Automated Workflow and Activity Recognition at the Pediatric ICU

Proposal
February 15th, 2012

Colin Lea
Mentors: Dr. Jim Fackler, Dr. Greg Hager, Dr. Russell Taylor

Project Summary

Our goal is to develop algorithmic tools for analyzing clinical workflow to identify directions for improving provider efficiency in an Intensive Care Unit. We believe that this may also lead to improved patient safety. The plan is to outfit a Pediatric ICU patient room with Xbox Kinects to gather video and 3D data to monitor the patient-staff and staff-device interactions. The focus over the course of this semester is to identify the actions that appear most frequently and to create a retrospective activity monitor to automatically log their occurrences. We will use a two-tiered approach for generating derived "low-level" signals and "high-level" activities that we believe will be adequate for workflow analysis.

Background and Relevance

If you talk to most staff at a hospital's Intensive Care Unit (ICU) you will get the same story. Process is inefficient, everyone is very busy, and there is a lack of written direction. For every patient there are many hundreds of tasks that need to be completed by a team of up to fifteen staff members. These include actions like giving patients medication, emptying chest tubes, and documenting vital signs on a regular basis. By talking with doctors and nurses at the pediatric ICU it has become apparent that vision-based activity monitoring may help determine tasks being performed and provide insight into making the ICU a safer and more efficient workplace.

Activity monitoring in video has been prevalent in computer vision over the past twenty years and is becoming an even greater area of interest. The recent advent of inexpensive 3D imaging systems like the Xbox Kinect brings a new dimension of data without relying on complex multi-view camera networks. Currently, with a few notable exceptions, most activity recognition research is done working with 2D images or is done to classify out-of-context actions – that aren't representative of the real world – in a 3D framework. Our approach extends activity research using a dataset that has a cluttered environment and includes complex actions.

The overarching goal is to determine a high level understanding of patient and staff interactions. Three key questions are as follows:

- What are the most frequent patient-staff and staff-device interactions?
- What features are necessary to classify these actions automatically?
- With what accuracy can the top actions be recognized?

Technical Approach
Data Collection
Data will be collected at the pediactric ICU at JHMI using two Xbox Kinects. Figure 1 depicts the approximate location of the equipment. Green circle/triangles represent the Kinects and the gray squares in the corners represent the computers used to record the data.

A preliminary set of data will be recorded over a period of a few hours at the start of the project. The data will be hand annotated to assess the types of actions being performed and their frequency. Additional datasets will be recorded later in the semester on an as-needed basis.
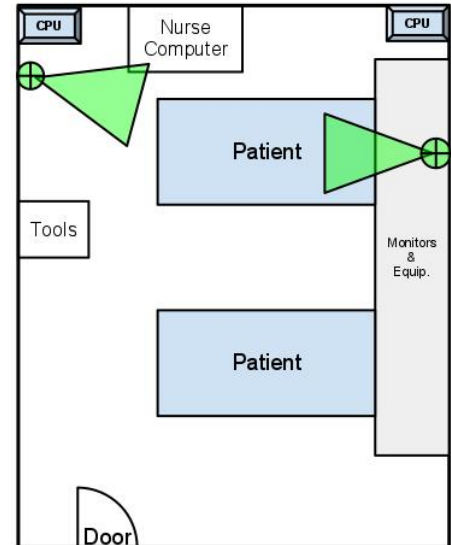

