Object Recognition:
Scale Invariant Feature Transform (SIFT) - based Approach,
in comparison with
CNN-based Approach

M. Goudarzi 5.12.2016

Object Recognition: an overview

- You meet a new person or an object, what makes you recognize them the next day?
- What helps our brain to first detect and then recognize we are meeting the same person again?
- Does our brain "tag" what it sees?

M. Goudarzi 5.12.2016

Object Recognition: an overview

- You meet a new person or an object, what makes you recognize them the next day?
- Is it the Facial Expression?
- Their Haircut?
- Their shape and size?
- etc.

M. Goudarzi 5.12.2016

Object Recognition: an overview

- You meet a new person or an object, what makes you recognize them the next day?
- Is it the Facial Expression?

[1]

M. Goudarzi 5.12.2016

Object Recognition: an overview

- You meet a new person or an object, what makes you recognize them the next day?
- Is it the Facial Expression?
- Is it the haircut?
- Is it a shape and size
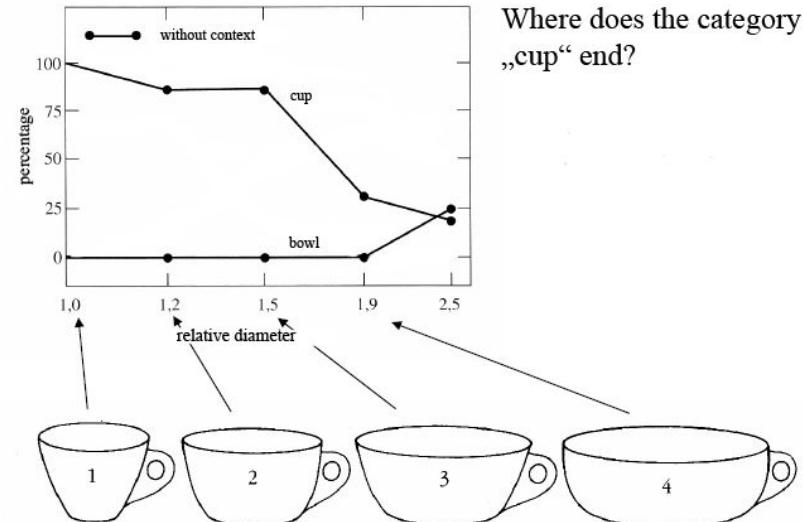
M. Goudarzi 5.12.2016   [3]

Object Recognition: an overview.

- What makes us detect, remember and recognize an object?
- Would we still remember an object/person when their disguised, re-colored, occluded , etc.
- What does our brains react to in terms of attention, detection and recognition.

M. Goudarzi 5.12.2016

Human Visual System

- Evolved over 500 million years.
- Adapted to the environment over time.
- Nuance Detection.
- HVS Model used in Computer Vision and Image Processing.

M. Goudarzi 5.12.2016

Object Recognition: Perspectives from Cognitive Psychology and Neuroscience

Some insightful reads:
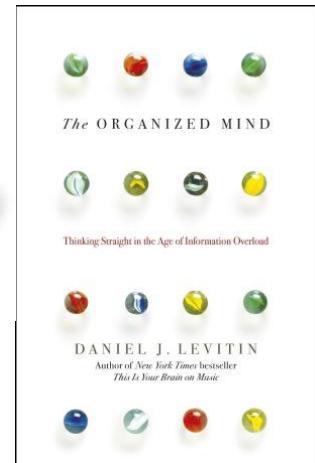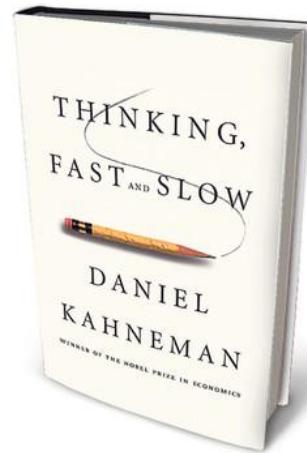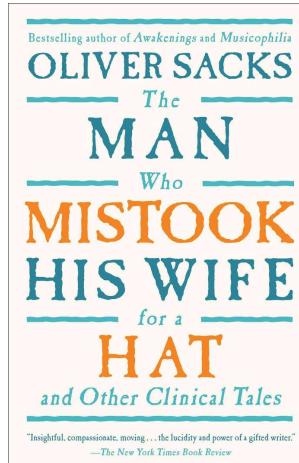
[4,5,6]

M. Goudarzi 5.12.2016

Human Visual System (HVS) model

- A **human visual system model** (HVS model) is used by computer vision experts to deal with biological and psychological processes that are not yet fully understood.

