

制約相互部分空間法を用いたジェスチャー認識

Gesture Recognition using the Constrained Mutual Subspace Method

西山 正志 † 山口 修 † 福井 和広 †

Masashi NISHIYAMA †, Osamu YAMAGUCHI †, Kazuhiro FUKUI †

†(株) 東芝 研究開発センター

†TOSHIBA Corporate Research and Development Center

E-mail: masashi.nishiyama@toshiba.co.jp

Abstract

従来のジェスチャー認識で用いられてきた HMM や DP マッチングとは異なり、本稿では、動作パターンが特徴空間でなす分布の構造を比較することで認識を行う手法を提案する。時系列に獲得されたパターンの分布を部分空間として近似し、部分空間同士を比較することができる制約相互部分空間法を用いる。提案手法の有効性を確認するために、顔向き動作の識別と唇動作による単語識別の実験を行った。

1 はじめに

動画像を用いたジェスチャー認識は、人間が機械に意志を伝えるために必要な技術であり多くの研究がなされている [1]。ジェスチャー認識は、対象とする動作の特徴量を動画像から抽出する部分と、得られた特徴量を用いて識別を行う部分の二つからなる。

動作を表す特徴量の一つとして、身体の各部位の位置の時間的な変化量が考えられる。例えば顔向き動作を識別する場合、目などの顔部品位置の時間的な変化量の特徴量として用いることができる。また、唇動作から発話内容を識別する場合、唇の輪郭上の点の位置関係の変化を特徴量として用いることができる。しかし、身体の各部位の位置関係のみを特徴量とすると、抽出誤差などのノイズに対して敏感になり識別率の低下を招く恐れがある。そこで本稿では、身体の各部位を含む濃淡画像パターンを特徴量として用いることで安定な認識を行うことをねらう。

従来のジェスチャー認識の識別部には HMM (Hidden Markov Model) や DP (Dynamic Programming) マッチングなどが用いられてきたが、本稿では異なるアプローチによりシンプルな認識系を構成する手法を提案する。図 1 に示すように、時系列に獲得された入力パターン

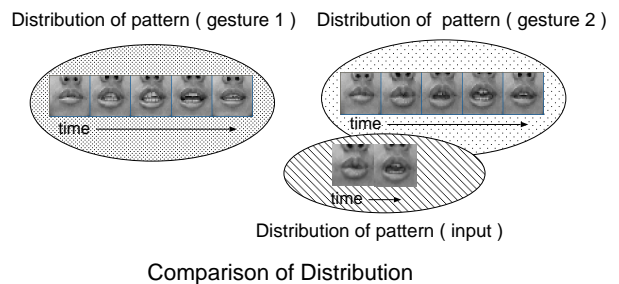


図 1 パターン分布を用いたジェスチャー認識の概念図

と辞書パターンによる、特徴空間上での分布構造を比較することでジェスチャー認識を行う。時系列に獲得されたパターンの分布構造を部分空間として近似し、部分空間同士を比較することができる制約相互部分空間法 [2] を適用する。これにより、分布同士の類似性が比較できるため、入力パターンと辞書パターンを個々に評価することなく認識を行うことができる。また、分布を部分空間として表現するため、抽出誤差などのノイズの影響を受けにくくなるという利点もある。さらに、本手法では設定すべきパラメータの数が少なく、簡便な計算で安定な認識系を構成することができる。

まず、2 章において、制約相互部分空間法をジェスチャー認識に適用する方法について概要を述べる。次に、提案手法の有効性を確認するために、顔向き動作について認識を行った結果を 3 章で、唇動作による単語認識を行った結果を 4 章で述べる。また、部分空間によるパターン分布の近似では、分布が同一であるがパターン間の順序関係が入れ換わることで別の動作が表される場合、識別することが困難となる。このような場合に対応するために、パターン間の順序関係を考慮した認識法について 5 章で述べる。

