IERG4190 / IEMS5707

Multimedia Coding and Processing / Multimedia Coding Applications

Chapter 1: Introduction

Bolei Zhou
Department of Information Engineering, CUHK
Spring 2021

Part of the materials courtesy of Dr. William Hui, Prof Tang Xiaoou, Prof. Liu JianZhuang

About myself

- http://bzhou.ie.cuhk.edu.hk/
- Research on computer vision and machine learning

Shen, Gu, Tang, Zhou. CVPR'20

https://genforce.github.io/interfacegan/

Shen and Zhou. arXiv

https://genforce.github.io/sefa/

Introduction

multimedia

/ˈmʌltɪmiːdɪə/

adjective

adjective: **multimedia**; adjective: **multi-media**

1. (of art, education, etc.) using more than one medium of expression or communication.
   - (of computer applications) incorporating audio and video, especially interactively. "multimedia applications"

What This Course is About:

- An introductory course to the wide area of multimedia coding, processing, and understanding
  - Multimedia: everything related to pictures, videos, sounds etc.
- Study of multimedia processing has a long history
  - Computers and digital systems started with limited multimedia

Multimedia? What Multimedia?

- Young Steve Jobs introduces the Macintosh in 1984

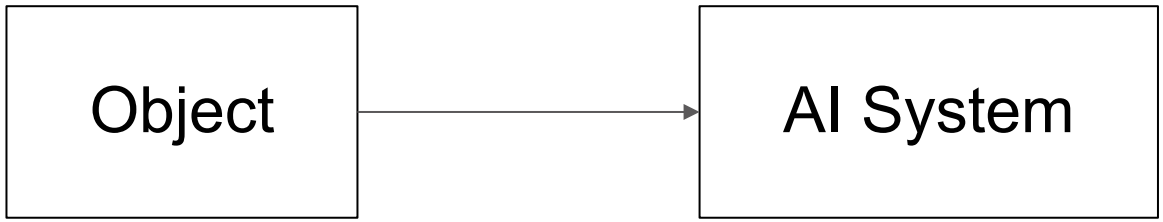
https://www.youtube.com/watch?v=2B-XwPjn9YY

What This Course is About:

- The Internet has transformed the landscape forever
  - "Who reads anyway?! Give me pictures and videos!"
  - Internet giants - Facebook, Google etc. are all big on multimedia
- To provide a proper treatment, we need to cover both the past and the present, as they are closely tied together

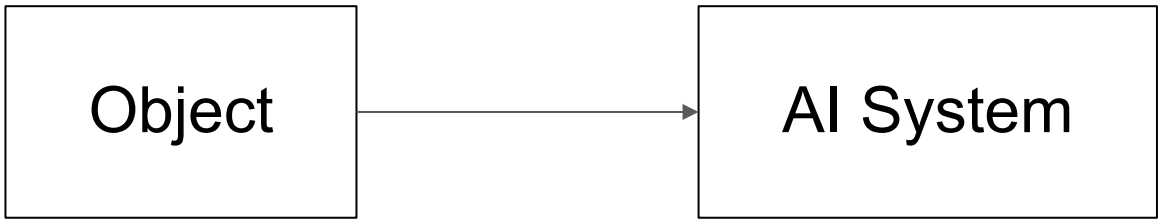
A Modern Computer Vision System

AI is hot! Everyone wants to know AI and computer vision. Is this a correct system diagram for a modern computer vision system?



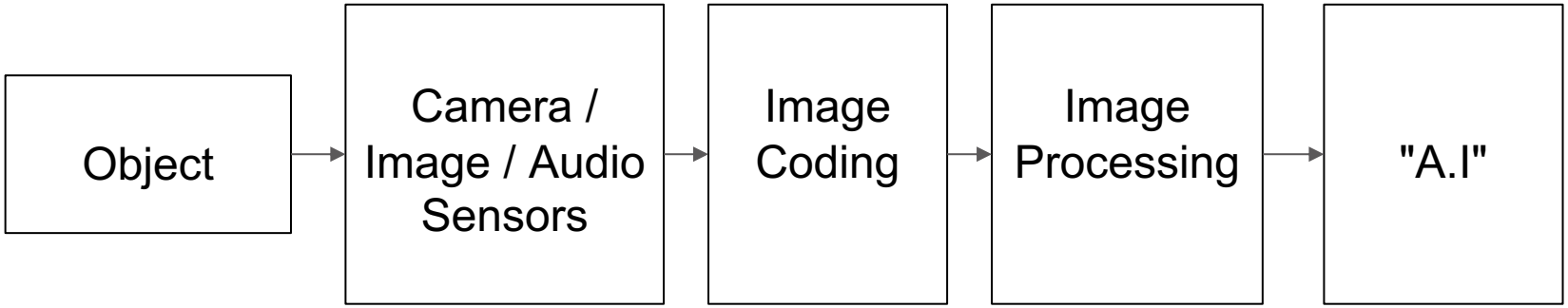
A Modern Computer Vision System

AI is hot! Everyone wants to know AI and computer vision. Is this a correct system diagram for a modern computer vision system?

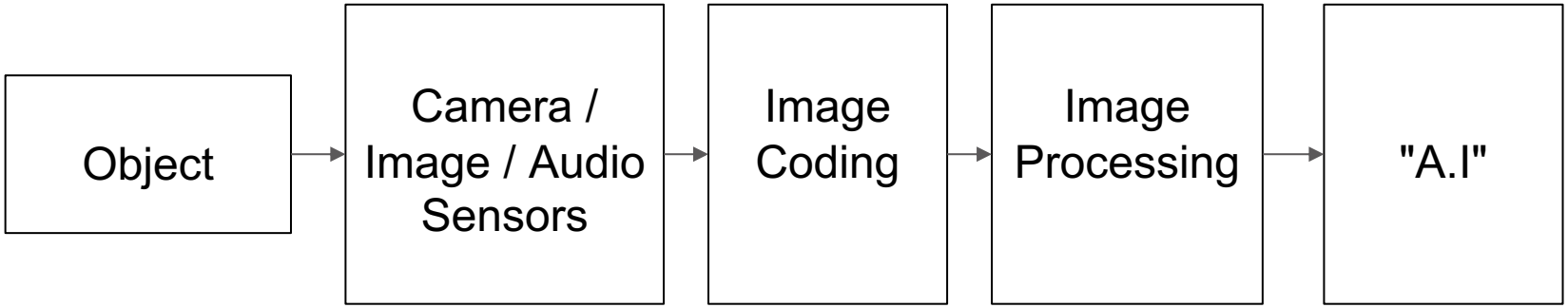


Capturing of image data?
Digitization?
Optimal data representation?
Efficiency? Cost?

A Modern Computer Vision System



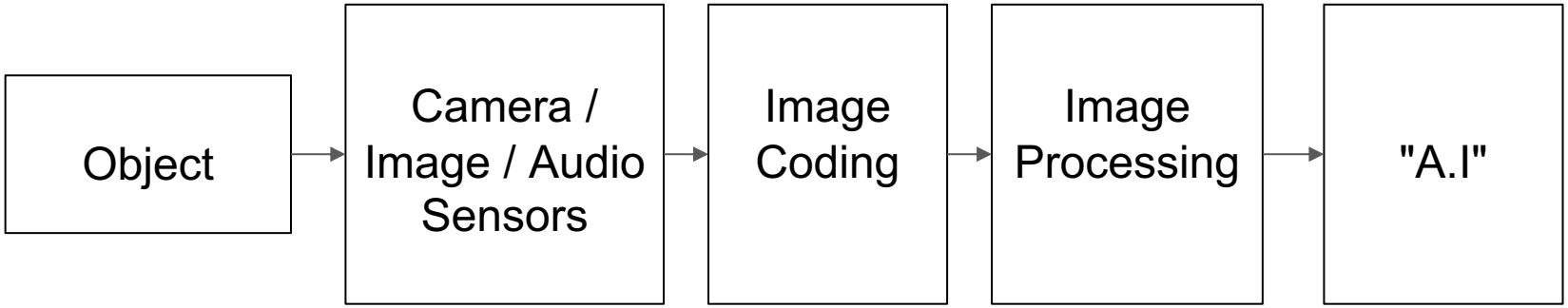
A Modern Computer Vision System



Capture visual and audio information and transform into digital data;
Still an area with innovations - increasing resolution and power

How's your phone's camera?

