用することができない。よって、この場合には図 A.1 に示すように、あらかじめ設定した探索範囲により部位の特定を行う。入力画像の左上を原点、全入力画像サイズを w £ h［pixel］とすると、領域内に存在する任意の座標点を (x_i,y_i) とすると、設定した探索範囲は次のようになる。

初期顔領域探索範囲

$$\frac{w}{4} < x_i < \frac{3w}{4}; \qquad 0 < y_i < \frac{h}{2}$$

初期右手領域探索範囲

$$0 < x_i < \frac{2w}{3}; \qquad \frac{h}{3} < y_i < h$$

初期左手領域探索範囲

$$\frac{w}{3} < x_i < w; \qquad \frac{h}{3} < y_i < h$$

ただし、これらの範囲は互いに重なっており、この範囲に抽出した領域が複数存在することがある。この場合、領域を特定する順序を顔、右手、左手とし、面積の大きい順に領域を決める。

また本研究では初期位置を右手は左手の右側に存在すると仮定するため、範囲が重なり、右手と左手の位置が逆になってしまった場合は、これを入れ替える。

領域特定アルゴリズム

前述の領域特定の初期化により部位を特定する手法では図 A.1 に示す位置に顔、右手、左手が常に存在するとは限らないため、謝った領域特定を行ってしまう可能性がある。このため、初期領域部位を決定した後は、過去フレーム情報を利用する領域特定アルゴリズムにより領域部位の決定を行う。この領域特定アルゴリズムでは、顔と手に対して異なった手法を用いる。

顔領域特定

顔領域特定アルゴリズムでは、1 つ前のフレームの顔領域の抽出結果、すなわち顔領域の位置と大きさを用いる。実際の顔の幅を w_r［cm］、最大移動速度を v_r［cm/sec］と

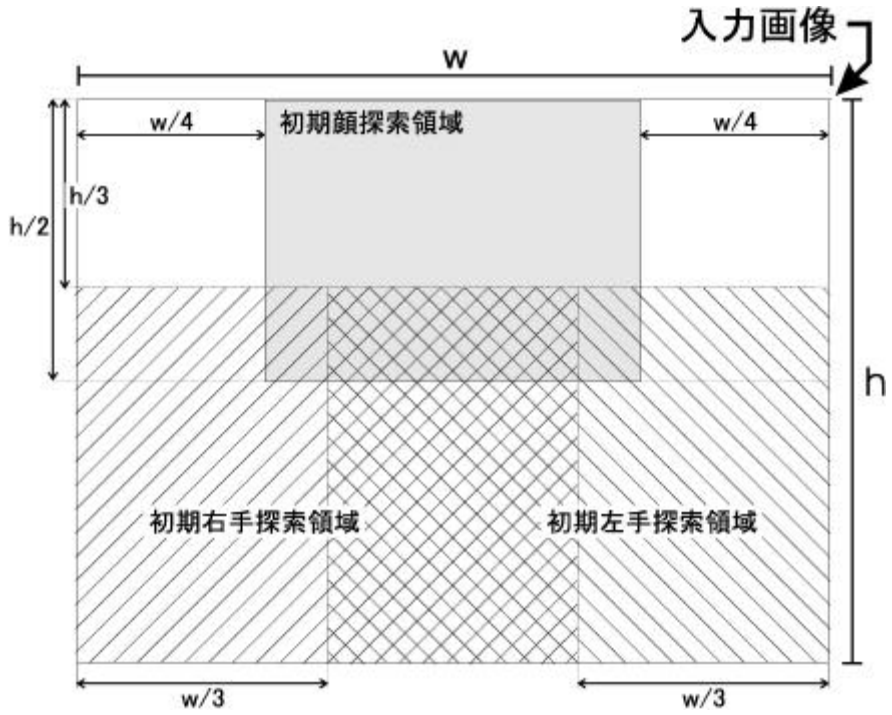

図 A.1: 領域特定の初期化時の探索範囲

すると、画像中での顔の最大移動量 d［pixel］は、過去フレームの顔の幅 w_p［pixel］、及び過去フレームからの経過時間 t［sec］により、次式のように求められる。

$$d = \frac{w_p}{w_r} v_r t \tag{A.1}$$

よって、前フレームの顔領域の左上の座標が $(x_p; y_p)$、大きさが幅 w_p、高さ h_p であった場合、顔領域の探索範囲は図 A.2 に示すように、左上の座標（$x_p - d; y_p - d$）、幅 $w_p + 2d$、高さ $h_p + 2d$ の矩形領域となる。本研究では、顔の幅を 15［cm］、最大移動速度を 50［cm/sec］と仮定して計算を行う。また、t については各フレームが取り込まれる際の時刻を記録しておき、前フレームと現在のフレームとの時刻の差をとっている。なお、この探索範囲内に複数の領域が存在する場合には、領域の中で最も面積が大きいものを顔領域とする。ただし、前フレームで顔領域の抽出に失敗している場合には、初期領域部位の決定と同様に、全画面からの顔領域を決める。