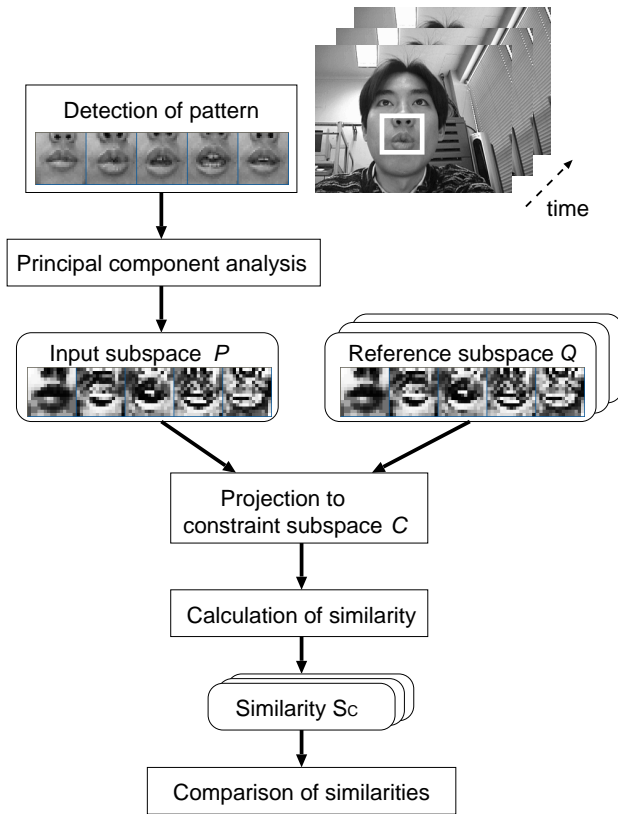


図2 制約相互部分空間法によるジェスチャー認識の流れ

2 制約部分空間を用いたジェスチャー認識

2.1 ジェスチャー認識の流れ

制約相互部分空間法を用いたジェスチャー認識の流れを図2に示す。まず、時系列に獲得した動画から濃淡画像のパターンを抽出する。パターンの抽出方法は認識対象の動作範囲によって異なり、本稿で対象とした顔向き動作は3.2節で、唇動作は4.2節で抽出方法を述べる。獲得されたパターンをベクトルに変換し、複数のベクトルを用いて主成分分析を行い入力部分空間を生成する。次に、入力部分空間と予め登録されている辞書部分空間を制約部分空間に射影することで特徴抽出を行い、入力部分空間と辞書部分空間のなす角度を測ることで類似度を求める。この類似度が最大になる辞書部分空間に対応する動作が入力された動作と判定する。2.2節より、それぞれの処理について詳細を述べる。

2.2 制約部分空間への射影

入力部分空間 P と辞書部分空間 Q から識別に有効な特徴を抽出するために、図3のように制約部分空間へ P, Q を射影する。制約部分空間は、識別に不用と考えられる個々の動作内のパターン変動は抑えつつ、識別に有効な動作間のパターン分布の違いを強調する空間

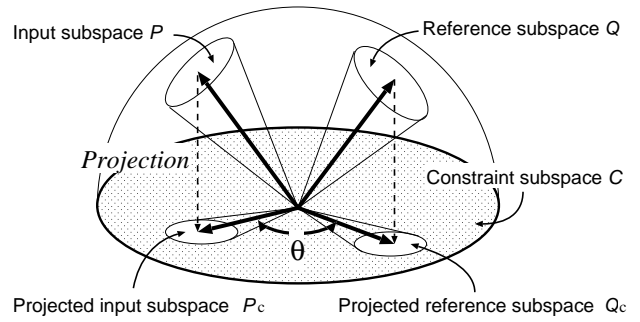


図3 制約相互部分空間法の概念図

である。制約部分空間の学習方法については2.4で述べる。ここでは、入力部分空間を制約部分空間に射影する方法について述べる。

- (i) 入力部分空間を張る N 個の基底ベクトルを制約部分空間上へ射影し射影ベクトルを求める。
- (ii) 各射影ベクトルの長さを正規化する。
- (iii) N 個の正規化ベクトルに対して Gram-Schmidt の直交化を施す。

