Motivations of this work

- **Business Analytics:** Emerging field of application of video surveillance to monitor stores to gather useful statistics (not for security reason), which may help owners in deciding the best layout for the store and finding the bestselling products.

- **Existing techniques are inadequate:** Standard infrared beams present shortcomings: limitation to entry and exit points; not easy to track people (no flow analysis).

- **Use of computer vision:** CV can overcome these limitations by permitting a continuous collection of statistics about the number and the position of people/customers in the store. Moreover, it allows to distinguish between adult people (i.e., potential customers) and other objects (children, trolleys, strollers, animals, etc.).

Overview of the method

- **Phase 1 – Depth computation:** thanks to a calibrated stereo camera the depth map is computed using the semi-global block-matching algorithm described in [1].

- **Phase 2 – 2D Object segmentation:**
  a) The depth map is then mapped on the ground plane to obtain the D-BEV (Depth Bird-Eye View). The number of pixels mapped to a given position (x, y) is counted and thresholded to discard both noise and too-small objects which are of no interest for our purposes.
  b) To remove static objects from D-BEV, a background suppression method based on median temporal modelling is employed. After background differencing, the resulting D-BEV is processed with morphological operations. The size of the structuring elements is adapted to the depth. The remaining points are back-projected in 3D.

- **Phase 3 – 3D Object segmentation:** The resulting 3D point cloud is further segmented directly in 3D. Since the current number of objects is unknown, a mean shift clustering with a cylindrical kernel is performed.

- **Phase 4 – People classification:** Distinguish between “adult person” or “other”. Fitting of a cylindrical model to each object’s point cloud. This fitting considers both the volume of the point cloud and the spatial distribution of points within the cylinder. Pseudo-code is presented in the following.

References

### Pseudo-code for people classification

```
Require: step, threshold, r_p, h_p, PCL
Ensure: class_people

j ← 0
for all object ∈ PCL do
    dir ← PCA(x, y)
    centers ← sample_points(dir, step)
    i ← 0
    for all (x, y) ∈ centers do
        cylinder ← create_cylinder(x, y, 0, r_p, h_p)
        filling[i] ← object_filling_space(object, cylinder)
        people[i] ← object ∩ cylinder
        i ← i + 1
    end for
    best_fill ← max(value(filling))
    best_fill_index ← max_index(filling)
    if best_fill ≥ threshold then
        class_people[j] ← people[best_fill_index]
        j ← j + 1
    end if
end for
```

### Phase 1 – Depth computation

![Example of D-BEV](image)

### Phase 2 – 2D Object Segmentation

**Example of D-BEV**

![D-BEV background subtraction](image)

**Back-projection on 3D**

### Phase 3 – 3D Object Segmentation

### Phase 4 – People classification

- **Point-Cloud**
- **PCA**
- **Fitting**
- **Montecarlo**
- **Best fit**

### Experimental results

**Accuracy results**

<table>
<thead>
<tr>
<th>Method</th>
<th>Setup 1: Faculty hall</th>
<th>Setup 2: Store</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Video 1</td>
<td>Video 2</td>
</tr>
<tr>
<td>Method [24]</td>
<td>82.94%</td>
<td>76.14%</td>
</tr>
<tr>
<td>Method [24]</td>
<td>95.95%</td>
<td>98.04%</td>
</tr>
<tr>
<td>Our method</td>
<td>88.85%</td>
<td>81.08%</td>
</tr>
<tr>
<td>Our method</td>
<td>97.58%</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

**Average F-measure in comparison with [24]**

### Conclusions

- New method for detecting objects in complex scenes and to classify adult people among them.
- Approach is robust to occlusions and partial or tilted view of a person.
- Real-time performance (more than 10 fps) in Matlab.
- **Future direction**: object tracking (given the good results of the detection, simple methods can be used).

### References