Background Subtraction in Video using Bayesian Learning with Motion Information

Suman K. Mitra

DA-IICT, Gandhinagar

suman_mitra@daiict.ac.in

Bayesian Learning

Given a model and some observations, the prior distribution of the parameters of the model is updated to a posterior distribution.

$$p(\theta \mid x) = \frac{l(\theta; x)\, p(\theta)}{\int l(\theta; x)\, p(\theta)\, d\theta}$$

Evaluation of posterior distribution, except in very simple cases, requires

- Sophisticated numerical integration
- Analytical approximation

The problem of relating prior distribution to the posterior via likelihood function has been addressed by Smith and Gelfand [American Statistician, 1992] from a sampling re-sampling perspective.

A. Smith and A. Gelfand, Bayesian statistics without tears:A sampling re-sampling perspective, *The American Statisticians*, 42, 1992.

Sampling Re-sampling Technique

1. Obtain a sample of observations { $\theta_1,\theta_2,\dots\dots,\theta_n$ } from a starting prior distribution

2. Compute weights q_i for each sample θ_i

$$q_i = \frac{l(\theta_i;x)}{\sum_{j=1}^{n} l(\theta_j;x)}$$

3. Resample { $\theta_1,\theta_2,\dots\dots,\theta_n$ } as { $\theta_1^*,\theta_2^*,\dots\dots,\theta_n^*$ }by placing mass q_i on θ_i

4. Repeat Steps 2 and 3 for a sufficiently large number of times. At the end, the sample { $\theta_1^*,\theta_2^*,\dots\dots,\theta_n^*$ } leads to the required posterior di

Detection and Tracking

- Many models exist for intelligent object detection and tracking from video. But all assume high contrast between background and object.
  - 'Pfinder' uses statistical model (single Gaussian) per pixel.
  - Ridder et al. : Model each pixel as a Kalman Filter.
  - Stauffer et al.: Gaussian Mixture Model (GMM), with online k-means approximation.
  - Davies et al.: Small objects in low contrast conditions using Kalman Filtering.

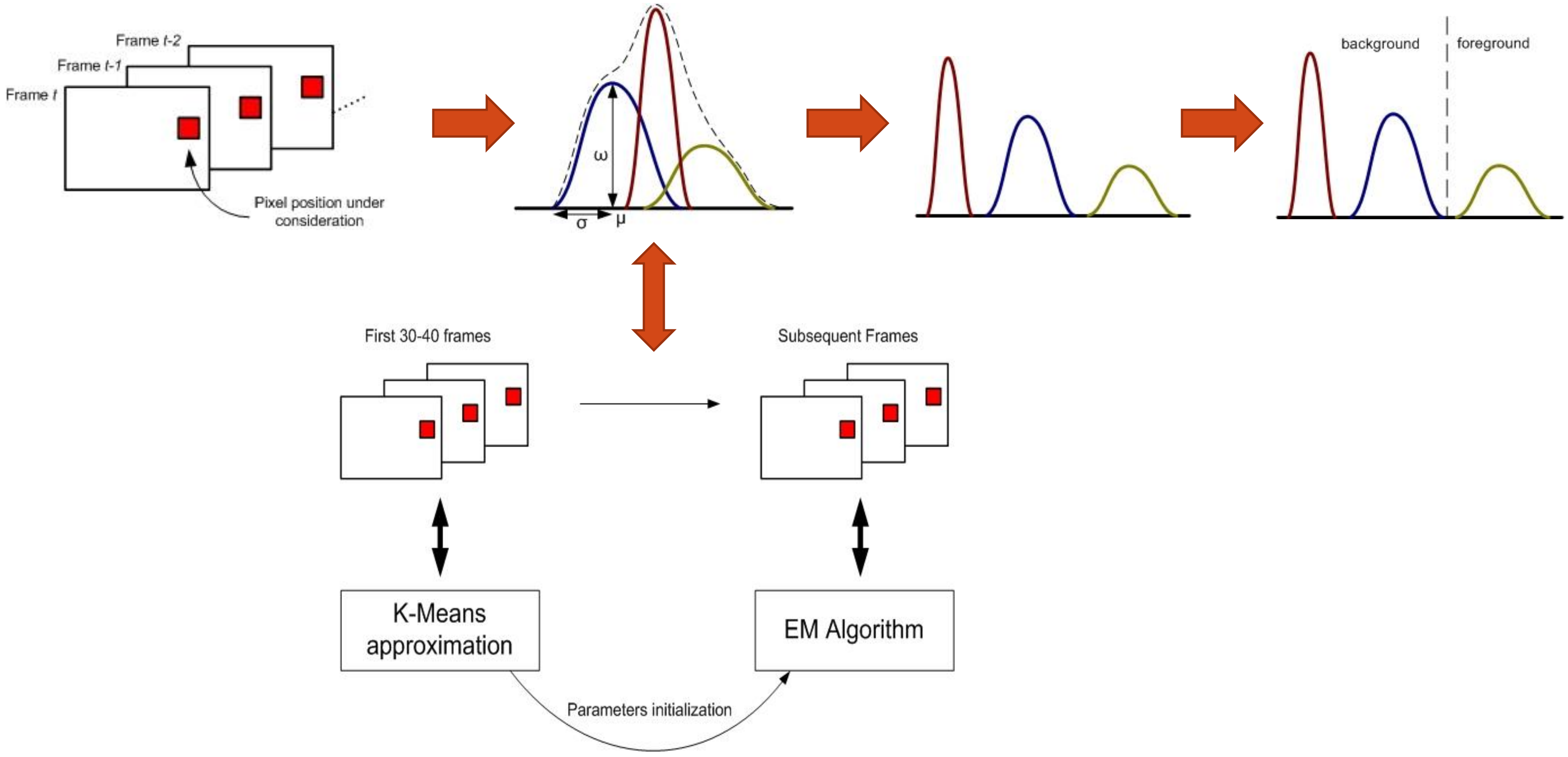
[3] effectively deals with problems of lighting variations and multimodal backgrounds. It however fails to detect low contrast objects.

