

MRI brain classification in the dissimilarity space

10th VIPS Advanced school on
Computer Vision and Pattern
Recognition

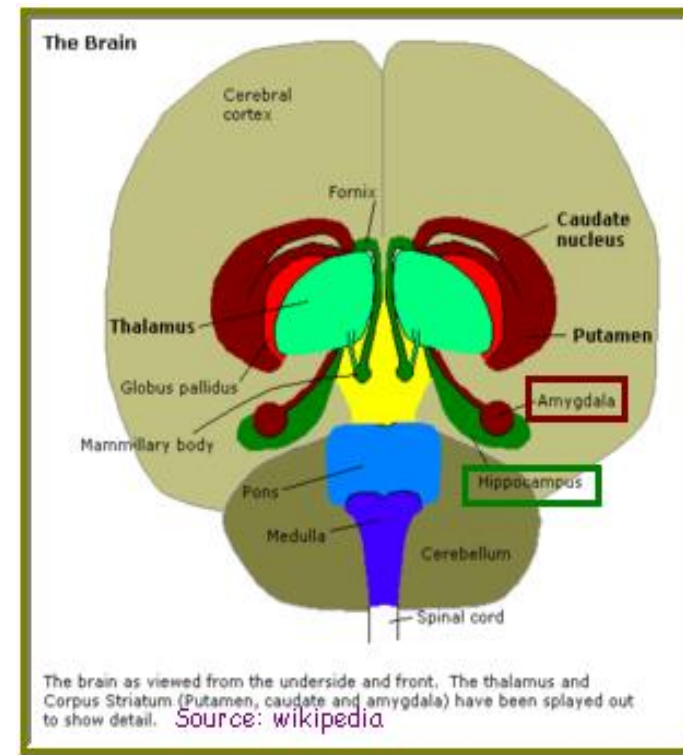
Verona 23-26 September 2013

The problem

- Characterizing healthy / mentally ill subjects based on the observation of the brain structure of the brain
- 2-class classification of schizophrenic (*patients*) and healthy (*controls*) people
- Challenging problem
 - Currently not diagnosed from MRI images
- Encouraged by medical studies

Regions of Interest (ROIs)

- The analysis is performed using a ROI-based approach
- Use a subpart of the brain instead of the whole brain to classify between patients and controls
- Abnormalities in ROIs are known to affect cognitive processes

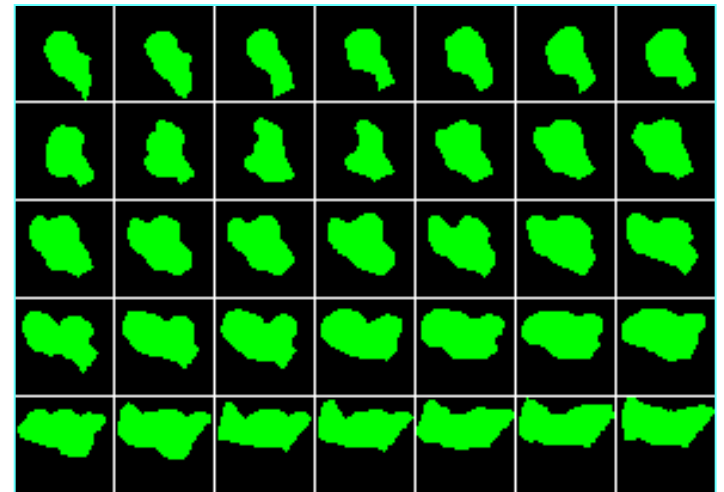
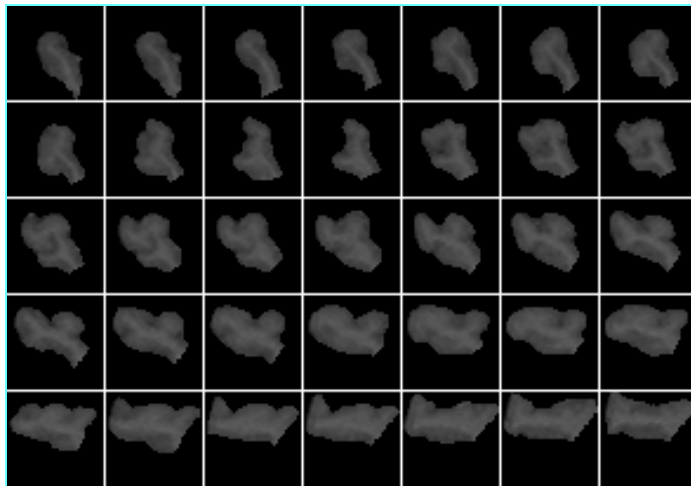


