Moving Image Perception

Prof. Ioannis Pitas
Aristotle University of Thessaloniki
pitas@csd.auth.gr
www.aiia.csd.auth.gr

Version 2.6.2

VML

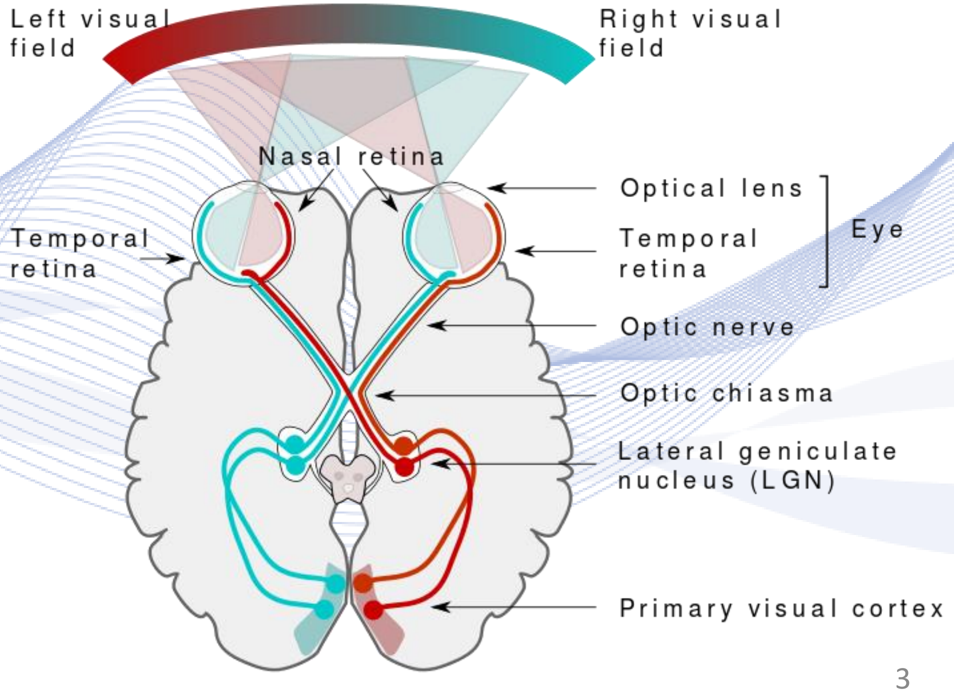
Artificial Intelligence & Information Analysis Lab

Moving Image Perception

- **Human Vision Modeling**
- Video Frequency Content
- Spatiotemporal HVS Models
- Video Quality Assessment

Human Vision Modeling

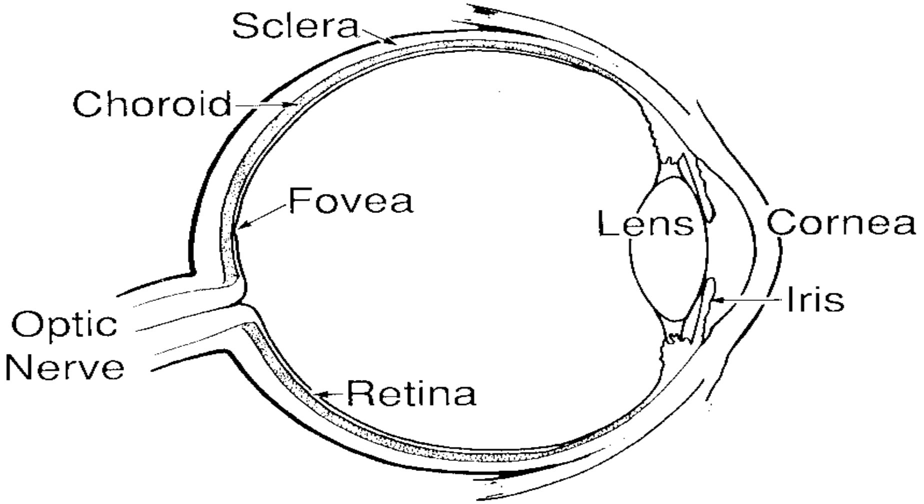
- One of digital image and video processing aims is image quality improvement.
- ***Human Visual System*** (***HVS***) modeling is difficult, because of its complex structure.



Human visual system [HVP].

Human Vision Modeling

- ***Human eye***: spherical shape with a diameter of 20 mm.
- Light enters the ***pupil*** of the ***iris*** (diameter 2 - 8 mm).
- It passes through ***lens***, ***vitreous humour*** and on the ***retina***.

