PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Heatmap generation for emergency medical procedure identification

Richard A. Paris, Paul Sullivan, Jamison Heard, Diedre Scully, Candace McNaughton, et al.

Richard A. Paris, Paul Sullivan, Jamison Heard, Diedre Scully, Candace McNaughton, Jesse M. Ehrenfeld, Julie A. Adams, Joseph Coco, Daniel Fabbri, Bobby Bodenheimer, "Heatmap generation for emergency medical procedure identification," Proc. SPIE 10951, Medical Imaging 2019: Image-Guided Procedures, Robotic Interventions, and Modeling, 1095130 (8 March 2019); doi: 10.1117/12.2513122

SPIE.

Event: SPIE Medical Imaging, 2019, San Diego, California, United States

Heatmap Generation for Emergency Medical Procedure Identification

Richard A. Paris[a], Paul Sullivan[a], Jamison Heard[a], Diedre Scully[a], Candace McNaughton[b], Jesse M. Ehrenfeld[b], Julie A. Adams[c], Joseph Coco[b], Daniel Fabbri[b] and Bobby Bodenheimer[a]

[a]Vanderbilt University
[b]Vanderbilt University Medical Center
[c]Oregon State University

ABSTRACT

Ideal treatment of trauma, especially that which is sustained during military combat, requires rapid management to optimize patient outcomes. Medical transport teams 'scoop-and-run' to trauma centers to deliver the patient within the 'golden hour', which has been shown to reduce the likelihood of death. During transport, emergency medical technicians (EMTs) perform numerous procedures from tracheal intubation to CPR, sometimes documenting the procedure on a piece of tape on their leg, or not at all. Understandably, the EMT's focus on the patient precludes real-time documentation; however, this focus limits the completeness and accuracy of information that can be provided to waiting trauma teams. Our aim is to supplement communication that occurs en route between point of injury and receiving facilities, by passively tracking and identifying the actions of EMTs as they care for patients during transport. The present work describes an initial effort to generate a coordinate system relative to patient's body and track an EMT's hands over the patient as procedures are performed. This 'patient space' coordinate system allows the system to identify which areas of the body were the focus of treatment (e.g., time spent over the chest may indicate CPR while time spent over the face may indicate intubation). Using this patient space and hand motion over time in the space, the system can produce heatmaps depicting the parts of the patient's body that are treated most. From these heatmaps and other inputs, the system attempts to construct a sequence of clinical procedures performed over time during transport.

1. DESCRIPTION OF PURPOSE

The purpose of this work is to automatically identify the clinical procedures that emergency medical technicians (EMTs) perform during transport using off-the-shelf passive sensors such as video cameras and EMT-worn accelerometers (e.g., the Apple Watch). A passive system, in which no active input is required, is necessary to avoid distracting the EMT away from patient care activities. Current documentation and communication to receiving medical teams includes hand-written notes and brief verbal reports, respectively. In both forms, the information presented to the receiving team can be incomplete and inaccurate. Supplementing these existing communication methods with an automatically produced list of clinical procedures with time stamps has the potential to more adequately prepare for the triage and downstream management of trauma cases.

2. METHODS

In this work, the system uses a single type of data, video data feeds, to identify clinical procedures. The video feeds are processed with a computer vision system OpenPose,[1–3] which analyzes each frame to identify people in the frame and calculate their skeletons. The skeletons include 18 different key point positions including hands, feet and the head. These key points designate where in each frame the people and their extremities are. In our data, the people in the frame consist of an EMT and a patient. Given these key points, the system first identifies the patient using simple heuristics such as being centered in the frame and having minimal movement. Next, the system identifies the EMT as the person closest to the patient. Once the patient and the EMT are identified, the system constructs a 'patient space', which is a geometric space relative to the patient's body. The system then tracks the EMT's hands in the patient space (i.e., hands over the head or over the leg).

To simulate real-world trauma transport, the team compiled a list of procedures that typically occur in an emergency setting, as seen in Table 1. The set of procedures were determined by analyzing military tactical

Medical Imaging 2019: Image-Guided Procedures, Robotic Interventions, and Modeling,
edited by Baowei Fei, Cristian A. Linte, Proc. of SPIE Vol. 10951, 1095130 · © 2019 SPIE
CCC code: 1605-7422/19/$18 · doi: 10.1117/12.2513122

Table 1. List of procedures and number of times each subject was supposed to complete each procedure. *Indicates that this procedure took place on the left or right arm randomly.