手領域特定

手領域の特定では顔領域とは異なり、前のフレームでの領域の状態により次の 3 通りに分けて考える必要がある。ただし、右手、左手とも別々に領域特定を行う。

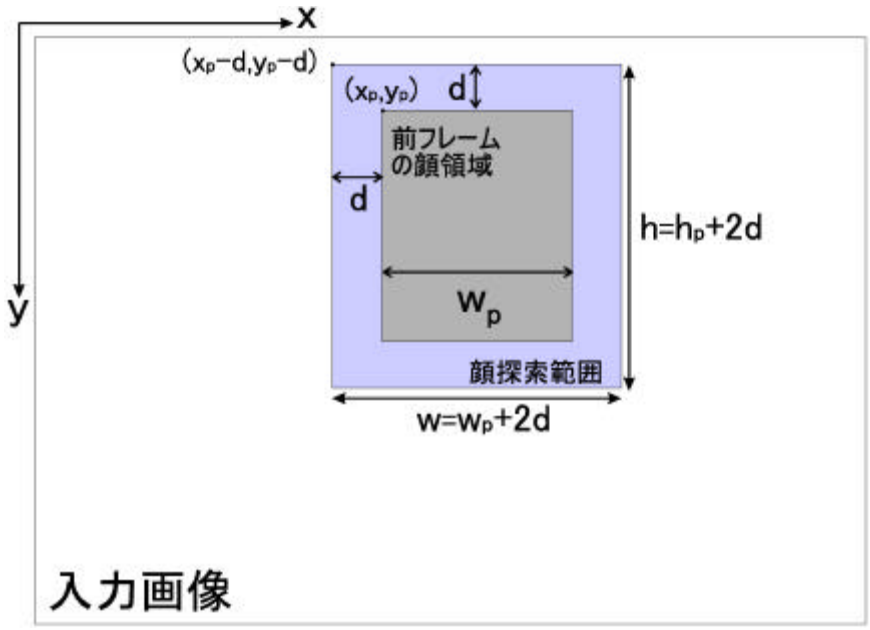

図 A.2: 顔領域の探索範囲

1. 前 2 フレームの領域情報がある場合

2. 前 2 フレームの領域情報がない場合

3. 前フレームの領域が他の領域と重複している場合

1. 前 2 フレームの領域情報がある場合

手領域の場合、顔と異なり移動速度が大きく、顔領域と同様の探索範囲を設定することができない。

このため、前 2 フレームで算出した重心点からの動きベクトルにより、探索範囲を設定する。2 フレーム前の重心点座標を $(x_{p2}; y_{p2})$、1 フレーム前の重心点座標を $(x_{p1}; y_{p1})$ とすると、動きベクトル M_1 は、$(x_{p1} - x_{p2}; y_{p1} - y_{p2})$ で表せ、フレーム間での移動距離 $\overline{M}_1$ は、$\sqrt{(x_{p1} - x_{p2})^2 + (y_{p1} - y_{p2})^2}$ で表される。よって、フレーム間の経過時間を t_1［sec］とすると、1 フレーム前の手の移動速度は次式で表される。

$$v_{p1} = \frac{\sqrt{(x_{p1} - x_{p2})^2 + (y_{p1} - y_{p2})^2}}{t_1} \tag{A.2}$$

よって、前フレームでの速度を元にした、重心点の予測移動距離 d［pixel］は、前フレームからの経過時間を t［sec］とすると次に示す式（A.3）のように求められる。

