

## フーリエ変換を用いた小さな手振りの検出

Detection of Small Hand-Waving Movements using Discrete Fourier Transform

浅野 秀胤†, 織茂 達也†, 高橋 真人‡, 寺林 賢司‡, 太田 睦†, 梅田 和昇‡

Hidetsugu Asano†, Tatsuya Orimo†, Masahito Takahashi‡,

Kenji Terabayashi‡, Mutsumi Oota†, Kazunori Umeda‡

†: パイオニア株式会社, {hidetsugu\_asano,tatsuya\_orimo,mutsumi\_oota}@post.pioneer.co.jp

‡: 中央大学, takaha@sensor.mech.chuo-u.ac.jp, {terabayashi,umeda}@mech.chuo-u.ac.jp

概要: 新たなジェスチャインタフェースとして, カメラを用いた小さな手振りの検出手法を提案する. 入力画像を時間方向にフーリエ変換し, 特定周波数の振幅スペクトルの差分を蓄積していくことで, 遠い位置でも精度の良い検出を可能にしていく. また効果を実験により確認する. この手法を用いることでより自然なヒューマン・マシン・インタフェースの実現が可能である.

### 1. はじめに

人と機械をつなぐ, いわゆるヒューマン・マシン・インタフェースとして, 近年はタッチパネルを用いたものが盛んに用いられているようになっていく. この操作の直感性をさらに推し進めた, 身振り・手振りによる機器操作であるジェスチャインタフェースは, 操作者がリモコン等のデバイスを持つ必要も, 機器に触れる必要もないという点で優れていると考えられる.

このジェスチャインタフェースを実現するための手法として, 撮像画像から直接に操作者もしくはその身体の一部を切り出して, その動作を認識させるもの[1][2]や, 距離画像を用いて操作を認識するものがある. 距離画像は TOF センサなどの専用デバイスを用いて求めているもの[3]や, ライトフィールド情報から求めているもの[4]などがある. しかしいずれの手法も, 操作者が撮像デバイスの正面の特定の位置に立って大きな身振りをする必要があるため, 自然なインタフェースと言うには難点がある.

我々はこのに対して, 部屋の中に複数台のカメラを設置して操作者の手振りを検出し, この検出位置でジェスチャ操作を行うシステムを提案して操作位置の制約を無くした[5][6]. しかし, 依然として大きな手振り動作を継続的に行う必要があり, 操作としては不自然さが残っていた.

本研究では時間方向の周波数解析を詳細に行い, ごく小さな手振りでも確実な検出を行えるようにシステムの改善を図り, その効果を検証する.

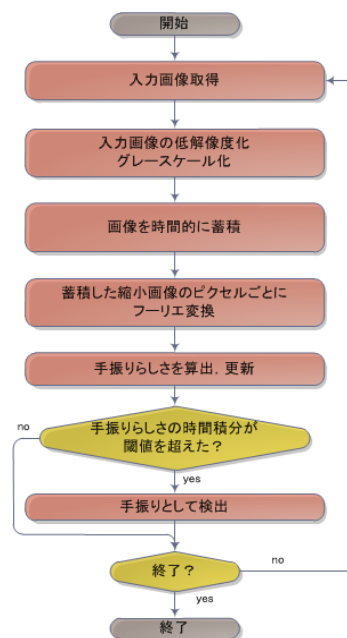


図 1 提案する手振り検出手法のフローチャート

### 2. 手振り検出の手順

我々が提案している手振りの検出手法は以下の手順で行われる(図 1, 2 参照).

- ・ 演算量の低減のため入力画像の解像度を下げ, グレースケール化する.
- ・ 低解像度画像の各画素に対して時間方向のフーリエ変換を行う. 画素の時間的な微小変動による影響を低減するために, ウィンドウ・フーリエ変換を用いる.
- ・ 得られた周波数成分を用いて手振りの有無を判断し, 検出する.