A Modern Computer Vision System

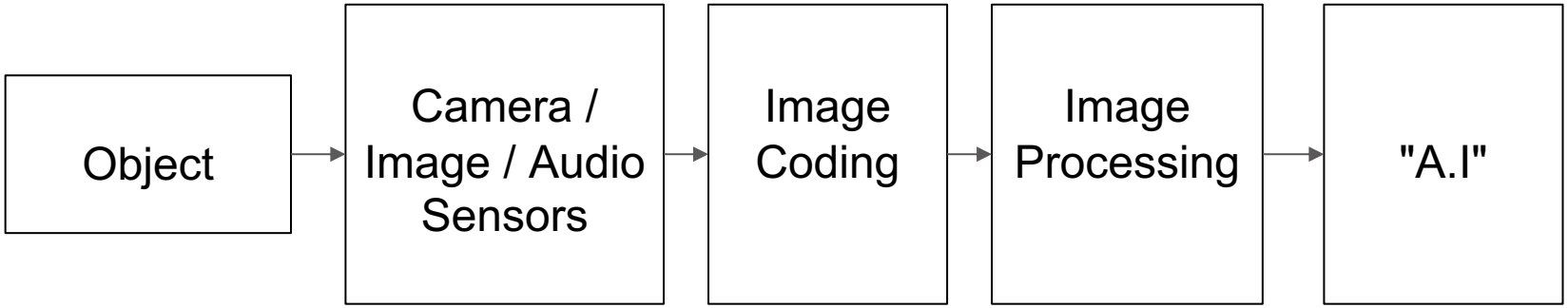


Capture visual and audio information and transform into digital data;
Still an area with innovations - increasing resolution and power

As signals, multimedia data need to be coded efficiently and conform to standards

Raw image or audio data are often not suitable for immediate use.
Noises?
Normalization?

A Modern Computer Vision System



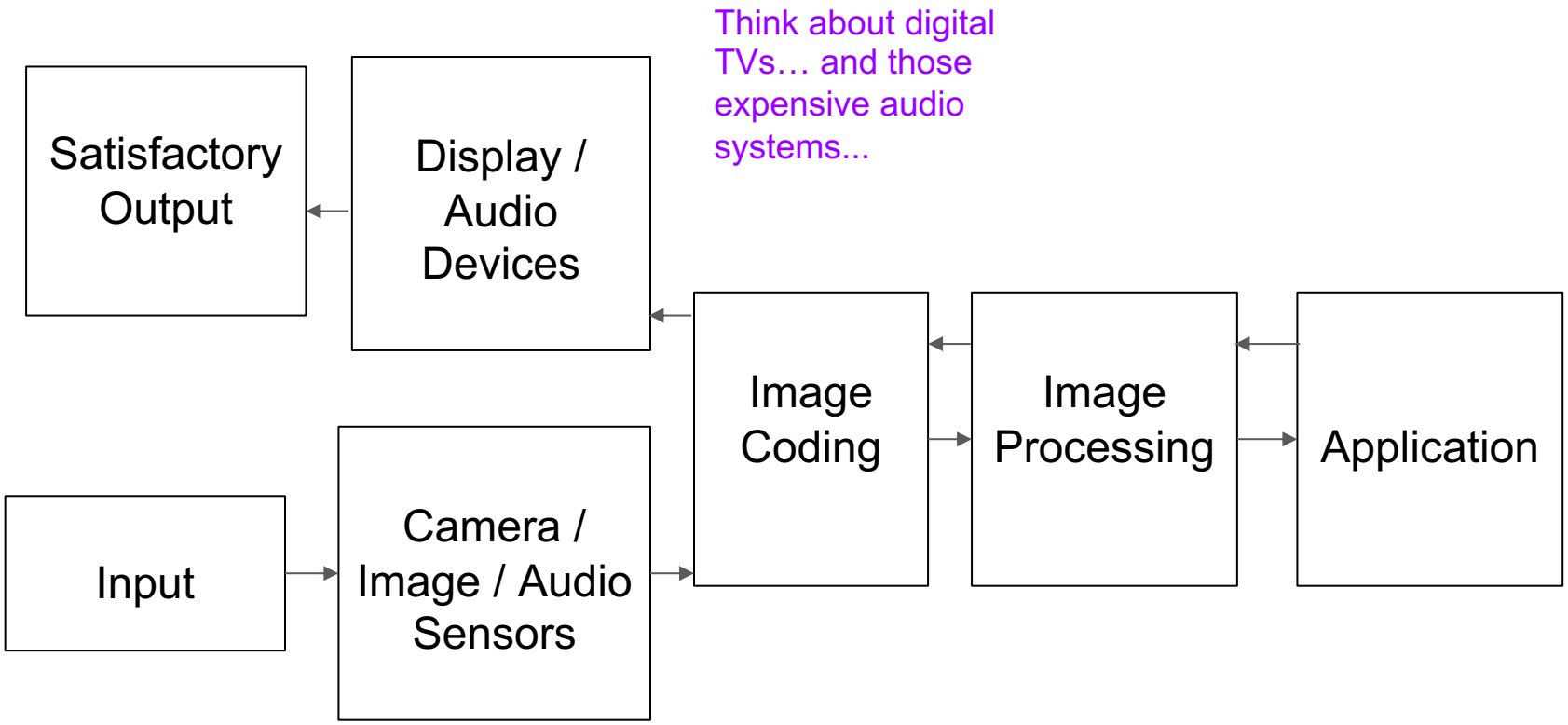
Capture visual and audio information and transform into digital data;
Still an area with innovations - increasing resolution and power

As signals, multimedia data need to be coded efficiently and conform to standards

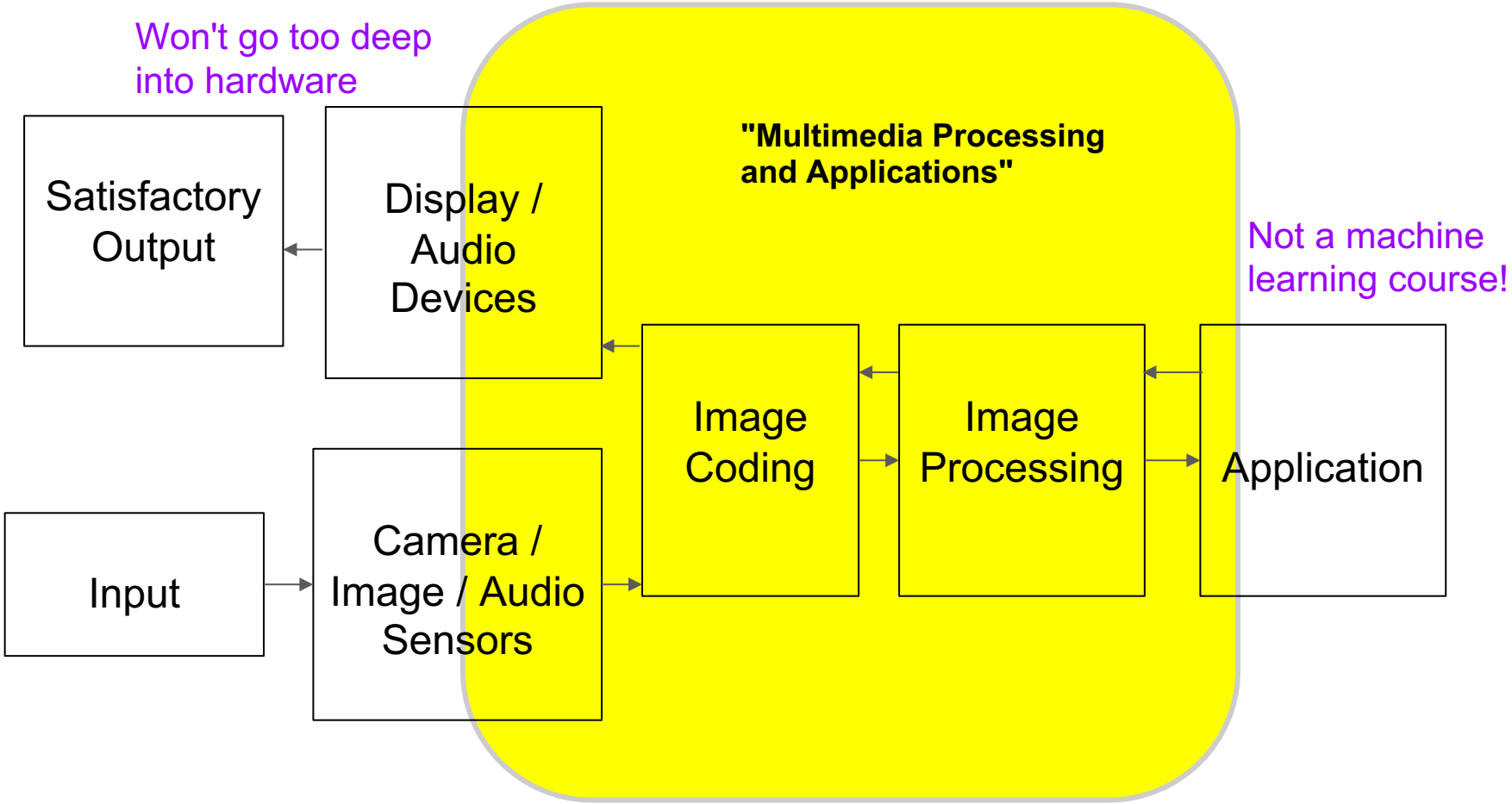
Raw image or audio data are often not suitable for immediate use. Noises? Normalization?