Medical Procedure	Times Completed	Medical Procedure	Times Completed
Adminster IM Medication*	5	Place an Oral Airway	10
Adminster IO Medication*	5	Place Blood-Pressure Cuff	5
Adminster IV Medication*	6	Place ECG Leads	5
Bagging	3	Place IO Line	5
Combat Gauze on Arm*	3	Place Pulse-Ox Monitor	5
Combat Gauze on Head	3	Splint Arm*	3
Combat Gauze on Leg*	3	Splint Leg*	3
Perform Chest Decompression	5	Take out ETT Tube	2
Perform CPR	1	Take out King Airway	2
Perform Intubation	2	Take out Oral Airway	10
Perform King Airway	2	Tourniquet on Arm*	3
Place an IV Line (Left Arm)	3	Use Stethoscope to check vitals	5
Place an IV Line (Right Arm)	3		

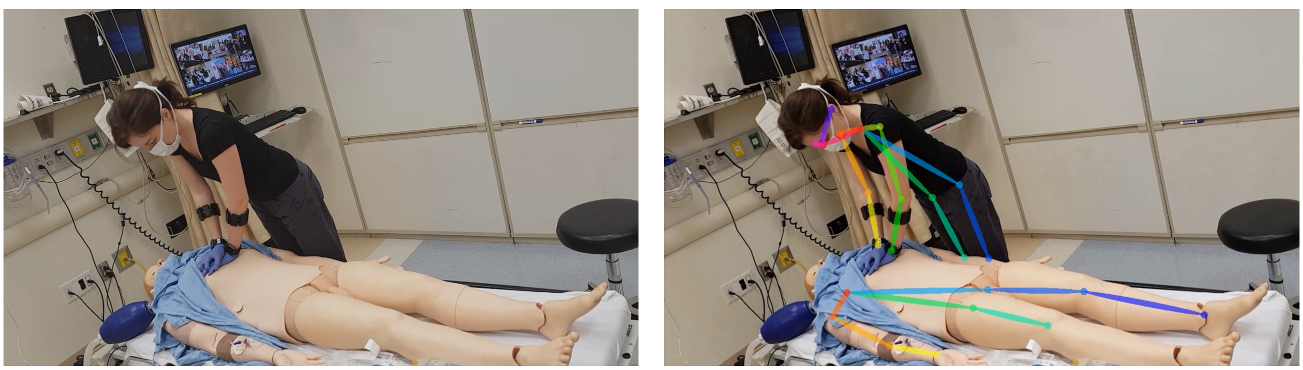

Figure 1. Still image taken from video data (left) and the same frame with OpenPose generated data overlaid to form a skeletal representation of both the patient and EMT (right).

combat care guidelines[4–6] and interviewing paramedics and trauma staff. The list of procedures includes a span of procedure types including airway management, medication administration, and stabilization. Video of four subjects with various medical and emergency response training was then recorded of each subject performing the procedures in a simulation lab (Figure 1). Each of these subjects performed a number of iterations of each procedure. Repetition allowed for the detection of individual differences as well as repetitive differences.

The video data collection system was configured as follows. Video was recorded with four Apeman A20 4K action cameras, which record 3840 by 2160 pixels at 24 fps. Video data were collected from four angles for 3D reconstruction. The positioning of the cameras relative to the patient is shown in Figure 2. Each of these cameras are at a height of 2m to ensure that the patient and subject are visible in each camera. Camera 2 was selected so that the patients body would be centered in the frame and so that screen space would roughly correspond to a 2D plane directly over the body. The Apeman cameras generate a series of 181 second videos with one second of overlapping frames between clips in the series. The final one second was removed (24 frames) of overlapping video so that no duplicate processing is completed by OpenPose. Each 3 minute video was analyzed with OpenPose, which was running on an NVidia Docker virtual machine using two GeForce GTX Titan X GPUs.