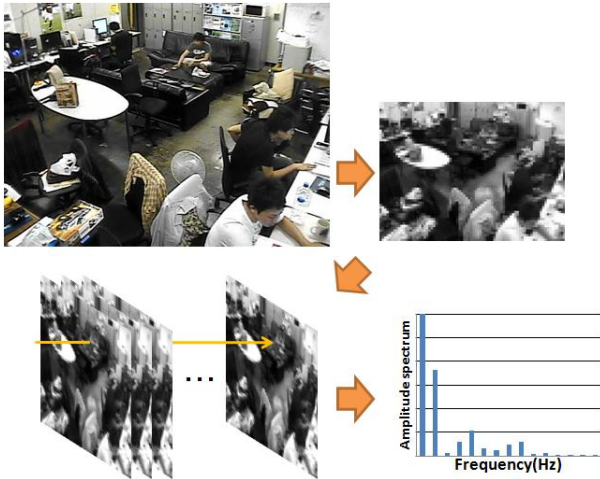
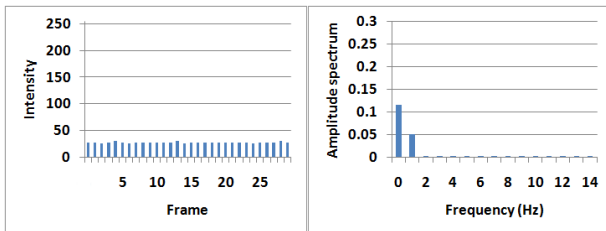
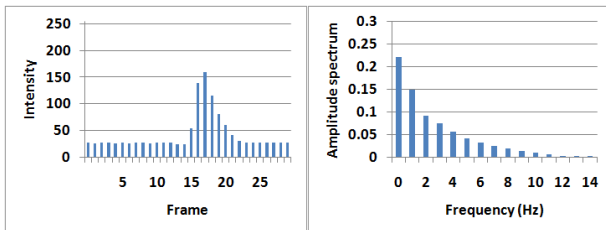


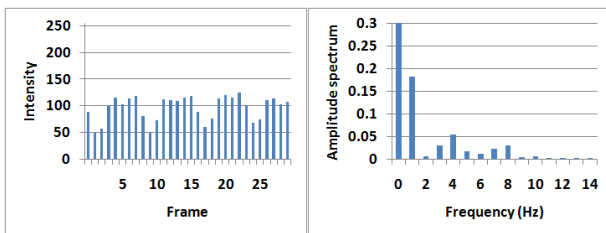
図 2 入力画像の低解像度化・グレースケール化と時間方向のフーリエ変換



(a)定常状態の輝度変化



(b)通過時の輝度変化



(c)手振りの輝度変化

図 3 低解像度画像中のあるピクセルの輝度変化とフーリエ変換結果

### 3. 手振り部分の周波数成分

低解像度画像中の、あるピクセルの輝度値の時間的な変化を並べたものを図 3 左に、これをフーリエ変換し、振幅スペクトルを取得した結果を図 3 右に示す。ここでは 30fps のカメラで 1 秒分のデータを用いた。(a)は輝度変化がほとんどない定常状態、

