Computer Vision, CS 763

Ajit Rajwade,
CS 763, Spring 2017,
IITB, CSE department

Why take this course?

- Recommended if you want to **do research work with us in the ViGIL group** in computer vision or image processing
- Inherently **interdisciplinary** subject: numerous application areas - remote sensing, photography, visual psychology, archaeology, surveillance, etc.
- Fast becoming a popular field of study in India: scope for R&D work in numerous research labs (In India: Samsung, GE, Phillips, Siemens, Microsoft, HP, TI, Google; DRDO, ICRISAT, ISRO, etc.)

Why take this course?

- India has numerous conferences in computer vision and related areas: ICVGIP, NCVPRIPG, SPCOM, NCC.
- International vision conferences: CVPR, ICCV, ECCV, many vision papers in NIPS, ICML.
- Other more specialized vision conferences: ICBA, IGARRS, ICCP

Computer Vision and Image Processing: What's the difference?

- Difference is **blurry**
- "Image processing" typically involves processing/analysis of (2D) images without referring to underlying 3D structure
- Computer vision – typically involves inference of **underlying 3D structure** from 2D images
- Many computer vision techniques also aim to infer properties of the scene directly – without 3D reconstruction.
- Computer vision – direct opposite of computer graphics

Course web-page

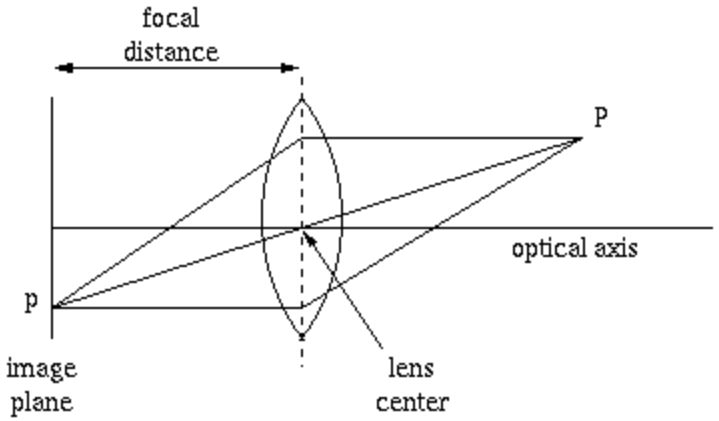
http://www.cse.iitb.ac.in/~ajitvr/CS763_Spring2017/

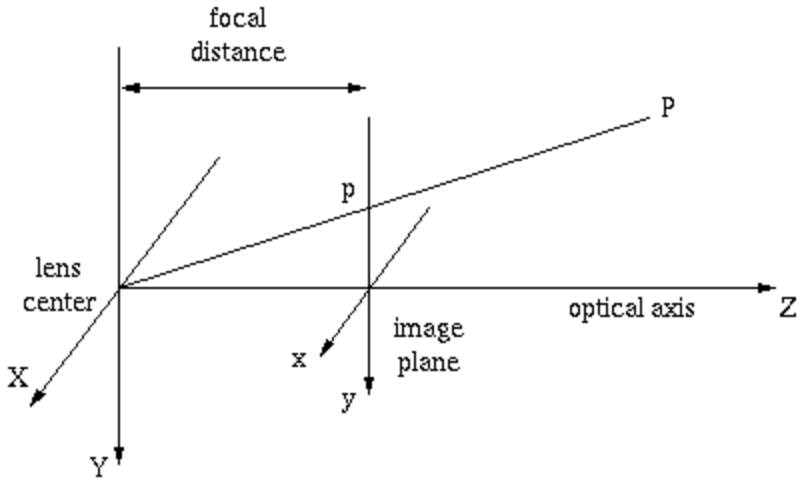
What will we study in this course?

