

# 解説

## インテリジェントルームにおけるビジョン技術

Robot Vision Technologies in an Intelligent Room

梅田 和昇\* 入江 耕太\* 寺林 賢司\* \*中央大学/JST CREST  
Kazunori Umeda\*, Kota Irie\* and Kenji Terabayashi\* \*Chuo University/CREST, JST

### 1. はじめに

本解説でいただいたテーマは、家庭での活動を支援するビジョン技術のうち環境型のもを、ということであった。環境や空間を知能化する・ロボット化するという方向の重要性が認識され、これまでに多くの研究が行われてきているのは周知のとおりである。我が国においても、パイオニア的な研究である佐藤・森らのロボティックルーム [1][2] をはじめ、様々なニーズや目的に応じて、多くの優れた研究事例があり (例えば文献 [3]~[7]), 本学会誌においても vol.25, no.4 (2007年5月) で環境知能化の特集が組まれている。

これらの環境知能化・空間知能化においては、モニタリングやインタラクションなどのため、人をセンシングすることが本質的に必要である。このためのセンシング手段として、様々なものが利用されている。例えば、超音波、RFID、圧力センサアレイ、マイクロフォン (音声認識)、人感センサ (焦電センサ) などがある。また、広い意味ではビジョンに含まれるが、距離画像センサ、サーモグラフィーも有用である。三菱電機のエアコンは、8個の焦電センサが縦に並んだセンサを横方向にスキャンすることで温度画像を計測し、エアコンのコントロールに活用している。そして、環境知能化・空間知能化のためのセンシングに最も多く利用されているのが、ビジョン技術であると言える。ちなみに、ビジョンには、監視されることに対する抵抗感やプライバシーの問題がについて回る。それもあってビジョンを一切用いない選択をしている研究もある [6]。ただし、家庭での使用に限定された場合には、屋外の公共の場に設置された監視カメラに比べるとプライバシーの問題は生じにくい (ただし室内カメラが外部からアクセスできない場合)。いずれにしても、環境知能化・空間知能化のためにビジョン技術の重要性が大きいことは間違いない。

本解説では、誌面の関係もあり、これまでの研究例を広

くサーベイすることはせず、我々が構築してきたインテリジェントルーム [8][9] において用いられているビジョン技術を例として紹介しながら、家庭での活動を支援するビジョン技術に関して考察する。

### 2. インテリジェントルームの概要

我々が構築してきたインテリジェントルームの概念図を図1に示す。室内に複数台のパン・チルト・ズーム機能を持つカメラを取り付け知能化した部屋である (なお、インテリジェントルームという用語は、“知能化した部屋”との意味の普通名詞として利用している、すなわち an intelligent room である)。対象は、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ことに特化している。一般に、部屋の知能化では、人の無意識の行動をモニタリング・蓄積したり理解したりすることを対象とすることが多いが、我々のインテリジェントルームでは、人 (操作者) の意図的な動作の認識を対象としている。操作者は、室内の任意の場所で、リモコン、センサグローブやマイクなどを用いず、手・指による自然なジェスチャでこれらの機器の操作を行うことができる。

図2に処理の流れを示す。操作者は、操作する意図を伝えるために手振り動作を行う。2台以上のカメラで手振りを検出すると、ステレオ計測により手振りの三次元位置を得ることができる。この位置情報に従ってカメラをパン・

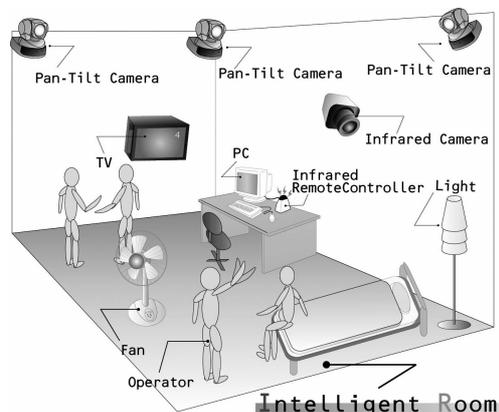


図1 インテリジェントルーム概念図

原稿受付 2009年4月27日

キーワード: Intelligent Room, Robot Vision, Operation of Home Appliances, Gesture Recognition, Human-Machine-Interface

\*〒112-8551 文京区春日 1-13-27

\*Bunkyo-ku, Tokyo

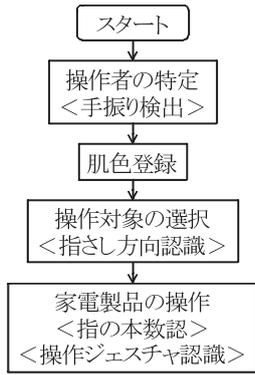


図2 インテリジェントルームの処理の流れ

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。なお、手振り検出時にはズームは広角側としており、部屋の広い範囲で手振りの検出を行えるようにしておく。続いて、検出された手振りを行っている手の色を登録する。こうすることで、手領域の抽出を個人差や照明環境による肌の色の变化に対してロバストに実現できる。

