

# フーリエ変換を用いた指振り検出と機器操作への応用\*

浅野秀胤\*\* 永易 武\*\*\* 織茂達也\*\* 寺林賢司† 太田 睦\*\* 梅田和昇††

Detection of Finger-Waving Movements using Discrete Fourier Transform and its Application for Equipments Control

Hidetsugu ASANO, Takeshi NAGAYASU, Tatsuya ORIMO, Kenji TERABAYASHI, Mutsumi OHTA and Kazunori UMEDA

In this paper, a method to detect slight finger-waving from images is proposed. The proposed method converts input images to low-resolution intensity images. To obtain the amplitude spectrum, the method applies windowed-Fourier transform to time series of each pixel in the low-resolution images. The difference value of the amplitude spectrum between 2Hz and the peak of 3 to 6Hz is accumulated as an evaluation value. If the evaluation value becomes above the threshold value, the pixel is detected as a point with finger-waving movement. The proposed method can detect even slight finger-waving. A gesture interface system using the proposed method with multiple cameras is constructed. The system obtains 3D position of the finger-waving. The position is related to some command and the system controls the equipment.

**Key words:** gesture interface, human machine interface, image processing, Fourier transform, Intelligent Room, network camera

## 1. はじめに

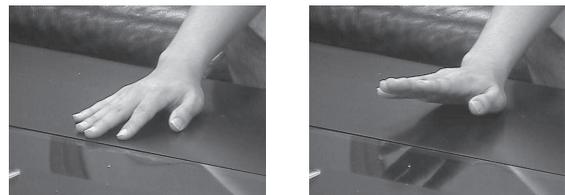
近年、人間の住環境に存在する機械が複雑化・多種多様化し、より簡易に機器を操作できるヒューマン・マシン・インタフェースへの要求がますます高まっている。中でも、身振り・手振りによる機器操作であるジェスチャインタフェースは、操作者がリモコン等のデバイスを持つ必要も機器に触れる必要もないことから、いつでも・どこでも・直感的な動作で簡単に機器を操作するインタフェースを実現できるという点で優れている。

ジェスチャインタフェースを実現するための手段として、様々な手法が提案されている<sup>1)~8)</sup>。画像を用いた手法として、肌色情報を予め与え、肌色領域の形状や動きからジェスチャを認識するもの<sup>1)2)</sup>や、オプティカルフローを用いたもの<sup>3)</sup>、背景差分などにより人物を抽出して動作を推定するもの<sup>4)</sup>などがある。また、距離画像を用いた手法として、距離の分布から人物を抽出し、ジェスチャの認識を行うもの<sup>5)</sup>、手領域の重心位置にのみ着目した手法<sup>6)</sup>などがある。これら画像を用いた手法では、操作者は操作デバイス等を持つ必要がなく、いつでも機器操作が可能であるが、カメラから特定の範囲内で、カメラに正対してジェスチャを行う必要があり、場所についての制限が依然存在している。位置センサや加速度センサを用いたものとして、グローブ型のデバイスを装着し、手の空間位置を取得することでジェスチャを認識する手法などがある<sup>7)8)</sup>。これらの手法では、操作者はいつでも・どこでも機器操作が可能であるが、センサを装着する必要がある。

我々は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、続けて行われるジェスチャを検知するシステムを提案して、いつでも・どこでも・直感的に行う操作を実現した<sup>9)10)</sup>。しかし、大きな手振り動作を継続的に行う必要があり、自然なインタフェースと言うには難点がある。また、一つの操作を行



(a) An example of target image



(b) Close-up images of a finger-waving movement

Fig. 1 Target images

うにも複数のジェスチャを特定の順序で行う必要があり、簡易なインタフェースとは言いがたい。

本論文ではより小さな動作による自然なインタフェースとして、指を上下する程度のごく小さな動きである指振りの検出手法を提案する(図1)。周波数解析を詳細に行い、従来手法を指振りのような小さな動作であっても確実な検出が行えるように改善し、その効果を検証する<sup>11)</sup>。また、改善手法を複数カメラに適用し、検出結果を組み合わせることで三次元空間位置を取得して、特定空間内で指振りが行われた際に家電を操作する簡易なジェスチャインタフェースシステムを構築し、その性能を検証する<sup>12)</sup>。

