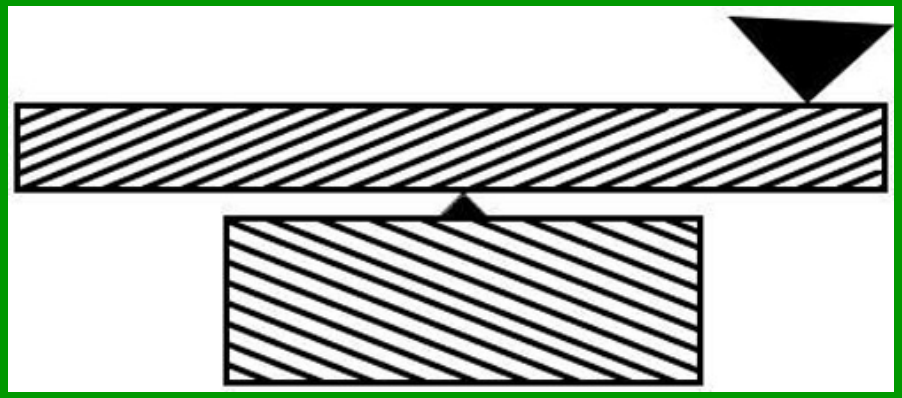
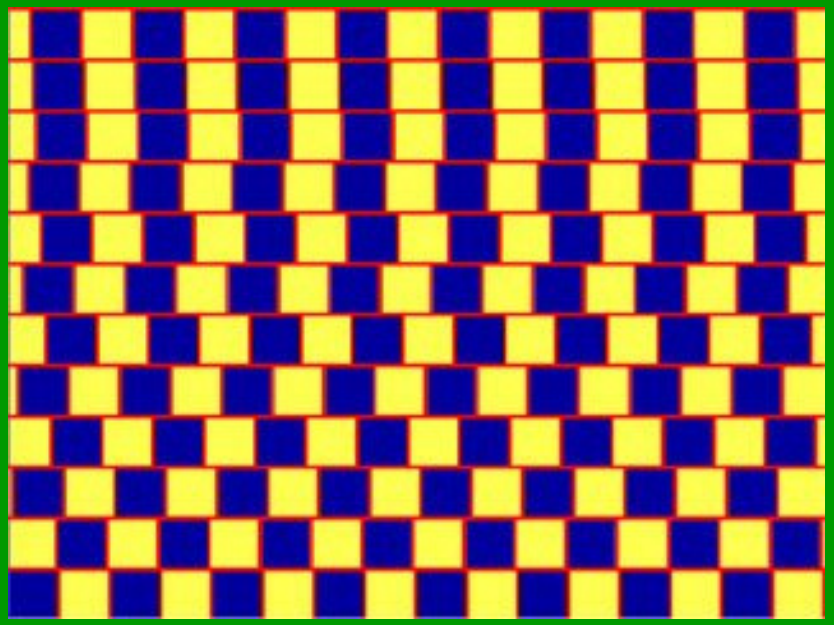
Module 8
Image Analysis II

- Models of Visual Cortex
- Oriented Pattern Analysis
- Range Finding by Stereo Imaging
- Advanced Applications:
  Iris and Face Recognition

QUICK INDEX

More Orientation Illusions





An actual building in Paris

Photo is not doctored.

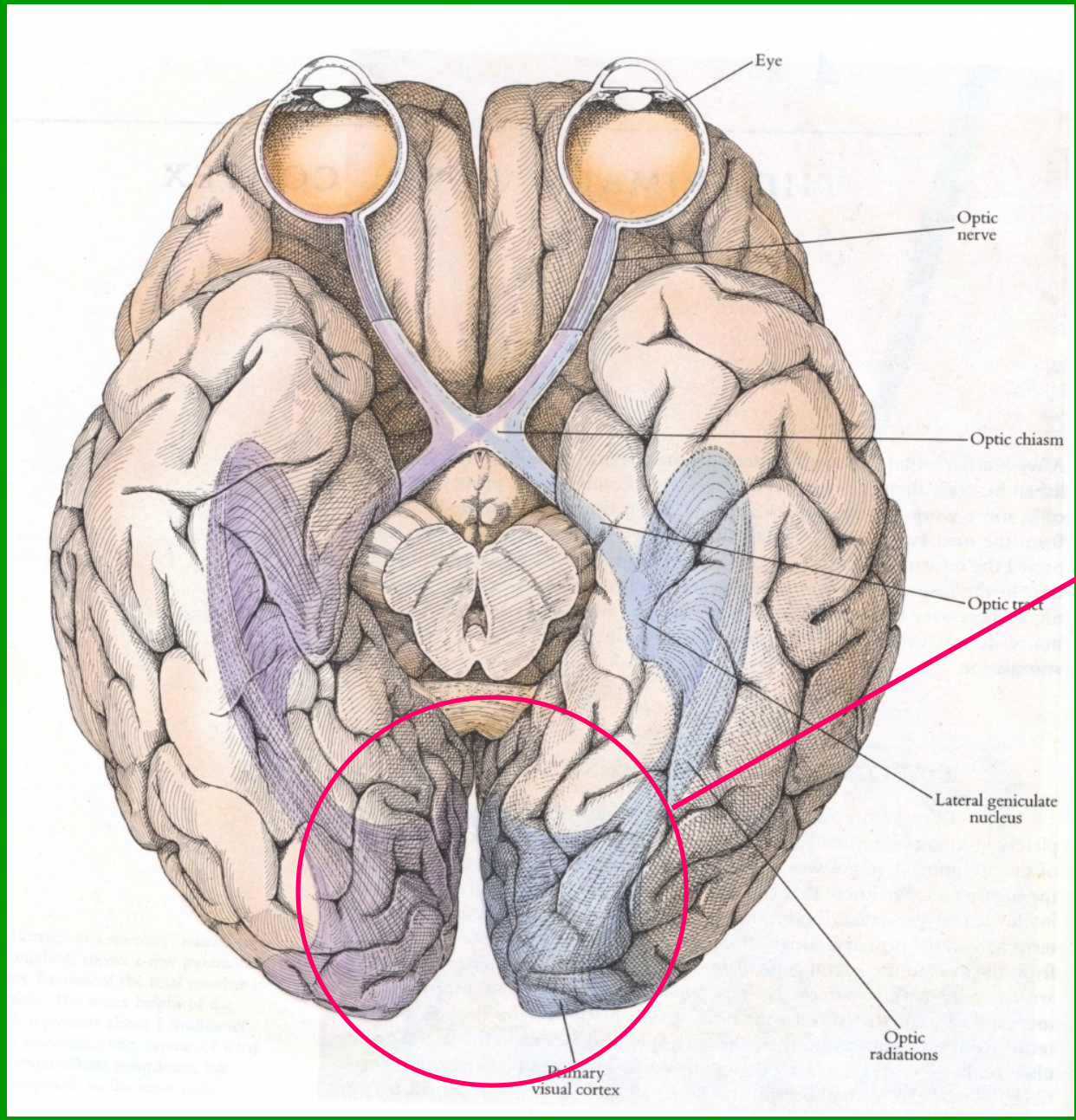
Let’s revisit the visual brain…

Primary Visual Cortex

- Also called **Area V1** or **striate cortex**. Much goes on here.

- **Visual signals** from the ganglion cells of the eye are **further decomposed** into orientation- and scale-tuned **spatial** and **temporal channels**.

- These are passed on to other areas which do **motion analysis, stereopsis,** and **object recognition.**



Primary visual cortex (from below)

From D. Hubel, *Eye and Brian*

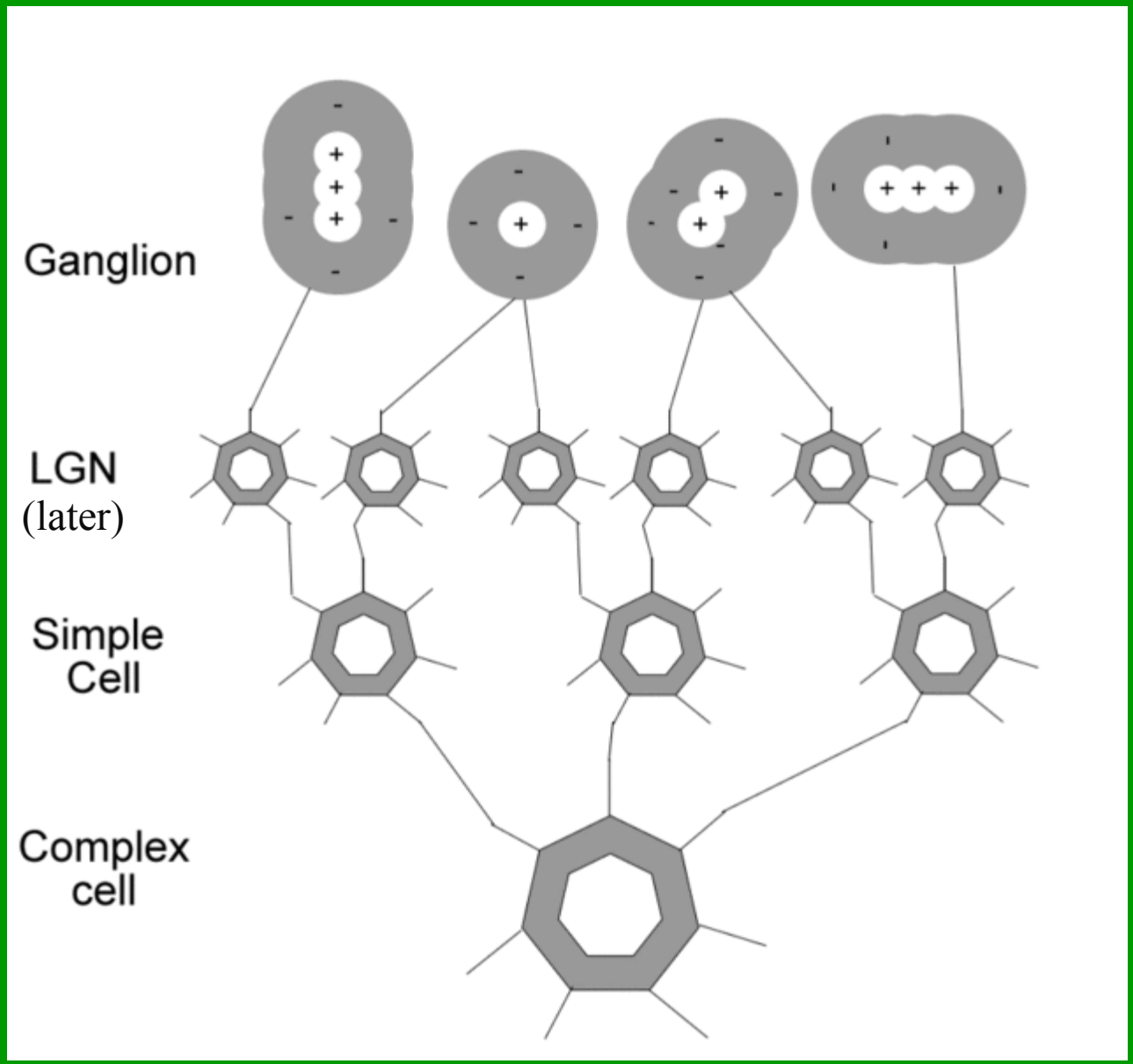
Types of Cortical Neurons

- Early 1960's, Nobel laureates **D. Hubel and T. Wiesel** performed visual experiments on cats.
- Inserted electrodes into **visual cortex**, presented patterns to animals' eyes, **measured neuronal responses.**
- Found **two general types of neurons** termed **simple cells** and **complex cells.**
- Defined by their **spatial responses to images.**

Simple and Complex Cells

- **Simple cells** are well-modeled as **linear.** We can model their outputs by linear convolution with the image signal.

- **Complex cells** receive signals from the **simple cells.** Their responses are **non-linear.**



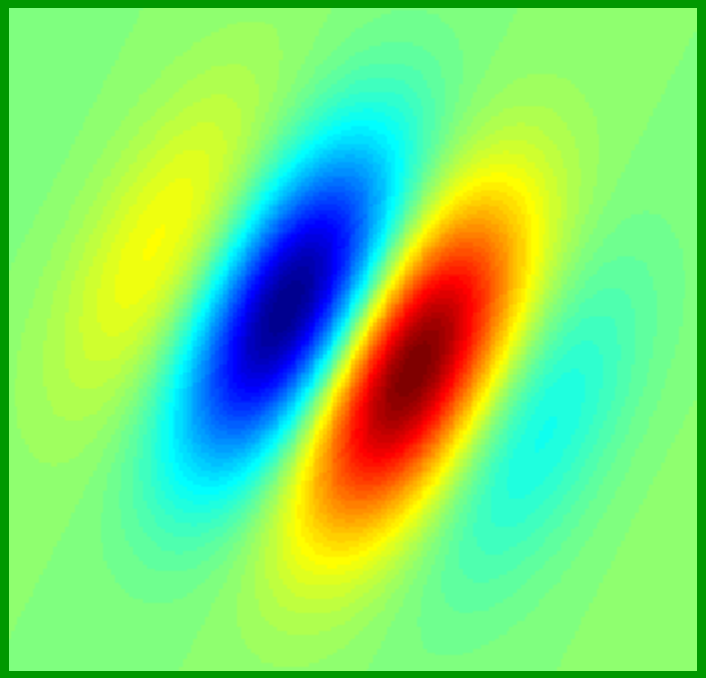
Stages on the visual pathway from retina to cortical cells.

Simple Cell Responses

- **Simple cells** respond to (at least) **three signal aspects:**
    - Spatial **pattern orientation, frequency** and **location**
    - Spatial **binocular disparity**
    - Spatio-temporal pattern **motion**

Simple Cell Spatial Response

- Simple cell **responses** have excitatory and inhibitory regions – multiple lobes.
- Appear to match spatial edges (shown) and bars.

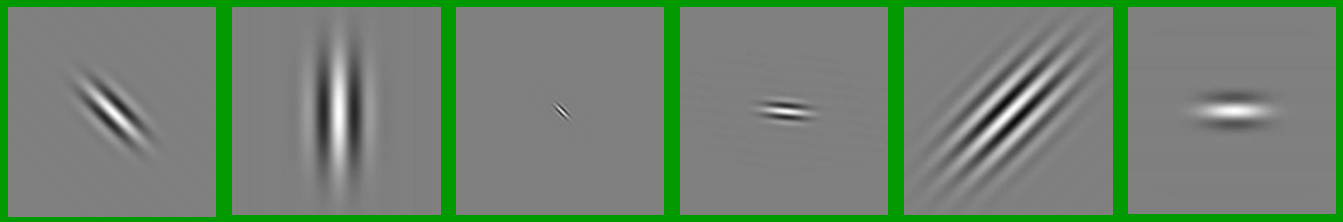


Red: excitatory

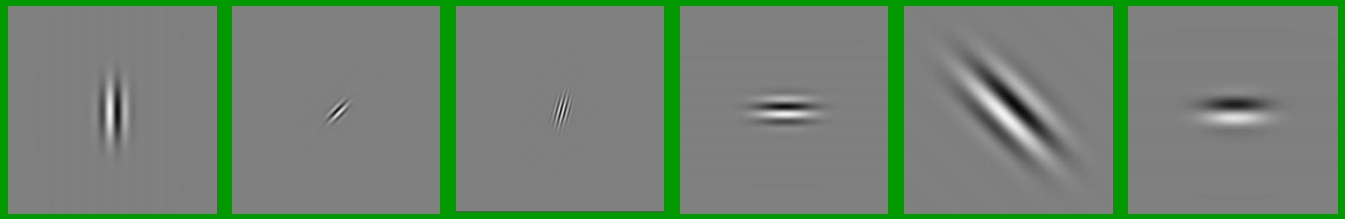
Blue: Inhibitory

Simple Cell Distribution

- The **simple cells** spatial responses correspond to a wide range of **orientations**, **lobe separations,** and **sizes.**



"Bar-sensitive" simple cells



"Edge-sensitive" simple cells

- Seemingly "spatially tuned" to detect edges and bars, thus representing images that way. They are well modeled as ...

2-D Gabor Function Model

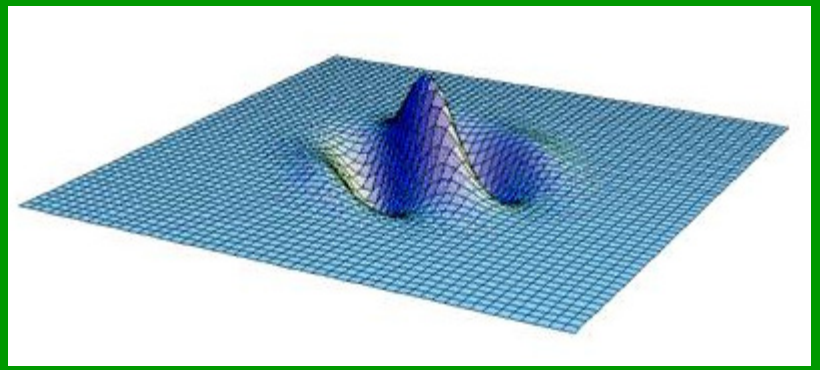
- **Gabor functions** are **Gaussian functions** that **modulate** (multiply) **sinusoids.** λ controls **elongation.**

$$g_c(x,y) = \frac{1}{2\pi\lambda\sigma^2} e^{-\left[\left(x/\lambda\right)^2 + y^2\right]/2\sigma^2} \cos\left[2\pi\left(\frac{u_0}{N}x + \frac{v_0}{M}y\right)\right]$$

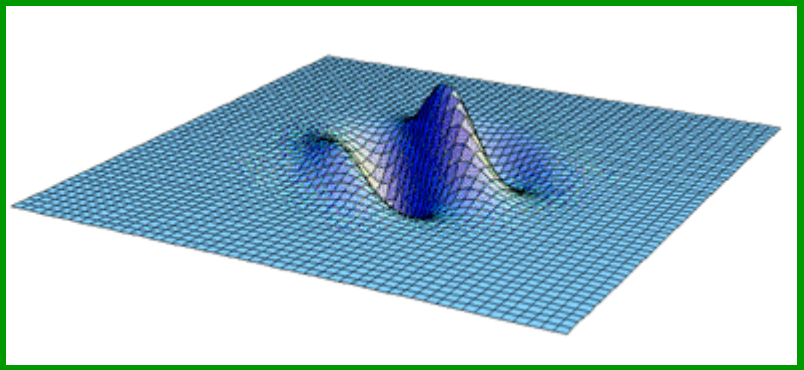
$$g_s(x,y) = \frac{1}{2\pi\lambda\sigma^2} e^{-\left[\left(x/\lambda\right)^2 + y^2\right]/2\sigma^2} \sin\left[2\pi\left(\frac{u_0}{N}x + \frac{v_0}{M}y\right)\right]$$

- **"Cosine and "sine" Gabors** are in 90 deg quadrature and model the bar and edge simple cells, respectively

Phase Quadrature



"Edge-sensitive"



"Bar-sensitive"

- The **simple cells** appear to occur in **phase-quadrature pairs** – 90 deg out of phase

Quadrature Gabor Pair

"Bar-sensitive" cosine Gabor "Edge-sensitive" sine Gabor

- The simple cells appear with widely variable **orientations**, **elongations**, and **sizes.**

Phasor Form of Gabor

- **Easy** to **manipulate**

$$g(x,y) = g_c(x,y) + \sqrt{-1}\, g_s(x,y)$$

$$= \frac{1}{2\pi\lambda\sigma^2} e^{-\left[\left(x/\lambda\right)^2 + y^2\right]/2\sigma^2} e^{2\pi\sqrt{-1}\left(u_0 x/N + v_0 y/M\right)}$$

- **Natural representation** of **simple cell pairs**
- **Fourier transform** is simply a **shifted Gaussian**

$$\tilde{G}(u,v) = e^{-2(\pi\sigma)^2\left[(u-u_0)^2\lambda^2 + (v-v_0)^2\right]}$$

Digital Form

$$g(i,j) = g_c(i,j) + \sqrt{-1}\, g_s(i,j)$$

$$= \frac{1}{2\pi\lambda\sigma^2} e^{-\left[\left(i/\lambda\right)^2 + j^2\right]/2\sigma^2} e^{2\pi\sqrt{-1}\left(u_0 i/N + v_0 j/M\right)}$$

- Similar considerations as designing LoG

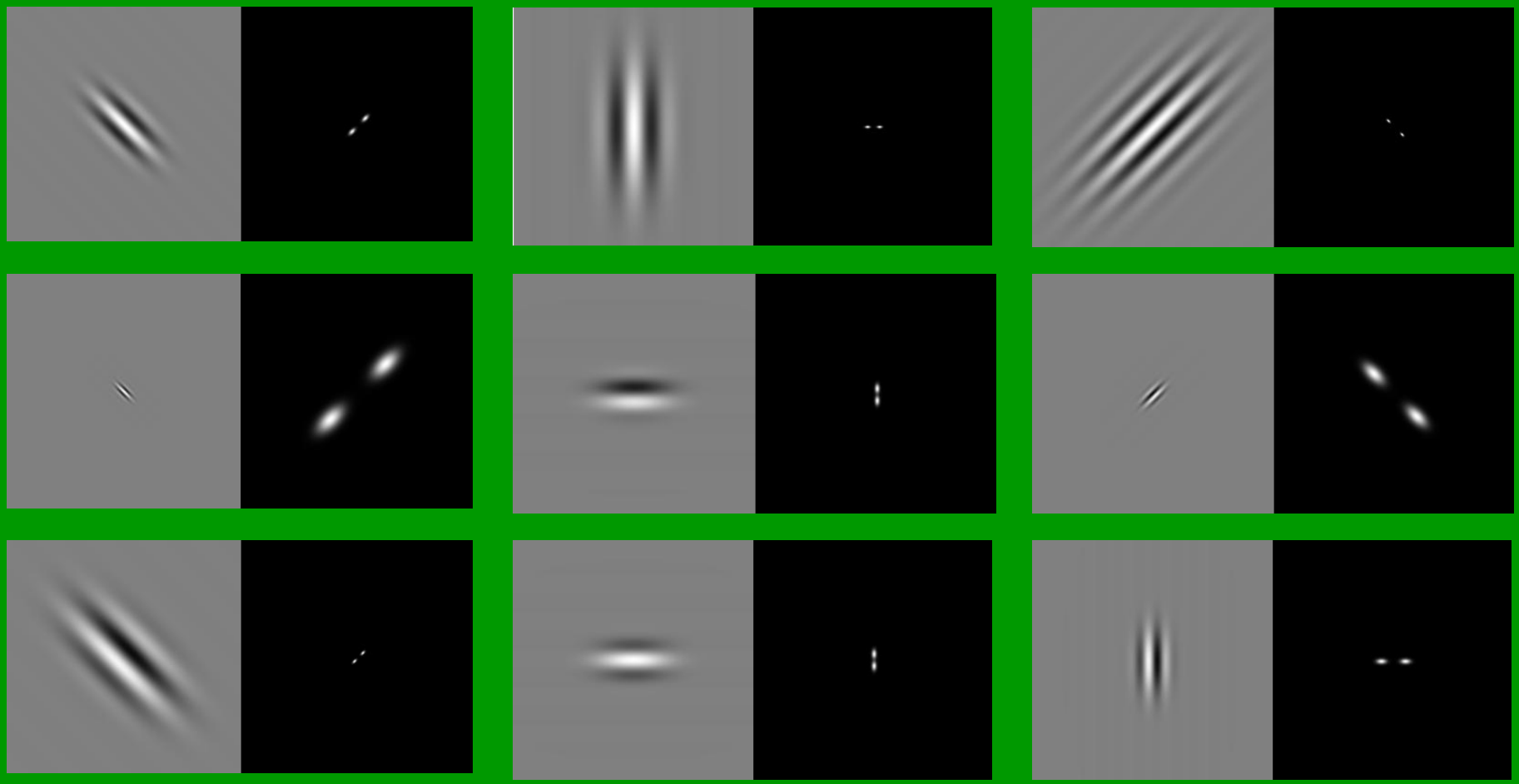
Minimum Uncertainty

- **Amongst all complex functions** and in any dimension, Gabor functions uniquely minimize the **uncertainty principle:**

$$\left(\frac{\int |xf(x)|^2\,dx}{\int |f(x)|^2\,dx}\right)\left(\frac{\int |uf(u)|^2\,du}{\int |f(u)|^2\,du}\right) \geq \frac{1}{4}$$

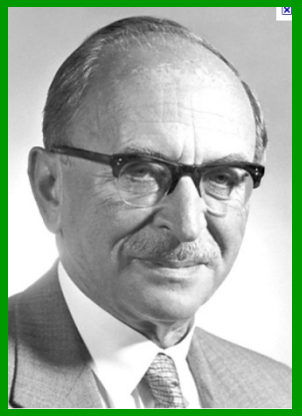
- Similar for y, v.
- They have **minimal simultaneous space-time duration.**
- Given a bandwidth, can perform the **most localized spatial analysis.**

Fourier Transform Magnitudes of Gabors



Sine and cosine Gabors have same FT magnitudes. **Phasor form only half plane**.

