

服装を考慮したポーズ認識による家電制御支援システムの研究

川野卓也† 山本和彦† 加藤邦人† 本郷仁志‡ 丹羽義典‡
(†岐阜大学工学部 ‡HOIP, ソフトピアジャパン/科技団)

抄録

本稿では高齢者や寝たきりの人に対する電化製品を操作するための新しい入力支援システムを提案する。本システムは、利用者が覚えやすいポーズを取ることで電化製品を操作することができる。今回は操作の対象を TV とした。処理の流れは、ベッドで寝ている人物をカメラで観測し、顔領域の重心点を求め、この重心点を基に顔周辺矩形を作成する。次に肌色候補領域のエッジ部分を抽出した後、顔周辺矩形内にあるエッジから高次局所自己相関特徴を抽出し、線形判別分析により識別を行う。利用者が長袖を着用している場合は、矩形内に手と顔の肌色領域の相対情報を持たせることによりポーズを識別できるようにした。これにより、服装に依存されないシステムを構築した。

Research of the Electric Appliances Control Support System by Pose Recognition

Takuya KAWANO † Kazuhiko YAMAMOTO † Kunihito KATO †
Hitoshi HONGO ‡ Yoshinori NIWA ‡
(† Faculty of Engineering, Gifu University ‡ HOIP, Softopia Japan/JST)

Abstract

In this paper, we propose a new input support system that can control electric appliances for aged and bedridden people. This system can operate electric appliances using the poses that are familiar and memorable to them. This time, we operated TV. First, the face and hand region of the user is detected by using the skin color from the input image. Our system focuses the recognition region on the center of gravity of the face region. Next, the edges are extracted within the recognition region and then, the higher order local autocorrelation is extracted. The linear discriminant analysis creates a coefficient matrix that can optimally distinguish among learning data. When the user was wearing a long sleeve, a pose is discriminated by giving the relative information of a hand and a face within the recognition. We built the system, which is not depended on clothes.

1. はじめに

本稿では、高齢者や寝たきりの人を対象とした入力方法の一形態として、電化製品の入力支援システムを提案する。今回はTVの操作を対象とした。現在、様々な家電製品を操作するのにリモコンを使うのが一般的となっている。リモコンは遠隔操作できるので便利ではあるが、手元になければ使えない。例えば、何かの拍子でリモコンがベッドから落ちてしまった、誰かが使って離れた場所へ置いてしまったという場合、高齢者や寝たきりの人などのように、ベッドから

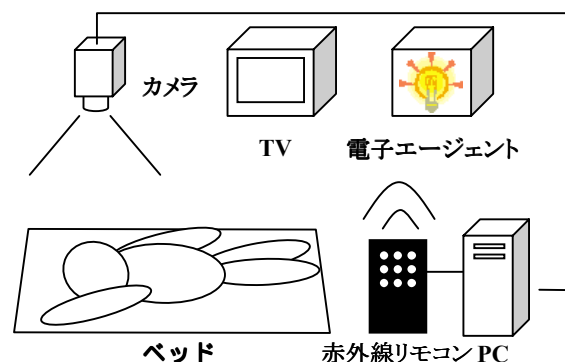


図 1. システム構成

離れられない人達にとっては不便な装置となつてし



(1)ノーマルポーズ



(1)スイッチオン

(2)チャンネルアップ



(4)ボリュームアップ

(3)チャンネルダウン

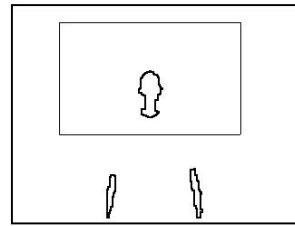


(5)ボリュームダウン

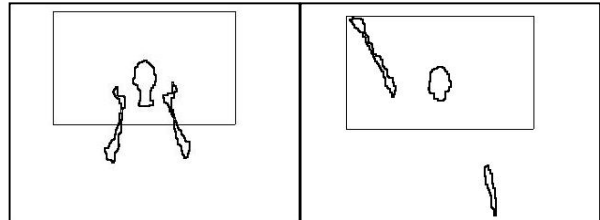
(6)スイッチオフ

図 2. 命令実行ポーズ

まう。この問題を解決するために、ジェスチャー認識 [1][2]、手形認識[3]、音声認識[4]、視線検出[5]などが考えられる。ところが、身振り・手振りといった動きのあるジェスチャーを認識する方法は、ジェスチャーを始めるところと終わるところを認識するのが困難であり、ジェスチャーの動きは利用者に依存される。また利用者はジェスチャーをマスターし、システムがジェスチャーを正確に認識するまで繰り返さなければならないため、利用者に負荷をかけてしまう。手形認識は、手をズームし、高解像度で画像を獲得しなければならない。音声認識は利用者の声がTVの音にかき消されてしまうことが考えられる。このため、スピーカーのように音の出るものを操作する場合、音声認識以外の方法で操作することが望ましい。視線検出

