Pattern Recognition

Kuan-Wen Chen

2022/3/10

Introduction

- Examples of PR in our daily life:

  - recognize a face
  - read handwritten characters
  - understand spoken words
  - identify car keys in the pocket by feel
  - decide whether a fruit is ripe by its smell
  - etc.

The Goal of PR

- To design and build machines that can recognize pattern. (as an engineering field)

- To gain deeper understanding and appreciation for PR systems in the natural world – particularly in human. (as a science)

Applications

- Face detection, recognition, and verification
- Speech and speaker recognition
- Fingerprint identification
- OCR and document analysis
- Industrial Inspection
- Medical diagnostics
- DNA sequence analysis
- etc.

An Example

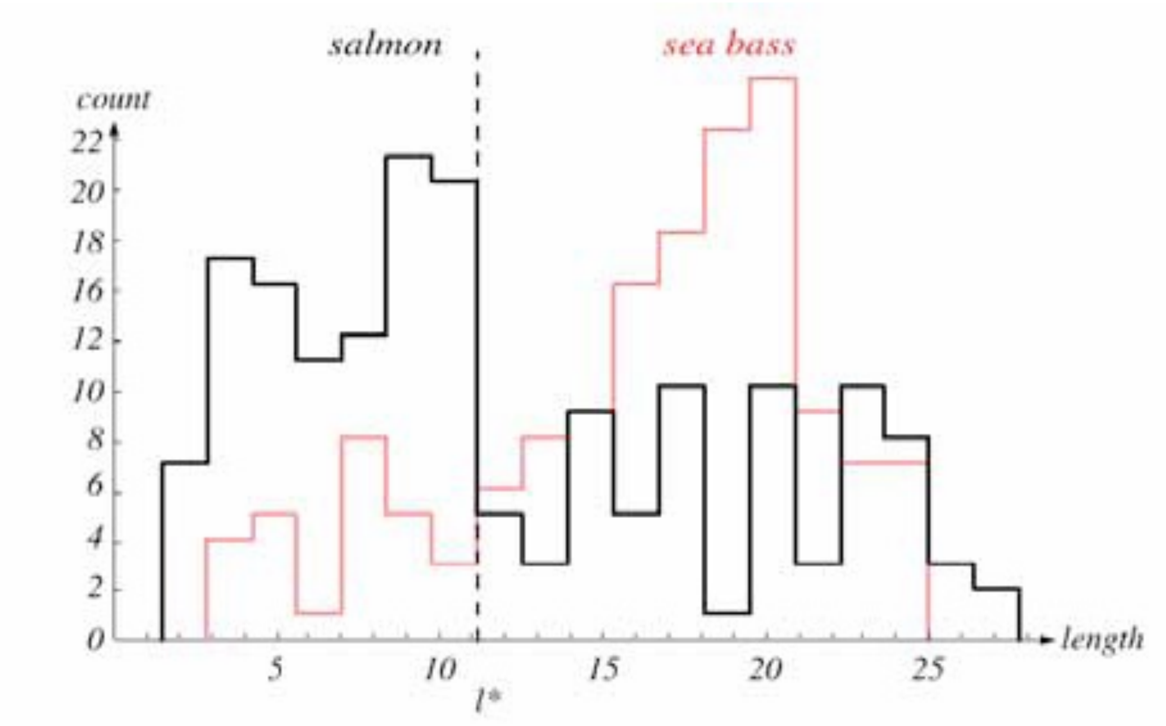
- A pilot project: separate sea bass from salmon.

- Physical differences between the two types of fishes: length, lightness, width, number and shape of fins, position of the mouth, etc.

→ possible features to be used in the classifier

An Example

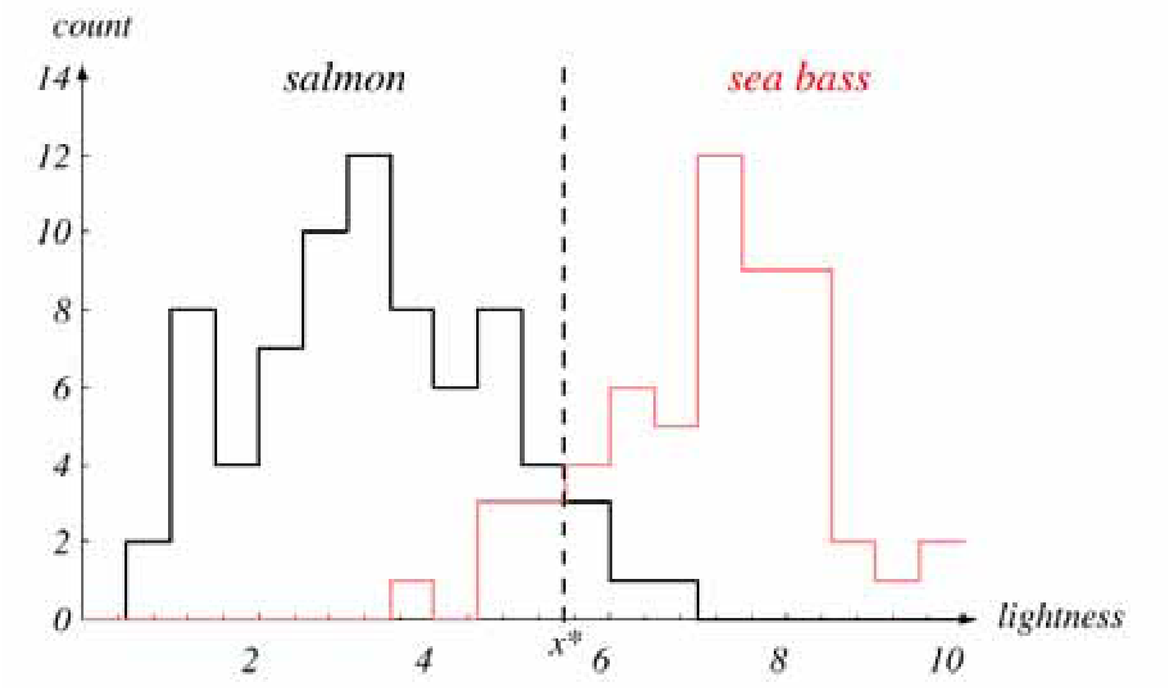
- Histogram for the length feature - obtained with some training samples