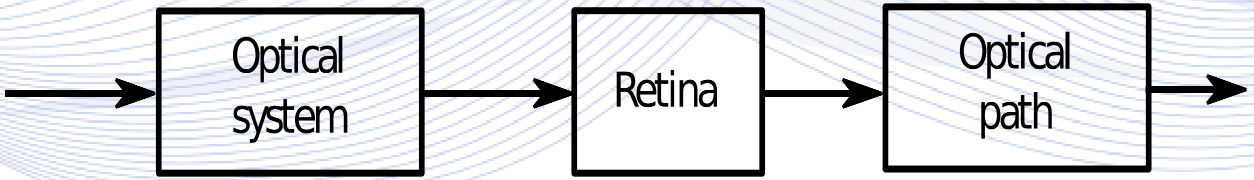


Human eye.

Human Vision Modeling

Retina light detectors: cones and rods.

- ***Cones***: sensitive to color.
  - ***Photopic*** (high brightness, daylight) vision.
- ***Rods***: sensitive to light intensity, not color.
  - They create a general idea of the visual content.
  - ***Scotopic*** (night) vision.



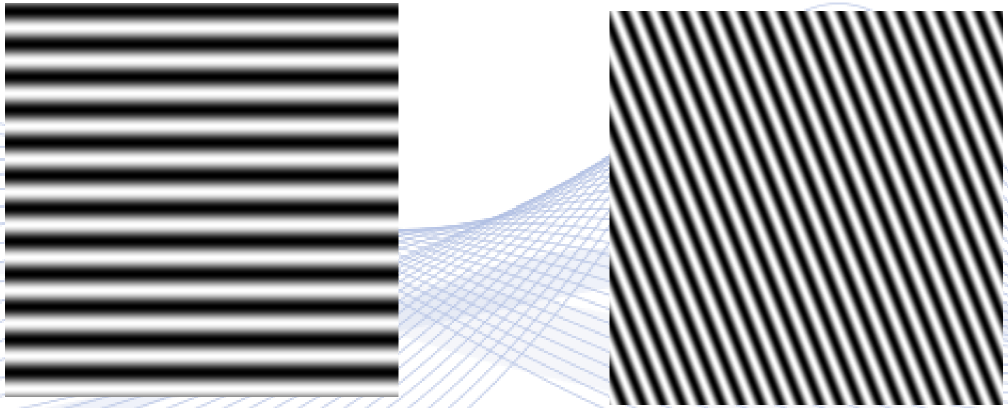
Human visual system model.

Moving Image Perception

- Human Vision Modeling
- **Video Frequency Content**
- Spatiotemporal HVS Models
- Video Quality Assessment

Video Frequency Content

- A frequency F is linked with angular frequency $\Omega = 2\pi F$.
- Spatial frequencies (video content changes along x, y axes):
  - $\Omega_x = 2\pi F_x$ and $\Omega_y = 2\pi F_y$.



2D sinusoidal signals: a) $(F_x, F_y) = (0,6)$; b) $(F_x, F_y) = (10,4)$.

Video Frequency Content

Spatial frequencies F_x, F_y:

- They show spatial luminance changes on the image plane.
- Local frequency vector $\mathbf{\Omega} = [\Omega_x, \Omega_y]^T$ is colinear to local image content change direction (perpendicular to edge direction).
- Spatial frequencies can be defined along different orthogonal axes than (x, y).
- They are measured in cycles per unit length:
  - e.g., a 2D sinusoidal spatial pattern $f(x, y) = \sin(20\pi y)$ has a frequency (0,10).

Video Frequency Content

Temporal Frequency F_t:

- Video signal: moving image (2D video frames changing over time).
- The temporal frequency F_t depends on image content motion.
- The video content motion is due to:
  - camera motion and/or
  - object(s) motion.