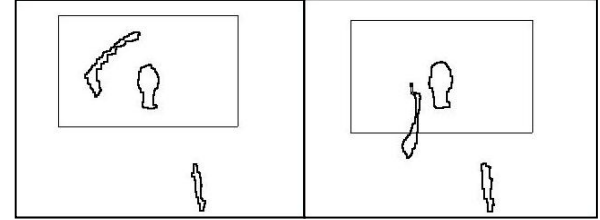


(1)ノーマルポーズ



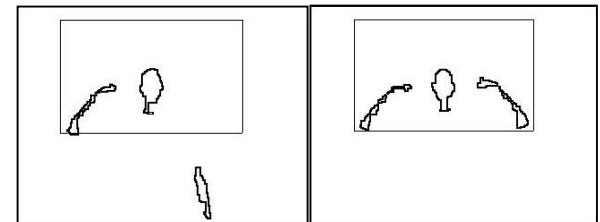
(1)スイッチオン

(2)チャンネルアップ



(4)ボリュームアップ

(3)チャンネルダウン



(5)ボリュームダウン

(6)スイッチオフ

図 3. エッジ画像に顔周辺矩形を当てはめたもの

の場合、システムが複雑な上に、キャリブレーションを行わなくてはならないため、利用者に負荷がかかってしまう。

電化製品を制御するためには、利用者が伝えたいことを確実に伝えることが重要である。このため、静止して確実に命令を送ることができるポーズを用いることが適していると考えられる。しかし、そのポーズを覚えるのに一苦労するようなものであるならば、利用者にやさしいとはいえない。このため日常生活に結びつく非常に身近なもので、覚えやすい動作でなければならない。また、家電製品などの制御を簡単なジェスチャーで実現することにより、リモコンに比べ操作を習得する負担を大幅に軽減することができる[6]。

2. ポーズによるコントロールシステム

2.1 システム構成

システムの構成を図 1 に示す。本システムは人物のポーズを入力するカメラ、TV の誤動作を防ぐための電子エージェント、ポーズの識別及び TV と電子エージェントの制御コマンドを送るコンピューター、TV と電子エージェントの制御信号をあらかじめ学習させておいた赤外線リモコンからなる。

大まかな流れとしては、ベッドで寝ている人物をカメラで観測し画像を取り込む。これを図 2 に示す。そこで得られた画像から肌色情報を用いて被験者の顔と手を抜き出す。次に顔領域の重心点を求め、この重心点を基に顔周辺矩形を生成し、肌色候補領域のエッジ部分を抽出する。これを図 3 に示す。この後、高次局所自己相関特徴を抽出し、得られた特徴量に対して線形判別分析を行い、入力された画像がどのポーズに当てはまるかを識別し、あらかじめ登録されているポーズとの距離が近くなった場合にそのポーズに割り当てられた命令を実行する。

2.2. 命令実行ポーズ

各種の命令を実行するために割り当てられたポーズを図 2 に示す。利用者はベッドの上に横になった状態で TV のチャンネルやボリュームを操作することができる。それぞれのポーズは利用者のしたいことに合わせて動機付けがなされている。これに加え、TV を見ているだけで、何もしていないポーズを用意した。このポーズをノーマルポーズとする。図 2 に示すポーズと、システムに送られる命令の関係を以下にまとめる。

- (1) ベッドに何もせず横たわっているポーズ・・・TV を見ているだけの状態。
- (2) 身体の正面で両腕を折りたたみ両耳にかざすような状態・・・TV を聞きたいポーズ。及び電子エージェントを立ち上げるポーズ。
- (3) 右腕を右斜め上方向にまっすぐ伸ばした状態(利用者からみて時計回りにチャンネルを回すポーズ)・・・TV のチャンネルを一つずつ上げるポーズ
- (4) 右腕を右斜め上方向に伸ばしながら肘より

