

濃淡値の時系列変化を利用した周期ジェスチャの認識

Periodic Gesture Recognition Based on Time Series of Intensity

○ 学 高橋 真人 (中央大)

入江 耕太 (中央大/JST CREST)

荻島 秀敏 (中央大)

正 寺林 賢司 (中央大/JST CREST)

正 梅田 和昇 (中央大/JST CREST)

Masahito TAKAHASHI, Chuo University, takaha@sensor.mech.chuo-u.ac.jp

Kota IRIE, Chuo University/CREST, JST

Hidetoshi OGISHIMA, Chuo University

Kenji TERABAYASHI, Chuo University/CREST, JST

Kazunori UMEDA, Chuo University/CREST, JST

In this paper, we propose a method to recognize periodic gestures from images as a key technology for human machine interface. The proposed method uses power spectrum and phase spectrum that are obtained by applying FFT to time series of intensity images. FFT is applied to each pixel of the low-resolution images. The method consists of 2 steps. First, the method detects periodic motion regions from power spectrum. Second, the system uses the phase spectrum in the detected periodic motion region to classify the gestures. The proposed method is robust to lighting condition and individual difference of skin color, because it does not use color information. To demonstrate the effectiveness of the proposed method, we have conducted several experiments and show that support vector machines are well suited for obtaining discriminate hyperplanes for recognition of periodic motion.

Key Words: Image Processing, Gesture Recognition, Periodic Motion, Human Machine Interface, FFT

1. はじめに

ロボットを始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(Human machine interface - HMI)が重要である。我々の日常的なコミュニケーションを考えた場合、意識的あるいは無意識的に身振り、手振りなどの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 HMI の 1 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、ジェスチャ認識を用いた HMI の研究が、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、3 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て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手法が研究されてきた[9] ~ [11]。画像処理に基づくジェスチャ認識の手法は、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 3 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では、入力画像から抽出された手などの領域に対して、動き、形状変化などからマッチングを行い、ジェスチャを認識している[13]。手などの領域を抽出するためには、2 つの手法が一般的に用いられている。1 つは、カラー画像からユーザの肌の色に近い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 1 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る、使用可能な環境が限定されるという問題が生じる。

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]。この手法では、色情報を用いていないため、照明条件や肌色の個人差に対してロバストであることが示されている。しかしながら、手を左右に振るなどの繰り返し動作を認識するのみで、その繰り返し動作がジェスチャであることを識別できないという問題がある。

本論文では、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 を行い、周期性を伴う時系列ジェスチャ認識手法を提案する。具体的には、FFT を行うことにより得られるパワースペクトルを用いて周期運動を検出する。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操作、自然なインタラクションを可能にする。また、提案手法は、あらかじめ手領域を抽出するといった画像処理は不必要であり、高速な計算が可能な FFT[17]を適用することで、極めて簡潔な処理を実現している。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。

2. 提案手法

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を検出する (2.1 節)。次に、その周期運動領域から手の動作領域を抽出する (2.2 節)。最後に手の動作領域を基にジェスチャの種類を識別する (2.3 節)。図 1 にこの処理手順を示す。

