

要旨

非接触型ユーザインターフェースを実現させる場合、身振りは一般ユーザにとって有効な手段であると考えられる。本報では、動画像を入力とし、身振りを腕全体の形状の変化と考えて動作を認識する手法を提案し、実験結果からその有効性について検討する。

1.はじめに

現在、人間と機械（コンピュータ）との接点であるインターフェースにはキーボード、マウスといった接触型ユーザインターフェースが主流となっているが、コンピュータの高機能化や一般ユーザへの普及に従い、より使い易いユーザインターフェースが必要とされてきている。また、様々な状況下におけるロボットとの協調作業においては、非接触型ユーザインターフェースの実現は極めて重要であろう。このような要求に応える究極のユーザインターフェースには人間と同じ入出力チャンネルが望まれるであろう。視覚はそうした入力チャンネルの1つであり、もしコンピュータによる視覚が実現されたならばここから豊富な情報を得ることができる。

そこで本研究では、人間の身振りの動作を動画像(連続した画像)から認識することを目的とする。身振りが腕全体の形状(移動領域図形)の変化と、その位置変化により表されると考える。そのため、動画像から移動領域を抽出し、移動領域図形特徴を求め、その時間的変化情報を用いて動作をモデル化し、認識する方法を提案する。

2.システム構成

システムの全体構成を図1に示す。本システムは2種類の処理系から構成される。一方は、標準パターンと呼ばれる動作を認識する際のテンプレート(基準)を作成する処理系である。入力画像系列(入力動画像)から入力動作のモデル化(移動領域図形抽出・特徴抽出)を行い各動作のモデルの時空間ベクトル(標準パターン)を作成する。この処理は動作認識をする前に認識させたい動作の数だけあらかじめ行なっておく。

もう一方は動作を認識する処理系である。入力画像系列から入力動作のモデル化を行い、入力動作の時空間ベクトルを作成し、あらかじめ作成してある標準パターンとの類似度を計算して認識結果を出力する処理系である。

2.1.入力画像系列

本研究ではビデオカメラにより撮影した動画像を、1画素256階調のRGB(赤緑青)表色系(約16万色)の320×240のカラービットマップを用いて保存し、入力画像系列として用いる。

2.2.移動領域図形抽出

移動領域図形を求める為のシステムを図2に示す。また、実画像を用いた例を図3に示す。

1)モザイク化

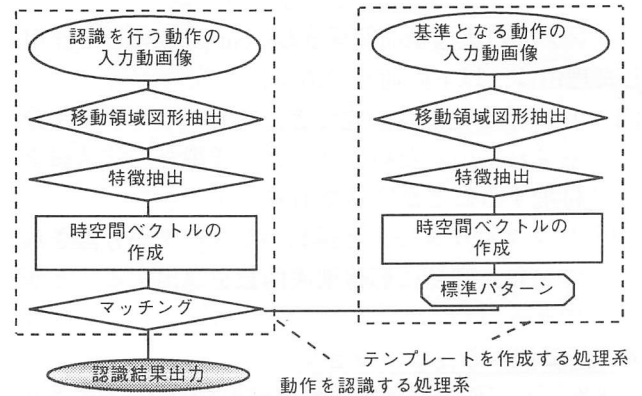


図1 システム構成

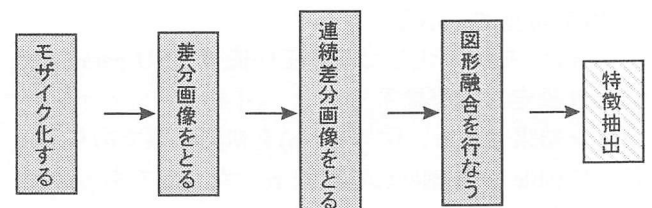


図2 移動領域図形抽出から特徴抽出への流れ

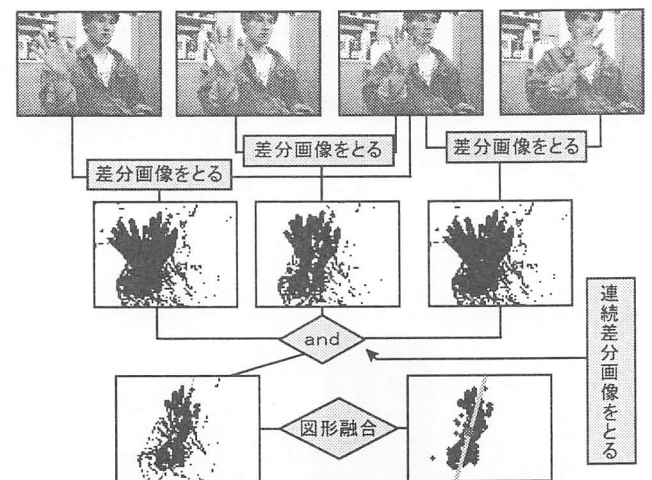


図3 特徴抽出実例

本手法では、RGBそれぞれについてモザイク化を行なう。モザイク化とはロバスト性を持たせるために $m \times n$ 画素を平均化し、1画素にまとめることである。これにより、画面の微妙なちらつき等の影響を減らすことができる。

2)差分画像

本手法では移動領域を求めるために差分画像を用いる。

それは2枚の画面の差を利用する方法である。画面間での差を利用して、画像を定常領域と変化領域に分割する。これを用いて入力画像を2値化(定常領域は0,変化領域は1にする)。RGBそれぞれについて行い、最後に論理和をとる。差分画像には次の利点がある。