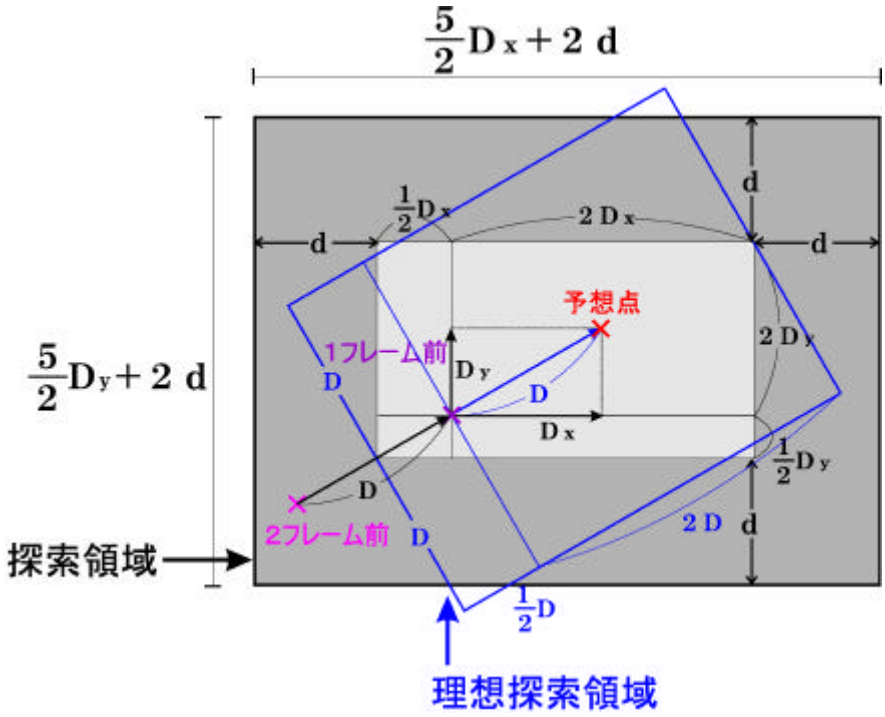

図 A.3: 手領域の探索範囲（2 フレーム前の情報がある場合）

$$D = v_{p1}t = \frac{\sqrt{(x_{p1} - x_{p2})^2 + (y_{p1} - y_{p2})^2}}{t_1}t \qquad (A.3)$$

ただし、この予測移動距離は速度が不変であると仮定しているが、実際は大きく変化している。また、速度が 0 であった場合に予測移動距離が 0 となってしまう問題もある。

このため、手領域の探索範囲としては、図 A.3 に示すように、動きベクトルの方向に 2D、それと逆方向に $\frac{1}{2}$D、ベクトル方向と垂直な方向に D とし、また顔領域と同様に式（A.1）で算出する手の最大移動量 d を加えた範囲とする。

ただし、本研究では処理の高速化を図るために、図 A.3 で示すように予想移動距離を x 軸方向、y 軸方向でそれぞれ式（A.4）で求め、矩形領域を探索範囲として用いる。

$$D_x = (x_{p1} - x_{p2})\frac{t}{t_1}; \qquad D_y = (y_{p1} - y_{p2})\frac{t}{t_1} \qquad (A.4)$$

抽出した肌色領域で、顔領域に決定した領域を除くものの重心点を算出し、重心点がこの範囲にあるものを手領域とする。ただし、この範囲に複数領域が存在する場合には、1 フレーム前の領域との重心点間距離が最も小さい領域を目的の手領域とする。

2. 前 2 フレームの領域情報がない場合

手が身体や物の影になり、前 2 フレームの抽出できなかった場合には、1. の動きベクトルは利用できない。この場合は、図 A.4 に示すように、まず最後に領域情報を取得できたフレーム上の重心点 P の座標を取得しておく。次に、抽出した領域の重心点と P との距離を算出しておき、顔、手と特定された領域との距離をそれぞれ r_f、r_h とした場合、式（A.5）に示す条件を満たす P を中心とした半径 r 内を探索範囲とする。

$$Min(r_f ; r_h) > r \qquad (A.5)$$

この探索範囲内に重心点をもつ領域のうち、その距離が最小のものを目的の手領域とする。

3. 前フレームの領域が他の領域と重複している場合

顔と手、もしくは両手が重なる場合には、抽出領域が重複した状態になる。この重複状態とは、複数の探索範囲が重複し、かつ、重複範囲内での抽出した領域が 1 つである場合である。重複している状態から分離する場合について、図 A.5 に示すような探索範囲において、図 A.6 に示す分離条件を満たすとき、重複状態から分離する。

