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

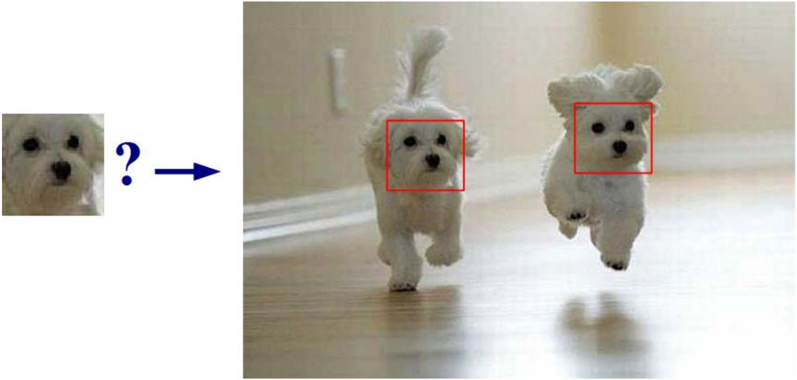
[2] Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE), Mexico.

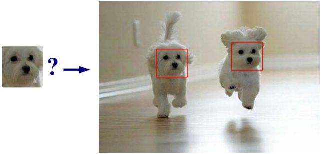
Our Problem: Counting People at Service Lines in Grocery Stores

Our Solution:

Computer Vision
+
Machine Learning

Computer Vision: Template Matching

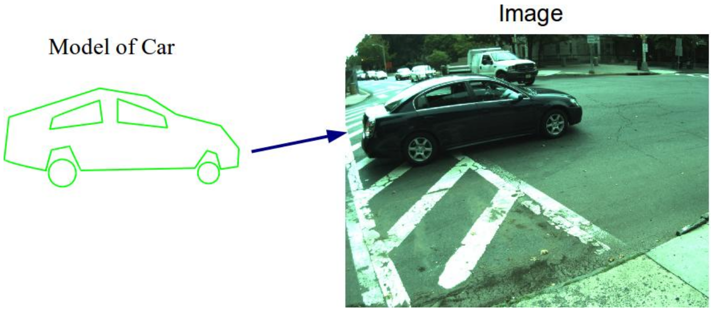


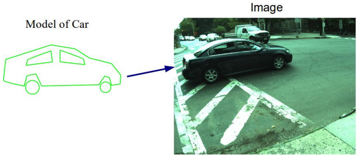


Computer Vision: Problems

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

Computer Vision: Geometric Models

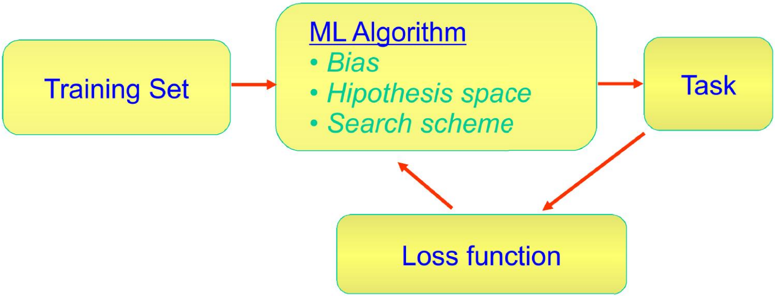




Computer Vision: Problems

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

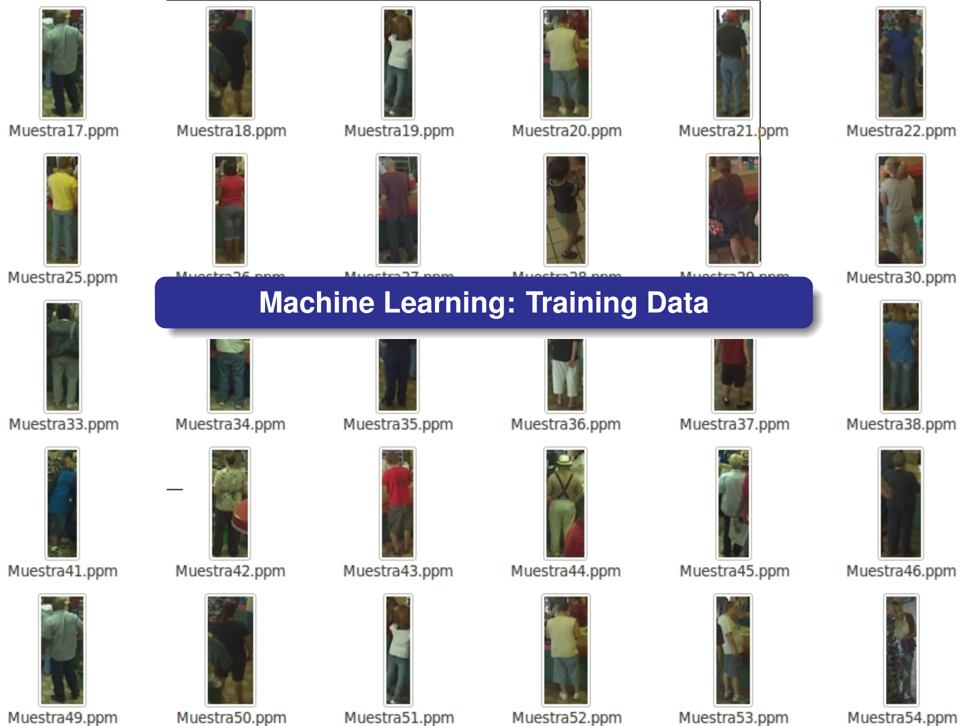
Machine Learning



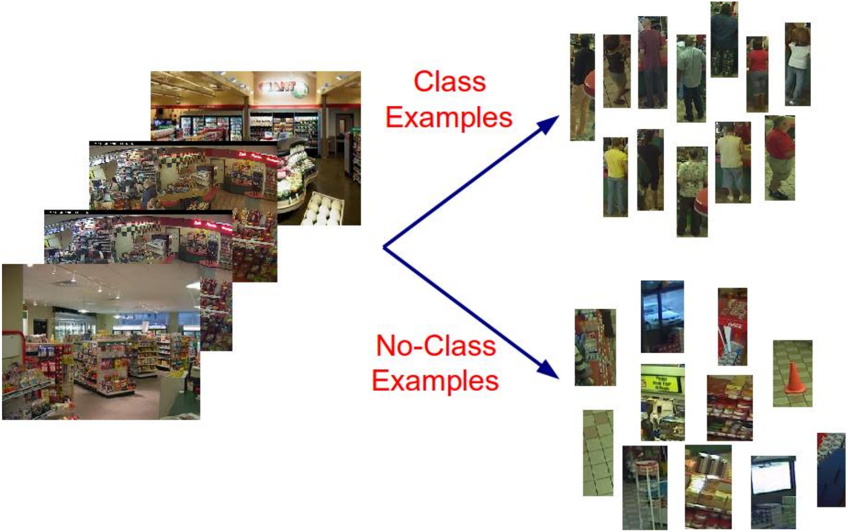
Main goal is to learn a concept (task) by finding a model that minimizes expected loss using observed data.

Machine Learning: Training Data

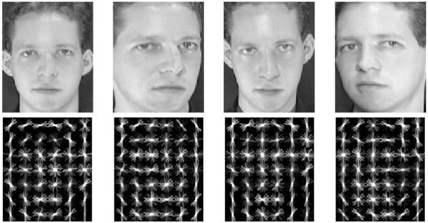
Machine Learning: Training Data

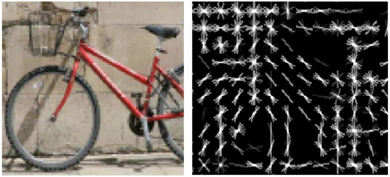


Binary Classification



Appearance Models
Representation: Visual Features





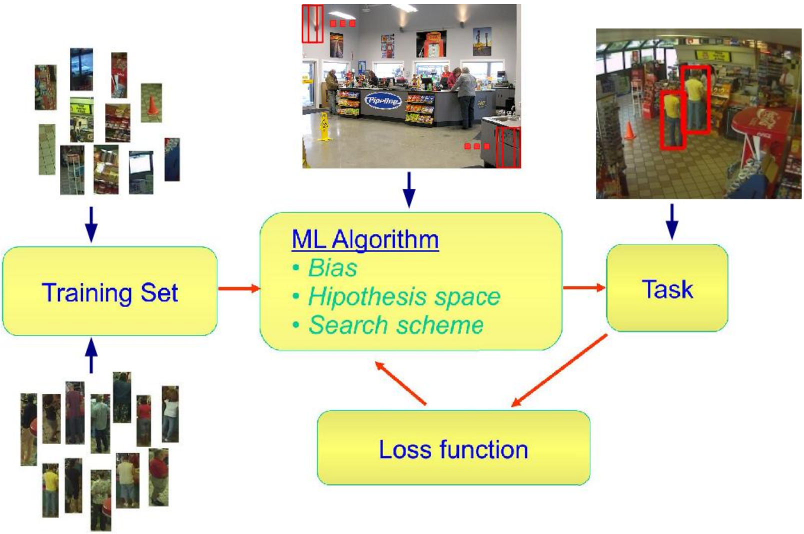
- Use of information invariant to some visual problems.
- Use of statistical information.

