コンピュータビジョンと
イメージメディア 112－2
（1998．9．17）

ジェスチャー動画像データベースの開発

†向井理朗，‡山下浩生，†岡隆一

†新情報処理開発機構、‡メディアドライブ(株)

†茨城県つくば市竹園 1-6-1 つくば三井ビル 13 階

†TEL:0298-53-1641 FAX:0298-53-1640

あらすじ

人間の身振りを計算機システムにいかに理解させ，より円滑で自然な対話を支援するかは重要な課題である．こうしたジェスチャー認識システムの開発を進めるための共通的な基盤となる人間の身振りのデータベース整備はその対象となる身振りが非常に多様であることやデータ整備自身についての知見の不足からこれまで十分には行われてこなかった．そこで我々はジェスチャーデータベースの企画，仕様決定，作成を行ったので，データベースの収録方法，収録データ等についてを報告し，今後のデータベース整備の概要を報告する．

RWC Database -Gesture Motion Image Database-

†Toshiro Mukai, ‡Hiroumi Yamashita, and †Ryuichi Oka

†Real World Computing Partnership and Mediadrive Co.

†Tsukuba Mitsui Building 1-6-1 Takezono, Tsukuba-shi, Ibaraki

†TEL:+81-298-53-1641 FAX:+81-298-53-1640

Abstract

In this paper, we describe about Gesture Database. It is important to be understood human gesture by computer. A common database is necessary to develop gesture recognition system. We developed gesture database. We used sign language as a gesture. Sign language includes a rule of movement. Therefore, we use sign language as the data which don't depend on recognition system. We describe specification and recording method of database and describe preparation of future database. And, we intend to show this database to a general researcher.

1 はじめに

「ジェスチャー認識システム」の研究開発において，人間の身振りを計算機システムにいかに理解させ，より円滑で自然な対話を支援するかは重要な課題である．身振りや音声の認識によるマルチモーダル対話機能の実現などの新しいコミュニケーションの実現にはさまざまな技術的課題を解決する必要がある．

こうした研究を進めるにあたり，共通的な基盤となるデータベースが必要である．しかし，人間の身振りのデータベースの整備はその対象となる身振りが非常に多様であることやデータ整備自身についての知見の不足からこれまで十分には行われてこなかった．

また、ジェスチャー認識システムを評価するための一つの評価方法として，認識可能なジェスチャーの種類とその認識率があげられるが，評価を行うための共通の動画像データベースは例がなく，多くのジェスチャーを収録したデータベースの開発が必要であった．

本稿ではジェスチャー認識システムの評価及び新しい認識方法の開発を目的とし、共通に用いることができる動画像データベースを開発したので、これについて報告する。今回、収録の対象としたのは人間のジェスチャーである．あらかじめ収録対象とするジェスチャーを被験者に見せ，ものまねをしてもらい，それを収録した．データは時間的な対応関係をつけた一連の画像ファイルに収められている．

2 データベース設計方針

2 .1 収録対象

収録対象とする身振りの選定にあたっては，人工的なシステムに依存する身振り（システムにあわせた身振り）になることを避けるため，身振りに規則がある身振りを選定する必要がある。意図伝達を行うための身振りはいわゆるジェスチャーと呼ぶ身振りによって表現するものと手話のように文法が存在するものとに分類することができる．

a) ジェスチャー
人間が勝手に身振りを決めることが出来る．身振りに規則がない．

b) 手話
国・地域により多少の違いがあるが，身振りに規則がある．

収録対象とする身振りの選定にあたっては，身振りに規則がある手話単語を選定した．収録手話単語の選定には RWCP マルチモーダル日立研究室の協力により，手話を行う際の手の動きの領域に関するデータをもとに、動きの大きな手話単語を選定し，この中から手話単語 400 種類を選定した．また，この中の一部の単語を用いて 300 種類の短文(単語 2-5 個で構成）を作成し，これを収録した

2 .2 収録方法

新情報処理開発機構内に設置されているマルチモーダルデータベースワーキンググループの活動による知見により，ものまねによる収録が収集コストが小さく，得られるデータのばらつきも少ないことがわかっている．このことからジェスチャーデータの収録においてもものまね方式を採用した．ただし，収録対象が手話単語であるため，被験者には手話に関する予備知識がある健常者を選定した．

