Module 9
Digital Video I

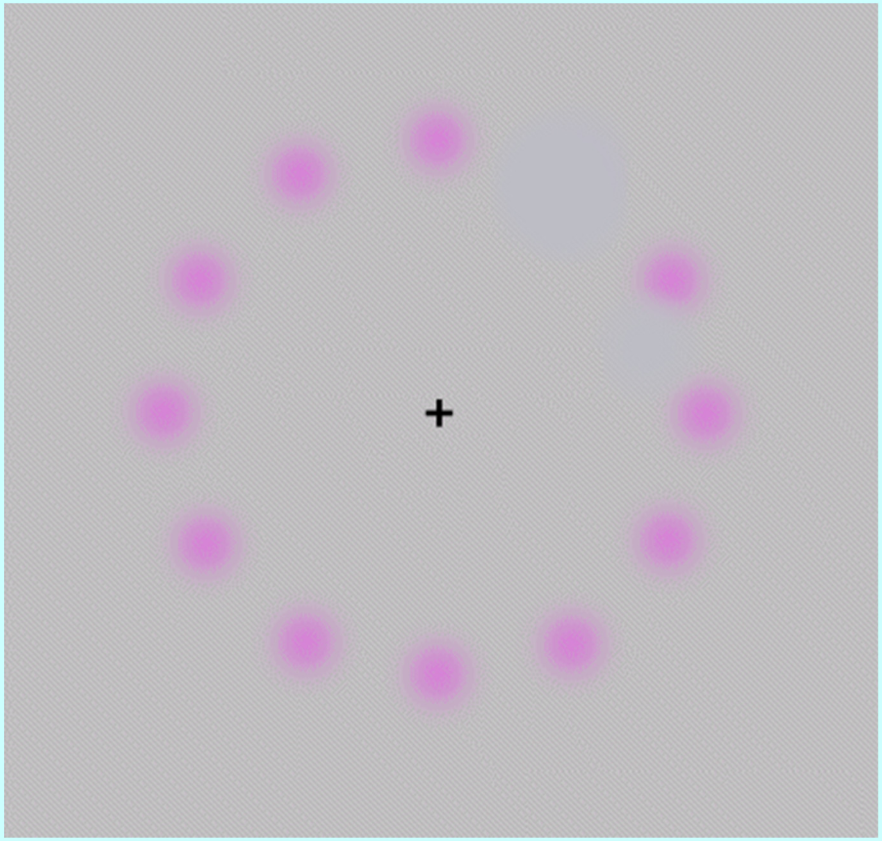
- Analog and Digital Video Formats
- Video Scanning
- Video Sampling
- Optical Flow

QUICK INDEX

Illusions of Motion

"The Lilac Chaser"

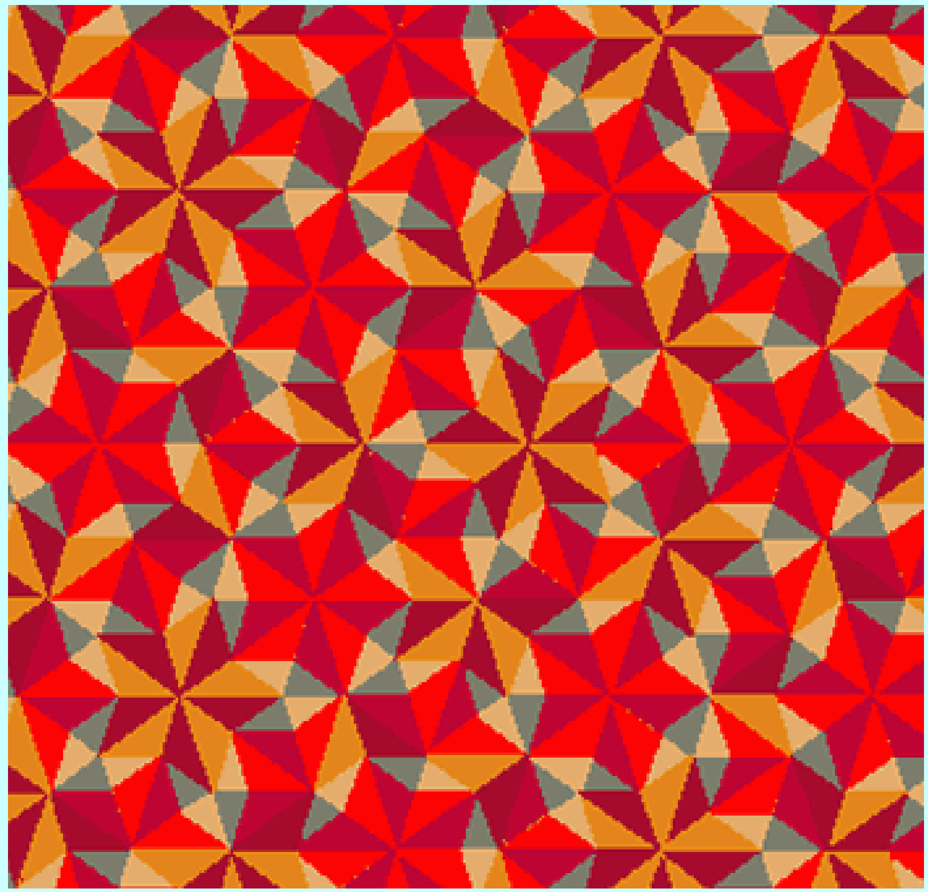
(Jeremy Hinton)



Stare at the center dot

Rotating Spiral Adaptation

Watch this!



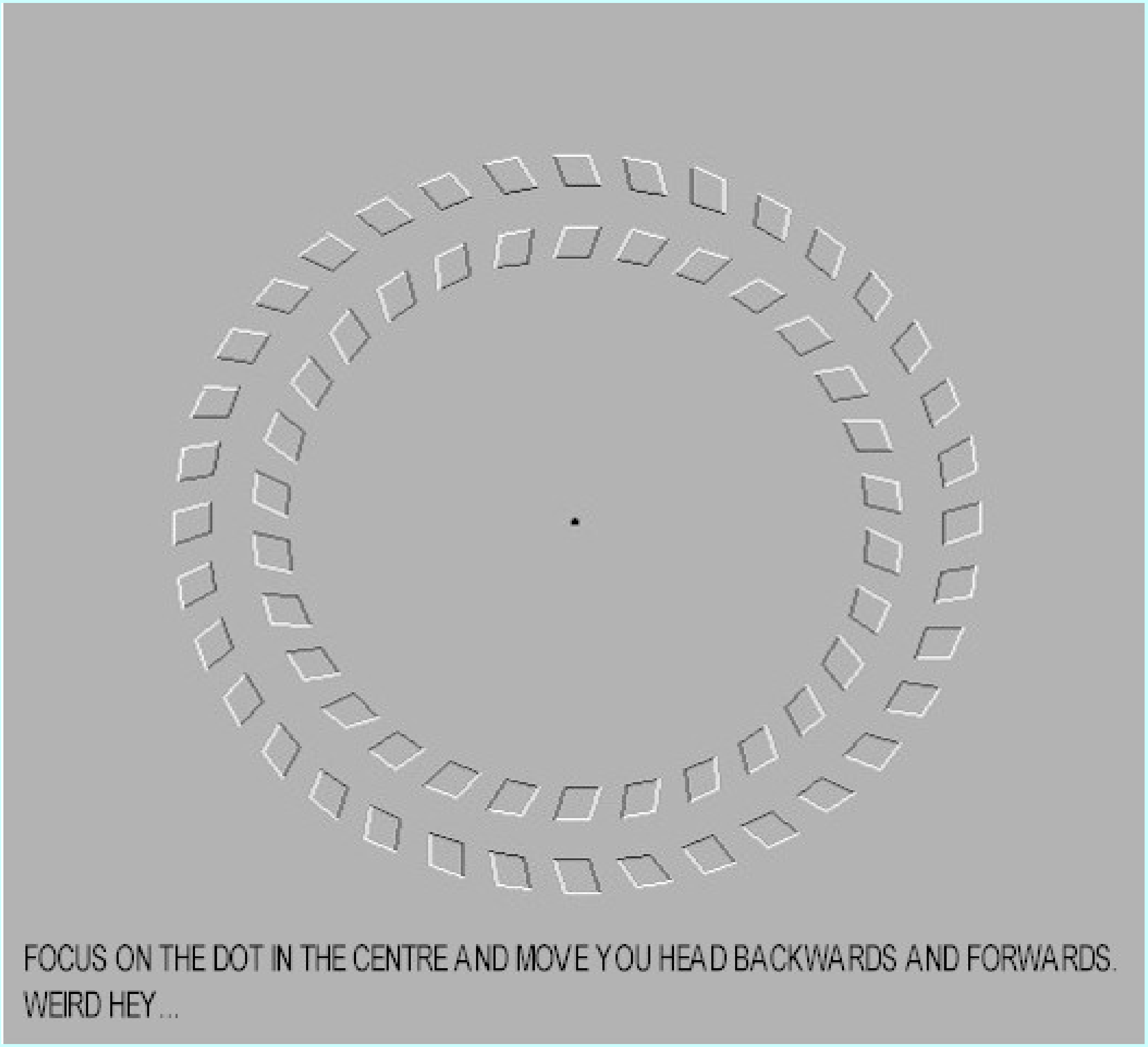
… then stare at the middle of this…

That was fun, let's try it again!

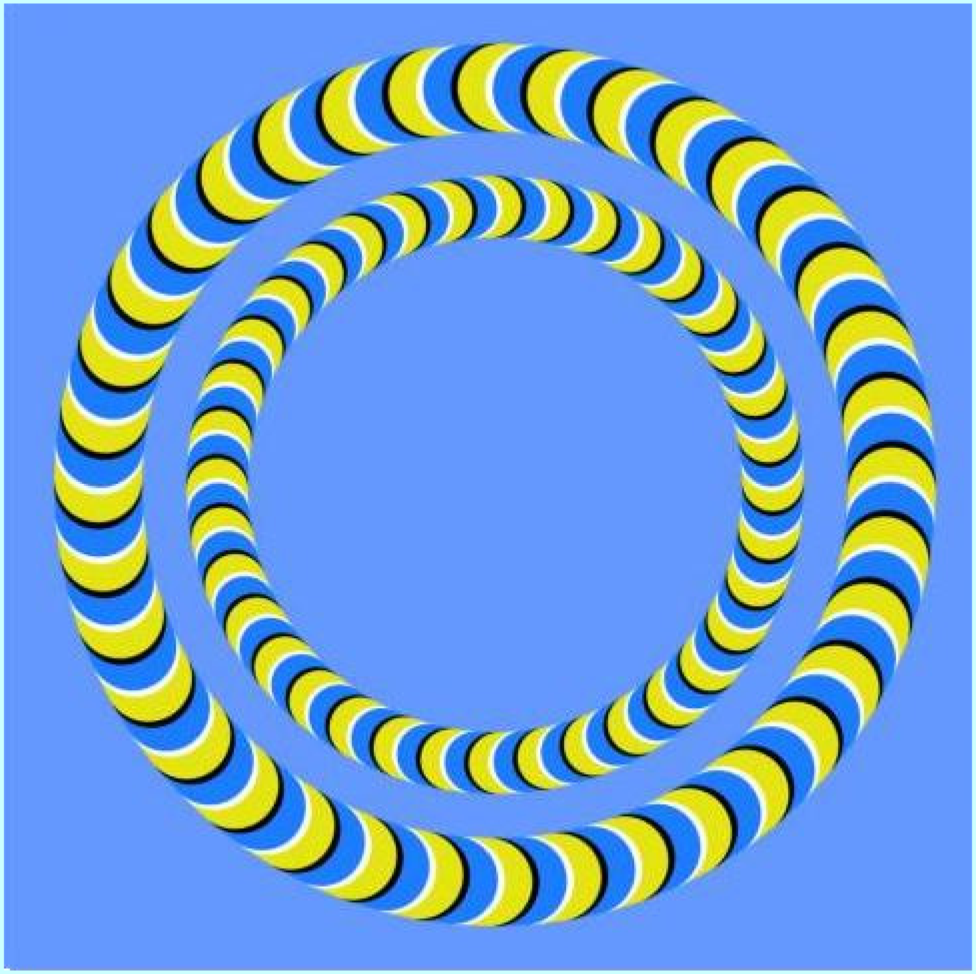
- **Stare at this video** … when it tells you to **look away** then stare at your **neighbor's face!** Video

