RobotVision@ImageCLEF 2014 Task overview

Organizers: J. Martínez Gómez, I. García-Varea, M. Cazorla and V. Morell
The behaviour of a robot \( r \) at timestamp \( t \) depends on:

- Not the topological \(<x,y,z>\) location
- ... But
- the semantic category of the place
- &
- the objects that are suitable for manipulation
Robot Vision - Motivation

I am at the kitchen and I can see a Fridge and a table.

I am placed at position $<1.52, 2.57, 3.51>$.
Robot Vision - The Task

- Proposed in 2009 (6th edition)
  - ImageCLEF 2009 – Greece
  - ImageCLEF@ICPR 2010 – Turkey
  - ImageCLEF 2010 – Italy
  - ImageCLEF 2012 – Italy
  - ImageCLEF 2013 – Spain
  - ImageCLEF 2014 – England

- Considerable attention
  - ImageCLEF 2009 – 7 groups
  - ImageCLEF@ICPR 2010 – 9 groups
  - ImageCLEF 2010 – 7 groups
  - ImageCLEF 2012 – 8 groups
  - ImageCLEF 2013 – 6 groups
  - ImageCLEF 2014 – 4 groups

- Organizers
  - Jesus Martínez Gómez, Ismael García Varea, Miguel Cazorla and Vicente Morell
Robot Vision - The Task

Evolution on the Participant Number

<table>
<thead>
<tr>
<th>Year</th>
<th>Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td>7</td>
</tr>
<tr>
<td>2010a</td>
<td>9</td>
</tr>
<tr>
<td>2010b</td>
<td>7</td>
</tr>
<tr>
<td>2012</td>
<td>8</td>
</tr>
<tr>
<td>2013</td>
<td>6</td>
</tr>
<tr>
<td>2014</td>
<td>4</td>
</tr>
</tbody>
</table>
# Robot Vision - Organizers

<table>
<thead>
<tr>
<th>Organizer</th>
<th>2009</th>
<th>2010</th>
<th>2010</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
</tr>
</thead>
<tbody>
<tr>
<td>B. Caputo</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>A. Pronobis</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>P. Jensfelt</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>H.I. Christensen</td>
<td>-</td>
<td>X</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>M. Fornoni</td>
<td>-</td>
<td>-</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>J. Martínez-Gómez</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>I. García-Varea</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>M. Cazorla</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>V. Morell</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>X</td>
</tr>
</tbody>
</table>
Multimodal information retrieval
- Two sources of information
  - Visual Images
  - Range Images
- Two problems to solve
  - Presence or lack of objects in the scene
  - Semantic category of the scene

In between computer vision and robotics
Robot Vision – The Task

- Supervised classification problem
  - Participants are provided with labelled training sequences

- Each training frame contains
  - Visual Image
  - Range Image (.pcd format)
  - Semantic category of the scene where the frame was acquired from
  - List of objects appearing in the scene

- Training and test sequences
  - Different buildings but with similar structure
Robot Vision – The Task

- The problem
  - Place classification and object detection
  - Question a. - where are you?
  - Question b. - list the objects that are in the scene

- 10 room categories
  - Corridor, Hall, Professor Office, Student Office, Technical Room, Toilet, Secretary, Visio conference, Elevator area area and Warehouse

- 8 objects
  - Extinguisher, Phone, Chair, Printer, Urinal, Bookshelf, Trash and Fridge
Robot Vision – The Task

- **Performance evaluation**
  - **Place classification** – 1 nominal problem
    - Category correctly classified → +1 point
    - Category wrongly classified → -0.5 point
    - Category not classified → +0.0 points
  - **Object detection** – 8 binary problems
    - Each object correctly detected (TP) → +1.0 points
    - Each object incorrectly detected (FP) → -0.25 points
    - Each object detected as not present (TN) → +0.0 points
    - Each object incorrectly detected as not present (FN) → -0.25 points
Robot Vision – The Task

Corridor  Hall  ProfessorOffice  StudentOffice

TechnicalRoom  Toilet  Secretary  VisioConference

ElevatorArea  Warehouse
Robot Vision – The Task

Exting.  Chair  Printer  Bookshelf  Urinal  Trash  Phone  Fridge
# Robot Vision – Major Changes

<table>
<thead>
<tr>
<th>Topics</th>
<th>2009</th>
<th>2010</th>
<th>2010</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
</tr>
</thead>
<tbody>
<tr>
<td>Place Classification</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>2 subtasks</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Stereo Images</td>
<td>-</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Unknown Labels</td>
<td>-</td>
<td>X</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Object Detection</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>3D Images</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>New Buildings (test)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>X</td>
</tr>
</tbody>
</table>
Robot Vision – Performance Eval