difficult to choose a good threshold, *l**

An Example

- Histogram for the lightness feature - average lightness of the fish scales



easier to choose a good threshold, *x**

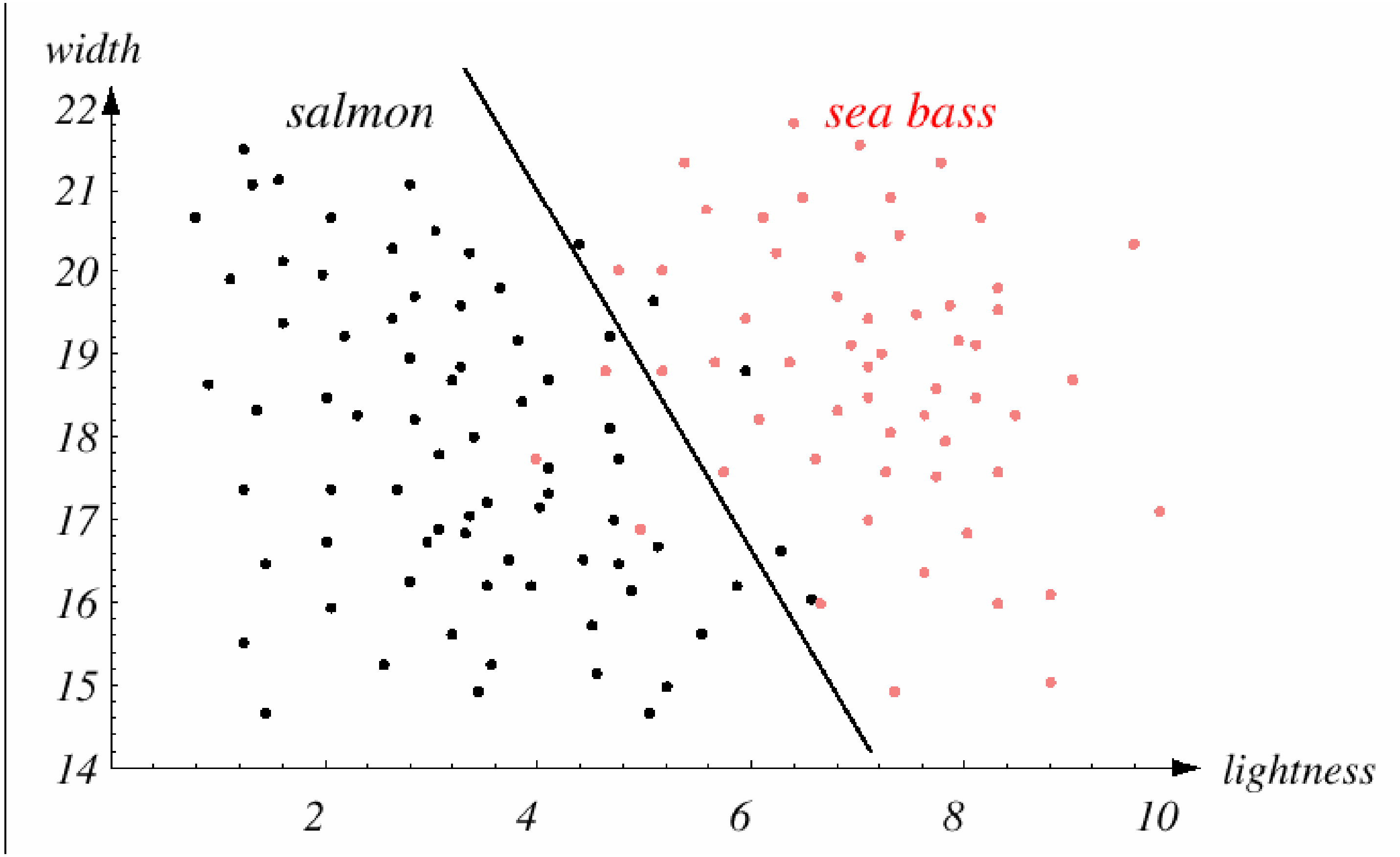
An Example

- Our task is to find the decision rule (or to set the decision boundary) that can minimize an overall cost. decision theory

- To improve the performance, try to use more features simultaneously.
  - e.g., sea bass are typically wider than salmon
  - choose lightness and width as features

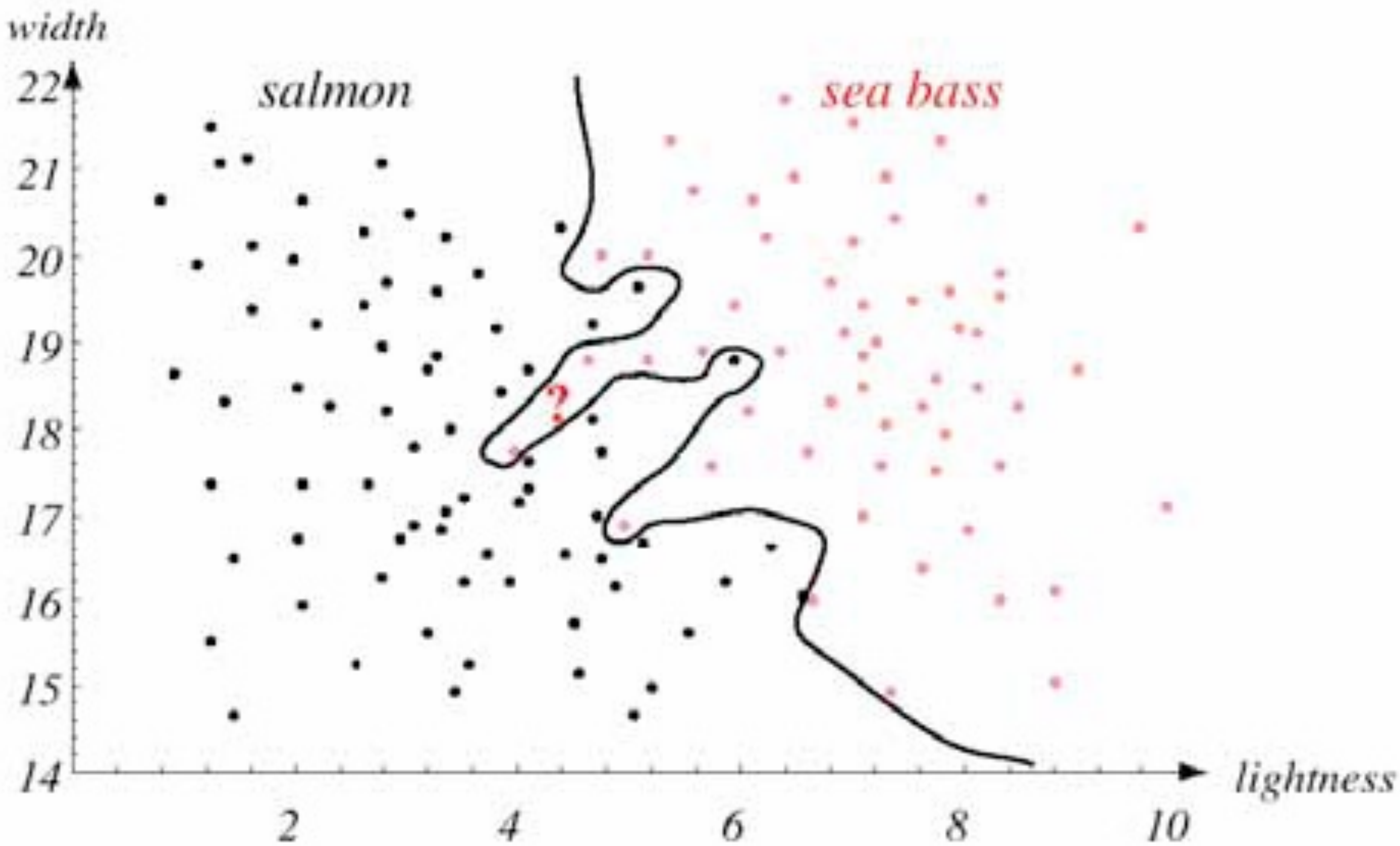
An Example

- Two-dimensional feature space



Generalization

- Our goal is to design a classifier to suggest actions when presented with *novel* patterns, i.e., fish not yet seen.