Gabor Function History



- **Gabor functions** first studied in context of information theory by Nobel laureate **D. Gabor.**

- In 1980 **S. Marcelja** noted that the simple cell receptive fields measured along 1D are well-modeled by **1-D Gabor functions** which have **optimal space-frequency localization.**

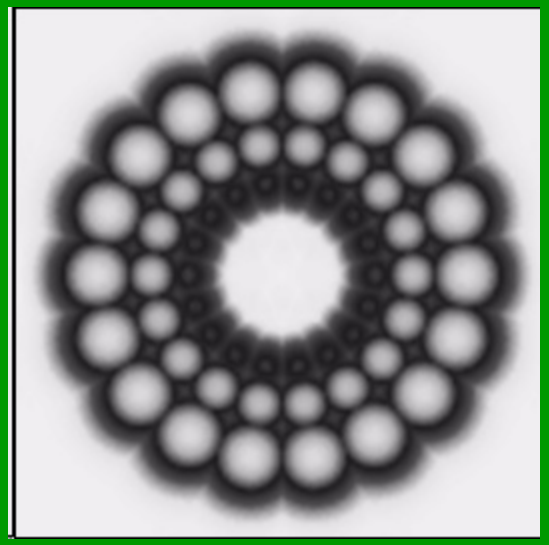
- In 1981 **D.A. Pollen and S.F. Ronner** showed that the simple cells often appear in **phase quadrature pairs,** which are natural for Gabor functions.

- In 1985 **J. Daugman** observed that receptive field profiles fit **2D Gabor functions** and have **optimal 2D space-frequency localization.**

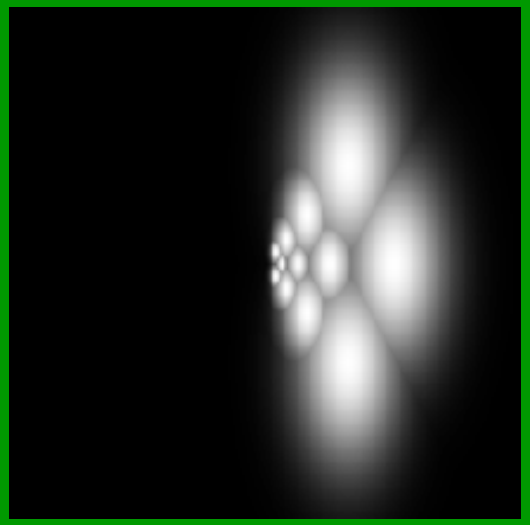
Gabor Filterbanks

- Exact arrangement of simple cells not known, but all orientations are possible **(horizontal/vertical more common)**.
- The cortex is layered. Simple cells that are in **different layers** but **similar positions** come from the same eye and are responsive to similar orientations. Called **ocular dominance columns.**
- Bandwidths in the range 0.5 octave – 2.5 octave are common.
- In engineering, **constant-octave filterbanks** covering the plane are common.

Gabor Filterbanks



- **Constant-octave** Gabor filter bank (cosine form)
- Nine orientations, four filters/orientation



- Constant-octave Gabor filter bank (phasor form)
- Three orientations, four filters/orientation

The Image Modulation Model

(where we model an image as a sine wave!)

$$I(x, y) = A\cos\left[\phi(x, y)\right]$$

FM Approximation

- We'll do this in 1-D for simplicity. A 1-D continuous Gabor:

$$g(x) = g_c(x) + \sqrt{-1}\, g_s(x)$$

$$= \frac{1}{\sqrt{2\pi\sigma}} e^{-x^2/2\sigma^2} e^{2\pi\sqrt{-1}(u_0 x)} \quad \Longleftrightarrow \quad \tilde{G}(u) = e^{-2(\pi\sigma)^2 (u-u_0)^2}$$

- 1-D FM image: $I(x) = A\cos[\phi(x)]$ Assuming the **instantaneous frequency** $\phi(x)$ changes slowly, then

$$g(x) * I(x) \approx A \cdot \tilde{G}[\phi'(x)] \cdot \exp\left[\sqrt{-1}\phi(x)\right]$$

- Contains responses of both cos Gabor and sin Gabor filters

2D FM Approximation

- 2-D continuous Gabor:

$$g(x,y) = g_c(x,y) + \sqrt{-1}\, g_s(x,y) \quad \Longleftrightarrow \quad \tilde{G}(u,v)$$

- 2-D FM image: $I(x,y) = A\cos\left[\phi(x,y)\right]$ Assuming the instantaneous frequencies $\phi_x(x,y)$, $\phi_y(x,y)$ change slowly, then

$$g(x,y) * I(x,y) \approx A \cdot \tilde{G}\left[\phi_x(x,y), \phi_y(x,y)\right] \cdot \exp\left[\sqrt{-1}\phi(x,y)\right]$$

- Contains responses of both **cos** Gabor and **sin** Gabor filters

FM Demodulation

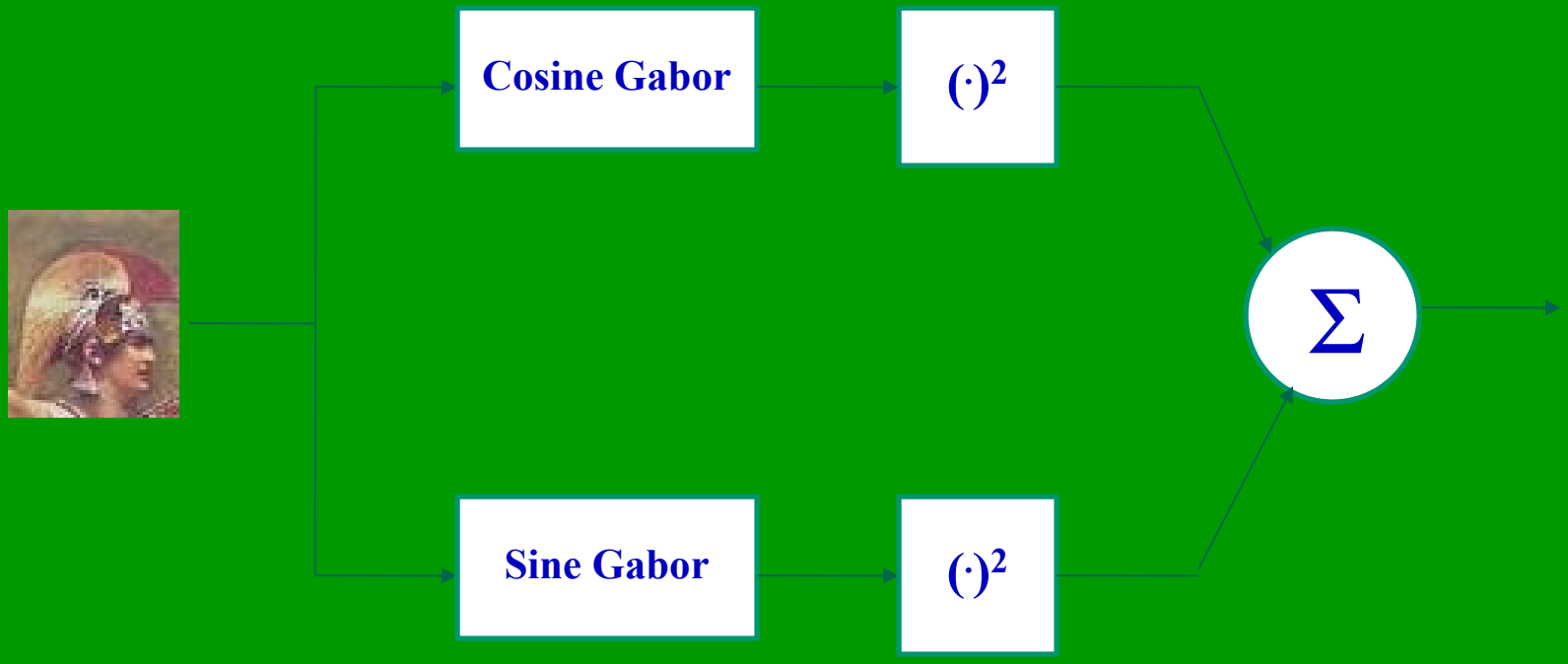
- 1-D Gabor-filtered FM image demodulation:

$$\left|g(x) * I(x)\right| \approx A \cdot \tilde{G}\left[\phi'(x)\right]$$

- 2-D Gabor-filtered FM image demodulation:

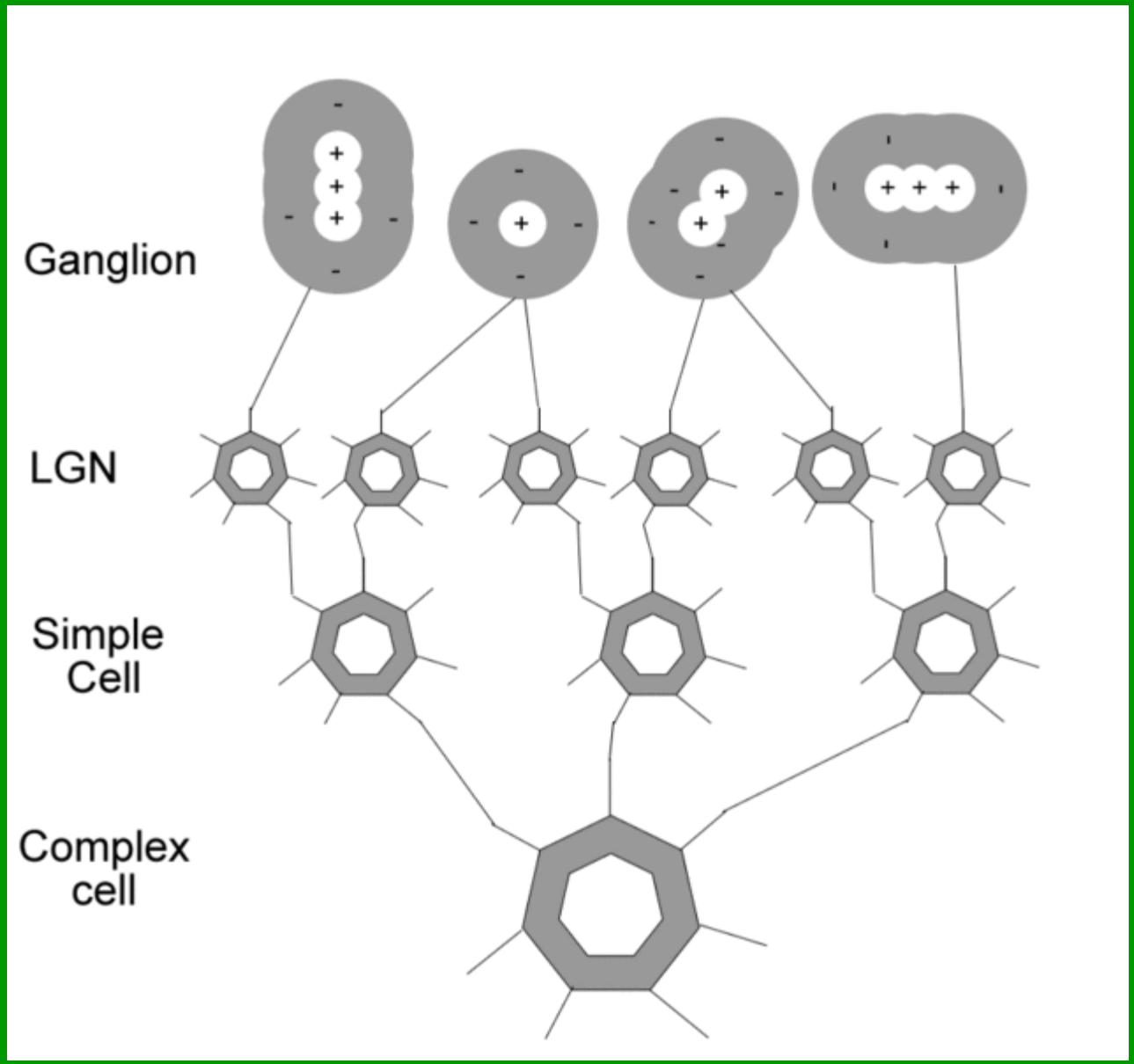
$$\left|g(x,y) * I(x,y)\right| \approx A \cdot \tilde{G}\left[\phi_x(x,y), \phi_y(x,y)\right]$$

Energy Model for Quadrature Simple Cell Demodulation



Complex Cells

- **Less well-understood.** Poorly understood and nonlinear responses**.**
- Viewed as accepting **simple cell responses as inputs,** and **processing them** in ways that are not yet understood.
- Simplest task might be the demodulation process on the **previous slide.**



While we do not know how the vision system processes the spatial simple cell (Gabor-like) responses, we can certainly use the model creatively for image processing.

Using V1 Cortical Cell Models

- **Massive decomposition** of **space-time visual data** provides optimally localized atoms of **pattern, pattern change, and pattern disparity information.**

- It's estimated that >80% of the functionality of the simple cells **remains unknown!**

- However, it is known that much of this information is passed to **other brain centers** – to accomplish **other processing**.

Spatial Orientation Estimation

- Now assume an image has a **strong oriented component** modeled as an **FM function**

$$I(\mathbf{x}) = A \cos [\phi(\mathbf{x})], \mathbf{x} = (x, y)$$

- Like **sine wave gratings** discussed earlier, but **local frequencies** and **orientations change.**

- **Goal:** Estimate $\nabla \phi(\mathbf{x}) = \left[\phi_x(\mathbf{x}), \phi_y(\mathbf{x}) \right]$

Frequency Modulation

- The **instantaneous frequencies** define the **direction** and **rate of propagation** of the pattern:

$$\left|\nabla \phi(\mathbf{x})\right| = \sqrt{\phi_x^2(\mathbf{x}) + \phi_y^2(\mathbf{x})}$$

$$\nabla \phi(\mathbf{x}) = \mathrm{Tan}^{-1}\left[\phi_y(\mathbf{x}) / \phi_x(\mathbf{x})\right]$$

Demodulation

- Simple approach:

  (1) Pass image through **bank** of complex Gabor filters

$$t_i(\mathbf{x}) = I(\mathbf{x}) * g_i(\mathbf{x});\ i = 1, \ldots, K$$

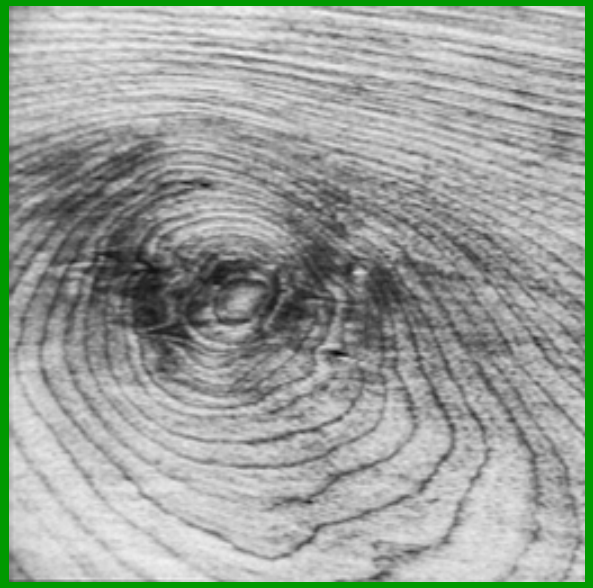
  (2) At each **x** find largest magnitude response (**demodulate**)

$$i^* = \arg\max_i |t_i(\mathbf{x})|$$
$$t^*(\mathbf{x}) = t_{i^*}(\mathbf{x})$$

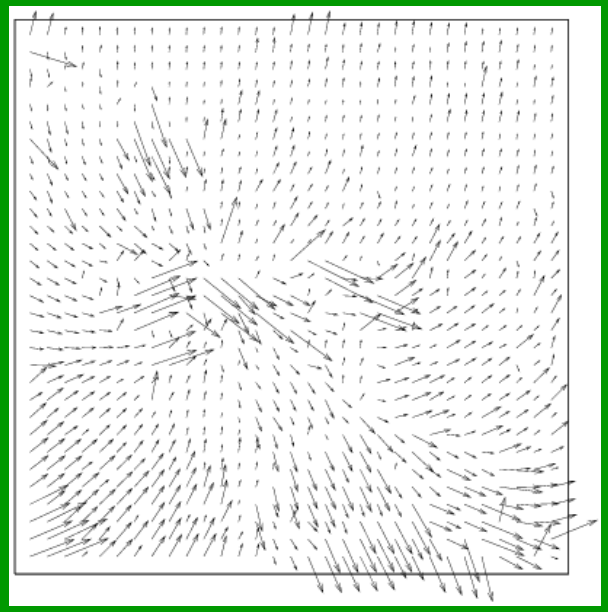
  (3) Then

$$\nabla\phi(\mathbf{x}) \approx \mathrm{Re}\left[\frac{\nabla t^*(\mathbf{x})}{\sqrt{-1}\cdot t^*(\mathbf{x})}\right]$$

Example



Image

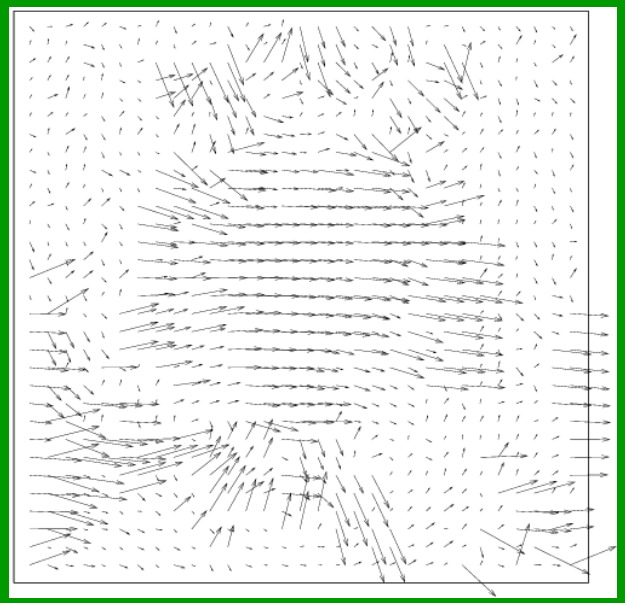


Orientation and frequency needle map

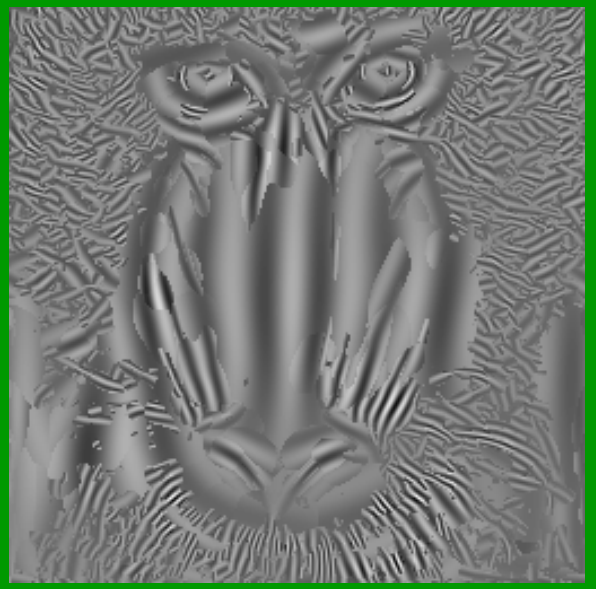
Reconstructed FM component

Example

Image



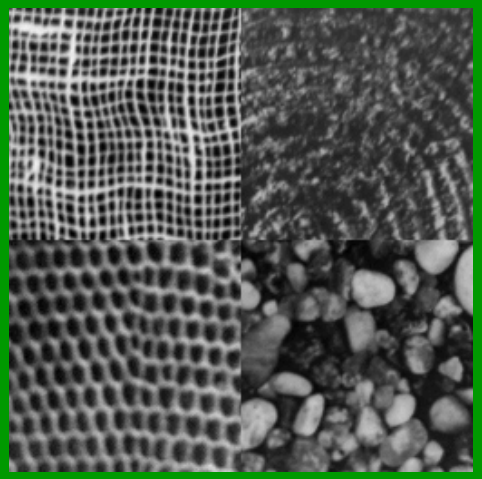
Orientation and frequency needle map

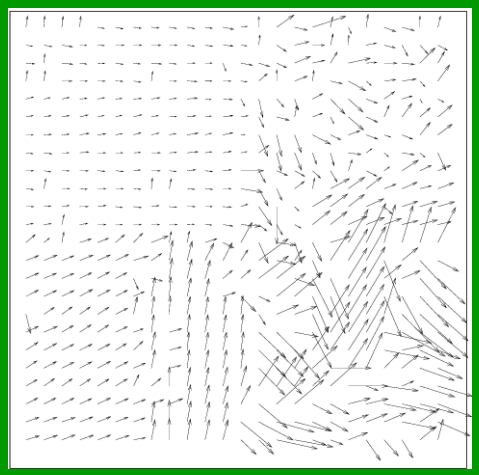


Reconstructed FM component

Segmentation by Clustering

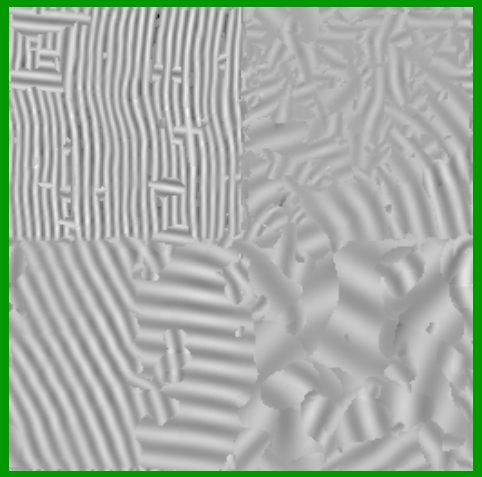
Multi-partite image





Orientation and frequency needle map

Reconstructed FM component



Segmentation by *k*-means clustering

Comments

- **Preceding algorithm** utilizes the linear filter responses which closely **model V1 simple cell responses.**

