

## 三次元情報に基づく環境変動にロバストな ジェスチャ認識システム

### 1 はじめに

キーボードやマウスといった伝統的な入力デバイスに代わるものとして、Natural User Interface(NUI)と呼ばれる、ジェスチャによる直感的なヒューマンインタフェースが注目を集めている。近年、Kinect センサに代表される三次元距離センサの低価格化に伴い、ジェスチャ認識の分野においても時空間情報を用いた研究が報告されている。代表的な従来法 [1,2] では、手や腕の動き特徴として、時間的な手の位置の変化量を用いている。しかし、個人差（被験者の体型）や撮影環境（カメラからの距離）などにロバストではない。さらに、手形状特徴として、手領域の輪郭や面積などといった 2 次元情報のみを用いているため、多様な形状の手を識別することが困難である。

本論文では、多様な手の形状を識別するため、Kinect センサより得られた距離画像から手領域を検出し、局所的な勾配情報に基づく手形状の識別に有効な特徴を提案する。また、時空間情報から得られる骨格モデルに基づき、個人差や環境変動にロバストな手や腕の移動特徴を提案する。認識には HMM を用い、従来法と比較することで提案手法の有効性を示す。

### 2 ジェスチャ認識システム

提案するジェスチャ認識手法の概要を図 1 に示す。

#### 2.1 手領域検出部

手領域抽出では、Kinect センサから得られる骨格モデルに基づいて左右の手、肘と肩の三次元座標を取得する。次に、得られた手の座標値を用いて、図 2(a) に示すような Kinect センサより得られるユーザ距離画像から手領域のみを検出する。図 2(a) から手領域のみを切り出した結果を図 2(b) に示す。しかし、図からも明らかなように、切り出した手領域には頭や体などといった認識に不必要な領域も含まれている。この問題を解決するため、ジェスチャは主に体の前（センサ側）で手を変化させることに着目し、距離値のヒストグラムに対して大津の閾値処理を用いることによって、図 2(c) のような手領域のみからなる画像を得る。

#### 2.2 特徴抽出部

ジェスチャ認識に用いられる特徴は、手や腕の動き特徴と手形状特徴に大別される。まず、手形状特徴として、局所的な勾配情報に基づく手形状の識別に有効な特徴を提案する。図 2(c) のように切り出した手は、センサからの距離や位置などによって、スケール変換と回転変換の影響を受ける。そのため、手形状特徴として、それらスケールや回転変換にロバストな特徴が望まれる。

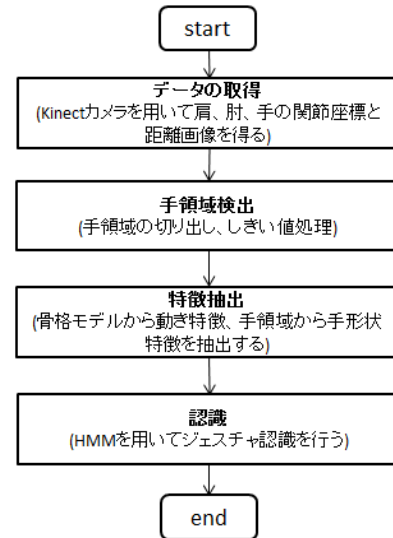


図 1: 提案手法の概要

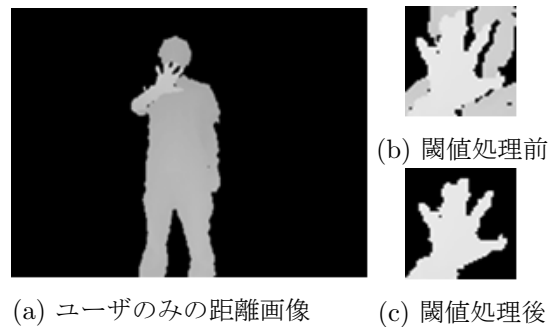


図 2: しきい値処理による手領域の検出

我々は、近年注目されている特徴量記述アルゴリズムである Speed Up Robust Feature (SURF) のキーポイント検出と特徴記述のアルゴリズムに着目し、高速かつ環境変動にロバストな特徴を提案する。一般的な SURF における特徴量記述処理では、回転に関する特徴であるオリエンテーション計算が含まれているが、回転の角度は図 3 に示すように手一肘間の関係から推定可能であるため、ここから得られる値  $\theta$  をオリエンテーションとして用いる。また、従来の SURF の場合は、特徴量記述処理において正方形領域を対象として局所勾配情報の特徴を記述するが、我々は図 2(c) で切り出した手領域全体を一つの矩形として考え、64 個のグリッド領域に分けて、そこから以下に示す SURF 特徴を抽出する。ここで、 $sh \times sw$  は各グリッド領域のサイズであり、特徴ベクトルのスケール変化に対応するための正規化処理に用いられる。