Figure 1 Equipment Layout in the PICU

Activity Recognition
The activity recognition approach we take consists of a pipeline that inputs a raw signal from the Kinect, creates "derived signals," and outputs an action label. This is depicted in the following figure.
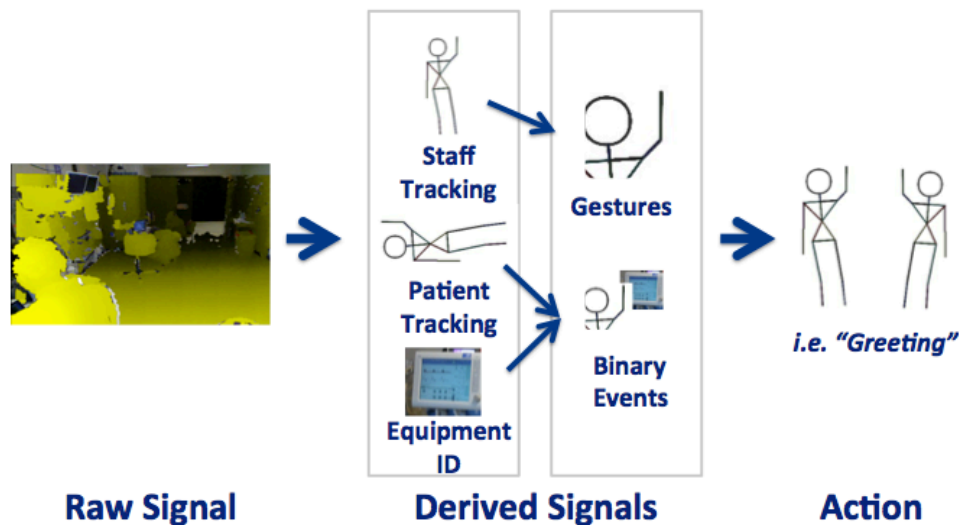

Figure 2 Activity Recognition Pipeline

Raw signal consists of the data collected using our recorder and includes video, depth images, and potentially Kinect skeletons. From this data we will derive more useful signals such as the locations of the staff. Each derived signal requires implementing an algorithm to take the raw data and return a more usable result. Each signal will be developed as a "module" that we can plug into our system. Additional activity recognition modules will be developed that use whatever signals are available as inputs.

Potential modules:
Staff location tracker – Determine the positions of visitors within the room. This can be used to initialize the body pose tracker.

Staff body pose tracker – The position of joints, such as hands, elbows, and the head will tracked. This will be used for gesture recognition.

Patient tracker – It may be useful to know if a staff member is touching a particular part of the patient. For example, it may be useful to know if the head or chest is being touched to distinguish between types of interactions.

Gesture recognizer – Gestures will be used to help identify actions. This requires the body pose tracker. I recently developed a technique based on PCA that performs with high accuracy on simple gestures.

Binary event recognizer – This module may track actions such as "Is a nurse touching a body?" or "Is a nurse touching a device?"

Activity Recognition – A simple technique will be implemented early in the project based on the gesture recognizer. Significantly more time will be spent on developing a more complex method, most likely using a type of linear dynamic system (ie Conditional Random Field, Switching Linear Dynamic System, or Hidden Semi-Markov Model).

Deliverables

There are four key components to this project, as shown in the following figure. Tasks with a green circle are considered "minimum" deliverables, yellow stars are "expected" and black diamonds are the "maximum." The final deliverable for each part is either a set of code for each module or a set of data for the experiments. A final report will also be written documenting the project.

Recorder – A general-purpose recorder that de-anonymizes the data for collection in the ICU. This includes face removal, which will be done with OpenCV's Viola Jones-based face detection technique, as well as patient remove from the color images.

Experimental Data – A minimum of one set of de-anonymized data from the pediatric ICU. This requires that our proposal is accepted by the Institutional Review Board. We anticipate that this will happen within the next few weeks. After collection we will annotate the video to determine the most frequent actions that happen so we can better gear our recognition algorithms.

Derived Signal Modules – Minimally a staff location tracker and body pose tracker will be used. The location component will be relatively simple to implement. However, the body pose tracker will likely be much more difficult. If possible we will use other research code for body pose.

Activity Recognition Modules – A gesture recognition system and a time-series recognition model will be developed to log the interactions in the ICU. The gesture recognition system is already largely completely but may require changes once we get a better understanding of what the ICU data will look like.

Recorder

Task	Difficulty	Notes
General Purpose Recorder	Easy	In Progress
Face removal	Medium	In Progress
Blackout stretcher	Easy	
Video tracking failure	Medium	

Experimentation

Task	Difficulty	Notes
Get IRB Approval	Easy	Pending
Record preliminary data	Easy	
Ensure de-identification	Easy	
Hand annotate activities	Medium	
Additional recording	Easy	

Derived Data

Task	Difficulty	Notes
Location tracker	Easy	
Face direction tracker	Easy	CSIRO Software
Staff body pose estimation	Hard	
Patient tracker	Medium	
Equipment Identification	Hard	

Activity recognition

Task	Difficulty	Notes
Gesture recognition	Medium	In Progress
Retrospective analysis	Hard	
Patient Tracking	Medium	
Multi-camera models	Medium	
Evaluation	Easy	

Figure 3 Task Breakdown

Milestones & Dates

The following milestones will be used as a guide for completing this project. Note that these dates are contingent to dependencies such as getting access to the IRB. The modular structure of the project means that even if this dependency is not filled in a timely manner, other sections can still be worked on.