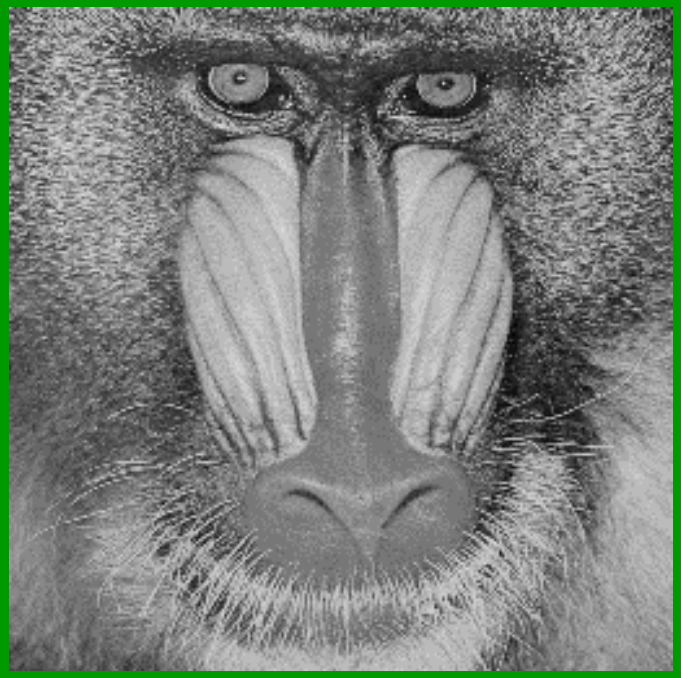
- The algorithms also use **nonlinear processing of many simple cells responses** which could be **similar to V1 complex cell processing.**

- The **image modulation model** can be made much more general:

$$I(x, y) = \sum_{i=1}^{K} A_i \cos\left[\phi_i(x, y)\right]$$

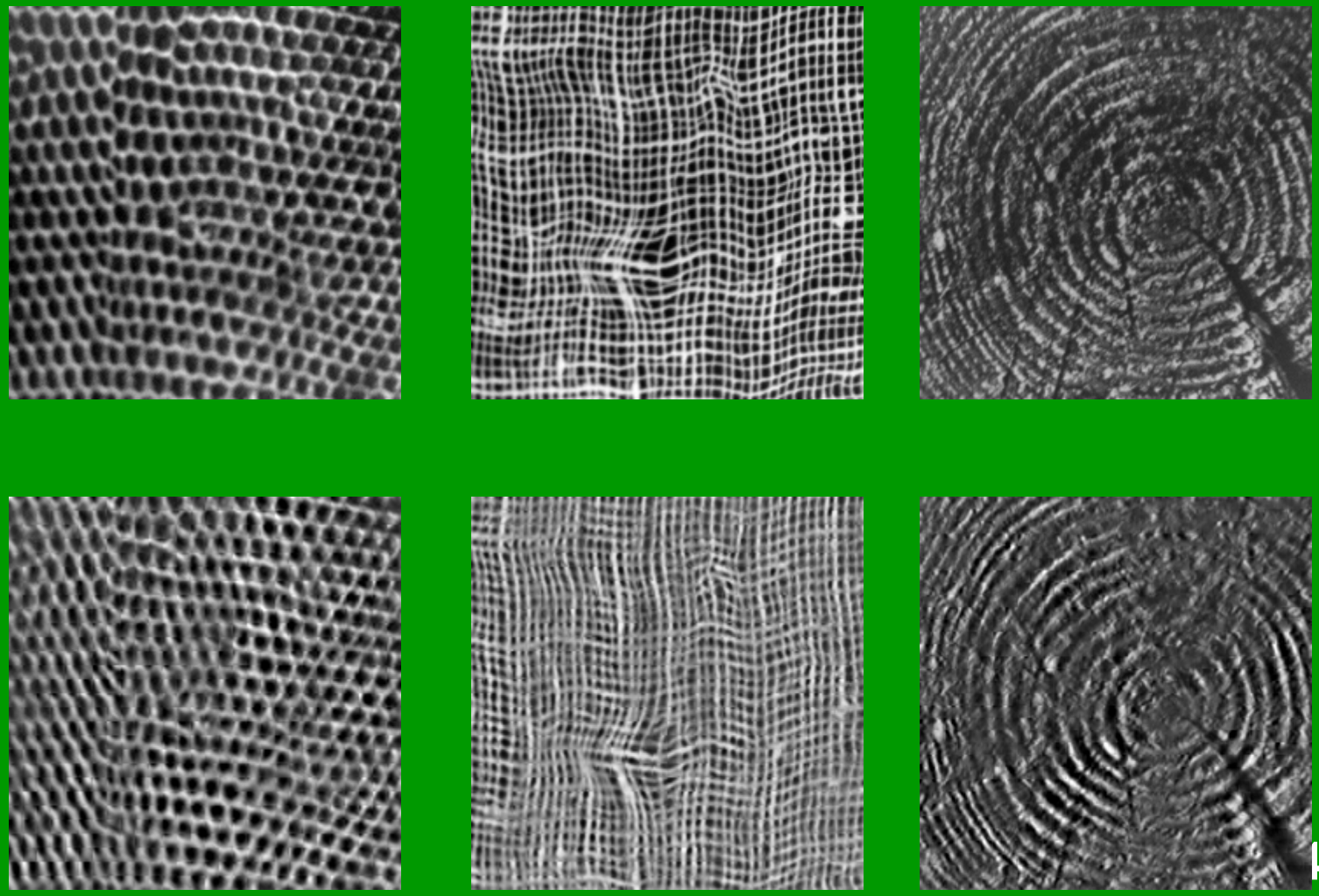
K=43 Components

K=43 Components

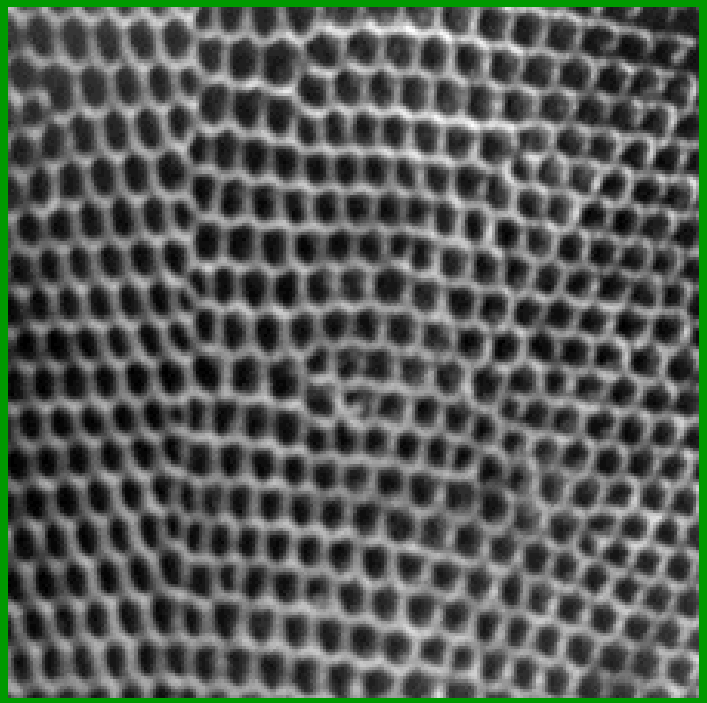


K=43 Components

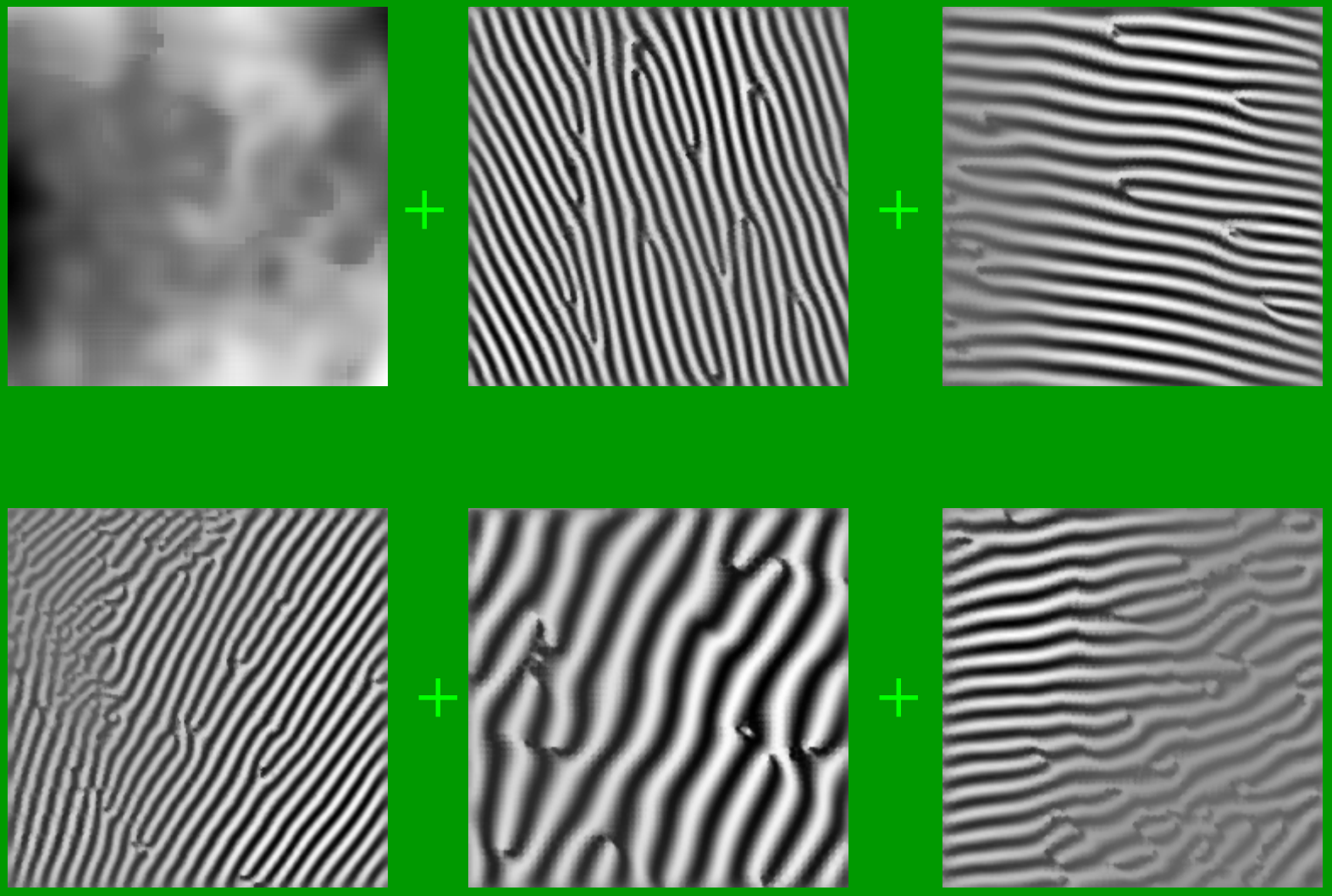
K=43 Components



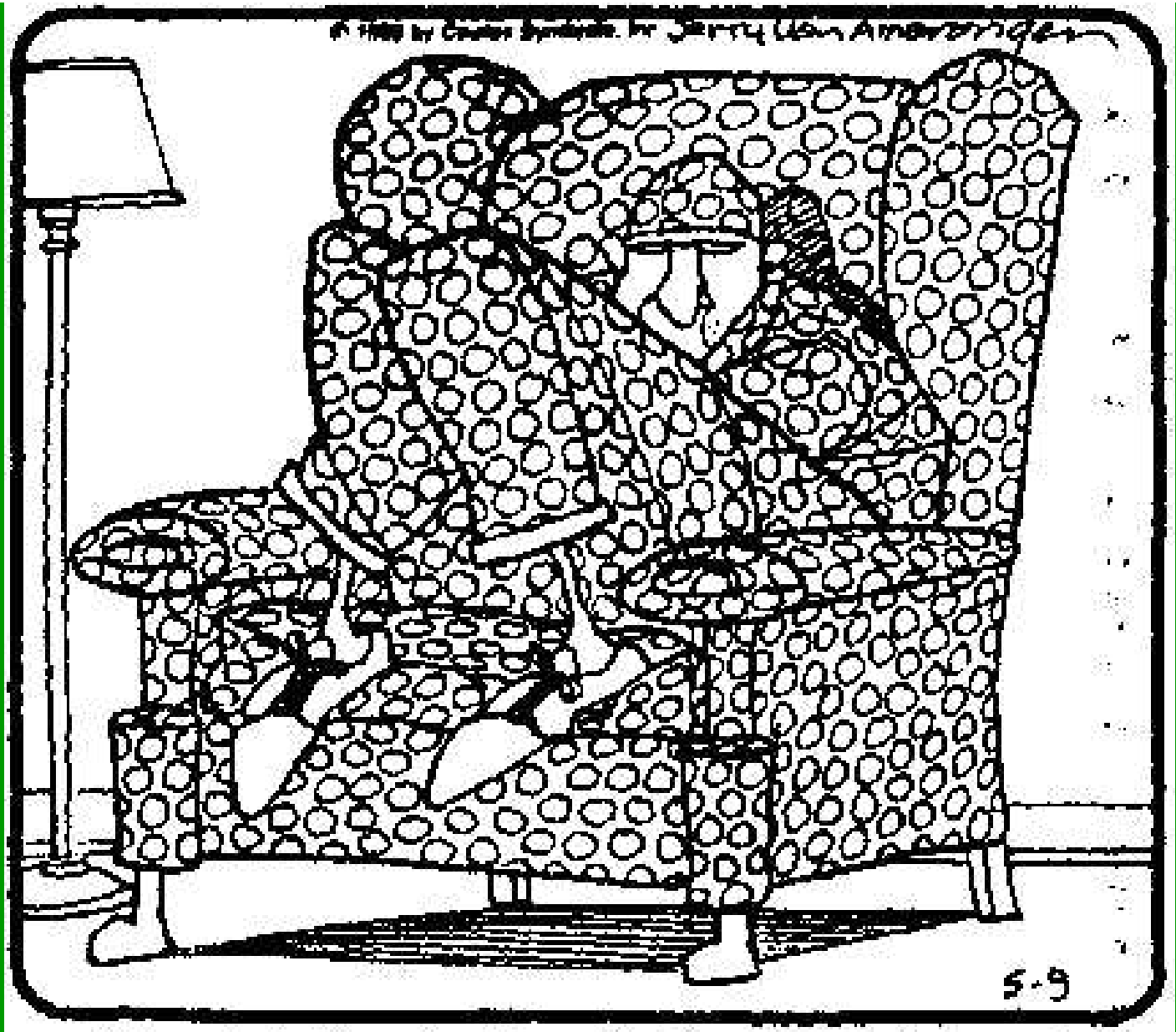
K=6 (only!) Components



The 6 Components

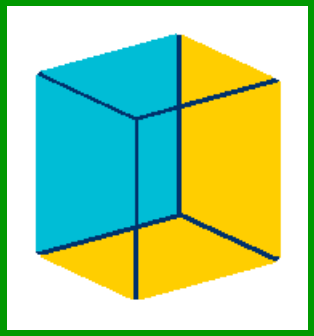


When the monster came, Lola, like the peppered moth and the arctic hare, *remained motionless and undetected.* Harold, of course, was immediately devoured.



You get the feeling Bob's not going out and grabbing life by the throat anytime soon.

Illusions of The Third Dimension



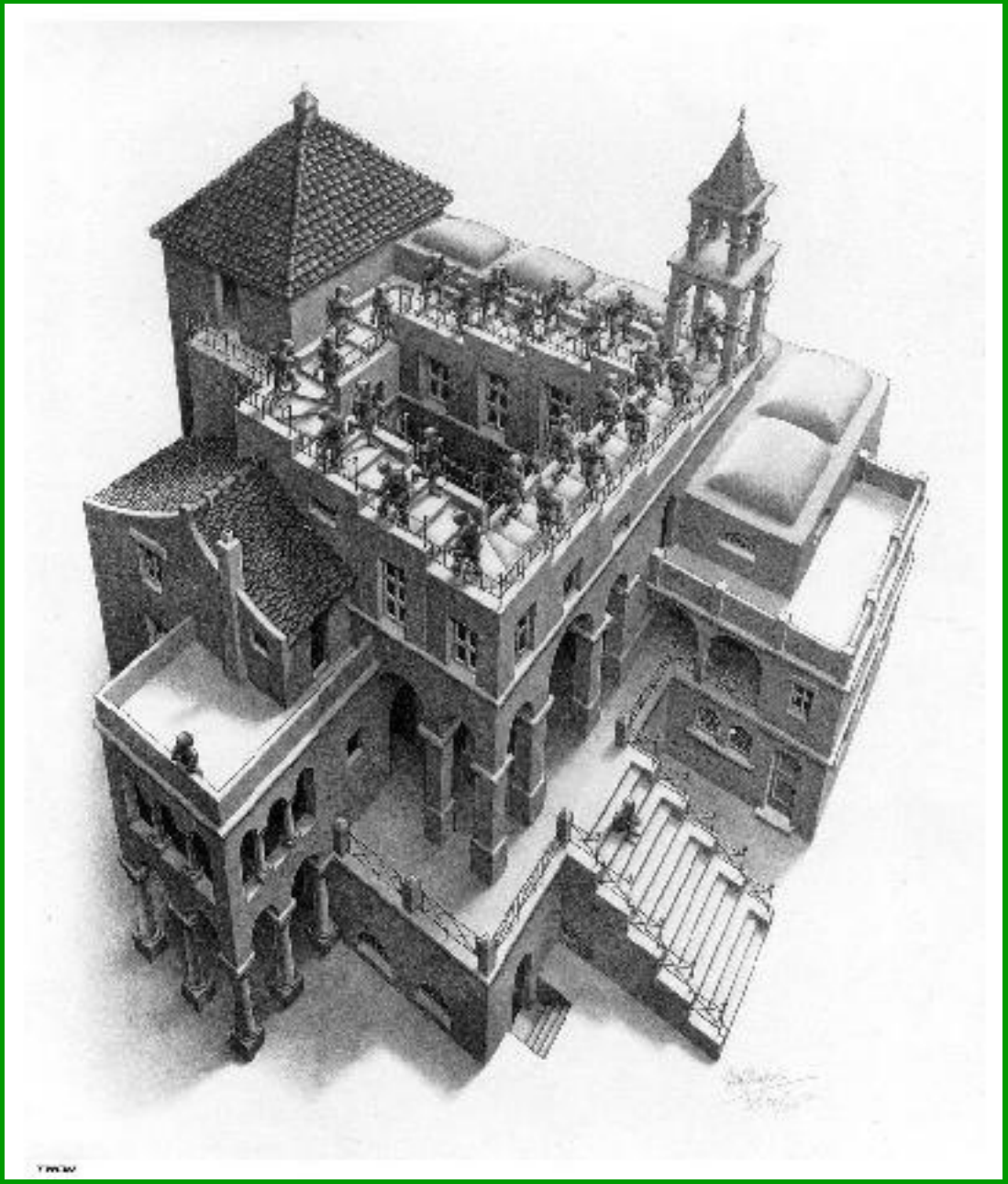
Which face is blue?

Tile Floor?

Mobius?

Three models (sizes) of cars?


Schmeckt
hervorragend



Ascending and Descending

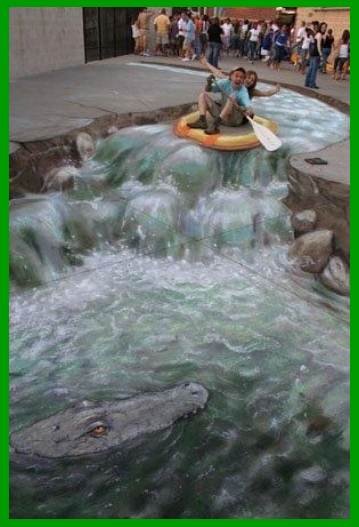
M.C. Escher

Amazing 3D Art Illusions

- All of the following are **two-dimensional** sidewalk paintings!
- They are all done with **chalk!**
- The technique is called **“anamorphosis”**

Perspective Illusions in 3-D Art

Perspective Illusions in 3-D Art

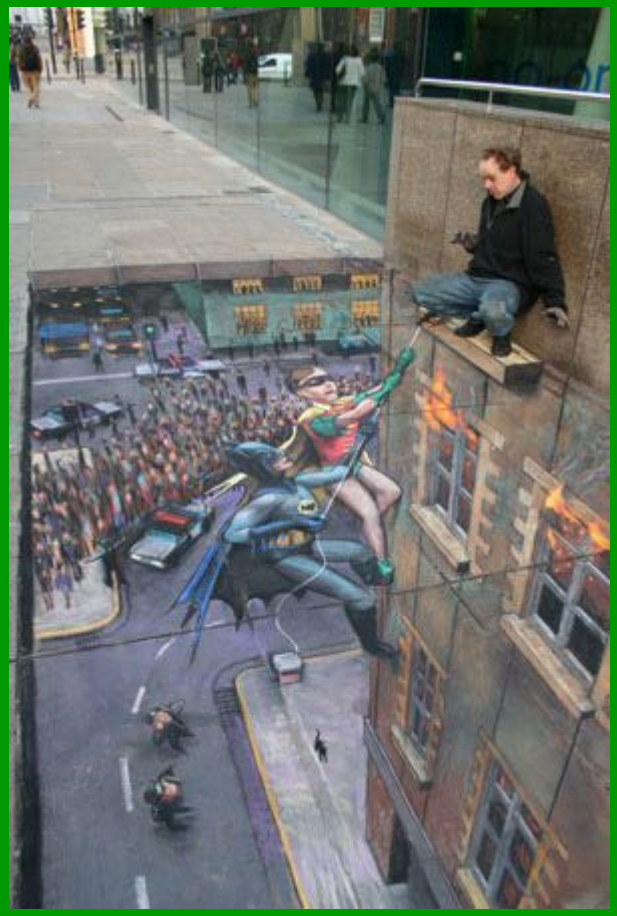


These Two Revealed

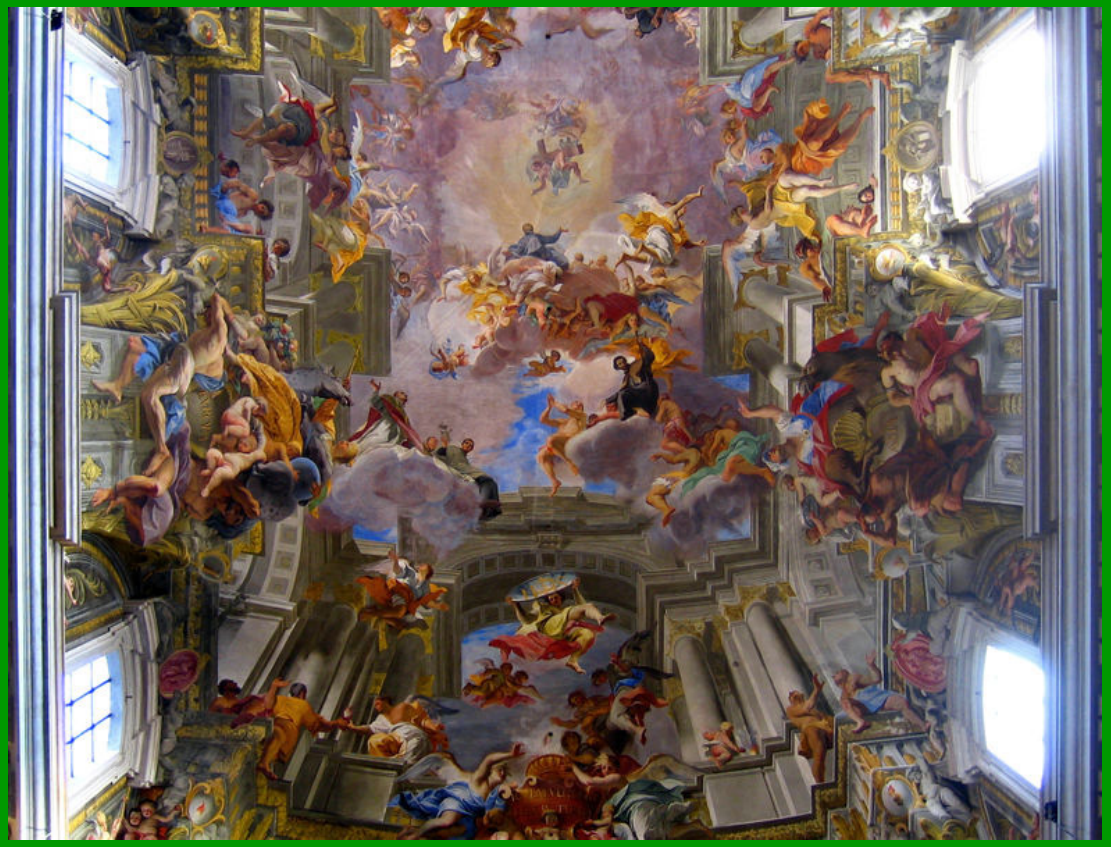
By English Artist Julian Beever

Sequence of Four

Last Three (Good Ones)