Four major components

- Camera geometry
- Shape from X
- Motion Estimation
- Machine learning in computer vision

(1-A) Camera Geometry



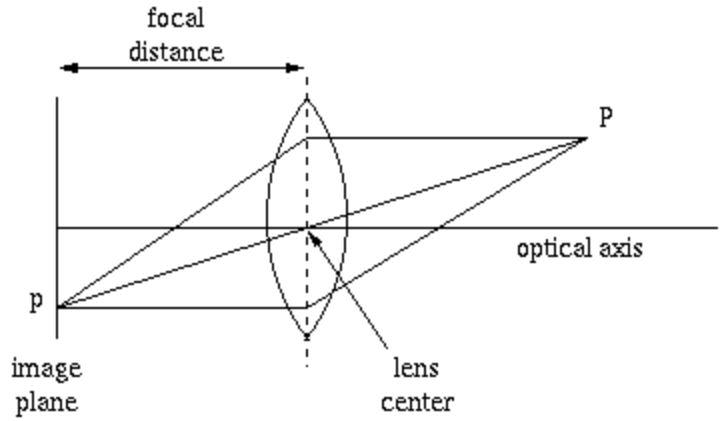


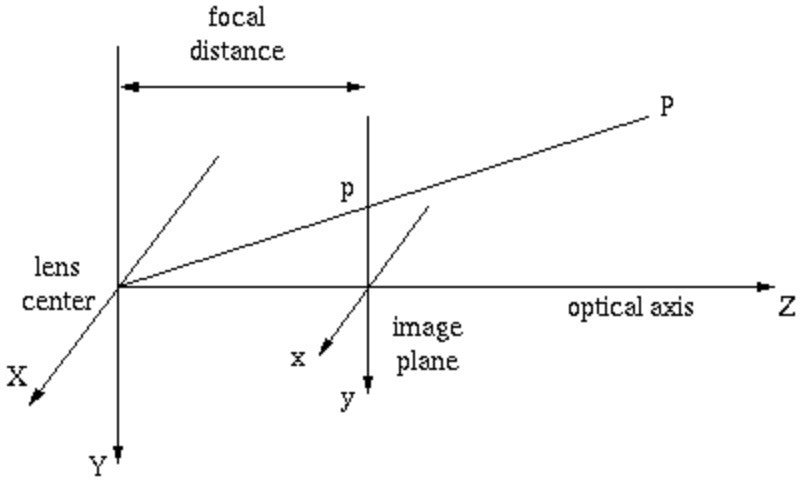
- Relationship between object coordinates (given by a vector **P** in 3D) and image coordinates (given by vector **p** in 2D)

- Effect of various intrinsic camera parameters (focal length of lens, nature of the lens, aspect ratio of detector array, etc) on image formation

- Effect of various extrinsic camera parameters on image formation

(1-A) Camera Geometry (continued)



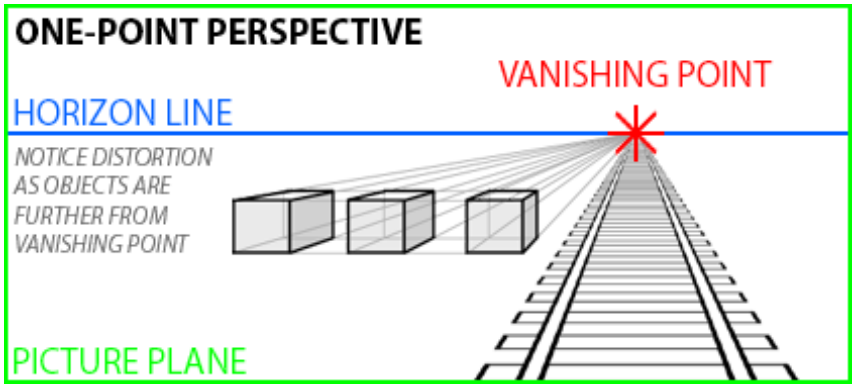


• Let's say you take a picture of a simple object of known geometry (example: chessboard, cube, etc.).