Assumptions need to be made:

- Low-Pass filter characteristics. (Mach Bands)
- Lack of color resolution
- Motion sensitivity
- Integral face recognition
- etc.

M. Goudarzi 5.12.2016

Human Visual System (HVS) model

[7]
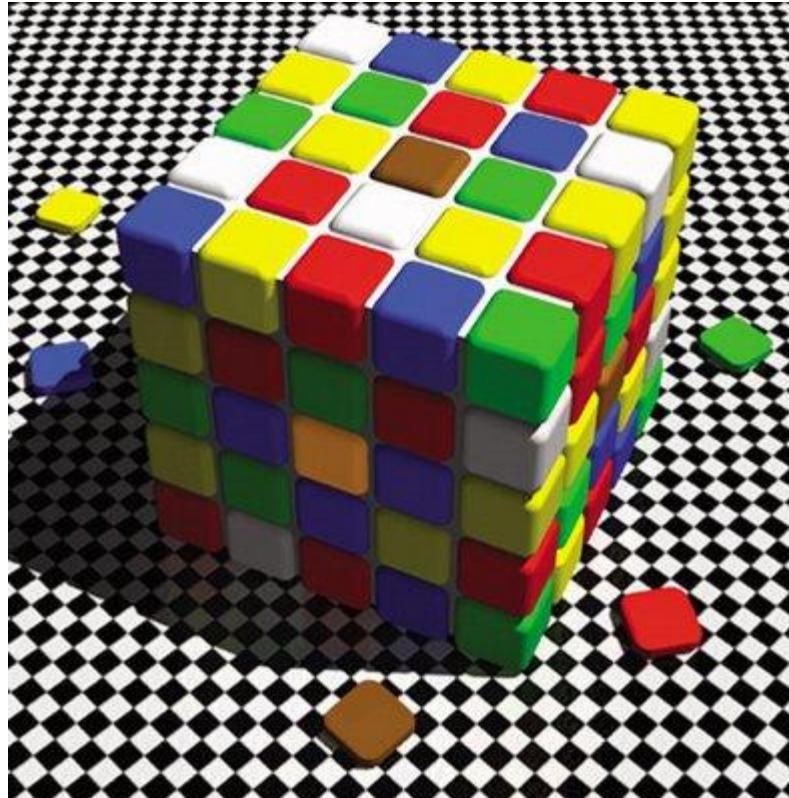
M. Goudarzi 5.12.2016

Human Visual System (HVS) model

[8]
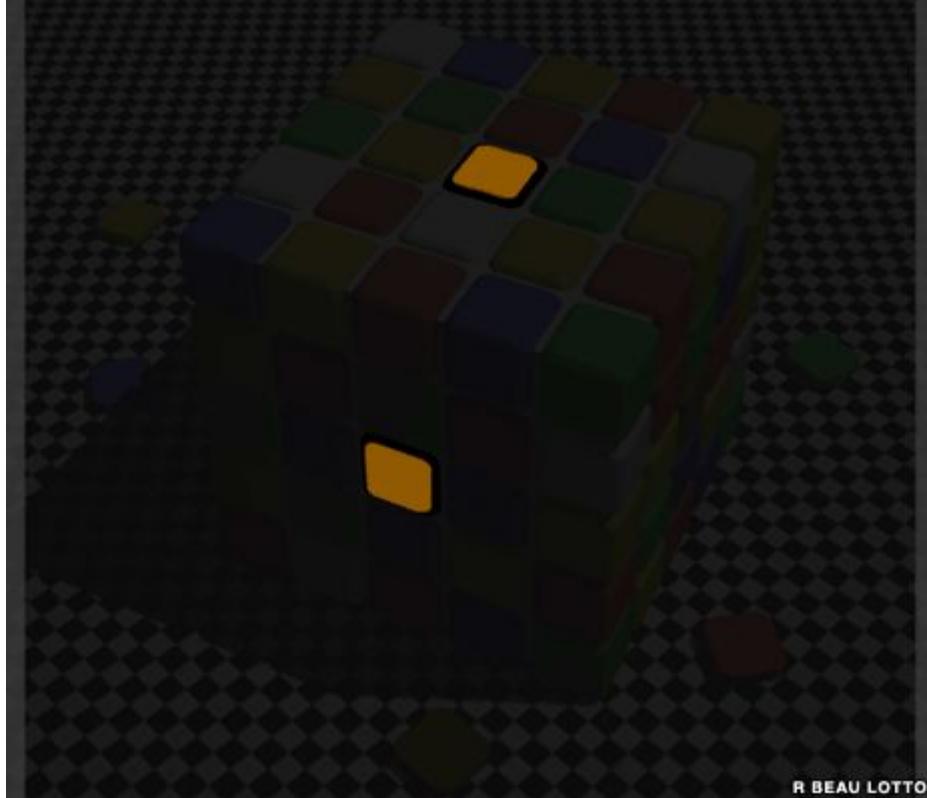
M. Goudarzi 5.12.2016

Man vs. The Machine

- Human object recognition vs. computer-based Object recognition system
- Fundamental Difference in semantics.

M. Goudarzi 5.12.2016

Man vs. The Machine

Man vs. The Machine

M. Goudarzi 5.12.2016   [9]

Man vs. The Machine

- To human beings, this is not just "a boy sitting with a pair of shoes"
- Context matters to us.

[11]

M. Goudarzi 5.12.2016

WERFEL, A 6-YEAR-OLD AUSTRIAN ORPHAN, BEAMS WITH UNBOUNDED JOY AS HE CLASPS A NEW PAIR OF SHOES PRESENTED TO HIM BY THE AMERICAN RED CROS

Man vs. The Machine

