J-022

GCN を用いた指差し指示位置の推定

Target Point Estimation for Pointing Gesture Using Graph Convolutional Network

中川 莉那 † 中井 満 †

Rina Nakagawa Mitsuru Nakai

1. はじめに

近年，ジェスチャインタフェースの発展により，指差し動作による機器の操作が可能になってきた．指差し動作で指示位置を推定する従来研究として，目・肩・肘・手首・指の根本・指先の 6 点を入力とした多層パーセプトロン (MLP) で推定する手法 [1] がある．本稿では，3 次元空間の座標情報を用いて，グラフ畳み込みネットワーク (GCN) で 2 次元指示位置を推定する手法を提案する．

2. システムの構成

提案法のシステム構成を図 1 に示す．RGB カメラより骨格情報を推定する．その関節に対して Depth カメラを用いて深度情報を取得し，3 次元空間の座標を得る．関節をノード，接続関係をエッジとした姿勢グラフを GCN の入力とし，指差しで指示したスクリーン上の位置座標 (x, y) を推定する．

図 1: 指差し指示位置推定システムの構成

2.1 姿勢グラフの作成

Google が提供している MediaPipe[2] を利用して RGB 画像中の手および全身のランドマークを検出する．提案法で取得するランドマークは図 2 に示す赤い丸のとおり，右手 21 個と全身のうちの 12 個の部位である．RGB 画像から取得できるランドマークの座標は身体の中心からの相対座標である．また，Depth カメラを用いることでカメラから見えるランドマークまでの正しい深度が得られる．この 2 つの情報を統合することで，取得したランドマークのワールド座標が得られる（図 3）．これらのランドマークを接続して姿勢グラフとする．

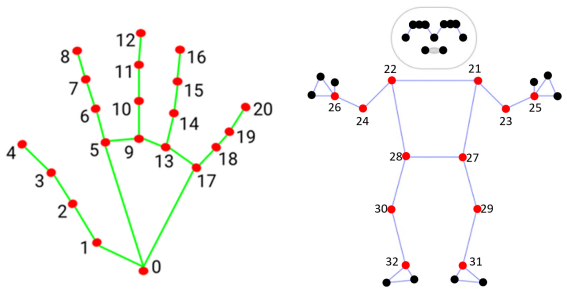

図 2: 右手と全身のランドマーク [2]

†富山県立大学, Toyama Prefectural University

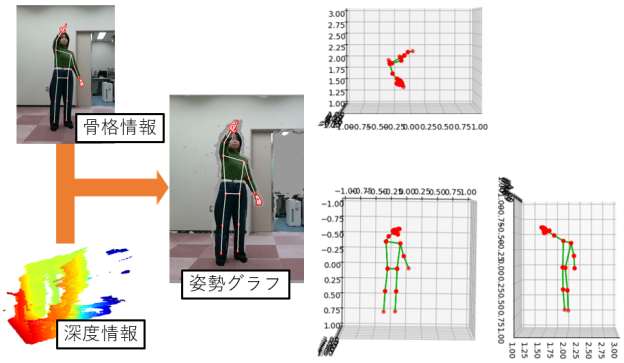

図 3: 姿勢グラフの作成

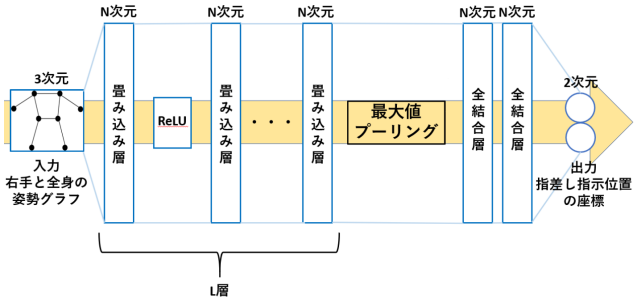

図 4: グラフ畳み込みネットワークの構成

2.2 GCN による指差し指示位置推定

姿勢グラフから指差し位置座標を推定する GCN を図 4 に示す．姿勢グラフは右手および全身のランドマークの 33 ノードからなり，各ノードは 3 次元のワールド座標 (x, y, z) を持つ．これを GCN の入力情報とする．グラフ畳み込み層では，各ノードに入力された情報にグラフのエッジで接続される隣接ノードの情報を畳み込んで更新する．1 層目では畳み込みの後，$3 \times N$ 次元の重み行列によって，各ノードの 3 次元の情報を N 次元の情報に拡張する．畳み込み層からの出力は，活性化関数 ReLU を通して，次の畳み込み層に入力する．2 層目以降は重み行列が $N \times N$ になることを除いて，1 層目と処理は同じである．これを L 層まで繰り返す．その後，全てのノードの情報を集約して N 次元の特徴量を抽出する．ここでは最大値プーリングによって，各次元の最大値を選択する．最終的に 2 つの全結合層を通して，2 次元の指示位置座標を出力する．

