Counting People Waiting in Service Lines Using Computer Vision and Machine Learning Techniques

Domingo Mery(1), Enrique Sucar(2), Alvaro Soto(1)

(1) Department of Computer Science, Pontificia Universidad Católica, Chile
(2) Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE), Mexico.
Our Problem: Counting People at Service Lines in Grocery Stores
Our Solution:

Computer Vision
+
Machine Learning
Computer Vision: Template Matching
Computer Vision: Problems

- Changes in illumination
- Pose variations
- Scale variations
- ...

![Image showing changes in illumination, pose, and scale variations]
Computer Vision: Geometric Models

Model of Car

Image
Computer Vision: Problems

- Changes in illumination
- Pose variations
- Scale variations
- Intra-class variations
- Occlusion
- Deformations
- Background clutter
- ...

Model of Car

Image
Main goal is to learn a concept (task) by finding a model that minimizes expected loss using observed data.
Machine Learning: Training Data
Use of information invariant to some visual problems.
Use of statistical information.
Image Segmentation vs Sliding Window

Image Segmentation

Sliding Window
Approach

- **ML Algorithm**
  - Bias
  - Hipothesis space
  - Search scheme

- Training Set

- Task

- Loss function
New Visual Feature: Saliency Map

Center Surround Patterns:
Biological Cells

Center Surround Patterns:
Approximation
New Visual Feature: Saliency Map
New Visual Feature: Saliency Map

a) Itti et al, 1998 (iVVT)  b) Frintrop, 2007 (VOCUS)  c) Our Approach (*)
Integral Image

\[ ii(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y') \]

\[ s(x, y) = s(x, y - 1) + i(x, y) \]

\[ ii(x, y) = ii(x - 1, y) + s(x, y) \]
Integral Image and Saliency Map

\[
\text{surround}(x, y, \varsigma) = \frac{\text{rectSum}(x - \varsigma, y - \varsigma, x + \varsigma, y + \varsigma) - i(x, y)}{(2\varsigma + 1)^2 - 1} \\
\text{center}(x, y) = i(x, y)
\]

\[
\text{Int}_{\text{On}, \varsigma}(x, y) = \max\{\text{center}(x, y) - \text{surround}(x, y, \varsigma), 0\} \\
\text{Int}_{\text{Off}, \varsigma}(x, y) = \max\{\text{surround}(x, y, \varsigma) - \text{center}(x, y), 0\}
\]
New Visual Feature: Learning Approach

Pixel and neighborhood

LBP

Binary: 10011000
Decimal: 52

threshold against 5

DTLBP

Decision tree

n<br>5<4<br>y<br>5<8<br>5<2<br>0<br>1<br>2<br>3<br>4
build_tree($\mathcal{X}$) ≡
{Recursively build DT-LBP tree}
if terminate then
  return LeafNode
else
  $m \leftarrow$ choose_split($\mathcal{X}$)
  left $\leftarrow$ build_tree($\{(c_i, n_i, y_i) \in \mathcal{X} \mid c_i \geq n_{im}\}$)
  right $\leftarrow$ build_tree($\{(c_i, n_i, y_i) \in \mathcal{X} \mid c_i < n_{im}\}$)
  return SplitNode($m$, left, right)
end if

choose_split($\mathcal{X}$) ≡
{Choose most informative pixel comparison}
for $d = 1$ to $S$ do
  $\mathcal{X}_L \leftarrow \{(c_i, n_i, y_i) \in \mathcal{X} \mid c_i \geq n_{id}\}$
  $\mathcal{X}_R \leftarrow \{(c_i, n_i, y_i) \in \mathcal{X} \mid c_i < n_{id}\}$
  $\Delta H_d \leftarrow H(\mathcal{X}) - |\mathcal{X}_L| H(\mathcal{X}_L) - |\mathcal{X}_R| H(\mathcal{X}_R)$
end for
return $\arg\max_d \Delta H_d$
**New Visual Feature: Learning Approach**

