COLUMBIA UNIVERSITY

An example of
Hybrid Vision

Brain State Decoding for Rapid Image Retrieval

Shih-Fu Chang

Department of Electrical Engineering

Columbia University

joint work with

Jun Wang, Eric Pohlmeyer, Barbara Hanna
Yu-Gang Jiang, and Paul Sajda

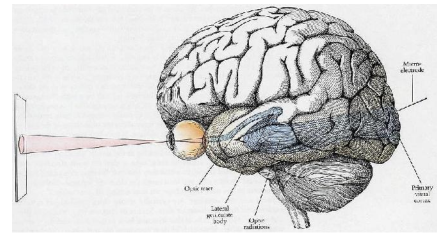
supported by funding from DARPA

Image Retrieval in the "Blink of an Eye"

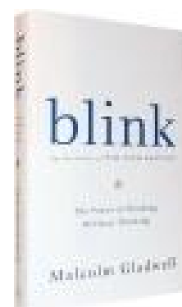
- Volumes of imagery and media
- Can we smartly "triage" information?

Image Retrieval in the “Blink of an Eye”

- Volumes of imagery and media
- Can we smartly “triage” information?
- Human Vision:
  Superb by quick “gist”
  But limited in throughput (5-10Hz)



(Hubel, 1995)



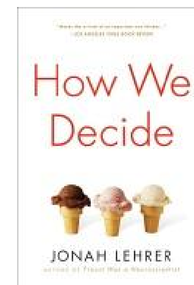
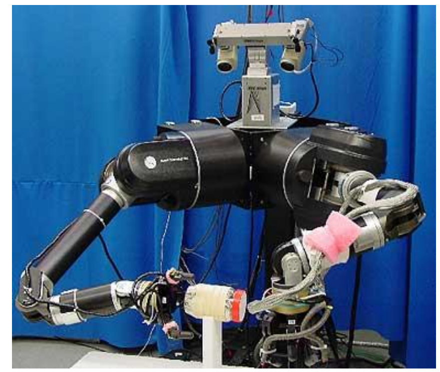


Image Retrieval in the “Blink of an Eye”

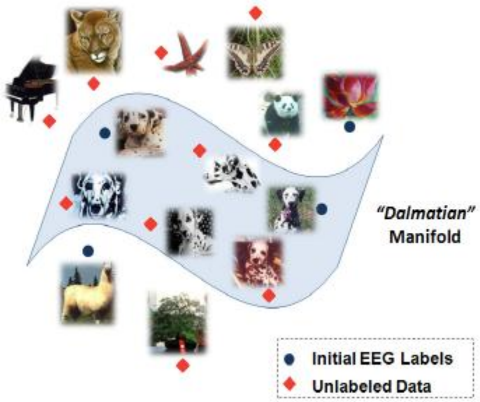
- Volumes of imagery and media
- Can we smartly “triage” information?
- Human Vision:
  Robust and quick “gist”
  But still limited in throughput
- Computer Vision:
  Can be very fast but suffer from
  sensitivity to variance and low accuracy



www.cs.umass.edu

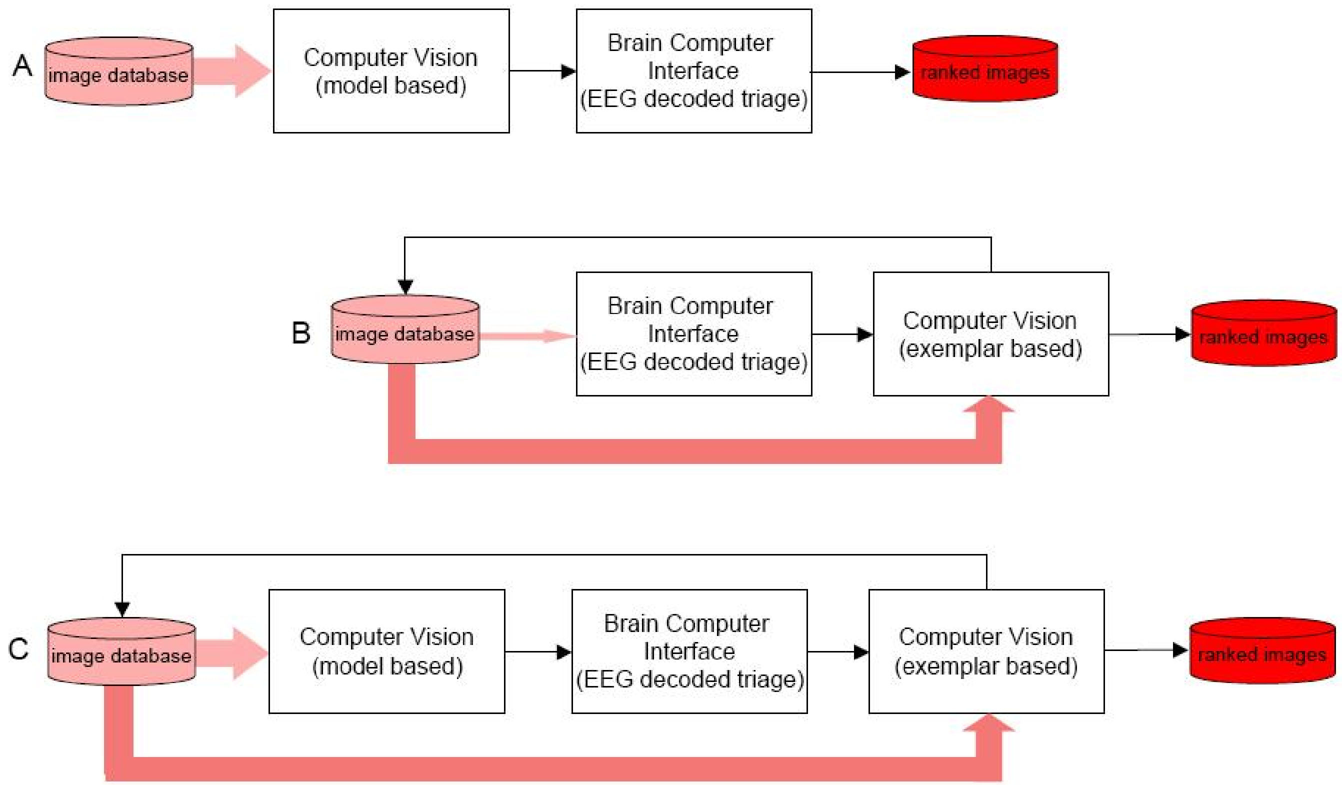
Image Retrieval in the "Blink of an Eye"

- Goal: optimally integrate neuro-vision and computer vision/machine learning to maximize information throughput and retrieval accuracy of image content



