Computer Vision and Machine Learning

(Introduction)

Bhabatosh Chanda
bchanda57@gmail.com

Outline

- Introduction
  - Motivation
    - Visual pattern recognition
  - Applications
  - Illusion
  - Image formation
    - Perspective projection
    - Photometric model

Motivation

- Introduction of digital computer in mid-20th century
- The machine turns out to be superior to human being in number crunching.
- If it can mimic the ability of human being in pattern recognition and decision making.
- Birth of new subjects like *computational intelligence*, *machine intelligence*.

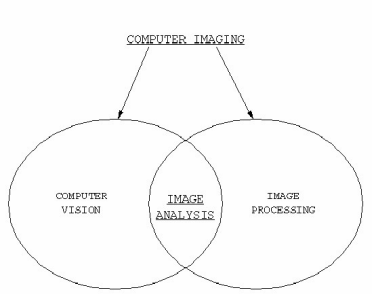
Visual pattern recognition

- In real life, human beings perform the pattern recognition tasks based on the information acquired from the ***environment*** through sensors like ears, eyes, nose, tongue and skin.
- Most of the ***information*** is acquired through ***eyes***.
- ***Visual pattern recognition***, thus, turns out to be most important activity in this context.
- This leads to development of ***Computer vision and Pattern recognition***.

Visual data processing

- Visual or pictorial data processing by digital computer may be grouped into three categories:
  - **Input:** textual description; **Output:** an image
    Computer Graphics
  - **Input:** an image; **Output:** an image
    Image Processing
  - **Input:** an image; **Output:** textual description
    Image analysis and recognition \ Computer Vision

Visual data analysis



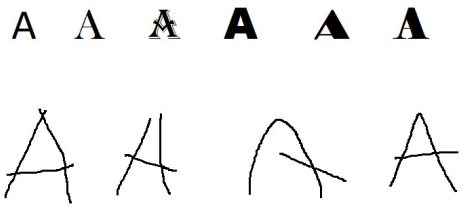
Pattern recognition: Learning

. . . . . .

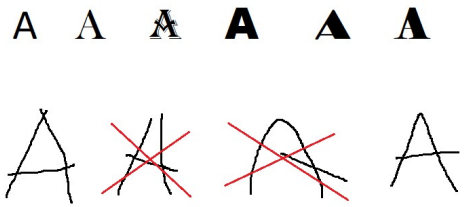
Pattern recognition: Learning

A A A A A A

Pattern recognition: Learning



Learning

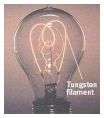


Natural images: examples

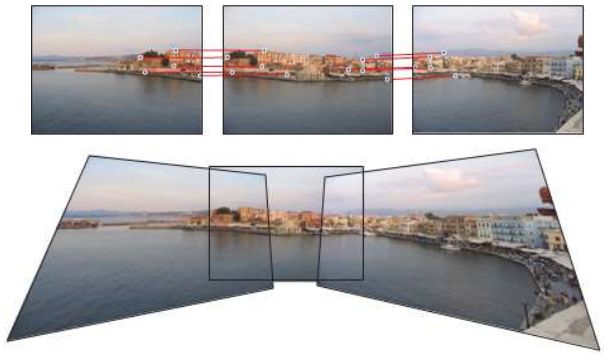
Outdoor scene Indoor scene Micrograph

Applications

- Intelligent machines (AI)
- Industrial inspection
- Automotive
- Security
- Image/video retrieval / Digital Library
- Robot navigation
- Activity analysis and so on ….



Panorama



Face Detection

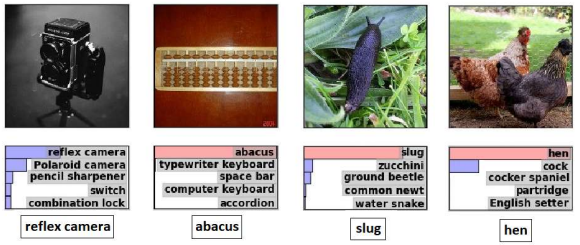
- Real-time face detection on most phones /cameras now
  - Use to set exposure
  - Also input for face *recognition* system

Object recognition

- Referred to by three tasks:
- Image classification
  - Assigning a class label to an image.
  - May be multiple labels with confidence levels.
- Object localization
  - Drawing a bounding box around the targeted object.
- Object detection
  - Drawing a bounding box around each object of interest and assigning class label to those.