An Older One

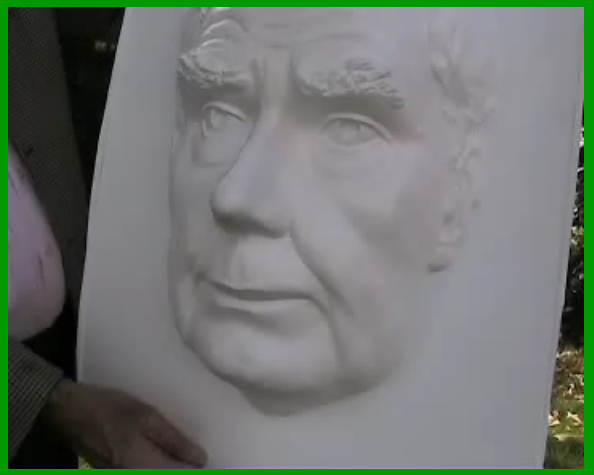


Fresco with the *Apotheosis of St. Ignactius*
Church of Sant'Ignazio, Rome

You don’t have to use paint…

Illusions in 3-D

The Dragon

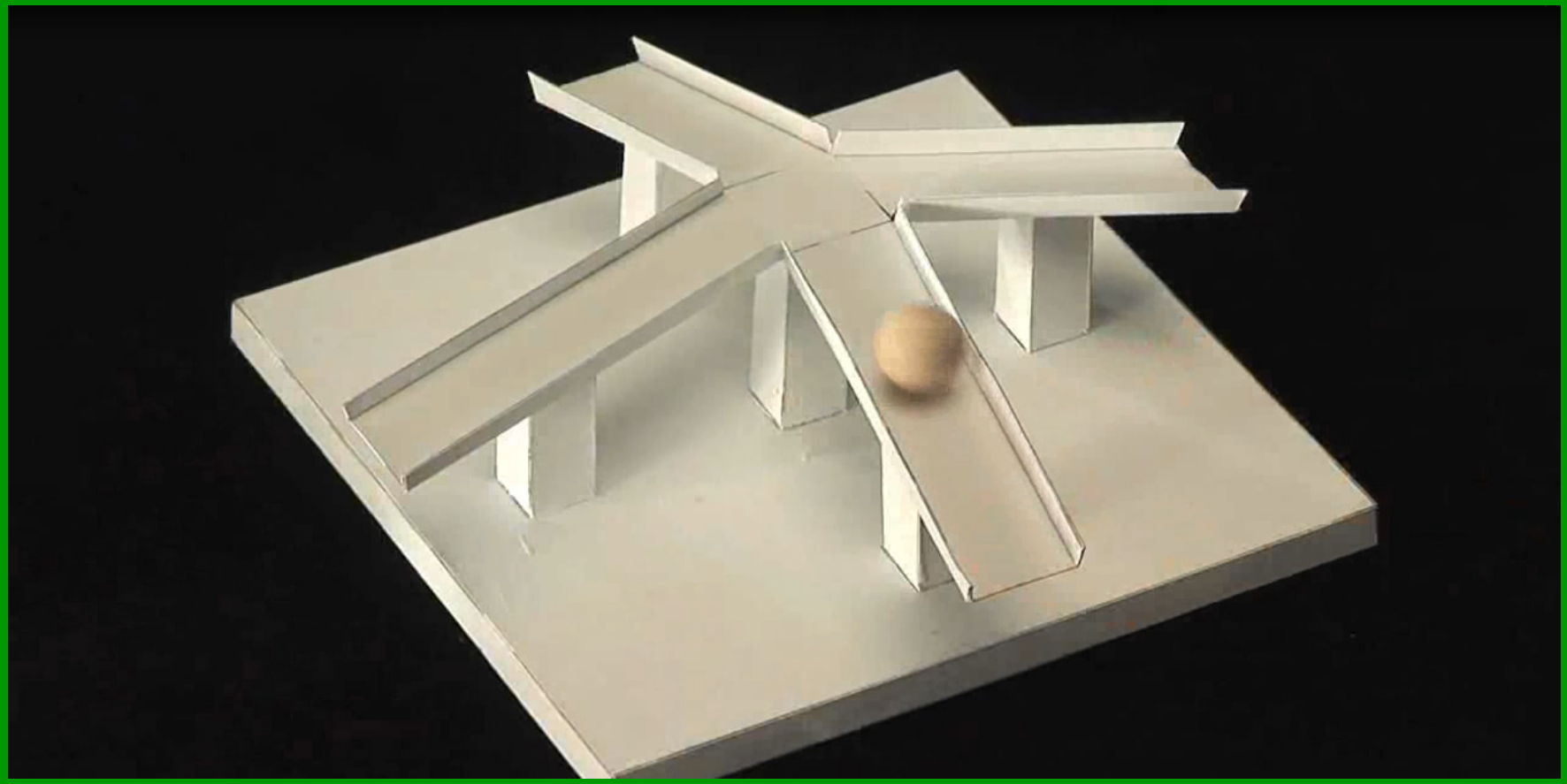


Richard Gregory

3D Sculpture on Park Avenue, NYC by Rafael Barrios

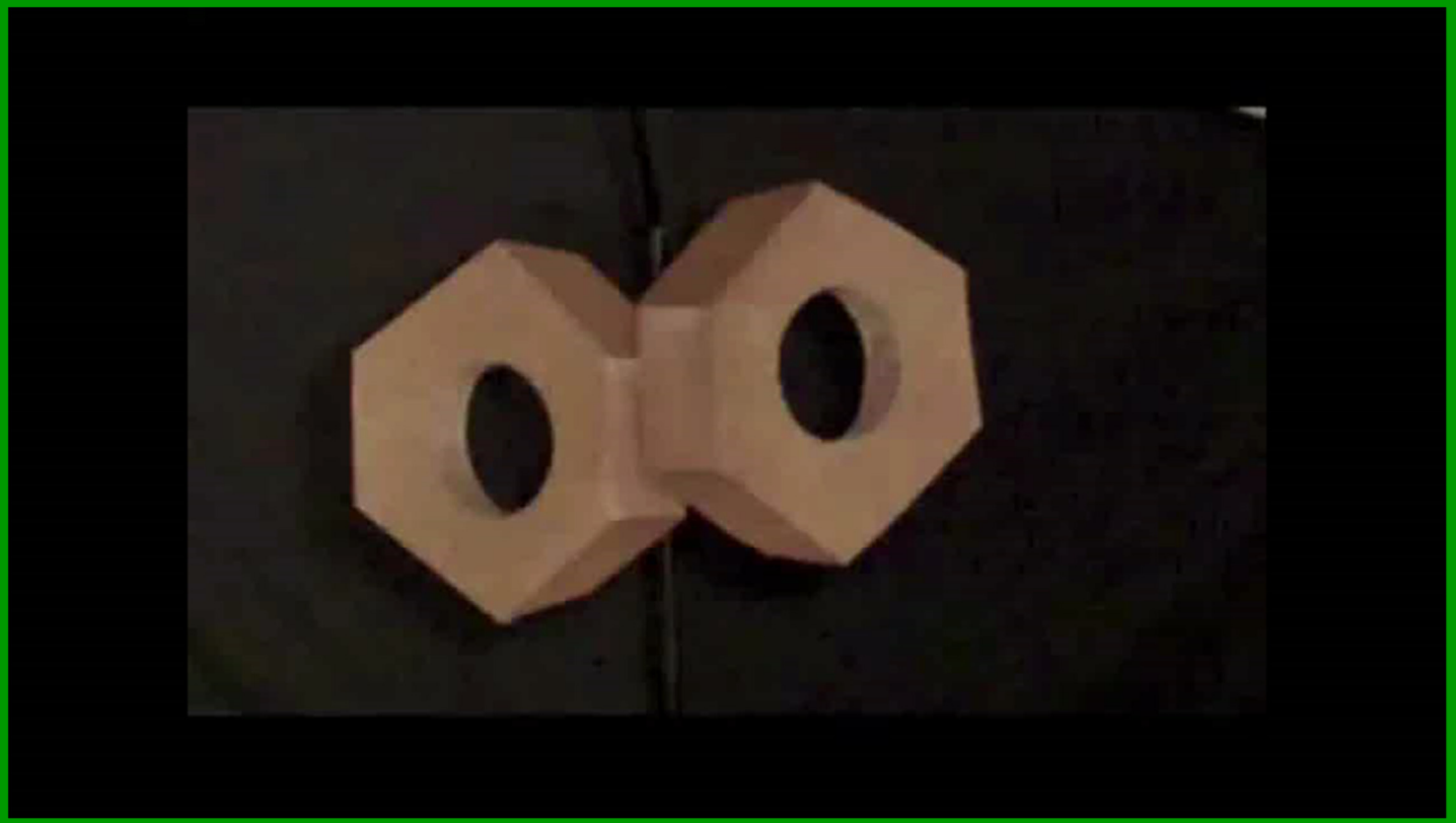
Play it

Perspective Illusion



Play it

“Crazy Nuts”



Created by magician Jerry Andrus

Seeing in 3-D

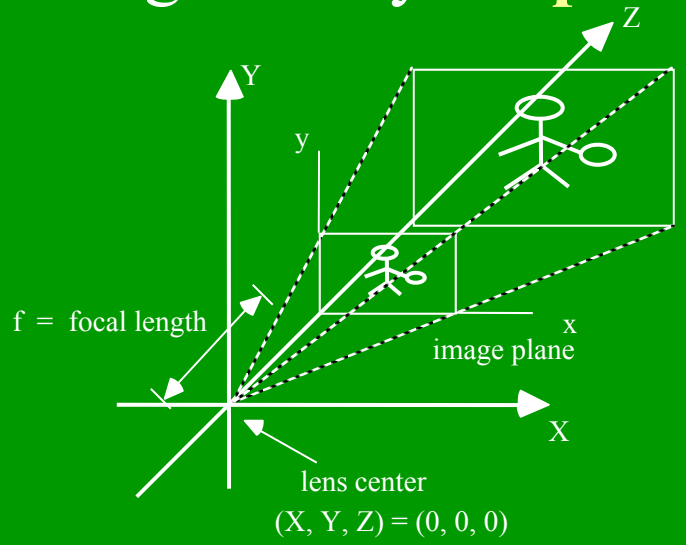
- **Three-dimensional vision** involves many **modes** of perception:
  - Relative size
  - Stereo vision
  - Motion parallax
  - Occlusion
  - Foreshortening, convergence of lines, etc etc etc

- A famous illusions on this theme … the **Ames Room** … the eye-brain can be fooled by 3D cues!

- We will study a main mode: **stereo vision**

Stereo Camera Geometry and Triangulation

- Begin with the geometry for **perspective projection**:



- (X, Y, Z) are points in **3-D space**
- The **origin** (X, Y, Z) = (0, 0, 0) is the **lens center**
- (x, y) denote **2-D image** points
- The x - y plane is parallel to the X - Y plane
- The **optical axis** passes through both origins
- The image plane is one **focal length** f from the lens

Relationship Between 3-D and 2-D Coordinates

- We found that a point (X, Y, Z) in 3-D space projects to a point

$$(x, y) = \frac{f}{Z}(X, Y)$$

  in the 2-D image, f = focal length, = magnification factor.

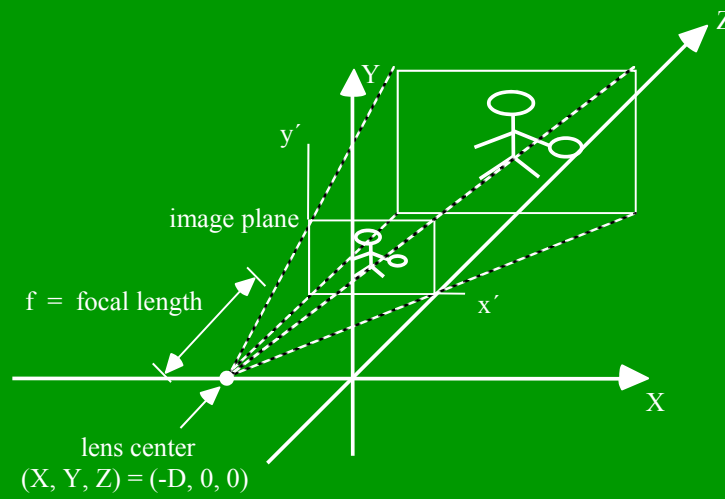
- We will now modify the camera geometry. First, we'll shift the camera to the **left** and **right** along the X-axis (in real space).

- We will then consider the case where we have **two** cameras, equidistant from the origin (X, Y, Z) = (0, 0, 0).

- By **relating** the projections (images) from the two cameras, we will find that it is possible deduce **3-D information** about the objects in the scene.

Shifting the Camera to the Left

- Suppose that we shift the camera (lens center) along the X-axis by an amount -D:



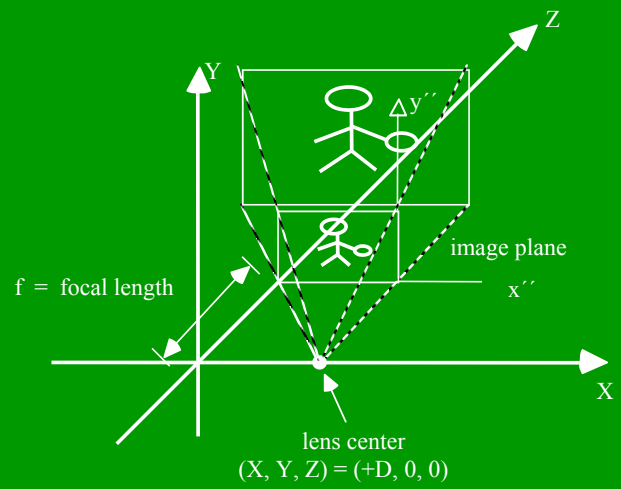
- **Note**: the coordinates of the image are now denoted (x', y').
- Now a point (X, Y, Z) in 3-D space projects to a point

$$(x', y') = \frac{f}{Z}(X+D, Y)$$

in the left-shifted image.

Shifting the Camera to the Right

- Suppose we shift the camera (lens center) along the X-axis by amount +D:



- The coordinates of the image are now denoted (x'', y'').
- Now a point (X, Y, Z) in 3-D space projects to a point

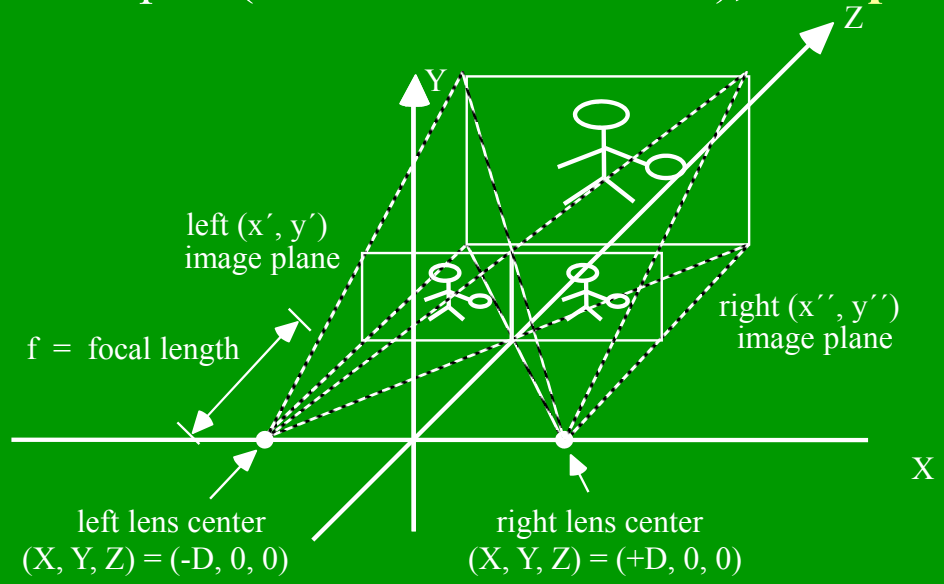
$$(x'', y'') = \frac{f}{Z}(X\text{-}D, Y)$$

in the right-shifted image.

- **Note** that the optical axis is still **parallel** to the Z-axis.

Binocular Camera Geometry

- Now suppose that we place **two cameras** (with the **same focal lengths** f), at a distance 2D apart (the **baseline distance**), with **parallel optical axes:**



- This is a **parallel** or **nonconvergent binocular camera geometry**.

- **Suppose** we are able to identify the **images** of an object feature (a **point**) whose coordinates are (X_0, Y_0, Z_0) in 3-D space.

- We could do this, e.g., **interactively**, by pointing a mouse at the image of the object feature in both of the images and determining the coordinates.

Triangulation

- The projections of the point (X_0, Y_0, Z_0) in the two camera images are

  and
  $$(x_0', y_0') = \frac{f}{Z}(X_0 + D, Y)$$
  $$(x_0'', y_0'') = \frac{f}{Z}(X_0 - D, Y)$$

- The **horizontal** and **vertical disparities** between images of (X_0, Y_0, Z_0) are:
  $$\Delta x_0 = x_0' - x_0'' = (f/Z_0)\,[(X_0 + D) - (X_0 - D)] = 2Df/Z_0$$
  $$\Delta y_0 = y_0' - y_0'' = (f/Z_0)\,(Y_0 - Y_0) = 0$$

- In a non-convergent system, the vertical disparity is **always zero**.

- The **horizontal disparity** is extremely useful. We can solve for Z_0 in terms of it:
  $$Z_0 = \frac{2Df}{\Delta x_0}$$

Ramifications

- The **triangulation equation**

$$Z_0 = \frac{2Df}{\Delta x_0}$$

  implies that we can compute the **distance**, from the baseline (hence to anywhere), to **any point** in the scene that is **visible** to **both** cameras ....

- ... **provided** we can find its horizontal **image coordinates** x_0' and x_0''.

- This follows since the camera **focal length** f is **known,** the **baseline** separation 2D is **known, and** the (X, Y, Z) coordinate system is **known** (it's defined by the camera placement)

- This approach is used in **aerial stereo-photogrammetry** using two cameras a known distance apart (e.g., on the wingtips).

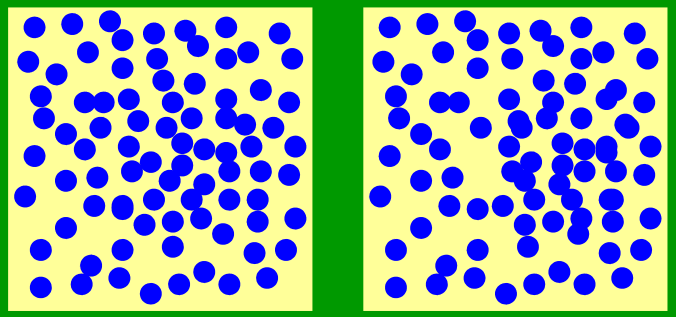
- Historically, a human operator would painstakingly **identify matching points** between the images, measure their coordinates, and compute **depths** via triangulation.

- There is an old device called a **stereoplotter** that was used for this.

