CS 3570

Final Project

2015

Ready for challenges ?

Part 1 (5/18)

General Guideline

2~4 students form a team

Select a topic from our suggestion list, or think a new one yourself

1-page proposal on iLMS, before 5/28, including team name, members, project name, brief description and goal

If your proposal is not approved, you will know before 6/1

Please hand in another one before 6/4

General Guideline

Each topic difficulty is rated as stars

Team with more members are recommended to take a challenging topic

Hand in your report before presentation

Every team member needs to participate the presentation

Project Requirement

- Realize the algorithms (Essential)
- Implementation (Essential)

  Using open source code and toolbox are permitted
- Comparison and analysis (Suggested)
- Improvement (Quality, accuracy, speed, or others)(Optional)
- System for demonstration (Optional)

Presentation in class

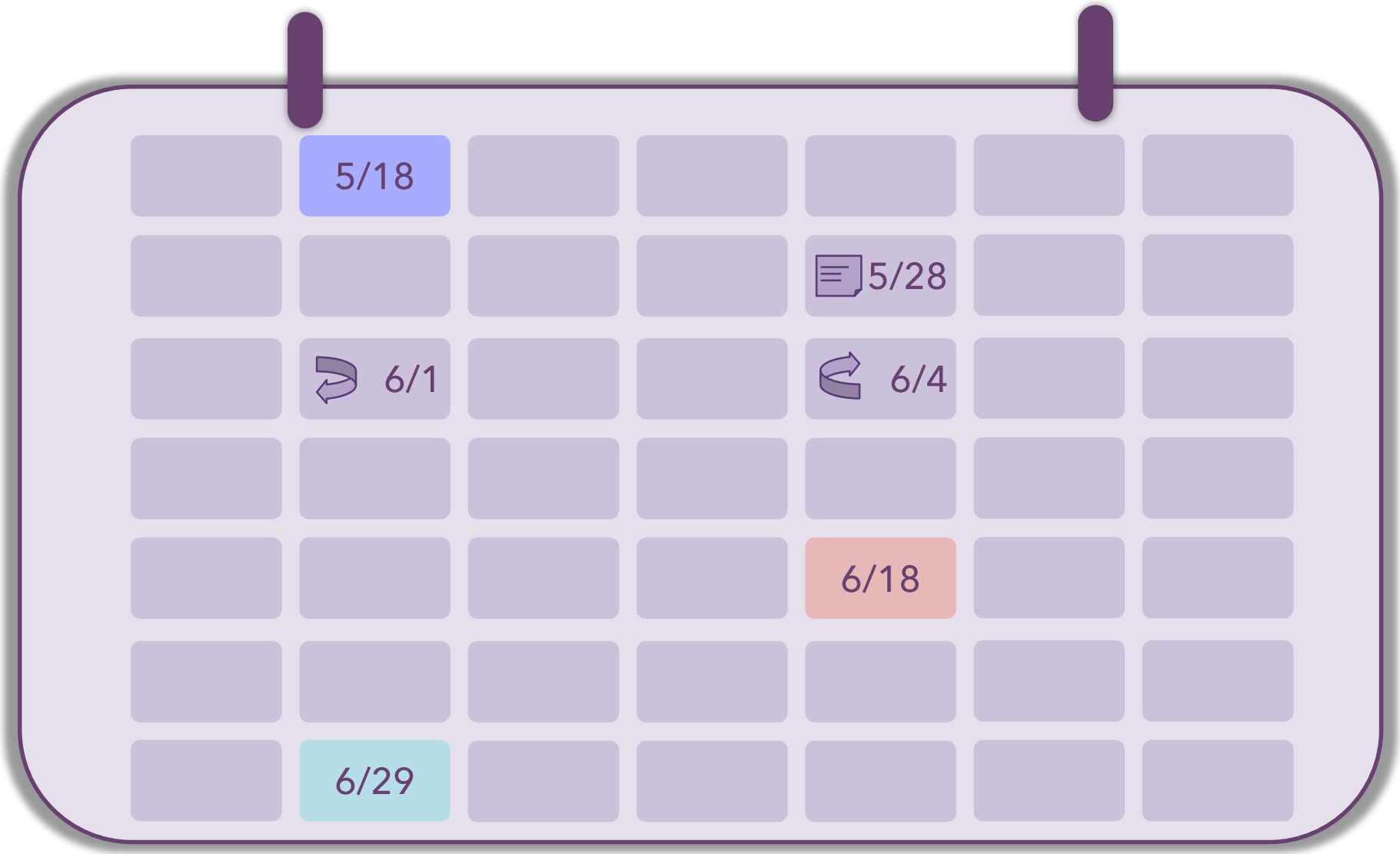
Afternoon in class

6/18 for the seniors

6/29 for the undergraduates

Best 3 teams get bonus points

Vote for presentation day



Project Categories

- Image Processing
- Machine Learning
- Graphics
- Audio Processing
- Video Processing

Image Processing

★ ★	Image denoising
★	Bilateral Filtering
★ ★ ★	Image matting
★ ★	Image inpainting

Poisson blending	★ ★
Seam carving	★ ★ ★
Special effects camera app	★ ★ ★
Rectangle rectification for document capture	★ ★

Machine Learning

★ ★ ★ Face Recognition

★ ★ ★ ★ Content Based Image Retrieval

★ ★ ★ ★ Image Classification

Graphics

★★★★ Vectorization

★★★★ OpenGL/Direct3D based game / application

Audio Processing

★★ DTMF Recognition

★★★ Music genre classification

★★★ Chord Transcription

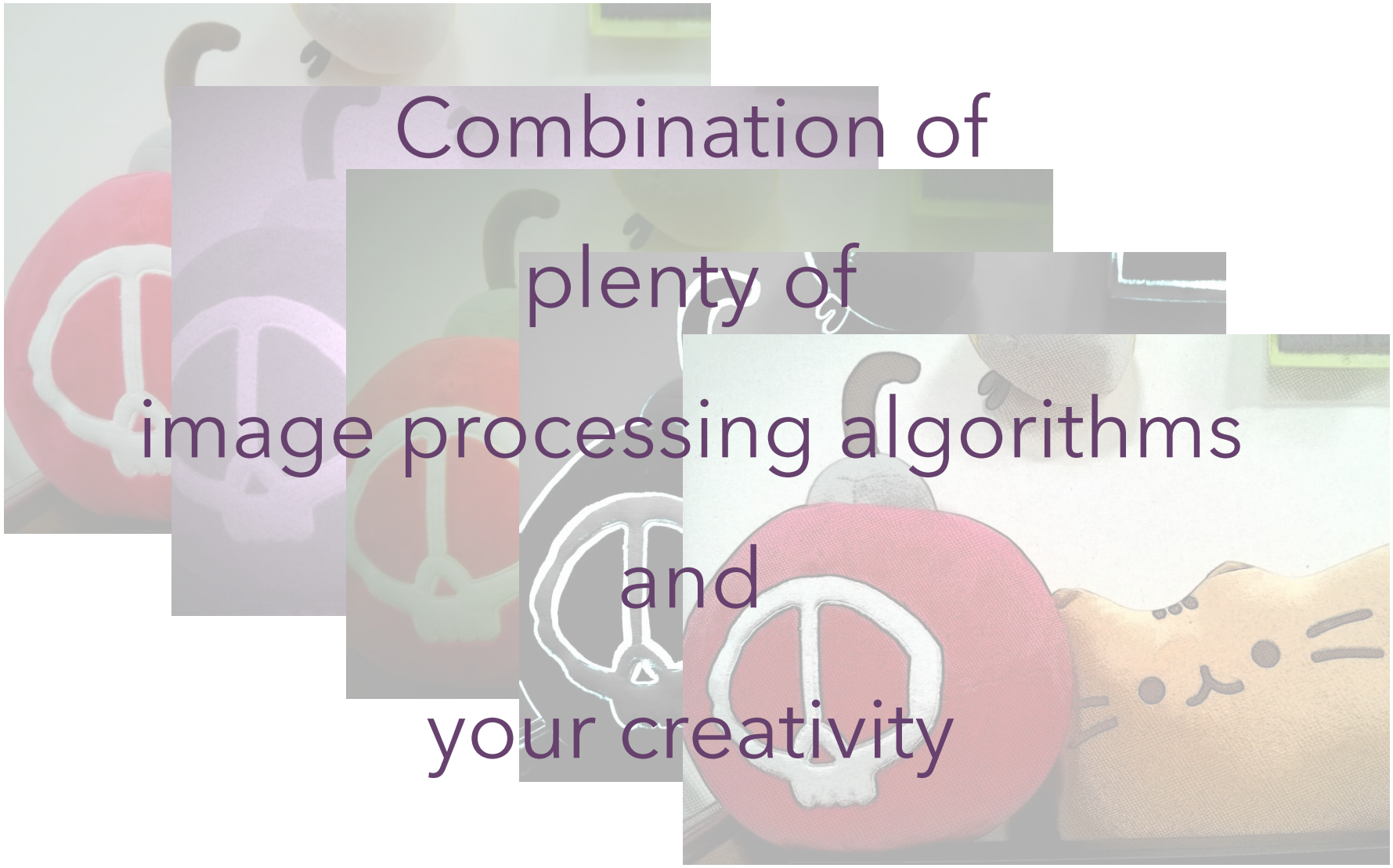
Video Processing

★ ★ ★ Background subtraction

★ ★ ★ Video denoising

★ ★ ★ Video stabilization

Special effects camera app



Special effects camera app

Original Photo

Powered by Camera 2

Special effects camera app

Old time

Spotlight effect

Color Transfer

Salt-and-pepper noise

Powered by Camera 2

Special effects camera app

Lomo

Spotlight effect

Color Transfer

Powered by Camera 2

Special effects camera app

Scientific

RGB to gray

Contrast Enhancing

Edge detection

Powered by Camera 2

Special effects camera app

Comic
(Non-photorealistic rendering)

