

インテリジェントルームの構築

-直感的なジェスチャを用いた家電製品の操作-

若村 直弘[†] 鈴木 健一郎[†] 入江 耕太^{††} 梅田 和昇^{†††}

[†] 中央大学大学院 理工学研究科 〒 112-8551 東京都文京区春日 1-13-27

^{††} 株式会社 日立製作所 〒 312-8503 茨城県ひたちなか市大字高場 2520

^{†††} 中央大学 理工学部 〒 112-8551 東京都文京区春日 1-13-27

E-mail: [†]{wakamura,ksuzuki}@sensor.mech.chuo-u.ac.jp, ^{††}kota.irie.ja@hitachi.com,
^{†††}umeda@mech.chuo-u.ac.jp

あらまし 本論文では、ジェスチャ認識に基づいて操作位置に依らないインテリジェントルームを提案している。まず手振りを検出することで操作者の意思及び位置を特定し、カメラを制御することでジェスチャ認識処理を行うための候補領域を絞る。肌色情報に基づいて手領域を抽出し、指の本数や方向、手領域の動きから様々なジェスチャの認識を行う。直感的なジェスチャを用いてテレビなどの家電製品の操作を行っている。

キーワード インテリジェントルーム、ジェスチャ認識、マンマシンインタフェース、画像処理、家電製品の操作

Construction of an Intelligent Room

-Operation of Electric Appliances with Intuitive Gestures-

Naohiro WAKAMURA[†], Kenichiro SUZUKI[†], Kota IRIE^{††}, and Kazunori UMEDA^{†††}

[†] Graduate School of Science and Engineering, Chuo University 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan

^{††} Hitachi Ltd. 2520 Oazatakaba, Hitachinaka-shi, Ibaraki 312-8503, Japan

^{†††} Faculty of Science and Engineering, Chuo University 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan

E-mail: [†]{wakamura,ksuzuki}@sensor.mech.chuo-u.ac.jp, ^{††}kota.irie.ja@hitachi.com,
^{†††}umeda@mech.chuo-u.ac.jp

Abstract This paper proposes an intelligent room that is free of operator's position based on gesture recognition technologies. Intention and position of an operator are recognized by detecting hand waving, and pan-tilt cameras are zoomed and focused on the operator. The hand region is extracted using color information, and direction and number of fingers and motion of the hand region are detected. Home appliances such as a television set are controlled by using the intuitive gestures.

Key words Intelligent Room, Gesture Recognition, Man-Machine Interface, Image Processing, Operation of Home Appliances

1. 序 論

現在、様々な技術の発展に伴い、我々の生活環境の情報化、インテリジェント化が進んでいる。その一例として、身近な機器である家電製品のネットワーク化が実現し、利便性の向上が図られている。一方、多機能化することで、操作が複雑化するという問題が生じる。我々の周囲にある機器の多くは、ボタ

ンやリモコンを用いて操作される。しかし、ボタン操作では、操作する位置が限定されるため、不便な場合がある。また、リモコン操作では、離れた場所から操作できる利点があるが、リモコンの位置が常に特定されないため、リモコンを探してから操作を行うというように二度手間になる欠点もある。特に家電製品の様な日常生活で頻繁に使用する機器は、ユーザーが直感的に操作できることが望ましい。さらに、テレビなどの遠隔操

作に利便性を有する機器の操作に関しては、操作する位置の拘束をうけないことも要求される。このことから、人間の自然な行動を利用し、かつ非接触のインタフェースが有効であると考えられる。我々の日常的なコミュニケーション手段を考えた場合、身振り手振りなどのジェスチャを頻繁に用いている。そこで、直感的でかつ非接触での操作を可能にするマン・マシン・インタフェースの一つとして、ジェスチャが挙げられる。これまでに、動画像からジェスチャを認識する研究が数多く報告されている [1]~[5]。また、それらのジェスチャ認識技術を用いて部屋全体を智能ロボット化したインテリジェントルームの研究も行われている [6]~[9]。しかし、これらの多くは、ジェスチャを認識できる場所が限定されているため、生活空間での実用性に欠ける。背景差分の適用により動作者の特定せずにジェスチャ認識を行う手法 [10] など提案されているが、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。一方、入江ら [11] は手振り動作を認識することで、複数の人物が存在する環境下においてもロボットの動作者及び動作者の位置の特定を行っている。そこで本研究では、入江らの手法を用いて動作者の位置の特定し、動作者の位置情報からカメラを制御することで、操作位置に依らないシステムを実現させる。利便性の向上を図るために直感的なジェスチャの認識手法を提案し、ジェスチャをインタフェースとしたインテリジェントルームの構築を行う。

2. インテリジェントルームの概要

本研究で扱うインテリジェントルームとは、室内にカメラを取り付け、部屋全体を智能ロボット化した部屋であり、一般的なオフィスや家庭などへの適用を想定している。具体的な機能としては、テレビや照明機器などの家電製品をジェスチャで操作するものとする。

