

周期運動検出に基づくジェスチャ認識

高橋 真人^{*1} 入江 耕太^{*2} 寺林 賢司^{*3} 梅田 和昇^{*3}

Gesture Recognition Based on Detection of Periodic Motion

Masahito Takahashi^{*1}, Kota Irie^{*2}, Kenji Terabayashi^{*3} and Kazunori Umeda^{*3}

In this paper, we propose a method to recognize periodic gestures from images. The proposed method uses an amplitude spectrum and a phase spectrum that are obtained by applying Fast Fourier Transform (FFT) to a time series of intensity images. FFT is applied to each pixel of low-resolution images. The method consists of 2 steps. First, the method detects periodic motion regions from the amplitude spectrum. Secondly, the method uses the phase spectrum in the detected periodic motion region to classify the gestures. The proposed method is robust to lighting conditions and individual differences in skin color because it does not rely on color information. Additionally, the proposed method is possible to recognize periodic gestures when multiple target regions, for example, two or more individuals are observed. Several experiments are performed to demonstrate the effectiveness of the proposed method.

Key Words: Gesture Recognition, Periodic Gestures, FFT, Image Processing, Discriminant Analysis, Support Vector Machine

1. はじめに

ロボットをはじめとする機械が我々人間のパートナーとなった安心・安全・豊かな暮らしを実現するために、人間から機械への意思伝達を自然に行える柔軟なヒューマン・マシン・インタフェースが重要である。我々の日常生活を考えた場合、意識的あるいは無意識的に身振り、手振りなど様々なジェスチャを頻繁に用いている [1]。そこで、直感的かつ非接触での操作を可能にするヒューマン・マシン・インタフェースの一つとして、ジェスチャを用いることが考えられる。

これまでに、ジェスチャ認識に関する研究が、人工現実感やコンピュータビジョンなどの分野で数多く報告されている [2]~[5]。人工現実感では、三次元位置センサや加速度センサを搭載したグローブ型のデバイスを用いることにより空間的な手の配置や動きを検出してジェスチャを認識する手法が提案されている [6] [7]。ほかにも、情報機器や情報家電機器の操作を実現する指装着型のウェアラブルデバイスが提案されている [8]。このような装着型の入力デバイスは人工現実感システムにおける重要な入力手段として用いられてきたが、自然なインタラクシ

ョンという観点からは、着脱の手間やケーブルなどによって動きに制約があるなどの問題から必ずしも適しているとはいえない。このようなことから、ケーブルなどを必要としない非接触型のアプローチとして画像処理に基づくジェスチャ認識手法が研究されてきた [9]~[11]。これらの手法は、マーカを利用する手法と利用しない手法に分けられる。前者のマーカを利用する手法は、例えばユーザの手足に装着したマーカの位置をステレオ法などにより三次元的に追跡し、ジェスチャを認識する [12]。しかし、マーカなどを手足に装着することは、自然なインタラクションを実現するうえでは望ましくない。これに対して、後者のマーカを利用しない手法には、入力画像から抽出された手などの領域に対して、動き、形状変化などからジェスチャを認識しているものがある [13]。手などの領域を抽出するためには、二つの手法が一般的に用いられている。一つは、カラー画像からユーザの肌色領域を抽出することにより手領域を特定する手法である [14]。しかしこの手法では、光源環境の変化によりユーザの肌色も変化するため、安定した領域抽出は困難である。もう一つは、画像間の差分を利用する手法である [15]。この手法では、背景が複雑である、背景が一定でない、対象領域と背景が同様の色を持つ、などの場合には対象領域だけを安定して抽出することが困難となる。そのため、背景を一定の状態や色に制限する必要があり、使用可能な環境が限定される。マーカなどを用いず、かつユーザの手に対応する画像領域を抽出するという前処理を行うことなくジェスチャを認識する手法も提案されている [16] [17]。これらの手法は、時系列の動画像データから抽出した特徴量を用いることでジェスチャを認識しているが

原稿受付 2009年6月9日

^{*1}中央大学大学院理工学研究科

^{*2}中央大学大学院理工学研究科/JST CREST

^{*3}中央大学理工学部/JST CREST

^{*1}Graduate School of Science and Engineering, Chuo University

^{*2}Graduate School of Science and Engineering, Chuo University

^{*3}Faculty of Science and Engineering, Chuo University

■ 本論文は学術性で評価されました。

人物が複数存在するなど移動領域が複数の場合、適用が困難である。

上述した研究では、身体の複雑な動きや多様な手のジェスチャを認識できる反面、特殊なデバイスを装着しなければならない、実用環境によって認識が不安定であるといった課題が残る。一方、イエスやノー等の単純な意思伝達や単純な機械の操作を行ううえで、必ずしも身体の複雑な動きや、多様な手のジェスチャを認識する必要はない。このような観点から、我々は単純な周期的動作である手振り動作に着目し、動作者の位置を機械に伝えることを目的とした画像からの手振り動作認識手法を構築した [18]。しかしながら、この手法は手を左右に振るなどの繰り返し動作の有無を認識するのみで、伝達可能なジェスチャは 1 種類のみである。

本論文では、この手法を拡張し、複数の単純な意思伝達を可能にする周期ジェスチャ認識手法を提案する。具体的には、低解像度化した濃淡画像の各画素に対して時間軸方向の FFT [19] を行うことにより得られる振幅スペクトルを用いて周期運動領域を検出する。さらに周期運動領域から手の動作領域を抽出し、その領域内の位相スペクトルを用いて周期ジェスチャの種類を識別する。提案手法は、直感的かつ非接触での動作を可能にし、濃淡画像を用いることで、照明環境や肌色の個人差に対してロバストに周期ジェスチャを認識できる。さらに、あらかじめ肌色情報を用いて手領域を抽出するといった画像処理は必要なく、簡潔な処理を実現できる。したがって、システムの小型化、ハードウェア化が期待でき、汎用性のあるインタフェースが実現できると考えられる。

以下、2 章で周期ジェスチャ認識手法を提案する。3 章で提案手法についての評価実験結果を示し、4 章でまとめと今後の展望を述べる。

2. 周期ジェスチャ認識手法

周期性を伴う時系列ジェスチャを認識するため、まず画像から周期運動領域を検出する。次に、その周期運動領域から手が動作している領域を抽出する。最後に手の動作からジェスチャの種類を識別する。順に 2.1、2.2、2.3 節で手法を示す。

2.1 濃淡値の時系列変化を利用した周期運動領域の検出

周期ジェスチャを行った場合、濃淡画像中の動作領域の画素において、動作部分の濃淡値と背景の濃淡値とで変化が生じる。この濃淡変化は一定の周期を持つため、FFT を用いて定量化することで周期運動領域の検出を実現できる。また、色情報を用いず画素の濃淡情報のみを用いることで、照明環境や肌色の個人差に対してロバストな処理が可能となる。