- Our perception of images changes with the surrounding context, including those including sound and rhythms.

[12]

M. Goudarzi 5.12.2016

SIFT Features

- Introduced by David Lowe in 1999
- Published in 2004

[13]

M. Goudarzi 5.12.2016

SIFT Features

- Goal: Extracting distinctive features which are invariant to common image transformations.
- Invariance to image rotation and scale.
- Local operation
- Close to real-time performance
- Robust w.r.t :
  - Affine Transformation
  - Noise
  - Viewpoint Change

M. Goudarzi 5.12.2016

SIFT Features Steps of key-point extraction:

- Scale-space peak selection
  - Potential feature locations
- Key-point localization
  - Locating key-points accurately
- Orientation assignment
  - Orientation assignment
- Key-point descriptor
  - Vectorizing key-point descriptions

M. Goudarzi 5.12.2016

SIFT feature:
Blob detection
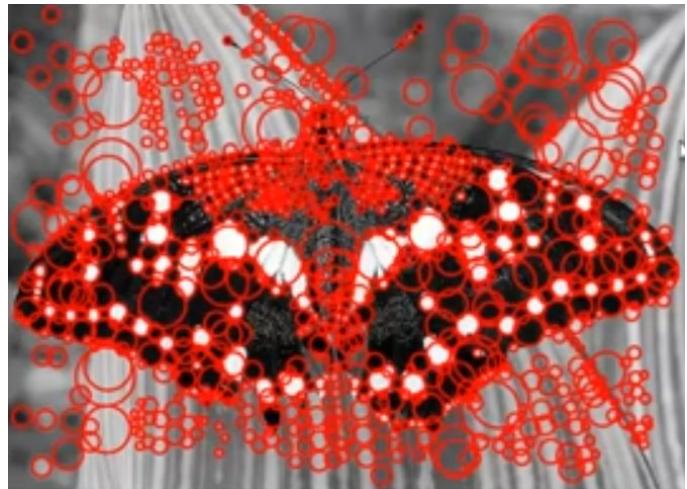
[14]