以上のような処理を行うことで、抽出した対象領域から、顔、右手、左手の部位を特定する。

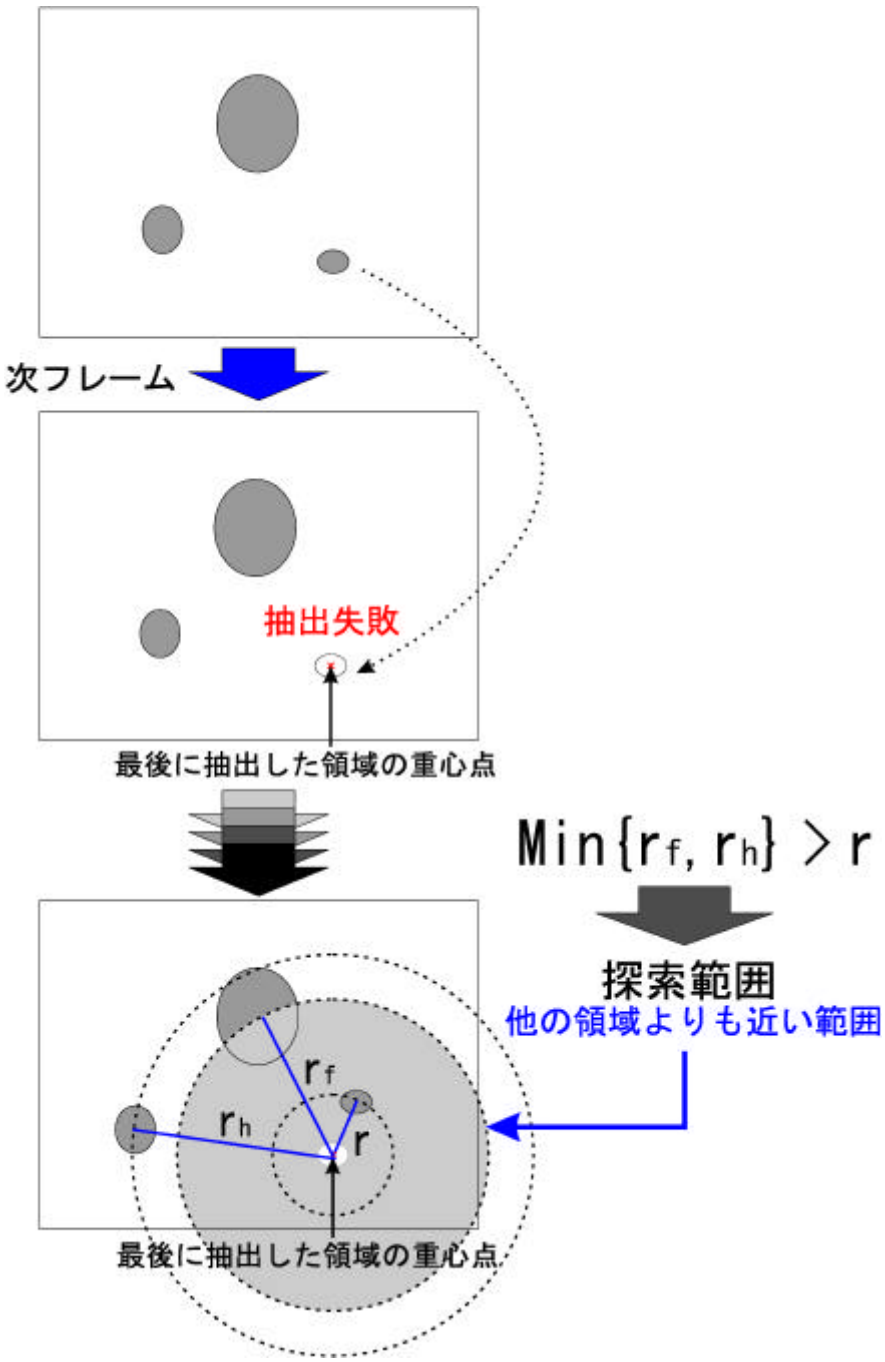

図 A.4: 手領域の決定手法（2 フレーム前の情報がない場合）

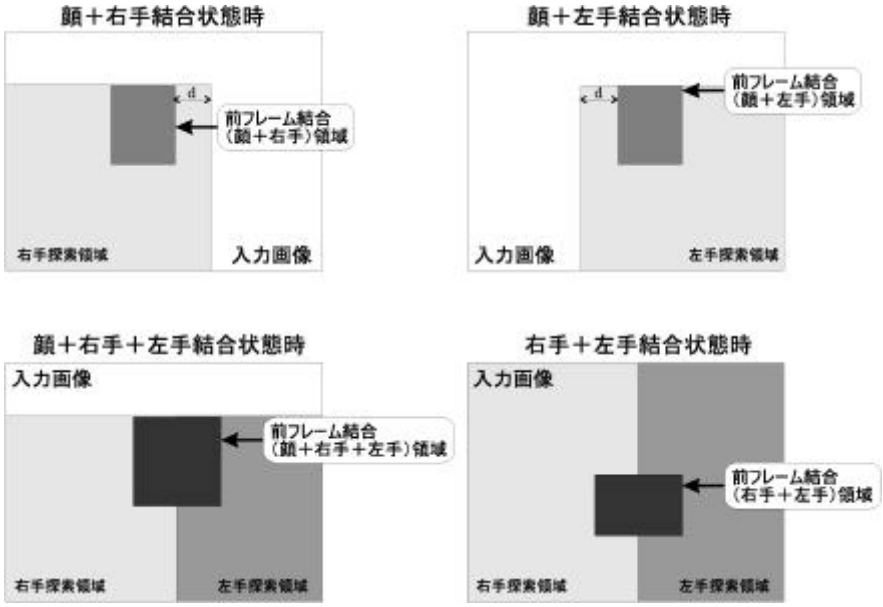

図 A.5: 手領域の探索範囲（前フレームが重複状態である場合）

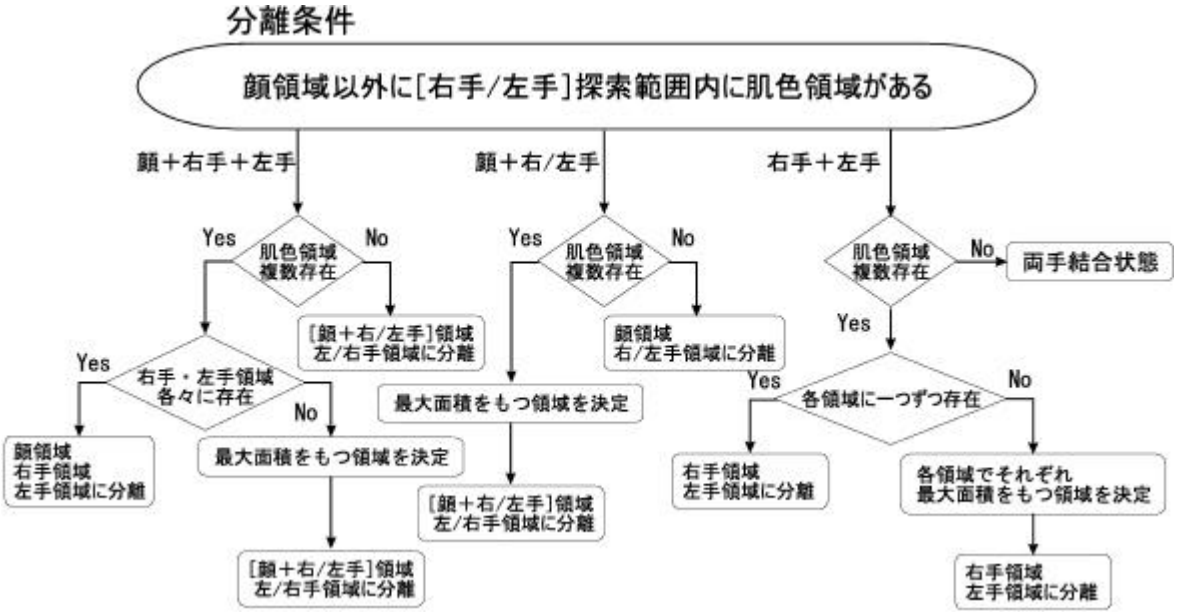

図 A.6: 重複状態から分離する条件

付録 B 身振りの分類機能評価実験結果

5.2 で述べた身振りの分類機能評価実験の結果を表 B.1～B.8 に示す。

クラスタ	分類数	分類した動作
A	5	右手を胸まで上げる
B	4	両手をちょっとだけ上げる
C	2	左手を頭まで上げる
D	5	右手をちょっと上げる
E	10	左手を胸まで上げる
F	2	両手を膝の上でもじもじ
G	6	両手でそれはおいといて(右へ)
H	2	右手を胸の高さで指さし
I	3	左手を胸の中央で指さし
J	1	左手を顎に当てる
K	1	右手を顎に添える
L	2	右手を顔の高さまで上げる
M	1	両手を右にぱっ左にぱっ
