HOGgles

Visualizing Object Detection Features

C. Vondrick, A. Khosla, T. Malisiewicz, A. Torralba
ICCV, 2013.

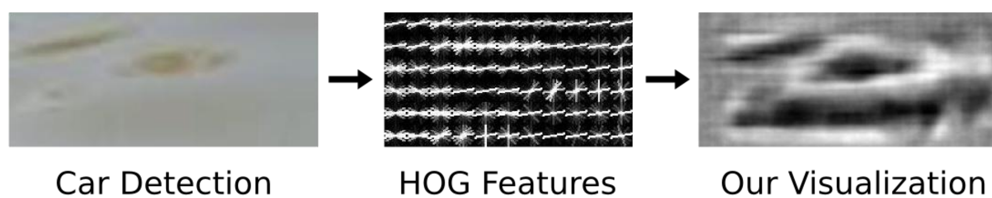
presented by

Ezgi Mercan

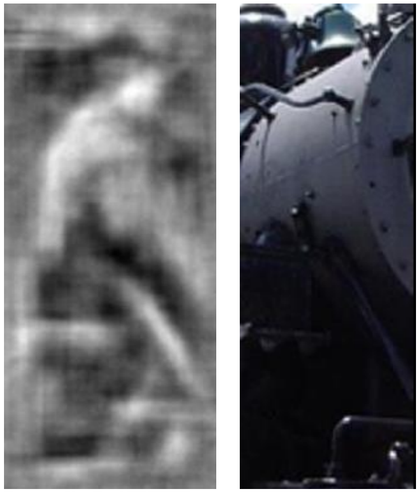
Object Detection Failures

A high scoring car detection from DPM.

- Why do our detectors think water looks like a car?
  - Data set?
  - Machine learning method?
  - Features?

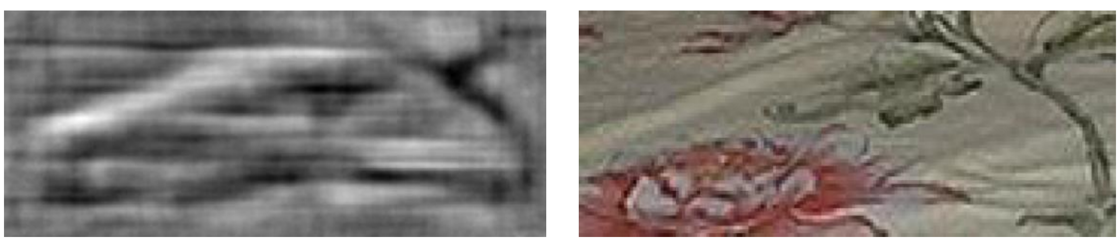


Object Detection Failures



Person

Chair

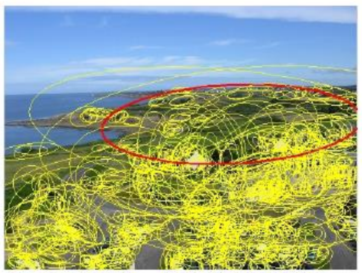


Car

<u>Object Categories</u>

- Aeroplane
- Bicycle
- Bird
- Boat
- Bottle
- Bus
- Car
- Cat
- Chair
- Cow
- Table
- Dog
- Horse
- Motorbike
- Person
- Potted plant
- Sheep
- Sofa
- Train
- TV/ monitor

