J-030

GCN-LSTM を用いた指差し指示位置の推定

Target Point Estimation for Pointing Gesture Using GCN-LSTM

中川 莉那 † **中井 満** †

Rina Nakagawa Mitsuru Nakai

1. はじめに

人同士のコミュニケーションにおいて，ジェスチャは多様に使われており，特に指差し動作は日常的な場面で頻繁に用いられる．本研究では、この指差し動作を応用し，指差しで指示した軌跡でロボットを誘導することを目的とする。そのためには指差し指示位置の推定が必要となる．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．我々は，3 次元空間の座標情報を用いて，2 次元の指差し指示位置を推定する手法を提案する．既報では，静止した点を指差す静止画像から推定した [2]．本報では，指示位置を動かすことを目的とする．そのため，Graph Convolutional Networks(GCN) および Long Short-Term Memory(LSTM) を組み合わせた GCN-LSTM を用いて，動画像から指示位置の軌跡を推定する手法を提案する．

2. システムの構成

構成を図 1 に示す．ユーザはスクリーンを指差して，指示位置を動かす．RGB カメラよりユーザの身体特徴点を推定し，同時に Depth カメラから深度情報を取得する．これらを統合して，特徴点の 3 次元位置座標および移動軌跡を得る．特徴点の接続関係を表した姿勢グラフと特徴点の移動軌跡を GCN-LSTM の入力とし，指差しで指示したスクリーン上の位置座標の軌跡 (x_t, y_t) を推定する．

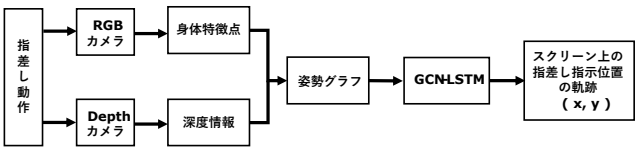

図 1: 指差し指示位置推定システムの構成

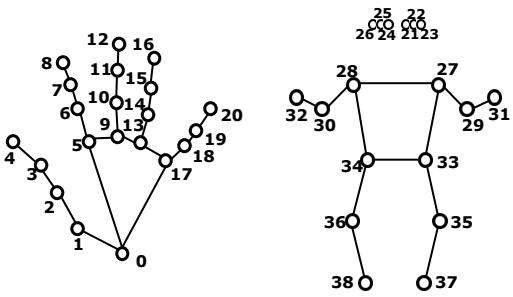

図 2: 利用する右手と全身のランドマーク

2.1 姿勢グラフの作成

Google が提供している MediaPipe を利用して RGB 画像中の手および全身の身体特徴点（ランドマーク）を検出する．利用するランドマークは図 2 のとおり，右手 21 個・目 6 個・全身のうちの 12 個，合計 39 個である．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身体特徴点として接続し，グラフ構造とすることで姿勢グラフとなる．

2.2 GCN-LSTM による指差し指示位置推定

姿勢グラフから指差し指示位置座標の移動軌跡を推定する GCN-LSTM を図 4 に示す．2.1 節の姿勢グラフの各ノード i は 3 次元のワールド座標の時系列情報 $(x_{it}, y_{it}, z_{it})(t = 1, 2, 3...)$ を持ち，これを GCN の入力情報とする．特徴量を 3 次元から 512 次元に拡張し，2 つの GCN に通し，最大値プーリングによって各次元の最大値をとる．その後，2 つの LSTM 層に入力し，最終的に 2 つの全結合層を通して，時刻 t 毎に 2 次元の指差し指示位置の座標を出力する．

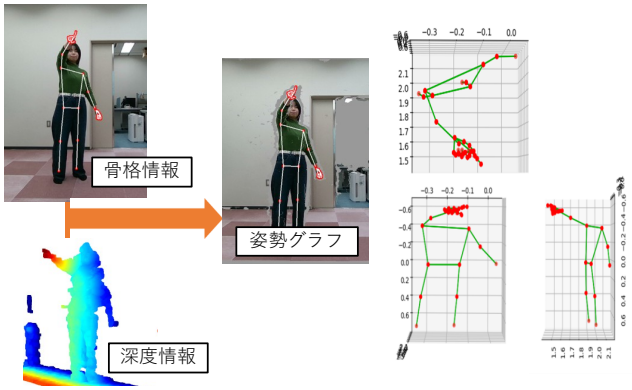

図 3: 姿勢グラフの作成 [2]