\* 原稿受付 平成 24 年 10 月 18 日

\*\* バイオニア株式会社 (神奈川県川崎市幸区新小倉 1-1)

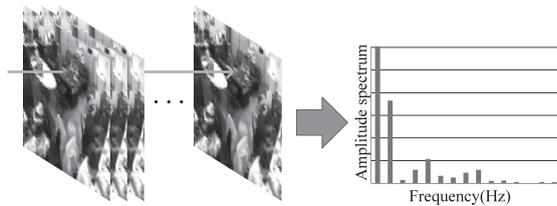
\*\*\* 中央大学大学院 (東京都文京区春日 1-13-27)

† 正 会 員 静岡大学 (静岡県浜松市中区城北 3-5-1)

†† 正 会 員 中央大学



(1) Input image (2) Convert to low-resolution intensity image



(3) Apply windowed-Fourier transform to time series of each pixel to obtain an amplitude spectrum

Fig. 2 Converting process from input image to amplitude spectrum

2. では指振りの特徴と検出アルゴリズムについて説明し、提案手法の評価を行う。3. ではネットワークカメラを用い、指振り検出を用いたジェスチャインタフェースシステムを提案し、性能を評価する。また、4. ではまとめと今後の展望について述べる。

## 2. 指振りの特徴と検出アルゴリズム

我々が提案している指振り検出は、画像から手や指の検出を行うことはせず、指振りにより起きる輝度の周期的な変化に着目し、これを検出することにより実現する。このため、以下の手順で検出処理を行う (図 2 参照)。

- (1) 入力画像をグレースケール化し、解像度を下げる。解像度を下げることで演算量とノイズの低減が可能であるが、下げ過ぎた場合指振り検出の感度が低下する。一方で解像度をあまり下げない場合、遠くの指振りなど細かい動きを検出可能となるが、ノイズに敏感になり誤検出が増加する。このため、低解像度化する割合は、入力画像のノイズ量と対象とする指振りの大きさや距離によって適切に設定する必要がある。
- (2) 輝度変化の周期性解析のため、低解像度画像の各画素で独立して時間方向の窓フーリエ変換を行う。過去  $N$  フレームの低解像度画像の、同画素位置のデータに窓フーリエ変換を適用し、振幅スペクトルを取得する。 $N$  は検出対象とする指振りの周波数に応じて適切に設定する。
- (3) 得られた振幅スペクトルから、指振りの評価値を更新し、閾値による検出処理を行う。

### 2.1 周波数軸での指振りの特徴

画像内で起こる、特徴的な状態を図 3 の白丸で示す。図 3(a) は輝度変化がほとんどない定常状態、図 3(b) は手が一度通過した状態、図 3(c) は指振りが行われている状態である。低解像度画像の白丸内のある 1 画素が取る輝度値の時間的な変化を図 4 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



(a) Static region (b) Region with a passing hand



(c) Region with a finger-waving movement

Fig. 3 Various regions (white circle)

を図 4 右に示す。ここでは 30fps のカメラで 1 秒分のデータを用い、 $N=30$  とした。この値は実験的に定めた。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。
- 指振りの場合、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 4(c) の場合およそ 4Hz と推定できる。

