# COMBINING BRAIN COMPUTER INTERFACE WITH VISION FOR OBJECT CATEGORIZATION

SOUMYA SARA JOHN
SC17M048
Dept. of Avionics
Indian Institute of Space Science and Technology

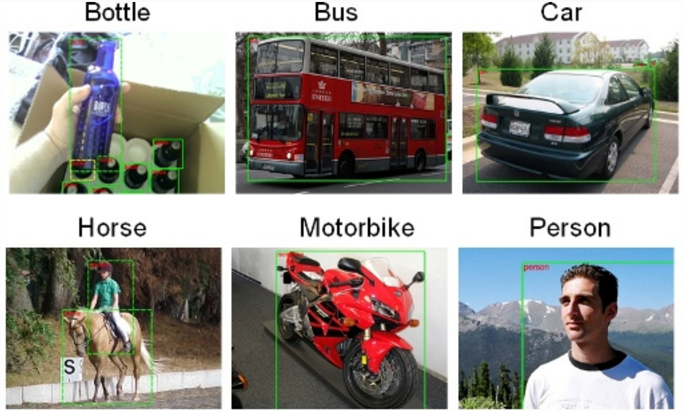
21 May 2018

# TABLE OF CONTENTS

# INTRODUCTION

# INTRODUCTION

- Detect objects in an image and detect the object categories
- Existing methods : Bag of Visual Words , Kernel methods , CNN-based ,etc.



# CHALLENGES

- Illumination condition , Occlusions , Clutter , Different camera view point

# OUR BRAIN

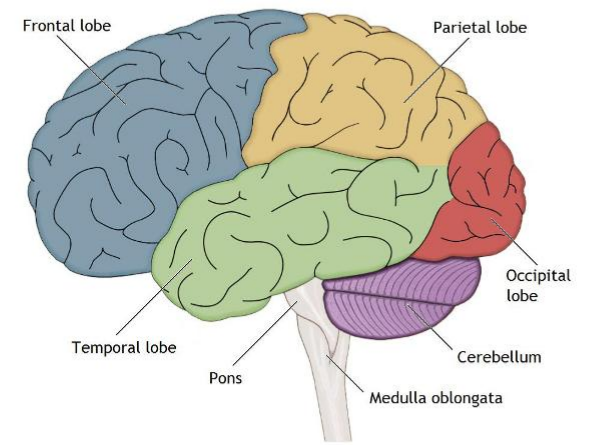



Figure 1: Brain lobes

Parts of the brain:

- Cerebrum:
  - Frontal Lobe : Emotions,judgement
  - Parietal Lobe : Pain,vision,speech
  - Occipital Lobe : Interprets vision
  - Temporal Lobe : Memory, sequencing,language
- Cerebellum: Coordinates muscle movements,maintain posture,balance
- Brainstem:
  - Pons : Relay center
  - Medulla Oblongata : breathing,body temperature

## OUR BRAIN-VISUAL PATHWAY

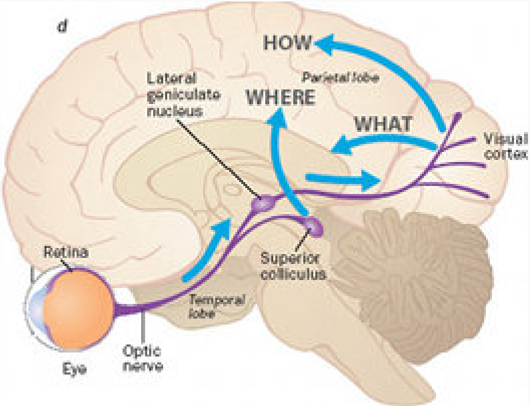



Figure 2: Visual Pathways in Brain

# ELECTROENCEPHALOGRAPHY(EEG)

# EEG - INTRODUCTION

- Invented by Hans Berger in 1924
- EEG data is a crude representation of brain activity
- Neurons - electrically charged by membrane transport proteins that pump ions across their membranes
- Action potential or nerve impulses
- Greatest advantage is its speed
- Applications:
    - Monitor alertness,brain death and coma
    - Locate areas of damage
    - Computer Vision

