

# インテリジェントルームにおける家電機器操作の高度化

## Improvement of Operation of Home Appliances in an Intelligent Room

学 鈴木健一郎 (中央大学) ○学 和田正樹 (中央大学) 正 梅田和昇 (中央大学)

Kenichiro Suzuki, Chuo University, ksuzuki@sensor.mech.chuo-u.ac.jp

Masaki Wada, Chuo University, wada@sensor.mech.chuo-u.ac.jp

Kazunori Umeda, Chuo University, umeda@mech.chuo-u.ac.jp

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. 緒言

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。その一例として、身近な家電機器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、機器の多機能化や複雑化が進み、ユーザーの直感的な操作が困難であるといった問題が生じている。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、操作位置が限定され、リモコン操作では、リモコンの位置が常に特定されないといった欠点がある。家電機器の様に日常生活で頻繁に使用する場合、ユーザーが直感的に操作でき、かつ操作位置の拘束がないことが望ましい。このことから、ボタンやリモコンの代わりに人間の自然な行動をインタフェースとして用いることが考えられる。我々は日常的に自然と身振り手振りで相手に意思を伝えている。そこで、ヒューマン・マシン・インタフェースの一つとして、ジェスチャが考えられる。これまでに、動画像からジェスチャを認識する研究<sup>[1][2]</sup>やそれらを用いたインテリジェントルームの研究<sup>[3][4]</sup>も数多く報告されている。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<sup>[5]</sup>などもあるが場所が限定されている場合や、人物が複数存在する場合に適用が困難であるといった問題点が挙げられる。一方、入江ら<sup>[6]</sup>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ストに動作者及び動作者の位置の特定を行っている。若村ら<sup>[7]</sup>は、入江らの手法を用いて動作者の位置を特定し、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家電機器操作システムを実現している。利便性や操作性を考慮し、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っている。

本論文では、手領域抽出に用いる色空間を新たに提案するなどにより、若村らの構築したインテリジェントルームを高度化する。

### 2. インテリジェントルームの概要<sup>[7]</sup>

本論文で扱うインテリジェントルームとは、室内に複数台の CCD カメラを取り付け、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知能ロボット化した部屋であり、一般的な家庭のような生活空間への適用を想定している。インテリジェントルームの概念図を図 1 に示す。

まず、2 台以上のカメラの画像から「手振りの検出」を行い、操作者を発見するとともに三次元位置情報を取得する。得られた三次元位置情報から算出されるカメラからの方向、距離に応じてカメラの Pan-Tilt-Zoom 制御を行い、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、画像内の手領域を検出し、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、「肌色情報の登録」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。手振りの三次元位置情報と肌色情報を用いて各ジェスチャの認識を行う。操作する機器の決定には「指差し方向の認識」を用いる。機器の具体的な操作には「指本数の認識」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し、操作者に伝える。操作対象となる家電機器は、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。

さらに、操作者への直接的な動作支援や、ジェスチャ認識の補助的役割を担うために、移動ロボットの導入を行う。

以下、若村ら<sup>[7]</sup>の手法からの改良点を主に詳細を記述する。

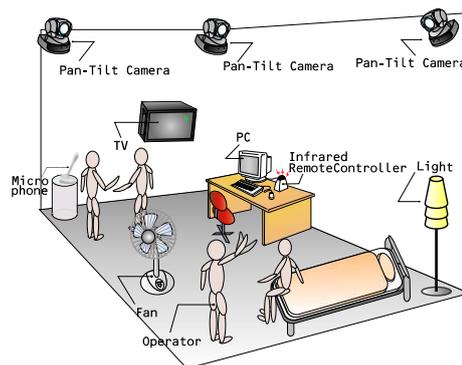


Fig.1 The conceptual figure of our intelligent room

### 3. 手領域の検出

肌色登録を行うために、Pan-Tilt-Zoom 制御した画像内の手領域を検出する。検出にはテンプレートマッチングを用いる。テンプレート画像には、肌色情報の多く含まれている R 画像を用い、入力画像の R 画像を対象にマッチングを行う。最もマッチングした画素の画素値から検出された領域が手領域であるかの判定を行う。テンプレート画像は両手の画像を予め

