

FFT を利用した周期ジェスチャ認識

○ 高橋真人†, 入江耕太†, 寺林賢司†, 梅田和昇†

○ Masahito TAKAHASHI †, Kota IRIE †, Kenji TERABAYASHI and Kazunori UMEDA †

†: 中央大学/JST CREST, takaha@sensor.mech.chuo-u.ac.jp, umeda@mech.chuo-u.ac.jp

本論文では、濃淡値の時系列変化を利用した画像からの周期ジェスチャ認識手法を提案している。提案手法では、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を用いて周期運動領域を検出し、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識別を行う。

<キーワード> Gesture Recognition, Periodic Gestures, Image Processing, FFT, SVM

1. はじめに

人間から機械への意思伝達を自然に行うために、ジェスチャ認識を用いたヒューマン・マシン・インタフェースの研究が盛んに行われている[1]~[3]。これらの多くは、ユーザの肌色情報を基に抽出し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[4]。そのため、照明環境の変化により手領域がうまく抽出できない場合、ジェスチャの認識が困難である。画像間の差分を利用することで手領域を抽出しジェスチャを認識する手法も提案されている[5]が、背景が複雑である、背景が一樣でない、などの場合適用が困難であり、環境が限定される。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、実用環境によって認識が不安定であるといった課題が挙げられる。そこで我々は、単純な周期的動作である手振り動作に着目し、複数の単純な意思伝達を目的とした周期ジェスチャ認識手法を構築してきた[6]。

本論文では、より汎用的な手法の構築を目指し、手の動作領域を抽出するための手法と各周期ジェスチャを識別するための手法を拡張した新たな周期ジェスチャ認識手法を提案する。具体的には、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、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。また、各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 を適用する。

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、周期運動領域の検出を実現できる。

2.1.1 濃淡値の時系列変化に対する FFT

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ンが得られるといった効果が期待できる。

つぎに、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。

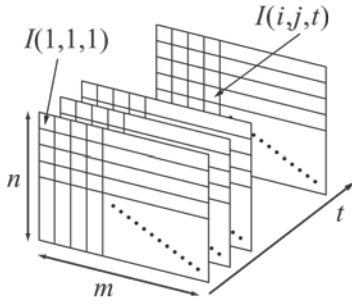


Fig.1 Time series of low-resolution images

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

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

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

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。

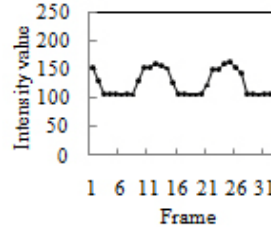
$$\max(I_{t-u+1}, \dots, I_t) - \min(I_{t-u+1}, \dots, I_t) \geq I_{dif}. \quad (4)$$

Fig.2(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.2(b)に、Fig.2(a)の正規化後の濃淡値変化を示す。また、Fig.2(c)に振幅スペクトル、Fig.2(d)に位相スペクトルを示す。

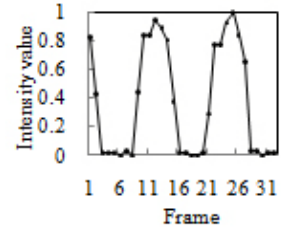
2.1.2 周期運動領域の検出

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毎に抽出できる。パターン認識にはSVM[7]を適用する。SVMは近年、顔検出[8]、歩行者検出[9]、動作認識[10]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。

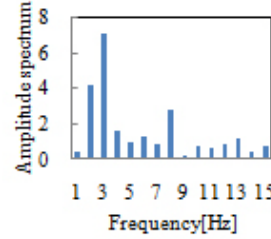
周期運動領域の検出のための特徴量として、前項で得た振幅スペクトル A の最大値、平均値などの要約統計量を用いる。そして、本手法では、ソフトマージン法とカーネルトリックをSVMに適用する。周期運動とそれ以外の動作の特徴空間を分離する超平面は以下の条件付き最適化問題の最適解から求める。