2.1.1 濃淡値の時系列変化に対する FFT

低解像度化した濃淡画像の各画素の濃淡値の時系列変化に対して FFT を行うことで、周期ジェスチャを特徴づける振幅スペクトルと位相スペクトルを取得できる。

具体的には、まず取得した濃淡画像を $m \times n$ 画素に低解像度化し、全画素の濃淡値 $I(i, j, t)$ を時系列で取得する (Fig. 1 参照)。 $i = 1, 2, \dots, m$, $j = 1, 2, \dots, n$ は画像座標、 t はフレーム数である。画像を低解像度化することにより、計算量の軽減やノイズに対するロバスト性の向上、滑らかな濃淡変化パター

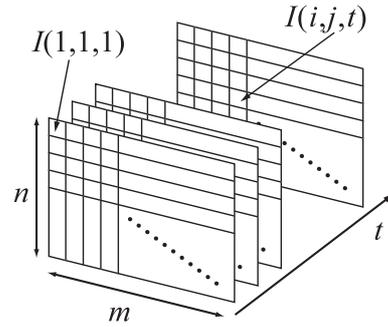
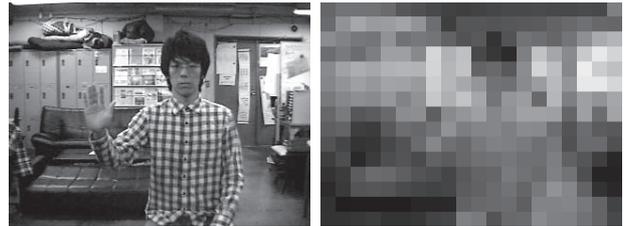


Fig. 1 Time series of low-resolution images



(a) Original image (b) Low-resolution image

Fig. 2 Conversion of image to low resolution

ンが得られるといった効果が期待できる。 Fig. 2 (a) に取得画像、 Fig. 2 (b) に低解像度化した画像の例を示す。

次に、取得した各画素における濃淡値の時系列変化に対して、前 u フレームにおける最大値と最小値の差を用いて正規化処理を施す。さらに、正規化した濃淡値変化に対して FFT を行い、振幅スペクトル A と位相スペクトル $\arg F$ を取得する。

$$F = \sum_{k=0}^{N-1} I_k \mathbf{W}_N^{nk}, \quad (1)$$

$$A = |F|, \quad (2)$$

$$\arg F = \tan^{-1} \frac{\text{Im}(F)}{\text{Re}(F)}. \quad (3)$$

N はサンプリングフレーム数、 \mathbf{W} は DFT (Discrete Fourier Transform) における回転子を表す。 $\text{Re}(F)$, $\text{Im}(F)$ はそれぞれ F の実部と虚部を表す。なお、計算量の軽減やノイズによる誤認識を防ぐために、濃淡値の最大値と最小値の差が I_{dif} 以上ある画素に対してのみ正規化処理を施し FFT を行う。

$$\max(I_{t-u+1}, \dots, I_t) - \min(I_{t-u+1}, \dots, I_t) \geq I_{dif} \quad (4)$$

Fig. 3 (a) に、周期ジェスチャを行った場合の、濃淡画像中の手の動作領域における 1 画素の濃淡値の時系列変化を示す。 Fig. 3 (b) に、 Fig. 3 (a) の正規化後の濃淡値変化を示す。また、 Fig. 3 (c) に振幅スペクトル、 Fig. 3 (d) に位相スペクトルを示す。

2.1.2 周期運動領域の検出

得られた振幅スペクトルを用いてパターン認識を行い、周期運動領域を画素ごとに抽出できる。パターン認識には SVM [20] を適用する。SVM は近年、顔検出 [21] [22]、歩行者検出 [23]、動作認識 [24] など様々な分野に適用され、優れた認識性能を持つことが示されている。SVM は基本的には 2 クラスを識別す

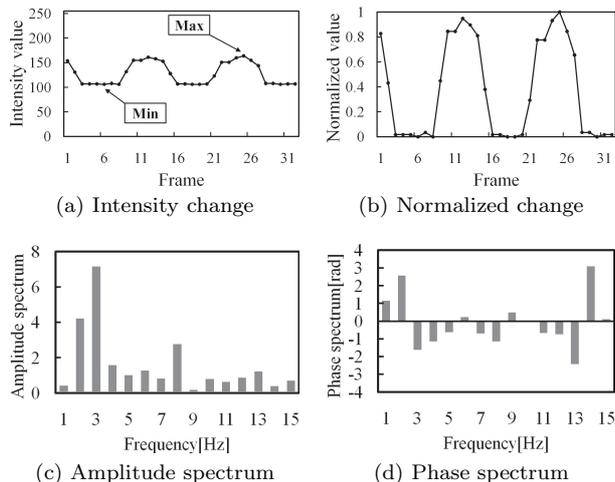


Fig. 3 Application of FFT to time series of intensity values

る識別器を構成するための学習法である。周期運動領域の検出は、周期運動とそれ以外の動作に分類する2クラスの識別問題に帰着されるので、SVMを適用することが可能である。

周期運動領域の検出のための特徴量として、前項で得た振幅スペクトル A の最大値、平均値などの要約統計量を用いる。Fig. 4に、周期運動とそれ以外の動作を行って得られた振幅スペクトルの最大値と平均値のデータを示す。周期運動の検出問題では、Fig. 4に示すように学習データが線形分離不可能であるため、特徴空間において非線形の識別面が必要となる。そこで本手法では、ソフトマージン法とカーネルトリックをSVMに適用する[25][26]。周期運動とそれ以外の動作の特徴空間を分離する超平面は以下の条件付き最適化問題の最適解から求める。

$$\text{minimize } L(\mathbf{w}, \zeta) = \frac{1}{2} \|\mathbf{w}\|^2 + \gamma \sum_{a=1}^M (\zeta_a) \quad (5)$$

$$\text{subject to } \zeta_a \geq 0, s_a(\mathbf{w}^T \mathbf{x}_a + b) \geq 1 - \zeta_a. \quad (6)$$

$(a = 1, 2, \dots, M)$

\mathbf{w} は重みベクトル、 γ は識別エラーに対するペナルティ定数、 M は特徴量の数、 $\zeta_a (\geq 0)$ は不完全な分離を許容するパラメータ、 s_a はクラスラベル、 \mathbf{x} は特徴ベクトル、 b はバイアス項と呼ばれるパラメータである。求められた値を用いて周期運動が行われているか否かの判別を行う。Fig. 5に手を縦に振る動作を行い、周期運動領域を検出した画像の例を示す。Fig. 5(b)において、画素ごとに四角の枠で囲われた領域が周期運動の検出された領域である。ジェスチャの動作領域とそれに伴い周期的に振動している領域が検出されている。