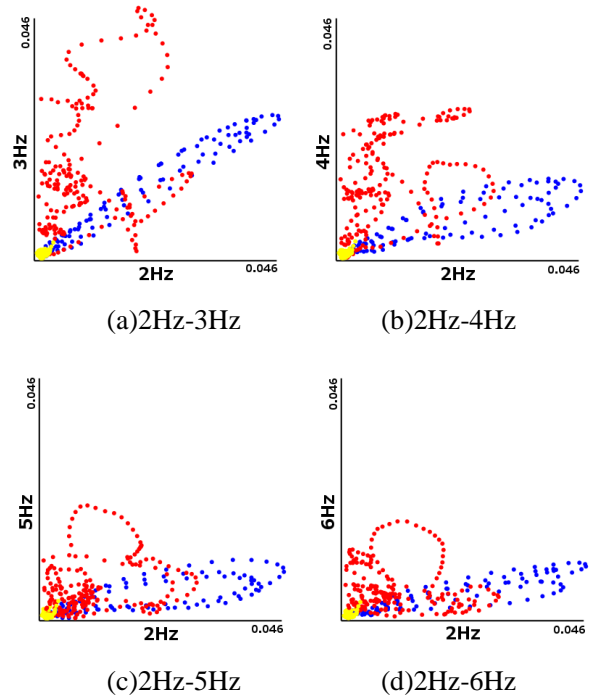


図 4 定常状態(黄)・通過時(青)・手振り状態(赤)での、2Hzと3~6Hzの振幅スペクトルの関係

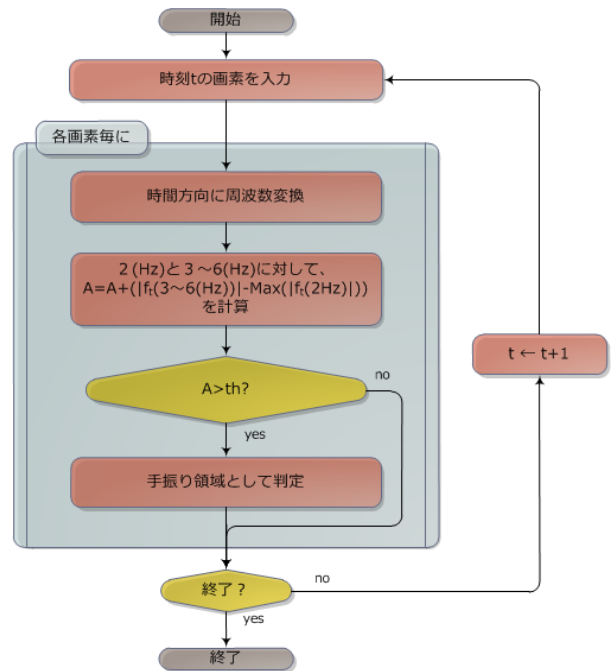


図 5 画素毎の手振り検出フローチャート

(b)は手が一度通過した状態、(c)は手振りが行われている状態である。この図から以下のことが分かる。

- ・ 定常状態でも振幅スペクトルは広がりを持つが、1Hz以下に集中している。
- ・ 手が一度通過した場合は、高周波に向かって減衰したスペクトルが発生する。

- 手振りのような繰り返し動作の場合、0Hz・1Hzを除くと、その繰り返し動作固有の周波数に強いピークが発生する。図 3 の場合おおよそ 4Hz と推定できる。

そこで、上記の(a),(b),(c)の状態を発生させ、振幅スペクトルの関係がどのようになるかを周波数毎に詳細に比較した。図 4 は 2Hz と 3~6Hz の振幅スペクトルをそれぞれプロットしたものであり、点の色はそれぞれ(a)定常状態(黄)・(b)手が一度通過した状態(青)・(c)手振り状態(赤)を示す。ここから以下のことが言える。

- 定常状態では、すべての周波数で振幅スペクトルが 0 近辺に集中する。
- 手が一度通過すると 2Hz 成分が他の成分より大きい状態が現れる。
- 手振り状態では、2Hz 成分より他の成分が大きい状態が継続する。

これらの特性を踏まえて、以下のアルゴリズムを考案した(図 5 参照)。

- 3~6Hz での振幅のピーク強度を求める。
- この強度と、2Hz での振幅強度の差を求める。
- この値を時間的に累積していく。
- これを評価値 A とし、値が閾値を超えた場合手振りがあったと判断する。

この処理を全てのピクセルで独立して行ない、手振りの検出を行う。

#### 4. 実験と考察

以上の手法を実装し、性能評価を行った。実験に用いたカメラのパラメータは表 1 の通りである。約1秒以上手振りが継続した箇所を検出するよう、先の評価値 A に加算する値に制限を設けている。低解像度化は入力画像を縦横それぞれ 1/8 の大きさに縮小した。検出処理は GPU (GeForce GTX 285)で行ない、処理時間は入力画像 1 枚あたり約 5ms であった。

