Digital Images and Videos

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

Version 3.1

Artificial Intelligence & Information Analysis Lab

2D data types: images

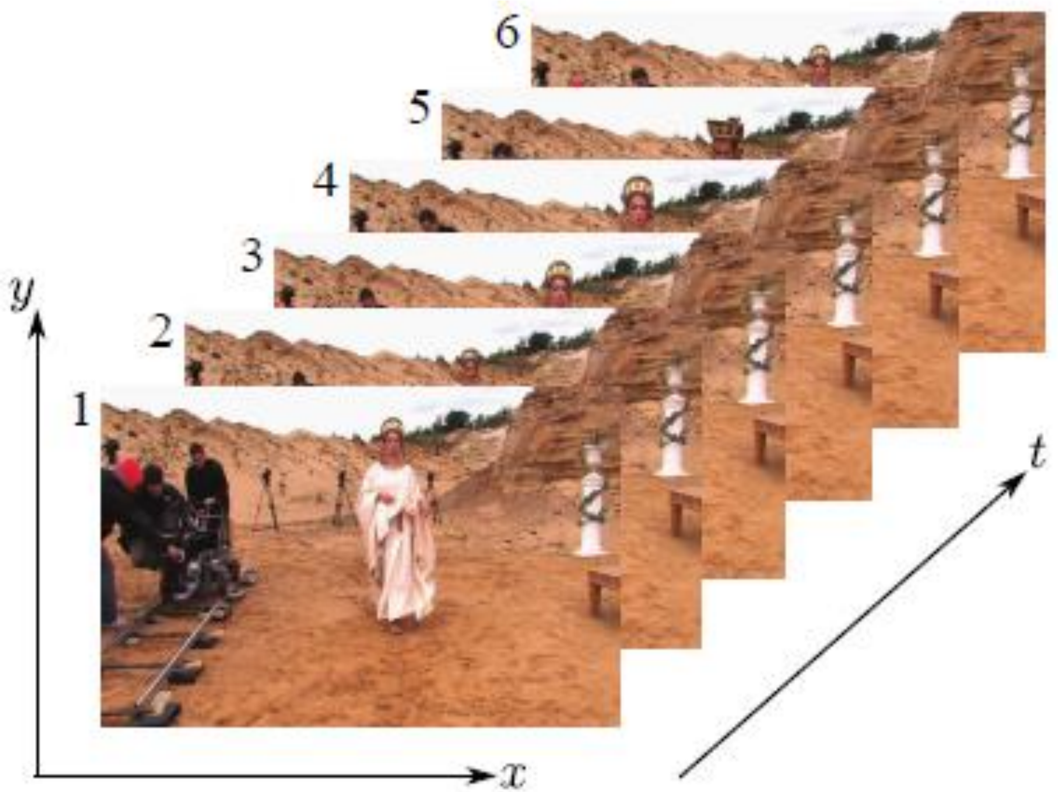
Spatial coordinates x, y.

2D data types: images

- ***Still images/pictures***: spatial 2D signals of the form $f(x,y): \mathbb{R}^2 \rightarrow \mathbb{R}$, having:
  - domain $\mathbb{R}^2$ and codomain $\mathbb{R}$.
  - two spatial coordinates x, y.
- ***Image sampling/digitization*** transforms continuous coordinates images to ***digital images***:

$$f(i,j): \mathbb{Z}^2 \rightarrow [0, \dots, 2^B - 1].$$

3D data types: video



3D data types: video

- Moving images: spatiotemporal 3D signals of the form: $f(x, y, t): \mathbb{R}^3 \rightarrow \mathbb{R}$, having:
  - domain $\mathbb{R}^3$ and codomain $\mathbb{R}$.
  - the time t coordinate has a different nature than the spatial coordinates x, y.
- ***Video scanning***: the process for obtaining an 1D analog video signal, by sampling the time-varying images (luminance or RGB channels) along the vertical axis y and time t.

Artificial Intelligence & Information Analysis Lab

3D data types: video

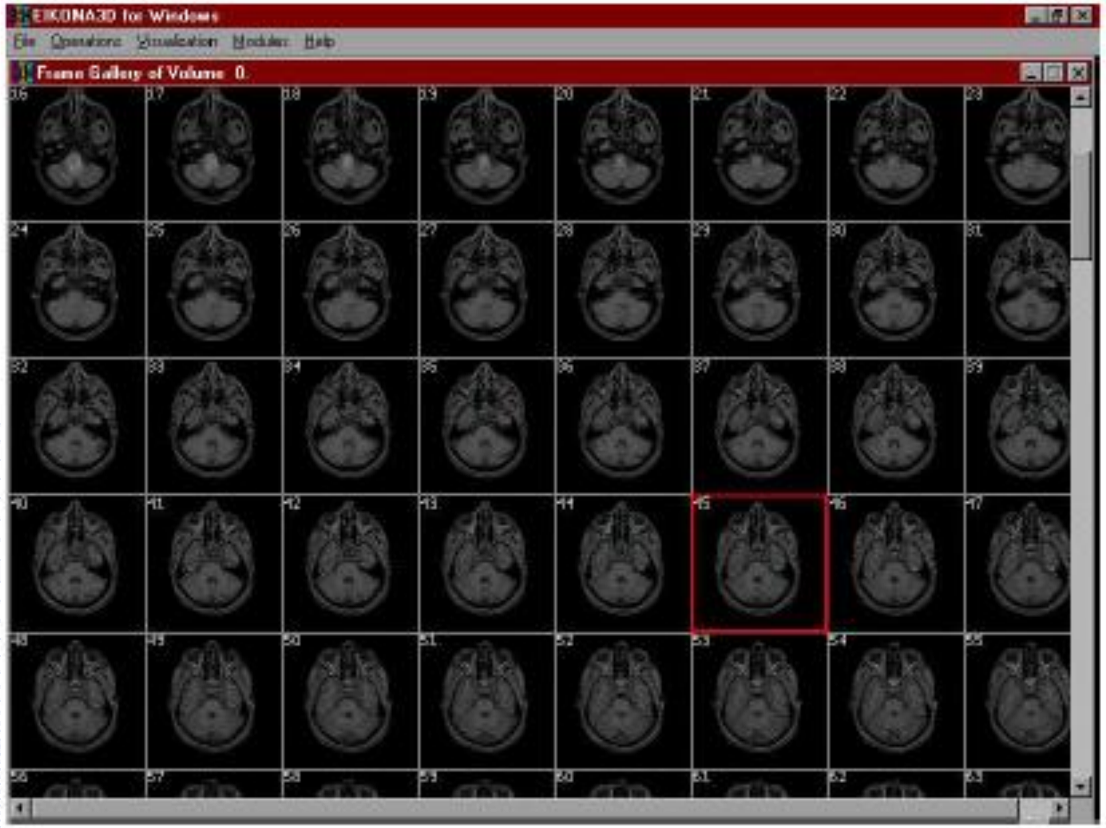
- ***Analog video signal*** $f(x, j\Delta y, k\Delta t)$: $\mathbb{R} \times \mathbb{Z}^2 \rightarrow \mathbb{R}$.
  - discrete along y and t axes
  - continuous along x axis.
- ***Digital video signal*** $f(i\Delta x, j\Delta y, k\Delta t)$: $\mathbb{Z}^3 \rightarrow \mathbb{R}$.
- Spatial sampling intervals $\Delta x, \Delta y$ define ***image resolution***: the smaller they are, the smaller the pixel size is.
- Temporal sampling interval Δt defines the ***video frame rate*** in frames per second (fps).

3D data types: volumetric images

- ***3D volumetric images***: 3D signals of the form $f(x, y, z): \mathbb{R}^3 \rightarrow \mathbb{R}$.

- Discrete versions (defined on a Euclidean grid $\mathbb{Z}^3$) : $f(n_1, n_2, n_3): \mathbb{Z}^3 \rightarrow \mathbb{R}$.
  - $x = n_1 \Delta x, \ y = n_2 \Delta y, \ z = n_3 \Delta z$
  - $\Delta x, \Delta y, \Delta z$: ***spatial sampling intervals*** defining 3D ***image resolution***
  - each *voxel* is a real number.

3D data types : volumetric images



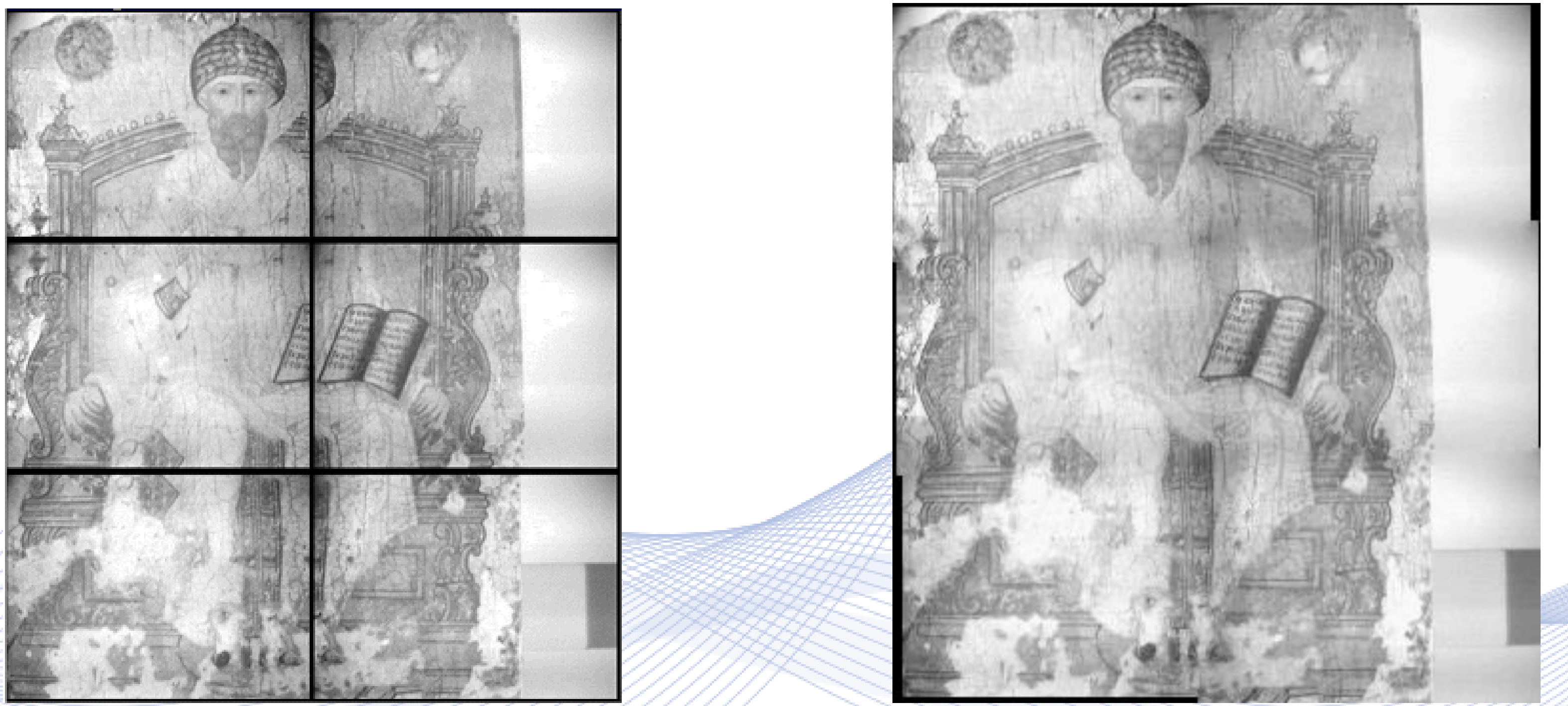
3D data types: multispectral images

- Multispectral/multichannel (n -channel) images have the form: $\mathbf{f}(x,y)\colon \mathbb{R}^2 \rightarrow \mathbb{R}^n$.
  - color images ($n=3$): $\mathbf{f}(x,y) = [f_R(x,y), f_G(x,y), f_B(x,y)]^T : \mathbb{R}^2 \rightarrow \mathbb{R}^3$.
  - digital color images (assigning 8 bits per color channel to each voxel): $\mathbf{f}(n_1,n_2)\colon \mathbb{Z}^2 \rightarrow \{0,\ldots,255\}^3$.
  - Hyperspectral images: $f(x,y,\lambda)\colon \mathbb{R}^3 \rightarrow \mathbb{R}$
    - λ wavelength.