2.2 手の動作領域抽出

画素ごとに抽出された周期運動領域には、手振り動作領域以外にジェスチャを行う際同時に振動する頭や肩、服なども含まれてしまうことがある (Fig. 5(b) 参照)。この場合、ジェスチャを行っている手の動作領域が的確に抽出できず、複数のジェスチャの識別が困難となる。そこで、まず抽出された周期運動領域を補正する。次に、画像中の手の動作領域に周期運動領域が密集していることに着目し、補正周期運動領域の重心座標を利用して手の動作領域を抽出する。

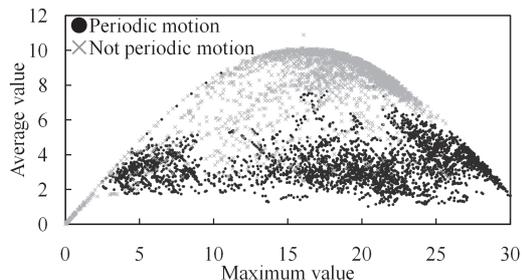


Fig. 4 Distribution of features of the amplitude spectrum



(a) Original image (b) Low-resolution image

Fig. 5 Detection of periodic motion region



(a) Periodic motion region (b) Moving hand region

Fig. 6 Extraction of moving hand region

具体的には、周期運動が認識された全画素に対してモルフォロジー処理を施すことで、頭や肩などのジェスチャを識別するうえで不要な周期運動領域を除去する。ラベリング処理で最大連結領域のみ選択することで、服などの振動領域を除去し補正周期運動領域を抽出する。補正周期運動領域の重心座標 $G(x_G, y_G)$ と補正周期運動領域の全画素を内包する矩形の (x, y) 座標の最小値 (x_{min}, y_{min}) と最大値 (x_{max}, y_{max}) を求める。そして、手の動作領域を次式から求める。

$$X = \min((x_{max} - x_G), (x_G - x_{min})) \quad (7)$$

$$Y = \min((y_{max} - y_G), (y_G - y_{min}))$$

ここで、 X, Y は重心座標 $G(x_G, y_G)$ から手の動作領域までの長辺、短辺までの距離である。Fig. 6(a)の周期運動領域から、上記処理を施し手の動作領域を抽出した結果を Fig. 6(b)に示す。

2.3 周期ジェスチャの認識

抽出された手の動作領域内の位相スペクトルを利用して、周期運動からジェスチャの種類を識別する。

まず、抽出された手の動作領域の全画素を内包する矩形領域を生成する。その矩形領域に内接する楕円周上の8点の画素の位相スペクトルを取得する。8点は楕円周を8等分する画素とする。この8点の画素間から位相スペクトル差を12次元算出する。Fig. 7に位相スペクトル差算出の様子を示す。ここで、識

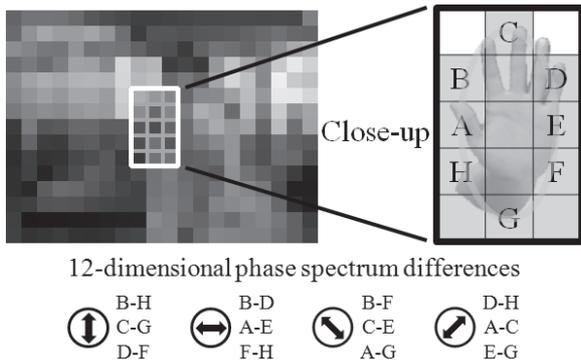


Fig. 7 Regions in which the component of the phase spectrum can be obtained

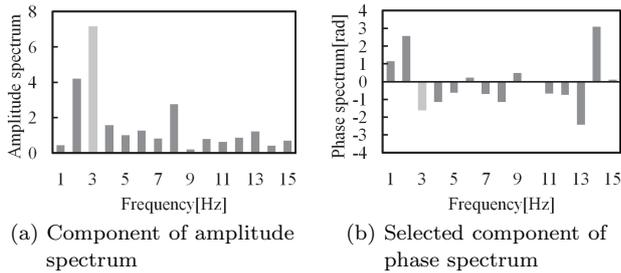


Fig. 8 Selection of a component of phase spectrum

別に用いる位相スペクトルは振幅スペクトルが最大となるピーク周波数の位相スペクトルとする。例えば Fig. 8 では、振幅スペクトルが最大となるピーク周波数は 3 [Hz] なので、3 [Hz] の位相スペクトルを取得する。

次に、この位相スペクトル差を特徴量とし、周期ジェスチャの種類を識別する。周期運動領域の検出と異なり多クラスの識別であるため、2クラス識別器である SVM をここでそのまま適用することはできない。そこで、SVM を多クラス分類問題に適用するための標準的な方法の一つである one-versus-one を採用する [27]。one-versus-one では、すべての 2 クラスの組み合わせに対応する $k(k-1)/2$ 個の 2 クラス識別器を作成する。そして、 $k(k-1)/2$ とおりの 2 クラス識別結果の多数決によって周期ジェスチャの種類を決定する。さらに、より確実な認識を行うために、 l フレーム間ジェスチャの識別を行い、識別回数が最も多い周期ジェスチャを認識ジェスチャとして確定する。

以上で提案した周期ジェスチャ認識処理の流れを Fig. 9 に示す。

3. 実験

提案手法の有効性を検証するため、実験を行った。

3.1 節で、本実験で扱う周期ジェスチャについて述べる。3.2 節で、本実験の条件および提案手法で用いる SVM のパラメータ設定について述べる。3.3 節で、周期運動領域の検出に用いるパターン認識手法を比較検証する。3.4 節で、位相スペクトルを用いた周期ジェスチャ識別手法の有効性を検証する。また、周期ジェスチャ識別に用いるパターン認識手法として、多クラスに対応させた SVM と判別分析の代表的な手法の性能を比較する。3.5 節で、提案手法が、照明環境の変化に対してもロバ