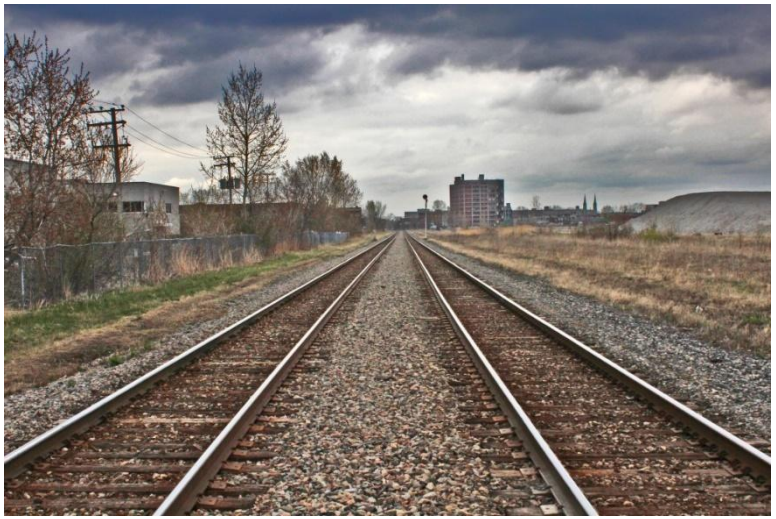
• Given the 3D coordinates of *N* points on the object, and their corresponding 2D coordinates in the image plane, can you determine the camera parameters such as focal length?

•Answer is yes you can. This process is called as **camera calibration**.

(1-B) Camera Geometry (Vanishing points)