- Now try it with this one! Video

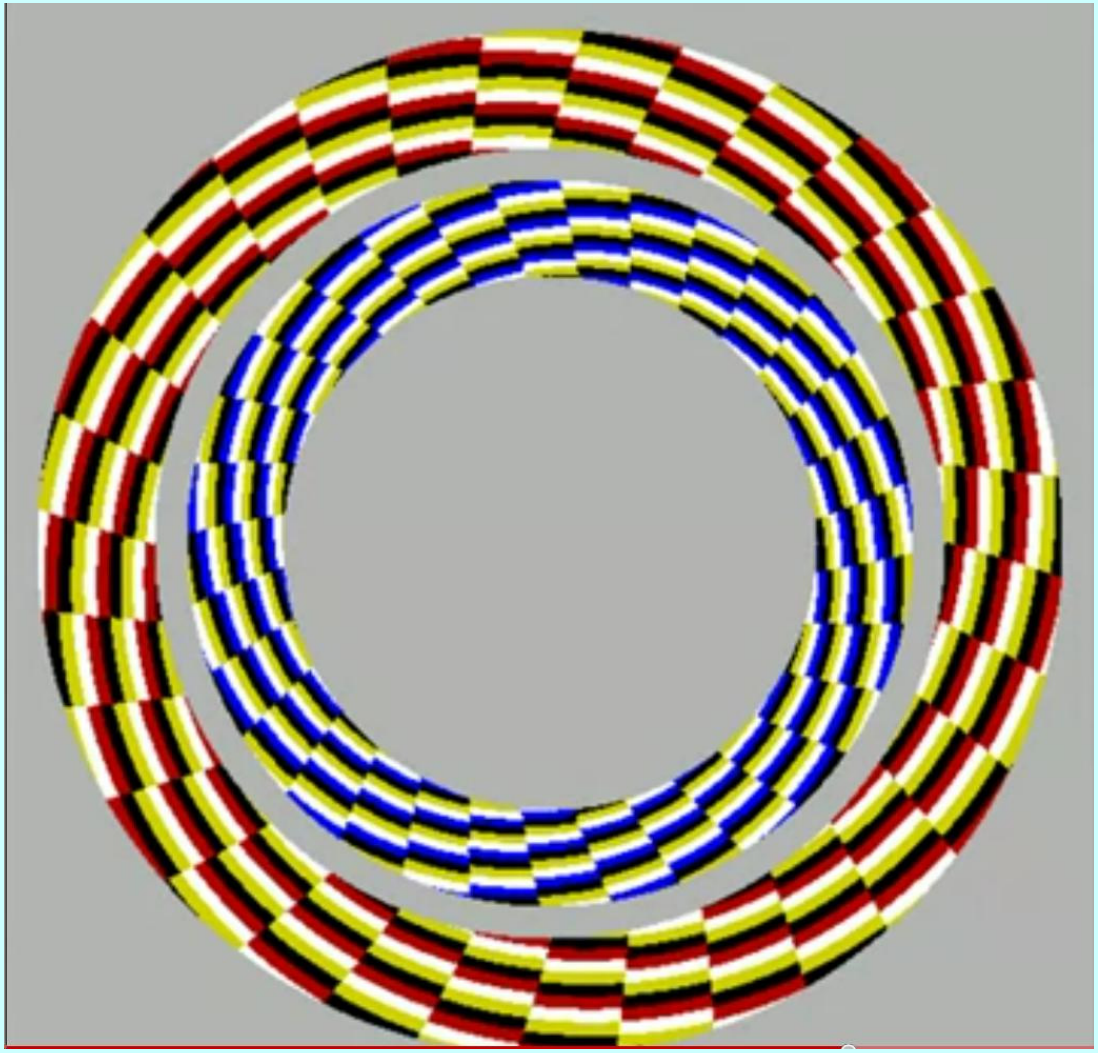
Induced Motion Effects

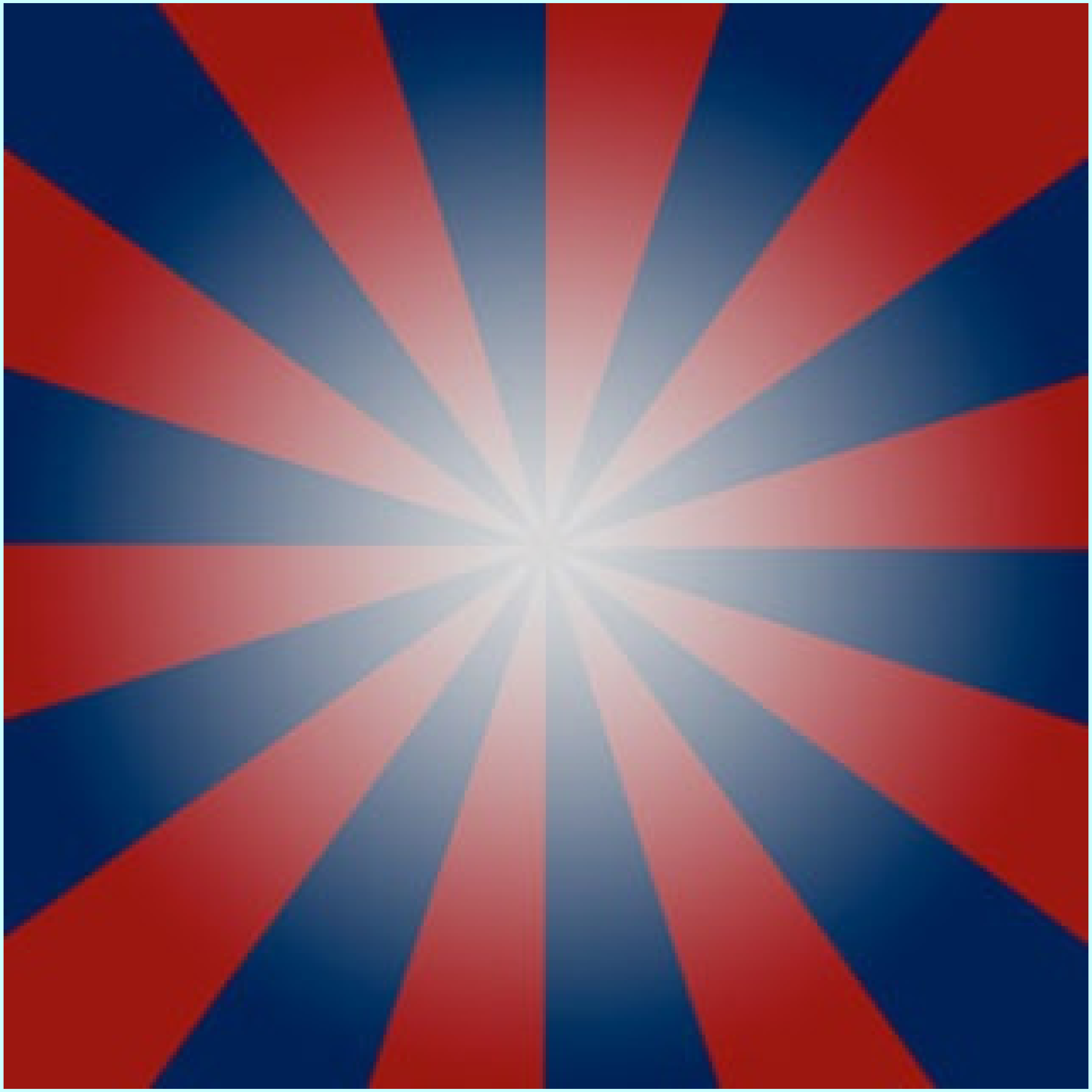


Don’t have to move your head for this one…

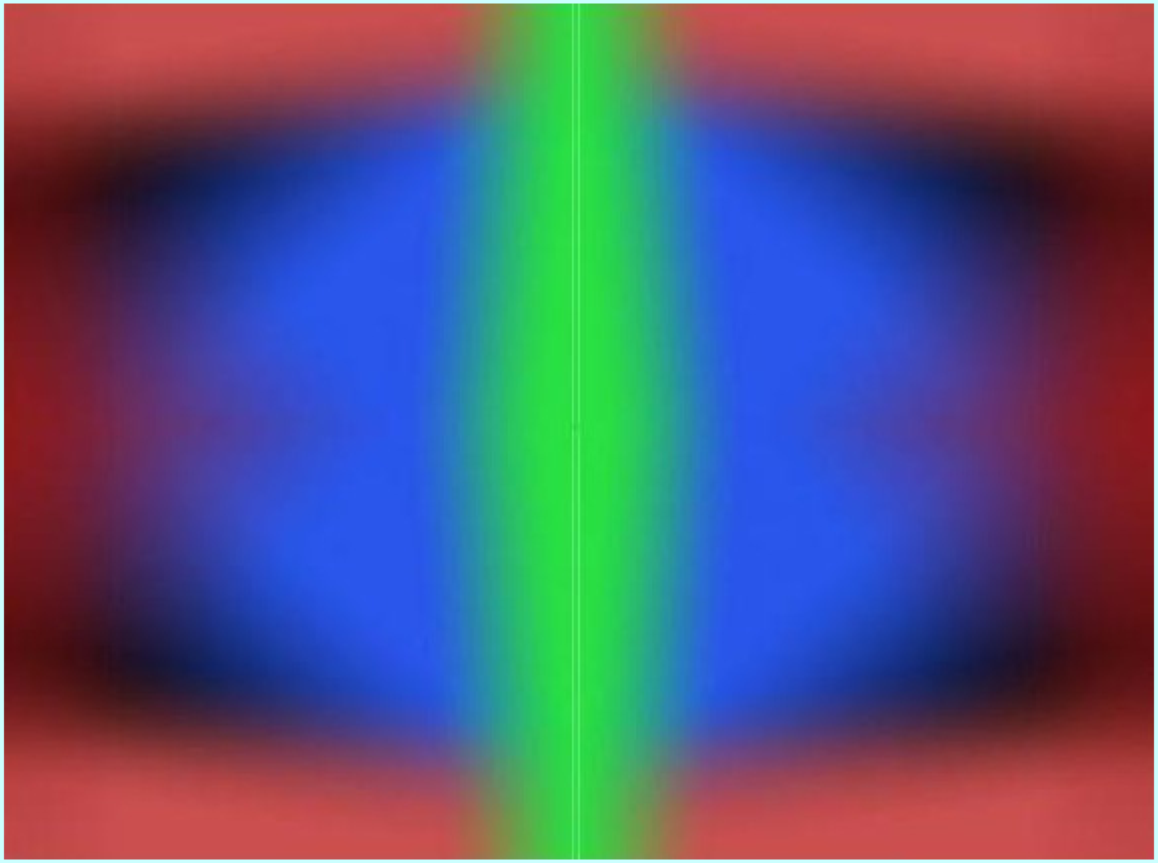


… or this one



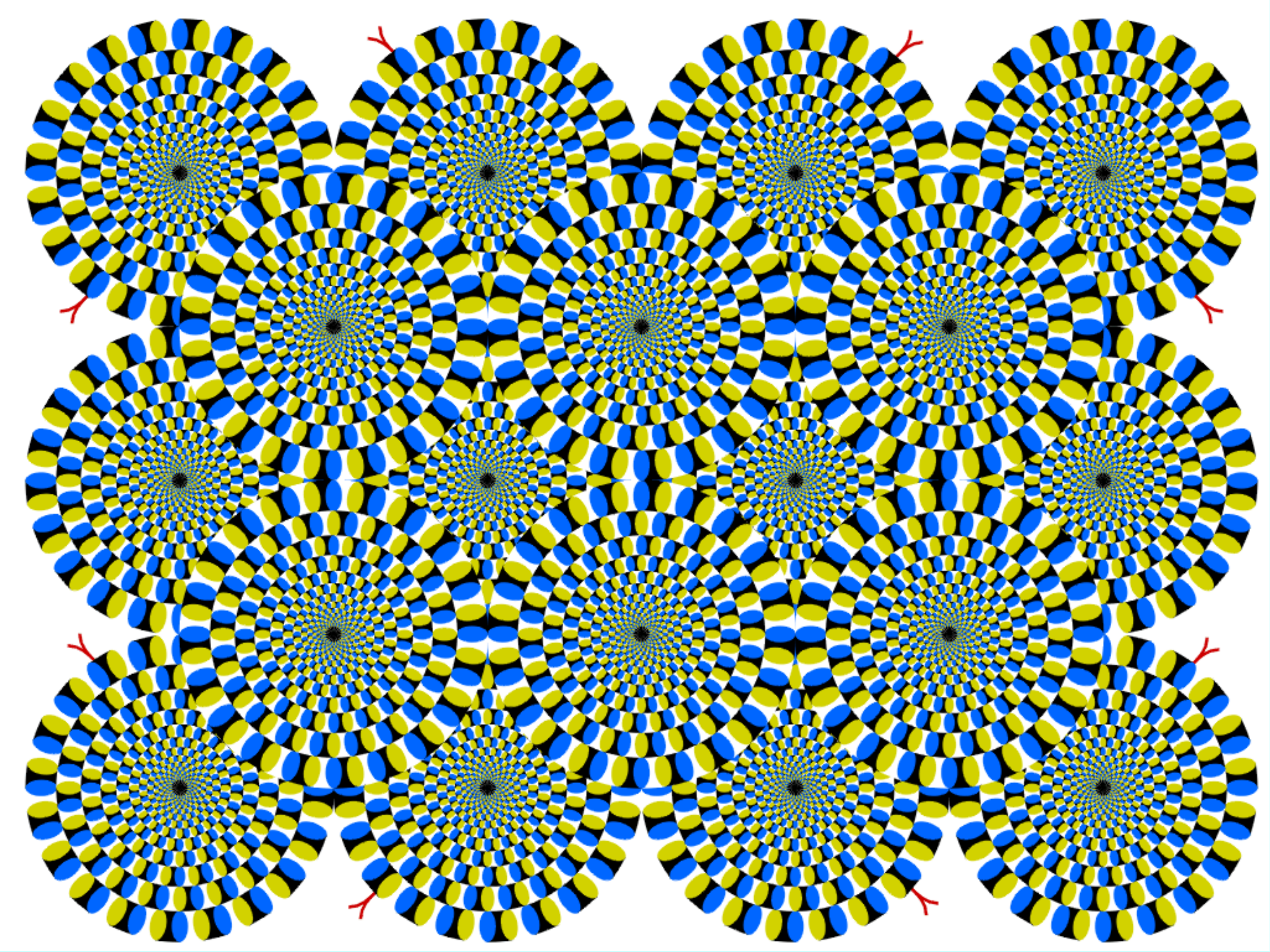


Move your head forward and backwards for this one also

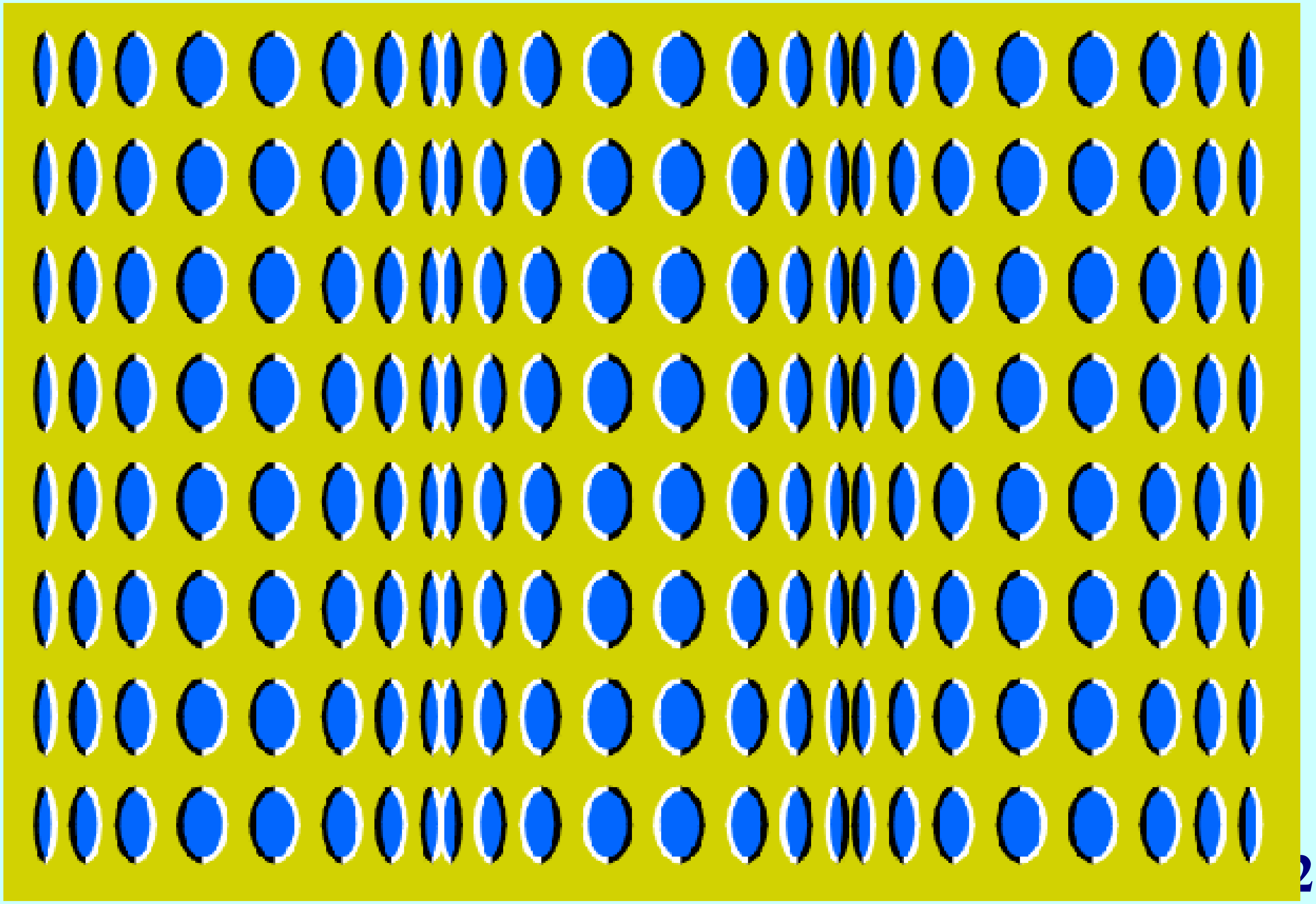


Forward and backwards for this too

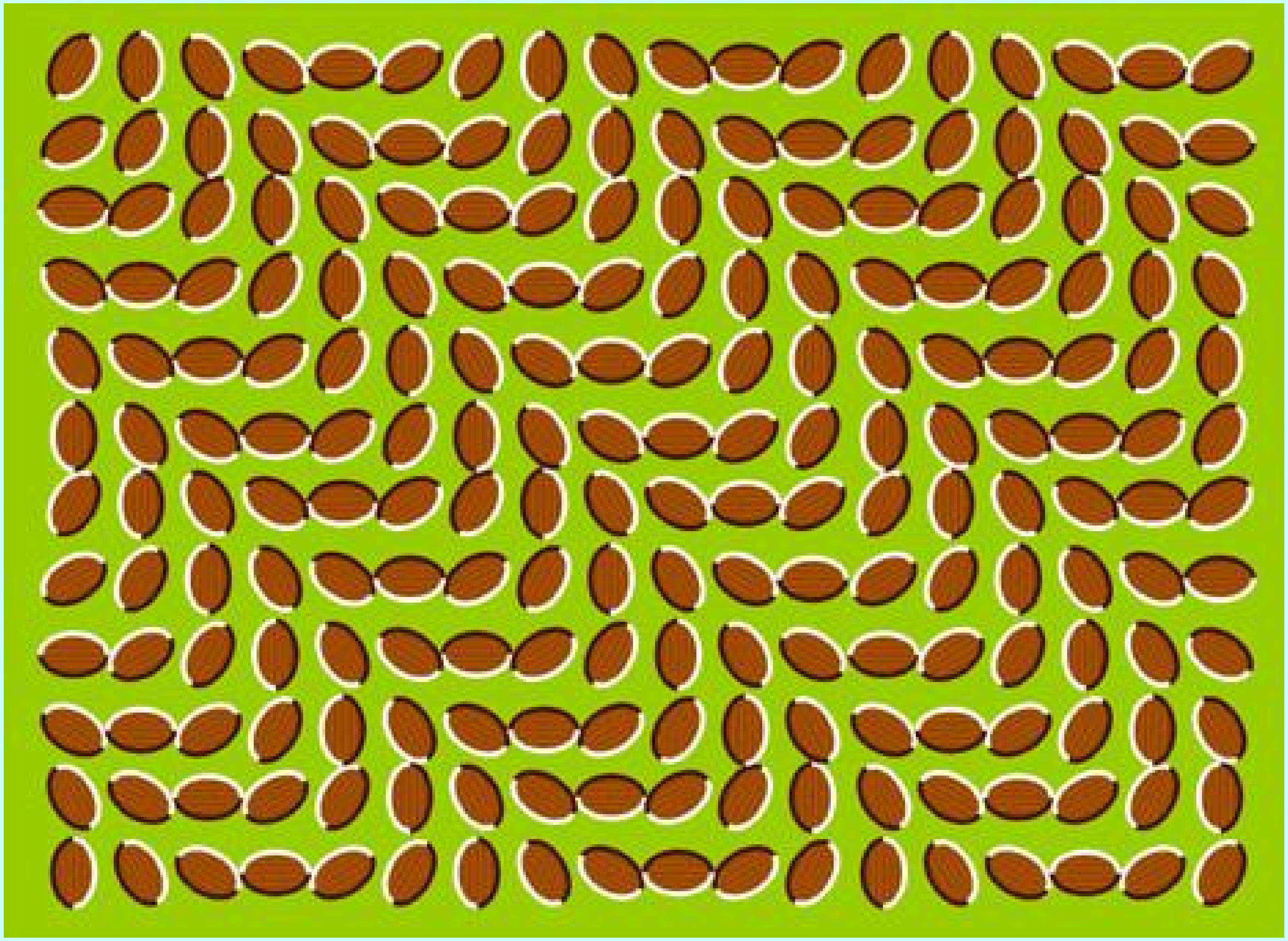
You thought that was bad…



