Towards Vision-Based Smart Hospitals: A System for Tracking and Monitoring Hand Hygiene Compliance

Albert Haque[1]   AHAQUE@CS.STANFORD.EDU

Michelle Guo[1]   MGUO95@STANFORD.EDU

Alexandre Alahi[1,2]   ALEXANDRE.ALAHI@EPFL.CH

Serena Yeung[1]   SYYEUNG@CS.STANFORD.EDU

Zelun Luo[1]   ZELUNLUO@STANFORD.EDU

Alisha Rege[1]   AMR6114@STANFORD.EDU

Jeffrey Jopling[3]   JJOPLING@STANFORD.EDU

Lance Downing[3]   LDOWNING@STANFORD.EDU

William Beninati[4]   BILL.BENINATI@IMAIL.ORG

Amit Singh[5]   AMSINGH@STANFORDCHILDRENS.ORG

Terry Platchek[5]   TPLATCHEK@STANFORDCHILDRENS.ORG

Arnold Milstein[3]   MILSTEIN@STANFORD.EDU

Li Fei-Fei[1]   FEIFEILI@CS.STANFORD.EDU

[1]*Department of Computer Science, Stanford University, United States*

[2]*Department of Civil Engineering, École Polytechnique Fédérale de Lausanne, Switzerland*

[3]*Clinical Excellence Research Center, School of Medicine, Stanford University, United States*

[4]*TeleCritical Care, Intermountain Healthcare, United States*

[5]*Division of Pediatric Hospital Medicine, Lucile Packard Children's Hospital Stanford, United States*

Abstract

One in twenty-five patients admitted to a hospital will suffer from a hospital acquired infection. If we can intelligently track healthcare staff, patients, and visitors, we can better understand the sources of such infections. We envision a smart hospital capable of increasing operational efficiency and improving patient care with less spending. In this paper, we propose a non-intrusive vision-based system for tracking people's activity in hospitals. We evaluate our method for the problem of measuring hand hygiene compliance. Empirically, our method outperforms existing solutions such as proximity-based techniques and covert in-person observational studies. We present intuitive, qualitative results that analyze human movement patterns and conduct spatial analytics which convey our method's interpretability. This work is a first step towards a computer-vision based smart hospital and demonstrates promising results for reducing hospital acquired infections.

1. Introduction

In recent years, smart hospitals have garnered large research interest in the healthcare community (Noury et al., 2008; Wu et al., 2006; Nugent and Augusto, 2006). Smart hospitals encompass a variety of network-connected products for controlling, automating and optimizing workflows in the hospital (Sánchez et al., 2008). Surgery checklists are known to
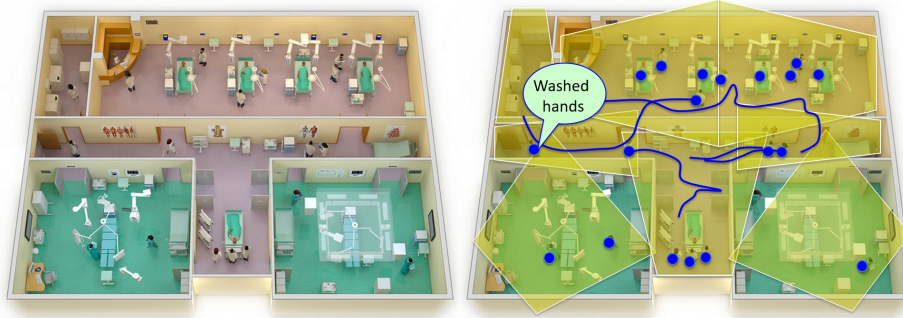
Figure 1: Illustration of our proposed vision-based smart hospital. (Left) Top-view of a hospital unit. (Right) Our system that track fine-grained activities such as hand hygiene compliance. Yellow polygons indicate areas covered by our sensors. Blue dots indicate pedestrian detections and blue lines indicate pedestrian movements over time.

improve clinical outcomes (Weiser et al., 2010) and smart sensors can be used to automate safety procedures (Yu et al., 2012). With automatic and continuous monitoring, smart hospitals can locate inventory, identify patients, track healthcare workers, and increase operational efficiency (Fisher and Monahan, 2008; Lenz and Reichert, 2007), ultimately improving patient care quality (Gao et al., 2006).

One significant problem that smart hospitals can help address is the prevention of hospital acquired infections (HAIs) (Cook and Schmitter-Edgecombe, 2009). One in twenty-five patients admitted to a hospital suffers from HAIs (Centers for Disease Control and Prevention, 2016), costing hospitals billions of dollars per year (Zimlichman et al., 2013). Proper hand hygiene is an important part of preventing such HAIs, and smart hospitals that are constantly aware of the dynamic environment can track the movements of healthcare workers, monitor hand hygiene compliance, and provide feedback.