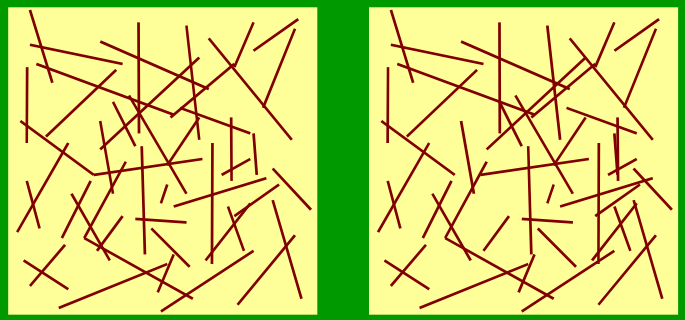
Simple Cell Disparity Responses

- Back to visual cortex….
- A **small percentage** of simple cells appear to be sensitive to **disparities between** the signals coming from the two retinae.
- These may accept input from **simple cell pairs** tuned to **corresponding locations** on the **two retinae.**
- Likely this is used in **3D depth perception.**

Stereopsis is Computational



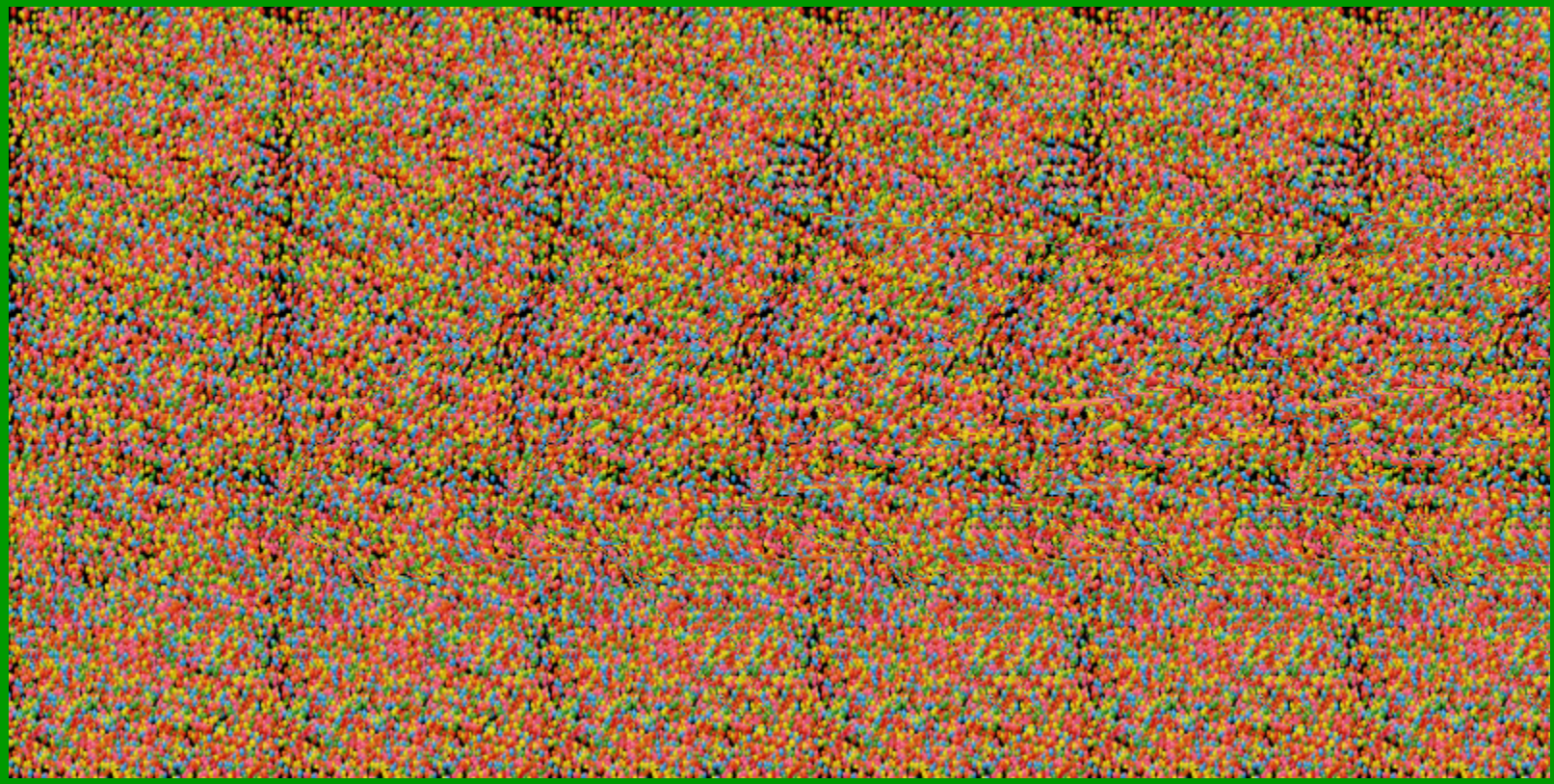
random-dot stereogram



random-line stereogram

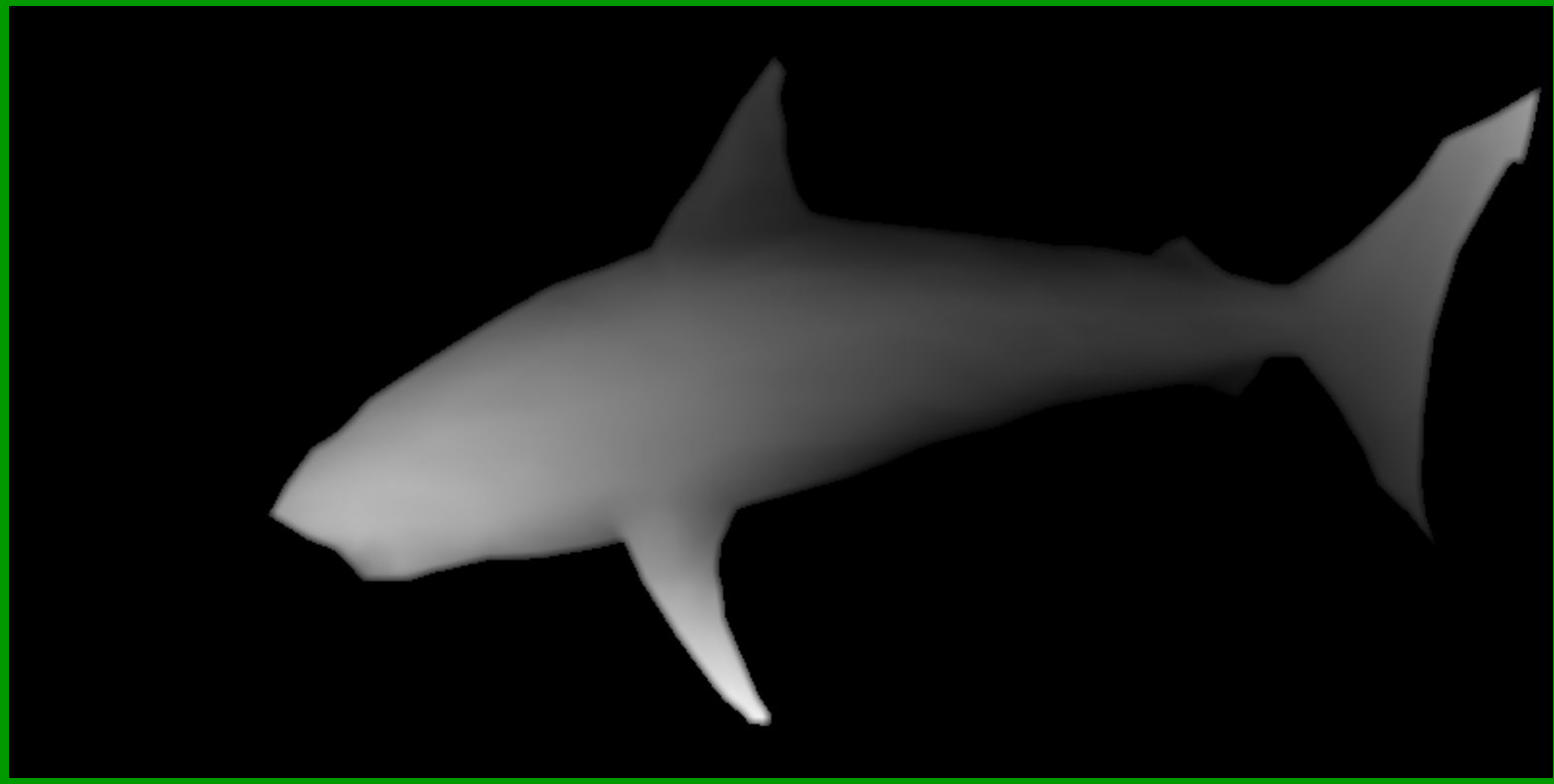
- As proved by B. Julesz' random dot and line stereograms.
- **A similar visual aid … click here.**

An Autostereogram

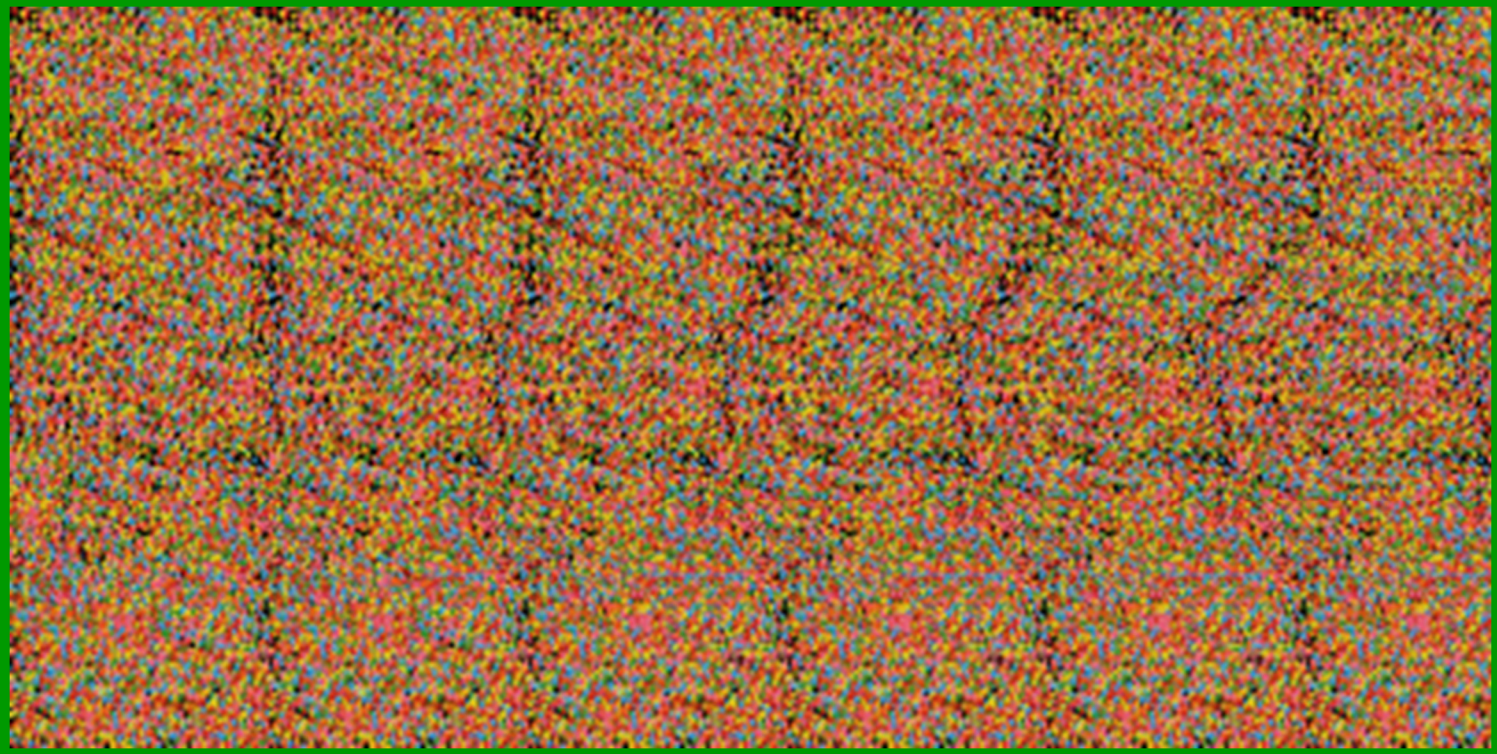


Remember those “Magic Eye” books?

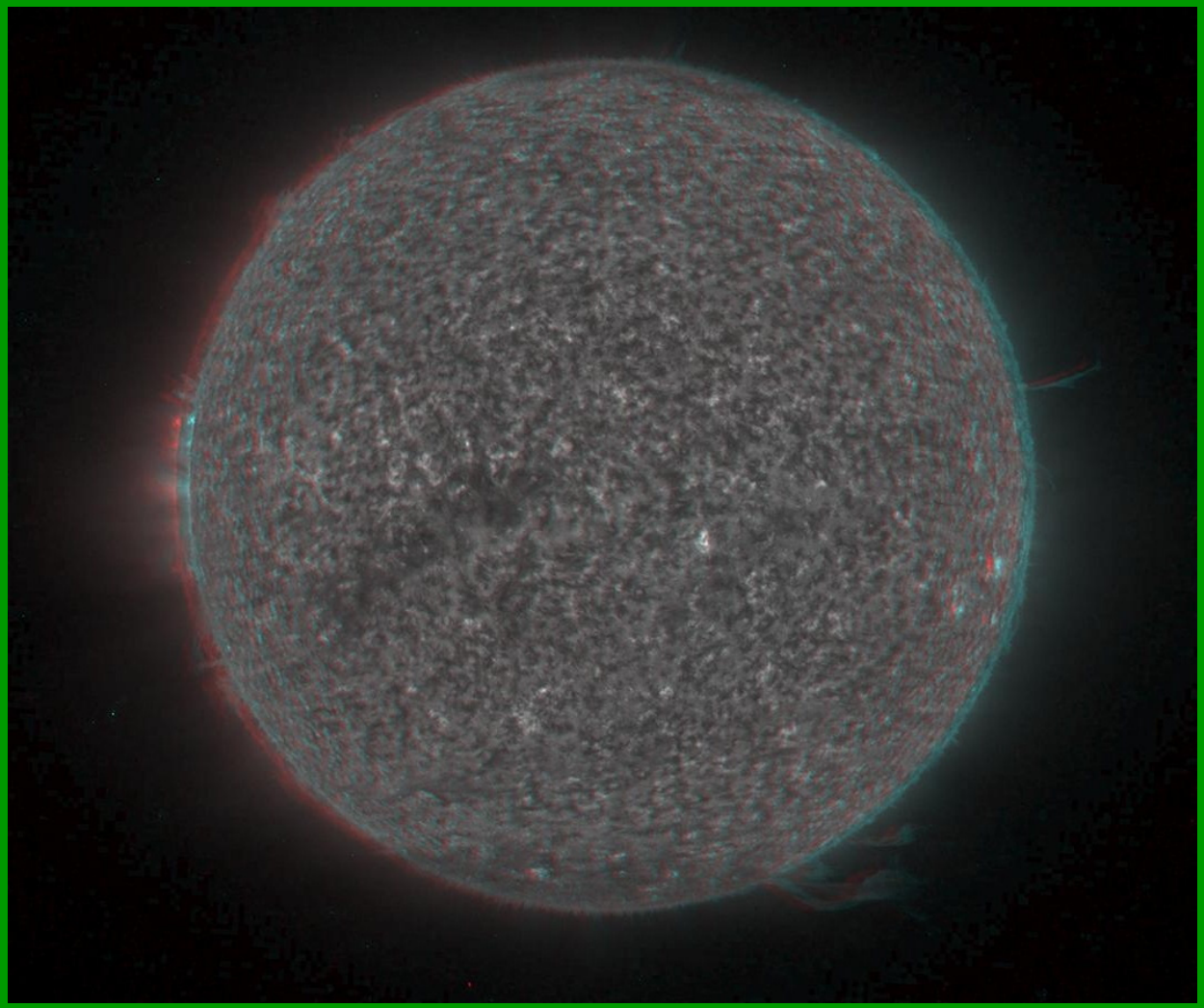
The Depth Map



In Motion!



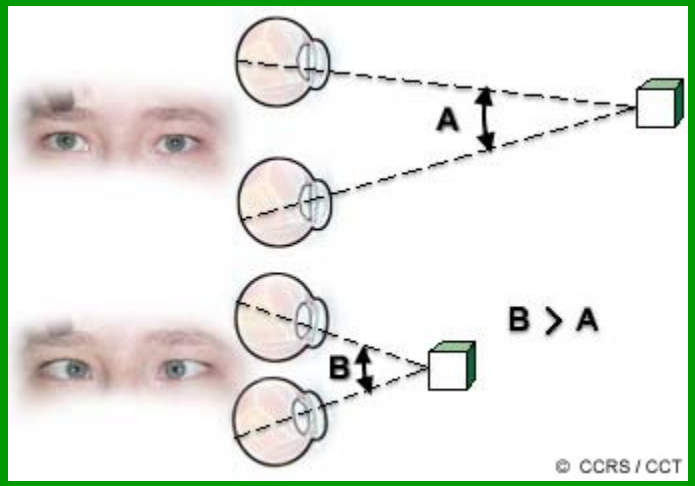
Red-Green Anaglyph



First 3D Image of the Sun

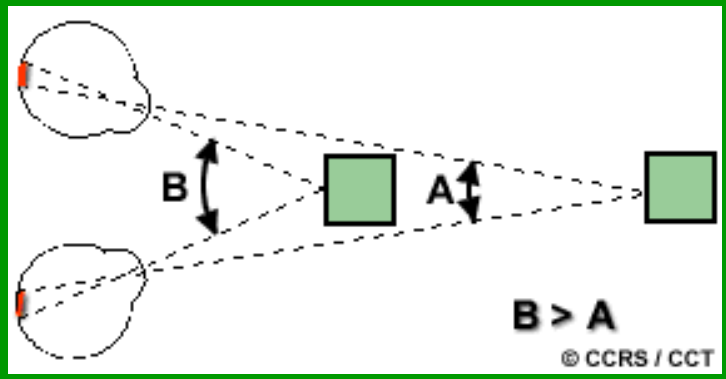
Stereo (Binocular) Geometry

- **Vergence angle** depends on object distance.

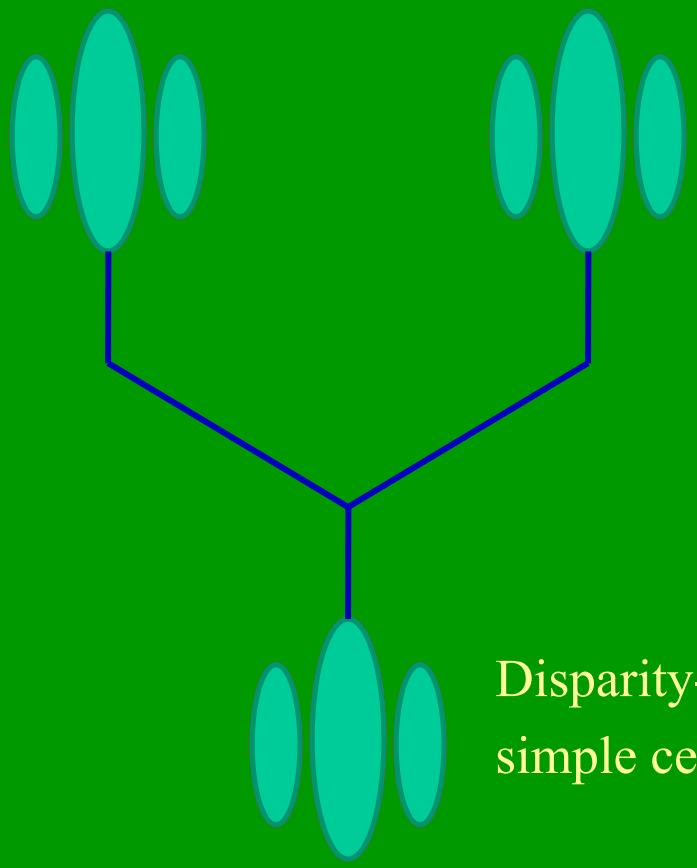


- Shown are **angular** and **positional disparities.**

- The simple cells appear to respond to the **positional disparities.**



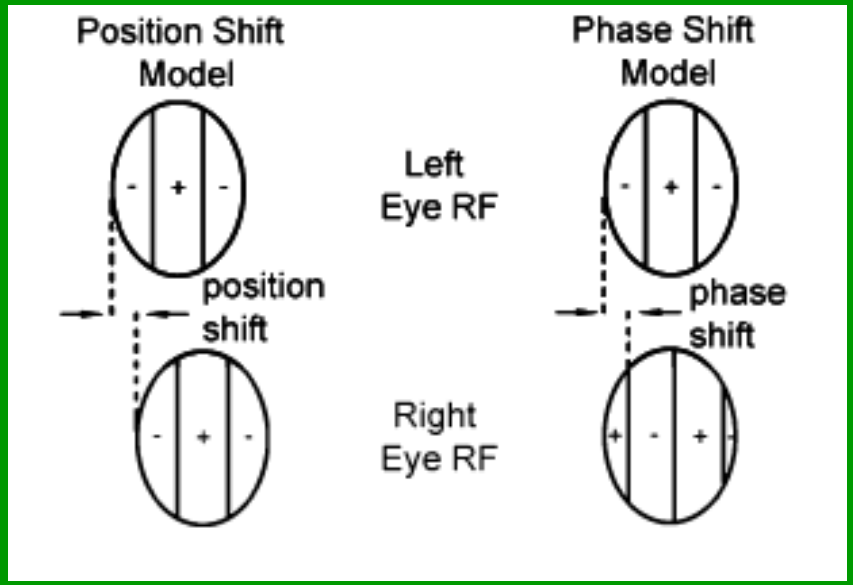
Disparity-Sensitive Simple Cells



Spatial-only
simple cells

Disparity-sensitive
simple cells

Position and Phase Shift Models



- In the **position shift model,** the responsive simple cells are relatively displaced.
- In the **phase shift model,** the stimulus is shifted but the simple cells are not.
- Both indicate a positional disparity **or input phase shift.**

Phase-Based Stereo

- Recall 2D Gabor filter model:

$$g(\mathbf{x}) = Ke^{-\left[\left(x/\lambda\right)^2 + y^2\right]/2\sigma^2} e^{2\pi\sqrt{-1}(u_0 x + v_0 y)}$$

- Find filterbank responses

$$t_i(\mathbf{x}) = I(\mathbf{x}) * g_i(\mathbf{x});\ i = 1, \ldots, K$$

- Largest magnitude response

$$i^* = \arg\max_i |t_i(\mathbf{x})|$$

$$t^*(\mathbf{x}) = t_{i^*}(\mathbf{x})$$

Left and Right Responses

- Find largest response for **both** left and right camera images:

$$t_L^*(\mathbf{x}) \approx A \cdot \tilde{G}\left[\nabla\phi(\mathbf{x}_L)\right] \cdot \exp\left[\sqrt{-1}\phi_L(\mathbf{x}_L)\right]$$

$$t_R^*(\mathbf{x}) \approx A \cdot \tilde{G}\left[\nabla\phi(\mathbf{x}_R)\right] \cdot \exp\left[\sqrt{-1}\phi_R(\mathbf{x}_R)\right]$$

- Demodulate to obtain $\phi_L(\mathbf{x}_L)$ and $\phi_R(\mathbf{x}_R)$

- Stereo assumption

$$\phi_L(\mathbf{x}_L) - \phi_R[\mathbf{x}_L - \Delta\mathbf{x}(\mathbf{x}_L)] = 0$$

Left and Right Responses

- Here is (truncated) 2-D Taylor's formula:

$$f(x_0 - \Delta x, y_0 - \Delta y) = f(x_0, y_0) - f_x(x_0, y_0)\Delta x - f_y(x_0, y_0)\Delta y$$