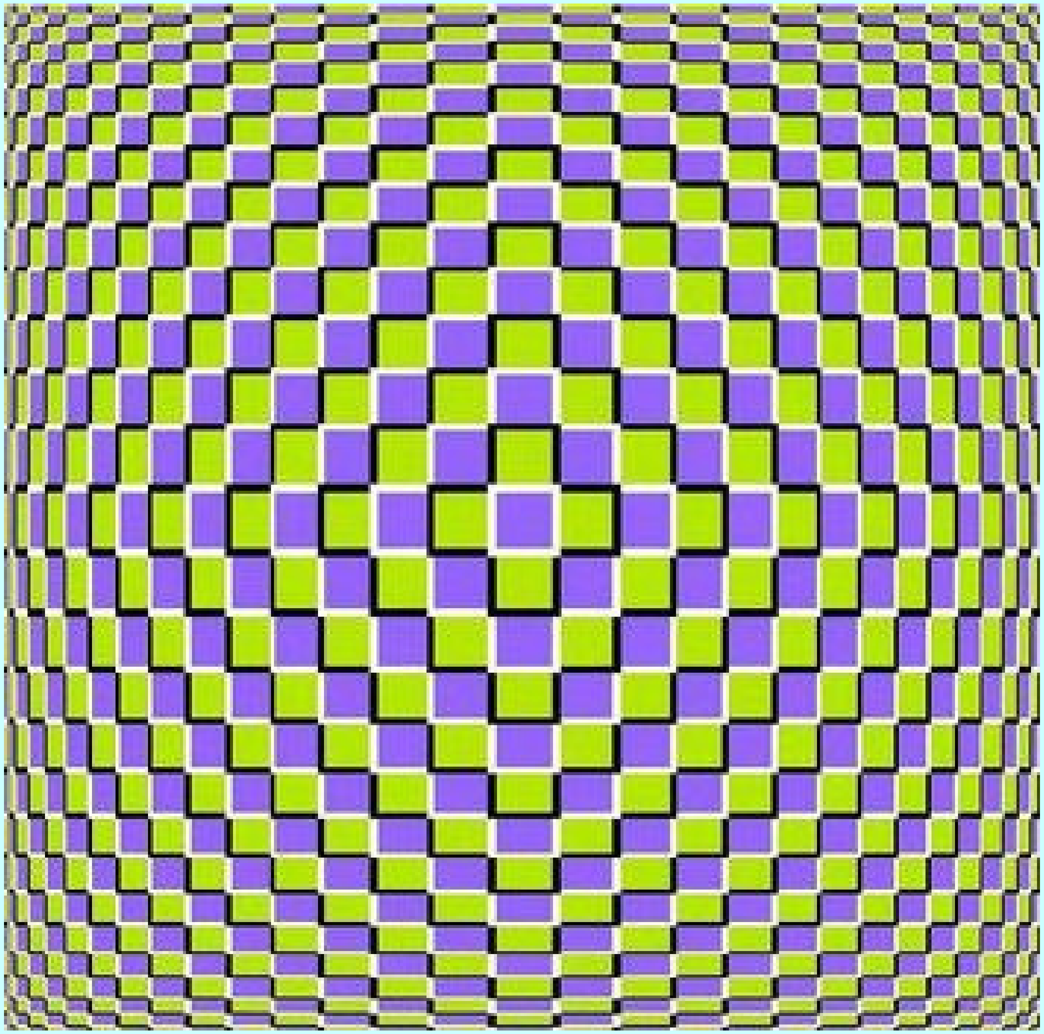
…. and how about this



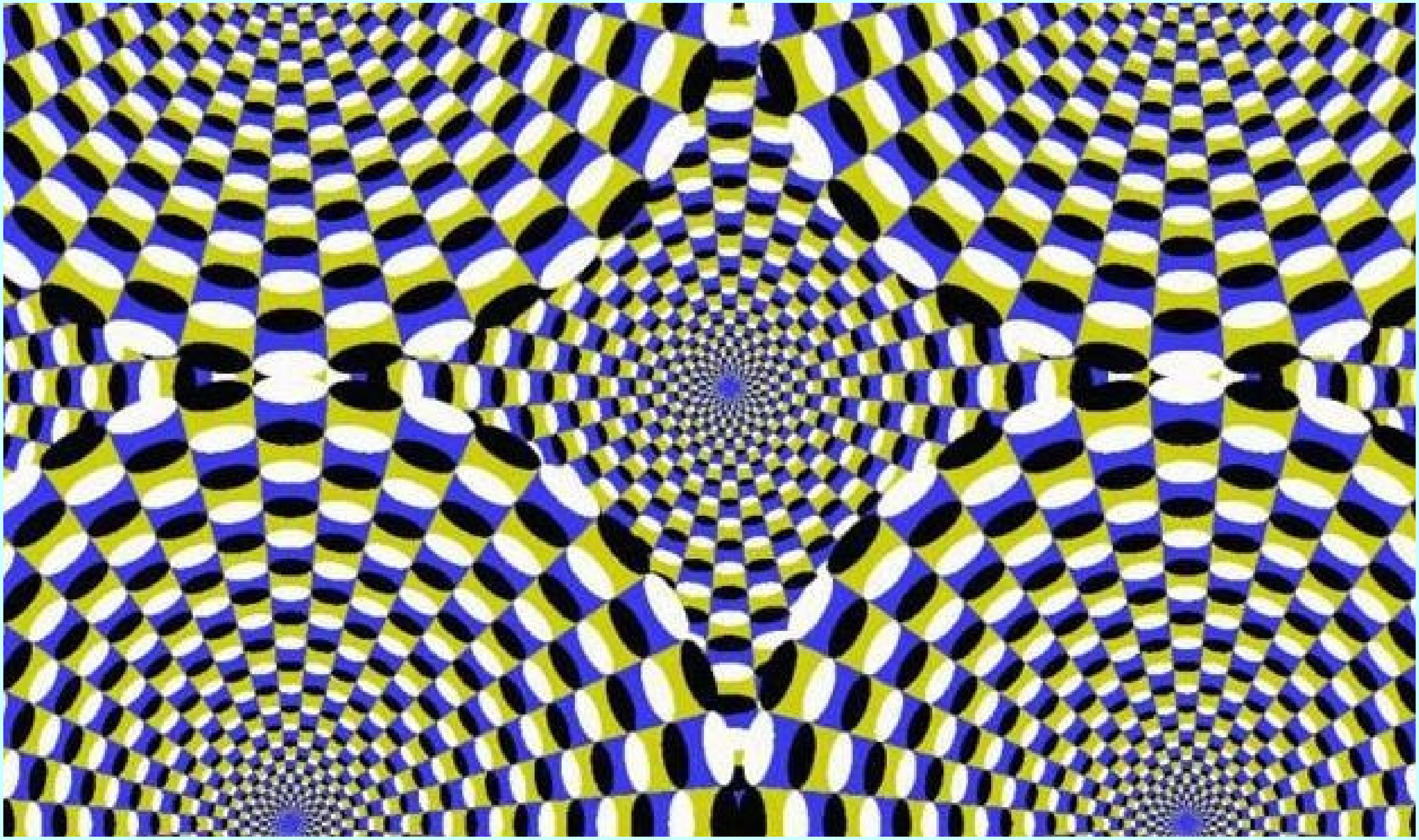
…. and this

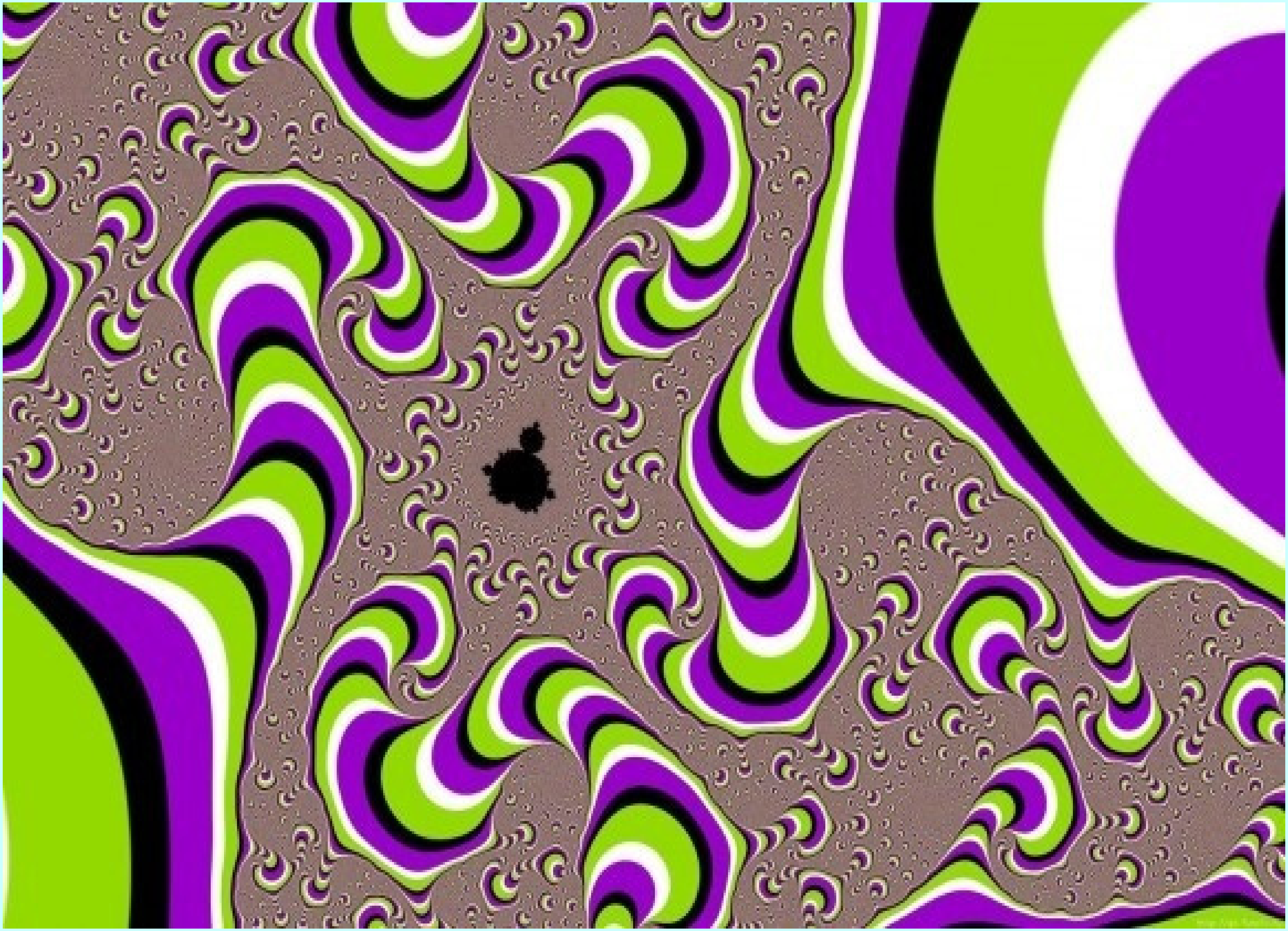


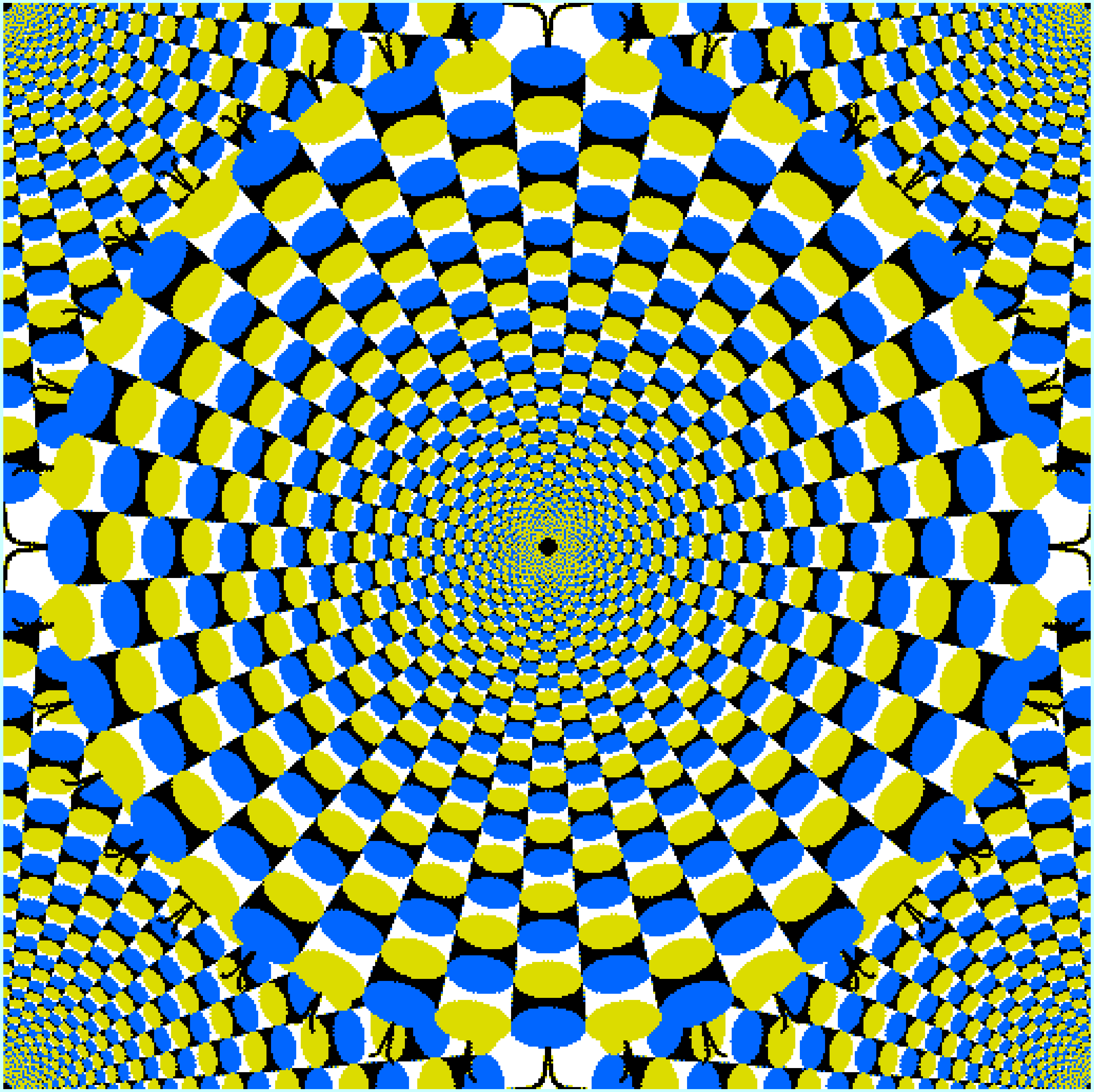
…. and this



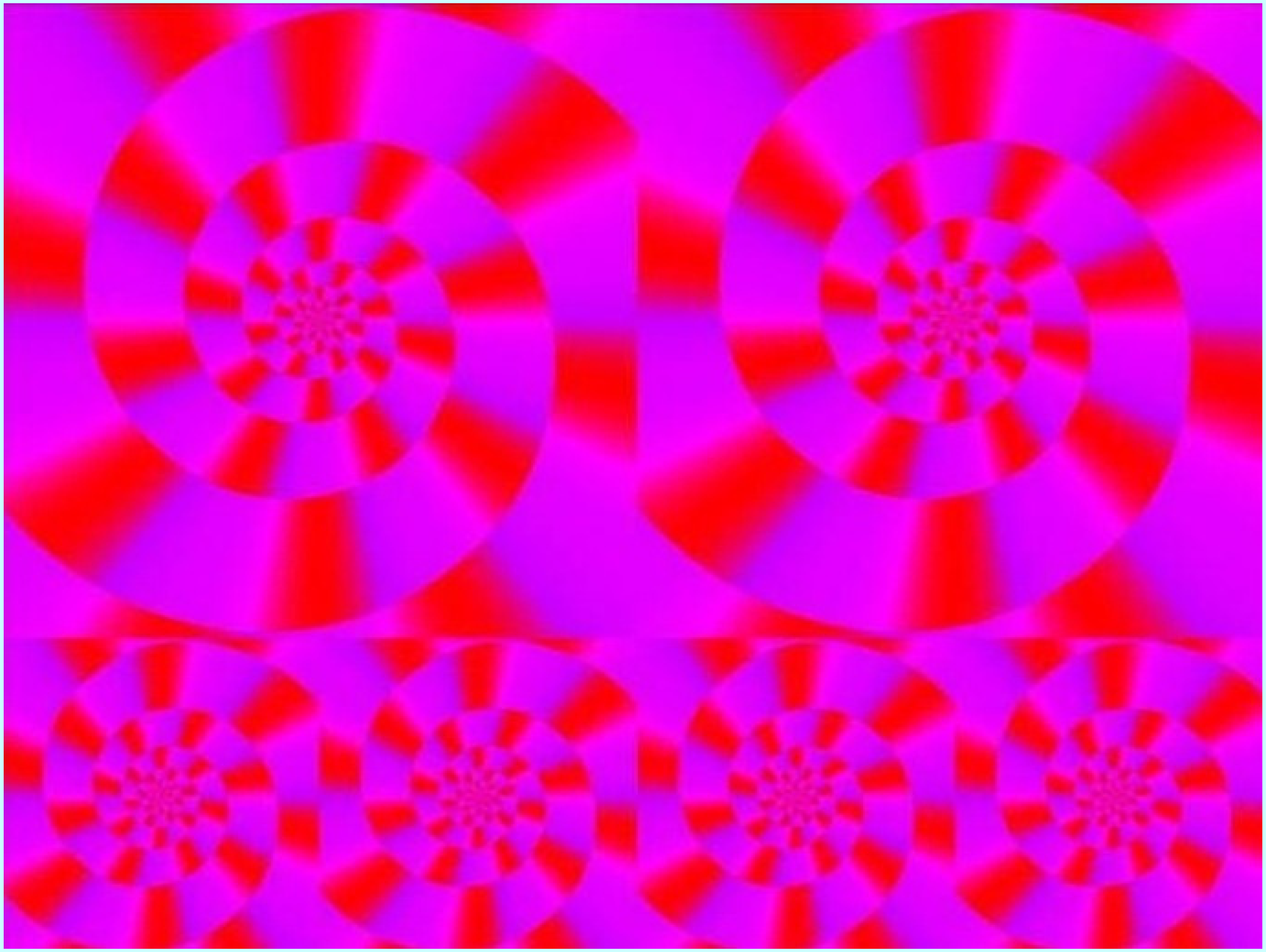
…. feeling OK?



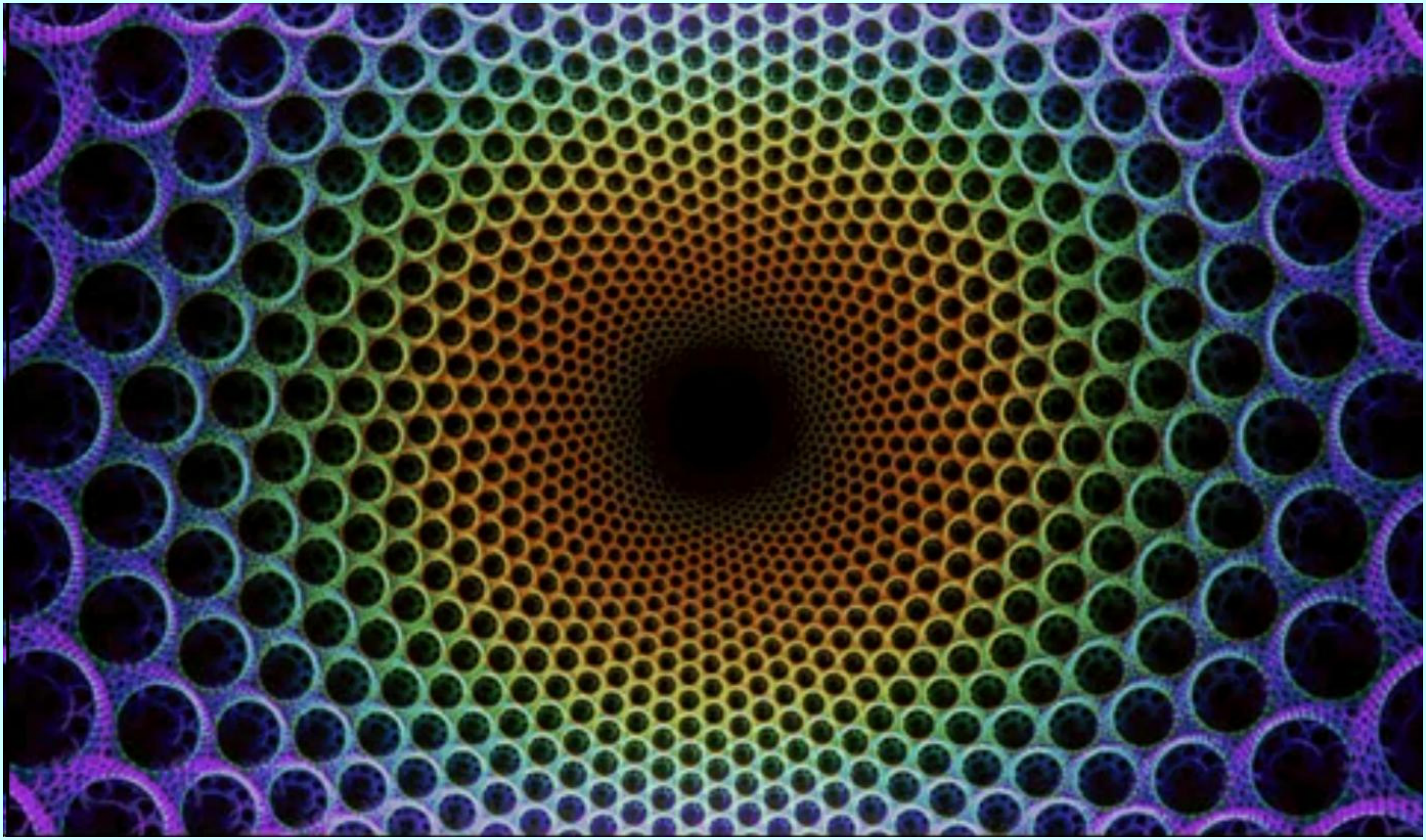




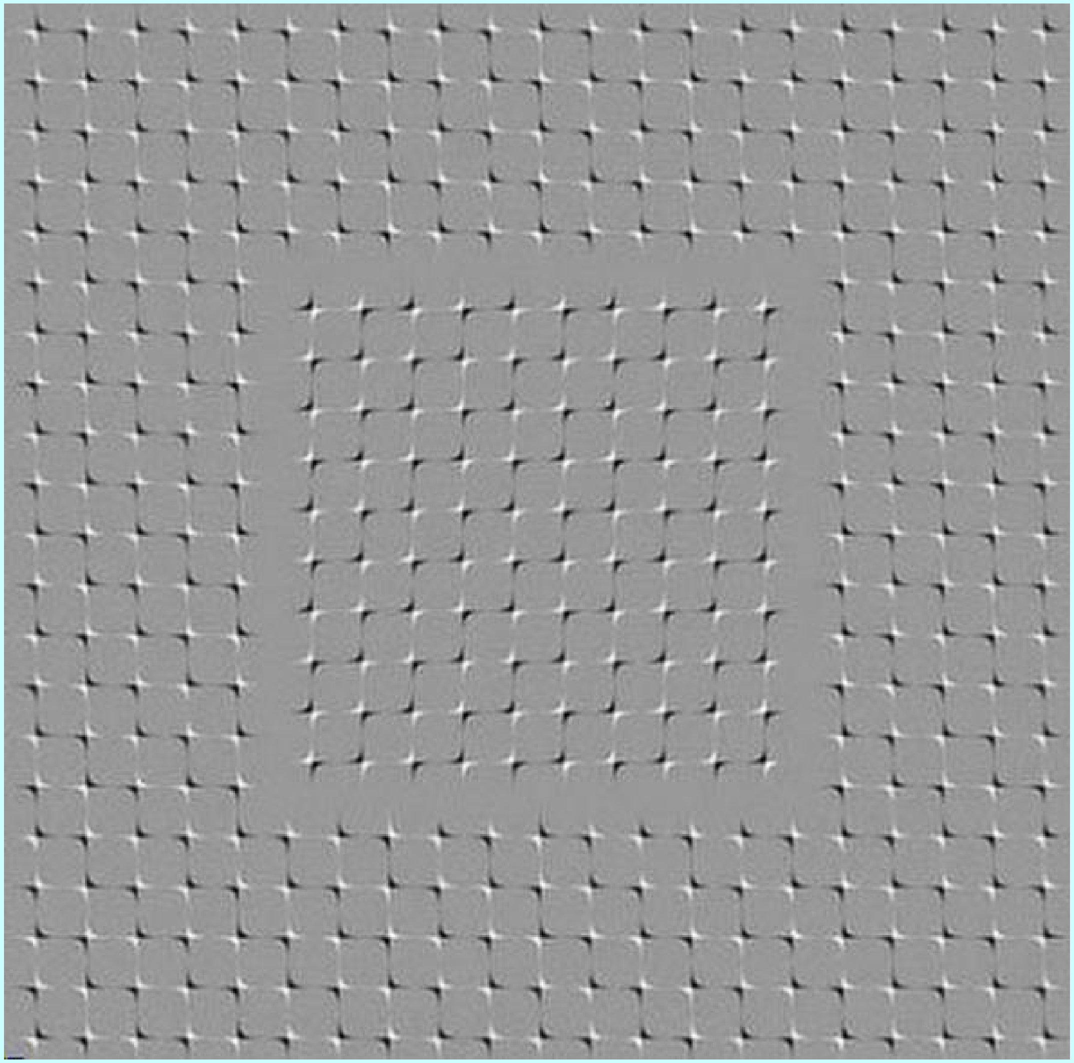
Move your eye back and forth

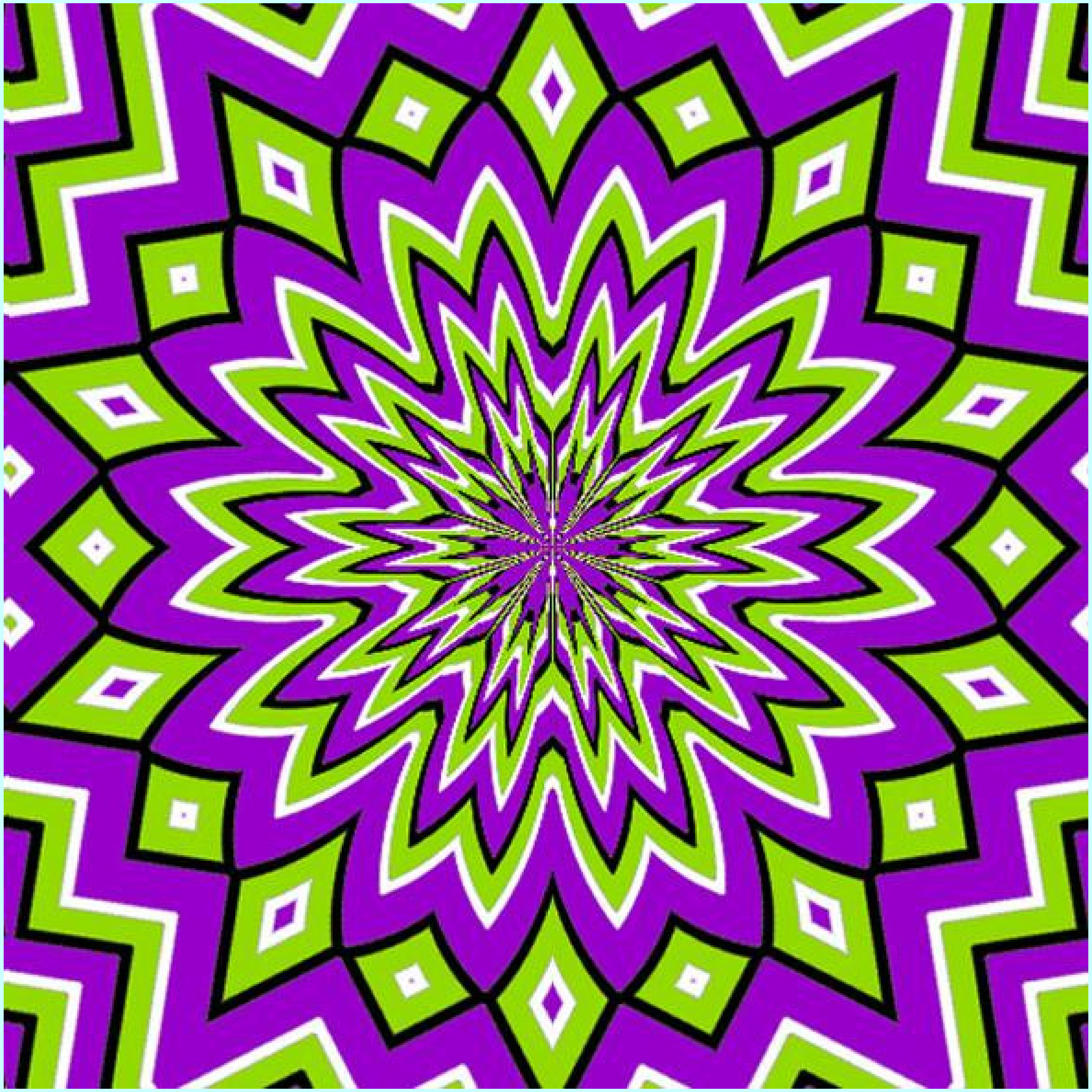


Don’t fall in … (but you can’t stop…)