An Example

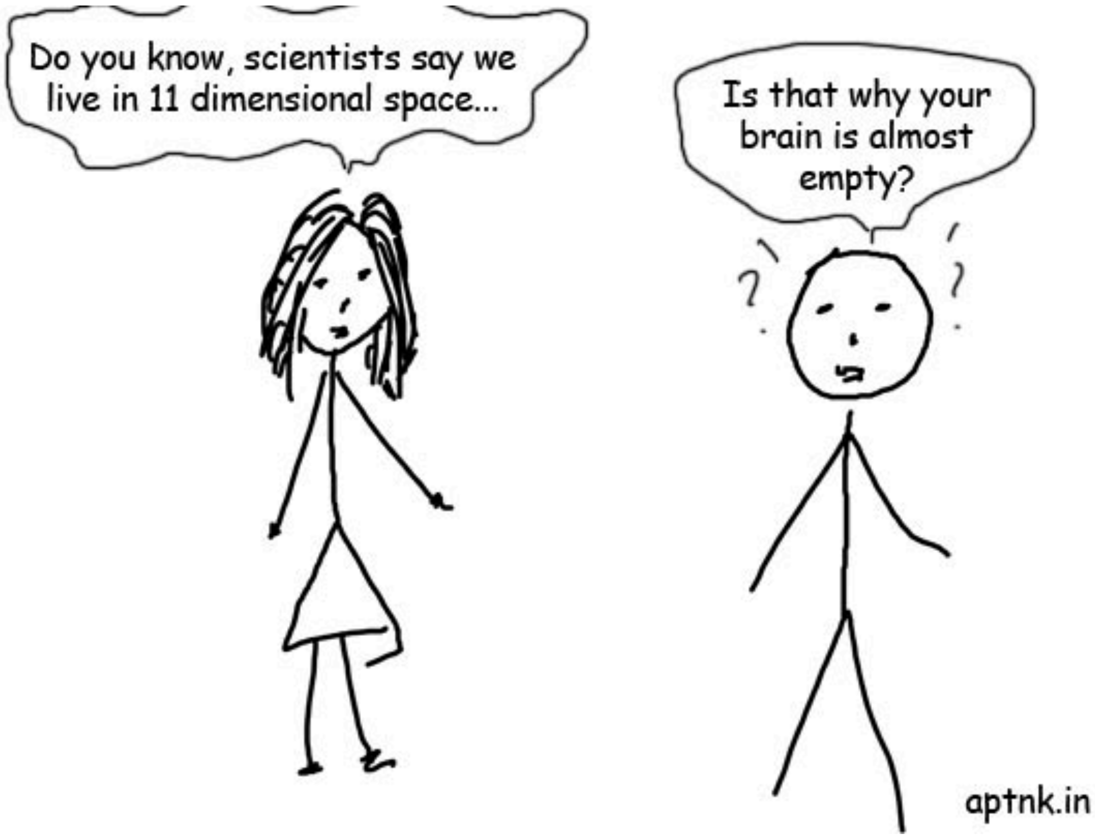
- To improve the performance, try to use more features simultaneously.
  - e.g., sea bass are typically wider than salmon
  - choose lightness and width as features

Does it more features always lead to better results?

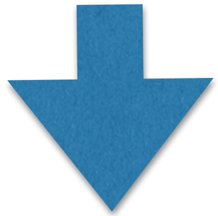
No!!

"Curse" of dimensionality

Curse of dimensionality



100 x 100



10,000-dimensional vector

Dimension reduction

PR Systems

- A typical PR system consists of the following five components:

  - Sensing (e.g., cameras, microphones.)
  - Preprocessing (Segmentation and Grouping)
  - Feature Extraction
  - Classification
  - Post Processing

The Design Cycle

- Collect Data
- Choose Features
- Choose Model
- Train Classifier
- Evaluate Classifier

How deep learning helps for PR:
- Feature
- Model
- Classifier

Bayesian Decision Theory

- Maximum *a posteriori* probability (M.A.P.) classifier is the minimum-error-rate classifier

The Bayes rule:

$$P(\omega_j \mid x) = P(x \mid \omega_j)\, P(\omega_j) \,/\, P(x)$$

$$\text{Posterior} = \frac{\text{Likelihood} * \text{Prior}}{\text{Evidence}}$$

Learning

- Supervised Learning
  - Parametric Approach
  - Non-Parametric Approach
    - Parzen Windows, k-NN Estimation

- Unsupervised Learning *(clustering)*
  - Parametric Approach
  - Non-Parametric Approach
    - Clustering

- Reinforcement Learning (learning with a critic)

Foreground Detection

Foreground Detection

- Foreground detection (also known as **background subtraction**) is to detect changes in image sequences.
- All detection techniques are based on modelling the background of the image, i.e. set the background and detect which changes occur.
- Defining the background can be very difficult when it contains shapes, shadows, and moving objects.

Foreground Detection

- The rationale in the approach is that of detecting the moving objects from the difference between the current frame and a reference frame, often called “background image”, or “background model”.

- A reliable and robust background subtraction algorithm should handle:

  - Sudden or gradual illumination changes
  - High frequency, repetitive motion in the background (such as tree leaves, flags, waves, . . .)

  - Long-term scene changes (a car is parked for a month).

Purpose of Background Subtraction

- Segment the image into foreground and background
- Reduce problem set for further processing
  - Only process part of picture that contains the relevant information
- Add a virtual background

Encountered Problems

- Lighting
  - Shadows
  - Gradual/Sudden illumination changes

- Camouflage

- Moving objects
  - Foreground aperture

- Foreground object becomes motionless

Encountered Problems

- Lighting
    - Shadows
    - Gradual/Sudden illumination changes

Simple Approach

Image at time t:
$I(x, y, t)$
$\Downarrow$

Background at time t:
$B(x, y, t)$
$\Downarrow$

$|$

$-$

$| > Th$

1. Estimate the background for time t.
2. Subtract the estimated background from the input frame.
3. Apply a threshold, Th, to the absolute difference to get the **foreground mask**.

Problem: **how to have a clear background?**

Frame Difference

- Background is estimated to be the previous frame. Background subtraction equation then becomes:

$$B(x, y, t) = I(x, y, t-1)$$
$$\Downarrow$$
$$|I(x, y, t) - I(x, y, t-1)| > Th$$

- Depending on the object structure, speed, frame rate and global threshold, this approach may or may **not** be useful (usually **not**).

$|$

$-$

$| > Th$

Frame Difference

$Th = 25$

$Th = 50$

$Th = 100$

$Th = 200$

Mean Filter

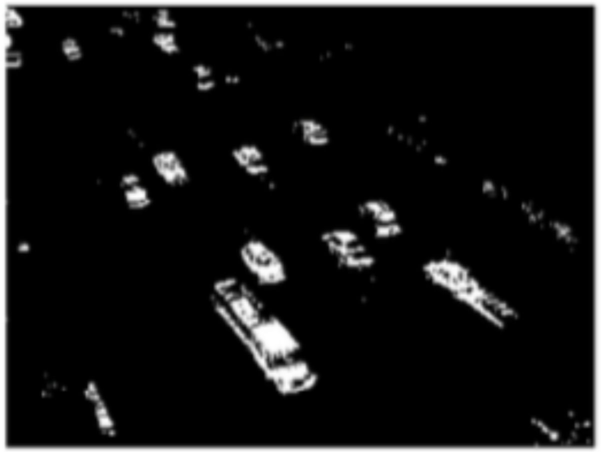
- In this case the background is the mean of the previous n frames:

$$B(x,y,t) = \frac{1}{n}\sum_{i=0}^{n-1} I(x,y,t-i)$$
$$\Downarrow$$
$$|I(x,y,t) - \frac{1}{n}\sum_{i=0}^{n-1} I(x,y,t-i)| > Th$$

- For $n = 10$:

Estimated Background

Foreground Mask



Mean Filter

- For $n = 20$:

Estimated Background

Foreground Mask

- For $n = 50$:

Estimated Background

Foreground Mask

Median Filter

- Assuming that the background is more likely to appear in a scene, we can use the median of the previous n frames as the background model:

$$B(x, y, t) = median\{I(x, y, t - i)\}$$
$$\Downarrow$$
$$|I(x, y, t) - median\{I(x, y, t - i)\}| > Th \text{ where}$$
$$i \in \{0, \ldots, n - 1\}.$$

- For $n = 10$:

Estimated Background

Foreground Mask

Median Filter

- For $n = 20$:

Estimated Background

Foreground Mask

- For $n = 50$:

Estimated Background

Foreground Mask

Advantages & Shortcomings

Advantages:

- Extremely easy to implement and use!
- All pretty fast.
- Corresponding background models are not constant, they change over time.

Disadvantages:

- Accuracy of frame differencing depends on object speed and frame rate!
- Mean and median background models have relatively high memory requirements.
  - In case of the mean background model, this can be handled by a **running average**:

$$B(x, y, t) = \frac{t-1}{t} B(x, y, t-1) + \frac{1}{t} I(x, y, t)$$

or more generally:

$$B(x, y, t) = (1-\alpha) B(x, y, t-1) + \alpha I(x, y, t)$$

where α is the learning rate.