2 .2 .1 収録条件

被験者の上半身をデジタルビデオカメラで撮影した．カメラの配置と撮影の角度については，視線をできるだけそろえることと，

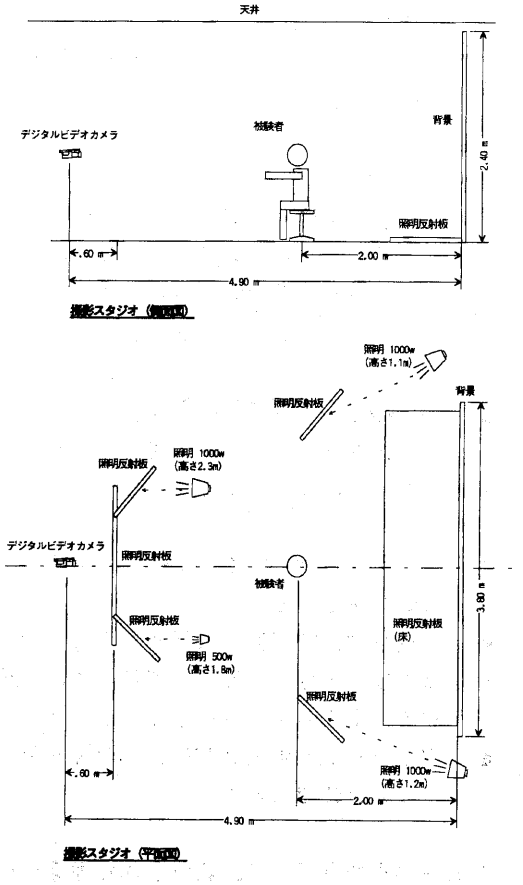

図1： 撮影スタジオ

ジェスチャーを画面内におさめることに留意した．

認識アルゴリズムによっては手首、ひじ、肩の位置を知る必要があるので、それぞれの部位におおまかな位置を知るためのマーカーをつけることにした．色の分離を考えて，腕に赤，ひじに橙，肩に白のマーカーをつけ，シャツは黄緑とした．

照明については，ビデオ撮影用の照明を用いた．図１に撮影スタジオのレイアウトを示す。被験者用に１０００ｗと５００ｗ照明を各１基，背景用に１０００ｗ照明を２基用いた．直接照明とせず，反射板に反射させる間接照明とした．被験者の顔及び背景の明るさがなるべく一様になるように，照明器具と反射板を配置した．また、背景は画像認識を簡単にするため，単一色（青）とした．

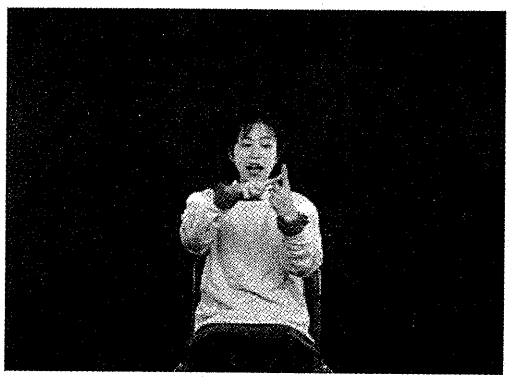

図2：収録した映像

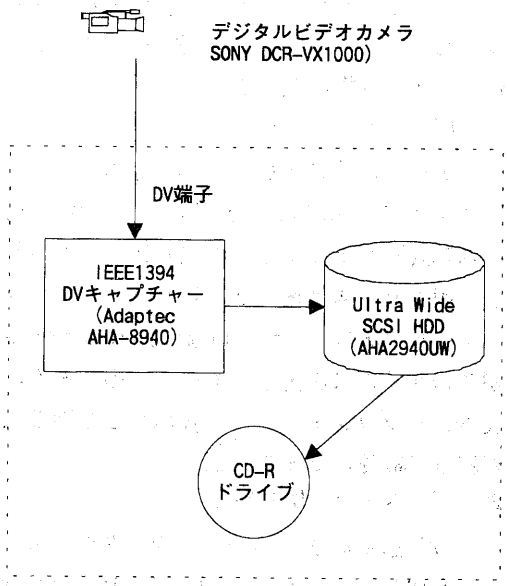

図3：ファイル化のシステム構成

2.2.2 被験者

被験者にはほぼ同じ体型であることが望ましい。そこで今回は同世代でかつ体型もほぼ同じである 20 代の女性２人を選定した．この２名の被験者はいずれも手話に関する予備知識（過去に手話の勉強をしたことがある程

