2020 19th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)

# Poster Abstract: A Weakly Supervised Tracking of Hand Hygiene Technique

Abdelwahed Khamis
UNSW, Australia
Data61, CSIRO, Australia
a.khamiss@unsw.edu.au

Branislav Kusy
Data61, CSIRO, Australia
brano.kusy@data61.csiro.au

Chun Tung Chou
UNSW, Australia
c.t.chou@unsw.edu.au

Marylouise McLaws
UNSW, Australia
m.mclaws@unsw.edu.au

Wen Hu
UNSW, Australia
wen.hu@unsw.edu.au

## ABSTRACT

Each year, hundreds of thousands of people contract Healthcare Associated Infections (HAI). Poor hand hygiene compliance among healthcare workers is thought to be the leading cause of HAIs and methods were developed to measure compliance. Surprisingly, human observation is still considered the gold standard for measuring compliance by World Health Organization (WHO). Moreover, no automated solutions exist for monitoring hand hygiene techniques, such as "how to hand rub" technique by WHO. In this work, we introduce RFWash; the first radio-based device-free system for monitoring Hand Hygiene (HH) technique. On the technical level, HH gestures are performed back-to-back in a continuous sequence and pose a significant challenge to conventional two-stage gesture detection and recognition approaches. We propose a deep model that can be trained on unsegmented naturally-performed HH gesture sequences. RFWash evaluation demonstrates promising results for tracking HH gestures, achieving gesture error rate of $< 8\%$ when trained on 10-second segments, which reduces manual labelling overhead by $\approx 67\%$ compared to fully supervised approach. The work is a step towards practical RF sensing that can reliably operate inside future healthcare facilities.

## KEYWORDS

device-free sensing, deep learning, sequence-to-sequence

## 1 INTRODUCTION

Healthcare Associated Infections (HAIs) find their way to one in twenty five patients admitted to hospitals [2] and continue to lead to increased patient mortality and healthcare cost [2]. Proper hand hygiene protocol, i.e. frequent and thorough hand cleaning, is an effective way to combat HAIs. This leads to the question of how one can monitor hand hygiene (HH) adherence in an hospital environment. nThe conventional approach for HH adherence monitoring is to employ a team of observers (e.g., overt nurse trained auditors) to record Hand Hygiene Opportunities (HHOs) and the number of times health care workers (HCWs) comply with the protocol. Today, this is considered to be the gold standard for measuring compliance by the World Health Organization (WHO).

Attempts to implement automated alternatives for monitoring HH had a limited success so far. For example, electronic counters [5] and RFID [6] simply count hand washing activities. These tools

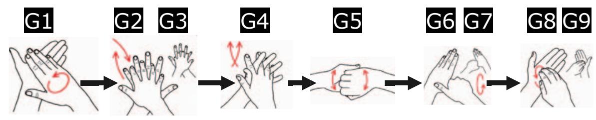


**Figure 1: Alcohol-based handrub procedure recommended by the WHO. The 9 steps are marked by labels G1, G2 etc.**

cannot reveal whether hand hygiene technique — such as the nine-step procedure for applying alcohol-based handrub recommenced by WHO [1], see Figure 1 — has been thoroughly adhered to. Although there are commercial camera systems for training HCWs to learn the correct HH technique, to the best of our knowledge, there exists no solution for automated monitoring of the HH technique in healthcare facilities.

Our work proposes to utilize commercial-off-the-shelf mmWave sensors to monitor the HH technique in Fig. 1. Our vision is to embed these sensors at the alcohol-based handrub dispensers, which are distributed throughout the hospitals, to monitor whether HCWs have adhered to HH technique. Our vision will therefore enable much more fine-grained monitoring of HH adherence.

The main *challenge* in HH tracking is that the entire procedure is performed without a pause between consecutive gestures. Contiguous sequences of gestures have not been investigated in RF sensing literature before. In fact, previous RF-based sensing approaches [4] rely on pauses between gestures, which are employed as physical markers identifying the start and end of each motion segment. This approach trivially achieves accurate segmentation and the problem reduces to gesture classification. *Without enforcing the pauses, joint segmentation and classification becomes a challenging task.*

## 2 PROPOSED MODEL

We address the problem of tracking back-to-back gestures by introducing RFWash; a segmentation-free approach for recognizing HH gestures sequence. We draw inspiration from modern end-to-end speech recognition systems, which are similar to our problem because it is difficult to label continuous speech data. Of particular relevance to our problem are weakly supervised methods that can learn directly form data without requiring explicit data segmentation and full annotation. To this end, we develop a model that can be trained on back-to-back gesture sequences (Figure 2) without requiring gesture segmentation, which can also reduce