本研究で提案するインテリジェントルームのシステムは、視覚となる CCD カメラを設置し、自律的に操作者の特定ならびにジェスチャの認識を行う。まず、2 台以上のカメラにより「手振りの検出」を行い、操作者を特定するとともに 3 次元位置情報を取得する。3 次元位置情報から得られる距離情報に従ってカメラをズームさせ、ジェスチャ認識処理を行う領域を注視する。続いて、色情報を用いて手領域の抽出を行う。個人差や照明環境による肌色の変化に対するロバスト性を実現させるために、検出された手の「肌色の登録」を行い、登録された肌色情報を用いて手領域を抽出する。こうして得られる手領域に対して、ジェスチャ認識を行う。まず、操作する機器を決定するために「指差し方向の認識」を行う。肌色情報を用いて指差し領域を抽出し、抽出領域の主軸ベクトルから、指差し方向を決定する。次に、決定された機器を操作するために「指の本数認識」や「手の動作認識」を行う。認識された結果は、PC のモニタ上での表示とスピーカからの音声により確認することができ、インタラクションが可能である。認識された操作内容に応じて機器を制御する。本論文で構築するシステムでは、PC に接続された赤外線リモコンを用いて機器の制御を行う。インテリジェントルームの概念図及びインテリジェントルーム内での

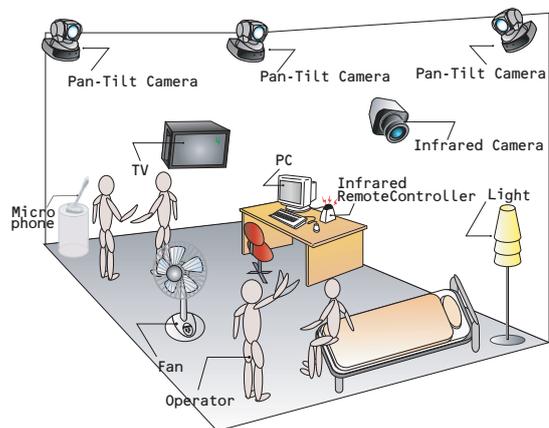


図 1 インテリジェントルームの概念図

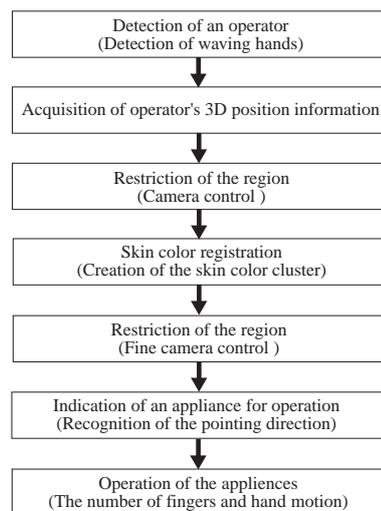


図 2 インテリジェントルーム内での処理過程

処理過程を図 1、図 2 に示す。

3. 手振り検出による操作者の特定

3.1 手振り検出の概要

インテリジェントルームにおいて操作者を特定するために、手振り動作を用いることとする。手振り動作は、人と人とのコミュニケーションにおいても呼びかけや注意の喚起に日常的に用いられているため、操作者にとってわかりやすいジェスチャであると言える。手振り動作の検出を行うことで、室内に複数の人間が存在する場合でも、操作の意図を持つ人物を特定することが可能となる。

手振り動作の検出には文献 [11] の手法を用いる。本手法の概要は次の通りである (図 3 参照)。まず画像を低解像度化する (b)。次に、低解像度化された画像の各画素の濃淡値に対して、時間軸方向の FFT を行う (c)。手振りが行われている領域では、低解像度画像の濃淡値が正弦波に近い形で振動する。よって、パワースペクトルを評価することにより (c)、容易に手振りの認識を行うことが可能である。

以上の手法は、画像の濃淡値のみを利用し色情報が不要であるため、個人差や照明環境による肌色の変化の影響を受けない。また、手領域を抽出するなどの画像処理の必要がなく、極めて

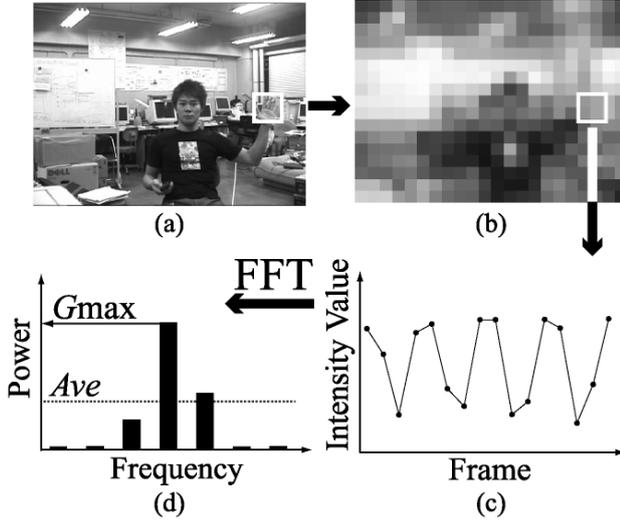


図 3 FFT を用いた手振り検出

簡潔な処理である。

3.2 ステレオ視による 3 次元位置計測

2 台の CCD カメラで手振りが検出されれば、ステレオ視によって手振りの 3 次元位置を計測することが出来る。

まず事前にカメラキャリブレーションを行い、各カメラの射影行列を求めておく。本論文では、文献 [12] の手法を用いて既知パターンを観測することによってカメラの内部パラメータと外部パラメータとを求め、射影行列を得る。