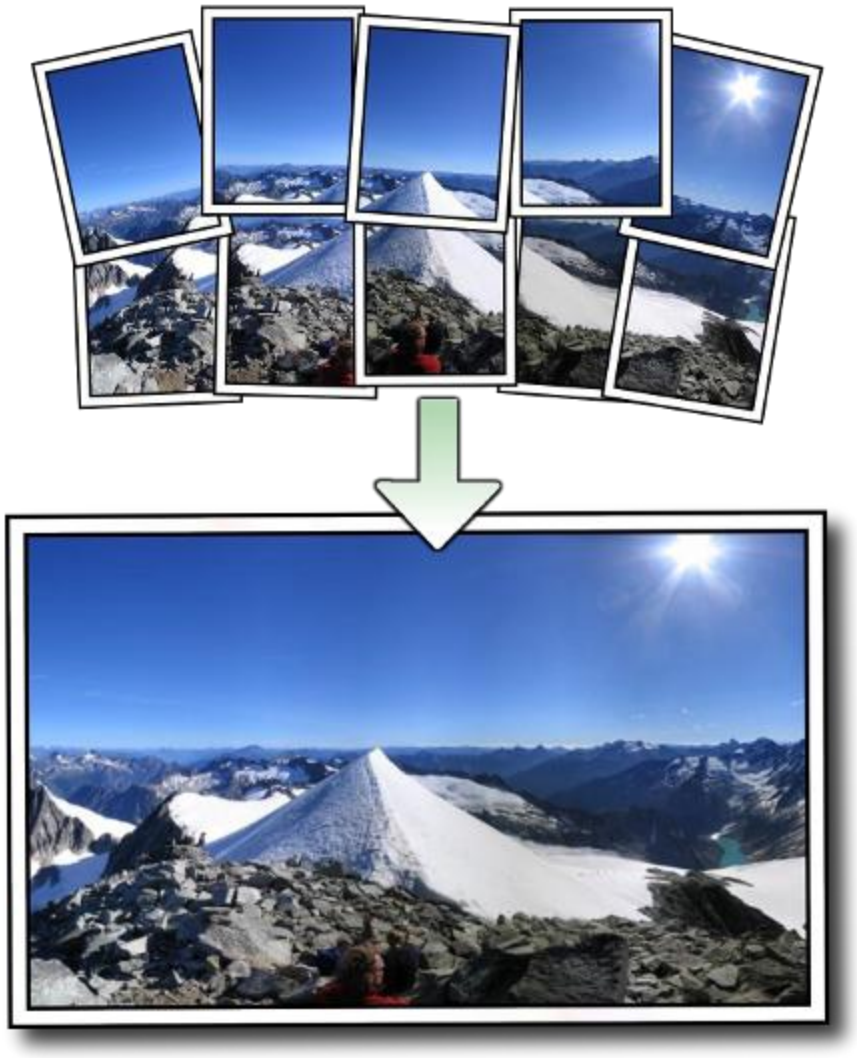
http://www.atpm.com/9.09/design.shtml



http://www.cns.nyu.edu/~david/courses/perception/lecturenotes/depth/depth-size.html

http://www.vertice.ca/index.php/2012/sonic-vanishing-points/

(1-C) Image Mosaicing/Panoramas



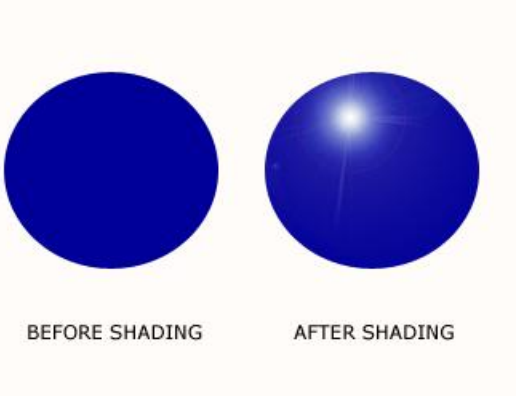
We will study an end-to-end technique for generating a panorama out of a series of pictures of a scene from different viewpoints.

http://cs.bath.ac.uk/brown/autostitch/autostitch.html

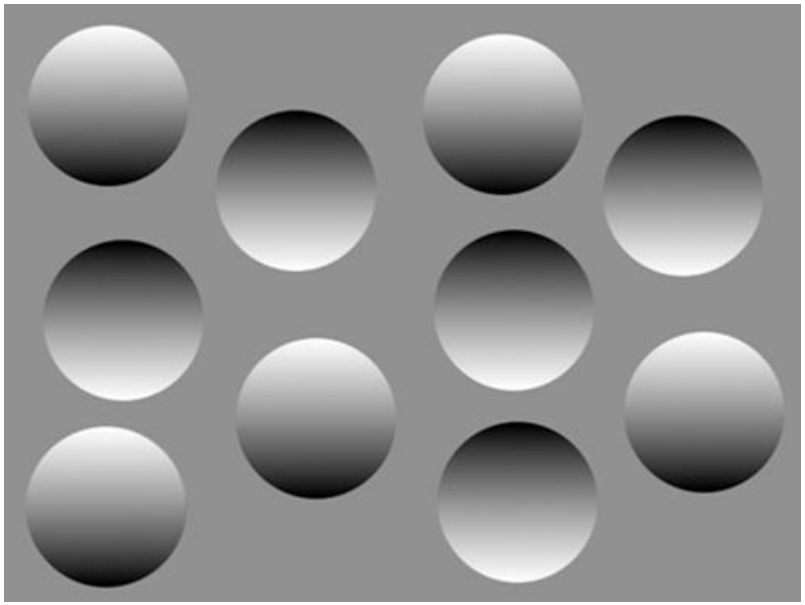
(2) Shape from ‘X’

- An image is 2D. But most underlying objects are 3D.
- Can you guess something **about the 3D structure** of the underlying object just given the **2D image**?
- The human visual system does this all the time.
- We want to reproduce this effect computationally (the “holy grail” of computer vision)