Color Quantization

Add mesh texture in shadow regions

Edge detection

Powered by Camera 2

Special effects camera app

Face detection

Face pose estimation

Render some template image onto the face

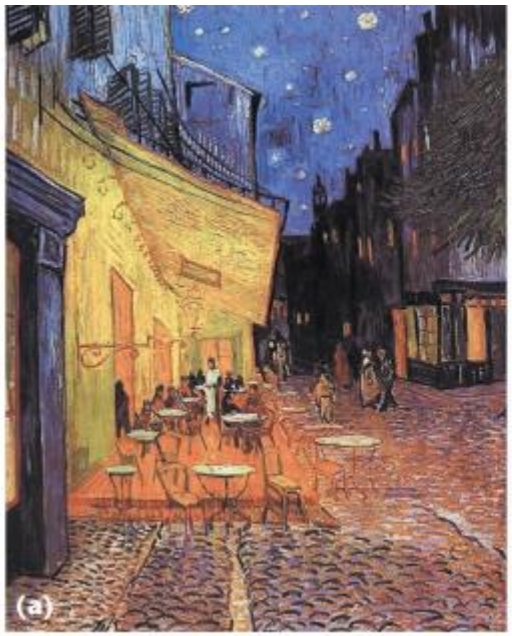
Powered by Hangouts

Color transfer

Target image

Reference image

Result image



Reference image

Target image

Result image

Color transfer

- Target image divided into RGB channels
- Reference image divided into RGB channels too.
- Use histogram equalization to let the histogram of target image be similar to reference image.

Special effects camera app

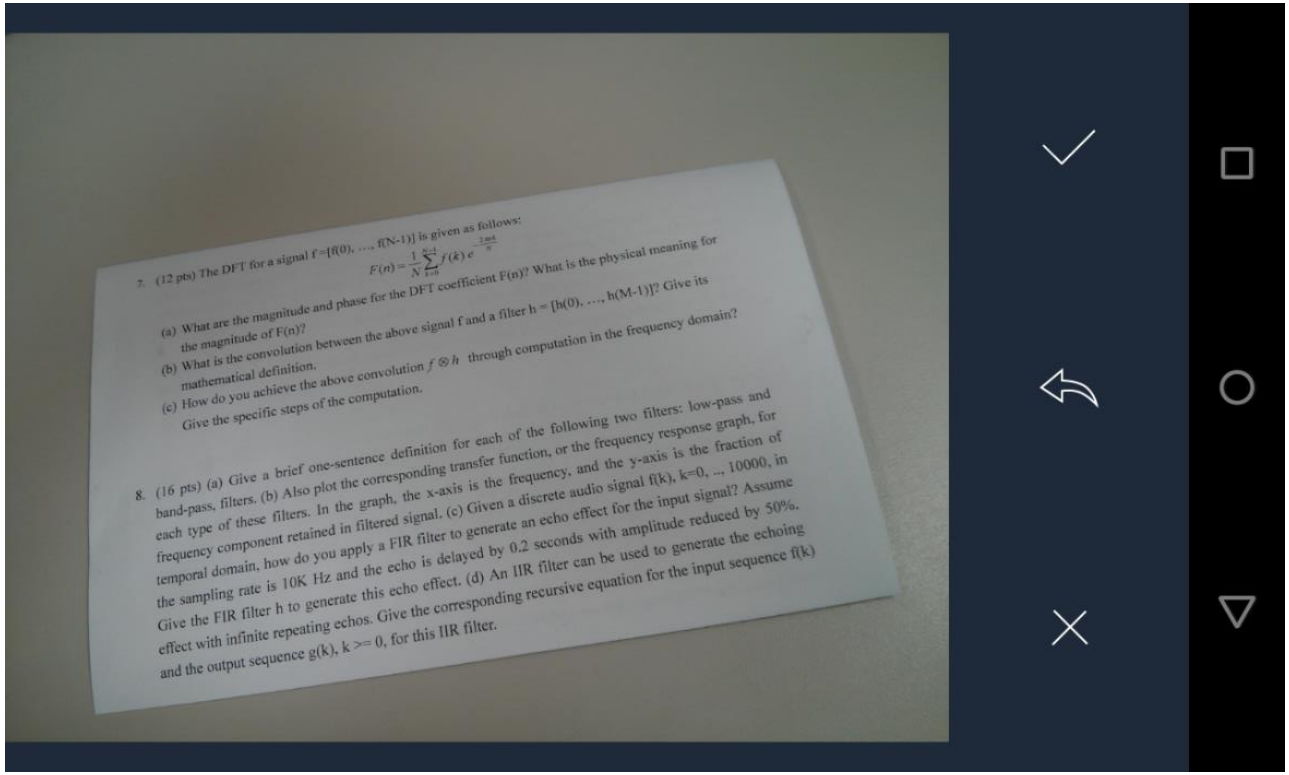
An interactive app or application with user interface