先を左斜め上方向に曲げた状態(利用者から見て反時計回りにチャンネルをまわすポーズ)・・・TV のチャンネルを一つずつ下げるためのポーズ。

- (5) 右腕を身体の正面で折りたたむようにして聞き耳を立てているような状態(音が良く聞こえないので聞き耳を立てているポーズ)・・・ボリュームを少しずつ上げ、音を大きくするポーズ。
- (6) 右腕を身体の右側で折り曲げ耳を塞ぐような状態(音がうるさいので片耳を塞ぐポーズ)・・・ボリュームを少しずつ下げ、音を小さくするポーズ。
- (7) 両腕を使って両耳を塞ぐような状態・・・TV 及び電子エージェントの電源を切るポーズ。

実際には、まず(2)のポーズで TV の電源が入る。次にもう一度(2)のポーズを入力することで、電子エージェントを立ち上げチャンネルやボリュームを操作できる状態にする。これは、いつでも TV が操作できる状態にあれば、睡眠中にボリュームが上がったりする誤動作が起こる可能性があるためである。ランプが点灯している間だけ命令受け付け可能となる。エージェントが呼び出されているときのみ(3)(4)(5)(6)のポーズをすることで、チャンネル・ボリュームの上げ下げが可能である。従ってエージェントが呼び出されていないときには、装置は作動しないため安心して TV を見ていることができる。(2)のように普通に寝ているだけでは、起こりにくいポーズをスイッチにすることによって誤動作を避けることができる。TV 鑑賞中においても比較的自由的な動きがとれ利用者に負担をかけない。そして、(7)のポーズでチャンネルやボリュームを操作できない(電子エージェントを消した)状態にし、さらに(7)のポーズを入力することで TV の電源を切ることができる。また、(3)(4)(5)(6)の入力がなされないまま一定の時間が経過すると、自動的にエージェントは消され装置は作動しなくなる。

3. ポーズの識別

3.1 肌色抽出

背景が単純であることから、高速に肌色領域を抜き出すことができるRGB画像のR値とB値を用いた処理を行った。R値-B値の計算をすることにより肌色らしい領域の画素値が比較的高く、それ以外の領域の画素値が低い画像が得られる。そこで閾値を設定して2値化処理を行うことにより、肌色領域とその他の領域とに分割した。

求められた領域に対して、膨張処理を行い目や口が抜け落ちた部分を埋め、収縮処理により比較的大きな肌色領域すなわち顔領域の重心点を算出する。求めた点を中心とした矩形を原画像に適用させ、顔周りの矩形内から特徴量を抽出する。

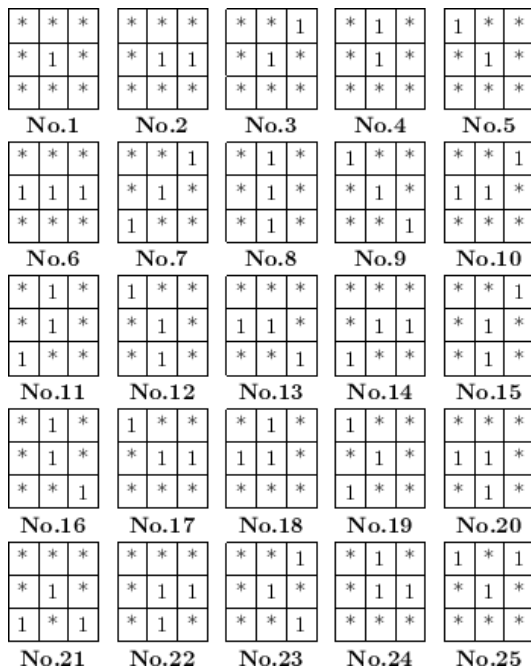


図 4. 高次局所自己相関特徴のマスクパターン

3.2 高次局所自己相関特徴

特徴抽出には、25種類の局所的な3×3ピクセルサイズのマスクを画像上に操作させ、そのパターンと一致した領域の数を加算することで得られる高次局所自己相関特徴[7]を用いた。これにより入力画像から平行移動不変な特徴を高速に抽出することが可能である。このためベッドのどこに寝ても同じ特徴量が

抽出される。得られる特徴の分布パターンは個性を反映した固有なものとなり、識別はこの分布の違いを利用して行っている。この処理は図3に示したような肌色領域のエッジ画像から生成される。2値化画像に対して施される高次局所自己相関特徴の25種のマスクパターンを図4に示す。

3.3 線形判別分析

判別には線形判別分析を用いた。線形判別分析は、与えられたクラスをより低次元で良好に分類する線形写像を構成する多変量解析手法である。ここでは最適な係数行列を学習データから算出している。未知画像の識別では、初期特徴ベクトルと係数行列から新特徴を計算し、学習辞書として作成された平均ベクトルとの比較において決定することで識別を行う。

