

距離画像センサを用いた操作者の姿勢を考慮に入れた ジェスチャ認識システムの構築

Construction of a gesture recognition system that considers an operator's posture
using a range image sensor

○ 武田泰幸 (中央大) 浅野秀胤 (Pioneer)
正 寺林賢司 (中央大) 正 梅田和昇 (中央大)

Yasuyuki TAKEDA, Chuo University, takeda@sensor.mech.chuo-u.ac.jp
Hidetsugu ASANO, Pioneer
Kenji TERABAYASHI, Chuo University
Kazunori UMEDA, Chuo University

These days, home appliances in our living environment are becoming full of functions. On the other hand, the increase of their functions makes their operation complicated. Therefore, an intelligent room which recognizes gestures and support operators is required in various places. In this paper, we construct a gesture recognition system that considers an operator's posture using a range image sensor. The proposed system uses Support Vector Machine (SVM) to recognize gestures. Several experiments are performed to demonstrate the effectiveness of the proposed method.

Key Words: Gesture Recognition, Range Image Sensor, SVM, Image Processing

1. 序論

近年, 日常生活において欠かすことのできない家電製品が, 多機能化, 高機能化している. しかしその一方で, 操作の複雑化という問題も生じている. 家電製品のように人に身近な製品を, 人のジェスチャを用いて直感的に操作することを目的とした研究[1]が多くなされている. 我々は, 図 1 に示すように, 部屋の四隅にカメラを設置し, 操作者のジェスチャを認識して家電製品の操作を行うインテリジェントルームを構築している[2]. 従来システム[3] [4]では操作者がジェスチャ操作を行う際に, カメラに正対するという前提条件がある. そのため, カメラに正対していない場合だと, ジェスチャ認識をするための特徴量を安定して取得することができないという問題がある. 実際の居住環境を考えると, 操作者の部屋での位置・姿勢は多種多様であり, より利便性を高めるためには操作者の位置・姿勢の制約のないジェスチャ認識システムの構築が必要だと考えられる. そこで本研究では, 対象物体までの距離情報が取得可能な距離画像センサである Kinect(Microsoft)を用いて操作者の位置・姿勢を推定し, それらを基にしてジェスチャを認識するシステムの構築を行う. 構築した本システムに対して, 実験を行いその評価をする.

2. 姿勢の推定

操作者の検出, 姿勢の推定に OpenNI (Prime Sense)を用いる. OpenNI では, 人物の検出や検出した人物の各パーツ (関節等) を検出し, その動きをトラッキングする関数がライブラリとして公開されている[5]. このトラッキングは一般的にスケルトントラッキングと呼ばれている. スケルトントラッキングの例を図 2 に示す. 図 2 の赤丸が検出した関節の位置である. 本研究では, 操作者の姿勢推定のために操作者の両肩, 腰の 3 点の座標を用いる. そのため, スケルトントラッキングを行う領域を上半身のみとする. 上半身のみとすることでカメラに近い距離でもスケルトントラッキングを行うことができる.



Fig.2 Skelton tracking

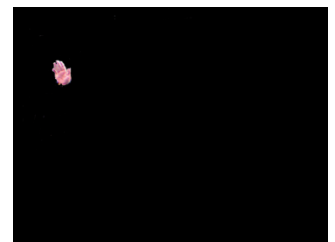


Fig.3 Hand region

3. 手領域抽出手法

ジェスチャを認識するために手領域の抽出を行う. 前章のスケルトントラッキングで得た情報を基に手領域の抽出を行う. 具体的には, スケルトントラッキングで得た手の点からの奥行方向の距離が $\pm 20[\text{cm}]$ の範囲内の点を抽出する. 距離計測する画像領域の大きさは, スケルトントラッキングで得た手の点を中心に, 上下左右にそれぞれ $\pm 50[\text{pixel}]$ の範囲とした. ただし, 腕や肩等, ジェスチャ認識の妨げになる領域も含まれてしまうため, 肌色抽出を行い手領域のみを抽出する. 抽出した手領域の抽出例を図 3 に示す.

4. ジェスチャ認識のための特徴量

抽出した手領域から特徴量を時系列で取得し, Support vector machine (SVM)を用いてジェスチャの認識を行う.

