Course Motivation

Directly-captured images
Computational images
Historical timeline
Peaking into the future
Digital image processing?
Digital image communication?

What is an image?

What is an image?

- A 2-dimensional (2-D) array of data representing some quantity captured by a *light-sensitive device*.
- This array **is not** merely a set of numbers – it is a set of numbers with some meaning (depending on what physical property being imaged).
- One must be aware of this property and its meaning to carry out effective image processing or analysis.

What is an image?

Captured by physical device or manipulated?

- A 2-dimensional (2-D) array of data representing some quantity captured by a *light-sensitive device*.
- This array **is not** merely a set of numbers – it is a set of numbers with some meaning (depending on what physical property being imaged).
- One must be aware of this property and its meaning to carry out effective image processing or analysis.

Directly-captured images

A photograph, TV picture, movie frame, document scan, X-ray film,...

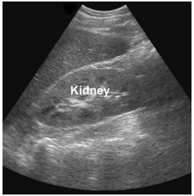
Cameraman (Copyright MIT)

X-ray (Copyright J. Cluett, M.D.)

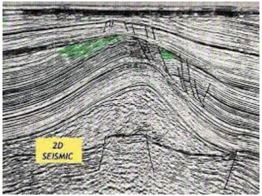
In this course, we will be primarily concerned with visual images, i.e., directly captured by a sensor, like the images shown above.

Computational images

Data captured by a sensor in: *ultrasound exam*, *seismology*, *computerized tomography* (CT), *magnetic-resonance imaging* (MRI), *positive-emission tomography* (PET), *mini microscopy*, ... **are not visually meaningful without additional computations**. The raw data must be processed by sensor-specific algorithms to produce visual images like this:



Ultrasound image



Seismological data

3D mini microscopy (Prof. Tian's lab)

Due to the tight integration of sensing and computing this is called **computational imaging**. We will touch on the topic but will not cover in depth – consider EC522 "Introduction to Computational Imaging" with Prof. Lei Tian.

Why image processing AND communication?

Images/video are captured for various applications:

- family archives (general photography, selfies),
- entertainment (YouTube, Netflix, TV, cinema),
- surveillance (airports, highways, power stations),
- medical diagnostics (X-ray, ultrasound, CT, MRI, PET),
- atmospheric/geological/ecological/oceanographic analysis,
- military, ...

Why image processing AND communication?

Images/video are captured for various applications:

- family archives (general photography, selfies),
- entertainment (YouTube, Netflix, TV, cinema),
- surveillance (airports, highways, power stations),
- medical diagnostics (X-ray, ultrasound, CT, MRI, PET),
- atmospheric/geological/ecological/oceanographic analysis,
- military, ...

that require various forms of *image processing and analysis*, such as:

- re-sizing, noise reduction, quality enhancement,
- segmentation, object detection and recognition, ...

Why image processing AND communication?

Images/video are captured for various applications:

- family archives (general photography, selfies),
- entertainment (YouTube, Netflix, TV, cinema),
- surveillance (airports, highways, power stations),
- medical diagnostics (X-ray, ultrasound, CT, MRI, PET),
- atmospheric/geological/ecological/oceanographic analysis,
- military, ...

that require various forms of *image processing and analysis*, such as:

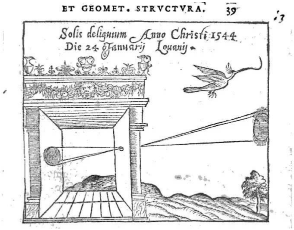
- re-sizing, noise reduction, quality enhancement,
- segmentation, object detection and recognition, ...

and *image communication* (to transmit data to a destination), such as:

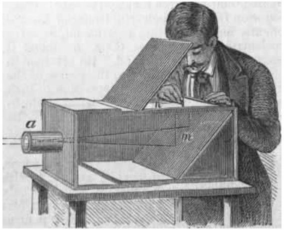
- compression, bit-rate scaling, transmission-error concealment, ...

