手話学習者のためのパーティクルフィルタによる手話の自動認識への試み

Challenge to Automatic Recognition by Particle Filter for Learners of Sign Language

浦川　由輝子、向　直人
Yukiko URAKAWA, Naoto MUKAI
椙山女学園大学 文化情報学部
Department of Culture-Information Studies, Sugiyama Jogakuen University
Email: uya11da018@st.sugiyama-u.ac.jp

あらまし: 本研究は手話学習者の補助を目的とし、対象者の手話を画像処理によって実時間認識することを試みる。実時間認識を可能にするため、高速に推定が可能なパーティクルフィルタを用いる。実験では、各画素の RGB 値と検出対象となる RGB 値との距離を基に、手話者の手の動きの軌跡を追う。検出された軌跡と手話のパターンを比較することで、学習者は正確な手の動きを学ぶことができる。
キーワード:　手話学習、手話認識、パーティクルフィルタ

1.　はじめに

現在、日本の多くのろう学校では、手話の授業は行われず、読唇から言葉を理解する口話法が採用されている。しかし、2011 年の法改正により、手話は言語として見直され、民間やろう学校の現場において手話教育への興味､関心が高まっている｡そこで、本研究では、手話学習者の補助を目的とし、パーティクルフィルタ[1,2]を用いた手話の認識を試みる。従来、物体追跡の方法として､「背景差分法」や近赤外線を用いた「ドットパターン投影方式」などのアルゴリズムが提案されてきたが、中でもパーティクルフィルタは他のアルゴリズムに比べ、比較的実装が容易であり、不規則な動きに対応しやすいという特徴を持つ。しかし、肌色を基準に手の動きを追跡すると、顔や背景に影響され推定位置に誤差が生じやすいという欠点がある。また、奥行きを用いた表現が必要な手話の認識も難しい。そこで､「磁気センサー」や「データグローブ」を用いた手法や、ゲームデバイス「Kinect」の近赤外線センサーを用いた研究が進められている[3,4]。Kinect に搭載されている可視光センサーと深度センサーによって観測された三次元的な距離から、手の肌色領域のみを抽出することができる。これにより、三次元空間における手の動きを追跡することで高精度な手話の認識が可能となる。

本研究では、これから手話の学習を始めようとしている人を対象に、手話認識を活用して学習をサポートすることを目的とする。本稿では、手話認識の第一歩として、特別なセンサーを用いず RGB 画像のみを用いた手話認識を試みる。

2.　手話動画の撮影

スマートフォンのカメラを用いて「待つ」「大丈夫」「寝る」「わかる」の 4 パターンの手話を撮影する。手話者には色のついた手袋を装着してもらい、画素の RGB 値の変化から手の動きの軌跡を認識する。

本稿で対象とする手話の特徴は以下の通りである。「待つ」は、図 1 に示すように、親指以外の手を軽く折り曲げ、顎の下に当てるという動作からなる。動きがほぼ垂直のみである点と上部で一時停止する点が特徴である。「大丈夫」は、図 2 に示すように、左右の鎖骨に手を当てる動作からなる。動作の軌跡が三角形であることが特徴である。「寝る」は、図 3 に示すように、拳をこめかみに当て目をつむるという動作からなる。この動きは、動画上部で一時停止する「待つ」と特徴が似ているが、中心からやや右にそれた場所で一時停止していることがもうひとつの特徴としてあげられる。「わかる」は、図 4 に示すように、胸に手を当てながら下になでおろすという動作からなる。