Video Frequency Content

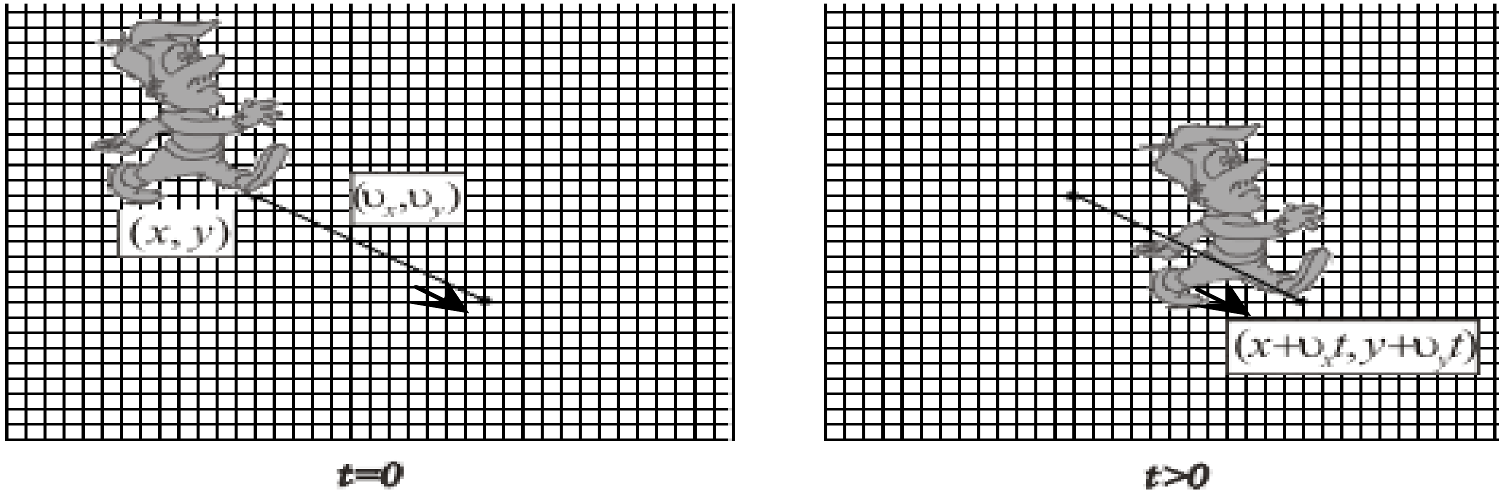
Constant velocity ***2D linear object motion***:

- $f_0(x,y) = f(x,y,0)$: object image at time $t = 0$.
- $\mathbf{v} = [v_x, v_y]^T$: object motion vector.
- v_x, v_y: horizontal/vertical speed.

- Object image at time t (for homogeneous image background):

$$f(x,y,t) = f\left(x - v_x t, y - v_y t, 0\right) = f_0(x - v_x t, y - v_y t).$$

Video Frequency Content



Linear constant 2D object motion.

Video Frequency Content

- ***Spatiotemporal Fourier Transform***:

$$F(\Omega_x, \Omega_y, \Omega_t) = \iiint f(x, y, t) e^{-i(\Omega_x x + \Omega_y y + \Omega_t t)} dx\, dy\, dt =$$
$$F_0(\Omega_x, \Omega_y)\delta(\Omega_t + \Omega_x v_x + \Omega_y v_y).$$

  δ: delta function.

- Non-zero spectrum $F(\Omega_x, \Omega_y, \Omega_t)$ only on the plane:

$$\Omega_t + \Omega_x v_x + \Omega_y v_y = 0.$$

- Temporal frequency due to object motion:

$$\Omega_t = -\Omega_x v_x - \Omega_y v_y = -\mathbf{\Omega}^T \mathbf{v}.$$

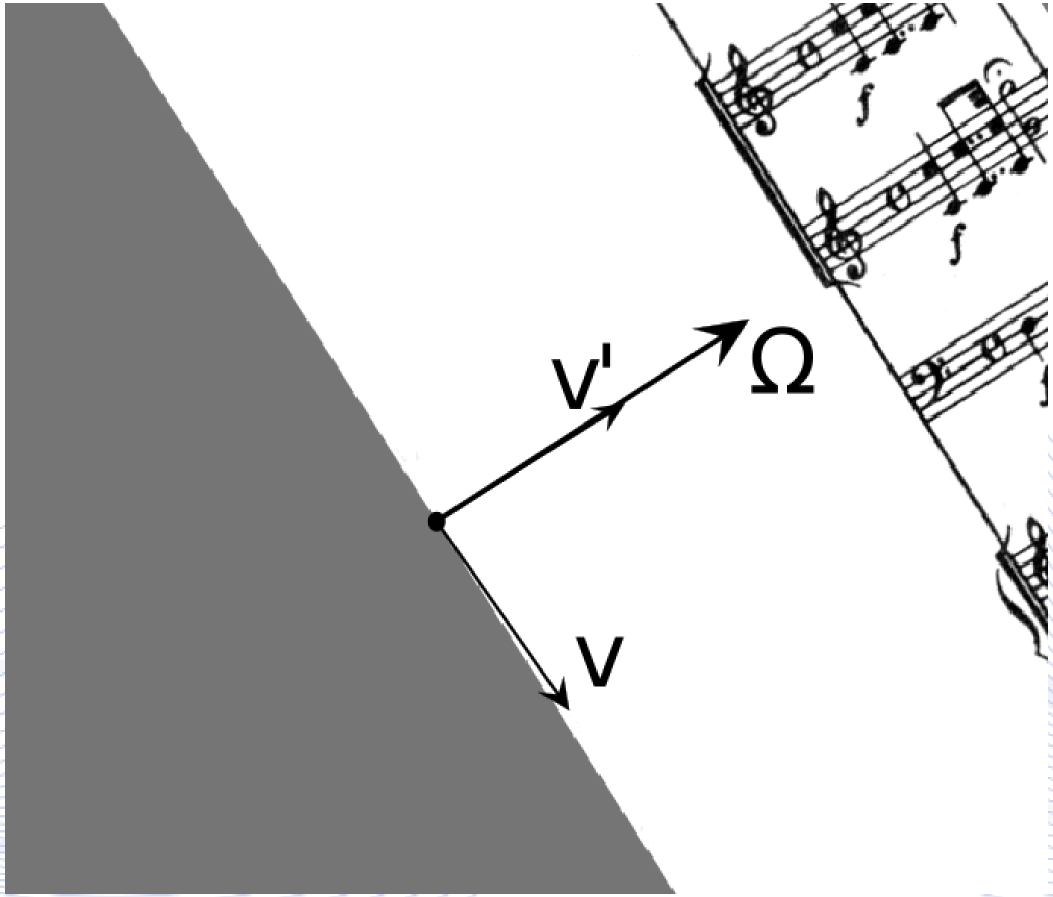
Video Frequency Content

- When $\Omega_x = \Omega_y = 0$ (image is blank, has no content), then $\Omega_t = 0$, regardless of image plane motion.

- Temporal frequency $\Omega_t = 0$, if:
  - Motion vector $\boldsymbol{v}$ is orthogonal to the spatial frequency vector $\boldsymbol{\Omega} = [\Omega_x, \Omega_y]^T$ or
  - $\boldsymbol{\Omega} = \mathbf{0}$ (no motion).

Artificial Intelligence & Information Analysis Lab

Video Frequency Content



Motion vector perpendicular to local image frequency vector.

Moving Image Perception

- Human Vision Modeling
- Video Frequency Content
- **Spatiotemporal HVS Models**
- Video Quality Assessment

Spatiotemporal HVS models

- For the ***spatiotemporal modelling*** of human vision, dynamic models of neurons must be used:
  - It is particularly difficult.

- The eye is a dynamic system:
  - pupil diameter changes with light intensity,
  - the human eye can rotate and perform smooth pursuit movements.
- Spatiotemporal image perception experiments.

Spatiotemporal HVS models

Temporal sensitivity of the human vision:

- HVS temporal frequency response refers to HVS sensitivity to temporal video content variations.

- ***Display flicker***.

Spatiotemporal HVS models

Kelly experiments: determination of the necessary video frame rate (fps).

- Observers were presented a flat screen whose luminance changed sinusoidally:

$$f(x,y,t) = C(1 + s\cos 2\pi F_t t).$$

- C: constant luminance,
- F_t: temporal frequency,
- s intensity modulation level.

Spatiotemporal HVS models

- Intensity modulation levels s were changed for constant luminance C.
- Observers were asked to identify lowest observable modulation level s_{min}.

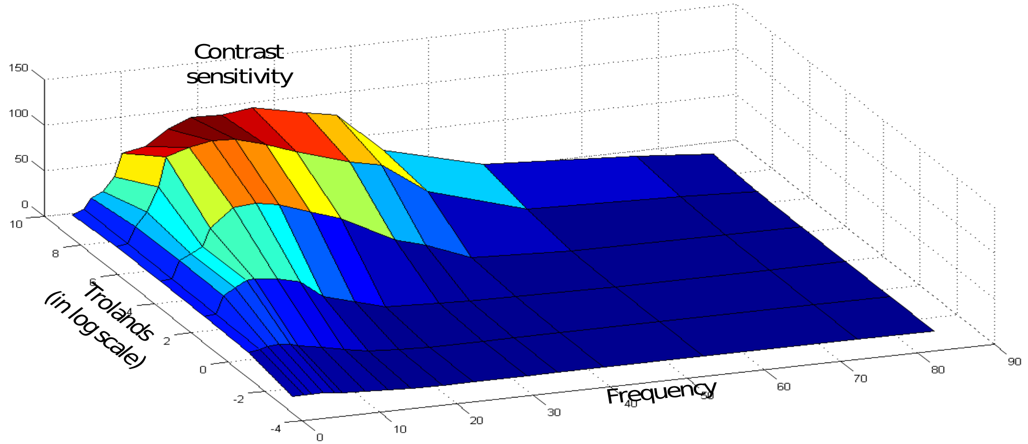
- Observer ***sensitivity*** to temporal luminance variations at frequency F_t:

$$s_e = 1/s_{min}.$$

Spatiotemporal HVS models

Contrast Sensitivity Function (***CSF***) $s_e(C, F_t)$.