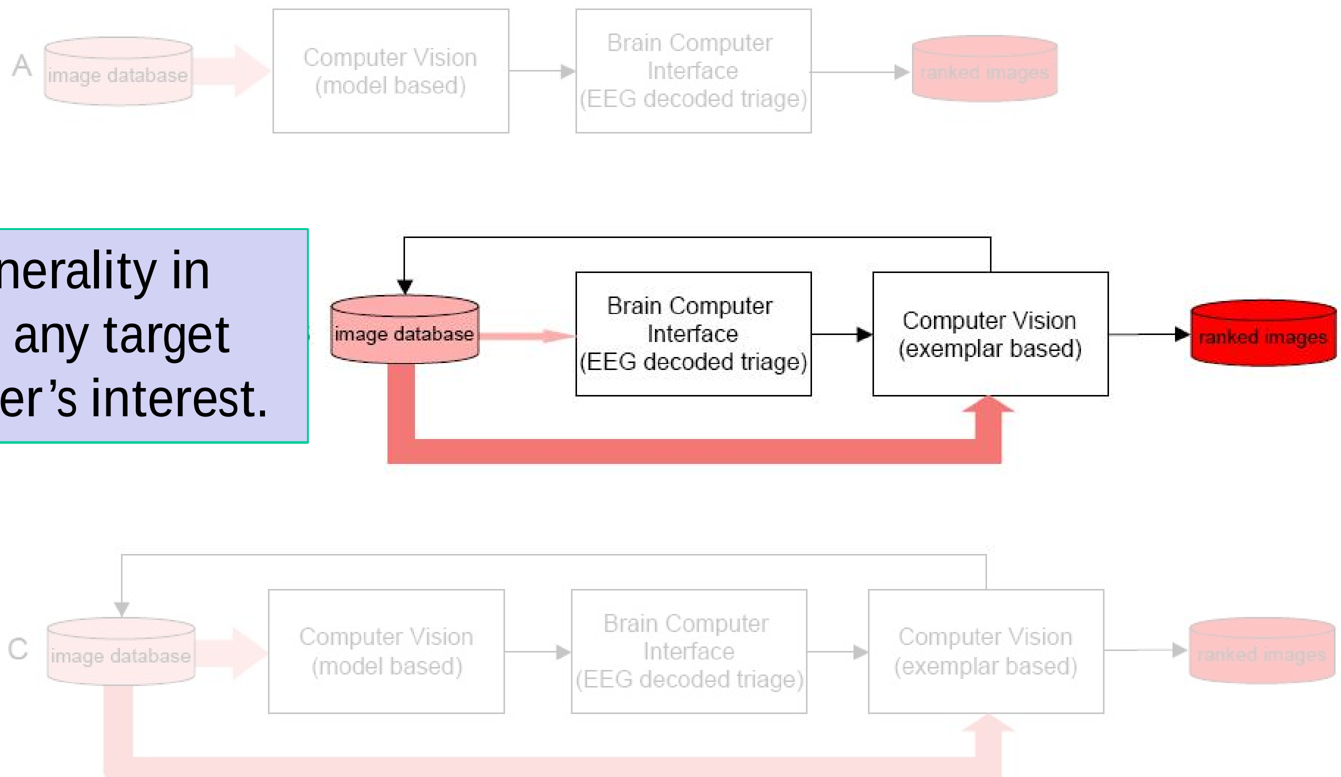
Graph-Based Visual Pattern Discovery

System Designs: Integration of HV (EEG) and CV



COLUMBIA UNIVERSITY

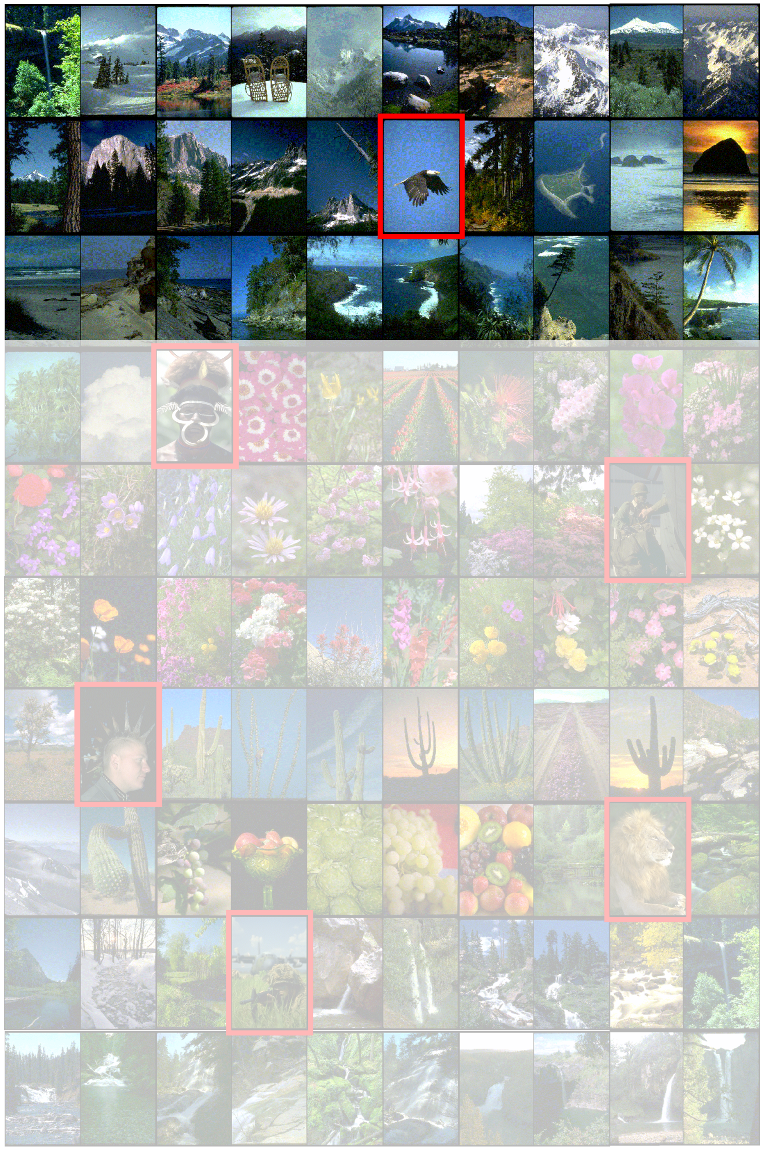
System Designs: Integration of HV (EEG) and CV



Goal: generality in detecting any target catching user's interest.

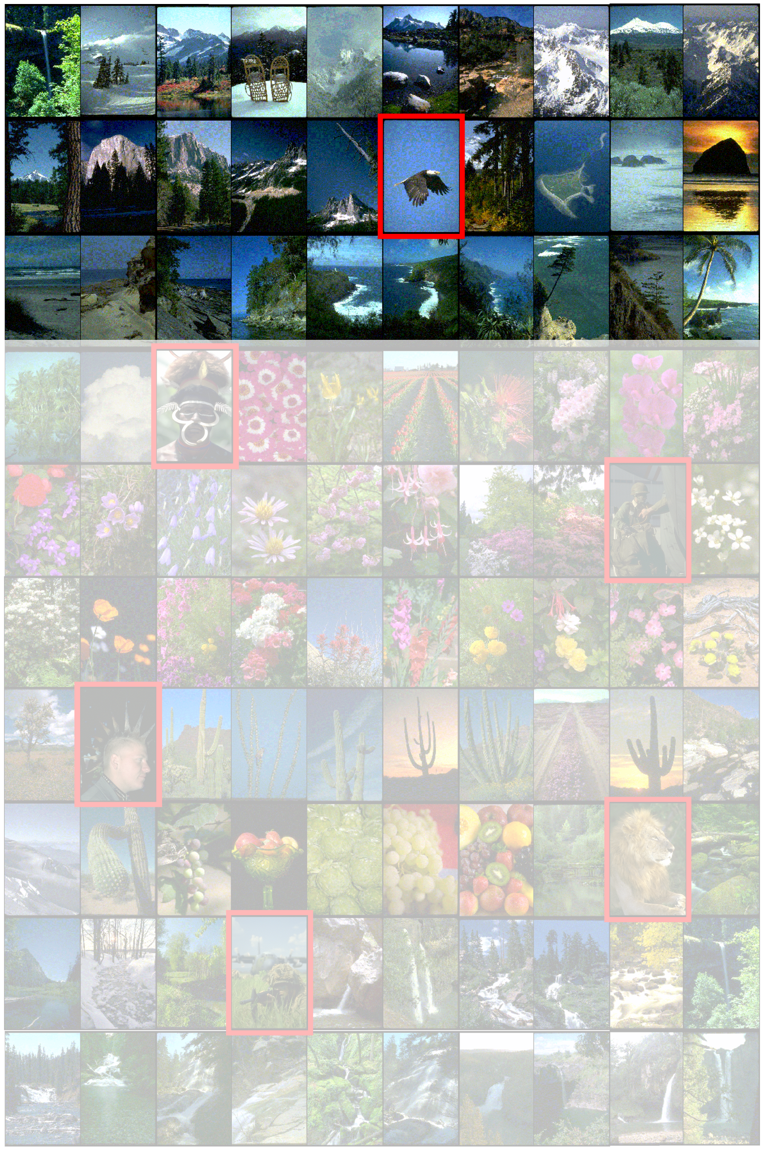
COLUMBIA UNIVERSITY

The Paradigm

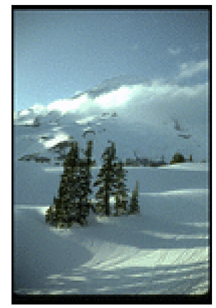


Database (any target that may interest users)