Infrared images

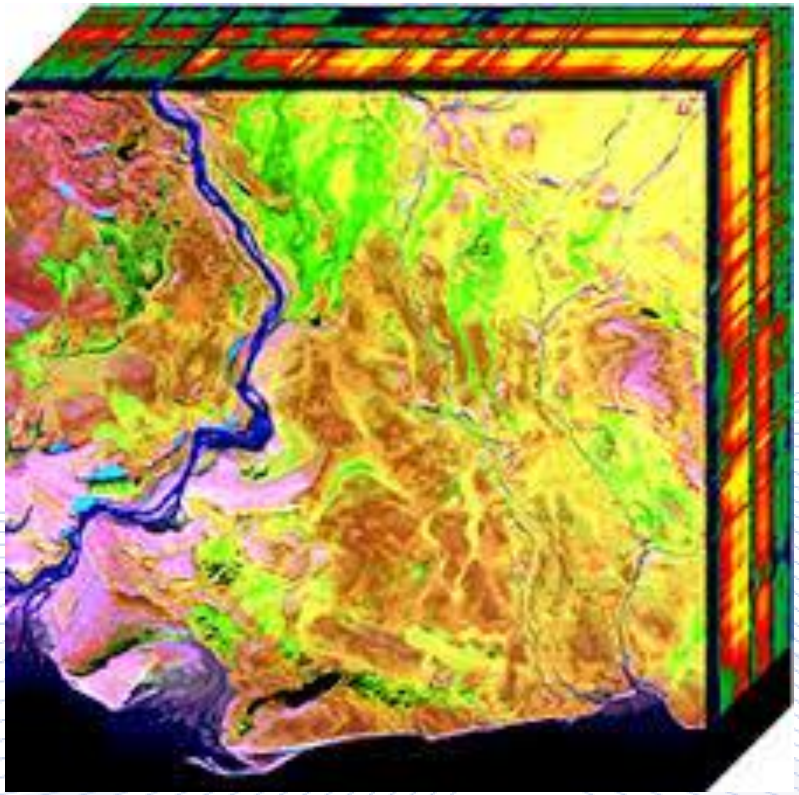
[www.Infrareddiagnostic.com]

Reflectography



a) IR image tiles of a painting; b) mosaiced IR image.

Hyperspectral images



[Wikipedia]

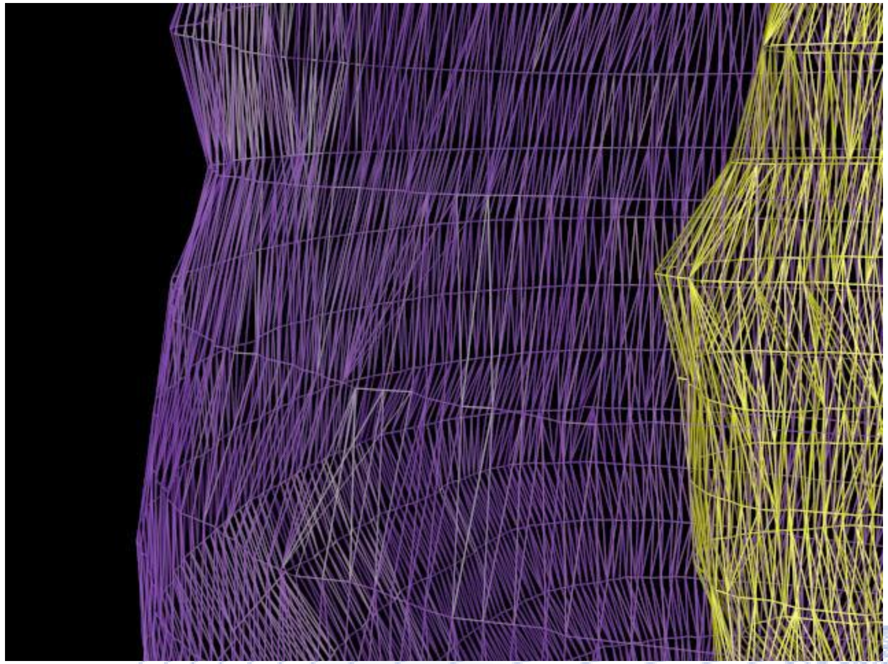
3D data types

- ***Multiview images***: images of an object or set, taken from different view points, typically using different cameras.
  - ***Stereo images***: a special case, employing only two cameras (left and right).
- They both carry only implicit geometrical information about the visualized 3D object.
  - They are not 3D data.
  - 3D object geometry can be derived using stereo or multiview 3D geometry reconstruction techniques.

3D data types

Multiview video: captured by synchronized video-cameras.

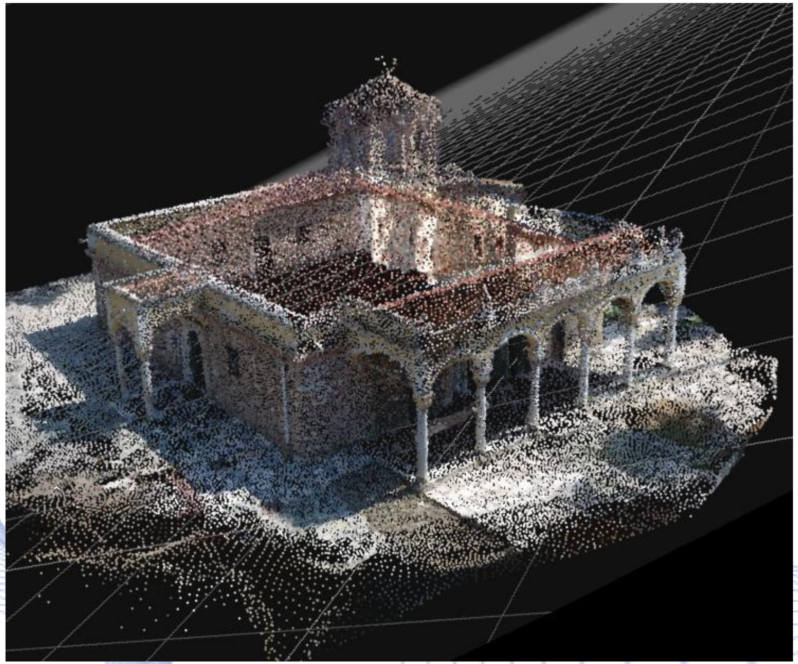
3D data types



a) 3D surface $\mathcal{S} \subset \mathbb{R}^3$ (expressed, e.g., by a triangular mesh)

b) 3D surface texture: $\mathbf{f}(X,Y,Z)$: $\mathcal{S} \subset \mathbb{R}^3 \rightarrow \mathbb{R}^3$.

3D data types



a) 3D surface $\mathcal{S} \subset \mathbb{R}^3$ (expressed, e.g., by a triangular mesh)

b) 3D surface texture: $\mathbf{f}(X, Y, Z)$: $\mathcal{S} \subset \mathbb{R}^3 \rightarrow \mathbb{R}^3$.

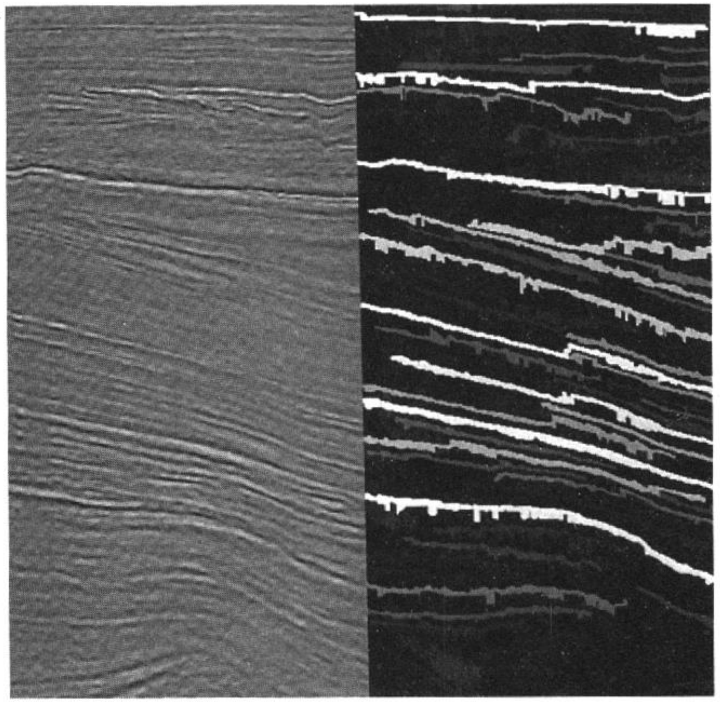
3D data types

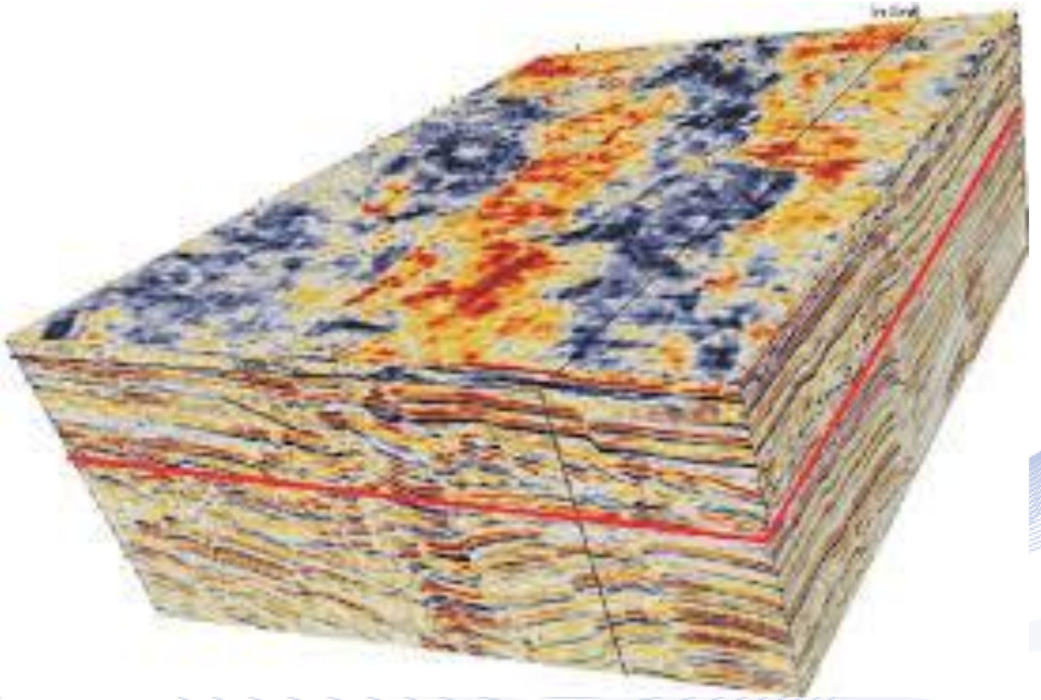
- RGB-D images have: a) RGB channels and b) D (depth) channel.
- They are acquired by RGB-D cameras.

