

# 連続 DP に基づくジェスチャの早期認識の検討

森 明慧\* 内田 誠一\*\* 倉爪 亮\*\* 谷口 倫一郎\*\* 長谷川 勉\*\* 迫江 博昭\*\*  
 (九州大学 \*工学部電気情報工学科 \*\*大学院システム情報科学研究科)

## 1 まえがき

ジェスチャ認識については、マンマシンインタフェースの提供を目的として数々の研究が行われている。これまでの手法は、入力完了後に結果を出力する形式が多い。このためジェスチャの認識結果提示までに遅延が発生していた。

本稿では、連続 DP を用いた認識手法において、入力ジェスチャが開始して間もないような過渡的な段階であっても、曖昧性が無いと判定されれば直ちに、そのジェスチャを認識結果とすることで、遅延を短縮する手法を提案する。以下に、その具体的な実現方法と、それに対する実験の結果を報告する。

## 2 早期認識の手法

### 2.1 連続 DP による従来の認識手法

連続した入力に対して、各時刻での認識結果を出力する方法に連続 DP [1] がある。連続 DP は実時間処理に適しており、ジェスチャ認識にも応用されている [2,3]。以下に連続 DP の漸化式を示す。

$$g_{c,t}(\tau) = \min \begin{cases} g_{c,t-1}(\tau-1) + 3d_{c,t}(\tau) & \text{(a)} \\ g_{c,t-1}(\tau-2) + 2d_{c,t}(\tau-1) + d_{c,t}(\tau) & \text{(b)} \\ g_{c,t-2}(\tau-1) + 3d_{c,t-1}(\tau) + 3d_{c,t}(\tau) & \text{(c)} \end{cases} \quad (1)$$

ここで、 $d_{c,t}(\tau)$  はフレーム  $\tau$  が標準パターン  $c$  のフレーム  $t$  に対応づけられた時のコストである。認識結果は、標準パターン長  $T$  として、累積コスト  $g_{c,T}(\tau)$  が最小であるような  $c$  である。ただし、最小の累積コストが閾値  $\theta_1$  以下で無ければ、そのフレーム  $\tau$  は過渡期として認識結果を出力しない。この手法を用いると、原理的にジェスチャ開始後  $T/2$  フレーム以上経たないと認識結果は出力されない。

### 2.2 サポートを用いた早期認識

本手法では早期認識のために (1) 式と併せて、以下の論理値 (サポート)  $h_{c,t}(\tau)$  を計算する。

$$h_{c,t}(\tau) = \begin{cases} h_{c,t-1}(\tau-1) \cap q(d_{c,t}(\tau)) & \text{if(a)} \\ h_{c,t-1}(\tau-2) \cap q(d_{c,t}(\tau-1)) \cap q(d_{c,t}(\tau)) & \text{if(b)} \\ h_{c,t-2}(\tau-1) \cap q(d_{c,t-1}(\tau)) \cap q(d_{c,t}(\tau)) & \text{if(c)} \end{cases}$$

記号 (a)-(c) は (1) 式中の記号に対応している。また、 $q(x)$  は  $x$  がある値  $\theta_2$  以下の時に真値を返す論理関数である。従って、 $h_{c,t}(\tau)$  が真値をとるのは、経路上の全てのマッチングにおいて、標準パターン  $c$  との誤差が一定値  $\theta_2$  以下となる場合である。このように、サポートは連続 DP の累積コストよりも厳しい基準である。

あるフレームでサポートが真となる標準パターンが唯一ならば、曖昧性が無いと判断して、そのパターンを認識結果として出力する。具体的な認識の手順を以下に示す。

1. まず、従来の連続 DP による認識を行う。従って、最小の累積コストが  $\theta_1$  以下ならば、その結果を出力する。 $\theta_1$  以上ならば過渡期と考え、ステップ 2 へ進む。
2. そのフレーム  $\tau$  における全てのサポートの値を調べる。もし、サポートがある唯一の標準パターン  $c$  で真値をとるならば、曖昧性が無いとして、現在の認識結果を  $c$  とする。それ以外の場合は、認識結果を出力せず、次のフレームへ進む。

## 3 実験

従来の連続 DP と、サポートを併用した場合とについて、比較実験を行った。まず、実験に使用するデータとして「万歳」、「ばいばい」、「指さし」という 3 種類のジェスチャをそれぞれ 10 パターンずつ用意した。パターンの平均フレーム長は約 250 フレームである。これら 30 個のデータに対して一つ抜き法を使って、閾値  $\theta_1, \theta_2$  を変化させながら認識実験を行い、各閾値での、最初の認識結果が出るまでの平均フレーム長と、その誤認識率を調べた。ただし、どの標準パターンも  $\theta_1$  を越えなかった場合 (つまり、リジェクトされた場合) も、誤認識として扱っている。

実験結果を図 1 に示す。今回の場合、従来法で最も認識率が良いのは  $\theta_1 = 250$  付近であった。このとき結果が出力されるまでに平均 87 フレームかかった。

これに対してサポートを用いた場合は、仮に  $\theta_1 = 250, \theta_2 = 190$  と設定したとき、平均 31 フレームで結果が出力できた。また、認識率も従来法とほぼ同様になっている。 $(\theta_2$  については、値を前後に 20 程度変化させても同様の結果が得られる。) つまり、サポートを用いることで、従来法と同程度の認識率を保ったまま半分以下の時間で認識できたといえる。

## 4 まとめ

本稿では、連続 DP での過渡期において、入力ジェスチャに曖昧性が無ければ認識までの時間を短縮できるとした。結論としては、連続 DP にサポートを併用することで、認識率を保ったまま従来より早期に認識できることを示した。謝辞 本研究の一部は総務省戦略的情報通信研究開発推進制度の支援を受けた。

参考文献 [1] 岡. 日本音響学会音声研究会資料, S78-20, 1978.

[2] 高橋他. 信学論, vol. J77-DII, no.8, pp.1552-1561, 1994.

[3] 太田他. 精密工学会誌, vol.63, no.6, pp.812-818, 1997.

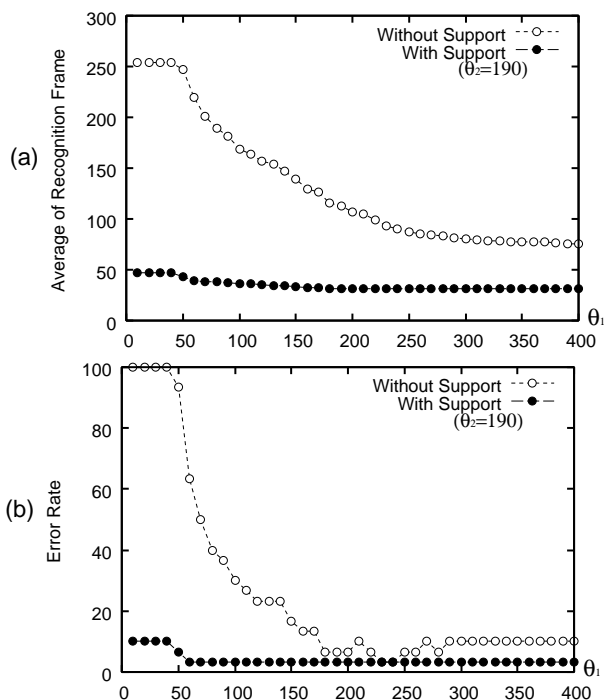


図 1:  $\theta_1$  を変化させたときの、認識までの平均フレーム長 (a) と、誤認識率 (b)