直感的なジェスチャの認識を利用したインテリジェントルームの構築

Construction of an Intelligent Room Using Intuitive Gesture Recognition

学　若村　直弘（中央大学）　○学　鈴木　健一郎（中央大学）
正　梅田　和昇（中央大学）

Naohiro WAKAMURA, Chuo University
Kenichiro SUZUKI, Chuo University
Kazunori UMEDA, Chuo University

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 Appliances

1.　序論

現在，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．一例として，身近な家電製品のネットワーク化が実現している．一方，多機能化することで，操作が複雑化するといった問題も生じる．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，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．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．我々は日常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．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)(2)]．また，それらのジェスチャ認識技術を用いて部屋全体を知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[(3)-(5)]．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場所を特定せずにジェスチャ認識を行う手法[(6)]なども提案されているが，人物が複数存在するなどで移動領域が複数の場合，適用が困難である．一方，入江ら[(7)]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バストに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，部屋全体を知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，3 台の CCD カメラの画像から「手振りの検出」を行い，操作者を発見するとともに 3 次元位置情報を取得する．得られた 3 次元位置情報から算出されるカメラからの距離に応じてカメラをズームさせ，ジェスチャ認識処理を行うための候補領域を絞る．続いて，手振り位置の色情報を用いて各ジェスチャの認識を行う．個人差や環境による肌色の変化に対するロバスト性の向上を図るために，前処理として「肌色の登録」を行う．これにより，安定した手領域抽出が可能になると考えられる．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 のモニタによる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．操作対象となる家電製品は，赤外線リモコンで操作可能な機器とし，各機器への制御信号は，PC に接続された赤外線リモコンを用いて送信する．インテリジェントルームの概念図を図 1 に示す．

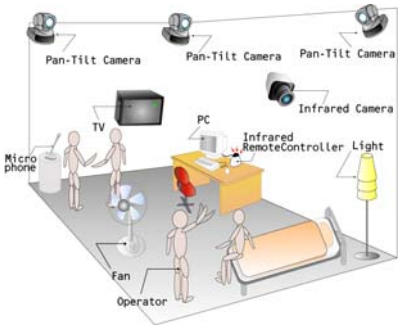


Fig.1　conceptual figure of our intelligent room

3.　操作者特定のための手振り検出[(7)]

インテリジェントルームにおいて，操作者を特定する手法として，手振り動作の検出を行う．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，その中で手振り領域の検出を行う．本手法は，カラー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．さらに，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，極めて簡潔な処理である．2 つの CCD カメラを用いて手振りを検出することで，手振り位置の 3 次元位置を計測する．

4. 肌色登録[5]

個人差や環境による肌色の変化に対するロバスト性の向上を図るために，手振り検出処理で得られた手領域に対して肌色の登録を行う．肌色登録における色空間は，H（色相），S（彩度），I（明度）空間を用いる．輝度成分を含まない H，S 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．

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 2 参照）．

(1) 肌色情報に基づき，手の領域を抽出する．
(2) 手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(2)の領域の差をとり，領域を分離し，残った領域から形状や面積の特徴から指領域のみを抽出する．

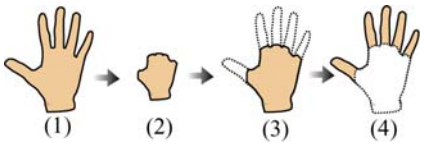


Fig.2 Recognition of number of fingers

5.2 指差し方向認識[8]

2 台のカメラから取得した画像に対し，肌色情報に基づき手領域の抽出を行い，抽出された手領域の主軸(長軸)ベクトルを求める．それぞれの主軸ベクトルを 3 次元空間に投影し，投影された 2 平面の交線を指差し方向とする．

5.3 手の動作認識

手の動作認識には DP マッチング[9]を用いる．肌色情報に基づき，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 3 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，