for live demonstration is necessary

Seek existing special-effect camera for inspiration

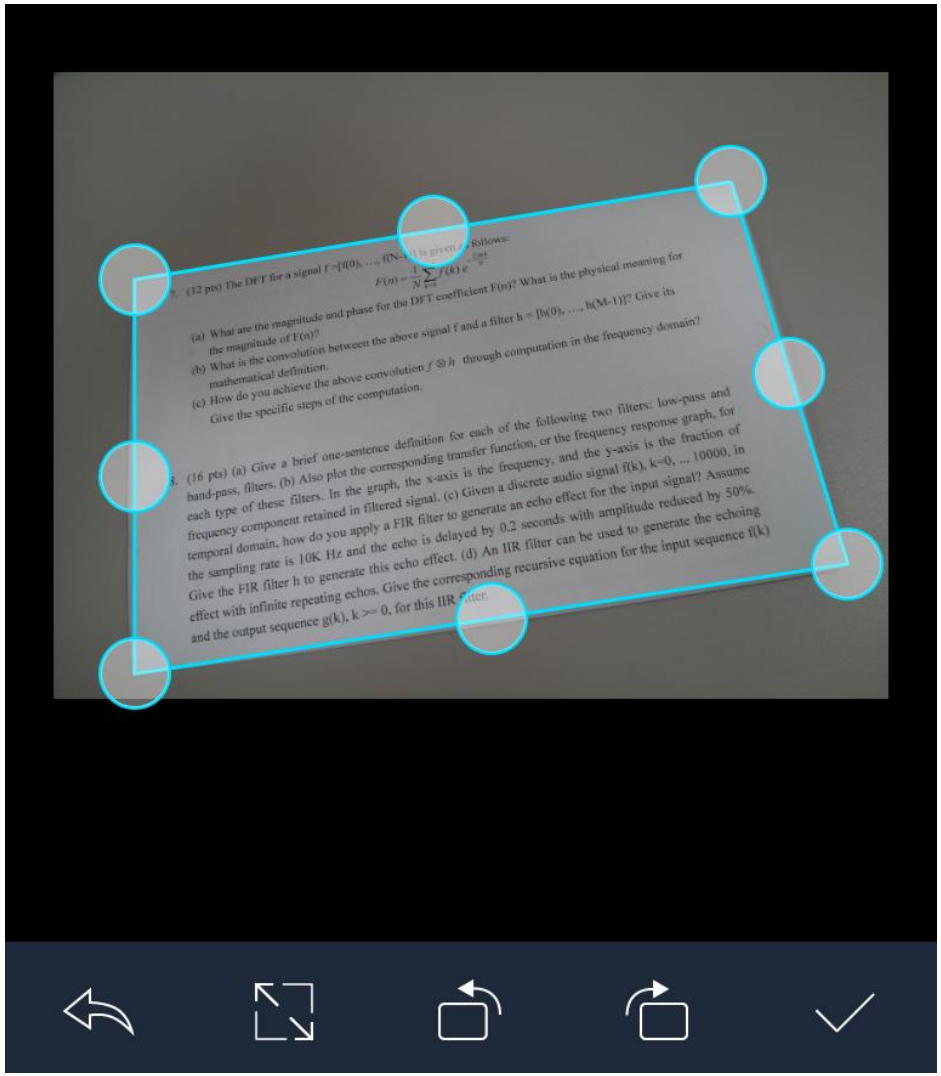
Realize your creativity

Rectangle rectification for document capture



Take a photo of a document

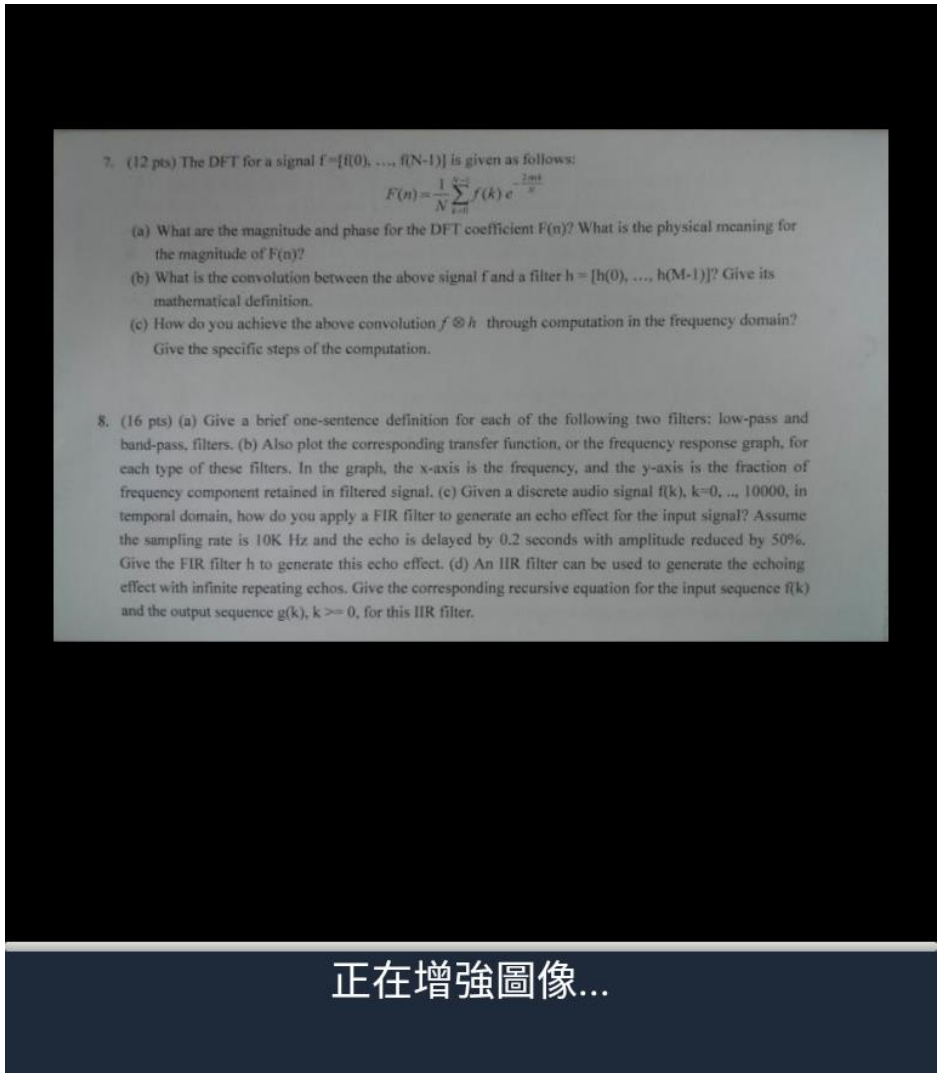
Rectangle rectification for document capture



Select 4 corners

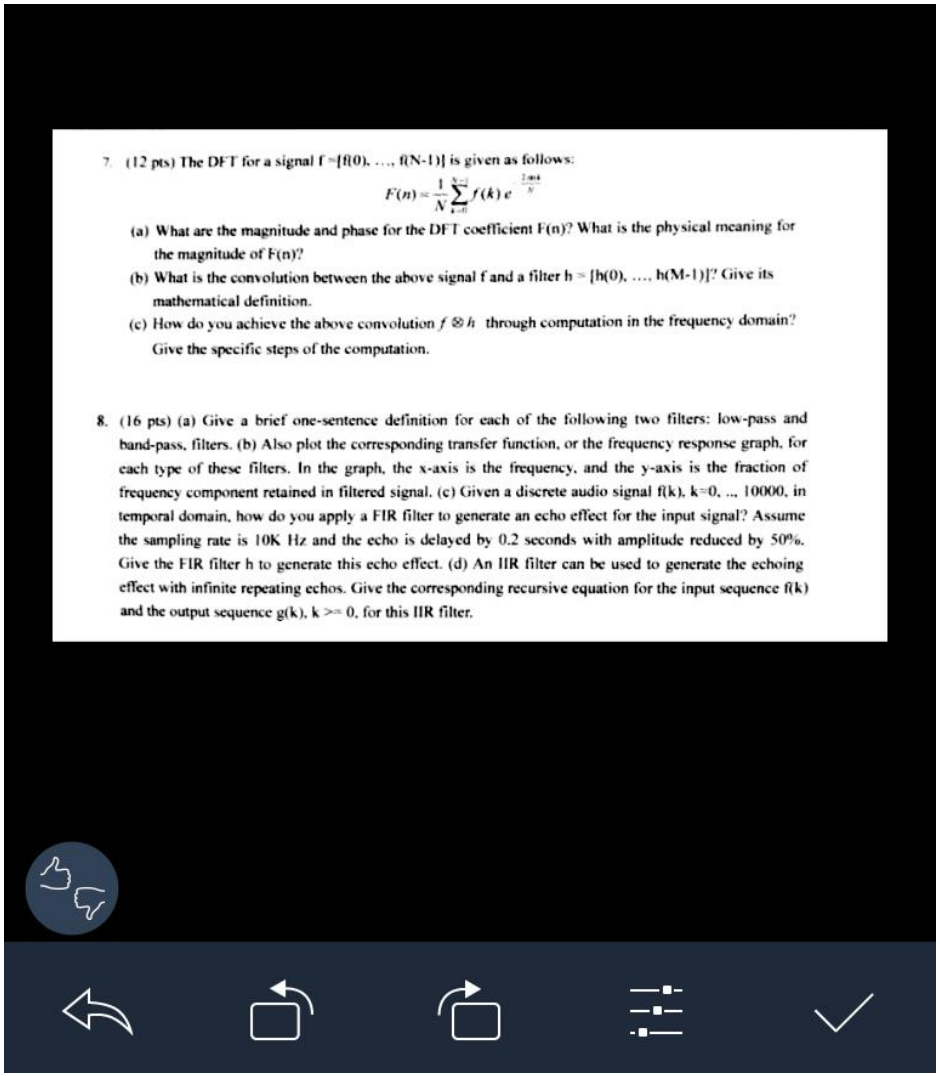
of this document

Rectangle rectification for document capture



Apply Homography for rectification

Rectangle rectification for document capture



7. (12 pts) The DFT for a signal f =[f(0), …, f(N-1)] is given as follows:

$$F(n) = \frac{1}{N}\sum_{k=0}^{N-1} f(k)\, e^{-\frac{2\pi nk}{N}}$$

(a) What are the magnitude and phase for the DFT coefficient F(n)? What is the physical meaning for the magnitude of F(n)?

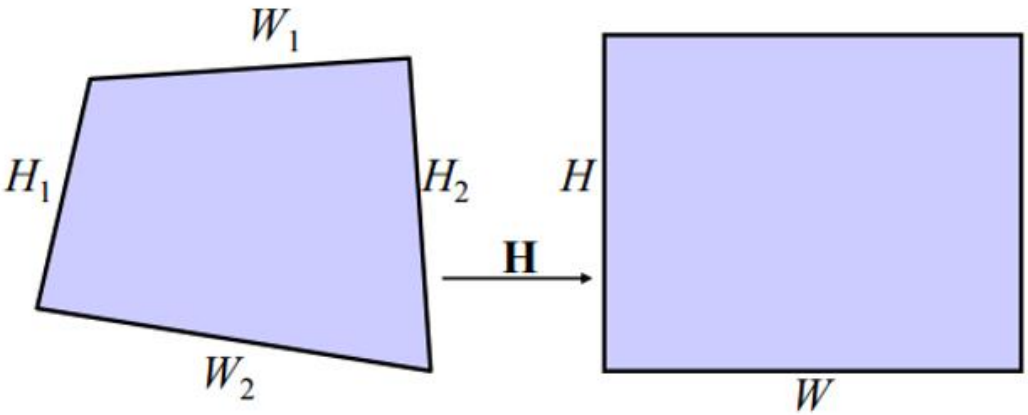
(b) What is the convolution between the above signal f and a filter h = [h(0), …, h(M-1)]? Give its mathematical definition.

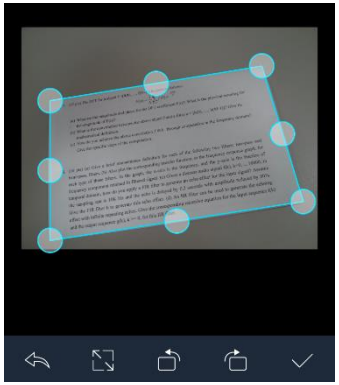
(c) How do you achieve the above convolution $f \otimes h$ through computation in the frequency domain? Give the specific steps of the computation.

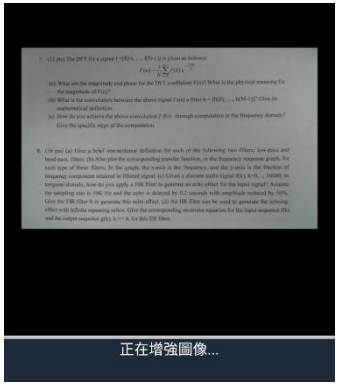
8. (16 pts) (a) Give a brief one-sentence definition for each of the following two filters: low-pass and band-pass, filters. (b) Also plot the corresponding transfer function, or the frequency response graph, for each type of these filters. In the graph, the x-axis is the frequency, and the y-axis is the fraction of frequency component retained in filtered signal. (c) Given a discrete audio signal f(k), k=0, .., 10000, in temporal domain, how do you apply a FIR filter to generate an echo effect for the input signal? Assume the sampling rate is 10K Hz and the echo is delayed by 0.2 seconds with amplitude reduced by 50%. Give the FIR filter h to generate this echo effect. (d) An IIR filter can be used to generate the echoing effect with infinite repeating echos. Give the corresponding recursive equation for the input sequence f(k) and the output sequence g(k), k >= 0, for this IIR filter.