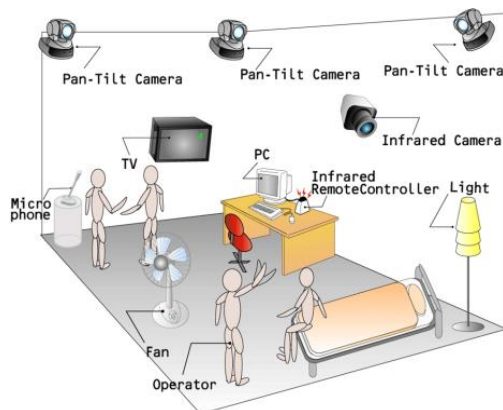


Fig.1 Conceptual figure of our intelligent room

4.1 特徴量の取得

ジェスチャを認識するために用いる特徴量として、手領域の重心、輝度値、距離値それぞれの時系列変化の3種類を用いる。

4.2 重心の時系列変化

手領域の重心の3次元座標(u, v, d)の値は u, v は[pixel]で、 d は[mm]で取得している。そのため、重心の3次元座標をカメラ空間座標に変換する。カメラ座標系における3次元座標を(X, Y, Z)とすると変換は式(1)で求められる。

$$(X, Y, Z) = d \left(\frac{u - C_u}{f}, \frac{v - C_v}{f}, 1 \right) \quad (1)$$

ここで C_u, C_v は画像中心座標の u, v 座標、 f はカメラの焦点距離である。これらのカメラパラメータの値[6]を表1に示す。式(1)により求めた3次元座標の時系列の変化を取得する。

Table 1 Camera parameters [pixel]

C_u	313.7
C_v	259.0
F	526.4

4.3 輝度値・距離値の時系列変化

3章で取得した手領域画像をグレースケール化し、正規化した全画素の輝度値 $I(i, j, t)$ を時系列で取得する(図4参照)。ただし、 $i=1, 2, \dots, m, j=1, 2, \dots, n$ は画像座標、 t はフレーム数である。図5にグレースケール化した手領域画像の取得例を示す。各画素における輝度値の時系列変化を取得する。距離値に関しても同様に時系列変化を取得する。

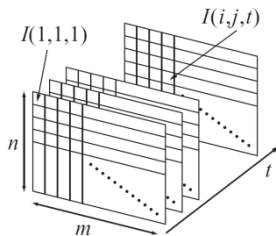


Fig. 4 Time series of images

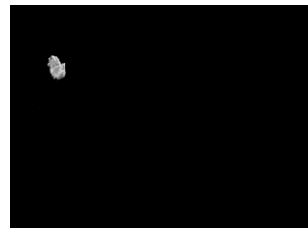


Fig.5 Gray scale image

5. ジェスチャ認識手法

本研究ではジェスチャの認識手法としてSVMを用いる。SVMは近年分類学習モデルの一つとして注目されている手法である。本研究では、ソフトマージン法・カーネルトリックをSVMに適用し学習する際、RBF(Radial Basis Function: 動的基底関数)カーネルを用いる[7][8]。このカーネルはガウスカーネルとも呼ばれている。RBFカーネルは式(2)によって求める。ここで β は重みで、本研究では0.09とした。

$$K(x, y) = \exp(-\beta * |x - y|^2) \quad (2)$$

このカーネルを用いてジェスチャの学習を行った。実験に用いたジェスチャは、Up, Down, Right, Left動作の4種類である。ジェスチャ例を図7に示す。ジェスチャをそれぞれ10

回ずつ行い学習器を作成した。1回のジェスチャには20フレーム分のデータを用いた。

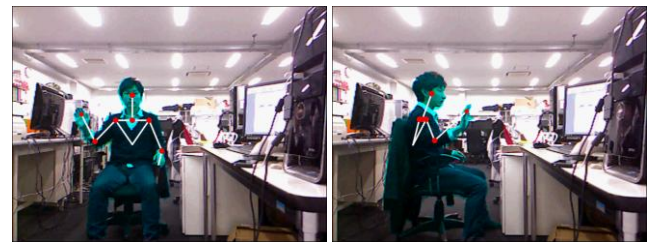
6. 認識実験

構築したシステムの有用性を検証するため実験を行った。Kinectの主な仕様[5]は、画像サイズはVGA、計測範囲は0.5~10.0[m]である。画像処理ソフトはOpenCV, OpenNIを用いた。ジェスチャを認識する手法としてSVMを用いた。スケルトントラッキングの可能な範囲が約1.0~4.0[m]であり、鉛直方向の変化だとスケルトントラッキングの精度が良くなかったため、実験はカメラから1.0, 2.0, 3.0, 4.0[m]離れた場所で、カメラに正対した場合、カメラに対し水平方向に90[°]向いた場合で、それぞれ30回ずつ行った。実験風景を図6に示す。実験結果を表2, 3, 4, 5に示す。