[4] fails to address multimodal backgrounds and lighting variations – focus is mostly on small object detection.

- Applications such as followings require low contrast detection technique
  - Detection of camouflaged objects
  - Tracking of balls in sports events

1. Wren C., Azarbayejani A., Darrell T. and Pentland A., Pfinder:Real time tracking of the human body, IEEE PAMI, 19, 1997.
2. Ridder C., Munkelt O. and Kirchner H., Adaptive background estimation and foreground detection using Kalman filter, Proceedings ICRAM, 1995
3. Stauffer C. and Grimson W., Adaptive background mixture model for real time tracking, Proceedings IEEE conference on CVPR, 1999.
4. Davies D., Palmer P. and Mirmehdi M., Detection and tracking of very small low-contrast objects, BMVC, 1998

Initial Approach (using GMM)



Experimental Results

Original frames (left column), frames segmented using k-means approximation on GMM [Stauffer et al.] (center column), frames segmented using our approach (right column). It can be seen that our approach works well for low contrast portions of moving bodies.

Stauffer C. and Grimson W., Adaptive background mixture model for real time tracking, Proceedings IEEE conference on CVPR, 1999.

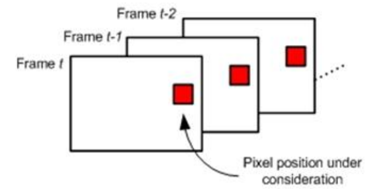
Where we stand?

- Low Contrast Object Detection and Tracking successfully addressed!
- Experiments have shown good results. However yielding a little high false alarm.
- No clue on the selection of number of Gaussian.
- Computationally (time) expensive.

Bayesian Learning Approach

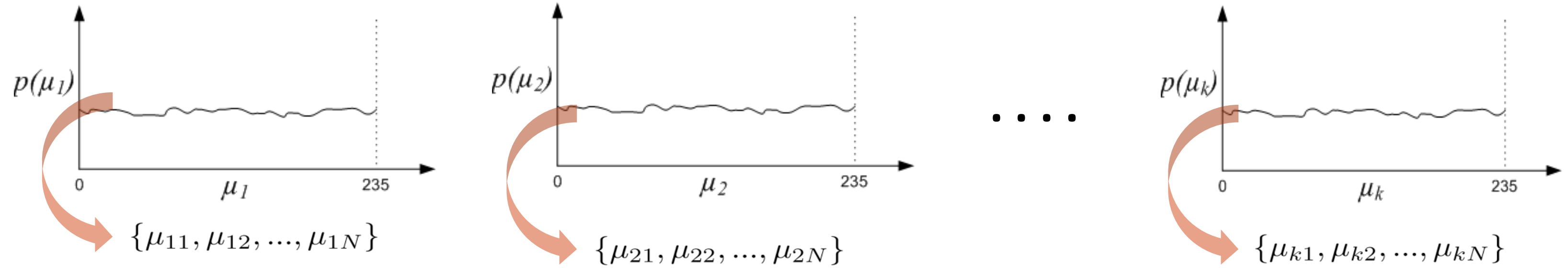
At every pixel position:

A pixel process
(observations coming in one by one)



Steps for Bayesian Learning

Step 1 We draw N samples each, from the distributions of the means



Step 2 When an observation is made, we compute the sum of likelihoods for all samples, from each cluster.

$$L_r = \sum_{i=1}^{N} l(\mu_{ri}; x) \quad \text{where } r = 1, 2, ..., k.$$

Gaussian distribution with a small variance is assumed for computing likelihoods. Variance of the Gaussian distribution is the **Model Variance**.

Steps for Bayesian Learning

Step 3

Determining the cluster to which the observation belongs:

Distribution having the highest value of L_r
(Maximum likelihood)

Step 4

Updating this prior (existing) distribution of the cluster mean to a posterior one:
(converting prior samples to posterior samples)

1. Compute weights for each sample of the prior distribution as follows:

$$q_i = \frac{l(\mu_{ri}; x)}{L_r}, \quad i = 1, 2, ..., N$$

2. Resample $\{\mu_{r1}, \mu_{r2}, ..., \mu_{rN}\}$r attaching weights $\{q_1, q_2, ..., q_N\}$.

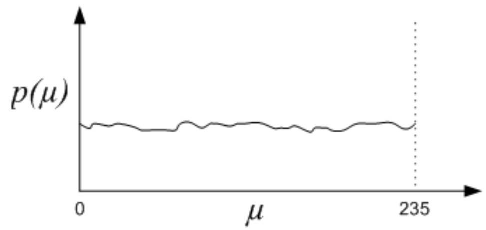
The resultant samples $\{\mu^*_{r1}, \mu^*_{r2}, ..., \mu^*_{rN}\}$he required posterior samples (samples drawn from the posterior distribution)

For every new observation, repeat steps 2 to 4.

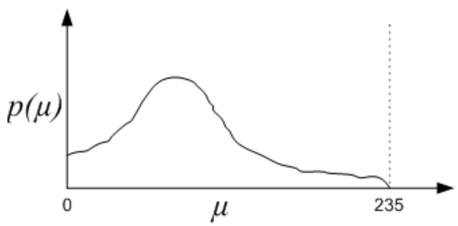
Model Variance

Effect of changing Model variance

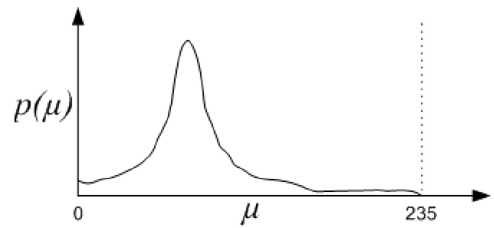
‘Model Variance’ affects likelihood of parameters, hence it affects the weights.



Prior distribution



Posterior distribution with a high ‘Model Variance’



Posterior distribution with a low ‘Model Variance’

Distribution is narrower. Allows for finer clustering.

Good when background-foreground clusters are close (low contrast conditions)

Identifying foreground pixels

Classification of pixels (into background and foreground) is done after 40-50 frames of Bayesian Learning steps. This allows a stable model to be built before classification steps can be used.

Basis of classification

Simple principle – Background clusters would typically account for a much larger number of observations. Prior weight of background clusters would be much higher.

1. Clusters are arranged according to their prior weights.

2. Based on a threshold, certain number of low weight clusters are considered as foreground clusters.

3. Based on the sum of likelihoods value, we can determine which cluster an observation belongs to. If this is a foreground cluster, the current observation belongs to foreground.

Foreground identification is done!

Computational (time) Cost

The entire Bayesian Learning steps need to be carried out for all pixel positions.

Computationally expensive!

- *Typically only a small fraction of the entire frame contains motion at any instant.*
- *It's a waste applying the Bayesian learning steps at all locations*.

Block Matching

- We use a simple block matching technique to get a rough idea of blocks that may have motion in them.
- Information from *Motion Vectors* in MPEG videos can also be used to the same effect.

Much faster processing!

Experimental Results

Results 'seem' to be much better than the previous approach.

<u>Much less False Alarm</u> Rate and <u>faster processing</u> speed.