ある点の 3 次元空間での同次座標を $\tilde{M} = [X, Y, Z, 1]^T$ 、対応する画像平面上での点の同次座標を $\tilde{m} = [u, v, 1]^T$ とすると、中心射影における射影方程式は、

$$s\tilde{m} = P\tilde{M} = A[R, t]\tilde{M} \quad (1)$$

で表される [13]。ここで、

$$A = \begin{bmatrix} \alpha_u & -\alpha_u \cot \theta & u_0 \\ 0 & \alpha_v / \sin \theta & v_0 \\ 0 & 0 & 1 \end{bmatrix}, R = \begin{bmatrix} \mathbf{r}_1^T \\ \mathbf{r}_2^T \\ \mathbf{r}_3^T \end{bmatrix}, t = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \quad (2)$$

である。A の $\alpha_u, \alpha_v, u_0, v_0, \theta$ がカメラの内部パラメータ、R (回転行列)、t (並進ベクトル) が外部パラメータである。

ある点 (本論文の場合は手振り位置) が 2 台のカメラで観測されると、その点の 3 次元位置は両カメラの射影行列から求めることができる。両カメラの射影行列を P_1, P_2 、両カメラの画像平面上での点の座標を $\mathbf{m}_1 = [u_1, v_1]^T$ 、 $\mathbf{m}_2 = [u_2, v_2]^T$ とすると、

$$s_1\tilde{m}_1 = P_1\tilde{M}, s_2\tilde{m}_2 = P_2\tilde{M} \quad (3)$$

が成り立つ。 s_1, s_2 は任意のスカラーである。この 2 式から

$$B\mathbf{M} = \mathbf{b} \quad (4)$$

が導かれ、3 次元座標 M は

$$\mathbf{M} = B^+\mathbf{b}. \quad (5)$$

により求まる [13]。

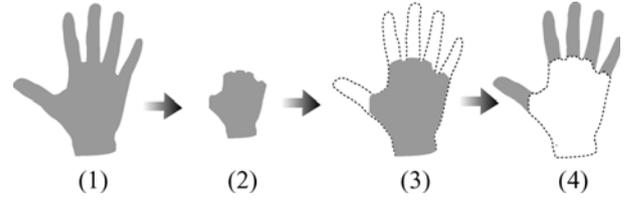


図 4 指の本数認識

4. 肌色登録

検出された手に対して肌色の登録を行う。これにより、事前に肌色の登録を行っておくのに比べ、個人差や照明環境による肌色の変化の影響を受けない口バストな手領域の抽出が実現される。肌色登録における色空間には、明るさの影響を受けにくい H (色相)、 S (彩度)、 I (明度) 空間を用いる。 HSI 空間において、明度 I を除いた HS を特徴空間とし、肌色のクラスタを形成する。認識時には、各画素で得られた H と S の値からマハラノビス距離を求め、設定した閾値以内の画素を手領域として抽出する。肌色のクラスタの平均ベクトルを $\mathbf{M} = [M_H, M_S]^T$ とし、共分散行列を \mathbf{V} とする。計測された特徴量ベクトル \mathbf{X} とのマハラノビス距離 d_M は

$$d_M^2 = (\mathbf{X} - \mathbf{M})^T \mathbf{V}^{-1} (\mathbf{X} - \mathbf{M}). \quad (6)$$

で与えられる。

5. ジェスチャ認識

5.1 指の本数認識

指の本数認識手法を示す。モフォロジー処理を利用して指領域を抽出し、その数をカウントする (図 4 参照)。

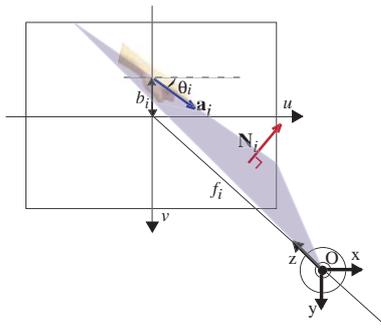
- (1) 肌色情報を用いて手の領域を抽出する。
- (2) (1) の領域に収縮処理を行う。
- (3) (2) の領域に膨張処理を行う。
- (4) (1) の領域から (3) の領域の差をとり、領域を分離し、残った領域の形状や面積の特徴から指領域のみを抽出する。

5.2 指差し方向認識

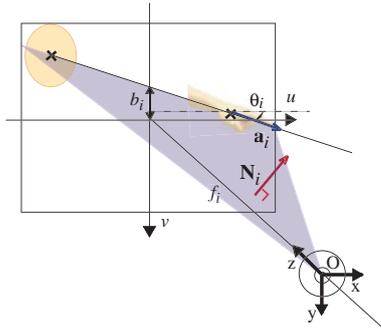
指差し方向を求めるには、カメラから取得した画像中から手領域を抽出し、手領域の主軸ベクトルを求めればよい。しかし、指差し方向認識を行う前には手領域を抽出しやすくカメラ制御を行っているが、カメラから取得した画像において手領域だけでなく顔領域も含まれている場合がある。また、指差しを行うに当たって、カメラの光軸方向に向かって指す場合やカメラに背を向けて指す場合も考えられる。そこで、カメラから取得した画像に対して顔領域が含まれている場合は、手領域だけでなく顔領域の抽出も同時に行う。

まず、カメラから取得した画像に対して、肌色情報に基づき肌色抽出を行う。抽出領域が 1 つの場合は手領域、2 つ以上の場合は領域面積の 1 番大きい領域を顔、2 番目に大きい領域を手とする。

