A Fragment-Based Memory Model of Saccadic Vision

Defragmenting the Face

Luke Barrington, ECE Dept. UC San Diego
lbarrington@ucsd.edu
Garrison Cottrell, CSE Dept. UC San Diego
gary@ucsd.edu

Motivation

Visual perception is based on non-uniform samples (saccades) that are discrete in space and time.

Scan paths are not repeated so memory can not work by template matching.

A kernel density estimate built from observed visual fragments can be used to probabilistically assess how new samples drawn from the current environment could have been generated by previous observations.

We present a reformulation of Lacroix et. al's (2006) saccade-based Natural Input Memory model as a kernel density estimator that uses a different salience map and feature vectors and apply it to a number of facial memory tasks.

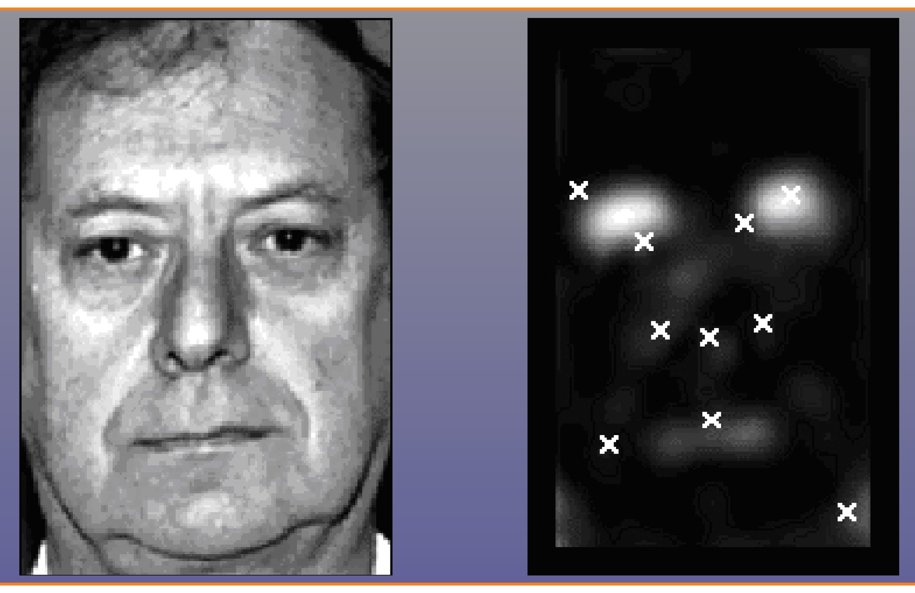
Saccade Selection

Salience levels are computed from the rotational variance of 8 low-resolution Gabor filters (Yamada & Cottrell, 1995):

$$Salience\,(i,j) = \frac{\sum_{n=1}^{8}\left(G\left(i,j,\frac{\pi n}{8}\right)-\mu_G\right)^2}{8}$$

This variance is large at edges and corners.

Fixation points are chosen randomly according to this salience distribution and sampled locations are suppressed.



Feature Transform

Gabor filter responses at 8 orientations and 4 frequencies form our biologically-motivated, V1 processing model.

At each fixation point, we extract a square patch of these filter responses that corresponds to a foveated representation of the image fragment

To improve generalization and increase capacity, the dimensionality of the representation is reduced to 80 components using PCA.

Natural Input Memory

We apply our saccade selection and feature extraction methods to the NIM model.

"Memories" are feature vectors stored in a high-dimensional space; conceived as patterns of neural activation.

