九州大学学術情報リポジトリ
Kyushu University Institutional Repository

カメラを用いた手話認識における見えの違いを考慮した手話の局所特徴認識

今川, 和幸
松下電器産業株式会社九州マルチメディアシステム研究所 | 九州大学大学院システム情報科学研究科

谷口, 倫一郎
九州大学大学院システム情報科学研究科

有田, 大作
九州大学大学院システム情報科学研究科

松尾, 英明
松下電器産業株式会社九州マルチメディアシステム研究所

他

https://hdl.handle.net/2324/5706

出版情報：映像情報メディア学会誌 : 映像情報メディア. 54 (6), pp.848-857, 2000-06-20. 映像情報メディア学会
バージョン：
権利関係：

KYUSHU UNIVERSITY

The Institute of Image Information and Television Engineers

カメラを用いた手話認識における見えの違いを考慮した手話の局所特徴認識

Recognition of Local Features for Camera-based Sign-Language Recognition System

学生会員 今川和幸[†,††], 正会員 谷口倫一郎[††], 有田大作[††],
松尾英明[†], 呂 山[†††], 猪木誠二[†††]

Kazuyuki Imagawa[†,††], Rin-ichiro Taniguchi[††], Daisaku Arita[††], Hideaki Matsuo[†], Shan Lu[†††] and Seiji Igi[†††]

Abstract A sign-language recognition system should use information from both global features, such as hand movement and location, and local features, such as hand shape and orientation. We designed a system that first selects possible words by using the detected global features, then narrows the choices down to one by using the detected local features.

In this paper, we describe an adequate local feature recognizer for a sign-language recognition system. Our basic approach is to represent the hand images extracted from sign-language images as symbols corresponding to clusters by using a clustering technique. The clusters are created from a training set of extracted hand images so that images with a similar appearance can be classified into the same cluster in an eigenspace. Experimental results showed that our system can recognize a signed word even in two-handed and hand-to-hand contact cases.

キーワード：画像認識，ジェスチャ，手話，KL 展開，クラスタリング，局所特徴

1. まえがき

現在，新たなヒューマンインタフェースや，聴覚障害者と健聴者の間のコミュニケーション支援を目指して，カメラで撮影した動画像から，手振りやジェスチャ，手話を認識する研究が盛んに行われている．その中でも手話は最も構造がはっきりしており，手の位置および動きといった大局特徴と，手の形状と向きといった局所特徴を用いて表記化することが可能であるといわれている[1)]．

従来，手振りやジェスチャを実時間で認識する場合，撮影した画像から画像全体もしくは腕の振りといったおおまかな特徴（大局特徴）を抽出することにより，認識することが有効であるといわれていた[2)~4)]．しかしながら，手話の場合，手の形状，動き，位置によって数多くの単語を表現している．そのため，体全体の動きとして似た動作が複数存在し，大局特徴を抽出するだけでは，それぞれの手話単語を明確に区別することが不可能である．したがって，手話を認識する場合，大局特徴（手の位置，動き）と局所特徴（手形状，向き）の両方の特徴を抽出し，認識することが必要である．

さて，現在までに動画像から手話を認識するための多くの手法が提案されている．しかしながら，これらの手法は，大局特徴のみから手話を認識する手法[5)6)]，もしくは，カメラに向かって提示する指文字等の片手形状を認識する手法[7)~9)]のいずれかであり，大局的な動き（腕・全身の動き）と局所的な動き（手首・指の動き）が混在する中で，大局特徴と局所特徴の両方を抽出し，総合的に認識することは行われていない．これは，大局動作と局所動作が混在する中から局所特徴を抽出する場合，カメラに向かって提示する指文字等の片手形状を認識する場合に比べ，以下の課題が存在するためである．

（1） 手話の場合，片手動作，両手動作の両方の場合が存在する．また，単語によって手が互いに接触する場合や顔と接触する場合がある．

（2） 手話動作中に，無意識のうちに手の握り具合や

1999 年 11 月 26 日受付，2000 年 4 月 25 日最終受付，2000 年 5 月 15 日採録

† 松下電器産業株式会社 九州マルチメディアシステム研究所
（〒 820-0067 飯塚市川津 693-47, TEL 0948-21-2609）
†† 九州大学大学院 システム情報科学研究科
（〒 816-8580 春日市春日公園 6-1, TEL 092-583-7618）
††† 郵政省 通信総合研究所
（〒 184-0015 小金井市貫井北町 4-2-1, TEL 042-327-6229）
† Kyushu Multimedia Systems Research Laboratory, Matsushita Electric Industrial Co., ltd.
(693-47 Kawazu, Iizuka-shi, Fukuoka 820-0067, Japan)
†† Graduate School of Information Science and Electrical Engineering, Kyushu University
(6-1, Kasuga-Koen, Kasuga-shi, Fukuoka 816-8580, Japan)
††† Communications Research Laboratory
(4-2-1, Nukui-Kitamachi, Koganei-shi, Tokyo 184-8795, Japan)

映像情報メディア学会誌 Vol. 54, No. 6, pp. 848～857 (2000)

NII-Electronic Library Service

The Institute of Image Information and Television Engineers

指の張りの違いによって手形状が微妙に変化したり，大局動作（腕の振り等）のぶれにより手の方向が変わることがある．その結果，カメラで捉えた手の「見え」が変わってしまう．特に両手動作の場合，同一単語においても，手が接触する場合としない場合がある．

本論文では，上記課題に対処した局所特徴抽出およびそれを用いた手話認識手法を提案する．我々の手話認識システムでは，認識対象動作が与えられた場合，まず大局特徴によって意味の候補を求め，局所特徴によって候補を絞り込むことを目指している．本システムにおける局所特徴認識では，「見え」による認識方式に基づき，同一大局動作を持つ手話単語の，開始と終了時の手領域の画像（手画像）をもとに局所特徴を定義する．その場合，手の「見え」のぶれに対処するために，1単語あたり複数回の動作を撮影して得た手画像群を，固有空間内で似た「見え」の画像を同一クラスタとする形でクラスタリングする．そして，学習・認識双方において，クラスタリング結果に従い，手画像と各クラスタとの距離を求めることにより手話を認識する．

以降，2章では，手話言語学における手話の音韻表記について紹介し，3章では，我々のシステムの概要を述べる．そして，4章において，本論文の主題である局所特徴認識手法について記述し，5章の実験によって有効性を検証する．

2. 手話の音韻表記

手話言語学における研究[1]によると，手話は，分節で区切られた連続体として表現される．1分節はほぼ単語に相当し，各分節は，手の動き，位置，向き，手型，非手指動作といった属性に従い記述される．動きについては手が動いた方向・軌跡に基づき，位置については身体部位に対する手の位置に従い記述される．そして，手の向きについては手の軸の方向と掌の向きに従い，手型については，手話で使われる手型を有限個数に分類し，各手型にそれぞれシンボルが割り当てることにより記述される．なお，手以外の表情や体の動きが非手指動作として表現される．