1 つの領域が抽出された場合は、図 5(a) に示すように抽出された手領域の画像座標系での主軸 (長軸) ベクトル \mathbf{a}_i を求める。



(a) “手領域のみ”



(b) “手と顔領域”

図 5 抽出領域の主軸ベクトルの算出

また、2 つ以上の領域が抽出された場合は、図 5(b) に示すように手と顔領域のそれぞれの重心を結ぶ直線の傾きを θ_i として画像座標系での主軸ベクトル a_i を求める。カメラ座標系の z 軸はカメラの光軸と一致しているので、求めた主軸ベクトルをカメラ座標系に変換すると、

$$a_i = [\sin \theta_i, \cos \theta_i, 0]^T \quad (7)$$

となる。ここで、主軸ベクトルを含む直線の v 軸の切片を b_i とすると、カメラ座標の原点 O およびこの直線を通る平面の法線ベクトル N_i は、

$$N_i = [f_i \tan \theta_i, -f_i, b_i]^T \quad (8)$$

となる。 f_i は各カメラの焦点距離である。指差しは、図 6 に示すように、2 台のカメラで得られるこの平面の交線として求めることができる。法線ベクトル N_1, N_2 をカメラ座標系から世界座標系に変換し外積を求めることで得られるベクトル方向が指差し方向となる。

なお、本手法に近い手法として、Yoda ら [14] は腕によるポインティングでの機器の指示手法を提案している。彼らの手法では市販のステレオカメラを複数用いているが、本研究では複数の単眼カメラでステレオ計測を行うことによって指差しの 3 次元方向の認識を実現している。

5.3 手の動作認識

手の動作特徴の抽出には DP マッチング [15] を用いる。肌色

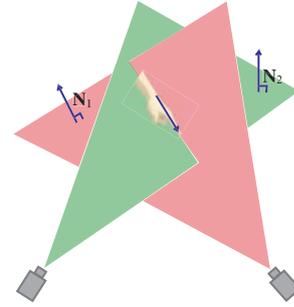


図 6 世界座標系における指差し方向

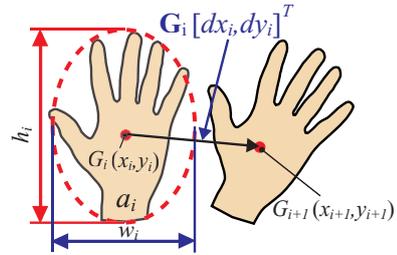


図 7 手の動作認識に用いる特徴量

情報に基づき、各フレームにおいて手領域を抽出する。抽出された手領域の形状、重心から、重心移動ベクトル dx, dy 、手領域の幅と高さの変化量 dw, dh 、面積変化量 da を求め、特徴量として用いる。図 7 に示すように、 x_i, y_i, w_i, h_i, a_i をそれぞれ i フレームにおける x 座標重心位置（水平成分座標）、 y 座標重心位置（垂直成分座標）、手領域の幅、高さ、面積とすると、各特徴量は

$$dx_i = x_{i+1} - x_i \quad (9)$$

$$dy_i = y_{i+1} - y_i \quad (10)$$

$$dw_i = w_{i+1} - w_i \quad (11)$$

$$dh_i = h_{i+1} - h_i \quad (12)$$

$$da_i = a_{i+1} - a_i \quad (13)$$

で算出される。式 (9)-(13) で算出される各特徴量を用いて DP マッチングを行い、登録されているモデルとの類似度を算出する。入力パターンを X 、ジェスチャモデルのパターンを Y とすると、類似度は X と Y のベクトル間の距離を計算することで算出することができる。なお、 X と Y の成分数は異なっても良い。モデルの全てのジェスチャとの類似度を求め、最小となる類似度がある閾値以下のとき、入力動作をそのジェスチャであると判定する。本研究で用いるジェスチャモデル及び各モデルの特徴量の時系列における変化を図 8 に示す。

5.4 判定ジェスチャの認識

認識結果が正しいかどうかを指示する OK・Cancel ジェスチャの認識手法を示す。親指と人差し指で円を作った形状、もしくは親指を立てて残りの指を閉じた形状を OK ジェスチャ、手を左右に振る動作を Cancel ジェスチャとする（図 9 参照）。肌色情報に基づいて手領域を抽出し、抽出された手領域の形状や移動量などの特徴量から認識する。それぞれのジェスチャの特徴量を表 1 に示す。