Most recent advances; actually an application of multimedia; receives most media attention

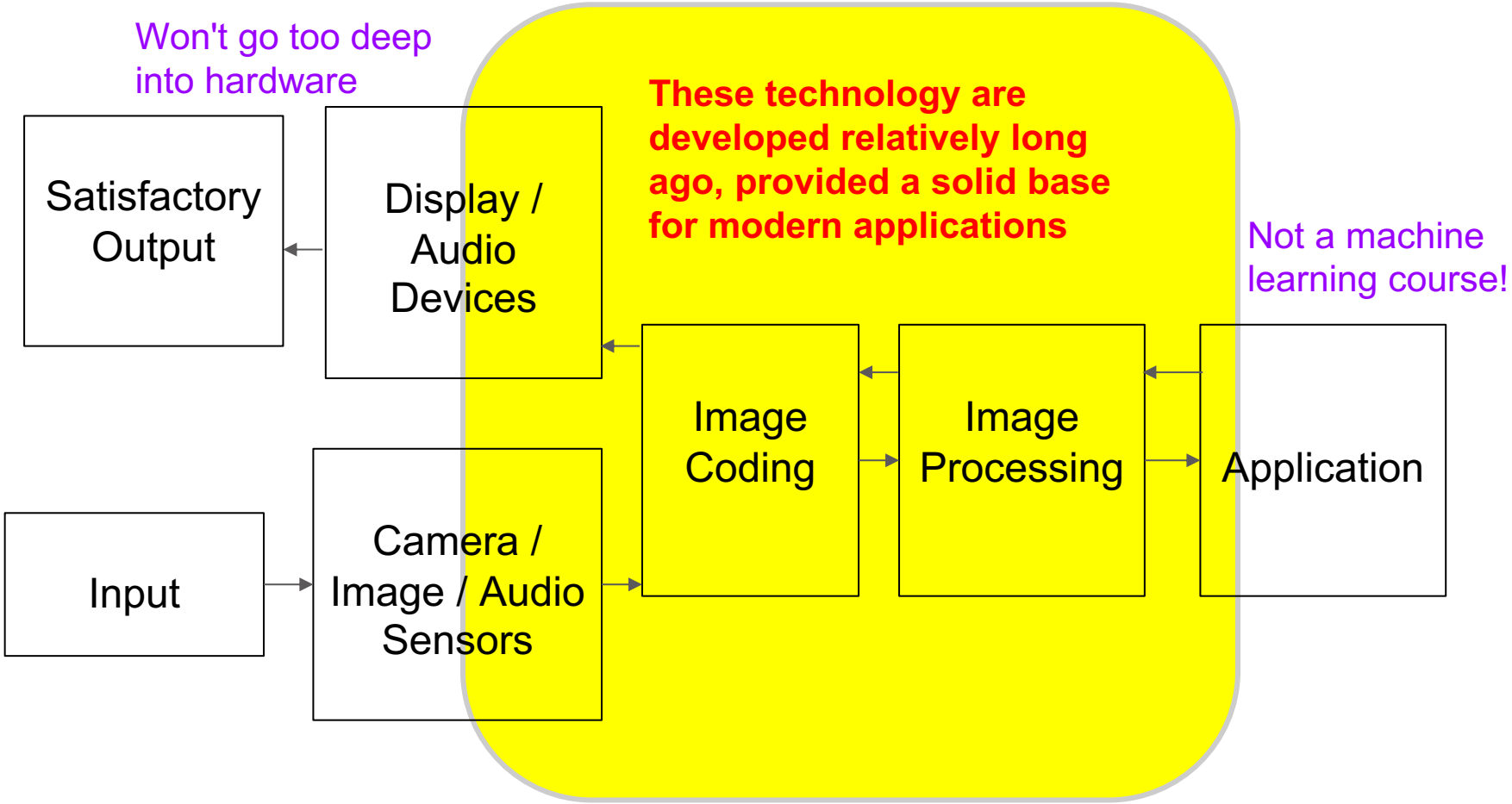
A Typical Multimedia System



Scope of Course

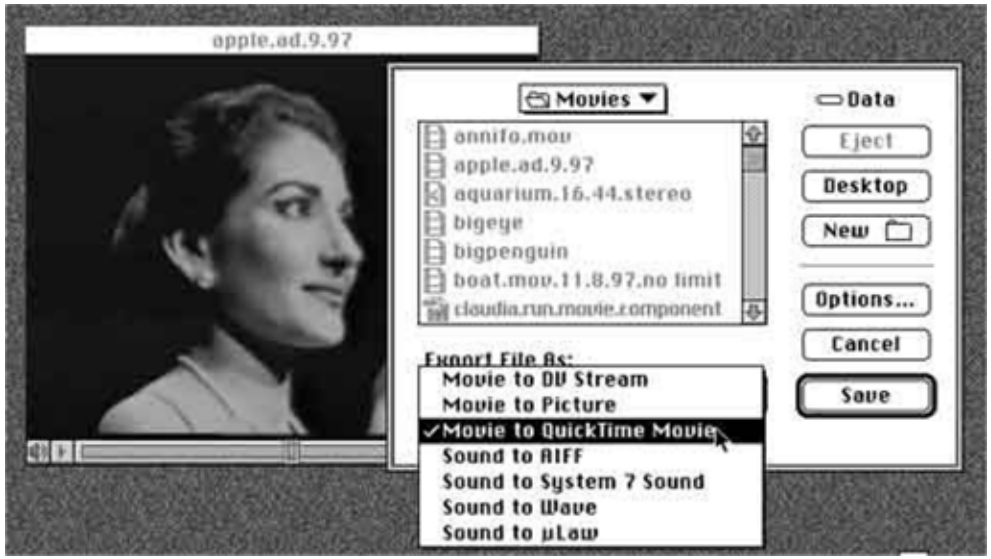


Scope of Course



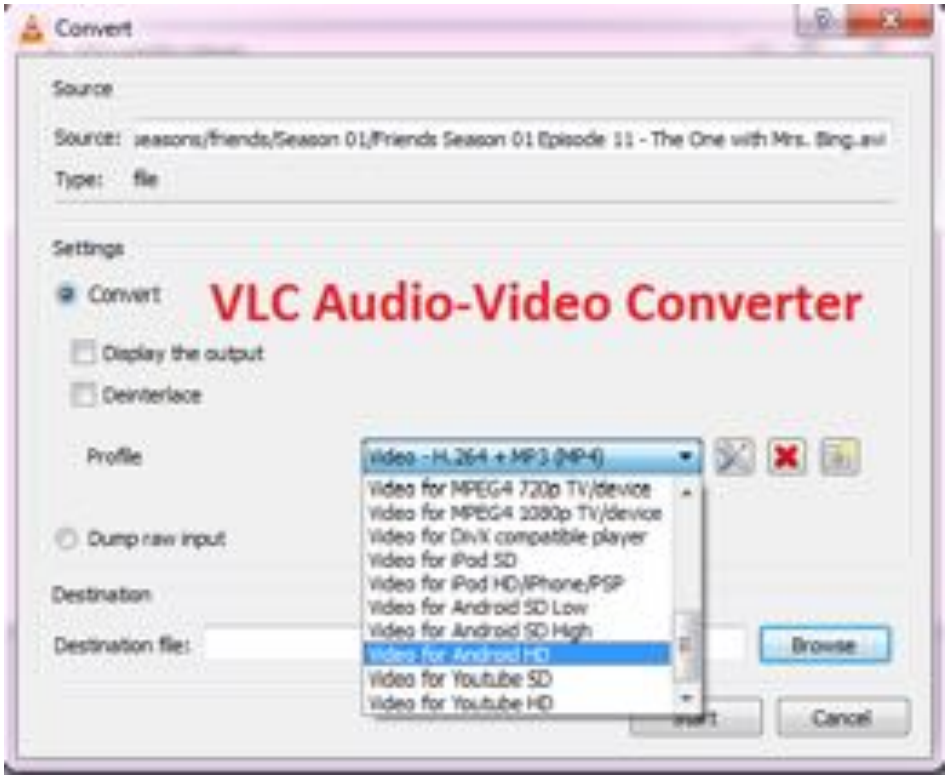
*"A **chronological journey** through the history of development of **multimedia processing** technologies (and **modern applications**)"*

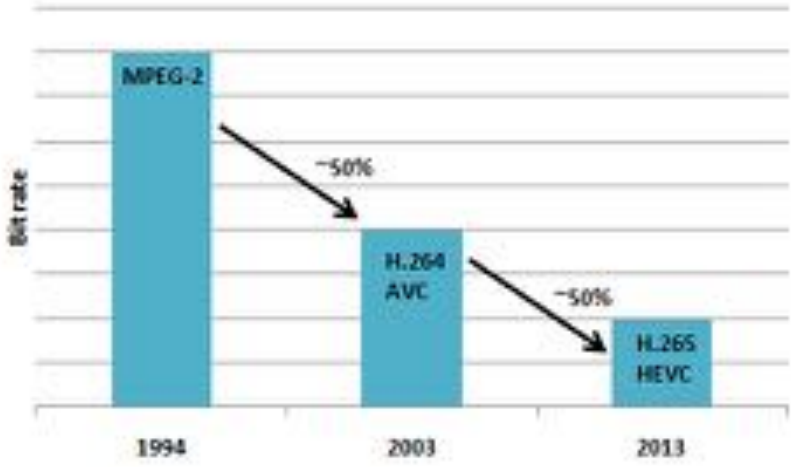
Long Time Ago...



Apple Quicktime (1994), appleinsider.com

Now

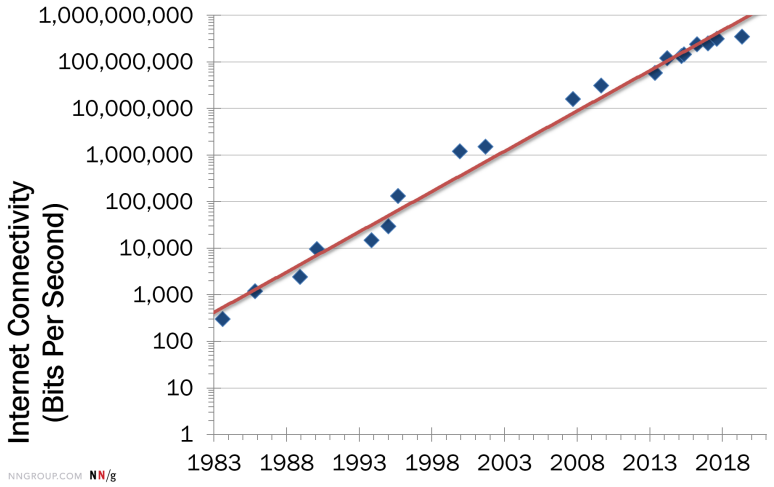

Convert
Source
Source: seasons/friends/Season 01/Friends Season 01 Episode 11 - The One with Mrs. Bing.avi
Type: file
Settings
Convert
VLC Audio-Video Converter
Display the output
Deinterlace
Profile
Video - H.264 + MP3 (MP4)
Video for MPEG4 720p TV/device
Video for MPEG4 1080p TV/device
Video for DivX compatible player
Video for iPod SD
Video for iPod HD/iPhone/PSP
Video for Android SD Low
Video for Android SD High
Video for Android HD