度）があった．その上で見本を見せ、これを物まねした状態を収録した。

2.2.3 収録内容

あらかじめ選定しておいた単語４３１種類，短文３０１種類の手話をしてもらい，それを収録した．収録ジェスチャのリストは http://www.rwcp.or.jp/wswg/rwcdb/gesture 以下に付録として掲載してあるので、こちらを参照していただきたい。

2.3 ファイル化

デジタルビデオカメラで撮影した映像を，パーソナルコンピューターに取り込みファイル化し，CD-Rに記録した． 図３ にファイル化のシステム構成を示す．

1. DVキャプチャー

デジタルビデオカメラからPCのハードディスクに，ＤＶ端子経由で，映像をAVI(DV soft圧縮形式) ファイルとして取り込む．このときの画像サイズは横７２０ピクセル縦４８０ピクセルである．これはDV形式AVIファイルがこのサイズに固定されているためである。実際にパソコンモニターに表示すると縦横比が狂って横長になっている．

2. ジェスチャー切り出し

取り込んだ AVI ファイルから各ジェスチャーを切り出し，さらに，横 320 ピクセル縦 240 ピクセルのサイズに縮小する（この時に横長の画像を適正な比に修正している）．

3. 連番画像への変換と Mpeg ファイルの作成

切り出した AVI ファイルを，各フレーム毎の連番画像に変換し，これをＺＩＰ圧縮する．さらに，参照用としてMpegムービーファイル（Mpegl）を作成した．

4. タグ付け

収録した手話にタグ（ラベル）を付けた．タグには手話の開始，終了フレームなどを記述した．１種類の手話で１つのタグファイルとした．タグの例を表１に示す

2.4 収録メディア及び形式の検討

データベースとして広く利用していただくにあたって以下の点に考慮した．

- 広く用いられているフォーマット及び，媒体であること
- 作成及び複製が廉価に行われること
- 媒体は，小さく，軽く，衝撃に強い，郵送等による頒布にも適していること
- 収録データが磁気などによる影響を受けにくい媒体であること

これらの点を考慮して，収録された画像データの内容を示すための配布媒体として，CD-ROMを用いることにした． ファイル化の際データを記録しておいたCD-RをISO9660フォーマットで複製して配布することにした．

また大量の動画像データを収録することが想定されるため，CD-ROM内に書き込まれる，

- 収録データ（画像）のスペック
- データファイルの圧縮形式
- プレビュー用データのファイル形式

について検討した．これらの項目を決定する

表1：タグ付けの例

（開始フレーム）	（終了フレーム）	（単語の漢字表記）
0003	0055	バスケットボールの
0057	0073	試合を
0078	0103	応援する

表 2：連番画像形式

圧縮形式 ：	非圧縮
画像サイズ(横X縦)：	320 X 240 ピクセル
色深度 ：	２４ビット/ピクセル（ＲＧＢ各８ビット）
原点 ：	左上
フレームレート ：	３０フレーム／秒 （1フレームが1ファイルに相当）

にあたって以下の点に考慮した.

- 想定される認識実験に対して,適切な大きさのデータ量であること
- 圧縮形式は, UNIX, Windows95, MacOS, DOSなどいずれのOS環境でも解凍することが可能,圧縮率が高く,圧縮解凍速度が速いものであること
- 参照用データの形式は,圧縮形式同様,いずれのOS環境でもデータ内容を簡単に確認できること.

以下に収録データの書式、ファイル名の付け方等について述べる。

2.4.1 収録データの書式

収録データは連番画像、参照用ムービー、タグファイルである。連番画像のスペックを表 2 に示す。連番画像は次節の命名ルールに基づき、一つのファイルとして ZIP 形式に圧縮、収録されている。参照用ムービーは:MPEG-1形式のムービーファイルとして収録した。

2.4.2 ファイル名の命名ルール

以下の規則にしたがってファイル名を命名した.