# EEG - INTRODUCTION

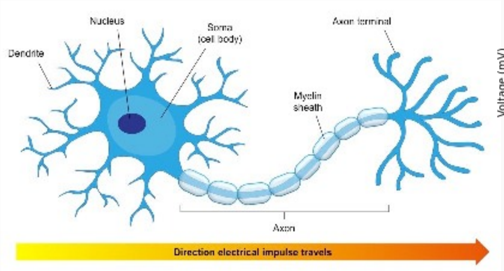



Figure 3: Structure of a Neuron

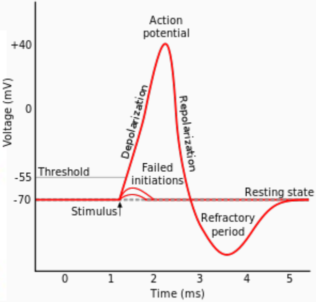



Figure 4: Action Potential

- Advantages:
    - non-invasive ,passive measuring device and hence safe to use
    - Does not require a highly skilled operator or medical procedure
- Two ways of analysing EEG data:
    - Spectral power of the signal
    - Inspecting the ERPs

## EEG ANALYSIS - SPECTRAL POWER

- Spectral power in different frequency bands corresponding to the neural activity
- Use FFT , Eigen Methods
- 5 types of brain waves:
    - Alpha waves (8-13 Hz) : seen in adults at rest with eyes closed
    - Beta waves (13-30 Hz) : seen in alert wake adults
    - Theta waves (4-8 Hz) : seen in children during sleep
    - Delta waves (<4 Hz) : seen in young babies during deep sleep
    - Gamma waves (>35 Hz) : reflects the mechanism of consciousness

- Also called as Event Related Potentials (ERPs)
- Significant voltage fluctuations resulting from evoked neural activity

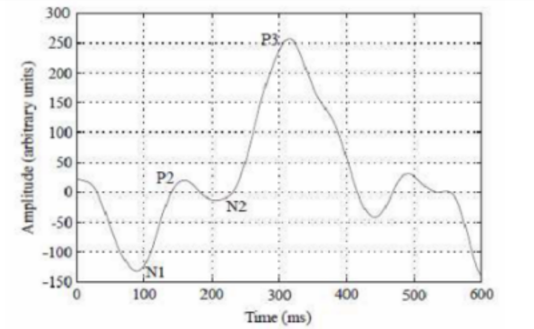



Figure 5: ERP waveform

- Use grand average ERPs
- In image categorization task, ERPs show two phases:
  - Early features ($100 < t(ms) < 220$)
  - Late features ($350 < t(ms) < 550$)
- Features of ERP response is highly sensitive to faces