登録しておく．図2にテンプレート画像を示す．



Fig.2 Template image of both hands

#### 4. 肌色登録

個人差や環境による肌色の変化に対するロバスト性向上のために，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．肌色登録に用いる色空間は，H (色相)，S (彩度)，I (明度) 空間を輝度で補正した空間を用いる．式(1)-(4)に RGB 空間から HSI 空間への変換式を示す．式(2)より H に輝度成分は含まれていないが，式(3)より S には輝度成分が含まれてしまうため，式(5)のように S を I で除して補正する． $k$  はスケールリングのための定数で，本論文では 64 とした．補正された S 値を  $Sc$  とする．H,  $Sc$  を特徴空間とし，肌色のクラスタを形成する．この特徴空間を用いれば，輝度の変動の影響を受けず，安定に色の認識が可能である．認識時には，各画素で得られた H と  $Sc$  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．図3に補正 HSI 空間の画像を示す．

$$\begin{bmatrix} M_1 \\ M_2 \\ I_1 \end{bmatrix} = \begin{bmatrix} 2/\sqrt{6} & -1/\sqrt{6} & -1/\sqrt{6} \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

$$H = \tan^{-1}(M_2/M_1) \quad (2)$$

$$S = \sqrt{(M_1)^2 + (M_2)^2} \quad (3)$$

$$I = I_1/\sqrt{3} \quad (4)$$

$$Sc = k \cdot S/I \quad (5)$$



(a) Hue image(H) (b) Saturation image(Sc) (c) Intensity image(I)

Fig.3 Corrected HSI space

#### 5. ジェスチャ認識

##### 5.1 手の動作認識

手の動作認識には DP マッチング<sup>[8]</sup>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4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，式(6)-(10)で算出される．

$$dx_i = x_{i+1} - x_i \quad (6)$$

$$dy_i = y_{i+1} - y_i \quad (7)$$

$$dw_i = w_{i+1} - w_i \quad (8)$$

$$dh_i = h_{i+1} - h_i \quad (9)$$

$$da_i = a_{i+1} - a_i \quad (10)$$

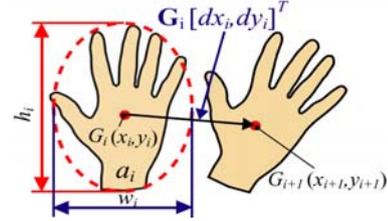


Fig.4 Features of a hand region for gesture recognition

式(6)-(10)で算出された各特徴量を用いて 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  $\mathbf{X}$ ，ジェスチャモデルのパターンを  $\mathbf{Y}$  とすると，類似度は  $\mathbf{X}$  と  $\mathbf{Y}$  の距離から算出できる．DP マッチングを用いるため， $\mathbf{X}$  と  $\mathbf{Y}$  の成分数は異なっても良い．入力パターンと全てのモデルとの類似度を求め，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．本論文で用いるジェスチャモデルを図5に示す．全てのジェスチャが同じ手の位置から開始できるため，操作性が向上している．

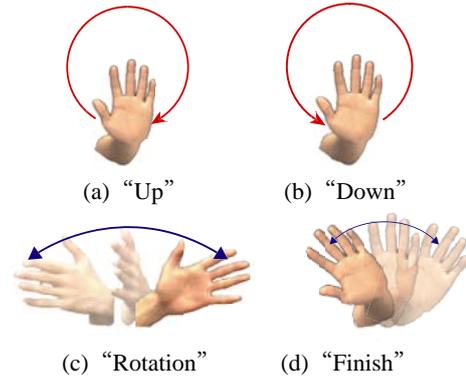


Fig.5 Gesture models

##### 5.2 判定ジェスチャ

認識結果に対して正誤判定を行う．これまででは，“正”，“誤”それぞれにジェスチャを割り当てていたが，操作が煩雑であった．そこで，一定時間内に Cancel ジェスチャが認識されなければ“正”，Cancel ジェスチャが認識されれば“誤”と判定することとする．Cancel ジェスチャは図5(d)と同様に手を左右に振る動作とする．認識には手領域重心の水平方向移動量のみに着目し，時間内に移動量の総和が一定値を超えれば Cancel と認識する．

#### 6. ジェスチャによる家電機器の操作

##### 6.1 操作対象の決定<sup>[7]</sup>

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」より三次元空間内の直線を求め，この直線と操作候補の機器との距離が最も小さいものを操作対象と決定する．本論文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1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、「指の本数認識」により“1”，“5”，“2”の順番で指の本数を認識する。指の本数よりテレビのチャンネル  $N$  を確定した後，5.2 節で示した「判定ジェスチャ」で認識結果の正誤判定を行う。さらに，音量の操作や，チャンネルの再操作は，5.1 節で示した「手の動作認識」を用いる。

