

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*

入江 耕太^{*1}, 若村 直弘^{*2}, 梅田 和昇^{*3}

Construction of an Intelligent Room Based on Gesture Recognition

Kota IRIE, Naohiro WAKAMURA and Kazumori UMEDA^{*4}^{*4} Department of Precision Mechanics, Chuo University,
1-13-27 Kasuga, Bunkyo-ku, Tokyo, 112-8551 Japan

This paper proposes an intelligent room in which a person can operate home appliances naturally by gestures. Cameras with pan-tilt and zoom functions are set in the room, and the system is realized by using image information. Intention of operation and 3D position of the operator are recognized by detecting his/her hand waving, and the cameras are zoomed and focused on the operator's waved hand and the skin color is registered. The hand and finger regions are extracted by using the color information and gestures with the hand and fingers are recognized. This method is robust for the change of skin color because color information is not necessary at the stage of detecting hand waving and the skin color at the moment is used. A target appliance is chosen by recognition of pointing direction by the hand and operations of the appliances such as switching on/off, turning up/down of volume and changing channels are realized by recognizing the number of fingers and the hand's motion.

Key Words: Recognition, Human Interface, Man-Machine-Communication, Image Processing, Operation of Home Appliances, Intelligent Room

1. 序 論

現在, 種々の技術の発展に伴い, 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。我々に身近な機器である家電製品も, ネットワーク化など利便性の向上が図られている。一方で, これらの機器には, 多機能化することで操作が複雑化するという問題も生じる。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。特に家電製品のように老若男女を問わず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。そこで, 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。我々は日常的なコミュニケーションにおいて, ノンバーバルインタフェース, その中でも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, マン・マシン・インタフェースの一つとしても, ジェスチャは有効であると考えられ, 実際, 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行われている^{(2)~(5)}。しかし, ジェスチャ認識の研究の多くは, ジェスチャを認識可能な場所がカメラ正面などに限定される。背景差分の適用により動作者の場所を特定せずにジェスチャ認識を行う手法なども提案されている⁽⁵⁾が, 人物が複数存在するなど移動領域が複数の場合には適用が困難である。

操作する位置の拘束を受けず, かつ複数の人物が存在する環境で操作者にとって自然でインタラクティブなインタフェースを実現するためには, 部屋そのものを知能化することが有効であると考えられる。近年, 部屋などの空間を知能化するアプローチが注目され, いくつかの興味深い研究が行われている^{(6)~(16)}。特にMITでは, 数多くのプロジェクトが行われており, スマートルーム⁽⁶⁾, インテリジェントルーム⁽⁷⁾, インテリジェントスペース⁽⁸⁾が提案され, 人の行動追跡やMixed Reality技術によるインタラクションが実現

* 原稿受付 2006年6月14日。

^{*1} (株)日立製作所 (☎ 312-8503 ひたちなか市大字高場 2520)。^{*2} 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) [現: 東日本旅客鉄道(株)]。^{*3} 正員, 中央大学理工学部 (☎ 112-8551 東京都文京区春日 1-13-27)。

E-mail: umeda@mech.chuo-u.ac.jp

されている。また Microsoft, IBM においてもそれぞれ EasyLiving⁽⁹⁾, DreamSpace⁽¹⁰⁾ というシステムが提案されている。また、佐藤、森らが提案しているロボティックルームがよく知られている⁽¹¹⁾⁽¹²⁾。ロボットセンサ群が人をさりげなく見守っており、必要なときに分散配置されたアクチュエータ群が支援するというシステムである。Lee ら⁽¹³⁾は、移動ロボットを物理エージェントとして用いるインテリジェントスペースを提案している。また、ジェスチャ認識は実現されていないものの、現状の様々な IT 技術を組み合わせた情報家電モデルハウスの試作例もある⁽¹⁴⁾。さらに、人とのインタラクティブ性を実現するのではなく、室内での人の日常生活を観察するシステムも提案されている⁽¹⁵⁾⁽¹⁶⁾。

