

## CNN-LSTM を用いた非接触型呼吸推定手法の開発\*

小尾 賢生<sup>†1</sup> 船越 孝太郎<sup>‡2</sup><sup>1</sup> 東京工業大学 工学院<sup>§</sup><sup>2</sup> 東京工業大学 科学技術創成研究院<sup>¶</sup>

## 1 はじめに

非接触による呼吸推定は、測定者への特別な器具装着を必要とせずに測定者の呼吸状態を推定する問題である。一般的に呼吸推定(計測)は測定者の健康状態評価を目的として行われ、非接触での推定は測定に伴う器具装着の不快感を軽減させる。我々は、非接触呼吸推定のマルチモーダル対話システムへの応用を目指している。本稿では、1名の被験者の動画像を入力として、3次元畳み込みニューラルネットワーク(3DCNN)およびConvolutional Long Short-Term Memory(ConvLSTM)を用いた推定手法を適用した結果を報告する。

## 2 関連研究

非接触での呼吸推定としては従来、呼吸に伴う気流変化の推定[4]、遠隔フォトプレステモグラフィ[7]、胸腹部における呼吸運動の推定[2, 6]などを用いた手法などが研究されてきたが、測定に特殊な機材が必要である、測定開始時に関心領域の選択が必要となるといった問題がある。これらの問題に対処するべく、安価なカメラを用いたオプティカルフローによる推定も行われた[3]が、測定者の身体動作により精度の低下が生じた。

呼吸状態の対話への応用については、呼吸状態と発話の関係性[5]から、接触式センサにて取得した呼吸状態による次話者予測の研究が行われている[8, 9]。

## 3 呼吸データセット

推定手法の予備的評価を行うためのデータセットを作成した。データの収録は次のように行い、収録後に呼吸波形データに対する平滑化と正規化、補間によるRGB動画像取得時刻の波形値取得を行った。撮影画像の例を図1に示す。

**RGB 動画像:** Kinect v2 を用いて解像度 1980 × 1080, 30fps にて収録。mp4 形式にて保存。フレーム毎の収録時刻を記録。



図 1: RGB 画像例

**呼吸波形:** 胸部へのセンサベルト装着により、呼吸運動に伴う胸部の拡張収縮を測定。csv 形式にて保存。測定値毎の収録時刻を記録。

RGB 動画像と呼吸波形は同一の PC で記録したため、両データの時刻情報は同期されている。著者1名について、安静静止状態、静止発話状態に分けて収録した。

**安静静止状態:** 身体を動かさずに静止着座状態で収録。

**静止発話状態:** 身体を動かさずにスマートスピーカーとの会話をし、定期的に発話をする状態で収録。

## 4 推定手法

RGB 連続画像を入力とする 3DCNN および 3DCNN-ConvLSTM を用いた呼吸状態推定モデル構築し、連続画像の末尾時点における呼吸波形値の推定を行った。Almina らにより提案された振動監視モデル[1]を基に、入出力に相当となるように 3DCNN/ConvLSTM モデルのカーネルやストライドのサイズ変更、FC レイヤーの変更を行った。

なお、入力として用いた RGB 画像は解像度 320 × 180 へトリサイズ後、10 フレーム分の連続画像をスタックし、3次元データとした上で用いている。

## 5 実験

## 5.1 実験設定

本実験では作成した呼吸データセットにおける安静静止状態、静止発話状態について、25 分程度分の連続収録データを用いた。各データの先頭 1 割程度をテストデータとして予め分割し、残りを訓練データとし

\*Non-contact Respiration Monitoring Using CNN-LSTM

†Takao Obi

‡Kotaro Funakoshi

§Tokyo Tech, School of Engineering

¶Tokyo Tech, Institute of Innovative Research

手法	安静静止状態	静止発話状態
オプティカルフロー	0.603 (0.260)	-0.591 (0.381)
3DCNN	0.617 (0.255)	0.109 (0.271)
3DCNN-ConvLSTM	<b>0.722 (0.211)</b>	<b>0.697 (0.170)</b>

表 1: 各手法における  $R^2$  および MAE (括弧内)