Table 1 Operations of home appliances

Recognition of number of fingers $n_i$ ( $i$ -th number)			
Operation	Device	Channel $N$	Gestures
Channel	TV	$N=1-4$	$N=n_1+n_2$ ( $n_1=0, n_2=1-4$ )
		$N=5-9$	$N=n_1+n_2$ ( $n_1=5, n_2=0-4$ ) ( $n_2=5$ : Retry)
		$N=10-14$ $N=20-24$ $N=30-34$ $N=40-44$	$N=10n_1+n_2$ ( $n_1=1-4, n_2=1-4$ )
		$N=15-19$ $N=25-29$ $N=35-39$ $N=45-49$	$N=10n_1+n_2+n_3$ ( $n_1=1-4, n_2=5, n_3=0-4$ ) ( $n_3=5$ : Retry)
Recognition of hand motion			
Operation	Device	Gestures	
Volume	TV, Fan	Up, Down	
Rotation	Fan	Rotation	
Power OFF	TV, Fan, Light	Finish	

## 7. 移動ロボットの導入

### 7.1 インテリジェントルームにおける移動ロボット

現在，本論文におけるインテリジェントルームは家電機器を操作することを対象としている。このとき，直接操作者の動作支援を行うことが可能で，またインテリジェントルームに対する新たなインタフェースとなり得る移動ロボットの導入が有効である。現在，手振りを行った操作者のところに移動ロボットを移動させ，ジェスチャ認識を行うシステムの構築を行っている。

### 7.2 移動ロボットの位置の算出

インテリジェントルーム内での移動ロボットの位置計測手法を示す。

室内の移動ロボットを検出するために，移動ロボットにマーカーとして LED を配置する。図 6 に LED を搭載した移動ロボットを示す。



Fig.6 Mobile robot with LED marker

LED の床面からの高さ  $h$  が既知であるとする。このとき，移動ロボット (LED) の位置  $(x_0, y_0)$  は，以下の式を用いて算出される。

$$\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \frac{h-t_{34}}{t_{31}\omega_u u + t_{32}\omega_v v + t_{33}f} \begin{pmatrix} t_{11}\omega_u u + t_{12}\omega_v v + t_{13}f \\ t_{21}\omega_u u + t_{22}\omega_v v + t_{23}f \end{pmatrix} + \begin{pmatrix} t_{14} \\ t_{24} \end{pmatrix} \quad (11)$$

ここで， $f$  はレンズの焦点距離， $\omega_u, \omega_v$  は 1 画素の縦，横の長さ， $u, v$  は画像座標系での LED の位置， $t_{ij}$  はカメラの位置・姿勢を表す同次変換行列  ${}^0T_n$  の  $(i,j)$  成分をそれぞれ表す。

## 8. 実験

### 8.1 実験システム

画像処理には HALCON(MVTec)を用い，各種演算処理は PC(DELTA, Pentium 4, 3.2GHz)で行った。Pan-Tilt-Zoom 機能を搭載した 4 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)で合成された画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，家電機器を操作するために，PC で制御可能な学習型赤外線リモコン(SUGIYAMA ELECTRON クロスサム 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。さらに，移動ロボットは ER-1(EvolutionRobotics)を用い，移動ロボットに搭載する LED は赤色砲弾型(5φ 19000mcd)に拡散キャップを取り付けて使用した。

### 8.2 カメラパラメータの設定

本論文では色情報を用いてジェスチャ認識を行うため，環境が変化するたびにホワイトバランスや露光などの設定が変化することは望ましくない。そこで，毎回カメラパラメータを表 2 のように初期化する。図 7 にパラメータ設定前と設定後の画像を示す。図 7(a)では画像全体が明るすぎるため，色情報が飽和している部分が存在するのに対し，図 7(b)では全体的に明るさが抑えられ，肌色抽出ならびにジェスチャ認識に適した画像となっている。

Table 2 Initial parameter of camera

Pan-Tilt	Home
Auto focus sensitivity	High
Digital zoom	Off
Backlight correction	Off
Exposure correction	Off
Shutter speed	1/100 [sec]
Exposure iris value	F2.8
Exposure gain value	0 [dB]
White balance	5800 [K]



