A Vision-based Affective Computing System

Jieyu Zhao

Ningbo University, China

Outline

- Affective Computing
- A Dynamic 3D Morphable Model
- Facial Expression Recognition
- Probabilistic Graphical Models
- Some related topics
- Conclusions and future work

What is Affective Computing?

- affective – producing emotional response,

Affective Computing – ability for the computer to recognize and express emotions as humans do

a. Recognize emotions

b. Express emotions

c. ‘Have’ emotions

(Rosalind Picard, MIT, 1997)

Recognize Emotions

- Facial expression
- Polygraph, Multimodal

  skin response, heartbeat, blood pressure…
- Which emotion:

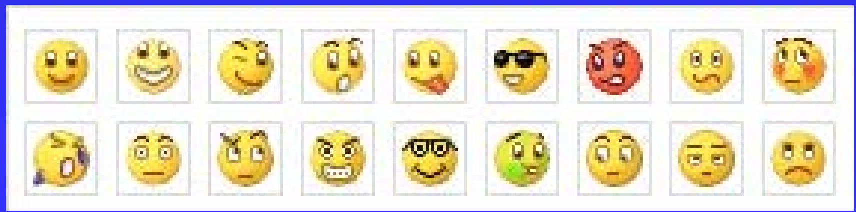
  happiness, angry, fear, surprise, sadness…
- Person dependent
- Person independent

Express Emotions

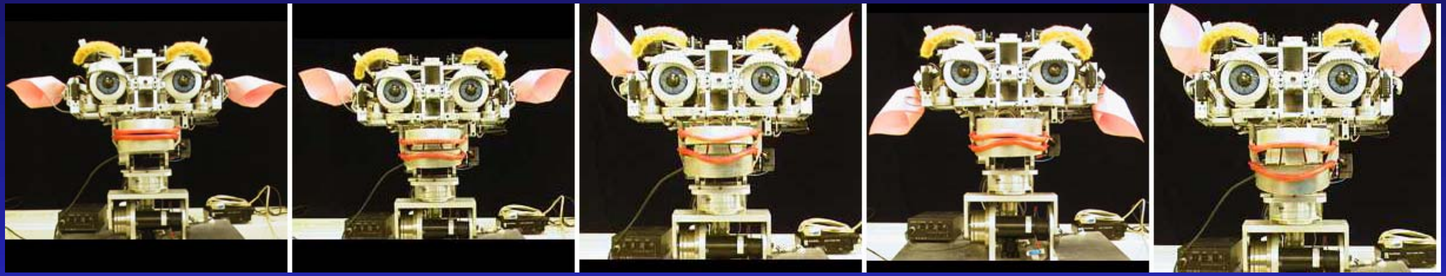
- **Emotional expression for communication and social co-ordination**
- **Emotion for organisation of behaviour (action selection, attention and learning)**

- **Emotion conveys information,**

  "Hello!" ☺



- Kismet: www.ai.mit.edu/projects/humanoid-robotics-group/kismet/kismet.html



- **Computer Graphics, 3D face model (FaceGen)**

Having Emotions

Emotions are Physical and Cognitive

- Emergent Emotions and Emotional Behavior
- Fast Primary Emotions
- Cognitively Generated Emotions
- Emotional Experience
- Body-Mind Interactions

- Emotional Intelligence?
- Can machines feel?
- How would we know?

Why Affective Computing?

- Humans naturally communicate affectively, expression identified 50% of the time
- Human-Computer Interaction – Frustration, mouse clicking behaviour, slow, debugging…

  We need more friendly HCI
- Applications: Hands-free computing, Social interfaces, Virtual sales agent, Internet banking, Distance education ……

Why vision based interface?

Visual cues are important in communication!

Useful visual cues

– Presence

– Identity (and age, sex, nationality, etc.)

– Facial expression

– Attention (gaze direction)

– Lip movement

– Gestures, Body language

– Location, Activity

Elements

- Hand tracking, Hand gestures
- Arm gestures
- Body tracking
- Activity analysis