Reconstructing an Image



calculate SIFT

→

elliptic
region of interest

→

affine normalization
to square patch

→

$$R_i = \{\mathbf{v}_i, \mathbf{x}_i, s_i, o_i, A_i\}$$

local descriptor

copied patches

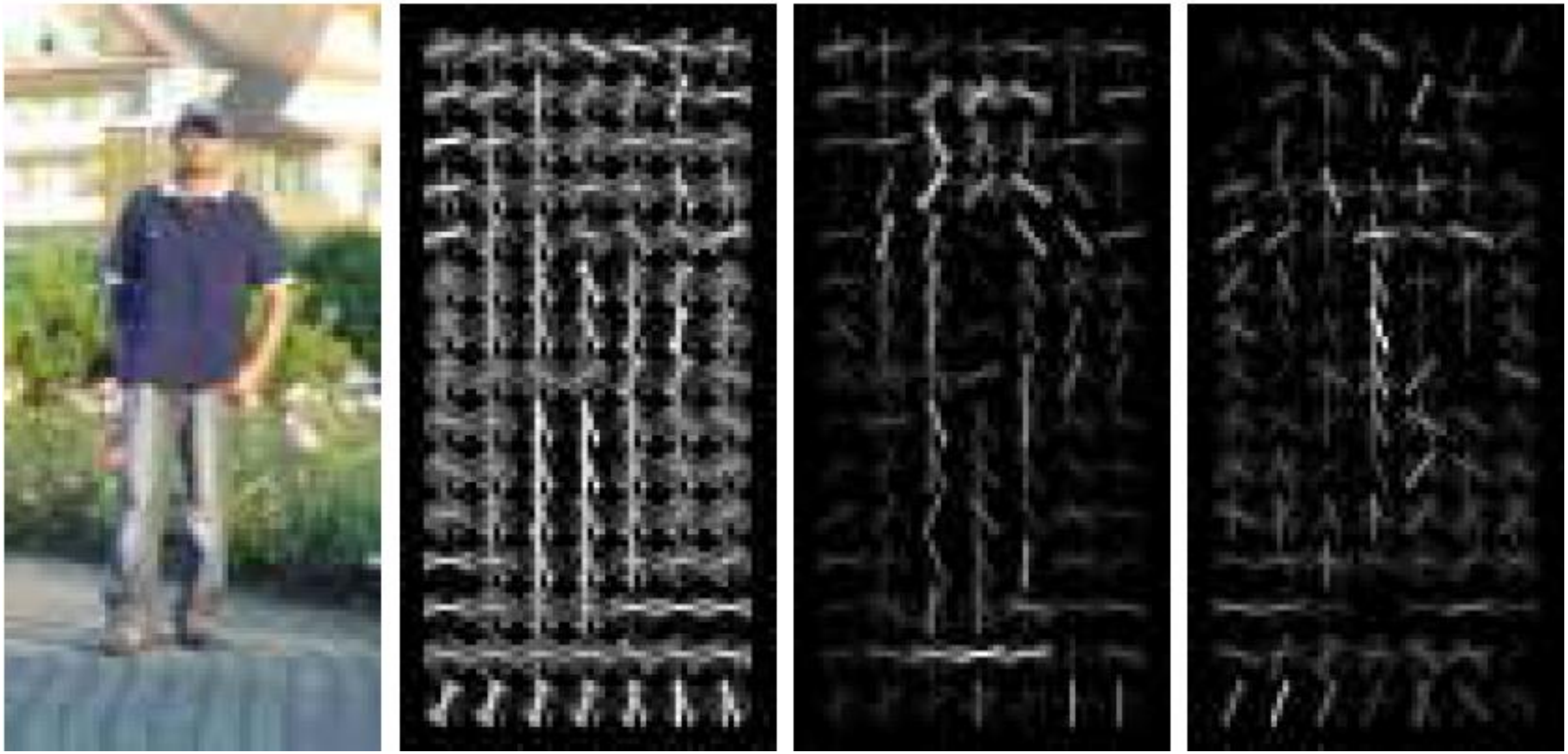
blending

interpolation

original image

Reconstructing an image from its local descriptors, Philippe Weinzaepfel, Hervé Jégou and Patrick Pérez, Proc. IEEE CVPR’11.

Histogram of Oriented Gradients



original image | R-HOG descriptor (HOG glyph) | R-HOG descriptor weighted by positive SVM weights | R-HOG descriptor weighted by negative SVM weights

Histograms of oriented gradients for human detection, Dalal, N., Triggs, B., CVPR'05.