これ以降の処理は、色情報を用いて抽出される手領域に対して行う。操作者は、操作したい機器を選択するために指さしを行う。続いて操作者は選択された家電製品をジェスチャによって操作する。具体的には、指の本数で数を指示したり、直感的なジェスチャによりコマンドを与えたりする。認識された操作内容に応じて機器を制御する。現状では、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っている。また、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るようになっている。

### 3. インテリジェントルームにおけるビジョン技術

我々のインテリジェントルームで利用されているビジョン技術を以下に示す。基本的にはそれ程特殊な処理はなく、一般的な画像処理手法やジェスチャ認識手法の活用を行っている。

#### 3.1 手振り検出による操作者の特定

すでに述べたように、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。その理由は、まず、手振りが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、操作者にとって意図を示すのに自然な動作であること、そして手振りをロバストに検出可能な手法を構築できた[10]ことである。また、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選択することが可能となる。

手振り動作の検出手法の概要は次のとおりである(図3参照)。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

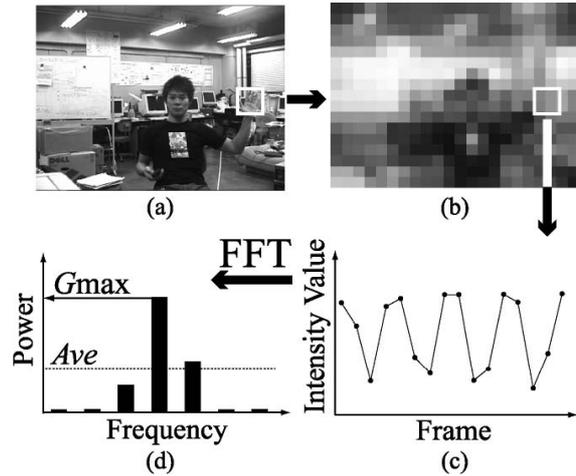


図3 FFTを用いた手振りの検出

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の色の变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、極めて簡潔な処理である。

なお、現在、位相情報も用いることで、手振り以外の周期ジェスチャの識別を行うことができるように改良しつつある[11]。

#### 3.2 色情報を用いた手領域の抽出

肌色登録は、現状のシステムではズーム後操作者がいったん手を静止してから行っているが、現在、手振りを検出しながら登録するように改良しつつある[12]。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)空間を用いている。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。このとき、モルフォロジー処理などでノイズの除去を行う。

なお、HSIのうちSは、定義にもよるが一般的には明るさの影響を受ける。同様に、よく利用されるYCbCrやYUVのCb、CrやU、Vもその定義から明るさの影響を受ける。すなわち、R、G、Bがすべてk倍になれば、SもCb、CrあるいはU、Vもすべてk倍になる。これはカラー画像処理において時々誤解されている点なので注意が必要である。

#### 3.3 指さし方向認識

画像中から手領域を抽出し、その主軸ベクトルを求めることで画像中での指さし方向が決定できる(図4参照)。2台のカメラの画像で指さし方向が決定できれば、指さしの三次元座標での位置・向きが決定できる。しかしながら、手

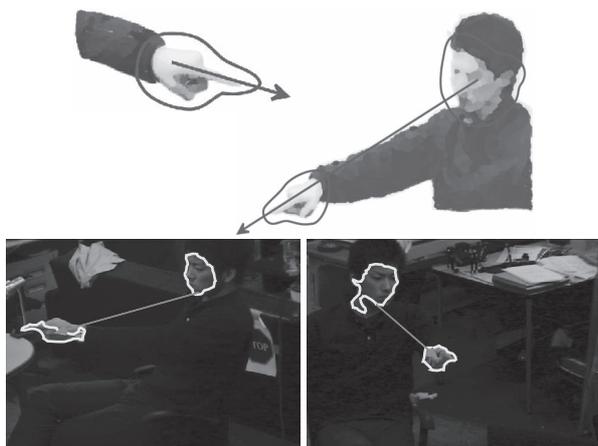


図4 指さし方向の決定(左:手のみ, 右:手と顔)および2台のカメラでの手と顔を用いた指さし方向決定例

領域の主軸ベクトルの算出は必ずしも安定ではなく(特に指さし方向がカメラの光軸方向に近い場合), この方法での指さし方向認識は誤差が大きい. 一方, 手領域だけでなく顔領域も検出されれば, 指さし方向を顔領域から手領域の方向に近似することで, 比較的安定な算出が可能である(図4参照). 構築したシステムでは, 抽出領域が一つの場合は手領域, 二つ以上の場合には領域面積の1番大きい領域を顔, 2番目に大きい領域を手としている. 現状では, 近年盛んに利用されている顔検出手法は利用していないが, その導入は有効であると考えている.