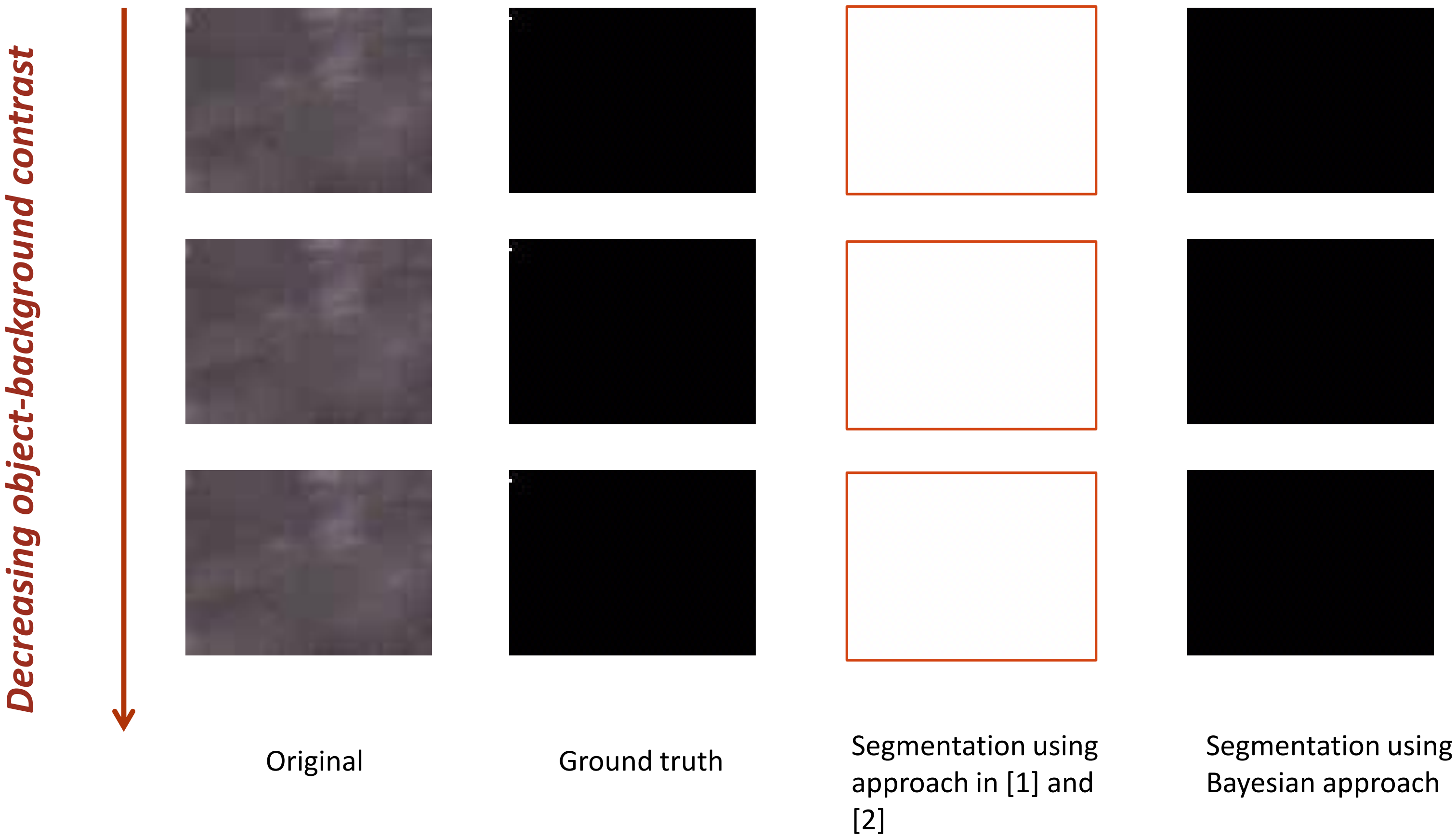
Original low contrast video

Segmentation using our earlier approach in [1] and [2]

Segmentation using the currently proposed technique

1. A. Singh, P. Jaikumar, S.K.Mitra and M.V.Joshi, Low contrast object detection and tracking using gaussian mixture model with split –and-merge operation, International Journal of Image and Graphics, 2008 (Submitted).
2. A. Singh, P. Jaikumar, S.K.Mitra, M.V.Joshi and A. Banerjee. Detection and tracking of objects in low contrast conditions. In Proceedings of NCVPRIPG 2008, pp. 98-103, January 2008.