そこで、上記の (a),(b),(c) について、振幅スペクトルの関係を周波数毎に詳細に比較した。図 5 は、低解像度画像の 1 画素上でそれぞれの状態を起こして取得した振幅スペクトルを、2Hz と 3~6Hz でそれぞれプロットしたものであり、点はそれぞれ定常状態:丸、手が一度通過した状態:三角、指振り状態:四角を示す。ここで、3~6Hz は指振りが起こる輝度変化の周波数である。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強度が 0Hz 近辺に集中する。
- 手が一度通過すると 2Hz の振幅強度が 3Hz 以上の振幅強度より大きい状態が現れる。
- 指振り状態では、3~6Hz のピーク強度が 2Hz の振幅強度と比較して大きい状態が継続する。

### 2.2 指振り検出アルゴリズム

以上の特性から、3~6Hz のピーク強度と 2Hz の振幅強度の関係で指振りの検出が可能であると考えられる。そこで、以下のアルゴリズムを考案した (図 6 参照)。

- (1) 3~6Hz での振幅強度のピークを求める。

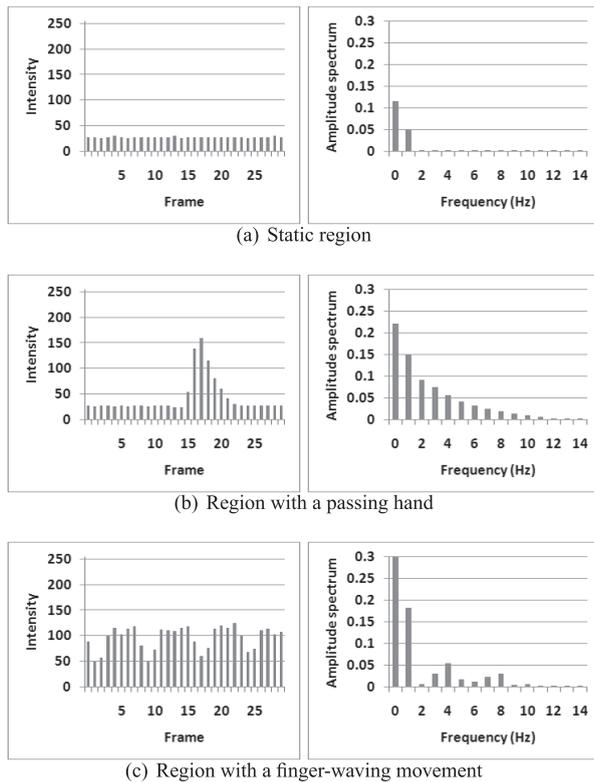


Fig. 4 Time sequence of a pixel intensity and the results of Fourier transform of a pixel in low-resolution images

- (2) このピーク強度と、2Hz での振幅強度の差を求める。この値はピーク強度が大きい場合には正の値を、2Hz の値が大きい場合には負の値を取る。
- (3) この値を時間的に累積していく。
- (4) 累積値が閾値を超えている場合指振りが行われていると判断する。

この累積値を評価値 A とする。上記の処理を全てのピクセルで独立して行ない、指振りの検出を行う。本手法は継続して動作するシステムを想定しているため、フレーム毎に A の値を更新し、検出判定を行う。

### 2.3 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は Axis 233D で、主なパラメータは表 1 の通りである。約 1 秒以上指振りが継続した箇所を検出するよう、先の評価値 A の上昇値に制限を設けている。一般的な成人の指の長さが 6~8cm であることと<sup>13)</sup>、用いたカメラの画角から、6m 程度離れた位置で指が 1 画素程度の大きさになるよう、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。評価値 A は 0 から 1 の間を取るものとし、検出の閾値を実験的に 0.95 に設定した。また、画面内の手の大きさは、画面全体の 1/10000 から 1/100 程度の面積を想定している。検出処理を GPU (GeForce GTX 285) で行なったところ、処理時間は 1 フレームあたり約 5ms であった。

間隔を空けて三度手を大きく動かし、その後小さく指を振った時のある画素での振幅スペクトルと評価値 A の遷移を図 7 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