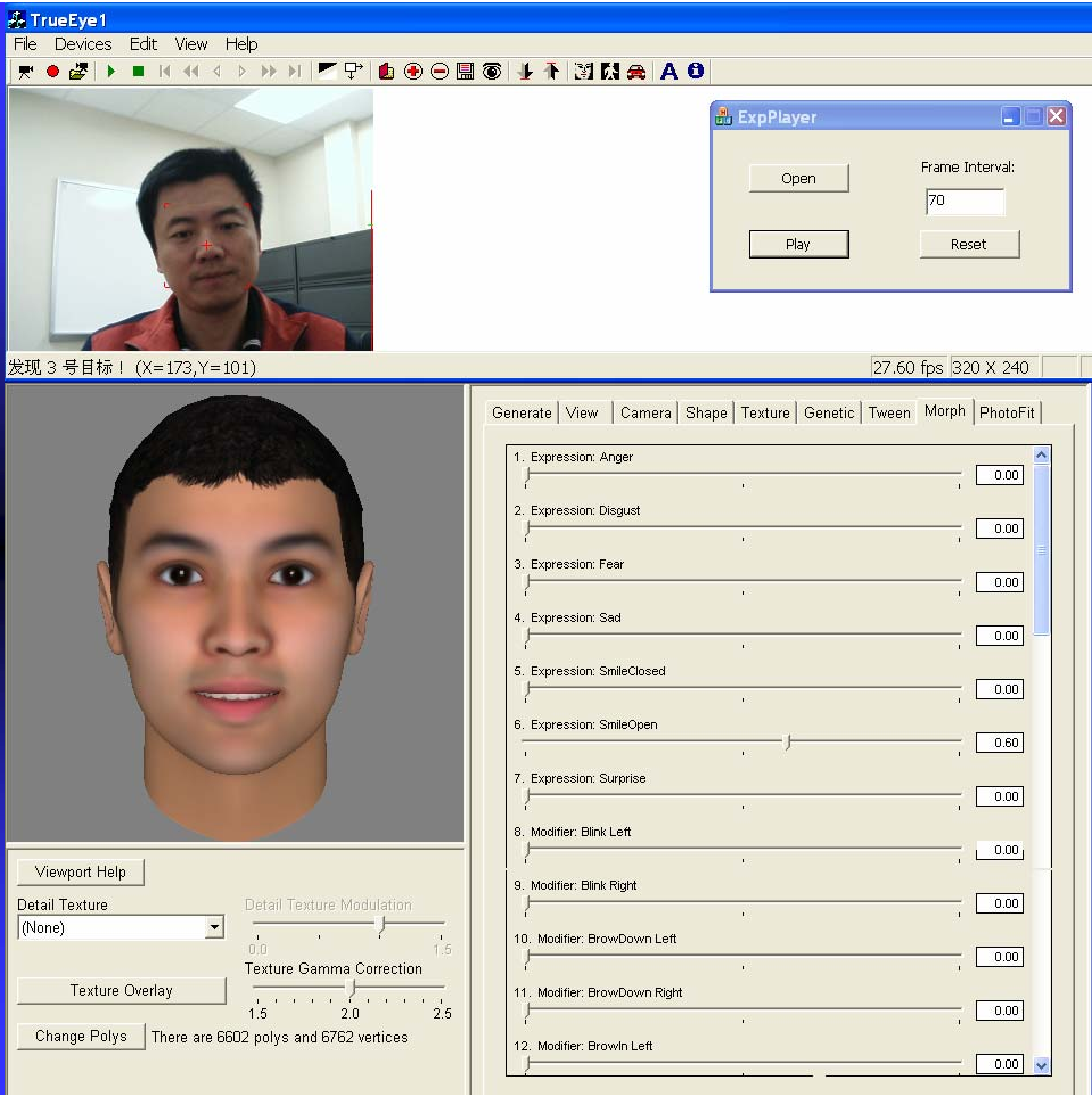
We focus on:

- Head tracking
- Face recognition
- Facial expression
- Lip movement
- Gaze

A System Developed

- Facial expression player based on FaceGen
- Moving object tracking system, gaze
- Running in real-time, interactive
- Face recognition and expression recognition
- Foreground/background discrimination

Demo


TrueEye1
File Devices Edit View Help
ExpPlayer
Open
Frame Interval:
70
Play
Reset
发现 3 号目标！ (X=173,Y=101)
27.60 fps
320 X 240
Generate
View
Camera
Shape
Texture
Genetic
Tween
Morph
PhotoFit
1. Expression: Anger
0.00
2. Expression: Disgust
0.00
3. Expression: Fear
0.00
4. Expression: Sad
0.00
5. Expression: SmileClosed
0.00
6. Expression: SmileOpen
0.60
7. Expression: Surprise
0.00
8. Modifier: Blink Left
0.00
9. Modifier: Blink Right
0.00
10. Modifier: BrowDown Left
0.00
11. Modifier: BrowDown Right
0.00
12. Modifier: BrowIn Left
0.00
Viewport Help
Detail Texture
(None)
Detail Texture Modulation
0.0
1.5
Texture Gamma Correction
Texture Overlay
1.5
2.0
2.5
Change Polys
There are 6602 polys and 6762 vertices

How we did it

- Programming in VC++.net and Direct X SDK
- A facial expression player,
  designed to play back facial expression files
- Moving object tracking in real time
  directShow, live video capture,
  moving object recognition, image pyramids
- Eye blink and movement control


TOPSPEED N6 - TrueEye1
File
Devices
Edit
View
Help
Ready
7.66 fps
640 X 480

Facial Expression Recognition

Challenges:

- Large variability

  rotation, scaling, illumination change,…
- Complex nonlinear manifold

  distance measure
- High dimensionality

  80x100 image, but relatively small sample size

Possible solutions:

- **Geometric feature based approaches**
  - **2D & 3D face model**
- **Statistical approaches**
  - **PCA (Principal Component Analysis),**
  - **ICA (Independent Component Analysis),**
  - **LDA (Linear Discriminant Analysis)**
  - **Kernel methods**
  - **Bayesian methods**
  - **Probabilistic Graphical models**

