

手の形状認識における弾性テンプレートモデルの提案

吉池 紀子[†] 武藤 佳恭^{††}

本論文では、自己組織化手法として知られている弾性ネット手法を拡張した弾性テンプレートモデルを用いた手の形状認識手法を提案する。本手法はトップダウンな制約であるテンプレートからの入力とボトムアップな制約である入力画像からの入力を相互作用させることにより画像中から目的のパターンを抽出する。この相互作用の効果により、対象物の位置ずれや歪みを吸収した認識が可能になった。本論文では、数種類の手の形状パターンを用いて、弾性テンプレートモデル、ボトムアップ手法、トップダウン手法の3種類の実験結果を比較し、リアルタイム性と認識精度における本手法の有効性を示した。

Hand Pose Recognition using Elastic Templates

NORIKO YOSHIKE[†] and YOSHIYASU TAKEFUJI^{††}

We present a hand pose recognition system using Elastic Template model which is an extension to self-organization algorithm. Elastic Template model searches an approximated optimum solution of a cost function that consists of bottom-up and top-down restriction. This model enables to detect a target object from an input image including position changes and distortions. The statistical analysis of experimental results shows that our method is practical for hand pose recognition in terms of real-time and robust performance.

1. はじめに

本研究では、非接触型のジェスチャー認識システムを構築することを目的としている。非接触型のジェスチャー認識システムとは、カメラからの画像入力を用いた手ぶりジェスチャーによる入力を可能にするシステムである。ジェスチャー認識システムの構築は、より人にやさしいインターフェイスの開発のため必要とされている。特に、高齢者や障害者の意志伝達のためにジェスチャーを情報家電のコントロールコマンドに変換するシステムとして有効である。

ジェスチャー認識の研究は近年多く研究されており、ジェスチャーの動きに関するパターン認識の研究では、2), 3), 4) などが代表的である。また、形状に関する研究では、1) などが代表的である。しかしながら、現在実用なアプリケーションとしてジェスチャー認識システムが実装されているものはほとんどない。これは、複雑な背景下や異なる照明環境での認識が難しいこ

と、人の手のような複雑に変形する形を対象とした認識が難しいことがあげられる。これらの複雑な背景下での認識問題に対しては確立した方法が与えられていない。

本研究では、ジェスチャー認識システムを手の形状の認識の処理と動作の認識の処理の2段階に分けている。本論文では、ジェスチャー認識問題の中で手の形の認識問題を扱う。提案する手法は、自己組織化手法として知られている弾性ネット手法を拡張した弾性テンプレートモデルを用いた新しい手の形状認識手法である。弾性テンプレートモデルは、画像からの入力とテンプレートの拘束条件からなるコスト関数を自己組織化を用いて近似的に最小化する。このモデルにより、対象物の位置ずれや歪みを吸収する大域的な特徴抽出を行うことが可能になった。従来位相保持マップを用いた手法⁶⁾⁷⁾と比べて、位相保持制約の代わりに傾き保持の制約を定義することにより、1次元の演算素子配列による高速な認識モデルを構築している。本手法の有効性を示すため、数パターンの手の形状を弾性テンプレートモデルを用いた場合の認識結果と弾性ネット手法を用いた場合の認識結果、テンプレートマッチング手法を用いた場合の認識結果を比較する。

[†] 慶応義塾大学大学院政策・メディア研究科
Graduate School of Media and Governance Keio University

^{††} 慶応義塾大学環境情報学部
Faculty of Environmental Information

2. 手の形状認識システム

2.1 システムの概要

全体のシステムは大きく分けて、特徴抽出、学習、認識の3段階に分かれる。特徴抽出の段階では、テンプレートを持たない弾性ネットモデルにより、動作中の手の形状を抽出し、学習用の手の形状データを生成する。学習の段階では、内積型のLVQ⁵⁾により、手の形状の汎化学習を行なう。認識の段階では、弾性テンプレートモデルを用いた特徴抽出を含む認識を行ない、動作中の手の形状を特定する。以下に、弾性テンプレートモデルを用いた認識手法の詳細を述べる。

2.2 弾性テンプレートモデルを用いた認識

2.2.1 位置推定

特徴抽出の段階と同様、オプティカルフローベクトルを用いた取り込みと対象物の位置推定を行なう。オプティカルフローとは、連続する2枚の画像の差分から予測された、動物体の動き方向ベクトルである。ここでは、8)の手法を用いた。一定以上の大きさを持つオプティカルフローベクトルが一定の数以上出現した時に静止画の取り込みを行なう、一定以上の大きさを持つオプティカルフローの生じた矩形領域を画像中から切り出す。