Video for Youtube SD
Video for Youtube HD
Dump raw input
Destination
Destination file:
Browse
Cancel


Bit rate
MPEG-2
~50%
H.264
AVC
~50%
H.265
HEVC
1994
2003
2013

Multimedia Coding

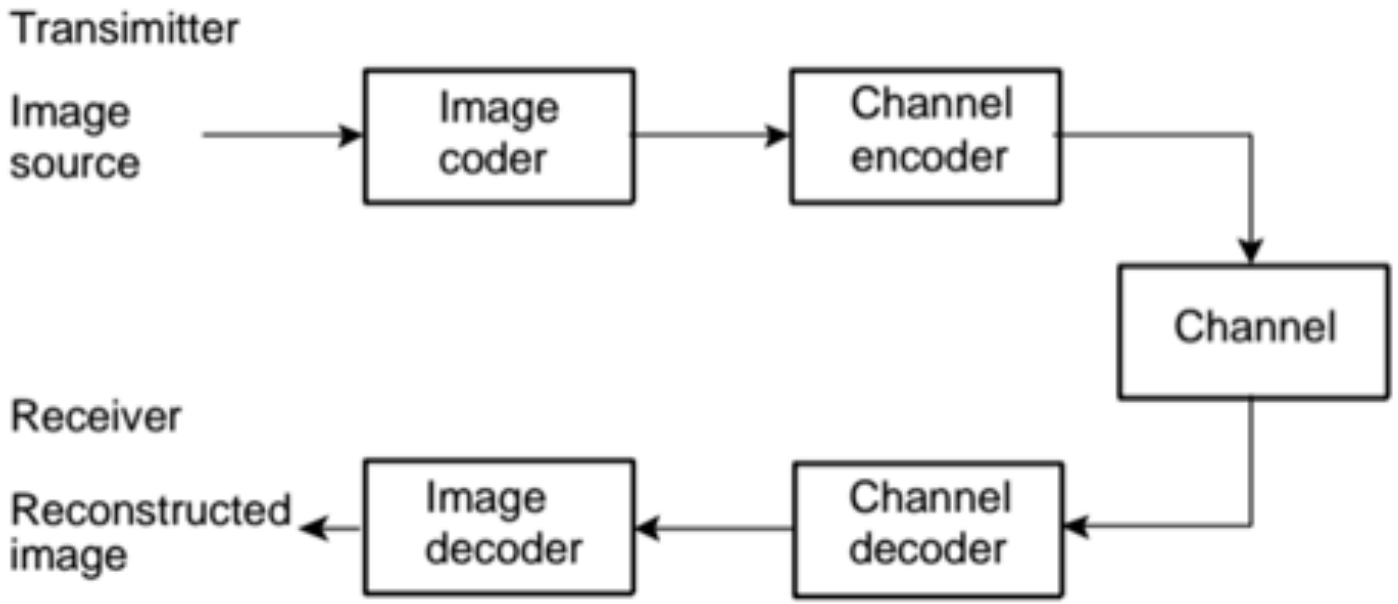
- In 90s and 80s...
- Multimedia data were pushing the limit of personal computing hardware
- Network bandwidth was also very limited
- Reducing storage/transmission cost was a priority
- Main engineering challenge is **multimedia coding**

Dial-in Internet access
56K modem ~20s for 1MB data



Multimedia Coding

- Objective: represent a media of acceptable quality with as few bits as possible (possibly lossy compression)



Multimedia Coding on the Internet

- The Internet as main driving force for multimedia coding standards
- Image compression standards:
  - Lossless: GIF (1987), PNG (2003)
  - Lossy: JPEG (1992)
  - Approximate year of standardization in brackets
- Video compression standards:
  - QuickTime (1991)
  - MPEG-1 (1993)
  - And many more…
- Also audio coding standards...