Advantages & Shortcomings

Disadvantages:

- There is another major problem with these simple approaches:

$$|I(x,y,t) - B(x,y,t)| > Th$$

  - There is one global threshold, *Th*, for all pixels in the image.
  - And even a bigger problem: **this threshold is not a function of *t***.

- So, these approaches will not give good results in the following conditions:
  - if the background is bimodal,
  - if the scene contains many, slowly moving objects (mean & median),
  - if the objects are fast and frame rate is slow (frame differencing),
  - and if general lighting conditions in the scene change with time!

Advantages & Shortcomings

To represent the background by single value could be unstable especially for non-stationary scenes.

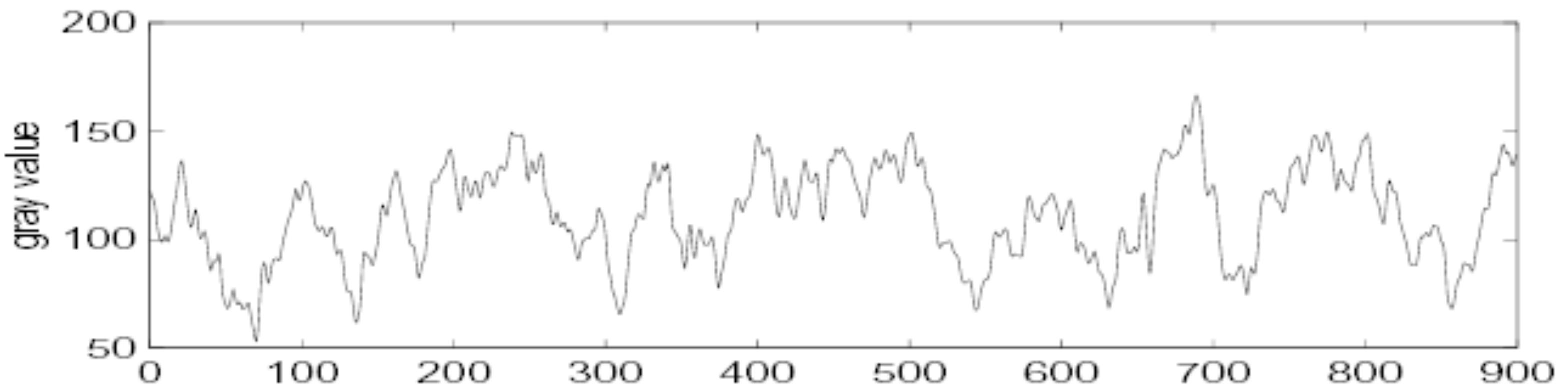


Figure 1: Intensity value overtime

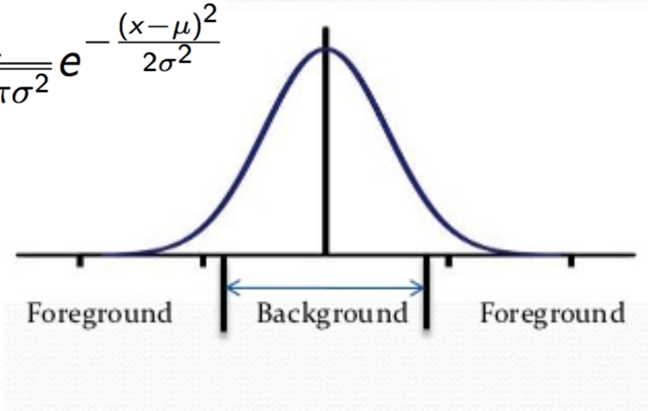
Simple background model

Single Gaussian Model: (parametric model)

- C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland. Pfinder: Real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):780–785, July 1997.

The background scene is modelled by representing each pixel by a Gaussian model

$$\mathcal{N}(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



The mean (ϕ) and covariance U(ϕ) of each pixel ϕ can be recursively updated as follows

$$\boldsymbol{\mu}^t(\phi) = (1-\alpha)\boldsymbol{\mu}^{t-1}(\phi) + \alpha I^t(\phi),$$

$$U^t(\phi) = (1-\alpha)U^{t-1}(\phi) + \alpha\nu(\phi)\nu(\phi)^T$$

α is the learning rate and $\nu(\phi) = (I^t(\phi) - \boldsymbol{\mu}^t(\phi))$

Simple background model

Single Gaussian Model: (parametric model)

- C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland. Pfinder: Real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):780–785, July 1997.

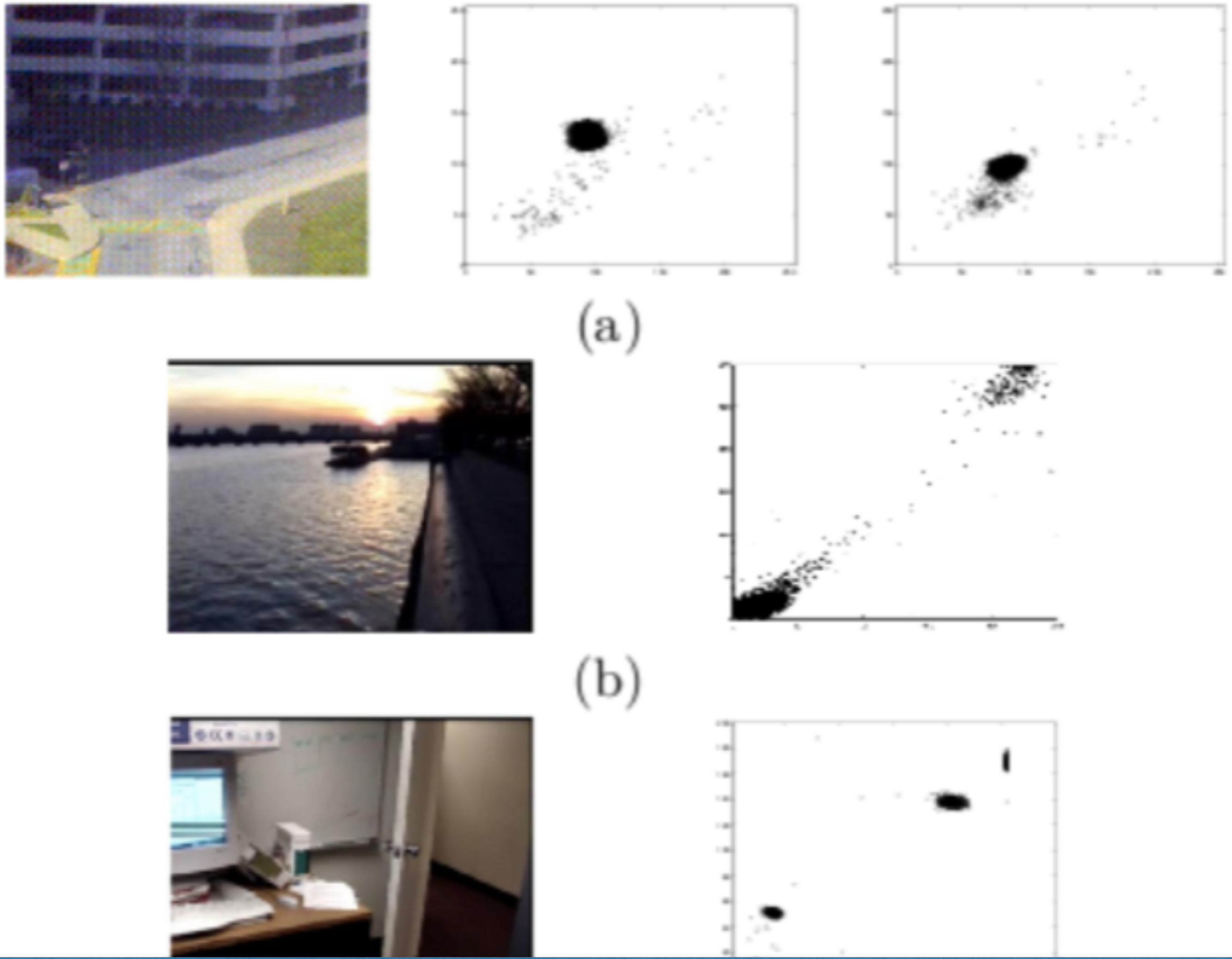
Advantages:

- Easy and fast!