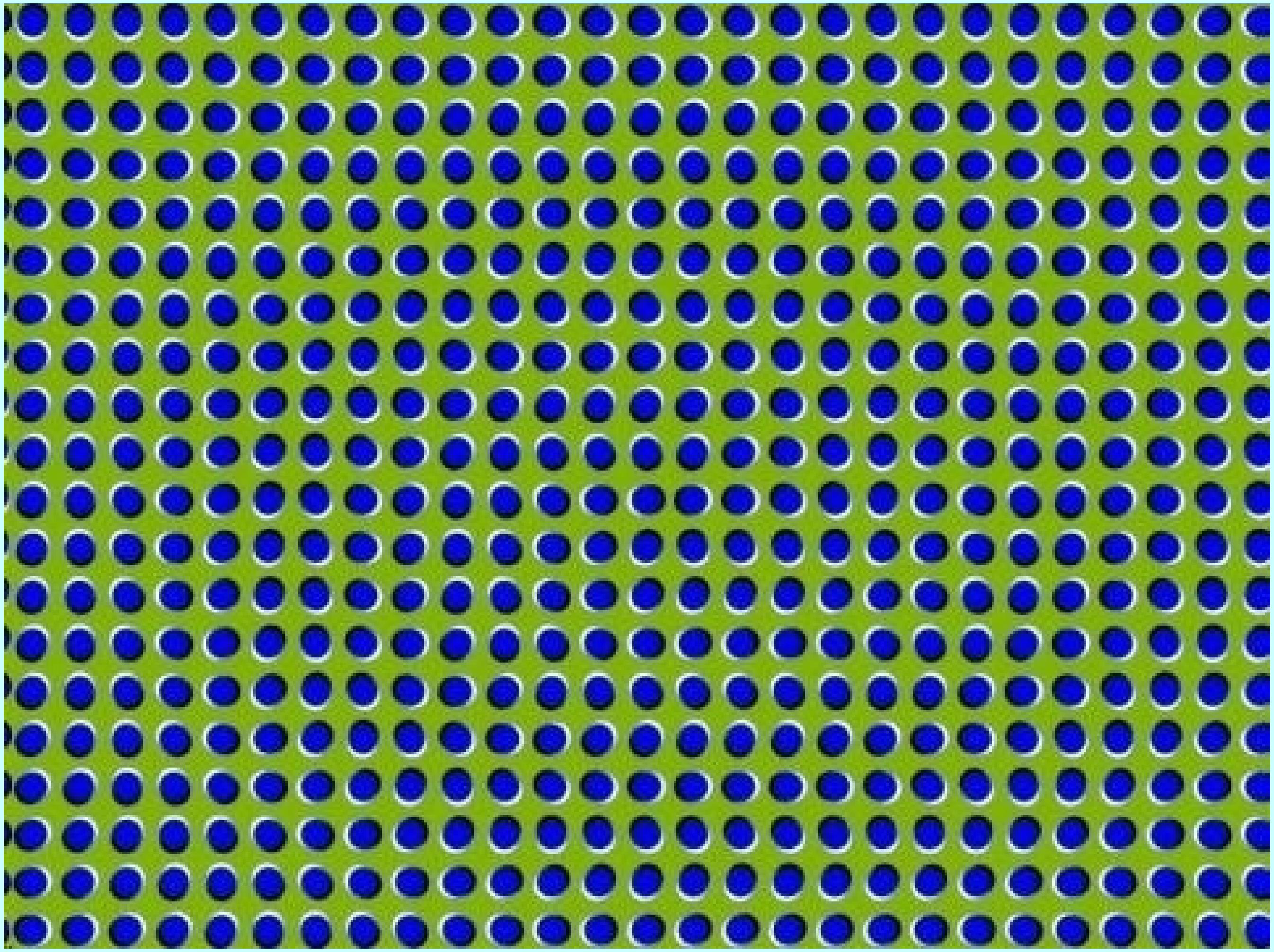


Which way is it moving?

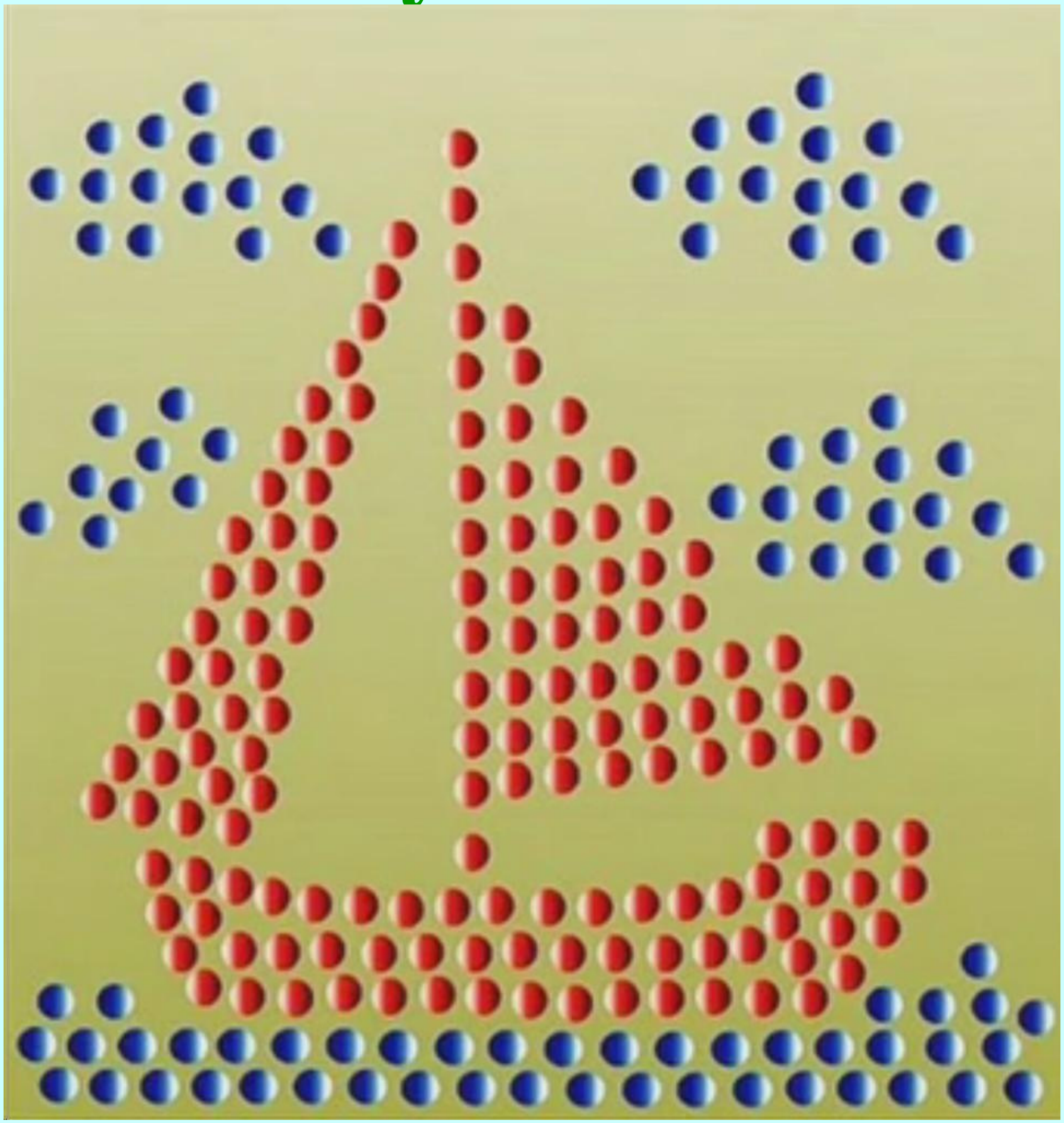




The "Nauseator"

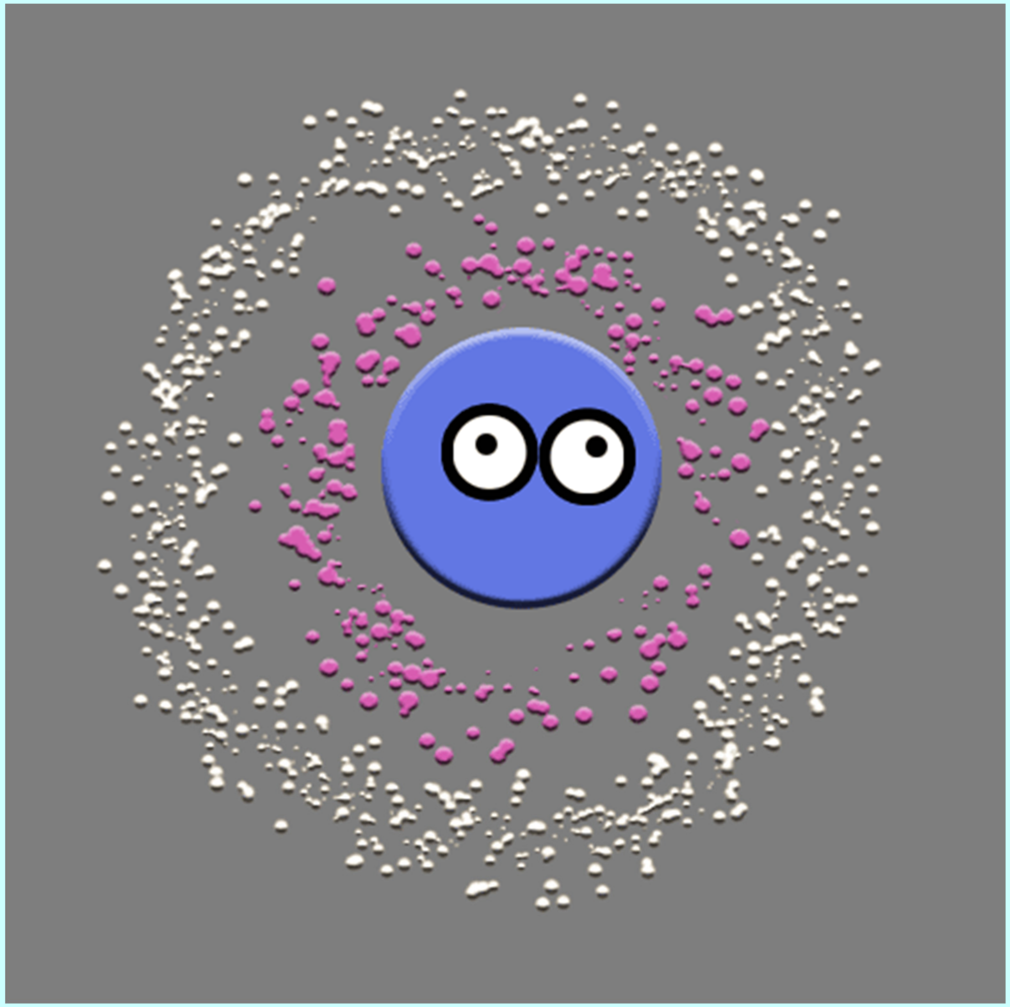


One of my favorites….



Stare at the middle spot

Yet another motion effect



Opposite Effect …. Motion Masking

- <u>**Luminance change masking by motion**</u>
- <u>**Hue change masking by motion**</u>
- <u>**Size change masking by motion**</u>
- <u>**Shape change masking by motion**</u>

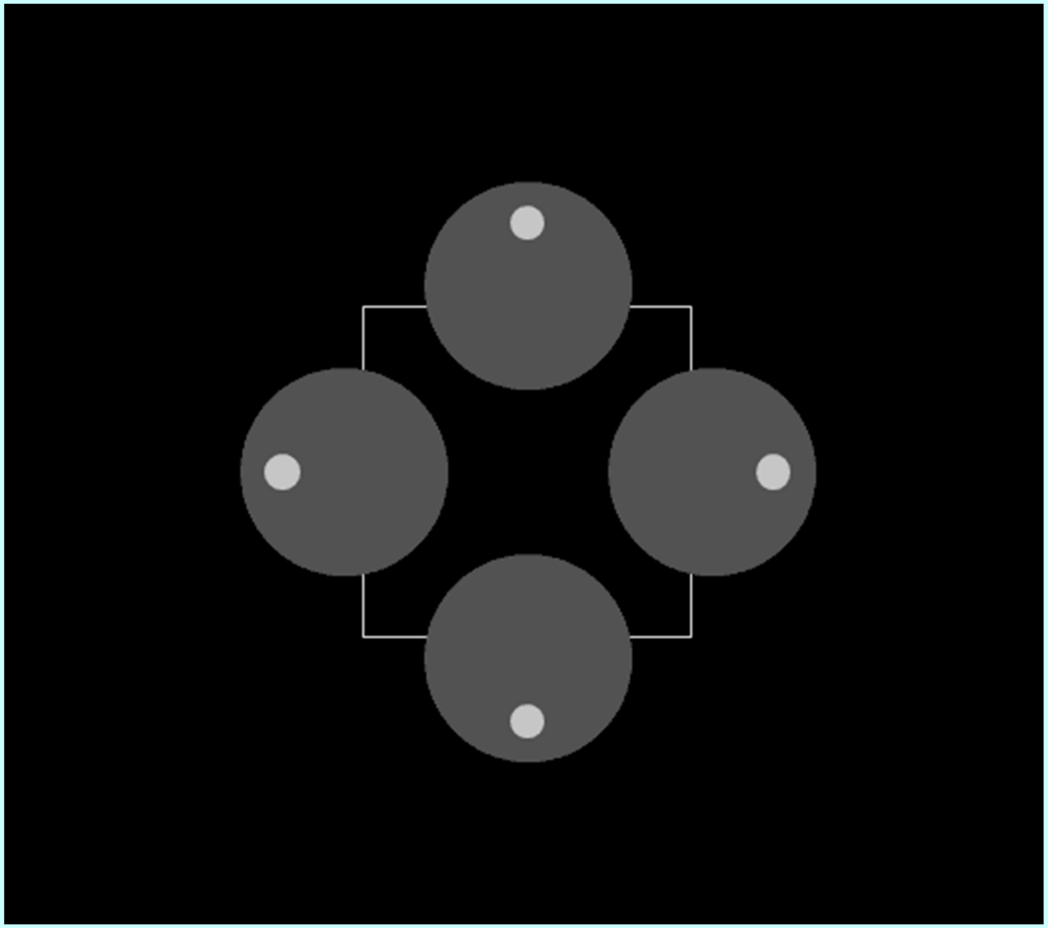
Stare at the center dot throughout each movie…

More Motion Masking

Another Moving Cylinder?

- **Which way is the cylinder turning?**
- **Is a dot on the front or rear face?**
- **Does your perception change over time?**

Moving Illusory Contour



- Stare at the dots
- Then stare at the center
- See the same thing?

INTRO TO DIGITAL VIDEO

- **Digital Video** is the **frontier** and the **future** of image processing.

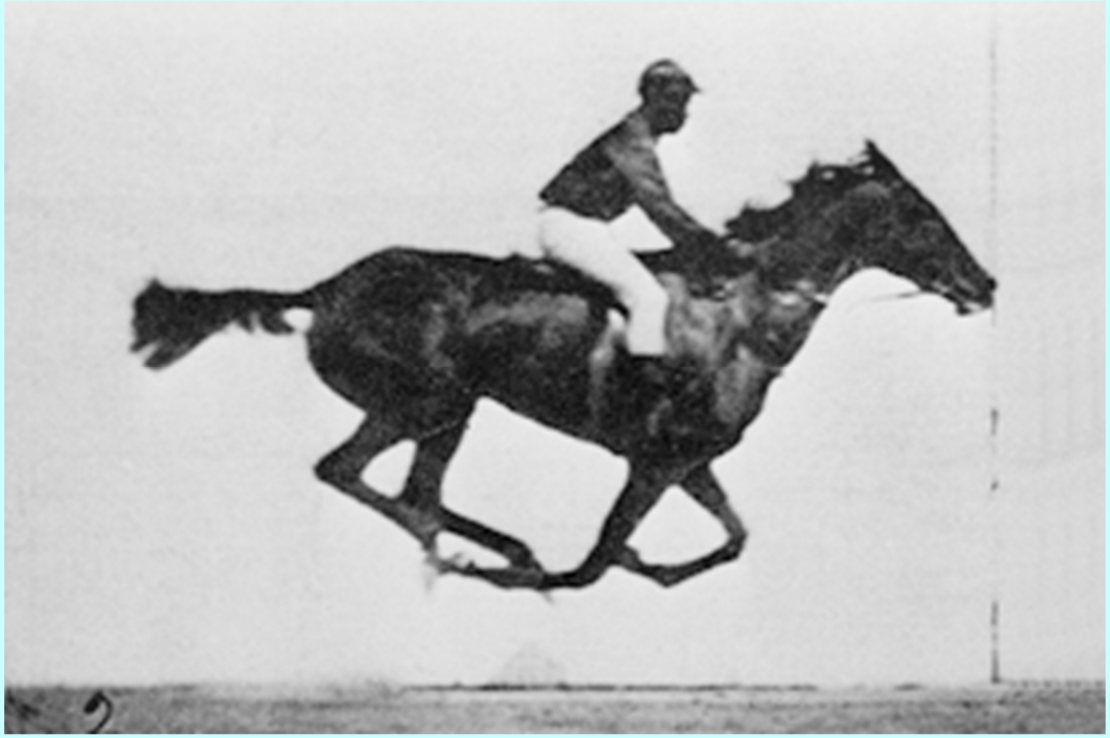
- The **applications** of **digital video** are **immense**:

  - Digital Video Teleconferencing
  - Digital TV
  - High-Definition Television (HDTV)
  - Video Instruction
  - Digital Cinema
  - Wireless and Mobile Video

- A **multibillion-dollar industry** is exploiting new advances in **theory** and **hardware** to make these technologies **publicly available**.

- **Important standards** for **video** exist.

The First Motion Picture



The Horse In Motion (1878)

Leland Stanford and Eadweard Muybridge

In 1872, former **Governor of California Leland Stanford** made a $25,000 this popular question of the day:
whether **all four of a horse's hooves are off the ground** at the same time during the trot!

Stanford believed they were **NOT** calling it “unsupported transit,” and hired famed photographer Muybridge to settle the question.

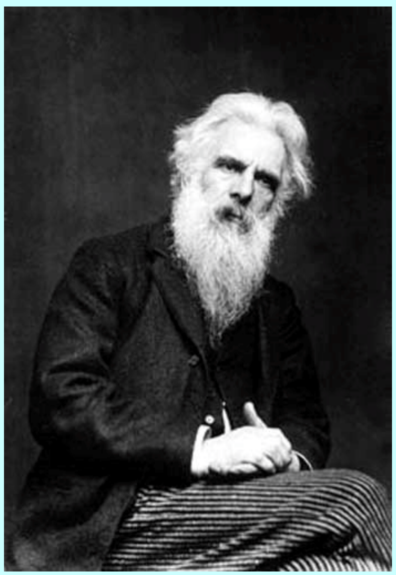
Over the years Muybridge solved the problem by placing **24 cameras at intervals along a racetrack**. These were **triggered in sequence by tripwires** fired by the passing of the racehorse, thus settling the question.