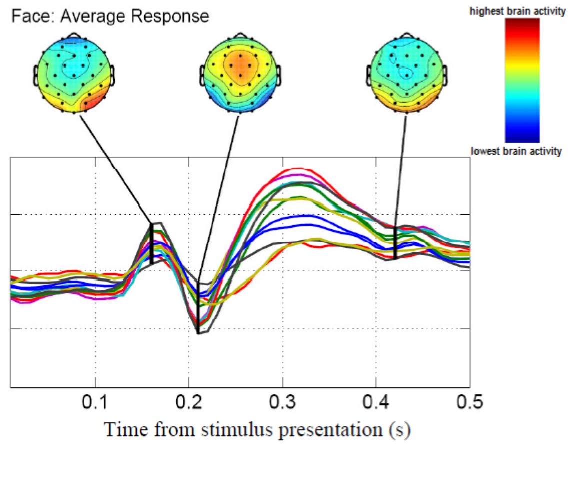



**Figure 6:** ERP waveforms when the subject sees a face [3]

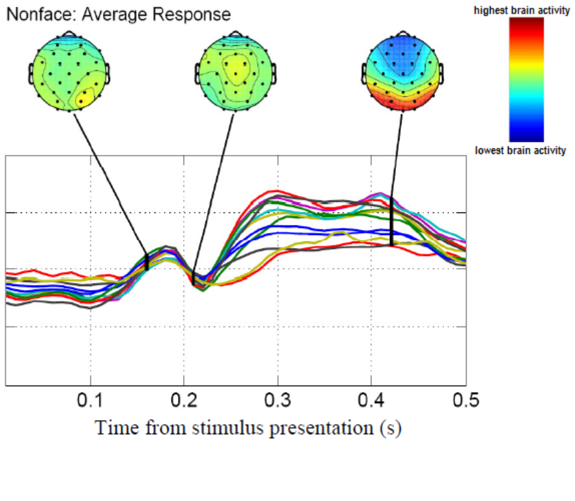



Figure 7: ERP waveforms when the subject sees an image with no face [3]

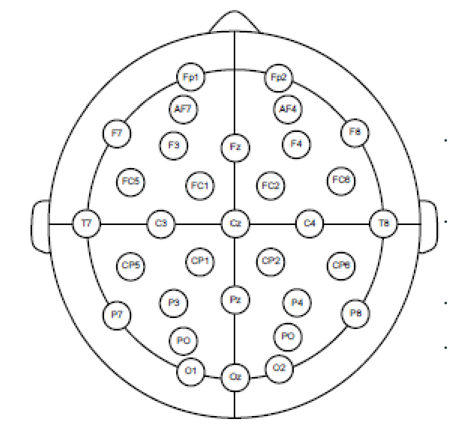



Figure 8: Internationally accepted 10-20 electrode placement

- Used 10-20 electrode placement
- Channels T7, T8, P3, PZ, P4, P7, P8, PO3, PO4, O1, Oz, O2
- Downsampled to 100 Hz
- Filtering

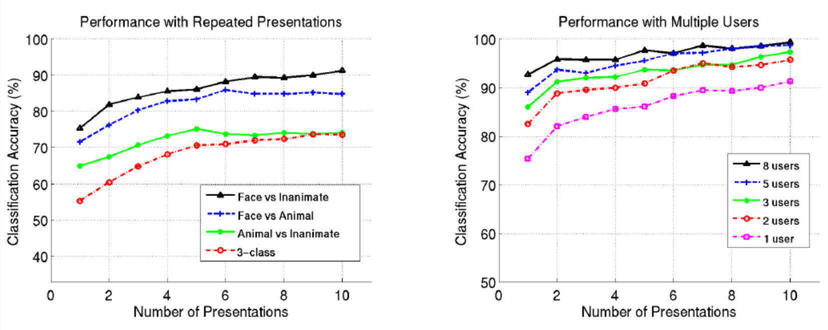



**Figure 9:** (a) Accuracy for various classes (b) Accuracy with multiple users for face vs inanimate object classification[1]

[1] P. Shenoy and D. Tan. ,"Human-aided computing: Utilizing implicit human processing to classify images" , In ACM - CHI 2008

# PYRAMID MATCH KERNEL

- Object Categorization steps , in general
    - Keypoints in image (Dense , Hessian , Harris corner detectors)
    - Features at these keypoints (HoG , SIFT)
    - Match the features (distance measure)
- Indexing features with visual vocabularies (Bag of Visual Words)

- Problems with existing techniques:
    - Designed to operate on fixed length vector inputs
    - Requires solving for explicit correspondences between features
    - Computationally costly
    - Impractical for large set sizes