以上の属性は，2つのカテゴリーに分類することができる．1つが大局特徴でありもう1つが局所特徴である．大局特徴とは，腕や体全体の動きによって現れる特徴であり，手の動き，位置，非手指動作がこれにあたる．局所特徴とは，手首から先の姿勢による特徴であり，手の向きと手型がこれにあたる．なお，局所特徴については，分節の開始時と終了時の情報が特に重要であり，その間の手型，姿勢の変化については単なる遷移情報として記述される．

聾者が手話を認識する場合，主にこの大局特徴をもとに読み取り，局所特徴を用いて意味を確実にするといわれている．これは，手型を一定の型（拳型）に固定した手話を用いた場合の聾者による手話理解において，語彙レベル，文脈レベル双方に対し予想されるより理解度が高いという事実[10]や，時間・空間解像度が異なる手話映像の手話知覚において，映像の時間解像度が空間解像度よりも優勢であるという実験結果[11]からも容易に想像できることである．

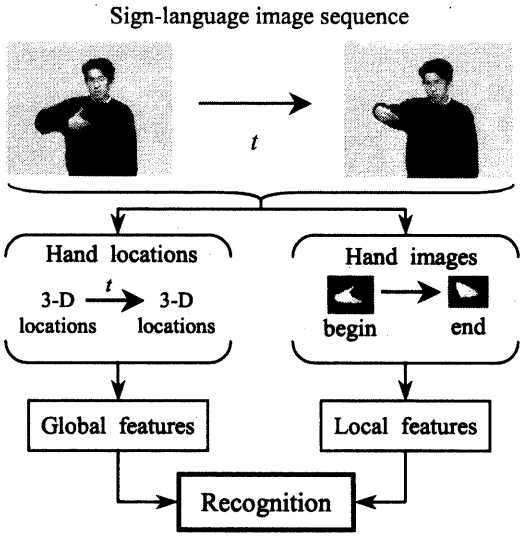

図 1 手話認識方式の概要
Overview of our system architecture.

3. 手話認識システムの概要

手話認識システムを構築する場合，前述の手話言語学で培われた知見に基づき構築することが重要である．ここでは，我々の手話認識システムの概要について記述する．**図1**に，我々の目標とする手話認識システムの概念図を示す．本システムは，カメラにより非接触で大局特徴と局所特徴の両方を抽出し手話を認識することを目標としている．

まず，カメラで撮影した手話の画像から手領域を切り出し，局所的な情報（手の周辺画像）と大局的な情報（手領域の位置）とに分離する．手話画像から手の周辺画像を切り出すために，我々は，肌色情報を用いて手と顔領域を抽出して2値化，ラベリングを行い，手領域を抽出，追跡する[12]．

一般に，肌色情報は周囲の照明条件や個人差によって変動するため，手話認識を行う場所や対象者に応じて調整する必要がある．通常，手話認識システムは屋内に設置されると予想されるため，1回の手話動作撮影時間内に照明等の撮影条件の変動は少ないと考えられる．そこで，色空間を表現する3次元ルックアップテーブル（3-D LUT）を用意し，認識前に対象者の画像を取得して顔領域を求め，求めた領域の色を肌色としてLUTに設定する．一般に，手よりも顔の方が，変形が少なく目や口等の部位が存在し，色以外の情報による検出・追跡が容易

NII-Electronic Library Service

The Institute of Image Information and Television Engineers

であると考えられるため，対象者の顔をまず検出することにより周囲の照明条件や個人差に対処した肌色情報を取得することが可能になると考えられる．なお，本論文では，手話の局所特徴を用いた認識が主題であるため，抽出すべき手・顔領域を手作業で指定することにより求めた色領域を肌色としてLUTに設定する．

次に，大局的な情報から手の位置，動きといった大局特徴を求める．また，局所的な情報から，手型や手の向きに関する局所特徴を求める．そして，大局・局所双方の特徴を用いて手話認識を行う．その場合に，前述の聾者に関する知見に基づき，最初に大局特徴に従い認識を行い，局所特徴は同一大局特徴を持つ分節の識別のために利用する．なお，手話の場合，1分節はほぼ単語に該当するため，以降では，「分節」を単に「単語」と表現する．

大局特徴による手話認識においては，複眼式のカメラによって取得した手の重心位置の3次元軌跡を，上，下，前，後，斜め方向といった45°刻みの方向に量子化し，時間方向に対しても同じ方向の動作が続く区間を圧縮して同一のコードとすることにより時間方向の動きのずれを吸収する．また，手の位置に関しても身体部位との位置関係を利用し，体格差を吸収する形で正規化することにより認識を行う．現在までに筆者らは，上記手法を用いて40単語の手話認識を行い，その有効性を示した[6)13)]．

さて，手話動作からの認識の場合，動作を単語や渡り（手話単語間の遷移動作）の区間に分割する「セグメンテーション」が重要であるといわれている[14)~16)]．一般に，データグローブ等の手袋型入力装置を用いる場合，手や指の動作が極小となる時点を分割位置としている[14)15)]．一方，画像からのセグメンテーションでは，オクルージョンなどにより手の動作が極小である時点を正確に検出することは難しい．そこで，筆者らは，聴覚障害者同士の自由対話における手話動作の画像の解析を行い，セグメンテーションの情報として，手の動作が極小であるという情報に加え，腹部より下の位置に手が存在するか否かという情報（腹部より下の場合無意味であり，その部位からの移動は渡りの区間である）や，両手同士の接触，顔との接触，顔のうなづきといった情報を検出することにより，手話動作のセグメンテーションが可能であることを示した[16)]．なお，本論文では，手話の局所特徴を用いた認識が主題であるため，1回の動作において1単語の動作を行い，その意味を認識する単語認識を対象とする．そして，動作から単語の区間を切り出すためのセグメンテーション位置を，手の3次元位置の変位がある一定距離より小さく停止して見える時点および両手が接触する時点とし，動作開始から単語の開始までと，単語の終了から動作終了までの渡りの区間を認識対象区間から排除する．

次に，本論文の主題である手話の局所特徴認識について述べる．前述したように，手話音韻論においては，各単語毎の開始・終了時の情報が特に重要であり，その間の手型，姿勢の変化については単なる遷移情報として記述される．また，人間の視覚的な認知特性を考えた場合，運動中の細かい手形状変化はあまり意味を持たないと考えられる．また，データグローブによる手話認識においても，運動による手部の移動形態と，その前後の静止時の手形状が大きな意味を持つとし，運動の前後の手の関節角度データを手形状の特徴（局所特徴）として認識している[17)]．そこで，我々のシステムにおいてもセグメンテーションで得られた単語の開始・終了時の手画像から局所特徴を求める．

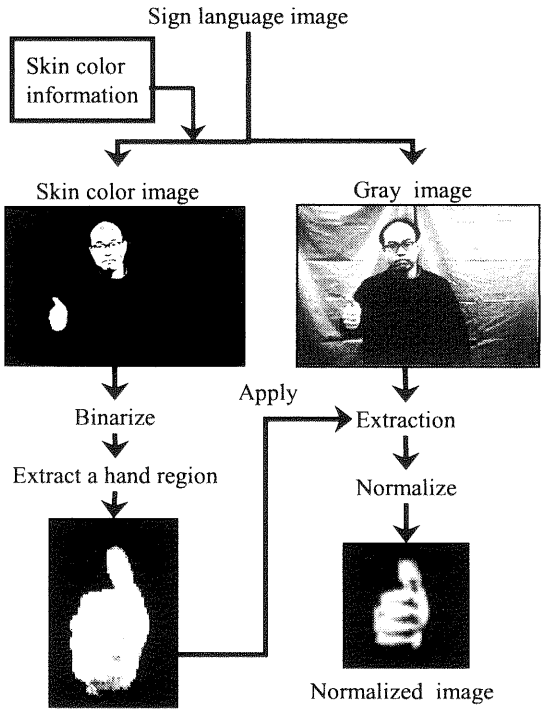

図2　手画像の取得
Acquiring a hand image.