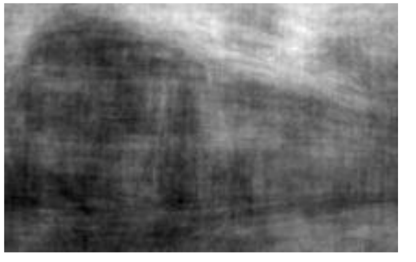
Algorithms

Image: $x \in \mathbb{R}^D$

HOG descriptor: $y = \emptyset(x) \quad \in \mathbb{R}^d$

HOG inverse: $\emptyset^{-1}(y)$

- Exemplar LDA



$$\emptyset_A^{-1}(y) = \frac{1}{K} \sum_{i=1}^{K} z_i$$

- Ridge Regression

$$\emptyset_B^{-1}(y) = \Sigma_{XY} \Sigma_{YY}^{-1}(y - \mu_Y) + \mu_X$$

Algorithms

- Direct Optimization

Image: $x \in \mathbb{R}^D$
HOG descriptor: $y \in \mathbb{R}^d$
Image basis: $U \in \mathbb{R}^{D \times K}$
Coefficients: $\rho \in \mathbb{R}^K$; $x = U\rho$

$$\emptyset_C^{-1}(y) = U\,\rho^*$$

$$\rho^* = \underset{\rho \in \mathbb{R}^K}{\operatorname{argmin}} \|\emptyset(U\rho) - y\|_2^2$$