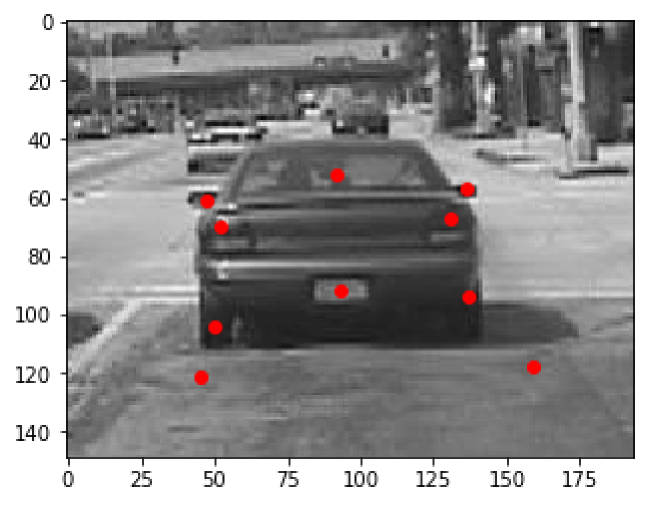



Figure 10: Car image and its keypoints

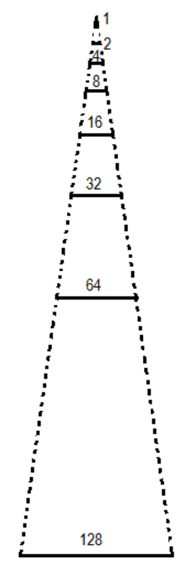



Figure 11: Histogram Pyramid

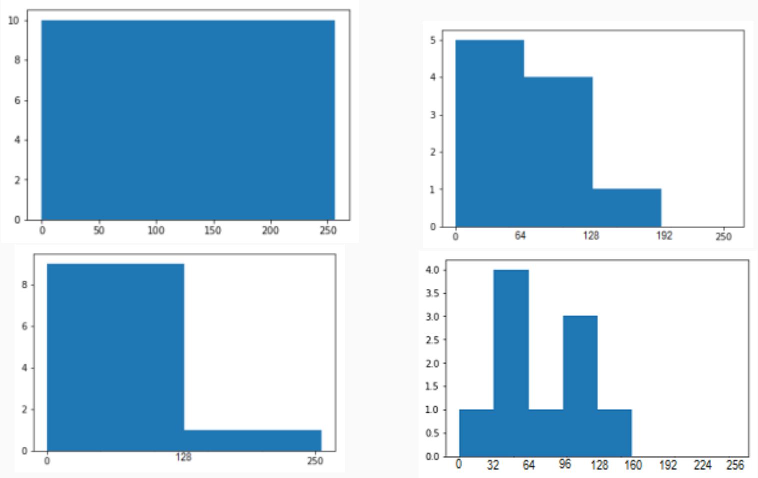



**Figure 12:** Histograms H(0),H(1),H(2),H(3) from pyramid

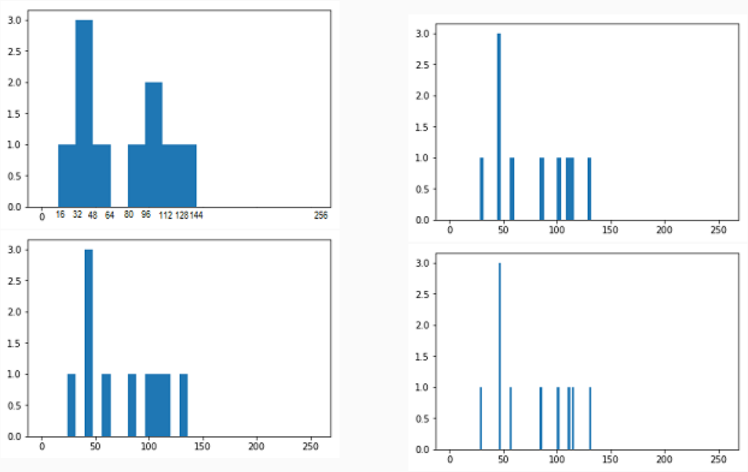



Figure 13: Histograms H(4), H(5), H(6), H(7) from pyramid

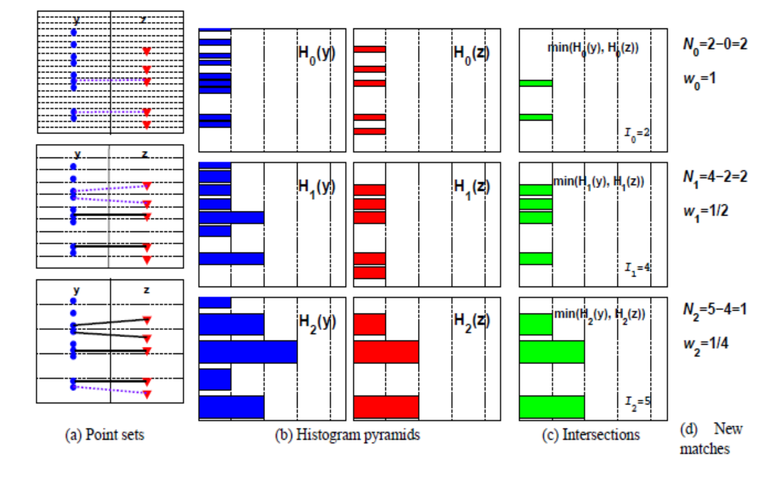



Figure 14: Pyramid match example

- By Kristen Grauman and Trevor Darrell in 2007
- Each object is a collection of some features
- Measures partial match similarity of features
- Maps unordered feature sets to multiresolution histograms
- The histogram pyramids are then compared using a weighted histogram intersection
- Robust to clutter

- Feature space F of d-dimensional vectors
- Point sets come from input space S :

$$S = \left\{ \mathbf{X} | \mathbf{X} = \{\mathbf{x_1}, \mathbf{x_2}, ...\mathbf{x_m}\} \right\} \tag{1}$$

where $\mathbf{x_i} \in F$ and $m = |\mathbf{X}|$

- Given point sets $\mathbf{X}$ and $\mathbf{Y}$,where $m = |\mathbf{X}|$ and $n = |\mathbf{Y}|$; $(m \leq n)$, a partial matching :

$$M\left(\mathbf{X}, \mathbf{Y}; \pi\right) = \left\{ \left(\mathbf{x_1}, \mathbf{y_{\pi_1}}\right), \left(\mathbf{x_2}, \mathbf{y_{\pi_2}}\right), ..., \left(\mathbf{x_m}, \mathbf{y_{\pi_m}}\right) \right\} \tag{2}$$

which pairs each point in $\mathbf{X}$ to some unique point in $\mathbf{Y}$ as per $\pi = [\pi_1, \pi_2, ..., \pi_m]$ and $1 \leq \pi_i \leq n$