- ・差分計算だけであるので計算が容易である。

3)連続差分画像を用いた移動領域図形抽出

1枚の差分画像では2画面間の変化量しか求められないので、移動物体の輪郭(移動領域図形)は求められない。その為、複数の差分画像を用いて、その論理積をとる(連続差分画像)。それにより、移動物体の輪郭を求めることができる。また、複数の差分画像を用いる為、以下の利点がある。

- ・ノイズの影響を受けづらい。

4)図形融合

連続差分画像を用いて求められた移動領域図形にはノイズも多く含まれる。その除去のために収縮・拡散の図形融合を行なう。図4にその例を示す。

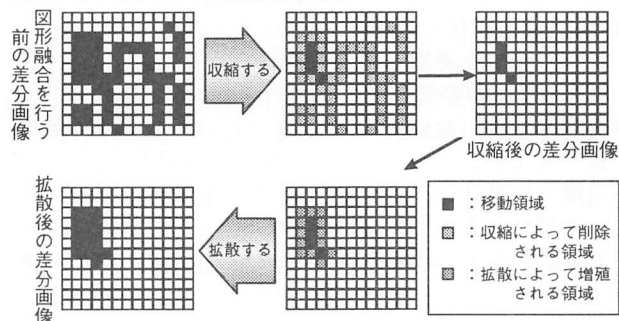


図4 図形融合

2.3.特徴抽出

移動領域図形特徴として以下のものを用いる。

I)動画像フレーム番号

II)モーメント特徴(図形の重心座標、慣性の主軸方向)

モーメント特徴(m_{pq})は図形(S)の各画素が濃度値 $f(x, y)$ の重さを持つものとして次の (p, q) の組み合わせにより、式(1)で定義される。

$$m_{pq} = \sum_{(x,y) \in S} x^p y^q f(x, y) \quad (1)$$

本研究では2値画像をとり扱うので、 $f(x, y) = 1$ となり、図形の重心 (x_G, y_G) と慣性の主軸方向 (θ) は、式(2)と式(3)のようになる。具体例を図5に示す。

$$x_G = m_{10} / m_{00}, \quad y_G = m_{01} / m_{00} \quad (2)$$

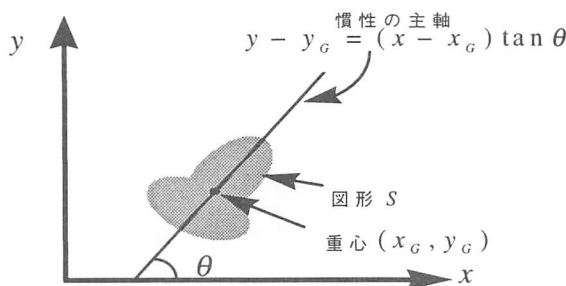


図5 移動領域図形特徴

$$\tan^2 \theta + \frac{m_{20} - m_{02}}{m_{11}} \tan \theta - 1 = 0 \quad (3)$$

図形の重心座標により位置を、慣性の主軸方向により姿勢を、フレームの番号により時間情報を考慮した特徴量になっている。

2.4.時空間ベクトルの作成

求められた特徴量から、以下のように時空間ベクトル(\mathbf{B})を作成する。 ω は重み係数とする。

$$\mathbf{B} = (\mathbf{b}^1, \mathbf{b}^2, \dots, \mathbf{b}^N) \quad (4)$$

$$\mathbf{b}^k = (\Delta X_G^k, \Delta Y_G^k, \theta^k, \omega \times k) \quad (5)$$

\mathbf{b}^k : (第 k フレームの特徴ベクトル)

2.5.マッチング

時空間ベクトルの類似度を式(6)により算出する。

$$D_i = \sum_{n=1}^N \min_{m \in M} |\mathbf{b}^n_{input} - \mathbf{b}^m_{template_i}| \quad (6)$$

\mathbf{b}^n_{input} : 入力の特徴ベクトル

$\mathbf{b}^m_{template_i}$: 標準パターン(i)の特徴ベクトル

1 動作の入力画像系列に対して全ての標準パターンとマッチングを行い、1番距離の少ない標準パターンの動作を入力画像系列の動作であると出力する。

3.動作認識実験・結果

1)認識動作(3種類)

- ・右手を上上げる動作。
- ・机上の目の前にある物体を右手で右に動かす動作。
- ・机上の右にある物体を右手で目の前に動かす動作。

2)入力画像

- ・0.15秒毎に1フレームの割合で撮影を行なった。
- ・被験者とカメラの距離は約3メートルとした。
- ・3動作について4人の被験者(背景は異なる)の入力画像を用いた。

3)実験結果

全ての入力(12の入力画像系列)に対して正しい動作認識結果が得られた。

4.終わりに

本研究では以下のことを提案した。

- 1)身振り動作の縦と横の動作の判別が可能である。
- 2)動作の方向(右に動いているのか、左に動いているのか)の判別が可能である。
- 3)ノイズや誤検出に影響を受け難い手法を提案した。
- 4)背景に影響されない手法を提案した。

【参考文献】

- [1]長谷川純一：画像処理の基本技法〈技法入門編〉、技術評論社、pp.25-61 (1986)
- [2]谷内田正彦：コンピュータビジョン、丸善株式会社、pp.167-190 (1990)