Probabilistic Graphical Model

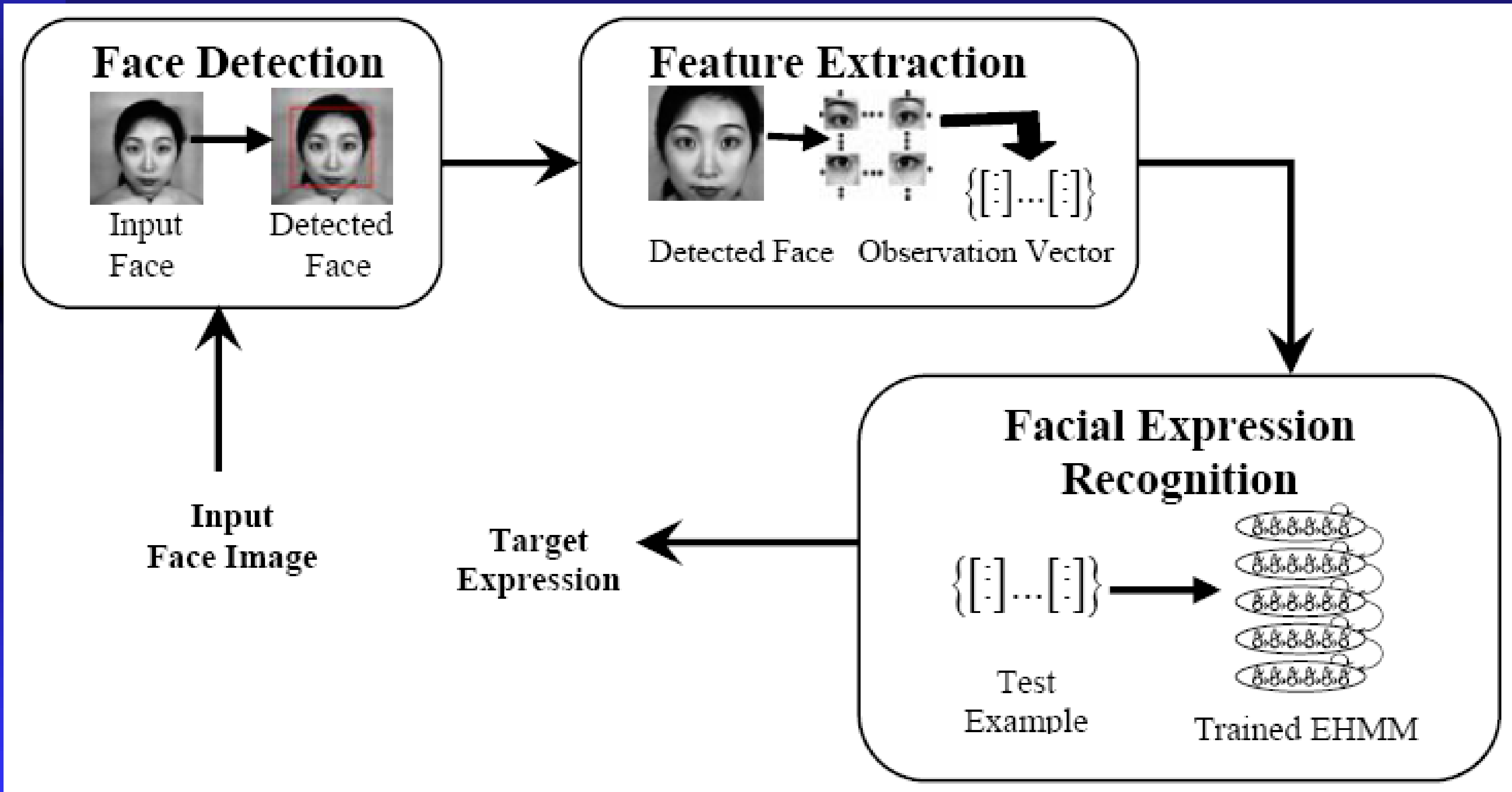
Probability Theory + Graph Theory

a natural tool for image representation, learning and inference

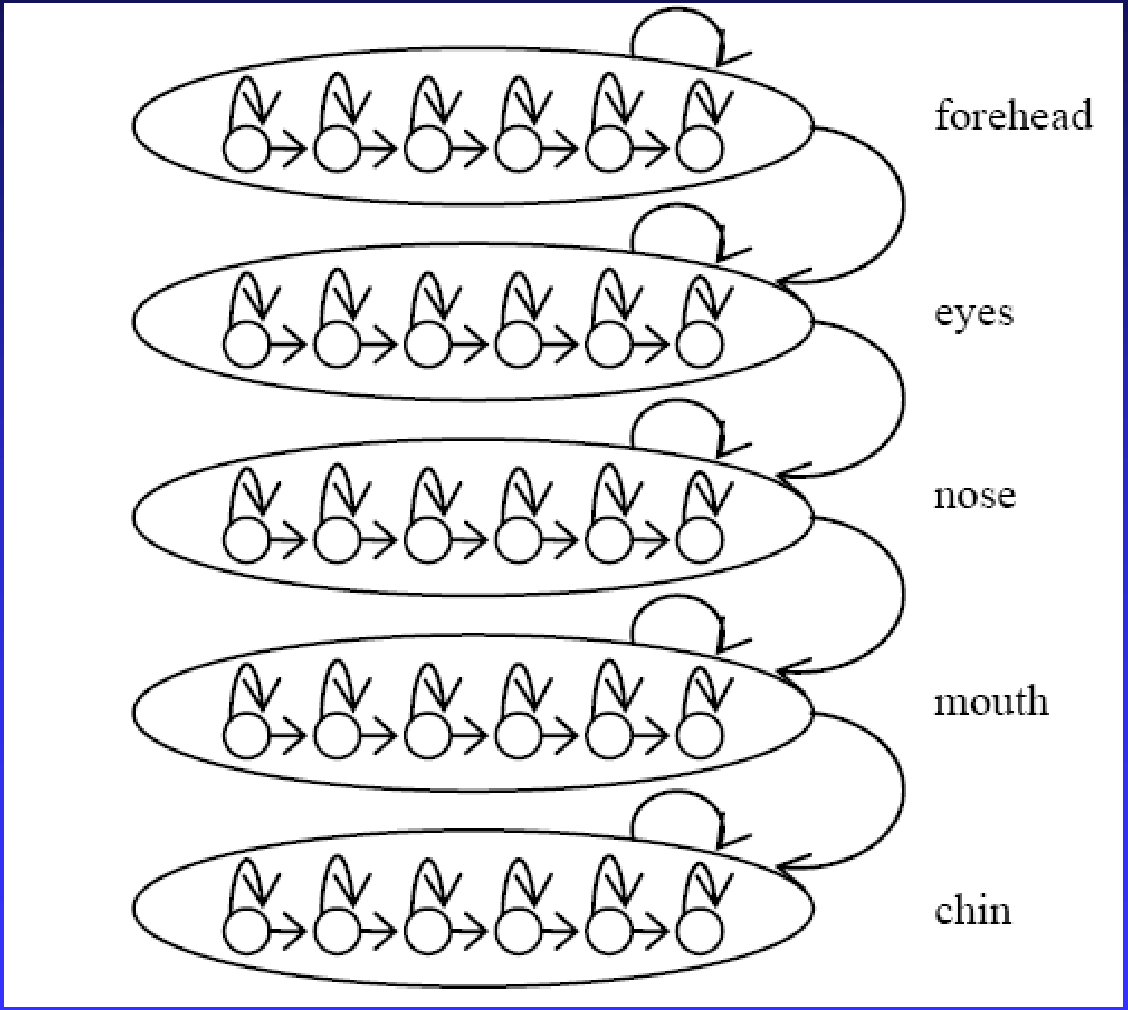
Various models:

