

# 研究テーマ 手サインインタフェースに関する研究

研究者 渡辺博己  
本郷仁志

財団法人ソフトピアジャパン  
財団法人ソフトピアジャパン

雇用研究員  
雇用研究員

## フェーズ I

### 1 研究の概要

人間と機械とが円滑なコミュニケーションを行うためには、ノンバーバル情報を無視することはできない。特に、人間の意図を機械に伝えるためには、手サインの認識による手サインインタフェース技術が重要である。そこで、画像から手領域を抽出し、抽出した手領域の形状を認識する手サイン認識手法の開発を行った。本研究では、単純背景下で抽出した肌色領域から四方向面特徴を用いて手領域の特徴量を抽出し、線形判別分析を用いて手サインの認識を行った。その結果、11種類の手サインに対し、平均93.6%の認識率を得ることが可能となった。

### 2 研究の目標

手サインインタフェース技術を構築するために、安定した手領域の抽出とロバストな手サインの認識が可能な手法を開発する。

### 3 実施内容

#### 3.1 手領域の抽出

手領域の抽出に当たっては、色情報を用いて肌色領域を抽出することとした。色を表現する表色系には様々な表色系が存在するが、本研究では、人間の知覚色差が幾何学的に近似できるとされているLUV表色系を用いて肌色の範囲を決定した。

安定した肌色領域を抽出するためには、各画素を肌色画素と非肌色画素とに分類する必要がある。そこで、以下のような方法で各画素を肌色画素と非肌色画素とに分類した。まず、入力画像の各画素からUV値を軸とした2次元ヒストグラムを生成し、予め設定しておいた肌色の取り得るUV値の範囲内で、最も画素数の多いUV値を基準肌色UV値として求めた。図1に、UV平面における設定した肌色範囲と基準肌色UV値の例を示す。次に、基準肌色UV値と入力画像の各画素におけるUV値との色差ヒストグラムを生成し、判別分析法により閾値を算出した。図2に、色差ヒストグラムと算出された閾値の例を示す。最後に、閾値に基づいて肌色画素と非肌色画素とに分類し、肌色画素をグループ化することにより肌色領域を抽出した。図3に抽出された肌色領域の例を示す。白色の矩形で囲まれた領域が抽出された肌色領域である。本研究では、手領域の抽出を簡単にするために、カメラと人物との位置関係を固定とすることで、画像中心に近い肌色領域を顔領域、その周辺にある肌色領域を手領域として抽出した。

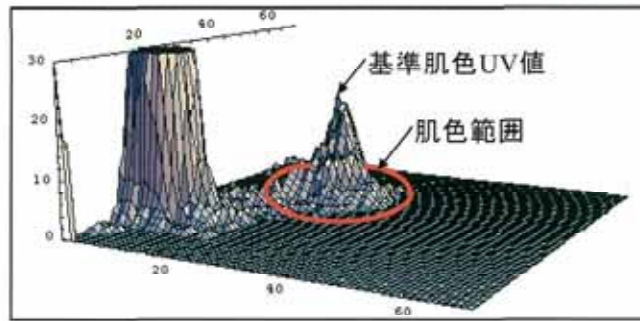


図1 UV ヒストグラムにおける肌色範囲と基準肌色 UV 値

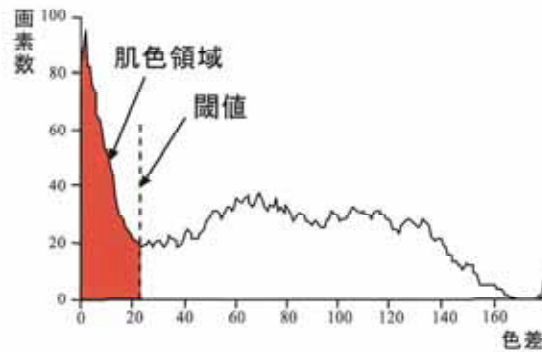


図2 色差ヒストグラムと算出された閾値



図3 抽出された肌色領域

### 3. 2 特徴量の抽出

前節で得られた手領域に対し、4方向面特徴を用いて手サインの特徴量を抽出した。4方向面特徴は文字認識の分野において開発され、有用性が証明された手法の一つである。特徴量は以下の方法で抽出した。まず、得られた手領域から Prewitt オペレータを用いて、水平、垂直、右上がり、右下がりの4つのエッジ方向面を抽出した。そして、これらの方向面をそれぞれ正規化、低解像度化し、低解像度化された各方向面の画素の濃淡値を特徴量として抽出した。抽出された4つのエッジ方向における特徴面の例を図4に示す。この特徴量は4つの方向面に分けてから解像度を低くするため、入力画像の解像度を直接低くする場合よりも、高解像度でのエッジ情報が保持されている。また、特徴面の濃淡値を特徴量としているので、空間配置情報を特徴量として保持している。