The Paradigm



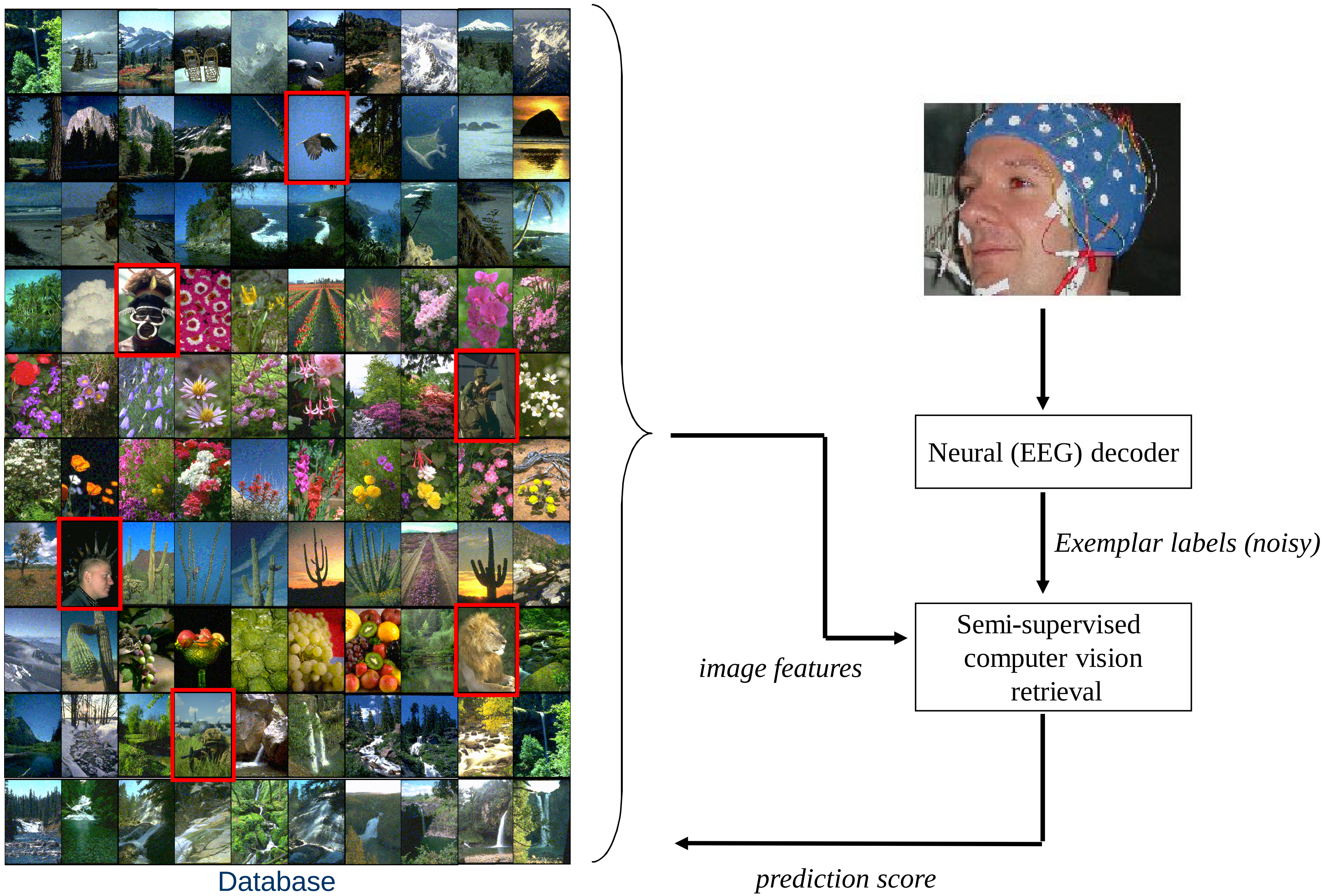
Database



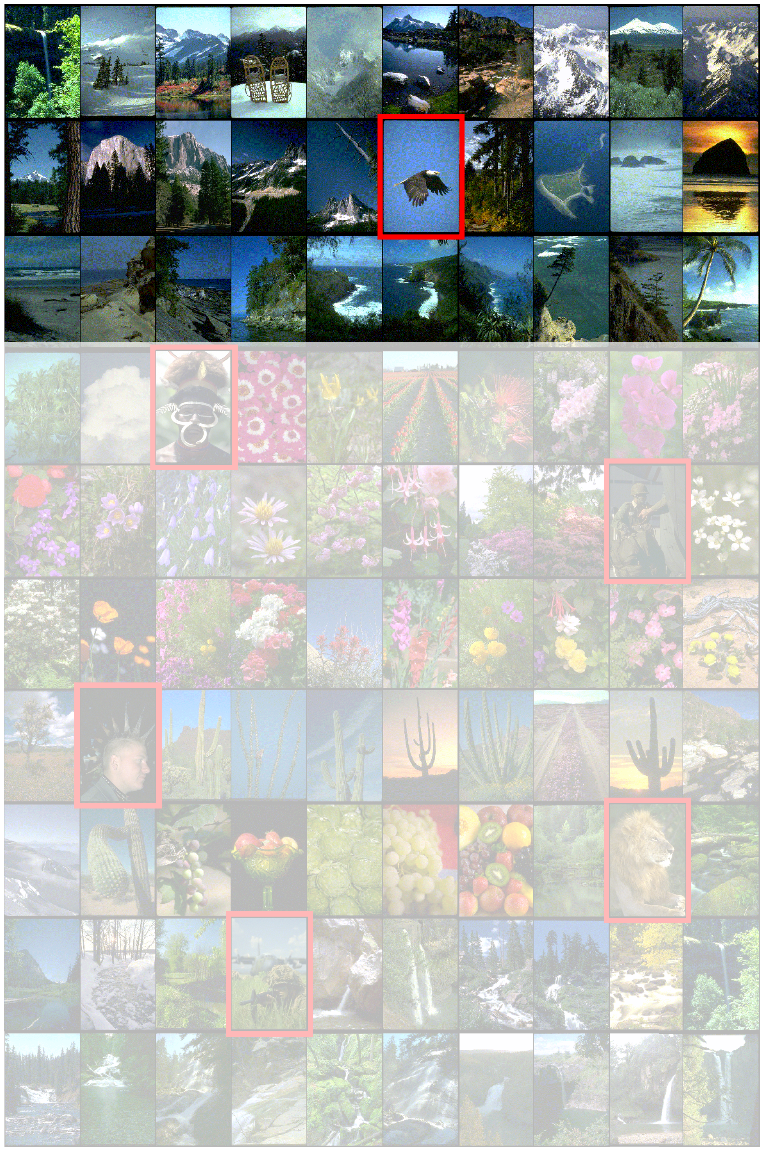
Neural (EEG) decoder

EEG-scores

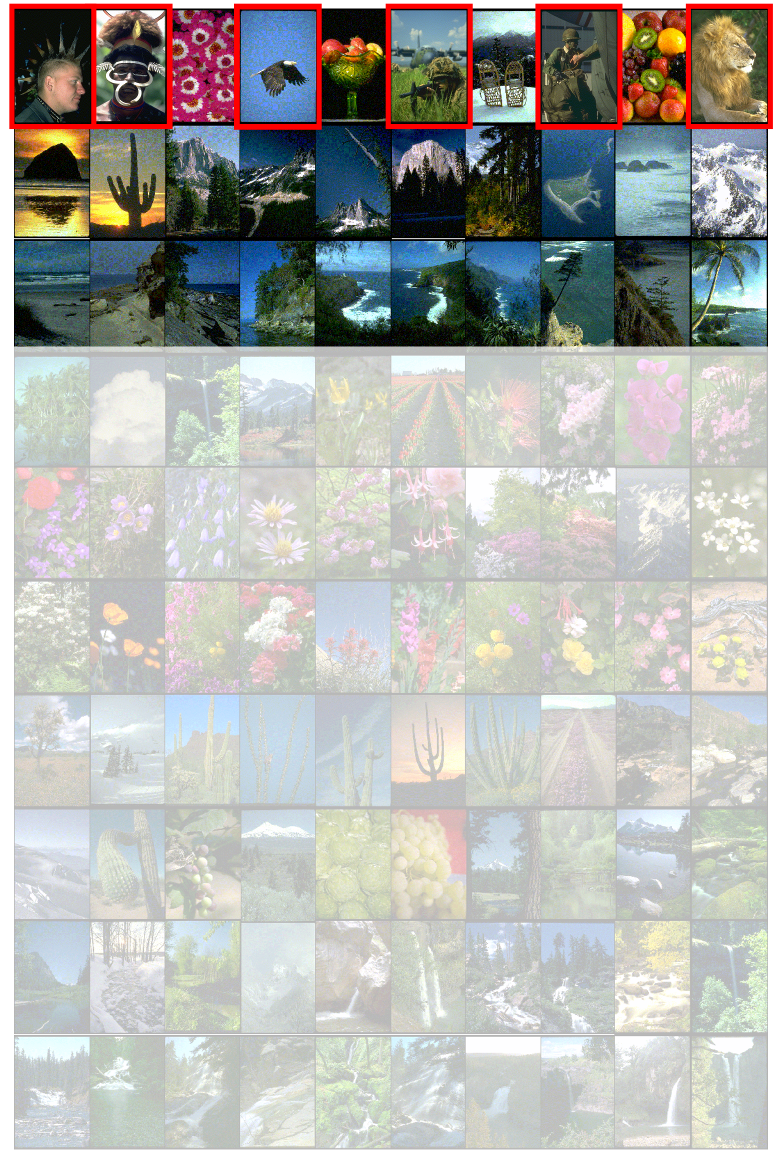
The Paradigm



The Paradigm

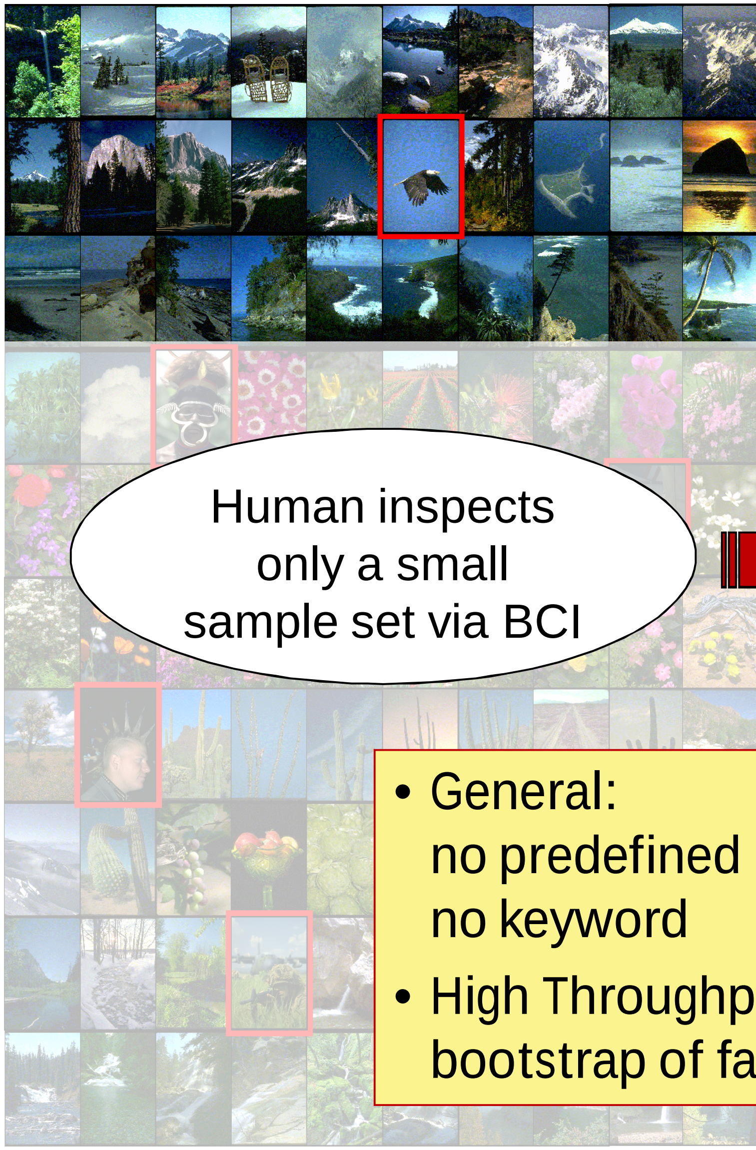


Pre-triage



Post-triage

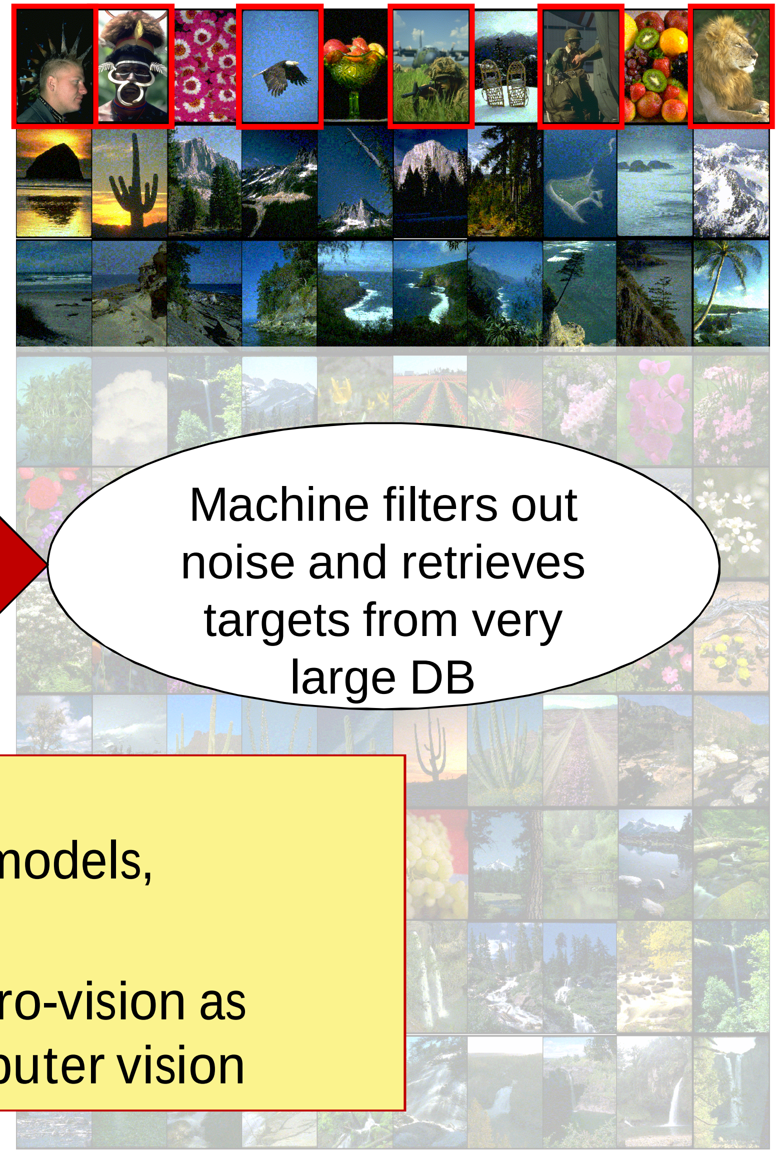
The Paradigm



Pre-triage

Human inspects only a small sample set via BCI

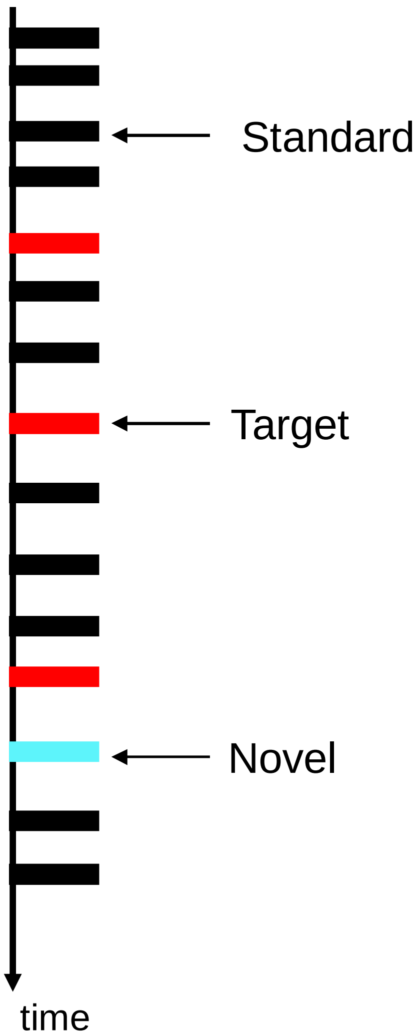
Machine filters out noise and retrieves targets from very large DB

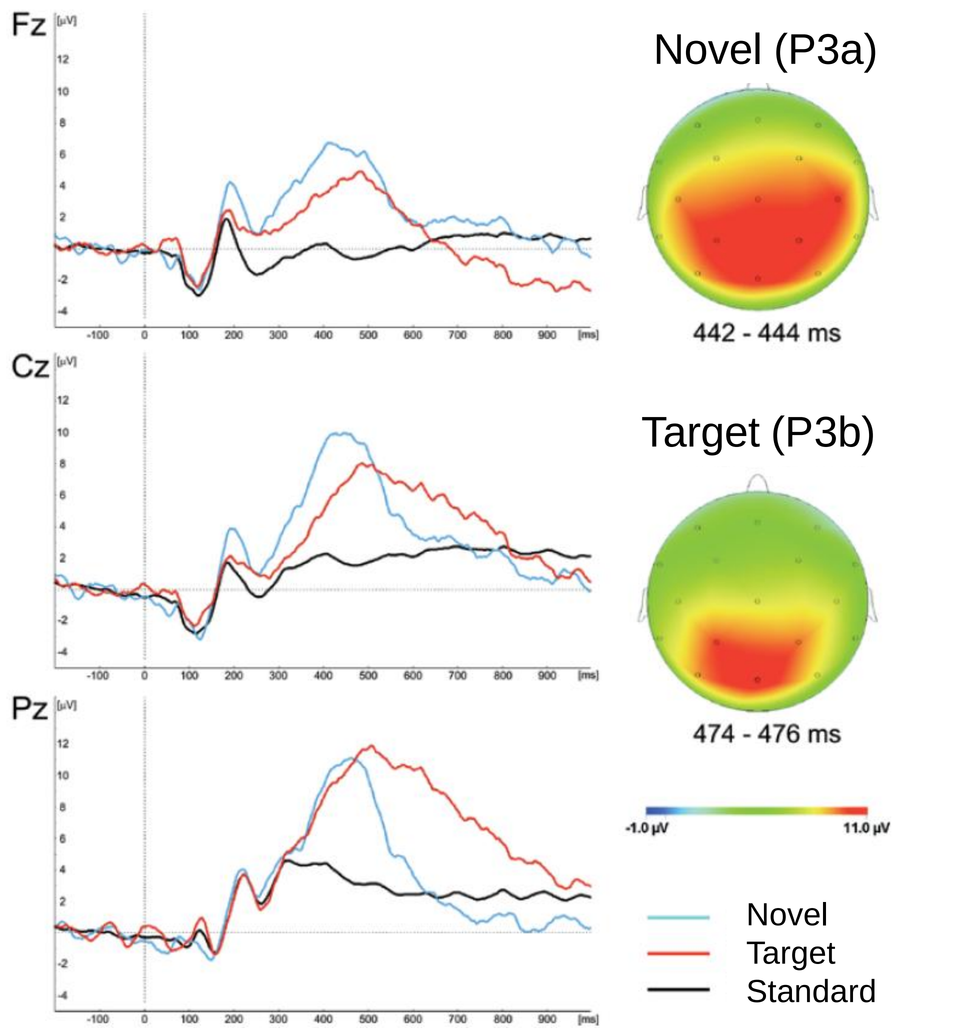


Post-triage

- General:
  no predefined target models,
  no keyword
- High Throughput: neuro-vision as bootstrap of fast computer vision

The Neural Signatures of “Recognition”

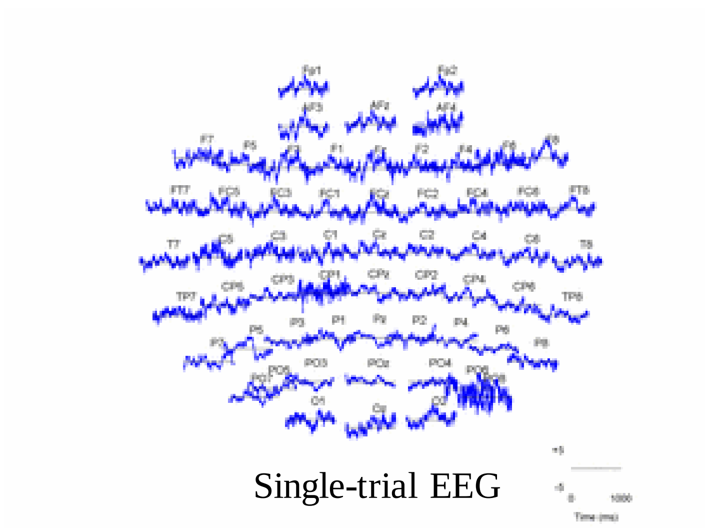




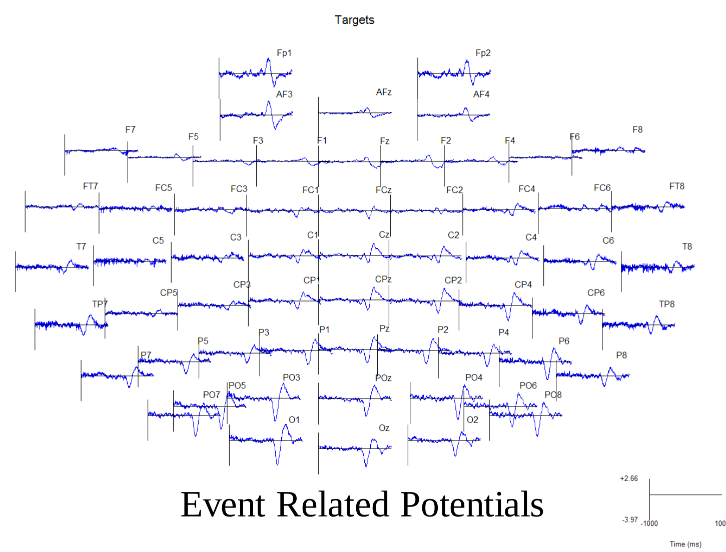
From D. Linden, 2005

Single-trial EEG Analysis