N	1	両手を膝の上であわせる
O	8	両手を膝の上でなでなで
P	1	右手を胸の高さで小刻みに左右に振る
Q	1	両手を膝の上でぱんぱん
R	1	右手をちょっと上げて真ん中に移し下げる
XX	6	複雑でわからない
	62	

図 B.1: 分類結果（映像 1：被験者 SK）

クラスタ	分類数	分類した動作
A	3	右手を顔の高さまで
B	1	身体を前方に傾ける
C	4	左手で目を拭う
D	10	右手で前方を指さす
E	4	右手で胸まで挙げる
F	1	両手でものをつかむまね
G	2	左手で鼻をさわる
H	6	右手で口おおう
I	1	両手で寝るしぐさ
J	2	右手をふる
K	4	右手をちょっと挙げる
L	4	両手ではくしゅ
M	2	右手で目を拭う
N	3	左手を顔の高さまで挙げる
O	4	両手で顔をさわる
P	3	両手を交互にちょっと挙げる
Q	1	両手を上下にあわせる
R	1	右手あいーん＋左手あいーん
S	2	両手をひざのうえで上下させる
T	5	両手を胸まで挙げ、左右に動かす
U	1	右手で左手をつかむ
V	1	右手で数回下方を指さす
W	1	両手を胸の前でクロスする
X	1	右手で鼻を触る
Y	1	左で指さす
Z	1	左手で膝をたたく
	69	