Existing technologies for smart hospitals have been used to tackle some of these problems. Radio-frequency identification (RFID) systems have been used for tracking people in hospitals and are a common approach towards building a smart hospital (Fuhrer and Guinard, 2006). RFID systems are generally cheap and easy to deploy (Coustasse et al., 2013). Despite these benefits, objects and humans of interest must be manually tagged and remain physically close to a base station for the system to correctly locate them. For people, this generally requires that hospital staff wear special wristbands, gloves, or external badges to register their position at specific events of interest (e.g., before entering a patient room) (Simmonds and Granado-Villar, 2011; Yao et al., 2010). However, this provides us with a non-continuous location history. Combined with the noisy localization accuracy of indoor radio localization methods (Whitehouse et al., 2007; Alahi et al., 2015; Marra and Edmond, 2014; Zanca et al., 2008), this makes RFID systems ill-suited for smart hospitals. We need a more scalable solution to track both objects (e.g., portable computers, trash cans) and humans with higher precision at finer resolution (e.g., actions, body poses) and in a non-intrusive fashion (e.g., no devices on people).

We argue that computer vision-based approaches offer the ability to address these challenges. Recently, computer vision has pushed state-of-the-art in tracking applications such as self-driving cars (Cho et al., 2014) and sports analytics (Halvorsen et al., 2013; Pers

and Kovacic, 2000). They have the advantage of operating on a continuous data stream (i.e., videos) and have the ability to locate objects and pedestrians in both two- and three-dimensional space. However, if we are to fully understand the hospital environment we must track more than just human locations. Computer vision systems allow us to observe the world at high resolution, enabling the characterization of fine-grained motions such as handwashing or surgical procedures.

In this work, we develop a non-intrusive vision-based system for tracking peoples activity in hospitals, specifically for the problem of hand hygiene. While computer vision has shown promising results for tracking (Zhang et al., 2008), a number of unsolved challenges remain. First, due to differences in building construction, hospitals may contain sparse networks of sensors with minimal overlapping field of views. This can make it difficult for trajectory prediction and data association (Leibe et al., 2007) across entire hospital units. Second, hospitals exhibit variances in visitor and staff traffic patterns leading to extremely busy and crowded scenes. Third, due to privacy laws surrounding personal health information, our system is limited in the visual data it can use. We must resort to de-identified depth images, losing important visual appearance cues in the process. This makes it especially difficult for our system to continuously locate and track personnel over time. We must be able to link events of interest (e.g., exiting a patient room) with specific people.

In this paper, our contributions are as follows. First, we propose a system for analyzing people's activities in hospitals. We demonstrate our algorithm's performance on the task of automatic hand hygiene assessment in the context of a classification task and a tracking task. Not only can this improve hand hygiene compliance but it opens the doors for future smart hospital research. Second, we present results analyzing physical movement patterns and hospital space analytics. Our system's interpretable visualizations can be used to optimize nurse workflows, improve resource management (Fry and Lenert, 2005), and monitor hygiene compliance. Interpretability can help accelerate adoption time and build more trust between the computer, clinicians, and patients.

2. Data

Privacy regulations such as HIPAA[1] and GDPR[2] limit the use of cameras in healthcare settings. In cases where cameras are allowed, access to recordings is strictly controlled, often preventing the use of computer vision algorithms. To comply with these regulations, we use *depth images* generated from depth sensors instead of standard color cameras.

Depth Images. A *depth image* can be interpreted as a de-identified version of a standard color photo (see Figure 2). In color photos, each pixel represents a color, often encoded as combination of red, green, and blue (RGB) values as unsigned integers. In depth images, each pixel represents how far away the "pixel" is, often encoded in real world meters as floating point numbers. While we lose color information in depth images, we are able to respect patient privacy and data sharing protocols. Notice that in depth images, we (humans) are unable to see the color of the person or the environment, we still understand the semantics of the scene (e.g., group of people standing in a hallway). This is because depth images convey volumetric 3D information while color images convey colored

---

1. Health Insurance Portability and Accountability Act of 1996 (HIPAA)
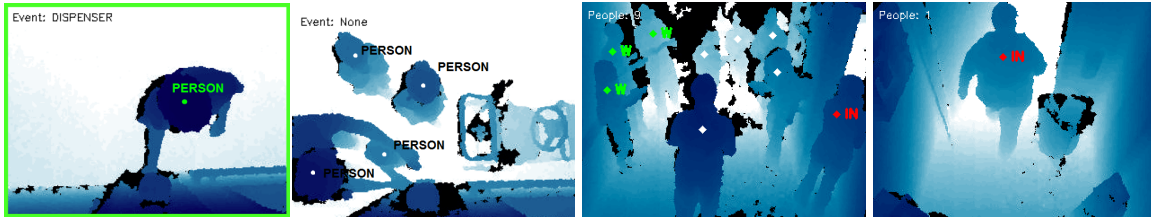2. General Data Protection Regulation (GDPR); Regulation (EU) 2016/679

Figure 2: Example depth images with human annotations. White dots represent different people. Green dots or "W" denote people performing handwashing events. Red "IN" denotes a person entering a room. The green box denotes a binary classification label.

appearance. It is this 3D spatial information that allows our computer vision algorithm to reason about hand hygiene activity.

