Look At This One : Detection sharing between modality-independent classifiers for robotic discovery of people

Joris Guerry¹, Bertrand Le Saux¹, David Filliat²

¹ONERA - DTIS, ²ENSTA Paristech - U2IS

During an autonomous exploration, a robot can find itself in very different situations. This variability has a direct impact over the images acquired by the embedded cameras which can lower the object detection performances. In particular, people detection becomes harder because of its complex aspects: modality of the studied object, global situation (e.g. in the dark), point of view (low-angle shot camera from small robots), sunshine...

ONERA ROOM : A NEW DATASET

ONERA ROOM is a new data set with 27 sequences captured by several RGBD sensors (Kinect v1, Xtion and RealSense):
- 15 sequences containing labeled people, representing 1213 Regions of Interest (ROIs),
- some sequences acquired in the dark,
- some sequences blurred by the robot motion,
- cases of unconscious people on the ground.

Available for downloading at: http://joriuguerry.fr/ONERA.ROOM

Parking lot with sunshine noise onto depth map

The environment is too dark because of the luminosity contrast with the previous room

RGB modality loss

The yellow curve represents the mean intensity of the image. We can see from left to right the RGB image and the RGB expert detections, the Depth image and the Depth expert detections, and finally, the X-fusion detections.

RESULTS ONERA.ROOM

Comparison with the state of the art

<table>
<thead>
<tr>
<th>Method</th>
<th>Source</th>
<th>Precision/Recall</th>
<th>AP</th>
<th>EER</th>
<th>IoU</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>RGB-D</td>
<td>[3]</td>
<td>87.4</td>
<td>80.3</td>
<td>82.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RGB-D</td>
<td>[2]</td>
<td>87.2</td>
<td>80.4</td>
<td>82.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RGB-D</td>
<td>[1]</td>
<td>87.2</td>
<td>80.4</td>
<td>82.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HOD</td>
<td>[3]</td>
<td>87.2</td>
<td>80.3</td>
<td>82.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RGB-D-Optical flow</td>
<td>[1]</td>
<td>87.2</td>
<td>80.4</td>
<td>82.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HGE</td>
<td>[3]</td>
<td>87.2</td>
<td>80.3</td>
<td>82.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gating Net.</td>
<td>[2]</td>
<td>RGB-D-Optical flow</td>
<td>89.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fr RCNN [1]</td>
<td>D</td>
<td>87.2</td>
<td>80.4</td>
<td>82.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fr RCNN [1]</td>
<td>RGB</td>
<td>87.2</td>
<td>80.4</td>
<td>82.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fr RCNN [1]</td>
<td>X</td>
<td>87.2</td>
<td>80.4</td>
<td>82.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fr RCNN [1]</td>
<td>RGB-D</td>
<td>87.2</td>
<td>80.4</td>
<td>82.4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

It is interesting to notice that the X-fusion (left column, last line) has a new ROIs. This is allowed by the intermediate NMS post-RPN: the Depth expert found an object and was unable to classify it, however, it indicated it to the RGB expert who was able to classify it even if it did not initially find this object. The U-fusion (right column) does not allow this pre-classification detection exchange but allows to reorder, by score, the ROIs after the final NMS. In this example, the ROI from the Depth expert is dismissed in favor of the RGB expert central ROI, which allows the Depth expert to provide a previously ignored ROI (in yellow).