- Typically EEG is averaged over trials to increase the amplitude of the signal correlated with cortical processes relative to artifacts (very low SNR)
- High-density EEG systems were designed without a principled approach to handling the volume of information provided by simultaneously sampling from large electrode arrays.
- Our solution: identifying neural correlates with individual stimuli via single trial EEG analysis.
- We apply principled methods to find optimal ways for combining information over electrodes and moments in time contained in individual trials



Single-trial EEG

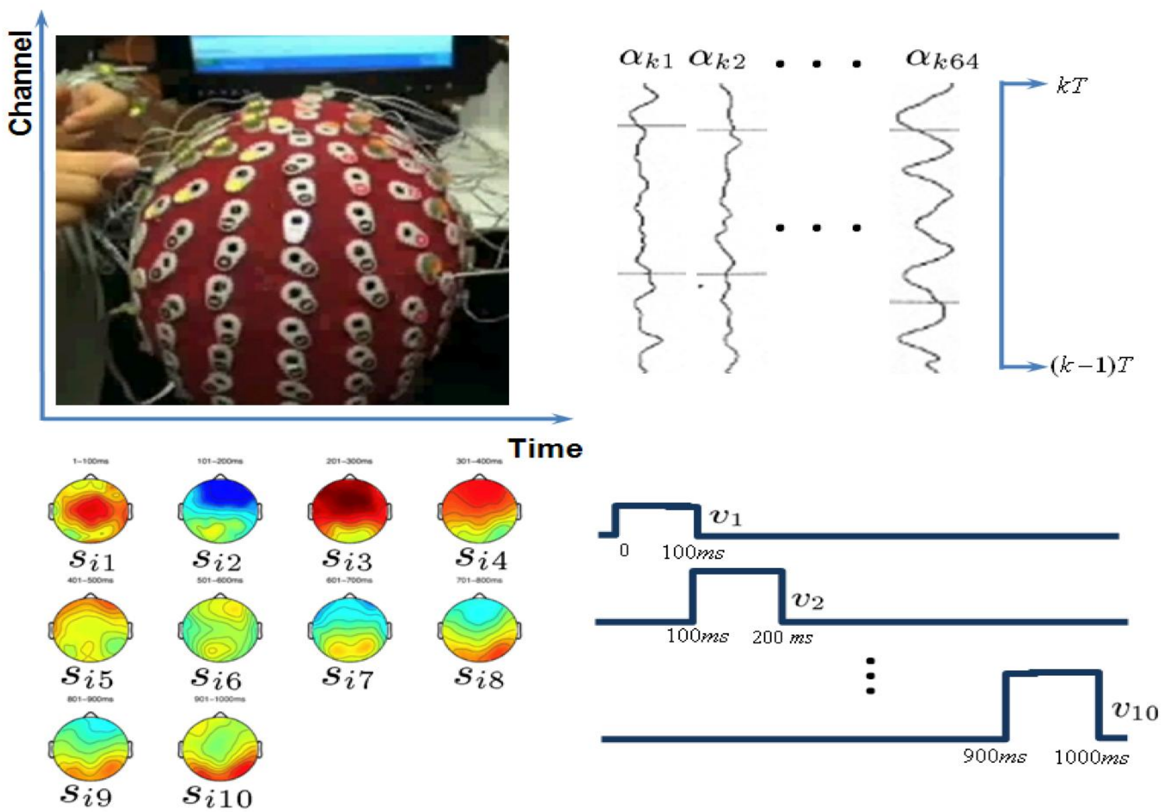


Event Related Potentials

Identifying Discriminative Components in the EEG Using Single-Trial Analysis

LDA or Logistic Regression is used to learn the contributions of EEG signal components at different spatial-temporal locations