Image Segmentation vs Sliding Window

Image Segmentation

Sliding Window

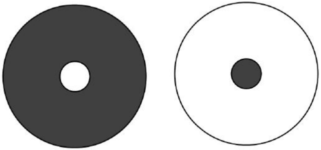
Approach



ML Algorithm
• Bias
• Hipothesis space
• Search scheme
Training Set
Task
Loss function

New Visual Feature: Saliency Map

Center Sorround Patterns:
Biological Cells



Center Sorround Patterns:
Approximation

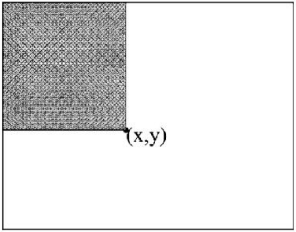
New Visual Feature: Saliency Map

New Visual Feature: Saliency Map

a) Itti et al, 1998 (iNVT) 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

$$surround(x, y, \varsigma) = \frac{rectSum(x-\varsigma, y-\varsigma, x+\varsigma, y+\varsigma) - i(x, y)}{(2\varsigma+1)^2 - 1}$$

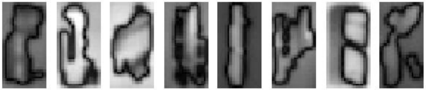
$$center(x, y) = i(x, y)$$

$$Int_{On,\varsigma}(x, y) = \max\{center(x, y) - surround(x, y, \varsigma), 0\}$$

$$Int_{Off,\varsigma}(x, y) = \max\{surround(x, y, \varsigma) - center(x, y), 0\}$$

New Visual Feature: Saliency Map

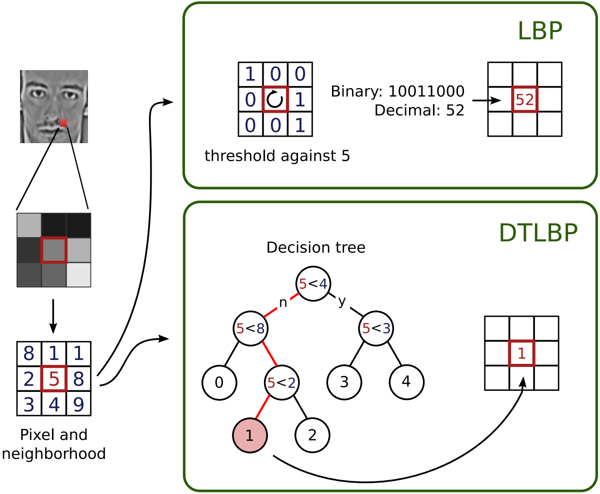
a) Visual salience features: human images



b) Visual salience features: non-human images

S. Montabone and A. Soto, "Human Detection Using a Mobile Platform and Novel Features Derived From a Visual Saliency Mechanism". Image and Vision Computing, vol. 28, No. 3, pp. 391-402, 2010.

New Visual Feature: Learning Approach



$\texttt{build\_tree}(\mathcal{X}) \equiv$
{Recursively build DT-LBP tree}
if $\texttt{terminate}$ **then**
  return $\texttt{LeafNode}$
else
  $m \leftarrow \texttt{choose\_split}(\mathcal{X})$
  $\texttt{left} \leftarrow \texttt{build\_tree}(\{(c_i, \mathbf{n}_i, y_i) \in \mathcal{X} \mid c_i \geq n_{im}\})$
  $\texttt{right} \leftarrow \texttt{build\_tree}(\{(c_i, \mathbf{n}_i, y_i) \in \mathcal{X} \mid c_i < n_{im}\})$
  return $\texttt{SplitNode}(m, \texttt{left}, \texttt{right})$
end if

