PEDESTRIAN DETECTION WITH THE MICROSOFT KINECT

NATMEC, 2014

YINHAI WANG
YINHAI@UW.EDU

XIAOFENG CHEN
CHENXF@U.WASHINGTON.EDU

KRISTIAN HENRICKSON
HENR2237@UW.EDU
HUMAN DETECTION: NOT JUST FOR DEMAND ESTIMATION

• Estimate demand for:
  • infrastructure investment
  • Safety treatments
• Analyze pedestrian movement and interaction with public spaces
• Actuated pedestrian signals
• Advertising
• Vehicle on-board pedestrian avoidance features
HOW ARE PEDESTRIANS AND CYCLISTS DETECTED?

- Manual count
- Pedestrian push buttons
- Infrared
- Inductance loops
- Pressure and acoustic mats
- Video image processing
CURRENT STATE OF VIDEO IMAGE PROCESSING

- Human detection in video imagery is a long-standing computer vision challenge
- A great deal of current work is focused on feature-based detection
- Train machine learning classifiers for identifying local image features corresponding to humans or body parts
- Example: Histogram of Oriented Gradients (HOG)
- A number of algorithms have been developed for resolving occlusion, still a persistent challenge
PEDESTRIAN DETECTION UNDER OCCLUSION

Easy detection
- Face and limbs clearly visible
- Distinct from background

Difficult detection
- Obscured by environment or other people
- Noisy environment
MICROSOFT KINECT® SOLUTION?

- **INFRARED DEPTH SENSOR ARRAY**
- **RGB (COLOR) CAMERA**
- **MOTORIZED TILTING BASE**

**ALSO:** ACCELEROMETER AND MICROPHONE ARRAY
KINECT SPECIFICATIONS

• 43° and 57 ° vertical and horizontal field of view respectively
• 30 frames per second (FPS) depth and color streams
• Structured light depth sensing technology
• 640 x 480 color images, up to 1280 x 960 at reduced frame rate
• 320 x 240 depth images
• 4 microphones in directional array
• 2G accelerometer range with 1° upper limit accuracy
• Microsoft SDK available for windows, open source development tools also available
• Version 1 Cost: $150.00 - $200.00
RECENT WORK IN DEPTH-BASED HUMAN DETECTION

• Feature based
  – Histogram of Oriented Depth (HOD)\textsuperscript{1,2}
  – Histogram of Depth Difference (HDD)\textsuperscript{3}
  – Augmented Histogram of Oriented Gradients (HOG)\textsuperscript{4,2}
  – Part based depth feature descriptors

• Microsoft Skeleton tracking algorithm: for gaming interface\textsuperscript{5}

• Current work not proven in crowded environments where occlusion occurs frequently
OUR APPROACH TO RGB-D HUMAN DETECTION

• Background subtraction to extract pedestrian contours from RGB image – simple and well studied
• Morphological processing to reduce noise and clutter in binary image
• Fuse RGB and Depth images
• Search for depth discontinuities within pedestrian blobs to resolve occlusion
• Pattern matching for people tracking
• Update count when people cross a depth threshold
• Implemented in C# with EMGU OpenCV 2.4 and Microsoft Kinect SDK 1.6
Pedestrians Counts: 15

STARLab, UW, Jun 2010
Scene 2:
Staircase landing

Scene 3:
Wide open courtyard

Note depth difference in occlusion instance
## TESTING RESULTS

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Test length</th>
<th>Manual counts</th>
<th>Under counting</th>
<th>Over counting</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5 min</td>
<td>56</td>
<td>0</td>
<td>3</td>
<td>94.7</td>
</tr>
<tr>
<td>2</td>
<td>5 min</td>
<td>60</td>
<td>4</td>
<td>0</td>
<td>93.3</td>
</tr>
<tr>
<td>3</td>
<td>5 min</td>
<td>58</td>
<td>4</td>
<td>0</td>
<td>93.1</td>
</tr>
</tbody>
</table>

Scene 1: STAR Lab, cluttered indoor scene

Scene 2: Staircase landing direct sunlight

Scene 3: Open courtyard cloudy
WHAT HAVE WE ACCOMPLISHED?

• Developed a RGB-D pedestrian detector using a low-cost consumer grade sensor
• Address the occlusion issue by fusing depth and color images
• Demonstrated good counting accuracy in both indoor and outdoor environments
• Demonstrated the utility of the Kinect outside of the manufacturer specified distance range
Sensor locations

Possible Applications?
FUTURE WORK

- Adapt current algorithm to measure speed
- Differentiate between travel modes (i.e. walk, bike)
- Investigate applications for new generation of consumer 3-D sensors
  - Kinect Version 2
  - Prime Sense Capri
- Other detection scenarios
  - Lingering crowd detection
  - Pedestrian presence detection for actuated signals
KINECT V2 SPECIFICATIONS

• Most notable: time of flight IR depth sensing technology
• Active IR technology for improved performance in varying light conditions
• 60° and 70 ° vertical and horizontal field of view respectively
• Full HD 1920 x 1080 color images at 30 FPS
• 512 x 424 Depth stream at 30 FPS
• Reduced latency and noise, increased useable depth range compared to v1
• Non-motorized adjustable tilt
• Microsoft SDK available soon for Windows
• Version 2 cost: $200.00
THANK YOU!

This work was supported by
The Pacific Northwest Transportation Consortium (PacTrans)

FOR MORE INFORMATION CONTACT:

YINHAI WANG
YINHAI@UW.EDU

XIAOFENG CHEN
CHENXF@U.WASHINGTON.EDU

KRISTIAN HENRICKSON
HENR2237@UW.EDU
IMAGE CREDITS

• Manual pedestrian Count: https://www.flickr.com/photos/yoavlerman/
• Bike counter: https://www.flickr.com/photos/wv/
• Pedestrian pushbutton: https://www.flickr.com/photos/katsrcool/
• Pedestrian counter: https://www.flickr.com/photos/giltay/
• Occlusion: https://www.flickr.com/photos/frerieke/
• Mall scene: https://www.flickr.com/photos/postsumptio/
• Burke Gilman trail: Google Earth