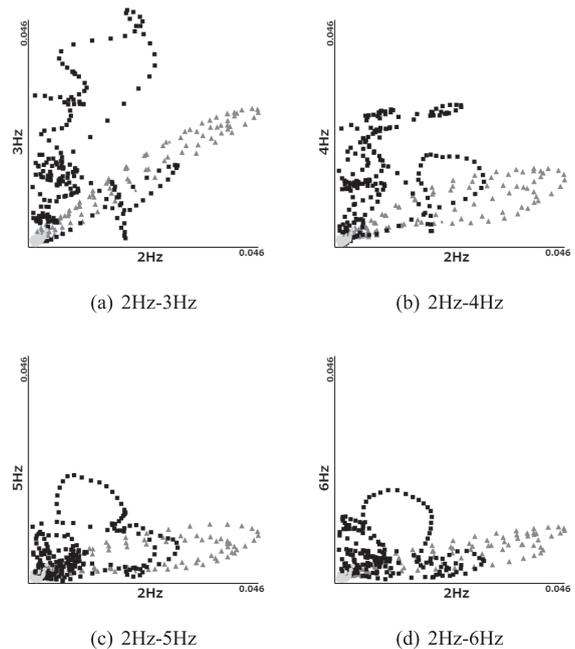


Fig. 5 Relation between amplitude spectrum of 2Hz and 3-6Hz in static region (circle), region with a passing hand (triangle), region with a finger-waving movement (rectangle)

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、小さく指を振った区間では指振り開始後若干遅延して 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指振りのみの検出が可能であると考えられる。続いて閾値を設定し、成人男性が手首をテーブル等に乗せた状態で、指を上下する動きを検出させた。上下する大きさは約 5cm であった。被験者はメトロノームの音に合わせて指振り速度を調整した。カメラからの距離を変えながら、3・4・5Hz の指振りをそれぞれ 10 回検出させたところ、全ての場合で検出に成功し、指振り検出が実現できていることが確認できた。また、検出までに要した時間の平均は表 2 の通りである。提案手法では入力画像を低解像度化しており、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、検出時間が最大で 69% 延びるのみで、検出性能に影響はなかった。カメラから遠くなることで、低解像度画素 1 ピクセルに対して指領域が占める割合が低下するため、指振りに起因する低解像度画素の輝度変化が小さくなる。このため、距離が遠くなると振幅強度全体が小さくなり、1 フレーム当たりの評価値 A の増加量が小さくなる。この結果、検出に要する時間が延びている。しかし、ピーク強度と 2Hz の振幅強度の大小関係は変化しないため、検出率には影響を与えていない。

続いて、部屋の明るさによる検出精度の比較を行った。部屋の明るさの変化のうち、明るくなった場合については、カメラの絞り等で容易に対応が可能であるが、暗くなった場合については対応が困難である。このため、部屋が暗くなった場合についてのみ実験を行った。カメラの露光を、画像中央付近の明るさが 450lx の時適切となるよう設定し、部屋の明るさを変化させた。指振りは画像中央付近で行い、カメラからの距離は 5m であった。指振り箇所の明るさは 450lx, 315lx, 200lx, 160lx, 70lx, 43lx, 9lx, 4lx の 8 通りで、それぞれ 20 回実験を行った。明るさによる入力画像の変化を図 8 に、検出時間と標準偏差の

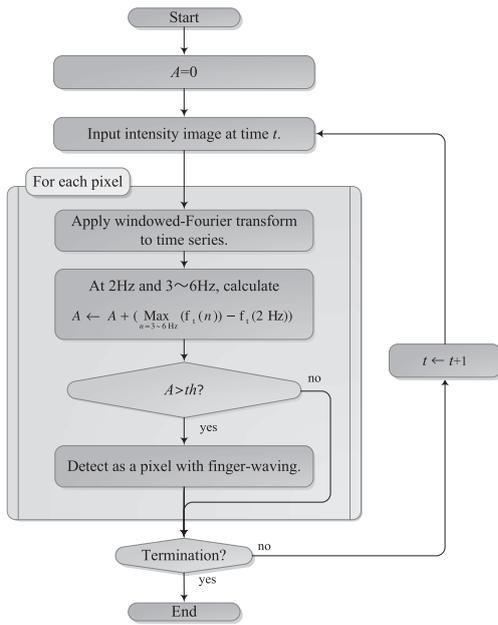


Fig. 6 The flowchart of finger-waving detection in each pixel.  $f_i(n)$  shows amplitude spectrum of  $n$ Hz.