M. Goudarzi 5.12.2016

SIFT feature: Laplace of Gaussian : LoG

Interest points:
Local maxima in scale space of Laplacian-of-Gaussian

$L_{xx}(\sigma) + L_{yy}(\sigma) \rightarrow$

⇒ List of (x, y, σ)
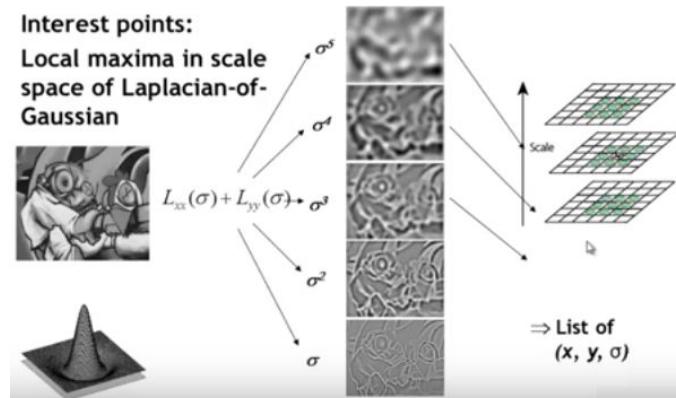
M. Goudarzi 5.12.2016   [15]

SIFT Features: LoG approximation with DoG

$$\frac{\partial G}{\partial \sigma} = \sigma \Delta^2 G \qquad \text{Heat Equation}$$

$$\sigma \Delta^2 G = \frac{\partial G}{\partial \sigma} = \frac{G(x, y, k\sigma) - G(x, y, \sigma)}{k\sigma - \sigma}$$

$$G(x, y, k\sigma) - G(x, y, \sigma) \approx (k-1)\sigma^2 \Delta^2 G$$

M. Goudarzi 5.12.2016   [16]

SIFT Features: Orientation Assignment
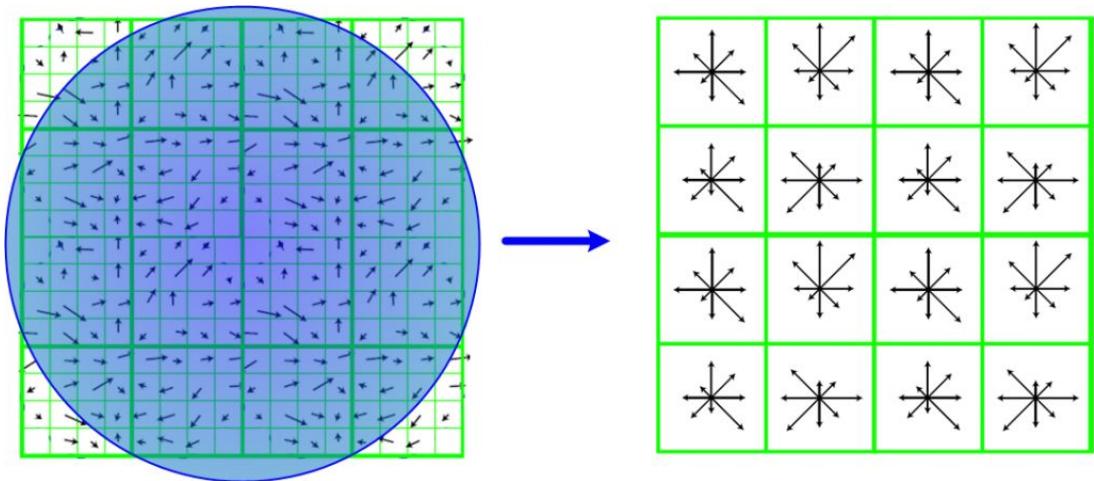
[16]