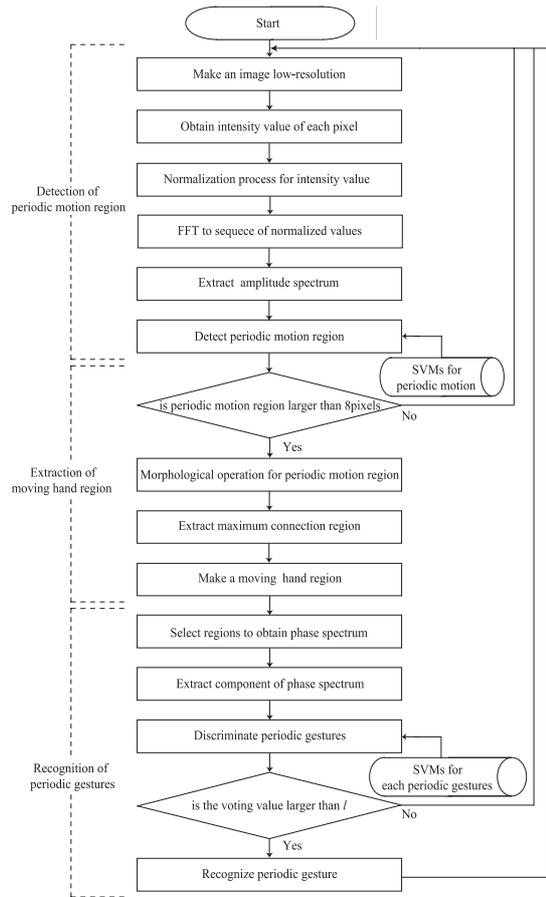


Fig. 9 Flow of recognition of a periodic gesture

トに周期ジェスチャを認識することを確認する。3.6 節で、ジェスチャの速さが認識率に与える影響を検証する。3.7 節で、提案手法が、人物や移動領域が複数存在する環境においても周期ジェスチャを認識できることを検証する。

3.1 本実験で扱う周期ジェスチャ

本実験で扱う周期ジェスチャは、手を横や縦に振る動作、時計回りや反時計回りに回転させる動作の 4 種類とした。これらは、簡単な機械を操作することを想定し、決定や中止、可変数の変更などの意味を表現できるものとして決定した。Fig. 10 にこれらの 4 種類の周期ジェスチャを示す (Horizontal: H, Vertical: V, Clockwise: CW, Counterclockwise: CCW)。

3.2 実験条件およびパラメータの設定

実験システムは PC (DELL Intel Core2Quad Q6600), USB カメラ (ELECOM UCAM-E130), キャプチャソフト (Microsoft DirectShow), 画像処理ソフト (Intel OpenCV) により構成した。実験環境は大学の研究室とし、被験者は 20 名とした。

カメラから被験者までの距離は 2 [m] とし、画像の解像度は、 20×15 とした。FFT を行う際に用いるサンプリングフレーム数 N は周波数分解能を考慮し決定する。周波数分解能は次式で与えられる。

$$R_f = \frac{R_t}{N} \quad (8)$$

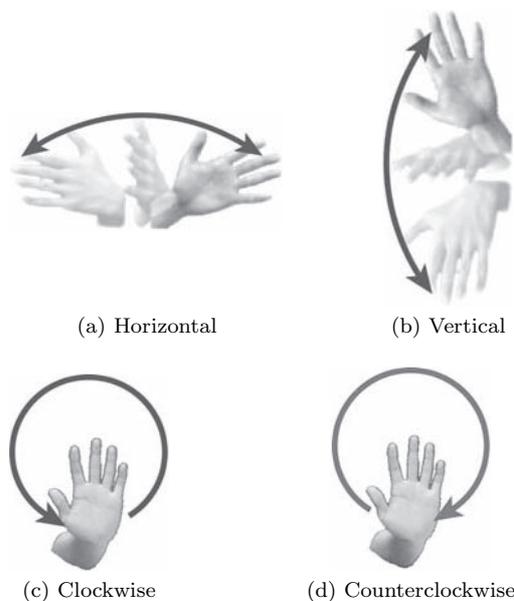


Fig. 10 Four periodic gestures

R_t はフレームレートである。本論文で用いるカメラのフレームレートは 30 [fps] であり、FFT を行うためのサンプリング数は 2 の累乗でなければならないためサンプリングフレーム数 N は 32 とした。正規化フレーム数 u も 32、周期ジェスチャの種類を確定するためのフレーム数 l は 20 とした。

本手法では、カーネルトリックを SVM に適用し学習する際、一般的に用いられる Gauss カーネルを用いた。

$$K(\mathbf{x}_1, \mathbf{x}_2) = \exp\left(-\frac{\|\mathbf{x}_1 - \mathbf{x}_2\|^2}{\sigma^2}\right) \quad (9)$$

ここで、ソフトマージン法で用いる式 (5) の γ と、カーネルトリックで用いる式 (9) の σ の二つの未知数を決定する方法として Grid Search を適用した。特に本手法では、学習にかかる時間の削減のため、最初は粗く次第に細かく値を求めていく Coarse Grid Search を適用した。また、オーバーフィッティングを防ぐために、学習データを 5 グループに分け、Cross Validation [28] を用いてパラメータを決定した。

3.3 周期運動領域の検出

周期運動領域を検出するためのパターン認識手法として SVM と、判別分析の代表的な手法である線形判別法 [29]、マハラノビス距離 [30] を比較し、検討を行った。実験を行うための SVM ツールとして LIBSVM [31] を利用した。実験環境は室内の照明をすべてつけた状態とした。

まず、2 名の被験者が 4 種類の周期ジェスチャと室内を前後左右に歩行する動作や室内における無造作な動作を行った。そして、周期運動とそれ以外の動作に対してそれぞれ 4,000 フレーム分の振幅スペクトルを取得し学習用データとした。次に、異なる 2 名の被験者が周期ジェスチャを行い、それぞれ 1,000 フレーム分の振幅スペクトルを取得しテスト用データとした。

Table 1 に周期運動領域の検出実験の結果を示す。マハラノビス距離と SVM は、線形判別法に比べ高い認識率を得ている。一方、SVM とマハラノビス距離は、わずかに SVM の認識率

Table 1 Detection rate of periodic motion (%)

Method	Detection rate
Linear discriminant analysis	89.9
Mahalanobis distance	91.5
Support vector machine	91.6