Reconstructed document

Homography







Left : original shape ; Right : rectified shape

Homography example

Left points Right points

H =

0.0084	0.0009	0.6518
0.0047	0.0052	-0.7583
0.0000	0.0000	0.0042

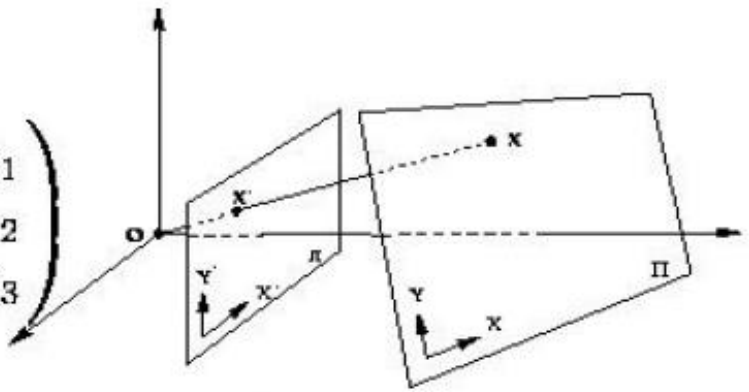
Backward warping
(bilinear interpolation)

input

output

Homography

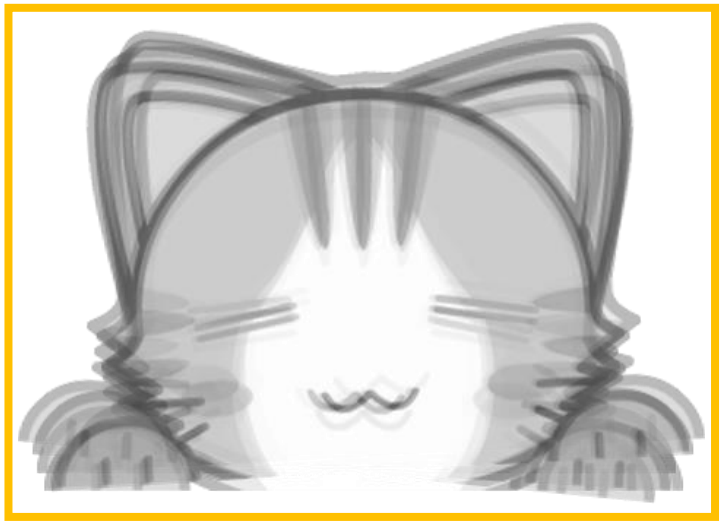
$$\begin{pmatrix} x_1' \\ x_2' \\ x_3' \end{pmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$



or $\mathbf{x}' = \mathtt{H}\mathbf{x}$, where $\mathtt{H}$ is a 3×3 non-singular homogeneous matrix.

- This is the most general transformation between the world and image plane under imaging by a perspective camera.
- It is often only the 3×3 form of the matrix that is important in establishing properties of this transformation.
- A projective transformation is also called a "homography" and a "collineation".
- $\mathtt{H}$ has 8 degrees of freedom.

Video Stabilization



Shaky Video

Stabilized Video

Video Stabilization

Why your video is shaky ?

Your camera is shaky

So we need to detect how camera shakes?