Naturally Stanford (the loser of the bet) and Muybridge has a falling out. Stanford (whom the University is named after), later wrote a book entitled *The Horse in Motion,* but **left all credit to Muybridge out**, even though he used his photographs! **Muybridge sued and lost**.

During all this drama, Muybridge was **acquitted for the murder of his wife’s lover**, one Major Harry Larkyns, on the grounds of “justifiable homicide.” His defense was **paid for by Stanford**.

Muybridge also created the **first movie projector**, called the **Zoopraxiscope**.

He used his devices in later years to study the motions of thousands of humans and animals, and helped to found the sciences of **biomechanics** and the **mechanics of athletics**.



Muybridge

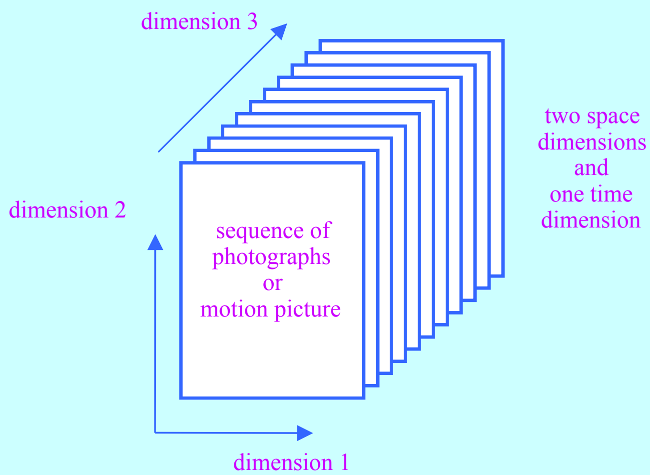
Stanford

Zoopraxiscopes


COPYRIGHT 1893 BY EADW

What is Video?

- Earlier we defined **video** as **time-indexed images**.



- Can **sample**in all three dimensions, yielding **discrete video**. This is always **quantized**, which is **digital video**.

- In principle, **analog video** is continuous in all three dimensions. However, **in practice** analog video is sampled along one space dimension and along the time dimension.

Analog Video

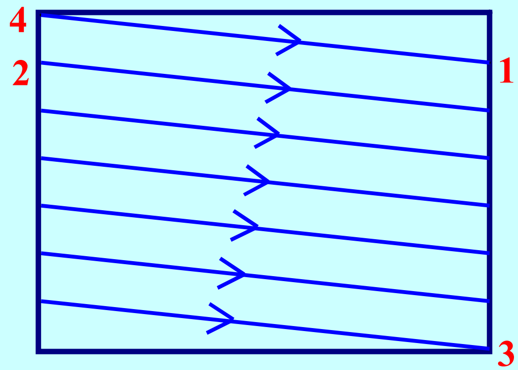
- An **optical** analog video signal is a function $\mathbf{I}_C(x, y, t)$ of space and time.

- Practical video systems, such as **television** and monitors, represent analog video as a **one-dimensional electrical signal** V(t).

- A 1-D signal V(t) is obtained by sampling $\mathbf{I}_C(x, y, t)$ along the vertical (y) direction and along the time (t) direction. This is called **scanning** and the result is a series of **scan lines**.

Progressive Scanning

- A progressive scan traces a complete picture, or **frame**, line-by-line from top-to-bottom, every Δt seconds. For hi-res computer monitors, Δt = 1/72 second.



- After reaching the end of a scan **1**, the electron gun "spot" snaps back to **2**. A blank signal is sent during the snap.
- After reaching the end of a frame **3**, the "spot" snaps back to **4**. A synchronization pulse sent at this time signals the start of a frame.

Refresh Rate

- The **refresh rate** is the **frame rate** at which video is displayed.
- A problem is **flicker**. Flicker occurs when the video display is not **refreshed** sufficiently quickly.
- The **human eye** detects flicker if the refresh rate is less than about **50 frames/second**.
- **Current workstation monitors** exceed this by almost 50%.
- **Such fast refresh rates** are not always possible (w/o losing resolution) - most notably, **old-style analog television**. Too much bandwidth!

Temporal Contrast Sensitivity Function

- **Temporal frequency:** Number of flicker cycles/period.
- In photopic (well-lit) vision, **temporal contrast sensitivity** peaks in the range 8-15 Hz, declines more rapidly at higher than at lower temporal frequencies.
- Frequencies **higher than 50 Hz are undetectable** even at maximum contrast.

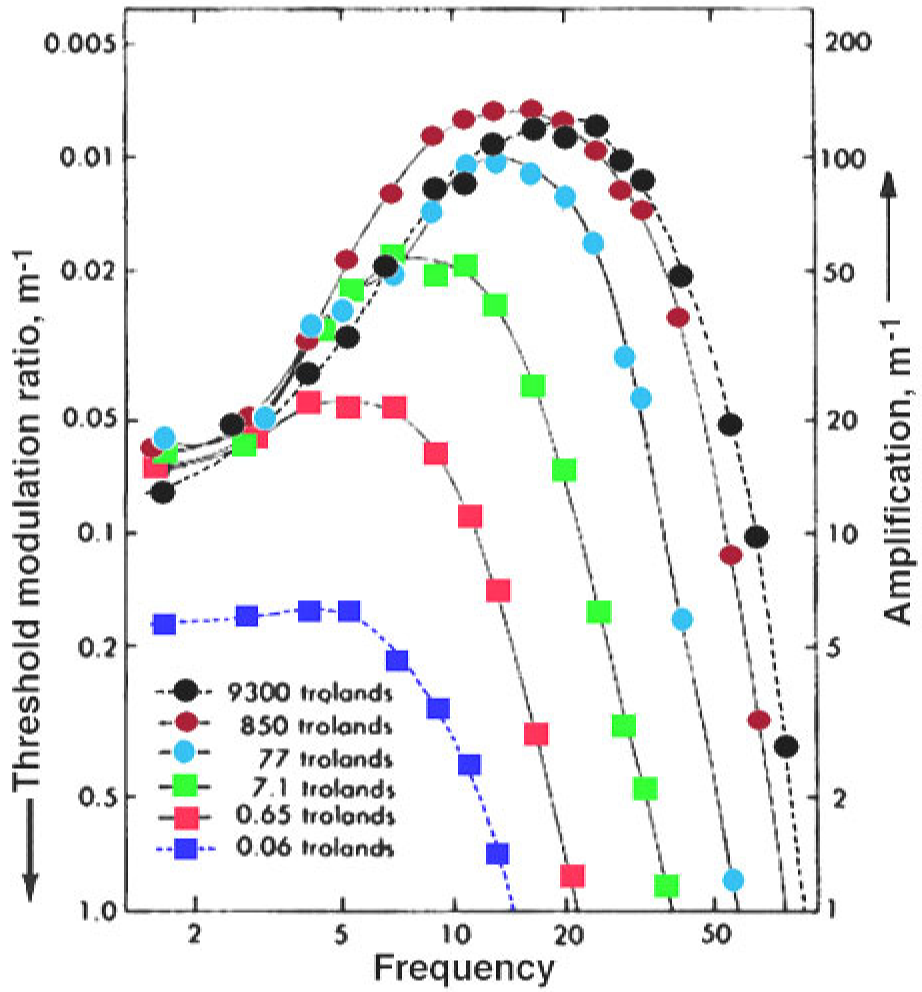
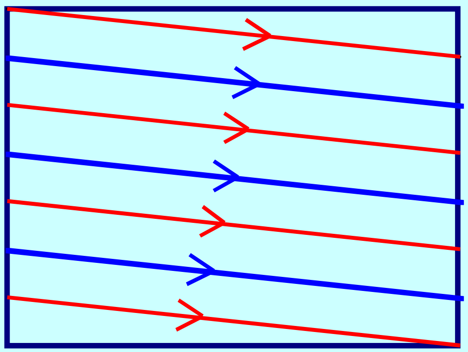


Fig. 11. Temporal Contrast Sensitivity Function (TSF) for various adapting fields. Kelly's data from Hart Jr, W. M., The temporal responsiveness of vision. In: Moses, R. A. and Hart, W. M. (ed) Adler's Physiology of the eye, Clinical Application. St. Louis: The C. V. Mosby Company, 1987.

- **Troland** = unit of retinal illumination (luminance x pupil area)

Interlaced Scanning

- **Interlaced scanning** is a solution to this. In N:1 interlacing, every Nth line is refreshed at each frame refresh. The most common is 2:1 interlacing, as in television:



- In this way, **flicker is reduced** without reducing resolution so much, for a given bandwidth.

Temporal Aliasing

- We won't get into sampling theory again (similar for 3-D as 2-D) but **temporal aliasing** can occur.

- The human eye actually deals with it **exceedingly well**. For example a **fast-moving football** trajectory **looks smooth** to the eye **even if severely aliased**.

- But **temporal aliasing** can produce **ambiguities** and fool the eye. A classic example is a **spinning bicycle wheel** which **appears** to be **spinning** "**the other way**."

Temporal Aliasing

Analog Video Formats

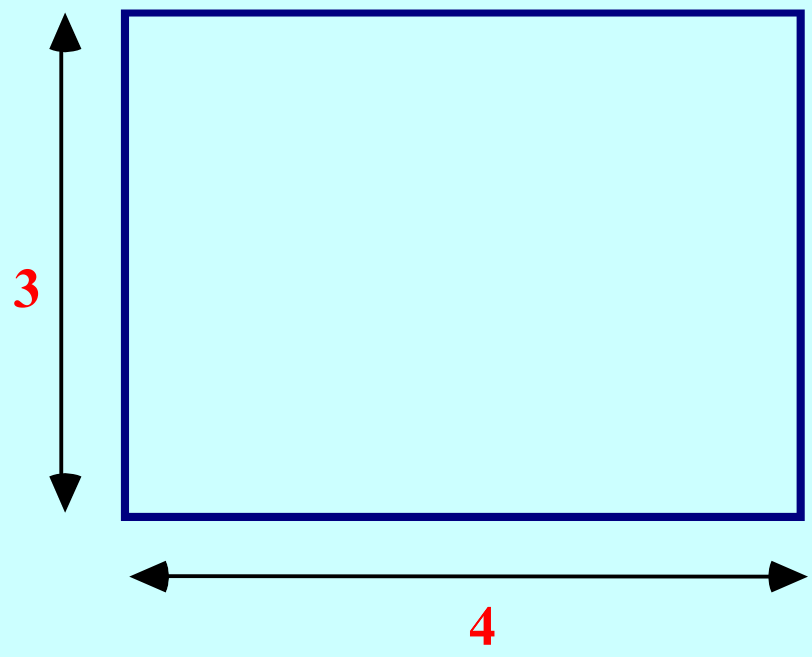
Video Formats

- **NTSC** (National Television Systems Committee):
- 2:1 interlaced
- 525 lines per frame (262.5 / refresh) - 485 active, 40 blanked
- 60 refreshes / second
- Used heavily in Japan and North America
- It's what **your TV** used to use.

- **PAL** (Phase Alternation Line) and **SECAM**
- 2:1 interlaced
- 625 lines per frame
- 50 refreshes per second
- Used heavily in **Europe**
- **Color** handled differently than NTSC

Aspect Ratio

- All of the above standards utilize the same horizontal/vertical **aspect ratio** of 4:3



Color Video

- Any color may be represented as a mixture of Red (R), Green (G) and Blue (B). The **RGB representation** codes color video as three separate signals: R, G, and B.

- The **YIQ representation** combines the information according to perceptual criteria:

  $$Y = \text{luminance} = 0.299R + 0.587G + 0.114B$$

  chrominance signals:

  $$I = 0.596R + 0.275G - 0.321B$$

  $$Q = 0.212R - 0.523G + 0.311B$$

- Cr-Cb: an **alternate** chrominance signal representation:

  $$Cr = R - Y$$

  $$Cb = B - Y$$

  This is used in modern video codecs.

- Why use Y-I-Q or Y-Cr-Cb? **Reduced bandwidth**. **Chrominance information** can be sent in a **fraction** of the bandwidth of luminance **information**.

Older Digital Video Formats

- **Modern** color video cameras supply RGB outputs - separately digitized in space and time.

  I = R + G + B

  Digital storage & **bandwidth** (BW) are a huge consideration.

Current Digital Video Standards:

- **CCIR 601 NTSC**
- 8 bits / pixel
- 720 pixels / line luma, 360 pixels / line chroma (Cr, Cb)
- 480 lines / frame both luma & chroma
- 2:1 interlacing
- 60 frames / second
- Luminance BW = (8 x 720 x 480 x 60) / 2 ≈ **82.9 Mbps**
- Chroma BW = (2 x 8 x 360 x 480 x 60) / 2 ≈ **82.9 Mbps**

Total BW ≈ **165.9 Mbps**

- **CCIR 601 PAL/SECAM**
- 8 bits / pixel
- 720 pixels / line luma, 360 pixels / line chroma (Cr, Cb)
- 576 lines / frame both luma & chroma
- 2:1 interlacing
- 50 frames / second
- Luminance BW = (8 x 720 x 576 x 50) / 2 ≈ **82.9 Mbps**
- Chroma BW = (2 x 8 x 360 x 576 x 50) / 2 ≈ **82.9 Mbps**

Total BW ≈ **165.9 Mbps**

Older Digital Video Formats

- **Common-Intermediate Format (CIF) Standard** (H.261, H.264)
- 8 bits / pixel
- 360 pixels / line luma, 180 pixels / line chroma (Cr, Cb)
- 288 lines / frame luma, 144 lines / frame chroma
- 1:1 interlacing (**progressive**)
- 30, 15, 10, or 7.5 frames / second
- Luminance BW = (8 x 360 x 288 x 30) / 1 ≈ **24.9 Mbps**
- Chroma BW = (2 x 8 x 180 x 144 x 30) / 1 ≈ **12.4 Mbps**

Total BW ≈ **37.3 Mbps**

- **Low-Bitrate CIF (QCIF)** (H.261, H.264)
- 8 bits / pixel
- 180 pixels / line luma, 90 pixels / line chroma (Cr, Cb)
- 144 lines / frame luma, 72 lines / frame chroma
- 1:1 interlacing (**progressive**)
- 30, 15, 10, or 7.5 frames / second
- Luminance BW = (8 x 180 x 216 x 30) ≈ **6.2 Mbps**
- Chroma BW = (2 x 8 x 90 x 72 x 30) / 1 ≈ **3.1 Mbps**

Total BW ≈ **9.3 Mbps**

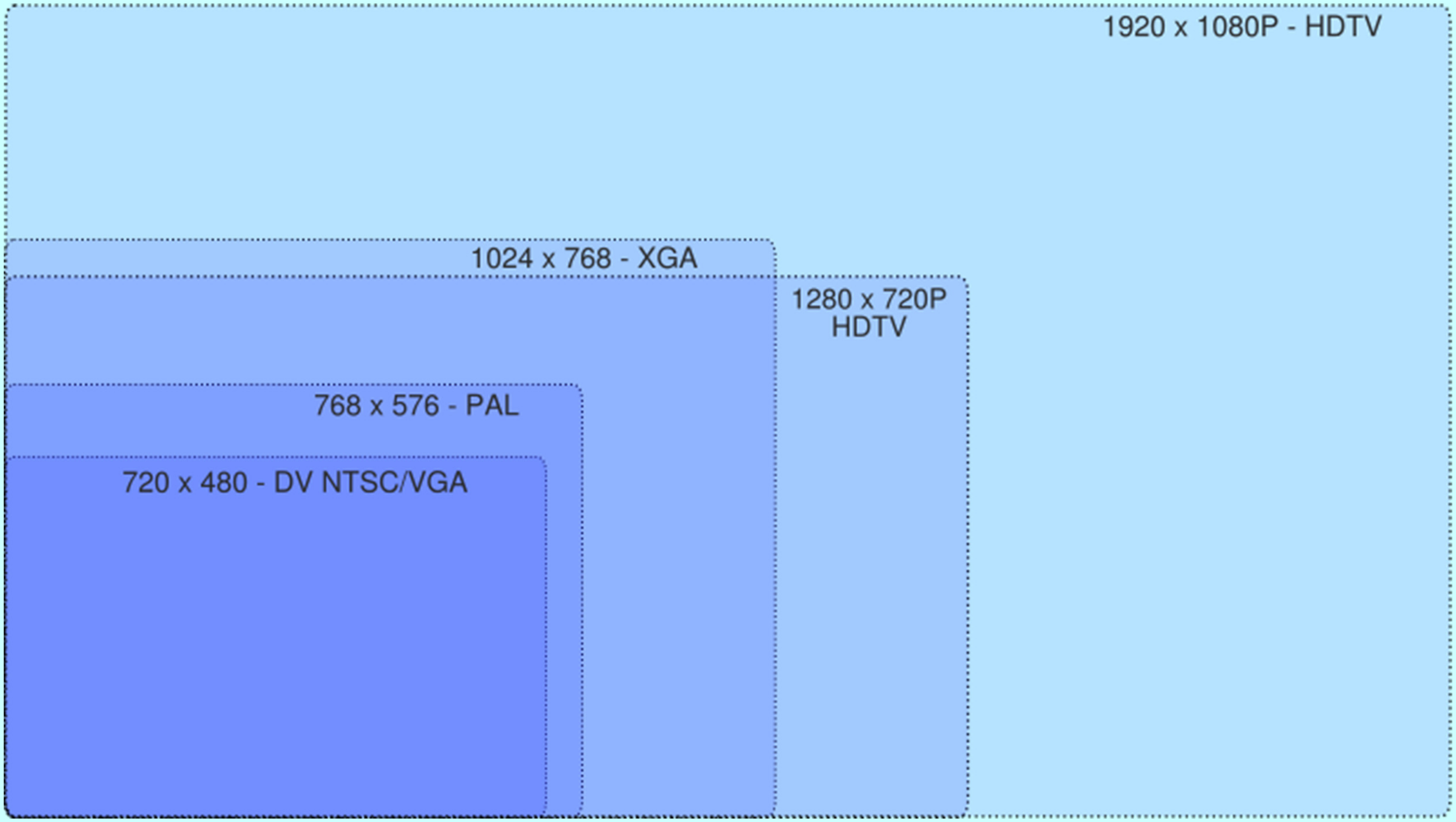
- **Standard Input Format (SIF) Standard** (MPEG-1,2)
- 8 bits / pixel
- 352 pixels / line luma, 176 pixels / line chroma (Cr, Cb)
- 240 lines / frame luma, 120 lines / frame chroma
- 1:1 interlacing (**progressive**)
- 30 frames / second
- Luminance BW = (8 x 352 x 240 x 30) ≈ **20.3 Mbps**
- Chroma BW = (2 x 8 x 176 x 120 x 30) / 1 ≈ **10.1 Mbps**

Total BW ≈ **30.4 Mbps**

HDTV Standard Format (Raw)

- Has **interlaced** and **progressive** modes.
- **720p:** progressive, 1280 × 720 pixels, 60 frames per second. Raw BW (24 bits/pixel): **1.3Gbps**
- **1080i**: interlaced encoding, 1920 × 1080 pixels, 50 fields (25 frames) per second. Raw BW (24 bits/pixel): **1.2Gbps**
- **1080p**: progressive, 1920 × 1080 pixels, 59.94 frames per second. Raw BW (24 bits/pixel): **2.98Gbps**
- **Aspect ratio** is 16:9 instead of 4:3.
- Typically **compressed**: MPEG-2 or MPEG-4 AVC (later!)

Comparison of Standard Formats



Common Video File Formats

- **.mov**: standard Apple Quicktime media file container – video (h.26x, mpeg1, 2, 4) and audio
- **.avi**: standard Microsoft media file container – video (h.26x, mpeg1, 2, 4) and audio
- **.mpg**: video in MPEG-1 or MPEG-2 format
- **.mp4:** video in MPEG-4 format. Also in .mov or .avi.
- **.m1v:** video-only MPEG-1 format (no audio)
- **.m2v:** video-only MPEG-2 format (no audio)

Current Network Capacities

- Conventional telephone: 56 kbps
- ISDN: p x 64 kbps, $1 \le p \le 30$ (Integrated Services Digital Network)
- T-1, HDSL dedicated lines: 1.5 Mbps
- Ethernet: Up to 10Gbps currently
- Future Wireless: 3G 1-10Mbps, 4G 100Mbps+

The Need for Compression

- Video data rates severely test the capabilities of communication networks, so **video compression** is of extreme interest.