Visual inspection by trained personnel in conjunction with specific *a priori* criteria (such as two fingers on

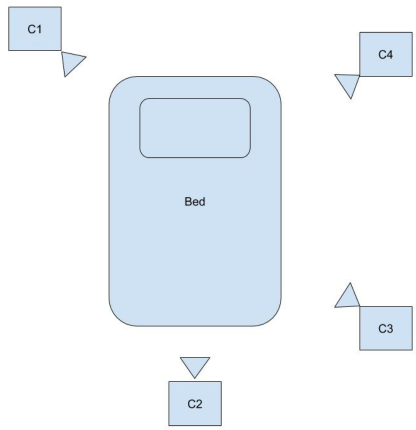

Figure 2. Positioning of the four cameras used during data collection. Each camera shot 4K video at 24 fps.

the wrist) are used to determine the beginning and end points of a procedure. To determine the exact frame at which each procedure begins and ends, the trained personnel visually inspect the recording and tag frames. These beginning and end points are used to split the data into smaller procedure-specific chunks to be analyzed.

Given these procedure-specific video chunks, hand key points of the EMT are extracted and used to generate a Gaussian field around them. This extraction process is done for every frame and summed over all frames in the chunk. By summing intensities of the fields over all chunks and frames, a heatmap is generated over the body showing the most frequently occurring positions of the hands for the chunk (Figure 3). These heatmaps will be used as training data for a convolutional neural net (CNN) classifier, which is intended to classify procedures.

These heatmaps represent one way to visualize where the hands are relative to the body. The second visualization method we present turns these heatmaps into bar charts. These bar charts show how much time the EMT's hands spend in close proximity to each of the 18 skeletal points. To create these bar charts we take each frame during the procedure and calculate the closest skeletal point to each hand. We calculate distance using pixels as reported by OpenPose. This data can be used separately or in conjunction with the heatmap data as training data for a CNN classifier. Every distribution graph is limited to the 6 most common body parts.

3. RESULTS

3.1 Across Procedure Comparison

Figure 3 shows the heatmaps generated from three procedures: intubation, insertion of an IV and splinting a leg. The background image represents the patient's body and the colors represent the position of the EMT's hands over the patient's body in patient space. The yellow color represents the areas above and around the patient where the EMT hands are located most often. Visually these heatmaps indicate that we can identify the body part which is being worked on, which will help in determining which procedure is being performed. Figure 4 shows the equivalent bar graphs. These procedures have very different signatures which is visible from the most common nearest body part.

3.2 Within Subject Validation

To reliably classify procedures a large amount of data is required. We obtain this data in two ways. The first way is to have a single participant repeat procedures multiple times. Figure 6 shows the distributions of three instances in which a single subject administered medication via an IV. These distributions show the six most frequent body areas visited by the EMT out of the 18 body areas in three instances, arranged in decreasing order from left to right. All show similar trends. The most common body parts are the elbows and wrists of each arm. Note that the OpenPose system occasionally reverses the patient's left and right (Figure 5). This can cause our system to misidentify the EMT's hands as being closest to the incorrect arm.

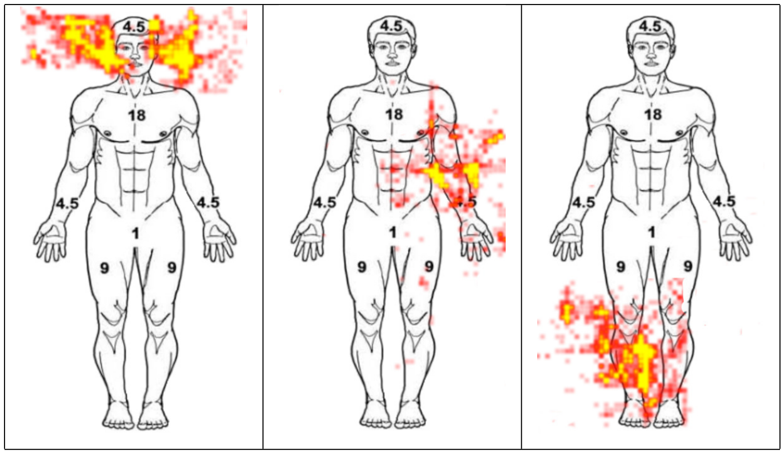

Figure 3. Heatmap showing the position of both the EMT's hands over the patient's body during a single instance of intubation, insertion of an IV, and splinting of a leg (from left to right).