4. 局所特徴認識

前述の手話認識システムの流れに従い手話動作がセグメンテーションされ，単語毎に同一大局特徴を持つ意味の候補が出力される．そして，局所特徴認識では，単語の開始・終了時の手画像をもとに認識を行う．

本手法は，学習と認識の2つのステージに分けられる．学習時には，認識対象とする手話単語を撮影し，開始・終了時の手画像を取得する．なお，撮影時には，一定の間接照明のもとで，ユーザの個人差および大局動作や手の形状・方向の無意識の変動から生じる手の見えの違いに対応できるように，複数人が1つの手話単語につき複数回動作を行う．次に，得られた画像群を用いて固有空間を作成し，これらの画像群を固有空間に投影し，似た見えの手画像を同一クラスタとする形でクラスタリング

NII-Electronic Library Service

The Institute of Image Information and Television Engineers

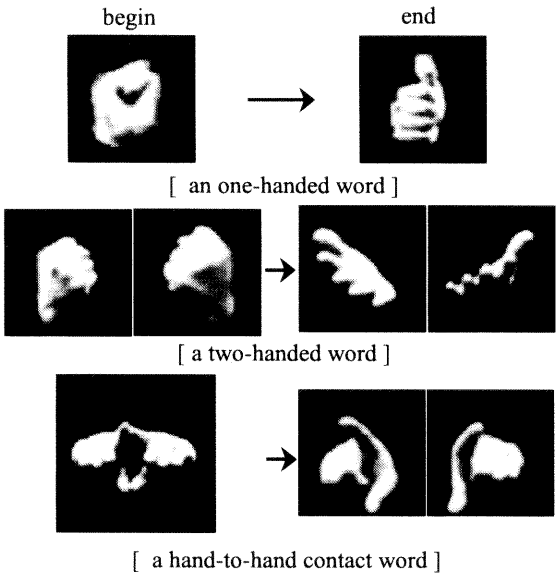

図 3 手話単語の開始・終了時における手画像例
The examples of hand images at the beginning and end of the word.

を行う．そして，クラスタリング結果に従い，学習動作から抽出した手画像から最尤クラスタを求め，その結果を学習する．また，認識時には，対象動作から学習時と同様の手法で手画像を抽出し，その手画像と学習パターンとして登録したクラスタとの距離を求めることにより手話を認識する．以下にそのアルゴリズムの詳細について順を追って説明する．

4.1 手画像の取得

撮影した単語の開始・終了時の手領域を切り出すことにより，局所特徴認識のための手画像を作成する．**図 2** に手画像の取得方法を示す．まず，先に述べたように，肌色情報を用いて手と顔領域を抽出して2値化，ラベリングを行い，手を追跡する[12]．次に，追跡結果を用いて，抽出した肌色領域から手の存在する領域を選び，その領域を外接矩形により切り出す．一方，同一時点の手話画像を白黒に変換する．そして，切り出した2値化画像をマスクとし，白黒画像からマスクと同位置の領域を切り出し，マスクを重ね合わせることにより，背景の影響を除去した白黒の手画像を取得する．次に，求めた手画像が一定の大きさになるように拡大縮小を行い，画像内での平均輝度と分散が一定になるように手画像を正規化する．

さて，手話の場合，片手で行う場合（片手手話）と両手で行う場合（両手手話）が存在する．特に，両手手話の場合，両手が接触する場合がある．**図 3** に各場合に該当する手話単語の開始・終了時における手画像例を示す．手が接触する場合，双方の手が複雑に絡み合っている場合もあるため，画像中の両手を分離することは不可能である．なお，大局特徴からある時点における手の接触の有無を判別することは可能である．そこで，手画像を作成する場合に，片手画像と両手接触画像とに分類し，片手画像の場合は 64×64 画素に，両手接触画像の場合は 80×80 画素に正規化し，各手画像群に対して，それぞれ個別に処理を行う．

4.2 特徴空間の作成

手画像を表現する特徴空間として，KL展開に従い手画像群の分布を最良近似する部分空間（固有空間）を求める．今，$m \times n$ 画素の画像 $\{f(i,j) \mid 0 \leqq i < m, 0 \leqq j < n\}$ をラスタ走査して得たベクトルを$\boldsymbol{U}$とし，p 個の画像が属する画像群を

$$\{\boldsymbol{U}_1, \boldsymbol{U}_2, \boldsymbol{U}_3, \ldots, \boldsymbol{U}_p\}$$

で表現する．

次に，この画像群の平均ベクトル$\boldsymbol{c}$を求め，画像群の要素$\boldsymbol{U}_i$から平均ベクトルを引いたものを列ベクトルとする行列を$\boldsymbol{X}$とする．そして，$\boldsymbol{X}$の共分散行列$\boldsymbol{Q}$を計算し，固有方程式

$$\boldsymbol{\lambda}_i = \boldsymbol{Q}\boldsymbol{e}_i \tag{1}$$

を解く．そして，上位 k 個の固有値 $(\lambda_1 \geqq \ldots \geqq \lambda_k)$ に対応する固有ベクトル $(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_k)$ を基底ベクトルとする k 次元の部分空間を特徴空間とする．そして，ある手画像 x が与えられた場合

$$\boldsymbol{u}_x = [\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_k]^T (\boldsymbol{U}_x - \boldsymbol{c}) \tag{2}$$

により求めたベクトル$\boldsymbol{u}_x$を特徴ベクトルとする．

4.3 識別部の設計

手画像群を固有空間に投影し特徴ベクトルを求めた後，特徴ベクトルの分布に従い，「見え」が似ている画像が同一クラスになるように手画像群のクラスタリングを行い識別部を設計する．さて，一般にクラスタリングを行う場合，最終クラスタの数（クラス数）を定義する必要がある．局所特徴認識では，単語の開始，終了時の手画像により識別部を設計するため，クラス数は認識対象とする単語の個数により，ある程度決定される．しかしながら，実際には手の「見え」には変動や個人差が存在するため，学習させる同一単語の手画像を同一クラスではなく，別のクラスにする方が良い場合がある．また，開始，終了のいずれかの時点で，同一の手型と向きを示す別単語の例もあり，そのような場合には，同一のクラスにする方が望ましい．

そこで，クラスタリングアルゴリズムとして，非階層クラスタリングを代表する手法であるISODATAを用いる[18]．ISODATAは教師なしクラスタリングの一種であり，K平均法によるクラスタリングに加え，最適なクラス数になるようにクラスタの分割と統合を行う手続きからなる．本識別部では，認識対象単語の開始・終了時の手画像から，左右の手に対し各1画像ずつ取り出して初期クラスタとし，K平均法によるクラスタリングを行う．

NII-Electronic Library Service

The Institute of Image Information and Television Engineers

表 1　識別辞書例
The example of the classification dictionary.