図 1　「待つ」

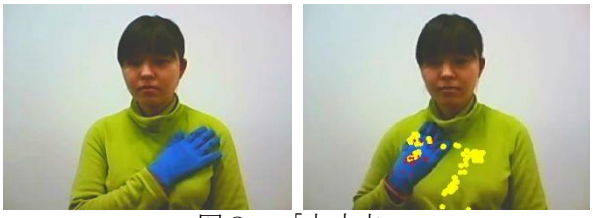

図 2　「大丈夫」

3.　パーティクルフィルタによる手話認識

パーティクルフィルタを用いて手袋の軌跡を検出する。ここでは、図 5 に示す 5×5 ピクセルの画像を例に用いて対象物体の検出プロセスを説明する。検

出プロセスは「尤度の更新」「リサンプリング」「予測」の 3 つのステップで構成される。

図 3　「寝る」

図 4　「わかる」

3.1 尤度の更新

画像上に配置された各パーティクル（粒子）の尤度を計算する（初期状態では、対象物体が特定できていないため、一様にパーティクルを配置する）。ここで、尤度とはパーティクルの置かれた画素が対象物体である尤もらしさを表し、画素の RGB 値を基に定める。対象物体の RGB 値を(r_o, g_o, b_o)、パーティクルの画素の RGB 値を(r, g, b)とすると、画素間の距離Dは下記の式(1)で求めることができる。

$$D = \sqrt{(r_o - r)^2 + (g_o - g)^2 + (b_o - b)^2} \qquad (1)$$

次に、平均0、標準偏差σの正規分布$N(0, \sigma^2)$を考え、距離Dを確率変数としたときの確率密度を尤度Lとする。よって、尤度Lの取り得る範囲は$0 < L < 1$となる。距離Dが小さいほど尤度Lは高く、逆に距離Dが大きいほど尤度Lは低い値となる。例えば、図 5 において画素(3,3)が対象物体だとすると、画素(3,3)のパーティクルの尤度が最も高くなる。また、色の似た画素(2.2)、画素(4,3)の尤度はそれよりも低く、色の異なる画素(5.1)、画素(2,5)の尤度は最も低くなる。

3.2 リサンプリング

パーティクルに設定された尤度に基づき復元抽出する（パーティクルの重複を許す）。ここで、パーティクルは“ルーレット選択”に基づき抽出する。よって、尤度の高いものは高確率で選択され、尤度の低いものは淘汰されることになる。図 5 において、5 つのパーティクルの尤度Lの和は 1 となるため、画素 (3,3) のパーティクルが選択される確率は$0.4/1 = 40\%$となり、画素(5.1)のパーティクルが選択される確率は$0.1/1 = 10\%$となる。

3.3 予測

動画の各フレームにおいてパーティクルの平均位置を求め対象物体の推定位置$(\bar{x}, \bar{y})$とする。図 5 においては、$\bar{x} = \frac{2+2+3+4+5}{5} = 3.2$、$\bar{y} = \frac{1+2+3+3+5}{5} = 2.8$が、対象物体の推定位置となる。また、前フレームとの推定位置の座標の差分（つまり対象物体の速度(v_x, v_y)）を求めておき、リサンプリングされたパーティクルの座標にこの差分とノイズを加えることで、次フレームのパーティクルの位置とする。

図 5　対象画像（5×5 ピクセル）

4. 実験結果

実験により認識された手話の軌跡が図 1、2、3、4 の右側である。「大丈夫」と「寝る」は軌跡が特徴的であり、認識が容易であった。一方、「待つ」と「わかる」の軌跡は酷似していた。このため、軌跡だけでは「待つ」と「わかる」を区別することは難しかった。区別するためには手の角度を考慮する必要がある。

5. まとめ

今回の手法では、実際に日常で使用される手話を単語ごとに認識することは難しい。しかし、我々の目標は手話学習者が、日常会話によく使う単語を覚えやすいように支援することにある。将来的には、手話の様子を録画した動画をアップロードするとサーバー側で処理され、認識した文章が表示されるというウェブ・サービスの構築を考えている。

参考文献

[1] 北川源四郎：“モンテカルロ・フィルタおよび平滑化について”，統計数理, Vol.44, No.1, pp.31-48(1996)
[2] 鈴木いおり，西村洋介，堀内靖男，黒岩眞吾：“パーティクルフィルタと HMM による動画像からの手話認識に関する検討”，電子情報通信学会, pp.25-29 (2011)
[3] 西村洋介，今村大輔，堀内靖男，篠崎隆宏，黒岩眞吾：“Kinect とパーティクルフィルタを用いた HMM 手話認識手法の検討”，電子情報通信学会, pp.161-166 (2012)
[4] 森昭太，松雄直志，白井良明，島田伸敬：“手話認識のための距離情報を用いた隠蔽を含んだ顔・手領域抽出”，情報処理学会研究報告, pp1-6(2013)