Algorithms

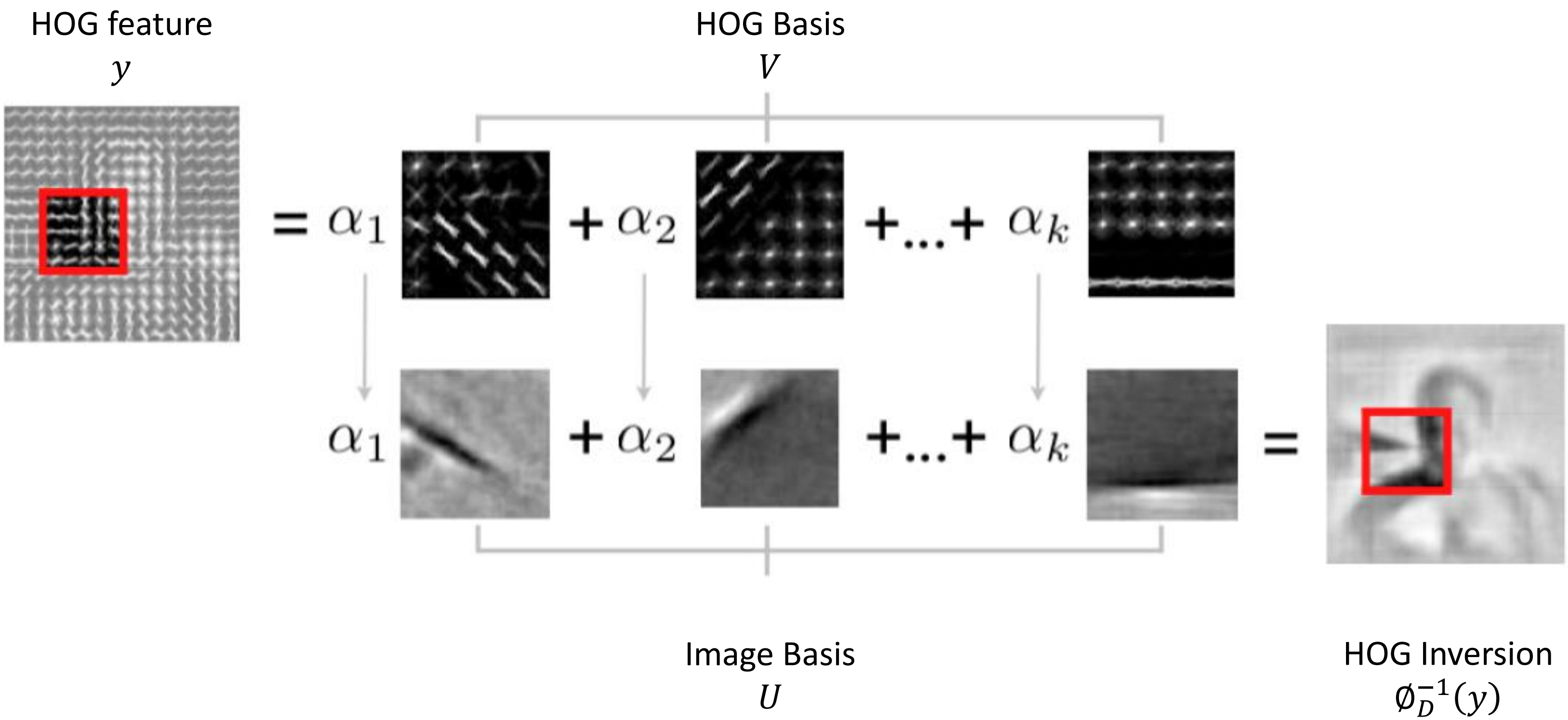
- Paired Dictionary Learning

Image: $x \in \mathbb{R}^D$
HOG descriptor: $y \in \mathbb{R}^d$
Image basis: $U \in \mathbb{R}^{D \times K}$
HOG basis: $V \in \mathbb{R}^{d \times K}$
Coefficients: $\alpha \in \mathbb{R}^K$; $x = U\alpha \; and \; y = V\alpha$

$$\emptyset_D^{-1}(y) = U\,\alpha^*$$

$$\alpha^* = \underset{\alpha \in \mathbb{R}^K}{\operatorname{argmin}} \|V\alpha - y\|_2^2 \qquad \text{s.t. } \|\alpha\|_1 \leq \lambda$$

Paired Dictionary Learning

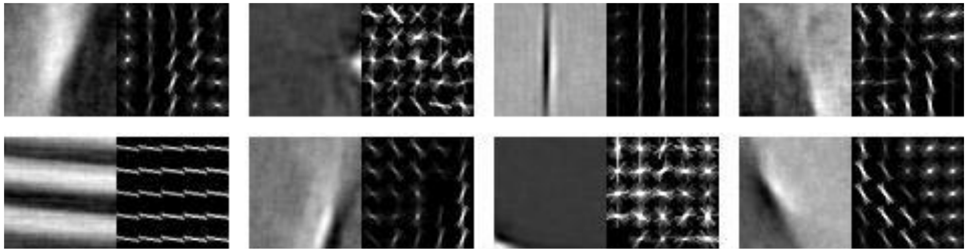


Paired Dictionary Learning

- Solving paired dictionary learning problem:

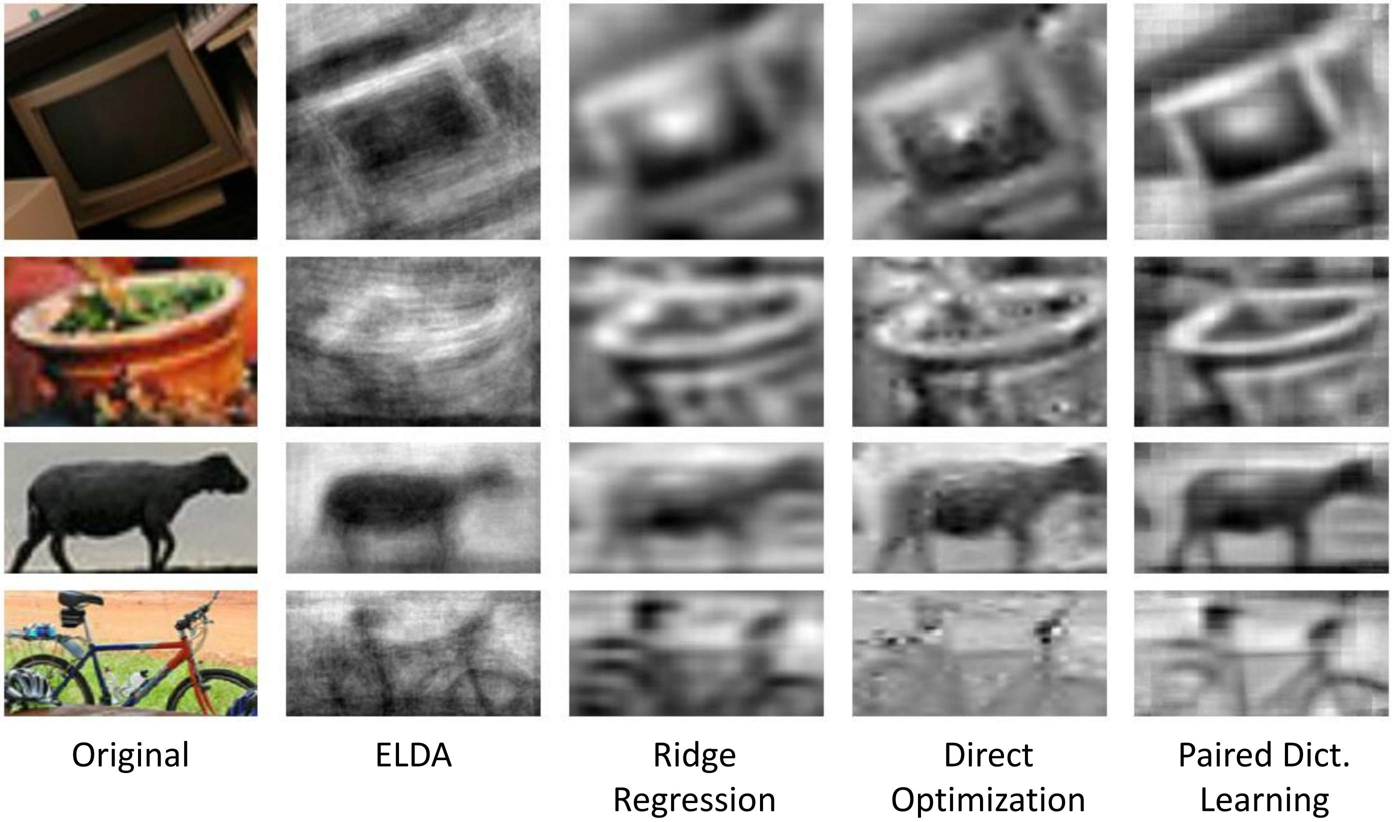
$$\underset{U,V,\alpha}{\operatorname{argmin}} \sum_{i=1}^{N} (\|x_i - Ua_i\|_2^2 + \|\phi(x_i) - Va_i\|_2^2)$$