単語	動作番号	クラスタ	
		開始	終了
A	1	h_1	h_{10}
	2	h_1	h_{10}
	3	h_2	h_{10}
B	1	h_{c1}	h_1, h_2
	2	h_{c1}	h_1, h_3
	3	h_{c0}	h_2, h_3

これにより,「クラス数は認識対象とする単語の個数により，ある程度決定される」という点が満たされる．次に，クラスタの分割と統合については，このクラスタの集合から構成される空間が多次元正規分布と仮定し，赤池の情報量基準 (AIC) に従いクラスタの分割・統合を制御する手法[19]を用いる．そして，この手法に従いクラスタリングを制御することにより，前述の最適なクラス数になるようにクラスタの分割・統合を繰り返す．

クラスタリングにおける距離尺度および AIC は次のように定義する．まず，クラスタ $h_i, 1 \leq i \leq c$ と特徴ベクトル$\boldsymbol{u}_x, 1 \leq x \leq p$ との間の距離尺度 D は，クラスタ h_iの共分散行列Σ_iおよび平均ベクトル$\boldsymbol{\mu}_i$を用いて

$$D(\boldsymbol{u}_x, h_i) = -\frac{1}{2}\ln|\Sigma_i| - \frac{1}{2}\chi^2(\boldsymbol{u}_x; \boldsymbol{\mu}_i, \Sigma_i) \qquad (3)$$

と定義する．ただし，$\chi(\cdot)$ はマハラノビス距離を示す．また，AIC の評価として，c 個のクラスタのパラメータ数 $m(c)$ および最大尤度の総和 $L[\hat{\psi}_c]$ を用いて

$$AIC = -2\ln(L[\hat{\psi}_c]) + 2m(c) \qquad (4)$$

と定義する．新たな単語が入力された場合，その手画像を固有空間へ投影し，(3) 式を用いることにより，最尤クラスタを決定する．以上の結果，手話単語の開始・終了時の手画像からその手画像に該当するクラスタが導出される．

4.4　学習と認識

本局所特徴認識では，対象動作の開始・終了時の手画像から導出したクラスタをもとに学習，認識を行う．

学習時には，学習対象の手話動作から得た手画像がどのクラスタに該当するかといった情報を識別辞書に登録する．**表 1** に識別辞書の例を示す．辞書には，各単語に対し，学習動作数（この場合 3 動作）だけ，各動作から導出した該当クラスタが登録されている．

表より,「単語 A」は片手手話の例である．この場合，片手のため，開始，終了において，各々 1 つのクラスタが登録されている．なお，3 動作のうち 1 動作において，開始時に手画像の変動が見られ，違うクラスタに割り当てられている．次に,「単語 B」は両手手話の例であり，両手が最初に接触して離反する動作を示す．この場合，開始時は両手接触画像であり，両手接触画像群から導出したクラスタ h_{c0}，h_{c1}が登録されている．また，終了時

表 2　大局特徴による単語分類規則
Word classification rules by global features.

項目	分類規則
大局的静止	身体・腕の動きが静止しているか否か．
手の数	動作する手の数（片手又は両手）．
手の接触	両手手話において，手の接触があるか否か．
動き方向	片手の場合，空間を 45 度刻み区切った場合の定性的な方向． 両手の場合，両手の相対関係（開く・閉じる等）．
手の位置	身体部位との関係により定性的に空間分割．

表 3　実験に使用したカテゴリー
Selected categories used in the experiment.

カテゴリ	大局動作	単語数	単語例
CR0	片手,（下）前動作	13	“言う”, “好き” ...
CR1	両手，下動作	8	“今日”, “重い” ...
CR2	両手接触から 横に開く動作	12	“普通”, “大きい”, “多い” ...

には，両手が別々の手画像となるため，それぞれの手に対するクラスタが登録されている．

一方，認識時には，認識動作 R から得られた開始・終了時の各手画像を固有空間へ展開し特徴ベクトルを得る．そして，識別辞書に登録した各登録動作 T に対し，開始・終了それぞれにおいて該当クラスタと特徴ベクトルとの距離を (3) 式より求め，その総和を登録動作との距離 $D_M(R,T)$ とする．そして，各単語に対し，登録動作との距離 $D_M(R,T)$ の平均を，その単語と認識動作との距離と定義し，この距離が最も近い単語を識別結果として出力する．

4.5　手と顔が接触する単語における学習と認識

次に，手と顔が接触して見える場合の手話単語における学習と認識について述べる．手話単語の中には，手と顔が接触する単語や顔の前で手を動作させる単語がある．そのような単語から手画像を取得する場合，肌色抽出画像を用いても顔と手を分離できない．そこで，次のようにして学習・認識を行う．

まず，学習時には，対象となる手話単語の開始・終了時の手画像を，顔と分離した形で取得する．前述のように，各手画像は実際の動作を撮影した映像から取得する．そのため，なんらかの形で顔と手を分離して手画像を取得する必要がある．この課題を解決する手段として，手に肌色とは別の色手袋を装着して手画像を取得し学習する手法，顔と手が混在する画像から手作業で手のみを切り出す手法等が考えられる．なお，今回は，本手法の顔の前での認識における有効性および課題抽出を第 1 の目的とするため，学習時に理想的な手画像を得ることが望ましい．そこで，学習者には，顔に布等を被せ顔を隠した形で手話動作をしてもらい，単語の開始・終了の手画像を取得し学習する．同時に，学習した開始・終了の手画像のうち，本来であれば顔の前にあると思われる手画像を選び，その画像に対する肌色抽出画像の OR を求めマスクを作成する．

一方，認識時には，顔と手が混在する形で手画像を取

NII-Electronic Library Service

The Institute of Image Information and Television Engineers

表 4　取得した手画像群
Hand images acquired in each category.