3D data types

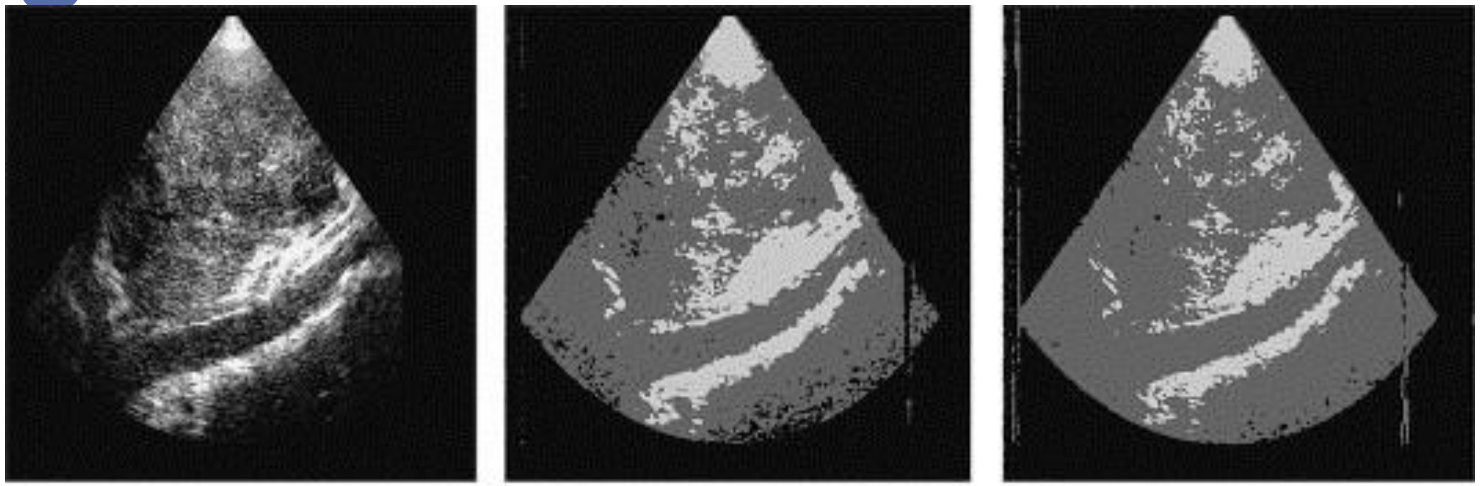
RGB-Depth image acquired from monocular video [APOLLO].

3D data types : seismic images and volumes

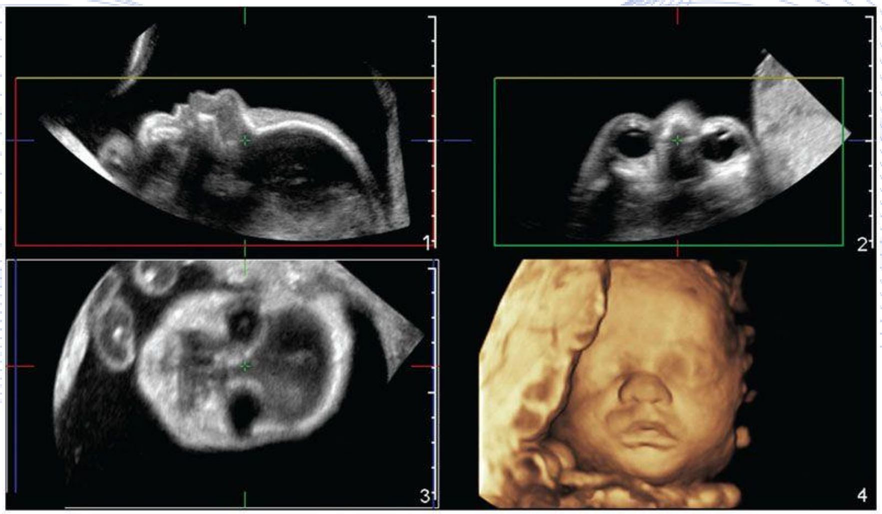




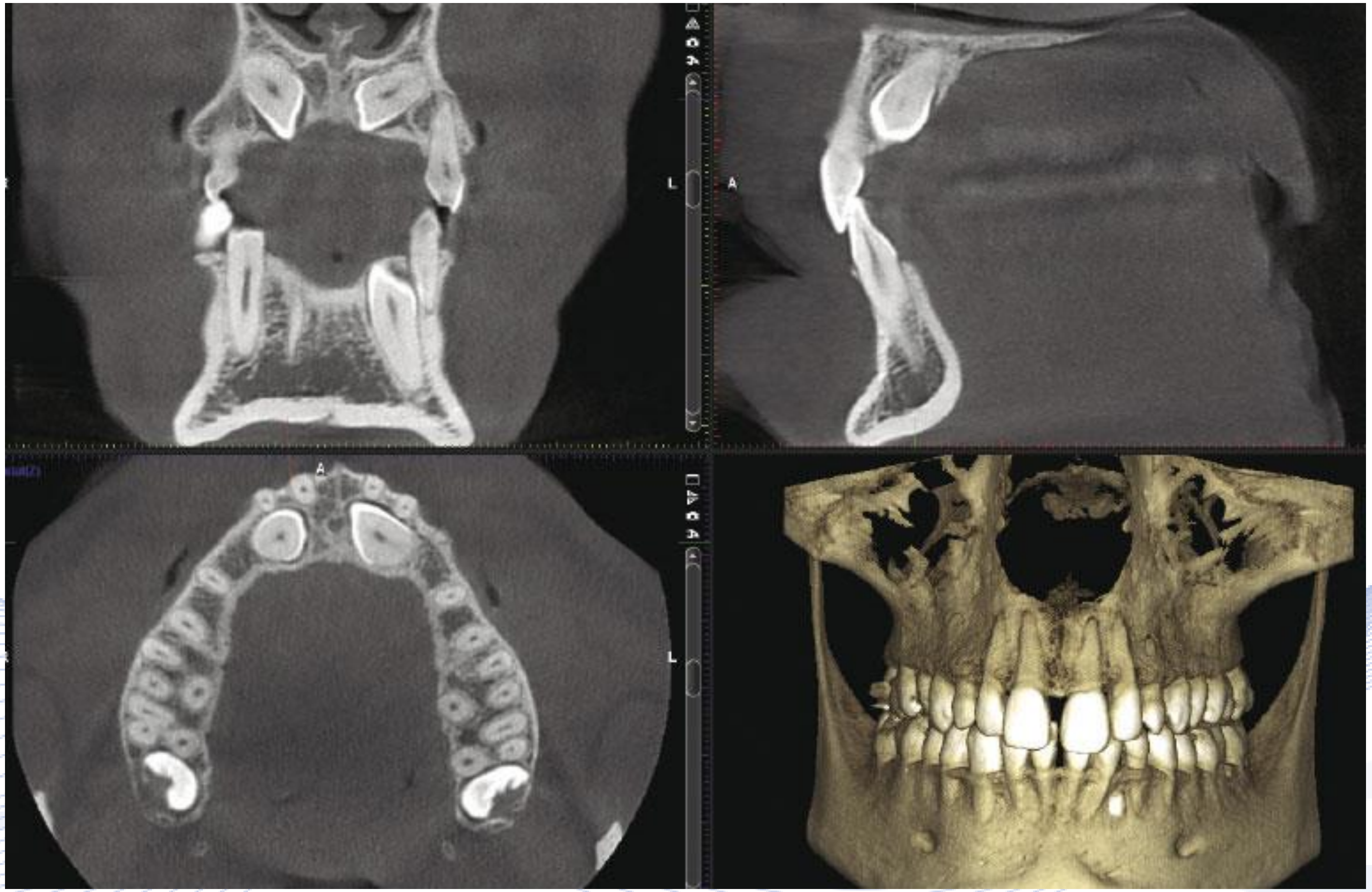
3D data types : ultrasound images and volumes



(a) (b) (c)

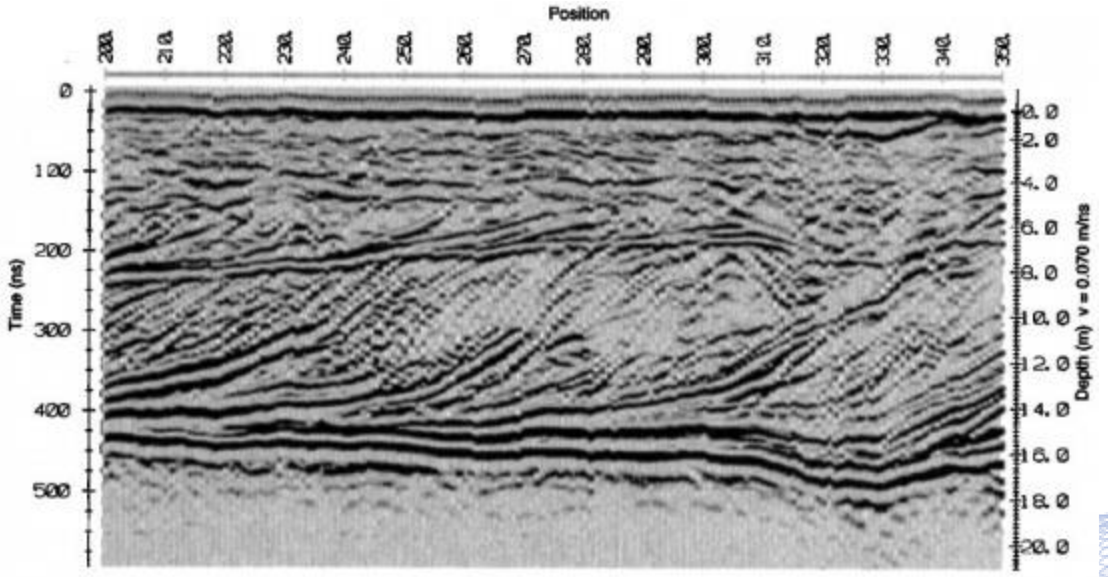


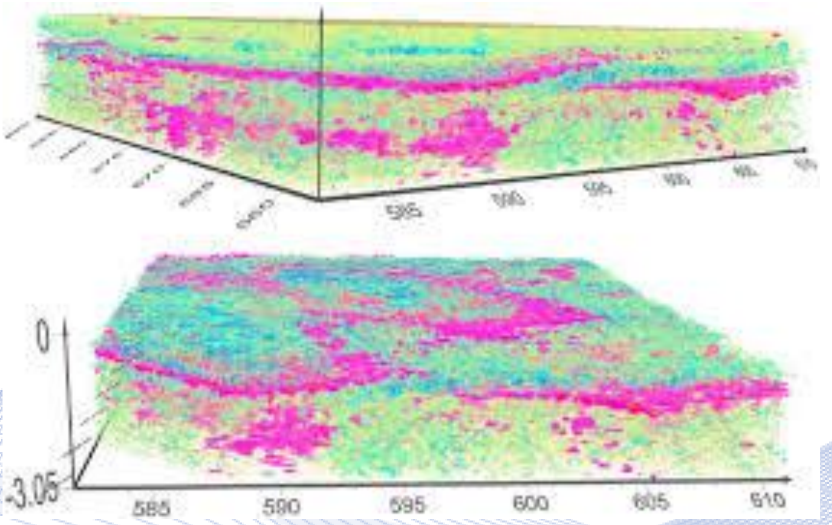
3D data types : x-ray images



a) Tooth X-ray; b) CBCT volume.