M. Goudarzi 5.12.2016

# SIFT Feature Matching
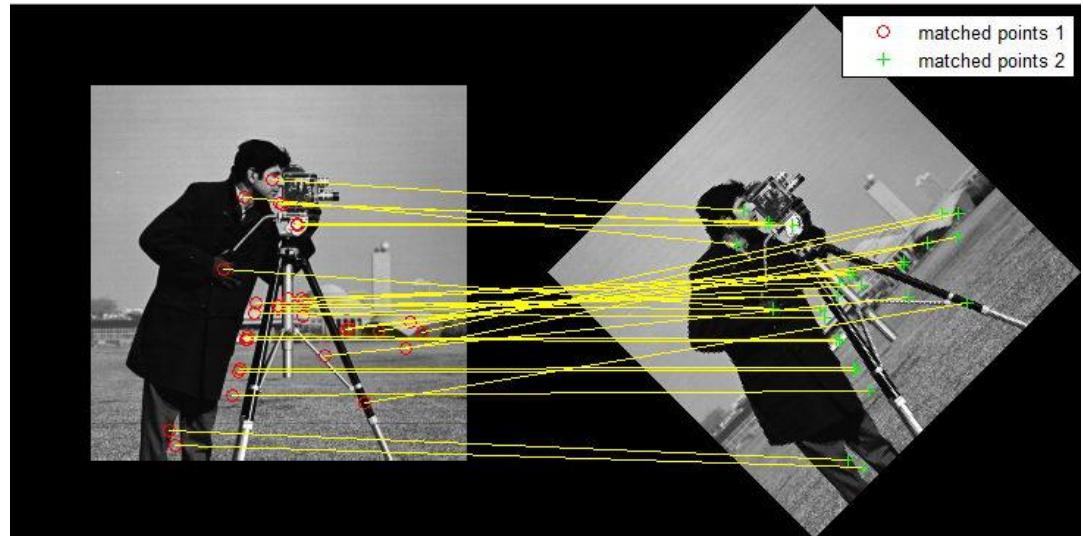
M. Goudarzi 5.12.2016   [17]

Bag of visual words (BoW) approach

- How to recognize an object from what has been already learned.

[18]
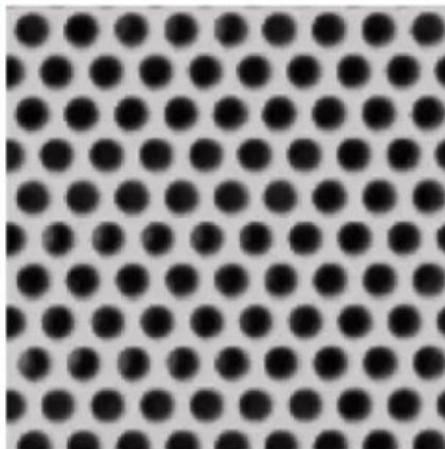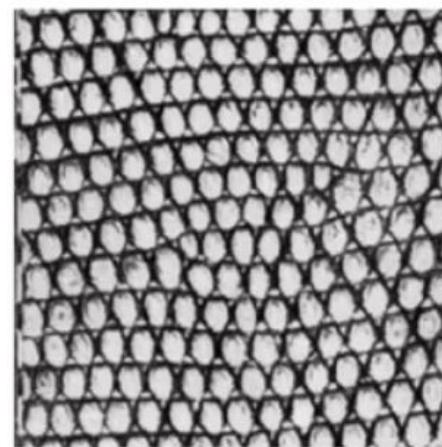
M. Goudarzi 5.12.2016

BoW - inspired by "Document Searching"
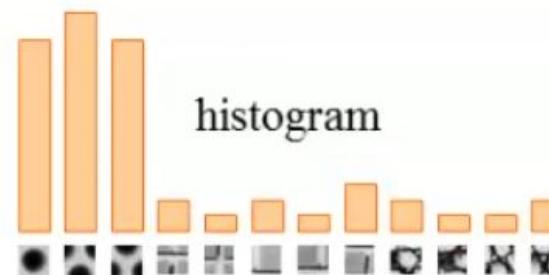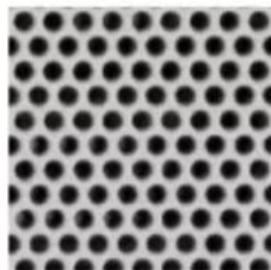
histogram