Data Sources. Images were collected from an acute care pediatric unit and an adult intensive care unit. A total of two hospitals participated in the data collection campaign. Sensors were installed in top-down views above alcohol-based gel dispensers, in side-views overlooking the corridors, inside patient rooms, and in corners of anterooms overlooking sinks and patient beds. The sensors have a field of view 58° horizontally, 45° vertically, and have a functional range between 0.8 and 4.0 meters. The images include objects such as garbage bins, computers, and medical equipment (see Figure 2).

Ground Truth. The ground truth, or *gold standard*, is considered the absolute correct account or labeling as determined by one or more human subject matter experts, also known as *annotators*. Similar to Armellino et al. (2012), the ground truth defines whether a person correctly followed hand hygiene protocol. In hospitals, an annotator determines the ground truth either (1) on-site in real-time as events occur or (2) after-the-fact by watching video recordings. A major benefit of annotating video recordings is the ability to move forward and backward in time. Additionally, the annotator can replay the same event from multiple sensor viewpoints to verify correctness. Because of these benefits, we chose to annotate our data using remote video recordings.

The ground truth was determined by ten annotators. This group includes students, professors, and practicing physicians. Each annotator was responsible for identifying the (x, y) position of all people present in the image by clicking with a computer mouse. Additionally, for each person present in the image, the annotator was asked to note whether the person was entering or exiting a patient room and if so, whether they washed their hands upon entry or exit (Figure 2). This was done with keyboard input (e.g., pressing 1 or 2). Clinicians trained each annotator on proper hand hygiene protocol with live demonstrations. Each depth image was annotated by one to three annotators and their ground truth assessments were cross-validated. On average, each annotator spent three seconds annotating each image.

Dataset Statistics. Ground truth hand hygiene compliance data was collected on a Friday from 12 pm to 1 pm. During this hour, which is peak lunch time, there were a large number of visitors in the hospital. A total of 351 ground truth tracks were collected and annotated, of which 170 involved a person entering a patient room, of which 30 were compliant (i.e., followed correct hand hygiene protocol). Of the 181 tracks exiting a room, 34 were compliant. The data used for the classifier (Section 3.3) consists of 150,400 images, of

which 12,292 images contained pedestrians using the dispenser. The training set contained 80% of the images with the remaining 20% allocated to the test set.

3. Method

Our goal is to develop a system to automatically detect, track, and assess hand hygiene compliance. The hope is that such a system can be used as part of a greater smart hospital ecosystem and ultimately prevent HAIs caused by improper hand hygiene. To compute the hand hygiene compliance rate, we must perform three tasks: (1) detect the healthcare staff, (2) track them in the physical world, (3) and classify their hand hygiene behavior. This will allow us to compute quantitative compliance metrics across the hospital unit.

3.1 Pedestrian Detection

The first step involves locating the 3D positions of the staff within the field-of-view of each sensor. We use the same sparsity-driven formulation as in Golbabaee et al. (2014). An inverse problem is formulated to deduce pedestrian location points (i.e., an occupancy vector x) given a sparsity constraint on the ground occupancy grid. Let y be our observation vector (i.e., the binary foreground silhouettes), and D the dictionary of atoms approximating the binary foreground silhouettes of a single person at all locations. We aim to solve:

$$x = \underset{x \in \{0,1\}}{\arg \min} ||y - Dx||_2^2 \quad \text{s.t.} \quad ||x||_0 < \epsilon_p, \tag{1}$$

where ϵ_p is an upper bound of the sparsity level. We use the Set Covering Object Occupancy Pursuit algorithm (Golbabaee et al., 2014) to infer the occupancy map x. At each iteration, the algorithm selects the atom of the dictionary which best matches the image.