Disadvantages:

- May not be sufficient for representation since some scenes and backgrounds are often dynamic.

Background modelling



A robust background subtraction algorithm should handle: lighting changes, repetitive motions from clutter and long-term scene changes.

Mixture-of-Gaussian

C. Stauffer and W.E.L. Grimson, "Adaptive Background Mixture Models for Real-Time Tracking," CVPR, 1998. (cited number: 10537 from Google)

C. Stauffer and W.E.L. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):747–757, August 2000. (cited number: 4866 from Google)

Algorithm Overview

- The values of a particular pixel is modelled as a mixture of adaptive Gaussians.
  - Why mixture? Multiple surfaces appear in a pixel.
  - Why adaptive? Lighting conditions change.

- At each iteration Gaussians are evaluated using a simple heuristic to determine which ones are mostly likely to correspond to the background.

- Pixels that do not match with the “background Gaussians” are classified as foreground.

- Foreground pixels are grouped using 2D connected component analysis.

If a pixel is categorized as foreground for a too long period of time, the background intensity in that location might have changed (because illumination has changed etc.). As a result, once the foreground object is gone, the new background intensity might not be recognized as such anymore.

Gaussian Mixture Model

- At any time t, what is known about a particular pixel, (x_0, y_0), is its history:

$$\{X_1, \ldots, X_t\} = \{I(x_0, y_0, i) : 1 \leq i \leq t\}$$

- This history is modelled by a mixture of K Gaussian distributions:

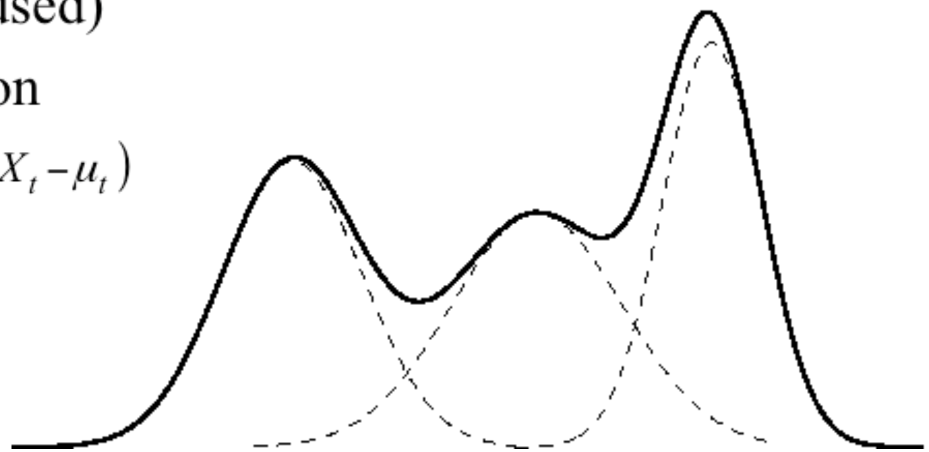
$$P(X_t) = \sum_{i=1}^{K} \omega_{i,t} \times G(X_t, \mu_{i,t}, \Sigma_{i,t})$$

e.g. R, G, B

$$\Sigma_{i,t} = \begin{bmatrix} \sigma_k^2 & 0 & 0 \\ 0 & \sigma_k^2 & 0 \\ 0 & 0 & \sigma_k^2 \end{bmatrix}$$

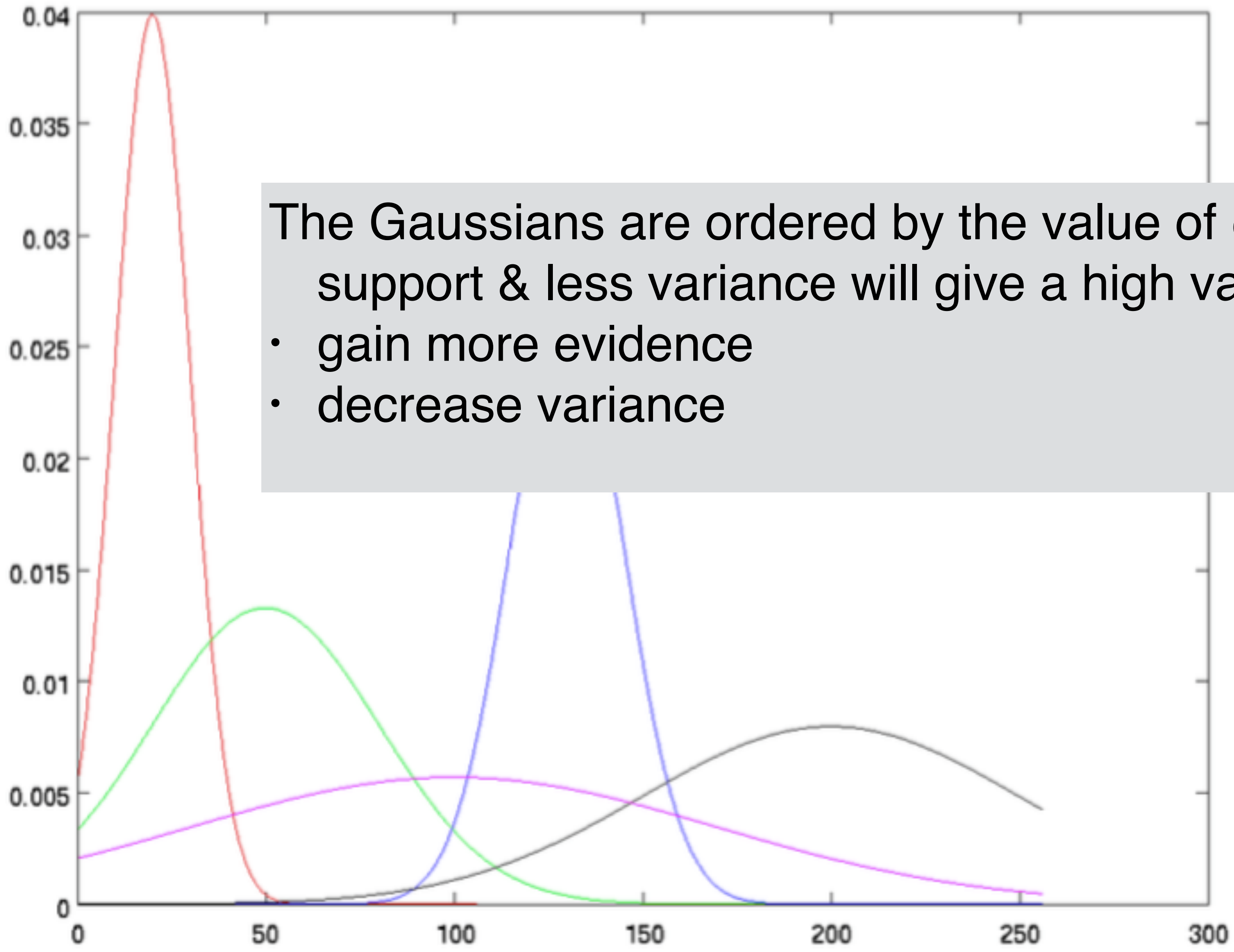
- $\mu_{i,t}$ is the mean value
- $\Sigma_{i,t} = \sigma_k^2 \mathbf{I}$ is the covariance matrix
- K is the number of distributions (3~5 are used)
- G is a Gaussian probability density function

$$G(X_t, \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(X_t - \mu_t)^T \Sigma^{-1} (X_t - \mu_t)}$$



Gaussian Mixture Model

If we assume gray scale images and set K = 5, history of a pixel will be something like this:



The Gaussians are ordered by the value of ω/σ (high support & less variance will give a high value)

- gain more evidence
- decrease variance

Model Update

An on-line K-means approximation is used to update the Gaussians.