本研究でも、このように部屋そのものを知能化するアプローチをとる。具体的な対象として、家電製品を自然に操作するシステムの構築を取り上げる。これは、実用上、有益な対象と考えられ、また前記の関連研究で陽に構築したものはなく、意味があると考ええる。部屋などの空間を知能化する研究では、画像だけでなく音声なども用いることによってインタラクティブ性を高めているものが多いが、本研究では画像のみでシステムの実装を行う。これは、システムを軽くするためと、音声に関しては、マイクを用いない場合の認識率の不十分さによる。一方、腕などに装着できるコントローラを使って操作するという手段も考えられるが、コントローラを身につける煩わしさがあり、細かいボタン操作も必要となる。その点、ジェスチャをインタフェースとして利用することで、感覚的な操作が実現できると考えられる。

以上より、本研究の目的を、家電製品を操作位置の拘束を受けずに画像のみを使ってジェスチャで自然に操作できるインテリジェントルームを構築することとする。

まず 2 章で、本研究で構築するインテリジェントルームのシステムの概要を述べる。3~6 章で、手振り検出による操作者の特定、肌色登録、ジェスチャ認識、家電製品の操作の順にシステムを構築する。7 章で実験結果を示し、8 章で本論文をまとめる。

2. 本研究におけるインテリジェントルームの概要

本研究で対象とするインテリジェントルームは、室内にカメラを取り付け知能化した部屋であり、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。操作者が日常生

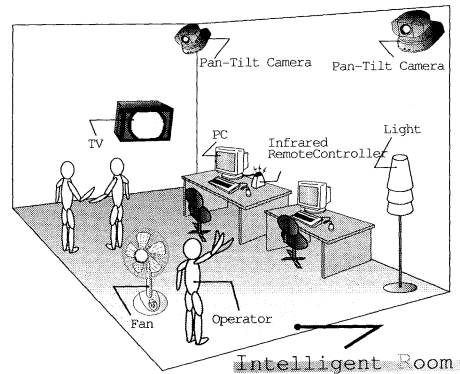


Fig. 1 The conceptual figure of our intelligent room

活を行っている状態で操作を行うことを想定し、センサグローブなどのウェアラブルデバイス⁽¹⁷⁾やマイクなどは装着しないとす。また、操作者は、操作する場所の拘束を受けず室内の自由な場所で操作可能とする。ジェスチャは手および指のみを用いて行うこととする。さらに、室内に複数の人間が存在する場合でも、インテリジェントルームが操作しようとする人を特定できるようにする。図 1 にインテリジェントルームの概念図を示す。

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その 3 次元位置情報を取得する。3 次元位置情報に従ってカメラをパン・チルト・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するために、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の「操作ジェスチャの認識」を行う。「操作ジェスチャ」は指の本数と手の往復運動とを組み合わせで実現する。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。

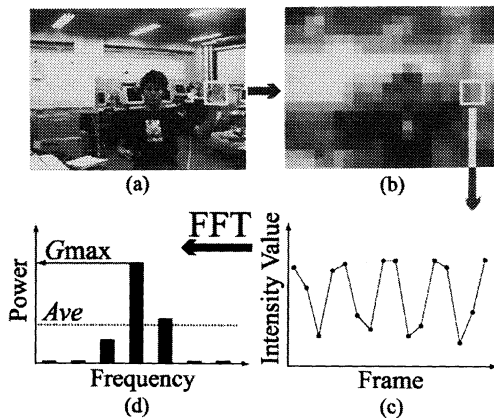


Fig. 2 Detection of hand waving using FFT

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⁽⁴⁾の手法を用いる。本手法の概要は次の通りである（図2参照）。まず画像(a)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(d)、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变化の影響を受けない。また、手領域を抽出などの画像処理の必要がなく、極めて簡潔な処理である。

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⁽¹⁸⁾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。さらに、ある点（本論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