カメラに正対している場合は輝度値の変化、距離値の変化共に、カメラからの距離によって認識率に大きな差はなかった。

90[°]向いた場合のRight, Left動作は、どの距離においても距離値の変化を用いている場合が最も認識率が高い結果となった。これは、このシーンだと、画像中の手領域の動きは他のジェスチャと大差が無いが、距離値では変化があるためだと考えられる。

どのシーンに対しても、重心の変化を用いた場合他よりも認識率が低かった。これは、肌色抽出の精度が安定しておらず、重心の変化に大きなばらつきがあったためだと考えられる。



(a) 0[°]

(b) 90[°]

Fig.6 Experimental scene

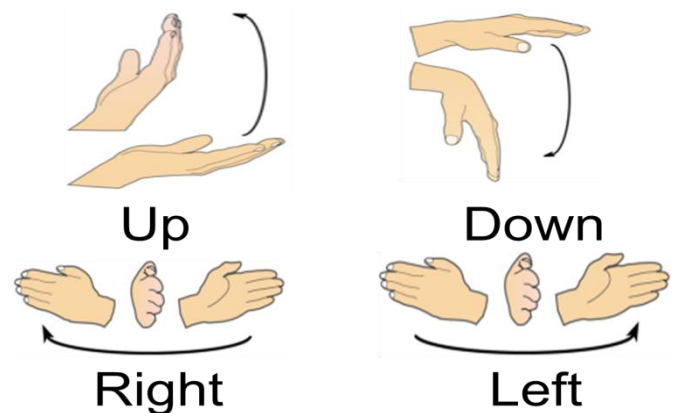


Fig.7 Four Gestures

文 献

- [1] T. Mori and T. Sato: "Robotic Room: Its Concept and Realization, Robotics and Autonomous Systems", Robotics and Autonomous Systems, Vol.28, No.2, pp.141-144, 1999.
- [2] 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.
- [3] 高橋真人, 入江耕太, 荻島秀敏, 寺林賢司, 梅田和昇, “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'09 予稿集, 1A1-C13, 2009.5.
- [4] 滝田清, 永易武, 浅野秀胤, 寺林賢司, 梅田和昇, “テンプレートマッチングを用いた口唇動作認識”, 日本機械学会ロボティクス・メカトロニクス講演会 2011 講演論文集, 2P1-Q12, 2011.5.
- [5] 中村薫: “KINECT センサープログラミング”, 秀和システム, 2011.
- [6] Tercel::Diary: Kinect プチハック (エピソード 3) <http://tercel-sakuragaoka.blogspot.com/2011/04/kinect3.html>
- [7] T. Okabe and Y. Sato, “Support Vector Machines for Object Recognition under Varying Illumination Conditions”, Asian Conf. Computer Vision, pp. 724-729, January 2004.
- [8] [opencv.jp](http://opencv.jp/opencv-1.0) - OpenCV-1.0 リファレンス マニュアル (日本語訳) <http://opencv.jp/opencv-1.0.0/document/>

Table 2 Experimental results (1m) [%]

Gesture Angle		Up	Down	Right	Left
0[°]	Intensity	80	73	73	70
	Distance	73	77	83	73
	Centroid	60	63	67	63
90[°]	Intensity	77	70	63	57
	Distance	70	77	77	80
	Centroid	73	70	60	50

Table 3 Experimental results (2m) [%]

Gesture Angle		Up	Down	Right	Left
0[°]	Intensity	80	83	80	77
	Distance	73	63	83	73
	Centroid	70	63	73	67
90[°]	Intensity	77	73	60	53
	Distance	63	77	90	80
	Centroid	73	70	60	50

Table 4 Experimental results (3m) [%]

Gesture Angle		Up	Down	Right	Left
0[°]	Intensity	70	73	67	77
	Distance	73	70	77	73
	Centroid	67	63	73	60
90[°]	Intensity	70	77	53	50
	Distance	73	77	80	73
	Centroid	67	60	63	67

Table 5 Experimental results (4m) [%]

Gesture Angle		Up	Down	Right	Left
0[°]	Intensity	63	70	67	73
	Distance	70	60	70	70
	Centroid	63	63	60	67
90[°]	Intensity	63	70	57	60
	Distance	67	80	73	73
	Centroid	57	63	63	57

7. 結論

本研究では、距離画像センサを用いて操作者の姿勢を推定し、それを基にジェスチャを認識するシステムを構築し、実験によりジェスチャ認識のための特徴量の比較・検証を行った。比較・検証の結果、距離値の変化が姿勢の変化に最も柔軟に対応できる特徴量であることがわかった。

今後は、さらに認識率を向上させるために、距離値の変化のみを使うのではなく、他の特徴量との組み合わせを行っていく。鉛直方向の変化にも対応できるように、ロバストな姿勢の検出・追跡手法の構築を行う。