4. 識別実験

学習データとしては、利用者が半袖を着用している場合と長袖を着用している場合を想定し、図2に示した7つの命令ポーズの各画像につき100フレーム分、半袖を着用している画像から700枚、長袖を着用している画像から700枚の画像を用いて、それぞれ学習辞書を作成した。学習データの識別正解枚数は半袖を着用している画像は700枚で識別率100%、長袖を着用している画像は695枚で識別率99.3%であった。

4.1 実験1

未知データとしては以下に示すような2つのデータセットを用意した。各データセットは(命令ポーズ7種類)×(100フレーム)で合計700枚の画像データで構成されている。学習データと撮影した日時が異なり、服装は異なるものを着用している。

- データセット1・・・利用者は半袖を着用し学習データを作成したときと同じ位置に寝て撮影した画像。このデータ例を図5に示す。
- データセット2・・・利用者は長袖を着用し学習データを作成したときと同じ位置に寝て撮影した画像。このデータ例を図6に示す。

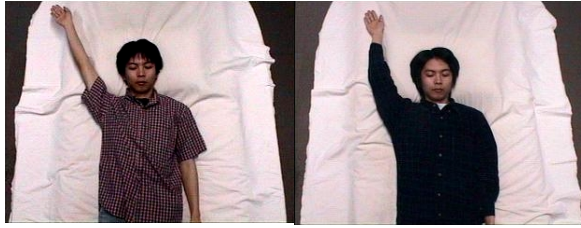


図5. データセット1

図6. データセット2

データセット 1 の識別率は 100%、データセット 2 の識別率は 80%であった。データセット 1 の場合では肌色で抜き出しているため服装の変化には影響されにくく、環境の変化もさほどなく、下腕のエッジが抽出されていたのでこのような結果になったと考えられる。ところがデータセット 2 の場合では利用者が顔の領域と手先の小さい領域だけしか得られないので、チャンネルアップがボリュームアップ、ボリュームダウンがチャンネルダウンといった誤りが目立ち、識別が困難であった。データセット 2 のポーズごとの識別率を図7に示す。グラフの見方は、縦軸が識別率、横軸がポーズの種類を示しており、左からノーマルポーズ、スイッチオン、チャンネルアップ、チャンネルダウン、ボリュームアップ、ボリュームダウン、スイッチオフを示す。

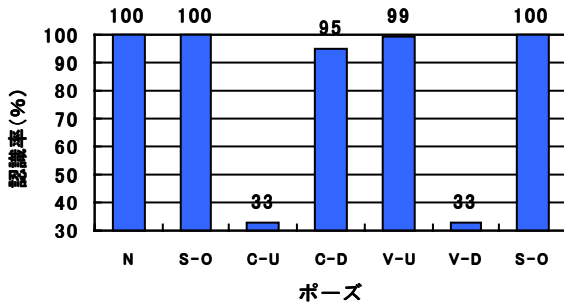


図7. ポーズごとの比較

利用者が半袖を着用している場合は下腕のほとんどが露出している状態である。このため腕の位置情報を用いなくても、肌色領域の長さや大きさ、腕の傾きといった情報のみで識別が可能であった。しかし、長袖のシャツを着た画像では肩から手首へかけての腕の大部分が隠されており、肌色領域を抽出することによって顔と手先の情報だけしか得ることができない。図 8 に利用者が半袖を着用した場合のエッジ画像を、図 9 に利用者が長袖を着用した場合のエッジ画像を示す。

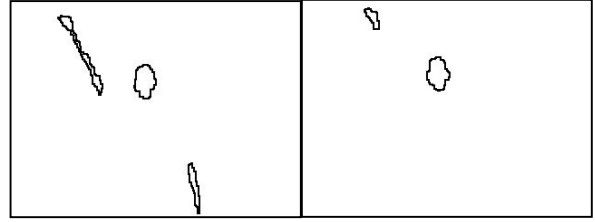


図8. エッジ画像
(半袖着用)

図9. エッジ画像
(長袖着用)

高次局所自己相関特徴はマスク走査によって局所的な特徴を積み上げることで平行移動不変な特徴を抽出するものであり、空間上の位置情報は保持していない。従って、ある物体の形や向きを変えずに位置を変えただけならば同じものとして認識される。このため長袖画像を適用した場合、識別が不安定になってしまうことが考えられる。

5. 空間情報を利用した識別手法

高次局所自己相関特徴の欠点を補うために空間的な位置の情報を取り込んだ特徴量を抽出しなければならない。このような問題を解決するために、空間的な情報を利用する情報として、顔と手の相対的な位置関係を持たせた自己相関特徴を抽出する。