直交化された N 個の正規化ベクトルが、制約部分空間へ射影された入力部分空間の基底ベクトルとなる。辞書部分空間も同様の手続きで制約部分空間へ射影する。

2.3 部分空間同士の類似度の算出

制約部分空間 C に射影された入力部分空間を P_C とし、辞書部分空間を Q_C とする。 P_C と Q_C 間の類似度 S_C は、正準角 [5] と呼ばれる二つの部分空間がなす角度 $\theta_i (i = 1 \dots N)$ から式 (1) で決定される。

$$S_C = \frac{1}{N} \sum_{l=1}^N \cos^2 \theta_l \quad (1)$$

ここで、 N は入力部分空間と辞書部分空間の次元数を表す。部分空間同士が完全に一致していれば $\theta_i = 0$ である。 $\cos^2 \theta_i$ は、以下の $N \times N$ の行列 \mathbf{X} の固有値 λ_i となる [5]。

$$\mathbf{X}\mathbf{a} = \lambda\mathbf{a} \quad (2)$$

$$\mathbf{X} = (x_{jk}), (j, k = 1 \sim N) \quad (3)$$

$$(x_{jk}) = \sum_{l=1}^N (\psi_j, \phi_l)(\phi_l, \psi_k) \quad (4)$$

$$\cos^2 \theta_i = \lambda_i (i = 1 \dots N, \lambda_1 \geq \dots \geq \lambda_N) \quad (5)$$

ここで、 ψ_j, ϕ_k は部分空間 P_C, Q_C の j, k 番目の正規直交基底ベクトルを表す。

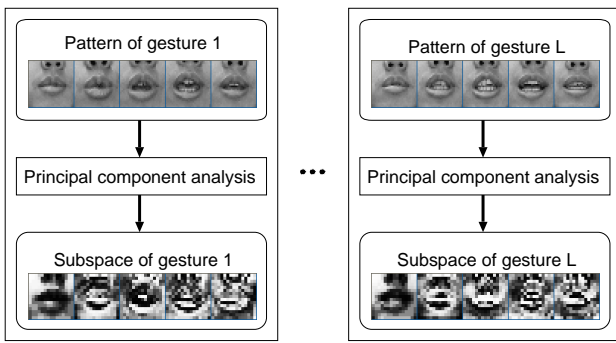


図 4 学習用部分空間の生成

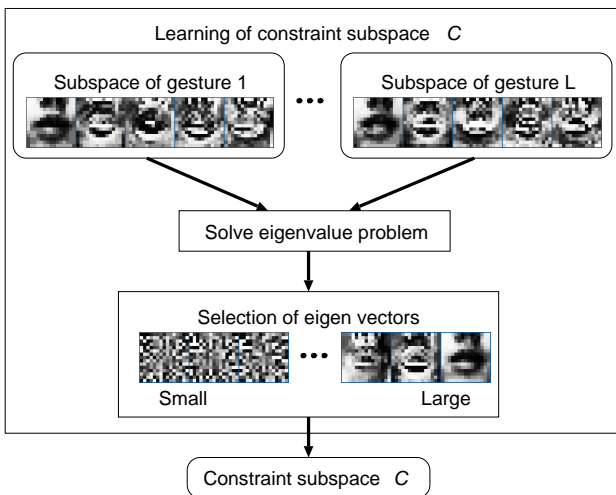


図 5 制約部分空間の学習の流れ

2.4 制約部分空間の学習

識別に有効な特徴抽出を行うための制約部分空間の学習方法について述べる．まず，図 4 のように， L 個の動作について，それぞれの動作のパターン変動をまとめて主成分分析を適用し学習用部分空間を用意する．制約部分空間の学習の流れを図 5 に示す．制約部分空間は， L 個の学習用部分空間が形成する和空間からこれらの主成分部分空間を取り除いた一般化差分部分空間を求めることで学習される [6]．学習用部分空間 $Y_i (i = 1 \dots L)$ の基底ベクトルを $\psi_{ij} (j = 1 \dots N_B)$ とすると，式 (6) の固有値問題を解くことで一般化差分部分空間を得ることができる．