(Parra, Sajda et al. 2002, 2003)

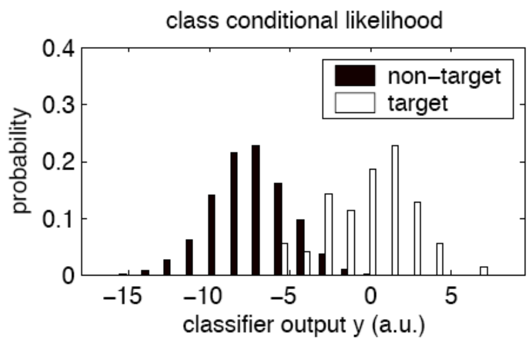


Optimal spatial filtering across electrodes within each short window (e.g., 100ms)

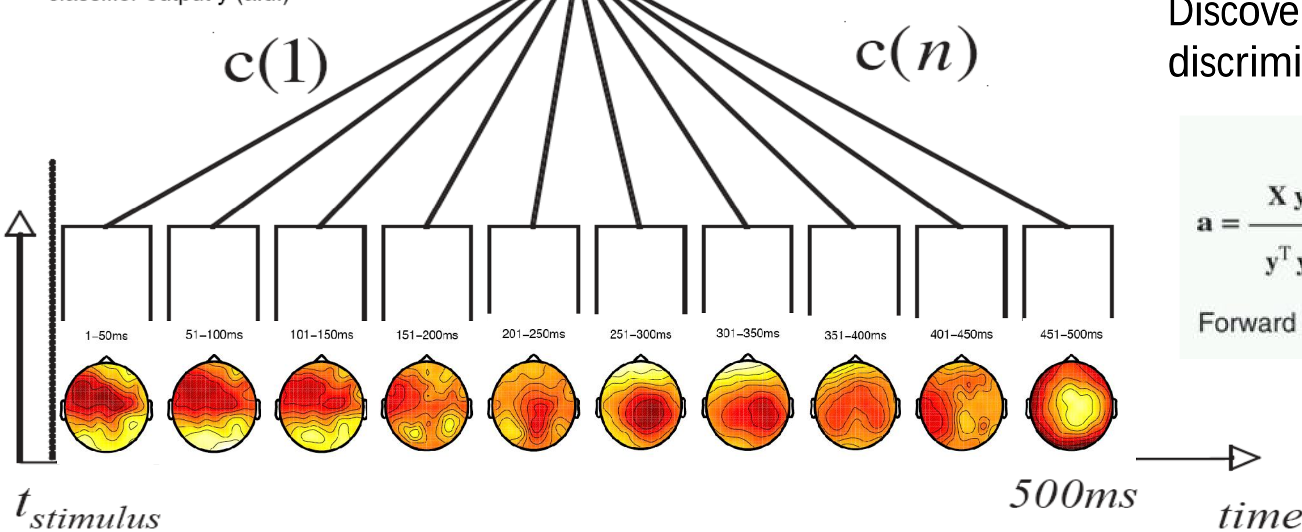
Optimal temporal filtering over time windows after onset

Hierarchical Discriminant Components

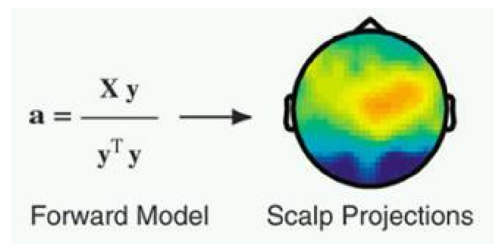
… use factorization to greatly reduce the number of parameters (100K -> 100)…



$$z = \mathbf{c}^T < \mathbf{y}(t) >$$

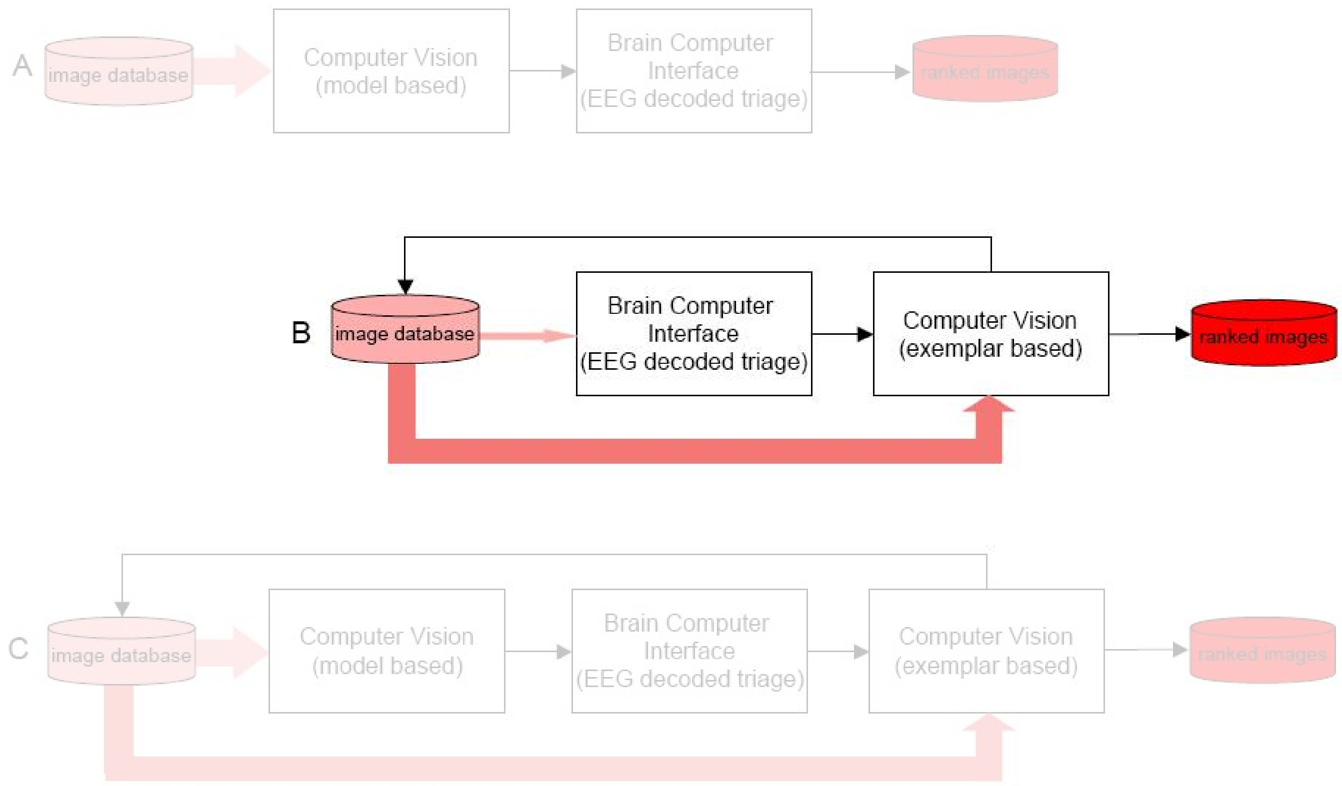


Discover distribution of discriminative components

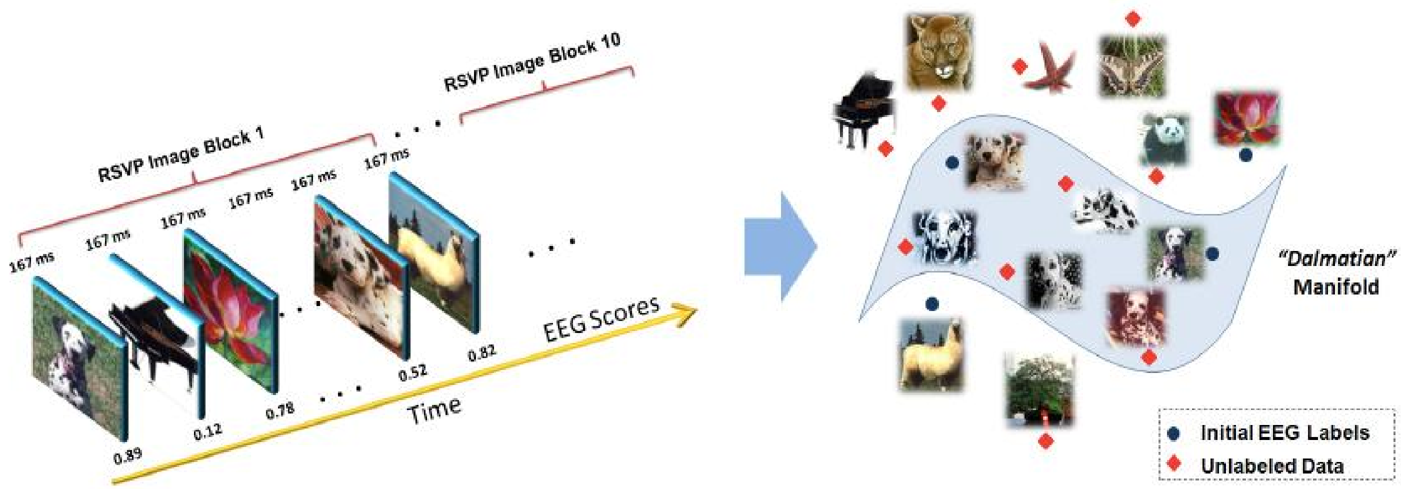


$$\mathbf{y}(t) = \mathbf{w}_\tau^T \mathbf{x}(t) \qquad \tau = \left\{ t_k - \frac{\delta}{2} \dots t_k + \frac{\delta}{2} \right\}$$