3D data types: Ground penetrating radar





Ground penetrating radar a) image; b) volume.

Color theory

- Visible light: an electromagnetic wave with wavelength λ varying in the range $380 - 780\ nm$.
- Perceived color: depends on the spectral content of the light.
  - Red light: a signal with energy concentrated around $700\ nm$.
  - White light: a signal with evenly distributed energy across the wavelength spectrum.
  - ***Monochromatic color***: a color with a very narrow spectral content (typically single-wavelength).

Color theory

- Multispectral/multichannel (n-channel) images have the form: $\mathbf{f}(x, y)\colon \mathbb{R}^2 \rightarrow \mathbb{R}^n$.
  - color images($n = 3$): $\mathbf{f}(x, y) = [f_R(x, y), f_G(x, y), f_B(x, y)]^T : \mathbb{R}^2 \rightarrow \mathbb{R}^3$.
  - digital color images (assigning 8 bits per color channel to each voxel): $\mathbf{f}(n_1, n_2)\colon \mathbb{Z}^2 \rightarrow \{0, \dots, 255\}^3$.
  - They can also be considered as 3D images: $f(n_1, n_2, i), i = 1,2,3$.
  - Hyperspectral images (3D images): $f(x, y, \lambda)\colon \mathbb{R}^3 \rightarrow \mathbb{R}$
    - λ wavelength.

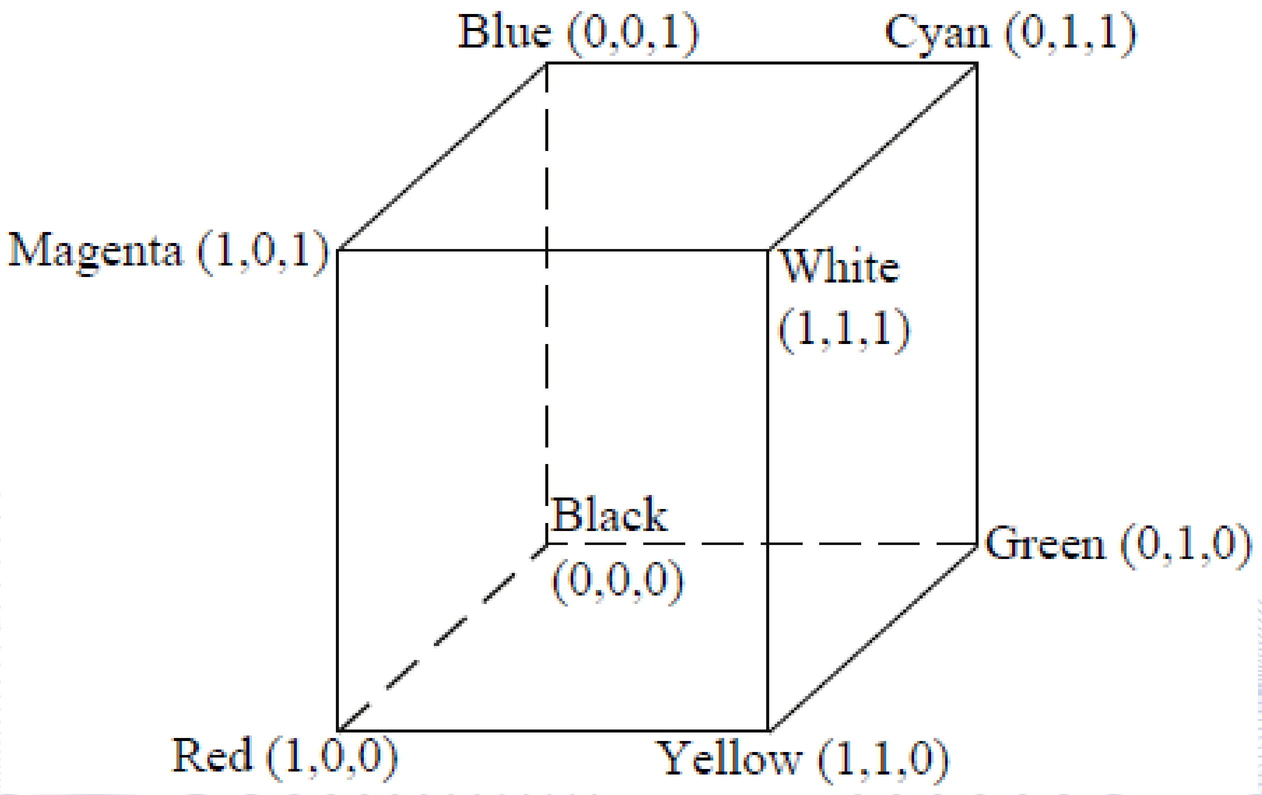
Color images

RGB color image.

Color coordinate systems

- Alternative color image representation:
  - ***Subtractive colors***: cyan, magenta, yellow (complementary of red, green, blue primary colors).
  - ***CMYK*** color system: subtractive color model complemented with black color (mainly used in color image printing).

Color coordinate systems



Color cube.

Color coordinate systems

- ***Human visual system*** is less sensitive to color than to luminance.
- RGB color space: the three colors considered equally important and stored at the same spatial resolution.
- More efficient color image representation: in the luminance-chrominance domain, allocating higher spatial resolution to luminance than to chrominance channels.

Color coordinate systems

- YC_bC_r color space:
  - It is an efficient color representation in analog and digital TV.
  - Y: the luminance channel:

$$Y = k_r R + k_g G + k_b B.$$

  - k: coefficients, with typical values $k_r = 0.299$, $k_g = 0.587$, $k_b = 0.114$.
    - Small weight in the B channel.
  - Chrominance information can be represented as:

$$C_b = B - Y, \qquad C_r = R - Y.$$

Color coordinate systems

- Advantages of the YIQ color space:
  - It guarantees backwards compatibility with monochrome television:

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & -0.322 \\ 0.211 & -0.523 & 0.312 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}.$$

  - Y: luminance component.
  - I, Q: image chrominance.

Color coordinate systems

- The ***Commission Internationale de l' Eclairage*** (***CIE***):
  - proposed the fundamental spectral system RGB to match the monochromatic fundamental sources of $R_{CIE}, G_{CIE}, B_{CIE}$.
    - White color reference: $R_{CIE} = G_{CIE} = B_{CIE} = 1$.
    - CIE RGB color space is unable to display all reproducible colors.
  - It proposed the XYZ color system:
    - Hypothetical coordinates X, Y, Z.
    - White reference color: $X = Y = Z = 1$.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.490 & 0.310 & 0.200 \\ 0.177 & 0.813 & 0.011 \\ 0.000 & 0.010 & 0.990 \end{bmatrix} \begin{bmatrix} R_{CIE} \\ G_{CIE} \\ B_{CIE} \end{bmatrix}.$$

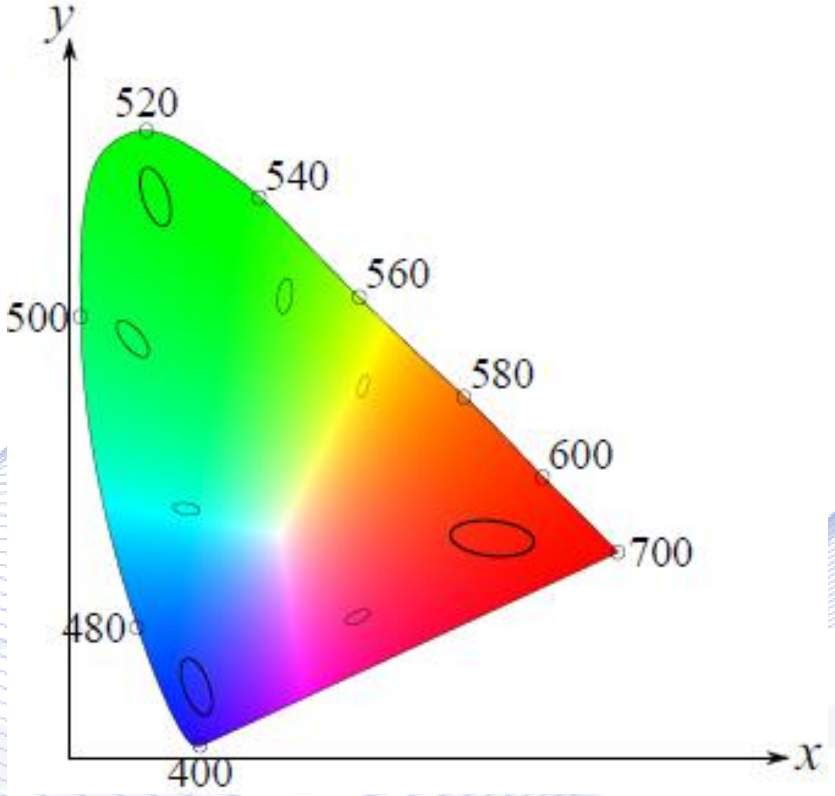
Color coordinate systems

- The color coordinates:

$$x = \frac{X}{X+Y+Z}, \quad y = \frac{Y}{X+Y+Z}$$

can be used to produce a chromaticity diagram.

- Ellipses correspond to colors which cannot be discerned by the human visual system.