- The cost of partial matching :

$$C\big(M(X, Y; \pi)\big) = \sum_{x_i \in X} ||x_i - y_{\pi_i}||_1 \tag{3}$$

- Optimal partial matching uses

$$\pi^* = \underset{\pi}{\operatorname{argmin}} C\big(M(X, Y; \pi)\big) \tag{4}$$

- Similarity of a partial matching

$$S\big(M(X, Y; \pi)\big) = \sum_{x_i \in X} \frac{1}{||x_i - y_{\pi_i}||_1 + 1} \tag{5}$$

- Uses a multidimensional multiresolution histogram pyramid
  - Finest resolution - bins are small
  - Grows in size at successive levels
- Two point sets are considered to be matched when two points from them begin to share a bin
- Each feature set is mapped to a multiresolution histogram
- Histogram pyramids are then compared using a weighted histogram intersection computation

- Feature extraction function $\psi$ for an input set $X$ is defined as:

$$\psi(X) = [H_0(X), H_1(X), ..., H_{L-1}(X)], \tag{6}$$

where $X \in S$, $L = \lceil \log_2 D \rceil + 1$, $H_i(X)$ is a histogram vector formed over points in $X$ using d-dimensional bins of side length $2^i$

- $H_i(X)$ has a dimension $r_i = (\frac{D}{2^i})^d$
- Pyramid match $P_\Delta$ between two point sets $Y$ and $Z$ is defined as :

$$P_\Delta(\psi(Y), \psi(Z)) = \sum_{i=0}^{L-1} w_i N_i \tag{7}$$

where $N_i$ signifies the number of newly matched pairs at level i and $w_i$ is a weight for matches formed at level i.