No, just how video shakes

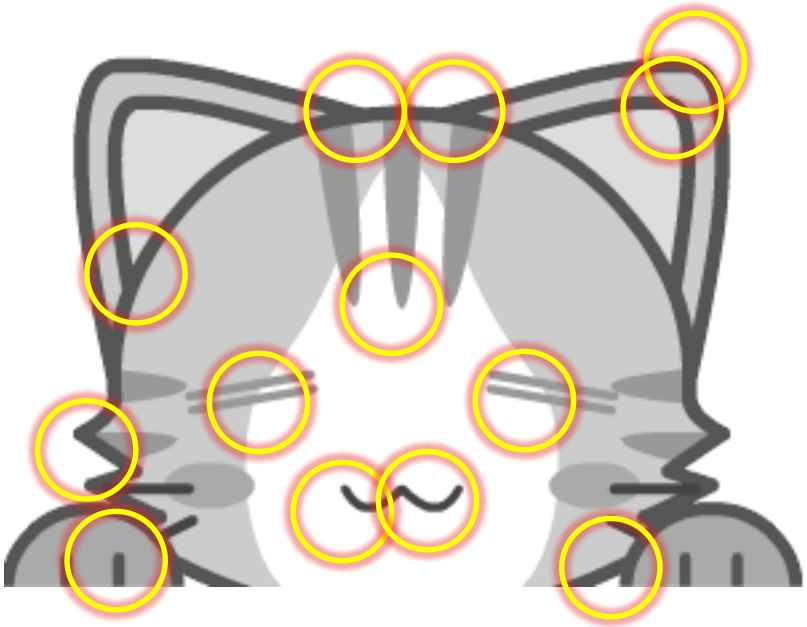
And compensate it

Video Stabilization

Youtube

Video Stabilization

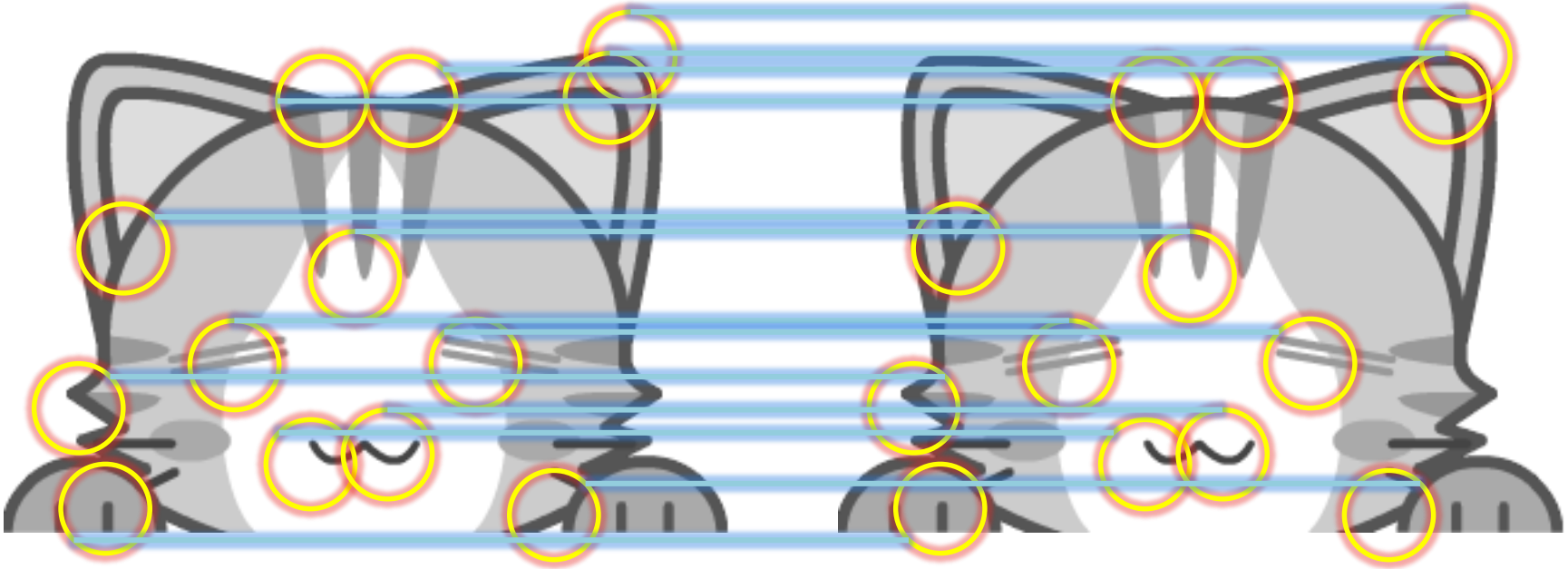
Local feature descriptor



Video Stabilization

Local feature descriptor

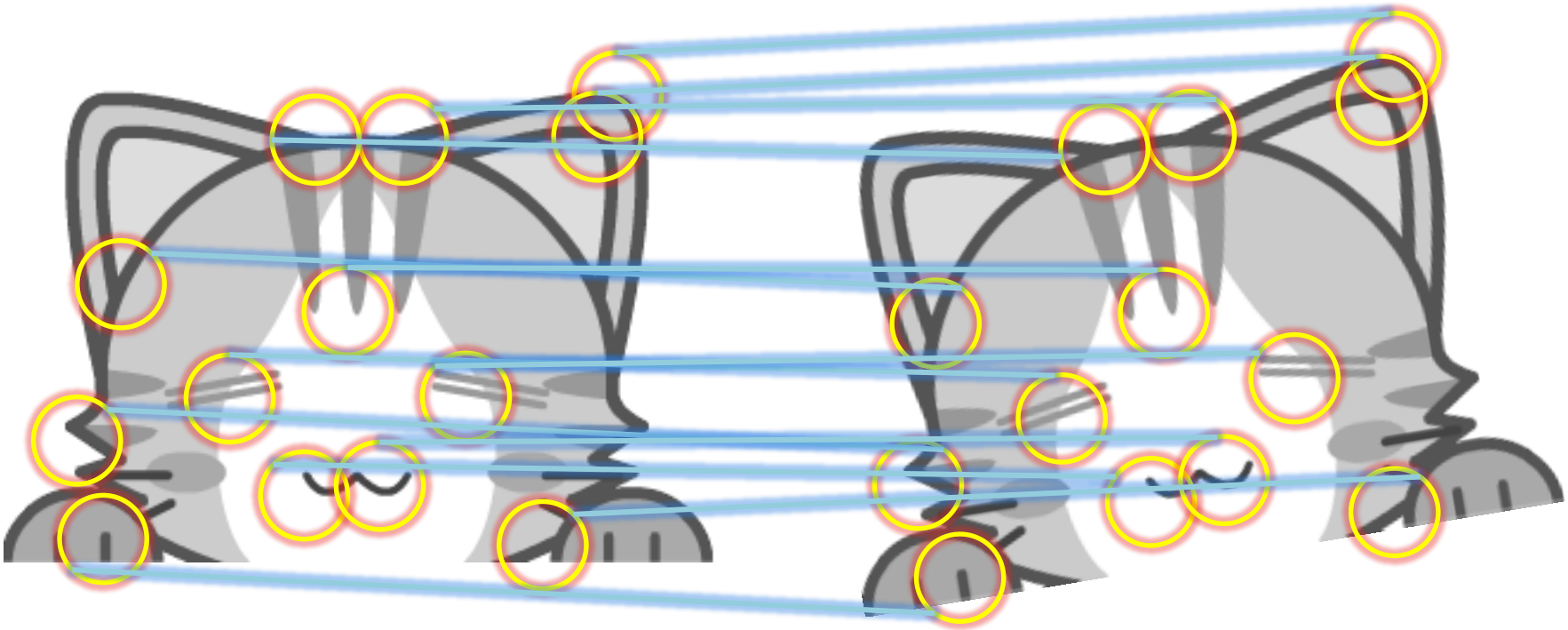
Feature correspondence



Video Stabilization

Local feature descriptor

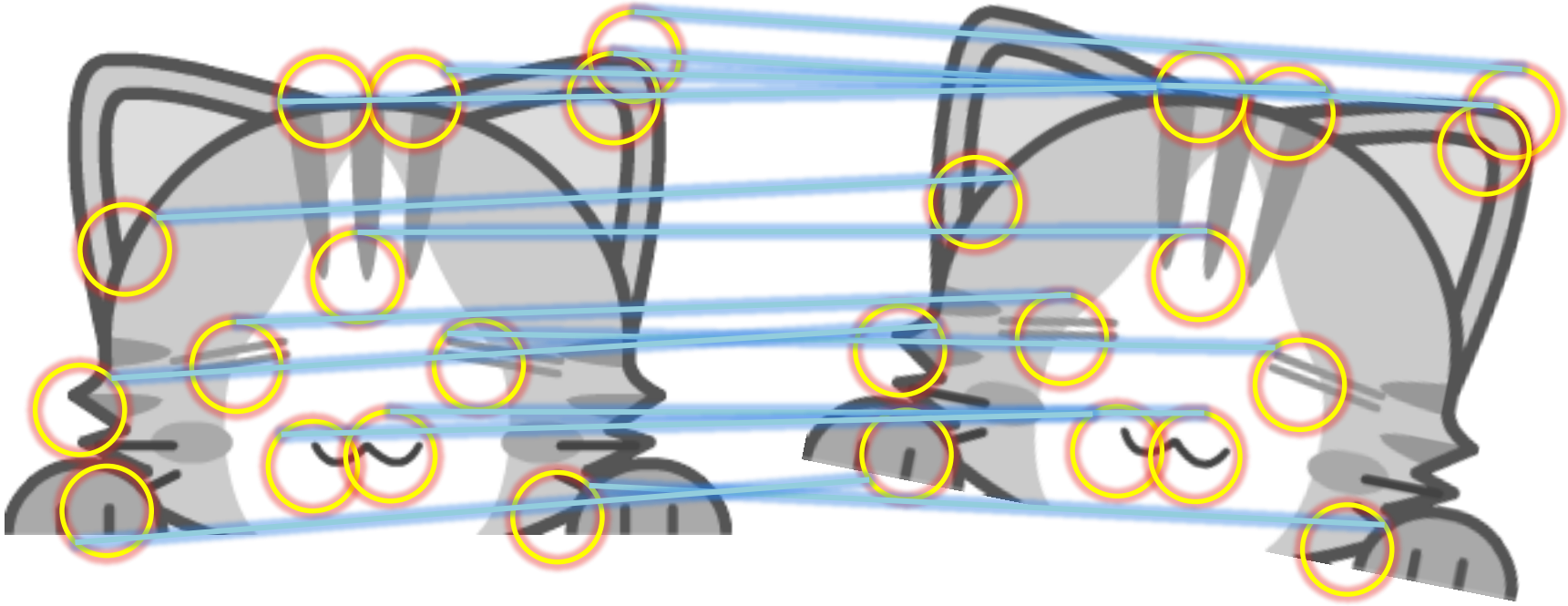
Feature correspondence



Video Stabilization

Local feature descriptor

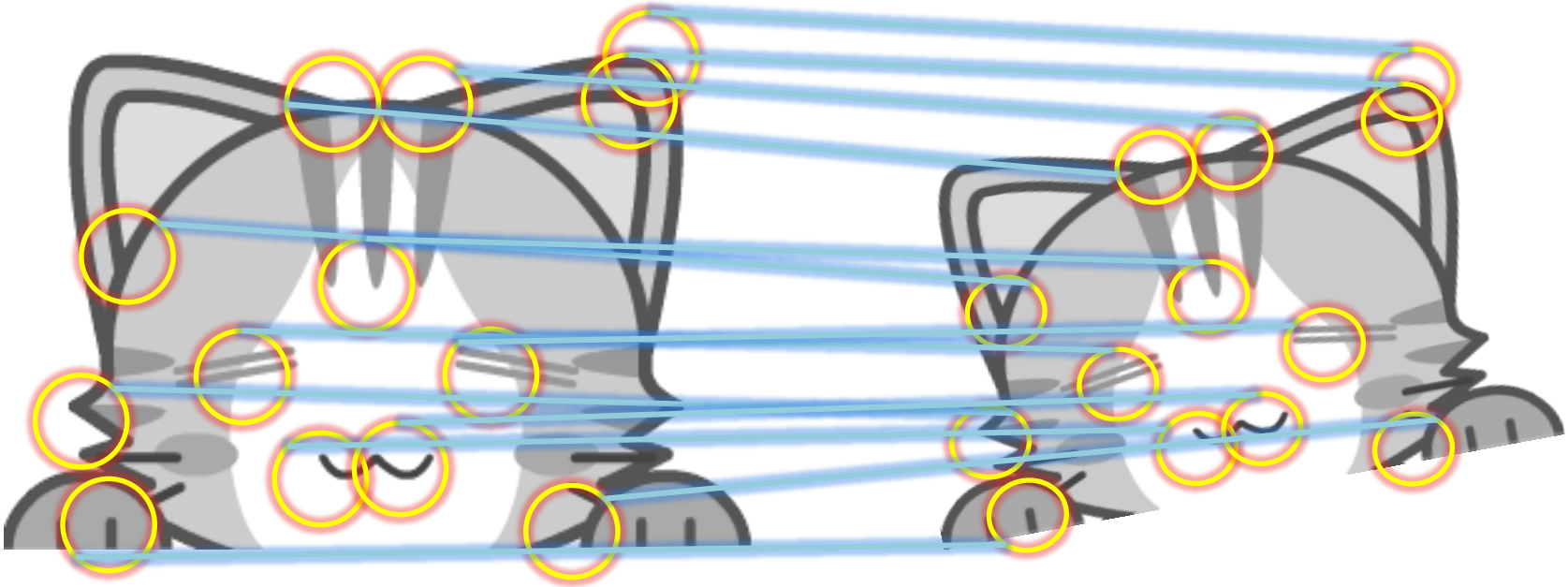
Feature correspondence



Video Stabilization

Local feature descriptor

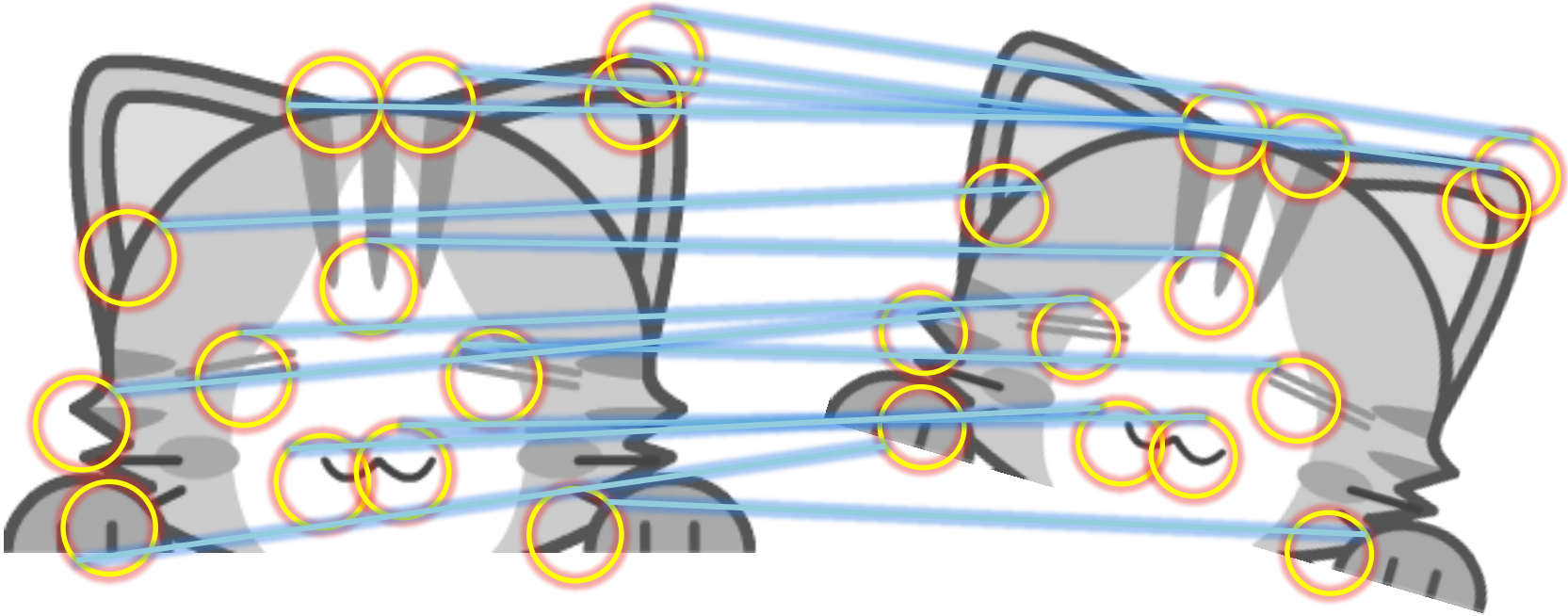
Feature correspondence



Video Stabilization

Local feature descriptor

Feature correspondence



Video Stabilization

Extract Local features

Match Features

Eliminate Outliers

Compute Homography

Warp Frames

MATLAB tutorial

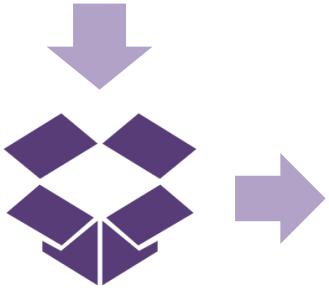
Content-based Image Retrieval

Machine Learning

Tag Based Image Retrieval

Content Based Image Retrieval

"Angry Birds"



Retrieve with tags