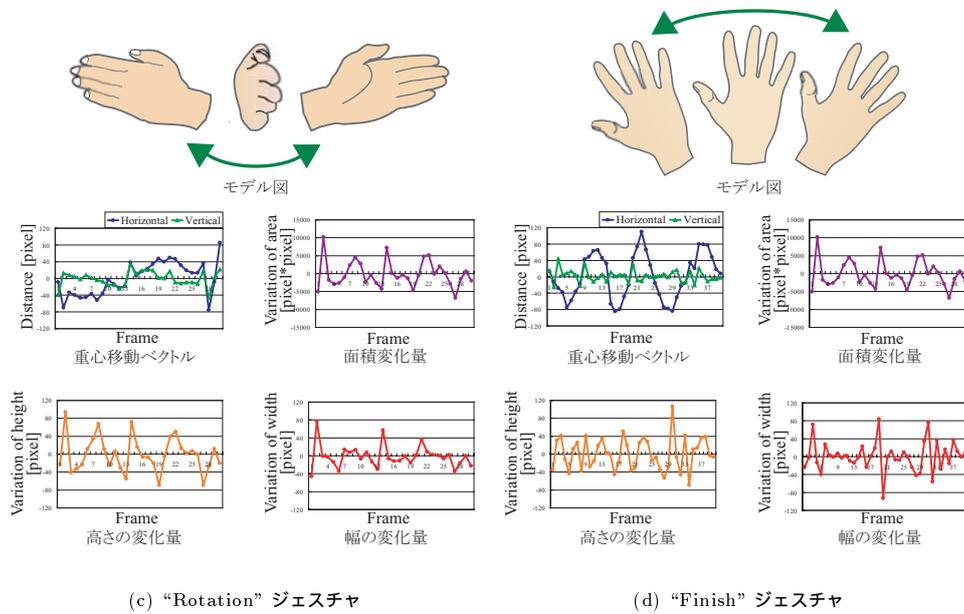
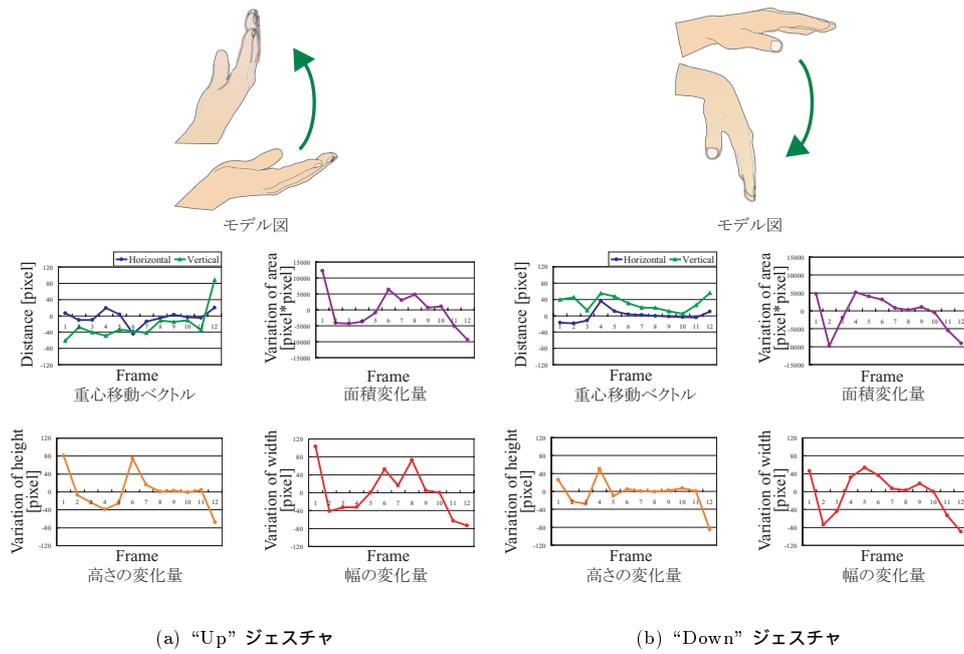


図 8 ジェスチャモデル

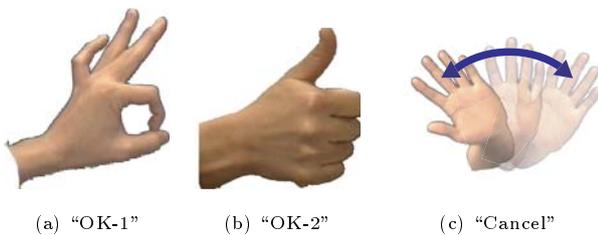


図 9 判定ジェスチャ

表 1 各判定ジェスチャの特徴

Gesture	Features
OK-1	<ul style="list-style-type: none"> Region surrounded by skin color region is circular (In case extracted continuously) Number of standing fingers are 3
OK-2	<ul style="list-style-type: none"> Number of standing fingers is 1
Cancel	<ul style="list-style-type: none"> Extracted skin color region vibrates horizontally (In case more than a threshold)

として以下のことが挙げられる。

- (1) 減衰時間が短い
- (2) 低周波から高周波まで周波数成分が存在する
- (3) 同じ音を出すのが困難である

5.5 手拍子の認識

手振りを検出する他に，暗い部屋でも操作の意思を伝えられる手段として，手拍子を認識することを考える．手拍子の特徴

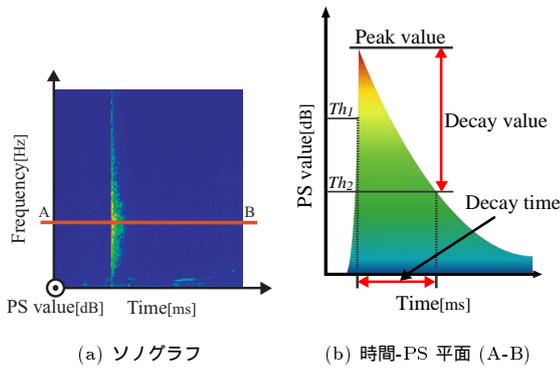


図 10 ソノグラムの概念図

以上の特徴を考慮し、手拍子の認識手法を提案する。手拍子の認識には WaveSono [16] で求められるソノグラム (図 10(a) 参照) を用いる。ソノグラムは入力波形のパワースペクトル (PS) 値を時系列に視覚的に表示したものである。WaveSono を用いる利点として、リアルタイムでの音の解析が可能、対象とする音の性質によりサンプリング周波数値の設定を適宜変更可能など、効率のよい使い方ができることが挙げられる。図 10(b) は図 10(a) を時間-PS 平面に投影した図である。これらの図から得られる以下の特徴量に対して閾値処理を行い、手拍子を認識する。