2.2.2 方位選択性ニューロンフィルタ

切り出した矩形領域に対し方位選択性のニューロンフィルタをかけることにより、特定の方向の輪郭に反応するニューロン層の出力とする。方位選択性ニューロンフィルタのモデル化には、9)の手法を参考にした。このモデルは、フィルタとしてモデル化されたオンオフ細胞の配置位置によって、特定の方向を持つエッジのゼロ交差を抽出する。

2.2.3 方位選択性テンプレートニューロン

汎化学習によって得られたテンプレート M_p に対しても、方位選択性テンプレートニューロン T_{ph} ($p = 1, \dots, pnum$), ($h = 1, \dots, 4$) により特定の4方向に対する出力を与える。

2.2.4 弾性テンプレートモデル

ここでは、画像からの入力とテンプレートからの入力を用いて、弾性テンプレートモデルにより、特徴抽出を含んだ認識処理を行なう。図(1)にニューロン間の結合関係を示す。図(1)に示すように、個々のニューロンは隣のニューロンとテンプレートニューロン、入力ニューロンとの結合を持っている。

2.2.5 動作式

エラスティックニューロン Y は、一致度を表す関数を C_{match} 、変形度を表す関数を C_{deform} とすると、

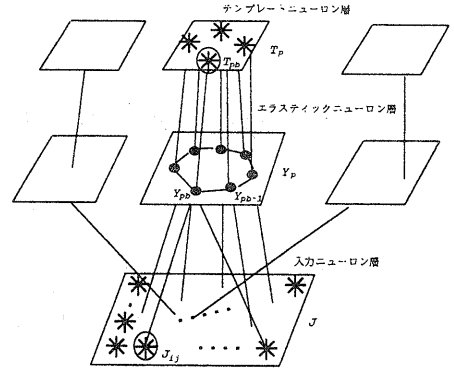


図1 ニューロン間の結合関係

エラスティックニューロン Y の変化分 ΔY は、式(1)-式(2)によって更新される。

$$Y_{pb}^{(t+1)} = Y_{pb}^{(t)} + \Delta Y_{pb} \quad (1)$$

$$\begin{aligned} \Delta Y_{pb} &= C_{match} + \kappa C_{deform} \\ &= \alpha \sum_k w_{bk} \cdot (Pos_{J_k} - Y_{pb}) \\ &\quad + \kappa(t) \cdot \beta \cdot (O_b - Y_{pb}) \end{aligned} \quad (2)$$

ここで、 Pos_{J_k} は入力ニューロン J_k の位置を表すベクトルである。また、 w_{bk} は式(3)によって決められる入力ニューロン J_k とエラスティックニューロン Y_b との結合の強さを示すパラメータである。

$$\begin{aligned} w_{pbk} &= \sum_h T_{phb} \cdot J_{hk} \\ \text{if}(k &= \max_{k'} (|Pos_{k'} - Y_{pb}| \times h(\sum_h T_{phb} \cdot J_{hk'}))), \\ &= 0 \quad \text{otherwise.} \end{aligned} \quad (3)$$

ここで、 $h(x)$ は、式(4)によって与えられる閾値関数である。

$$\begin{aligned} h(x) &= 1 \quad \text{if}(x > \theta_h), \\ &= 0 \quad \text{otherwise.} \end{aligned} \quad (4)$$

また、式(2)の O_b は、隣のエラスティックニューロン Y_{b-1} とテンプレートから相対的に求められる位置ベクトルであり、式(5)によって求められる。

$$O_b = Y_{b-1} + rs(t) \cdot rt \cdot M_{pb} \quad (5)$$

ここで、 $rs(t)$ は現在のエラスティックネット $Y^{(t)}$ の大きさと初期状態 $Y^{(0)}$ の大きさとの比率を表す。また、式(2)の $\kappa(t)$ は時定数である。

2.2.6 評価関数

パターン p に対する評価関数は式(6)によって与えられ、最終的に $E(p)$ の値の最も大きいパターン p が解として選ばれる。

$$\begin{aligned}
E(p) &= \lambda_1 \cdot E_{match} + \lambda_2 \cdot E_{deform} \\
&= \lambda_1 \sum_b \sum_h T_{phb} \cdot J_{hPosY_{pb}} \\
&\quad + \lambda_2 \sum_b \cos(\phi_{defb} - \phi_{tempb}) \quad (6)
\end{aligned}$$