処理の流れとしては、肌色領域を抽出し、顔領域を特定して、顔周辺矩形を求める。その矩形内を図 10 に示すように 3×3 の 9 領域に分割する。そして、顔周辺矩形を図 11 に示したような肌色領域エッジ画像に適用させ、分割された各々の領域から自己相関特徴を抽出するというものである。

このとき、今回用いた命令実行ジェスチャーに対して不必要な部分を除いた 6 つの領域のみを特徴抽出の対象とした。図 11 の右側に示すように顔周辺矩形を分割し、それぞれの領域に番号をつける。そして、特徴抽出の対象となる領域を、顔を含む領域を 0、チャンネルアップのとき右手が入ってくる領域 1、チャンネルダウンのとき右手が入ってくる領域 2、スイッチオフのとき左手が入ってくる領域 4、スイッチオンのときに両手が入ってくる領域 6、ボリュームを操作するとき右手が入ってくる領域 8 とした。この 6 つの領域から 25 次元の自己相関特徴を抽出する。これにより 1 枚の画像から得られる特徴次元数は、各

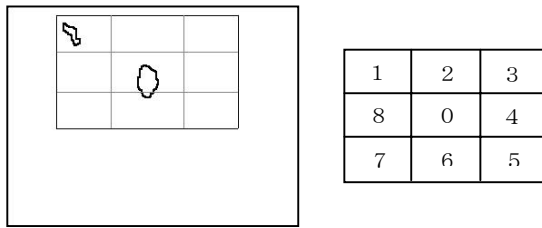


図10. 顔周辺矩形の3×3分割

領域の特徴量を用いることで $25 \times 6 = 150$ 次元ということになる。

これまでの特徴抽出法では画像全体での平行移動に対応しており、ベッド上における利用者の水平・垂直方向への位置ずれに対しては不変な特徴が得られていた。ところがこの方法では、領域を分割することによって、それぞれの領域内での平行移動普遍性は保たれるものの、領域全体としてみた場合にはこの特性が薄れることが懸念される。

しかし、空間を分割することで各領域から抽出された特徴量はその場所の相対位置関係を含んでいるために、腕のエッジが検出しにくくなった長袖を着用しての画像には有効であると考えられる。

実験 1 で用いたデータセットを用いて識別実験を行ってみた。利用者が半袖を着用しているデータセット 1、利用者が長袖を着用しているデータセット 2 ともに識別率 100%であった。

6. まとめ

高齢者や寝たきりの人に対する家電制御支援システムを構築した。本システムは、手と顔の相対情報を持たせることにより、服装の形状によらない識別が可能となった。今後は、被験者を増やした場合への対応、コマンドが増えた場合への対応、より小さいポーズへの対応、学習枚数についての検討、複雑背景への対応、特徴量についての検討、登録されていないポーズへの対応を行っていく予定である。

謝辞

本研究を支援いただいた JST、中部コンピューター(株)に感謝いたします。

参考文献

- [1] 渡辺, 李, 谷内田, : “インタラクティブシステム構築のための動画像からの実時間ジェスチャー認識手法—仮想指揮システムへの応用—”, 信学論(D-II), Vol.J80-D-II, No.6, pp.1571-1580, 1997
- [2] 高橋, 関, 小島, 岡, : ”ジェスチャー動画像のスポットティング認識”, 信学論(D-II), Vol.J77-D-II, No.8, pp.1552-1561, 1994
- [3] L.Xueyin, Z.Xiaoping and R.Haibing, “Hand Shape Extraction and Understanding by Virtue of Multiple Cues Fusion Technology”, Proc. ICMI, pp.103-110, 2000
- [4] B.J.Grosz, D.Appelt, P.Martin and F.Pereira, ”TEAM: An experiment in the design of transportable natural language interfaces,” Artif. Intell., vol.32, no.2, pp.173-244,1987
- [5] Y.Matsumoto and A.Zelinsky ”An Algorithm for Real-time Stereo Vision Implementation of Head Pose and Gaze Direction Measurement” Proc. 4th FG, pp.499-504, 2000
- [6] 伊藤, 山本, 本郷, 加藤, 丹羽, “寝たきりの人を支援するユビキタスインターフェースシステム(システムの提案と基礎実験)”, 知能情報メディアシンポジウム, pp.111-116, 1999
- [7] 栗田, “柔らかな情報処理のための統計的手法の応用に関する研究”, 電総研研究報告, 第 957 号, 1993