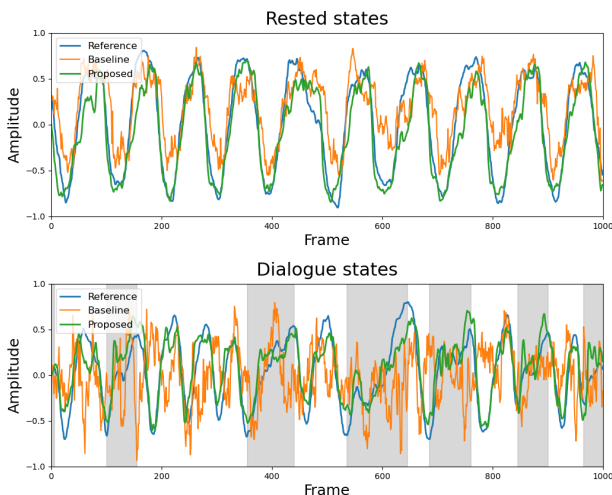


図 2: 安静静止/静止発話における推定結果

て学習に用いた。学習の最適化には確率的勾配降下法 (SGD) を用いた。Learning rate は 3DCNN モデルでは 0.001, 3DCNN-ConvLSTM モデルでは 0.01 とした。Early Stopping を用いて学習を行い, patience を 5 とした。

比較のため, オプティカルフローによる推定手法 [3] を基にベースライン手法を実装した。ベースライン手法では処理速度の向上を目的として一定間隔の画素ごとに処理を行うように変更を行っている。また, 変動量の基準となるベクトルの取得において最大ノルムとなるベクトルを使用している。安静静止状態の収録データにて, 各変更に伴う処理速度/推定精度の向上を確認している。

## 5.2 実験結果

各手法における安静静止状態, 静止発話状態での決定係数 ( $R^2$ ) と平均絶対誤差を表 1 に示す。  $R^2$  は 1 に近づくほど, 平均絶対誤差は 0 に近づくほど, 測定値と推定値が一致していることを示す。また, 測定により得られた呼吸波形 (Reference) とベースライン手法 (Baseline) および 3DCNN-ConvLSTM (Proposed) による推定呼吸波形の例を図 2 に示す。図 2 下では発話区間に背景の色付けを行っている。

## 6 考察

表 1, 図 2 より, 非接触による呼吸波形値の推定に 3DCNN および 3DCNN-ConvLSTM によるモデルを用いることでベースラインより精度の高い推定ができることを確認できた。また, 3DCNN モデルに ConvLSTM モデルを加えることで測定者の発話動作による影響が大きく軽減された。被験者 1 名のデータによる予備評価のため, 十分な評価が行われたとは言えないが, 軽微な動作を含む環境での RGB カメラを用いた呼吸状態の測定に期待できる。なお, RGB 画像を入力とした推定のため, 背景色と測定者の服装の色の類似による精度への影響が存在する。

## 7 まとめ

本稿では, RGB 連続画像を用いた 3DCNN と ConvLSTM による非接触による呼吸状態推定の予備的な検討を行った。今後は複数の測定者を対象としたデータセットの作成を行いつつ, より頑健で精度の高い推定手法の検討を行う。

## 参考文献

- [1] H. Liu A. Alimasi and C. Lyu. Low frequency vibration visual monitoring system based on multi-modal 3DCNN-ConvLSTM. *Sensors*, Vol. 20, No. 20, 2020.
- [2] H. Aoki, H. Nakamura, K. Fumoto, K. Nakahara, and M. Teraoka. Basic study on non-contact respiration measurement during exercise tolerance test by using kinect sensor. In *IEEE/SICE International Symposium on System Integration*, 2015.
- [3] A. Chatterjee, A. P. Prathosh, and P. Praveena. Real-time respiration rate measurement from thoracoabdominal movement with a consumer grade camera. In *IEEE EMBC*, 2016.
- [4] Y. Cho, N. B. Berthouze, and S. J. Julier. Deepbreath: Deep learning of breathing patterns for automatic stress recognition using low-cost thermal imaging in unconstrained settings. In *Seventh International Conference on Affective Computing and Intelligent Interaction*, 2017.
- [5] B. Conrad and P. Schönle. Speech and respiration. *Archiv für Psychiatrie und Nervenkrankheiten*, Vol. 226, No. 4, pp. 251–268, 1979.
- [6] L. Ge, J. Zhang, and J. Wei. Single-frequency ultrasound-based respiration rate estimation with smartphones. *Computational and Mathematical Methods in Medicine*, pp. 1–8.
- [7] M. Ghodratioghar, H. Ghanadian, and H. A. Osman. A remote respiration rate measurement method for non-stationary subjects using CEEMDAN and machine learning. *IEEE Sensors Journal*, Vol. 20, No. 3, 2020.
- [8] R. Ishii, K. Otsuka, S. Kumano, and J. Yamato. Using respiration to predict who will speak next and when in multiparty meetings. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, Vol. 6, No. 20, pp. 1–20, 2016.
- [9] M. Włodarczyk, K. Laskowski, M. Heldner, and K. Aare. Improving prediction of speech activity using multi-participant respiratory state. *Proc. Interspeech 2017*, pp. 1666–1670.