- HMM
- MRF and GRF
- Bayesian Network
- Kalman Filter, ICA, Factor Analysis

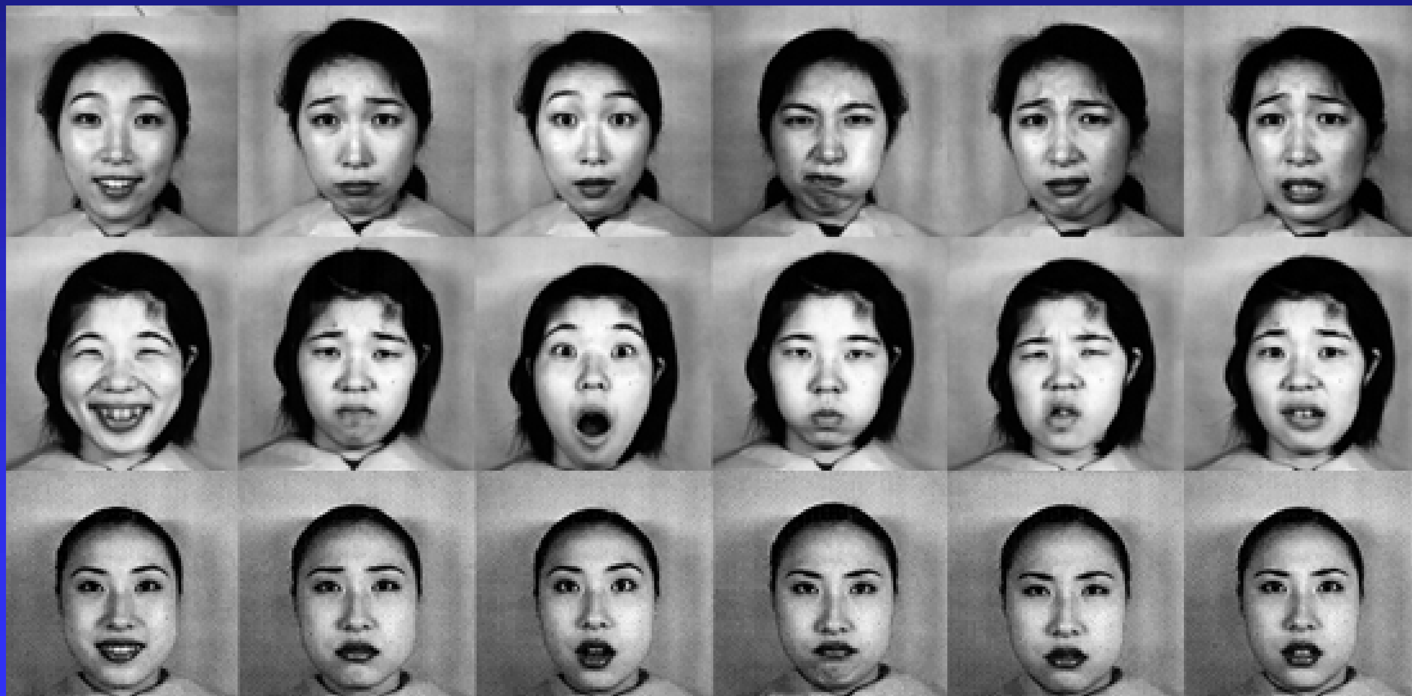
Facial Expression Recognition with Embedded HMM



An Embedded HMM



Small Database: 9 people, 6 expressions $\times$ 3, 256 $\times$ 256



Person-dependent

Expression	Anger	Disgust	Fear	Happiness	Sad	Surprise
Anger	**85.19**	7.40	3.70	0	3.70	0
Disgust	0	**88.89**	7.40	0	3.70	0
Fear	0	7.40	**93.60**	0	0	0
Happiness	0	3.70	7.40	**88.89**	0	0
Sad	0	3.70	11.11	0	**85.19**	0
Surprise	0	0	0	0	0	**96.30**

Person-independent

Expression	Anger	Disgust	Fear	Happiness	Sad	Surprise
Anger	**77.78**	22.22	0	0	0	0
Disgust	14.81	**62.97**	11.11	0	0	11.11
Fear	11.11	7.40	**51.85**	3.70	11.11	14.81
Happiness	0	0	14.81	**77.78**	7.40	0
Sad	7.40	7.40	18.51	0	**62.96**	0
Surprise	0	0	0	3.70	0	**96.30**

Gibbs Random Fields:

Gibbs distribution:

$$P(f) = \frac{e^{-E(f)/T}}{\sum_{f \in F} e^{-E(f)/T}}$$

where *E* is the energy function, *T* is the temperature.

