

動作位置の違いに頑健なジェスチャ認識

1 L-6

天田泰亨* , 鈴木基之*** , 後藤英昭** , 牧野正三***

東北大学 *大学院情報科学研究科 , **情報処理教育センター , ***大型計算機センター

1. はじめに

我々は、人間と協同作業を行うオフィスロボットを対象として、様々なコミュニケーション手段を統合的に扱う次世代ヒューマンインターフェイスの構築を目指している。オフィスロボットは、オフィス内で自律的に行動するロボットであり、人間の仕事の手助けや、人間との協同作業を行う。本報告では、その一機能である、動画像からジェスチャを認識するシステムを提案する。

従来は、カメラから得られる動画像全体が、学習した動画像と類似しているかによって認識する方法がとられた（例えば[1]）。しかし、このとき、同じ意味を示す同じ動作であっても、学習した画像と移動部分の相対的位置が異なると認識できなかった。つまり、顔の横で手を振っている画像で学習したモデルは、顔の上で手を振っている画像を認識できず、動作位置の違いなどを吸収できなかった。

この問題に対し、我々は、移動物体の動作を特徴化した、移動ベクトルを抽出する方法を提案した[2]。本手法は、迫江らが提案した単調連続2次元ワーブ法[3][4]を基にした方法を用いて抽出を行う。本報告では、その特徴量を連続DPで認識することで、このシステムが動作位置の違いを吸収できることを述べる。

2. 移動ベクトルの抽出

単調連続2次元ワーブ法（以下、2次元ワーブ法）とは2画像間の最大一致を実現する画素間のマッピング法で、単調連続性を制限として用いた2次元DPを用いている。この方法は、対象物が多少変形したような2画像に対して、それらの歪みが最小となることを基準として、それぞれの画素の対応点を探索する方法である。ある画素の対応点は、その近隣画素の対応点の近辺であるとした連続条件をDP法で実現する。移動ベクトルの取得には、前処理によって静止部分を消去した、時間的に連続である2つの画像に2次元ワーブ法を適用して、対応点を探索することで実現する。

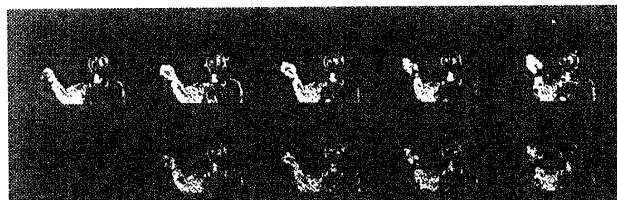


図1 ワーブ法による移動物体の追跡

2次元ワーブ法は、ある画素の対応点を近隣画素の対応点近辺としているが、背景と対象物体の境界では、この条件を満たさない。そこで、近隣画素が対象物体か、背景かで場合分けし、対象物体であれば近隣、背景であれば窓幅 w の範囲を探索範囲とした [2]。

図1は、上5フレームは、元になった動画像であり、下4フレームは得られた移動量を、前フレームに加えたものである。

2次元ワーブ法による、画像間の対応づけと移動量抽出が精度良く行われていることがわかる。

3. ジェスチャ認識部

認識は移動物体の動きを表す、移動ベクトル情報、位置情報や色情報を表す特徴ベクトルを用いて行う。特徴ベクトルはある特徴量空間を時間毎に移動するものと考えられるので、そのパスを認識の基準とする。つまり、認識対象から得られた画素毎のベクトル系列が通るパス、それぞれが、記憶した標準パターンのパスを含むか判断することで認識を行う。これは、音声認識における単語スポッティングに類似する。そこで、単語スポッティングに主に用いられる連続DPを用いて認識を行う。連続DPは、それぞれの時刻を開始点としてDP計算を行うことで、標準パターンと類似する系列を、テストパターンから抜き出すことができる。DP計算には以下の式を用いた。

$$f(i, j) = \min \left\{ \begin{array}{l} f(i-2, j-1) + 0.5 \times d(i-1, j) + 0.5 \times d(i, j) \\ f(i-1, j-1) + d(i, j) \\ f(i-1, j-2) + d(i, j-1) + d(i, j) \end{array} \right.$$

The Gesture Recognition Robust For Change of Gesture Position.