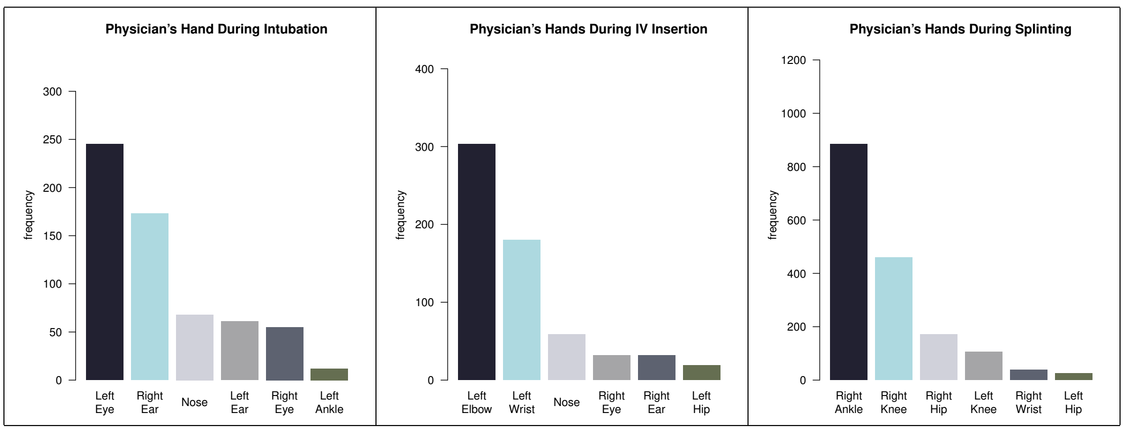

Figure 4. Bar graphs showing the distribution of both EMT's hands over the patient's body during a single instance of intubation, insertion of an IV, and splinting of a leg (from left to right).

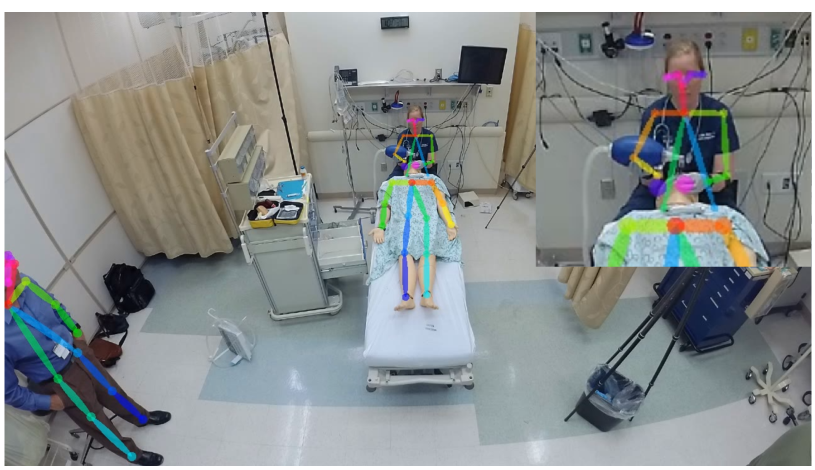

Figure 5. In some cases when the patient's face is not visible the left and right side of the patient can reversed. The blue/purple extension should indicate the left leg as shown by an observer and the EMT. However, in this frame the right leg is identified. This is also true of the left and right ear. The colors indicating which ear is which are reversed for the EMT and patient even though they are both facing toward the camera.