Color coordinate systems

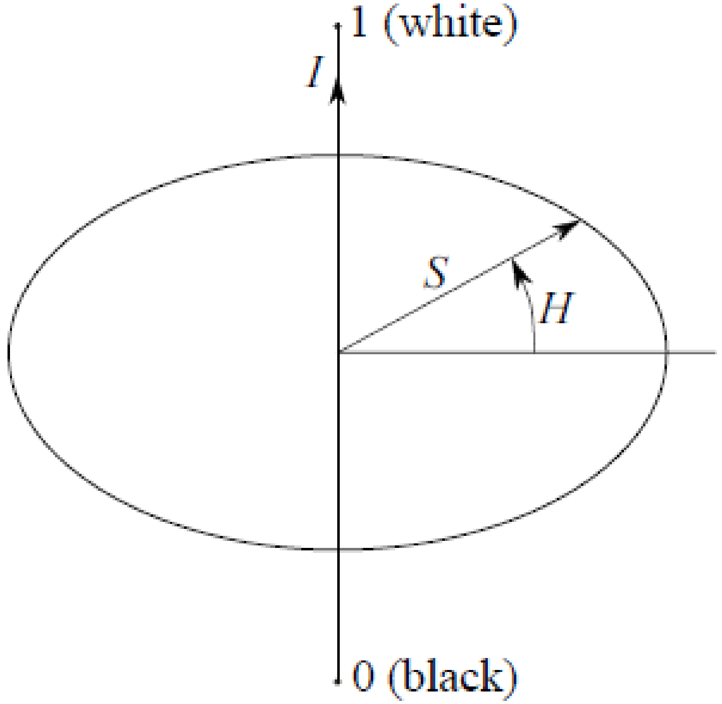
- Such systems can not approximate well the human visual perception of the following three color properties:

  - ***Hue***: it determines color redness, greenness, blueness.
  - ***Saturation***: it defines the percentage of white light added to a pure color.
  - ***Brightness***: it indicates the perceived light luminance.

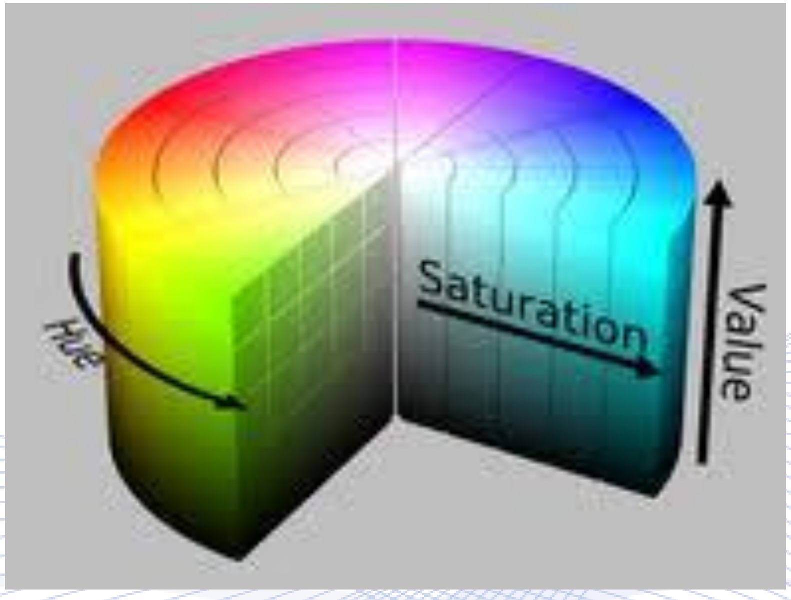
Color coordinate systems

Hue, saturation, brightness color coordinates define a cylindrical color coordinate system:

- Brightness I varies from pure black to pure white color.
- Saturation S ranges from pure gray ($s = 0$) to highly saturated colors ($s = 1$).
- Hue H is measured by the angle between the actual color vector and a reference pure color vector.



Color coordinate systems

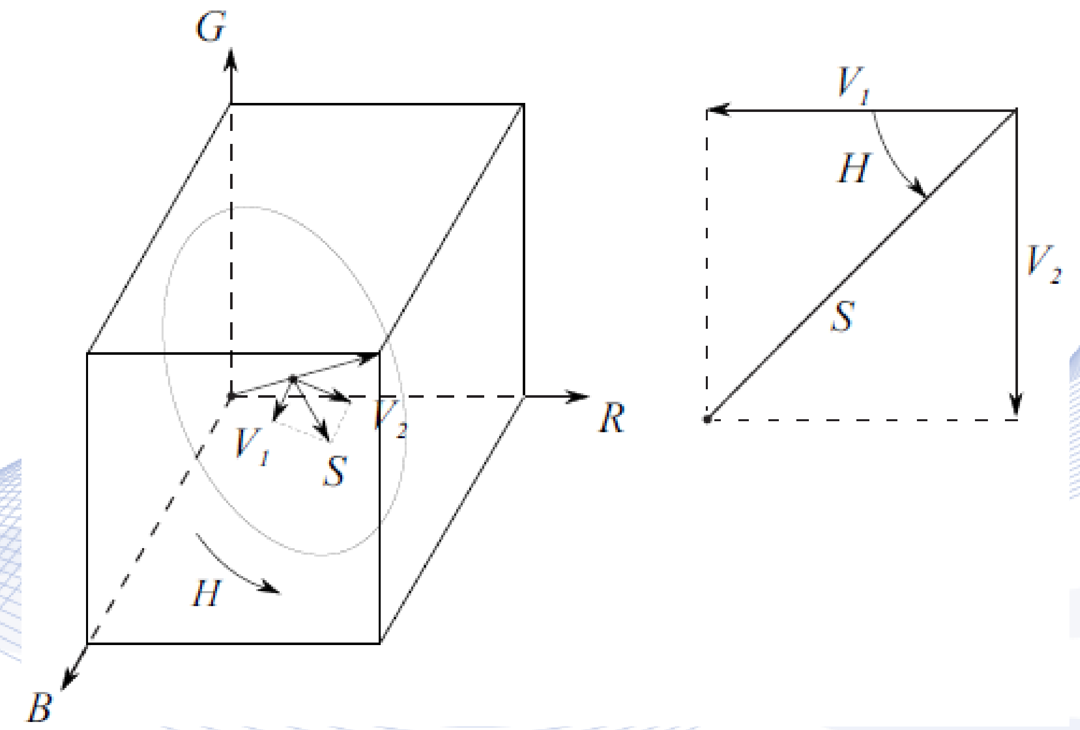


Color hue and saturation.

Color coordinate systems

HSI system (Hue, Saturation, Intensity):

- It is a cylindrical coordinate system with axes determined by the diagonal line $R = G = B$ in the RGB space.
- The colors in the HSI cylindrical coordinate system that are inside in the RGB cube can be displayed.



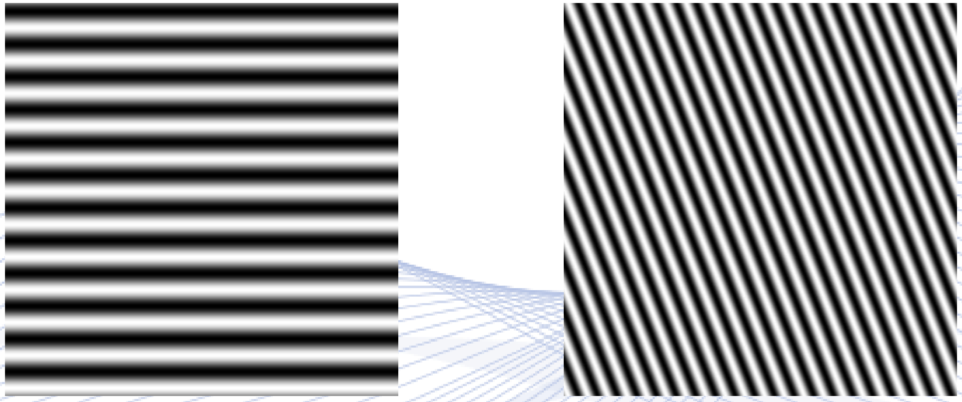
Spatial Frequency Content

- A (temporal) ***frequency*** F is linked to ***angular frequency*** $\Omega = 2\pi F = 2\pi/T$.

- F_x, F_y: 2D spatial frequencies representing how rapidly image luminance or chrominance changes on the image plane:
  - in ***cycles per unit length*** along a given axis,
  - in ***cycles per meter (cpm)*** in the metric measure system.
- $\Omega_x = 2\pi F_x,\ \Omega_y = 2\pi F_y$: respective angular frequencies.

Spatial Frequency Content

- 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) $(F_x, F_y) = (10,4)$.

Spatial Frequency Content

Image

$$f(x, y) = \sin(20\pi x + 8\pi y)$$

has frequencies $(F_x = 10, F_y = 4)$, $(\Omega_x = 20\pi, \Omega_y = 8\pi)$:

- 10 cycles per unit length along the horizontal direction,
- 4 cycles per unit length along the vertical direction.
- $F_s = \sqrt{F_x^2 + F_y^2} = 10{,}77 \cong 11$ cycles per unit length along the direction:

$$\theta = arctan(F_x / F_y) = 21{,}8^o.$$

Spatial Frequency Content

Any image $f_a(x,y)$ can be analyzed in many complex exponential components using Fourier transform:

$$F_a(\Omega_x,\Omega_y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_a(x,y)e^{-i(\Omega_x x+\Omega_y y)}\,dxdy$$

- $F_a(0,0)$: ***DC term*** that is equal to the average image luminance.
- Small spatial image change rate results to power spectrum $|F_a(\Omega_x,\Omega_y)|$ concentrated around the DC term, at low frequencies (Ω_x,Ω_y).
- Image edges and details correspond to higher frequencies (Ω_x,Ω_y) lying further apart from the DC term.

Spatial Frequency Content

a) Test image LENNA; b) periodogram of LENNA.

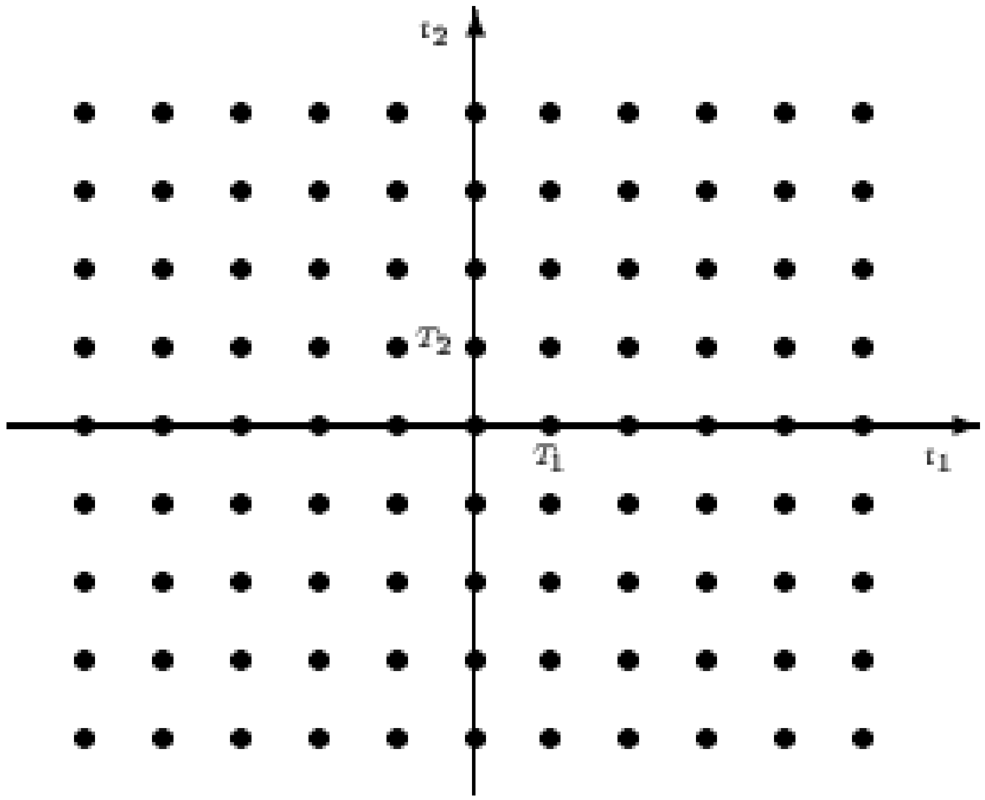
Image sampling

2D image digitization: ***uniform image sampling*** along axes x, y:

- $\Delta x, \Delta y$: the sampling intervals along the two axes (inversely proportional to the horizontal and vertical dpi).
- Usually $\Delta x = \Delta y$, so that image pixels are square.