- Test frame
  - Real labels (not provided)
    - Class: Technical Room
    - Objects present: Phone and Printer

- Participant decision
  - Class: Technical Room
  - Objects present: Phone and Trash
  - Objects not present: Extinguisher, Chair, Printer, Urinal, Booshelf and Fridge
# Robot Vision – Performance Eval

<table>
<thead>
<tr>
<th>Class</th>
<th>Exting.</th>
<th>Phone</th>
<th>Chair</th>
<th>Printer</th>
<th>Urinal</th>
<th>Books</th>
<th>Trash</th>
<th>Fridge</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Real labels</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technical Room</td>
<td>NO</td>
<td>YES</td>
<td>NO</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
</tr>
</tbody>
</table>

| **Participant Decision** |         |       |       |         |        |       |       |        |
| Technical Room           | NO      | YES   | NO    | NO      | NO     | NO    | YES   | NO     |

| **Right/Wrong Classification** |         |       |       |         |        |       |       |        |
| Hit                        | TN      | TP    | TN    | FN      | TN     | TN    | FP    | TN     |

| **Points**          |         |       |       |         |        |       |       |        |
| +1.0                 | 0.0     | +1.0  | 0.0   | -0.25   | 0.0    | 0.0   | -0.25 | 0.0    |

Total: \(1.0 + 1.0 -0.25 - 0.25 = 1.5\)
Unreleased frames from the ViDRILLO dataset (to be published)

- 3 sequences
  - Training: 5000 frames
  - Validation: 1500 frames
  - Test: 3000 frames
Robot Vision – The Data

- Unreleased frames from the ViDRIL dataset (to be published)
  - Extreme lighting conditions in the test sequence
    - Range images are highly recommended
## Robot Vision – The Data

<table>
<thead>
<tr>
<th>Room Category</th>
<th>Training (Building A)</th>
<th>Validation (Build. A+B)</th>
<th>Test (Build. B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corridor</td>
<td>1833</td>
<td>479</td>
<td>772</td>
</tr>
<tr>
<td>Hall</td>
<td>306</td>
<td>103</td>
<td>202</td>
</tr>
<tr>
<td>Professor Office</td>
<td>355</td>
<td>149</td>
<td>372</td>
</tr>
<tr>
<td>Student Office</td>
<td>498</td>
<td>174</td>
<td>419</td>
</tr>
<tr>
<td>Technical Room</td>
<td>437</td>
<td>110</td>
<td>242</td>
</tr>
<tr>
<td>Toilet</td>
<td>389</td>
<td>094</td>
<td>141</td>
</tr>
<tr>
<td>Secretary</td>
<td>336</td>
<td>102</td>
<td>245</td>
</tr>
<tr>
<td>Visioconference</td>
<td>364</td>
<td>113</td>
<td>159</td>
</tr>
<tr>
<td>Warehouse</td>
<td>174</td>
<td>081</td>
<td>201</td>
</tr>
<tr>
<td>Elevator Area</td>
<td>308</td>
<td>095</td>
<td>247</td>
</tr>
<tr>
<td>All</td>
<td><strong>5000</strong></td>
<td><strong>1500</strong></td>
<td><strong>3000</strong></td>
</tr>
</tbody>
</table>
Robot Vision – The Data

Training
- Corridor: 37%
- Hall: 6%
- Professor Office: 6%
- Students Office: 7%
- Technical Room: 10%
- Toilet: 7%
- Secretary: 9%
- Visio Conference: 7%
- Warehouse: 8%
- Elevator Area: 3%

Validation
- Corridor: 32%
- Hall: 7%
- Professor Office: 6%
- Students Office: 10%
- Technical Room: 7%
- Toilet: 12%
- Secretary: 6%
- Visio Conference: 5%
- Warehouse: 6%
- Elevator Area: 7%

Test
- Corridor: 26%
- Hall: 7%
- Professor Office: 8%
- Students Office: 5%
- Technical Room: 12%
- Toilet: 14%
- Secretary: 8%
- Visio Conference: 5%
- Warehouse: 8%
- Elevator Area: 6%
## Robot Vision – The Data