Universal texton dictionary

[19]
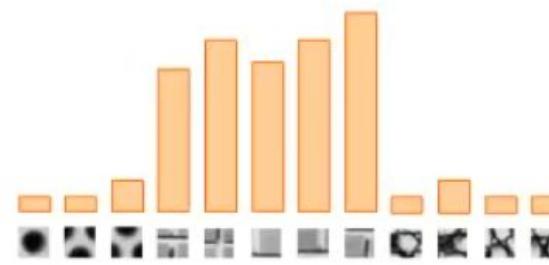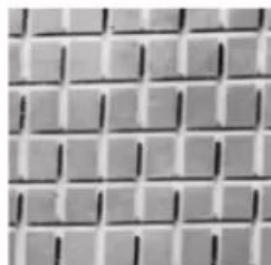
M. Goudarzi 5.12.2016

BoW Approach - using SIFT Features

[19]
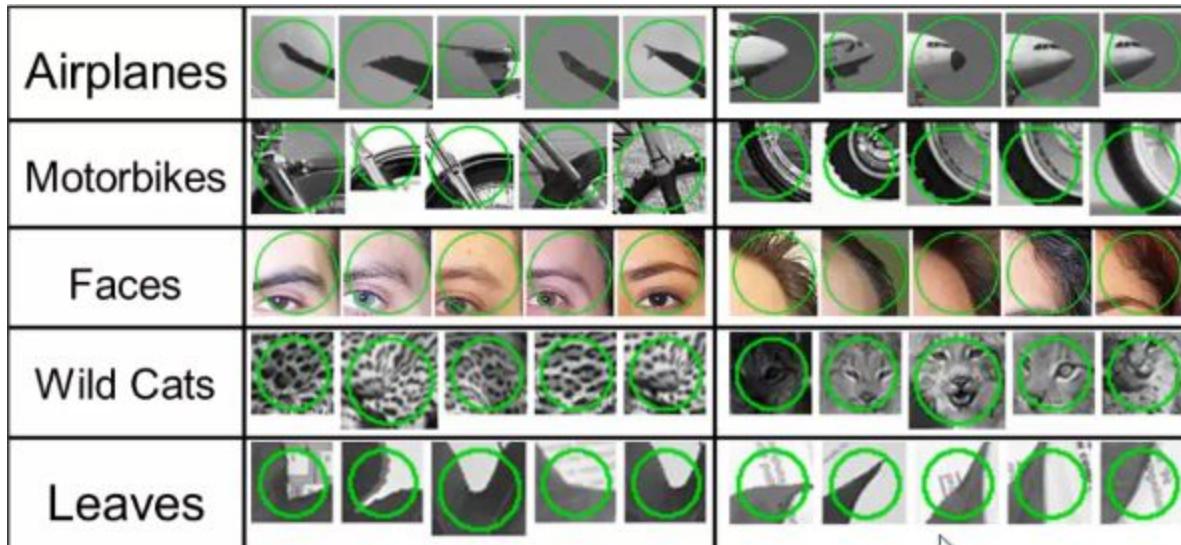
M. Goudarzi 5.12.2016

HMAX:
A CNN-based
bio-inspired
Object Recognition
approach

[20]