$$(Y_1 + Y_2 + \dots + Y_L)a = \lambda a \quad (6)$$

$$Y_i = \sum_{j=1}^{N_B} \psi_{ij} \psi_{ij}^T \quad (7)$$

固有値が小さい順番で選んだ N_C 本の基底ベクトルで張られる空間が一般化差分部分空間となる．得られた一般化差分部分空間を制約部分空間として用いる．なお， N_B, N_C は実験的に定める．

3 顔向き動作の認識実験

3.1 内容

提案手法の有効性を確かめるために，顔向き動作を対象としてジェスチャー認識を行った．正面向きから上下左右のどちらに向いたかを識別することを目的とした．カメラで撮影した動画の一部を図 6 に示す．顔を向ける方向を画面で指示し，指示した瞬間前後に撮影した動画を用いた．具体的なアプリケーションはカメラを用いたゲームシステム「あっち向いてホイ」[7]などを想定した．

3.2 顔パターンの抽出

動画の中から識別に用いるパターンを得るために，我々がこれまでに提案している手法 [4] で顔領域を抽出した．まず，大まかな顔領域の候補を部分空間法を用いたテンプレートマッチングにより検出した．顔領域の候補に対して，特徴点検出法を適用することで瞳，鼻孔の 4 点を検出することで顔領域を改めて決定した．顔向きが大きく変化すると，隠れにより特徴点が見つからないため，テンプレートマッチングにより顔領域のトラッキングを行った．特徴点検出法により得られた顔領域を初期テンプレートとして，フレーム毎にテンプレートを更新した．図 7 に動画より顔領域を抽出した例を示す．縦方向が顔向きの上下左右を表し，横方向が時間経過を表す．獲得されたパターンには，正面向きのパターンが各方向に共通に含まれており誤認識を招く恐れがある．そこで，正面向きのパターンを検出し削除するために，部分空間法を適用した．正面向きのパターンのみから生成した部分空間と，正面向き以外のパターンから生成した部分空間を用意し，入力パターンとそれぞれの部分空間の類似度を求めることで正面向きであるかを判定した．

3.3 実験結果

この実験では，(a) 使用者本人のパターンが事前に登録できる場合と，(b) 本人のパターンが登録できず他人のパターンから生成した辞書部分空間を用いる場合について評価した．動画は顔向きを指示する瞬間を含む約 1 秒間で撮影した 20 枚の画像とした．(a) では撮影対象を 1 人として，辞書用の動画を各向きに 5 回ずつ合計 20 回，入力用の動画を各方向に 10 回ずつ合計 40 回とした．(b) では撮影対象を 10 人として，5 人で辞書用の動画を各向きに 1 回ずつ合計 20 回，残りの 5 人で入力用の動画を各方向に 3 回で合計 60 回の撮影を行った．パターンは 12×12 pixels で，ベクトルの次元数は 144 次元とした．前処理としてベクトルの L^2 ノルムの正規化を行った．