For some Benchmark videos (Obtained from: *Advanced Computer Vision GmbH – ACV, Austria*)



1. A. Singh, P. Jaikumar, S.K.Mitra and M.V.Joshi, Low contrast object detection and tracking using gaussian mixture model with split –and-merge operation, International Journal of Image and Graphics, 2008 (Submitted).
2. A. Singh, P. Jaikumar, S.K.Mitra, M.V.Joshi and A. Banerjee. Detection and tracking of objects in low contrast conditions. In Proceedings of NCVPRIPG 2008, pp. 98-103, January 2008.

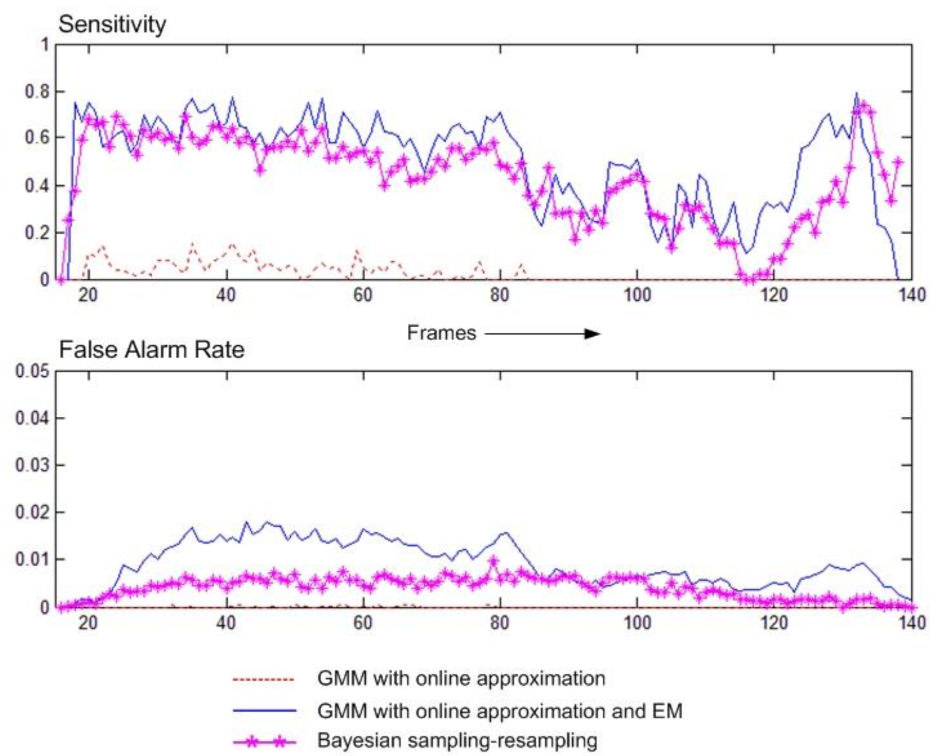
Quantitative analysis

- ***True Positive (TP):*** Number of pixels which are actually foreground and are detected as foreground in the final segmented image.
- ***False Positive (FP):*** Number of pixels which are actually background but are detected as foreground in the final segmented image.
- ***True Negative (TN):*** Number of pixels which are actually background and are detected as background in the final segmented image.
- ***False Negative (FN):*** Number of pixels which are actually foreground but are detected as background in the final segmented image.

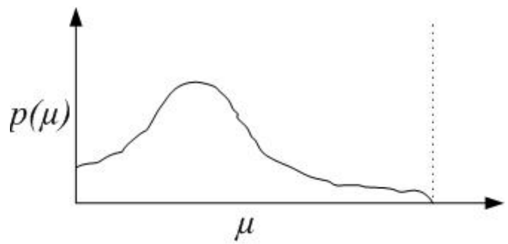
- *Sensitivity (S) = TP/(TP+FN)*

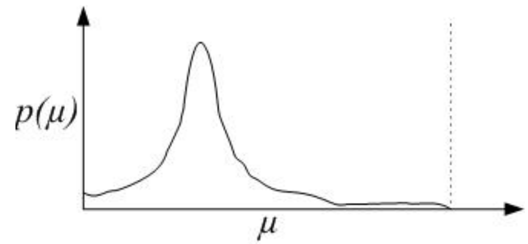
  It is the fraction of the actual foreground detected.