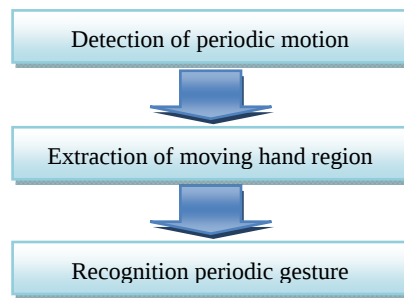


Fig.1 The process of periodic gesture recognition

2.1 周期運動の検出

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、定量化する手法としてFFTを用いる。

まず、取得した濃淡画像を $m \times n$ の画素数に低解像度化し、全画素の濃淡値 $I(i,j,t)$ を時系列で取得する (図2参照)。ただし、 $i=1,2,\dots,m$, $j=1,2,\dots,n$ は画像座標、 t はフレーム数である。図3(a)に取得画像、図3(b)に低解像度化した画像の例を示す。画像の低解像度化は、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、各画素における濃淡値の時系列変化を取得し、図4で示すように前 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。正規化した濃淡値変化に対してFFTを行い、図5に示すようなスペクトルを取得する。位相スペクトルはパワー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を取得する。得られた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する。特徴量にはパワースペクトル G を用いる。 G は式(1)で与えられる。なお、 N はサンプリング数、 W はDFT (Discrete Fourier Transform) における回転子を表す。図6に周期運動を検出した画像の例を示す。

$$G = \sum_{k=1}^{N-1} I_k W_N^k \quad (1)$$

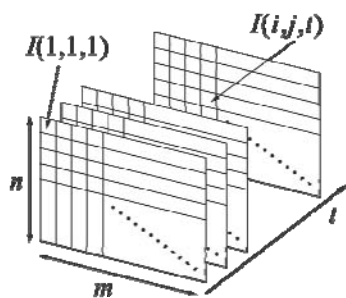
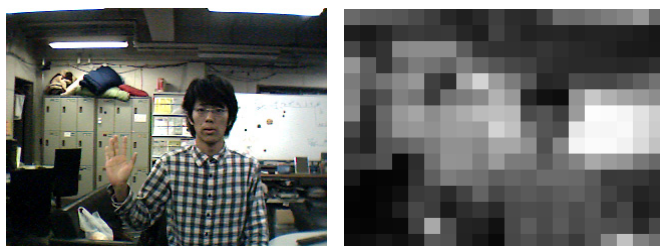
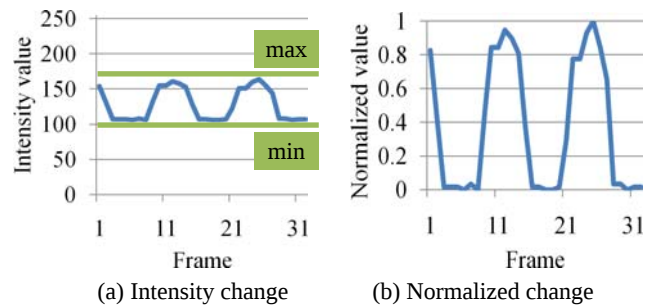


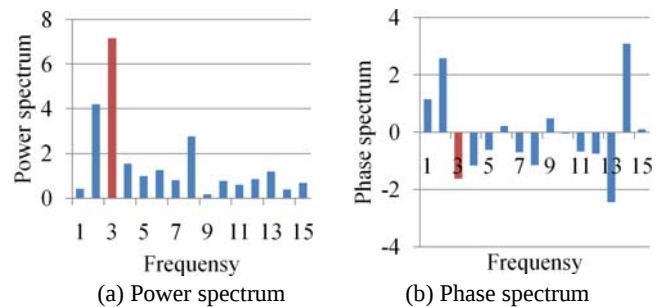
Fig.2 Time series of low-resolution images



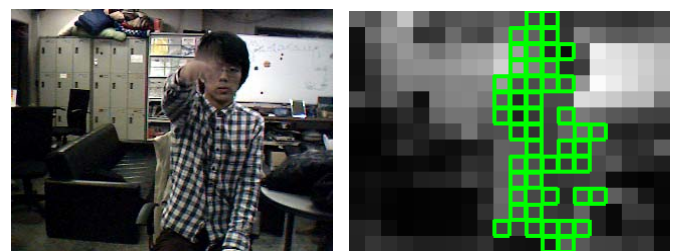
(a) Original image (b) Low-resolution image
Fig.3 Conversion of image to low-resolution



(a) Intensity change (b) Normalized change
Fig.4 Time series change of intensity values