Angry Birds

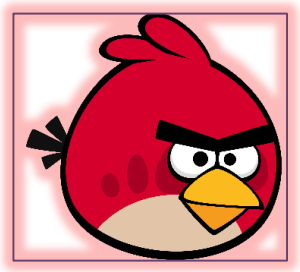
Angry Birds Rio

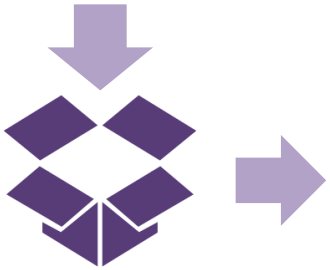
Angry Birds Space

Content-based Image Retrieval

Tag Based Image Retrieval

Content Based Image Retrieval





Retrieve with features

Similar

Less Similar

Content-based Image Retrieval

What is “similarity” ?

Similarity between images

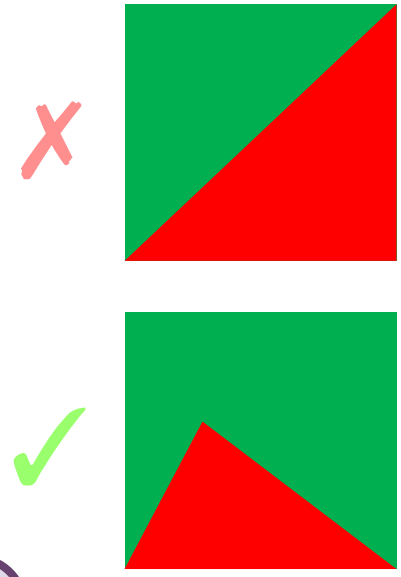
is

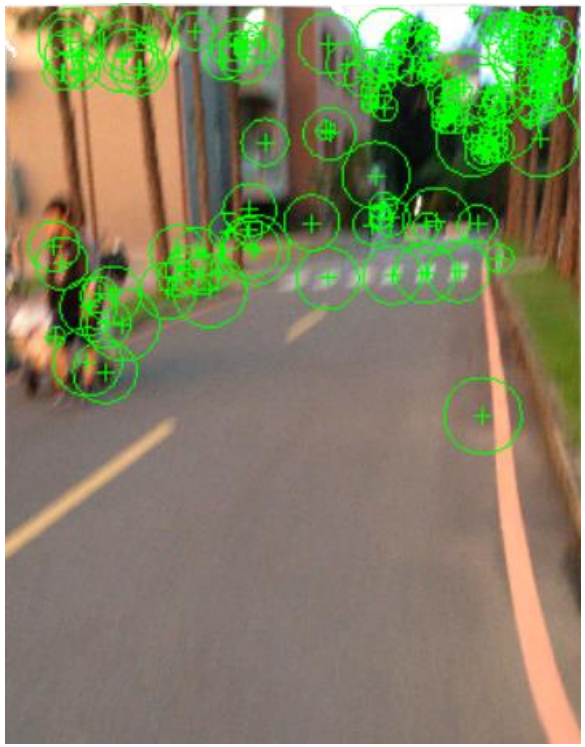
Similarity between features

Content-based Image Retrieval

Machine Learning

Each Image
in Dataset





Detect Key-points
(in gray scale)

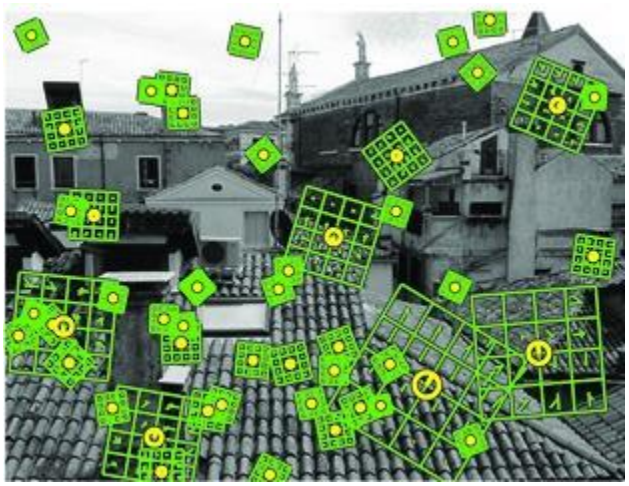
Position Scale

Content-based Image Retrieval

Machine Learning

Position | Scale | Rotation

Describe these Key-points
(For ex : using SIFT)



For example : SIFT

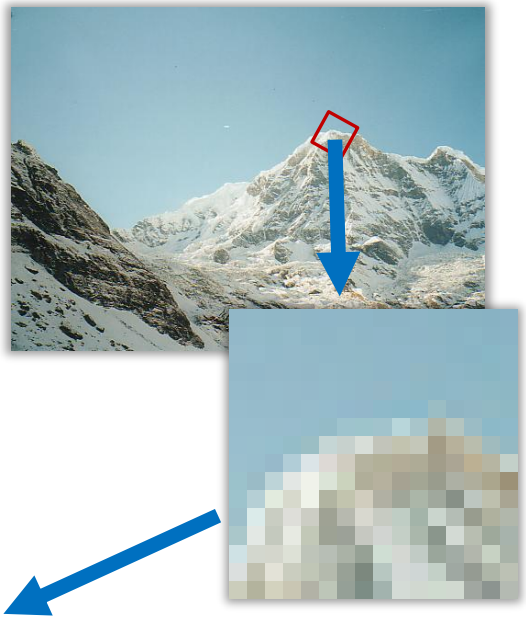
An SIFT feature is
a normalized 128-D floating-point vector

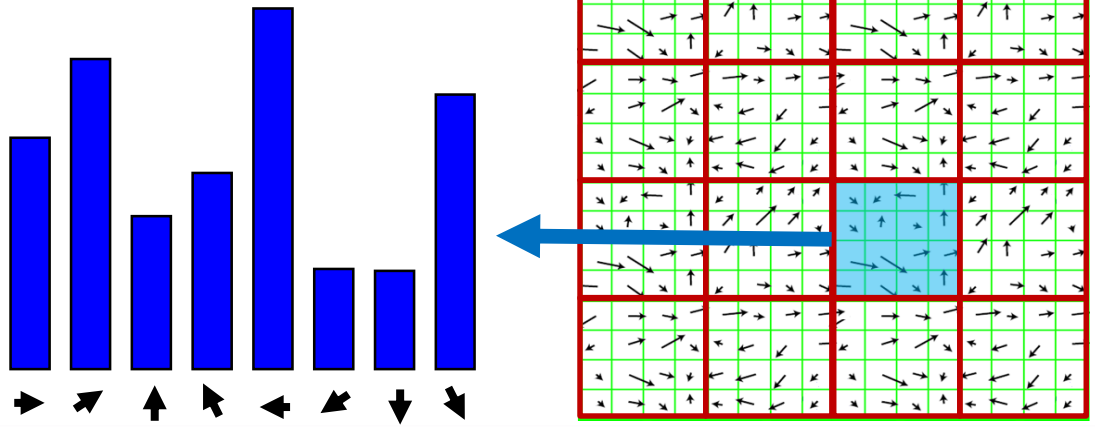
Content-based Image Retrieval

Machine Learning

Position | Scale | Rotation

Describe these Key-points
(For ex : using SIFT)