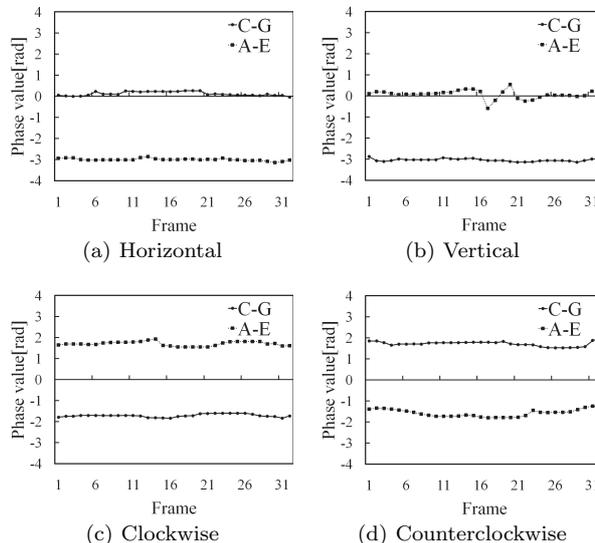


Fig. 11 Time series of difference of phase spectrum for each periodic gestures

が高いことがみとれるが、明確な差はみられなかった。

3.4 周期ジェスチャの認識

被験者 20 名を対象に、4 種類の周期ジェスチャ認識実験を行った。各周期ジェスチャの認識率を確認することにより、位相スペクトルを用いた提案手法の有効性を検証した。また、複数の周期ジェスチャを識別するためのパターン認識手法として多クラスに対応させた SVM と線形判別法、マハラノビス距離を比較し、検討を行った。周期ジェスチャ識別の前処理である周期運動領域の検出には、前節の実験結果より SVM を用いた。

まず、16 名の被験者が 4 種類の周期ジェスチャを行った。1 秒間に手を 2 往復以上周期的に動作させることとし、各被験者が振りやすい振り方でジェスチャを行った。各周期ジェスチャに対して 16,000 フレーム分の位相スペクトル差を取得し、合計 64,000 フレーム分の位相スペクトル差を用いて学習を行った。各周期ジェスチャを行って取得した、Fig. 7 の C-G 間と A-E 間に対応する位相スペクトル差の時系列変化を Fig. 11 に示す。

次に、学習データを提供した 16 名とは異なる 4 名の被験者が各周期ジェスチャの認識実験を行った。各被験者が各周期ジェスチャを 25 回ずつ行い、4 種類のジェスチャで合計 100 回の動作を行った。各被験者は 4 種類の周期ジェスチャを行うように求められ、求められたジェスチャと本システムが正しくジェスチャを認識した割合で評価を行った。実験環境は学習データ取得時、ジェスチャ試行時、ともに室内の照明をすべてつけた状態とした。

Table 2 と Fig. 12 に各周期ジェスチャの認識結果を、Fig. 13 に各周期ジェスチャを行い正しく認識された例を示す。

Table 2 Recognition rate of periodic gestures for each machine learning method (%)

(a) Linear discriminant analysis

Actual class	Predicted class			
	Horizontal	Vertical	Clockwise	Counter clockwise
Horizontal	71	4	2	23
Vertical	0	76	8	16
Clockwise	4	6	74	16
Counterclockwise	9	16	17	58

(b) Mahalanobis distance

Actual class	Predicted class			
	Horizontal	Vertical	Clockwise	Counter clockwise
Horizontal	70	4	1	25
Vertical	0	72	8	20
Clockwise	3	2	68	27
Counterclockwise	5	7	11	77

(c) Support vector machine

Actual class	Predicted class			
	Horizontal	Vertical	Clockwise	Counter clockwise
Horizontal	75	3	2	20
Vertical	0	74	6	20
Clockwise	3	2	71	24
Counterclockwise	5	8	16	71

周期ジェスチャを正しく認識する平均確率（表の対角要素の平均）は、線形判別法を用いた場合 69.75%、マハラノビス距離を用いた場合 71.75%、SVM を用いた場合 72.75% であった。どのパターン認識手法を用いた場合においても、約 70% のジェスチャ認識結果が得られた。各周期ジェスチャを識別するうえで位相スペクトル差を用いることが有効であるといえる。一方、誤認識の結果をみると、どのパターン認識手法、どのジェスチャを行った場合においても、反時計回りに誤認識する確率が高い。反時計回りのジェスチャは、同じ回転動作である時計回りや横・縦動作に比べ、周期的に適当な軌跡で動作を行うことが難しい。そのため、手の動作領域が的確に抽出できず、他のジェスチャと類似の位相スペクトル差が学習されていたことが考えられる。これにより、横・縦・時計回り動作のジェスチャを行った場合、反時計回りの動作と誤認識される確率が高くなったと考えられる。

各パターン認識手法の認識結果を比較してみると、SVM は横、線形判別法は縦と時計回り、マハラノビス距離は反時計回りの動作において最も高い認識結果となった。線形判別法は、他の二つの手法と比較して縦と時計回りにおいて最も高い認識結果を得たが、反時計回りの認識結果が極端に低いため、複数のジェスチャを識別するための手法として有効であるとはいえない。SVM とマハラノビス距離を比較すると、横、縦、時計回りにおいて SVM が高い認識結果を得ている。また、周期ジェスチャを正しく認識する平均確率においても、SVM は 72.75%、マハラノビス距離は 71.75% であることから、SVM は誤認識

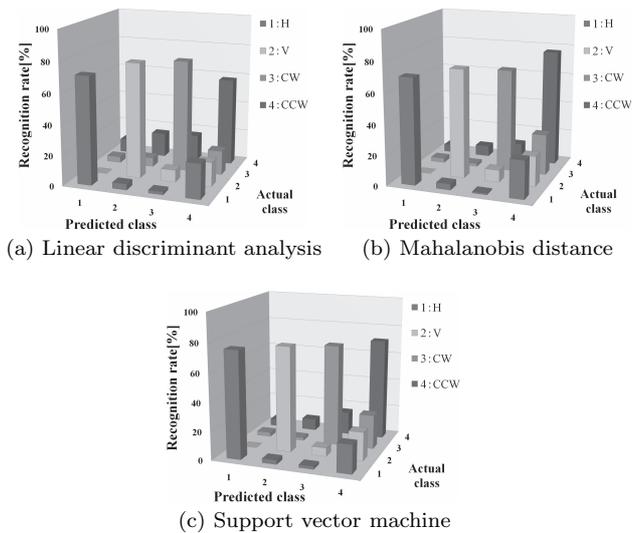


Fig. 12 Recognition rate of periodic gestures for each machine learning method (%)

率が低いことが分かる。

以上より、周期ジェスチャを認識するための手法として位相スペクトルを用いた提案手法が有効であるといえる。複数の周期ジェスチャを識別するためのパターン認識手法として、多クラスに対応させた SVM の周期ジェスチャを正しく認識する平均確率が最も高かったが、明確な差はみられなかった。以降の実験では、多クラスに対応させた SVM を用いる。

また、画像中の任意の場所でジェスチャを始めてから認識されるまでにかかる時間の平均は 1.5 [s] であった。この時間に関しては、ジェスチャを行う人間に負担や不快感を与えることなく、かつ自然なインタラクションを実現できる時間であると考えられる。