図 B.2: 分類結果（映像 2：被験者 SK）

システム	フレーム	被験者SK	実験時の動作メモ
1	182		
	650		
	723		
	1050		
	2792		
	8837		
	9277		
	10105	M	両手で右に「パッ」左に「パッ」
	12661		
	17271	N	両手を腰の上であわせる
	17927		
	18739	O	両手を腰の上でなでなで
	23577	Q	両手を腰の上でパンパン
	23949		
	24144	O	両手を腰の上でなでなで
2	1783	A	右手を胸まであげる
	1866		
	3515	B	両手をちょっとだけ上げる
	4535	E	左手を胸まで上げる
	4763		
	5008	D	右手をちょっとあげる
	5053		
	7150	H	右手で胸の高さで指さし
	8904	L	右手を顔の高さまで上げる
	9018		
	9047	A	右手を胸まであげる
	12616	G	両手で「それはおいといて」(右へ)
	15370	A	右手を胸まであげる
	16000	G	両手で「それはおいといて」(右へ)
	16628	D	右手をちょっとあげる
	17591	A	右手を胸まであげる
	17730	L	右手を顔の高さまで上げる
	18344	D	右手をちょっとあげる
	22211	G	両手で「それはおいといて」(右へ)
	23779	R	右手をちょっと上げ、真ん中に移し、戻す
	23871		
	26563		

システム	フレーム	被験者SK	実験時の動作メモ
3	2821	C	左手を頭まで上げる
	2909		
	4590	B	両手をちょっとだけ上げる
	4818	E	左手を胸まで上げる
	5185	E	左手を胸まで上げる
	7654	I	左手を胸の中央で指さし
	8004	I	左手を胸の中央で指さし
	8485		
	8524	J	左手を顎に当てる
	8636		
	9403	E	左手を胸まで上げる
	11121	E	左手を胸まで上げる
	11605	E	左手を胸まで上げる
	14803	I	左手を胸の中央で指さし
	18988	XX	複雑でわからない
	21040	XX	複雑でわからない
	21113		
	28829	E	左手を胸まで上げる
	28978	E	左手を胸まで上げる
4	5512	G	両手で「それはおいといて」(右へ)
5	5554		
6	5684		
	8057	XX	複雑でわからない
7	8254	XX	複雑でわからない
8	8707	K	右手を顎に添える
9	9171	C	左手を頭まで上げる
10	9829	D	右手をちょっとあげる
	20367	P	右手を胸の高さで小刻みに左右に振る
11	15005	G	両手で「それはおいといて」(右へ)
12	15948		
13	17071	H	右手で胸の高さで指さし
14	17158	XX	複雑でわからない
15	17238		
16	18763		
	23604		
17	22138	G	両手で「それはおいといて」(右へ)
18	23669	O	両手を腰の上でなでなで
19	25311	O	両手を腰の上でなでなで

図 B.3: 分類結果（映像 1：被験者 SK）

システム	フレーム	分類結果	実験時の動作メモ
1	440	A	右手を顔まで上げる
2	508	B	体を前方に傾ける
	882		
	1268	C	左手で目を拭う
	1320		
	1766		
	2342		
	3582		
	5446		
	8438		
	8487		
	8880	N	左手を顔まで上げる
	8944	N	左手を顔まで上げる
	8996		
	25173		
	25584		
	25631		
3	2413	A	右手を顔まで上げる
	9706	H	右手で口を覆う
4	4082	D	右手で前方を指さす
	6749	H	右手で口を覆う
	17813	D	右手で前方を指さす
	22882	D	右手で前方を指さす
5	4143	D	右手で前方を指さす
	4559		
	7613	K	右手をちょっと上げる
	10802		
	11118	D	右手で前方を指さす
	11390		
6	4223		
	4289	C	左手で目を拭う
	5184	F	両手でものをつかむ動作
	26420	G	左手で鼻を触る
7	4670	E	右手を胸まで上げる
	7000		
	7712	H	右手で口を覆う
	9549	A	右手を顔まで上げる
	13024		
	13078	O	両手の指で頬を押す
	15109	K	右手をちょっと上げる
	17945	K	右手をちょっと上げる
	24997	X	右手で鼻を触る
8	5597	G	左手で鼻を触る
	25893	C	左手で目を拭う
9	6875	I	両手で寝る仕草
10	7738	L	両手で拍手
	26543	L	両手で拍手
11	7761		
12	8371	H	右手で口を覆う
	9798		
	13349		
	13524	E	右手を胸まで上げる
	26597	O	両手で顔を触る
	26700		
13	8403		
14	8549	M	右手で目を拭う