3.2 Tracking Across Cameras

Once pedestrians are located in the field-of-view of each camera, to track their hygiene status, we must track them across cameras. Formally, we want to find the set of trajectories X, where each trajectory $x \in X$ is represented as an ordered set of detections, L_x, representing the detected coordinates of pedestrians. Similarly, $L_x = (l_x^{(1)}, ..., l_x^{(n)})$ is an ordered set of intermediate detections which are linked to form the trajectories. These detections are ordered by the time of initiation. The problem can be written as a maximum a-posteriori (MAP) estimation problem. Next, we assume a Markov-chain model connecting every intermediate detection $l_x^{(i)}$ in the trajectory X, to the subsequent detection $l_x^{(i+1)}$ with a probability given by $P(l_x^{(i+1)}|l_x^i)$. We can now formulate the MAP task as a linear integer program by finding the flow f that minimizes the cost C:

$$\min_{f} \quad C = \sum_{x_i \in X} \alpha_i f_i + \sum_{x_i, x_j \in X} \beta_{ij} f_{ij} \qquad \text{s.t} \qquad f_i, f_{ij} \in (0,1) \quad \text{and} \quad f_i = \sum_{j} f_{ij}, \tag{2}$$

where f_i is the flow variable indicating whether the corresponding detection is a true positive, and f_{ij} indicates if the corresponding detections are linked together. The variable β_{ij} denotes the transition cost given by $\log P(l_i|l_j)$ for the detection $l_i, l_j \in L$. The local cost α_i is the log-likelihood of an intermediate detection being a true positive. In our case,
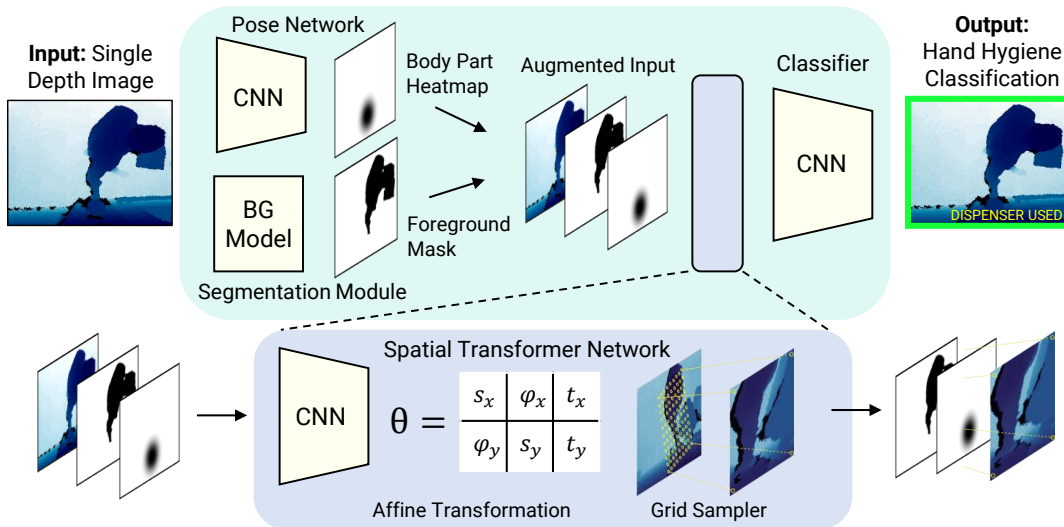
5

Figure 3: Hand hygiene activity classifier. Using pose and segmentation as mid-level cues, a spatial transformer projects the features into a viewpoint invariant feature space. θ denotes the learned transformation parameters: scale (s_x, s_y), skew (φ_x, φ_y), and translation (t_x, t_y).

we suppose that all detections have the same likelihood. This is equivalent to the flow optimization problem, solvable in real-time with k-shortest paths (Berclaz et al., 2011).

3.3 Hand Hygiene Activity Classification

At this stage, we have identified the tracks (i.e., position on the global ground plane) of all pedestrians in a hospital unit. The last step is to detect hand hygiene activity and link it to a specific track. That is, we must label each pedestrian track as *clean* or *not clean*. A person becomes *clean* when they use one of the various alcohol-based gel dispensers located throughout the hospital unit. We propose an activity classifier which accepts depth images as input and outputs a binary prediction of whether a person used the dispenser or not.

Learned Viewpoint Invariance. Deployment of sensors in real-world settings is often prone to error. Whether intentional or not, construction crews and maintenance technicians install sensors that vary in both angle and position. If our goal is to propose a non-intrusive vision-based system for tracking hand hygiene activity, our model must be robust to such variances. Towards this goal, the input to our classifier is a single depth image captured from any sensor viewpoint (Figure 3). We augment this depth image by estimating hand position estimates and a extracting a foreground body silhouette. The pose network is a two-layer convolutional network and the segmentation module uses optical flow to determine foreground-background changes. The output of the hand hygiene activity classifier is a binary prediction: whether a hand dispenser event occurred or not.