また, 指さし以外での物体指示も検討しており, 読唇による物体指示手法の導入を行いつつある[13].

### 3.4 指の本数認識

数を指示するのに指の本数を利用している. 図5に示すように, モルフォロジー処理[14]のオープニングを利用して指領域を抽出し, その数をカウントしている.

- (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.

なお, 片手の指の本数だと0~5しか入力できないが, 5以上の場合に2回入力することで, 0~9の入力が可能となり, さらにこれを繰り返すことで任意の桁数の数字の入力が可能である.

### 3.5 手のジェスチャ認識

手の直感的なジェスチャによって, いくつかの機器操作のためのコマンドを与えている[9]. ジェスチャ認識には, HMM (Hidden Markov Model) と並んでよく利用されるDPマッチング[15]を用いている. 肌色情報に基づき, 各フレームにおいて手領域を抽出し, 重心移動量(縦, 横),

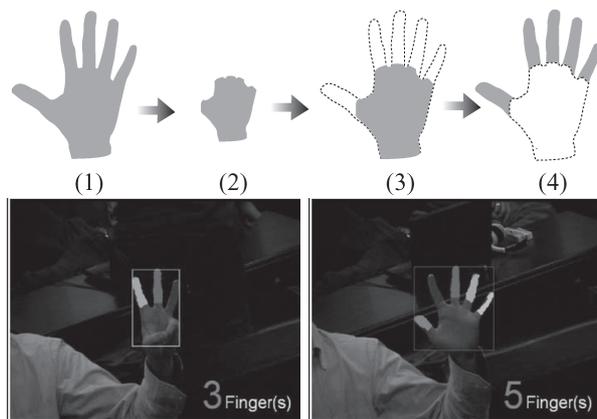


図5 モルフォロジー処理を利用した指の本数認識手法と認識例

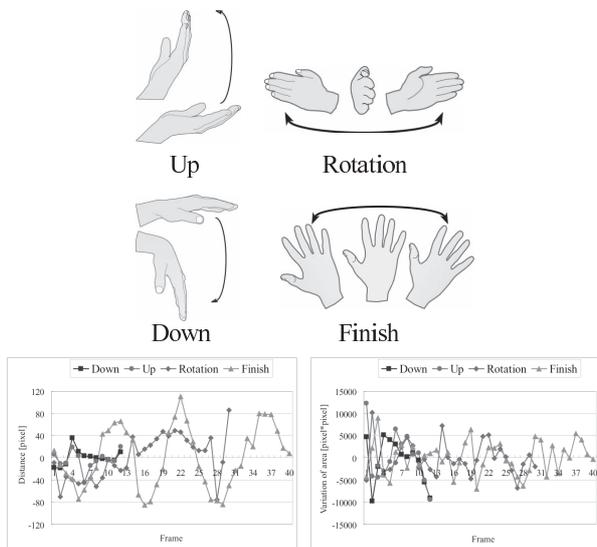


図6 手のジェスチャと登録されたモデルの特徴量の例(左:重心移動量(横), 右:面積変化量)

手領域の幅と高さの変化量, 面積変化量を求め, これら5変数の特徴量として用いている. 図6に用いているジェスチャを示す. これらのジェスチャで例えばテレビのボリュームのコントロールを行っている.

なお, 読唇[13]でもDPマッチングによる認識を行っている. こちらでは, 口唇部の $10 \times 10$ 画素, 計100次元の低解像度画像を特徴量として用いている.

### 3.6 実験例

構築したシステムを用いて家電製品の操作を行った実験例を示す. 実験を行った部屋は $6.9 \times 7.8$  [m<sup>2</sup>]の長方形で, カメラは部屋の角近くの高さ3.2[m]の位置に設置した. 画像処理にはHALCON (MVTec) およびVisual Cを用い, 各種演算処理はPC (DELL, Pentium 4 2.2 GHz)で行った. CCDカメラにはSONY製のEVI-D100を用い, 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に入力している.