間隔を空けて三度手を大きく動かし、その後小さく手を振った時の振幅スペクトルと評価値 A の遷移を図 6 に示す。手を大きく動かした時は手の通過に対応してスペクトルの三つの山があらわれるが、評価値 A はほぼ 0 で変化していない。一方、小さく手を振った区間では若干の遅延を持って A が上昇している。この結果より、適当な閾値を設定すれば手振りのみの検出が可能であると考えられる。

続いて適当な閾値を設定し、成人男性が手首をテーブル等に乗せた状態で、手を上下する動きを検出

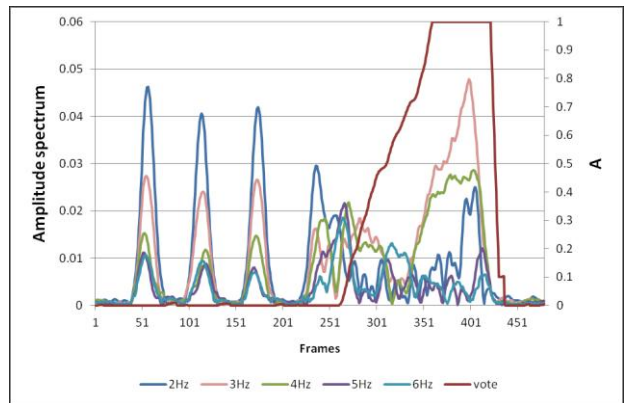


図 6 周波数毎の振幅スペクトルと、提案手法による評価値 A の変化

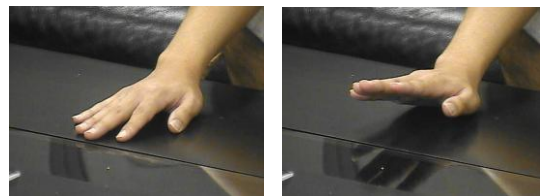
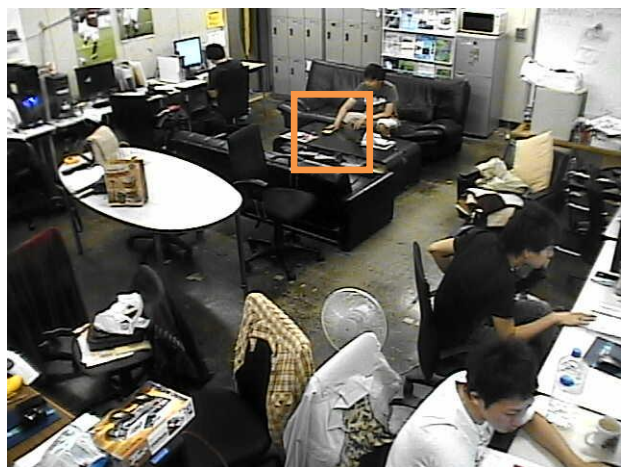


図 7 距離 6m での検出実験の様子と、検出対象である上下運動の様子

表 1 実験に用いたカメラのパラメータ

画素数	640x480
フレームレート	30fps
画角 (メーカー公表値)	水平 55.8° 垂直 43.3°

表 2 検出に要した時間(s)

カメラからの距離	手振りの速度(Hz)		
	3Hz	4Hz	5Hz
2m	1.47	1.39	1.51
4m	2.25	1.83	2.06
6m	2.36	2.30	2.55

させた。この様子を図 7 に示す。上下する大きさは約 5cm であった。カメラからの距離を変えながら、3・4・5Hz の手振りをそれぞれ 10 回検出させたところ、全ての場合で検出に成功し、小さな手振り検出が実現できていることが確認できた。この際、被験者はメトロノームの音に合わせて手振り速度を調整した。また、検出までに要した時間の平均は表 2 の通りである。提案手法は高速演算のために低解像度画像を用いており、カメラからの距離が遠くなれば検出性能が落ちることが予想されたが、検出時間が多少延びる程度であった。