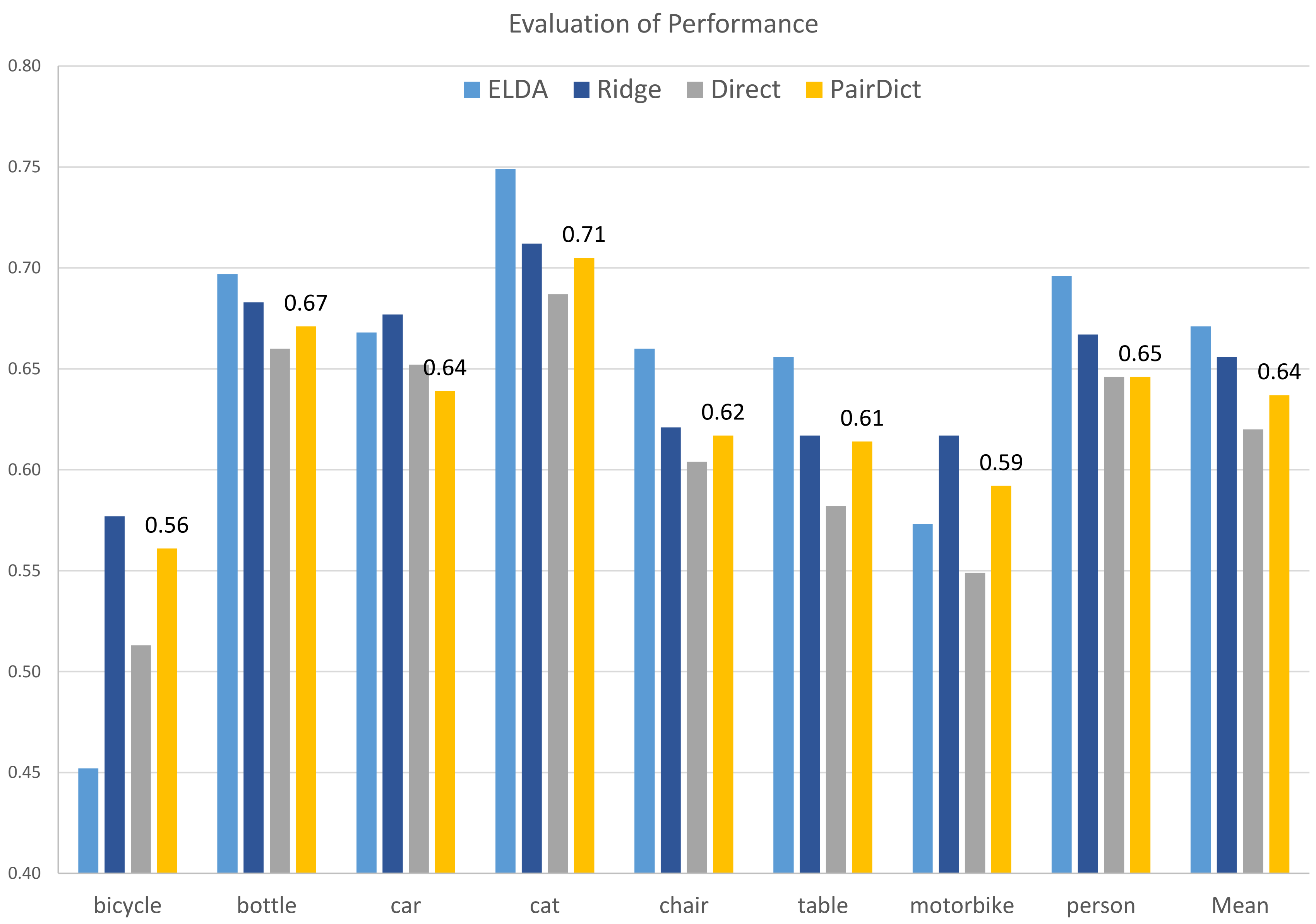
s.t. $\|a_i\|_1 \leq \lambda \, \forall i, \quad \|U\|_2^2 \leq \gamma_1, \quad \|V\|_2^2 \leq \gamma_2$