カテゴリー	手画像の種類	画像数	平均サイズ	正規化サイズ
CR0	片手	1950	72 × 72	64 × 64
CR1	片手	2296	72 × 64	64 × 64
CR2	片手	1816	72 × 64	64 × 64
	両手接触	872	104 × 68	80 × 80

得し，取得した領域をラスタ走査しながら作成したマスクを適用して領域を抽出する．そして，抽出した領域を正規化し，固有空間へ展開して該当クラスタと尤度を求め，最も尤度が高い位置を手の位置とし，その時のクラスタおよび尤度に従い単語認識を行う．

5. 実　験

5.1　単語認識実験

（1）　実験方法

本手法の有効性を検証するために，手話単語を撮影し局所特徴による認識実験を行った．なお，認識対象の手話単語は次のように定義した．

我々は，本システムが将来的に聾者と健常者の間の対話システムの一部として統合され，公共の窓口のような限定した状況の下で使われると仮定している[20]．限定した状況の下では，対象とする語彙の数が制限される．本実験では，郵便局窓口での対話を想定し，郵便局窓口の会話の解析に基づき対話で用いられる単語を抽出した．その結果，約160単語が抽出された．

さて，筆者らの提案する手話認識システムでは，最初に大局特徴により認識を行い，局所特徴は同一大局特徴を持つ単語の識別のために利用することを前提にしている．そのため，局所特徴認識の有効性を検証するためには，大局特徴だけでは識別不可能な手話単語に対し，局所特徴を用いることによって識別可能かどうかを検証することが必要である．そこで，手話単語映像データベース[21]を用いて該当単語の動作を観察し，**表 2** に示す大局特徴による単語分類規則に従い，人手により該当単語を同一大局動作の単語が同一カテゴリーに分類されるよう分類した．まず，片手手話，両手手話（接触なし），両手手話（接触あり）の大きく3つの様相に分類した．そして，それぞれの様相に対して「動き方向」,「手の位置」に関しては，大局動作における動作のぶれを考慮し，近傍の動作・位置まで含む形で分類した．以上の結果，片手手話，両手手話（接触なし），両手手話（接触あり）に対し，それぞれ，10カテゴリー，6カテゴリー，6カテゴリーに分類された．そこで，片手手話，両手手話（接触なし），両手手話（接触あり）のそれぞれに対し，最も単語数が多いカテゴリーを1つずつ選択し，そのカテゴリーに属する単語に対して認識実験を行った．**表 3** に選択したカテゴリーの概要を示す．また，大局特徴だけでは識別不可能な手話単語例として，CR2に分類された

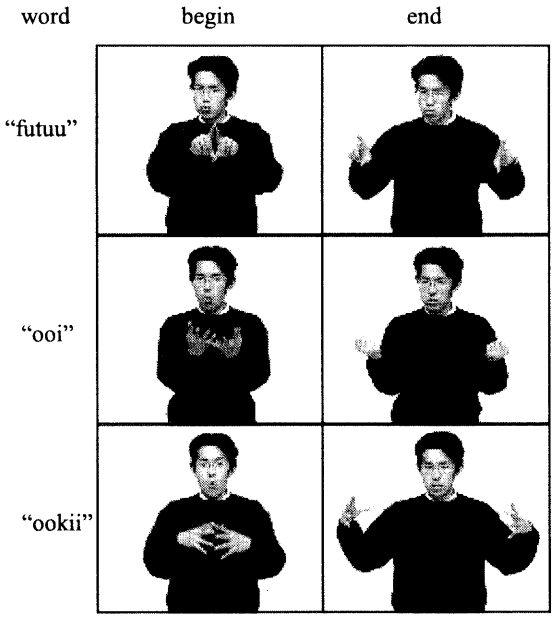

図 4　同一大局動作の手話単語例
The examples of the words which have same global features.

「普通」,「大きい」,「多い」という意味の手話単語を**図 4** に示す．

次に，対象データの取得のための被験者を決定した．一般に，手話に関する認識実験の場合，聾者によるデータを取得することが理想である．しかしながら，本実験においては単語認識ということもあり，手話というよりはジェスチャの認識実験に近い．そのため，手話の理解度の違いによる認識結果への影響については考慮しなくてよいと思われる．被験者として実際の聾者を用意するのは困難であるため，今回は健聴者を被験者として用意し，動作する際に対象単語の映像を見てもらいながら撮影した．また，簡便に動作の開始時と終了時のセグメンテーションを行うため，各被験者は単語の最初と最後に一定時間手を静止させた．なお，撮影環境は，通常の蛍光灯のもとで人物の上半身(腰から上)が画角に入る環境とし，取得した画像は，システムの構成上480 × 320画素の8 bit画像とした．

対象データとして，6名の被験者による1単語あたり10～12サンプルの手話画像から，開始・終了時の手画像を取得した．**図 5** に各カテゴリーにおける認識対象単語の手画像例を示す．認識実験として，学習と認識で用いる被験者が同一の場合と，異なる場合の2通りの実験を行った．学習と認識で同一の被験者を用いる場合，6名の被験者によるサンプルを半分に分割し，一方を学習に，もう一方を認識実験に使用した．そして，学習と認識のサンプルを入れ替えて同一の実験を行った．そのため学習と認識に用いたサンプルは重複しない．また，学習と認識に用いる被験者が異なる場合には，6名の被験者のうち，5名のサンプルを用いて学習し，認識時には残り