- (1) PS 値のピーク値の減衰量
- (2) 減衰時間 (PS 値が Th_1 を超えて Th_2 に減衰するまでの時間)
- (3) 高周波領域 (350 ~ 510Hz) におけるピーク値

6. ジェスチャ認識による家電製品の操作

6.1 操作対象の決定

機器を操作するためには、まず、操作対象を決定する必要がある。そこで、5.2 で述べた「指差し方向認識」を利用し、手振りの 3 次元位置を通り方向が指差し方向の直線を求め、操作候補の機器のうちこの直線との距離が最も小さいものを操作対象と決定する。本研究で構築する実験システムでは「テレビ」、 「扇風機」、 「照明機器」を操作対象とする。

6.2 家電製品の操作

操作対象を選択した後、ジェスチャ認識により各機器の操作を行う。表 2 に家電製品の操作手法をまとめる。例として、テレビを 17 チャンネルにする場合、指差し方向認識において“テレビ”を指差した後、“1”、“5”、“2”の順番で指本数認識を 3 回行う。

7. 実験

7.1 実験システム

画像処理には HALCON (MVTEC 製) を用い、各種演算処理は PC (DELL, Pentium 4 3.2GHz) で行った。Pan-Tilt-Zoom 機能を搭載した 3 台の CCD カメラ (SONY EVI-D100) で画像を撮像し、画像分割ユニット (Panasonic WJ-MS488) で合成された映像をキャプチャボード (Leutron PicPort Color) を用い

表 2 家電製品の操作

Recognition of number of fingers n_i (i -th number)		
Operation	Device	Gestures
Channel	TV	Channel $N=1 \sim 4$ $N=n_1+n_2$ ($n_1=0, n_2=N$)
		Channel $N=5 \sim 9$ $N=n_1+n_2$ ($n_1=5, n_2=N-5$)
		Channel $N=10 \sim 14$ $N=10n_1+n_2$ ($n_1=1, n_2=N-10n_1$)
		Channel $N=15 \sim 19$ $N=10n_1+n_2+n_3$ ($n_1=1, n_2=5, n_3=N-15$)
Recognition of hand motion		
Operation	Device	Gestures
Volume	TV	Up
	Fan	Down
Rotation	Fan	Rotation
Power OFF	TV	Finish
	Fan Light	and Handclap

て PC に入力した。また、家電製品を操作するために、PC で制御可能な学習型赤外線リモコン (SUGIYAMA ELECTRON クロッサム 2+USB) を使用した。手拍子の認識にはマイクロフォン (ELECOM MS-STM55) を用いた。

7.2 手振りの認識

計測距離を変化させた時の認識率に関しては文献 [11] に示してある。広い計測範囲で高い認識率で手振りの認識を実現している。

次に 2 台のカメラのステレオ計測による 3 次元位置計測の評価を行った。部屋のほぼ中心を原点に、 x 軸、 y 軸方向に 0.5[m] ずつ位置をずらして計測した。各点で 5 回、手を振る高さを変えて位置計測を行った結果を図 11 に示す。エラーバーは標準偏差を示す。カメラは y 軸方向に設置されている。この結果より、カメラから操作者までの距離が遠くなるにつれつきが小さくなっているものの、誤差の大きさは 0.1[m] 程度で部屋の中で操作者の位置を見つけるという本システムでの用途には十分な精度が実現出来ていると言える。

得られた距離に応じて、手の幅が 160[pixel] 程度になるように設定された倍率でカメラのズームを行い、以後の処理を行った。

7.3 指の本数認識

5.1 で述べた手法に基づき、被験者 5 人に対して指の本数認識を行った。まず、被験者ごとに 4. で述べた肌色登録を行い、続いて指の本数を 0 ~ 5 本に変化させ、それぞれの認識率を算出した。20 枚の画像で連続して認識を行い、認識回数が最も多かった本数を認識結果とした。指の本数ごとの認識率を表 3 に示す。全ての指の本数において良好な認識率が得られた。認識