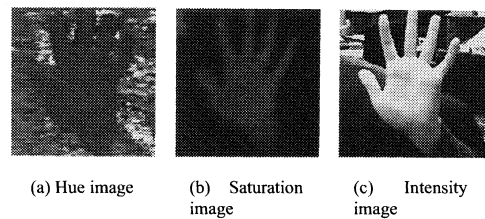


Fig. 3 HSI image

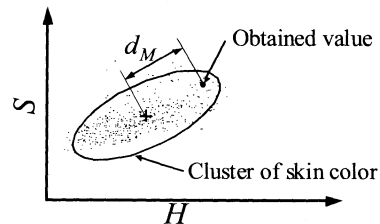


Fig. 4 Distribution of HS value in feature space

る⁽¹⁹⁾。

こうして得られた3次元位置情報に従ってカメラをパン・チルト・ズームさせる。このとき、パン・チルト・ズーム量に応じてカメラの内部パラメータ・外部パラメータを修正する。なお、現状のシステムではパン・チルト・ズームを行うのはこのタイミングのみとしている。

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变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、明るさの影響をうけにくい H （色相）、 S （彩度）、 I （明度）空間を用いる（図3参照）。 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 と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 $M = [M_H, M_S]^T$ とし、共分散行列を V とする。計測された特徴量ベクトル X とのマハラノビス距離 d_M は

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

で与えられる。図4に HS 特徴空間におけるマハラノビス距離の模式図を示す。

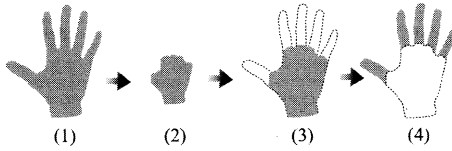


Fig. 5 Recognition of number of fingers

5. ジェスチャ認識

前章で登録された肌色情報を用いて手の領域の抽出を行い、指の本数の認識、指差し方向の認識、ならびに指の本数と手の動作特徴の認識による操作ジェスチャの認識を実現する。

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5参照）。

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 2台のカメラから取得した画像に対し、肌色情報を用いて手領域の抽出を行い、図6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 a 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、

$$a = [\sin \theta, \cos \theta, 0]^T \quad (2)$$

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 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 は、

$$N = [f \tan \theta, -f, b]^T \quad (3)$$

となる。 f は各カメラの焦点距離である。指差しは、図7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、Yamamoto ら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⁽²⁰⁾。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。

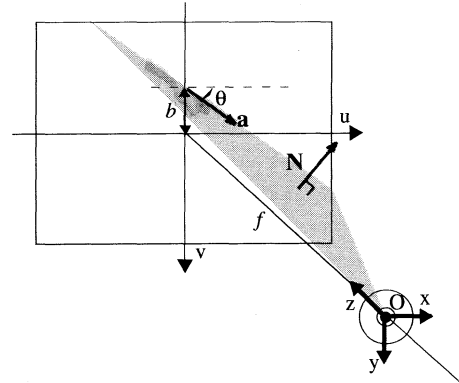


Fig. 6 Recognition of finger direction

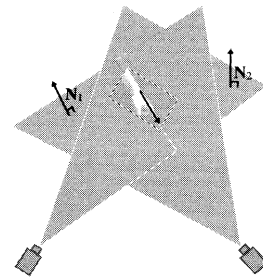


Fig. 7 The pointing direction in the world coordinate system

5.3 手の動作特徴の抽出 手領域の動作を表す特徴量として、重心移動ベクトルを利用する。重心移動ベクトル (du_i, dv_i) は、時系列で算出される重心座標 (u_i, v_i) の差をとることにより、

$$du_i = u_i - u_{i-1}, dv_i = v_i - v_{i-1} \quad (4)$$

により容易に求められる（図8参照）。

適当なフレーム間での du_i, dv_i の絶対値和を手の動作特徴とする。すなわち、 k をフレーム数として、

$$Mu_i = \sum_{j=i-k+1}^i |du_j|, Mv_i = \sum_{j=i-k+1}^i |dv_j| \quad (5)$$

の大きさにより、横方向あるいは縦方向の動作の大きさを評価できる。

5.4 操作ジェスチャの認識 機器を操作するためのジェスチャを「操作ジェスチャ」とし、指の本数と手の往復運動とを組み合わせることで実現する。認識結果が正しいかどうかを判定する OK・Cancel ジェスチャ、音量や風量などボリュームを調整する Up・Down ジェスチャを定める。OK ジェスチャは手を握って上下させる動作、Cancel ジェスチャは手を開いて左右に振る