978-1-7281-5497-8/20/$31.00 ©2020 IEEE
DOI 10.1109/IPSN48710.2020.00-21

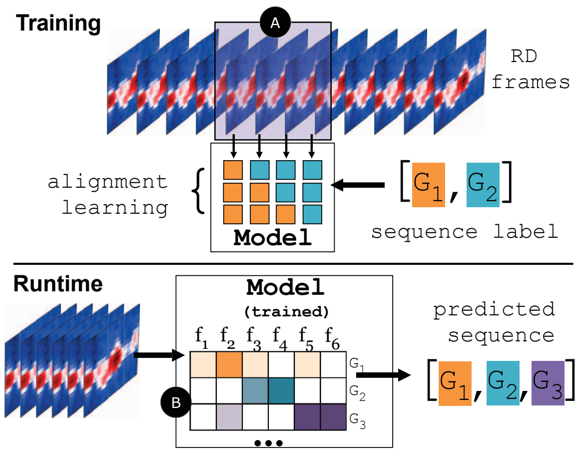


**Figure 2: RFWash is trained on continuous RF samples (A) of HH gestures and corresponding sequence labels. The model automatically learns which frames correspond to individual gestures (e.g. $G_1$ vs $G_2$) through "alignment learning". In runtime, per-frame gesture predictions (B) are produced and used to estimate the most likely gesture sequence .**

labelling overhead substantially compared to framewise labelling. The propsoed deep model consists of Spatiotemporal layers (CNN) followed by Bidirectional LSTM (BiLSTM) layers. Added to this is Connectionist Temporal Classification [3] which is employed to for aligning the gesture labels (i.e. $G_1 \rightarrow G_2 \rightarrow G_3$ ) with the corresponding RF samples. In this way, continuous input stream can be processed, without segmentation, by the model during training and run time.

**Data Augmentation:** With longer training sequence lengths, the labelling effort is reduced but fewer training samples becomes available and poor temporal alignment becomes a big issue. We employ "order preserving" concatenation of existing samples/labels to counter this. Despite the simplicity, the method is very effective and allowed the model generalize very well to unseen gesture sequence lengths as our experiments show.

## 3 EVALUATION

As the output of the mpodel is the complete gesture sequence, we report the error in terms of Gesture Error Rate (GER). GER is defined as the minimum number of gesture insertions, substitutions, and deletions needed to transform the predicted gesture sequence into the ground truth gesture sequence, divided by the number of gestures in the ground truth.

As shown in table 1, our algorithm maintains consistent performance regardless of the input length. For input sequence containing 6 gestures, about 0.45 substitutions, deletions or insertions are required to make the prediction match the ground truth. On average, 75 percent of the sequences predicted by the model are exactly matching the ground truth. Moreover, the proposed data augmentation technique allows the model to maintain superior exact match rate on sequences with length unseen during training. As shown in Fig. 3 (the model was originally trained on segments with length ≈ 6 seconds).

**Table 1: HH Sequence Recognition Accuracy**

| Segment | mean GER | median GER | Exact Match |
|---|---|---|---|
| 1s (2 gestures) | 16% | 0% | 75% |
| 5s (4 gestures) | 11% | 0% | 75% |
| 10s (6 gestures) | 7.41% | 0% | 76% |

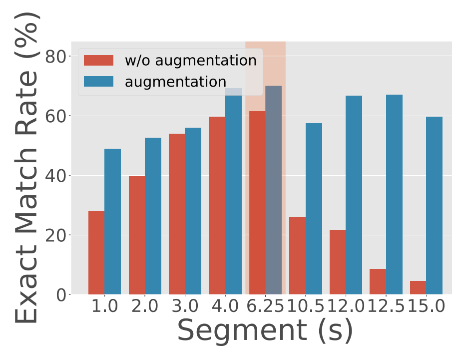


**Figure 3: Augmentation Impact on Unseen Sequence Length**

## 4 CONCLUSION AND FUTURE WORK

We introduced the first RF-based system for contact-free monitoring of healthcare workers performing Hand Hygiene techniques. Experiments show that the proposed model along with data augmentation technique achieve very accurate tracking and encourages us to further expand this work by collecting data at a larger scale in clinical facilities.

## REFERENCES


[1] 2009. WHO Guidelines on Hand Hygiene in Health Care. Published by World Health Organisation. Retrieved from: whqlibdoc.who.int/publications/009.pdf.

[2] 2016. Healthcare-associated infections: data and statistics. Published by Center for Disease Control and Prevention.

[3] Alex Graves, Santiago Fernández, Faustino Gomez, and Jürgen Schmidhuber. 2006. Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks. In *Proceedings of the 23rd international conference on Machine learning*. ACM, 369–376.

[4] Yongsen Ma, Gang Zhou, Shuangquan Wang, Hongyang Zhao, and Woosub Jung. 2018. SignFi: Sign language recognition using WiFi. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 23.

[5] AR Marra and MB Edmond. 2014. New technologies to monitor healthcare worker hand hygiene. *Clinical Microbiology and Infection* 20, 1 (2014), 29–33.

[6] Lisa L Pineles, Daniel J Morgan, Heather M Limper, Stephen G Weber, Kerri A Thom, Eli N Perencevich, Anthony D Harris, and Emily Landon. 2014. Accuracy of a radiofrequency identification (RFID) badge system to monitor hand hygiene behavior during routine clinical activities. *American journal of infection control* 42, 2 (2014), 144–147.