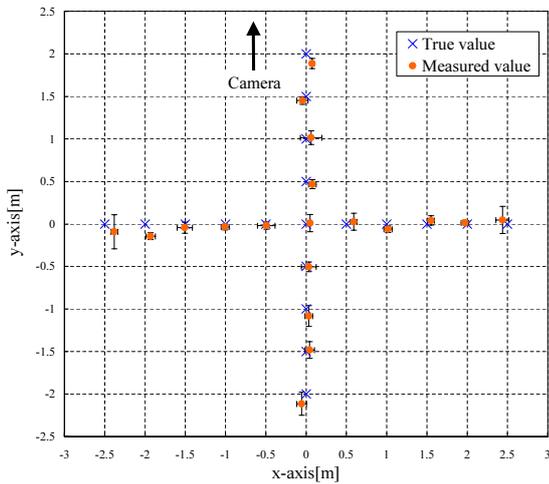


図 11 手振りの位置計測

表 3 指の本数の認識率

Number of fingers	0	1	2	3	4	5
Recognition rate	90%	98%	96%	82%	88%	88%

の失敗例として、掌が極端に下方または上方へ傾いている場合が挙げられる。

7.4 指差し方向認識

5.2 で述べた手法に基づき、指差し方向の認識実験を行った。被験者は図 12 に示すように真上から見て 2 台のカメラ (Camera2 と Camera3) 間の角度が 90[deg] になる位置に立ち、Camera3 の光軸となす角 α が 0[deg], 15[deg], 30[deg], ..., 180[deg] (15[deg] おき) で仰角 β が 0[deg] (床に水平) の方向を指差し、各角度において認識を 10 回行った。(3 次元での) 指差し方向は 2 台のカメラからの (画像上での) 指差し方向より求めるので、1 回の認識において 3 つの指差し方向 (Camera1 と Camera2, Camera2 と Camera3, Camera3 と Camera1) が得られる。3 つの結果の中で理論値に最も近い値を採用した場合の指差し方向の認識の位置精度を図 13, 表 4 に示す。手領域 (手領域と顔領域) が確実に抽出できれば、 α 方向に関しては理論値と平均との差は約 3[deg], ばらつきは 6[deg] 未満で十分な精度があるといえる。 β 方向に関しては、実際に指を差した角度よりも全体的に 7[deg] 程度低い値が得られているが、ばらつきは 4[deg] 未満で α 方向と同様、十分な精度があるといえる。 β 方向で全体的に結果が低いのは、手領域と顔領域が抽出された場合、手と顔の重心を結ぶ直線方向を指差し方向としていることが原因として考えられる。今後の課題として、1 回の認識で得られる 3 つの結果からどの結果を用いるかを考える必要がある。

7.5 手の動作認識

5.3 で述べた手法に基づき、被験者 5 人に対し各ジェスチャの認識実験を行った。ジェスチャモデルは特定の被験者のものを用いた。カメラ・被験者間の距離はモデル登録時と同じとした。まず、被験者ごとに 4. で述べた肌色登録を行い、続いて 4 種類のジェスチャに対し、認識率を算出した。各ジェスチャの認識率を表 5 に示す。誤認識の例として、手を速く動かしすぎ

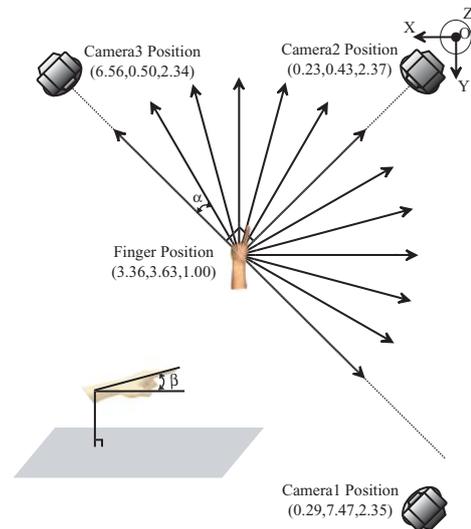


図 12 指差し方向認識の実験環境

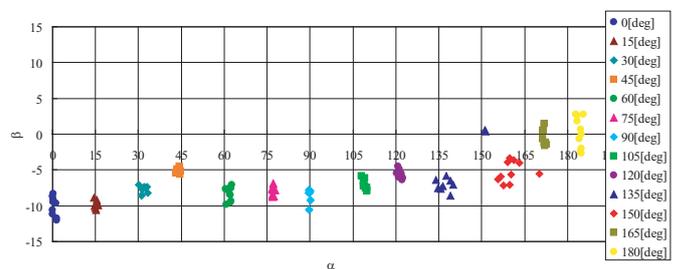


図 13 指差し方向の認識精度 (分布図)

表 4 指差し方向の認識精度 (平均及び標準偏差)

α [deg]	0	15	30	45	60	75	90
平均	0.5	15.2	31.8	44.0	61.8	77.1	89.8
標準偏差	0.5	0.5	0.9	0.5	0.8	0.3	0.2
β [deg]	0						
平均	-10.0	-9.8	-7.7	-5.1	-8.3	-7.9	-8.8
標準偏差	1.2	0.5	0.5	0.3	0.9	0.6	1.0
α [deg]	105	120	135	150	165	180	
平均	109.1	121.4	139.9	160.3	172.0	184.3	
標準偏差	0.6	0.7	5.9	3.9	0.4	0.8	
β [deg]	0						
平均	-7.1	-5.6	-5.6	-5.3	-0.6	0.3	
標準偏差	0.7	0.6	3.1	1.4	1.0	2.1	

表 5 手の動作の認識率

Gesture	Down	Up	Rotation	Finish
Recognition rate	96 %	99 %	86 %	89 %

て手領域を抽出できなかったことなどが挙げられる。

7.6 判定ジェスチャの認識

5.4 で述べた手法に基づき、被験者 5 人に対して判定ジェスチャの認識実験を行った。まず、被験者ごとに 4. で述べた肌色登録を行い、続いて OK ジェスチャ (2 種類), Cancel ジェスチャに対し、認識率を算出した。表 6 に各ジェスチャの認識率を示す。誤認識としては、OK ジェスチャに関して、ジェスチャの仕方 (円形や指の形状) に個人差があり、うまく抽出で