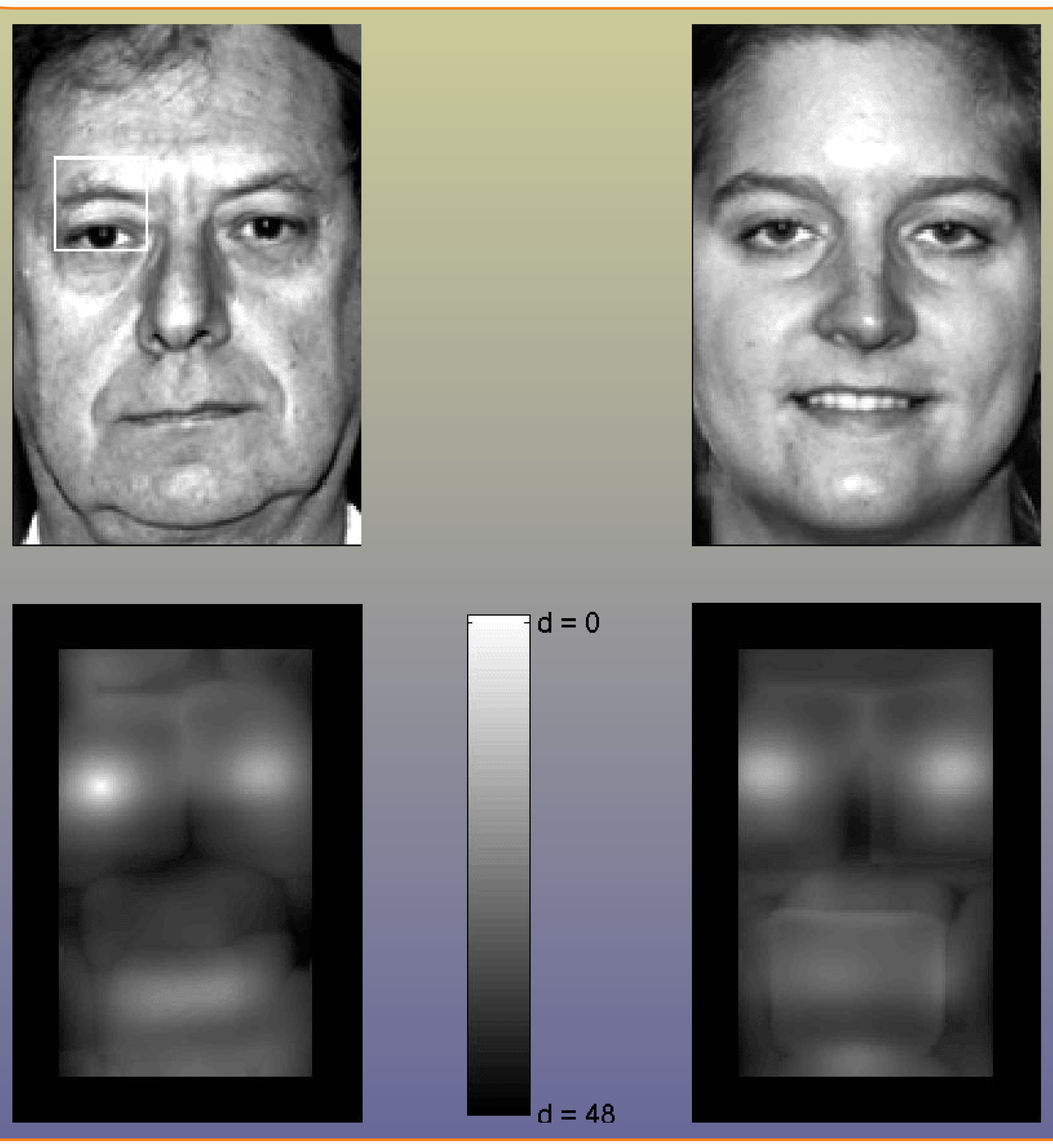
Retrieval takes place by comparing a new fragment to all the stored instances in memory - proximity in the memory space corresponds to perceptual similarity.

Familiarity is judged using a form of *kernel density estimation.*

For each of the S samples from a target image, count the number of stored memories, n_f that lie within a volume of radius r from the input fragment.