- ***Troland***: a unit to describe the light intensity entering the eye retina.
- CSF $s_e(C, F_t)$ is a band-pass function of both C, F_t.



Spatiotemporal HVS models

- Sharp reduction after a cutoff frequency.
- Peak frequency increases with the average image brightness.
- Human eye has low sensitivity to high temporal frequencies because it retains the sense of an image for a short time (***vision persistence***).
- It is caused by the temporal integration of the incoming light energy.

Spatiotemporal HVS models

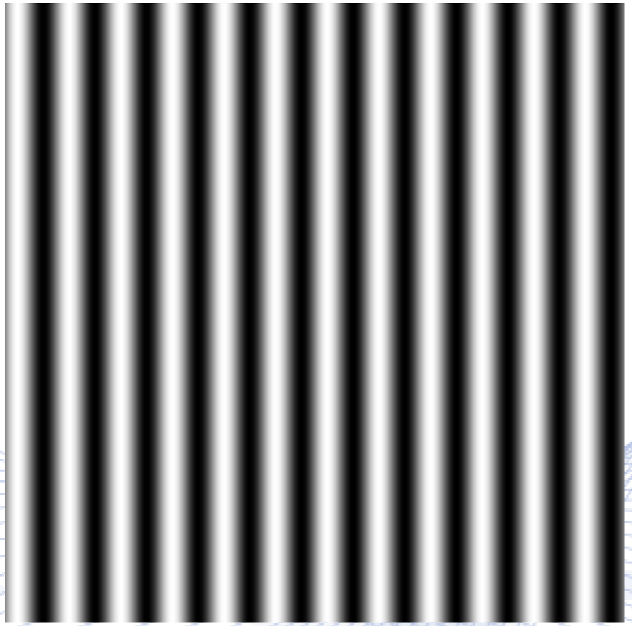
- ***Block's law*** states that the integration period is inversely proportional to that of the light intensity.
- The brighter the source is, the shorter is the retention period.
- Brighter screens (e.g., computer screens) need higher refresh rate (fps) to avoid flickering.
- For cinema (dark theater) and fps of $24\,Hz$ is enough to avoid flickering.

Spatiotemporal HVS models

HVS Spatial Frequency Response (***SFR***):

- Assumption: spatial HVS sensitivity is isotropic, the spatial frequency response can be measured along any arbitrary spatial axis.
- Spatial sensitivity normalization to the observation distance: SFR expressed as a function of the ***spatial angular frequency***.

Spatiotemporal HVS models



Horizontal 2D sinusoidal signals having $(F_x, F_y) = (6,0)$.

Spatiotemporal HVS models

Kelly experiments

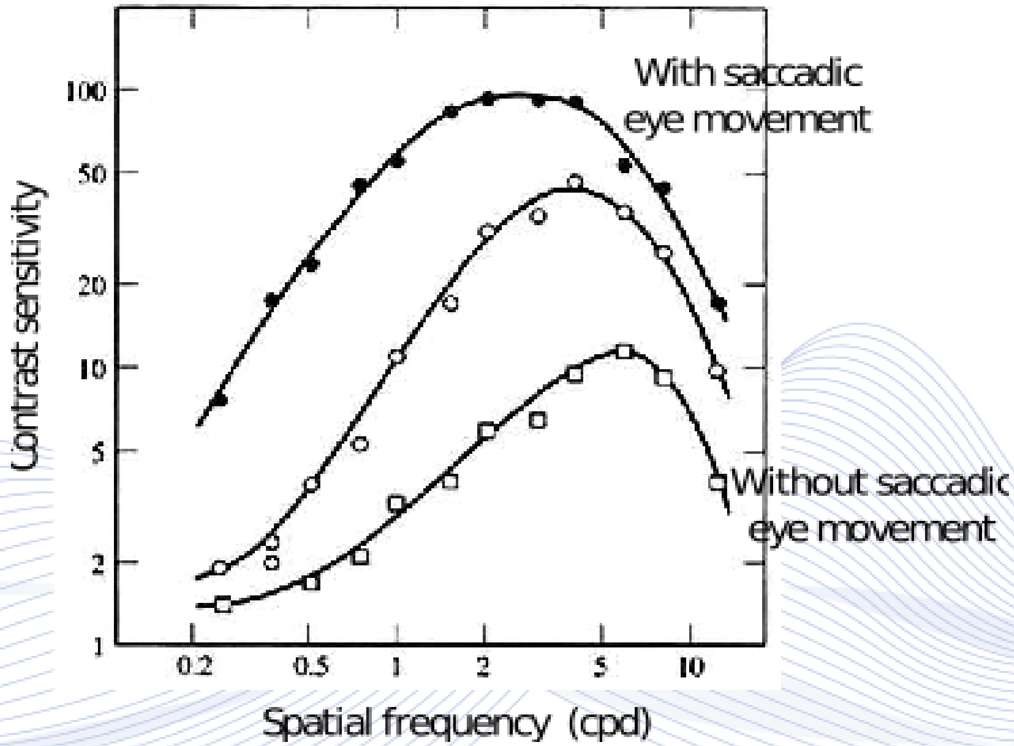
- Human observers observe vertical sinusoidal patterns of amplitude Cs and frequency F_x superimposed on a constant background having illumination C:

$$f(x,y) = C(1 + s\cos 2\pi F_x x).$$

- When they first observe the existence of the pattern having amplitude Cs_{min}, contrast sensitivity is defined by:

$$S_e = \frac{1}{s_{min}}.$$

Spatiotemporal HVS models



Spatial HVS frequency response.

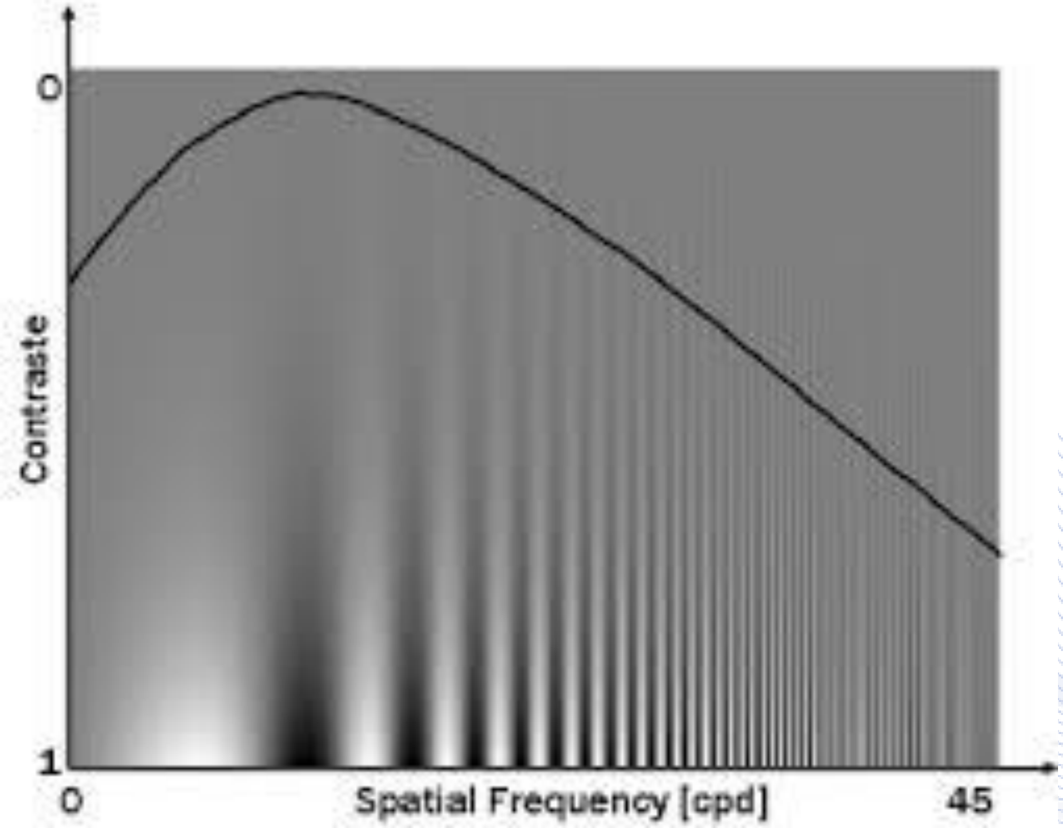
Spatiotemporal HVS models

- Peak SFR is observed at about 2-5 cycles/degree.
- Cutoff spatial frequency is 30 cycles/degree.

- No details can be observed on a fast moving object.

- SFR varies with eye motion configuration:
  - ***Saccadic motion*** allows much higher spatial HVS sensitivity.

Spatiotemporal HVS models



Contrast sensitivity.