An analog image is sampled on an ***orthogonal lattice*** resulting in a discrete image $f(n_1, n_2) = f_a(n_1 \Delta x, n_2 \Delta y)$.

2D signal sampling



Rectangular image sampling grid.

Image sampling

- Problems resulting from image sampling:
  - Relationship between the spectra of the continuous and the discrete images.
  - Reconstruction of the continuous images from the discrete one.

- Both can solved by the *2D Fourier transform* of the continuous image.

Image sampling

$$F_a(\Omega_x, \Omega_y) \triangleq \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_a(x,y) exp(-i\Omega_x x - i\Omega_y y) dx dy$$

$$f_a(x,y) = \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F_a(\Omega_x, \Omega_y) exp(i\Omega_x x + i\Omega_y y) d\Omega_x d\Omega_y$$

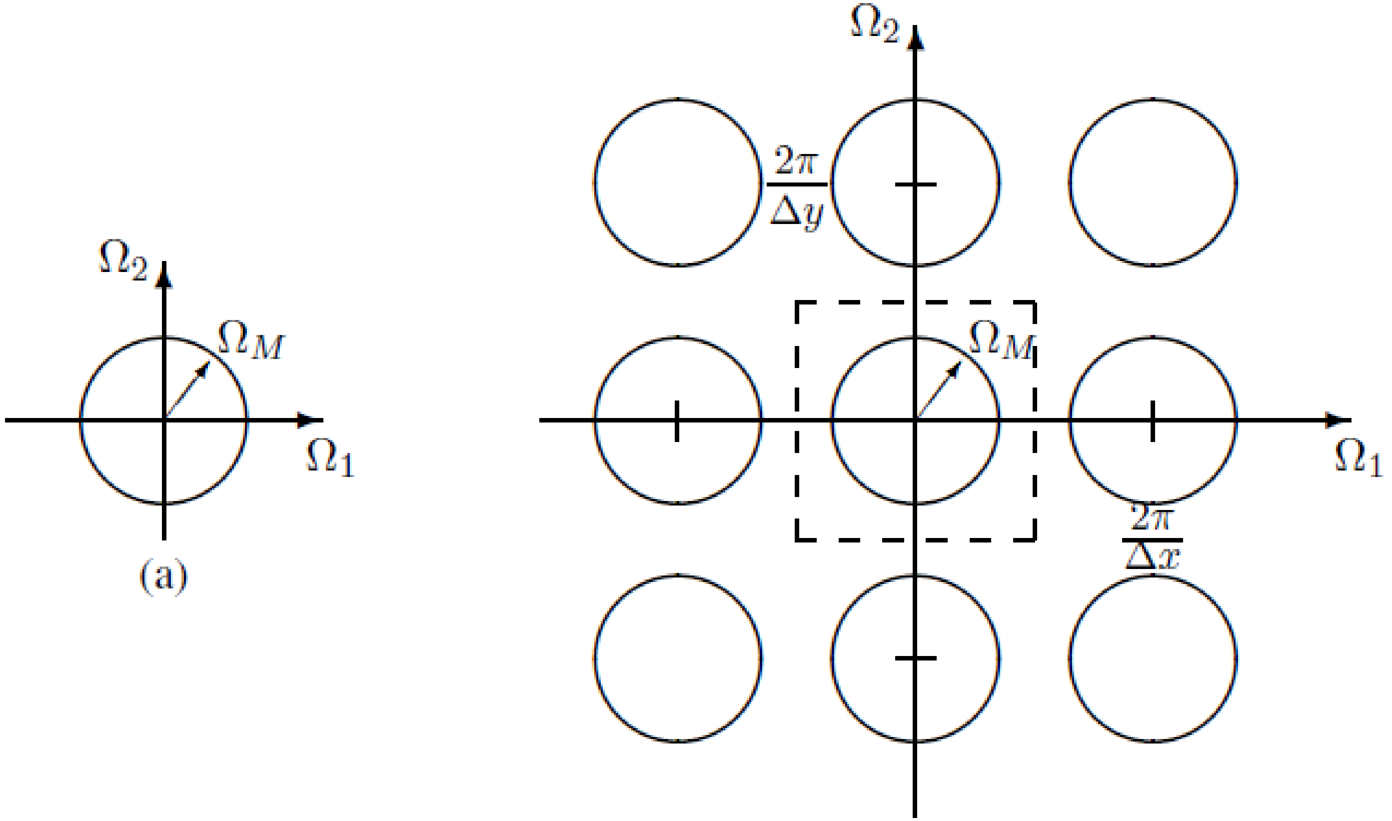
where $\Omega_x = 2\pi F_x,\ \Omega_y = 2\pi F_y$.

- Fourier transform $F_a(\Omega_x, \Omega_y)$ of the discrete image $f(n_1, n_2)$:

$$F(\Omega_x \Delta x, \Omega_y \Delta y) = \frac{1}{\Delta x \Delta y}\sum_{k_1}\sum_{k_2} F_a\left(\Omega_x - \frac{2\pi k_1}{\Delta x}, \Omega_y - \frac{2\pi k_2}{\Delta y}\right)$$

Artificial Intelligence & Information Analysis Lab

Image sampling



Discrete image spectrum: a 2D periodic extension of the continuous image spectrum.

Image sampling

- Spectrum $F_a(\Omega_x, \Omega_y)$ of a low-pass image $f_a(x, y)$ with spectrum contained in a region of the (Ω_x, Ω_y) plane around $(0,0)$ and sufficiently small intervals $\Delta x, \Delta y$:

$$F_a(\Omega_x, \Omega_y) = \Delta x \Delta y\, F(\Omega_x \Delta x, \Omega_y \Delta y),\ |\Omega_x| \leq \frac{\pi}{\Delta x}, |\Omega_y| \leq \frac{\pi}{\Delta y}$$

- Reconstruction of the continuous image $f_a(x, y)$:

$$f_a(x, y) = \sum_{n_1} \sum_{n_2} f(n_1, n_2) \frac{\sin \frac{\pi}{\Delta x}(x - n_1 \Delta x)}{\frac{\pi}{\Delta x}(x - n_1 \Delta x)} \cdot \frac{\sin \frac{\pi}{\Delta y}(y - n_2 \Delta y)}{\frac{\pi}{\Delta y}(y - n_2 \Delta y)}$$

Image sampling

- Analog (continuous-space) image reconstruction from its pixels:
  - It occurs when projecting or displaying a digital image on screen.
  - It is essentially a 2D interpolation operation.
- Other interpolation forms:
  - Zero order polynomial interpolation
  - Linear interpolation.

Image sampling

- ***Nyquist-Shannon sampling theorem***: An accurate reconstruction of a continuous image from the discrete one is possible when the sampling frequencies $\Omega_{sx} = 2\pi/\Delta x$, $\Omega_{sy} = 2\pi/\Delta y$ satisfy:

$$\Omega_{sx} \geq 2\Omega_{xmax}, \ \Omega_{sy} \geq 2\Omega_{ymax}.$$

- Ω_{xmax}, Ω_{ymax}: maximal image frequencies along x, y axis.
- ***Nyquist sampling rate***: sampling intervals should satisfy:

$$\Delta x \leq \frac{\pi}{\Omega_{xmax}}, \qquad \Delta y \leq \frac{\pi}{\Omega_{ymax}}.$$

Image sampling

- Image ***aliasing***:
  - It is caused by alterations of the spectrum, due to its periodic repetitions, primarily in high frequencies, when:
    - the image is not low-pass.
    - the sampling intervals $\Delta x, \Delta y$ are not sufficiently small.
  - It renders exact reconstruction of the continuous image impossible.

Image sampling

- Orthogonal sampling can be extended to other types of sampling lattices by:

$$\mathbf{x} = \mathbf{V}\mathbf{n}$$

where:

$$\mathbf{x} = [x, y]^T, \ \ \mathbf{n} = \left[n_x, n_y\right]^T, \ \mathbf{V} = [\mathbf{v}_1 | \mathbf{v}_2] = \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix}$$

- Vectors $\mathbf{v}_1, \mathbf{v}_2$ must be linearly independent.

Image sampling

- Continuous analog image sampling:

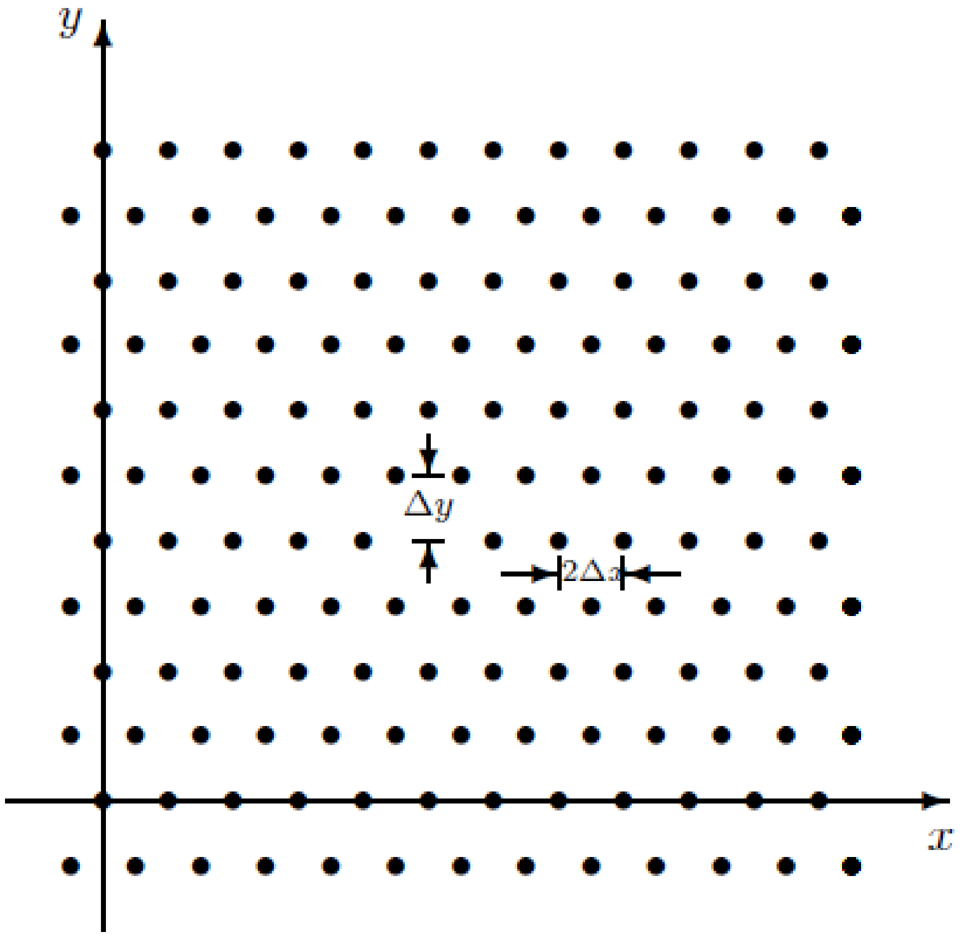
$$f(\mathbf{n}) \triangleq f_\alpha(\mathbf{V}\mathbf{n}).$$

- Rectangular sampling matrix:

$$\mathbf{V} = \begin{bmatrix} \Delta x & 0 \\ 0 & \Delta y \end{bmatrix}.$$

- In a square lattice with $\Delta x = \Delta y = 1$, the eight neighbors of a pixel may be 1 or $\sqrt{2}$ apart.