- To calculate $N_i$, use histogram intersection function I as follows:

$$I(\mathbf{A},\mathbf{B}) = \sum_{j=1}^{r} \min\left(\mathbf{A}^{(j)}, \mathbf{B}^{(j)}\right), \tag{8}$$

where $\mathbf{A}$ and $\mathbf{B}$ are histograms with r bins and $\mathbf{A}^{(j)}$ denotes the count of the $j^{th}$ bin of $\mathbf{A}$

$$N_i = I(H_i(\mathbf{Y}), H_i(\mathbf{Z})) - I(H_{i-1}(\mathbf{Y}), H_{i-1}(\mathbf{Z})) \tag{9}$$

where $H_i$ represents the $i^{th}$ component histogram generated by $\psi$

- Number of new matches found at each level in the pyramid is weighted according to the size of that histogram's bins
- Applying a geometric bound in terms of $L_1$ norm , at level i in pyramid : bin length = $d2^i$
- $w_i = \frac{1}{d2^i}$
- Pyramid match then is:

$$P_\Delta(\psi(Y), \psi(Z)) = \sum_{i=0}^{L-1} w_i \Big( I(H_i(Y), H_i(Z)) - I(H_{i-1}(Y), H_{i-1}(Z)) \Big) \tag{10}$$

- No pairwise distance computation between features - $O(m^2)$ savings
- Computational complexity is $O(dmL)$
- If multiple (T) pyramid matches are there $O(TdmL)$

# BCI BASED OBJECT CATEGORIZATION

- Use an EEG device to measure the brain activity when the subjects are shown images
- Assign a kernel for the EEG data based on similarity
- Pyramid Match Kernel for those images
- Combine the kernels via kernel alignment
- Classify using SVM

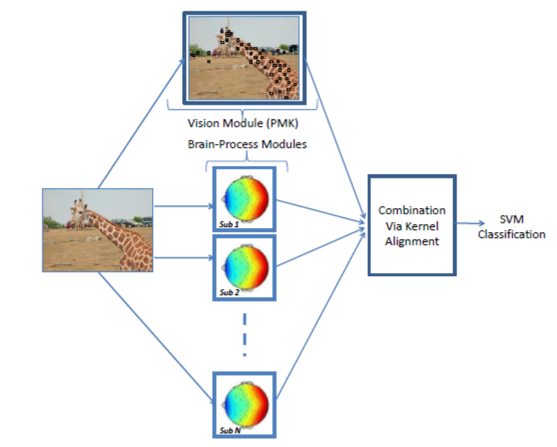



Figure 15: Framework to combine vision computations with human brain processing for visual category recognition[1]

- Feature fusion methods
    - Boosting
    - Bagging
- Decision level fusion methods [2]
    - Combining decisions from multiple modalities
    - eg: majority vote, sum, product, maximum, minimum

[2] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas. ”On combining classifiers ”Pattern Analysis and Machine Intelligence,1998

# BOOSTING ALGORITHM

- Given : $(x_1, y_1), \ldots\ldots(x_m, y_m)$ where $x_i \in X$ , $y_i \in Y = \{-1, +1\}$
  Initialize $D_1(i) = 1/m$
  For t=1,2,...T :
  - Train a weak learner using distribution $D_t$.
  - Get a weak hypothesis $h_t : X \to \{-1, +1\}$ with error,