A little history of images: from wall to film

- 400 BC - observation of pinhole effect (Mozi, China)
- 1502 - explanation of pinhole effect (da Vinci, Italy)
- late 1500s - lens replaces pinhole (camera obscura)
- 1840 - photographic film (L. Daguerre, France, W. Talbot, USA)
- 1895 - first public motion picture (Lumière brothers, France)
- 1920s - motion picture with sound
- 1930s - color movies
- 1960s - multichannel sound
- 1970s - huge-screen cinema (large field of view), Imax Corp., Canada



Homework #1: build your own pinhole camera



Camera obscura

Paintings before pictures: a Venice canal by Canaletto, 1725

Very accurate perspective, incredible detail for the era - how?

Paintings before pictures: a Venice canal by Canaletto, 1725

Very accurate perspective, incredible detail for the era - how?

The artist used *camera obscura* to produce a small-scale view to be traced by hand.

A little history of images: from electrons to bits

- 1920s - first TV experiments (Nipkov disk, Russia)
- 1939 - first regular B/W TV service (USA)
- 1954/67 - introduction of color TV (USA/Europe)
- 1970s - video cassette recorder or VCR (USA)
- early 1980s - consumer (analog) laser disk player (Japan)
- late 1980s - analog direct broadcasting by satellite (Europe)
- mid 1990s - digital direct broadcasting by satellite (USA)
- late 1990s - high-definition (digital) TV (USA, Europe)
- 1999 - digital cinema (USA)
- 2000s - digital cameras, video streaming, smartphones, ...

1929 Western Television (USA)
Scanning Disc Television

Laser disk player

Lessons from the historical time-line

- **Rapid expansion** of imaging technologies in the last three decades
- **Turning point**: introduction of **digital capture**, **storage** and **transmission** of images
- **New applications** of imaging will arise in the near future due to:
  - continuing advances in digital capture (multi-camera, multi-lens, multi-focus, multi-modal, ...),
  - huge advances in AI/deep learning fueled by GPU computing and vast databases,
  - rapidly increasing transmission and storage capabilities,
  - flexibility in manipulating digitally, as compared to film,
  - human ingenuity.

What is this?

Lessons from the historical time-line

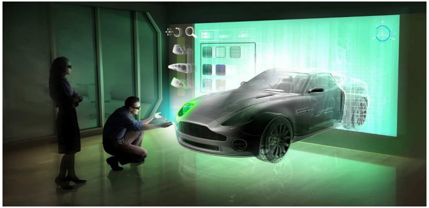
- **Rapid expansion** of imaging technologies in the last three decades
- **Turning point**: introduction of **digital capture**, **storage** and **transmission** of images
- **New applications** of imaging will arise in the near future due to:
  - continuing advances in digital capture (multi-camera, multi-lens, multi-focus, multi-modal, ...),
  - huge advances in AI/deep learning fueled by GPU computing and vast databases,
  - rapidly increasing transmission and storage capabilities,
  - flexibility in manipulating digitally, as compared to film,
  - human ingenuity.

What is this?

Click to learn

Today's challenges

- **Visual scene understanding:** How to recognize objects and environments for navigation, tourism, entertainment?
- **Visual surveillance:** How to distinguish anomalies from normality, predict intent?
- **Visual databases:** How to retrieve images based on appearance instead of meta-data (keywords)?
- **Video over IP or mobile:** How to preserve visual quality when internet connection gets congested or smartphone "hides" between buildings?
- **Natural user interfaces (NUIs):** How to interact with a system by gestures instead of mouse and keyboard?
- **3-D visual communications:** How to capture, communicate and display 3-D surroundings?



3-D cinema
Virtual reality
Augmented reality

Tomorrow's solutions?

Surveillance: Pervasive, intelligent camera networks

- Stationary versus mobile (drone)
- Wired versus wireless
- Grid-powered, battery-powered, solar-powered

Goal: each camera equipped with sufficient intelligence to perform inference for indoor monitoring, urban surveillance, natural habitat monitoring, ... without transmitting images.

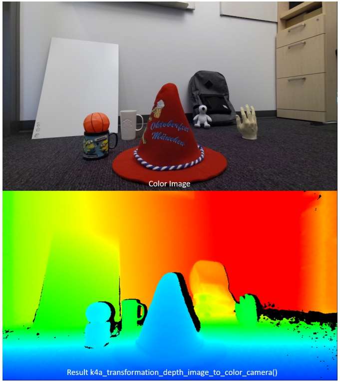
http://vsn1-ids.bu.edu, http://vsn2-ids.bu.edu
http://ptz1-ids.bu.edu, http://ptz2-ids.bu.edu

Tomorrow's solutions?