Difficulties of Multimedia Coding

- It's not a pure compression problem
- We want to minimize the distortion in quality
  - Tied to human perception
  - Also dependent on usage
  - Never truly a clean, clear-cut solution
- e.g. Audio
  - Speech? Music? Sound? Appliance? Network condition? Computation time?
  - http://en.wikipedia.org/wiki/Comparison_of_audio_coding_formats

Another Example: JPEG

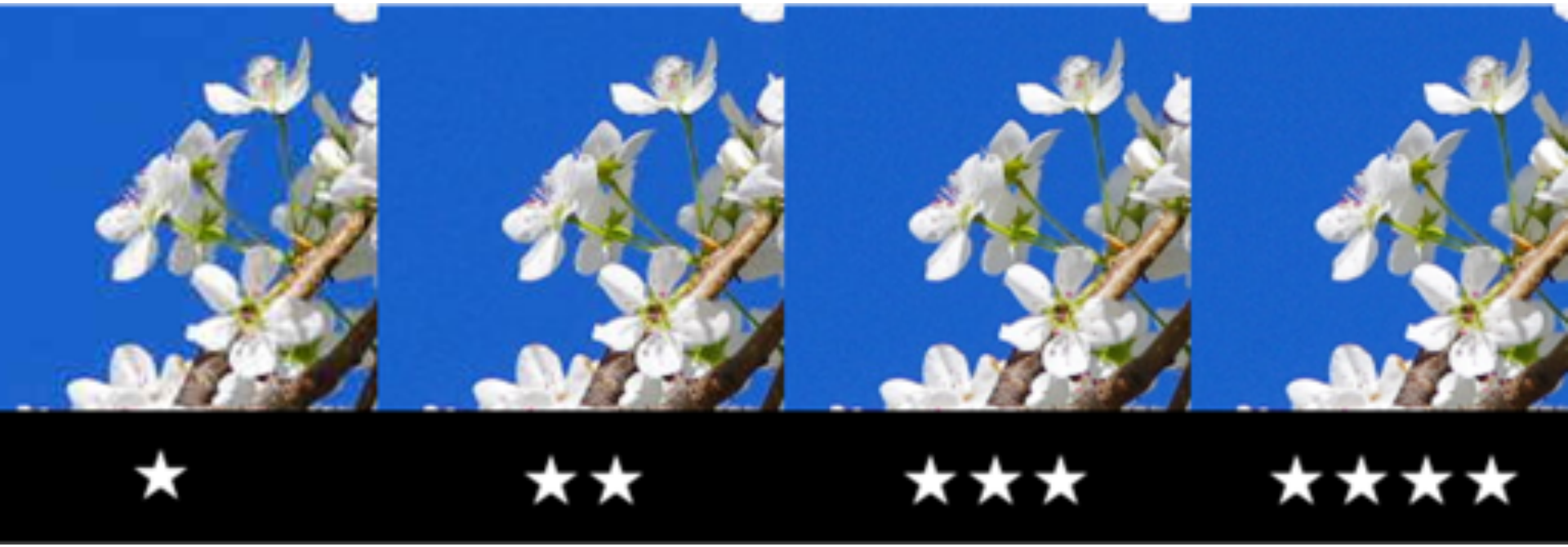


Image from Google Images. Rights belong to respective owners.

People Wanted More...

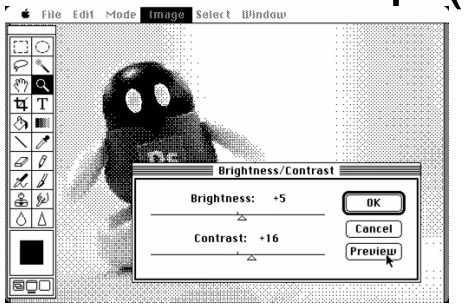
- Decades of efforts on multimedia coding have put multimedia into the hands of billions of people...
- With increased consumption of multimedia content, people wanted more….

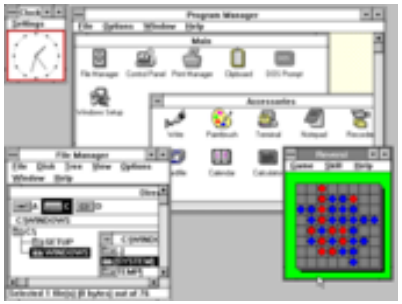
People Wanted More

From Google Images. Rights belong to respective owners.

Multimedia Processing

- Raw multimedia materials aren't always presentable
- Multimedia processing became necessary, at least first for the professionals
  - Photoshop (1990)





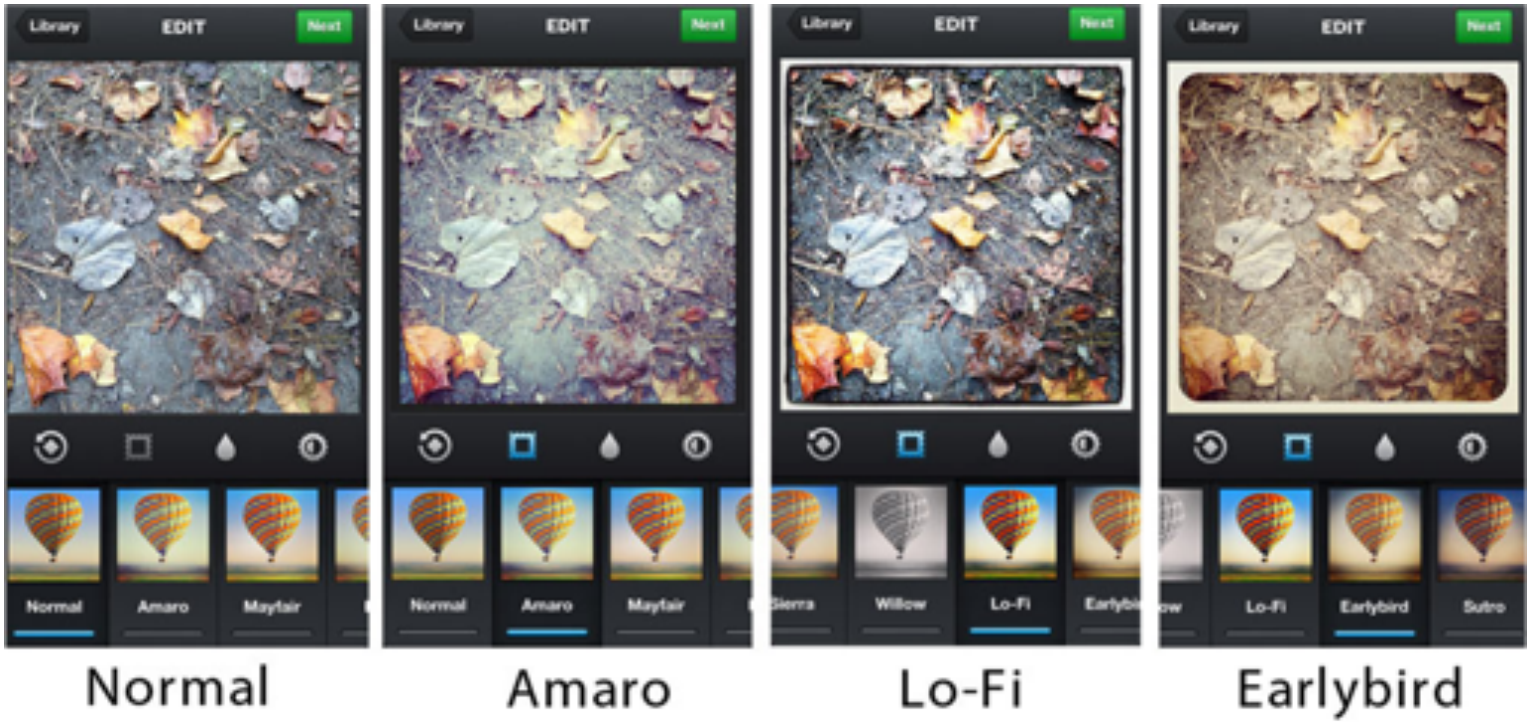
Windows 3.2 (1990)

Multimedia Processing

- Application of signal processing theories has allowed many applications in image, audio and video processing
  - Video editing, interleaving of images and sound, synchronization
  - Synthesis of speech
  - Enhancement of photos of natural objects and humans

Multimedia Processing Everywhere

Now a part of a consumer's everyday computing...