実験結果を表 1 に示す．表中の認識手法 COR では，正規化相関を用いた．入力パターンと辞書パターンの全組合せについて類似度を求め，その平均値を類似度



図 6 顔向き動作認識に用いる画像の一部

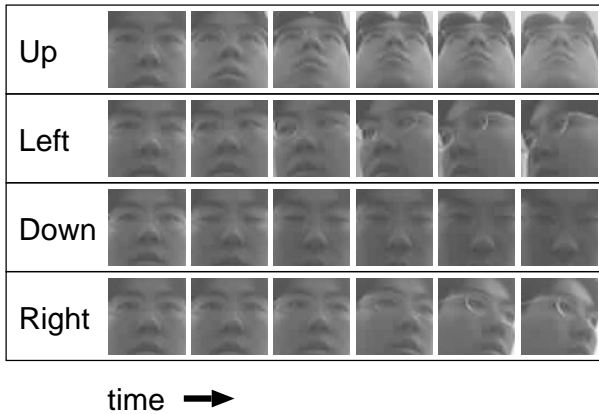


図 7 顔向き動作認識に用いるパターンの例

とした。

DP では、入力パターンと辞書パターンを図 8 のように時間順に並べて DP マッチングを適用した。辞書用の動画の数だけ DP マッチングを適用して類似度を算出し、その平均値を類似度とした。

CMSM では、制約相互部分空間法を用いた。入力部分空間と辞書部分空間は、動画から獲得したパターンをベクトルに変換したのちに、主成分分析により生成した。主成分分析を適用する際に、平均ベクトルを引かず求めた自己相関行列 [3] を用いる。 $N_C = 120$ 次元の制約部分空間で特徴抽出を行い、入力部分空間と辞書部分空間の次元数 N は 7 とした。制約部分空間を学習するために、辞書パターンから 4 動作の学習用部分空間を生成し、それぞれの学習用部分空間の $N_B = 40$ 本の基底ベクトルを用いた。

表 1 の結果より、制約相互部分空間法を用いた顔向き動作の認識は、他の手法と比較して有効であるといえる。処理時間は PentiumIII 1.0GHz の PC で画像の取り込みに要する時間を除いて約 100 ミリ秒であった。また、使用者本人のパターンが事前に登録できない場合 (b) においても、CMSM の正解率が高いことから、ヒューマンインターフェースにおける要素技術として制約相互部分空間法が活用できる可能性があるといえる。

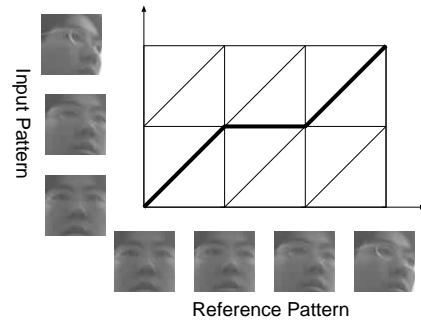


図 8 DP マッチングによる顔向き動作認識の例

表 1 顔向き動作の認識結果

辞書パターン	手法	正解率 (%)
(a) 本人	COR	92.5
	DP	90.0
	CMSM	97.5
(b) 他人	COR	71.7
	DP	85.0
	CMSM	96.7

4 唇動作の認識実験

4.1 内容

唇動作を対象として、発音された単語を動画から識別する実験を行った。識別の対象とする単語は、「うえ」、「した」、「ひだり」、「みぎ」、「はい」、「いいえ」の 6 個とした。カメラで撮影した動画の一部を図 9 に示す。単語を発音する前後は唇を閉じた状態とし、発話する瞬間の前後を撮影した。具体的なアプリケーションは、騒音が激しい環境や、美術館のように発声しにくい環境での音声認識の補助などを想定している。

4.2 唇パターンの抽出

識別に用いるパターンを得るために、[8] で述べられている手法で唇輪郭を抽出し唇領域を決定した。瞳、鼻孔の 4 点から大まかに推定した初期位置を与え、動的輪郭モデルを用いてエネルギー関数が最小になるように唇輪郭を求めた。図 10 に動画より唇領域のパターンを抽出した例を示す。縦方向が単語の種類を表し、横方向が時間経過を表す。また、唇を閉じた状態のパターンは各単語の発話に共通に含まれているために、3 章と同様に部分空間法を用いて削除した。ここでは、唇を閉じた状態のパターンのみから生成した部分空間と、唇を開いている状態のパターンから生成した部分空間を用いた。