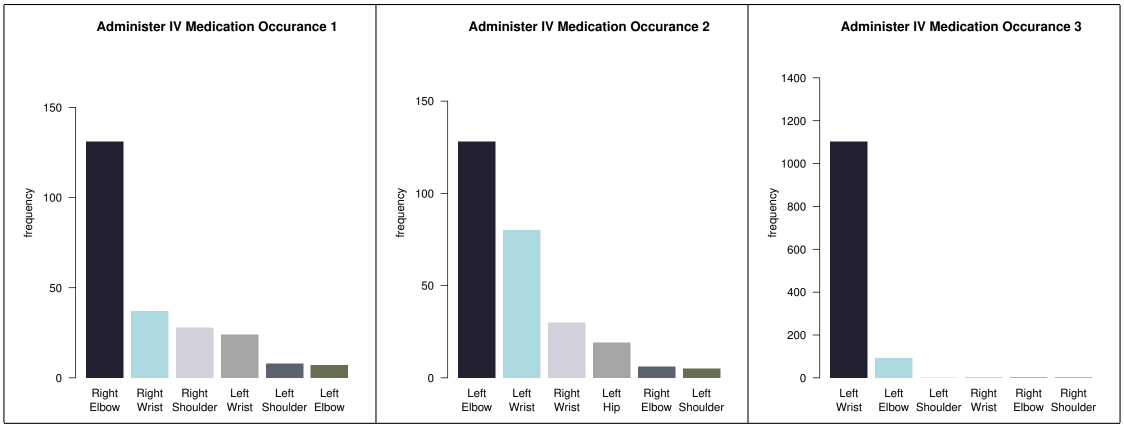

Figure 6. Bar graphs showing the distribution of both EMT's hands over the patient's body during three different instances of the EMT inserting an IV. The six body areas (out of 18) with the highest occurrence frequency of hands over the body area are shown in decreasing order from left to right.

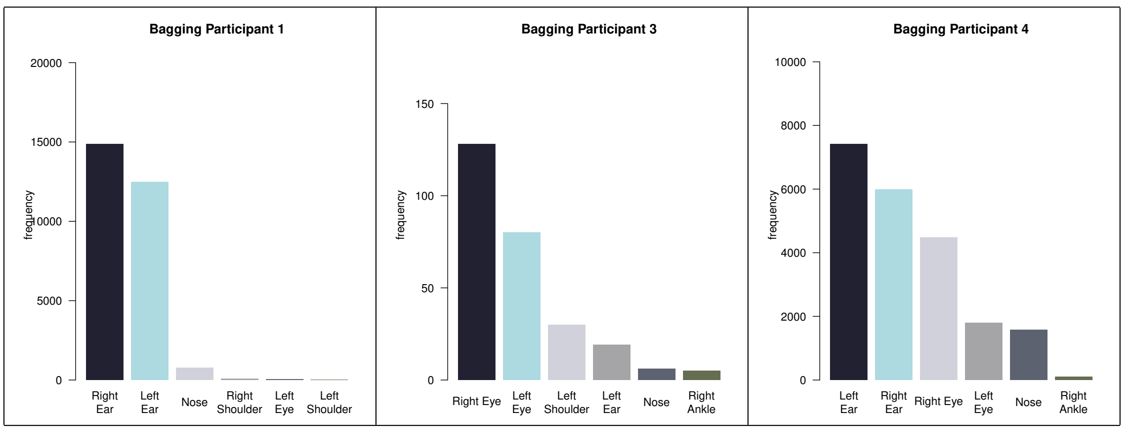

Figure 7. Bar graphs showing the distribution of each of three EMT's hands over the patient's body during a single instance of bagging.

3.3 Across Subject Validation

The second way to obtain the necessary data is by having multiple subjects. Here we explore the variability across subjects performing the same procedure. Figure 7 shows an instance of three different EMTs bagging the patient. Each of these bar graphs show that the hands are predominantly near the head (either the eyes, ears, or nose).

3.4 Across Camera Comparison

Up to this point all results and figures have been from the perspective of a single camera, camera 2 (Figure 2). OpenPose uses only a single camera angle to generate its 2D representations. However, we have viewpoints from multiple cameras. Figure 8 shows the distributions for cameras 2 and 3 (Figure 2) when the EMT checks the vitals of the patient. Likewise, Figure 9 shows the distributions when the EMT administers an IV to the patient. For both procedures, the distributions from both cameras are substantially similar. The point here is that additional perspectives do not provide additional information for the purpose of training, but multiple cameras do provide the ability to secure information in the event of occlusion, as when the EMT blocks the view of the patient from one camera.