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

$$dy_i = y_{i+1} - y_i \quad (2)$$

$$dw_i = w_{i+1} - w_i \quad (3)$$

$$dh_i = h_{i+1} - h_i \quad (4)$$

$$da_i = a_{i+1} - a_i \quad (5)$$

で算出される．

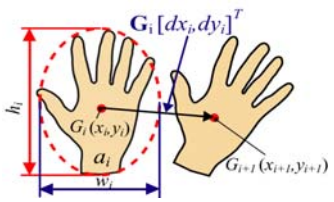
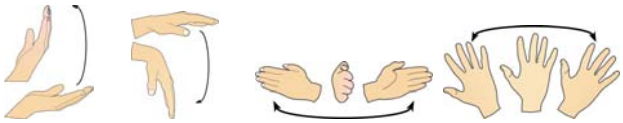


Fig.3 Features of a hand region for gesture recognition

式(1)-(5)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いても良い．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を図 4 に示す．



(a)“Up” (b)“Down” (c)“Rotation (d)“Finish”

Fig.4 Gesture models

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 5 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．

Table 1 Features of judgment gestures

Gesture	Features
OK-1	・Region surrounded by skin color region is circular. (In case extracted continuously) ・Number of standing fingers are 3.
OK-2	・Number of standing fingers is 1.
Cancel	・Extracted skin color region vibrates horizontally. (In case more than a threshold)

(a)“OK-1” (b)“OK-2” (c)“Cancel”

Fig.5 Judgment gestures

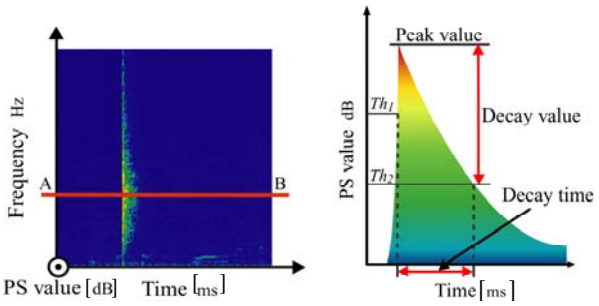
5.5 手拍子の認識

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

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[10]で求められるソノグラフ(図 6(a)参照)を用いる．ソノグラフ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．図 6(b)は図 6(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) 高周波領域におけるピーク値



(a)Sonograph (b) Time-PS plane (A-B)

Fig.6 Basic concept of sonograph

6. ジェスチャ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 節で述べた「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”の順番で指の本数認識を 3 回行う．

Table 2 Operations of home appliances

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

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

3 章で述べた手法に基づき，手振りの検出実験を行った．被験者 5 人を対象に，計測距離を変化させて認識率を算出した．(1)画像中の任意の場所で手振りを約 2[s]間行う．(2)手振りを止めてから約 2[s]おいて次の手振りを再開する．以上の動作を 20 回試行し，そのうち約 2[s]以内で手振りを検出した場合を成功とした．距離による認識率を表 3 に示す．広い計測範囲で高い認識率が得られた．認識失敗例として，手振りの幅が小さい場合が挙げられる．

Table 3 Recognition rate of hand waving

4m	5m	6m	7m	8m
96%	96%	97%	92%	83%

7.3 指の本数認識

5.1 節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 章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，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 4 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．認識の失敗例として，掌が下方または上方へ大きく傾いている場合が挙げられる．

Table 4 Recognition rate of number of fingers

0	1	2	3	4	5
90%	98%	96%	82%	88%	88%

7.4 指差し方向認識

5.2 節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 7 に示すように真上から見て 2 台のカメラ間が 90°になる位置に立ち，カメラ 1 となす角α が 0°，15°，…，180°(15°ずつ)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 8，表 5 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．