Spatiotemporal HVS models

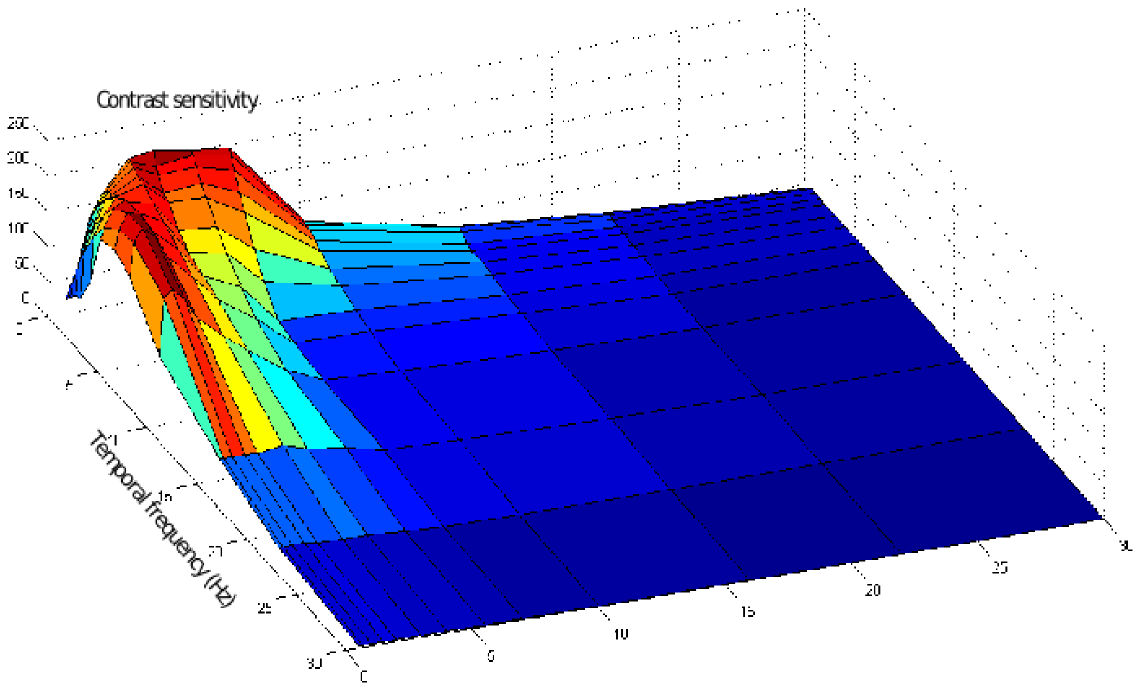
HVS spatiotemporal frequency response experiments:

- Test pattern:

$$f(x,y,t) = C(1 + s\cos(2\pi F_x x)\cos(2\pi F_t t)).$$

- For a fixed pair of F_x and F_y the modulation level s was changed.
- The observer was requested to determine the minimal observable modulation level s_{min}.

Spatiotemporal HVS models



Spatiotemporal HVS contrast sensitivity as a function of F_x , F_t for unconstrained eye motion.

Spatiotemporal HVS models

- When the temporal(spatial) frequencies are close to zero, the spatial (temporal) HVS frequency response has band-pass characteristics.
- At high temporal (spatial) frequencies, the spatial (temporal) frequency HVS response is low pass:
- Its peak response frequency reduces, when the temporal (spatial) frequency increases.

- When an object moves too quickly, the eye cannot distinguish object details (high spatial frequencies).