Image sampling



Hexagonal lattice.

Image sampling

- ***Hexagonal lattice***: every pixel has six equally distant neighbors.
  - Sampling matrix:

$$\mathbf{V} = \begin{bmatrix} \Delta x & \Delta x \\ \Delta y & -\Delta y \end{bmatrix}.$$

  - Curves are uniformly sampled on a hexagonal lattice.

Image sampling

- Discrete – continuous image spectrum relation:

$$F(\mathbf{V}^T\mathbf{\Omega}) = \frac{1}{|\det \mathbf{V}|}\sum_{\mathbf{k}} F_\alpha(\mathbf{\Omega} - \mathbf{U}\mathbf{k}),$$

where:

$$\mathbf{k} = [k_1, k_2]^T, \quad \mathbf{\Omega} = \left[\Omega_x, \Omega_y\right]^T, \quad \mathbf{U}^T\mathbf{V} = 2\pi\mathbf{I}.$$

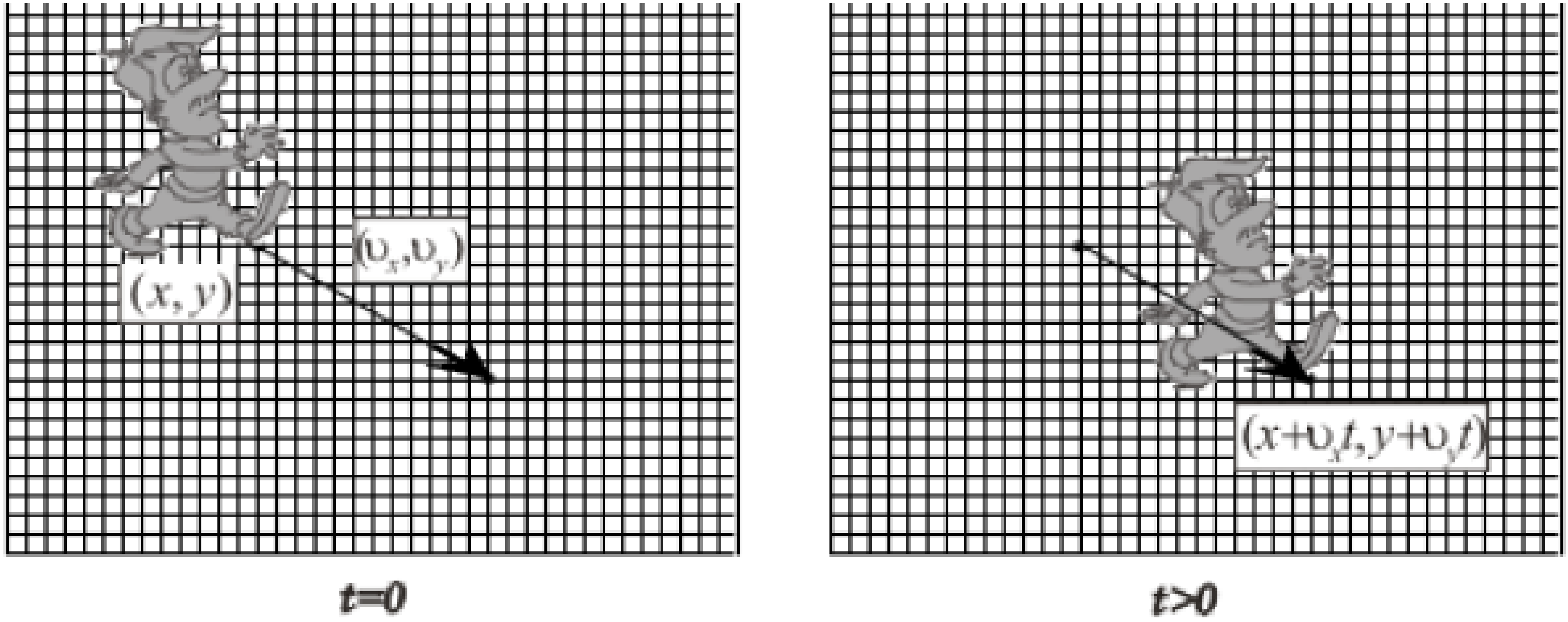
- Periodic spectrum with periodicity matrix **U.**
- Less severe aliasing problems.
- Better quality images for same CCD chip area.

Artificial Intelligence & Information Analysis Lab

Spatiotemporal Frequency Content

- In video signal, consisting of 2D video frames changing over time, we have spatiotemporal frequencies $\Omega_x, \Omega_y, \Omega_t$.
- Temporal frequency Ω_t depends on temporal video content changes, primarily due to object motion.
- Possible reasons of video content changes:
  - object motion,
  - camera motion,
  - illumination changes,
  - combination of all the above.

Spatiotemporal Frequency Content



Linear object motion.

Artificial Intelligence & Information Analysis Lab

Spatiotemporal Frequency Content

- Let $f_0(x, y) = f(x, y, 0)$ be an object image at time zero, and v_x, v_y be the object speed along the horizontal and vertical directions respectively.

- The object image at time t is given by:

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

Spatiotemporal Frequency Content

- The spatiotemporal Fourier transform $F(\Omega_x, \Omega_y, \Omega_t)$ of the video signal $f(x, y, t)$ is given by:

$$F(\Omega_x, \Omega_y, \Omega_t) = F_0(\Omega_x, \Omega_y)\delta(\Omega_t + \Omega_x v_x + \Omega_y v_y).$$

- $\delta(\cdot)$: the delta Dirac function.
- The spectrum $F(\Omega_x, \Omega_y, \Omega_t)$ is nonzero only on the plane:

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

Spatiotemporal Frequency Content

Therefore:

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

$\boldsymbol{\Omega} = [\Omega_x, \Omega_y]^T$: frequency vector.

$\mathbf{v} = [v_x, v_y]^T$: motion vector.

- The temporal frequency depends not only on motion speed, but also on the spatial object frequencies.

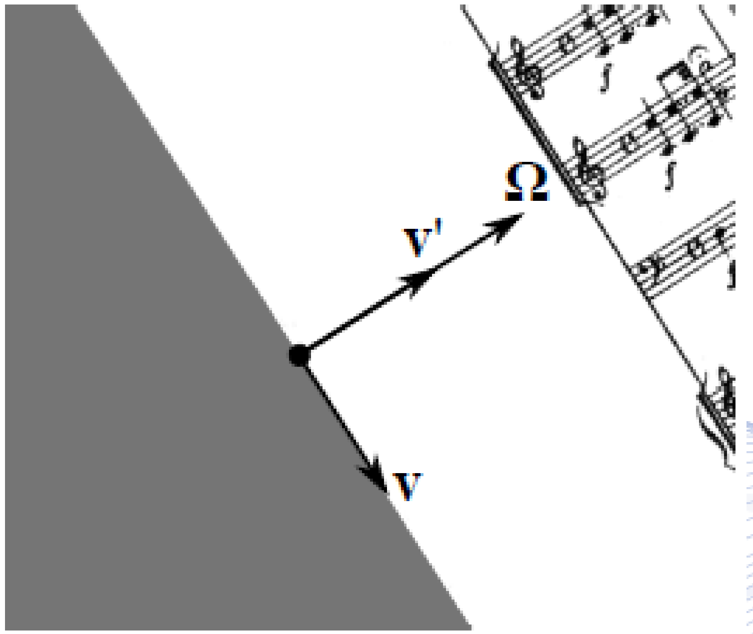
Spatiotemporal Frequency Content

- When $\Omega_x = \Omega_y = 0$, $\Omega_t = 0$ regardless of v_x, v_y.
  - If the object has uniform luminance, no temporal variation can be observed when it moves.

- If the motion vector $[v_x, v_y]^T$ is orthogonal to the spatial frequency vector $\mathbf{\Omega}$, then $\Omega_t = 0$.
  - The direction of the maximal spatial luminance variations is the same as the direction of $\mathbf{\Omega}$, i.e., perpendicular to the local image edges.

Artificial Intelligence & Information Analysis Lab

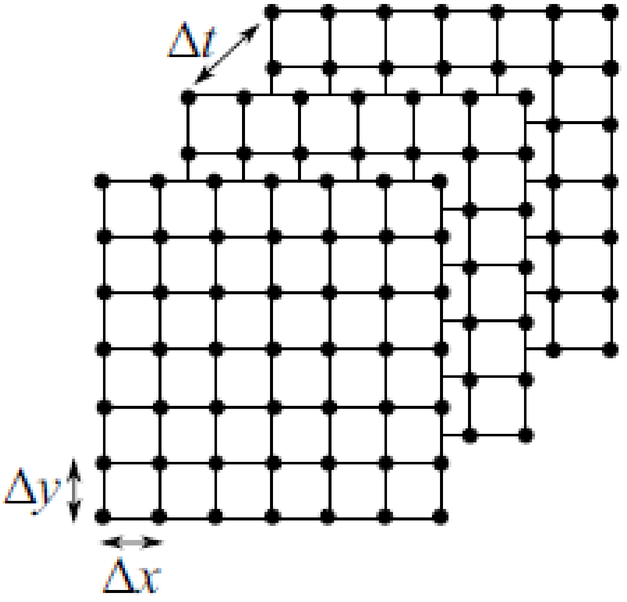
Spatiotemporal Frequency Content



- If an object moves along a direction and the object pattern does not change, it will not produce any temporal variation.

- Temporal frequency is maximal when the object moves along a direction where spatial luminance change is the greatest.

Video sampling



- Analog video signal:
  - a time-varying image of the form $f(x, j\Delta y, k\Delta t)$
  - It is obtained by *video scanning*: sampling the time-varying image luminance along the y and t axis.
- Digital video $f(i\Delta x, j\Delta y, k\Delta t)$ can be obtained by:
  - sampling the analog video along the horizontal scan lines or
  - using the existing discrete two-dimensional sampling grip that is inherent in several photoelectrical sensors (e.g., CCD chips).