ϕ_{defb} は変形後のエラスティックニューロン Y_{pb} と $Y_{p(b-1)}$ をつないだベクトル $Y_{pb} - Y_{p(b-1)}$ の傾き角、 ϕ_{tempb} はテンプレート M_{pb} の傾き角を表す。

式(6)の評価関数は、エラスティックニューロンの最終状態におけるエネルギーと対応するものである。

3. 実験

3.1 学習結果

トップダウン手法、ボトムアップ手法と本手法において、内種型LVQによる共通の学習結果を実験中に用いた。学習用のパターンは背景ノイズや位置ずれを含まない理想的な条件で得られた画像を用いた。特徴抽出には弾性ネット手法を用いた。得られたパターンベクトルに対し、3種類のグー、チョキ、パーの手の形状それぞれ20パターンのベクトルを学習させた。結果として得られたベクトルを図(2)に示す。

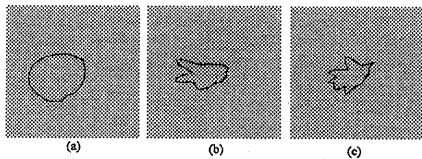


図2 学習後のベクトル

3.2 認識率による比較

ここでは、弾性テンプレートモデルを用いた手の形状認識の実験結果を示す。比較1から比較4では、認識の際に以下の4通りの条件において、動作中の3種類の手の形状認識実験を行なった。

- 実験1. 形の歪みがあり、背景ノイズが無い。
- 実験2. 形の歪みがあり、背景ノイズがある。
- 実験3. 形の歪みがあり、位置ずれがある。
- 実験4. 回転による歪みがある。

これらの条件下で得た画像に対し、それぞれ、トップダウン手法のテンプレートマッチングによる認識、ボトムアップ手法の弾性ネットによる認識、自己組織化手法の弾性テンプレートモデルの認識結果を統計的に比較する。結果を図(3)に棒グラフとして示す。

また、理想条件の入力画像に対し、弾性テンプレートマッチング手法を適用した結果を図(4)に示す。

実験では、Visual C++言語を用いて本システムを

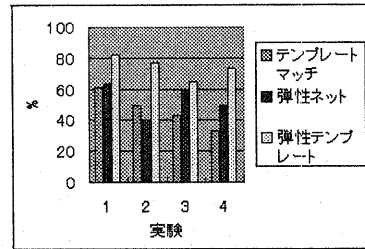


図3 実験結果の比較

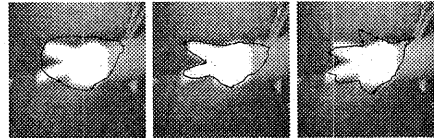


図4 理想条件での入力画像(b)に対する弾性テンプレート

コーディングし、Celeron 300MHzのマシン上で実験を行なった。入力には、解像度 64×64 、8ビットグレースケールの画像と解像度 32×32 のオブティカルフローを用いた。また、すべての実験を通して、以下の共通のパラメータを用いた。

$$\begin{aligned}
\theta &= 500, \alpha = 0.02, \beta = 1.0, \\
\theta_h &= 0.5, \lambda_1 = 0.5, \lambda_2 = 0.1, \\
\kappa(t) &= 0.25 - 0.025 \times t/10.
\end{aligned}$$

3.3 実行時間の計測

実験5では、画像中に手を含む領域のサイズの変化による実行時間の比較を行なう。

結果を図(5)に散布図として示す。

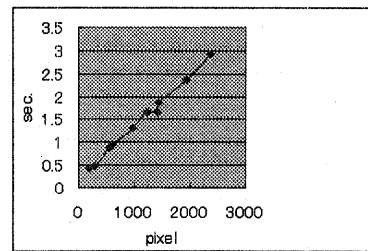


図5 入力画素数と実行時間の散布図

4. 考察

4.1 モデルの柔軟性

実験1の結果から、本論文で提案する弾性テンプレート手法は、他の手法と比較して、理想条件の時に対象物の歪みに最も良く適応したとすることができる。実験2では、弾性ネット手法がノイズのある環境条