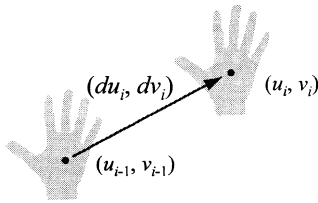


Fig. 8 Movement vector of center of gravity

Table 1 Correspondence of operating gestures to features

	OK	Cancel	Up	Down
Number of finger(s)	0	5	1	0
Motion feature	Mv_i	Mu_i	Mv_i	Mv_i

動作, Up ジェスチャは指を 1 本立てて手を上下させる動作, Down ジェスチャは手を握って上下させる動作 (OK ジェスチャと同じ) とする。

前節で述べた指の本数認識, 手の動作特徴の抽出を順に行い, これらの操作ジェスチャを認識する。表 1 に各操作ジェスチャと特徴量との対応を示す。

6. ジェスチャ認識による家電製品の操作

5 章で述べた手法を利用して各種の操作を実現する。

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, 「指差し方向認識」を利用して手振りの 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 に操作手法をまとめる。

Power ON 操作は, 前節の手法で操作対象が決定された時, すなわち操作対象が指差された時に自動的に行うとする。Power OFF 操作も同様で, 操作対象がオンの状態の時に指差された時に Power OFF 操作の指示であるとみなす。Power ON/OFF 操作はテレビ, 扇風機, 照明機器すべてに共通である。

Channel 操作はテレビに特有の操作であり, 「指の本数認識」によってチャンネルを入力する。次に, チャンネルの入力結果が意図通りであるかに応じて「OK・Cancel ジェスチャ」を行う。これにより, それぞれ入力結果の確定, 入力のやり直しを行うことができる。なお, 1 回目の指の本数 n_1 を 1 本と認識した場合に

Table 2 Operations of home appliances

Operation	Device	Gesture	Number of finger(s)
Power ON/OFF	TV Fan Light	Recognition of number of fingers	1
Channel	TV	Recognition of number of fingers n_i (i -th number)	Channel $N=1\sim 4$ $\rightarrow N=n_1$
			Channel $N=5\sim 9$ $\rightarrow N=n_1+n_2$ ($n_1=5, n_2=N-5$)
			Channel $N=10\sim 12$ ($n_1=1, n_2=N-10, n_1$)
Volume	TV Fan	Recognition of number of fingers and motion features	Up : 1 Down : 0

は, 続けて 2 回目の指の本数 n_2 の認識を行う。一定の時間内に n_2 が 0~2 本と認識された場合は, チャンネルを「10~12」と判断する。一方, 一定の時間内に n_2 が認識されない (例えば手を下げるなどで) 場合にはチャンネルは「1」と判断する。

Volume 操作はテレビと扇風機に対する操作で, 「Up・Down ジェスチャ」によりボリュームの調整を行う。

7. 実験

7.1 実験システム 画像処理には HALCON(MVTec Ver.6.1) を用い, 各種演算処理は PC(DELL, Pentium 4 2.2GHz) で行った。パン・チルト・ズーム機能を搭載した 2 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) を用いて PC に入力した。本 CCD カメラは 10 倍光学ズームにより, 視野 65° から 6.6° までを実現している。また, キャプチャされる画像の画素数は $640 \times 482[\text{pixel}]$ である。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。

部屋は $6.9 \times 7.8[\text{m}^2]$ の長方形で, カメラは $6.9[\text{m}]$ の辺の両隅近くの高さ $3.2[\text{m}]$ の位置に設置した。

7.2 手振りの認識 3 章で述べた手法に基づき, 手振りの検出を行った。このとき, ズームは最も広角側に設定し, 部屋の任意の場所での手振りの検出を可能としている。計測距離を変化させた時の認識率に関しては文献⁽⁴⁾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 $0.5[\text{m}]$ ずつ位置をずらして計測した。各点

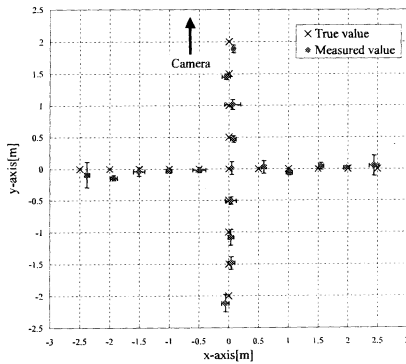


Fig. 9 Measurement of position of hand waving

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9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ばらつきが大き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140[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。これにより、以後の処理は距離の違いの影響を受けずに行うことができる。