3.5 異なる照明環境での周期ジェスチャ認識

照明環境の変化に対してロバストに周期ジェスチャを認識できることを確認するため、実験を行った。実験環境は蛍光灯のみの照明とした。

本節の学習データとして、前節で取得した 16 名の被験者の学習データを用いた。この学習データは室内の照明をすべてつけた状態で取得された。室内の照明をすべてつけた場合、手領域付近の鉛直面照度は約 250~315 [lx] であった。

室内の照明を約 1/3 にして、学習データを提供した 16 名とは異なる 4 名の被験者が、各周期ジェスチャの認識実験を行った。このとき、手領域付近の鉛直面照度は約 60~190 [lx] であった。Fig. 14 に室内の照明をすべてつけた場合と約 1/3 にした場合の画像例を示す。

Table 3 に室内の照明が暗い場合における各周期ジェスチャの認識結果を示す。Table 2 と同様に各ジェスチャが認識できていることが分かる。以上のことから、提案手法は照明環境の変化に対しても有効であるといえる。

3.6 速さの異なる周期ジェスチャの認識

ジェスチャの速さに対してロバストに周期ジェスチャを認識できることを確認するため、実験を行った。実験環境は学習データ取得時、ジェスチャ試行時、ともに室内の照明をすべてつけ



Fig. 13 Image sequence of each periodic gesture and recognition result (from top to bottom)

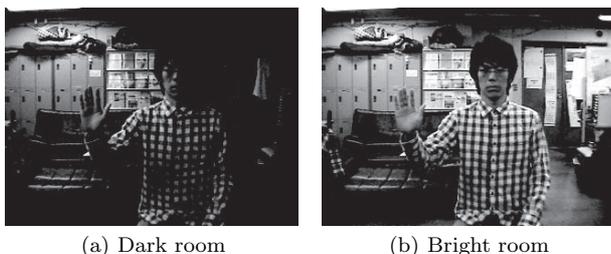


Fig. 14 Difference of illumination

た状態とした。

本節の学習データとして、3.4節で取得した16名の被験者の学習データを用いた。学習データを提供した16名とは異なる4名の被験者が、各周期ジェスチャを行う速さを変えて認識実験を行った。ジェスチャを行う速さは、1秒間に手を2往復、3往復、4往復させる3種類とした。時計回りや反時計回りに回転させる動作は、適当な軌跡で1秒間に手を4往復させるのが困難であるため2往復と3往復のみを行った。なお、本節の実験にはメトロノームを使用し、ジェスチャを行う速さを確認しながら実験を行った。

Table 4 (a), (b) に、1秒間に手を2往復させた場合、3往復

Table 3 Recognition rate of periodic gestures in the dark room (%)

Actual class	Predicted class			
	Horizontal	Vertical	Clockwise	Counter clockwise
Horizontal	73	1	3	23
Vertical	0	84	2	14
Clockwise	3	4	66	27
Counterclockwise	1	10	13	76

させた場合の各周期ジェスチャの認識結果を示す。Table 4 (c) に1秒間に手を4往復させた場合の横振りと縦振りジェスチャの認識結果を示す。Fig. 15 にこれらの結果をグラフで示す。

Table 4 からジェスチャの速さが異なる場合においても各周期ジェスチャが認識できていることが分かる。2 [Hz] の周期ジェスチャをみると、横振りと縦振りにおいて高い認識結果が得られている。これは、手の動作領域が安定的に抽出され、的確な位相スペクトル差が取得できた結果と考えられる。一方、時計回りと反時計回りの認識結果が低い理由として、手の動作軌道が安定していなかった場合、画素によっては周期的な濃淡変化が得られず周期運動領域が検出されないことがあり、それに伴い手

Table 4 Recognition rate of periodic gestures for different hand speeds (%)

(a) 2 [Hz]

Actual class	Predicted class			
	Horizontal	Vertical	Clockwise	Counter clockwise
Horizontal	88	0	4	8
Vertical	0	80	5	15
Clockwise	16	7	59	18
Counterclockwise	8	18	10	64

(b) 3 [Hz]

Actual class	Predicted class			
	Horizontal	Vertical	Clockwise	Counter clockwise
Horizontal	76	0	6	18
Vertical	0	83	0	17
Clockwise	7	2	69	22
Counterclockwise	6	11	8	75

(c) 4 [Hz]

Actual class	Predicted class			
	Horizontal	Vertical	Clockwise	Counter clockwise
Horizontal	62	2	6	30
Vertical	0	84	5	11

の動作領域が的確に抽出されなかったことが挙げられる。3 [Hz] の周期ジェスチャをみると、周期ジェスチャを正しく認識する平均確率が 75.75% と 2 [Hz] の場合 (72.75%) と比較して高い認識結果が得られている。時計回りと反時計回りに関しては、3 [Hz] の手振りを行った場合のほうが安定的に手の動作領域が抽出され、2 [Hz] と比較し認識結果がよいといえる。4 [Hz] の周期ジェスチャをみると、横振りの認識結果が 2 [Hz] や 3 [Hz] と比較して低い。これは、4 [Hz] で横振りをする場合、上半身全体が揺れてしまい、的確に手の動作領域が抽出できなかったことが原因である。

以上をまとめると、提案手法は様々な速さのジェスチャに対しても有効であるといえる。ただし、ジェスチャの速さによって認識率が低下するので改善する必要がある。

3.7 人物が複数存在する環境における周期ジェスチャ認識

人物や移動領域が複数存在する環境においてもロバストに周期ジェスチャを認識できることを確認するため、実験を行った。実験環境は学習データ取得時、ジェスチャ試行時、ともに室内の照明をすべてつけた状態とした。

本節では、3.4 節で取得した 16 名の被験者の学習データを用いた。学習データを提供した 16 名とは異なる 1 名の被験者が、横振り動作を 50 回 5 セット行った。1 名の被験者が手を振っているときに、他の人物は以下に示す 5 ケースの行動を行った。

- ・ Case1: 被験者から離れた場所において、画像内を横切るように 1 名が歩行
- ・ Case2: 被験者から離れた場所において、画像内を横切るよ

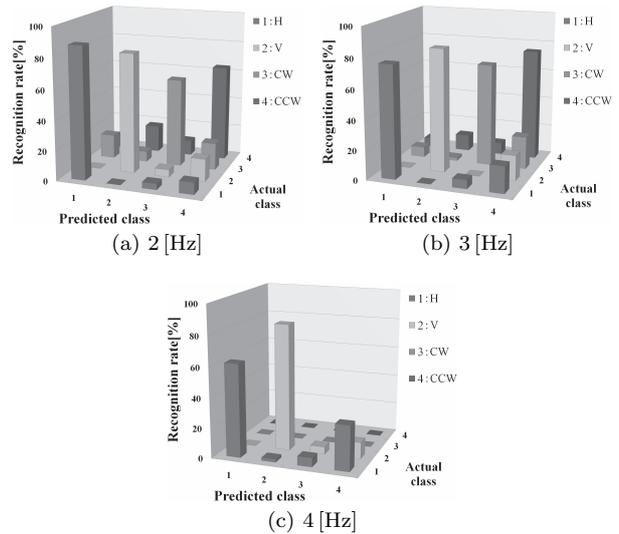


Fig. 15 Recognition rate of periodic gestures for different hand speeds (%)