システム	フレーム	分類結果	実験時の動作メモ
15	8681	H	右手で口を覆う
16	8717	J	右手をふる
17	9771	O	両手で顔を触る
18	9912	M	右手で目を拭う
	10531	H	右手で口を覆う
19	9978		
20	10222	P	両手を交互にちょっと上げる
21	10476		
	13661		
22	10586	Q	両手を上下に合わせる
23	10621	R	右でアイーン、左でアイーン
24	10653		
25	10735	S	左手で腹を触る
26	10831	K	右手をちょっと上げる
	10895		
	10924		
	12800		
	14662	D	右手で前方を指さす
	15454		
	15541		
	15592		
	16461		
	16575	V	右手で前方を指さす
	16898		
	16923		
	17270		
	25320	D	右手で前方を指さす
	26973	L	両手で拍手
27	11264	C	左手で目を拭う
28	11616	S	両手を膝の上で上下させる
	28		
	26874	Z	左手で膝をたたく
29	12133	T	両手を胸の高さで左右に動かす
30	12185		
31	12622	S	両手を膝の上で上下させる
	12922	P	両手を交互にちょっと上げる
32	12948		
33	13298	O	両手で顔を触る
	18041	W	両手を胸の前でクロス
34	13696	D	右手の指さし顔の高さ
35	14982	T	両手を胸の高さで左右に動かす
36	16702		
37	17149	N	左手を顔まで上げる
38	18436	T	両手を胸の高さで左右に動かす
39	18464		
40	19620	T	両手を胸の高さで左右に動かす
	19914	E	右手を胸まで上げる
	25405	D	右手で前方を指さす
	25967		
41	19669		
	21255	T	両手を胸の高さで左右に動かす
42	25055		
43	25443	L	両手で拍手
44	25935	D	右手で前方を指さす
45	26736		
46	26764	Y	左手で指さす

図 B.4: 分類結果（映像 2：被験者 SK）

クラスタ	分類数	分類した動作
A	3	右手を上げて振り下ろす
B	3	両手を浮かせる
C	1	右手で頭をかく
D	8	右手を上げる（相手を指す感じ）
E	1	左手を上げて手を振る
F	3	左手を上げて振り下ろす
G	2	右手で相手を人差し指で指す
H	4	左手を首まで上げる
I	2	両手を浮かせて左へ
J	5	左手を上げて人差し指で指す
K	1	左手を顎に添える
L	2	右手を顎に当てる
M	2	左手を上げて（複雑）
N	3	右手で左手首をいじる
O	2	両手を浮かせて右へ
P	1	右手を上げて下を指す（ここ）
Q	1	左手を顔まで上げて泣くふり
R	4	右手を上げて（複雑）
S	1	左手を上げて後ろを指す
T	1	両手を上げて広げる
XX	3	複雑でわからない
	53	

図 B.5: 分類結果（映像 1：被験者 OY）

クラスタ	分類数	分類した動作
A	2	右手を鼻にやる
B	5	左手を手にやる
C	2	右手を目にやる
D	5	右手で相手を指す
E	2	右手で自分の胸を軽くたたく
F	5	相手の方に右手を突き出す
G	1	両手を軽く前に出す
H	2	左手で鼻に手をやる
I	5	右手を口にやる
J	1	両手で寝るしぐさ
K	2	右手を横に振る
L	3	両手をたたく
M	1	右手を髪にやる
N	1	左手を口にやる
O	1	右手を挙げて振り下ろす
P	2	両手を口にやる
Q	2	右手でどこかを指さす
R	1	左手を挙げて振り下ろす
S	1	両手を軽く胸の前にやる
T	1	両手を軽く突き出す
U	1	人差し指で口の両端を押す
V	1	両手を顔にやる
W	1	両手を左右に振る
X	1	右手で軽く太股をたたく
Y	1	右手を左手首にやる
Z	1	右手で下を指す
AA	1	左手でどこかを指す
AB	1	右手で自分を指す
AC	1	右手を挙げて前後に振る
AD	1	両手で口を覆う
AE	1	左手でどこかを指す
AF	1	左手で左太股をたたく
XX	9	複雑でわからない
	66	

図 B.6: 分類結果（映像 2：被験者 OY）

システム	フレーム	被験者OY	実験時の動作メモ
1	182		
	650		
	723		
	1050		
	2782		
	8837		
	9277		
	10105		
	12661		
	17271	N	右手で左手首の服の裾をいじる
	17927		
	18739		
	23577		
	23949		
	24144		
2	1793	A	右手を上げて振り下ろす
	1866		
	3515	B	両手を浮かす
	4535	D	右手を上げる
	4763		
	5006	D	右手を上げる
	5053	D	右手を上げる
	7150	G	右手を上げて相手を指す
	8904	L	右手を上げて手を振る
	9018	D	右手を上げる
	9047	D	右手を上げる
	12616	O	両手を浮かせて右の方へやる
	15370	A	右手を上げて振り下ろす
	16000		
	16628	D	右手を上げる
	17591	R	右手を上げて複雑な動きをする
	17730	R	右手を上げて複雑な動きをする
	18344	D	右手を上げる
	22211	R	右手を上げて複雑な動きをする
	23779		
	23871		
	26563		