System Designs: Integration of HV (EEG) and CV



EEG “Teacher” and Manifold Learning



Rapid Serial Presentation of Caltech 101 Images

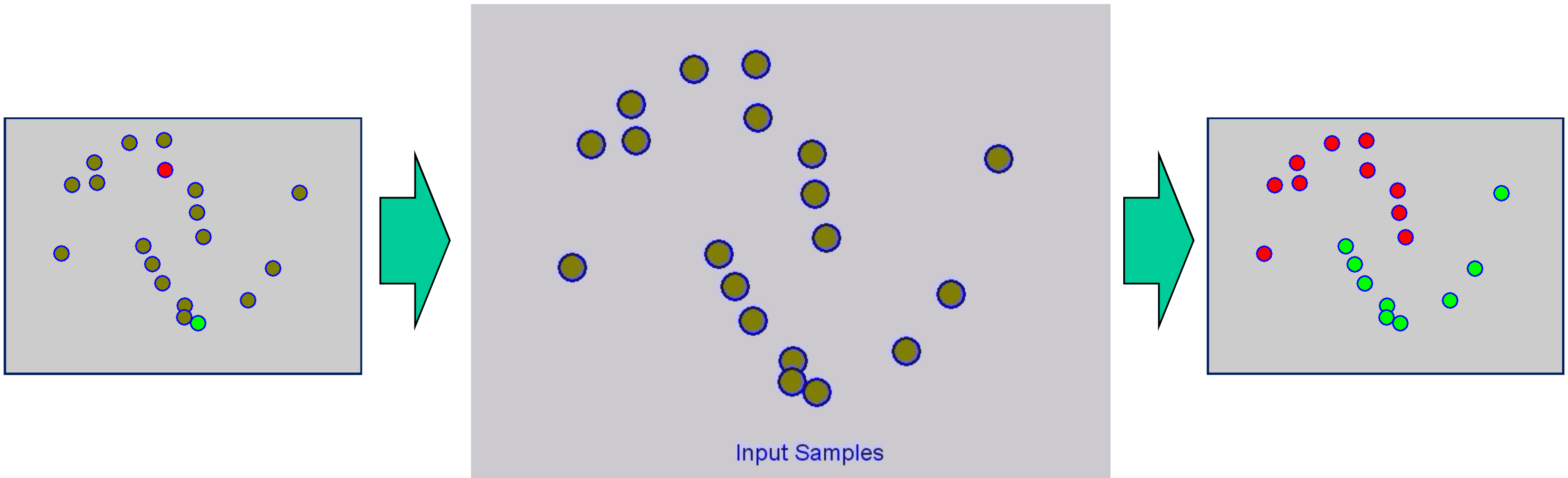
Graph-Based Visual Pattern Discovery

Opportunities and Issues:

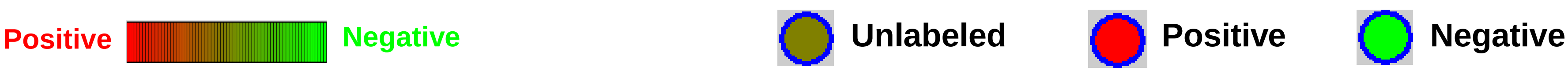
- EEG results used as exemplars indicating user interest
- Propagate “interest” scores over manifolds in the image space
- But EEG labels are noisy and limited
- No prior knowledge about target models

Graph-based Semi-Supervised Learning in the Image Space

- Given few noisy labels and a large # of unlabeled data



Input samples with sparse labels | Label propagation on graph | Label inference results



$$f^* = \min_f Q(f, y, \mathcal{G}(V, W))$$

$\mathcal{G}$ -- graph

V -- graph node

W -- weight matrix

Q -- risk function

f -- prediction

y -- label matrix

A hot topic in Machine Learning

- Given initial labels, Y, find predictions F over all nodes

Prediction smoothness

Fitness with given labels

$$\begin{aligned} Q(F) &= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\left\|\frac{F_{i\cdot}}{\sqrt{D_{ii}}}-\frac{F_{j\cdot}}{\sqrt{D_{jj}}}\right\|^2+\mu\sum_{i=1}^{l}\|F_{i\cdot}-Y_{i\cdot}\|^2 \\ &= \mathrm{tr}\{F^\top LF+\mu(F-Y)^\top(F-Y)\} \end{aligned}$$

(Zhou, et al NIPS04)

Gaussian fields & Harmonic functions (*Zhu et al ICML03*)

$$\boxed{Q(F)=\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\|F_{i\cdot}-F_{j\cdot}\|^2}$$

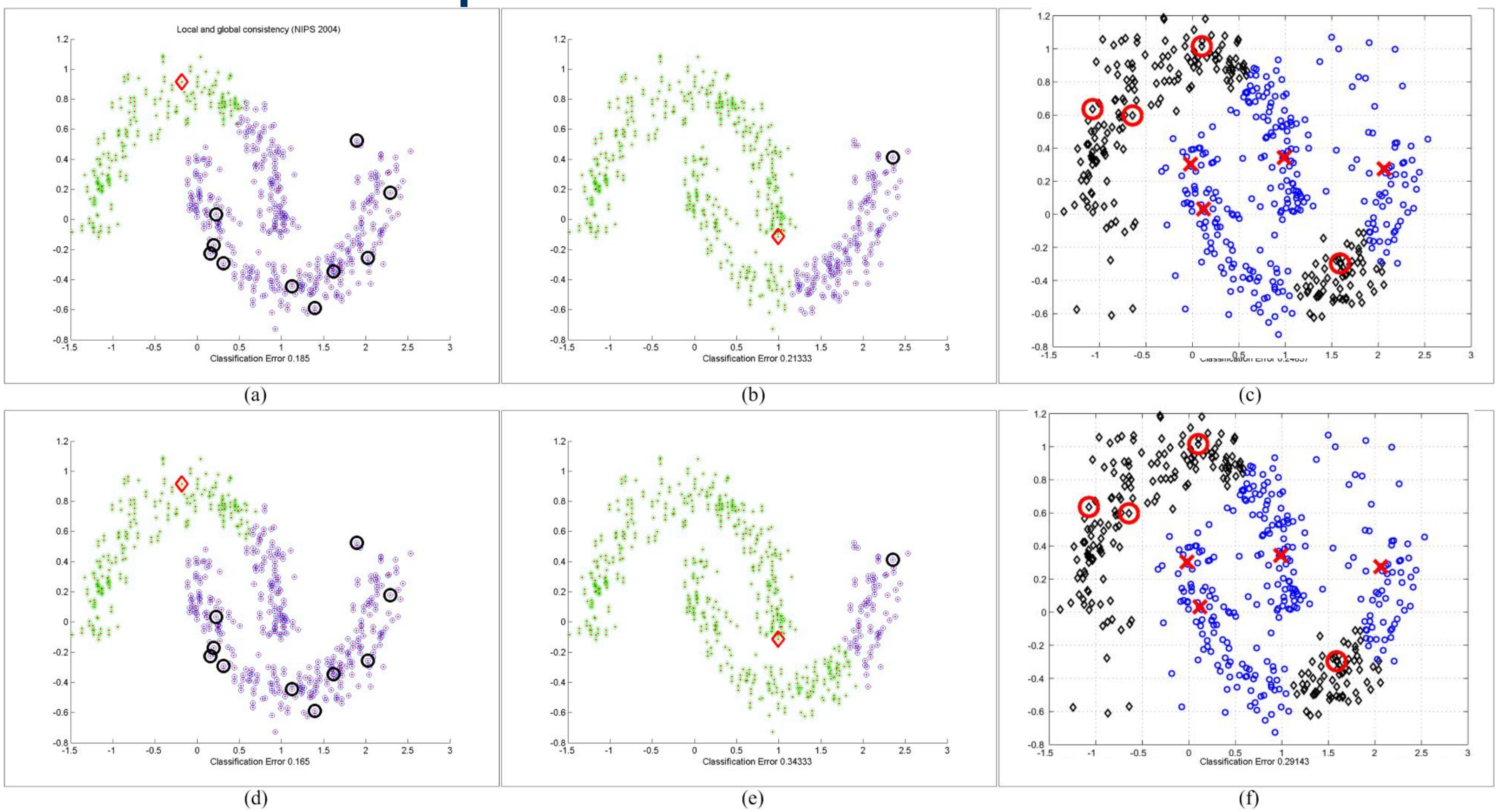
1) $\triangle F = 0$ on unlabeled data, where $\triangle = D - W$ is the graph Laplacian;

2) $F_{i\cdot} = Y_{i\cdot}$ on labeled data.

Important Issues

LGC Method

GFHF Method



Unbalanced Labels

Bad Label Locations

Noisy Labels

Graph Transduction via Alternating Minimization (GTAM)

(Wang, Jebara, Chang, ICML08) (Wang and Chang, CVPR09)

-- Bivariate Optimization over Labels (Y) and Prediction (F)

$$\boxed{\mathcal{Q}(\mathbf{F},\mathbf{Y}) = \frac{1}{2}\text{tr}\left\{\mathbf{F}^T\mathbf{L}\mathbf{F} + \mu(\mathbf{F}-\mathbf{V}\mathbf{Y})^T(\mathbf{F}-\mathbf{V}\mathbf{Y})\right\}}$$

- Label Tuning Step
  - Iteratively remove bad labels and add good labels

$$\mathcal{Q}(\mathbf{Y}) = \frac{1}{2}\text{tr}\left(\mathbf{Y}^T\mathbf{V}^T\left[\mathbf{P}^T\mathbf{L}\mathbf{P} + \mu(\mathbf{P}^T-\mathbf{I})(\mathbf{P}-\mathbf{I})\right]\mathbf{V}\mathbf{Y}\right)$$

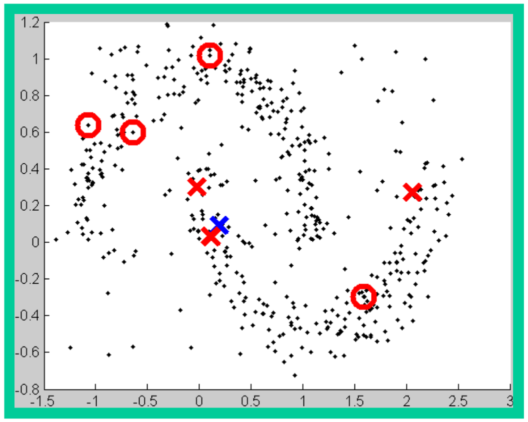
- Propagation Step
  - Given label (Y), find prediction F over graph

$$\frac{\partial \mathcal{Q}}{\partial \mathbf{F}^*} = 0 \Rightarrow \mathbf{F}^* = (\mathbf{L}/\mu + \mathbf{I})^{-1}\mathbf{V}\mathbf{Y} = \mathbf{P}\mathbf{V}\mathbf{Y}$$