(a) Power spectrum (b) Phase spectrum
Fig.5 FFT spectrum

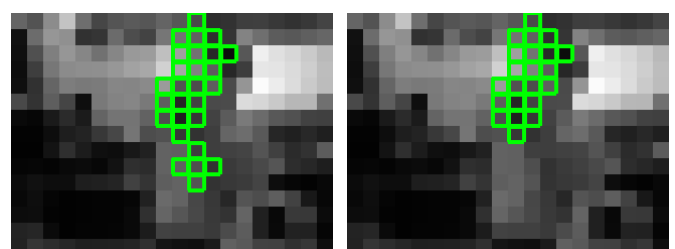


(a) Original image (b) Low-resolution image
Fig.6 Detection of periodic motion

2.2 手の動作領域抽出

検出された周期運動を識別する。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してしまう頭や肩、服なども含まれている (図6(b))。このため、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、ジェスチャの識別が困難である。そこで、抽出された周期運動領域を補正する。

具体的にはまず、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 (図7(a))。さらに、ラベリング処理で最大連結領域のみ選択することにより、服などの振動領域を除去する (図7(b))。このように補正された周期運動領域を手の動作領域とする。



(a) Corrected region (b) Moving hand region
Fig.7 Extraction of moving hand region

2.3 周期ジェスチャの識別

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し、その領域内における画素間の位相スペクトル差を算出する。この位相スペクトル差を特徴量とし、パターン認識を行い周期ジェスチャを識別する。多クラスの識別であるため、ここでは識別法にマハラノビス距離を用いる。より確実な認識を行うために、1フレーム数以上同一のジェスチャが識別された場合に最終結果として周期ジェスチャの種別を確定する。

3. 周期ジェスチャ認識実験

濃淡値の時系列変化を利用した周期ジェスチャ認識手法の有効性を確認するために、認識実験を行った。

まず、本手法が照明環境に対してロバストであることを確認するために、照明環境を変えて周期運動の検出実験を行った(3.1節)。周期運動を検出するためのパターン認識の手法として3つを比較し、検討を行った。1つ目は線形判別法で、パターンの分布全体を考慮して識別面を決定する代表的な手法である[18]。2つ目は識別を誤る確率を考慮したマハラノビス距離による識別法である[19]。3つ目は、Vapnikによって提案され、近年分類学習モデルの一つとして注目を集めている Support Vector Machine (SVM)である[20]~[22]。次に、位相スペクトルを用いた周期ジェスチャ認識手法の有効性を確認するために、4種類の周期ジェスチャ認識実験を行った(3.2節)。図8に実験で扱った4種類の周期ジェスチャを示す。

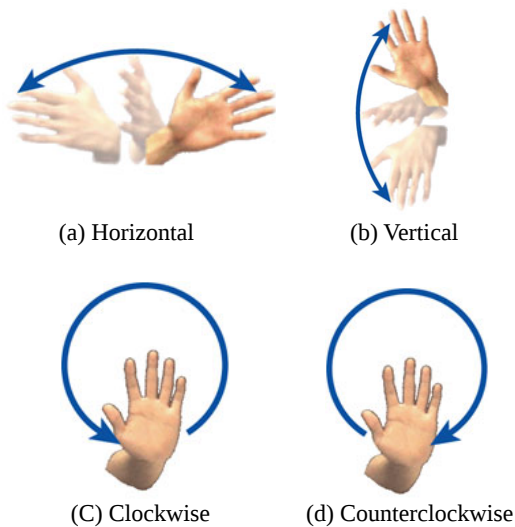


Fig.8 Four periodic Gestures

サンプリング数 N は周波数分解能から 32 とし、正規化フレーム数 u は 32、周期ジェスチャの種別を確定するためのフレーム数 l は 10、解像度は 20×15 とした。また、カメラから被験者までの距離は 2[m] とした。2名の被験者に周期ジェスチャとそれ以外の動作を行ってもらい、スペクトルを取得して学習を行った。本手法では、ソフトマージン法・カーネルトリックを SVM に適用し学習する際、Gauss カーネルを用いる。

$$\text{Subject to } \zeta_i \geq 0, y_i(w^T x_i + b) \geq 1 - \zeta_i, (i=1, \dots, N) \quad (2)$$

$$\text{Minimize } L(w, \zeta) = \frac{1}{2} \|w\|^2 + \gamma \sum_{i=1}^N \zeta_i \quad (3)$$

$$\text{Gauss カーネル } K(x_1, x_2) = \exp\left(-\frac{\|x_1 - x_2\|^2}{2\sigma^2}\right) \quad (4)$$