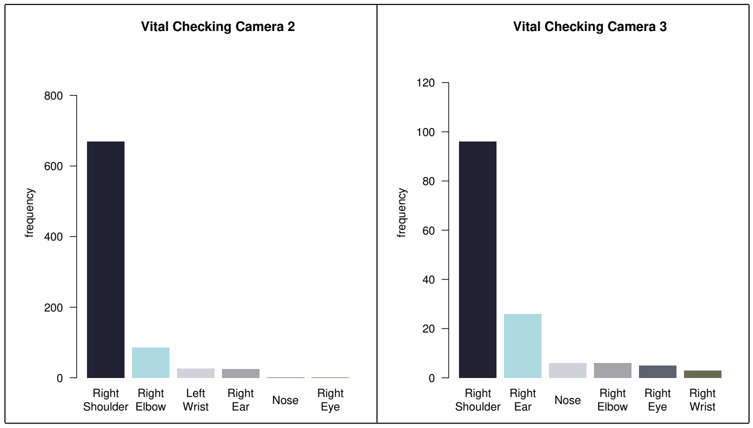

Figure 8. Bar graphs showing the distribution of an EMT's hands while checking the patient's vitals from camera 2 (left) and camera 3 (right).

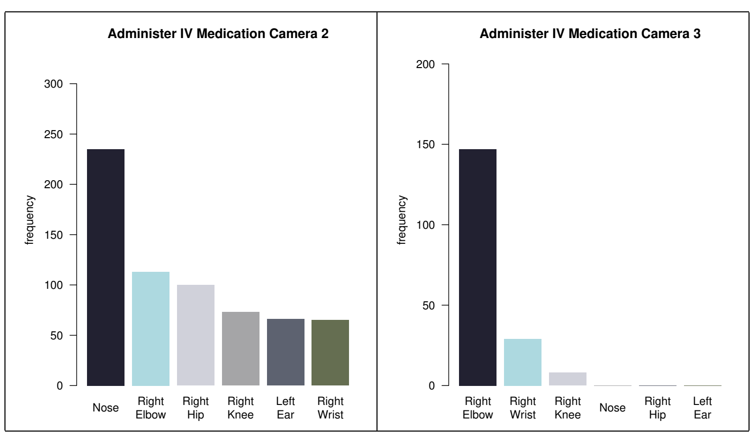

Figure 9. Bar graphs showing the distribution of an EMT's hands while administering IV medication to the patient from camera 2 (left) and camera 3 (right).

4. CONCLUSIONS

This work presents a method to video record training data of medical procedures and visualize heatmaps of those procedures. This visualization allows inspection of a given set of procedures. These heatmaps can be used as training data to classify procedures and may aid in computer identification of the procedure as it is being performed in emergency conditions. Since we intend to use this in conjunction with activity data gathered from other sensors, this work shows a first step in how computer vision and machine learning can be used to help further identify the procedure being performed.

In this work we demonstrated that we can track an EMT's body and identify the location on the body in which treatment is being administered. During one specific procedure, checking the patients vitals, multiple camera angles yield similar hand location signatures. This finding shows that we can use multiple cameras to confirm relative hand location. Figure 6 shows repeatability between procedures performed by the same medic. Additionally Figure 7 shows the repeatability between the same procedure being performed by different medics. These two taken together show the potential strength of a CNN in identifying procedures based on the hand distributions.

ACKNOWLEDGMENTS

This work was supported by the Department of Defense Contract Number W81XWH-17-C-0252 from the CDMRP Defense Medical Research and Development Program.

REFERENCES

[1] Cao, Z., Simon, T., Wei, S.-E., and Sheikh, Y., "Realtime multi-person 2d pose estimation using part affinity fields," in [*Computer Vision and Pattern Recognition*], (2017).

[2] Simon, T., Joo, H., Matthews, I., and Sheikh, Y., "Hand keypoint detection in single images using multiview bootstrapping," in [*Computer Vision and Pattern Recognition*], (2017).

[3] Wei, S.-E., Ramakrishna, V., Kanade, T., and Sheikh, Y., "Convolutional pose machines," in [*Computer Vision and Pattern Recognition*], (2016).

[4] Carlson, J. N., Karns, C., Mann, N. C., E. Jacobson, K., Dai, M., Colleran, C., and Wang, H. E., "Procedures performed by emergency medical services in the united states," *Prehospital Emergency Care* **20**(1), 15–21 (2016).

[5] El Sayed, M., Tamim, H., and Mann, N. C., "Description of procedures performed on patients by emergency medical services during mass casualty incidents in the united states," *The American journal of emergency medicine* **33**(8), 1030–1036 (2015).

[6] Butler, F., [*Tactical Combay Casulty Care Handbook (17-13), Version 5*], U.S. Army Center for Army Lessons Learned, Fort Leavenworth, KS (May 2017).