Using Thermal Imaging to Measure Hand Hygiene Quality

Chaofan Wang1,.*, Weiwei Jiang2, Kangning Yang3, Zhanna Sarsenbayeva4, Benjamin Tag5, Tilman Dingler3, Jorge Goncalves3, and Vassilis Kostakos3

1Faculty of Industrial Design Engineering, Delft University of Technology, Delft, Netherlands
2Department of Computer Science and Technology, Anhui Normal University, Wuhu, China
3School of Computing and Information Systems, The University of Melbourne, Melbourne, Australia
4School of Computer Science, University of Sydney, Sydney, Australia
5Faculty of Information Technology, Monash University, Melbourne, Australia

*Corresponding author: Chaofan Wang
Email: C.Wang-16@tudelft.nl
Present address: Faculty of Industrial Design Engineering, Delft University of Technology, Delft, Netherlands

Keywords: thermal camera; ultraviolet light; hand hygiene; world health organization

Word count: 3382
ABSTRACT

Objectives: Hand hygiene has long been promoted as the most effective way to prevent the transmission of infection. However, due to the low compliance and quality of hand hygiene reported in previous studies, constant monitoring of healthcare workers’ hand hygiene compliance and quality is crucial. In this study, we investigate the feasibility of using a thermal camera together with an RGB camera to detect hand coverage of alcohol-based formulation, thereby monitoring handrub quality.

Methods: We recruited 32 participants for this study. Participants were required to perform four types of handrubs to produce different hand coverage of alcohol-based formulation. After each task, participants’ hands were photographed under a thermal camera and an RGB camera, while an ultraviolet (UV) test was used to provide the ground truth of hand coverage of alcohol-based formulation. Then, a U-Net was used to segment areas covered with alcohol-based formulations from thermal images, and the system performance was evaluated by comparing coverage differences between thermal images and UV images regarding accuracy and Dice coefficient.

Results: The system yields promising results in terms of accuracy (93.5%) and Dice coefficient (87.1%) when observations take place 10 seconds after performing handrub. In addition, we also examine the system performance change over a 60-second observation period, and the accuracy and Dice coefficient still remain at about 92.4% and 85.7% when observation happens at the 60-second time point.

Conclusions: Given these encouraging results, thermal imaging shows its potential feasibility in providing accurate, constant, and systematic hand hygiene quality monitoring.
1. INTRODUCTION

Hand disinfection — a major component of appropriate hand hygiene — is the most effective way to prevent Healthcare-Associated Infections (HAIs) and reduce their transmissions [1–3]. HAIs are one of the most crucial patient-safety challenges in healthcare settings [4]. They dramatically increase patients’ length of stay, costs, morbidity, and mortality [5, 6]. In 2009, the World Health Organization (WHO) issued the “WHO guidelines on Hand Hygiene in Health Care” providing a thorough review of evidence on hand hygiene in healthcare and specific recommendations to improve practices in healthcare settings [2]. The guidelines recommend two standard hand hygiene procedures, including “Handwash with Soap and Water” for visibly soiled hands and “Handrub with Alcohol-based Formulation” for routine decontamination of hands for all clinical indications [2].

However, research has found that hand hygiene quality in healthcare settings is generally unsatisfactory [7–9]. Szilágyi et al. reported that only 72% Healthcare Workers (HCWs) could adequately clean all hand surfaces immediately after hand hygiene training [9]. Since handrub with an alcohol-based formulation has been widely adopted into routine clinical practices [2], precisely measuring handrub quality and providing HCWs with feedback regarding their performance are essential to promote good hand hygiene in healthcare environments.

Yet, measuring handrub quality in healthcare settings is challenging [10]. Direct observation by trained auditors is considered the gold standard for monitoring both hand
hygiene compliance and quality, but its rigor is limited by personnel time and expense, insufficient sample size, and the Hawthorne effect [11]. Recently, researchers measured hand hygiene quality by tracking HCWs’ compliance with the WHO 6-step hand hygiene procedures through environmental and wearable sensors [7, 12–17]; however, their techniques were proxy measures to detect HCWs’ hand hygiene quality. Nevertheless, researchers attempted to quantify the hand coverage during handrub and handwash procedures [18, 19]. Ultraviolet (UV) tests have been widely used for medical hand hygiene training to highlight insufficiently cleaned regions after HCWs handrubbing. However, since UV tests require fluorescent dye, which often leaves residual which affects follow-up measurements, and UV lamps, it cannot be easily incorporated into HCWs’ daily routines.