(2-A) Shape from Shading



http://www.famouslogos.org/the-basics-of-three-dimensional-design



http://www.psychol.ucl.ac.uk/vision/Lab_Site/Demos.html

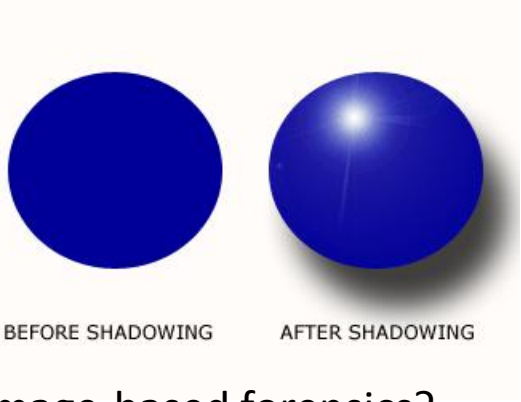
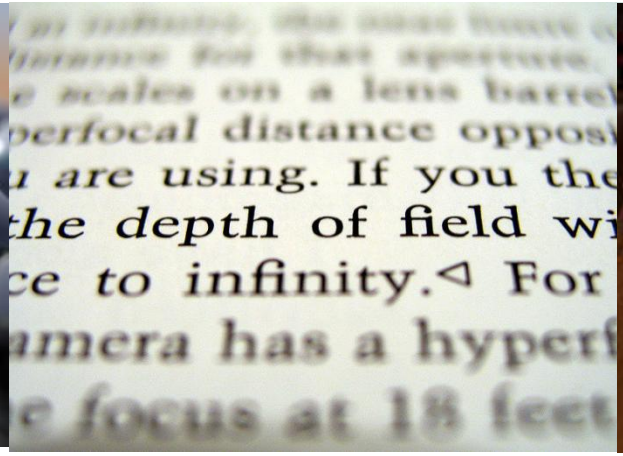
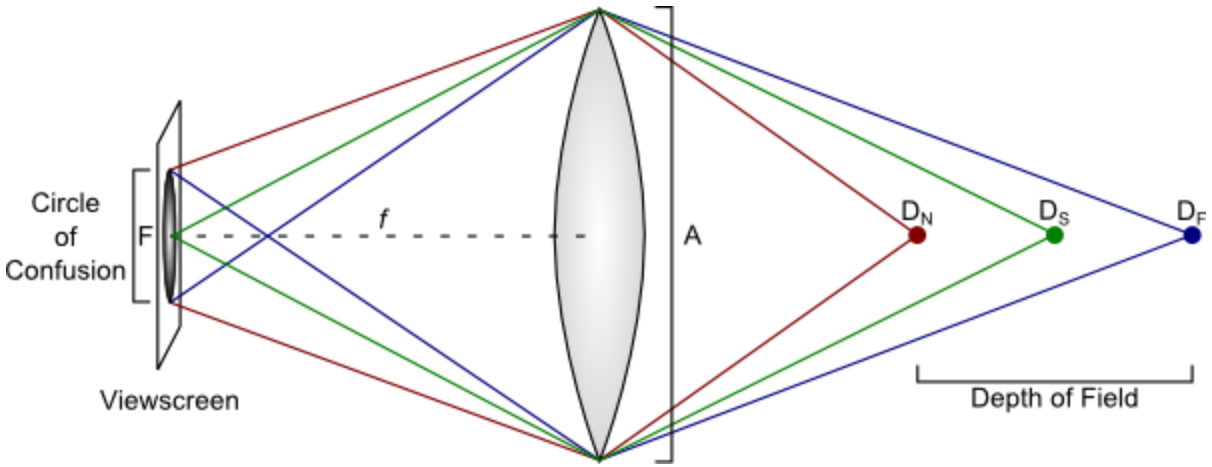


Image-based forensics?