3D capture and interaction: Novel cameras

Various structured-light or time-of-flight technologies:

- Microsoft Kinect
- Asus Xtion
- Creative Senz3D
- Intel RealSense,
- Google Tango project
- recent smartphones with depth sensor
- ...



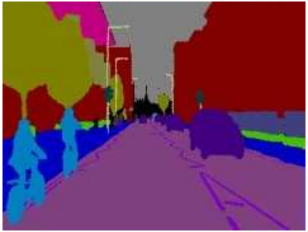
are already used for gaming, authentication, navigation. What's next?

Tomorrow's solutions?

Autonomous navigation: Novel algorithms

Sophisticated AI (deep learning) algorithms to:

- detect and/or recognize vehicles,
- detect pedestrians, cyclists,
- identify landmarks (buildings, monuments),
- ...



See this video: *Mask R-CNN*

Goal 1 of this class: Digital Image Processing and Analysis

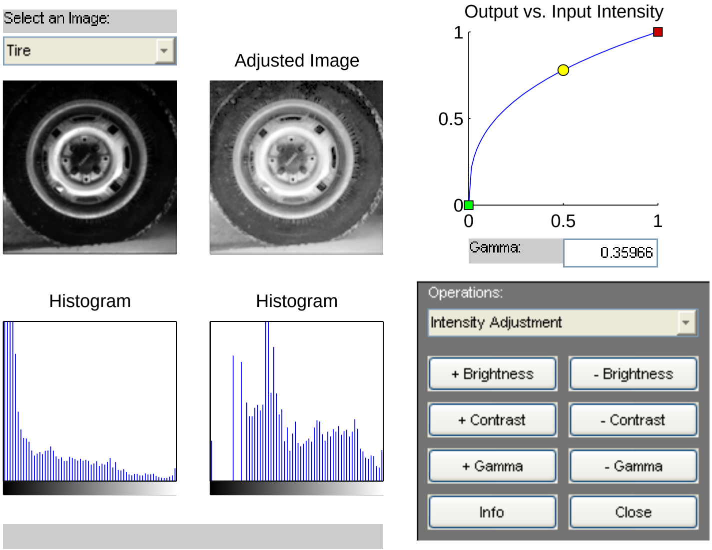
Development of numeric algorithms to perform (among others):

- image resizing (zoom-in, zoom-out),
- change of image resolution (level of detail),
- suppression of noise, quality improvement (enhancement),
- extraction of features from the image (edges, boundaries),
- analysis of the image (segmentation, object detection),
- understanding of the image (object recognition),

often under constraint of **lowest complexity**.

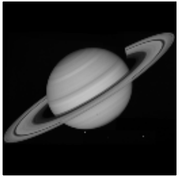
In many tasks there is no objective/numerical criterion! For example, in image segmentation, there is no *single* correct answer.

Example I: Contrast enhancement



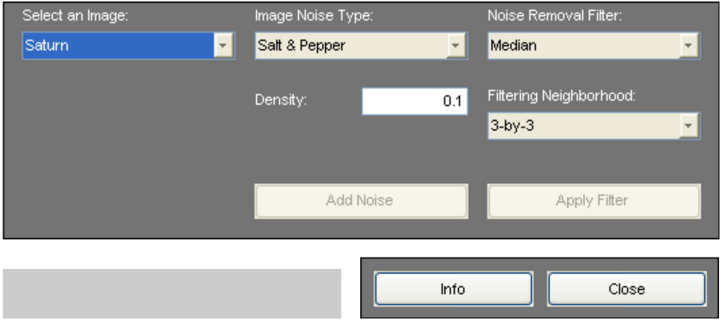
Example II: Noise reduction

Original Image

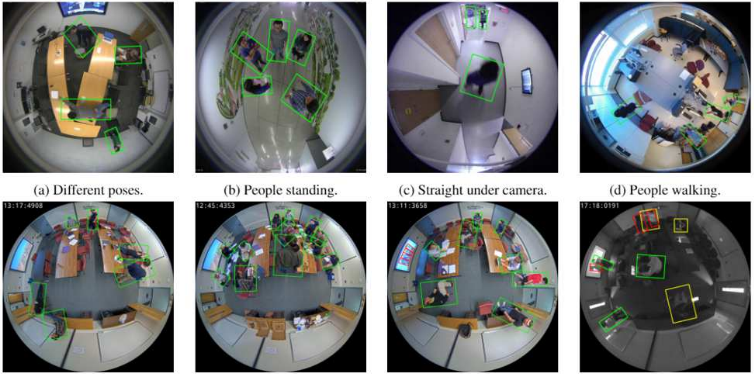


Corrupted Image

Filtered Image



Example III: Image understanding – finding people



(a) Different poses. (b) People standing. (c) Straight under camera. (d) People walking.