Table 1 Camera parameters in the experiment

Resolution	640 × 480	
Frame rate	30fps	
Angle of view	horizontal	55.8°
	vertical	43.3°

変化を表 3 に示す。4lx の環境下では画像に変化が生じず、指振りを検出することが出来なかったが、それ以外の場合については全て検出することが出来た。この結果から、本手法は明るさの変化に頑健であると言える。特に 9lx の環境下では、指領域がほぼ見えないにもかかわらず、指を振り続けることで検出が可能であった。また暗くなった際には、振幅強度が小さくなるため検出にかかる時間が伸びるが、検出率が変わらず、明るさが検出の限界に達すると急激に検出率が下がる、という点が本手法の大きな特徴である。

また、様々な実験の中で誤検出が発生したケースは、チェック・ボーダー等の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。カメラの近くを人物が通過する例として、カメラから 2m の位置の人が 5m の位置まで移動し、指振りを行った後 2m の位置まで戻るといふ 443 フレームのシーケンスにおいて本手法を適用したところ、135 フレームの指振り動作のうち 105 フレームの検出に成功したが、全体で 23 点誤検出が発生した。本手法では、低解像度化した画像の画素毎に、時間方向のみの処理を行っているため、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、対象とする指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。このため、誤検出の低減には別の手法などと組み合わせる必要がある。

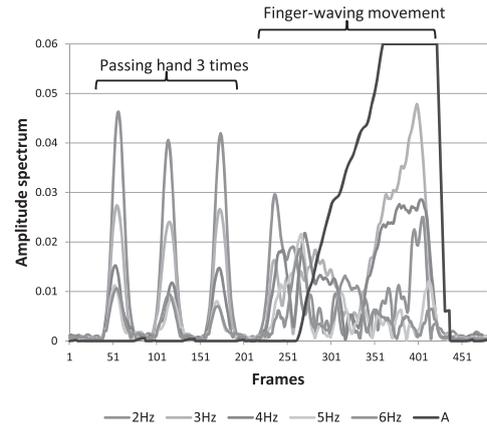


Fig. 7 Time sequence of amplitude spectrum for each frequency and evaluation value A

Table 2 Detection time for different measurement distance s

	Finger-waving frequency			
	3Hz	4Hz	5Hz	
Distance from camera	2m	1.47	1.39	1.51
	4m	2.25	1.83	2.06
	6m	2.36	2.30	2.55

### 3. 指振りの空間位置検出とそれを用いた機器操作

2. で示した指振り検出手法を用い、ジェスチャインタフェースを実現する。特定の空間内で手振りが行われた際に機器操作を行うという、極めて簡易であるが容易に利用可能なジェスチャインタフェースシステムを構築する<sup>14)</sup>。

#### 3.1 ネットワークカメラを用いた空間位置検出

指振り検出を部屋に設置した複数のカメラで並行して行い、指振りが行われた三次元位置を算出する。

三次元位置の算出に必要なカメラパラメータを予め取得しておく。それぞれのカメラ画像で指振り検出を行う。2 台以上のカメラで、互いのエピポーラ線にある検出点の組を選択し、三角測量の原理でその三次元位置を算出する<sup>15)</sup>。