Artificial Intelligence & Information Analysis Lab

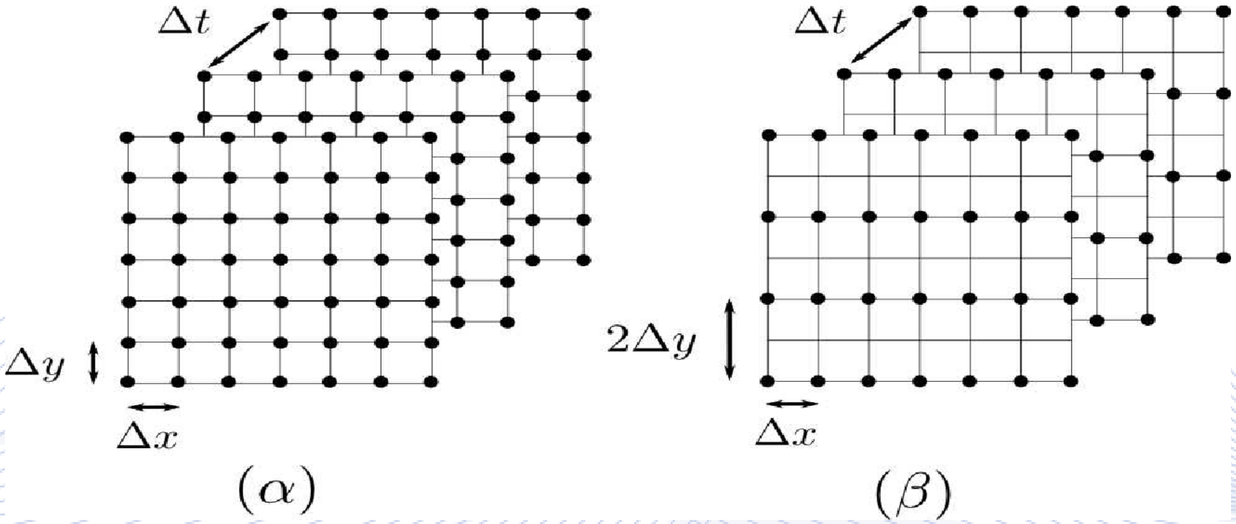
Spatiotemporal HVS models

- HVS can discern object details (higher spatial frequencies) much better, when the object image is stationary.
- Similarly, at higher spatial frequencies the temporal response becomes low pass.

- Implication of this inverse relationship between spatial and temporal HVS responses:
  - ***spatial video resolution can be swapped with temporal one and vice versa***.

Spatiotemporal HVS models

- 2:1 Interlaced video takes advantage of HVS properties.



Sampling grids for: a) progressive; b) 2:1 interlaced video.

Spatiotemporal HVS models

2:1 interlaced video:

- fast moving scenes:
  - They can be visualized at a limited frame rate, by dividing a frame into two temporally adjacent fields, each with half the number of frame lines.
- slow moving scenes:
  - the lines in two consecutive fields are perceived as one high spatial resolution video frame.

Spatiotemporal HVS models

Smooth eye pursuit movement:

- Limited spatial sensitivity for fast moving objects, but eye does not.
- Observed image on eye retina, when eye moves:

$$\tilde{f}(x,y,t) = f\left(x + v_x t, y + v_y t, t\right),$$

  - $f(x,y,t)$: displayed image,
  - (v_x, v_y): eye motion velocities.

Spatiotemporal HVS models

- Eye motion cancels object motion on retina.
- Details of a fast moving object can be seen by eye pursuit movement:
  - Increased spatial HVS sensitivity.

Moving Image Perception

- Human Vision Modeling
- Video Frequency Content
- Spatiotemporal HVS Models
- **Video Quality Assessment**

Video Quality Assessment

- In many cases, humans are the final video consumers.
- Perceived video quality must be quantified.
- ***Video Quality*** (***VQ***) is influenced by:
  - Acquisition noise;
  - Compression effects;
  - Transmission errors.

- VQ assessment can help meet video storage and transmission requirements.

Artificial Intelligence & Information Analysis Lab

Video Quality Assessment

Subjective video quality assessment:

- Ask humans to watch the video and assess its quality.
- ***Mean Opinion Score*** (***MoS***): scale $[1, \ldots, 5]$.
- 1: worst, 5: best quality.
- Labor intensive and expensive.
- A large number of viewers is needed to lower score variability and provide statistical certainty.
- Impossible to assess all videos before broadcasting.
- Useful in providing a golden standard for automated VQA methods.

Video Quality Assessment

Objective video quality assessment:

- No human observers involved.
- ***Full reference VQA algorithms*** operate on distorted video, while employing the original video reference for comparison.
- VQA measures:
  - Mean Square Error,
  - Peak Signal to Noise Ratio,
  - SSIM.

Video Quality Assessment

- ***Reduced reference VQA algorithms*** use distorted video and some original video information, e.g., edge locations.

- ***Blind VQA algorithms***: no knowledge about distortion nor reference video.

Bibliography

[PIT2017] I. Pitas, "Digital video processing and analysis" , China Machine Press, 2017 (in Chinese).
[PIT2013] I. Pitas, "Digital Video and Television" , Createspace/Amazon, 2013.
[PIT2021] I. Pitas, "Computer vision", Createspace/Amazon, in press.
[NIK2000] N. Nikolaidis and I. Pitas, "3D Image Processing Algorithms", J. Wiley, 2000.
[PIT2000] I. Pitas, "Digital Image Processing Algorithms and Applications", J. Wiley, 2000.

Q & A

Thank you very much for your attention!

More material in
http://icarus.csd.auth.gr/cvml-web-lecture-series/

Contact: Prof. I. Pitas
pitas@csd.auth.gr