Traditional convolutional network architectures (Krizhevsky et al., 2012) are generally not viewpoint invariant (Jaderberg et al., 2015). We address this issue by introducing an implicit attention mechanism: a fully differentiable spatial transformer network that can spatially transform any input feature map (Jaderberg et al., 2015). Given the augmented input $U \in \mathbb{R}^{H \times W \times 3}$, the localisation network $f(\cdot)$ predicts the two-dimensional transforma-

6

tion parameters θ needed to transform the input. This performs a geometric transformation on U using the affine transformation matrix A_θ (also see Figure 3):

$$A_\theta = \begin{bmatrix} \theta_1 & \theta_2 & \theta_3 \\ \theta_4 & \theta_5 & \theta_6 \end{bmatrix} = \begin{bmatrix} s_x & \varphi_x & t_x \\ \varphi_y & s_y & t_y \end{bmatrix}. \tag{3}$$

With the transformation parameters θ predicted by the localisation network, the grid generator computes a *sampling grid* $G \in \mathbb{R}^{H \times W \times 3}$. This grid identifies the points in the augmented input which will be used for the transformed output. To perform a warping on U, a sampler applies a sampling kernel at each grid point $G_i = (x_i^t, y_i^t)$ to produce V, where (x_i^t, y_i^t) are the target coordinates in V. We use a bilinear sampling kernel, where the kernel $k = \max(0, 1 - |\cdot|)$. The output feature map V is thus defined as

$$V_i^c = \sum_n^H \sum_m^W U_{nm}^c \max(0, 1 - |x_i^s - m|) \max(0, 1 - |y_i^s - n|) \tag{4}$$

where (x_i^s, y_i^s) are source coordinates in the augmented input U that define the sample points. Because Equation 4 is a linear function applied to U, the spatial transformation network is differentiable, allowing loss gradients to flow through network (Jaderberg et al., 2015). At this point we have generated the transformed input V, which we will pass into a convolutional network classifier for predicting hand hygiene dispenser events.

3.4 Fusing Tracking with Hand Hygiene Activity Classification

Our goal is to create a system with a holistic understanding of the hospital. In isolation, an activity detector is insufficient to understand human behavior over time and a tracker is unable to understand fine-grained motions on its own. It is necessary to fuse the outputs from these components. This becomes a spatio-temporal matching problem where we must match hand hygiene activity detections with specific tracks.

For each hand hygiene classifier detection (i.e., dispenser is being used), we must match it to a single track. A match occurs between the classifier and tracker when a track $\mathcal{T}$ satisfies two conditions:
1. Track $\mathcal{T}$ contains (x, y) points $\mathcal{P}$ which occur at the same time as the hand hygiene detection event $\mathcal{E}$, within some temporal tolerance level.
2. At least one point $p \in \mathcal{P}$ is physically nearby the sensor responsible for the detection event $\mathcal{E}$. This is defined by a proximity threshold around the patient's door.

If there are multiple tracks that satisfy these requirements, we break ties by selecting the track with the closest (x, y) position to the door.

Final Output. The final output of our model is a list T of tracks, where each track consists of an ordered list of (t, x, y, a) tuples where t denotes the timestamp, x, y denote the 2D ground plane coordinate, and a denotes the latest action or event label. From T, we can compute the compliance rate or compare with the ground truth for evaluation metrics.

4. Experiments

Our goal is to build a computer vision system for automatically assessing hand hygiene compliance across an entire hospital unit. We conducted two primary experiments. First,

we evaluate our system's ability to assess hand hygiene compliance. Our system's detections are compared against the ground truth, including a comparison to existing hand hygiene assessment solutions. Second, we evaluate pur hand hygiene activity classifier. Formulated as a binary classification task, we show quantitative and qualitative results.

4.1 Baselines

Covert Observation. Today, many hospitals measure hand hygiene compliance using *secret shoppers* (Morgan et al., 2012). Secret shoppers are trained individuals who physically go to hospitals and measure compliance rates in-person without revealing the true purpose of their visit. We refer to this as a *covert observation*, as opposed to an *overt observation* performed by someone openly disclosing their audit. The purpose of covert observations is to minimize the Hawthorne effect (Adair, 1984).

In this work, we conducted two covert observational studies: (i) a single auditor responsible for monitoring the entire unit and (ii) a group of three auditors with a collective responsibility of the entire unit. We refer to these groups as *covert1* and *covert3*, respectively. The motivation for having two different covert groups is to measure the performance gains of having additional people for in-person observations. Both covert groups were disguised as hospital visitors. The group of three was spread out over the unit, remaining stationary, while the individual auditor constantly walked around the unit while monitoring hand hygiene compliance.

Proximity Algorithm. Existing RFID solutions can be interpreted as proximity algorithms. In some radio-based hand hygiene compliance assessments, a heathcare worker must *badge in* before or after washing their hands upon entering or exiting a patient a room. If a healthcare worker approaches a radio base station within some threshold, the RFID will activate indicating a localization event. Since the standard error for radio-based positioning is one meter (Whitehouse et al., 2007; Alahi et al., 2015; Zanca et al., 2008), we use this as our activation threshold. The proximity algorithm simulates this process. If a person approaches within one meter of a patient's door or hand hygiene dispenser, they are considered compliant with hand hygiene protocol. Using depth sensors, we can detect if a person approaches within one meter of a door or dispenser.