ここで、特定の三次元位置と家電機器の操作を関連付けておく。この三次元位置で指振りが行われた場合に家電操作コマンドを送信して、機器の操作を行う。3 台以上のカメラで指振りが検出された場合、複数の三次元位置が得られるが、複数のユーザが同時に機器操作を行う場合を考え、全ての組み合わせを算出し、家電機器の操作と関連付けられた位置かを判定する。

#### 3.2 評価実験と結果

以上の手法に基づき実験システムを構築し、評価を行った。本実験では 6.9m × 7.8m の大きさの部屋に、2.3 で用いたカメラを部屋の隅に配置した (図 9, 表 4)。この条件で、部屋の 5 箇所検出精度の検証を行った。複数の三次元位置が得られた場合、最も早く検出された点を検出位置とした。それぞれ 30 回検出を行った結果を表 5 に示す。本システムは数 cm の誤差で検出できており、高い精度で三次元位置が算出できている。

得られた指振りの三次元位置の検出精度と手の大きさから、ひとつの機器操作に割り当てる空間範囲は一辺約 30~40cm の

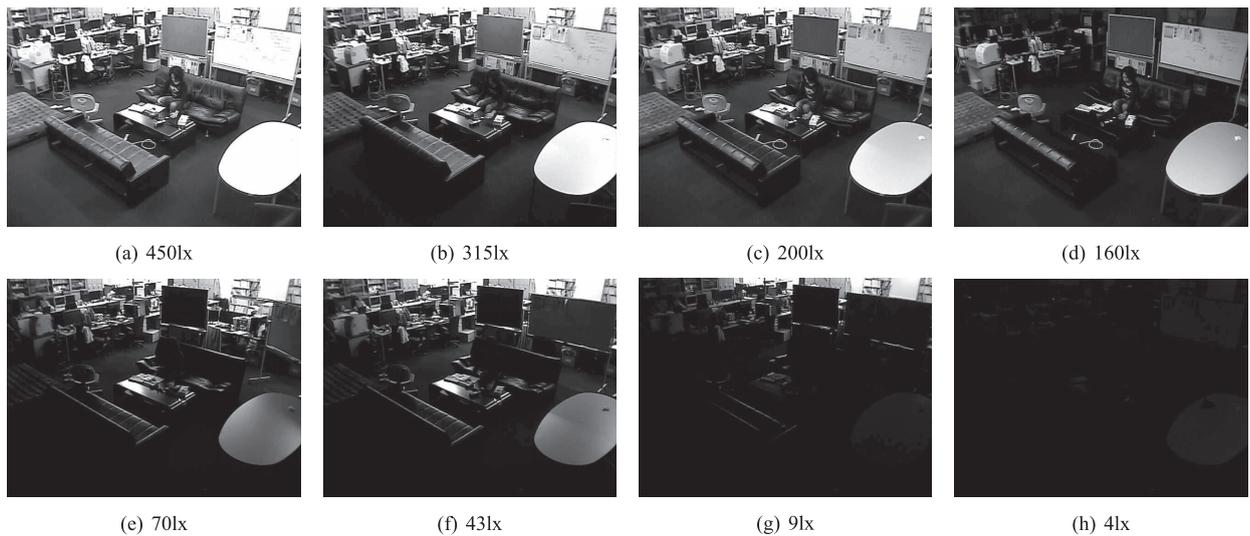


Fig. 8 Images with various brightness

Table 3 Detection rate and detection time for various brightness

Illuminance lx	Detection rate %	Average detection time s	Standard deviation s
450	100	1.2	0.1
315	100	1.2	0.1
200	100	1.2	0.1
160	100	1.3	0.1
70	100	1.9	0.2
43	100	2.9	0.6
9	100	10.9	3.2
4	0	-	-