If a new pixel value, X_{t+1}, can be **matched** to one of the existing Gaussians (within 2.5σ), that Gaussian's $\mu_{i,t+1}$ and $\sigma_{i,t+1}^2$ are updated as follows:

$$\mu_{i,t+1} = (1-\rho)\mu_{i,t} + \rho X_{t+1}$$

$$\sigma^2_{i,t+1} = (1-\rho)\sigma^2_{i,t} + \rho(X_{t+1} - \mu_{i,t+1})^2$$

where $\rho = \alpha \mathcal{N}(X_{t+1} | \mu_{i,t}, \sigma^2_{i,t})$ and α is a learning rate.

Prior weights of all Gaussians are adjusted as follows:

$$\omega_{i,t+1} = (1-\alpha)\omega_{i,t} + \alpha(M_{i,t+1})$$

where $M_{i,t+1} = 1$ for the matching Gaussian and $M_{i,t+1} = 0$ for all the others.

Model Update

An on-line K-means approximation is used to update the Gaussians.

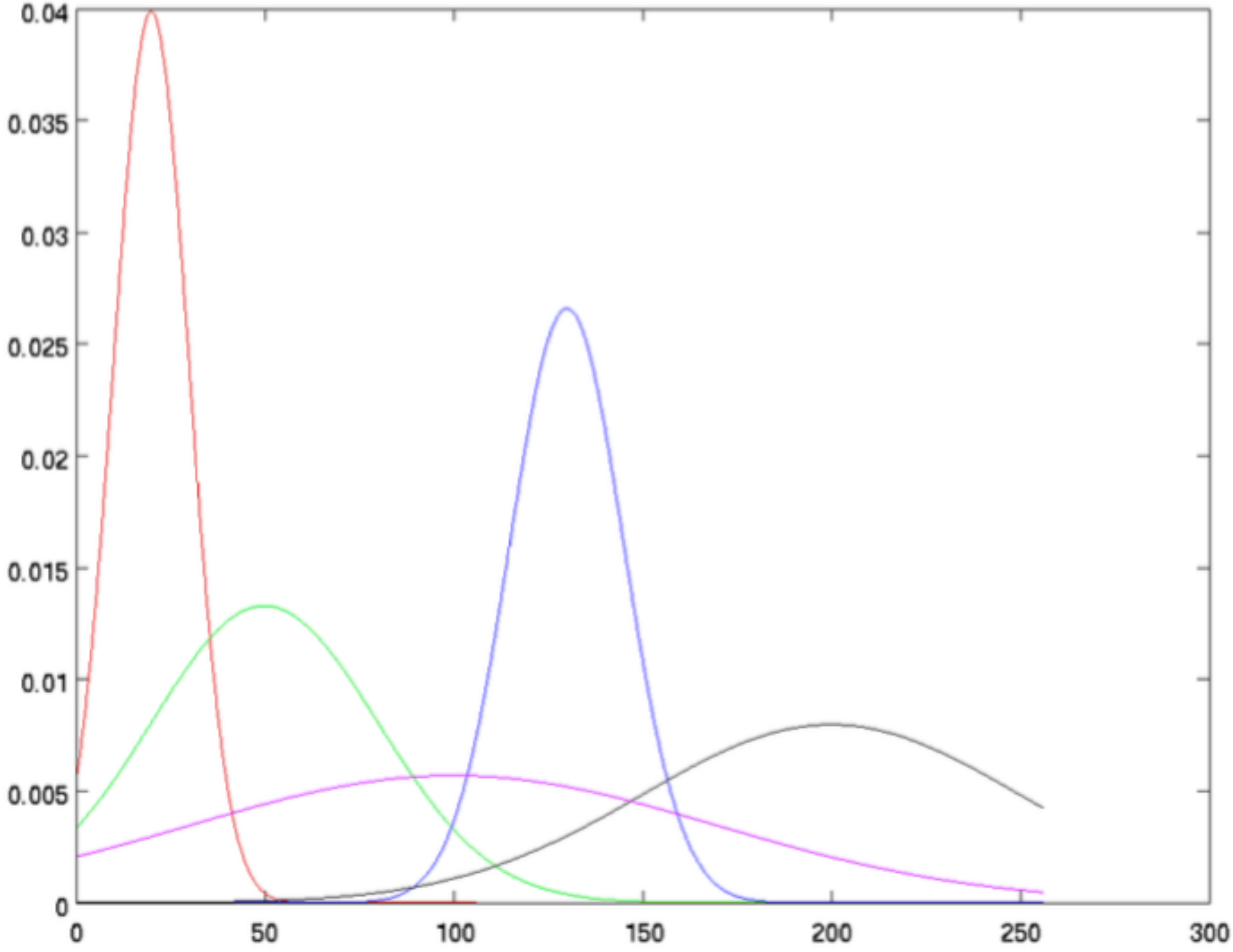
If X_{t+1} **do not match** to any of the *K* existing Gaussians, the least probably distribution is replaced with a new one.

The Gaussians are ordered by the value of ω/σ (high support & less variance will give a high value)

- gain more evidence
- decrease variance

New distribution has $\mu_{t+1} = X_{t+1}$, a high variance and a low prior weight.

Model Update



Background Model Estimation

- Heuristic: the Gaussians with the most supporting evidence and least variance should correspond to the background.

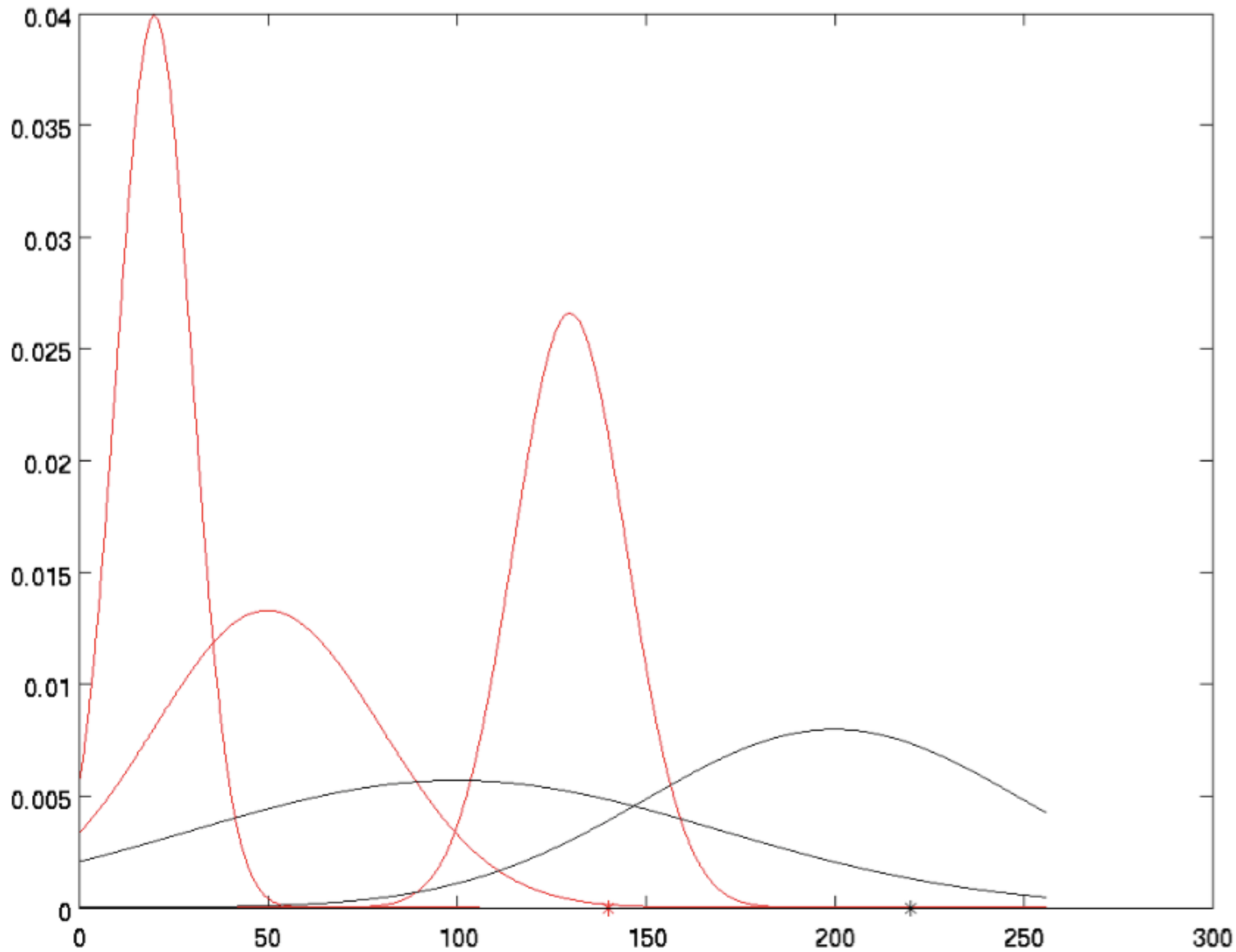
- The Gaussians are ordered by the value of ω/σ (high support & less variance will give a high value).