Image (object) recognition

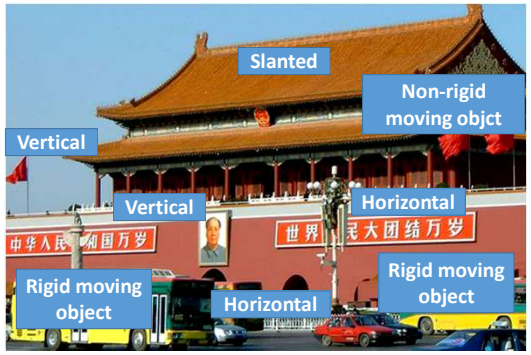
Pixels to class-label:



Optical character recognition (OCR)

Image to semantic information

Qualitative information

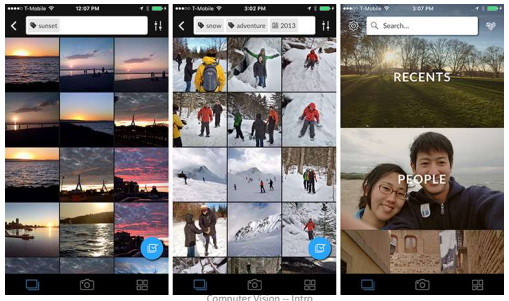


Object categorization

Scene and context categorization

Advanced Photo Search

- Content-based image search
  - that actually looks at image



http://clarifai-blog.s3.amazonaws.com/wp-content/uploads/2015/12/pic1.jpg

Data sources

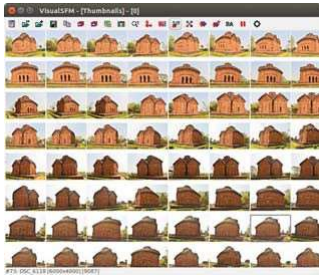
Vision-based interaction (and games)

Microsoft Kinect

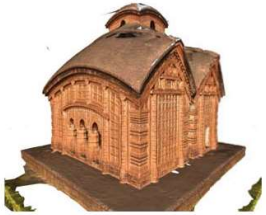
Virtual/Augmented Reality

- Tracking of user head with high accuracy
- Rendering realistic 3D scene in real-time
- Oculus / HTC / Hololens

Geometry: 3D reconstruction (cond.)