直方体を目安とし、Xが2.0~3.2m, Yが5.0~5.6m, Zが0.3~1.0mの三次元空間位置に家電操作を配置した(図10)。なお、図中の座標系の原点位置は正確ではない。操作位置には目印として、A4サイズの紙に操作内容を印刷したものを用意している。このシステムで、それぞれの操作を50回行い、正しく検出した割合と検出にかかった時間を表6に示す。全体として高い割合で検出が実現できているが、Ch.3のみ誤検出が多く発生している。今回の実験で操作者はTV On/Off側に位置し、Ch.3の操作の際にはTV On/Offを超える形で腕を伸ばし、指振り動作を行った。この際に腕部分が振動し、TV On/Offと誤認識した事が誤検出の原因であると考えられる。また、平均検出時間は設定した通り1秒程度であった。コマンドにより検出時間に若干差異が出ているが、これはカメラまでの距離の違いによるものと考えられる。

以上の結果から、実験で用いた大きさの部屋であれば、一辺約30~40cmの立方体サイズ毎に一つの操作を配置し、指振り家電操作を行うシステムが実現可能であることが確認できた。

#### 4. 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時間方向にフーリエ変換し、特定の振幅スペクトルの差分を蓄積していくことで、指振りを検出可能な手法を提案した。この手法によ

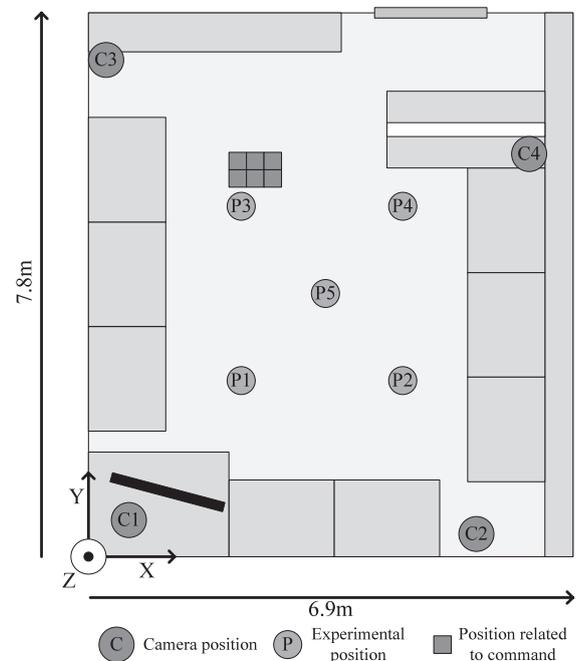


Fig. 9 Camera positions and experimental points of finger-waving detection

りカメラから遠い位置である場合や画像が暗い場合であっても指振り検出が可能であることを確認した。また、指振り検出手法を利用し、複数カメラで三次元位置を算出する事により、ごく小さな指振りでの家電操作システムを実現し、その精度を確認した。

本手法の課題としては、誤検出の低減が挙げられる。前述の通り、本手法では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、誤検出が発生する。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、ピラミッドレベルが高い画素で指振りを検出した場合はその箇所での検出を行わない、などと言った方法

Table 4 Camera positions m

	X	Y	Z
Camera1	0.59	0.53	2.11
Camera2	5.67	0.37	2.18
Camera3	0.13	7.15	2.47
Camera4	6.22	5.78	2.39

Table 5 Accuracy of position detected as finger-waving m

		True position	Average detected position	Average error	Standard deviation
Point1	X	2.30	2.34	0.04	0.04
	Y	2.60	2.59	0.01	0.05
	Z	1.00	0.99	0.01	0.03
Point2	X	4.60	4.65	0.05	0.03
	Y	2.60	2.56	0.04	0.03
	Z	1.00	0.95	0.05	0.02
Point3	X	2.30	2.22	0.08	0.03
	Y	5.20	5.15	0.05	0.01
	Z	1.00	1.00	0.00	0.02
Point4	X	4.60	4.59	0.01	0.02
	Y	5.20	5.16	0.04	0.03
	Z	1.00	0.93	0.07	0.03
Point5	X	3.45	3.40	0.05	0.03
	Y	3.90	3.81	0.09	0.06
	Z	1.00	0.97	0.03	0.06