Analyzed Region of Interest

- 14 different Regions of interest
 - Left/right amygdala
 - Left/right dorsolateral prefrontal cortex (dlpfc)
 - Left/right entorinal cortex (ec)
 - Left/right Heschl's gyrus (hg)
 - Left/right hippocampus
 - Left/right superior temporal gyrus (stg)
 - Left/right thalamus

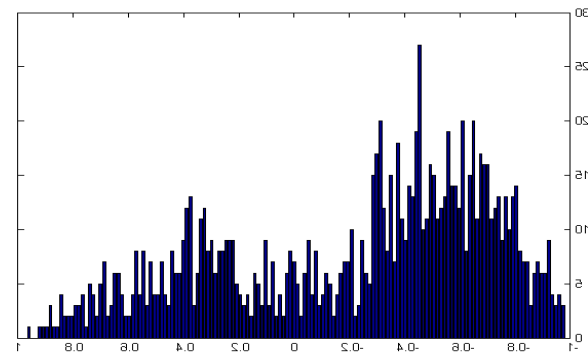
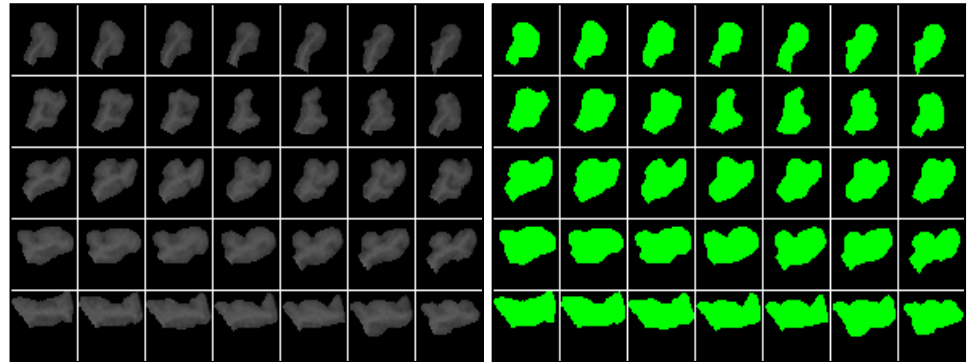
sMRI ROI dataset: details

- 124 sets of 14 ROIs (64 patients and 60 controls):
 - we have a set of slices for each subject
 - different number of images for different ROIs and different subjects