Recorder   2/29/2012
Metric: Fully functional recording software will be completed that captures depth, video, and the Kinect skeleton structure. A test will be done over a period of 1+ hour to ensure there are no problems with long-term recording.

Preliminary data collection   3/27/2012
Metric: I will have raw data from the Kinect in the ICU that I will hand annotate.

Derived Signal Modules
Metric 1: Gesture recognition module   2/16/2012
Metric 2: Body location module   3/30/2012
Metric 3: Body joint location module   4/12/2012

Metric 4: Scene analysis module   5/3/2012

Activity Recognition Modules
Metric 1: Preliminary recognition module   3/15/2012
Metric 2: Advanced recognition module   4/26/2012

Final documentation   5/10/2012

Milestone	Mid Feb	End Feb	Early March	Mid March	End March	Early April	Mid April	End April	Early May
Recorder		■							
Data					■				
Signal 1	■								
Signal 2						■			
Signal 3							■		
Signal 4									■
Activity 1				■					
Activity 2								■	
Docs									■

Figure 4 Milestone Completion Dates

Dependencies
The following graph depicts the primary dependencies for each part of the project. Additional dependencies are listed below. The red blocks in the graph are current unresolved dependencies and the green blocks are resolved dependencies.

In order to do data collection at the ICU we must have IRB approval and a data recorder. Currently I have multiple data recorders for extracting depth and video or the Kinect skeleton but not one robust system for doing both. This needs to be completed before full experiments can be run.

Note that we don't need to run experiments in the ICU to perform the recognition tasks. Using data available freely online through a CVPR competition as well as data I have collected on my own we can run our derived data and activity recognition algorithms. Currently there is one set of derived signals from a gesture analysis algorithm I developed, thus, for example, the activity component can be in development at any time.
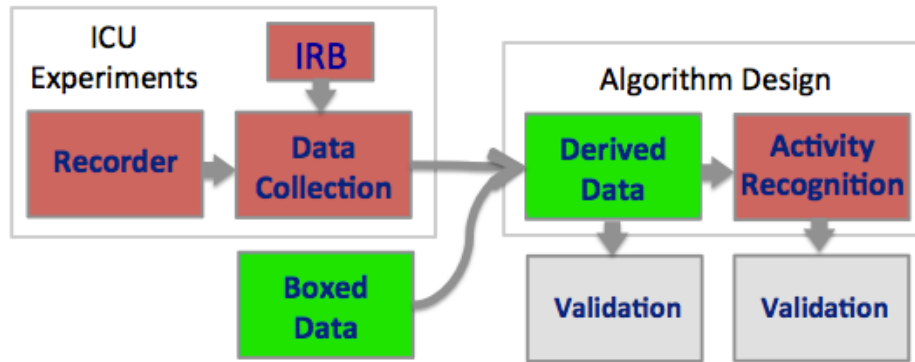
Figure 5 Dependency Graph

Project Dependencies
1. **Access to the ICU for data collection**
   Plan: A proposal has been submitted to the IRB to perform this study. It is currently tabled pending a few small changes to the language in our proposal.
2. **Meetings with mentors**
   Plan: I have emailed Dr. Fackler about a setting up a bi-weekly meeting but haven't heard back. I have also been meeting with Dr. Hager on a (roughly) bi-weekly basis.
3. **Equipment: I need two Xbox Kinects, Safety locks for the laptops being used, and a RAID drive to store all of the data.**
   Plan: APL has already donated the supplies.
4. **Alternative Kinect data in case the IRB doesn't go through**
   Plan: I have access to thousands of gesture recognition color+depth videos through the CVPR gesture recognition competition. Additionally, I can record my own video using one of the Kinects I already have.