Bernard et al. [20] discovered that surface temperature measurements can be affected by the presence of topically applied substances, which can be detected using infrared thermal imaging. Alcohol-based formulations typically consist of 60%-80% ethanol, which evaporates at room temperature and cools hand surfaces. As a result, several studies proposed to use of thermal imaging to assess hand hygiene quality. Boyce and Martinello [21] observed significant decreases in mid-palm, finger, and thumb temperatures after participants performed hand hygiene through thermal imaging. Similarly, Smieschek et al. [22] utilized a thermal camera to capture images of subjects’ hands, dividing each hand into 20 segments. By comparing the temperature differences between corresponding segments, their system estimated the segment-level surface coverage of alcohol-based formulations. However, a precise coverage comparison between thermal imaging and UV tests or microbiological tests has not yet been reported in the literature. Additionally, it remains
unclear within what time window thermal imaging can accurately detect the surface coverage of alcohol-based formulations, as hands initially cool down after a handrub and then gradually warm up again.

2. METHODS

2.1. study design and participants

We recruited 32 participants through our institution’s mailing lists and snowball recruitment with an equal number of women and men. All participants were students or staff at our institution, and their ages ranged between 18 and 27 (\( M = 22.8, \text{SD} = 2.1 \text{ years} \)). The majority of participants (27 out of 32, 84.4%) had not received formal hand hygiene training in the last three years and were not familiar with the formal hand hygiene procedures. The entire experimental procedure lasted for about 130 minutes, including briefing and debriefing. This study was approved by the anonymous institution’s Human Ethics Advisory Group.

Before the experiment, participants completed a questionnaire about allergic reactions to alcohol-based formulation and UV light. Upon arrival at our lab, we briefed participants on the study purpose and obtained their written consent agreeing to participate in our experiment. We subsequently provided training to our participants on how to perform the WHO 6-step handrub procedures. To achieve this, we first explained the procedure steps. They then watched an instructional video provided by the WHO [23] three times, and performed the handrub procedure along with the instructional video for training.
After the training session, participants proceeded to complete the experimental tasks. Before each task, participants rinsed their hands to remove residual fluorescent dye and dried their hands with tissues. Then, they rubbed their hands with hand warmers for one minute to rewarm their hands back to their body temperature [24]. We took both thermal and RGB images as baseline images before depositing alcohol-based formulation on participants’ hands. After that, participants performed the experimental tasks and placed their hands on the pegboards for 60 seconds while they were observed by the thermal camera. Finally, the UV lamp was turned on above the participants’ hands, and a photograph was taken using the RGB camera. After each participant, the devices used in the experiment will be sanitized for hygiene reasons. More details of the experiment flow are shown in Figure 1.

We designed 30 different tasks in line with four task types mentioned in Appendix Section B for both sides of participants’ hands, as well as the task descriptions for each task and the visualized surface coverage examples. However, since participants are required to keep their hands on the pegboards during the 60-second observation period, this would extend the experiment duration up to 180 minutes in total, and potentially cause participant fatigue. Therefore, we reduce the experiment duration by only observing one side of each participant’s hands (either palmar side or dorsal side) for the tasks of separated WHO handrub steps. As a result, each participant completed 21 out of the 30 possible tasks. Specifically, tasks 5-13 (individual handrub steps, dorsal) and 20-28 (individual handrub steps, palmar) were completed by half the cohort.
2.2. segmentation of RGB and thermal images

Here, we adopted a deep learning neural network – U-Net, one of the most widely-used biomedical image segmentation algorithms, to segment (1) hand areas from RGB images and (2) covered areas from thermal images.

We used the same network structure (shown in Appendix Figure 5) for both segmentation tasks. The inputs to the model for task 1 (segment hand areas from RGB images) are shown in Figure 2. In the meantime, the inputs of task 2 (segment areas covered by alcohol-based handrub from thermal images) will be combined with the baseline image, observation image, and their differences. The background may be noisy when adopting thermal imaging in healthcare settings; therefore, we further employ the segmented hand areas generated by task 1 to remove the background information from the inputs of task 2. More details are shown in Figure 2. To increase the size of the dataset, we flipped images horizontally for data augmentation.

