Toward Automated Assistance for Operating Home Medical Devices

Zan Gao, Marcin Detyniecki, Ming-yu Chen, Wen Wu, Alexander G. Hauptmann, Howard D. Wactlar, Anni Cai

Computer Science Department, School of Computer Science, Carnegie Mellon University, 15213, PA, USA
zangaonsh4522@gmail.com
Today’s healthcare environment is significantly more technologically sophisticated than ever before.

The average number of medical devices at the patient’s bedside has increased from 7 devices to 26.

In fact, the FDA (Food and Drug Administration) which regulates medical devices, calls that the market of medical device industry is expected to double in size by 2011.
Introduction (2/4)

- Besides healthcare facilities, many home medical devices are now frequently used in patient’s homes.

- Home-use devices range from simple equipment such as canes and wheelchairs to sophisticated items such as glucose meters, ambulatory infusion pumps and laptop-sized ventilators.

- The rapidly growing home health industry has raised new safety concerns about devices being used in the home setting.
Introduction (3/4) ----Device Use-error rate

- Device use-errors are those which occur as a consequence of user or operator error, poor user interface design, inadequate or incomplete product labeling, inadequate or incorrect user documentation, misuse or abuse of the device, and usage problems resulting from user-device interaction.

- A.T. Lee has studied a total of 65,826 base records (incidents) received by MAUDE in the year 2002 (Lee 2003). Of these, 2,792 incidents (4.2%) were identified as having use-errors as contributing or causal factors. The overall rate of adverse consequences to patients (illness, injury, or death) resulting from all types of incidents was 33.6%. 
Introduction (4/4) --Some Home Medical Devices

Respirators    oxygen pump    ventilators    nebulizers

Dialysis machines    mobility aids    Infusion pump    apnea monitors
Motivation

Create an intelligent Automated Assistance System
- to detect and recognize activities which users are interested in
- to detect the sequence error, and warn the patients
Related work –
A recent promising approach in CV

- Spatio-temporal interest point detector
- Spatio-temporal corners
  - Spatial corners with non-constant velocity
But: Rare, too sparse for many activities
Motion SIFT Framework

- MoSIFT extends SIFT to describe video interest points
- Describe both appearance and motion explicitly
- SIFT-like representation for both appearance and motion → 256 dimensions total
- Capture smooth movements

Pair of frames → Pyramids of many scales

Local Extremes of DoG → Single Pyramid

Candidate Points → Determining motion interest points

Pair of Pyramids → Compute Optical Flow

Sufficient Motion → List of points

Get feature descriptor
MoSIFT Interest Point Detection

- SIFT (Scale Invariant Feature Transform) interest point detection
  - Gaussian pyramids to analyze multiple scales
  - Local maxima and minima of DoG to detect high contrast points
- Motions identify activities
  - Calculate optical flow over pairs of frames
  - Discard interest points without sufficient motion
MoSIFT Feature Descriptor

- Resize patch around interest point to 40x40 pixels
- Orient so that dominant gradient(s) are vertical
- Create 4x4 array of orientation histograms
- 8 orientations x 4x4 histogram array
  = 128 dimensional features
- Optical flows are described in the same way
  = 128 dimensional features
Activity recognition framework

- Local interest point method
  - Interest point detection and extraction
  - Video codebook construction & bag of video words representation
  - Activity modeling and classification
The Experimental Layout of Cameras
The location of pump and plastic infusion bag
The 22-Step Infusion Pump Operation Protocol

- Power on the pump device
- Select the correct infusion program (note: this requires multiple steps)
- Open (un-cap) the arm port (entry tube into the body)
- Clean the port with sterile pads
- Open (un-cap) pump tube
- Clean the pump tube (port) with sterile pads
- Flush the port tube with saline syringe injections
- Connect arm port tube to the pump tube
- Start the infusion process on the device
- When pump signals end of infusion, stop pump program
- Disconnect arm port tube
- Clean pump port tube with sterile pads
- Cap pump port tube
- Clean arm port with sterile pads
- Flush the arm port tube with saline syringe injections
- Flush arm port tube with Heparin anti-clotting agent syringe injections
- Clean arm port tube
- Cap arm port tube
- Turn off pump
Different related actions were grouped into 6 classes of behaviors:

- Powering the pump On/Off,
- Pressing a button,
- Cleaning a port,
- Connecting/Disconnecting the tubes,
- Injecting the contents of a syringe,
- Removing or attaching a cap to a tube end.
Pump Dataset (5/6) -- Images from different cameras
Infusion Pump Recorded Test Data

- As the initial data, we recorded 9 subjects using the device correctly once each.
- We then recorded, for test purposes, the complete operation sequence again for 3 of the previous subjects.
- And then 15 videos are recorded for 1 subject.
- The scene was recorded using 4 standard camcorders for all our 9 participants, for a total of 26 complete operations.
SIFT and MoSIFT Interest Points

SIFT

MoSIFT

Carnegie Mellon
STIP vs MoSIFT Interest Points

Laptev

MoSIFT
The Experimental Method

- **Leave-one-Person-out (LoPo)**
  We trained the action classifier on all the subjects, but left completely one of the three test subjects and tested the accuracy of the action classifier on the two sequences of that person.

- **Leave-one-Sequence-out (LoSo),**
  We trained on all sequences from all subjects except for one sequence from one of the three test subjects. The accuracy test was performed on that held-out sequence.

- **How many Videos of patients we need? (Person Adapation)**
<table>
<thead>
<tr>
<th>In Total Nine Subjects</th>
<th>In Total Nine Subjects</th>
<th>In Total Nine Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training dataset:</td>
<td>Training dataset:</td>
<td>Training dataset:</td>
</tr>
<tr>
<td>Sequences from all “1-5” subjects, “6 &amp; 7” subjects, and one sequence from “9” subject</td>
<td>All Sequences from all “1-5” subjects, “6 &amp; 7” subjects, on sequence from “8” subject, and one sequence from “9” subject</td>
<td>Sequences from all “1-5 &amp; 6-8” subjects, two sequence from “9” subjects</td>
</tr>
<tr>
<td>Test Dataset:</td>
<td>Test Dataset:</td>
<td>Test Dataset:</td>
</tr>
<tr>
<td>Two sequences from “8” subject</td>
<td>One of sequences from “8” subject</td>
<td>Five Sequences From “9” Subjects</td>
</tr>
</tbody>
</table>

In Total Nine Subjects

Training dataset:
Sequences from all “1-5” subjects, “6 & 7” subjects, and one sequence from “9” subject
Test Dataset:
Two sequences from “7” subject

Training dataset:
Sequences from all “1-5” subjects, “7 & 8” subjects, and one sequence from “9” subject
Test Dataset:
Two sequences from “6” subject

Training dataset:
Sequences from all “1-5” subjects, “7 & 8” subjects, on sequence from “6” subject, and one sequence from “9” subject
Test Dataset:
One of sequences from “6” subject

Training dataset:
Sequences from all “1-5 & 6-8” subjects, one sequence from “9” subject
Test Dataset:
One of sequences from “9” subject

Training dataset:
Sequences from all “1-5 & 6-8” subjects, 10 sequence from “9” subject
Test Dataset:
Five Sequences From “9” Subjects

LoPo
LoSo
PersAdapt
## The Average Performance of Classifiers

<table>
<thead>
<tr>
<th>Action Class</th>
<th>Freq Count</th>
<th>Baseline (random)</th>
<th>Avg LoPo Accuracy</th>
<th>Avg LoSo Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power On/Off</td>
<td>12</td>
<td>0.08</td>
<td>0.44</td>
<td>0.77</td>
</tr>
<tr>
<td>Push button</td>
<td>49</td>
<td>0.34</td>
<td>0.94</td>
<td>0.96</td>
</tr>
<tr>
<td>Open/CapArm/Pump</td>
<td>24</td>
<td>0.17</td>
<td>0.69</td>
<td>0.74</td>
</tr>
<tr>
<td>Flush Green/Yellow</td>
<td>16</td>
<td>0.11</td>
<td>0.48</td>
<td>0.67</td>
</tr>
<tr>
<td>Clean Arm/Pump</td>
<td>31</td>
<td>0.22</td>
<td>0.45</td>
<td>0.50</td>
</tr>
<tr>
<td>Connect/Disconnect</td>
<td>12</td>
<td>0.08</td>
<td>0.06</td>
<td>0.00</td>
</tr>
<tr>
<td>(Total)Average</td>
<td>144</td>
<td>0.17</td>
<td>0.51</td>
<td>0.61</td>
</tr>
</tbody>
</table>
The Performance from Different Cameras (LoSo)