Data Dependencies
5. **Data collected at the ICU must be de-identified**
   Plan: The recorder will incorporate face removal and patient removal features. After recording the data in the ICU we must verify that the data is completely anonymized. If it is not anonymized correctly then either the color data will be completely removed or I will develop other software to help remove faces on pre-recorded data.
6. **Developing activity recognition algorithms**
   Plan: One derived signal – gesture recognition – has already been developed, thus a more complete activity recognition system can be started
7. **Validation of our algorithms**
   Plan: Some validation has been done for the gesture recognition component. An activity recognition system must be in place before any validation can be done on that component. Most importantly, validation of our system using the ICU data cannot be completed until we do the data collection phase and we annotate the data.

Management Plan

The plan is to meet with Dr. Fackler on a bi-weekly basis (pending the aforementioned dependency) and with Dr. Hager on a routine basis. There are no standing meetings but I have been talking with Dr. Hager on a (roughly) bi-weekly basis

Code management will be done by checking

Reading List

To read

[1]   L. Fei-Fei, R. Fergus, and P. Perona, "One-shot learning of object categories.," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 4, pp. 594-611, Apr. 2006.

[2]   K. P. Murphy, "Hidden semi-Markov models (HSMMs)," 2002.

[3]   N. Padoy, "Workflow Monitoring based on 3D Motion Features," in *International Conference on Computer Vision (ICCV)*, 2009.

[4]   J. Sung, C. Ponce, B. Selman, and A. Saxena, "Human Activity Detection from RGBD Images," *Artificial Intelligence*, 2005.

Recently read

Dynamical Systems:

[1]   T. V. Duong, H. H. Bui, D. Q. Phung, S. Venkatesh, R. Ave, and M. Park, "Activity Recognition and Abnormality Detection with the Switching Hidden Semi-Markov Model," *Artificial Intelligence*.

[2]   T. V. Kasteren, A. Noulas, G. Englebienne, and B. Kr, "Accurate Activity Recognition in a Home Setting," in *UbiComp*, 2008.

[3]   L. Li and B. A. Prakash, "Time Series Clustering : Complex is Simpler !," in *International Conference on Machine Learning (ICML)*, 2011.

[4]   D. L. Vail, "Conditional Random Fields for Activity Recognition," Thesis. Carnegie Mellon University. 2008.

[5]   D. L. Vail and J. D. Lafferty, "Conditional Random Fields for Activity Recognition Categories and Subject Descriptors," in *International Conference on Autonomous Agents and Multiagent Systems*, 2007.

[6]   H. M. Wallach, "Conditional Random Fields : An Introduction," 2004.

Gestures:

[1]   A. Bigdelou, T. Benz, and N. Navab, "Simultaneous Categorical and Spatio-Temporal 3D Gestures Using Kinect," in *3DUI*, 2012.

[2]   A. Bigdelou and L. A. Schwarz, "An Adaptive Solution for Intra-Operative Gesture-based Human-Machine Interaction," in *International Conference on Intelligent User Interfaces (IUI)*, 2012.

[3]   L. A. Schwarz, A. Bigdelou, and N. Navab, "Learning gestures for customizable human-computer interaction in the operating room.," *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2011.

[4]   S. B. Wang and D. Demirdjian, "Hidden Conditional Random Fields for Gesture Recognition," *Artificial Intelligence*.

Scene analysis

[1]   L. Spinello, M. Luber, and K. O. Arras, "Tracking People in 3D Using a Bottom-Up Top-Down Detector," in *International Conference on Robotics and Automation (ICRA)*, 2011.

[2]   C. Plagemann, "Real Time Motion Capture Using a Single Time-Of-Flight Camera," in *International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.

[3]   C. Plagemann and D. Koller, "Real-time Identification and Localization of Body Parts from Depth Images," in *International Conference on Robotics and Automation (ICRA)*, 2010.

Activity recognition:

[1]   M. Hamid, "A Computational Framework For Unsupervised Analysis of Everyday Human Activities A Computational Framework For Unsupervised Analysis of Everyday Human Activities," 2008.

[2]   R. Hamid, S. Maddi, A. Bobick, and I. Essa, "Structure from Statistics - Unsupervised Activity Analysis using Suffix Trees," *International Conference on Computer Vision (ICCV)*, 2007.

[3]   M. S. Ryoo and W. Yu, "One Video is Sufficient? Human Activity Recognition Using Active Video Composition," in *Workshop on Motion and Video Computing (WMVC)*, 2011, no. January.

[4]   E. H. Spriggs, F. De La Torre, and M. Hebert, "Temporal segmentation and activity classification from first-person sensing," *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Jun. 2009.