The *familiarity,* F_i of an image i is defined as:

$$F_i = \frac{1}{S}\sum_{f=1}^{S} n_{if} \qquad \forall \quad fragments, f \in image, i$$



Detectability Index, $d' = \dfrac{\mu_{F_T}-\mu_{F_L}}{\sqrt{\dfrac{\sigma_{F_T}^2+\sigma_{F_L}^2}{2}}}$

Acknowledgments

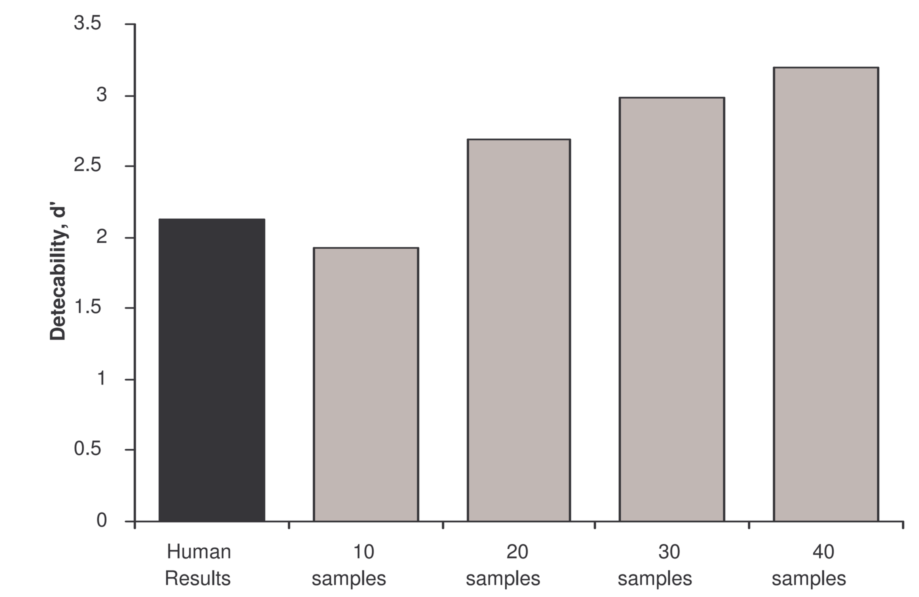
This research project was supported by NIMH grant MH57075 to GWC

References

Yamada, K. & Cottrell, G.W. (1995). A model of scan paths applied to face recognition. *Proceedings of Cog Sci Conference 1995.* 55-60.
Lacroix, J.P.W., Murre, J.M.J., Postma, E.O. & van den Herik, H.J. (2006). Modeling recognition memory using the similarity structure of natural input. *Cognitive Science, 30*, 121-145
Lewis, M.B. & Johnston, R.A. (1997). Familiarity, target set and false positives in face recognition. *European Journal of Cognitive Psychology*, 9, 437-459.