図9 唇動作認識に用いる画像の一部

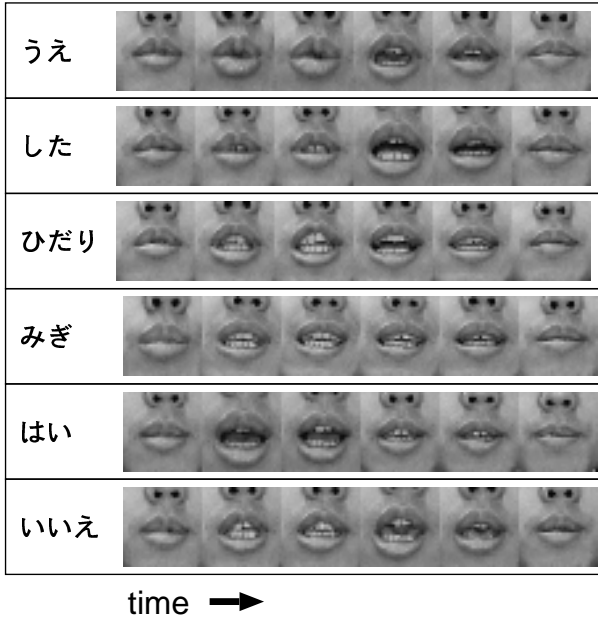


図10 唇動作認識に用いるパターンの例

4.3 実験結果

この実験では、使用者本人のパターンが事前に登録できる場合について評価した。動画は、単語の発話を指示した後の約 1.5 秒間で撮影した 45 枚の画像とした。撮影対象を 1 人として、辞書用の動画を各単語に 5 回ずつ合計 20 回、入力用の動画を各単語に 15 回ずつ合計 90 回とした。パターンは 15×15 pixels で、ベクトルの次元数は 225 次元とした。前処理としてベクトルの L^2 ノルムの正規化を行った。

実験結果を表 2 に示す。表中の COR, DP, CMSM は 3.3 節と同じ手法を表す。CMSM では、 $N_C = 160$ 次元の制約部分空間で特徴抽出を行い、入力部分空間と辞書部分空間の次元数 N を 7 とした。制約部分空間を学習するために、辞書パターンから 6 動作の学習用部分空間を生成し、それぞれの学習用部分空間の $N_B = 40$ 本の基底ベクトルを用いた。

表 2 の結果より、制約相互部分空間法は、他手法と比較して認識率が高いことから有効であるといえる。しかし、表 3 に示すように正解率が高い単語と低い単語が存在し、特に「ひだり」の正解率が低い。「ひだり」は、図 10 のように「みぎ」と類似したパターンで構成されており誤認識が発生したと考えられる。このような単

表 2 唇動作の識別結果

手法	正解率 (%)
COR	70.0
DP	70.0
CMSM	87.8

表 3 単語毎の正解率 (%)

うえ	した	ひだり	みぎ	はい	いいえ
100.0	93.3	66.7	80.0	86.7	100.0

語に対応するためには、パターン間の順序関係を考慮する必要があると考える。

5 パターン間の順序関係に対する一検討

5.1 内容

パターン間の順序関係を考慮した制約相互部分空間法の一つとして、順序シンボルを付加し認識を行うことを試みる。図 11(b) のように、元のパターン (a) の最下部に濃淡画像のパターンを追加することで、順序シンボルをパターンに付加させる。順序シンボルを表す画像は正規分布に従う画素値の変化によって表される。パターンを獲得した時間が進むにつれて、画像の左端から右端へ正規分布の中心を表す画素が移動する。

5.2 実験結果

実験データには、3 章の顔向き動作と 4 章の唇動作と同じものを用いた。表 4 に実験結果を示す。顔向き動作 (a) と顔向き動作 (b) では、144 次元のベクトルに 12 次元の順序シンボルを付加し、合計 156 次元のベクトルから入力部分空間と辞書部分空間を生成した。 $N_C = 110$ 次元の制約部分空間で特徴抽出を行い、入力部分空間と辞書部分空間の次元数 N は 10 とした。制約部分空間を学習するために、辞書パターンから 4 動作の学習用部分空間を生成し、それぞれの学習用部分空間の $N_B = 60$ 本の基底ベクトルを用いた。