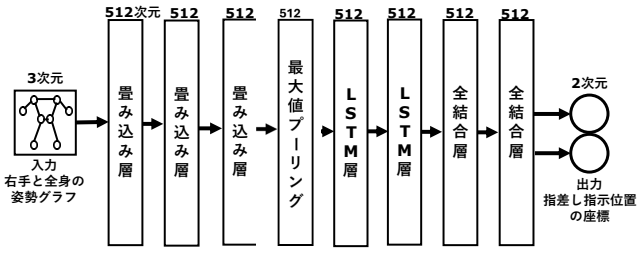

図 4: GCN-LSTM の構成

†富山県立大学, Toyama Prefectural University

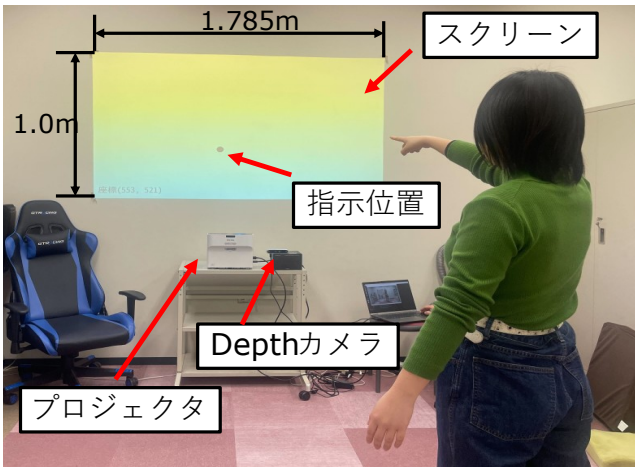

図 5: データ収集環境

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（指示位置）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 DepthCamera D455）はスクリーンの下に設置した．スクリーンから 2m 離れた位置に，特徴点が隠れないよう，カメラの斜め前方で立ってもらった．ポインタの始点と終点を 2 点間の距離が 0.6~1.2m の範囲になるようランダムに設定した．始点で 0.5 秒間静止した後，3 秒間で等速に終点まで移動し，終点で 0.5 秒間静止した．被験者はこの 4 秒間ポインタを指差して追従した．なお，データ収集中は指示位置の推定を行わないので，被験者は推定指示位置を見て指差し動作を修正することはできない．動画は 15fps で撮影し，1 サンプルは 60 フレームとなる．これを大学生 1 名から 50 サンプル収集した．

3.2 GCN-LSTM による推定精度

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．収集した 50 サンプルのうち，40 サンプルを学習に，10 サンプルを評価に用いた．LSTM 層の有効性を調べるため，図 4 の構成から LSTM 層ありのモデルとなしのモデルを作成して比較実験を行った．LSTM 層がない場合はフレームごとに独立に推定することと等価である．それぞれのモデルについて最大 1000 エポックで学習し，学習時の損失関数が最も低くなったときのモデルを評価に用いた．

評価には 3 つの指標を用いる．まず，フレーム毎に目標指示位置と推定指示位置の誤差を求めた．これを点座標誤差とする．2 つ目，目標指示位置の移動に対して指差しの追従が遅れたり進みすぎたりすることが考えられる．そこで，動的時間軸伸縮 (DTW) により，目標移動軌跡と推定移動軌跡の誤差が最小になる対応点から平均誤差を求めた．これを移動軌跡誤差とする．3 つ目，被験者は推定指示位置を見ることができないので，指差し動作を修正することはできない．そこで移動方向と移動量が正しいかを調べるため，それぞれの移動軌跡から平均値を減算し，相対座標において移動軌跡誤差を測る．これらの方法で求めた誤差を表 1 に示す．表 1 より，絶対座標では LSTM 層の優位性は見られないが，相対座標では LSTM 層がある方が推定誤差を小さくできることが分かった．

表 1: 推定誤差の平均

	点座標誤差	移動軌跡誤差	
		絶対座標	相対座標
LSTM なし	16.7cm	9.5cm	6.7cm
LSTM あり	17.4cm	10.7cm	6.0cm

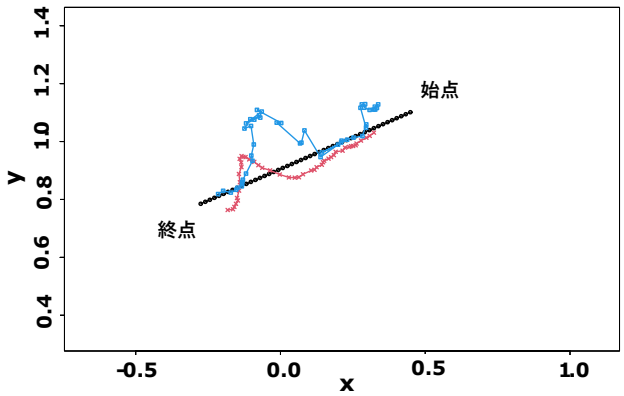

図 6: 指示位置の軌跡（黒）および推定指示位置の軌跡（青：LSTM なし　赤：LSTM あり）

評価に用いた 10 サンプルのうち，1 サンプルの目標移動軌跡および推定移動軌跡を図 6 に示す．黒色（○）は目標移動軌跡，青色（□）は LSTM 層なしの推定移動軌跡，赤色（×）は LSTM 層ありの推定移動軌跡である．LSTM 層がない場合，推定指示位置が大きく移動することや逆行することがあるが，LSTM 層があることによって，滑らかな移動軌跡を推定できていることが分かる．このため，あるフレームで推定を大きく誤ると，しばらく誤り続けるので，絶対座標による誤差は大きくなる．一方，移動方向や移動量が大きく変わることがないので相対座標による誤差は小さくなったと考えられる．したがって，指差しの軌跡で誘導するという用途では GCN-LSTM が有効であると考えられる．

4. まとめ

GCN-LSTM を用いて指差し動作の姿勢グラフからスクリーン上の動く指示位置を推定した．LSTM 層の有無で比較した場合，ないときの方が推定誤差が小さくなるが，相対的な移動軌跡を見た場合，LSTM 層が有効であることが分かった．今後は移動中の推定結果をユーザにフィードバックすることで，ユーザの指差し動作と推定誤差がどのように変わるかを調査する予定である．

謝辞　本研究は JSPS 科研費 24K15050 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] 中川莉那, 中井満, “GCN を用いた指差し指示位置の推定,” 第 23 回情報科学技術フォーラム, FIT2024.