$\texttt{choose\_split}(\mathcal{X}) \equiv$
{Choose most informative pixel comparison}
for $d = 1$ to S **do**
  $\mathcal{X}_L \leftarrow \{(c_i, \mathbf{n}_i, y_i) \in \mathcal{X} \mid c_i \geq n_{id}\}$
  $\mathcal{X}_R \leftarrow \{(c_i, \mathbf{n}_i, y_i) \in \mathcal{X} \mid c_i < n_{id}\}$
  $\Delta H_d \leftarrow H(\mathcal{X}) - \frac{|\mathcal{X}_L|}{|\mathcal{X}|} H(\mathcal{X}_L) - \frac{|\mathcal{X}_R|}{|\mathcal{X}|} H(\mathcal{X}_R)$
end for
return $\arg\max_d \Delta H_d$

New Visual Feature: Learning Approach

Method	FERET				CAS-PEAL	
	fb	fc	dup1	dup2	Expr.	Acc.
LBP	0.93	0.51	0.61	0.50	-	-
LGBP	0.94	0.97	0.68	0.53	0.95	0.87
LVP	0.97	0.70	0.66	0.50	0.96	0.86
LGT	0.97	0.90	0.71	0.67	-	-
HGPP	0.98	0.99	0.78	0.76	0.96	**0.92**
LLGP	0.97	0.97	0.75	0.71	0.96	0.90
$DTLBP^7_8$, no TT	0.98	0.44	0.63	0.42	0.96	0.80
$DTLBP^7_{10}$, no TT	0.98	0.55	0.65	0.47	**0.99**	0.87
$DTLBP^7_{12}$, no TT	**0.99**	0.63	0.67	0.48	**0.99**	0.88
$DTLBP^7_8$	0.98	0.99	0.79	0.78	0.95	0.89
$DTLBP^7_{10}$	**0.99**	0.99	0.83	0.78	0.98	0.91
$DTLBP^7_{12}$	**0.99**	**1.00**	**0.84**	0.79	0.98	**0.92**
$DTLBP^7_{13}$	**0.99**	**1.00**	**0.84**	**0.80**	0.98	**0.92**

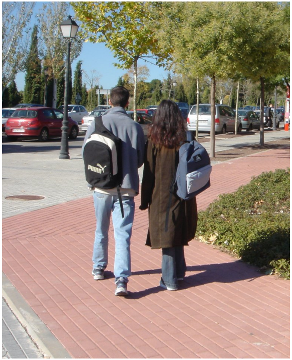
D. Maturana, D. Mery, and A. Soto, "Face Recognition with Decision Tree-based Local Binary Patterns". In Proc. of Asian Conference on Computer Vision (ACCV), 2010.
D. Maturana, D. Mery, and A. Soto, "Learning Discriminative Local Binary Patterns for Face Recognition". In Proc. of IEEE Conference on Face and Gesture Recognition (FG), 2011.

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)

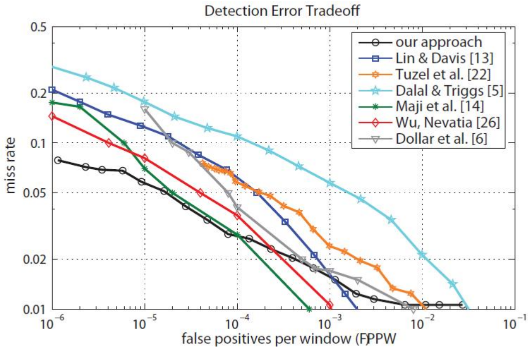


People Detection: Results

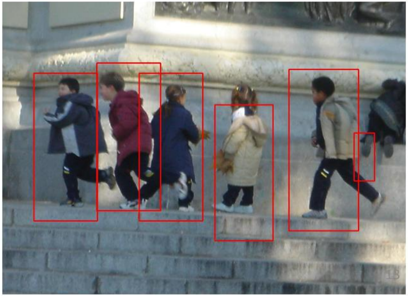
Inria Person Dataset (Dalal & Triggs, 2005)

- 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



People Detection: Results



People Detection: Results

People Detection: Results

Set	Nº Imágenes	Nº Personas	Precision	Recall	FPPI
1	32	77	0.89	0.77	0.22
2	51	63	0.88	0.78	0.14
3	50	78	0.92	0.76	0.16

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.