<table>
<thead>
<tr>
<th>Method</th>
<th>FERET</th>
<th>CAS-PEAL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>fb</td>
<td>fc</td>
</tr>
<tr>
<td>LBP</td>
<td>0.93</td>
<td>0.51</td>
</tr>
<tr>
<td>LGBP</td>
<td>0.94</td>
<td>0.97</td>
</tr>
<tr>
<td>LVP</td>
<td>0.97</td>
<td>0.70</td>
</tr>
<tr>
<td>LGT</td>
<td>0.97</td>
<td>0.90</td>
</tr>
<tr>
<td>HGPP</td>
<td>0.98</td>
<td>0.99</td>
</tr>
<tr>
<td>LLGP</td>
<td>0.97</td>
<td>0.97</td>
</tr>
</tbody>
</table>
| DTLBP
| 7, no TT            | 0.98 | 0.44 | 0.63 | 0.42 | 0.96 | 0.80 |
| DTLBP
| 10, no TT           | 0.98 | 0.55 | 0.65 | 0.47 | 0.99 | 0.87 |
| DTLBP
| 12, no TT           | 0.99 | 0.63 | 0.67 | 0.48 | 0.99 | 0.88 |
| DTLBP
| 7                  | 0.98 | 0.99 | 0.79 | 0.78 | 0.95 | 0.89 |
| DTLBP
| 10                  | 0.99 | 0.99 | 0.83 | 0.78 | 0.98 | 0.91 |
| DTLBP
| 12                  | 0.99 | 1.00 | 0.84 | 0.79 | 0.98 | 0.92 |
| DTLBP
| 13                  | 0.99 | 1.00 | 0.84 | 0.80 | 0.98 | 0.92 |

People Detection

- **Visual Features**: HoG, LBP, DT-LBP, Saliency, Sal-HOG y Sal-DTLBP.
- **Dimensionality reduction**: Partial Least Square (PLS) (Schwartz et al., 2009). Reduction: 20,000 to 20-30 features.
- **Classifier**: Support Vector Machine (Radial Basis Kernel).
- **Sliding window approach**: 8 pixels.
- **Scale**: Gaussian filtering, 8 scales.
- **Non-Maximal suppression**: >0.5.
People Detection: Results

Inria Person Dataset (Dalal & Triggs, 2005)
People Detection: Results

Inria Person Dataset (Dalal & Triggs, 2005)
- Train: 2416 positive crops from 614 images.
- Train: 1218 negative images.
- Bootstrapping: 5 iterations over negative set.
- Test: 1126 positive crops from 288 images
- Test: 453 negative images.
People Detection: Results

Detection Error Tradeoff

Graph showing detection error tradeoff with miss rate on the y-axis and false positives per window (FPPW) on the x-axis. The graph compares different approaches including:
- our approach
- Lin & Davis [13]
- Tuzel et al. [22]
- Dalal & Triggs [5]
- Maji et al. [14]
- Wu, Nevatia [26]
- Dollar et al. [6]
## People Detection: Results

<table>
<thead>
<tr>
<th>Set</th>
<th>Nº Imágenes</th>
<th>Nº Personas</th>
<th>Precision</th>
<th>Recall</th>
<th>FPPI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>32</td>
<td>77</td>
<td>0.89</td>
<td>0.77</td>
<td>0.22</td>
</tr>
<tr>
<td>2</td>
<td>51</td>
<td>63</td>
<td>0.88</td>
<td>0.78</td>
<td>0.14</td>
</tr>
<tr>
<td>3</td>
<td>50</td>
<td>78</td>
<td>0.92</td>
<td>0.76</td>
<td>0.16</td>
</tr>
</tbody>
</table>
People Detection: Problems
**Part Based Approach**

**IDEA:** instead of detecting whole body, detect parts, such as torso, head, etc. This increases robustness to occlusion and deformation problems.

**Problem:** What is a part?, Where are they located?

**Solution:** Unsupervised approach.
Part Based Approach

### Algorithm
- Detect **relevant** local areas of positive training examples.
- **Relevance?**: Use weights of linear SVM classifiers.
- Train a classifier using only image area defined by relevant part.
- Run classifier on training set and remove image areas that fire the classifier.
- Repeat steps above until obtaining a predefined number of parts.
Part Based Approach

Patch Relevance
Part Based Approach

Part Classifiers
Part Based Approach: Results

Part Detection Results
Current Research: Counting and Recognizing People in a Classroom
Current Research: Counting and Recognizing People in a Classroom

Sliding Cube

View 4

View 3

View 1

View 2
Current Research: Robot Navigation Using Vision
Current Research: Scene Recognition Through Object Detection

PlaceProbs

Office : 6.45%
Hall   : 0.00%
Conf   : 93.55%
Bath   : 0.00%

Screen: 73.71%
Clock: 75.07%
People
- 4 PhD Faculties
- 8 PhD Students
- 11 MSc Students

Research Areas
- Machine Learning
- Robotics
- Computer Vision
THANKS!