<table>
<thead>
<tr>
<th>Camera Action</th>
<th>Above</th>
<th>Front</th>
<th>Side</th>
<th>Very High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power On/Off</td>
<td>0.75</td>
<td>0.83</td>
<td>0.67</td>
<td>0.83</td>
</tr>
<tr>
<td>Push button</td>
<td>1.00</td>
<td>0.96</td>
<td>0.96</td>
<td>0.92</td>
</tr>
<tr>
<td>Open/Cap Arm/Pump</td>
<td>0.79</td>
<td>0.75</td>
<td>0.83</td>
<td>0.58</td>
</tr>
<tr>
<td>Flush Green/Yellow</td>
<td>0.75</td>
<td>0.56</td>
<td>0.88</td>
<td>0.50</td>
</tr>
<tr>
<td>Clean Arm/Pump</td>
<td>0.45</td>
<td>0.45</td>
<td>0.58</td>
<td>0.52</td>
</tr>
<tr>
<td>Connect/Disconnect</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td>0.62</td>
<td>0.59</td>
<td>0.65</td>
<td>0.56</td>
</tr>
</tbody>
</table>
The best camera and fusion scheme (LoSo)

<table>
<thead>
<tr>
<th>Action Class</th>
<th>Best Camera (Side)</th>
<th>Best Camera per Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power On/Off</td>
<td>0.67</td>
<td>0.83</td>
</tr>
<tr>
<td>Push button</td>
<td>0.96</td>
<td>1.00</td>
</tr>
<tr>
<td>Open/Cap Arm/Pump</td>
<td>0.83</td>
<td>0.83</td>
</tr>
<tr>
<td>Flush Green/Yellow</td>
<td>0.88</td>
<td>0.88</td>
</tr>
<tr>
<td>Clean Arm/Pump</td>
<td>0.58</td>
<td>0.58</td>
</tr>
<tr>
<td>Connect/Disconnect</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Average</td>
<td>0.65</td>
<td>0.69</td>
</tr>
</tbody>
</table>
The confusion matrix of best camera for each action in the LoSo

<table>
<thead>
<tr>
<th>Recognized as</th>
<th>Power On/Off</th>
<th>Push button</th>
<th>Open/Cap Arm/Pump</th>
<th>Flush Green/Yellow</th>
<th>Clean Arm/Pump</th>
<th>Connect/Disconnect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Truth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Power On/Off</td>
<td>0.83</td>
<td>0.00</td>
<td>0.08</td>
<td>0.08</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Push button</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Open/Cap Arm/Pump</td>
<td>0.04</td>
<td>0.04</td>
<td>0.83</td>
<td>0.00</td>
<td>0.08</td>
<td>0.00</td>
</tr>
<tr>
<td>Flush Green/Yellow</td>
<td>0.00</td>
<td>0.00</td>
<td>0.13</td>
<td>0.88</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Clean Arm/Pump</td>
<td>0.13</td>
<td>0.00</td>
<td>0.23</td>
<td>0.06</td>
<td>0.58</td>
<td>0.00</td>
</tr>
<tr>
<td>Connect/Disconnect</td>
<td>0.08</td>
<td>0.17</td>
<td>0.67</td>
<td>0.00</td>
<td>0.08</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Person Adaptation

The graph shows the accuracy of different frontiers as the number of adding videos increases. The y-axis represents accuracy, ranging from 0 to 100, and the x-axis represents the number of adding videos, ranging from 1 to 10. Different lines represent different frontiers: Front1, Front2, Front3, Front4, Front5, and Average.
Conclusion

- The presented approach (motion interest points combined with Square-2-kernelized SVM) provides good performance, with a general (single camera) recognition average of 61%.

- The performances from different cameras are similar, so 1-2 camera should be enough.

- If we can get some videos from the patients, it will improve the performance. (the average accuracy is 85%)
Future work

- Real time error detection

- Giving the reminding and warning

- Use these kinds of method to other types of home medical devices, with perhaps different required actions and protocols.