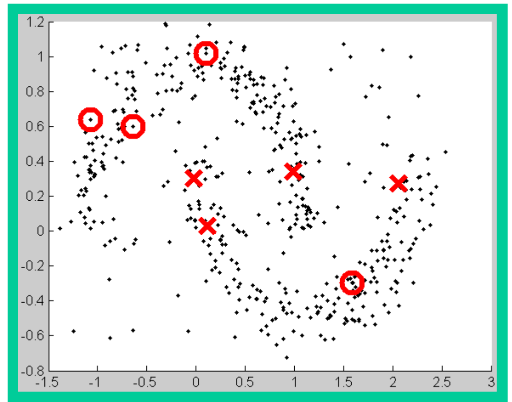
Label Diagnosis and Self Tuning

(Wang and Chang CVPR'09)

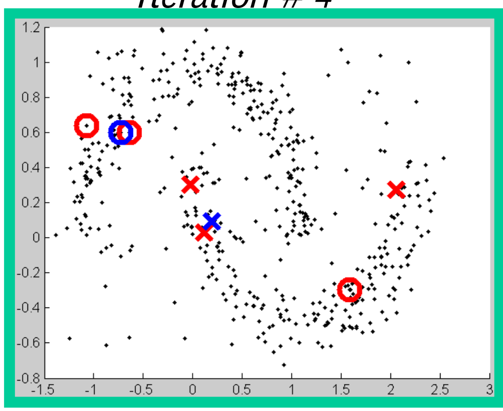
Iteration # 2



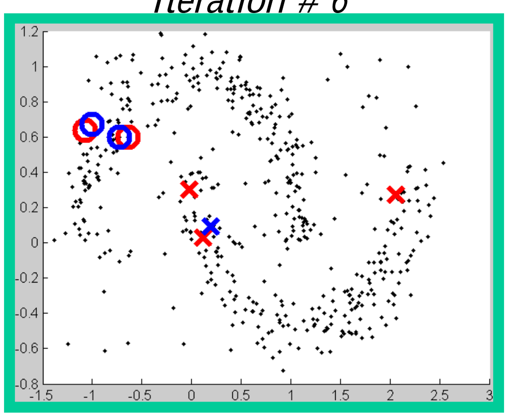
Initial Labels

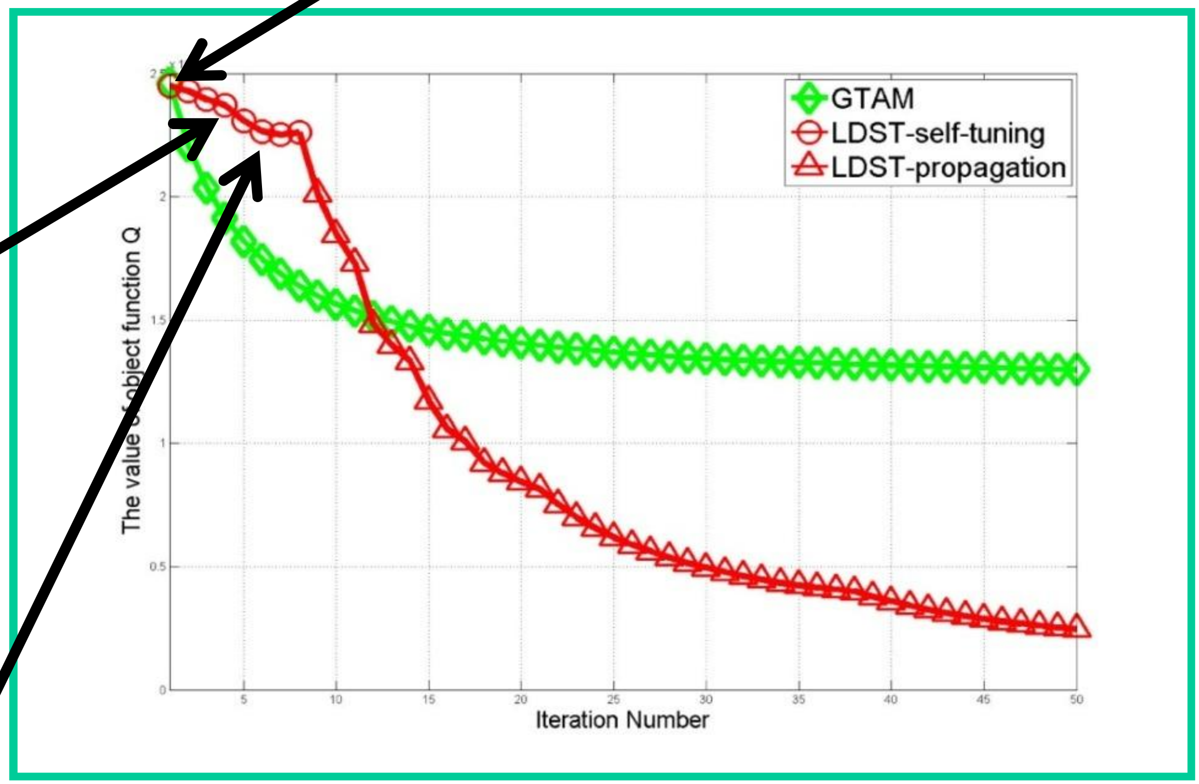


Iteration # 4



Iteration # 6

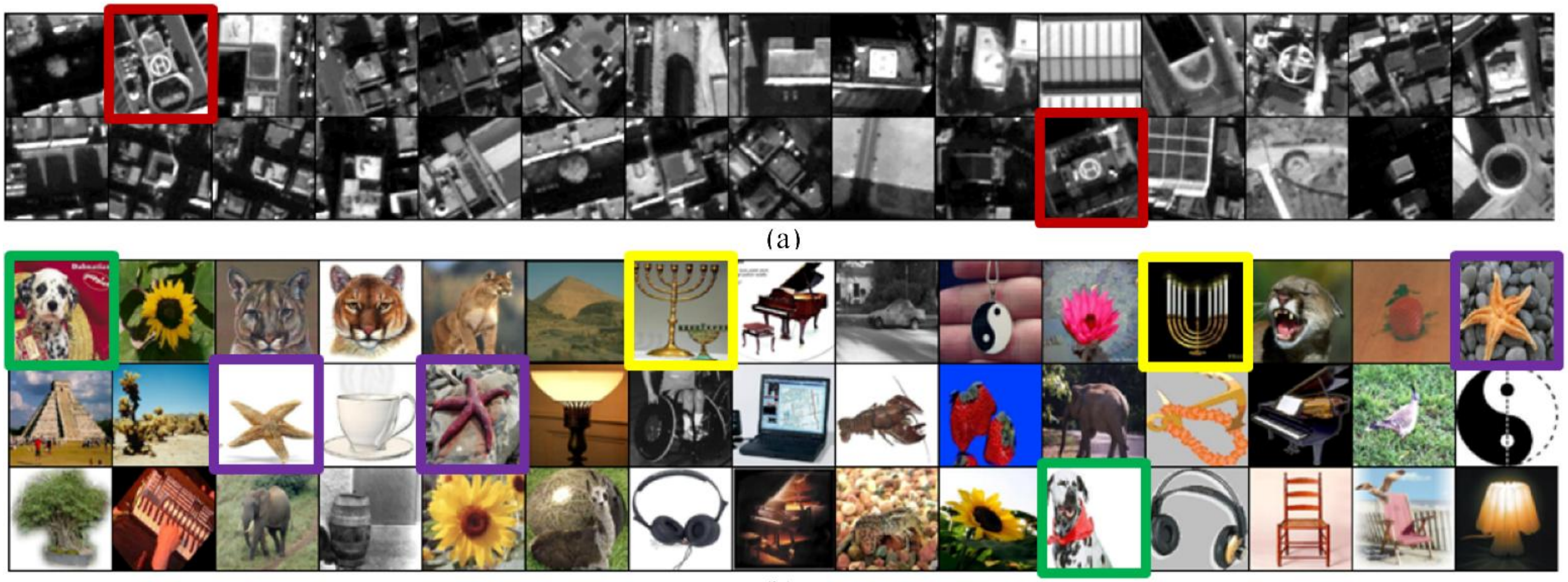




The values of the cost function Q during optimization procedure of LDST and GTAM methods.

Experiments

- CalTech101: 3798 images from 62 categories
  Satellite images
- EEG decoder trained per user using images *(Soccer Ball* or *Baseball Gloves)* from Caltech256
- A subset of 1000 images randomly sampled to construct 6-Hz RSVP sequence
- Initial Trials: 4 subjects, 3 targets (*Dalmatian*, *Chandelier/Menorah, & Starfish)*



(a)

(b)

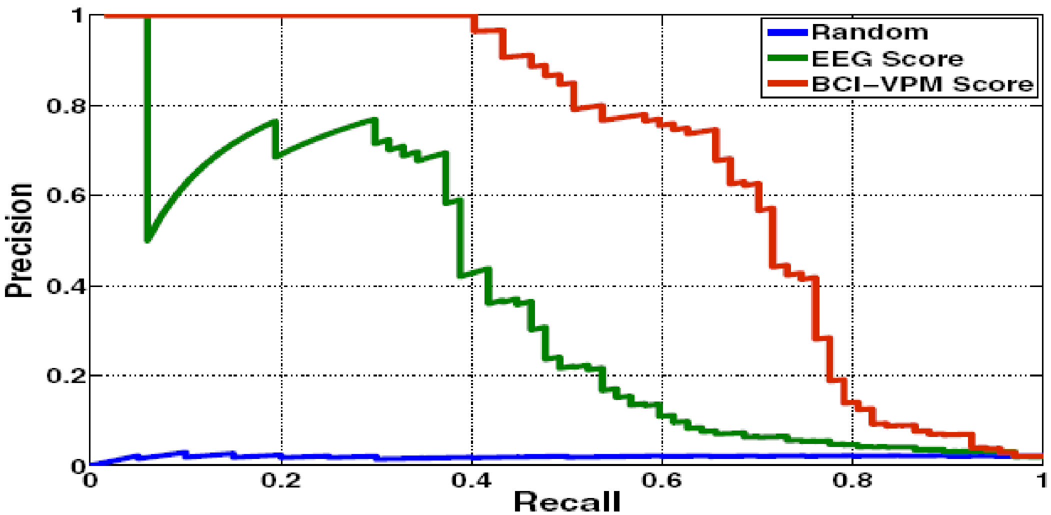
Example results

Top 20 results of EEG detection

(a)