Table 5 Recognition rate of periodic horizontal gesture when other subjects exist (%)

	Case1	Case2	Case3	Case4	Case5
Horizontal	96	90	84	54	100

うに 2 名が連続して歩行

- ・ Case3: 被験者から近い場所において、画像内を横切るように 1 名が歩行
- ・ Case4: 被験者から離れた場所から被験者に近づくように 1 名が歩行 (手の動作領域と移動人物領域が重なることあり)
- ・ Case5: 被験者から離れた場所から被験者に近づくように 1 名が歩行 (手の動作領域と移動人物領域が重なることなし)

Case1~Case4 はジェスチャを行う手の動作領域と移動人物領域が重なることがあり、Case5 は手の動作領域と移動人物領域が重なることがないものとする。なお本節では、人物や移動領域が複数存在する環境においても周期ジェスチャを認識できることを検証するのが目的のため、前節までの実験結果で最も認識結果の高かった 2 [Hz] の横振りで実験を行った。実験を行う被験者は、前節の 2 [Hz] の横振りにおいて 92% の認識率を得ている。

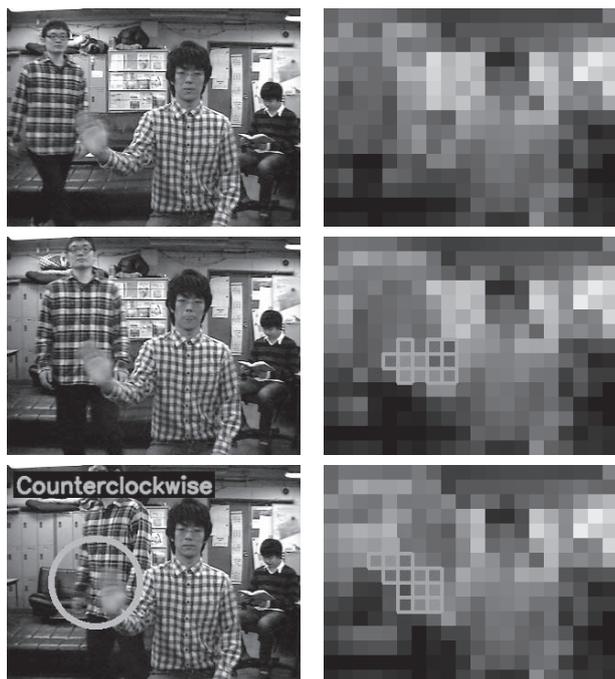
Table 5 に被験者以外の人物が行動している場合に、正しく横振りを認識した結果を示す。Case1 と Case5 においては人物や移動領域が存在しない環境と同様に横振りを認識することができた。Case2 と Case3 に関しては、横振りをしていない領域と移動人物領域が重なった場合において周期運動領域が検出されなくなる画素があり、それに伴い手の動作領域が的確に抽出できず誤認識する場合がみられた。Case4 に関しては、Case2 と Case3 と同様の理由から誤認識する場合がみられた。また、チェックの服 (移動や揺れにより周期的な濃淡変化がおきるもの) を着用している人物が手の動作領域と重なりながら被験者



(a) Original image

(b) Low-resolution image

Fig. 16 Image sequence of Case1 (from top to bottom)



(a) Original image

(b) Low-resolution image

Fig. 17 Image sequence of Case4 (from top to bottom)

に近づいてくる場合、手の動作領域を的確に抽出することが困難となり、正しく横振りを認識することができなかった。Fig. 16に Case1, Fig. 17に Case4の実験の様子を示す。Fig. 16では、横振りを行っている被験者の背後を人物が通過しているが、問題なく認識していることが分かる。Fig. 17では、被験者に近づくとチェックの服を着用した人物により、周期運動領域が検出

され手の動作領域が的確に抽出できないことで、反時計回りと誤認識していることが分かる。

以上をまとめると、ジェスチャを行う手の動作領域と移動人物領域が重なることがない場合は正しく周期ジェスチャを認識できるといえる。また、ジェスチャを行う手の動作領域と移動人物領域が重なる場合においても、正しく周期ジェスチャを認識できる。ただし、チェックの服を着用している人物が手の動作領域と重なりながら被験者に近づいてくるなどの特定のケースにおいては正しく周期ジェスチャを認識することができない。

4. おわりに

本論文では、画像からの周期ジェスチャ認識手法を提案した。提案手法は、低解像度化した濃淡画像の各画素に対して時間軸方向のFFTを行うことにより得られる振幅スペクトルと位相スペクトルを用いることで、周期ジェスチャを認識することができる。実験により、照明環境の変化や様々なジェスチャ速さに対応できることを示した。また、人物や移動領域が複数存在する環境において実験を行い、周期ジェスチャを認識できることを確認した。また、周期運動領域を検出するための手法として、SVM、線形判別法、マハラノビス距離について比較検討を行った。さらに、周期ジェスチャを識別するための手法として、多クラスに対応させたSVMと線形判別法、マハラノビス距離について比較検討を行った。

今後の展望として、インテリジェントルーム [32] などのジェスチャ認識を利用したシステムへの応用が挙げられる。また、距離画像センサなどを用いることにより人間がさらに直感的に利用できる3次元ジェスチャの認識について検討する。

参考文献

- [1] 黒川隆夫：ノンバーバルインタフェース。オーム社、1994。
- [2] D.J. Sturman and D. Zeltzer: "A survey of glove-based input," IEEE Computer Graphics and Applications, vol.14, no.1, pp.30-39, 1994.
- [3] T. Starner, J. Weaver and A. Pentland: "Real-time american sign language recognition using desk and wearable computer based video," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.20, no.12, pp.1371-1375, 1998.
- [4] A. Wilson and A. Bobick: "parametric hidden Markov models for gesture recognition," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.21, no.9, pp.884-900, 1999.
- [5] P. Hong, M. Turk and T.S. Huang: "Gesture Modeling and Recognition Using Finite State Machines," Proc. of IEEE Int. Conf. on Automatic Face and Gesture Recognition, pp.691-694, 2000.
- [6] C. Vogler and D. Metaxas: "ASL recognition based on a coupling between HMMs and 3D motion analysis," Proc. of IEEE int. Conf. on Computer Vision, pp.363-369, 1998.
- [7] 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", 情報処理学会論文誌, vol.39, no.5, pp.1325-1333, 1998.
- [8] 塚田浩二, 安村通晃: "Ubi-Finger: モバイル指向ジェスチャ入力デバイスの研究", 情報処理学会論文誌, vol.43, no.12, pp.3675-3684, 2002.
- [9] H. Wu, T. Shioyama and H. Kobayashi: "Spotting Recognition of Head Gestures from Color Image Series," Proc. of IEEE Int. Conf. on Pattern Recognition, pp.83-85, 1998.
- [10] R. Cutler and L. Davis: "View-based Detection and analysis of Periodic Motion," Proc. of IEEE Int. Conf. on Pattern Recog-