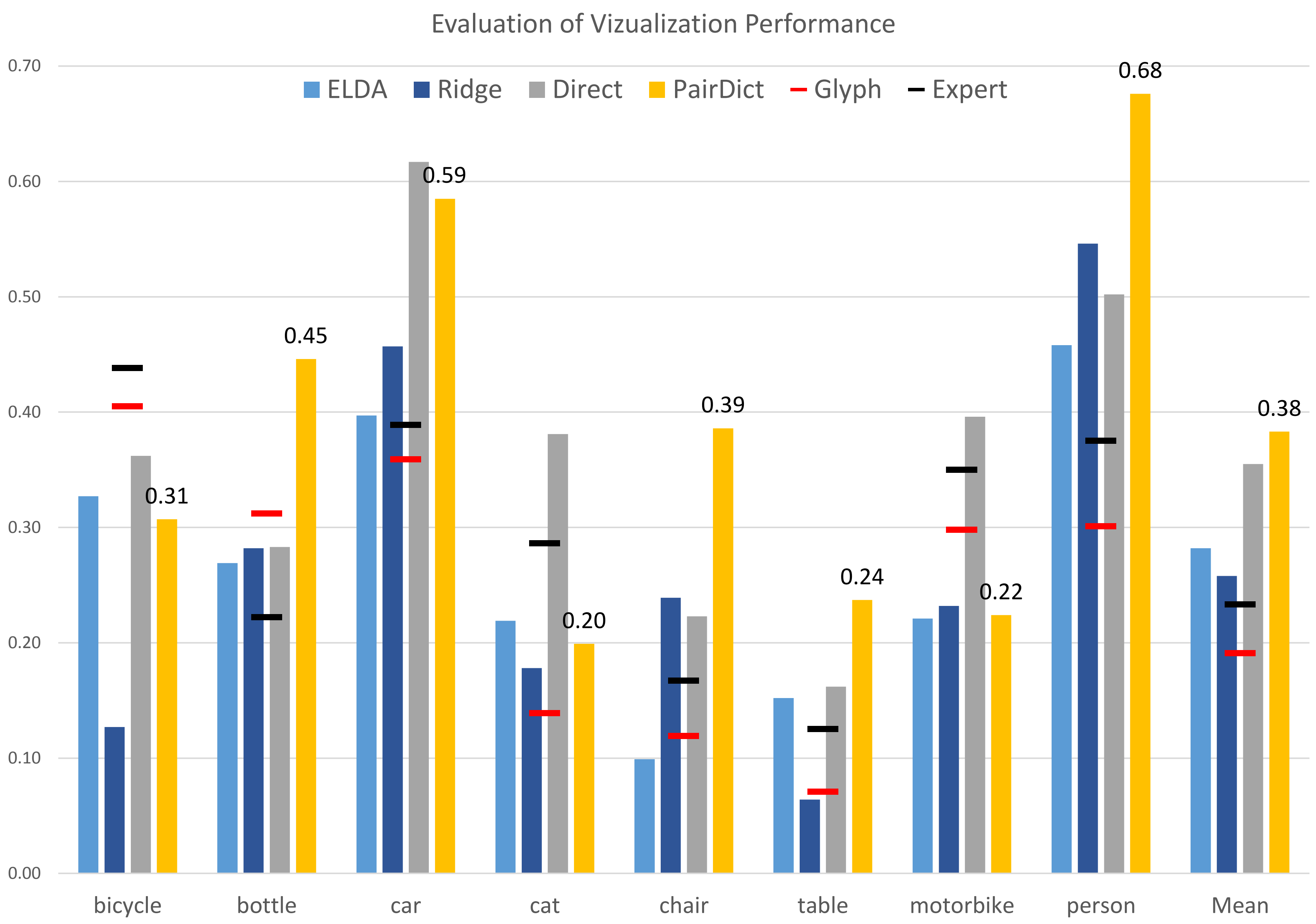


some learned pairs of dictionaries for U and V

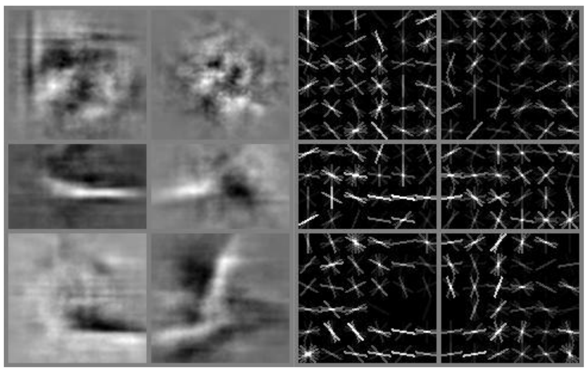
Feature Visualization



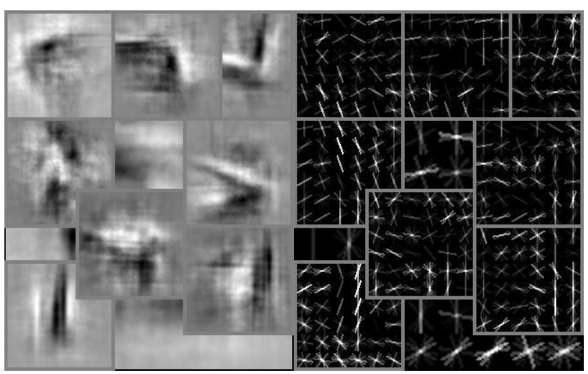

Evaluation of Performance
ELDA
Ridge
Direct
PairDict
0.80
0.75
0.70
0.65
0.60
0.55
0.50
0.45
0.40
0.56
0.67
0.64
0.71
0.62
0.61
0.59
0.65
0.64
bicycle
bottle
car
cat
chair
table
motorbike
person
Mean


Evaluation of Vizualization Performance
ELDA
Ridge
Direct
PairDict
Glyph
Expert
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0.31
0.45
0.59
0.20
0.39
0.24
0.22
0.68
0.38
bicycle
bottle
car
cat
chair
table
motorbike
person
Mean