7.3 指の本数認識 あらかじめ4章で述べた肌色登録を行い、5章で述べた手法に基づいて指の本数認識を行った。画面面と掌との角度を $0^\circ, 10^\circ, 20^\circ, 30^\circ$ と変化させ、それぞれの角度における指の本数0~5本の認識を20回行った。認識率ならびに検出結果の例をそれぞれ表3と図10に示す。画面面と掌との角度が 0° （カメラに対して正面）の場合、全ての指の本数において良好な認識率が得られている。認識を失敗したのは、掌が下方または上方へ大きく傾いている場合などであった。また、指の本数が3本の場合の誤認識原因として、3本の指（人差し指、中指、薬指）の角度がそれぞれ異なる（特に薬指が前傾する）ことにより、照明条件が不均一になり肌色抽出がうまく行えなかったことが挙げられる。さらに、指の本数が4本の場合の誤認識原因としては、指の開きが不十分で、指が重なってしまった場合が挙げられる。一方、画面面と掌との角度が 20° 以上かつ指の本数が3本以上の場合、指の領域が重なってしまい、正確な認識が行えなかった。しかしながら、人間の心理として、何らかの目標物に対して指示を与えるといった行為は比較的自然であると考えられる。よって、いずれかのカメラ

Table 3 Recognition rate of number of finger(s) [%]

Number of finger(s)		0	1	2	3	4	5
Angle	0°	100	100	100	80	85	100
	10°	85	80	75	60	65	85
	20°	100	100	50	20	0	0
	30°	100	90	40	0	0	0

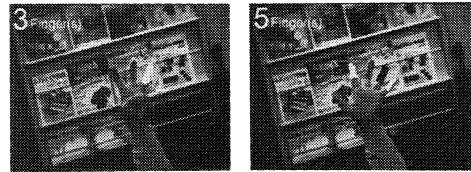


Fig. 10 Examples of recognizing number of fingers

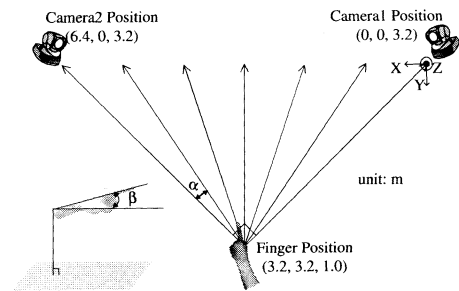


Fig. 11 Position of the subject for pointing direction

に向かってジェスチャを行うことを前提としても、システムの実現という観点では、それほど問題は無いと考えられる。

7.4 指差し方向認識 5章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11に示すように真上から見て2台のカメラ間の角度が 90° になる位置に立ち、カメラ1とのなす角 α が $0^\circ, 15^\circ, \dots, 90^\circ$ （ 15° おき）で仰角 β が 0° （床に水平）の方向を指差し、その場合の認識結果を調べた。図12に2台のカメラの各々の画像での指差し方向の認識例、図13、表4に指差し方向の認識結果を示す。いずれかのカメラとのなす角が小さい場合に位置精度が悪くなっていることがわかる。原因として、手領域抽出の際にカメラの光軸に近い方向に向いて指差している状態になり、領域が上手く抽出できないことが挙げられる。計測誤差はやや大きい、室内の近い位置に複数の操作対象がなければ、ほぼ間違えることなく正しい対象の指示が可能であると言える。

7.5 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2台のカメラで手振りを検出している様子を示す。画像中で四角で囲まれた箇所が