システム	フレーム	被験者OY	実験時の動作メモ
3	2821	C	右手で頭をかく
	2909		
	4590	E	左手を上げて手を振る
	4818	F	右手を上げて振り下ろす
	5185	H	左手を上げる
	7654	J	左手を上げて相手を人差し指で指す
	8004	J	左手を上げて相手を人差し指で指す
	8485	J	左手を上げて相手を人差し指で指す
	8524		
	8856	K	左手で額に手を添える
	9403	H	左手を上げる
	11121	H	左手を上げる
	11605	F	右手を上げて振り下ろす
	14803	J	左手を上げて相手を人差し指で指す
	16986	M	左手を上げて複雑な動きをする
	21040	S	左手を上げて後方を指す
	21113		
	26829	H	左手を上げる
	26978	J	左手を上げて相手を人差し指で指す
4	5512		
5	5554		
6	5664	I	両手を浮かせて左の方へやる
	8057	XX	両手を上げて複雑な動きをする
7	8254		
8	8707	L	右手で額に手を添える
9	9171	M	左手を上げて複雑な動きをする
10	9829	XX	両手を上げて複雑な動きをする
	20367	R	右手を上げて複雑な動きをする
11	15005	O	両手を浮かせて右の方へやる
12	15946	XX	両手を上げて複雑な動きをする
13	17071	P	右手の人差し指で下を指す
14	17158	Q	左手を目まで持っていって泣いてるふり
15	17238		
16	18763		
	23604		
17	22136	I	両手を浮かせて左の方へやる
18	23669		
19	25311		

図 B.7: 分類結果（映像１：被験者 OY）

システム	フレーム	被験者OY	実験時の動作メモ
1	440	A	右手で鼻に手をやる
2	508		
	882		
	1268	B	左手で目に手をやる
	1320		
	1766		
	2342		
	3582		
	5446		
	8438		
	8487	B	左手で目に手をやる
	8880	N	左手を口にやる
	8944	XX	複雑でわからない
	8996		
	25173		
	25584		
	25631		
3	2413	C	右手で目に手をやる
	9706	I	右手を口にやる
4	4082	D	右手で相手を指さす
	6749	I	右手を口にやる
	17813	AB	右手で自分を指す
	22882	XX	複雑でわからない
5	4143	D	右手で相手を指さす
	4559	E	右手で自分の胸のあたりを軽くたたく
	7613	F	右手で相手の方に手を突き出す
	10802		
	11118	Q	右手でどこかを指す
	11390		
6	4223	B	左手で目に手をやる
	4289		
	5184	G	両手を軽く前に出す
	26420	H	左手で鼻に手をやる
7	4670	F	右手で相手の方に手を突き出す
	7000	K	右手で横に手を振る
	7712	F	右手を口にやる
	9549	O	右手を上げて振り下ろす
	13024		
	13078	T	両手を軽く前に突き出す
	15109	X	右手で軽く右太股をたたく
	17945	F	右手で相手の方に手を突き出す
	24997	A	右手で鼻に手をやる
8	5597	H	左手で鼻に手をやる
	25893	AC	右手を上げて前後に振る
9	6875	J	両手を使って机の上で寝るような仕草
10	7738	L	両手で手をたたく
	26543	L	両手で手をたたく
11	7761		
12	8371	I	右手を口にやる
	9798		
	13349		
	13524	W	両手を顔の上にやり、すぐおろす
	26597	AD	両手で口を覆う
	26700		

システム	フレーム	被験者OY	実験時の動作メモ
13	8403		
14	8549	M	右手を腰にやる
15	8681	K	右手で横に手を振る
16	8717		
17	9771	P	両手で顔を覆う
18	9912		
	10531	I	右手を口にやる
19	9978	C	右手で目に手をやる
20	10222	XX	複雑でわからない
21	10476		
	13661	E	右手で自分の胸のあたりを軽くたたく
22	10586	XX	複雑でわからない
23	10621		
24	10653		
25	10735		
26	10831	F	右手で相手の方に手を突き出す
	10895		
	10924		
	12800		
	14662	Q	右手でどこかを指す
	15454		
	15541		
	15592		
	16461		
	16575	Z	右手で下を指す
	16898		
	16923		
	17270		
	25320	D	右手で相手を指さす
	26973	L	両手で手をたたく
27	11264	B	左手で目に手をやる
28	11616		
	11655		
	26874	AF	左手で左太股をたたく
29	12133	XX	複雑でわからない
30	12185		
31	12622		
	12922	R	両手を軽く胸の前あたりにやる
32	12948	S	右手を腰にやる
33	13298	U	両手の人差し指で口の両端を軽く押す
	18041	P	両手を交差させるように胸に当てる
34	13696		
35	14962	W	両手で左右に動かす
36	16702		
37	17149	AA	左手を上げて後ろを指さす
38	18436	XX	複雑でわからない
39	18464		
40	19620	XX	複雑でわからない
	19914	XX	複雑でわからない
	25405	D	右手で相手を指さす
	25967		
41	19669		
	21255	XX	複雑でわからない
42	25055		
43	25443		
44	25935		
45	26736		
46	26764	AE	左手でどこかを指す

図 B.8: 分類結果（映像 2：被験者 OY）