A Random field:

$$F = \{F_1, \ldots, F_m\}$$

Configuration: a value assignment

$$f = \{f_1, \ldots, f_m\}$$

Only consider the discrete case

Define a neighborhood system *N* and energy function

$$E(f) = \sum_{c \in C} Vc(f)$$

- The energy is a sum of clique potentials over all possible cliques *C*
- Clique: a subset in which every pair are neighbors of each other.

Markov random fields

Positive:

$$P(f) > 0, \forall f \in F$$

Markovian: state only depends on neighbors

$$P(f_i \mid f_{S-\{i\}}) = P(f_i \mid f_{N_i})$$

Homogenious: probability independent of positions of sites

Markov-Gibbs Equivalence

GRF -- global property (the Gibbs distribution)

MRF -- local property (the Markovianity)

The Hammersley-Clifford theorem [1971] establishes the equivalence of these two:

F is an MRF on S with respect to N if and only if F is a GRF on S with respect to N .

Bayesian Interpretation

- the Bayes risk $R(f^*) = \int_{f \in F} C(f^*, f) P(f \mid d) df$
- the Bayesian rule

$$P(f \mid d) = \frac{p(d \mid f) P(f)}{p(d)}$$

- define a cost function $C(f^*, f) = \begin{cases} 0 & if \mid f^* - f \mid \le \delta \\ 1 & otherwise \end{cases}$
- minimizing the risk is equivalent to maximizing the posterior

$$f^* = \arg\max_{f \in F} P(f \mid d)$$

(a) maximization of the posterior probability in the Bayesian framework

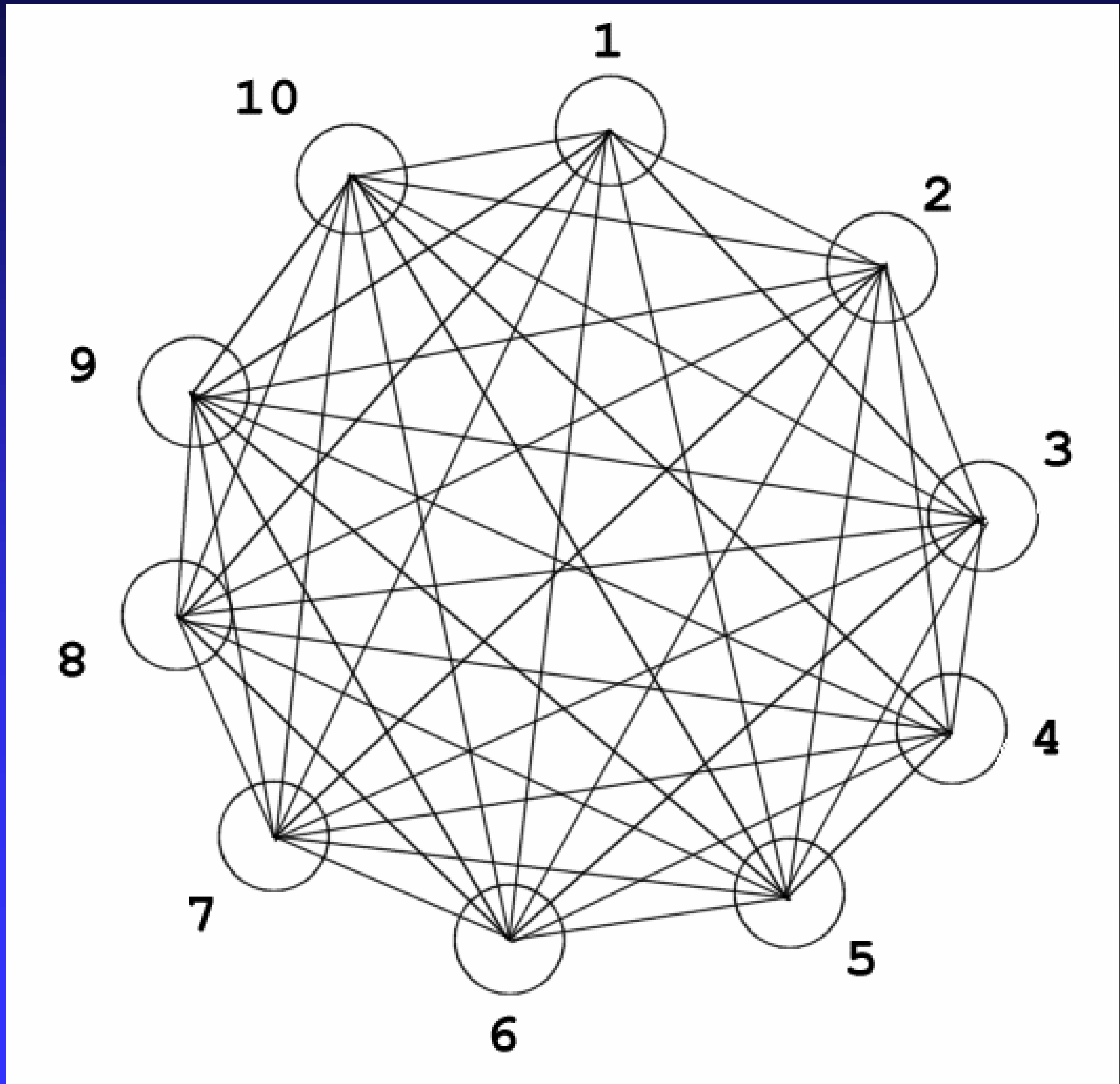
←→ (b) minimization of the posterior energy function of a MRF

←→ (c) minimization of the energy in a stochastic recurrent network

image restoration using MRF (S. Geman and D. Geman, 1984)

Bayesian labeling problem (Stan Z. Li, 2001)

A Recurrent Network



A binary network

A *recurrent stochastic binary network* *B(V,W,U)* is a pseudo-graph with vertex set *V* having state $S \in \{-1,+1\}^n$, edge set *W* of real value, a neighborhood structure *N*, and a dynamic updating mechanism *U*.