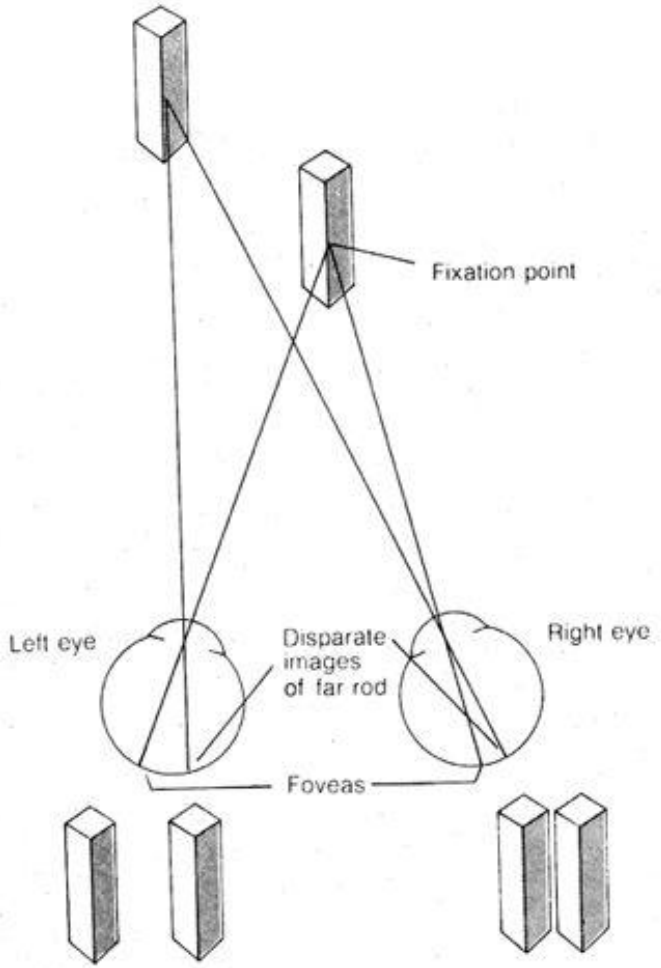
(2-B) Depth from Defocus





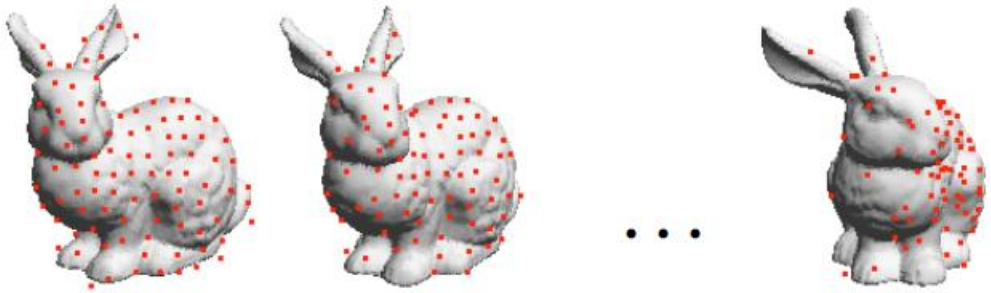
(2-C) Stereo and Disparity

Binocular disparity



http://www.cns.nyu.edu/~david/courses/perception/lecturenotes/depth/depth-size.html

(2-D) Structure from Motion



- **Input 1:** Video sequence of moving (translating + rotating) object taken from a still camera
- **Input 2:** Tracks of some *N* 2D salient points from each frame of the video sequence
- **Outputs:** 3D coordinates of each of those N points in each frame + 3D motion of the object!

https://www.youtube.com/watch?v=zdKX7Xo3Cb8&feature=player_detailpage#t=270

(3) Motion Estimation

- Input: a video sequence
- Desired Output: an estimate of the motion (2D) at all pixels in all frames
- Applications of such an algorithm: object tracking, facial expression analysis, video stabilization, etc.
- Typical assumptions: no change in illumination across frames, small motion between consecutive frames.