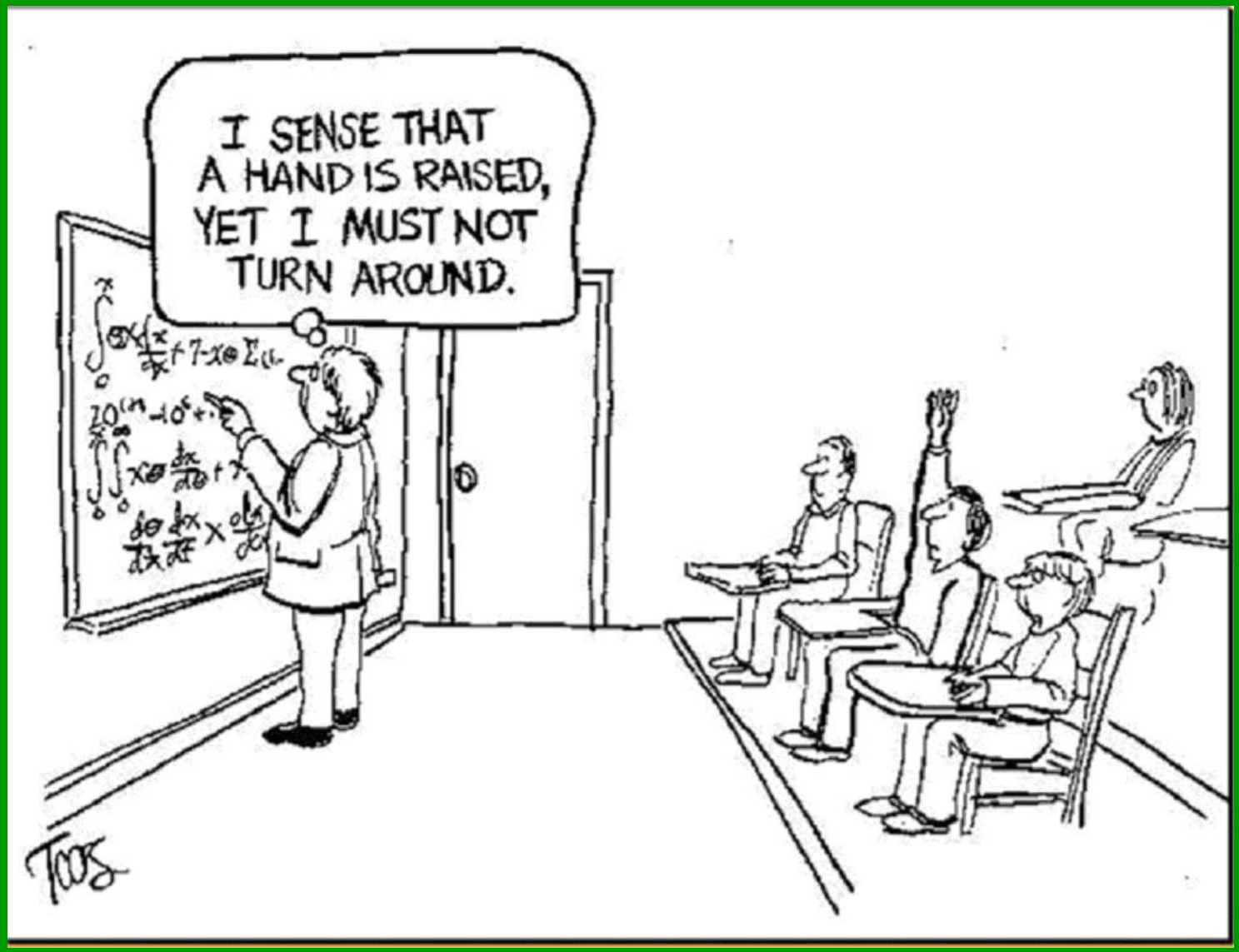
or

$$f(\mathbf{x}_0 - \Delta \mathbf{x}) = f(\mathbf{x}_0) - \Delta x^T \nabla f(\mathbf{x}_0)$$

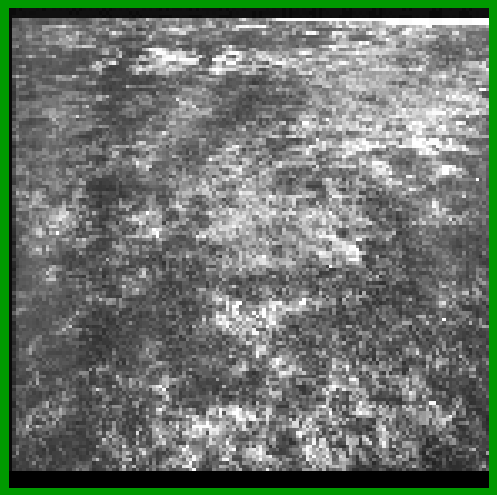
- Taylor Approximation [Note: $\Delta \mathbf{x}(\mathbf{x}_L) = (\Delta x(\mathbf{x}_L), 0)$] :

$$\phi_L(\mathbf{x}_L) - \phi_R(\mathbf{x}_L) \approx \Delta \mathbf{x}(\mathbf{x}_L)^T \nabla \phi_R(\mathbf{x}_L)$$

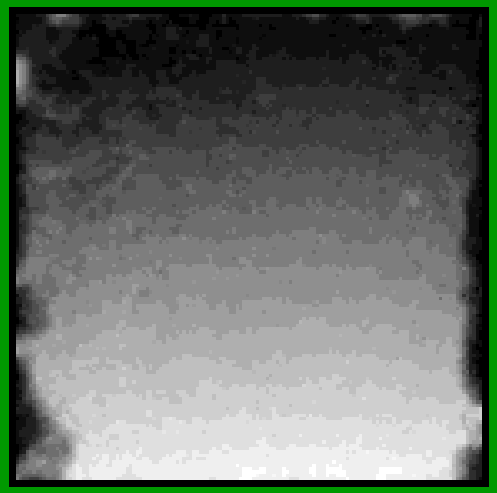
Algorithm! $\longrightarrow$ $$\Delta x(\mathbf{x}_L)^T \approx \frac{\phi_L(\mathbf{x}_L) - \phi_R(\mathbf{x}_L)}{\phi'_R(\mathbf{x}_L)}$$


I SENSE THAT A HAND IS RAISED, YET I MUST NOT TURN AROUND.
TOOS

Phase-Based Stereo Example



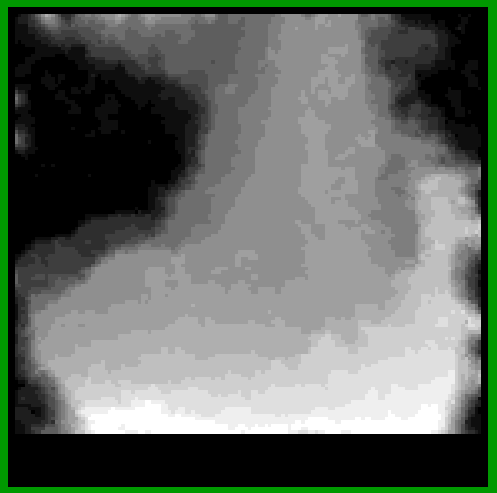
"Turf"



Brighter = nearer

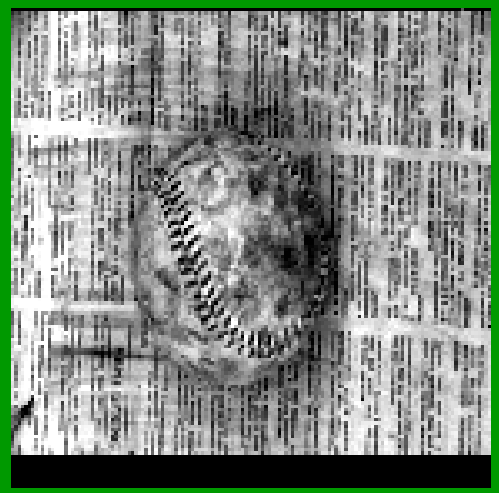
Phase-Based Stereo Example

"Tree trunk"

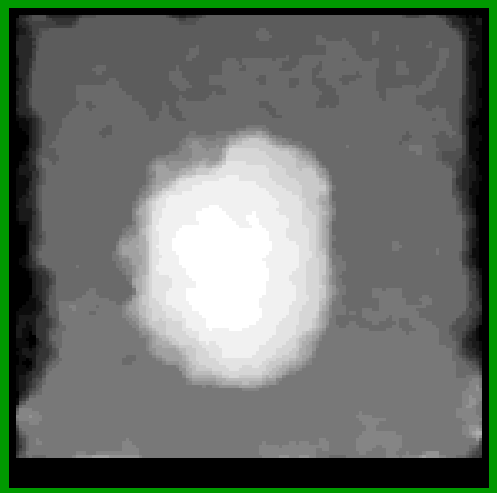


Brighter = nearer

Phase-Based Stereo Example

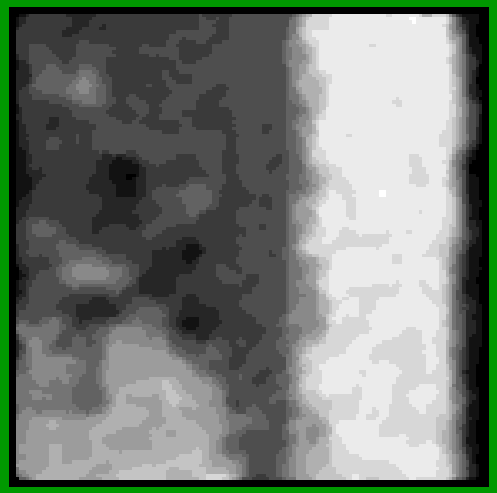


"baseball"



Brighter = nearer

Phase-Based Stereo Example



"Doorway"

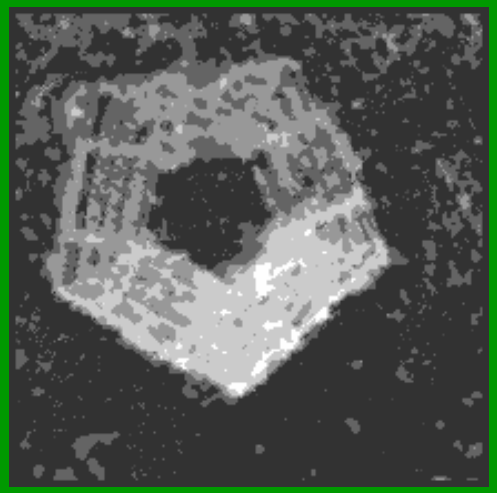
Brighter = nearer

Phase-Based Stereo Example

"Stone"

Brighter = nearer

Phase-Based Stereo Example



"Pentagon"

Brighter = nearer

The Future: Active, Vergent, Fixating Stereo Systems

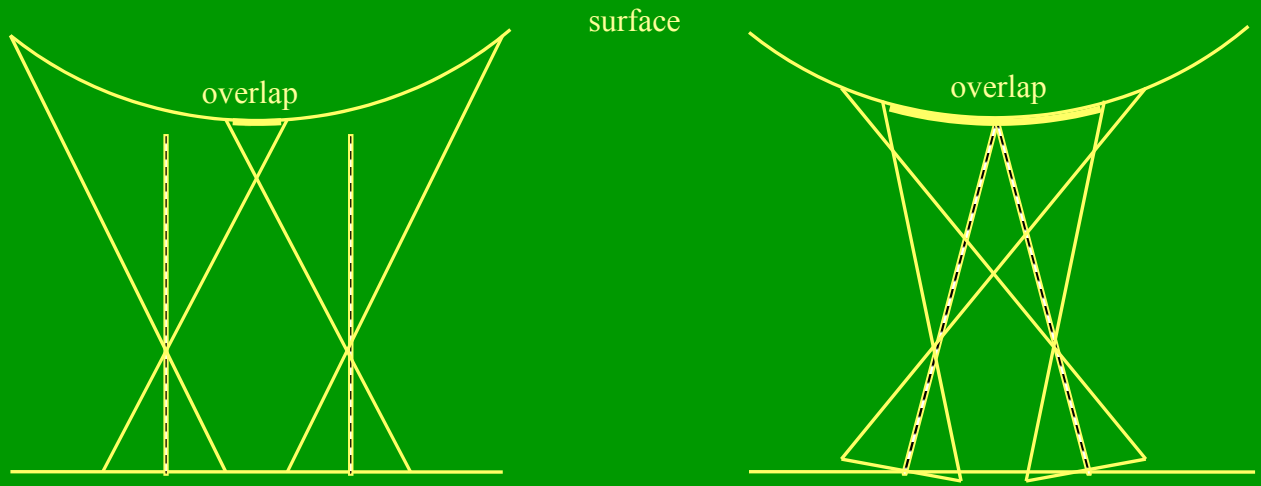
- **More sophisticated** stereo vision systems are attempting to incorporate more degrees of freedom:

vergence, **foveation**, and **camera motion**

Vergent Stereo

Vergent Stereo Systems

- **Vergent stereo systems** have non-parallel optical axes.



parallel vergent

Vergent Stereo Systems

Advantages

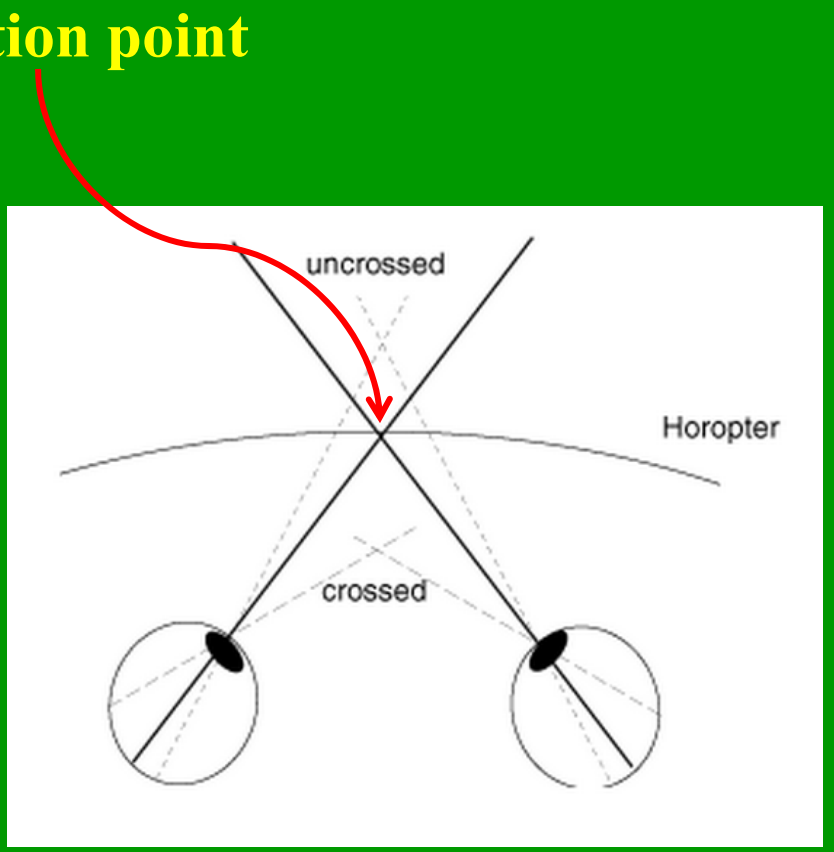
- **More overlap** between views
- **Smaller disparities**, especially near image center
- **Concentrated attention** at a point of interest

Disadvantages

- Epipolar assumption **cannot hold**
- More **complex computation**

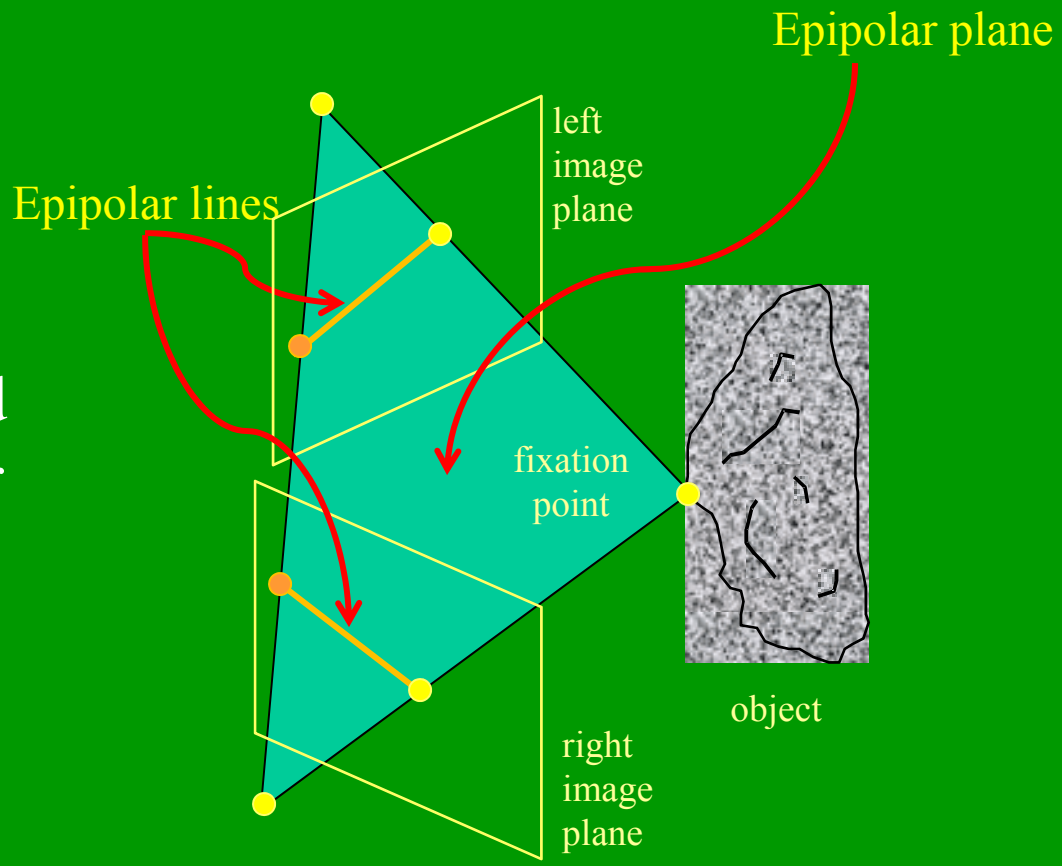
Horopter & Crossed/Uncrossed Disparity

- 3D points that produce **crossed disparity** lie in front of the fixation point.
- 3D points that produce **uncrossed disparity** lie behind the fixation point.
- The **horopter** is the locus of 3D points at the **same distance** as the fixation point.



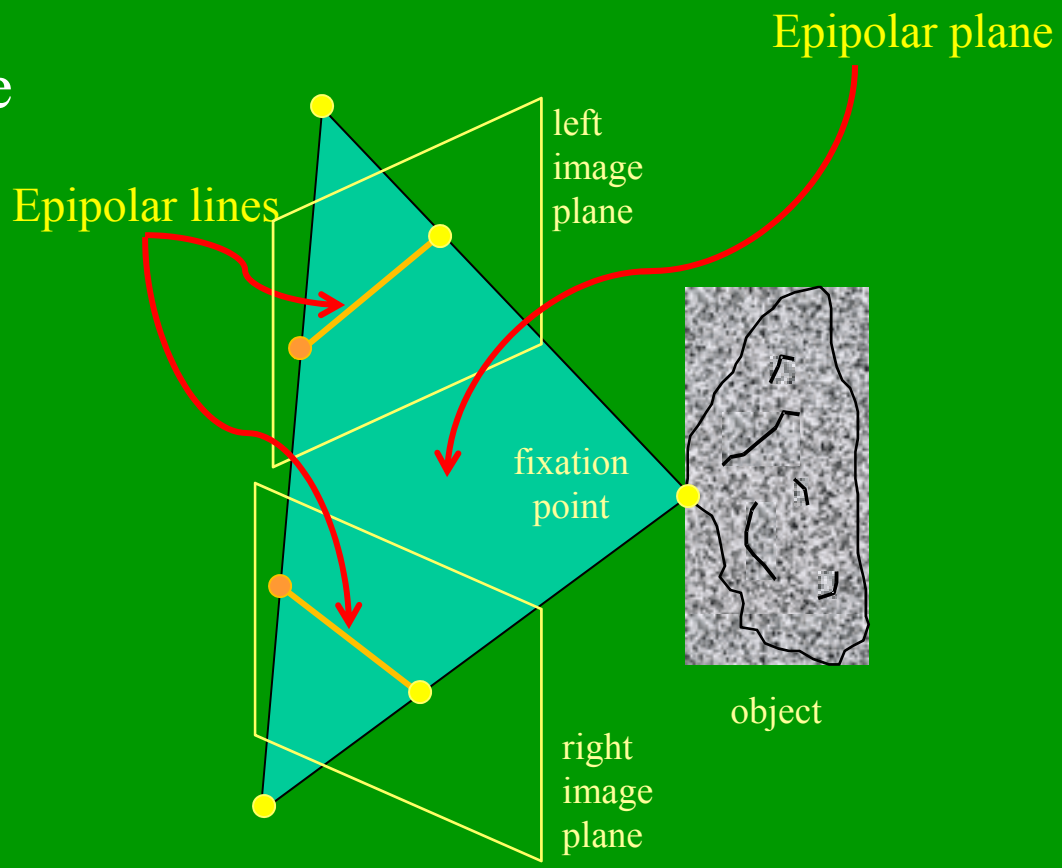
Epipolar Plane

- The **epipolar plane** is formed by a 3D point and its two projections.
- **Epipolar lines** are formed by intersection of epipolar planes with image planes.
- The **horopter** is the locus of 3D points at the **same distance** as the fixation point.



Epipolar Matching

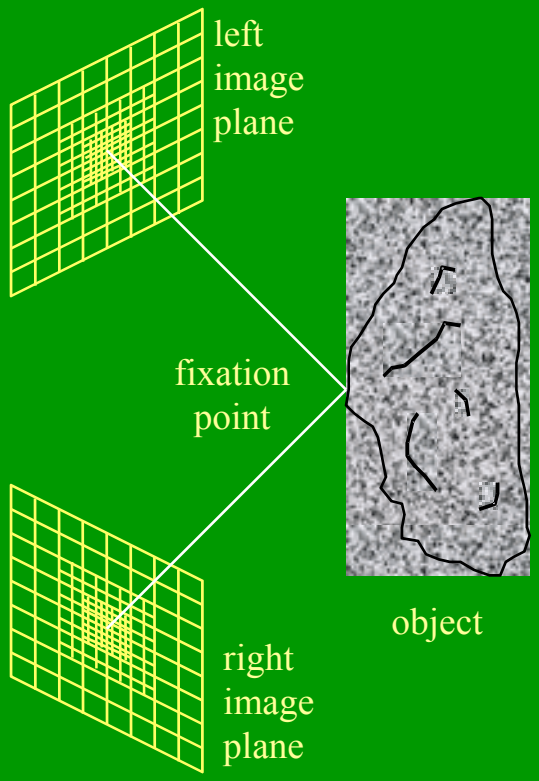
- The epipolar lines in the **non-vergent** geometry are **parallel** with each other.
- In the **vergent** case, they are **not parallel**.
- The **matching** problem is still done **along the non-parallel** epipolar lines.
- This is a bit more involved than we will get into.



Foveated Stereo

Foveated Stereo Systems

- A **foveated** imaging system uses a nonuniform sampling of image intensities - **more samples near image center** (the **fovea**) and fewer near periphery.



Foveated Stereo Systems

Advantages

- **Much less storage and much less computation**
- **Concentrated computation** near image center
- **Reduced importance of loss of epipolar assumption**

Disadvantages

- Loss of **peripheral resolution**
- **Must decide** where to **foveate/fixate/verge**
- **Must move cameras** to new fixations

Active Stereo Vision

- **Active** stereo vision systems use **camera motion** to **fixate**.
- **Deciding fixations** is application-specific and an open problem!
- **Foveation** is necessary to operate in anywhere near real time.

FOVEA

- **FOVEA** is (was) a mobile computer-controlled stereo vision system with several degrees of freedom: foveation, tilt, pan, rotation, baseline distance (and also focus and aperture).