From Google Images. Rights belong to respective owners.

Multimedia Processing Everywhere

Now a part of a consumer's everyday computing...

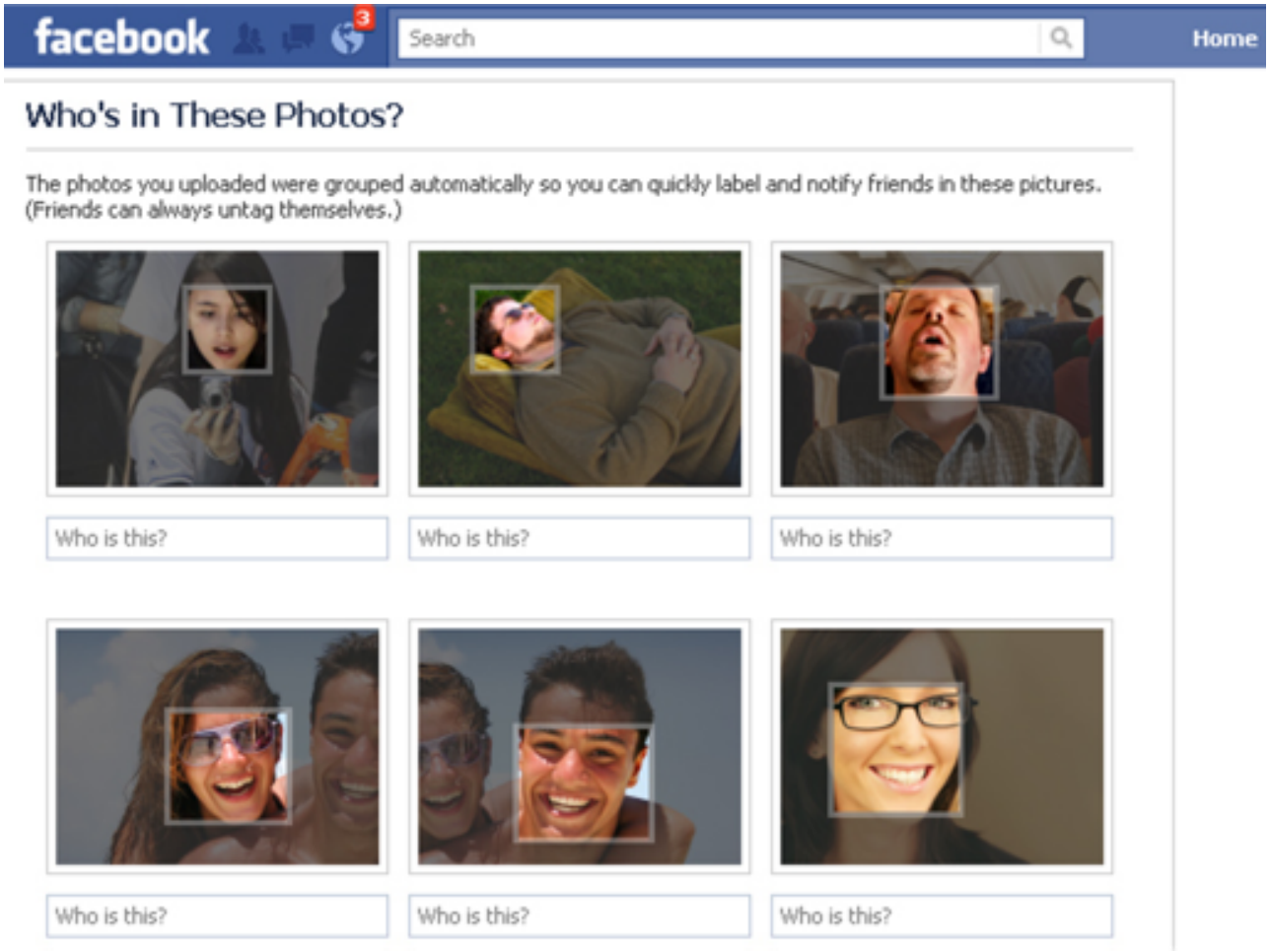
Multimedia Processing Everywhere

- So now what?
- So have we **finished** the multimedia revolution already?

- We produce significant amount of multimedia content - we are **inside** our multimedia
- People expect computers to **understand** them through these multimedia

From Processing to Understanding

From Processing to Understanding



From Processing to Understanding

Understanding is HARD….

From Processing to Understanding

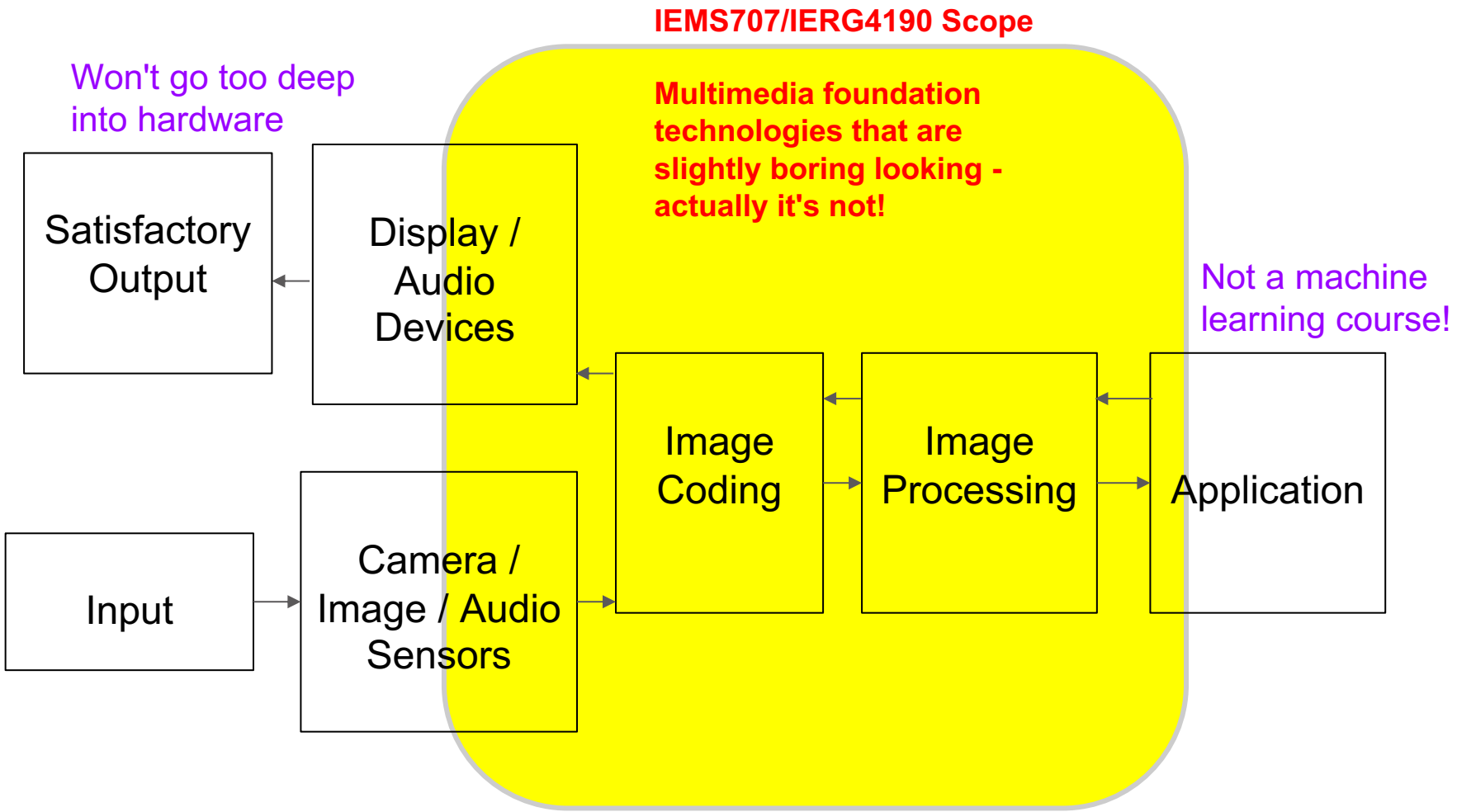
Even speech is surprisingly hard….