http://www.jonathanmugan.com/GraphicsProject/OpticalFlow/

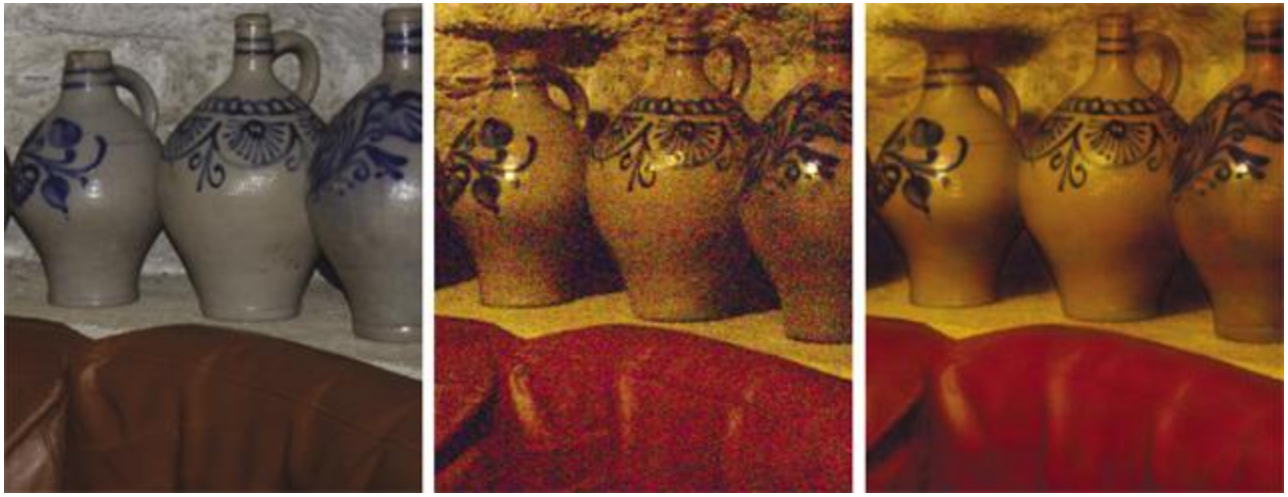
Aperture Problem:

http://en.wikipedia.org/wiki/File:Aperture_problem_animated.gif

(3) Motion Estimation

- Sometimes the motion between two images can be represented more compactly – eg: rotation, scaling, translation.

- We will look at methods to estimate such "parametric motion" – even if the images were acquired under different lighting conditions.