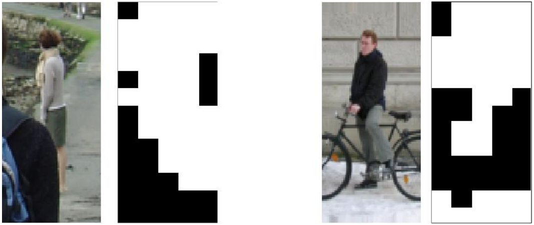
Part Based Approach

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.

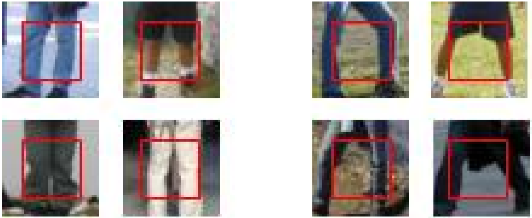
Part Based Approach

Patch Relevance



Part Based Approach

Part Classifiers

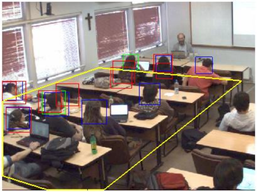


Part Based Approach: Results

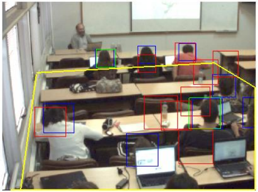
Part Detection Results

Current Research: Counting and Recognizing People in a Classroom

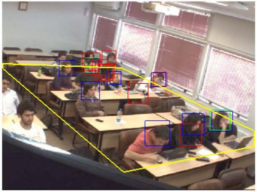
View 4



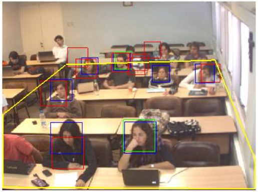
View 3



View 1

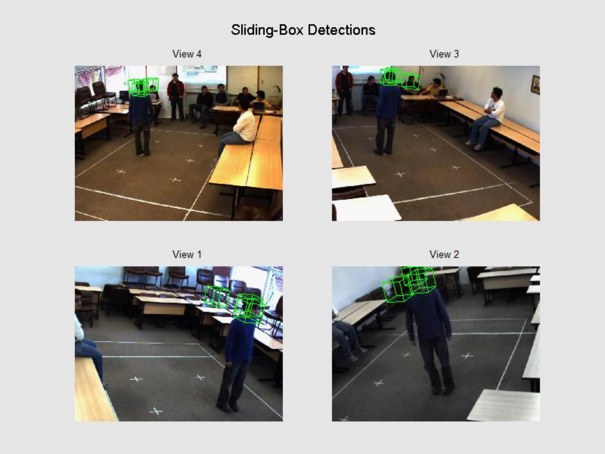


View 2

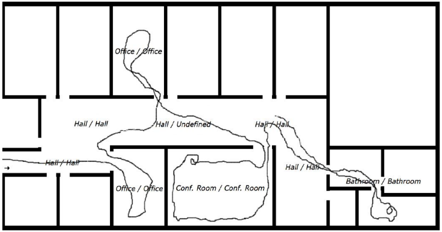


Current Research: Counting and Recognizing People in a Classroom

Sliding Cube

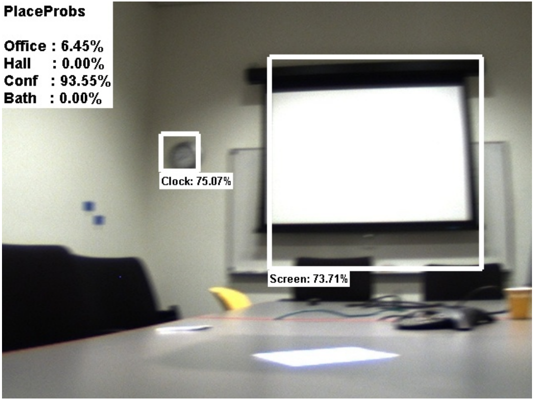


Current Research: Robot Navigation Using Vision



Detected place / Ground truth

Current Research: Scene Recognition Through Object Detection



GRIMA

Grupo de Inteligencia de Máquina

http://grima.ing.puc.cl

People

- 4 PhD Faculties
- 8 PhD Students
- 11 MSc Students

Research Areas

- Machine Learning
- Robotics
- Computer Vision

THANKS!