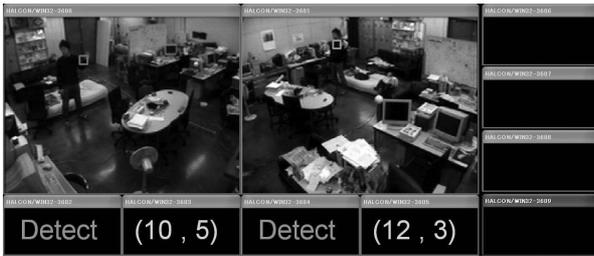


図7 手振りの検出



図8 TVチャンネルの操作

図7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手振りを検出した箇所である。画角が広いこと、および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。また、図8に指の本数認識を用いてテレビのチャンネルを操作しているときの画面を示す。指の本数が2本であることが認識され、すでに認識された1本と合わせて12チャンネルと入力されている。図7と比較すると、ズームにより手が拡大していることが示されている。

#### 4. おわりに

本解説では、家庭での活動を支援するビジョン技術の環境型の例として、我々が構築してきたジェスチャ認識によって家電製品を操作するシステム（インテリジェントルーム）におけるビジョン技術の紹介を行った。用いられているビジョン技術は比較的シンプルなものが多いが、それらを工夫して組み合わせることでそれなりに有効なシステムが組めるのだとの感想を持っていただけたとしたら幸いである。最近、CEATEC [16]で、カメラを搭載しジェスチャにより操作可能なTVが発表された。部屋の知能化に限らず、このような技術が今後普及していくであろう。

#### 参考文献

[1] T. Mori and T. Sato: "Robotic Room: Its concept and Realization," Robotics and Autonomous Systems, vol.28, no.2, pp.141-144, 1999.  
 [2] 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.  
 [3] J.H. Lee and H. Hashimoto: "Intelligent Space—concept and contents—," Advanced Robotics, vol.16, no.4, pp.265-280, 2002.  
 [4] JEITA ハウス, <http://www.eclipse-jp.com/jeita/>

[5] 西田佳史 他: "センサルームを用いた人の日常活動の頑健な観察とその応用", 情報処理学会研究報告 HI, no.106-006, pp.37-44, 2003.  
 [6] 森武俊, 野口博史, 佐藤知正: "センシングルーム—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.665-669, 2005.  
 [7] 上田博唯, 山崎達也: "ユビキタスホーム: 日常生活支援のための住環境知能化への試み", 日本ロボット学会誌, vol.25, no.5, pp.494-500, 2007.  
 [8] 入江耕太, 若村直弘, 梅田和昇: "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", 日本機械学会論文集 C 編, vol.73, no.725, pp.258-265, 2007.  
 [9] 若村直弘, 鈴木健一郎, 入江耕太, 梅田和昇: "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—", 画像の認識・理解シンポジウム (MIRU2005), pp.1074-1081, 2005.  
 [10] 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.  
 [11] 高橋真人他: "濃淡値の時系列変化を利用した周期ジェスチャの認識", 日本機械学会ロボティクス・メカトロニクス講演会 2009 講演論文集, 1A1-C13, 2009.  
 [12] 入江耕太他: "インテリジェントルームにおけるジェスチャ認識のための肌色登録手法", 日本機械学会ロボティクス・メカトロニクス講演会 2009 講演論文集, 2A1-D11, 2009.  
 [13] 中西達也, 寺林賢司, 梅田和昇: "インテリジェントルームのための DP マッチングを用いた口唇動作認識", 電気学会論文集 C, vol.129-C, no.5, 2009.  
 [14] 高木幹雄, 下田陽久 (編): "新編 画像解析ハンドブック", 東京大学出版会, 2004.  
 [15] 西村拓一, 向井理朗, 野崎俊輔, 岡隆一: "低解像度特徴を用いた複数人物によるジェスチャの単一動画からのスポットニング認識", 電子情報通信学会論文集, vol.J80-D-II, no.6, pp.1563-1570, 1997.  
 [16] CEATEC, <http://www.ceatec.com/>



梅田和昇 (Kazunori Umeda)

1989年東京大学工学部精密機械工学科卒業, 1994年同博士課程修了. 同年中央大学理工学部精密機械工学科専任講師, 2006年より同教授. 2003~2004年カナダNRC Visiting Worker. ロボットビジョンの研究に従事. 2004年MIRU長尾賞受賞, 博士(工学). 日本機械学会, 精密工学会, 電子情報通信学会, IEEE等の会員.  
 (日本ロボット学会正会員)



入江耕太 (Kota Irie)

2003年中央大学大学院理工学研究科精密工学専攻博士前期課程修了. 同年(株)日立製作所入社, 車載用画像認識カメラの開発を担当. 2006年中央大学精密工学専攻社会人博士後期課程に入学, 現在に至る. 画像処理を用いたヒューマンインタフェースの研究に従事. 2002年電気学会電子・情報・システム部門大会奨励賞受賞.  
 (日本ロボット学会正会員)



寺林賢司 (Kenji Terabayashi)

2004年北海道大学大学院システム情報工学専攻修士課程修了. 2008年東京大学大学院工学系研究科精密機械工学専攻博士課程修了, 博士(工学). 同年中央大学理工学部精密機械工学科助教, 現在に至る. ヒューマンインタフェース, ロボットビジョン等の研究に従事. 精密工学会, 日本機械学会, 日本バーチャルリアリティ学会等の会員.  
 (日本ロボット学会正会員)