(e) Various angles. (f) Crowded scene + occlusions. (g) People on the screen. (h) Low-light scenario.

- Localization of people from overhead fisheye cameras
- Deep-learning RAPiD algorithm: Z. Duan (MS, 2020), O. Tezcan, H. Nakamura
- Details at: RAPiD, Exhibition video demo

Goal 2 of this class: Digital Image Communication

Development of numeric algorithms to transform digital image into a string of bits for:

- transmission from a transmitter to receiver(s), or
- storage on a magnetic, optical, silicon or other medium,

subject to the following constraints:

- quality,
- rate (transmission) or volume (storage),
- algorithm complexity (speed, implementation cost).

Digital image processing algorithms are often exploited in digital image communication.

Possible objective/numerical criteria:

- maximum image quality for a given transmission rate, or
- minimum transmission rate for a given image quality.

Why is image communication research needed?

Picture is worth a thousand words

Color images (3 components: RGB) convey a lot of information but are storage-hungry.
Recall: 1 MB = 1 Mbyte, 1 Mb = 1 Mbit, 1 byte = 8 bits.

- Early flip-phone camera: 720×576×3 ⇒ 1.2 MB (10 Mb)
- Early smartphone camera: 1,600×1,200×3 ⇒ 5.8 MB (46 Mb)
- High-definition TV video frame: 1,920×1,080×3 ⇒ 6.2 MB (50 Mb)
- 4K video frame: 3,840×2,160×3 ⇒ 24.9 MB (199 Mb)
- Modern smartphone camera: 4,032×3,024×3 ⇒ 36.6 MB (293 Mb)
- Nikon D850 prosumer DSLR camera: 8,256×5,504×3 ⇒ **136.3 MB** (1,090 Mb)

10 Nikon D850 images → 1.36 GB storage needed !

How to compress ?

Images (and video) contain a lot of redundancy:

- **spatial redundancies** – neighboring pixels are similar,
- **temporal redundancies** (video only) – consecutive video frames are similar,
- **statistical redundancies** – some pixel values are more frequent than others and can be handled more efficiently.

Human vision is far from perfect (human eye is incapable of perceiving certain details):

- **psychovisual irrelevancies**.

Image communication exploits these redundancies/irrelevancies by applying digital processing techniques to reduce data size.

How to compress ?

Images (and video) contain a lot of redundancy:

- **spatial redundancies** – neighboring pixels are similar,
- **temporal redundancies** (video only) – consecutive video frames are similar,
- **statistical redundancies** – some pixel values are more frequent than others and can be handled more efficiently.

Human vision is far from perfect (human eye is incapable of perceiving certain details):

- **psychovisual irrelevancies**.

Image communication exploits these redundancies/irrelevancies by applying digital processing techniques to reduce data size.

Tradeoff: data size *versus* image quality

Example VI: Image compression

How to compress pictures:

- *effectively*, i.e., to produce high compression ratio ?
- *efficiently*, i.e., to do this in real time ?

Example: comparison of JPEG with JPEG-2000 ($\approx$50-fold compression)

Original = 238kB

JPEG (1992) = 5kB

JPEG-2000 = 5kB

Course outline

- Two-dimensional continuous signals and systems (straightforward extension of 1-D)
- Discrete image representation:
  - image sampling & quantization, human vision, color spaces
  - image models and transforms
- Digital image processing:
  - image re-sizing, color transformation
  - image enhancement and restoration
  - image analysis: boundary extraction, segmentation, recognition
- Digital image compression:
  - basic notions from information theory, entropy coding,
  - still image coding (JPEG, and foundations of JPEG-2000).

I will touch on Convolutional Neural Networks (CNNs) for image analysis but I will not go into details. Consider EC523 "Deep learning" to learn about CNNs.