- 被験者コード
  最初の１文字の「m」は男性を，「f」は女性を表わす．次の２桁の数字は通し番号を表わす．例えば「m02」は男性の２番，「f11」は女性の 11 番を表わす． 今回は女性２人なので「f01」と「f02」のみである.

●手話番号

単語手話は0001を開始とする4桁の通し番号で,文手話は1001を開始とする通し番号で表わしている.

●ファイル名

被験者コードと手話番号の組み合わせでファイル名を表わす．例えば，女性２番の文手話 35 番の場合，Mpeg ファイルは，f021035.mpg となり，zip ファイル(連番画像を zip 圧縮してアーカイブ化したファイル)は，f021035.zip となり，タグファイルは，f021035.tag となる．また，zip ファイルに格納されている連番画像ファイルは，各フレーム毎に f021035.000, f021035.001 , f021035.002 ... となる.

2.4.3 CD-ROM のディレクトリ構造

各被験者のデータファイルは複数枚に渡る．最初の数十枚に単語ジェスチャー(Zip ファイル)を入れ,次の数十枚に文ジェスチャー(Zip ファイル)を入れ，最後の１枚に Mpeg ファイルとタグファイルを入れた.

●単語ジェスチャー

１枚目～m枚目 Zip ファイル
f010001.zip, f010002.zip,

…, f010426.zip

●文ジェスチャー

m+1 枚目～n 枚目　　Zip ファイル

f011001.zip, f011002.zip,

…, f011301.zip

●タグファイル

n+1 枚目

f010001.tag, f010002.tag,

…, f010426.tag

- Mpeg ファイル

f010001.mpg, f010002.mpg,

…, f010426.mpeg

上記の要領で CD-R に収録した結果、今回収録したデータの場合では2名分で41枚のデータとなった。

3 評価及び今後の課題

今回整備したデータはジェスチャー動画像である。問題点として、

- 繰り返しがない。
  ジェスチャー＝手話単語となるデータには、繰り返しがなく、認識実験に用いるためには参照用データとテスト用データの最低2セットが必要である。
- 被験者が少ない。
  被験者が同性（女性）2名とデータとして不足している。
- 収録時に手話として正しいかの評価を加えていない。
- ステレオ情報がない。
  単眼のカメラによるデータであるため、奥行情報がなく、ステレオ画像による認識実験に用いることができない。
- データの量が多く、記録媒体の検討が必要である。

などがあげられる。これらの問題点を考慮し、今後のデータベース整備として、

- 2回以上の繰り返しがある。
- 被験者は4名以上（多い程よい）。
- ステレオ画像に対応するため、補助カメラを設置する。
- 収録メディアとして大容量の記録媒体であり、一般にも普及するものを選定する必要がある。

などの点を考慮してデータの収集整理を行う必要がある。

現在、今年度のデータベース収録についても検討を行っている。上記の問題を考慮し、今年度収録データとしては、被験者 4 名、繰り返し 2 回、語彙（ジェスチャ-）数 2～300 程度の規模でのデータの収集整備を計画している。収録メディアについても、現在検討を行っている。DVD ドライブの普及を考慮し、CD-ROM または DVD-ROM による収集を検討している。

4 むすび

本稿では RWCP データベースワークショップ人間動作理解データベースサブワーキンググループ（サブ委員会）の平成9年度の活動の一部として，企画，仕様決定，作成を行った大語彙ジェスチャーデータベースについて報告した.

ジェスチャー認識システムの評価を行うための統一的な動画像データベースとして動きに規則性がある手話単語を対象としてきた. しかしながら対象語彙数が 400 と手話単語を扱うには少なく，繰り返しのあるデータも少ないため，より多くの動画像データの収集が今後の課題として残されている.