From Processing to Understanding

- Object/face recognition
- Speech understanding
- Motion/action detection
- … many more
- **Multimedia processing** meets **machine learning**
- Frontier of AI research and future applications

Previous Lecturer of the Course

But Remember - Scope of Course



Limitations of the Journey

- More and more stuffs happen in multimedia processing
  - But we can't leave the past behind as they are still the foundations
  - Tougher and tougher to teach (!)
- The more we need to cover, the less depth we can achieve
  - For in-depth understanding of specific technologies, you will need to go deeper in other postgraduate courses

So….

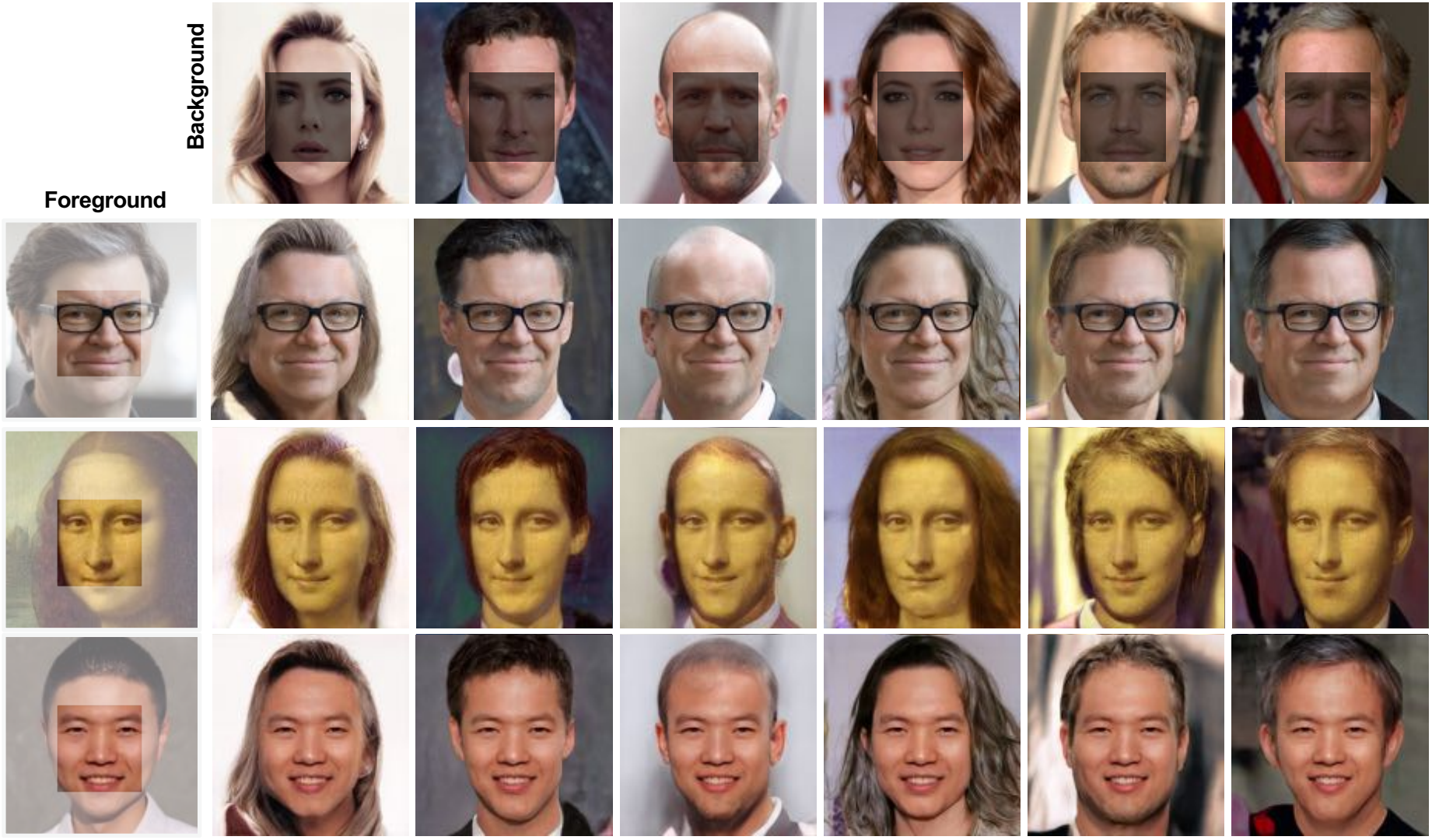
What exactly is this course about?