- There exist a variety of **standards** for video compression which we'll discuss later:

  - H.26x: H.261, H.263, H.264
  - MPEG-1, MPEG-2, MPEG-4

Standards Conversion (Transcoding)

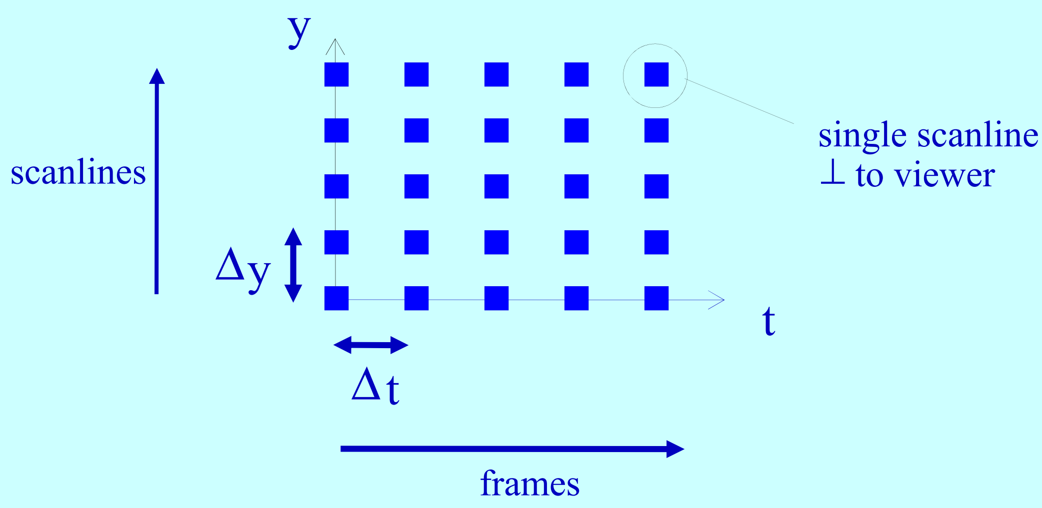
- With all of the existing/emerging video standards, there are a lot of methods for **converting** between them (**transcoding**).

- Because of the wide diversity of codecs, we won't get into this, but it's good to be aware of it.

Analog Video Sampling

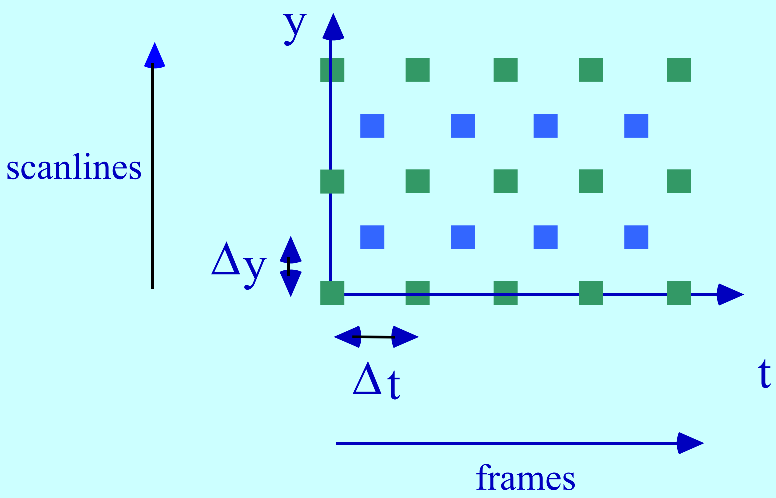
- A continuous 3-D space-time **intensity distribution** I(x, y, t) is **sampled** in one space dimension (y or vertical) and the time dimension. This is called **scanning**.

- **Progressive Analog Video** involves sampling row after row at intervals Δy and each frame at intervals Δt:



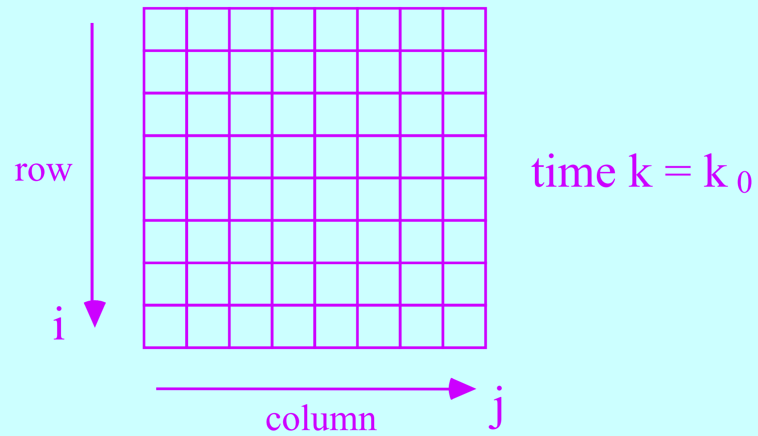
Analog Video Sampling

- **Interlaced Analog Video** involves sampling even and odd rows alternately:



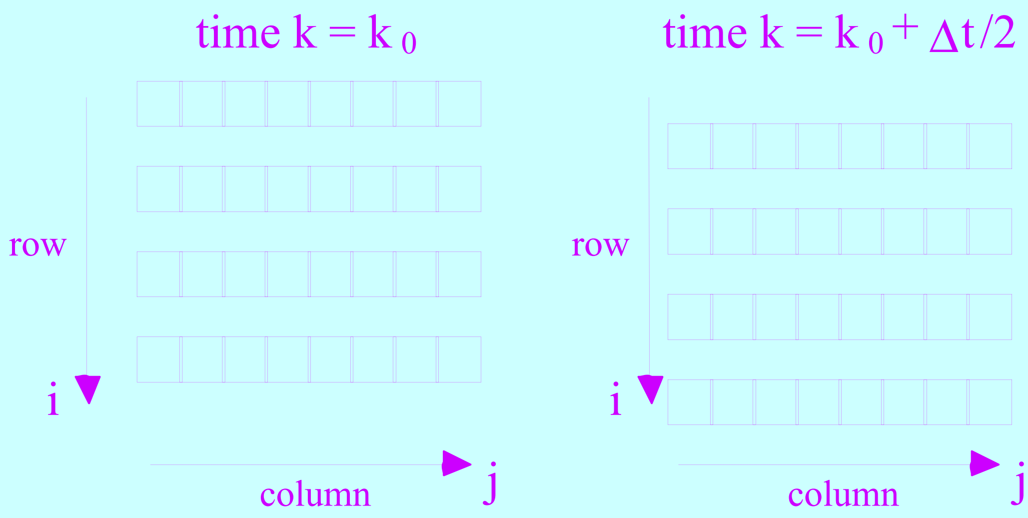
Digital Video Sampling

- Digital video is obtained either by sampling an analog video signal or by directly sampling the 3-D intensity distribution.
- If **progressive analog video** is sampled, **or** if digital video is **directly sampled** (e.g., via CCD array), then the sampling is **rectangular** and properly indexed:



Digital Video Sampling

- If **interlaced analog video** is sampled, then the digital is interlaced also and must be re-indexed:



OPTICAL FLOW

- Much of the remainder of this Module will deal with concepts of **motion**.
- Fundamental to the concept of motion, but nevertheless different, is **optical flow**.
- Optical flow is the **instantaneous motion of image intensities**. This is **not** the same as the motion of the objects being imaged:

image motion is not object motion

Optical Flow

- **Examples**:

  - An "off-camera" variable light source (which may be moving) illuminating a **stationary object**. A case of **optical flow without object motion**

  - A mirrored sphere that is spinning. A case of **object motion without** image motion (optical flow).

- **Still,** optical flow is all the motion information that the image supplies! So, most methods of **motion estimation**, **motion compensation**, etc. depend on it.

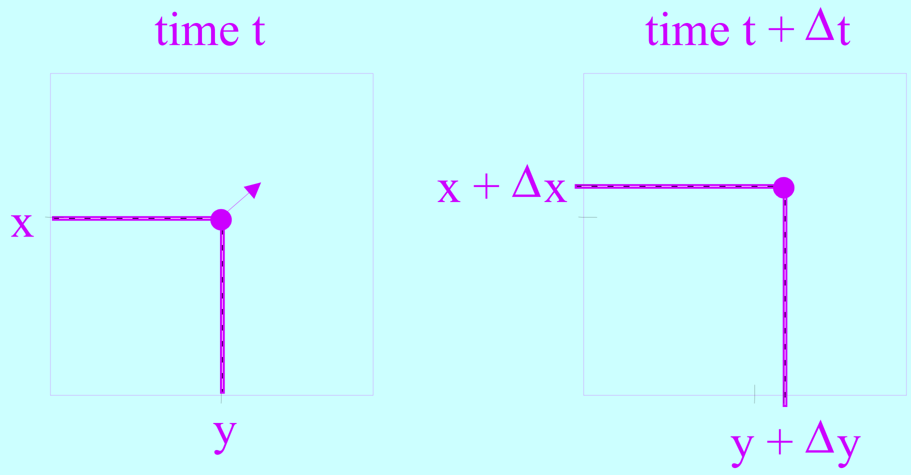
Continuous Formulation

- The image intensity at a point in space and time is I(x, y, t).

- After a sufficiently **small** time interval Δt, the intensity at (x, y) will move to a point $(x + \Delta x, y + \Delta y)$. In other words:

$$I(x + \Delta x, y + \Delta y, t + \Delta t) = I(x, y, t)$$

as depicted:



- This assumes that the intensity does not change, just its position.

Taylor Expansion

- Expanding the LHS in a Taylor's series:

$$I(x + \Delta x, y + \Delta y, t + \Delta t)$$

$$= I(x, y, t) + \Delta x \cdot \frac{\partial I}{\partial x} + \Delta y \cdot \frac{\partial I}{\partial y} + \Delta t \cdot \frac{\partial I}{\partial t} + \text{h.o.t.}$$

so that

$$I(x, y, t) + \Delta x \cdot \frac{\partial I}{\partial x} + \Delta y \cdot \frac{\partial I}{\partial y} + \Delta t \cdot \frac{\partial I}{\partial t} + \text{h.o.t.} = I(x, y, t)$$

- Letting the h.o.t. = 0 (assuming small time and motion), cancelling the term I(x, y, t) and dividing through by Δt:

$$\frac{\Delta x}{\Delta t}\frac{\partial I}{\partial x} + \frac{\Delta y}{\Delta t}\frac{\partial I}{\partial y} + \frac{\partial I}{\partial t} = 0$$

Optical Flow Constraint Equation

- Taking the limit $\Delta t \rightarrow 0$ yields:

$$\frac{\partial I}{\partial x}\frac{\partial x}{\partial t}+\frac{\partial I}{\partial y}\frac{\partial y}{\partial t}+\frac{\partial I}{\partial t}=0$$

- The **optical flow** components are:

$$u(x, y, t) = \frac{\partial x}{\partial t}(x, y, t) \text{ and } v(x, y, t) = \frac{\partial y}{\partial t}(x, y, t)$$

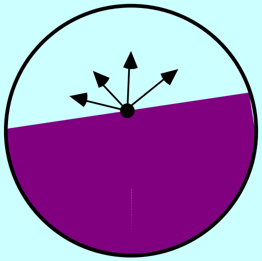
- Putting this together gives the **optical flow constraint equation** or **OFCE**:

$$I_x u + I_y v + I_t = 0.$$

- So-called since it does not solve for optical flow - it only **constrains** the optical flow vector (u, v) to lie on a line.

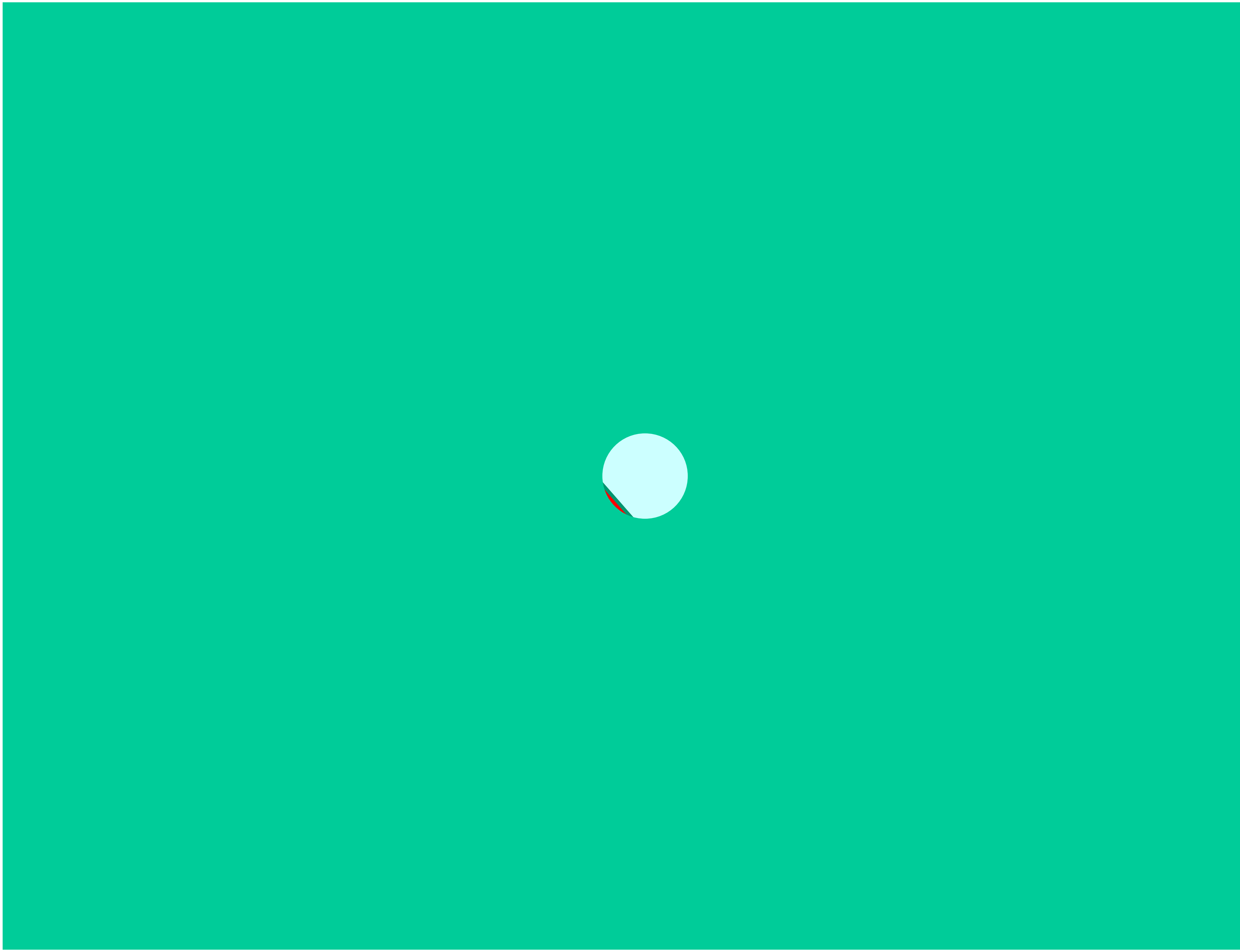
The Aperture Problem

- Even knowing I_x, I_y, I_t, the OFCE does not solve optical flow.
- This is the **aperture problem**.
- Imagine being able to view only a **small region** of the image that is in motion:



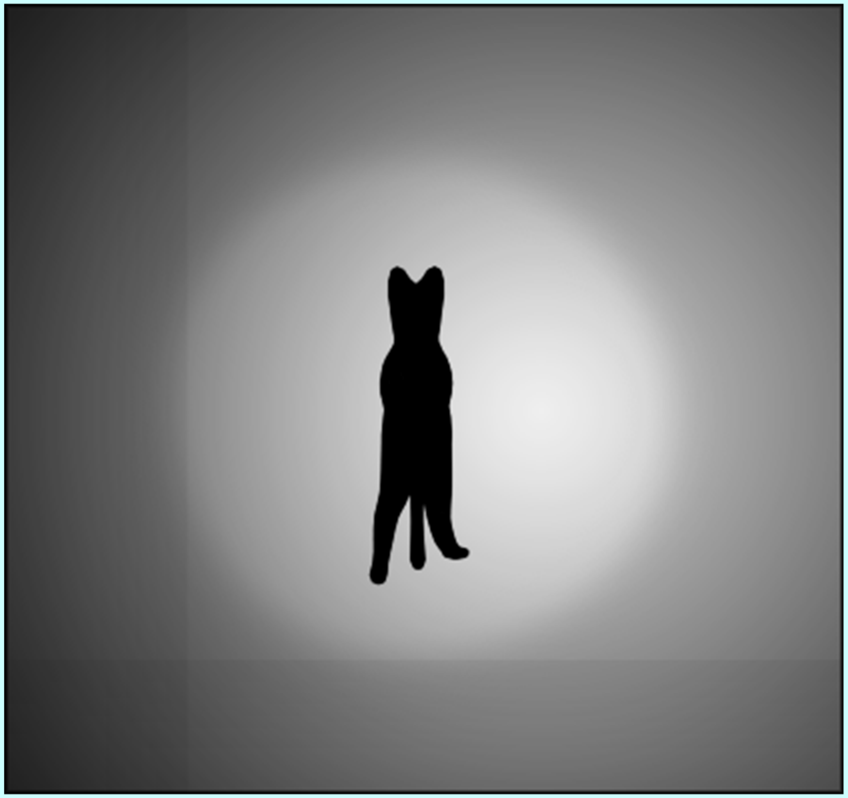
The Aperture Problem

- If the edge is sensed to be moving "up," the true motion could actually be in any of the indicated directions.
- In order to solve for optical flow, some other **physically meaningful** constraint(s) must be found or assumed.

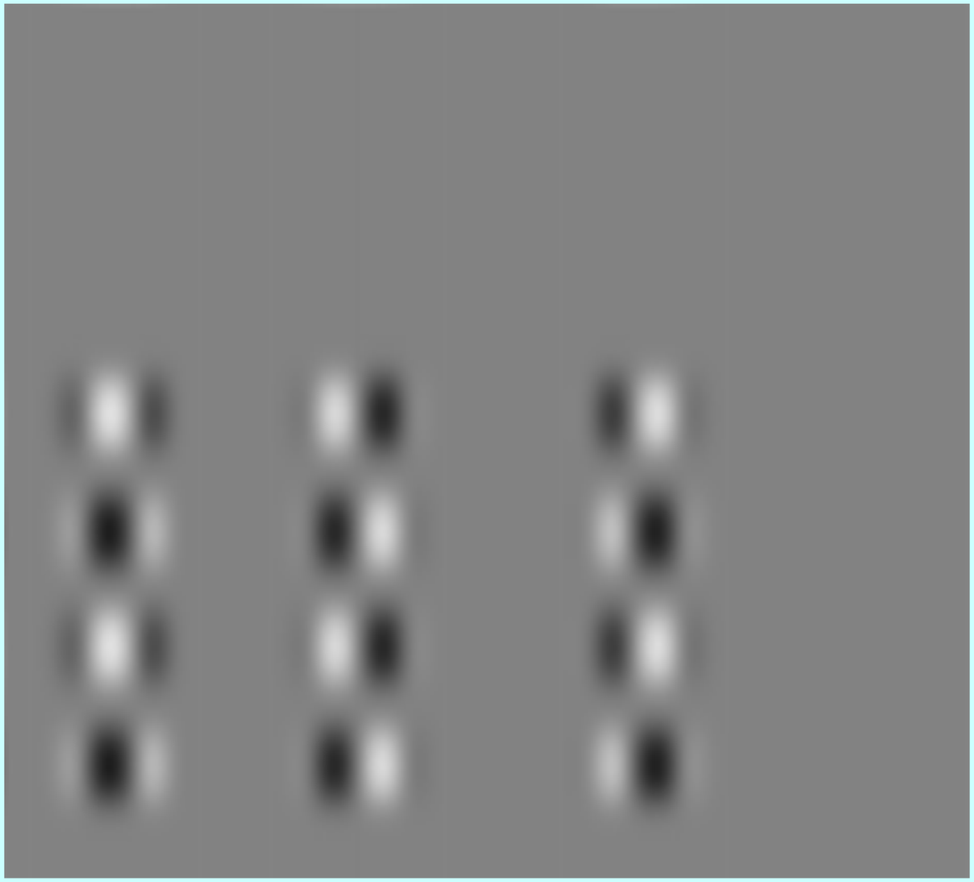


The Barber's Pole

A Bistable Motion Illusion



A Motion Direction Illusion



Smooth Optical Flow

- The assumption that is usually made is that optical flow is **smooth**. Smooth in the sense that the derivatives of u and v have small magnitudes.