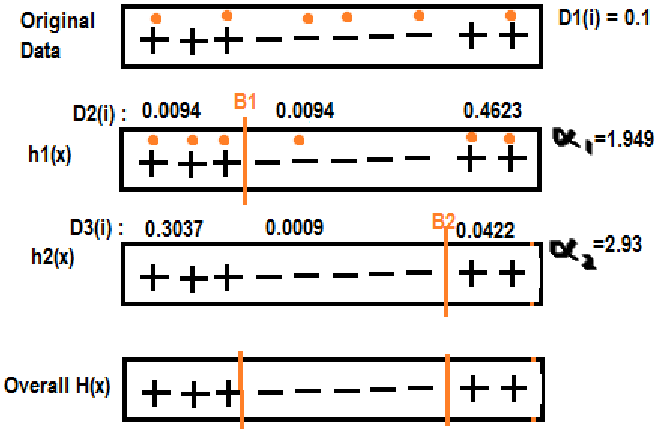
$$\epsilon_t = Pr_{D_i}[h_t(x_i) \neq y_i] = \sum_{i:h_t(x_i) \neq y_i} D_t(i)$$

  - Choose $\alpha_t = \frac{1}{2} \ln(\frac{1-\epsilon_t}{\epsilon_t})$
  - Update:

$$D_{t+1}(i) = \frac{D_t(i)\exp(-\alpha_t y_i h_t(x_i))}{Z_t}$$

    where $Z_t$ is a normalization factor
- Output of the hypothesis:

$$H(x) = \text{sign}\Big(\sum_{t=1}^{T} \alpha_t h_t(x)\Big)$$


Original Data
D1(i) = 0.1
D2(i) :
0.0094
B1
0.0094
0.4623
h1(x)
α1=1.949
D3(i) :
0.3037
0.0009
B2
0.0422
h2(x)
α2=2.93
Overall H(x)

- Given set L contains $\{(x_1, y_1), ......(x_m, y_m\})$ where $x_i \in X$ , $y_i \in Y = \{1, 2, ...j\}$
  Predictor $\psi(x, L)$
- Take repeated bootstrap samples $\{L^{(B)}\}$ from L and form $\{\psi(x, L^{(B)})\}$
- Vote to form $\psi_B(x)$

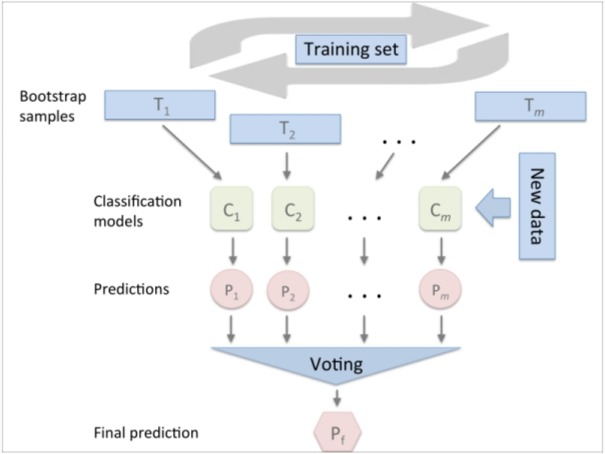



Figure 16: Bagging procedure

- Given a set of training images, corresponding EEG responses from k users , compute kernel $\mathsf{K}_{\epsilon i}$ for user i
- Compute Pyramid Match Kernel $\mathsf{K}_{\Delta}$
- Combined kernel $\mathsf{K}$ as

$$\mathsf{K} = \alpha_0 \mathsf{K}_{\Delta} + \sum_{i=1}^{k} \alpha_i \mathsf{K}_{\epsilon i} \tag{11}$$

  $\alpha = \{\alpha_0, ...\alpha_k\}$ are the parameters to be optimized
- Define an ideal kernel $\mathsf{A}$ such that $A_{ij} = 1$ iff, $i^{th}$ and $j^{th}$ image have the same visual category label else $A_{ij} = 0$

- Minimize objective function $L(\alpha)$ defined as

$$\underset{\boldsymbol{\alpha}}{\text{argmin}}||\mathsf{K} - \mathsf{A}||_F^2$$

subject to : $\alpha_i \geq 0$ for $i \in \{0, ...k\}$

- Solved using gradient descent procedure based BFGS method
- Gradients :

$$\frac{\delta L(\alpha)}{\delta \alpha} = 2.\text{sum}(\mathsf{K}_{\epsilon i} \circ (\mathsf{K} - \mathsf{A}))$$

where sum(·) means summation over all the elements of the matrix and $'\circ'$ denotes the Hadamard product