件では最も認識率が低くなっていることが分かる。これは、弾性ネット手法は形のテンプレートを持たないボトムアップ手法であり、背景ノイズに影響されやすいためである。また実験3では、テンプレートマッチング手法は最も位置ずれに適應できていないことが分かる。このような位置ずれを含む場合にはトップダウン手法であるテンプレートマッチング手法は入力に適應できないため、失敗が起る。これらの弱点に対し、自己組織化手法である弾性テンプレートは、実験2と実験3において他の手法と比較して良い結果を示している。これらの結果から弾性テンプレートモデルはトップダウン手法とボトムアップ手法の両手法の弱点をカバーしていることが分かる。これは、弾性テンプレートの更新式(2)がボトムアップ的な制約項とトップダウン的な制約項から成っており、双方の制約を満たすように動作しているためである。

実験4の結果から、弾性テンプレート手法は回転のある入力画像の場合も良い認識率を保っていることが分かる。テンプレートマッチング手法は、回転による歪みがあると対象物の位置を正しく推定できないことを反映していると考えられる。また、弾性ネット手法では、抽出したパターンとテンプレートとの内積計算が多少の回転歪みを吸収できない場合があることを反映した結果である。

4.2 実行時間

図(5)から、入力サイズにほぼ比例して実行時間が増加していることが分かる。また、Celeron 300MHzのマシンを用いて、遅くとも3秒以内で解を出すことが示されている。入力サイズに対して実行時間が指数関数的には増加しないため、ジェスチャー認識システムとしての実用的な応用が可能である。

この結果は、一次元の円結合をした少ない演算素子を用いているためである。ネットワークの更新式の中で、隣のニューロンからの相対的な位置関係に基づいた形の保持制約を行なったことにより、(式(5))一次元の円結合により形を保つことが可能になっている。

5. おわりに

5.1 結論

本論文で提案する弾性テンプレート手法は、ボトムアップ手法とトップダウン手法の両方の側面による長所を生かし、対象物の歪みやノイズ、位置ずれ、回転などの入力の変化に柔軟に適應できる手法である。本手法の認識の柔軟さは、従来のボトムアップ手法、トップダウン手法との比較実験により示された。

また、入力サイズが大きくなっても実行時間は指数

関数的に増加せず、実行時間は遅くとも3秒以内であるため、リアルタイム処理に向けた手法である。

5.2 今後の課題

今回の実験において、理想条件の場合に弾性テンプレートモデルを用いた結果は82%となっていた。失敗した入力画像の中には、動作中の手を撮影したために対象物である手がおれて形の情報が失われてしまった例が多く見られた。このような失敗を避けるためには、静止画の情報だけでなく他の情報を用いることを今後の課題としたい。例えば、グーで手を回すと「O.K.」パーで手を横に振ると「バイバイ」など、動作と手の形の2つの情報を用いて認識を行なうシステムである。このようなジェスチャーによる意志伝達は、より人のコミュニケーション形態に近いと、人に近いインターフェイスの開発により近付くと期待できる。

参考文献

- 1) 木下晶雄, 鶴岡信治, 三宅康二, 石田宗秋: カラー濃淡画像からの手領域の抽出と指先位置の特定方法, 信学技報, 99-104,01, 1997
- 2) 西村拓一, 岡隆一: ジェスチャー動画像の逆時間動作や静止動作をスポッティング認識するための Non-monotonic 連続 DP, 信学技報 06, 1996
- 3) 今川和幸, 呂山, 猪木誠二, 松尾英明: 肌色領域により隠れて見える場合を考慮した手話動画像からの手の実時間追跡, 電子情報通信学会論文誌 D-II No.8 pp.1787-1795 1998年8月
- 4) 佐川浩彦, 酒匂 裕, 大平栄二, 崎山朝子, 阿部正博: 圧縮連続 DP 照合を用いた手話認識方式, 電子情報通信学会論文誌 D-II Vol.J77-D-II No.4 pp.753-763 1994年4月
- 5) Kohonen T.: Self-Organizing Maps, Springer-Verlag Berlin Heidelberg New York, 1995
- 6) Kohonen, T.: Self-organized Formation of Topologically Correct Feature Maps, *Biol. Cybern.*, 43, 59-69(1982)
- 7) Bienenstock, E. and Doursat, R.: A shape recognition Model using Dynamical Links, *Network*, 5, 241-258(1994)
- 8) Freeman, W.T., Anderson, D.B., Beardsley, P.A., Dodge, C.N., Roth M., Weissman, C.D., Yerazunis, W.S., Kage, H., Kyuma, K., Miyake, Y. and Tanaka, K.: Computer Vision for Interactive Computer Graphics, *IEEE Computer Graphics and Applications*, 42-53, 18, 3, 1998
- 9) David Marr: VISION A Computational Investigation into the Human Representation and Processing of Visual Information, *W.H. Freeman and Company*, 1982