- Solution involves minimizing the overall **departure** from smoothness:

$$E_{smooth} = E_s = \iint_{image} \left[u_x^2 + u_y^2 + v_x^2 + v_y^2 \right] dx\, dy$$

- We also want the overall OFCE error to be small:

$$E_{constraint} = E_c = \iint_{image} \left[I_x u + I_y v + I_t \right]^2 dx\, dy$$

A Minimization Problem

- Minimize the **weighted sum**:

$$E = E_s + \lambda E_c$$

- A solution to this will always **exist** and be **unique**.
- For λ larger, the solution will more closely track the OFCE.
- For λ smaller, the solution will be forced smoother.
- Picking λ is a **hard problem** - not discussed here (Cross-Validation).
- We will not try to solve the **continuous** problem.

Discrete Optical Flow

- **Approximations to derivatives** of flow:

$$u_x \approx [u(i+1, j) - u(i, j)]/2 \qquad u_y \approx [u(i, j+1) - u(i, j)]/2$$

$$v_x \approx [v(i+1, j) - v(i, j)]/2 \qquad v_y \approx [v(i, j+1) - v(i, j)]/2$$

- Then:

$$E_s = \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} \Big\{ [u(i+1, j) - u(i, j)]^2 + [u(i, j+1) - u(i, j)]^2 + [v(i+1, j) - v(i, j)]^2 + [v(i, j+1) - v(i, j)]^2 \Big\}$$

and

$$E_c = \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} [I_x(i, j)u(i, j) + I_y(i, j)v(i, j) + I_t(i, j)]^2$$

where estimates $I_x(i, j)$, $I_y(i, j)$, $I_t(i, j)$ will be discussed shortly.

Intensity Gradient Estimation

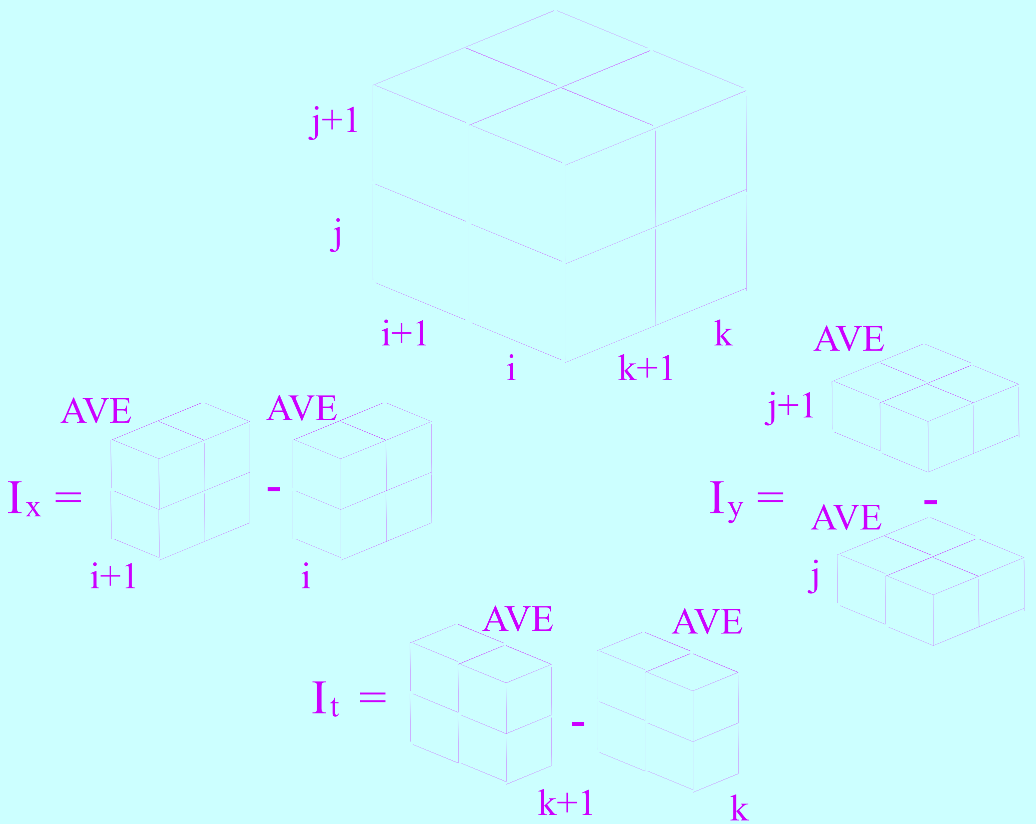
- The derivatives I_x, I_y, I_t can also be estimated as **differences-of-averages** across a 2 x 2 x 2 data cube:

$$I_x \approx [I(i+1,j,k) + I(i+1,j,k+1) + I(i+1,j+1,k) + I(i+1,j+1,k+1)] - [I(i,j,k) + I(i,j,k+1) + I(i,j+1,k) + I(i,j+1,k+1)]$$

$$I_y \approx [I(i,j+1,k) + I(i,j+1,k+1) + I(i+1,j+1,k) + I(i+1,j+1,k+1)] - [I(i,j,k) + I(i,j,k+1) + I(i+1,j,k) + I(i+1,j,k+1)]$$

$$I_t \approx [I(i,j,k+1) + I(i,j+1,k+1) + I(i+1,j,k+1) + I(i+1,j+1,k+1)] - [I(i,j,k) + I(i,j+1,k) + I(i+1,j,k) + I(i+1,j+1,k)]$$

Intensity Gradient Estimates



Discrete Optimization

- The goal is to **minimize**

$$E = E_s + \lambda E_c$$

- Take derivatives of E w.r.t. u(m, n), v(m, n) for $0 \le m \le N\text{-}1$, $0 \le n \le M\text{-}1$:

$$\frac{\partial E}{\partial u(m,n)} = 2[u(m,n) - u_{ave}(m,n)] + 2\lambda\,[I_x u(m,n) + I_y v(m,n) + I_t]\cdot I_x$$

$$\frac{\partial E}{\partial v(m,n)} = 2[v(m,n) - v_{ave}(m,n)] + 2\lambda\,[I_x u(m,n) + I_y v(m,n) + I_t]\cdot I_y$$

where the **local 4-averages**:

$$u_{ave}(m,n) = \frac{1}{4}[u(m+1,n) + u(m-1,n) + u(m,n+1) + u(m,n-1)]$$

$$v_{ave}(m,n) = \frac{1}{4}[v(m+1,n) + v(m-1,n) + v(m,n+1) + v(m,n-1)]$$

Discrete Solution

- The minima occur where the derivatives are zero:

$$\frac{\partial E}{\partial u(m, n)} = \frac{\partial E}{\partial v(m, n)} = 0$$

yielding linear equations:

$$(1 + \lambda I_x^2)u(m, n) + \lambda I_x I_y v(m, n) = u_{ave}(m, n) - \lambda I_x I_t$$

$$\lambda I_x I_y u(m, n) + (1 + \lambda I_y^2)v(m, n) = v_{ave}(m, n) - \lambda I_y I_t$$

- Solving for u(m, n) and v(m, n) yields:

$$u(m, n) = u_{ave}(m, n) - \lambda \frac{I_x u_{ave}(m,n) + I_y v_{ave}(m,n) + I_t}{1 + \lambda\left(I_x^2 + I_y^2\right)} \cdot I_x$$

$$v(m, n) = v_{ave}(m, n) - \lambda \frac{I_x u_{ave}(m,n) + I_y v_{ave}(m,n) + I_t}{1 + \lambda\left(I_x^2 + I_y^2\right)} \cdot I_y$$

Iterative Solution

- The solution for u(m, n) and v(m, n) suggests a numerical algorithm for actually computing them. The **relaxation algorithm** is:

$$u^{(p+1)}(m, n) = u_{ave}^{(p)}(m,n) - \lambda \frac{I_x u_{ave}^{(p)}(m,n) + I_y v_{ave}^{(p)}(m,n) + I_t}{1+\lambda\left(I_x^2 + I_y^2\right)} \cdot I_x$$

$$v^{(p+1)}(m, n) = v_{ave}^{(p)}(m,n) - \lambda \frac{I_x u_{ave}^{(p)}(m,n) + I_y v_{ave}^{(p)}(m,n) + I_t}{1+\lambda\left(I_x^2 + I_y^2\right)} \cdot I_y$$

- This technique of computing a "new" estimate from "old" estimates is a common technique in numerical analysis called **successive refinement**.

Initial Estimates

- The **initial estimates** $u^{(0)}(m, n)$, $v^{(0)}(m, n)$ might be taken from some **independent estimate** of u, v, or simply by taking

$$u^{(0)}(m, n) = v^{(0)}(m, n) = 0$$

in which case

$$u^{(1)}(m, n) = -\lambda \frac{I_t I_x}{1 + \lambda\left(I_x^2 + I_y^2\right)}$$

$$v^{(1)}(m, n) = -\lambda \frac{I_t I_y}{1 + \lambda\left(I_x^2 + I_y^2\right)}$$

Iteration Limit

- The **iterations** are continued either:

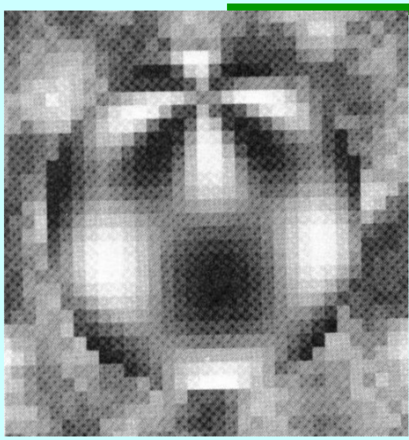
  - for a prescribed number P of iterations
  - until iterating doesn't change the solution much, e.g.,

$$\max_{(m,n)} |\, u^{(p+1)}(m, n) - u^{(p)}(m, n)| < TOL$$

  where TOL is a tolerance threshold.

- Although **in principle** it could take N iterations for the constraints to propagate across the image domain, **in practice** it takes just a few iterations due to the **localness** of image motion.

Example Four frames of a Rotating Sphere

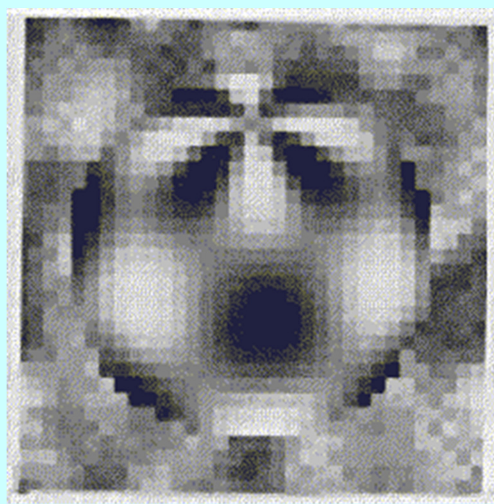


$t = t_0$

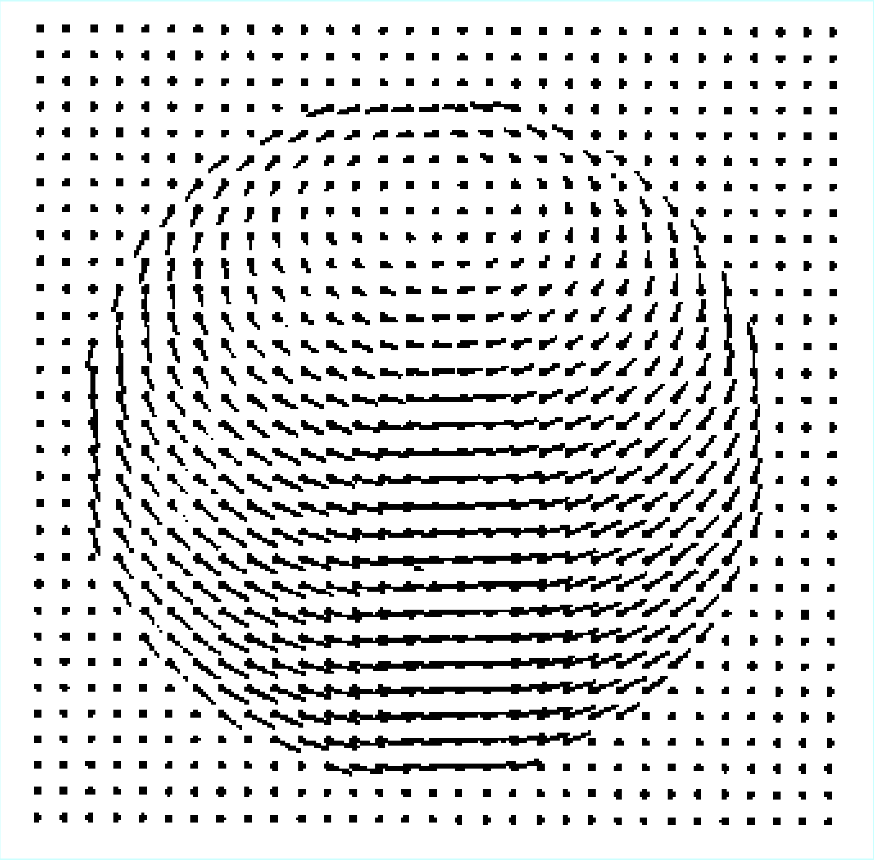
$t = t_0 + \Delta t$

$t = t_0 + 2\Delta t$

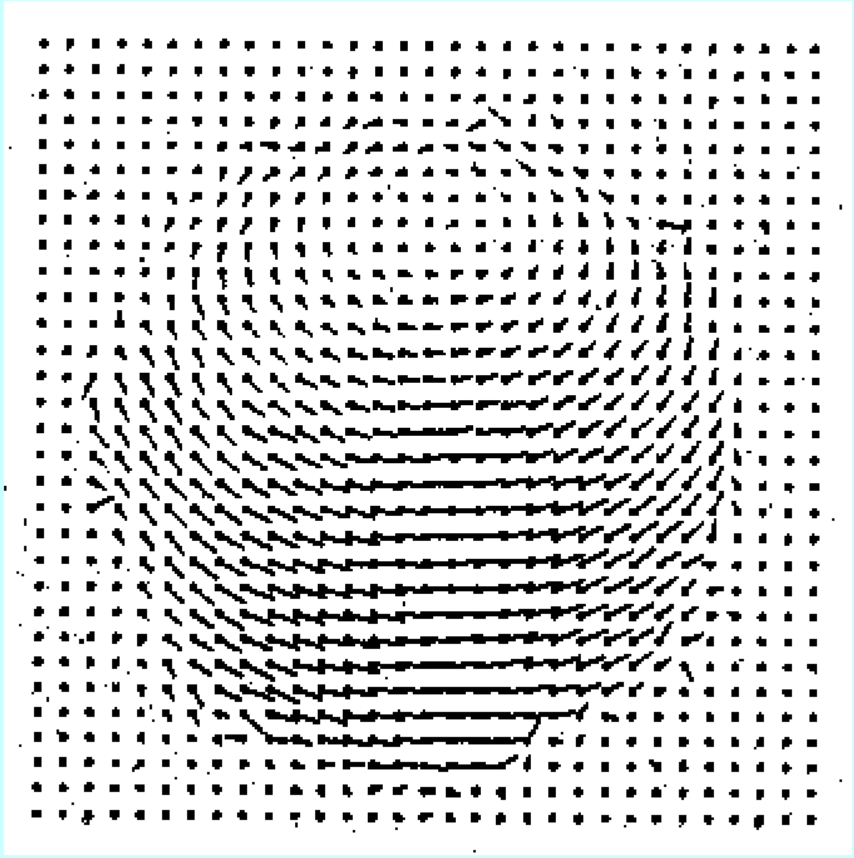
$t = t_0 + 3\Delta t$



Example Computed Flow



Actual optical flow field



Computed flow field – 32 iterations

Needle diagram: Arrow direction indicates flow direction
Arrow length indicates flow magnitude.

Optical Flow Results

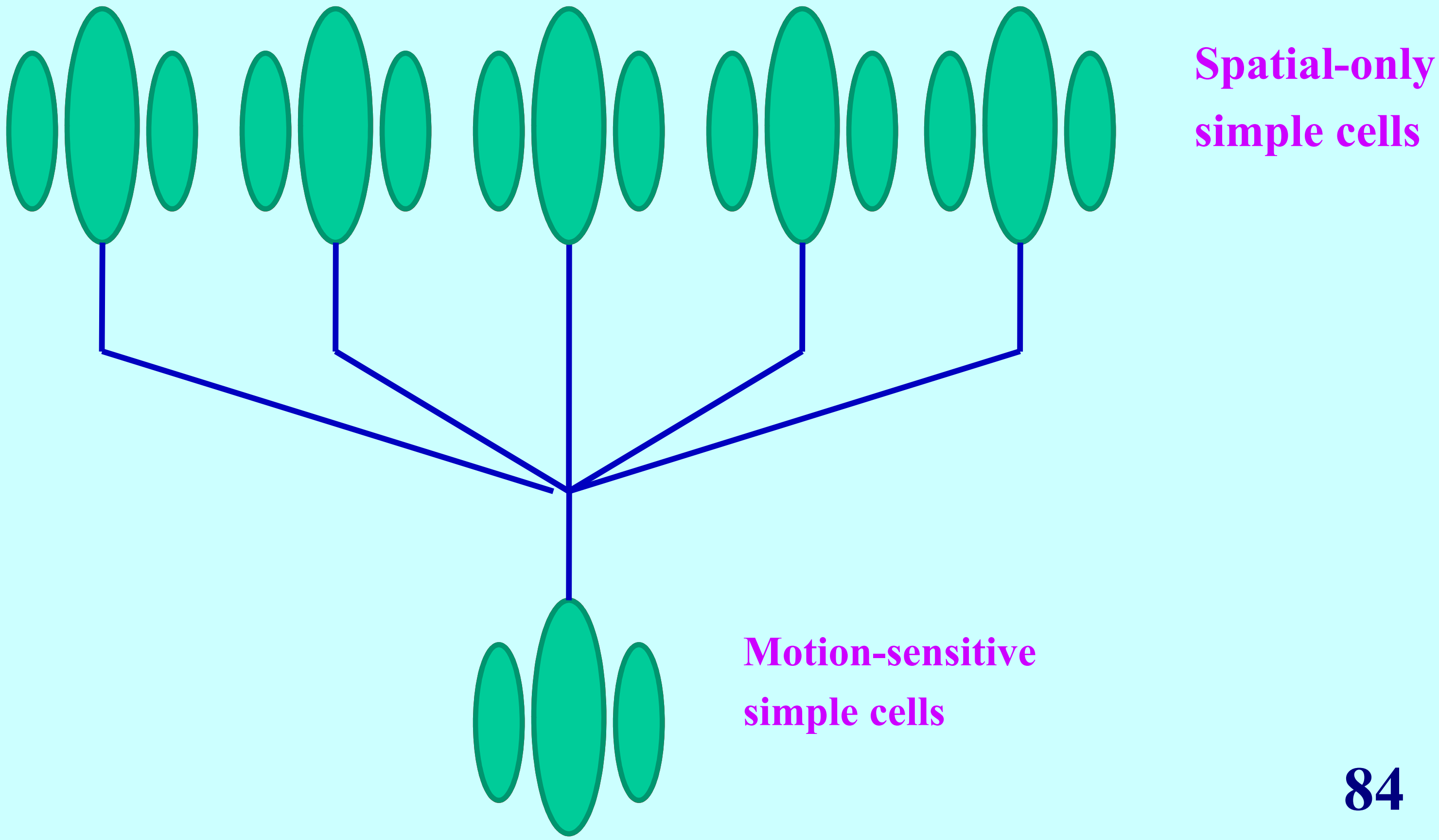
- Results are accurate in most places but **errors occur near the sphere boundary**.
- Errors occur near **flow discontinuities** - the smoothness constraint is inaccurate.
- This is the **Horn-Schunk Algorithm**, the first and still classic approach.
- Many more sophisticated techniques exist, e.g. that attempt to **find** flow **discontinuities**, then **disable** the **smoothness** constraint there.