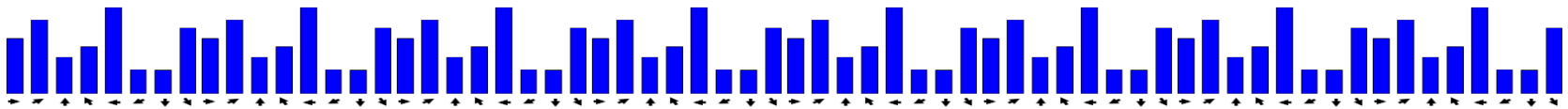


Content-based Image Retrieval

Machine Learning

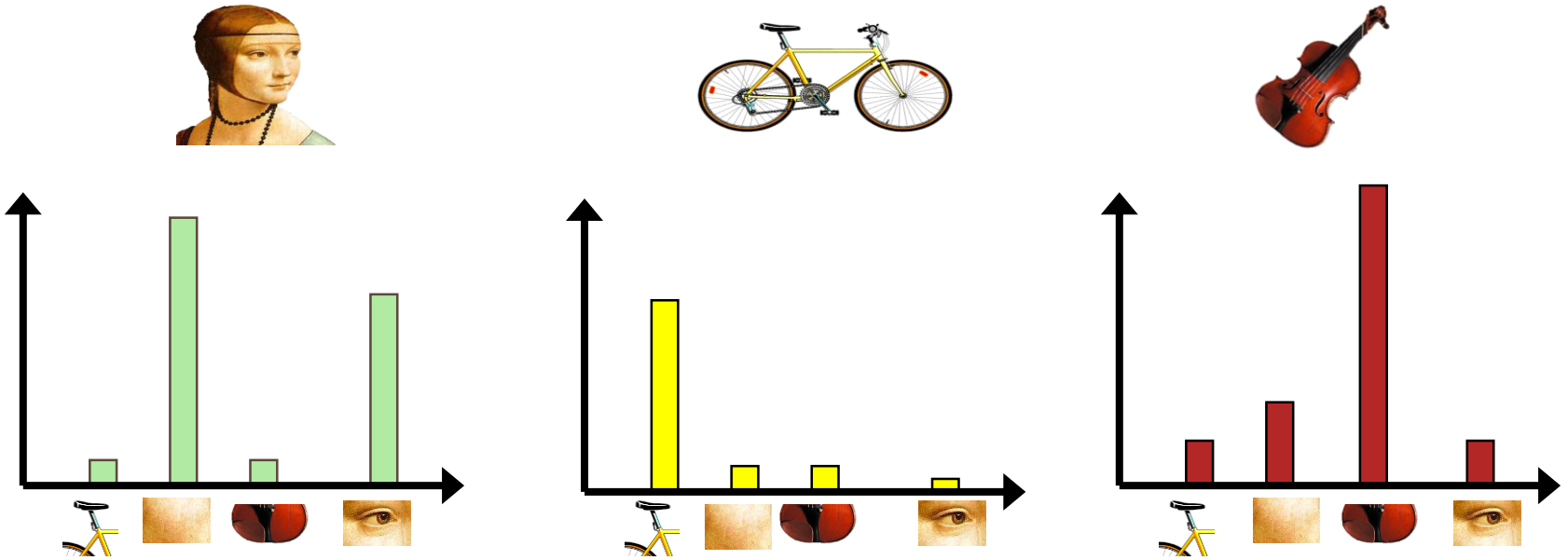
Position | Scale | Rotation

Describe these Key-points
(For ex : using SIFT)



An SIFT feature is
a normalized 128-D floating-point vector

Bag of (visual) words

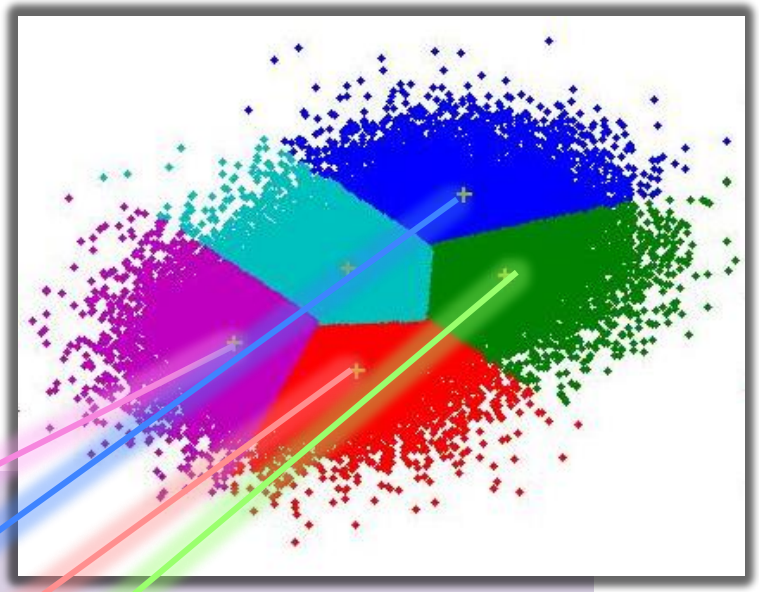


Content-based Image Retrieval

Machine Learning

128-D Vector ≡ point in a 128-D space

Quantization
(If using Bag of Words Model)



10k
~
1M

Visual Words 1
Visual Words 2
Visual Words 3
Visual Words 4
⋮

Dictionary

Content-based Image Retrieval

Machine Learning

Compare and Compute Similarity

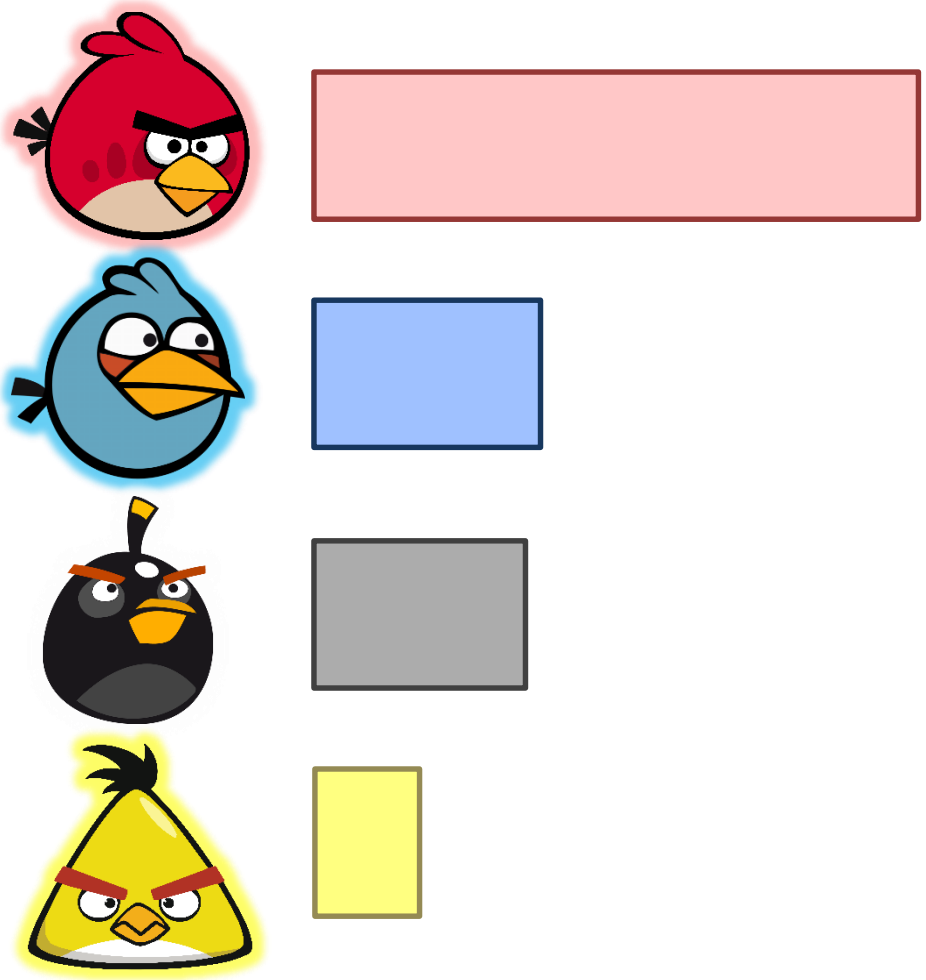
Euclidean Distance
Taxicab Distance
Hamming Distance
Cosine Similarity
…

Content-based Image Retrieval

Machine Learning

Sort Similarity Descending

query



Content-based Image Retrieval

Datasets

- INRIA Holidays
- Flickr 11k
- Caltech 101
- Caltech 256
- Bing Grand Challenge 2013

Image Classification

- How to do “Classification” ?