今回作成したジェスチャーデータベースについては，研究目的に限り、一般に公開す

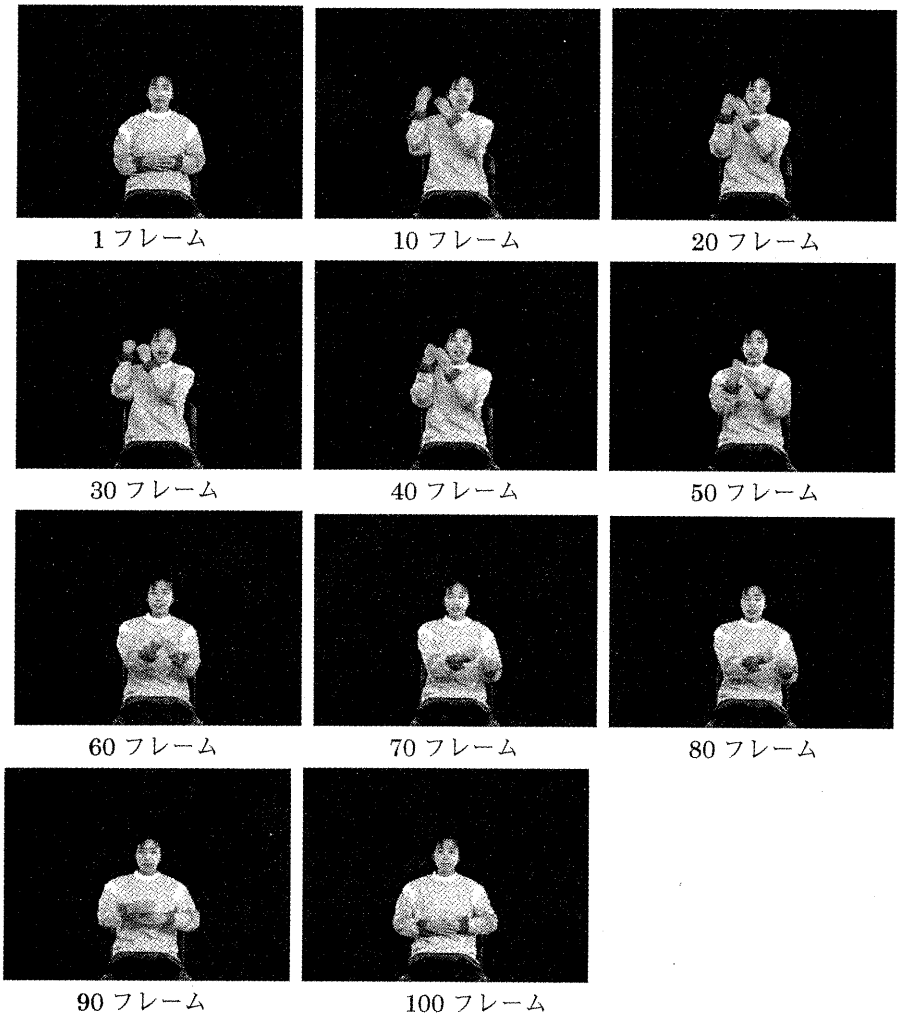

図4：収録ジェスチャーの例

る予定があり、関係者から広くご意見を頂けたら幸いである.

謝辞

本研究を行うにあたり，機会を与えてくださった，新情報処理開発機構つくば研究センター島田潤一所長に深く感謝いたします．本データベース構築にあたり，手話の動きに関するデータを提供していただいた RWCP マルチモーダル日立研究室の皆様に感謝いたします。また、ご助力をいただいた情報ベース機能つくば研究室の皆様、電総研坂上勝彦氏をはじめとするデータベース WG 委員の皆様に感謝いたします。また，被験者としてデータ収録にご協力いただいた方々に感謝の意を表します.

参考文献：

[1] 速水悟，他：“身振りと発話の RWC マルチモーダルデータベース”，人工知能学

会情報統合研究会，SIG-CII-9710，pp. 20-27(1997)
[2] 速水悟，他：“音声と画像情報による対話システムのためのデータベース設計”，人工知能学会全国大会 15-07，pp. 423-426(1996)
[3] 速水悟，他：“RWC マルチモーダルデータベース”,第3回知能情報メディアシンポジウム論文集 pp. 245-252(1997)
[4] 速水悟，他：”身振りと発話のマルチモーダルデータベース”,電子情報通信学会技術報告 PRMU97-95 pp. 1-8(1997)
[5] 鎌田一雄：“手話・身振りインタフェース構築の現状と課題”，情報処理学会研究報告 CVIM，Vol.97，No.114pp. 31-38(1997)
[6] http://www.rwcp.or.jp/wswg/rwcdb/