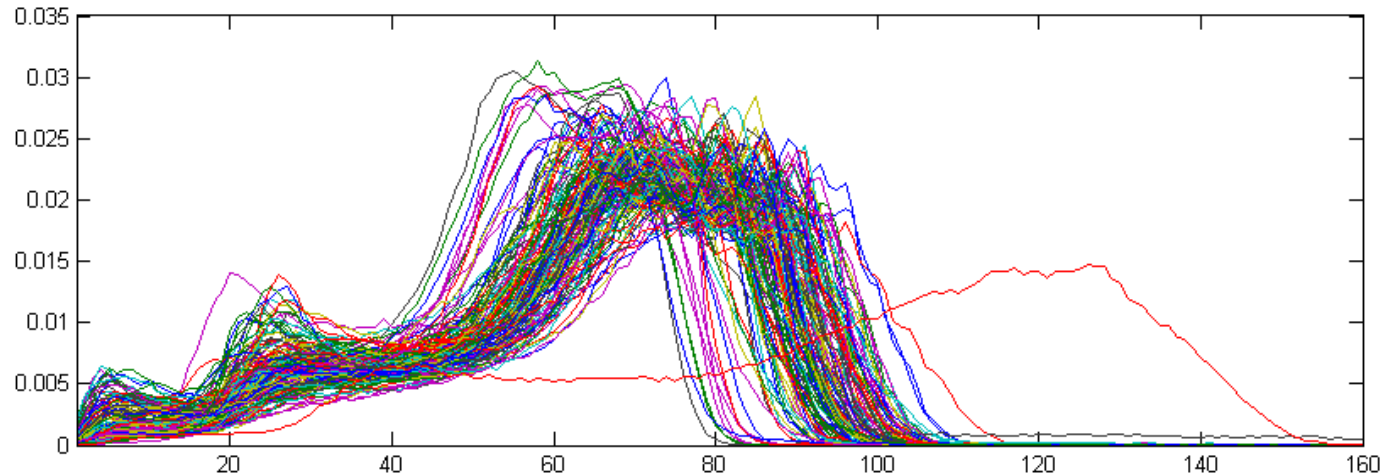


sMRI ROI dataset: details

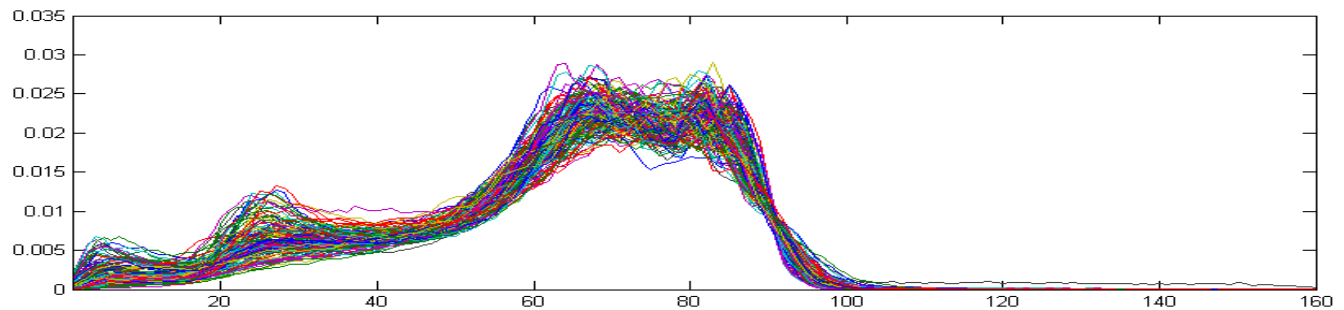
- Histograms of intensities (sMRI)
- Scaling based on matching landmarks on ICV (Intra Cranical Volume) histograms
- Histograms are also transformed to pdf



Histogram Rescaling



Original ICV histograms



Rescaled ICV histograms

(Dis-)Similarity measures

(Dis-)Similarity for histograms

- L2 (Euclid) distance
- L1 distance
- Histogram dissimilarities
 - χ^2 distance
 - Diffusion distance
 - EMD (Earth Mover Distance)
 - Histogram intersection

(Dis-)Similarity for pdf

- L1 and L2 distance
- Bhattacharyya
- EMD (Earth Mover Distance)
- K-L divergence based
 - Original K-L divergence (Not symmetric)
 - Average K-L divergence
 - Jensen-Shannon divergence

Notation

- Two histograms, S and M
- n bins S_i and M_i
- Corresponding pdfs p_i and q_i
- Total number of elements in histograms $|S|$ and $|M|$
- D is the distance between S and M .

χ^2 distance (histograms)

- This metric is based on the χ^2 test for testing the similarity between two distributions

$$D = \sum_{i=1}^n \frac{(S_i - M_i)^2}{S_i + M_i}$$

Diffusion distance (histograms)

- The distance between two histograms is defined as a temperature field. It is derived as the sum of dissimilarities over scales (Ling and Okada, 2006)
- Use the distance between the histograms as an initial value of a temperature field at $t = 0$

Diffusion distance

- Using the heat diffusion equation where r is an upper bound and k is a norm (L1):

$$\frac{\partial T}{\partial t} = \frac{\partial^2 T}{\partial x^2}$$

$$T(x, t) = T(x, 0) \times \phi(x, t)$$

$$\phi(x, t) = \frac{1}{(2\phi)^{1/2}t} \exp -\frac{x^2}{2t^2}$$

$$D = \int_0^r k(|T(x, t)|) dt$$

Histogram intersection (histograms)

- Is the number of intersecting values in each cell, normalized by the total number of elements in the histograms