NII-Electronic Library Service

The Institute of Image Information and Television Engineers

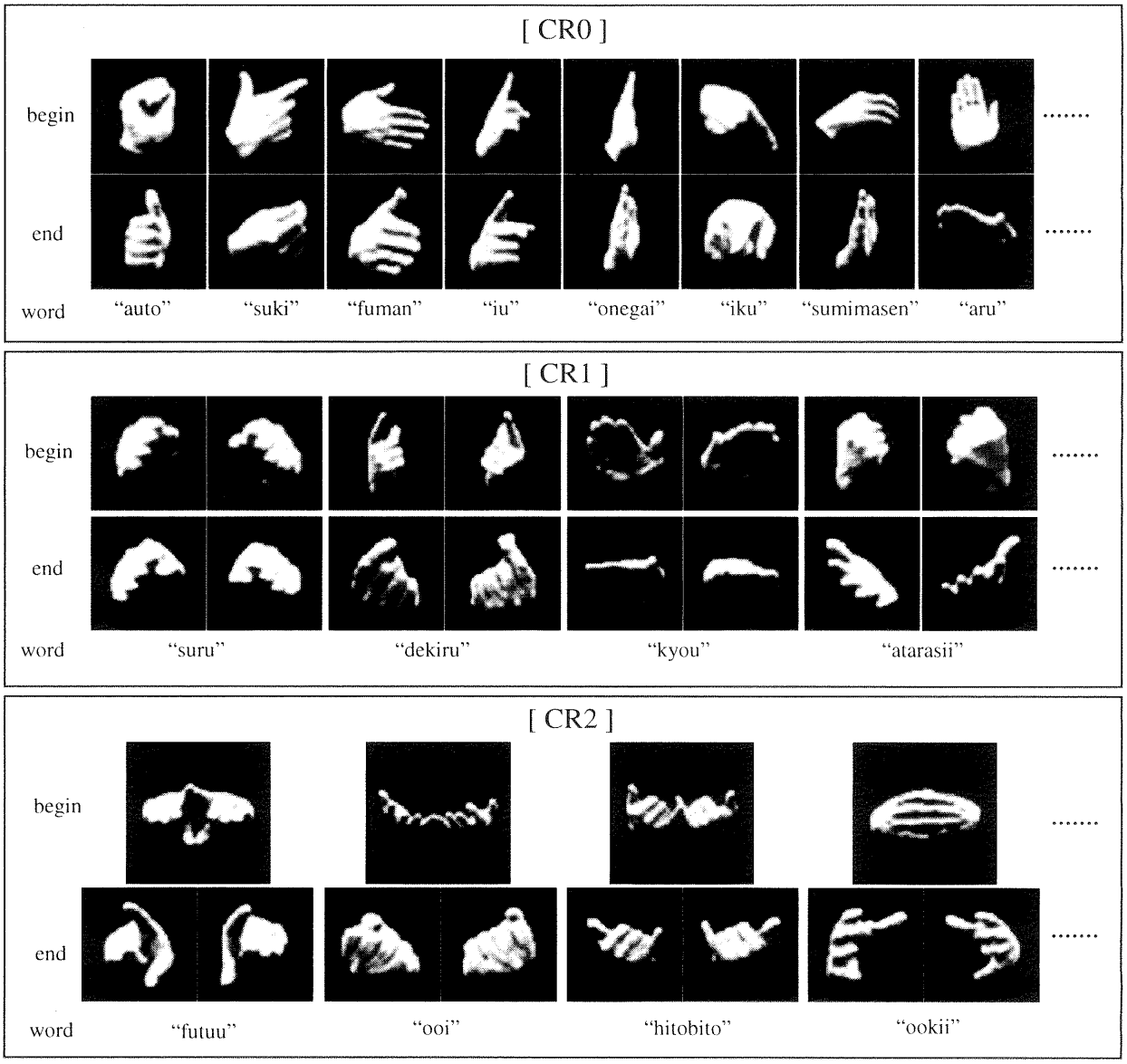

図 5 各カテゴリーにおける認識対象単語の手画像例
The examples of hand images at the beginning and end of words in each category.

1名のサンプルを用いた．そして，認識する被験者を変えながら6名分認識実験を行った．

各カテゴリーにおいて取得した手画像は**表 4** のようになった．CR0 の場合，片手手話のため，取得サンプルの2倍の数となる．CR1 の場合，両手手話（接触なし）のため，取得サンプルの4倍の数となる．CR2 の場合，開始時に両手接触があり，終了時に両手が離れている手話のため，片手画像と両手接触画像の両方が混在する．また，両手接触手話の場合，手の動きの無意識な変動により，手を完全に接触させなかった場合が 18 サンプル存在した（**図 6** 参照）．その結果，得られた手画像群は片手画像で 1816 個，両手接触画像で 872 個となった．それぞれの画像に対し，片手画像の場合 64 × 64 画素に，両手接触画像の場合は 80 × 80 画素に正規化した．なお，各カテゴリーの手画像群における固有空間次元については，実験結果を見ながら最も効率良く手画像を識別する次元数を決定した．結果として，今回の実験では各手画像群において 10 次元となった．

なお，本実験では，手の「見え」の変動への対処，および片手，両手および両手接触動作に対する有効性の検証を主眼としている．そこで，顔の前で手を動作させる手話においては，顔に布等を被ることによって取得した手のみが存在する手画像を学習・認識の双方のデータとして用いた．

（2） 実験結果と考察

表 5 に学習と認識で同じ被験者の場合における認識実験結果を示す．表より，すべてのカテゴリーにおいて，認識率が 93〜94% と高い値であり，各カテゴリー間の認識率に差異がないことがわかる．以上のことから，本手法により，手話のための局所特徴認識が充分可能である

NII-Electronic Library Service

The Institute of Image Information and Television Engineers

表 5 認識実験結果：学習と認識で同じ被験者の場合
Recognition rates by trainers.

カテゴリー	サンプル数	正解数	認識率
CR0	975	917	94%
CR1	574	534	93%
CR2	890	836	94%

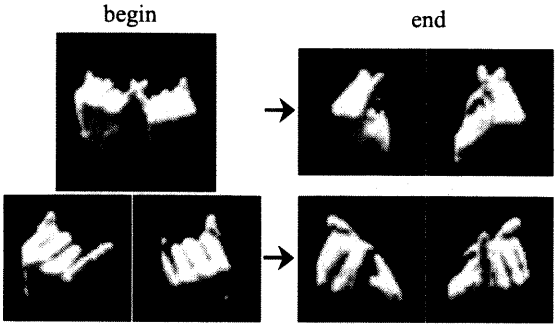

図 6 動作の変動により手が接触する場合と接触しない場合がある同一単語例
The example of the word which has involuntary changes of hand contact or non-contact.

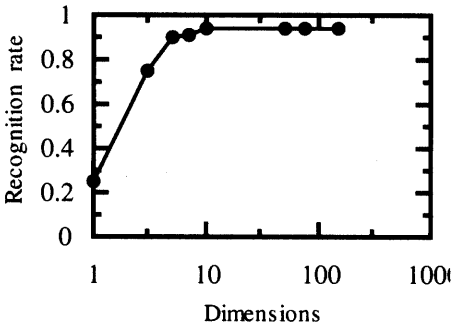

図 7 CR0 における固有空間次元と認識率の関係
Relation between eigenspace dimension and recognition rate in CR0.

ことがわかる．また，各カテゴリーに属する単語は，大局動作が近いため大局特徴のみでは識別が難しい単語である．したがって，従来技術である大局特徴のみで認識する手法では識別が難しい単語においても，本手法を付加することにより識別が可能になる．また，本実験で認識対象としたカテゴリーは，それぞれ，片手手話，両手手話（接触なし），両手手話（接触あり）の場合である．実験結果より，各カテゴリー間で認識率に差異がなく，片手・両手の違い，接触の有無にかかわらず同等の認識率であることから，従来技術では対象としなかった両手動作，両手接触動作の場合においても，片手動作の場合と全く同一の手法により認識することが可能になる．このことにより，例えば，図 6 に示したような手の動きの無意識な変動により，両手を接触する場合と接触させない場合の両方が存在するような場合においても，それらを区別することなく同一の手法で認識することが可能になる．

図 7 に，CR0 における固有空間次元を変えた場合の

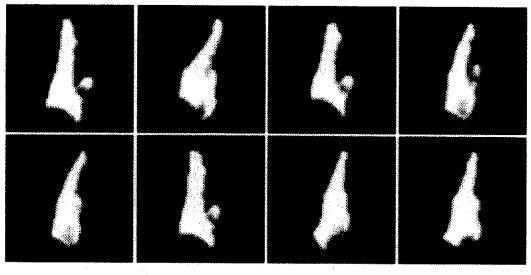

図 8 単語「お願い」の開始時の手画像例
The examples of hand images at the beginning of a word 'onegai'.

表 6 単純パターンマッチングによる認識実験結果
Recognition rates by the simple pattern matching.