4.2 Compliance Results

We evaluated the result of our computer vision system compared to covert observational studies and the proximity algorithm in Figure 4. Each compliance assessment method aims to identify when a person enters or exit a room. Additionally, each compliance assessment method must classify the person as *compliant* or *non-compliant*. The person is flagged as non-compliant if they do not obey proper handwashing protocol before entering or after leaving a patient room. If the compliance assessment correctly identifies a compliant or non-compliant behavior, it is considered a successful detection and will increase its accuracy. When compared to the ground truth, our model achieves a 75% accuracy. This is higher than both covert1 and covert3 observations. Although the covert3 observation achieves 72% accuracy, in practice, this type of observation is rare. Most covert observations planned by hospitals consists of a single covert observer. In our experiments, a single covert observer

Method	Accuracy
Proximity Algorithm	18%
Covert Observation (1 Person)	63%
Covert Observation (3 People)	72%
Our Model	**75**%

Figure 4: Comparison of hand hygiene assessment method. Each method must identify: (i) when a person enters or exit a room and (ii) classify the person as *compliant* or *non-compliant*. If the method correctly classifies a person within five seconds of the ground truth, the method scores a correct detection towards the accuracy metric.
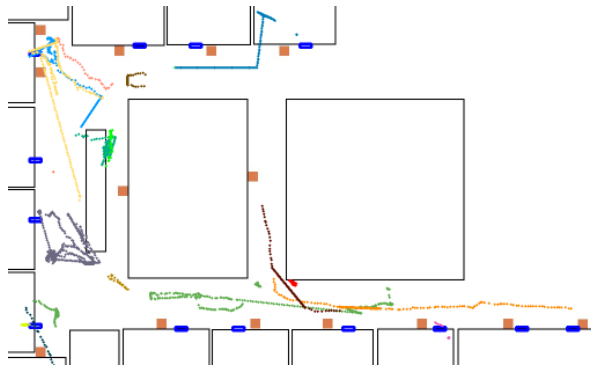

Figure 5: Top-down view of tracks. Blue rectangles are doors, orange squares are dispensers, and black lines are walls. Different track colors denote different people.

achieved a 63% accuracy. This is because the single observer was simply unable to monitor the entire hospital unit by themselves.

Traffic Patterns. Using our model's pedestrian tracking ability, we can derive insights about traffic patterns in the physical space. The method proposed in Section 3 allows us to convert between three-dimensional positions and two-dimensional coordinates for each person, at each point in time. Figure 5 shows several tracks (i.e., sequence of 2D points denoting a person's location over time) from a top view perspective. This type of analysis can be combined with hospital manifests to identify reasons for crowding around certain patient rooms and potentially identify areas prone to HAIs. With such a visualization, we are able to intuitively understand the location distribution of healthcare workers.

4.3 Hand Hygiene Activity Classification Results

When trained from scratch, our experiments show that ResNet-152 outperforms the other model architectures, due to the high number of layers and parameters of the model. We also show that augmenting the depth map with both the foreground segmentation and pose information also improves classifier performance. Finally, by jointly training the classifier with a spatial transformer network and making the classifier viewpoint invariant, we achieve a 1% boost in accuracy.

We now turn to a qualitative analysis of the hand hygiene classifier.. Figure 6 shows several before and after images produced by our model's transformation layer. Because the transformation is explicitly parametrized (see Equation 3), we can recover the bounding box and overlay it on the original image. Bounding boxes indicate the regions selected by our model, which are transformed and shown in Figure 6. Our model is capable of "zooming in" on areas of visual importance. The resulting transformation contains fewer extraneous and redundant information (i.e., floor and walls). We want to emphasize that we provide no information about sensor or dispenser placement to our model. The model learns how to transform each input image solely based on patterns present across and within viewpoints. Such a visualization greatly assists the interpretability of our model to clinicians.