ここで、ソフトマージン法で用いる、逸脱する範囲を決める定数である式(3)の γ と、カーネルトリックで用いる式(4) Gauss カーネルの σ の2つの未知数を決定する方法として grid search を用いる。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 を適用する。また、オーバーフィッティングを防ぐために、クロスバリデーション(Cross validation) を用いてパラメータを決定する。

3.1 周期運動の検出

パターン認識の手法として線形判別法、マハラノビス距離による識別法、SVM、それぞれを用いて周期運動の検出に関するシミュレーション実験を行った。シミュレーション実験を行うための SVM ツールとして LIBSVM[23]を利用した。

まず、2名の被験者に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てもらい、周期運動とそれ以外の動作に対してそれぞれ 4000 フレーム分のパワースペクトルを取得し学習用データとした。実験環境は室内の照明をすべてつけた状態で行った。

次に、2パターンの照明環境下で、周期ジェスチャを行ってもらい、2000 フレーム分のパワースペクトルを取得しテスト用データとした。実験環境は蛍光灯のみの照明とし、手領域付近の鉛直面照度は約 60~190[lx] (室内の照明を約 1/3 にした場合) と約 250~315[lx] (室内の照明をすべてつけた場合) で行った。図9に室内の照明をすべてつけた場合と約 1/3 にした場合の様子を示す。



Fig.9 Difference of illuminations

表1は周期運動検出実験結果である。異なる照明環境においても高い精度で周期動作が認識されていることがわかる。このことから、照明環境に左右されることなく周期運動を認識可能であることが確認できる。室内の照明が明るい場合と暗い場合の認識率を比較すると、どの手法を用いた場合でも照明を暗くした方が若干認識率が高い。これは照明が明るい場合、壁面が十分に照らされ、手と背景で濃淡の差が生じなかったことが考えられる。また、SVM が線形判別法やマハラノビス距離による識別法より優れていることがわかる。以上のことから、周期運動の検出に関して、濃淡値の時系列変化を利用した認識手法が有効であり、周期運動を検出するためのパターン認識の手法として、SVM が有効であるといえる。

Table 1 Detection rate of periodic motion

Method	60-190[lx]	250-315[lx]
LD	95.00%	88.65%
MD	93.90%	88.85%
SVM	95.15%	90.05%

3.2 周期ジェスチャの認識

前節では、本手法で照明環境に対してロバストに周期運動を検出できることを示した。またパターン認識の手法について検討し、SVM が有効であることを示した。本節では、SVM を用いて周期運動を検出し、各周期ジェスチャを認識する実験を行い考察する。

まず、2名の被験者に4種類の周期ジェスチャを行ってもらい、各ジェスチャに対して1000フレーム分の位相スペクトル差を取得した。このようにして集めた合計4000フレーム分の位相スペクトル差を用いて学習を行った。次に、学習データを提供した2名を含む4名の被験者により、各周期ジェスチャの認識実験を行った。各被験者が各ジェスチャを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。実験環境は室内の照明をすべてつけた状態で行った。各被験者は4つ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により判別されたジェスチャの一致度で評価を行った。