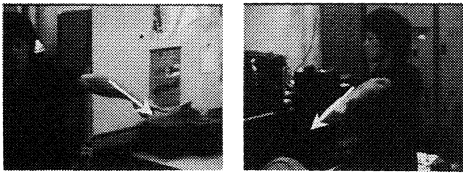


Fig. 12 Example of the recognition of pointing in each image

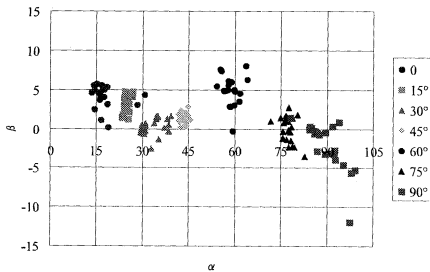


Fig. 13 Experimental results of pointing direction

Table 4 Average and standard deviation of pointing direction

α [deg]	0	15	30	45	60	75	90
Average[deg]	17.3	25.1	33.3	43.4	59.2	77.4	90.0
SD[deg]	5.0	1.1	3.5	1.0	2.7	3.0	10.6
β [deg]	0						
Average[deg]	5.9	3.8	0.2	1.4	5.0	-0.2	-3.0
SD[deg]	2.1	1.3	1.1	0.7	1.9	1.9	4.3

手振りを検出した箇所である。画像中で操作者がかなり小さくても手振りの検出が可能であることが示されている。

図 16 にテレビのチャンネルを操作している時の画面を示す。指の本数が 2 本であることが認識され、既に認識された 1 本と合わせて 12 チャンネルと入力されている。図 15 と比較すると、ズームにより手が拡大されていることが示されている。また、チャンネル操作ではカメラは一台しか必要でないため、二台目のカメラに対応する画面に画像が表示されていない。また、図 17 に操作ジェスチャの認識例として、Down と Cancel の例を示す。

以上に示すように、対象とする家電製品を操作するシステムを実現できている。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。

また、テレビの操作を例に、各操作にかかる時間を表 5 に示す。チャンネル操作では、目的のチャンネル

Table 5 Operating time

Operation	Operating time[s]
Detection of hand waving	6.1
Registration of skin color	8.3
Recognition of pointing direction	12.3
Operation of TV: Power ON	5.5
Operation of TV: Channel "12"	15.4

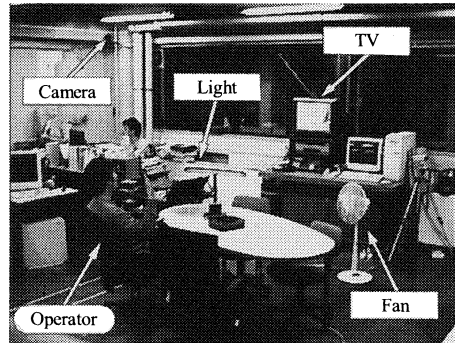
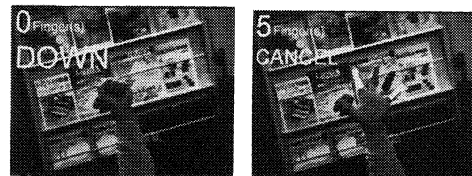


Fig. 14 Overview of experiments for operating home appliances



(a) Operation of "Down"

(b) Operation of "Cancel"

Fig. 17 Examples of recognition of operation gestures

を"12"としており、指の本数認識を 2 回 (10 の位と 1 の位) 行った時間である。肌色登録と指差し方向認識においては、カメラのパン、チルト、ズーム動作を含んだ時間である。本実験では、一連の操作にかかる合計時間は約 48[s] であり、実用化を想定した場合には、各認識処理の高速化やカメラ動作の高速化等の改善が必要であると考えている。

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、パン・チルト・ズーム機能を持つカメラを室内に複数台設置しジェスチャ認識によって家電製品を操作するシステムを構築した。手振り認識によって操作者の検出および 3 次元位置の計測を行ってズームアップし、指差し認識によって指示対象物を選択し、指の本数認識や手の縦ないし横の往復運

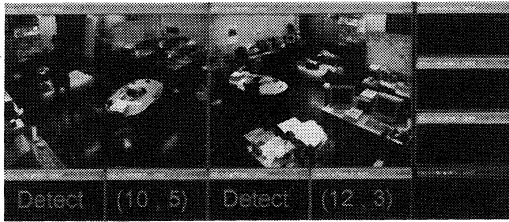


Fig. 15 Detection of hand waving



Fig. 16 Operation of TV channel

動の認識によってスイッチのオン・オフやボリュームのアップ・ダウン、テレビのチャンネルの選択などの機能を実現した。操作者は室内の自由な位置で機器を操作することが可能で、また人にわかりやすいジェスチャのみを用いており、人に優しいインタフェースを持つインテリジェントルームの一例を構築することが出来たと考える。また、システムとして見た時、以下の点に特に意義があると考ええる。