9

Architecture	D	F	P	STN	Accuracy	Precision	Sensitivity	Specificity
AlexNet	X	X	X	X	93.9	91.8	96.3	91.4
VGG-16	X	X	X	X	92.3	91.9	92.8	91.8
ResNet-152	X	X	X	X	**95.5**	94.6	**96.7**	**94.5**
ResNet-152	X	X			94.6	93.1	96.3	92.9
ResNet-152	X	X		X	95.4	**95.3**	96.6	94.2

Table 1: Classifier ablation study: Effect of different inputs and architectures. D, F, and P denote depth, foreground, and pose inputs, respectively. STN denotes the spatial transformer network. The training and test sets are balanced with a 50/50 class-split.
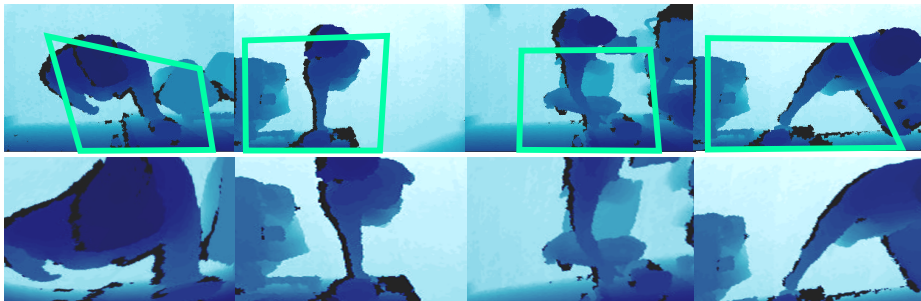

Figure 6: Examples before and after the spatial transformation. (Top) Input images with green bounding boxes from the grid generator. (Bottom) Transformed inputs. The model stretches, skews, and interpolates the input differently, depending on the scene contents.

5. Discussion & Related Work

There has been recent interest in creating smart hospitals with the aim of increasing operational efficiency and improving patient care (Fisher and Monahan, 2008; Gao et al., 2006). One use case of smart hospitals is in the prevention of HAIs (Cook and Schmitter-Edgecombe, 2009), or more specifically, for monitoring and tracking hand hygiene of hospital staff. Current technologies that track hand hygiene include RFID-based systems (Fuhrer and Guinard, 2006). However such systems are limited in resolution and precision (Alahi et al., 2015; Zanca et al., 2008). Computer vision-based tracking systems have shown promising results in non-clinical applications such as self-driving cars (Cho et al., 2014) and sports analytics (Halvorsen et al., 2013). We believe our work demonstrates the effectiveness of a computer vision-based method for tracking and monitoring hand hygiene compliance and identifies promising directions for future smart hospital research.

In this paper, we proposed a non-intrusive vision-based system for tracking people's activity in hospitals. We evaluated our method on measuring hand hygiene compliance and showed hand hygiene activity classification results. Our method outperforms existing solutions such as proximity-based techniques and covert in-person observational assessments. We presented intuitive, qualitative results that analyze human movement patterns and conduct spatial analytics which convey our method's interpretability. The system presented in this work is a first step towards a vision-based smart hospitals and demonstrates promising results for reducing HAIs and ultimately improve the quality of patient care.

References

John G Adair. The hawthorne effect: A reconsideration of the methodological artifact. *Journal of Applied Psychology*, 69(2):334, 1984.

Alexandre Alahi, Albert Haque, and Li Fei-Fei. Rgb-w: When vision meets wireless. In *International Conference on Computer Vision*, 2015.

Donna Armellino, Erfan Hussain, Mary Ellen Schilling, William Senicola, Ann Eichorn, Yosef Dlugacz, and Bruce F Farber. Using high-technology to enforce low-technology safety measures: the use of third-party remote video auditing and real-time feedback in healthcare. *Clinical infectious diseases*, 54(1):1–7, 2012.

Jerome Berclaz, Francois Fleuret, Engin Turetken, and Pascal Fua. Multiple object tracking using k-shortest paths optimization. *IEEE transactions on pattern analysis and machine intelligence*, 33(9):1806–1819, 2011.

Centers for Disease Control and Prevention. Healthcare-associated infections data and statistics. 2016.

Hyunggi Cho, Young-Woo Seo, BVK Vijaya Kumar, and Ragunathan Raj Rajkumar. A multi-sensor fusion system for moving object detection and tracking in urban driving environments. In *International Conference on Robotics and Automation*, pages 1836–1843. IEEE, 2014.

Diane J Cook and Maureen Schmitter-Edgecombe. Assessing the quality of activities in a smart environment. *Methods of information in medicine*, 48(5):480, 2009.

Alberto Coustasse, Shane Tomblin, and Chelsea Slack. Impact of radio-frequency identification (rfid) technologies on the hospital supply chain: a literature review. *Perspectives in Health Information Management*, (Fall), 2013.

Jill A Fisher and Torin Monahan. Tracking the social dimensions of rfid systems in hospitals. *International journal of medical informatics*, 77(3):176–183, 2008.

Emory Fry and Leslie A Lenert. Mascal: Rfid tracking of patients, staff and equipment to enhance hospital response to mass casualty events. In *AMIA*, 2005.

Patrik Fuhrer and Dominique Guinard. *Building a smart hospital using RFID technologies: use cases and implementation*. Department of Informatics-University of Fribourg, 2006.