カテゴリー	サンプル数	正解数	認識率
CR0	975	878	90%
CR1	574	499	87%
CR2	890	734	83%

認識率の遷移を示す．このグラフより，5 次元で 90%程度の認識率となり，10 次元を超えるとほとんど認識率に変化がないことがわかる．このことから，固有空間の計算において，手画像を識別するために必要な特徴が効率良く圧縮されていることがわかる．また，このように 10 次元程度の固有空間に投影することにより手画像の認識を行うことができることから，本手法は実時間処理においても有効な手法であることがわかる．

次に，「1 章．まえがき」で述べた 2 つめの課題である，大局動作や手の形状・方向の無意識の変動から生じる手の「見え」の違いへの対処について検証する．例えば，手の「見え」の変動を考慮しない場合の従来手法として，学習サンプルに対し，各単語の開始・終了時の手画像の平均画像を求め，その平均画像とのテンプレートマッチングにより認識することが考えられる．**表 6** に従来手法による単語認識結果を示す．この結果，すべてのカテゴリーにおいて，我々の手法の結果の方が 5%以上高かった．これは，本手法の場合，「4.3 節．識別部の設計」に示したように，最適なクラス数になるようにクラスタリングを行ったことによると考えられる．**図 8** にカテゴリー CR0 に属する単語「お願い」の開始時における手画像例を示す．このように，同じ単語の開始時においても，親指の位置や手の方向などによって，手の「見え」にぶれが生じる．そのため，単純なパターンマッチングの場合，別の単語の手画像と誤認してしまう場合がある．この場合，親指が出ているような画像では，単語「言う」の開始時の画像と誤って認識してしまう．しかしながら，我々の手法のように，「見え」によるクラスタリングを用いることで，このような手の「見え」のぶれに対し，最適化されたクラスタリングがなされ結果として従来手法より高い認識率が達成される．

表 7 に，学習と認識を違う被験者で行った場合の認識

NII-Electronic Library Service

The Institute of Image Information and Television Engineers

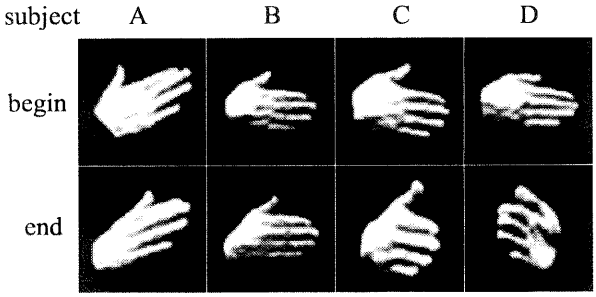

図 9 同一単語での個人毎の動作の違い
The difference of movements among individuals in the same word.

表 7 認識実験結果：学習と認識で違う被験者の場合
Recognition rates by a subject except trainers.

カテゴリ	サンプル数	正解数	認識率
CR0	975	683	70%
CR1	574	344	60%
CR2	890	543	61%

結果を示す．この場合，学習と認識で同じ被験者を用いる場合に比べ低い値となった．これは，手自体の個人差よりもむしろ，動作自体の個人差に起因する．例えば，腕の振りの大きさによる個人差は，結果的に手の向きが変わり，見えが変わってしまう結果となって現れる．また，人により同一の手話に対しても接触をする場合としない場合がある．提案手法においては，このような見えが変わる場合についても考慮した手法になっているが，それは，あらかじめ学習パターンとして近い見えの手画像が登録されている場合に限られる．そのため，今回のように学習者数が少ない場合，他の被験者が行う動作と同じ見えの画像を得ることは難しい．したがって，より汎用的な学習パターンを取得するために，学習者数を多くすることや，ある程度視点を変えた位置から撮影した手画像を取得する手段を用意する必要がある．

また，単語の開始・終了時における停止して見える時点の違いや接触・非接触の違いにより，セグメンテーション位置がずれる場合もある．例えば，**図 9** の場合，「不満」という単語の終了位置が，個人により大きくずれていることを示している．この場合，腕の振りの個人差によって生じるセグメンテーション位置のぶれを吸収する必要がある．今回の場合，本局所特徴認識では，手話言語学による知見に従い単語の開始・終了時の手画像から特徴を抽出し認識を行ったが，このような課題に対処するためには，今後，セグメンテーション位置を動的にずらしながらスポッティング認識を行うことも検討する必要がある．しかしながら，そのような場合においても本手法のポイントである，手画像の「見え」のぶれを吸収する形での特徴抽出においては，同等の手法により適用することが可能であると考えられる．

表 8 顔の前で動作する単語における認識実験結果
Recognition rates in words which a hand overlaps a face.

カテゴリー	単語数	サンプル数	認識率	
			顔あり	顔なし
CR0	5	360	72%	95%

5.2 手が顔の前で動作する単語の認識実験

（1） 実験方法

前述の認識実験では対象外とした，手が顔の前で動作する手話単語に対する認識実験を行った．開始もしくは終了時において手が顔の前にある手話は，CR0 に 5 単語存在する（「言う」，「お願い」，「すみません」，「明日」および「好き」の開始時）．そこで，これらの単語に対し，学習データとして，前述の単語認識実験で用いた 6 名の被験者による 1 単語あたり 10〜12 サンプルの手のみが存在する画像を用意した．そして，認識データとして，同じ 6 名の被験者による 1 単語あたり 10〜12 サンプルの手と顔が混在する画像を取得した．

（2） 実験結果と考察

顔の前で行う手話についての認識実験結果を**表 8** に示す．この場合，手の検出に成功し認識に成功した割合は，手のみを撮影した場合に比べ低く 72%であった．この結果に関し，**図 10** に，検出した手の位置を実際の画像上に矩形表示した結果を示す．(a) は認識に成功した例であり，(b) については認識に失敗した例である．この場合，影の影響により，検出位置を実際の手の位置より下と誤認識してしまい，認識結果としても手のみを撮影した場合とは違うクラスタに分類され，実際には (c) の形状のものを (d) と誤ってしまった．以上のことから，顔の前での検出および識別精度を向上させるためには，このような影の影響を消す必要がある．そのためにも，本手法により手の位置を探索する前に，より精度の高い手の位置の検出を行い，探索範囲を狭める必要がある．

なお，今回は，肌色情報と動き情報を利用して，手話動画像から手を追跡するシステムを用いて実験を行った[12]．そのシステムを用いると，顔の前であっても手のおおよその位置を追跡することが可能であるが，影の影響を消すまでには至らなかった．そこで，今後この課題を解決するために，本手法による手の検出と，色情報と動き情報を利用した追跡手法とを組合せ，より高い精度の手の追跡を行い，顔の前での認識率の向上を図っていきたい．

6. むすび

本論文では，我々の手話認識システムのアーキテクチャを示し，そのシステムで有効な手話における局所特徴認識手法について述べた．本手法は，学習時に同一大局動作を持つ手話分節の開始・終了時の手画像から固有空間を作成し，固有空間内で，手の「見え」の変動を考慮した形でクラスタリングを行う．そして，認識時には，

NII-Electronic Library Service

The Institute of Image Information and Television Engineers

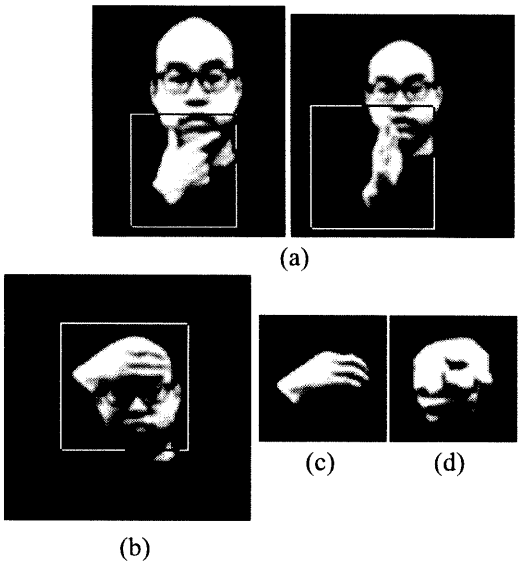

図 10 顔の前での手の検出と認識結果
Typical results of the locating and recognition when a hand overlaps a face.