We trained both models with Combo loss by combining Cross-Entropy loss and Dice loss [25]. We implemented and trained the models in Pytorch with a single Nvidia GeForce RTX 3080 super (12GB RAM). Both models were trained for 30 epochs and a batch size of 8. We resized the input images to $3 \times 483 \times 322$ pixels (16% of the original image). RMSprop optimization was used, with an initial learning rate of $10^{-5}$, weight decay of $10^{-8}$, and momentum of 0.9. We used the learning rate schedule on this basis: if the Dice coefficient on the validation set is not increased for two epochs, the learning rate will decay by a factor of 0.1.
2.3. statistical analysis

The accuracy and Dice coefficient of both segmentation tasks were measured to assess the performance of the proposed systems, and 5-fold cross-validation was used to check their generalizability. For task 1, we compared the U-Net segmented hand areas with manually segmented hand areas. For task 2, since previous work has shown that fluorescent dye highlights the areas of the hand surface that are adequately disinfected with acceptable accuracy (95% sensitivity and 98% specificity), we compared the coverage difference between results detected by UV images and thermal images. More details are mentioned in Appendix Section C.2 [26].

For each image (including both left and right hands), we calculated accuracy across all classes (task 1: hand areas and background, and task 2: hand areas covered or uncovered with alcohol-based formulation). In the meantime, we calculated the Dice coefficient for Region of Interest (ROI)s (task 1: hand areas, and task 2: hand areas covered with alcohol-based formulation). However, since hands in several tasks will be either fully covered or not covered by alcohol-based formulation (e.g., the task of Step 1 only; more details shown in Appendix Section B.2), 0% coverage will raise division by zero errors when calculating Dice coefficient. Therefore, we calculated \( Dice \ coefficient = \frac{2 \times True \ Positive \ (TP) + \epsilon}{2 \times TP + 2 \times False \ Positive \ (FP) + 2 \times False \ Negative \ (FN) + \epsilon} \), where \( \epsilon = 0.001 \) to prevent division by zero errors.
Due to heteroscedasticity in the data, we adopted the Welch Analysis of Variance (ANOVA) and Welch-Satterthwaite degrees of freedom [27]. Statistical analyses were conducted using Python (version 3.6.8) and statsmodels (version 0.9.0).

3. RESULTS

3.1. segmentation performance

We discarded the data of 3 participants (Participants 1, 13, and 22) because they did not precisely follow the study protocol (e.g., failed to wash out residual UV dye, or did not keep their hands still during the observation period). Considering the accuracy of the system, it is important to note that the timing of measurements matters. As the alcohol evaporates from participants’ hands, it causes a temporary temperature drop. This means that if the thermal observation happens a long time after the alcohol is applied, there may be no observable effect. Given these constraints, we report the accuracy and Dice coefficient at the 10-second observation time: the thermal imaging observation happens 10 seconds after participants place their hands on the pegboards, due to its highest accuracy. In the subsequent section, we report the effect of increasing this time window to up to 60 seconds.

For the model of task 1, the mean accuracy and Dice coefficients throughout 5-fold cross-validation are 99.6% (SD = 0.0003) and 97.2% (SD = 0.003), respectively. More details are shown in Figure 3a. Meanwhile, to validate the accuracy of using thermal imaging to detect the coverage of alcohol-based formulation, we summarize the performance (in terms of
accuracy and Dice coefficient) of thermal imaging across participants, hand sizes, and tasks (coverage ranges from 0% to 100%) to ensure the reliability and validity of the results. The system recognizes the hand areas covered by alcohol-based formulation with a mean accuracy of 93.5% (SD = 0.046) and a mean Dice coefficient of 87.1% (SD = 0.195) for all participants across all the experiments. We also group the accuracy for each participant for different tasks and present the results in Figure 3b. Of these, the highest mean accuracy is 96.0% for Participant 25, and the lowest mean accuracy is 84.9% for Participant 32. In the meantime, Participant 7 has the highest mean Dice coefficient (96.0%), and Participant 32 has the lowest mean Dice coefficient (64.0%).

Furthermore, we measure the effect of hand size on accuracy. Because we used a fixed amount of alcohol per participant task (3 ml as recommended by [28]), there will be less alcohol per unit area for participants with larger hands. Thus, we group participants in terms of hand length: 160 ≤ XS < 171 mm, 171 ≤ S < 182 mm, 182 ≤ M < 192 mm, and 192 ≤ L < 204 mm, and each participant’s mean accuracy is considered as one data point (shown in Figure 3c). A one-way ANOVA does not show a significant difference in system accuracy for different hand size groups ($F_{3,25} = 1.302, P = 0.295$) as well as in Dice coefficient ($F_{3,25} = 0.662, P = 0.583$).

We also measure the mean accuracy for each task by summarizing results across all participants (shown in Figure 3d). Of these, Task 30 has the highest mean accuracy of 96.3%, and Task 8 has the lowest mean accuracy of 88.7%. Furthermore, Task 30 has the
highest mean Dice coefficient of 97.7%, and Task 5 has the lowest mean Dice coefficient of 17.6%.