Deformable Parts Model



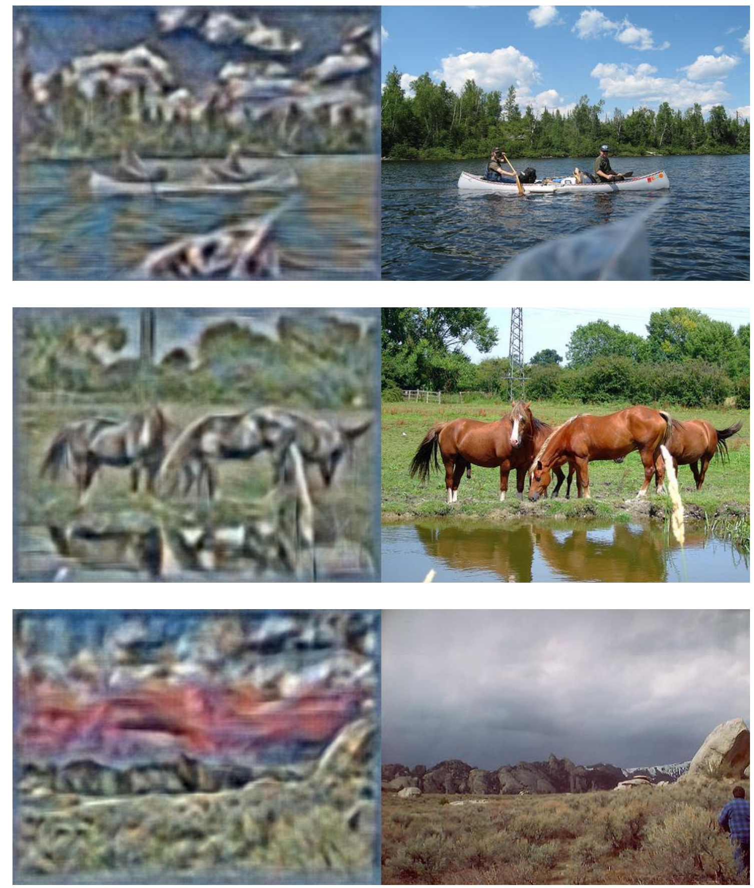
Potted Plant



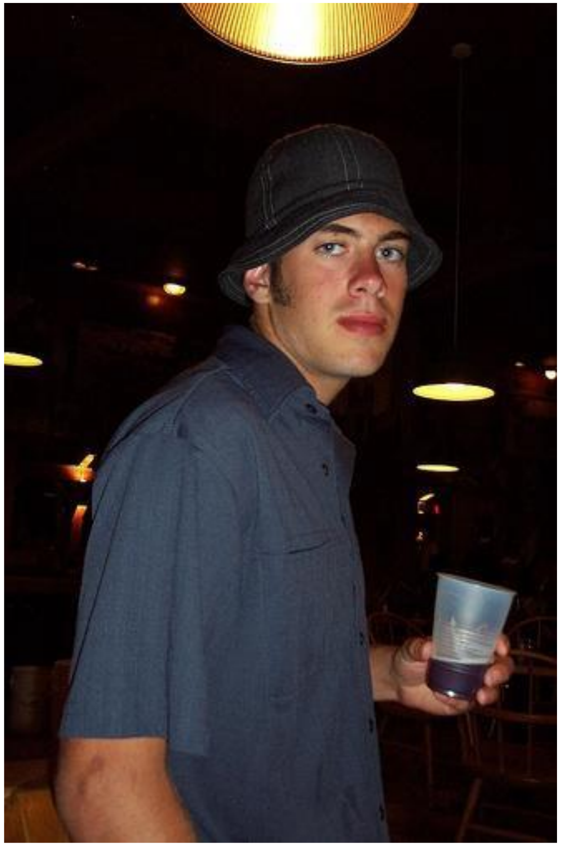
Chair

Object detection with discriminatively trained part-based models, P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan. PAMI, 2010.

Does HOG capture color?



Computers can see better than us



Questions?

Visit
web.mit.edu/vondrick/ihog/
for more cool stuff.