3. 実験

3.1 データ収集

スクリーン上に表示したポインタ（点）を目標として指差す姿勢のデータを収集した（図 5）．センサ（intel RealSense

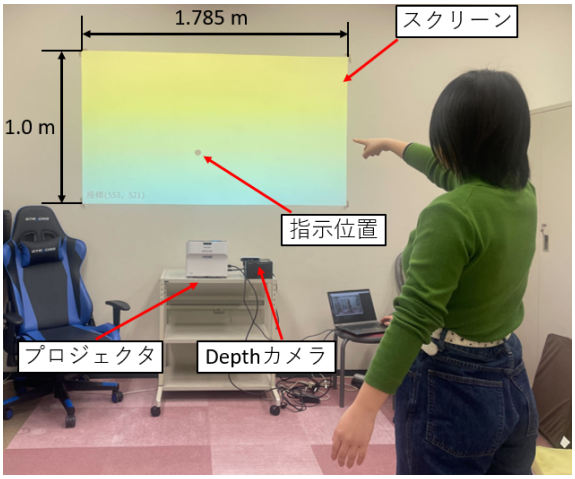

図 5: データ収集環境

DepthCamera D455）はスクリーンの下に設置し，RGB 画像と深度画像の静止画を正面から撮影した．スクリーンから 2.0〜3.5m の範囲で指差しやすい場所を立ち位置として選んでもらった．スクリーンにポインタをランダムに表示し，同じ立ち位置からの指差しを 1 分間繰り返した．これを 1 セットとし，休憩を挟んで立ち位置を変更した．大学生 22 名に対し，1 日 10 セットで 2 日間収集したところ，1 日目は合計 6304 サンプル，2 日目は 7040 サンプルとなった．

3.2 GCN による推定精度

推定精度は，スクリーンのポインタを目標指示位置として，推定指示位置との誤差で評価した．次元数 N および層数 L と指示位置の推定誤差の関係を調べた予備実験より，$N = 512$, $L = 3$ とした．そのパラメータで推定した結果を表 1 に示す．同日モデルでの評価サンプルはモデルの学習に使用したサンプルであり，学習できていることを確認するための実験である．別日モデルは服装が異なるなど実際に起こりうる状況を想定した評価実験である．別日モデルの誤差の平均は 0.232m となった．

4. GCN と MLP の比較

4.1 MLP による推定精度

グラフを使わない MLP との比較実験を行った．33 個のランドマークの 3 次元情報を 99 次元のベクトルとして入力した．パラメータを変えて実験したところ，中間層が 3 層，各層のノード数が 128 個のときに推定誤差が最小で 0.226m となった．これは GCN とほぼ同じ誤差であり，推定精度に有意な差はみられなかった．

表 1: 同日モデルおよび別日モデルでの推定誤差

		学習モデル	
		同日	別日
評価サンプル	1 日目	0.163m	0.237m
	2 日目	0.178m	0.226m
	平均	0.170m	0.232m

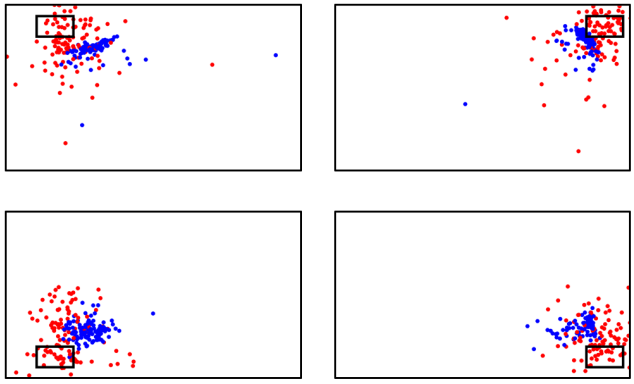

図 6: 四隅の推定指示位置（赤：GCN，青：MLP）

表 2: 領域ごとの平均推定誤差

	左上	右上	左下	右下
GCN	0.219m	0.154m	0.214m	0.181m
MLP	0.294m	0.156m	0.266m	0.242m

4.2 領域ごとの推定精度

スクリーンの四隅について評価を行った．6304 サンプルのうち，図 6 の四隅の黒枠内にポインタを表示したものは，左上 110，右上 117，左下 114，右下 92 サンプルあった．赤の丸は GCN で，青の丸は MLP で推定した指示位置である．MLP の推定は全体的に中央に寄る傾向があるということが分かった．また，図 6 の領域ごとの平均推定誤差を表 2 に示す．スクリーン全体では MLP と GCN に有意な差はなかったが，スクリーンの端では，四隅の黒枠内に推定できているものが多くあったことから，推定誤差が短く，GCN の方がよいと考えた．

5. まとめ

本稿では，指差し動作の姿勢グラフからスクリーン上のポインタ指示位置を推定する研究を行った．GCN を用いた提案手法では，スクリーンから 2.0〜3.5m 離れたとき，0.232m の誤差で推定できることが分かった．MLP と GCN を比較すると，MLP は全体的に中央寄りになるため，端の方においては GCN の方がよいことが分かった．

謝辞 本研究は JSPS 科研費 21K11998 の助成を受けて行った．

参考文献

[1] 山本龍平, 河合千春, 矢野良和, “機械学習による指差しの指示位置推定検証と内挿表現評価,” 第 33 回ファジィシステムシンポジウム, 2017.

[2] Google for Developers, “MediaPipe” https://developers.google.com/mediapipe, 2024/2/8 閲覧.