The state changes with updating rule

$$S_i = F\left(\sum_{j \in N} w_{ij} S_j\right)$$

where *F* is a random activation function.

Why a recurrent network?

- auto-associative memory

  can recall a memory with a corrupt or incomplete input
- sound theoretical basis in physics and math

  Ising model, Markov Random Fields,…
- powerful learning algorithms

Foreground/Background Discrimination

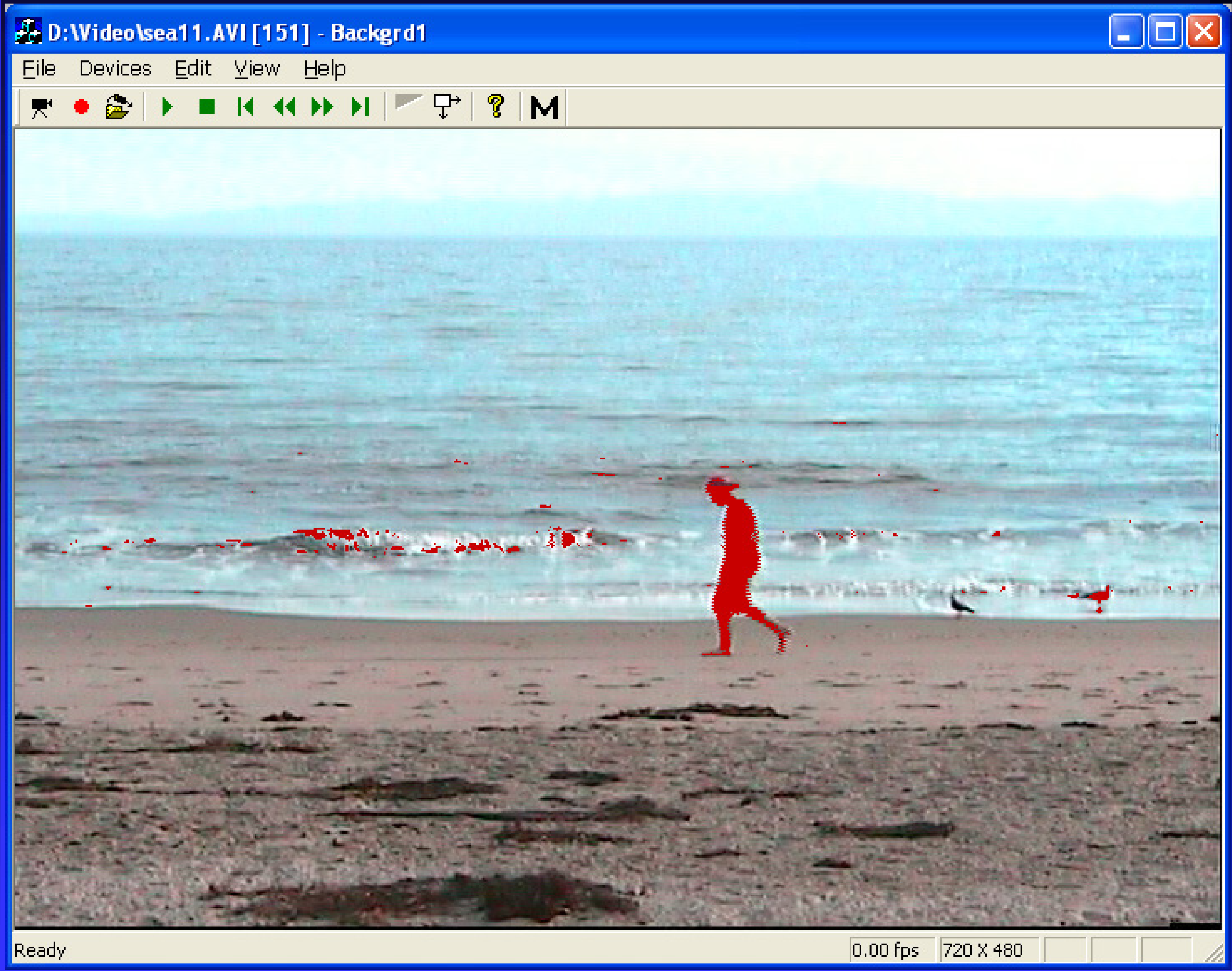
A recurrent binary network can be used to implement foreground/background discrimination

Find a right mapping:

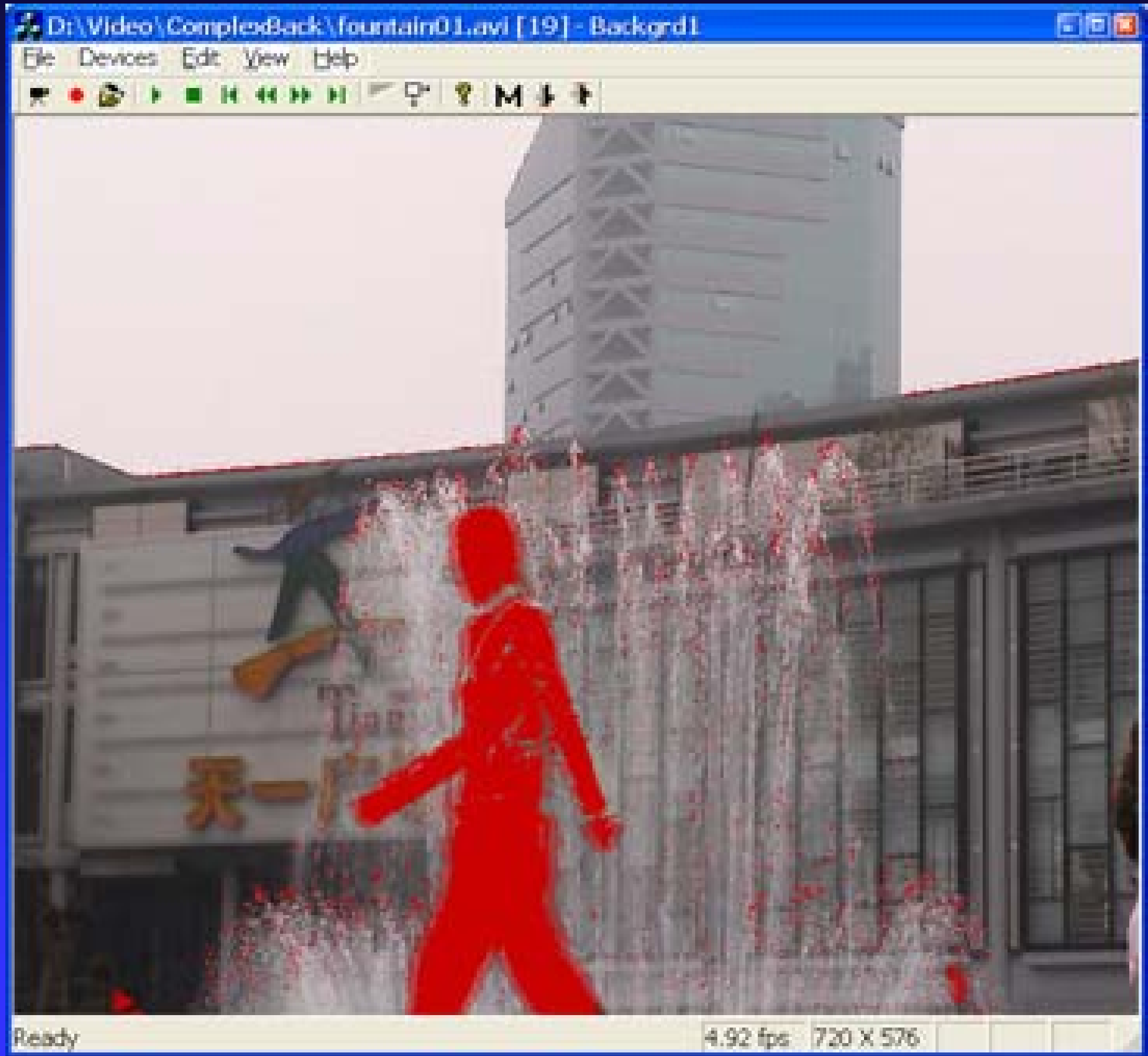
Segmentation $\leftrightarrow$ Energy Minimization

by appropriately setting connection weights,

Energy minimization with SA or BP



D:\Video\sea11.AVI [151] - Backgrd1
File Devices Edit View Help
Ready
0.00 fps
720 X 480


D:\Video\ComplexBack\fountain01.avi [19] - Backgrd1
File Devices Edit View Help
Tian
Ready
4.92 fps
720 X 576

Other Related Topics

- **Computer vision – Generative model or discriminative model ?**
- **Human vision – How we see?**
  “perceptual filling-in”

Generative Model

- Given a problem domain with variables $X_1,..,X_T$ system is specified with a joint pdf $P(X_I,..,X_T)$
- Called generative model since we can generate more samples artificially
- Given a full joint pdf we can

Marginalize

$$P(X_j) = \sum_{\forall X_{i,i \neq j}} P(X_1,..,X_n)$$

Condition

$$P(X_j \mid X_k) = \frac{P(X_j, X_k)}{P(X_k)}$$

By conditioning a joint pdf we can easily form

– Classifiers, regressors, predictors

Discriminative Model

- Make no attempt to model underlying distributions
- Only interested in optimizing a mapping from inputs to desired outputs
- Focuses model and computational resources on given task and provides better performance