Tia Gao, Dan Greenspan, Matt Welsh, Radford R Juang, and Alex Alm. Vital signs monitoring and patient tracking over a wireless network. In *Engineering in Medicine and Biology Society, Annual International Conference of the*, pages 102–105. IEEE, 2006.

Mohammad Golbabaee, Alexandre Alahi, and Pierre Vandergheynst. Scoop: A real-time sparsity driven people localization algorithm. *Journal of Mathematical Imaging and Vision*, 2014.

Pål Halvorsen, Simen Sægrov, Asgeir Mortensen, David KC Kristensen, Alexander Eichhorn, Magnus Stenhaug, Stian Dahl, Håkon Kvale Stensland, Vamsidhar Reddy Gaddam, Carsten Griwodz, et al. Bagadus: an integrated system for arena sports analytics: a soccer case study. In *Multimedia Systems Conference*, pages 48–59. ACM, 2013.

Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al. Spatial transformer networks. In *Advances in Neural Information Processing Systems*, pages 2017–2025, 2015.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105, 2012.

Bastian Leibe, Konrad Schindler, and Luc Van Gool. Coupled detection and trajectory estimation for multi-object tracking. In *International Conference on Computer Vision*, pages 1–8. IEEE, 2007.

Richard Lenz and Manfred Reichert. It support for healthcare processes–premises, challenges, perspectives. *Data & Knowledge Engineering*, 61(1):39–58, 2007.

A.R. Marra and M.B. Edmond. New technologies to monitor healthcare worker hand hygiene. *Clinical Microbiology and Infection*, 20(1):29 – 33, 2014.

Daniel J Morgan, Lisa Pineles, Michelle Shardell, Atlisa Young, Katherine Ellingson, John A Jernigan, Hannah R Day, Kerri A Thom, Anthony D Harris, and Eli N Perencevich. Automated hand hygiene count devices may better measure compliance than human observation. *American Journal of Infection Control*, 40(10):955–959, 2012.

N Noury, T Hadidi, M Laila, A Fleury, C Villemazet, V Rialle, and A Franco. Level of activity, night and day alternation, and well being measured in a smart hospital suite. In *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, pages 3328–3331. IEEE, 2008.

C Nugent and JC Augusto. A system for activity monitoring and patient tracking in a smart hospital. In *Smart Homes and Beyond: ICOST 2006: 4th International Conference on Smart Homes and Health Telematics*, volume 19, page 196. IOS Press, 2006.

Janex Pers and Stanislav Kovacic. Computer vision system for tracking players in sports games. In *Workshop on Image and Signal Processing and Analysis*, pages 177–182. IEEE, 2000.

Dairazalia Sánchez, Monica Tentori, and Jesús Favela. Activity recognition for the smart hospital. *IEEE intelligent systems*, 23(2), 2008.

Barbara Simmonds and Deise Granado-Villar. Utility of an electronic monitoring and reminder system for enhancing hand hygiene practices in a pediatric oncology unit. *American Journal of Infection Control*, 39(5):E96–E97, 2011.

Thomas G Weiser, Alex B Haynes, Gerald Dziekan, William R Berry, Stuart R Lipsitz, Atul A Gawande, et al. Effect of a 19-item surgical safety checklist during urgent operations in a global patient population. *Annals of Surgery*, 251(5):976–980, 2010.

Kamin Whitehouse, Chris Karlof, and David Culler. A practical evaluation of radio signal strength for ranging-based localization. *Mobile Computing and Communications Review*, 11(1):41–52, 2007.

B Wu, Z Liu, R George, and KA Shujaee. ewellness: Building a smart hospital by leveraging rfid networks. In *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, pages 3826–3829. IEEE, 2006.

Wen Yao, Chao-Hsien Chu, and Zang Li. The use of rfid in healthcare: Benefits and barriers. In *International Conference on RFID-Technology and Applications*, pages 128–134. IEEE, 2010.

Lei Yu, Yang Lu, and XiaoJuan Zhu. Smart hospital based on internet of things. *International Research Journal of Engineering and Technology*, 7(10):1654–1661, 2012.

Giovanni Zanca, Francesco Zorzi, Andrea Zanella, and Michele Zorzi. Experimental comparison of rssi-based localization algorithms for indoor wireless sensor networks. In *Proceedings of the Workshop on Real-world Wireless Sensor Networks*, pages 1–5. ACM, 2008.

Li Zhang, Yuan Li, and Ramakant Nevatia. Global data association for multi-object tracking using network flows. In *Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2008.

Eyal Zimlichman, Daniel Henderson, Orly Tamir, Calvin Franz, Peter Song, Cyrus K Yamin, Carol Keohane, Charles R Denham, and David W Bates. Health care–associated infections: a meta-analysis of costs and financial impact on the us health care system. *JAMA Internal Medicine*, 173(22):2039–2046, 2013.