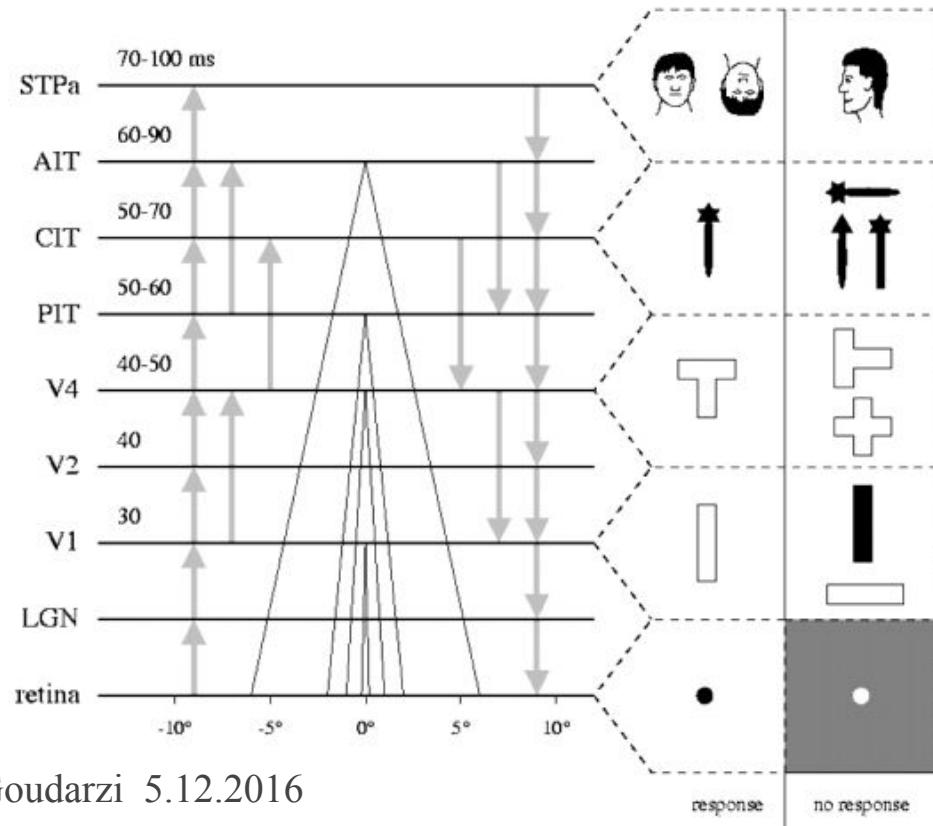
M. Goudarzi 5.12.2016

HMAX:
A CNN-based bio-inspired Object Recognition approach

[20]
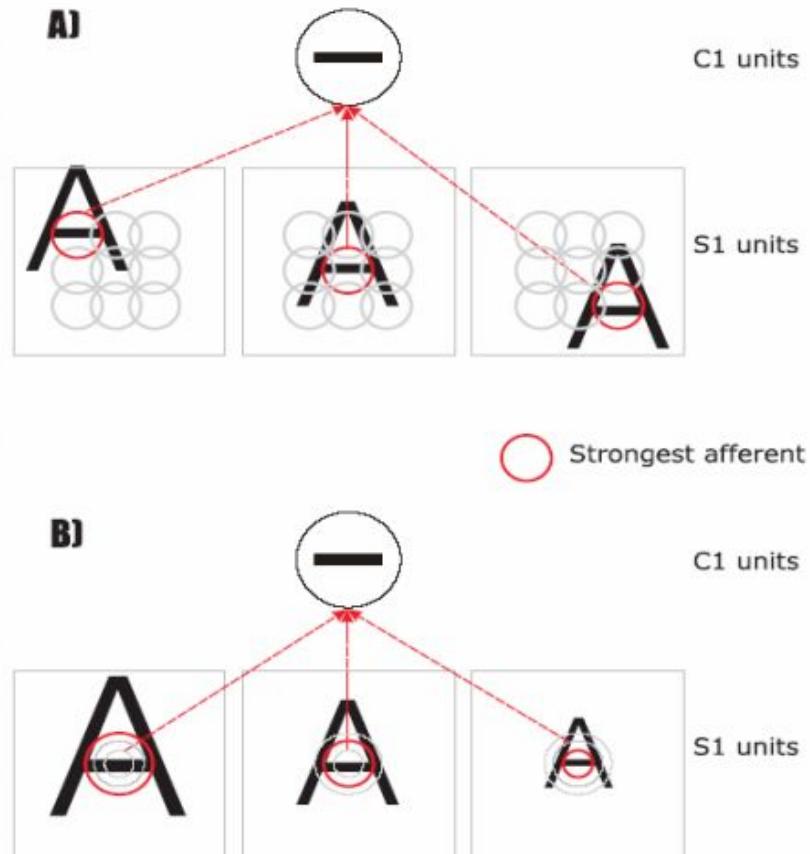
Convolutional Neural Network Real-Time Face Detection

M. Goudarzi 5.12.2016

[21]

Comparison between SIFT-based vs. HMAX Object Recognition Approach.
SIFT: pros and cons.