また、実験中誤検出が起きやすい事象は、チェック・ボーダー等、空間方向に高周波成分が強い服装の人物が移動している場合や、カメラの近くを人物が通過する場合であった。本手法は低解像度化した画像の画素毎に、時間方向のみに着目した処理を行っている。このため、空間的に周波数成分を持った物が連続して通過する状態と、対象とする手振り動作との違いを区別することが原理的に不可能である。しかしこれらの誤検出は、複数の画素で同時に発生する可能性が高い。このため、ガウシアンピラミッドのように入力画像を階層化して検出に制限を設けることで誤検出を排除できると考えられる。

## 5. まとめ

本研究では、低解像度化したグレースケール画像をフーリエ変換し、定常状態・手が一度通過した状態・手振り状態それぞれの振幅スペクトルを取得した。この振幅スペクトルの特徴を状態毎に分析し、特定の振幅スペクトルの差分を蓄積していくことで、小さな手振りをも検出可能な手法を提案した。また、この手法によりカメラから遠い位置であっても手振り検出が可能なことを確認した。

今後の展望としては、誤検出の低減、複数カメラによる手振りの空間位置取得などが挙げられる。

## 参考文献

- [1] VI Pavlovic, R Sharma, TS Huang: Visual interpretation of hand gestures for human-computer interaction: A review, IEEE Trans. on PAMI, vol. 19, no. 7, July 1997.
  - [2] M Bayazit, A Couture-Beil, G Mori: Real-time Motion-based Gesture Recognition using the GPU, IAPR Conf. on MVA 2009.
  - [3] E Kollorz, J Penne, J Hornegger, A Barke: Gesture recognition with a Time-Of-Flight camera, Int. J. Intell. Syst. Technol. Appl. 5, 3/4 (Nov. 2008), 334-343.
  - [4] M Hirsch, D Lanman, R Raskar, H Holtzman: BiDi screen: depth and lighting aware interaction and display, International Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 2009.
  - [5] 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.
  - [6] 高橋真人, 入江耕太, 寺林賢司, 梅田和昇: 周期運動検出に基づくジェスチャ認識, 日本ロボット学会誌, vol.28, no.6, pp.756-765,2010.
- 浅野秀胤: 2000 年横浜国立大学工学部電子情報工学科卒業, 2002 年同大学院電子情報工学専攻修士課程修了. 2002 年よりパイオニア(株)研究開発部に所属し, 映像処理関連の研究に従事.
- 織茂達也: 1984 年明治大学工学部電子通信工学科卒業, 同年日本電気ホームエレクトロニクス(株)入社. 開発研究所にて HDTV 関連の研究開発に従事. 1996 年以降, PDP 用映像信号処理 ASIC 開発に従事. 現在, パイオニア(株)研究開発部に所属し, 映像処理関連の研究に従事.
- 高橋真人: 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. 日本機械学会, 日本ロボット学会等の学生会員.
- 寺林賢司: 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, 日本ロボット学会, 日本機械学会, IEEE等の会員.
- 太田睦: 1981 年京都大学工学部数理工学科卒, 1983 年同修士課程卒, NEC C&C 研究所等で画像信号処理, 主として圧縮符号化の研究を行う. 2000 年東京大学より博士(工学)の学位取得. 現パイオニア株式会社研究開発部. 電子情報通信学会, 映像情報メディア学会, IEEE 会員.
- 梅田和昇: 1989 年東京大学工学部精密機械工学科卒, 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 年同教授, 現在に至る. ロボットビジョン, 画像処理の研究に従事. 博士(工学). 日本ロボット学会, 精密工学会, 日本機械学会, 電子情報通信学会, IEEE等の会員.