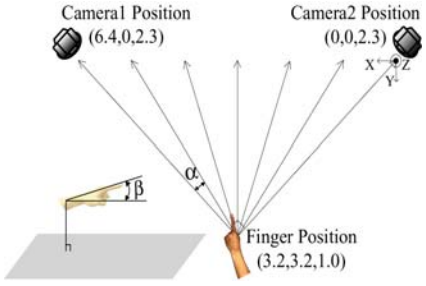


Fig.7 Position of the subject for pointing

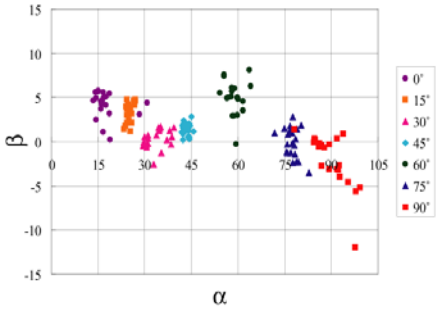


Fig.8 Experimental results of pointing direction

Table 5 Average and standard deviation of pointing direction

α[deg]	0	15	30	45	60	75	90
Ave.[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						
Ave.[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

7.5 手の動作認識

5.3 節で述べた手法に基づき，被験者 5 人に対して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 章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 6 に示す．誤認識の例として，手を速く動かしすぎて手領域を抽出できなかった場合などが挙げられる．

Table 6 Recognition rate of hand motion

Down	Up	Rotation	Finish
96%	99%	86%	89%

7.6 判定ジェスチャの認識

5.4 節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 章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 7 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．

Table 7 Recognition rate of judgment gestures

OK-1	OK-2	Cancel
85%	85%	97%

7.7 手拍子の認識

5.5 節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回数を変化させて,それぞれの場合における認識率を算出した．表 8 に手拍子の認識結果を示す．距離が遠くなるにつれて，また回数が多くなるにつれて認識率が低くなっていることがわかる．認識の失敗例として，距離が遠くなることで手拍子の音が小さくなったことや手拍子の音を上手く出すことができなかったことなどが挙げられる．

Table 8 Recognition rate of handclap

拍数	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%

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 9 に示す．図 10 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だった．

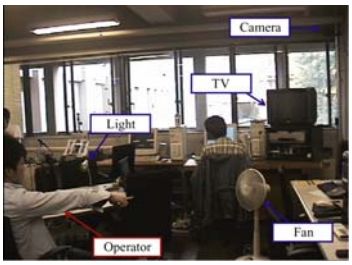


Fig.9 Situation of an experiment

Fig.10 Operation of “12ch”

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いる．

文 献

(1) P. Hong, M.Turk, T. S. Huang, Gesture Modeling and Recognition Using Finite State Machines, IEEE Int. Conf. on Automatic Face and Gesture Recognition (2000) pp.691-694.

(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 (1998) pp.83-85.

(3) Taketoshi MORI and Tomomasa SATO, Robotic Room: Its concept and Realization, Robotics and Autonomous Systems, Vol.28 No.2 (1999) pp.141-144.

(4) I. Yoda, K. Sakaue, and Y. Yamamoto, ArmPointing Gesture Interface Using Surrounded Stereo Cameras System, Proc. International Conference on Pattern Recognition, Vol.4 (2004) pp.965-970.

(5) 入江耕太，若村直弘，梅田和昇，ジェスチャ認識を用いたインテリジェントルームの構築 - 手のジェスチャによる家電製品の操作 - ，第 21 回日本ロボット学会学術講演会予稿集 (2003) 2J15.

(6) 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 (2001) pp.2212-2223.

(7) 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 (2003)pp.923-931.

(8) Kota Irie, Naohiro Wakamura, Kazunori Umeda, Construction of an Intelligent Room Based on Gesture Recognition -Operation of Electric Appliances with Hand Gestures-, Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (2004) pp.193-198.

(9) 酒井幸市, デジタル画像処理入門, CQ 出版社, 東京(2002) pp.133-136.

(10) 田辺義和，Windows サウンドプログラミング，翔泳社，東京 (2001) pp.143-149.