Advantage
accuracy, stability, scale & rotational invariance
affine invariance
Efficiency / speed
reliable pose models
better error tolerance with fewer matches
reliability

Disadvantages
Fundamentally different from human brain mechanism.
Loses spatial information
Requires careful tweeting
If not used carefully can include noises into features

M. Goudarzi 5.12.2016

Comparison between SIFT-based vs. HMAX Object Recognition Approach.
CNN: pros and cons.

Advantages	Disadvantages
Use of shared weight for C-layer	Requires intensive computational power and taking too long to train
Independent from human effort	Too much of a "Black Box"
Invariance to certain features	Difficult to add training samples later on
Closer to human brain mechanism	Difficult to use properly, more knowledge demanding

References:

[1] http://kanigas.com/donald-trump-2/

[2] David Labov, http://www.skilja.de/2012/classification-and-context/, last accessed
http://www.skilja.de/wp-content/uploads/2012/03/Labov-Cups-2.png

[4] Sacks, O. (1985). *The man who mistook his wife for a hat and other clinical tales*. New York: Summit Books. Photo available via
http://t3.gstatic.com/images?q=tbn:ANd9GcT1idlXjD7CkbIAv3Kk2-riy_Tk_8RiUE3mnlfU55KQUnslhyEa

[5] Levitin, D. J. (2014). *The organized mind: Thinking straight in the age of information overload*. New York, NY: Dutton. Photo available via:
http://blogs.lse.ac.uk/impactofsocialsciences/files/2015/01/9780670923106-1.jpg

[6] Thinking, Fast and Slow. (2015). *College Music Symposium, 55*. doi:10.18177/sym.2015.55.ca.10990. Photo available via:
http://2.bp.blogspot.com/-f7SFFKhuXn0/UfIzrpGguSI/AAAAAAAAAG0/0X-W0YZp7rw/s1600/Thinking+Fast+and+Slow.jpg

M. Goudarzi 5.12.2016

References (cont.)

[7] optical illusion cube
 http://www.nerdist.com/wp-content/uploads/2015/02/DressIllusion_3.jpg

[8] optical illusion cube revealed
http://news.bbcimg.co.uk/nol/shared/bsp/hi/dhtml_slides/10/illusion3/img/illusion_dhtml_7_v2.gif

[9] Fei Fei, Stanford, TED Talk.
https://www.youtube.com/watch?v=40riCqvRoMs&t=217s

[11] Austrian Child embracing shoes - 1946
http://65.media.tumblr.com/tumblr_mcb4x5GoH61qgwmzso1_r1_1280.jpg
https://dl.dropboxusercontent.com/u/4001169/TUMBLR/BLOG%20-%20FROM%20A%20TO%20B/PHOTOS/34117773241_gerald_waller_LARGE.jpg
First published in LIFE Magazine.
[12]
https://www.youtube.com/watch?v=vAEFmurII-A

M. Goudarzi 5.12.2016

References (cont.)

[13] David Lowe
http://www.cs.ubc.ca/~lowe/photoCredit.html

[14] Object Recognition using SIFT
http://www.di.ens.fr/willow/teaching/recvis10/assignment1/

[15] VLFEAT SIFT
http://www.vlfeat.org/overview/sift.html

[16] http://homepages.inf.ed.ac.uk/rbf/HIPR2/log.htm

[17] Open CV - SIFT Features.
http://docs.opencv.org/trunk/d5/d3c/classcv_1_1xfeatures2d_1_1SIFT.html

M. Goudarzi 5.12.2016

[18] Bag of Visual Words model
http://www.robots.ox.ac.uk/~az/icvss08_az_bow.pdf

[19]   Serre, T. and Riesenhuber, M. (2004)

[20]   https://www.quora.com/What-are-the-pros-and-cons-of-neural-networks-from-a-practical-perspective.

[21]   http://maxlab.neuro.georgetown.edu/hmax.html

[20]   https://www.quora.com/What-are-the-pros-and-cons-of-neural-networks-from-a-practical-perspective.

[21] https://www.youtube.com/watch?v=ptzpJwtbPp0

M. Goudarzi 5.12.2016