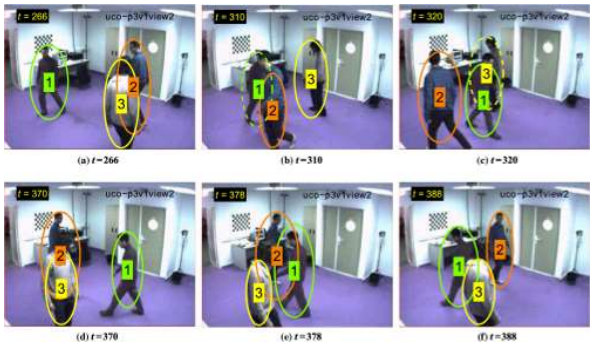


Geometry: 3D reconstruction

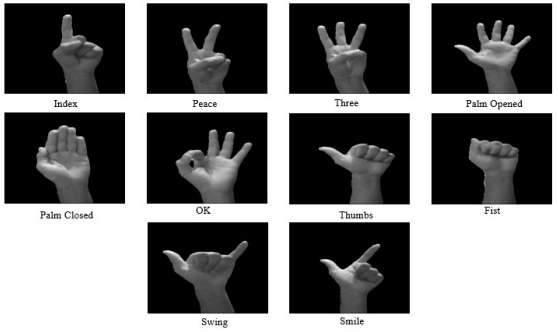


Motion detection and tracking

Activity analysis: tracking



Activity analysis: hand gesture



Activity analysis: events

Other applications with DL

Image classification

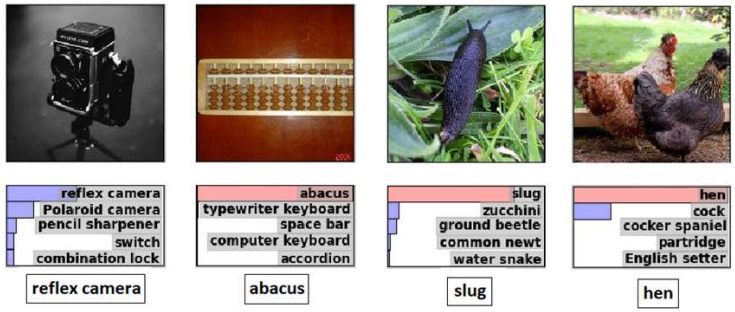


Image generation

Image generation: Progress

Aging face image

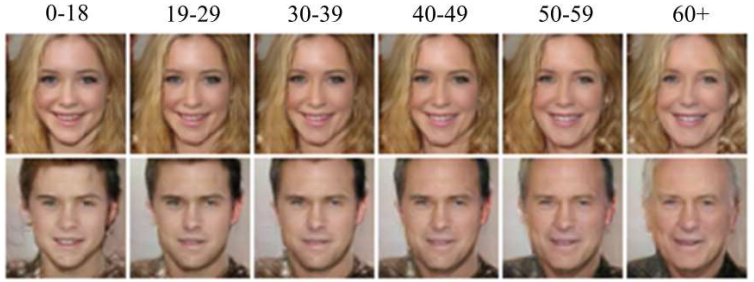
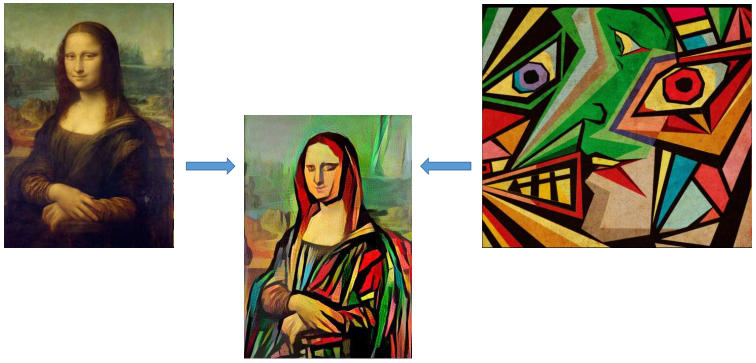
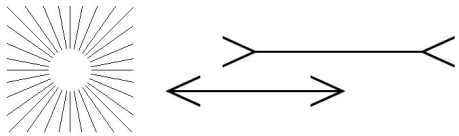


Image colorization

Style transfer

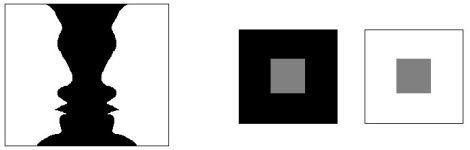


Illusion: Examples



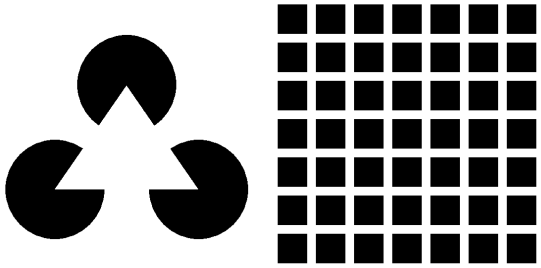
(a) (b)

Illusion: Examples



(c) (d)

Illusion: Examples



(e) (f)

Illusion: Examples

Illusion: Examples

Illusion: Examples

Scene types

Scenes are of two types:

- **3-D**, example: usual outdoor and indoor scenes
- **2-D**, example: document page, micrograph, satellite image, x-ray image, etc.