$$Sim(S, M) = \frac{\sum_{i=1}^n \min(S_i, M_i)}{\min(|S|, |M|)} .$$

$$D = (1 - Sim(S, M))$$

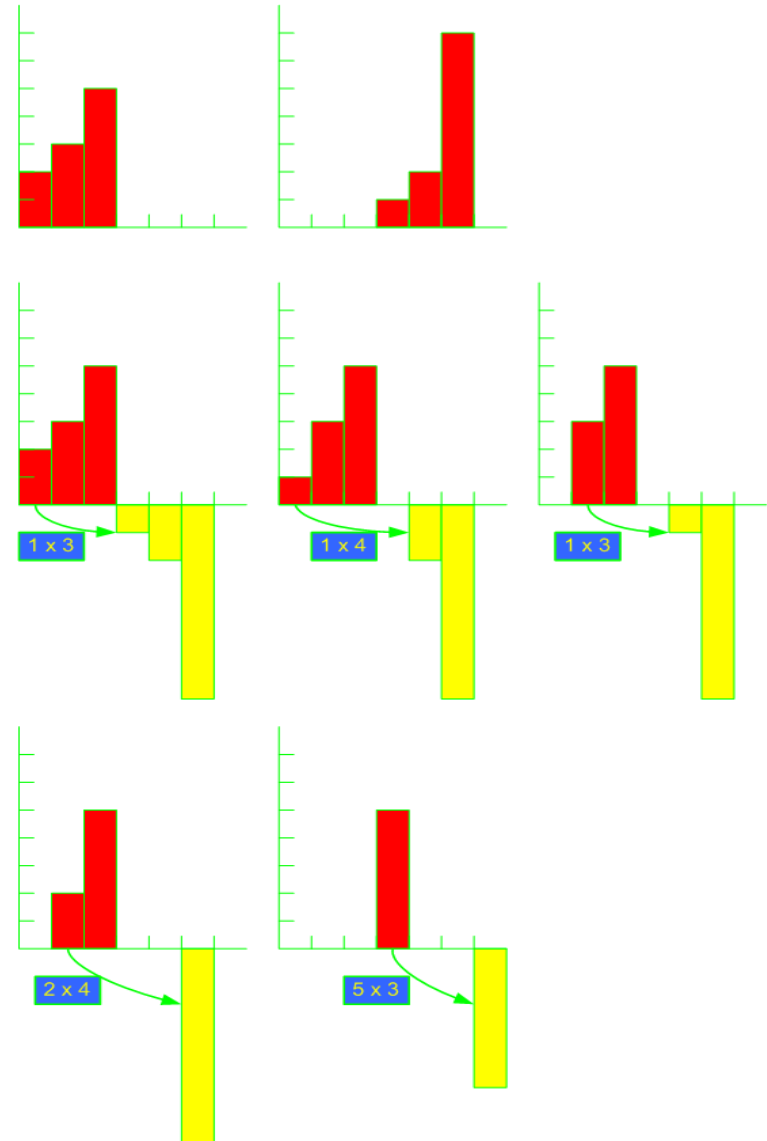
EMD (histograms and pdfs)

- EMD (Rubner et al., 2000) is defined as the cost to transform one distribution into another
- Linear optimization by defining it as a transportation problem. In 1D:

$$C_i = \left| \sum_{j=1}^i (S_j - M_j) \right|, D = \sum_{i=1}^n C_i.$$

EMD

- Treat one distribution as hills, the other as holes and find the minimum earth to be moved to fill the holes
- Formulated as a supply demand problem



Bhattacharyya (pdfs)

- Is used to measure the similarity of two discrete probability distributions

$$D = -\log BC(p, q)$$

$$BC(p, q) = \sum_{x \in X} \sqrt{p(x)q(x)}$$

Kullback–Leibler divergence (pdfs)

$$D(p, q) = \sum_{i=1}^n q_i \log \frac{q_i}{p_i}$$

$$D = D(p, q) + D(q, p)$$

$$D = \frac{1}{2}D(p, r) + \frac{1}{2}D(q, r)$$

$$r = \frac{p + q}{2}$$

Summary

- 14 different Regions of interest
 - (1, 2) Left/right amygdala
 - (3,4) Left/right dorsolateral prefrontal cortex (dlpfc)
 - (5,6) Left/right entorinal cortex (ec)
 - (7,8) Left/right Heschl's gyrus (hg)
 - (9,10) Left/right hippocampus
 - (11,12) Left/right superior temporal gyrus (stg)
 - (13,14) Left/right thalamus

Summary

13 different dissimilarities

- *his-euclid*: Euclidean distance between histograms.
- *his-l1*: L1 distance between histograms.
- *his-intersect*: Intersection between histograms.
- *his-diffusion*: Diffusion distance between histograms.
- *his-chi*: χ^2 distance between histograms.
- *his-emd*: Earth mover's distance between histograms.
- *pdf-euclid*: Euclidean distance between pdfs.
- *pdf-l1*: L1 distance between pdfs.
- *pdf-emd*: Earth mover's distance between pdfs.
- *pdf-bs*: Bhattacharyya distance between pdfs.
- *pdf-kl*: Symmetrized KL divergence between pdfs.
- *pdf-kl_orig*: Original, asymmetric KL divergence.
- *pdf-js*: Jensen-Shannon divergence between pdfs.

Summary

- In summary we have defined:
 - 14 ROIs
 - 13 (Dis-)Similarity measures
 - 6 for histogram and 7 for pdf
- In total we have $14 \times 13 = 182$ dissimilarity measures