表 6 判定ジェスチャの認識率

Gesture	OK-1	OK-2	Cancel
Recognition rate	84.5%	85.0%	97.0%

表 7 手拍子の認識率

Times	0.1m	1.0m	2.0m	3.0m
1	98%	98%	90%	88%
2	98%	94%	90%	86%
3	96%	96%	86%	82%
4	92%	98%	90%	92%

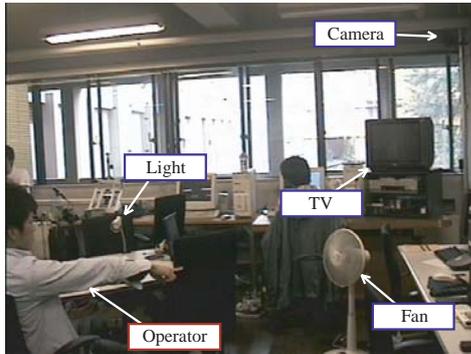


図 14 家電製品の操作の実験風景

きなかったことが挙げられる。また Cancel ジェスチャに関しては、手を振る移動量が少なかったことが挙げられる。なお、OK ジェスチャを行ったときに Cancel ジェスチャ(もしくはその逆)と認識されることはなかった。

7.7 手拍子の認識

5.5 で述べた手法に基づき、手拍子の認識実験を行った。被験者 5 人を対象に計測距離及び手拍子の拍数を変化させて、それぞれの場合における認識率を算出した。表 7 に手拍子の認識結果を示す。全体的に高い認識率が得られていることがわかる。またマイクと被験者の距離や手拍子拍数の増加に伴う著しい認識率の低下はみられない。

7.8 家電製品の操作実験

構築した実験システムを用いて家電製品の操作実験を行った。実験の様子を図 14 に示す。図 15 にテレビのチャンネルを操作しているときの PC の画面を示す。既に指の本数が“1”、“2”の順で認識が行われ、12 チャンネルと表示されており、判定ジェスチャの“OK”ジェスチャが認識されている。誤動作に関しては、指の本数の誤認識が主な原因であった。

8. 結 論

本研究では、知能を持つ部屋であるインテリジェントルームの一例として、Pan-Tilt-Zoom 機能を持つカメラを室内に複数台設置し、直感的なジェスチャの認識によって家電製品を操作するシステムを構築した。また、実験により各ジェスチャ認識の有効性を確認した。今後の展開として、インテリジェントルーム自体の能力を向上させるために、他のモダリティ特に音声認識の導入、ならびにインタフェースとしてあるいは人間の動作支援として移動ロボットの導入を行っていくことを考えて

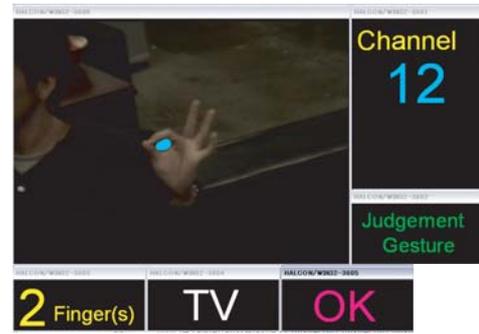


図 15 12 チャンネルの操作

いる。

文 献

- [1] V. I. Pavlovic, R. Sharma, T. S. Huang, Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review, Trans. PAMI, vol.19, no.7, pp.677-695, 1997.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] 西川敦, 大西映生, 西村正典, 平野敦士, 小荒健吾, 宮崎文夫, “局所相関演算に基づくオプティカルフローを用いた身振り動作の認識手法”, “情報処理学会論文誌”, Vol.40, no.8, pp.3118-3133, 1999.
- [6] A. Pentland, Looking at People: Sensing for Ubiquitous and Wearable Computing, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.22, no.1, pp.107-118, 2000.
- [7] Michael Coen, The Future Of Human-Computer Interaction or How I learned to stop worrying and love my Intelligent Room, IEEE Intelligent Systems, vol.14, no.2, pp.8-19, 1999.
- [8] Brumitt, B., Meyers, B., Krumm, J., Kern, A. and Shafer, S., EasyLiving: Technologies for Intelligent Environments, Proc. Int. Symposium on Handheld and Ubiquitous Computing, pp.12-27, 2000.
- [9] 佐藤知正, 森 武俊, 原田達也, “ロボティックルームの知能-ユービキタス知能-”, “日本ロボット学会誌”, vol.20, no.5, pp.482-486, 2002.
- [10] 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-DII, no.10, pp.2212-2223, 2001.
- [11] 入江耕太, 梅田和昇, “濃淡値の時系列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.
- [12] Fest Project 編集委員会, 新実践画像処理, リンクス, 神奈川, 2001.
- [13] 徐剛, 辻三郎, 3 次元ビジョン, 共立出版, 東京, 1998.
- [14] I. Yoda, K. Sakaue and Y. Yamamoto: “Arm-Pointing Gesture Interface Using Surrounded Stereo Cameras System”, Proc. International Conference on Pattern Recognition (ICPR 2004), vol.4, pp.965-970, 2004.
- [15] 酒井幸市, デジタル画像処理入門, pp.133-136, CQ 出版社, 東京, 2002.
- [16] 田村義和, Windows サウンドプログラミング, pp.143-149, 翔泳社, 東京, 2001.