  1. Preprocessing
  2. Feature extraction
  3. Modeling
     - Description of each class in mathematical form
  4. Classification
     - The classifier divides the feature space into class regions

Image Classification

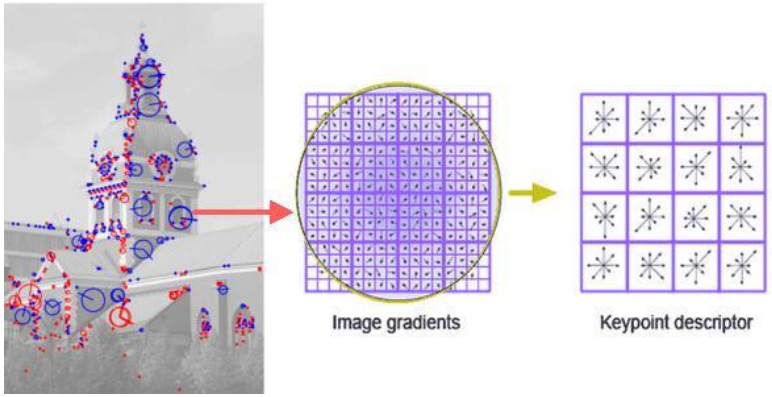
- Preprocessing

  Cluster: K-means, Mean-shift…

  Segmentation: Snake, Graph cut…

- Feature extraction

  Feature: color intensity, SIFT, LBP…



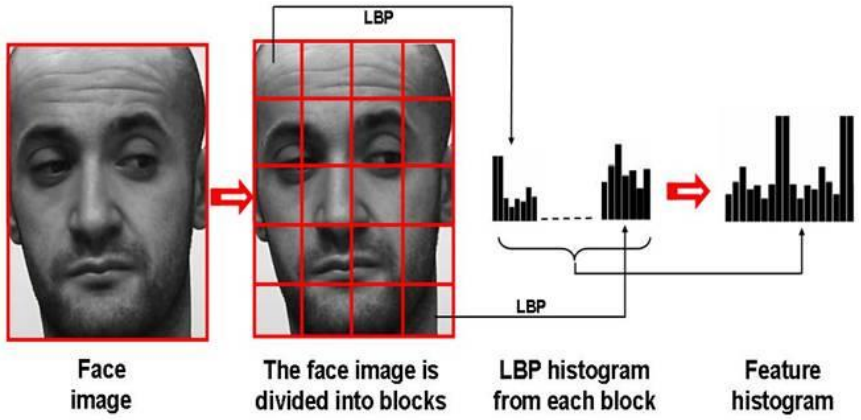


Image Classification

- Feature
  - A vector to describe a sample
  - e.g. Apply BoW to get a 10000-D feature
  - ≡ a point in 10000-D space, called feature space
- Label
  - A scalar, indicate the class to which a sample belongs
  - e.g. for a 10-class classification problem, labels: 0, 1, …, 9

Image Classification

- Represent a sample as features and label
- Sample 1
  - Feature: 0, .3, 0, 0, 0, 0, 0, .1, 0, …, .2, 0, 0
  - Labeled 4
- Sample 2
  - Feature: 0, 0, .1, .1, 0, 0, 0, 0, 0, …, .2, 0, 0
  - Labeled 5
- …

Image Classification

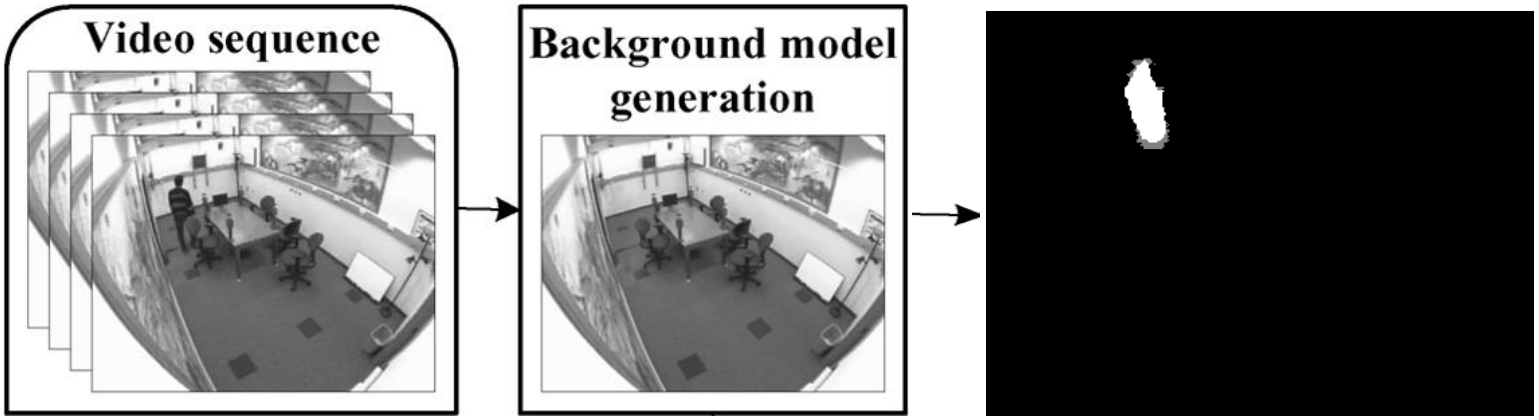
- Classification problem is
- Training on plenty of samples
  with features and labels
- To get a model, such that
- The model can classify samples with features
  into correct classes
- Not 100% correctness, but to maximize the
  probability of correct classification

Image Classification

- Classifiers
  - Probability based or Geometry based
- Bayesian classifier
- Support Vector Machine (LibSVM)
- Kernel tricks (“plug-in” of classifiers)

Background Subtraction

- **Background subtraction** is a technique in the fields of computer vision wherein an image's foreground is extracted for further processing
- BS is a widely used approach for detecting moving objects in videos from static cameras, e.g. surveillance tracking, human poses estimation…

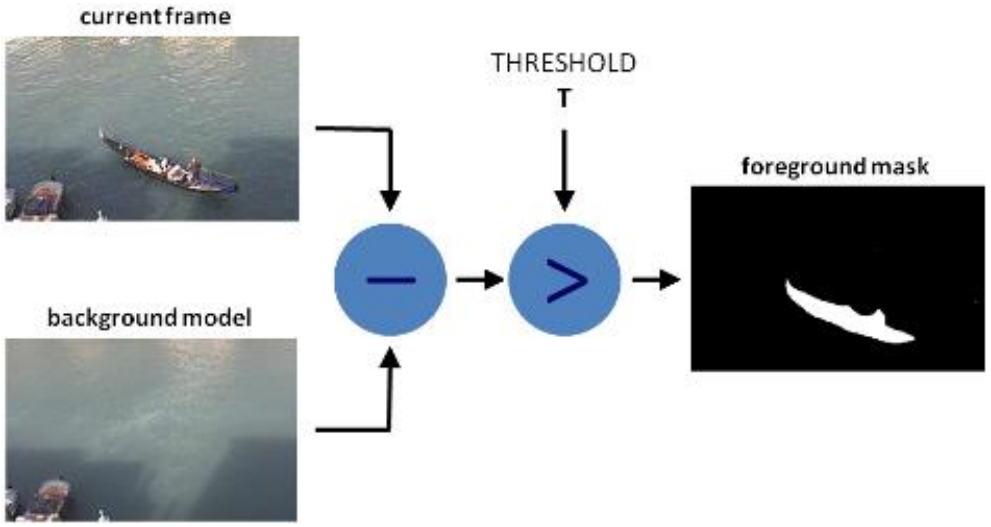


Background subtraction

- The simplest way is that segment out the objects by using image subtraction technique
- For each pixels in I(t), take the pixel value P[I(t)] and subtract it with the corresponding pixels at the same position on the background image P[B].

$$P[F(t)] = P[I(t)] - P[B]$$

$$|P[F(t)] - P[F(t+1)]| > \text{Threshold}$$



Background subtraction

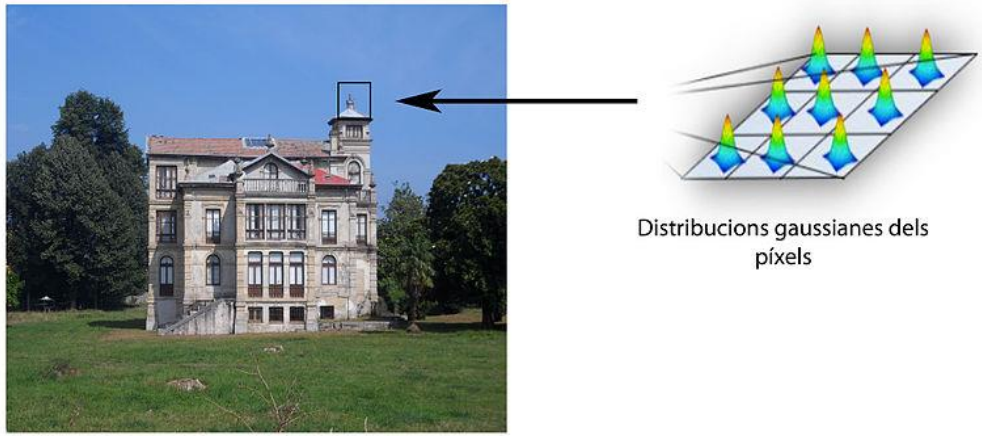
- Running Gaussian average adapt the background which changes over time (e.g. illumination changes or non-static background objects)
- mean μ_t and variance σ_t at time t

  I_t is the value of the pixel's intensity at time t

  the parameter k is a free threshold

$$\frac{|(I_t - \mu_t)|}{\sigma_t} > k \longrightarrow Foreground$$

$$\frac{|(I_t - \mu_t)|}{\sigma_t} \leq k \longrightarrow Background$$



Background subtraction - Example

Background subtraction

- Better approach: BackgroundSubtractorMOG

http://opencv-python-tutroals.readthedocs.org/en/latest/py_tutorials/py_video/py_bg_subtraction/py_bg_subtraction.html

- Related paper & book

https://sites.google.com/site/backgroundsubtraction/overview

- BS dataset

http://bmc.iut-auvergne.com/?page_id=24

To be continued

More topics to be reviewed in the next class