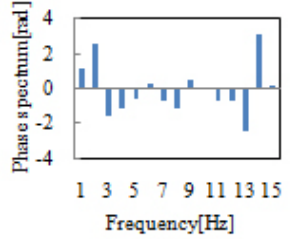
(a) Intensity change



(b) Normalized change



(c) Amplitude spectrum



(d) Phase spectrum

Fig.2 Application of FFT to time series of intensity values

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

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

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

$\mathbf{w}$ 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、 $\delta_a (\geq 0)$ 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.3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.3(b)において、画素毎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

画素毎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.3(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、抽出された周期運動領域を補正した後に手の動作領域を抽出する。

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



(a) Original image (b) Low-resolution image
Fig.3. Detection of periodic motion region

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。さらに、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 $G(x_G, y_G)$ を求める。そして、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 $(x_{\max}, y_{\max})$ を用いて手の動作領域を次式から求める。

$$\begin{aligned} X &= \min((x_{\max} - x_G), (x_G - x_{\min})) \\ Y &= \min((y_{\max} - y_G), (y_G - y_{\min})) \end{aligned} \quad (7)$$

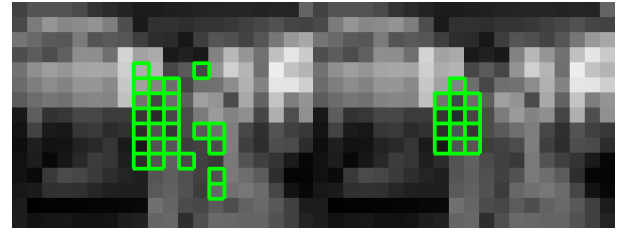
ここで、 X, Y は重心座標 $G(x_G, y_G)$ から手の動作領域を内包する矩形 (長辺, 短辺) までの距離である。Fig.4 (a) の周期運動領域から、上記処理を施し手の動作領域を抽出した結果を Fig.4 (b) に示す。

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.5 に位相スペクトル差算出の様子を示す。ここで、識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.2 (c) 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、Fig.2 (d) の 3[Hz] の位相スペクトルを取得する。

つぎに、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[11]。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。



(a) Periodic motion (b) Moving hand region
Fig.4. Extraction of moving hand region

Fig.4. Extraction of moving hand region

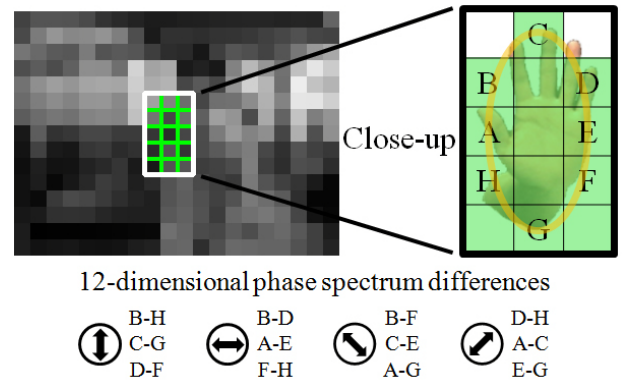


Fig.5. Regions in which the component of the phase spectrum can be obtained

そして、 $k(k-1)/2$ 通りの 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、1 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。

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、本実験の条件及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、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。3.4 節で、提案手法が照明環境の変化に対してもロバストに周期ジェスチャを認識することを確認する。

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、可変量の変更などの意味を表現できるものとして決定した。