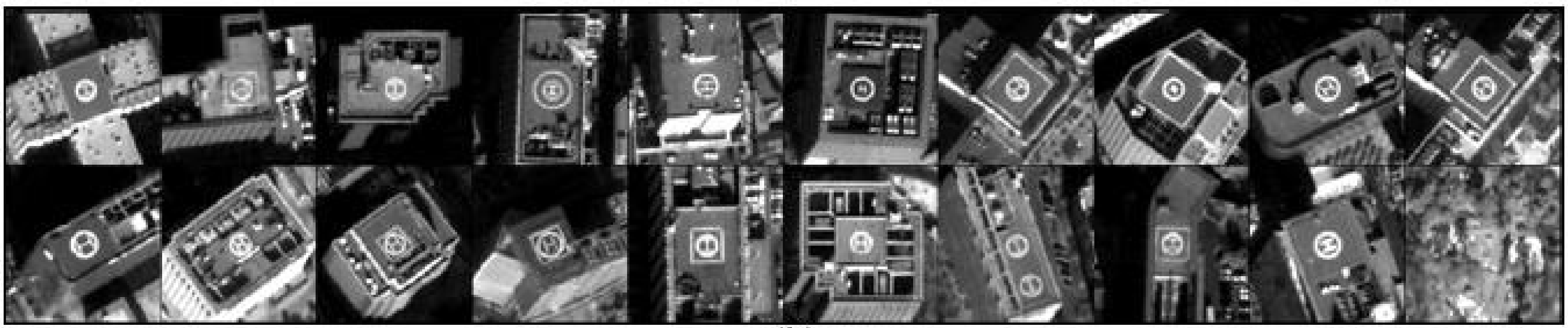
Top 20 results of Hybrid System (BCI-VPM)

(b)



Retrieval on Satellite Imagery

(a)



(b)

The experimental results of "*helipad" target RSVP, showing the top 20 ranked images* .
a) ranking by original EEG scores; b) ranking by the BCI-VPM refined interest score.

Performance Evaluation

- Neuro (EEG detector) vs.
  Hybrid Neuro-Computer System (BCI -VPM)

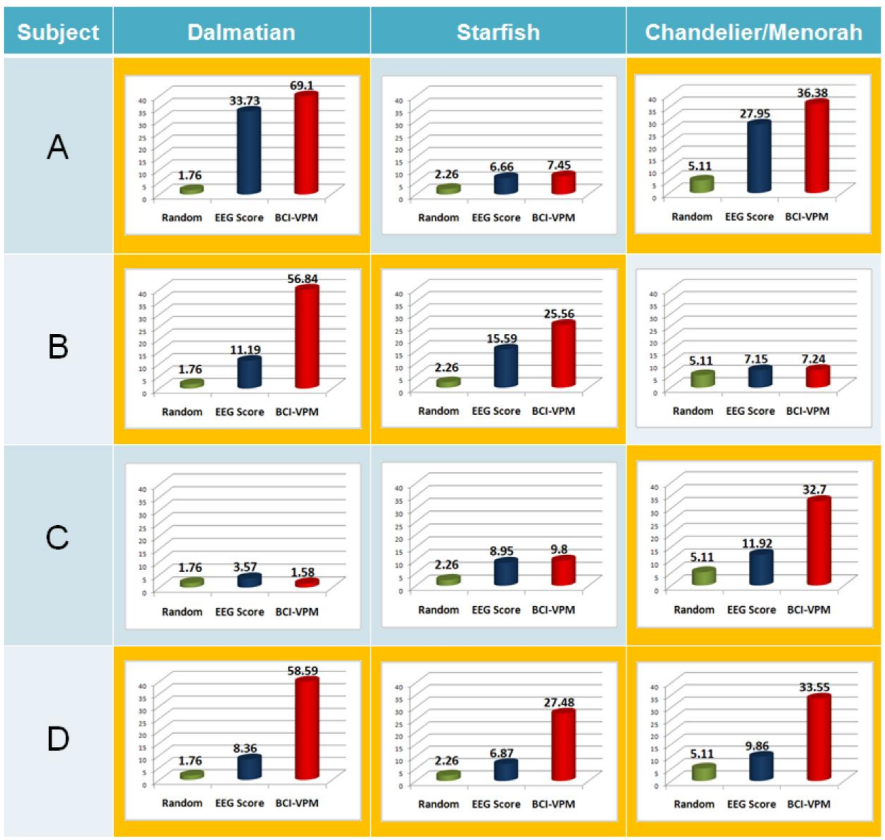
Subject	Method	AP-30	AP-60	AP-100	AP-ALL
A	EEG score	19.76	15.83	14.9	30.97
	BCI-VPM score	**50.19**	**32.65**	**25.46**	**37.89**
B	EEG score	8.76	9.79	9.56	23.71
	BCI-VPM score	**87.31**	**63.29**	**46.07**	**57.41**
C	EEG score	12.70	19.58	16.54	29.62
	BCI-VPM score	**90.82**	**63.30**	**41.21**	**53.66**
D	EEG score	10.68	11.87	11.75	24.45
	BCI-VPM score	**91.87**	**60.62**	**40.24**	**52.70**

(Varying depths on the ROC curve)

Dependency of Neuro & CV Components

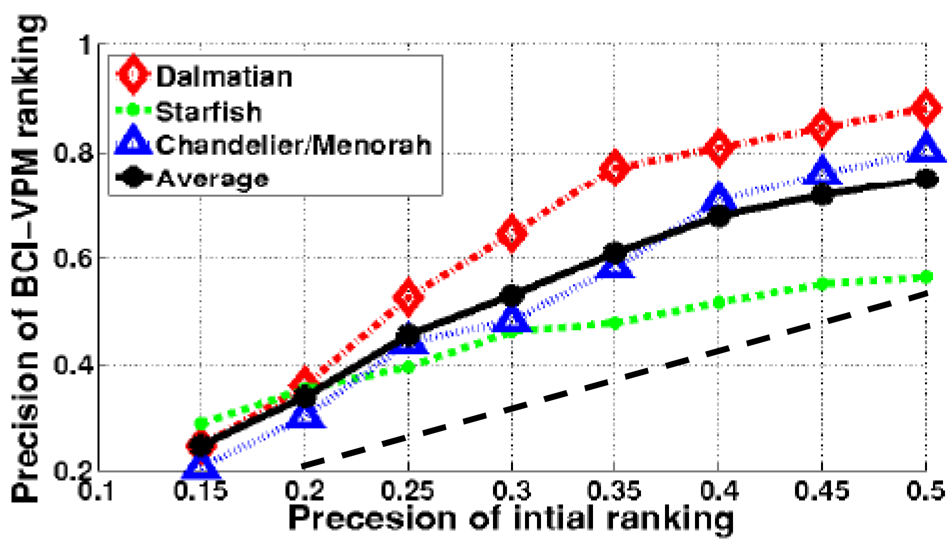
… not every case improves …

- Among 12 cases (4 subjects & 3 targets), 8 cases are clearly improved. When the EEG decoder fails, the hybrid system also fails.



Question:

- what's the required EEG accuracy for the hybrid system to work?
- are some categories more difficult?



EEG accuracy

What's Next

- Consider case when we do not instruct the subject to look for a particular class of image. **Pick an image, any image…**
  - "Readout" what has grabbed a user's attention/interest without prior cuing or instruction--the "mind reading" trick.
  - Deliver visually relevant content based on what is "interesting" -- can be subjective, based on **user's intent.**
  - Can integrate non-visual information (such as metadata).
- Current experiments are for one iteration. What if there is feedback?
  - Co-learning between subject and machine
    - Human learner
    - EEG machine learner
    - Computer Vision machine learner
- Systems for consumer use
  - Low cost EEG systems that are usable (wireless, simple to setup) - gaming applications
- Detector improvement in both neuro and computer vision components

References

- J. Wang, E. Pohlmeyer, B. Hanna, Y.-G. Jiang, P. Sajda, and S.-F. Chang, “Brain State Decoding for Rapid Image Retrieval,” ACM Multimedia, 2009.
- P. Sajda, E. Pohlmeyer, J. Wang, L. C. Parra, C. Christoforou, J. Dmochowski, B. Hanna, C. Bahlmann, M. K. Singh, and S.-F. Chang, "In a Blink of an Eye and a Switch of a transistor: Cortically Coupled Computer Vision," Proc of the IEEE, 98(3), pp 462-478, 2010.
- L. Parra, C. Christoforou, A. Gerson, M. Dyrholm, A. Luo, M. Wagner, M. Philiastides, and P. Sajda. Spatiotemporal linear decoding of brain state: Application to performance augmentation in high-throughput tasks. IEEE Signal Processing Magazine, 25(1):95–115, January 2008.
- J. Wang, Y.-G. Jiang, S.-F. Chang, “Label Diagnosis through Self Tuning for Web Image Search,” IEEE Computer Vision and Pattern Recognition (CVPR), 2009.
- J. Wang, T. Jebara, S.-F. Chang, “Graph Transduction via Alternating Minimization.,” International Conference on Machine Learning (ICML), 2008.

Comparison with Other Systems

- Comparison of BCI based image retrieval systems

System	**C3Vision** [8]	**HAC** [31]	**HAC-CV**[15]	**BCI-VPM**
System Structure	pure BCI	pure BCI	hybrid BCI+CV	hybrid BCI+CV
Neural Signal Trials	single trial single subject	single/multiple trials single/multiple subjects	multiple trials single/multiple subjects	single trial single subject
Object Class	people vs. background	face vs. animal animal vs. inanimate	face, animal, and inanimate	general object class
Target Frequency	2%	25%	50%	$\sim 2\%$
Manual Labels	No	Yes	Yes	No
Image Presentation Speed	$5-10$ HZ	$1-2$ HZ	$1-2$ HZ	$5-10$ HZ
Learning Method	unsupervised	supervised	supervised	unsupervised

***C3Vision*: Cortically-Couple Computer Vision;**
***HAC*: Human Aided Computing;**
***HAC-CV*: HAC with computer vision component (spatial pyramid matching tech)**
***BCI-VPM*: Brain computer interface and visual pattern Mining**