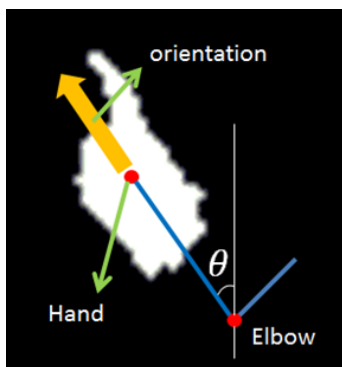


図 3: オリエンテーションの推定

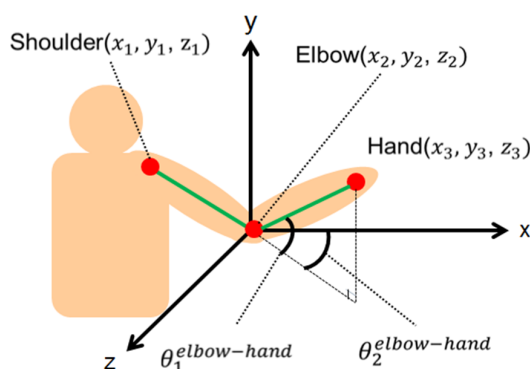


図 4: 角度特徴の概要

$$\frac{1}{sh \times sw} \sum_{(x,y)} dx(x,y) \quad (1)$$

$$\frac{1}{sh \times sw} \sum_{(x,y)} dy(x,y) \quad (2)$$

$$\frac{1}{sh \times sw} \sum_{(x,y)} |dx(x,y)| \quad (3)$$

$$\frac{1}{sh \times sw} \sum_{(x,y)} |dy(x,y)| \quad (4)$$

一方、手や腕の動き特徴として、Kinect センサから得られる手や肘、肩といった関節の三次元情報から各関節間の角度とそのフレーム間の変化量に着目し特徴として定義した。

### 2.3 認識部

本論文では、これまでに述べた動き特徴や手形状特徴に対して、登録されるジェスチャごとに HMM によりモデル化することで認識を行う。HMM を用いたジェスチャ認識では、各ジェスチャに対応する HMM モデルを作成し、特徴ベクトルから出力確率が最も高くなるように HMM のパラメータを学習を行う。

## 3 評価実験

提案手法の有効性を検証するため、実データを用いたジェスチャ認識実験を行った。実験データとして、11 個の日本語手話単語を用いる。この中には、手形状特徴の識別性を評価するため、腕の動きは類似しているが手形状が異なる単語が複数含まれている。また、従来研究 [2] の実験で用いられている手話単語も比較のために用いることとした。撮影条件として、被験者にはセンサから 2000mm, 2500mm, 3000mm の位置に立って手話を行っていただいた。ただし、学習データには距離 2000mm データのみを用いる。被験者は普段手話を使わない健聴者 4 名である。なお、各被験者に対して、1 単語につき 3 回撮影を行った。また、HMM は状態数 13 の Left-to-Right 型であり、10 個ガウシアン混合モデルベースで構成した。

提案手法と従来手法 [2] との結果を表 1 と表 2 に示す。表 1 は学習データとテストデータが両方とも同じ場所で撮影したデータを用いた認識率であり、表 2 は異なる場所で撮影したデータを用いた認識率である。結果より、提案手法が従来法に比べて有効であることが確認できた。

表 1: 同じ場所で撮影したデータを用いた認識率 (学習 2000mm, テスト 2000mm)

| (%)  | 全て   | 腕動き  | 手形状  |
|------|------|------|------|
| 提案手法 | 81.8 | 54.8 | 82.4 |
| 従来手法 | 77.6 | 41.3 | 82.2 |

表 2: 異なる場所で撮影したデータを用いた認識率 (学習 2000mm, テスト 2500, 3000mm)

| (%)  | 全て   | 腕動き  | 手形状  |
|------|------|------|------|
| 提案手法 | 81.9 | 56.1 | 72.3 |
| 従来手法 | 68.4 | 17.2 | 68.1 |

## 4 おわりに

本論文では、環境変動や個人差にロバストなジェスチャ認識を実現するため、Kinect センサより得られる時空間情報を用い、手形状の識別に有効な三次元特徴を複数提案した。また、三次元情報から得られる人の骨格モデルに基づき、複数の関節間の位置と動きを手と腕の特徴として提案した。

## 参考文献

- [1] 東山和弘, 小野智司, 王宇, 中山茂: “3 次元テンプレートマッチングによる指文字認識”, 電学論 C, Vol. 125-C, No. 9, pp. 1444-1454 (2005)
- [2] 佐藤新, 篠田浩一, 古井慎熙: “ToF カメラによる 3D 手話認識”, 画像の認識・理解シンポジウム (MIRU2010), IS3-44, pp. 1861-1868 (2010)