で誤検出を排除できると考えられる。

今後の展望として、より使いやすいインタフェースの実現が考えられる。今回はインタフェースの簡易性を重視し、空間に操作を関連付けてシステムを構築した。これにより極めて簡易なジェスチャでの機器操作を実現することが出来たが、操作位置の自由が失われている。このため、人物検出や追跡と組み合わせ、人物周囲に操作を関連付けるなど、特定の空間位置にとられないインタフェースの実現が望ましい。

参考文献

- 1) V. I. Pavlovic, R. Sharma, T. S. Huang: Visual interpretation of hand gestures for human-computer interaction: A review, IEEE Trans. on PAMI, **19**, 7, (1997) 677.
- 2) S. M. Dominguez, T. Keaton and A. H. Sayed: A Robust Finger Tracking Method for Multimodal Wearable Computer Interfacing, IEEE Trans. on Multimedia, **8**, 5, (2006) 956.
- 3) M. Bayazit, A. Couture-Beil, G. Mori: Real-time Motion-based Gesture Recognition using the GPU, IAPR Conf. on MVA 2009, **11**, (2009) 9.
- 4) H. Li, M. Greenspan: Multi-scale Gesture Recognition from Time-Varying Contours, Proceedings of the Tenth IEEE International Con-

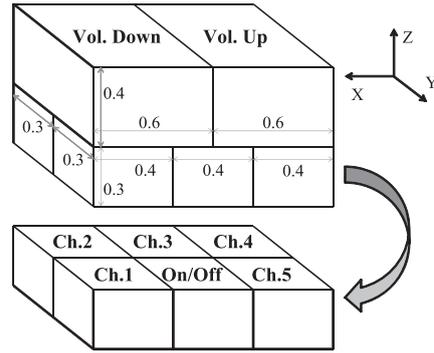


Fig. 10 3D zones related to the operations [m]

Table 6 Detection rate and detection time for each operation

	Detection rate %	Average detection time s
TV On/Off	100	1.19
Ch. 1	98	1.13
Ch. 2	100	1.21
Ch. 3	72	1.08
Ch. 4	100	1.10
Ch. 5	98	1.05
Vol. Up	100	0.92
Vol. Down	100	1.00

- ference on Computer Vision (ICCV'05), **1**, (2005) 236.
- 5) X. Liu, K. Fujimura: Hand gesture recognition using depth data, IEEE International Conference on Automatic Face and Gesture Recognition, (2004) 529.
- 6) 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, **13**, (2009).
- 7) C. Vogler, D. Metaxas: ASL recognition based on a coupling between HMMs and 3D motion analysis, Proc. of IEEE int. Conf. on Computer Vision, (1998) 363.
- 8) 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, **39**, 5, (1998) 1325.
- 9) 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, **21**, 8, (2003) 923.
- 10) 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, **28**, 6, (2010) 756.
- 11) 浅野 秀胤, 織茂 達也, 高橋 真人, 寺林 賢司, 太田 睦, 梅田 和昇: フーリエ変換を用いた小さな手振りの検出, ビジョン技術の実利用ワークショップ 2010(ViEW2010), (2010) 264.
- 12) 浅野 秀胤, 織茂 達也, 永易 武, 寺林 賢司, 太田 睦, 梅田 和昇: 小さな手振り検出を用いた家電操作システムの構築, 映像情報メディア学会年次大会 2011, (2011).
- 13) 河内 まき子: AIST 日本人の手の寸法データ, <http://riodb.ibase.aist.go.jp/dhbodydb/hand/index.html> 2012.10.18 アクセス
- 14) 新妻 実保子, 橋本 洋志, 橋本 秀紀: 空間メモリ: 知識活用を支援する空間知能化, 計測自動制御学会論文誌, **42**, 4, (2006) 367.
- 15) 徐 剛, 辻 三郎: 3次元ビジョン, 共立出版, (1998).