表2に周期ジェスチャの判別結果を示す。各周期ジェスチャを正しく認識する確率 (Probability of correct recognition - PCR) は80.25%である。PCRは表の対角要素の平均である。横動作に関しては、96%と高い認識率が得られていることがわかる。さらに、同じ回転動作である時計回りと反時計まわりに関しても高い認識率が得られていることから、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ことが確認できる。反時計まわりの回転動作の方が若干認識率が低くなっているが、これは個人により利き手があるように回転させる方向にも得意、不得意などの個人差があったと考えられる。また、縦動作の認識率が他のジェスチャと比較し低いことがわかる。これは、周期運動検出後に精度良く手の動作領域を抽出できなかったため、的確に位相スペクトル差を取得出来なかったことが原因と考えられる。

Table 2 Recognition rate of periodic gesture

Input Output	Horizontal	Vertical	Clockwise	Counter Clockwise
Horizontal	96%	0%	1%	3%
Vertical	11%	68%	18%	3%
Clockwise	14%	1%	82%	3%
Counter Clockwise	13%	0%	12%	75%

4. おわりに

HMIとしての利用を目的とした、画像からの周期ジェスチャ認識手法を提案した。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を行い、認識にはパワースペクトルと位相スペクトルを用いた。色情報を用いていないため、照明条件に対してロバストな処理を実現した。周期性を伴う時系列ジェスチャの認識実験を行い、濃淡値の時系列変化を利用した提案手法の有効性を示した。また、周期運動を検出するためのパターン認識手法として、線形判別法、マハラノビス距離による識別法、SVMについて比較検討を行い、その結果SVMが有効であることを示した。

今後の展望として、多様な情報の伝達を実現するため新た

な周期ジェスチャの追加が考えられる。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ような3次元ジェスチャの認識を実現することも目指す。

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, D. Zeltzer, “A survey of glove-based input,” IEEE Computer Graphics and Applications, vol.14, 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. Pattern Analysis and Machine Intelligence, vol.20, pp.1371-1375, 1998.
- [4] J. Sherrah, and S. Gong, VIGOUR, “A System for Tracking and Recognition of Multiple People and their Activities,” Proc. of the International Conference on Pattern Recognition”, pp.179-183, 2000.
- [5] P. Hong, M. Turk, T. S. Huang, “Gesture Modeling and Recognition Using Finite State Machines,” 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,” Sixth International Conference 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 the International Conference on Pattern Recognition, pp.83-85, 1998.
- [10] R. Cutler and L. Davis, “View-based Detection and analysis of Periodic Motion,” Proc. of the International Conference on Pattern Recognition, 1998.
- [11] R. Cutler and L. Davis, “Robust Real-Time Periodic Motion Detection, Analysis, and Applications,” IEEE Trans. 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, J.-B. Durand, “Automatic handwriting gestures recognition using hidden Markov models”, Proc. of IEEE International Conference on Automatic Face and Gesture Recognition,” pp.403-409, 2000.
- [14] Y. Sato, M. Saito, H. Koike, “Real-time input of 3D pose and gestures of a user's hand and its applications for HCI,” The 2001 IEEE Virtual Reality Conference, pp.79-86, 2001.
- [15] J. Segen, S. Kumar, “Shadow gestures: 3D hand pose estimation using a single camera,” Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp.479-485, 1999.
- [16] 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.
- [17] W. Press, S. Teukolsky, W. Vetterling and B. Flannery, “Numerical Recipes in C,” Cambridge University Press, 1988.
- [18] R. O. Duda and P. E. Hart, “Pattern Classification and Scene Analysis.” John Wiley and Sons, 1973.
- [19] B. Ripley, Pattern Recognition and Neural Networks, Cambridge University Press.
- [20] V. Vapnik, “Statistical learning theory,” John Wiley and Sons, New York 1998.
- [21] T. Okabe and Y. Sato, “Support Vector Machines for Object Recognition under Varying Illumination Conditions”, Asian Conf. Computer Vision, pp. 724-729, January 2004.
- [22] T. Kurita, “Support vector machine and generalization,” Journal of Advanced Computational Intelligence, vol.8, no.2, pp.84-91, 2004.
- [23] C.-C. Chang, C.-J. Lin, LIBSVM, “A Library for Support Vector Machines.” <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>