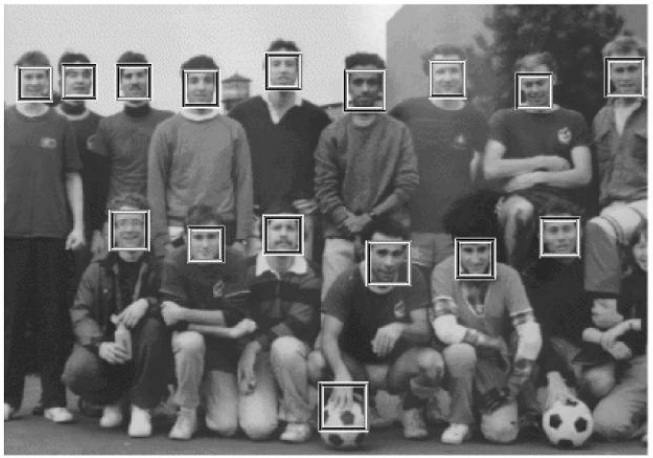
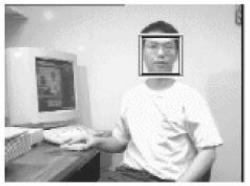
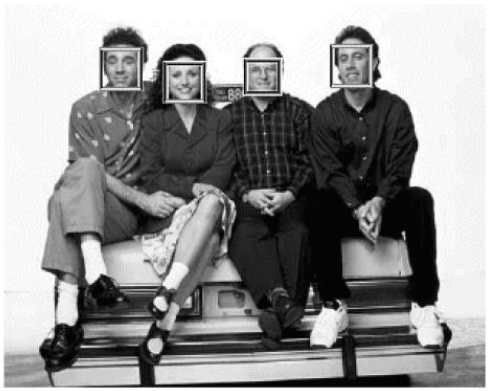
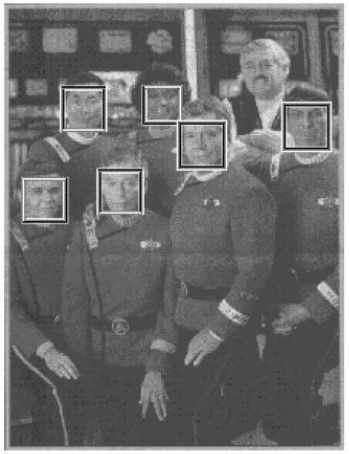
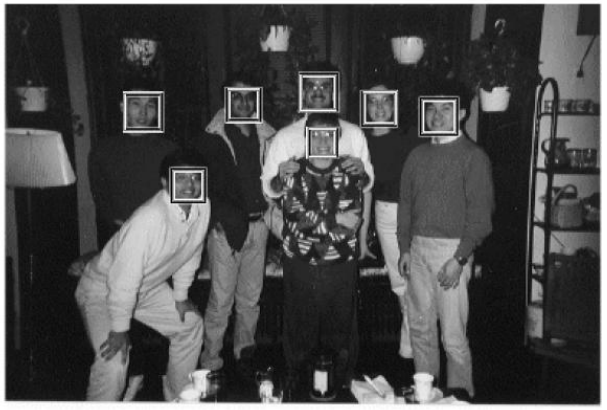


(3) Motion Estimation

- Or in cases where the general image motion is a translation or rotation, but there are various independently moving objects!

- This has some cool applications – such as video stabilization.

(4) Learning in Vision: Face Detection from Images

We will learn a machine learning technique called **Adaboost**. We will study how this technique is applied for one particular classification problem: does a small rectangular region in an image contain a face or not?

(4) Learning in Vision: Deep Neural Networks

- A flourishing sub-area in computer vision – with excellent empirical results on some vision problems.

- Will be taught by Prof. Arjun Jain (roughly after 20th March)

(4) Learning in Vision: Deep Neural Networks

- Will cover basics of neural nets, MLPs, back-propagation, stochastic gradient descent

- Will cover different architectures: convolutions neural nets, Siamese nets, triplet nets; compression of neural network architectures (for real-time or low-power applications)

- Applications in human pose estimation, correspondence estimation in 3D point clouds, neural art, etc.

Class Timings

- Tue and Fri 7:00 to 8:25 pm (slot 15) in SIC 201
- Roughly from 20th March to 14th April, this course will be taught by Prof. Arjun Jain.
- At that time, there will be lectures on Friday (as usual) and on Saturday (timing will be confirmed later). There will be no Tuesday lectures.

(+) Some “fundoo” topics along-side

- Image restoration in special settings. Example below.
- Consider an object submerged in a water tub/tank. The object is imaged from outside (camera is not in water). The water surface is wavy and shaky, leading to distortions in the pictures. Can you remove these distortions?


Jump into January
with great
stationery deal
Colgate


with great
stationery deal
Colgate

Mathematical Tools

- Numerical linear algebra (eigenvectors and eigenvalues, SVD, matrix inverse and pseudo-inverse) – **you are expected to know this.**
- Signal processing concepts: Fourier transform, convolution – **you are expected to know this.**
- Some machine learning methods (will be covered in class)

Programming tools

- **MATLAB and associated toolboxes**
- For the part to be covered by Prof. Arjun Jain, you should be willing to learn some packages such as lua and Torch7.
- OpenCV (open source C++ library)