- niton, 1998.
- [11] R. Cutler and L. Davis: "Robust Real-Time Periodic Motion Detection, Analysis, and Applications," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.22, no.8, pp.781-796, 2000.
- [12] 仁野裕一, 古川久雄, 丸家誠: "ユーザ位置によらないジェスチャ認識手法の提案と仮想空間インタフェースへの応用", 電子情報通信学会, 信学技報 PRMU, pp.1-8, 1999.
- [13] J. Martin and J.-B. Durand: "Automatic handwriting gestures recognition using hidden Markov models," Proc. of IEEE Int. Conf. on Automatic Face and Gesture Recognition, pp.403-409, 2000.
- [14] Y. Sato, M. Saito and H. Koike: "Real-time input of 3D pose and gestures of a user's hand and its applications for HCI," Proc. of IEEE Virtual Reality Conference, pp.79-86, 2001.
- [15] J. Segen and S. Kumar: "Shadow gestures: 3D hand pose estimation using a single camera," Proc. of IEEE Int. Conf. on Computer Vision and Pattern Recognition, pp.479-485, 1999.
- [16] J. Yamato, J. Ohya and K. Ishii: "Recognizing Human Action in Time-Sequential Images using Hidden Markov Model," Proc. of IEEE Int. Conf. on Computer Vision and Pattern Recognition, pp.379-385, 1992.
- [17] 西村拓一, 十河卓司, 小木しのぶ, 岡隆一, 石黒浩: "動き変化に基づく View-based Aspect Model による動作認識", 電子情報通信学会論文誌, vol.J84-D-II, no.10, pp.2212-2223, 2001.
- [18] 入江耕太, 梅田和昇: "濃淡値の時系列変化を利用した画像からの手振り検出", 日本ロボット学会誌, vol.21, no.8, pp.923-931, 2003.
- [19] W. Press, S. Teukolsky, W. Vetterling and B. Flannery: Numerical Recipes. C. Cambridge University Press, 1988.
- [20] V. Vapnik: Statistical learning theory. John Wiley and Sons, 1998.
- [21] T. Okabe and Y. Sato: "Support Vector Machines for Object Recognition under Varying Illumination Conditions," Proc. Asian. Conf. Computer Vision, pp.724-729, 2004.
- [22] E. Osuna, R. Freund and F. Girosi: "Training Support vector machines: An application to face detection," Proc. of Computer Vision and Pattern Recognition, pp.130-136, 2000.
- [23] M. Oren, C. Papageorgiou, P. Sinha, E. Osum and T. Poggio: "Pedestrian detection using wavelet templates," Proc. of Computer Vision and Pattern Recognition, pp.193-197, 1997.
- [24] 森武俊, 下坂正倫, 原田達也, 佐藤知正: "マージンに基づく問い合わせ学習を用いたオンライン動作認識", 日本ロボット学会誌, vol.24, no.7, pp.861-872, 2006.
- [25] N. Cristianini and J. Shawe-Taylor: An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge University Press, 2000.
- [26] 前田英作, 村瀬洋: "カーネル非線形部分空間法によるパターン認識", 電子情報通信学会論文誌, vol.J82-D-II, no.4, pp.600-612, 1999.
- [27] C.M. Bishop: Pattern Recognition and Machine Learning. Springer, 2006.
- [28] R. Kohavi: "A study of cross-validation and bootstrap for accuracy estimation and model selection," Proc. of the 4th Int. Joint Conf. on Artificial Intelligence, pp.1137-1143, 1995.
- [29] R.O. Duda and P.E. Hart: Pattern Classification and Scene Analysis. John Wiley and Sons, 1973.
- [30] B. Ripley: Pattern Recognition and Neutral Networks. Cambridge University Press, 1995.
- [31] C.-C. Chang and C.-J. Lin: LIBSVM: A Library for Support Vector Machines. <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.
- [32] 入江耕太, 若村直弘, 梅田和昇: "ジェスチャ認識に基づくインテリジェントルームの構築", 日本機械学会論文集 C 編, vol.73, no.725, pp.258-265, 2007.



高橋真人 (Masahito Takahashi)

2009 年中央大学理工学部精密機械工学科卒業。現在同大学大学院精密工学専攻修士課程前期課程在学中。画像処理を用いたヒューマンインタフェースの研究に従事。2009 年日本機械学会若手優秀講演フェロー賞受賞。(日本ロボット学会学生会員)



寺林賢司 (Kenji Terabayashi)

2004 年北海道大学大学院システム情報工学専攻修士課程修了。2008 年東京大学大学院工学系研究科精密機械工学専攻博士課程修了, 博士 (工学)。同年中央大学理工学部精密機械工学科助教, 現在に至る。ヒューマンインタフェース, ロボットビジョン等の研究に従事。精密工学会, 日本機械学会, 日本バーチャルリアリティ学会等の会員。(日本ロボット学会正会員)



入江耕太 (Kota Irie)

2003 年中央大学大学院理工学研究科精密工学専攻博士前期課程修了。同年 (株) 日立製作所入社, 車載用画像認識カメラの開発を担当。2006 年中央大学精密工学専攻社会人博士後期課程に入学, 現在に至る。画像処理を用いたヒューマンインタフェースの研究に従事。2002 年電気学会電子・情報・システム部門大会奨励賞受賞。(日本ロボット学会正会員)



梅田和昇 (Kazunori Umeda)

1989 年東京大学工学部精密機械工学科卒業, 1994 年同博士課程修了。同年中央大学理工学部精密機械工学科専任講師, 2006 年より同教授。2003~2004 年カナダ NRC Visiting Worker。ロボットビジョンの研究に従事。2004 年 MIRU 長尾賞受賞, 博士 (工学)。日本機械学会, 精密工学会, 電子情報通信学会, IEEE 等の会員。(日本ロボット学会正会員)