# EXPERIMENTAL DETAILS[1]

- Images from Caltech - 256 dataset
- Classes : Inanimate, face, animals
- Computed Pyramid Match Kernel using SIFT features
- EEG data measurement
    - 32 channels (10-20) ; used only 12 of them
    - Captured at 2kHz , downsampled to 100 Hz
    - Filtered using Butterworth filter in range 0.5 - 30 Hz
    - Restricted data to 100 - 500 ms
    - Concatenated them to a single vector
    - Gaussian kernel :

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\beta ||\mathbf{x}_i - \mathbf{x}_j||^2)$$

where $\beta = 10^{-5}$

| Method | Accuracy |
|---|---|
| PMK | 81.67% |
| EEG | 68.33% |
| PMK + EEG | 86.67% |
| PMK + EEG (Voting) | 91.67% |

**Table 1:** Comparison of classification accuracies

| | Recognized Class | | |
|---|---|---|---|
| True Class | Animals | Faces | Inanimate |
| Animals | 19 | 0 | 1 |
| Faces | 1 | 19 | 0 |
| Inanimate | 1 | 2 | 17 |

**Table 2:** Confusion Matrix for EEG + PMK

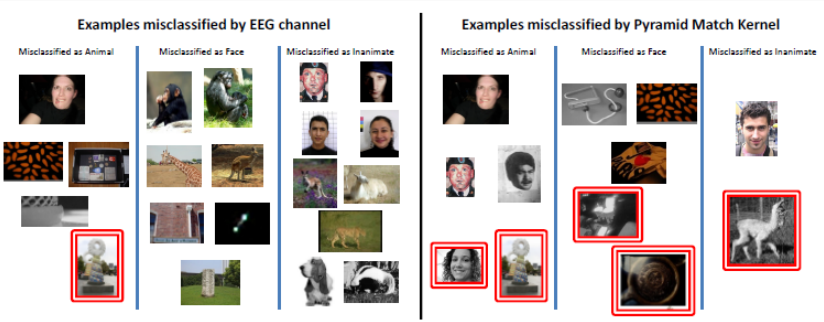



**Figure 17:** Examples misclassified by EEG and PMK but corrected in EEG + PMK ; the red doubled ones were misclassified even in EEG + PMK[1]

# CONCLUSION

- Testing using EEG
- Object Localization
- No independent assumption on PMK

# OTHER BCI RELATED WORKS

- BCI and Deep Learning
  - Motor Imagery[3]
  - On board BCI using CNN
- Medical Applications:
  - Bionic Eye
  - Wearable BCIs to assist communication in ICU’s

[3] S.Sakhavi, C. Guan and S. Yan ”Learning Temporal Information for Brain-Computer Interface Using Convolutional Neural Networks” , IEEE Trans. Neural Netw. Learn. Syst.

- Such a combination between vision and human brain processing can yield significant gains in accuracy for the task of object categorization
- Advantages:
    - Robust to clutter and partial occlusion
    - Lower number of parameters compared to CNN
    - Efficient method
- Disadvantages:
    - Difficult to choose the channels from 32 channels
    - EEG signals are subject dependent

# REFERENCES


1. A.Kapoor , P.Shenoy and D.Tan.,"Combining Brain Computer Interfaces with Vision for Object Categorization",In CVPR 2008
2. K. Grauman and T. Darrell., "The pyramid match kernel: Discriminative classification with sets of image features",In ICCV 2005
3. P. Shenoy and D. Tan. ,"Human-aided computing: Utilizing implicit human processing to classify images" , In ACM - CHI 2008
4. R. Schapire. "A brief introduction to boosting." In Proceedings of International Conference on Algorithmic Learning Theory, 1999.
5. L. Breiman." Bagging predictors", Machine Learning, 1996.

# Questions?

# Thank you