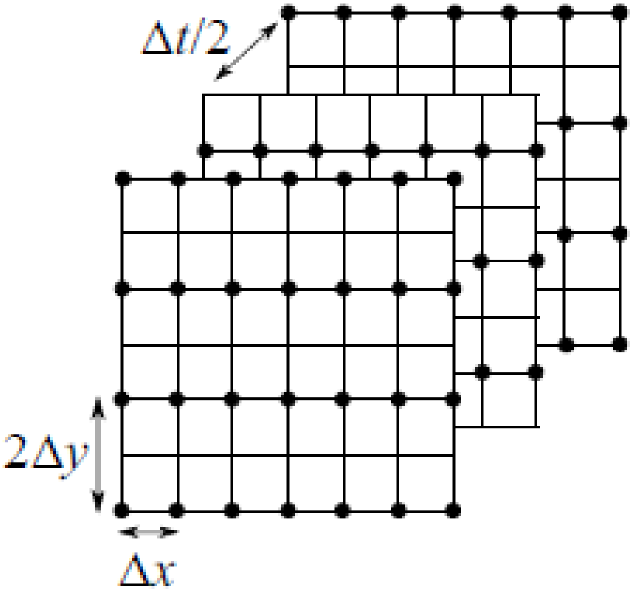
Video sampling

- ***Progressive video sampling grid***:
  - It is the simplest way to digitize an analog 3D video signal.
  - It leads to uniform spatiotemporal sampling along three space-time coordinates x, y, t.

- Progressive digital video consists of *video frames*:
  - SDTV PAL system: 25 fps, $\Delta t = 1/25$, video frame resolution 480×720 pixels.
  - SDTV NTSC system: 30 fps, $\Delta t = 1/30$, video frame resolution 576×720 pixels.

Video sampling

- HDTV digital video 1080p offers 1080×1920 pixels per video frame both for ATSC and DVB systems.
- Alternative form of digital video sampling: ***2:1 interlaced video***
  - It samples the odd- and even-numbered video lines alternatively.
  - It produces odd and even *video fields* at double the sampling rate per second.
  - Two video fields can form one video frame (when motion is small).



Video sampling

- Let $f_a(x, y, t)$ be the 3D analog continuous signal and $\Delta x, \Delta y, \Delta t$ be the ***sampling intervals*** along x, y, t axes.
- The discrete video is given by:

$$f(n_1, n_2, n_t) = f_a(n_1 \Delta x, n_2 \Delta y, n_t \Delta t).$$

- Forward 3D Fourier transform of continuous video:

$$F_a(\Omega_x, \Omega_y, \Omega_t) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_a(x, y, t) e^{-i\Omega_x x - i\Omega_y y - i\Omega_t t} dxdydt$$

Video sampling

- The inverse 3D Fourier transform of continuous video:

$$f_a(x,y,t) = \frac{1}{(2\pi)^3} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} F_a(\Omega_x, \Omega_y, \Omega_t) e^{i\Omega_x x + i\Omega_y y + i\Omega_t t} d\Omega_x d\Omega_y d\Omega_t$$

$\Omega_x = 2\pi F_x,\ \Omega_y = 2\pi F_y,\ \Omega_t = 2\pi F_t$ are the spatiotemporal frequencies describing video content variations along axes x, y, t.

Video sampling

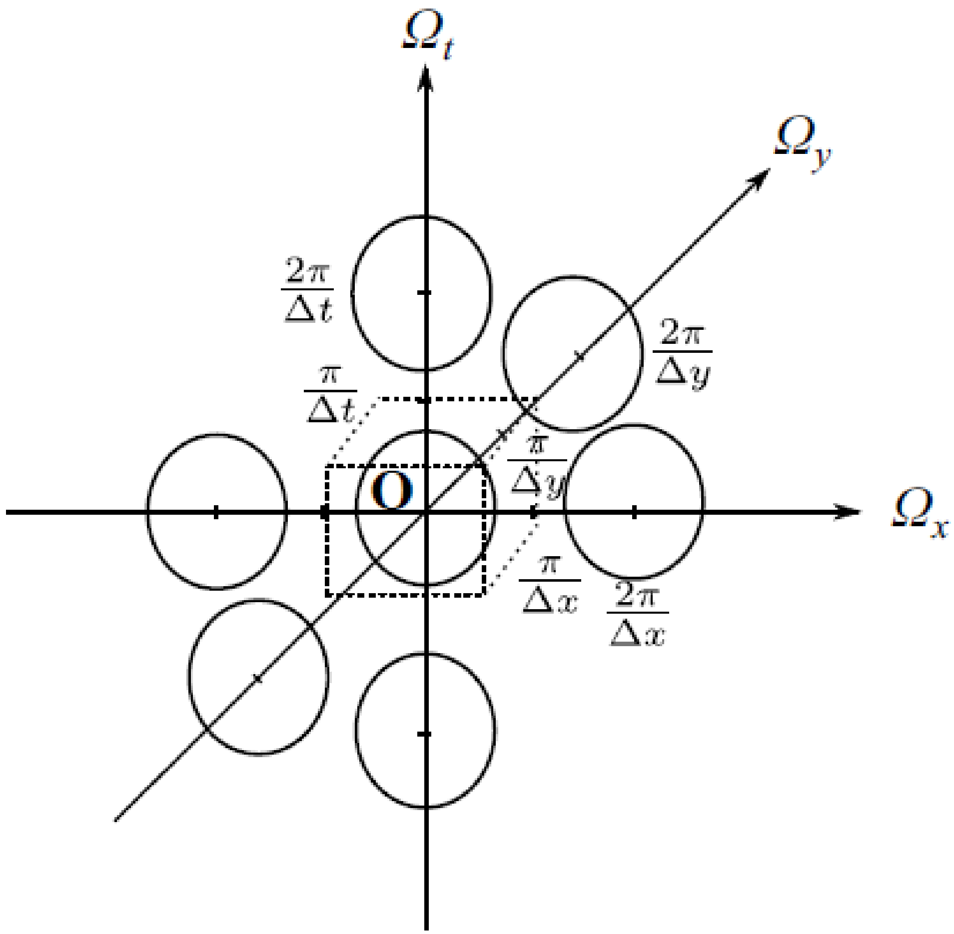
- Spatiotemporally smooth video content has power spectrum concentrated around the ***DC term*** $[\Omega_x, \Omega_y, \Omega_t]^T = [0,0,0]^T$.

- Fourier transform $F(\Omega_x, \Omega_y, \Omega_t)$ of the discrete video signal $f(n_1, n_2, n_3)$:

$$F(\Omega_x \Delta x, \Omega_y \Delta y, \Omega_t \Delta t)$$
$$= \frac{1}{\Delta x \Delta y \Delta t} \sum_{k_x} \sum_{k_y} \sum_{k_t} F_a\left(\Omega_x - \frac{2\pi k_x}{\Delta x}, \Omega_y - \frac{2\pi k_y}{\Delta y}, \Omega_t - \frac{2\pi k_t}{\Delta t}\right).$$

Video sampling

- Discrete progressive video spectrum: a 3D periodic translation of the continuous video spectrum.



Video sampling

- Spectral ***aliasing***:
  - caused by spectrum overlapping, due to its periodic repetitions, primarily in high frequencies (close to $\pm\pi/\Delta x, \pm\pi/\Delta y, \pm\pi/\Delta t$), if:
    - the video is not low-pass,
    - the sampling intervals $\Delta x, \Delta y$ are not sufficiently small.
  - It renders accurate reconstruction of the continuous video impossible.

Video sampling

- ***Nyquist criterion***: The accurate reconstruction of a continuous video from the discrete one is possible when the sampling periods $\Delta x, \Delta y, \Delta t$ satisfy:

$$\Delta x \leq \frac{\pi}{\Omega_{xmax}}, \quad \Delta y \leq \frac{\pi}{\Omega_{ymax}}, \quad \Delta t \leq \frac{\pi}{\Omega_{tmax}}.$$

- Therefore, sampling frequencies $F_{sx} = \frac{1}{\Delta x}, F_{sy} = \frac{1}{\Delta y}, F_{st} = \frac{1}{\Delta t}$ must be at least double the maximal spatial and temporal video frequencies: $F_{sx} \geq 2F_{xmax}$, $F_{sy} \geq 2F_{ymax}$, $F_{st} \geq 2F_{tmax}$.

Video sampling

- Let a low-pass signal $f_a(x, y, t)$ with spectrum contained in a region of the $[\Omega_x, \Omega_y, \Omega_t]^T$ plane around $\mathbf{0} = [0,0,0]^T$ and the intervals $\Delta x, \Delta y, \Delta t$ are sufficiently small, so that:

$$F_a(\Omega_x, \Omega_y, \Omega_t) = 0, \quad |\Omega_x| \geq \frac{\pi}{\Delta x}, \quad |\Omega_y| \geq \frac{\pi}{\Delta y}, \quad |\Omega_t| \geq \frac{\pi}{\Delta t}.$$

Video sampling

- Continuous video spectrum:

$$F_a(\Omega_x, \Omega_y, \Omega_t) = \Delta x \Delta y \Delta t \, F(\Omega_x \Delta x, \Omega_y \Delta y, \Omega_t \Delta t),$$

$$|\Omega_x| \leq \frac{\pi}{\Delta x}, \ |\Omega_y| \leq \frac{\pi}{\Delta y}, \ |\Omega_t| \leq \frac{\pi}{\Delta t}$$

- Continuous video reconstruction:

$$f_\alpha(x, y, t) = \sum_{n_1} \sum_{n_2} \sum_{n_t} f(n_1, n_2, n_t) \frac{\sin \frac{\pi}{\Delta x}(x - n_1 \Delta x)}{\frac{\pi}{\Delta x}(x - n_1 \Delta x)} \cdot \frac{\sin \frac{\pi}{\Delta y}(y - n_2 \Delta y)}{\frac{\pi}{\Delta y}(y - n_2 \Delta y)} \cdot \frac{\sin \frac{\pi}{\Delta t}(t - n_t \Delta t)}{\frac{\pi}{\Delta t}(t - n_t \Delta t)}$$

Bibliography

[PIT2019] I. Pitas, “Computer vision”, Createspace/Amazon, in press.
[SZE2011] R.Szelinski, “ Computer Vision ” , Springer 2011
[HAR2003] Hartley R, Zisserman A. , “ Multiple view geometry in computer vision” . Cambridge university press; 2003.
[DAV2017] Davies, E. Roy. “Computer vision: principles, algorithms, applications, learning ”. Academic Press, 2017
[TRU1998] Trucco E, Verri A. “Introductory techniques for 3-D computer vision”, Prentice Hall, 1998.
[PIT2017] I. Pitas, “Digital video processing and analysis ” , China Machine Press, 2017 (in Chinese).
[PIT2013] I. Pitas, “Digital Video and Television ” , Createspace/Amazon, 2013.
[PIT2000] I. Pitas, Digital Image Processing Algorithms and Applications, J. Wiley, 2000.
[NIK2000] N. Nikolaidis and I. Pitas, 3D Image Processing Algorithms, J. Wiley, 2000.
[APOLLO] http://apolloscape.auto/

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