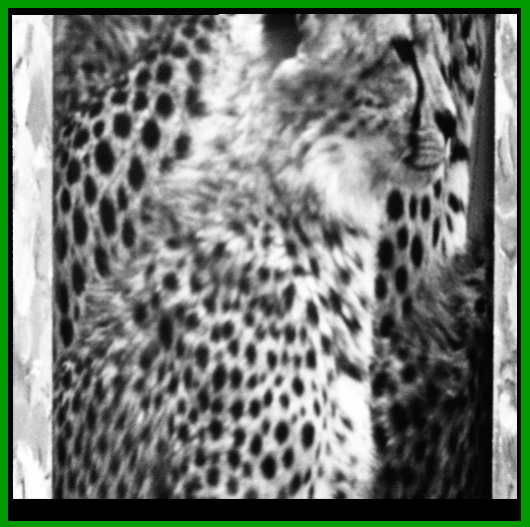
- **FOVEA's fixation points** are selected semi-randomly:
  - Edge points visited with high probability
  - Next fixation is neither very close nor very far in 2-D
  - Next fixation stays on 3-D surfaces with high probability

FOVEA's Mobile Camera System

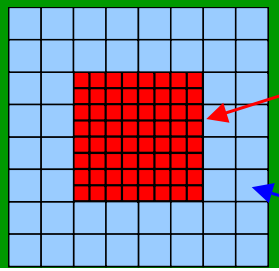
FOVEA Example

- FOVEA will "examine" this object:

A cylinder with a photo pasted on



One camera view



FOVEA IN ACTION

Stereo Would be Useful Here

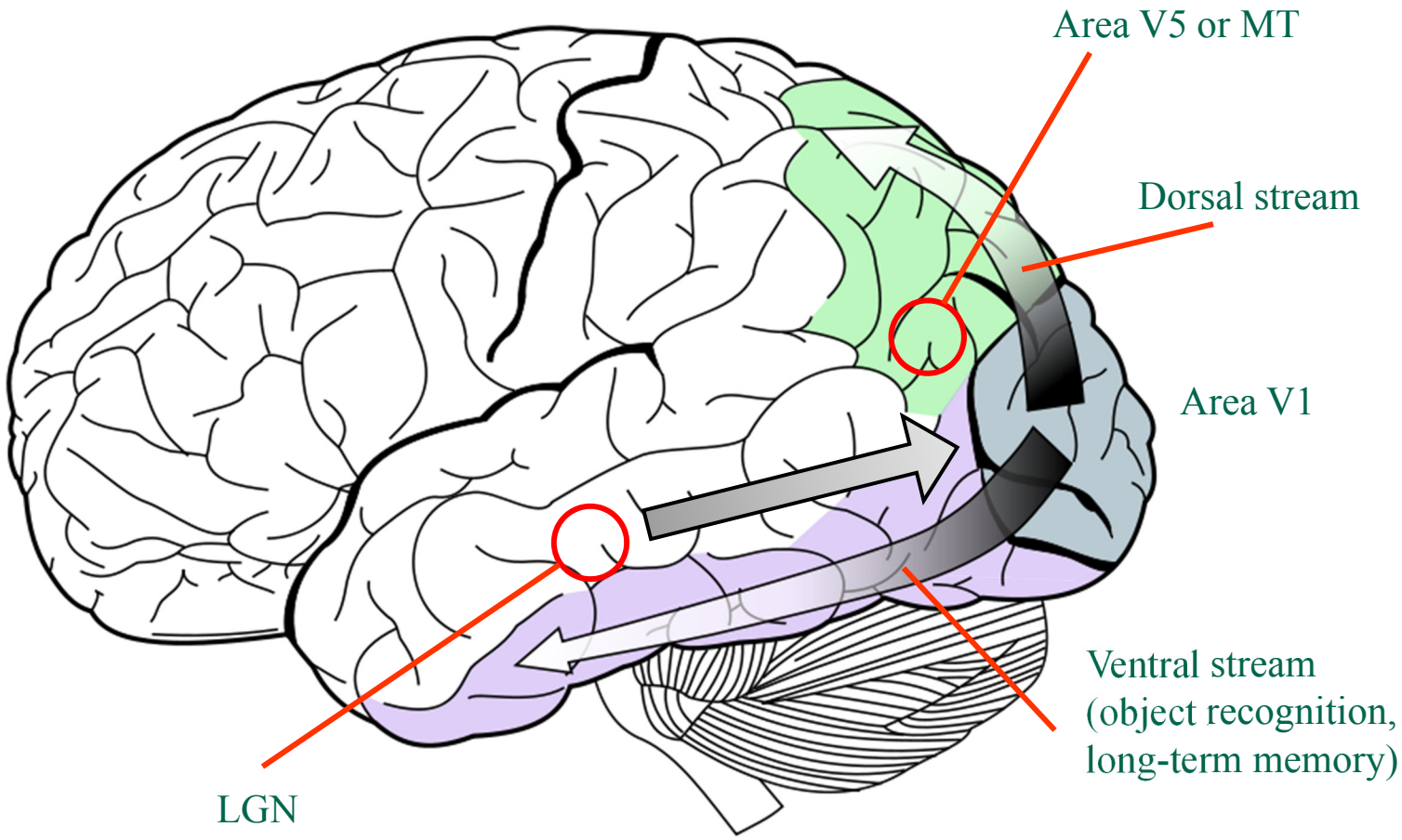
Teeny tiny lady in a chair

Here's Why…

Hmmm….

Object Recognition

Flow of Visual Data



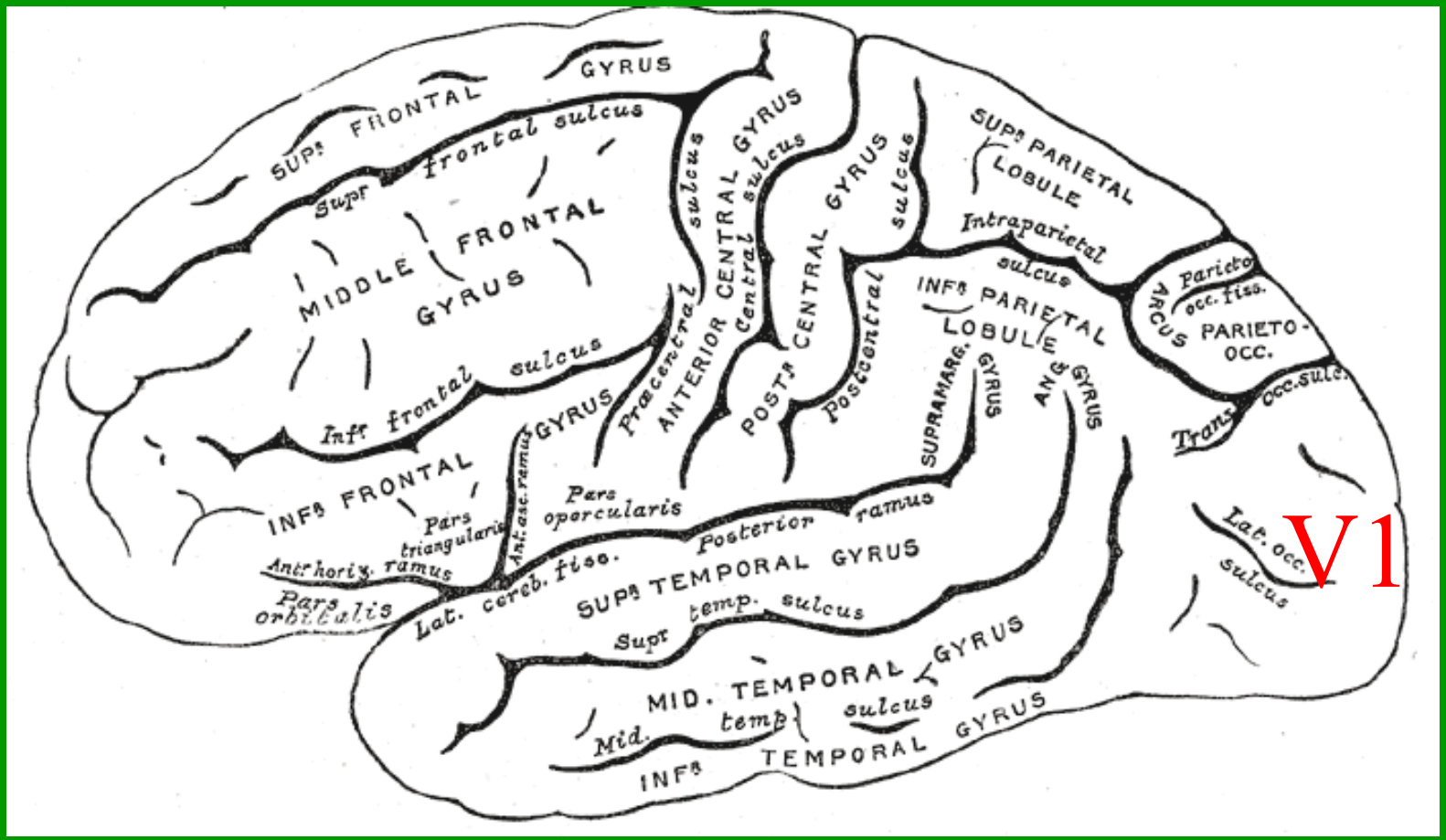
Ventral Stream

- The **ventral stream** (also called the **"what pathway")** begins with Area V1, and lands in the **inferior temporal gyrus**.

- Much **shared processing** occurs between V1, V2 (on the way to ITG), and the ITG.

- Areas devoted to **object recognition, memory,** and **closer** to **consciousness.**

Inferior Temporal Gyrus



Also called Inferior Temporal (IT) Cortex

Inferior Temporal Gyrus

(Fusiform Gyrus)

- Has available **entire retinotopic map** from Area V1.
- **Considerable feedback and forth from V1**
- Devoted to **visual memory** and **visual object recognition.**
- **Not much known** about "how" it's done.
- **Computational algorithms** for **object recognition** using V1 primitive models **have proved effective.**

Biometrics: Recognition of Iris

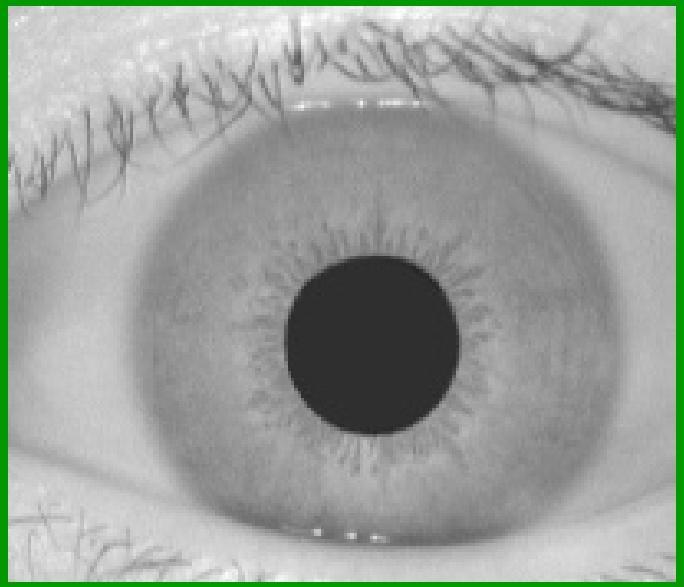
Daugman's Iris Recognition System

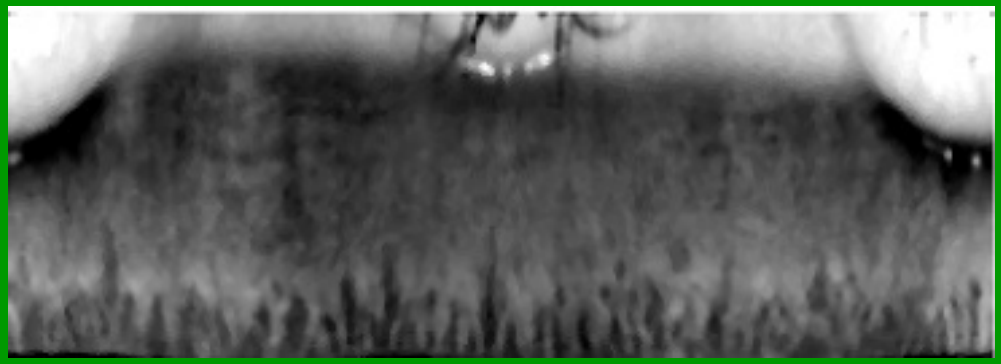
- One of the earliest approaches to recognition using V1 primitives was J. Daugman's **iris recognition system.**

- Uses spatial V1 (**Gabor**) models responses to perform important **biometric application: recognition** of the **iris** of the **eye**.

- Far-seeing invention is still the **industry standard** of **performance.**

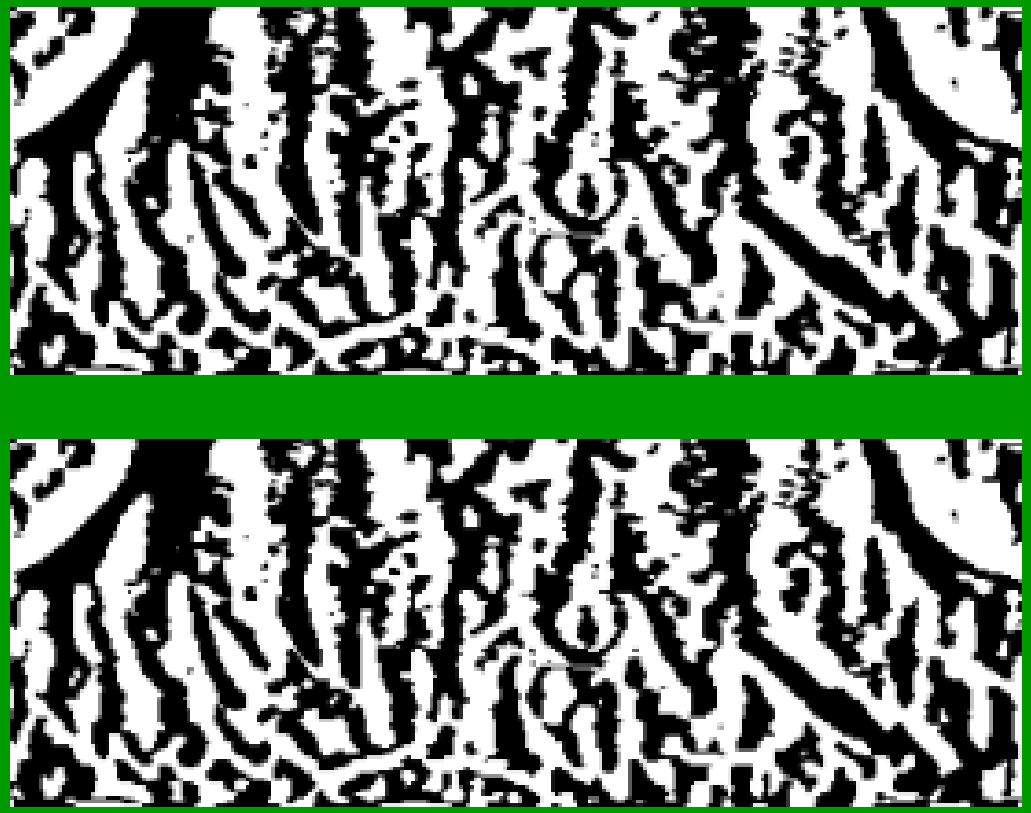
Iris Recognition





(1) Detect and extract iris and “unroll it”

Iris Recognition



(2) Filter each row with 1-D cosine (top) and sine Gabor filters, and binarize result.

Iris Recognition

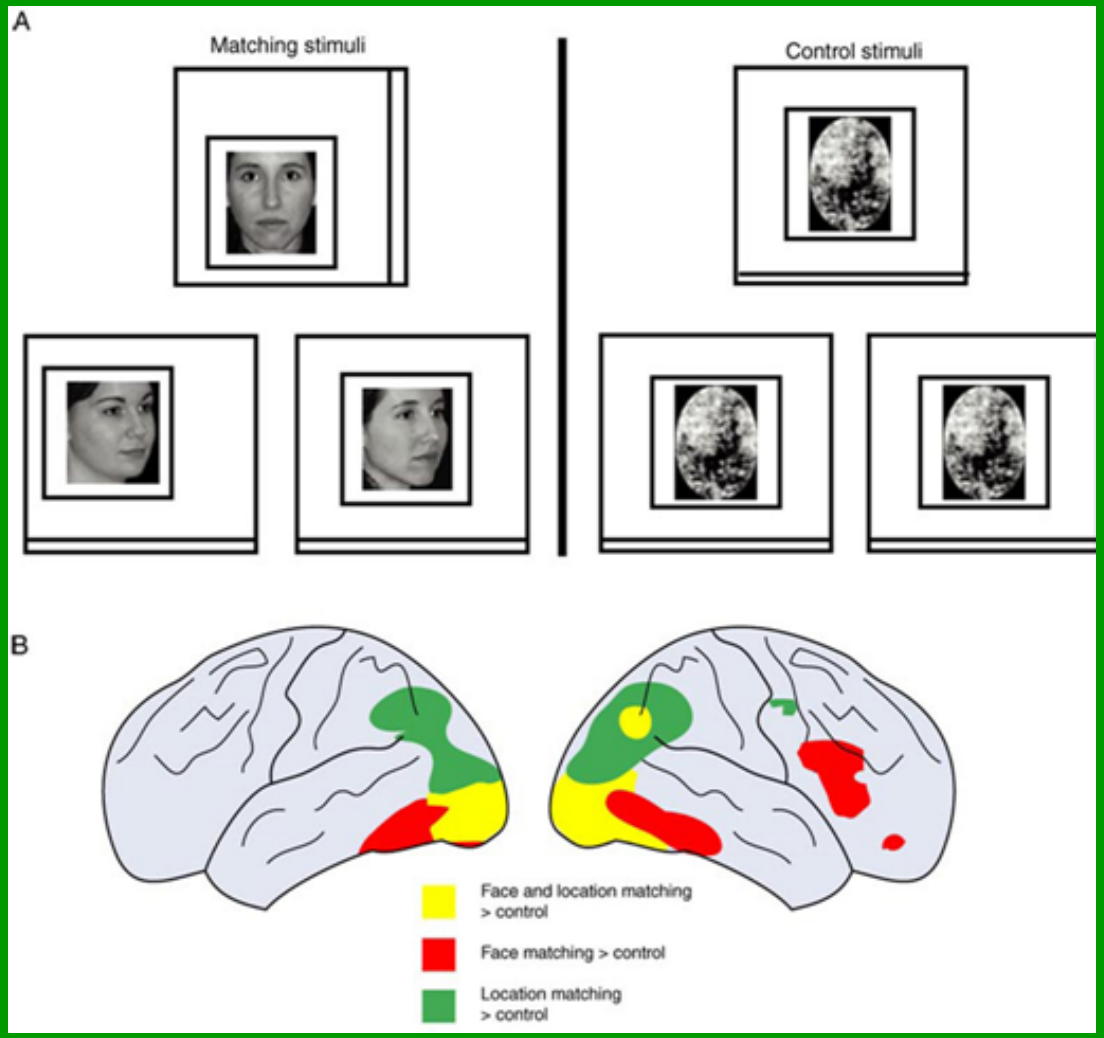
$$H(A,B)=\frac{\text{cardinality}(A \text{ XOR } B)}{\text{cardinality}(A)}<\tau$$

(3) Compare binary rows with binary **reference** of iris stored in database using **Hamming Distance**.

- **If A = B, then H(A, B) = 0.**
- **Random chance, H(A, B) = 0.5.**
- **For τ = 0.3, typical false positive rate is < 10^{-7}.**

Biometrics: Recognition of Face

Face Recognition



The **inferior temporal gyrus** appears to be heavily implicated in **face recognition**.

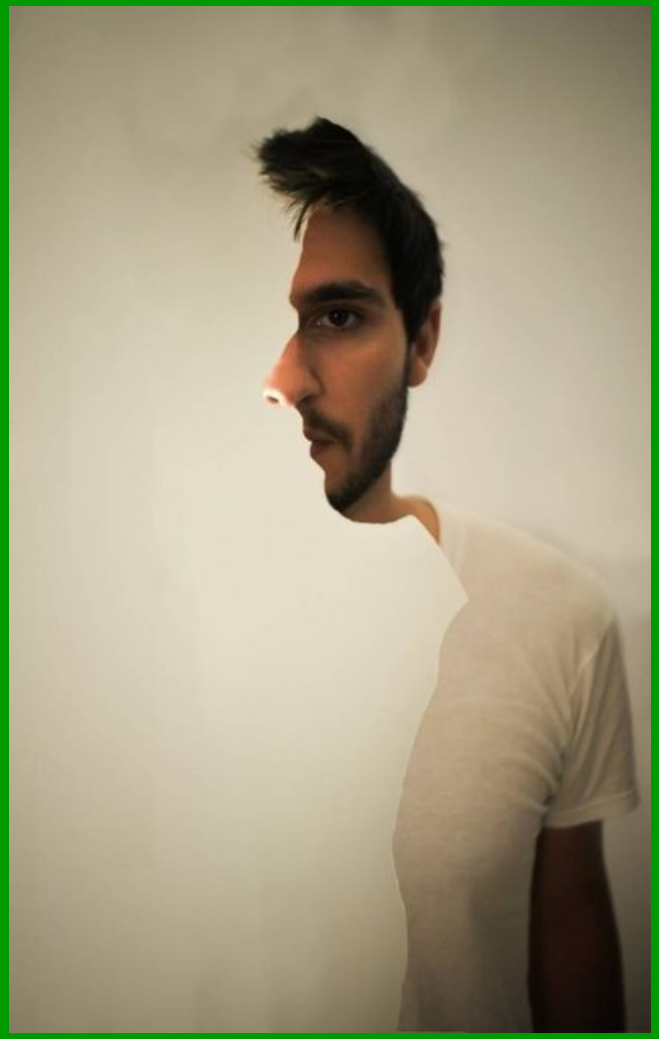
We **don't know** how the brain does it.

There is lots of **interplay** between the various visual areas of the brain.

Visual memory is heavily involved.

Very nice computational methods have been devised using **V1 primitive models**.

Could You Recognize Him?



Which Way is He Looking?

The "Iron Lady"

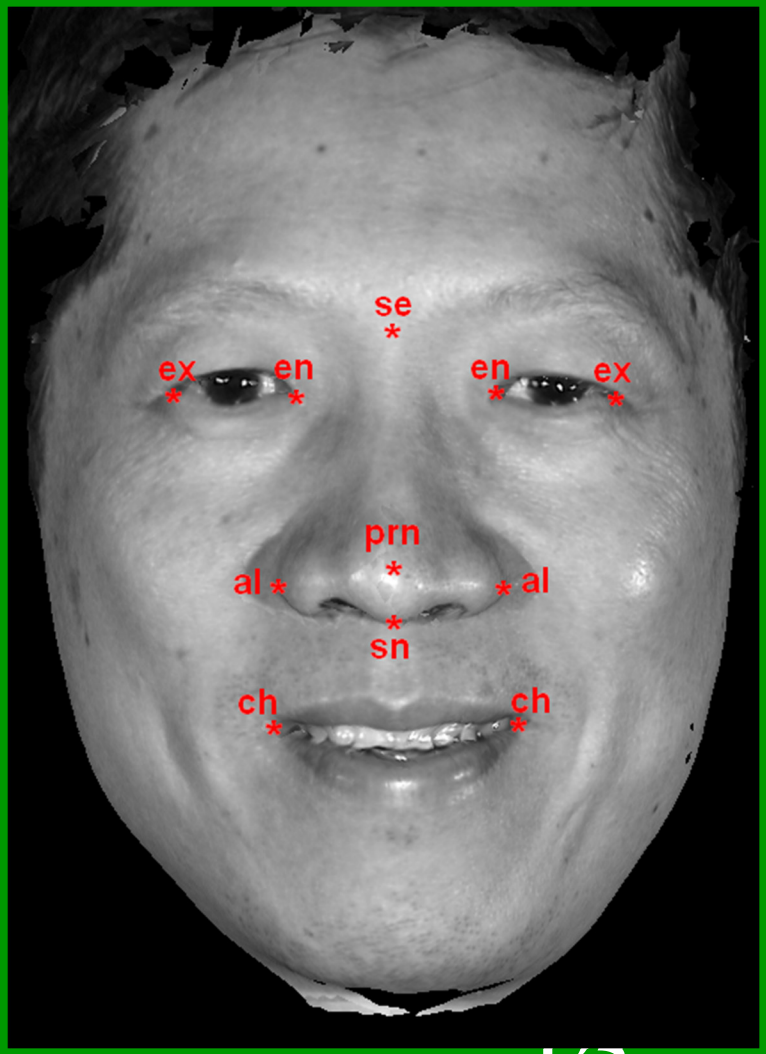
What's This Doing to My Brain?