唇動作では、225 次元のベクトルに 15 次元の順序シンボルを付加し、合計 240 次元のベクトルから入力部分空間と辞書部分空間を生成した。 $N_C = 180$ 次元の制約部分空間で特徴抽出を行い、入力部分空間と辞書部分空間の次元数 N は 10 とした。制約部分空間を学習するために、辞書パターンから 6 動作の学習用部分空間を生成し、それぞれの学習用部分空間の $N_B = 75$ 本の基底ベクトルを用いた。

顔向き動作 (a) と顔向き動作 (b) の正解率は、順序シ



図 11 パターンへの順序シンボルの付加

表 4 順序関係を考慮した認識結果

認識対象	正解率 (%)
顔向き動作 (a)	97.5
顔向き動作 (b)	96.7
唇動作	95.6

ンボルを付加した場合でも表 1 の CMSM と変わらなかった。唇動作の正解率は表 2 の CMSM と比べて向上し、表 5 に示す個々の単語の正解率でも表 3 と比べて向上していることから、順序シンボルが有効に働いたと考えられる。

6 まとめ

本稿では、入力パターンと辞書パターンが特徴空間でなす分布の構造を比較することで、従来の HMM や DP マッチングとは異なるジェスチャー認識の手法を提案した。時系列に獲得されたパターン分布を部分空間として近似し、部分空間同士を比較することができる制約相互部分空間法を用いた。これにより、簡便な計算で安定な認識系を構成することができることを示した。提案手法の有効性を、顔向き動作の識別実験と、唇動作による単語識別の実験により確認した。また、パターンに順序関係を表すシンボルを付加することで、制約相互部分空間法においてパターン間の順序関係を考慮にいれる一つの方法を検討した。

今後の課題として、順序シンボルを付加することで正解率が向上する理由の考察や、順序シンボルをパターンに付加するのではなく異なる方法で順序関係を考慮した制約相互部分空間法の検討などが必要である。

参考文献

- [1] Y. Wu, T. Huang, Vision-based gesture recognition: A review, In Proceedings of the International Gesture Workshop (GW'99), 1999.
- [2] 福井 和広, 山口 修, 鈴木 薫, 前田 賢一, 制約相互部分空間法を用いた環境変動にロバストな顔画像認識—照明変動の影響を抑える制約相互部分空間の学

表 5 順序関係を考慮した場合の単語毎の正解率

うえ	した	ひだり	みぎ	はい	いいえ
100.0	93.3	93.3	93.3	93.3	100.0

習一, 信学論 D-II Vol. J82-D-II, No.4, pp.613-620, 1999

- [3] 石井 健一郎, 上田 修功, 前田 英作, 村瀬 洋, わかりやすいパターン認識, オーム社, 1998.
- [4] 山口 修, 福井 和広, 顔向きや表情の変化にロバストな顔認識システム “Smartface”, 信学論 D-II Vol. J84-D-II, No.6, pp.1045-1052, 2001
- [5] 前田 賢一, 渡辺 貞一, 局所的構造を導入したパターンマッチング法, 信学論 D Vol. J68-D, No.3, pp.345-352, 1985
- [6] 福井 和広, 部分空間表現に基づく顔画像認識, 信学技報 PRMU2002-206 pp.33-40, 2003
- [7] 西山 正志, 山口 修, 福井 和広, 家庭用ゲーム機における顔画像認識システム, SICE SI2002 講演論文集 (I) pp.185-186, 2002
- [8] 若杉 智和, 西浦 正英, 福井 和広, 多次元分布間の分離度を用いたロバストな唇輪郭抽出, 信学技報 PRMU2003-278 pp.121-40, 2004