Types of **images**:

- Static image (single frame)
  - Black and white, colour
- Sequence of images (video/movies)

Types of digital image

Black-and-white image

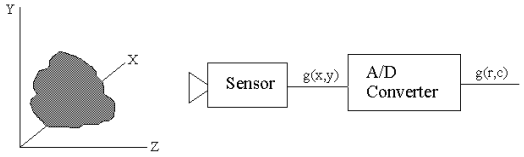
- Binary (two-tone) image
- Gray level (gray-tone) image

Color image

Another way of classifying images:

- Static image (single image)
- Image sequence (movie)

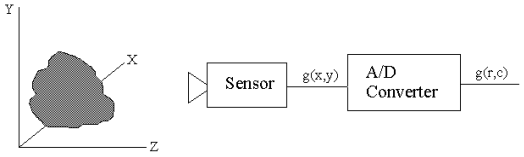
Image formation



Sensor grabs and forms two-dimensional intensity map *g(x,y)* of the 3D scene, which is continuous in space and value.

Example: **Camera, Scanner, Ultrasound sensor, Infra-red sensor, MRI, CT, etc.**

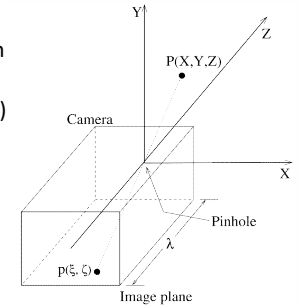
Image formation



- There are two transformations that simultaneously take place in the camera
  - Geometric transformation (perspective projection)
  - Photometric transformation (light transportation)

Geometric transformation

- Consider pinhole camera model.
- Camera coordinate system coincides with world coordinate system.
- Origin *(0,0,0)* is at pinhole (optical centre)
- *z*-axis is same as optical axis of the camera and is perpendicular to image plane.
- λ is the focal length of the camera.
- *P(X,Y,Z)* and *p(ξ, ζ)* are world points an image point respectively.



Geometric transformation (perspective projection)

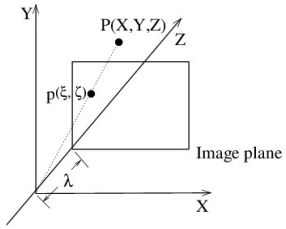
- Consider image plane in front of the pinhole (to avoid negative coordinate)

$$\xi = \frac{\lambda X}{Z} \text{ and } \zeta = \frac{\lambda Y}{Z}$$

- In matrix-vector form:

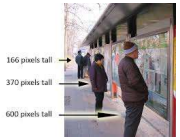
$$\begin{bmatrix} X \\ Y \\ Z/\lambda \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1/\lambda \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

- *(X, Y, Z/ λ)* is homogeneous coordinate in image plane, to get Cartesian coordinate divide first two (X and Y) by the third (Z/λ).



Properties of perspective projection

- Straight lines in 3D scene (world) map into straight lines in image.
- Distant objects (in scene) appear smaller in the image.
- A set of parallel lines in 3D, not perpendicular to z-axis maps to a set of concurrent lines in 2d image.
  - Point through which the concurrent lines pass is called *vanishing point*.
  - All the vanishing points lie on a straight line, called *horizon*.



Photometric model

- Suppose light intensity $\mathcal{I}$ is transported ideally to image plane *(ξ, ζ).*
- Scene reflectance $\mathcal{R}$ is also transferred appropriately to image plane.
- Thus image intensity mapped is ideal one and is given by

$$f(\xi, \zeta) = \mathcal{R}(\xi, \zeta)\mathcal{I}(\xi, \zeta)$$

- However, light energy transport to real image plane undergoes some transformation. This may be modeled by photometric transformation.

$$g(x,y) = \mathrm{T}[f(x,y)]$$

that satisfies the condition: $f(x,y) \geq 0$ and $g(x,y) \geq 0$

Thank you!

Any question?