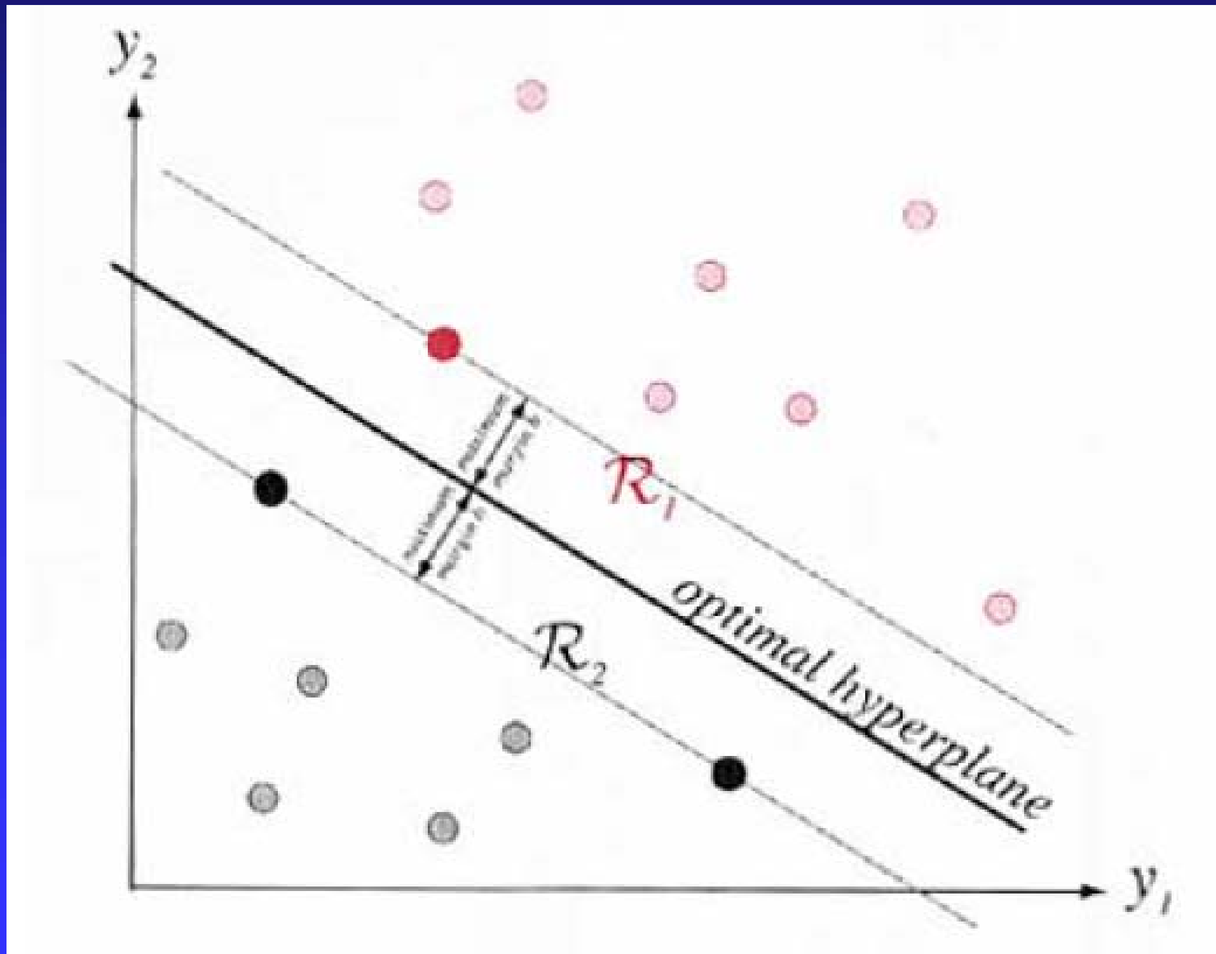
Examples:

– logistic regression, sigmoid

$$P(y=1 \mid X) = \frac{1}{(1+\exp(-\theta^T X))}$$

– SVMs

SVM finds hyperplane with maximum distance from nearest training patterns



Computer vision - generative or discriminative

Generative classifiers:

- learn the joint probability p(x,y), x-inputs, y-label
- calculate p(y|x), predict and pick the most likely

Pros: powerful; can handle missing data;
better performance with few data

Cons: complex, time consuming

Discriminative classifiers

- model the posterior p(y|x) directly.

Pros: efficient, higher accuracy

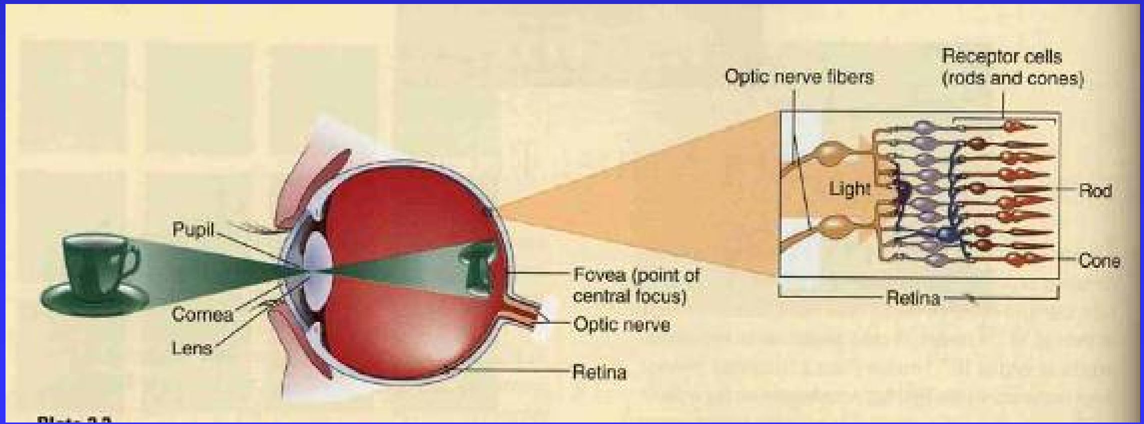
Cons: cannot handle missing data

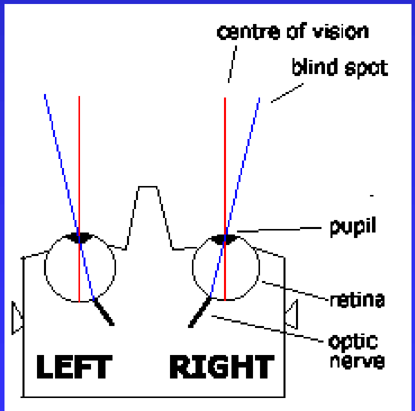
A hybrid model could be better

Human vision

- **Perceptual Filling-in**

  a famous visual illusion, the brain fills in the missing information across the physiological blind spot





So what we see is not strictly a reflection of the physical inputs (to the retina),

but instead it is highly dependent on the processes by which our brain attempts to interpret the scene.

Our brain is a very powerful generative model !

Conclusions and Future Work

- A platform developed
- Robust Facial Expression Recognition in real time is hard
- A powerful graphical model needed
- Applications

Thank You !