As seen in the aforementioned results, Dice coefficients show several dramatic drops for some participants and tasks even though their accuracy is still above 90%. This phenomenon may be associated with the small sizes of the areas that alcohol-based formulations cover. Therefore, we applied Spearman’s Correlation test to examine the correlation between Dice coefficients and sizes of the areas covered by alcohol-based formulations. For each participant, the mean percentage of areas covered with alcohol-based formulations across all tasks varies from 34.0% to 73.6%, and there exists a strong positive correlation with Dice coefficients (0.840, P < 0.001). For each task, the mean percentage of areas covered with alcohol-based formulations across all participants varies from 6.1% to 94.2%, and there is a strong positive correlation with Dice coefficients (0.967, P < 0.001) as well. These findings suggest that because of the limited resolution of thermal cameras and the imperfect alignment between baseline and observation images, thermal imaging may not be able to identify areas with small sizes and near edges.

3.2. effects of varying observation

In the experiment, participants were required to place their hands on the observation pegboards for 60 seconds after completing all 21 tasks. Throughout this 60-second observation period, we measured the effect of the gradual hand rewarming on the system performance. In this section, accuracy and Dice coefficient are calculated for each participant and then grouped by the observation time across all 30 participants. Due to the
computational requirements of the analysis, we analyze the thermal imaging results in every 5 seconds from 0 delay (images immediately captured after participants place their hands on the observation pegboards) up to a 60-second delay, as shown in Figure 4.

The maximum mean accuracy of 93.6% is displayed at 10 seconds, where all mean accuracy values are above 92% across the 60-second observation period. Meanwhile, all mean Dice coefficients are above 85% between 0 seconds and 60 seconds, with the highest mean Dice coefficient of 87.4% occurring at 35 seconds. More details are shown in Figure 4.

Across the 60-second observation period, accuracy and Dice coefficient gradually decrease slightly over time. Therefore, we evaluate the correlation between the observation time and system performance through Spearman’s correlation test. The correlation values between the observation time and accuracy and Dice coefficients are -0.275 (P = 0.026) and -0.034 (P = 0.785), respectively. Although it is recommended to collect thermal images as soon as possible because of the weakly negative correlation between observation time and system performance, thermal imaging continues to operate effectively during the 60-second observation period.

4. DISCUSSION

4.1. adopting thermal imagining in healthcare settings

This study primarily aimed to demonstrate the feasibility and functionality of the method, rather than its actual implementation in healthcare settings. In real-world deployments,
users would not need to remain motionless for 60 seconds on each side of their hands, as the evaluation could potentially be completed within a few hundred milliseconds. Furthermore, recognizing the impracticality of maintaining users’ hand positions between two observations, we propose three alternative approaches for adopting thermal imaging in healthcare settings to record baseline and observation images (illustrated in Figure 5).

The first approach is to provide a Graphical User Interface (GUI), which shows a wireframe RGB image with superimposed hand contours. The GUI would instruct HCWs to place their hands at the same position when taking baseline images and observation images after handrubbing (details shown in Figure 5 [29]). However, it may run into the issue of hand misalignment, which might result in classification mistakes and poorer system performance.

The second approach is to section hands into segments and then compare the temperature differences of the same segments between the baseline images and observation images (details shown in Figure 5 [22]). While HCWs are aware that a certain segment is exposed to alcohol-based formulation, they are unable to identify which parts within the segment are absent due to the lack of precise coverage information that comes with this solution.

The third approach is to avoid calculating the temperature differences for segments, but rather map hand segments to a standard hand drawing (details shown in Figure 5 [18]). The system first splits recognized hand regions and hand drawings into 18 segments based on the landmarks generated by MediaPipe and finger-web points [30]. Then, the segments
of the hand regions are matched and mapped to the corresponding segments of the hand drawings. After that, the system can calculate temperature differences without losing coverage information. As a result, HCWs can recognize missed areas, and the visual intervention could help them improve their hand hygiene. This approach may offer extensive information inside hand segments and is more resilient to hand misalignment than the other two aforementioned approaches.

4.2. Limitation

Since we needed to first validate thermal imaging in the context of measuring hand hygiene quality, we conducted the study in a controlled laboratory environment rather than a field study in healthcare settings. Nevertheless, we agree that when adopting thermal imaging in healthcare settings, more factors must be taken into account.