認識対象の手画像を固有空間へ投影し，最尤法に従って該当クラスタを求めることにより手話を認識する．

郵便局の対話から抽出した同一大局動作の手話単語に対する認識実験の結果，手話の大局動作の違い（片手，両手，両手接触等）に関係なく高い認識率が得られ，大局特徴のみで認識する手法では識別が難しい単語においても識別が可能となることを示した．更に，大局動作の違い（片手，両手，両手接触等）に関係なく本手法が適用可能であることを示した．

また，手の「見え」の変動を考慮した局所特徴抽出を行うことにより，手の「見え」の変動を考慮しない場合に比べ認識率が向上することを確認し，固有空間法を用いることにより実時間での認識が可能となる見通しを得た．

今後は，顔の前にある手の検出および認識の精度を高め，更に，大局特徴と局所特徴の両方を認識する手話認識システムとして統合し，実際の聾者による認識実験を行う予定である．

〔文　献〕

1) 神田 和幸："手話学講義"，福村出版 (1994)
2) 高橋 勝彦，関 進，小島 浩，岡 隆一："ジェスチャ動画像のスポッティング認識"，信学論，**J77-D-II**，8，pp.1552-1561 (1994)
3) 大和 淳司，倉掛 正治，伴野 明，石井 健一郎："カテゴリー別 VQ を用いた HMM による動作認識法"，信学論，**J77-D-II**，7，pp.1311-1318 (1994)
4) 畠 直志，岩井 儀雄，谷内田 正彦："動き情報と情報圧縮を用いたロバストなジェスチャ認識手法"，信学論，**J81-D-II**，9，pp.1983-1992 (1998)
5) T. Starner, J. Weaver, and A. Pentland: "Real-Time American Sign Language Recognition Using Desk and Wearable Computer Based Video", IEEE Trans. Pattern Anal. & Mach. Intell., **20**, 12, pp.1371-1375 (1998)
6) H. Matsuo, S. Igi, S. Lu, Y. Nagashima, Y. Takata, and T. Teshima: "Recognition Algorithm with Non-contact for Japanese Sign-Language using Morphological Analysis", Gesture Workshop, Bielefeld, German, pp.273-284 (1997)
7) C. Uras, and A. Verri: "On the Recognition of the alphabet of the sign language through size functions", Proc. 12th IAPR Int. Conf. on Pattern Recognition, Vol.2, pp. 334-338 (1994)
8) 平川 幹和子，松尾 英明，高田 雄二，手島 光隆："シルエット画像を用いた手形状認識"，信学技報，PRU95-232，3，pp.123-130 (1996)
9) 渡辺 賢，岩井 儀雄，八木 康史，谷内田 正彦："カラーグローブを用いた指文字の認識"，信学論，**J80-D-II**，10，pp.2713-2722 (1997)
10) 神田 和幸："「どらえもん手話」の実例とその認識"，第 14 回ヒューマンインタフェースシンポジウム，pp.143-148 (Sep.1998)
11) 鎌田 一雄，平間 淳子，山下 真希："手話映像知覚に対する時間・空間解像度の寄与に関する基礎検討"，画電学誌，**28**，3，pp.256-263 (1999)
12) 今川 和幸，呂 山，猪木 誠二，松尾 英明："肌色領域により隠れて見える場合を考慮した手話動画像からの手の実時間追跡"，信学論，**J81-D-II**，8，pp.1787-1795 (1998)
13) 松尾 英明，高田 雄二，手島 光隆，平川 幹和子："非接触による手話動作の認識アルゴリズムの開発"，Human Interface News and Report，**10**，1，pp.41-46 (Feb.1995)
14) E. Ohira, H. Sagawa, T. Sakiyama, and M. Ohki: "A Segmentation Method for Sign Language Recognition", IEICE TRANS., **E78-D**, 1, pp.49-57 (1995)
15) 佐川 浩彦，竹内 勝："手話認識における手動作セグメンテーション方式"，ヒューマンインタフェースシンポジウム'99，pp.749-754 (Sep.1999)
16) 高田 雄二，武藤 大至，関 宜正，長嶋 祐二，呂 山，猪木 誠二，松尾 英明："手話認識のための手話動画像セグメンテーション手法の検討"，第 14 回ヒューマンインタフェースシンポジウム，pp.163-168 (Sep.1998)
17) 澤田 秀之，橋本 周司，松島 俊明："運動特徴と形状特徴に基づいたジェスチャ認識と手話認識への応用"，情処学論，**39**，5，pp.1325-1333 (1998)
18) 高木 幹雄，下田 陽久："画像解析ハンドブック"，東京大学出版会 (1991)
19) C. Carman, and M. Merickel: "Supervising ISODATA with an Information Theoretic Stopping Rule", Pattern Recognition, **23**, 12, pp. 185-197 (1990)
20) S. Igi, S. Lu, T. Ebina, K. Imagawa, H. Sakato, T. Urata, and M. Nishida: Telecommunication Technologies for the Disabled - towards a Dialogue System between Sign Language and Voice -", J. Communications Research Laboratory, **45**, Nos. 1/2, pp.53-59 Japan (1998)
21) 呂 山，猪木 誠二："手話単語映像データベースの開発"，信学技報，MVE98-86，pp.41-47 (1999)

今川　和幸 1989 年，九州大学工学部電子工学科卒業．1991 年，同大学院修士課程修了．同年，松下電器産業 (株) 入社．1996～1997 年，郵政省通信総合研究所特別研究員．現在，九州大学大学院システム情報科学研究科博士後期課程在学中．松下電器産業 (株) 九州マルチメディアシステム研究所勤務．

谷口倫一郎 1978 年，九州大学工学部情報工学科卒業．1980 年，同大学院工学研究科修士課程修了．同年，同大学院総合理工学研究科助手．1988 年，同助教授．1996 年，同大学院システム情報科学研究科教授．工学博士．正会員．

有田　大作 1992 年，京都大学工学部情報工学科卒業．1994 年，九州大学大学院総合理工学研究科修士課程修了．1998 年，同大学院システム情報科学研究科博士後期課程単位修得退学．同年，同助手．

松尾　英明 1986 年，九州大学工学部生産機械工学科卒業．同年，凸版印刷 (株) 総合研究所入社．1990 年，松下電器産業 (株) 入社．九州マルチメディアシステム研究所勤務．

呂　山 1984 年，中国重慶大学卒業．1988 年，京都大学大学院工学研究科電気工学専攻博士後期課程単位修得退学．同年，日本電気 (株) 入社．1996 年，郵政省通信総合研究所入所．工学博士．

猪木　誠二 1973 年，名古屋工業大学工学部計測工学科卒業．1975 年，同大大学院博士前期課程修了．同年，郵政省電波研究所 (現 通信総合研究所) 入所．1990 年，太陽電波研究室長．1995 年，ユニバーサル端末研究室長．

NII-Electronic Library Service