図4 4方向面特徴

### 3.3 手サインの認識

手サインの認識には線形判別分析を用いた。線形判別分析は、より低い次元で、しかも、クラス間の分離を最大限強調する線形写像を構成する多変量解析手法である。

学習の際には、画像データから得られた特徴量を線形判別分析することにより、係数行列と各クラスの平均ベクトルを生成し判別空間を構成した。しかし、様々な特徴量をもつものを同一判別空間内に分布させると、局所的に分布が集中したり、個々の分布状態が拡散したりするため、よい判別空間を構成することは困難であると予想される。そこで、適度な分布状態を形成するため、認識対象物をカテゴリに分類し、その中で判別空間を構成するように階層化を行った。

認識実験では、未知の画像データから特徴量を抽出し、係数行列を用いた写像変換より生成されるベクトルとのユークリッド距離を測ることでパターン認識を行った。この時、階層化された判別空間による認識の際には2回判別式を計算することになる。

## 4 結果

### 4.1 認識実験

手サインの認識実験を行うに当たっては、図2のように手サインを提示している画像を、5人の被験者から11種類の手サインについて、それぞれ100フレーム分を収集した(計5,500フレーム)。収集した手サインは、「指0本(グー)」、「指1本」の順に「指5本(パー)」までの6種類と指文字の「あ」から「お」までの5種類である。これらの画像について、4方向面特徴を用いて線形判別分析を行い、図5のように3つの判別空間を構成した。ここで、複数の判別空間を構成したのは、判別空間の構成の違いにおける認識率の検証を行うためである。

- (1) **実験1** 手サイン提示画像の収集時に得られた5人の被験者の顔クラスと11種類の手サインクラスをすべて1つにまとめた、16クラスの判別空間を作成し実験を行った。
- (2) **実験2** 手サインクラスのみで構成した11クラスの判別空間を作成し実験を行った。
- (3) **実験3** 5人の被験者ごとに、手サインクラスのみで構成した11クラスの判別空間をそれぞれ作成し実験を行った。

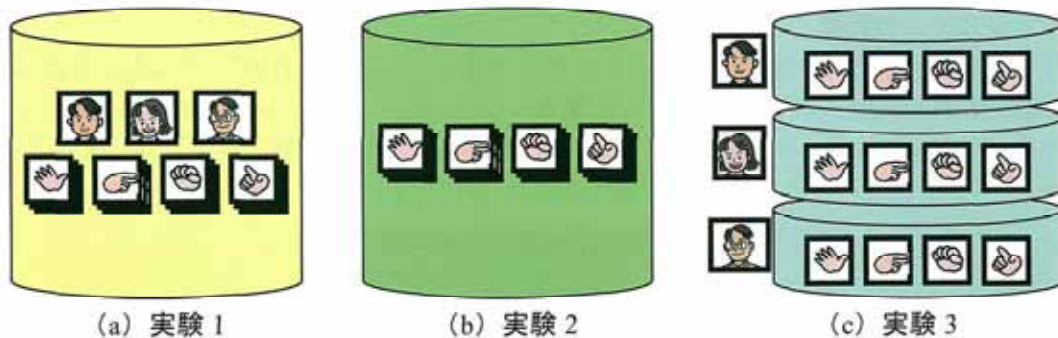


図5 判別空間の構成

### 4.2 結果と考察

表1に実験1~3における各手サインの認識率を示す。実験により、手サインのみで構成した判別空間は認識率を向上させることが分かった。実験3は、被験者間で手サインにバラつきがあることを予想し、被験者ごとに構成した判別空間は実験2における判別空間に比べ認識率を向上させることを期待して行ったが、平均認識率は実験2に比べ低下した。しかしながら、実験2において認識率の低かった手サインの認識率が実験3では向上していることから、被験者ごとの判別空間は有効であると考えられる。

表1 手サインの認識率 (単位: %)

	指0本	指1本	指2本	指3本	指4本	指5本	あ	い	う	え	お	平均
実験1	100	71.7	92.3	97.0	100	100	100	88.7	81.0	86.7	99.3	92.4
実験2	100	71.0	96.0	85.0	100	99.3	100	94.0	98.7	99.0	87.0	93.6
実験3	93.7	84.3	98.3	79.3	99.3	100	97.3	72.0	91.3	100	97.7	92.1

## フェーズII

### 1 研究の概要

人間と機械とが円滑なコミュニケーションを行うためには、ノンバーバル情報を無視することはできない。特に、人間の意図を機械に伝えるためには、手サインの認識による手サインインタフェース技術が重要である。そこで、フェーズIで開発した手サインインタフェース技術の応用例として、手サインによるAV機器の制御を可能とするシステムを構築した。これにより、家庭でリモコンにより操作していたAV機器を手サインで操作することが可能になると考えられる。

### 2 研究の目標

リモコンで操作していたAV機器を手サインにより操作するためのインタフェースを構築する。リモコンを手にとってボタンを押すまでの時間が約1秒程度であることから、手サインを出してAV機器が応答するまでの時間を1秒以内とする。