Firstly, the system uses thermal imaging to track temperature drops caused by alcohol-based formulations; hence, it is unable to evaluate the efficacy of handwashing with water and soap, because water might cause consistent temperature drops throughout all hand surfaces. Instead, to evaluate handwash quality, previous studies used sensors to track the HCWs’ adherence to WHO six-step hand hygiene techniques [13–17]. Moreover, the system’s performance may be affected by variations in hand hygiene techniques, such as changes in order and the inclusion of extra steps, which necessitates further investigation.

Secondly, since the system is based on thermal imaging, any temperature changes could affect the system’s performance. Of these, hand temperature is the most important and
diverse factor. In the experiment, we controlled participants’ hand temperature to start at around 36 degrees, using hand warmers to speed up the process of rewarming their hands between tasks. As HCWs’ hand temperatures can vary substantially in real-life scenarios, future studies need to investigate the effects between hand temperature and system performance.

Thirdly, we conducted the study in a laboratory setting, where the room temperature was controlled and did not vary substantially. However, the room temperature could affect the temperature of the alcohol-based formulation and its evaporation rate, impacting the system performance. Currently, our laboratory’s ambient temperature was set to a standard temperature (i.e., between 21 degrees to 24 degrees) defined by the “Guidelines for Construction and Equipment of Hospital and Medical Facilities” [31]. However, future studies are needed to examine its performance and applicability in real healthcare settings.

The fourth factor is the alcohol-based formulation itself. In the experiment, we mixed a colored antimicrobial hand gel (Microshield Angel Blue, Schülke & Mayr GmbH) with a fluorescent handrub formulation (Glitterbug Gel, OnSolution Pty Ltd), and both gels had been used in prior research or were implemented in healthcare settings [32–35]. Nevertheless, the correlation between the types of alcohol-based formulation and the system performance should be further explored. Meanwhile, different types of alcohol-based formulations may have different sterilization efficiency, which may result in insufficient disinfection of the areas covered by alcohol-based formulations. Thus, alcohol-
based formulations used in future studies should meet EN 1500 and ASTM E-1174 standards to ensure their sterilization efficiency [2].

5. CONCLUSIONS

In this paper, we showed the feasibility of using thermal imaging to detect hand coverage with alcohol-based formulation, thereby monitoring hand hygiene quality. In an evaluation with 32 participants, the system achieved promising results in terms of accuracy and Dice coefficient while being comparable to the gold standard for UV concentrate. Our study shows the potential flexibility of employing thermal imaging to monitor hand hygiene quality, which can be a step toward a continuous automated hand hygiene monitoring system that allows real-time monitoring without interrupting the HCWs’ daily routines.

CONFLICT OF INTEREST STATEMENT

None declared.

FUNDING

This work is partially funded by NHMRC grants 1170937 and 2004316, European Union’s Horizon 2020 research and innovation program (Grant agreement 957296), and PhD Write Up Award from Faculty of Engineering and Information Technology, the University of Melbourne.
REFERENCES


FIGURE LEGEND

Figure 1. Experiment flow.

Figure 2. Required inputs of the proposed models.

Figure 3. Classification accuracy and Dice coefficient of participants, hand sizes, and tasks.

Figure 4. System performance of accuracy and Dice coefficient across 60 seconds observation.

Figure 5. Alternative approaches for deploying thermal imaging in healthcare settings.

Approach (a) is based on [29], Approach (b) is based on [22], and Approach (c) is based on [18].

APPENDICES

Appendix A. HARDWARE SETUP

Appendix B. TASK DESIGN

Appendix C. DATA PREPROCESSING

Appendix D. MODEL STRUCTURE
(a) Task 1 – segmentation performance from RGB images

(b) Task 2 – segmentation performance across participants from thermal images

(c) Task 2 – segmentation performance across hand sizes from thermal images

(d) Task 2 – segmentation performance across tasks from thermal images
Rinse hands; dry hands with tissues;
rub hands to warm with hand warmers

Take baseline photos

Apply fluorescent handrub formulation by following tasks

2 Sides * 2 Shape combinations - (circle + hexagon) or (rectangle + star)
(use stamps)

2 Sides * 2 Equally Split combinations - (0% + 100%) or (50% top half + 50% lower half)
(use brush)

1 Side * 9 Separated WHO Handrub Steps - step 1 alone or step 1 with each other handrub steps
(participant follows WHO poster)

2 Sides * 2 Entire WHO Handrub Procedures (participant follows WHO poster)

Observed under a thermal camera (duration: 60 seconds)

Observed under a RGB camera without and with UV light from a 365 nm UV-A lamp

Finish 21 pre-defined tasks

Finish experiment