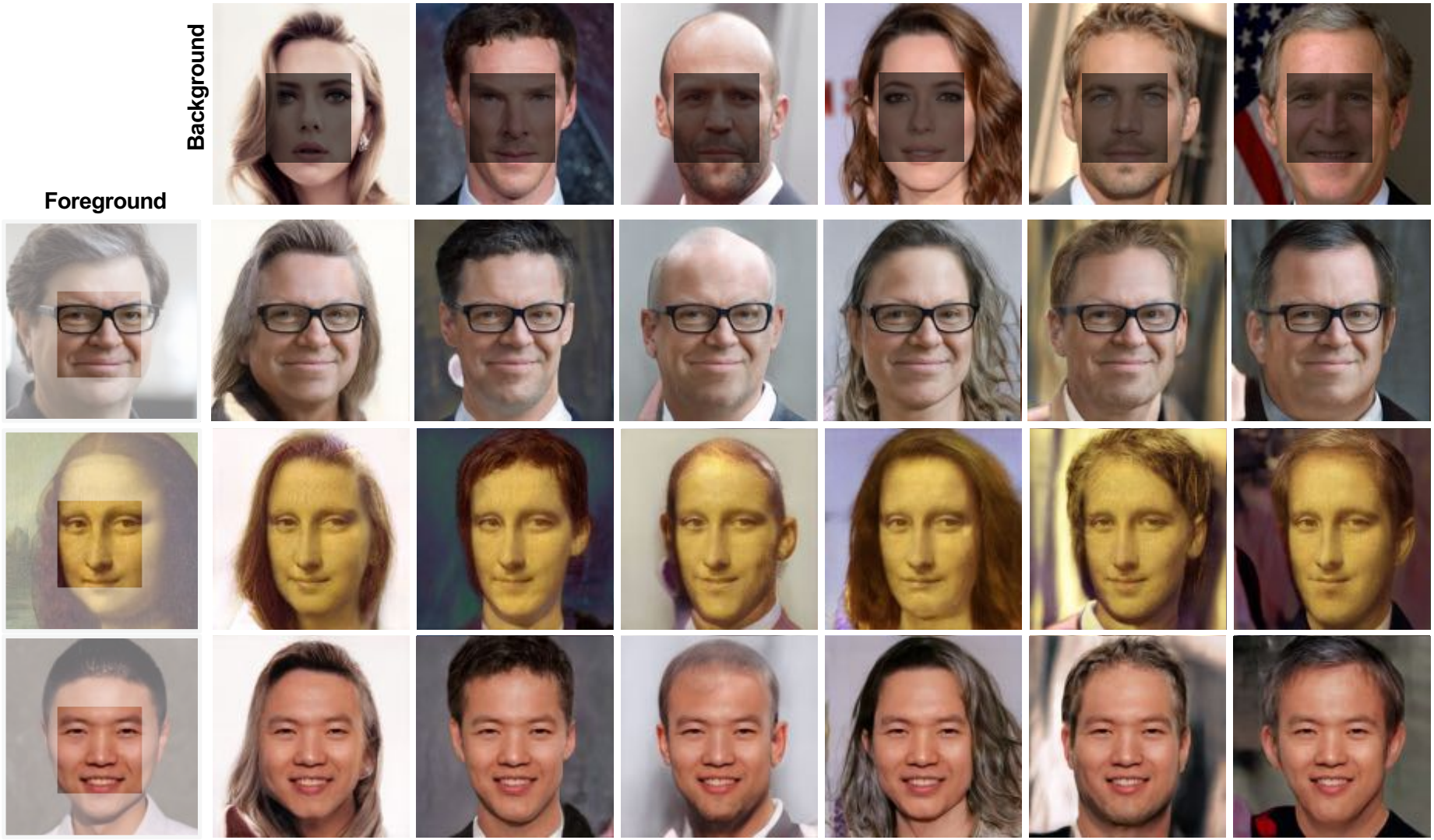
Three Parts of the Course

- Multimedia Coding
  - Some lightweight mathematical analysis
  - Basic signal processing knowledge required
  - Introductory, can't possibly cover all
- Multimedia Processing
  - Focus on image processing
  - Hands-on with Coding
  - e.g. Get a chance to "PS/Meitu" yourself with your own code - it's harder than you think!
- Applications + Multimedia Understanding
  - New coding standards, new processing technologies, multimedia understanding applications etc.

Background

Foreground





Try you self!
https://colab.research.google.com/github/genforce/idinvert_pytorch/blob/master/docs/Idinvert.ipynb

What is Expected on You?

- After the course…
  - understanding of **key concepts** behind multimedia coding and processing
  - understanding of the basic signal processing foundation
- Obtain foundation knowledge for all modern systems
  - Video streaming/Computer vision/Image generation...
- Can listen to an expert in the field and at least know what's going on

Who Should Take This Course

- If you are entering any field that make use of multimedia
  - Knows the basics and technicalities
- If you are interested in the emerging fields of computer vision or new multimedia processing topics
  - Some pointers for you to dive into
- Or if you find it fun!
  - Always a good reason, right?

What is Special About This Course?

- A multi-angle approach - some theory, some hands-on, some conceptual
- Hence we call this a **"Journey"**
- Probably not a familiar course format to most of you
  - ...but a good approach for an introduction to a very large field

What This Course is NOT About

- **NOT** about performing deep mathematical analysis on multimedia systems
  - Often unnecessary in modern age when media processing is supported by many great software and code libraries
  - Image, audio and video data are often too hard to operate by hand or mind alone
- **NOT** an introductory course on visual recognition or deep learning
  - As I said, NOT a computer vision course
  - Computer vision etc. is just ONE type of multimedia processing

Course Schedule

We go through the development of multimedia processing in **chronological** order…

Week 1-5: Multimedia Coding

Week 5-8: Multimedia Processing (some lab sessions, to be determined)

Week 9-12: Multimedia Processing Applications

Course Schedule (Tentative)

Week 1	Course Introduction
Week 2	Coding Basics and 2D Signal Processing
Week 3	Multimedia Coding (1)
Week 4	Multimedia Coding (2)
Week 5	Advanced Topics in Coding
Week 6	Image Processing (1)

Course Schedule (Tentative)

Week 7	Image Processing (2) Project Introduction
Week 8	Lab: Image Processing Hands-on Guest lectures
Week 9	Generative Modeling I
Week 10	Generative Modeling II
Week 11	Project Presentation
Week 12	Project Presentation
Week 13	Course Conclusion

Logistics

Webpage: http://bzhou.ie.cuhk.edu.hk/teaching/ierg4190iems5707/

ZOOM Link and **Master Doc File** are posted at Blackboard

Lectures:

- Tuesday: 10:30 am – 12:15 pm, ZOOM
- Wednesday: 11:30 am – 12:15 pm, ZOOM

My office hour: Thursday 8:30 am – 9:15 am

TA Tutorials (start in Week 3) and Office Hour (start in Week 2)

- Wednesday 6:00 pm – 7:00 pm

Assessment Scheme

- Attendance 10%
  - But **> 66% attendance to get full score**
- Assignments 30%
  - Three times, including more difficult questions for IEMS
- Final Exam 30%
- Project 30%
  - A self-research report (2 students in a group)
  - Will need to present to class

Teaching Team

- Lecturer:
  - Bolei Zhou (bzhou@ie.cuhk.edu.hk)
- TAs:
  - Xinge Zhu (zx018@ie.cuhk.edu.hk)
  - Yuanbo Xiangli (xy019@ie.cuhk.edu.hk)
  - Ceyuan Yang (yc019@ie.cuhk.edu.hk)

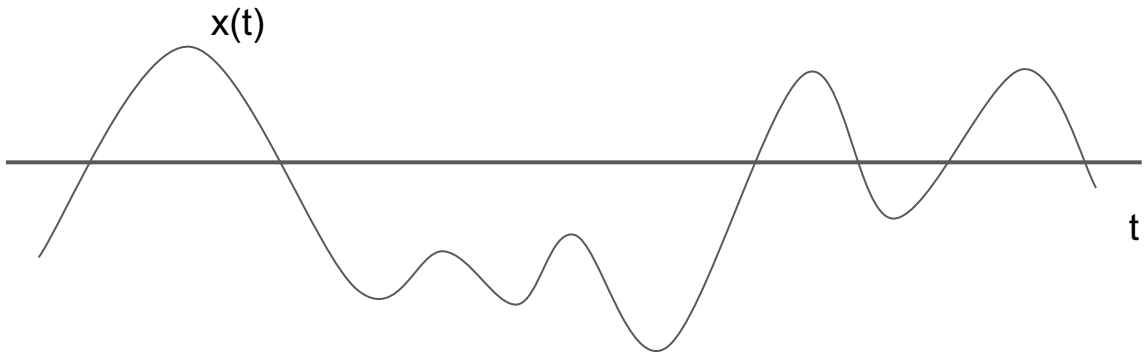
Teaching Assistants

Signal and System Basics

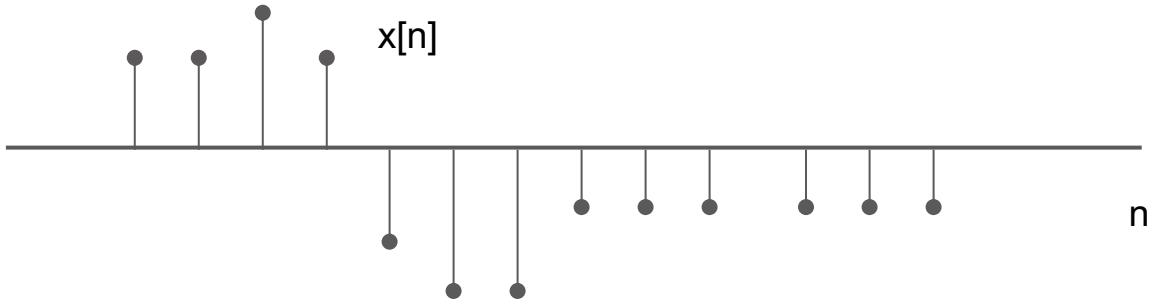
In case there are any misunderstanding on assumed background knowledge...

Signal

Continuous time (one-dimensional)

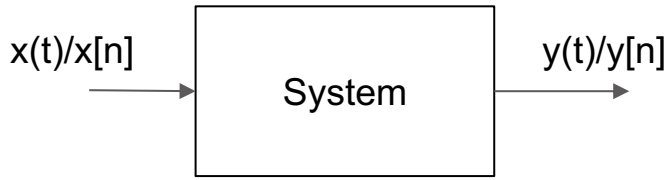


Discrete time (one-dimensional)



Signal Processing System

A Signal Processing System



We are particularly interested in Linear Time-Invariant (LTI) System

- What does that mean?

Other Concepts

- Frequency domain
- Impulse Response
- Laplace/Z Transform
- Convolution

Since this course is about **applied** signal processing you don't need to be good in calculating any transforms etc.

(But I assume you know what these concepts are)

Preparations

That is (supposedly) all you need to know minimally for this course.

If you lost some of your memory on them you can refresh your memory here:

https://en.wikipedia.org/wiki/Linear_time-invariant_theory