- Then simply the first B distributions are chosen as the background model:

$$B = argmin_b\left(\sum_{i=1}^{b} \omega_i > T\right)$$

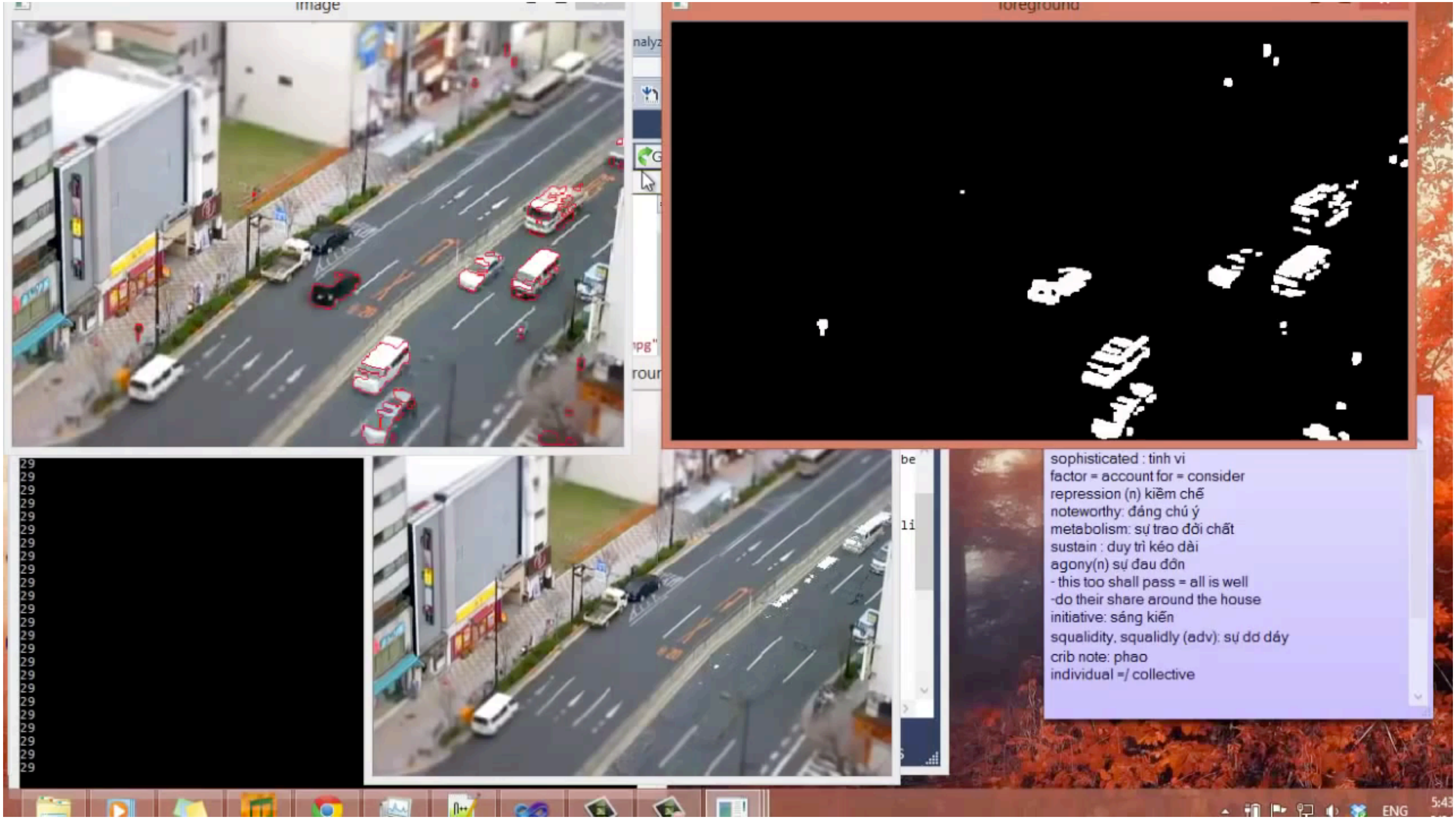
where T is minimum portion of the image which is expected to be background.

Background Model Estimation



After background model estimation **red** distributions become the background model and **black** distributions are considered to be foreground.

OpenCV MoG


sophisticated : tinh vi
factor = account for = consider
repression (n) kiềm chế
noteworthy: đáng chú ý
metabolism: sự trao đổi chất
sustain : duy trì kéo dài
agony(n) sự đau đớn
- this too shall pass = all is well
-do their share around the house
initiative: sáng kiến
squalidity, squalidly (adv): sự dơ dáy
crib note: phao
individual =/ collective
ENG

OpenCV MoG

Advantages & Shortcomings

Advantages:

- A different “threshold” is selected for each pixel.
- These pixel-wise “thresholds” are adapting by time.
- Objects are allowed to become part of the background without destroying the existing background model.
- Provides fast recovery.

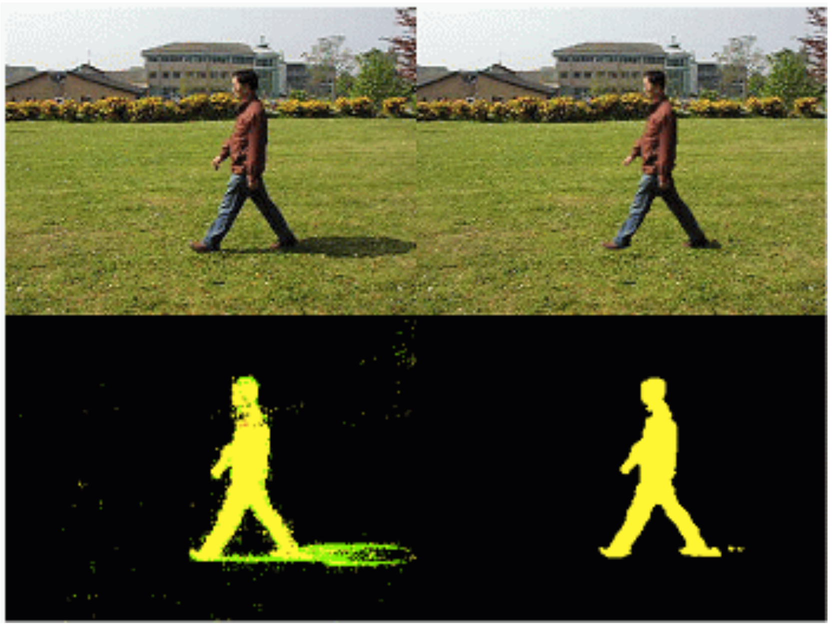
Disadvantages:

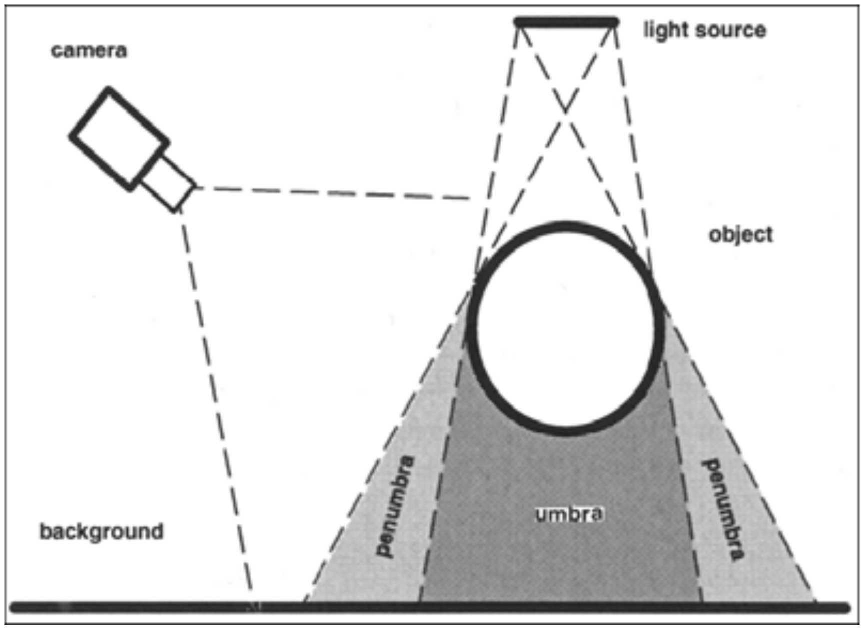
- Cannot deal with sudden, drastic lighting changes!
- Initializing the Gaussians is important (median filtering).
- There are relatively many parameters, and they should be selected intelligently.

Moving Cast Shadow Removal

Moving Cast Shadow

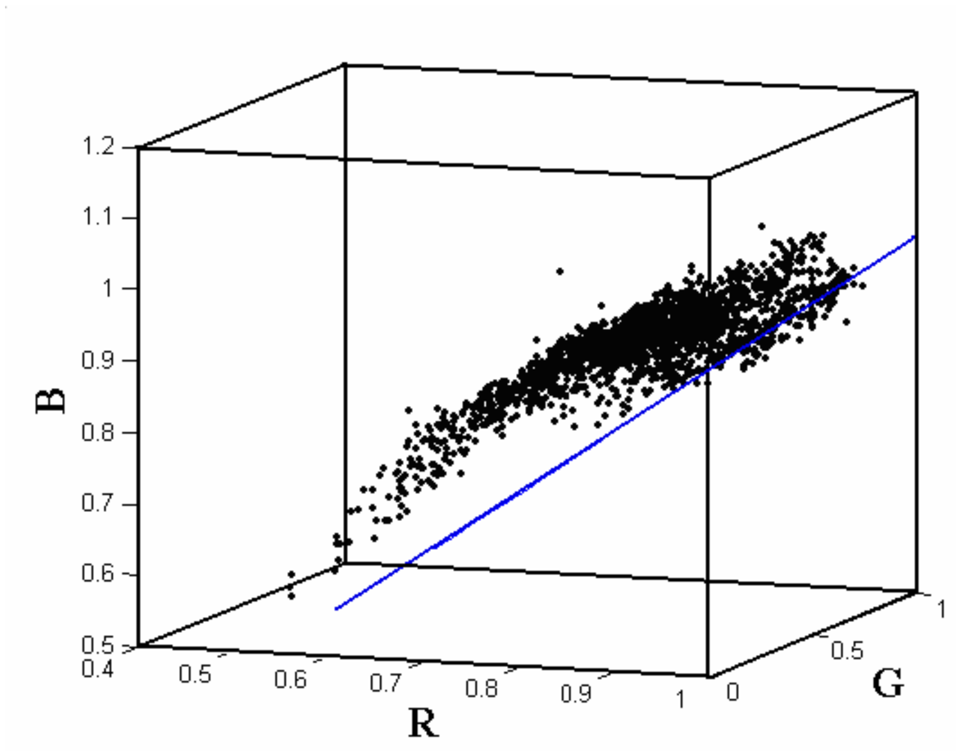
- Shadow is formed by the change of illumination conditions and shadow detection comes down to a problem of finding the illumination invariant features.
- From the viewpoint of geometric relationship, shadow can be divided into umbra and penumbra
  - the umbra corresponds to the background area where the direct light is almost totally blocked by the foreground object,
  - in the penumbra area, the light is partially blocked





Moving Cast Shadow

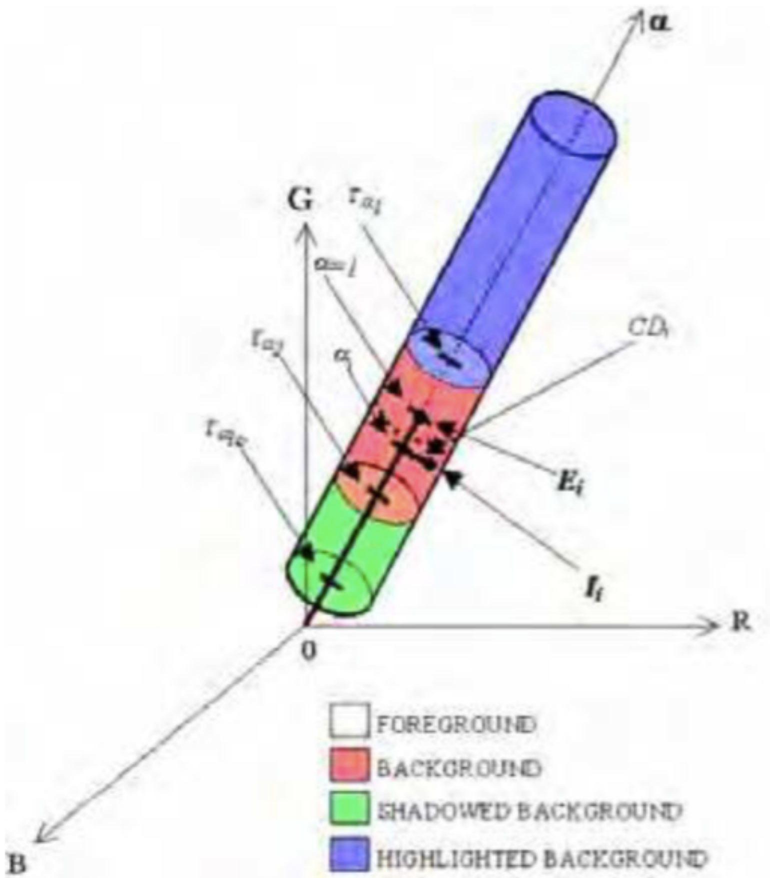
- **Color/Specturm-based shadow detection**
  - shadows change the hue component slightly
  - decrease the saturation component significantly.



A scatter plot in the color ratios space of a shaded pixels set. The line corresponds to the equal ratio in RGB components.

Moving Cast Shadow

- **Color/Specturm-based shadow detection**
  - shadows change the hue component slightly
  - decrease the saturation component significantly.



Pixels classification using the normalized color distortion and normalized brightness distortion: original background, shaded background, highlight background, and moving foreground objects.

Pattern Recognition

Kuan-Wen Chen

2022/3/10