(a) Before parameter setting (b) After parameter setting

Fig.7 Difference by parameter setting

### 8.3 テンプレートマッチング

3章で述べたテンプレートマッチングを用いた手領域の検出実験を行った。被験者10人に対し、左右それぞれの手について10回ずつ実験を行い、認識率を算出した。テンプレート画像は特定の被験者のものを使用した。1フレーム(約100ms)ごとに両手のマッチングを行い、2フレーム以内で手領域を検出した場合を成功とした。カメラ・被験者間の距離はテンプレート画像登録時と同じで、4.5mとした。認識率を表3に示す。左右それぞれで高い認識率を示した。誤認識の例として、カメラに向かって手が正対していない場合が挙げられる。

Table 3 Recognition rate of template matching

Right Hand	Left Hand
92%	90%

### 8.4 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で4.5m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、各10回ずつ実験を行い、認識率を算出した。各ジェスチャの認識率を表4に示す。誤認識の例として、手を速く動かす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 4 Recognition rate of recognition of hand motion

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

### 8.5 移動ロボットの位置算出

7章で述べた手法に基づき、室内の移動ロボットの位置計測実験を行った。カメラの位置は(0.23,0.43,2.32)[m]である。計測は室内の10点で各10回行った。図8に結果を示す。おおよその位置は求まっているが、移動ロボットの位置が画像中心から離れるほど誤差が大きくなっている。原因としてカメラの外部パラメータのキャリブレーションが不正確であることが考えられる。

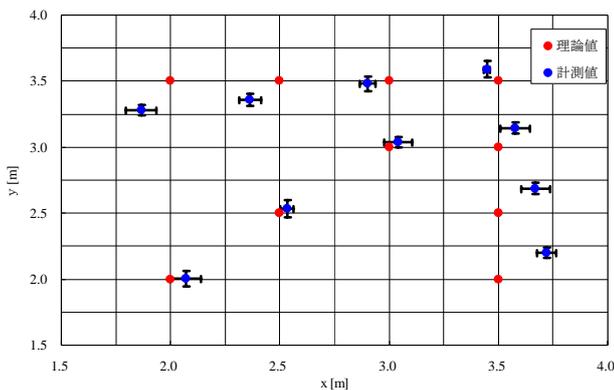


Fig.8 Experimental results of detecting mobile robot positions

### 8.6 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。図9に実験の様子を、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。指の本数が“5”、“3”の順で認識が行われ、8チャンネルと表示されている。

誤動作に関しては、指差し方向認識の時に意図した機器を選択できない場合が多かった。

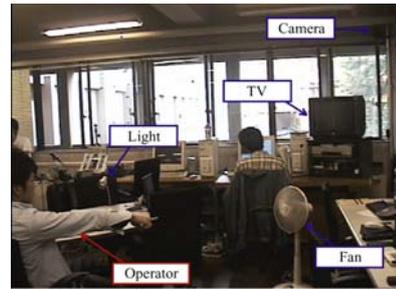


Fig.9 Situation of an experiment



Fig.10 Operation of "8ch"

## 9. 結言

本論文では、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CCD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、肌色領域の抽出のために本論文に適した色空間を提案した。さらに、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、提案したジェスチャの中には得意不得意のあるものが確認されたことから、操作者の判別を行い、操作者に合わせてジェスチャの使い方を変更するなど、ユーザーがより直感的に家電機器を操作できるようにすることや、移動ロボットのインテリジェントルームへの実装、また、画像だけでなく他のモダリティ特に音声の導入などが考えられる。

## 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, pp.691-694, 2000.
- [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] 森武俊, 野口博史, 佐藤知正, "センシングルーム 一 部屋型日常行動計測蓄積環境 第2世代ロボティックルーム", 日本ロボット学会誌, Vol.23, No.6, pp.25-29, 2005.
- [4] 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 一 手のジェスチャによる家電製品の操作一", 第21回日本ロボット学会学術講演会予稿集, 2J15, 2003.
- [5] 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.
- [6] 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.
- [7] 若村直弘, 鈴木健一郎, 梅田和昇, "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", 日本機械学会ロボティクス・メカトロニクス講演会, 2A1-N-051, 2005.
- [8] 酒井幸市, "デジタル画像処理入門", CQ 出版社, 東京, pp.133-136, 2002.