<table>
<thead>
<tr>
<th>Objects</th>
<th>Training</th>
<th>Validation</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extinguisher</td>
<td>770 / 4230</td>
<td>238 / 1262</td>
<td>566 / 2434</td>
</tr>
<tr>
<td>Chair</td>
<td>1304 / 3696</td>
<td>471 / 1029</td>
<td>1070 / 1930</td>
</tr>
<tr>
<td>Printer</td>
<td>473 / 4527</td>
<td>139 / 1361</td>
<td>265 / 2735</td>
</tr>
<tr>
<td>Bookshelf</td>
<td>802 / 4198</td>
<td>317 / 1183</td>
<td>896 / 2104</td>
</tr>
<tr>
<td>Urinal</td>
<td>162 / 4838</td>
<td>040 / 1460</td>
<td>060 / 2940</td>
</tr>
<tr>
<td>Trash</td>
<td>813 / 4187</td>
<td>323 / 1177</td>
<td>797 / 2203</td>
</tr>
<tr>
<td>Phone</td>
<td>267 / 4733</td>
<td>113 / 1387</td>
<td>303 / 2697</td>
</tr>
<tr>
<td>Fridge</td>
<td>190 / 4810</td>
<td>034 / 1466</td>
<td>047 / 2953</td>
</tr>
<tr>
<td><strong>All</strong></td>
<td>4781 / 35219</td>
<td>1675 / 10325</td>
<td>4004 / 19996</td>
</tr>
</tbody>
</table>
Robot Vision – The Data

**Training**
- Extinguisher: 17%
- Chair: 16%
- Printer: 27%
- Bookshelf: 10%
- Urinal: 4%
- Trash: 3%
- Phone: 6%
- Fridge: 2%

**Validation**
- Extinguisher: 19%
- Chair: 14%
- Printer: 28%
- Bookshelf: 8%
- Urinal: 7%
- Trash: 2%
- Phone: 2%
- Fridge: 3%

**Test**
- Extinguisher: 16%
- Chair: 17%
- Printer: 27%
- Bookshelf: 10%
- Urinal: 6%
- Trash: 4%
- Phone: 3%
- Fridge: 17%
Robot Vision – Participation

- 28 participants but only 4 submitted one run
  - NUDT: National University of Defense Technology, Changsha, China
  - UFMS CPPP: Federal University of Mato Grosso do Sul, Ponta Pora, Brazil
  - AEGEAN: University of the Aegean Karlovassi, Greece
  - SIMD: University of Castilla-La Mancha, Spain
    - Out of competition contribution using proposed techniques
Robot Vision – Script

The sequences of the challenge are released with a useful MATLAB script that can be used as template for participant proposals, including:

- Features extraction
- Descriptor generation
- Training
- Classification
- Performance evaluation
## Robot Vision – Overall Results

- Ranking of the best run submitted by group

<table>
<thead>
<tr>
<th>Rank</th>
<th>Group Name</th>
<th>Score Class</th>
<th>Score Object</th>
<th>Score Total (% Max)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>NUDT</td>
<td>1075.5</td>
<td>3354.75</td>
<td>4430.25 (63.25)</td>
</tr>
<tr>
<td>2</td>
<td>UFMS</td>
<td>219.0</td>
<td>1519.75</td>
<td>1738.75 (24.83)</td>
</tr>
<tr>
<td>3</td>
<td>SIMD*</td>
<td>67.5</td>
<td>186.25</td>
<td>253.75 (3.62)</td>
</tr>
<tr>
<td>4</td>
<td>AEGEAN</td>
<td>-405</td>
<td>-995.00</td>
<td>-1400 (&lt;0)</td>
</tr>
</tbody>
</table>

- Winner: NUDT group
Robot Vision – Detailed Results

![Comparison Chart]

- **NUDT**: High in Objects, moderate in Rooms.
- **UFMS**: Moderate in Objects, low in Rooms.
- **SIMD**: Low in both Objects and Rooms.
- **AEGEAN**: Low in Objects, high in Rooms.

Legend:
- **Objects**
- **Rooms**
Room classification remains as an open problem when generalization is requested
  - Current approaches
    - Perform well when the test environment has been previously image (past editions)

Object recognition is not affected by environment changes
Robot Vision – Proposals

- MIAR-ICT
  - Ranked 1st
  - Visual images
  - Bag of Words $\rightarrow$ SIFT and PHOG
  - Multiclass SVN (one versus all)
Robot Vision – Proposals

- UFMS
  - Ranked 2nd
  - Visual images
  - Bag of Words → dense SIFT descriptors
  - K-nearest neighbors
    - Classification
Again ... less attention than expected

None of the groups used 3D images

Problems with room classification

Room classification and object recognition problems were managed separately
RobotVision@ImageCLEF 2014 Task overview

Organizers: J. Martínez Gómez, I. García-Varea, M. Cazorla and V. Morell