Vision-Based Hand Hygiene Monitoring in Hospitals

Serena Yeung [1], Alexandre Alahi [1], Zelun Luo [1], Boya Peng [1], Albert Haque [1], Amit Singh [1,2], Terry Platchek [1,2], Arnold Milstein [1], Li Fei-Fei [1]

[1]Stanford University, [2]Lucile Packard Children's Hospital Stanford

serena@cs.stanford.edu, alahi@stanford.edu, zelunluo@stanford.edu, boya@stanford.edu, ahaque@stanford.edu, AmSingh@stanfordchildrens.org, TPlatchek@stanfordchildrens.org, amilstein@stanford.edu, feifeili@cs.stanford.edu

Abstract

Recent progress in developing cost-effective depth sensors has enabled new AI-assisted solutions such as assisted driving vehicles and smart spaces. Machine learning techniques have been successfully applied on these depth signals to perceive meaningful information about human behavior. In this work, we propose to deploy depth sensors in hospital settings and use computer vision methods to enable AI-assisted care. We aim to reduce visually-identifiable human errors such as hand hygiene compliance, one of the leading causes of Health Care-Associated Infection (HCAI) in hospitals.

1 Introduction

In recent years, much progress has been made in machine learning-based interpretation of clinical data for decision support as well as knowledge discovery. These works have reasoned on data sources such as electronic medical records and radiographic images, among others. However, a valuable source of clinical data that remains underexplored is visual data capturing patient experience and environment during health care episodes such as hospital stays. Such data can contain rich information about patient condition such as the appearance of distress, which has been described as the 6th vital [2]. It can also capture details about the occurrence and nature of clinical activities ranging from patient care to bundle compliance and hand hygiene.

In this work, we focus on interpreting visual clinical data for the specific application of health care-associated infection (HCAI) prevention, and present initial experiments towards this end. Studies have shown that HCAIs present a challenging and costly problem for hospitals in the United States [7, 9]. Proper hand hygiene is known to play a very important role in preventing HCAI, yet the monitoring and sustainment of good hand hygiene practices remains challenging in hospital environments with high levels of activity and large, constantly shifting populations of health care workers. Hospitals have attempted to monitor hand hygiene compliance with covert hand-written audits done primarily by nursing staff. However this typically is only able to obtain a low sample size, has been insufficient to provide accurate assessment, and furthermore requires staff to take time away from their normal workflow.

We introduce an approach for monitoring hand hygiene compliance using machine learning-based interpretation of visual recording of the environment. Specifically, we propose to deploy depth sensors in hospitals to capture the physical space near hand hygiene dispensers, and use computer vision methods to detect when a person performs the hand hygiene action. This action is defined by a person placing his or her hand under a hand hygiene dispenser and receiving soap. We note that we use only the depth modality in order to collect privacy-safe signals. We show experiments using a convolutional neural network (CNN) over the dispenser region in the depth images, that is able to detect when a hand hygiene action occurs, and additionally using a human pose-based approach that is able to tie the action to a specific person in the image.

Figure 1: Examples of depth images from our dataset. From left to right, the first two are positive instances of hand hygiene, and the last two are challenging negative instances.

2 Related Work

CNNs have demonstrated state-of-the-art performance for a wide range of visual recognition tasks [5, 4, 10]. In our work we study their application in the context of recognizing hand hygiene actions, using both RGB and depth modalities. For the task of human action recognition, pose-based approaches have also been shown to work well [6, 12, 13, 3, 1]. While methods for robust full-body pose estimation in depth data have been introduced [11], they typically assume a side view of the person instead of the top view that provides more utility for hand hygiene monitoring. In our work we therefore learn body and hand detectors using background subtraction and a CNN, respectively, and explore pose-based reasoning on top of these locations.

3 Approach

CNN-based hand hygiene detection. We trained a CNN to detect whether hand hygiene occurs in a video frame, and compare using the full frame image as input versus a cropped region containing the dispenser. The network architecture consists of 2 convolutional layers, each followed by a max pooling layer, and 2 fully connected layers. The output is a binary classification of whether the hand hygiene action is occurring, and we optimize a logistic loss function using stochastic gradient descent.

Pose-based approach. We additionally explore a pose-based approach that provides additional information of which person is performing the action. We first segment and detect humans in each frame using a background subtraction-based method, and then detect the hand of each human using a CNN-based hand detector trained on a large hands dataset [8]. Hand hygiene is considered to be performed if a hand is detected in the physical space immediately under the hand hygiene dispenser.

4 Experiments

We collected a dataset of 20 hours of depth signals (Fig. 1), and show results for our CNN-based hand hygiene detector in Table 1. For the completeness of our research, we also illustrate the performance if RGB signals were used. For privacy reasons, these experiments were conducted in the lab. The CNN was able to achieve strong performance using a cropped region around the dispenser.

We additionally evaluated our pose-based approach on the RGB signals, and achieved an average precision of 0.807. While lower than the CNN performance, this method has the benefit of being able to tie the action to the person performing it.

	RGB	Depth
CNN over full image	0.695	0.450
CNN over dispenser region	**0.956**	**0.937**

Table 1: Average precision of hand hygiene action detection.

5 Conclusion

We believe that the recent success of using machine learning techniques over depth signals to perceive the world could have an unprecedented impact in health care. We have shown that it is possible to detect hand hygiene compliance, an important component of reducing the cost associated with hospital-acquired infections. In our future work we plan to address other use cases in health care as well.

References

[1] Y. Du, W. Wang, and L. Wang. Hierarchical recurrent neural network for skeleton based action recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1110–1118, 2015.

[2] D. Howell and K. Olsen. Distress the 6th vital sign. *Current oncology*, 18(5):208, 2011.

[3] H. Jhuang, J. Gall, S. Zuffi, C. Schmid, and M. J. Black. Towards understanding action recognition. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 3192–3199. IEEE, 2013.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[5] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[6] W. Li, Z. Zhang, and Z. Liu. Action recognition based on a bag of 3d points. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pages 9–14. IEEE, 2010.

[7] S. S. Magill, J. R. Edwards, W. Bamberg, Z. G. Beldavs, G. Dumyati, M. A. Kainer, R. Lynfield, M. Maloney, L. McAllister-Hollod, J. Nadle, et al. Multistate point-prevalence survey of health care–associated infections. *New England Journal of Medicine*, 370(13):1198–1208, 2014.

[8] A. Mittal, A. Zisserman, and P. H. S. Torr. Hand detection using multiple proposals. In *British Machine Vision Conference*, 2011.

[9] W. H. Organization et al. *WHO guidelines on hand hygiene in health care: first global patient safety challenge. Clean care is safer care*. World Health Organization, 2009.

[10] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, pages 1–42, 2014.

[11] J. Shotton, T. Sharp, A. Kipman, A. Fitzgibbon, M. Finocchio, A. Blake, M. Cook, and R. Moore. Real-time human pose recognition in parts from single depth images. *Communications of the ACM*, 56(1):116–124, 2013.

[12] J. Wang, Z. Liu, Y. Wu, and J. Yuan. Mining actionlet ensemble for action recognition with depth cameras. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 1290–1297. IEEE, 2012.

[13] L. Xia, C.-C. Chen, and J. Aggarwal. View invariant human action recognition using histograms of 3d joints. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*, pages 20–27. IEEE, 2012.