Computed Face Recognition

- **Commercial opportunities** greatly increased interest in personal identification systems :
    - **Access Control**
    - **Surveillance and Security**
    - **Airport Screening**
- Other **biometrics** (**iris scans, fingerprints**) are very accurate but faces are:
    - **More accessible**
    - **Minimum cooperation**

Facial Features

- **Facial features**:
  - Inner and outer **eye corners** (en,ex)
  - **Mouth corners** (ch: Chelion)
  - **Nose lateral points** (al:Alare)
  - **Nose tip** (prn: pronasale)
  - **Nose bridge** (se:Sellion)
  - **Lowest** point of the nose (**sn:subnasale**)
  - Many more ….



Face Recognition By Gabor Features

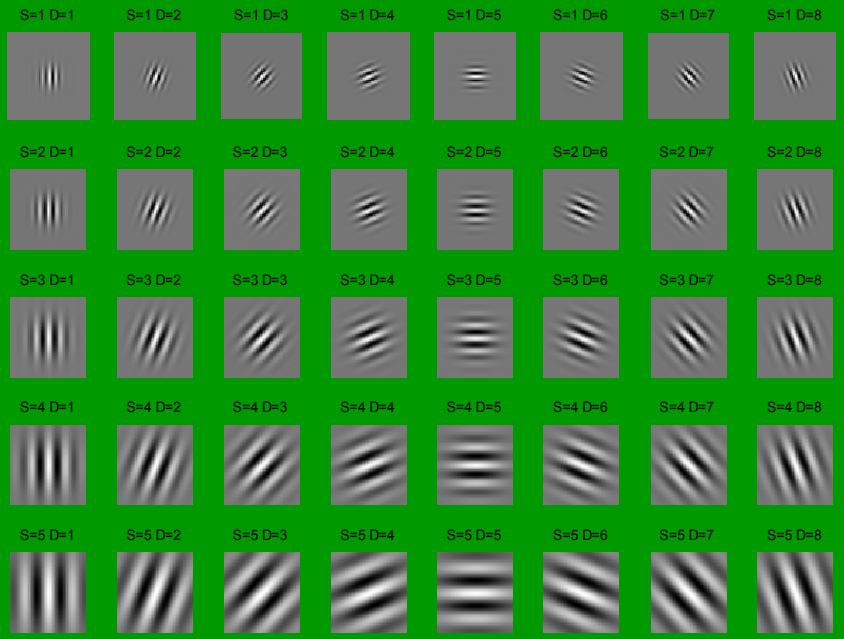
- Gabor responses encode appearance of 2D (**Portrait**) and/or 3D (**Range**) facial features.
- **Range data** acquired via stereo ranging.
- **Approach**:
  - **Train** using Gabor features on training sets(s)
  - **Apply** by comparing Gabor responses to training data

Gabor Wavelets

- Our familiar Gabor receptive field model:

$$g(\mathbf{x}) = \frac{K}{\sigma^2} e^{-\left[\left(x/\lambda\right)^2 + y^2\right]/2\sigma^2} e^{2\pi\sqrt{-1}(u_0 x/N + v_0 y/M)}$$

- Example Gabor filterbank of 40 filters



Gabor Jets

- Response of each complex Gabor:

$$J_j(\mathbf{x}) = g_j(\mathbf{x}) * I(\mathbf{x}) \approx A_j \cdot \tilde{G}_j\left[\nabla\phi_j(\mathbf{x})\right] \cdot \exp\left[\sqrt{-1}\phi_j(x, y)\right]$$

- Estimate and remove $\tilde{G}_j\left[\nabla\phi_j(\mathbf{x})\right]$

- Yields set of (e.g. 40) phase responses calculated at each pixel called a **Jet**.

$$\vec{J} = \{J_j, j = 1, \ldots, 40\} \qquad J_j = A_j \exp\left(\sqrt{-1}\phi_j(\mathbf{x})\right)$$

Gabor Jet Similarity

- Phase similarity between two jets:

$$S\left[\vec{J}(\mathbf{x}),\vec{J}'(\mathbf{x})\right]=\frac{\sum_{i=1}^{40}A_iA_i'\cos\left[\phi_i(\mathbf{x})-\phi_i'(\mathbf{x})\right]}{\sqrt{\sum_{i=1}^{40}A_i^2\sum_{i=1}^{40}A_i'^2}}$$

$$-1\le S(\vec{J},\vec{J}')\le+1$$

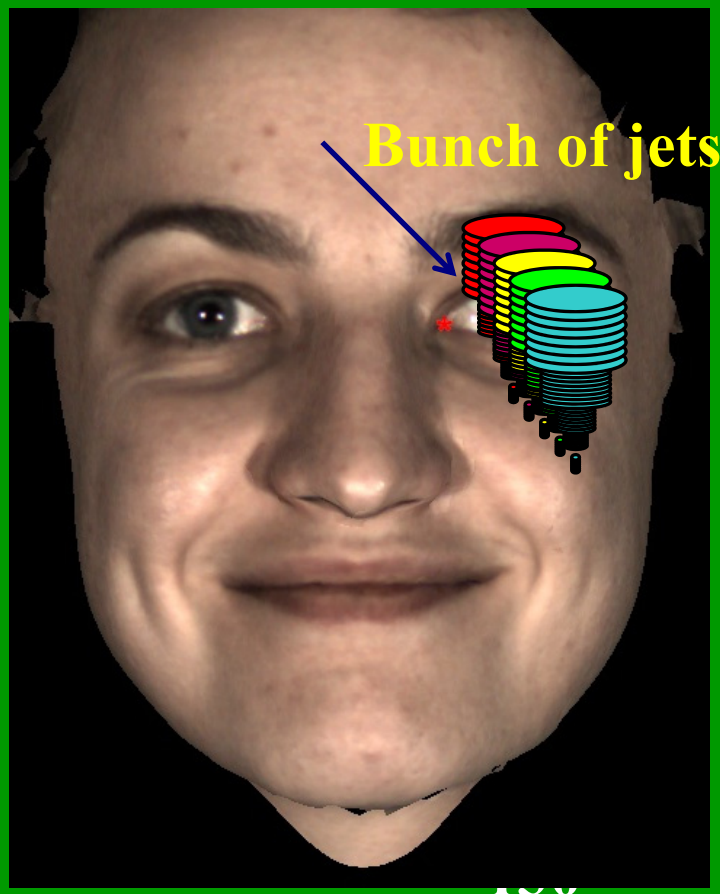
Gabor Bunches

- **Gabor jets** from multiple features bundled in **training stacks** called **Gabor Bunches**.

- Similarity between a **Jet** and a **Bunch**:

$$S_B(\vec{J}, \vec{\vec{B}}) = \max_{i=1} S(\vec{J}, \vec{B}_i)$$

- Find the **jet** $\vec{B}_i$ in the **bunch** $\vec{\vec{B}}$ having the **maximum similarity** to the jet $\vec{J}$.



A 2D/3D Face Dataset

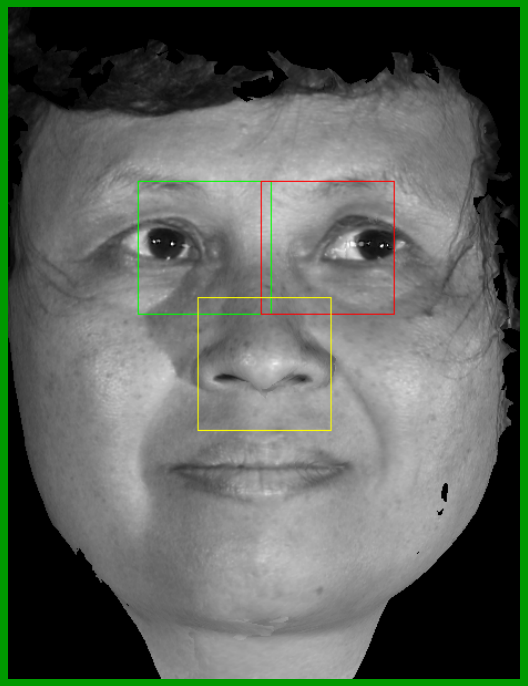
- There are **many** face recognition databases. Here is one described:
- 1196 pairs of **range** and **portrait images**, 119 subjects
- 251x167 images with 256 gray levels
- Both **neutral** and **expressive** images
- 50 **pairs** for **training** (bunch creation)
- 1146 **pairs** for testing

Training Phase

- **Mark 11 feature points on 50 example portraits**
- **Perfectly co-registered range** and **portrait** pairs
- **Jets** from a **specific feature point** (e.g. nose tip) and **modality** (range, portrait) create a **bunch representation**
- One **bunch per feature** point per **modality**

Application Phase

- **Anthropometric knowledge** restricts ROI
- 40x40 boxes put at **mean** feature **locales**



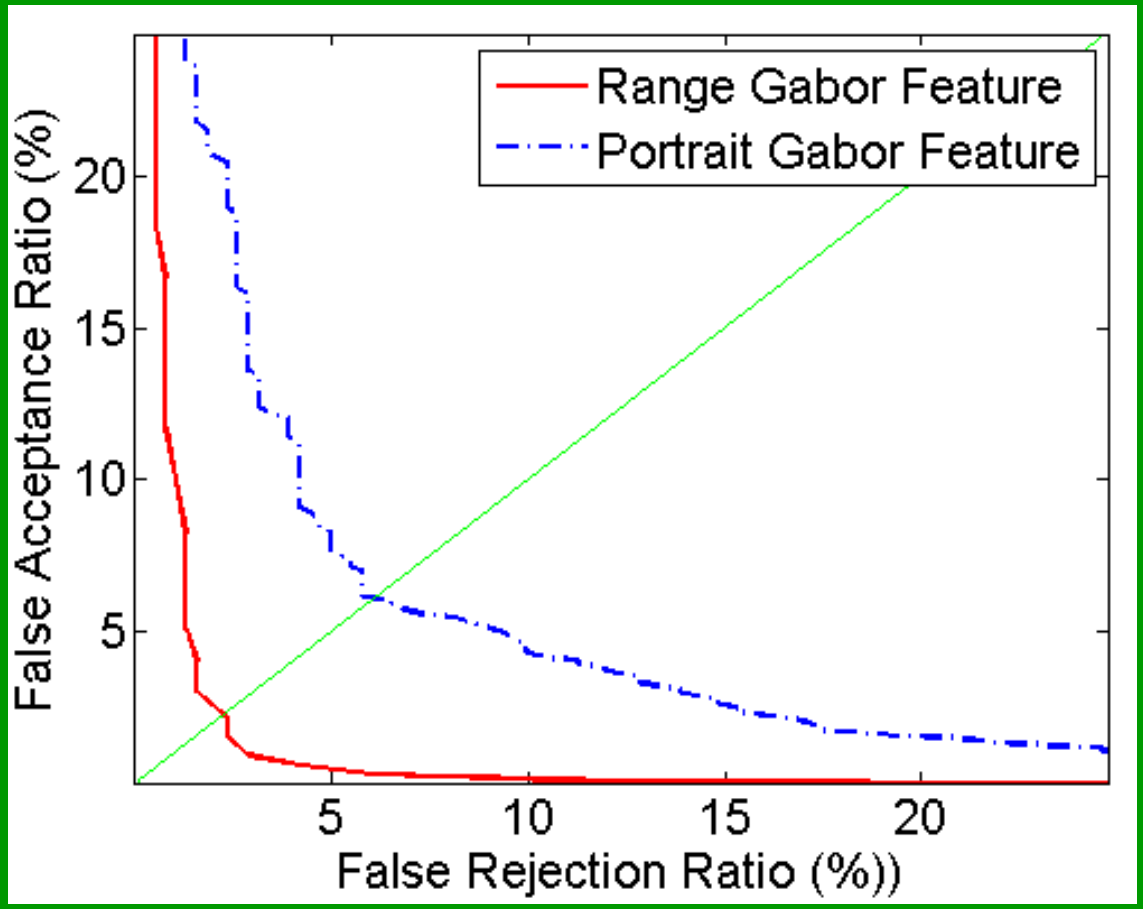
(1) Compute **"Portrait Jets"** and **"Range Jets"** for each modality.

(2) Compute **Jet-Bunch Similarity Maps** for **Range** and **Portrait**

(3) Find point in similarity maps with **highest value**

(4) Combine **Portrait** and **Range modalities**

Portrait vs. Range Features



Results

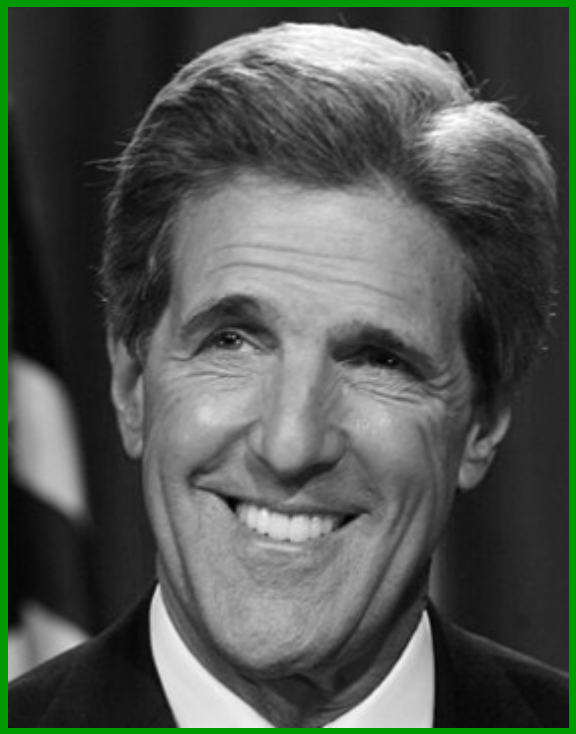
- Examples of performance: **2D vs. 3D**

	Algorithm	Modality	Test set	Detection Accuracy
Cristinacce *2004*	Shape Optimized Search (SOS)	Portrait	1521	85%
Pan *2003*	PCA and Hausdorf Distance	Range	360	93-95%
Gabor Jets	**Gabor Jets**	**Portrait & Range**	**1146**	**>98%**

Visual Adaptation to Faces

Mr. B

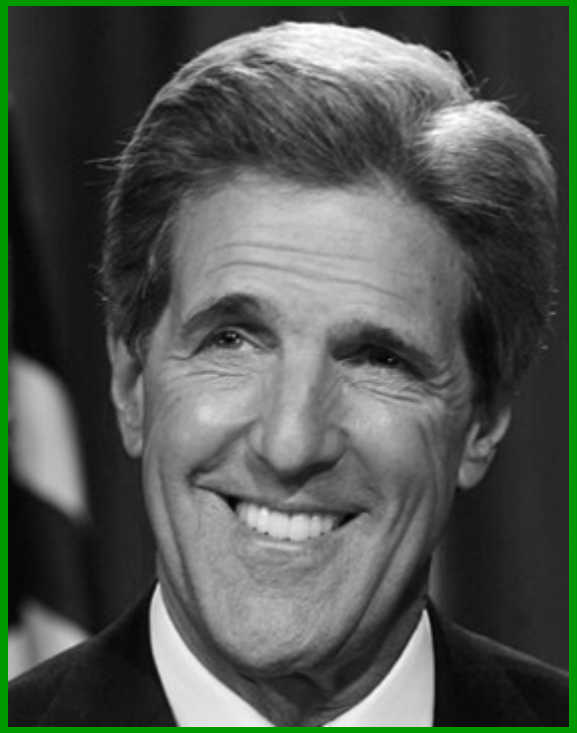
Morph of Mr.s B & K



Mr. K

Adapt to Mr. B

Morph looks like Mr. K



Adapt to Mr. K

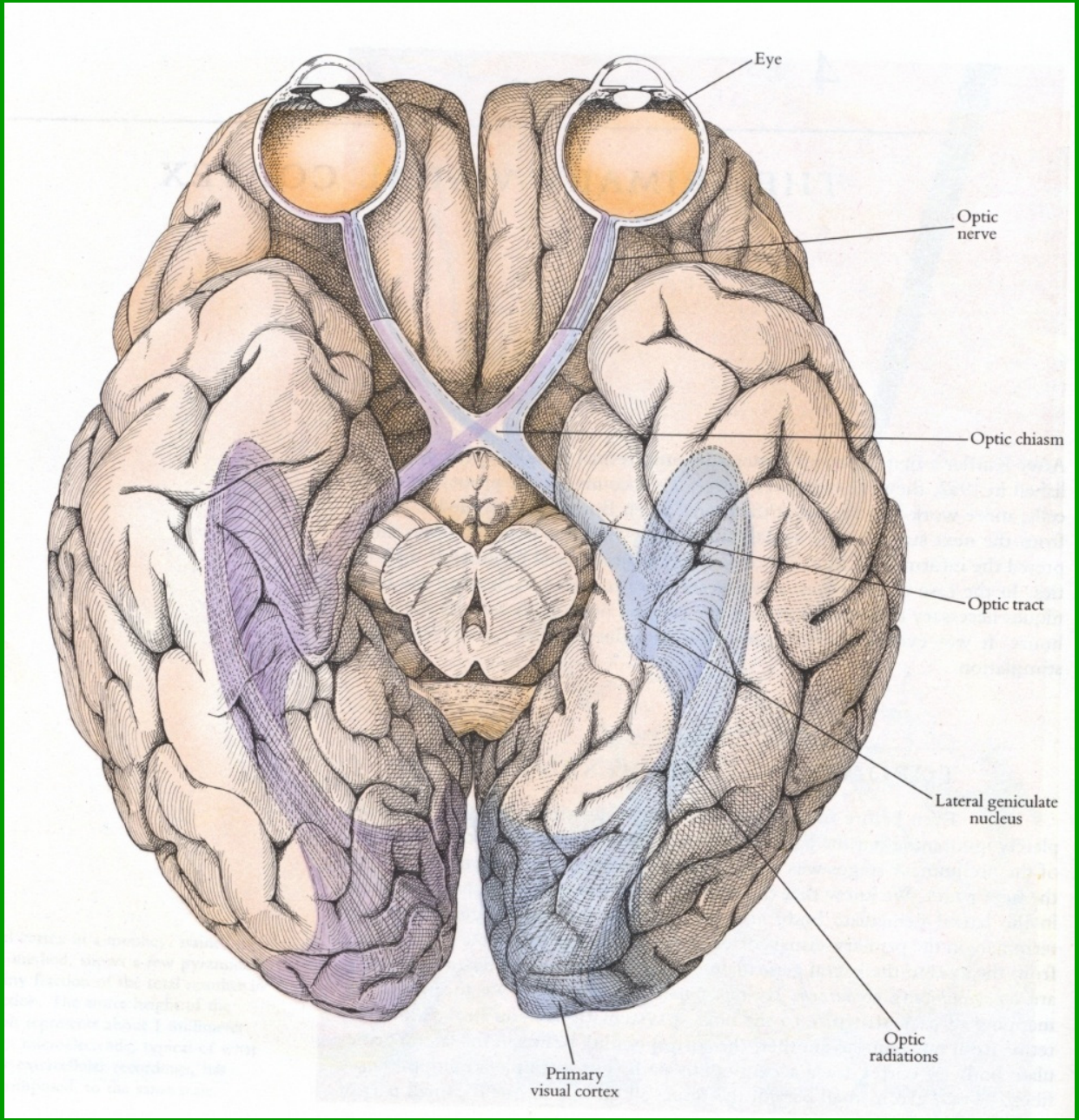
Morph looks like Mr. K

How About a Pig's Face?

Would you recognize me?

Grotesque Flashed Face Illusion

Nobody understands this remarkable illusion …

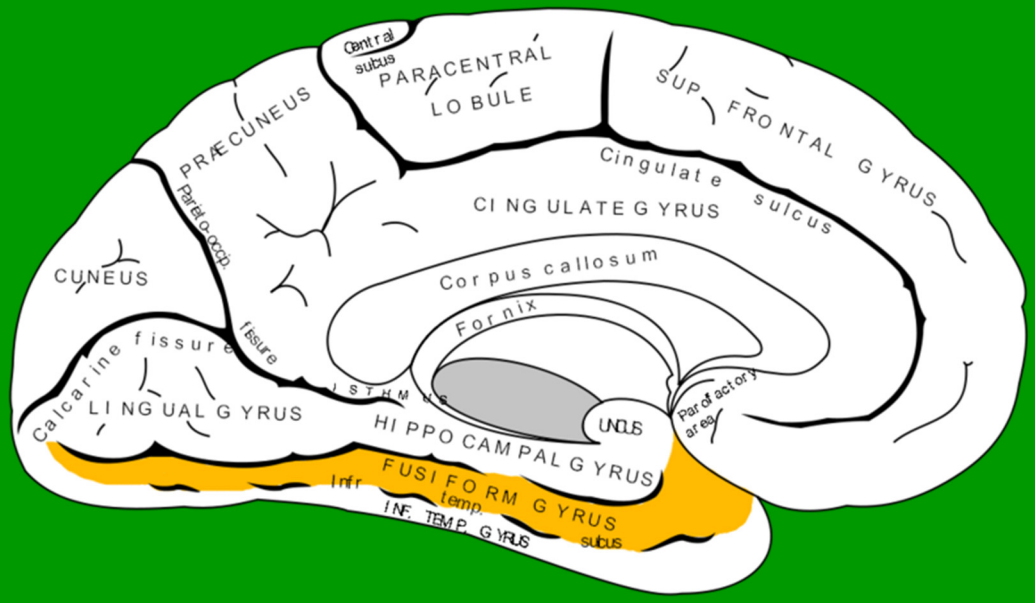


What we know about the **specific functions** of the **elements** of the visual system remains **quite limited.**

The Present

- Computational models of **visual perception** have **immeasurably enriched image processing** and **analysis.**
- I have only **scratched the surface** of **existing applications.**
- **Future applications** should find **remarkable advancement** and **breadth** as our **models improve.**

Fusiform Gyrus



- Where much of **face processing** occurs
- **Left fusiform gyrus**: **Recognition** of face-like features
- **Right fusiform gyrus**: **Identification** of "faceness"
- Damage here may lead to **prosopagnosia**

Object Recognition

- **General-purpose object recognition** remains quite **unsolved**
- Complicated by **huge diversity** of **objects**, poses, conditions
- And **other factors**:

Baron Caterpillar

Pygmy Seahorse

Comments

- We will now enter the next dimension and study **digital video** … on to **Module 9**.