RESULTS RGBD PEOPLE

Comparison with the state of the art

<table>
<thead>
<tr>
<th>Method</th>
<th>Source</th>
<th>Precision/Recall</th>
<th>AP</th>
<th>EER</th>
<th>IoU</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>RGB-D</td>
<td>[3]</td>
<td>87.4</td>
<td>80.3</td>
<td>82.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RGB-D</td>
<td>[2]</td>
<td>87.2</td>
<td>80.4</td>
<td>82.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RGB-D</td>
<td>[1]</td>
<td>87.2</td>
<td>80.4</td>
<td>82.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HOD</td>
<td>[3]</td>
<td>87.2</td>
<td>80.3</td>
<td>82.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RGB-D-Optical flow</td>
<td>[1]</td>
<td>87.2</td>
<td>80.4</td>
<td>82.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HGE</td>
<td>[3]</td>
<td>87.2</td>
<td>80.3</td>
<td>82.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gating Net.</td>
<td>[2]</td>
<td>RGB-D-Optical flow</td>
<td>89.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fr RCNN [1]</td>
<td>D</td>
<td>87.2</td>
<td>80.4</td>
<td>82.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fr RCNN [1]</td>
<td>RGB</td>
<td>87.2</td>
<td>80.4</td>
<td>82.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fr RCNN [1]</td>
<td>X</td>
<td>87.2</td>
<td>80.4</td>
<td>82.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fr RCNN [1]</td>
<td>RGB-D</td>
<td>87.2</td>
<td>80.4</td>
<td>82.4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

FUTURE APPLICATIONS

Mobility refinement (drones, autonomous car)

Assistance

REFERENCES


CONCLUSION

— Classifiers independence → robust to modality loss & independent trainings
— Performance gain by sharing detections
— A new dataset : ONERA.ROOM
— New state of the art results on the InOutPeople and RGBD People datasets

REALITY OF THE SITUATION

(3-2) solution

RGBD camera

the active Kinect c

fusion:

a strategy

modalities

robust

(as

The environment is too dark because of the luminosity contrast with the previous room

NIPS, 2015.

modality-independent classifiers for robotic discovery of people

ISCOINT:

and finally, the X-fusion detections.

Depth image and the Depth expert detections, the

RGB image and the RGB expert detections, the

ONERA.ROOM : a new dataset

Context :

RGB-D Detector

The basis : Faster R-CNN [1]

Fusion strategies: how to look at the good one?

U

X

Y

RGB MODALITY LOSS

Look at these examples

Look At This One : Detection sharing between modality-independent classifiers for robotic discovery of people

Joris Guerry¹, Bertrand Le Saux¹, David Filliat²

¹ONERA - DTIS, ²ENSTA Paristech - U2IS

During an autonomous exploration, a robot can find itself in very different situations. This variability has a direct impact over the images acquired by the embedded cameras which can lower the object detection performances. In particular, people detection becomes harder because of its complex aspects: modality of the studied object, global situation (e.g. in the dark), point of view (low-angle shot camera from small robots), sunshine...

Our solution : an active RGBD camera (as the Kinect®) - a robust modalities fusion strategy

RESULTS ONERA.ROOM (see below)

X-fusion allows to detect people where both experts can't, with instance differentiation.

X-fusion improves the ROIs position and is able to detect people on the ground in the dark.

RESULTS RUBG PEOPLE

Comparison with the state of the art

From left to right : RGB, Depth, X-Fusion

Parking lot with sunshine noise onto depth map

The environment is too dark because of the luminosity contrast with the previous room

ONERA ROOM : A NEW DATASET

ONERA ROOM is a new data set with 27 sequences captured by several RGBD sensors (Kinect v1, Xtion and RealSense):
- 15 sequences containing labeled people, representing 1213 Regions of Interest (ROIs),
- some sequences acquired in the dark,
- some sequences blurred by the robot motion,
- cases of unconscious people on the ground.

Available for downloading at: http://joriuguerry.fr/ONERA.ROOM

Parking lot with sunshine noise onto depth map

The environment is too dark because of the luminosity contrast with the previous room

ONERA ROOM : A NEW DATASET

ONERA ROOM is a new data set with 27 sequences captured by several RGBD sensors (Kinect v1, Xtion and RealSense):
- 15 sequences containing labeled people, representing 1213 Regions of Interest (ROIs),
- some sequences acquired in the dark,
- some sequences blurred by the robot motion,
- cases of unconscious people on the ground.

Available for downloading at: http://joriuguerry.fr/ONERA.ROOM