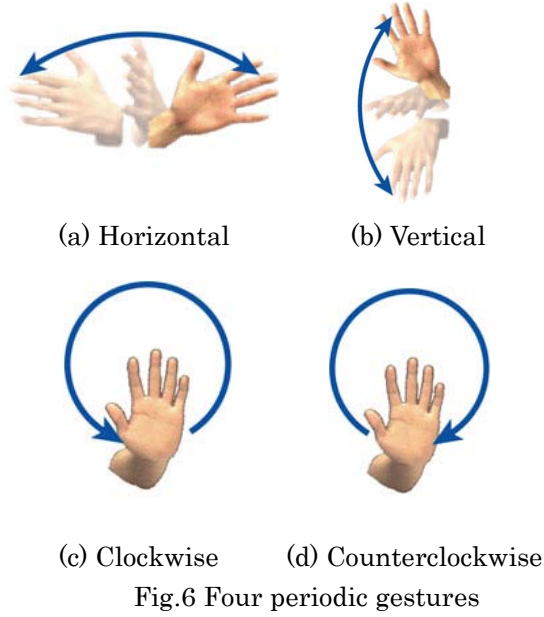


Fig.6 にこれらの 4 種類の周期ジェスチャを示す (Horizontal: H, Vertical: V, Clockwise: CW, Counterclockwise: CCW).

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(DELLE 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. 用いるカメラのフレームレートは 30[fps]である. 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数 N は 2 の累乗でなければならないため 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.

$$k(x_1, x_2) = \exp\left(-\frac{\|x_1 - x_2\|^2}{\sigma^2}\right) \quad (8)$$

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(8)の σ の 2 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 を用いてパラメータを決定した.

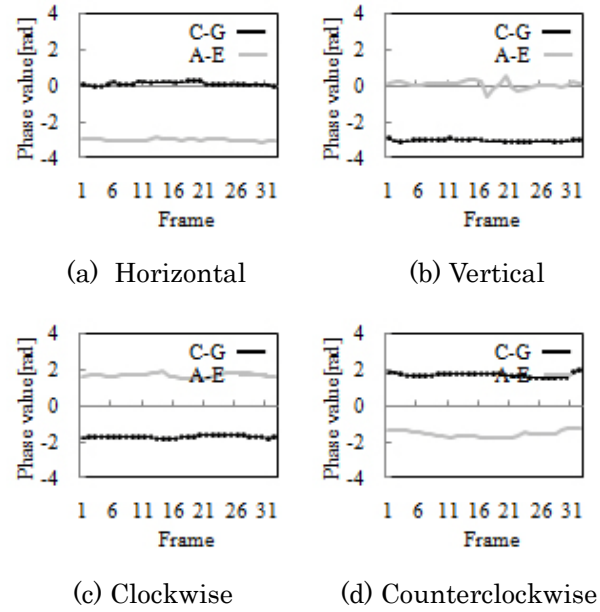


Fig. 7 Time series of difference of phase spectrum for each periodic gestures

3.3 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.

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 16000 フレーム分の位相スペクトル差を取得し, 合計 64000 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.5 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.7 に示す.

つぎに, 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, とともに室内の照明をすべてつけた状態とした.

Table 1 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

Table 1 に各周期ジェスチャの認識結果を示す。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、反時計回りのジェスチャ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



(a) Dark room

(b) Bright room

Fig.8 Difference of illumination

結果となった。線形判別法は、他の 2 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は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 のほうが高い認識率を得ている。

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わずかであるが最も高かった。なお、次節の実験では、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。

3.4 異なる照明環境での周期ジェスチャの認識

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.8 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 2 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 1 と同様に各ジェスチャが認識できていることがわかる. 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.

Table 2 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

4. おわりに

本論文では, 手の動作領域を抽出するための手法と各周期ジェスチャを識別するための手法を拡張した新たな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した. また, 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[12]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] D. J. Sturman, D. Zeltzer: “A survey of glove-based input,” IEEE Computer Graphics and Applications, vol.14, no.1, pp.30-39, 1994.
- [2] 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.

- [3] 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.
- [4] Y. Sato, M. Saito, 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.
- [5] J. Segen, 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.
- [6] 高橋真人, 入江耕太, 寺林賢司, 梅田和昇: “画像からの周期ジェスチャ認識”, 画像の認識・理解シンポジウム (MIRU2009), 2009.
- [7] V. Vapnik: Statistical learning theory. John Wiley and Sons, 1998.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] Christopher M. Bishop :Pattern Recognition and Machine Learning. Springer, 2006.
- [12] 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.