Facial Recall

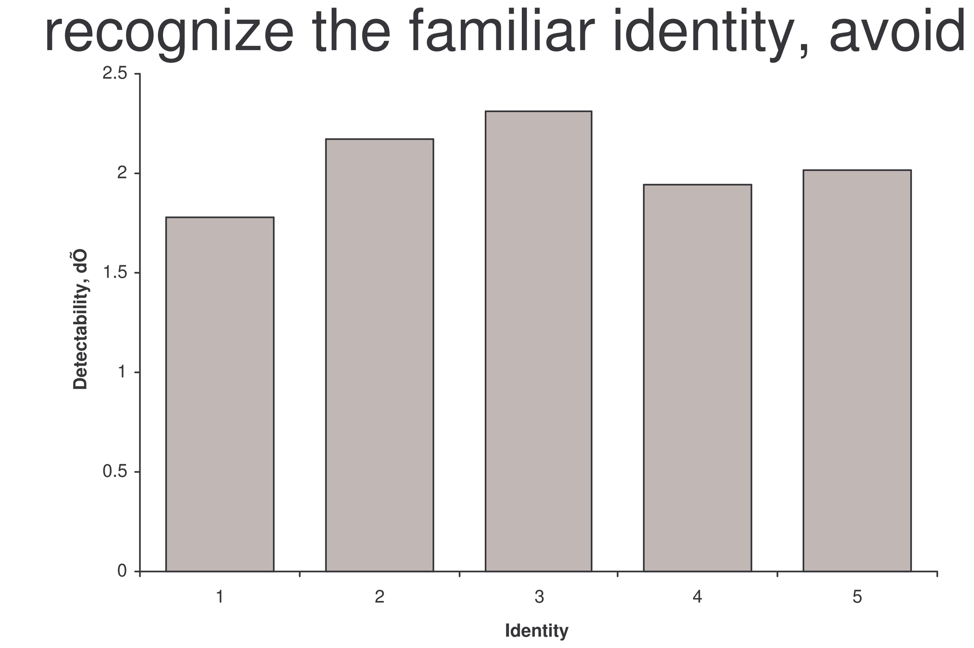
Training: study N(=20) briefly presented face images.
Testing: study 2N faces: N familiar targets, N lures.
Task: recognize the familiar faces, avoid the lures.



These results demonstrate that our changes to the NIM have not harmed the facial recall performance.

Identity Recognition

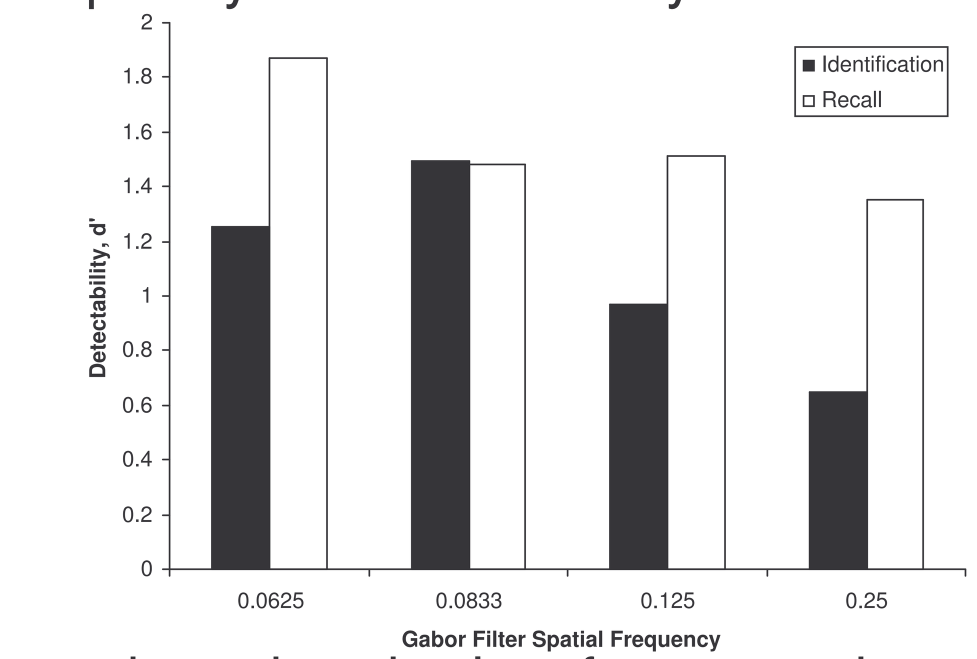
Training: study 6 different images of the same person
Testing: study 39 faces: 3 novel target images, 6 lure images from 4 different identities.
Task: recognize the familiar identity, avoid the lures.



The NIM model works for other visual memory tasks and can generalize to unseen images.

Multiple Resolutions

It is well known that different visual memory tasks use spatial frequency bands differently.



Here we show that the low frequency is most useful for recall but medium-low is best for identification while high frequency is actually detrimental.