Back to the visual brain…

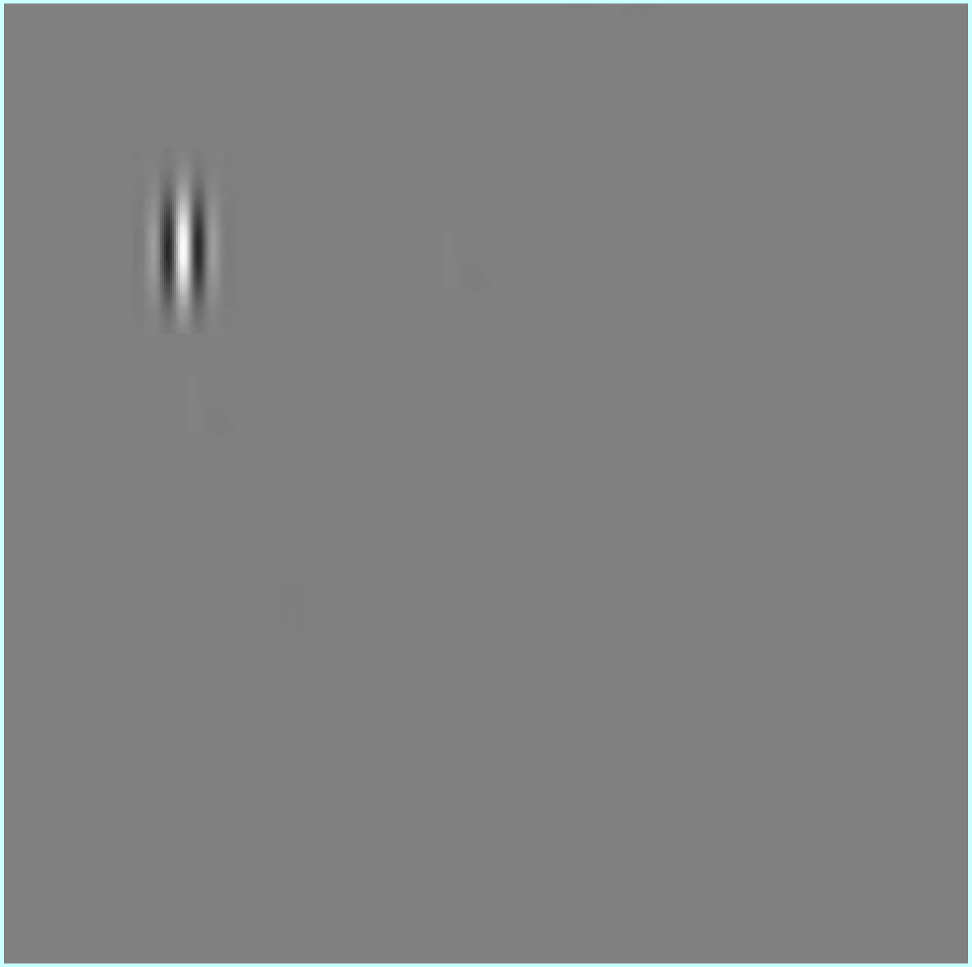
Simple Cell Temporal Responses

- **Motion** is perhaps the most important part of our **visual experience**.
- A small percentage of **simple** and **complex cells** are **responsive** to **motion.**
- **Motion sensitive simple cells** may use input from **multiple spatial simple cells.**
- The responsiveness of such simple cells appears to be largely **space-time separable.**

Plausible Motion Stimulus



Moving Patch Experiments



- Place subject in front of **moving pattern**
- Vary **speed, spatial frequency,** and **orientation**
- **Record responses** of **simple** and **complex cells**

Spatio-Temporal Filter Model

- A **separable spatio-temporal filter:**

$$f(t)g(x, y)$$

- In the brain, a **causal filter is perhaps desirable.** Several models exist.

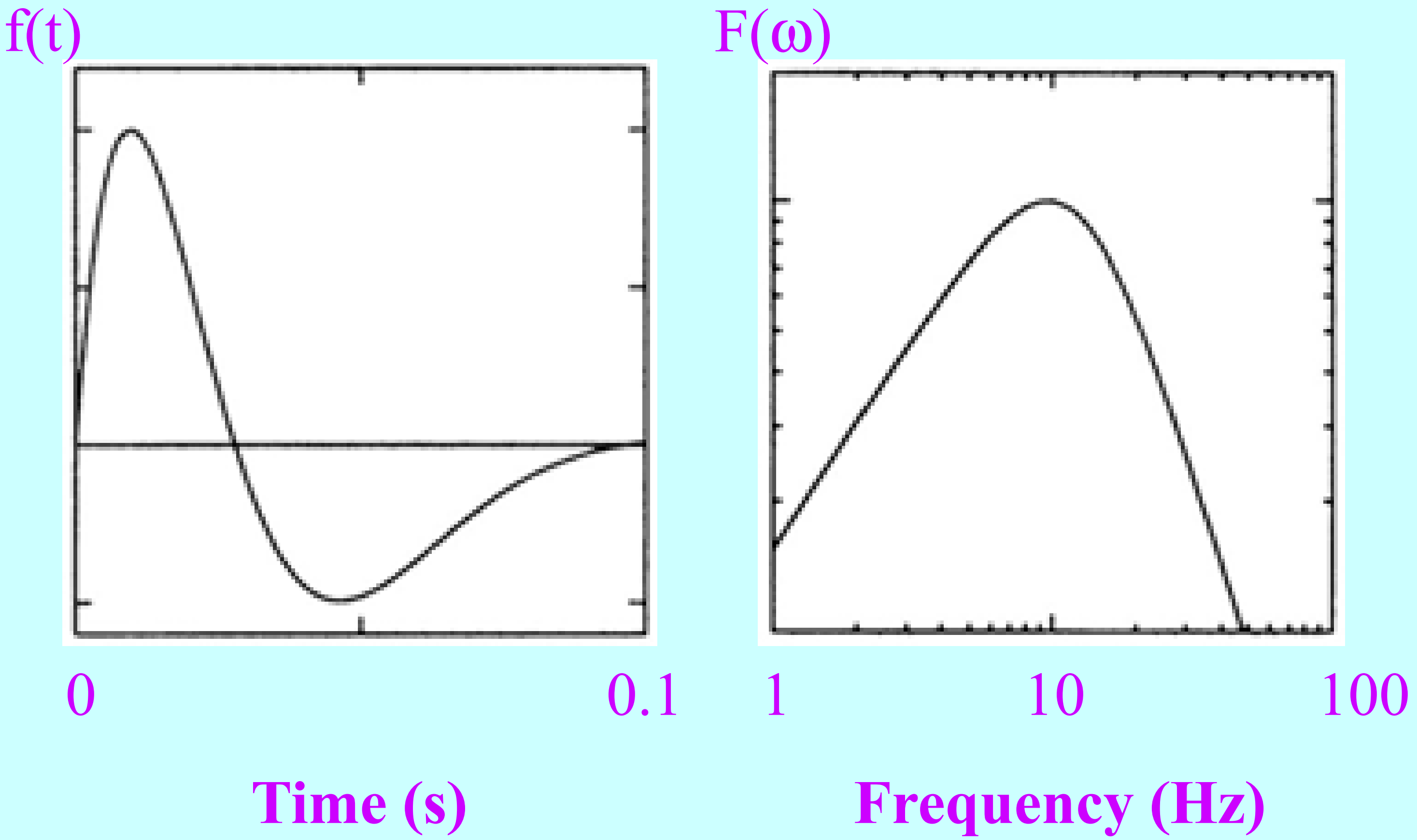
- Common causal filter model: **causal gamma filter**

Causal Temporal Model

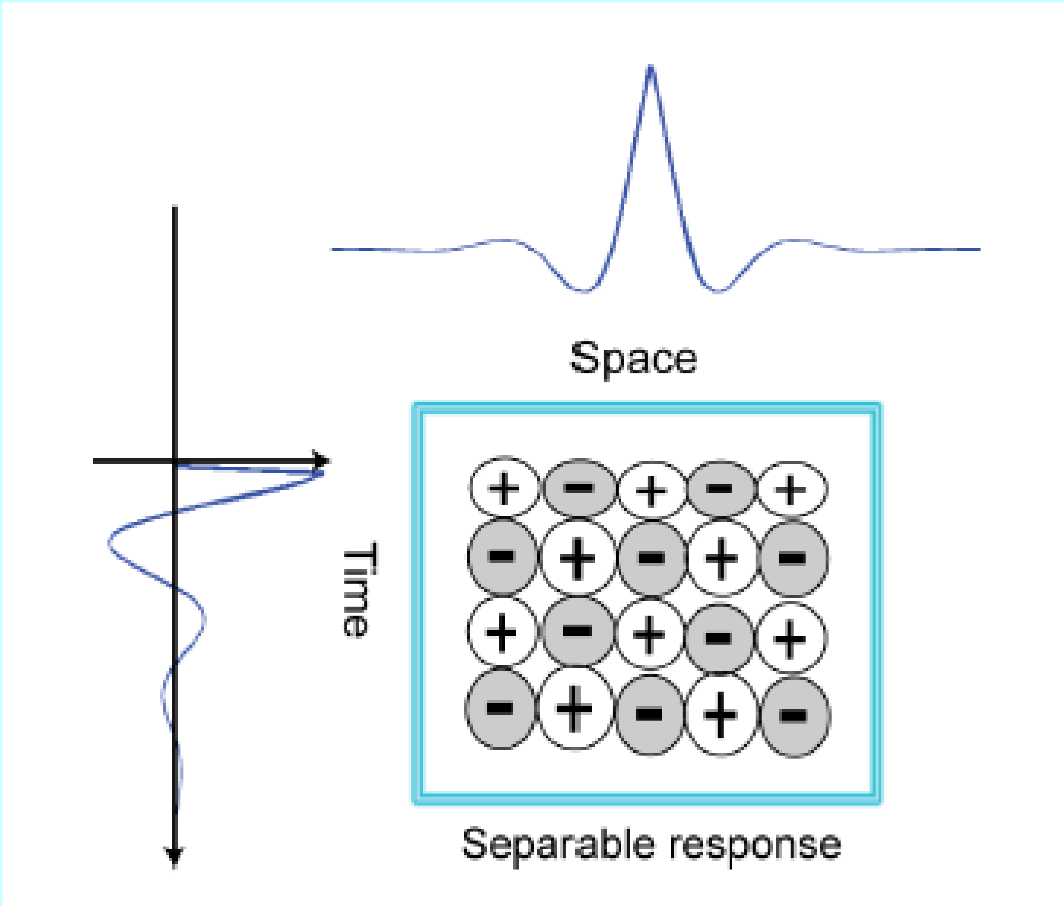
$$f(t)=\begin{cases} te^{-t/\tau}e^{\sqrt{-1}\omega_0 t} & ; t\geq 0 \\ 0 & ; t<0 \end{cases}$$

$$F(\omega)=K\left\{\frac{1}{\left[(1/\tau)+\sqrt{-1}\left(\omega-\omega_0\right)\right]^2}+\frac{1}{\left[(1/\tau)+\sqrt{-1}\left(\omega+\omega_0\right)\right]^2}\right\}$$

Bandpass Causal Temporal Model



Space-Time Receptive Field Response



Three-Dimensional Gabor Model

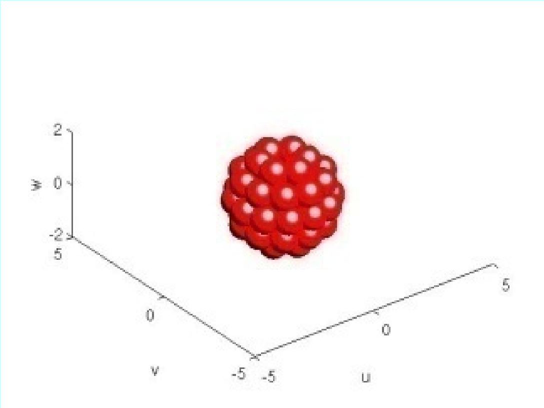
- **Causality** not needed if real-time processing not required.

- **Separable time** and **space Gabor filters:**

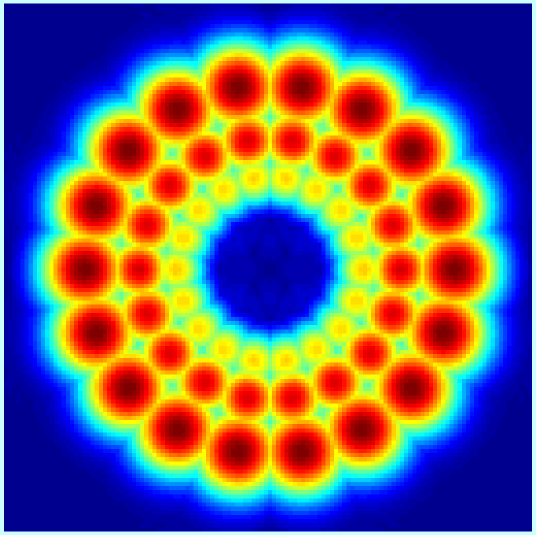
$$g(x,y,t) = K\left[e^{-t^2/2\gamma^2} e^{2\pi\sqrt{-1}(w_0 t)}\right] e^{-\left[\left(x/\lambda\right)^2 + y^2\right]/2\sigma^2} e^{2\pi\sqrt{-1}(u_0 x + v_0 y)}$$

- **Optimally localized** in space-time- frequency.

3D Gabor Filterbanks

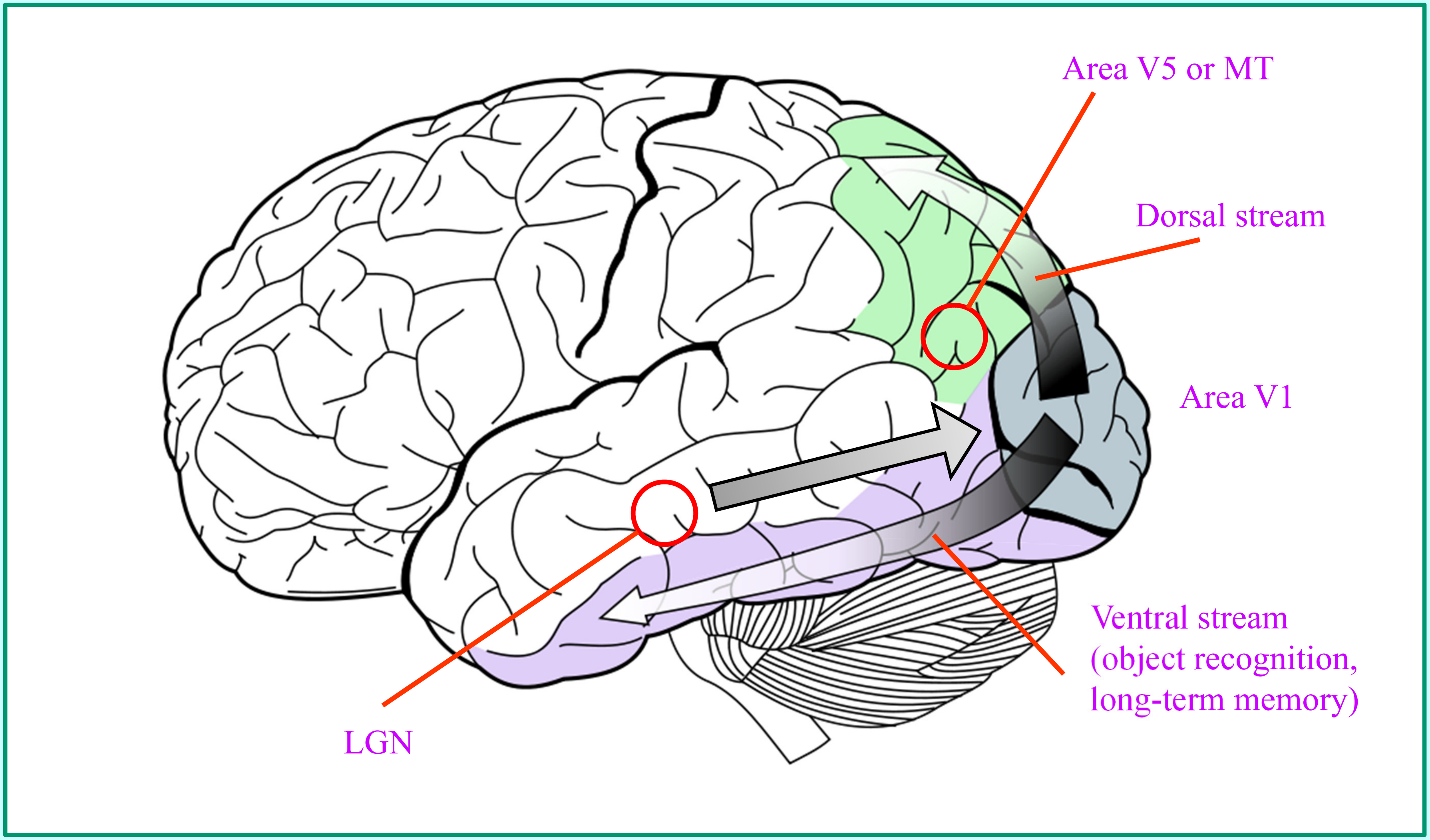


- One scale of a **3D Gabor filterbank** (iso-contours shown)



- Slice through entire filterbank (w = 0) showing multi-scales.

Flow of Visual Data



Dorsal Stream

- The **dorsal stream** (also called the **“where pathway”)** begins with Area V1, and lands in several areas devoted to hand-eye control, **motion perception,** location perception, **depth perception,** and **eye movements**.

- Much of this processing appears to occur in **Area V5,** also known as **Area MT.**

Visual Area MT

- **Some motion and disparity processing occurs in Area V1.**
- **A lot occurs in Area MT.**
- Apparently local motion signals are **integrated into global motion representations.**
- Apparently local disparity signals are **integrated into global depth perception.**
- Processing division between Area1 V1 and MT (and other areas) not clear.

Motion Computation

- **Don't know** how motion is computed in **Area MT.**
- **Do** know that **most of the MT cells are motion sensitive** and use the "data" from Area V1.
- We do know the **primitive information used.**
- And can construct **successful algorithms.**

3D Spatiotemporal Filtering

- 3D spatiotemporal filter model:

$$g(\mathbf{x},t) = K\left[e^{-t^2/2\gamma^2} e^{2\pi\sqrt{-1}(w_0 t)}\right] e^{-\left[\left(x/\lambda\right)^2 + y^2\right]/2\sigma^2} e^{2\pi\sqrt{-1}(u_0 x + v_0 y)}$$

- Find filterbank responses

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

- Largest magnitude response

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

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

Flow Computation

- "Maximizing" response has phasor form

$$t^*(\mathbf{x}, t) = \rho(\mathbf{x}, t)\exp[\sqrt{-1}\,\phi(\mathbf{x}, t)]$$

- Algorithm: compute motion vectors

$$\mathbf{v}(\mathbf{x}, t) = s(\mathbf{x}, t)\mathbf{n}(\mathbf{x}, t)$$

where

$$s(\mathbf{x}, t) = \frac{-\phi_t(\mathbf{x}, t)}{|\nabla\phi(\mathbf{x}, t)|} \qquad \mathbf{n}(\mathbf{x}, t) = \frac{\nabla\phi(\mathbf{x}, t)}{|\nabla\phi(\mathbf{x}, t)|}$$

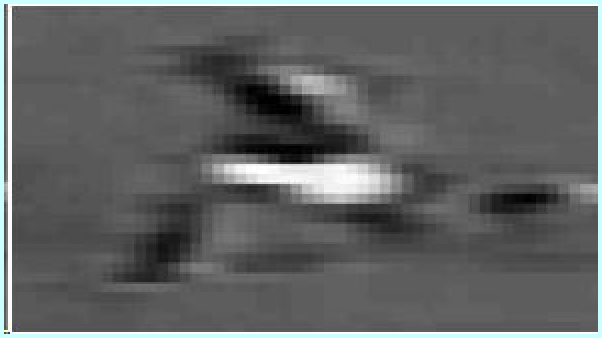
Speed of (temporal) phase change (normalized)

Direction of fastest (spatial) phase change

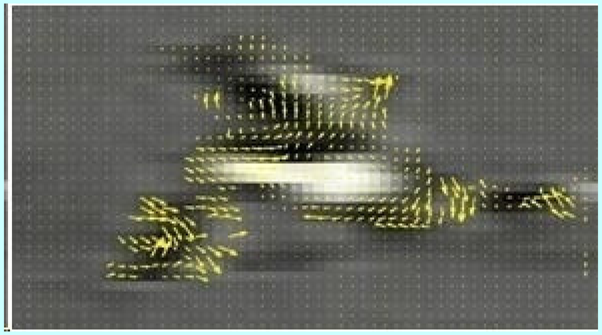
Phase-Based Optical Flow

- Algorithm (D. Fleet and A. Jepson) computes the **optical flow of phase contours** (of fastest change).
- **More stable** and **reliable** than **intensity flow.**
- Maximizing channel **improves SNR** greatly.
- A **plausible model** for Area V1 / MT processing

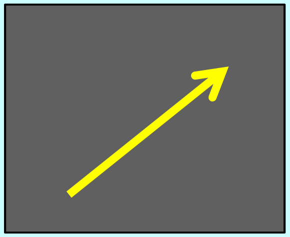
Optical Flow Examples



Frame from “running man” video



Frame showing flow needle diagram overlaid in yellow



Arrow length is magnitude of flow

Arrow orientation is direction of flow

Moving Robot Camera Computing Optical Flow

Global Motion

- We do NOT know how motion is put together to create the perception of **entire moving objects**.

- But it occurs:

  MOVIE 1

  MOVIE 2

- Although the vision system does not necessarily try to create rigid objects or parts of objects: The Blue Bowl

- Now let's look at **practical motion estimation methods….**

Onward….

- …. to **digital video compression standards.**