- 手振り検出手法の利用により操作者を安定に特定
- パン・チルト・ズームカメラを利用し、カメラの画角・撮像範囲を変化させることにより、任意の位置における操作者発見と指などの細かなジェスチャ認識の実現を両立
- 手振り検出後の肌色登録により、異なる照明環境や肌色の個人差に対するロバスト性を向上
- 認識処理の単純化により、安価な装置で実現することが可能であり、実用性の向上につながる

本研究では、2台のカメラだけでできる機能を実現しており、カメラだけでここまで出来ることを示した。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は、各認識処理の高速化や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の移動ロボットの導入を行っていくことを考えている。

文 献

- (1) Kurokawa, T., *Non-Verbal Interface* (in Japanese), Ohomsha, (1994).
- (2) Pavlovic, V.I., et al., *Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review*, Trans. PAMI, vol.19, no.7(1997), pp.677-695.
- (3) Nishikawa, A., et al., *Recognition of Human Gestures from Optical Flow Based on a Correlation Method Between Local Image Regions* (in Japanese), IPSJ Journal, vol.40, No.8(1999), pp.3118-3133.
- (4) Irie, K., Umeda, K., *Detection of Waving Hands from Images Using Time Series of Intensity Values* (in Japanese), Journal of the Robotics Society of Japan, vol.21, No.8(2003), pp.923-931.
- (5) Nishimura, T., et al., *Recognition of Human Motion Behaviors Using View-Based Aspect Model Based on Motion Change* (in Japanese), The Transactions of the Institute of Electronics, Information and Communication Engineers A, vol.J84-DII(2001), pp.2212-2223.
- (6) Pentland, A., *Looking at People: Sensing for Ubiquitous and Wearable Computing*, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.22, no.1(2000), pp.107-118.
- (7) Michael, C., *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(1999), pp.8-19.
- (8) <http://oxygen.lcs.mit.edu/E21.html>
- (9) Brumitt, B., et al., *EasyLiving: Technologies for Intelligent Environments*, Proc. Int. Symposium on Handheld and Ubiquitous Computing, (2000), pp.12-27.
- (10) <http://www.research.ibm.com/natural/dreamspace/>
- (11) Mori, T., Sato, T., *Robotic Room: Its concept and Realization* (in Japanese), Robotics and Autonomous Systems, Vol.28, No.2(1999), pp.141-144.
- (12) Sato, T., et al., *Intelligence of Robotic Room - Ubiquitous Intelligence -*, Journal of the Robotics Society of Japan, vol.20, no.5(2002), pp.482-486.
- (13) Lee, J.H., Hashimoto, H., *Intelligent Space - concept and contents -*, Advanced Robotics, vol.16, no.4(2002), pp.265-280.
- (14) <http://www.jeita.or.jp/>
- (15) Nisida, Y., et al., *Robustly Observing Human Activity by Sensor Room and Its Application* (in Japanese), IPSJ SIG Technical Reports, no.106-006(2003), pp.37-44, 2003.
- (16) Mori, T., et al., *Sensing Room: Distributed Sensor Environment for Measurement of Human Daily Behavior*, Proc. First International Workshop on Networked Sensing Systems (INSS2004), (2004), pp.40-43.
- (17) Tsukada, K., Yasumura, A., *Ubi-Finger: Gesture Input Device for Mobile Use* (in Japanese), IPSJ Transaction, vol.43, No.12(2002), pp.3675-3684.
- (18) FestProject committee, *New Practical Image Processing* (in Japanese), LinX Corporation, (2001).
- (19) Jo, G., Tsuji, S., *Three-Dimensional Vision* (in Japanese), Kyoritsu Shuppan, (1998).
- (20) Yamamoto, Y., et al., *Arm-Pointing Gesture Interface Using Surrounded Stereo Cameras System*, Proc. International Conference on Pattern Recognition (ICPR 2004), vol.4(2004), pp.965-970.