Y. Amada*, M. Suzuki***, H. Goto**, S. Makino***

*Graduate School of Information Sciences, Tohoku University, **Education Center for Information Processing, Tohoku University, ***Computer Center, Tohoku University
Aoba, Aramaki, Aobaku, Sendai, 980-8579, Japan.



図 2 スポットティング判断の結果

ただし、 i はテスト系列の時刻 i のベクトルを、 j は標準系列の時刻 j のベクトルを表し、 $d(i,j)$ はベクトル間の距離を示す。 $f(i,j)$ は、 i と j が対応するとしたときの距離の総和を示す。本報告では、ベクトルの移動量を移動方向・距離、画像中心と移動開始点との距離の対数、色度座標点(r,g,b)の6次元で表し、ベクトル間の距離にはユークリッド距離を用いた。但し、移動量と移動方向を中心に識別することを目的としたため、それらの重みを大きくした。

本手法では、連続DPは画素毎に存在するベクトルすべてに対して行われる。したがって、ノイズなどによる誤認識を考慮すれば、ジェスチャの有無は連続DPによる認識点が、閾値以上存在するかによって判断する必要がある。そこで、標準パターンの画素数を基準にスポットティング判断を行った。本報告では、認識点のうち、それらの中心点に近い順に8割をジェスチャの存在判断に用い、それらが、標準パターンの画素数の6割以上存在すれば、対象とするジェスチャが存在する、とした。

図2は、図1における手のひらの部分の移動ベクトルを標準パターンとしたときに認識されたテストパターンの1部である。○で囲った部分がスポットティング判断の結果を示している。

4. 認識実験

標準パターンは5フレームの動画像とし、右手を“右から左に振っている動画像”と“左から右に振っている動画像”の2シーンとした。テストパターンは15フレームの動画像とし、顔の横や上で手を振っている動画像12シーンと、手招きをしている動画像8シーンの合計20シーンに対して行った。但し、フレームのサンプリングレートは1/30秒とした。スポットティング判断の正解率を表1に示す。スポットティング判断の正解は、シーンにおいて、目的の動作区間をスポットティングできたとき、もしくは、目的の動作が存在しないシーンに対してスポットティングがなされているフレームが存在しないときとした。また、スポットティングは、連続DPによる認識画素数を基準とするため、動作の中心である手がスポットティング



図3 スポットティング判断箇所例

されるとは限らず、腕など移動ベクトルが類似する部分もスポットティングする可能性がある(図3)。そこで、本手法によるスポットティング判断が、実際に手を含んでいるか分類を行った(表1)。本実験では、手のほかに、腕の部分も抽出される場合が、多く確認された。

表1 スポットティング判断の結果

		右から左	左から右
スポットティング判断の正解率		100%	90%
内訳	手	3	1
	手とその他	1	4
	その他	0	2

(単位：正解率 %，内訳 シーン)

6. まとめ

2次元ワーブ法で取得した特徴量を連続DPで認識を行うことで、動作位置の違いを吸収してスポットティングが行えることを示した。今後は、アルゴリズムの高速化と精度の向上をめざす。また、オフィス内では、物品、または認識対象外の人など、様々な要因で認識対象が一部隠れてしまう場合が考えられる。これに対しても、移動ベクトルを基に認識する方法で解決する方法について検討を行う。

参考文献

- [1] 栗田ら「PACOR 画像の高次局所自己相関特徴を用いた背景変化および平行移動に強いジェスチャ認識」, PRMU96-213, pp.159-165 (1997)
- [2] 天田, 牧野ら「ジェスチャ認識のための単調連続2次元ワーブ法を用いた移動量検出」, 信全大 (1999-03)
- [3] 内田, 迫江「動的計画法に基づく単調連続2次元ワーブ法の検討」 信学論 vol.J81-D-II, No.6, pp.1251-1258 (1998)
- [4] 内田, 迫江「単調連続2次元ワーブ法によるオフライン手書きひらがな認識」 PRMU98-163, pp103-110, (1998-12)