### 3 実施内容

#### 3.1 システムの概要

本研究では操作するAV機器をデジタルビデオカメラ(以下、DVC)とした。これは、PCへの接続が容易であり、制御プログラムを構築する上でもインタフェースが公開されているためである。図6にシステム構成を示す。カメラを具備するノート型PC(ソニー㈱、PCG-C1XF、Pentiumu2 400MHz)にIEEE1394ケーブルを通してDVC(ソニー㈱、DCR-PC110)を接続した。



図6 システムの概観

#### 3.2 処理の流れ

本研究では、カメラから320x240のカラー画像を取得した後、フェーズIで構築した手サインインタフェース技術を応用して、以下の通り処理を行うシステムを構築した。なお、手サイン認識では、フェーズIで記述した実験3のように個人ごとに手サインの判別空間を生成し認識を行うことにした。

まず、取得した画像から肌色領域を抽出する。通常は顔領域のみ抽出されるが、手サインの提示時には、顔と手領域の2つの領域が存在する。そこで、人間が対象に向かって手サインを提示する時には、顔より前の位置で提示する傾向があることから、本システムでは、最も大きい肌色領域を1つ抽出することにした。

次に、得られた肌色領域を顔領域か手領域かに分類する必要がある。そこで、手サイン認識手法を応用し、4方向面特徴と線形判別分析を用いて顔クラスと手サインクラスの2つのクラスを持つ判別空間を生成し分類した。事前に行った実験結果を表2に示す。結果より、登録人物においては100%分類できることが確認できた。また、未登録人物においても顔領域で99.0%、手領域で92.0%という高い確率で分類できることが確認できた。

表2 顔と手領域の認識率(単位: %)

	顔	手サイン
登録人物	100	100
未登録人物	99.0	92.0

最後に、システムは顔領域か手領域かによって処理を切り替える。得られた肌色領域が顔領域の場合、システムは個人識別を行い、ユーザを特定する。ユーザが特定されたら、そのユーザに対応する手サインの判別空間を呼び出し、手サインの提示に備えると共に、構築したGUIのコマンドリストをユーザごとのリストに切り替える。また、手領域の場合、手サイン認識を行い、手サインに対応する制御コマンドをDVCに送信する。なお、顔や手サインの認識時には5フレーム中4フレームが同一結果である場合のみ、認識結果を処理することにした。これにより、80%以上の認識率であれば100%の認識結果が得られることになる。また、認識したことをユーザに知らせるために、認識結果のカラー表示機能を付加した。

### 3.3 システムの操作

システムを操作する上で重要になるのがGUIである。本研究では、ユーザは家庭内における限られた人物であると想定し、図7に示すようなGUIを構築した。以下、操作方法について述べる。

#### (1) DVCの操作

DVCを操作するためには、事前に登録した手サインの中から、操作したいコマンドに対応している手サインをカメラに向かって提示する。登録した手サインと対応するコマンドを並べて表示することで、ユーザに対し分かりやすいコマンドリストの実現を図った。



図7 構築したGUI

## (2) ユーザの登録

ユーザを登録する場合は、「ユーザ登録」ボタンをマウスでクリックする。すると、確認のダイアログが表示されるので、「収集」ボタンをクリックすると画像の収集を開始する。画像を100フレーム収集するため、その間ユーザは入力画像を見ている必要があるが、収集に要する時間は約4秒程度であるため、大きな負担にはならないと考えられる。収集後、判別空間を生成し、収集したユーザの顔画像が「ユーザー一覧」に追加される。

## (3) 手サインの登録

手サインを登録する場合は、「サイン登録」ボタンをマウスでクリックする。すると、どのユーザの手サインリストに登録するかの確認ダイアログが表示されるため、「ユーザー一覧」からユーザの顔画像をクリックし、確認ダイアログの「OK」ボタンをクリックする。その後、収集確認のダイアログが表示されるので「収集」ボタンをクリックすると、ユーザ登録時と同様に画像を100フレーム収集し、判別空間を生成後、「サイン一覧」に登録した手サインが表示される。この時、手サイン画像の上部にある制御コマンドのプルダウンリストには「未定義」と表示されるので、リストからコマンドを選択する必要がある。

## 4 結果

以上をシステムとして構成し、動作確認を行ったところ認識速度は約5フレーム/秒であった。その結果、最終的な認識結果を取得するまでに5フレーム分の認識結果を必要とするため、手サイン提示からコマンド送信までは1秒程度要した。目標は1秒以内であったが、PCの処理能力を考えると概ね達成できたと考えられる。認識については、5フレーム分の認識結果の中で同一結果が4フレーム以上の場合にのみ処理するとしたことで、ロバストな手サイン認識が可能となった。

## フェーズⅢ

### 今後の取り組み

これまでの研究において、手サインインタフェース技術とその応用システムの有効性が示された。今後の取り組みとしては、安定した顔や手領域の抽出を実現することが考えられる。応用システムを考えた場合、照明や背景に対する頑健性が必要となるため、肌色情報と画像パターン等を統合した顔や手の検出技術等が重要である。また、実際にシステムを運用する場合は、機器の選択方法を検討しなければならない。