- *False Alarm Rate = FP/(FP+TN)*



Effect of changing *Model Variance*





High *Model Variance* Low *Model Variance*

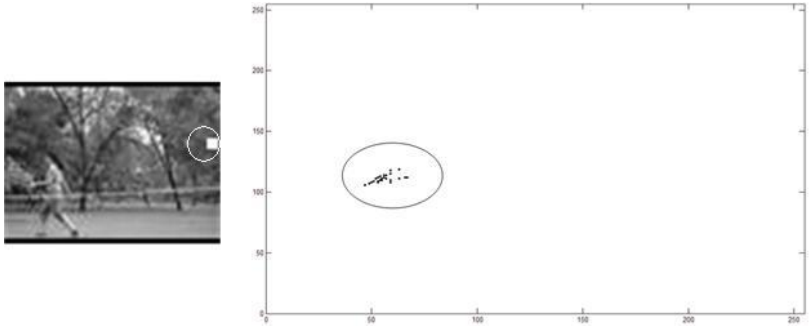
Model Variance can be used as a measure to control sensitivity of the system.

Low *Model Variance* leads to better results in low contrast conditions

Selecting number of clusters

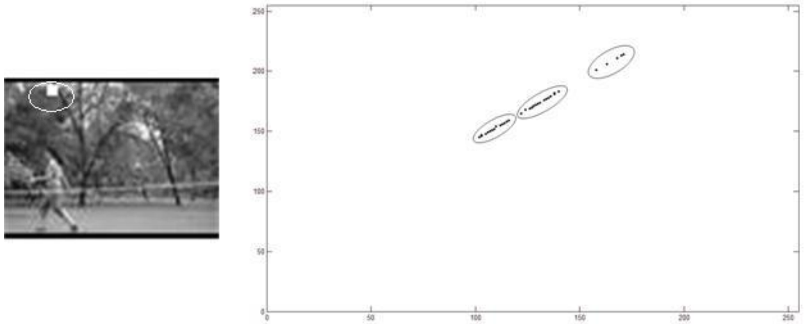
Different number of clusters automatically get formed at different pixel locations.

No need to predefine a fixed number of clusters for each pixel process.



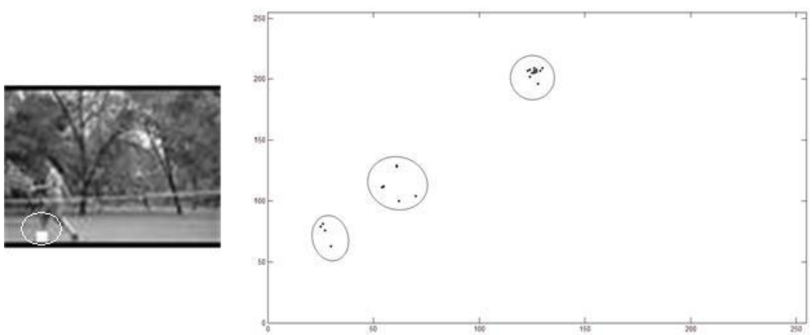
0	0	0	131	0	0

No. of pixel observations in each cluster



0	6	44	0	0	81

No. of pixel observations in each cluster



8	0	118	0	0	5

No. of pixel observations in each cluster

Results of Motion Region Estimation

COMPARISON OF SENSITIVITY, FALSE ALARM RATE AND CPU TIME FOR PROCESSING **Benchmark Video 1**

Approach used	Avg. Sensitivity (%)	Avg. False Alarm Rate (%)	CPU Time for 100 frames (min:sec)
GMM with online approx.	36	0.05	1:25
GMM with online approx. and EM	87	1.8	28:30
Bayesian Learning with Motion Region Estimation	78	0.46	1:55

The values were obtained by implementing the techniques on 128x96 pixel videos, in Matlab 7.2 using a 1.7 Ghz processor. Note that these are time taken for running computer simulations of the techniques, meant for comparative purposes only. Actual speeds on optimized real time systems may vary.

Number of Gaussian still a Problem?

Constraint: number of Gaussians (k) needs to be known beforehand.

Lower value of *k* :

Clustering ability is compromised

Higher value of *k*:

Needless increase in computational cost

Ray and Turi [ICAPRDT 1999]**:**

Clustering is done for all values of k from 2 to K_{max}. The results are checked against some criteria to determine optimum k.

Global k-means algorithm [The Journal of Pattern Recognition Society 2003]**:**

Starts with just one cluster. Keeps increasing the number of clusters until optimum number is reached.

Yang and Zwolinski [IEEE Trans. PAMI 2001]**:**

Start with just K_{max} clusters. Keeps decreasing the number of clusters until optimum number is reached.

Criteria:

- Ratio of intra and inter cluster distance
- Mutual information between two classes

One simple Solution

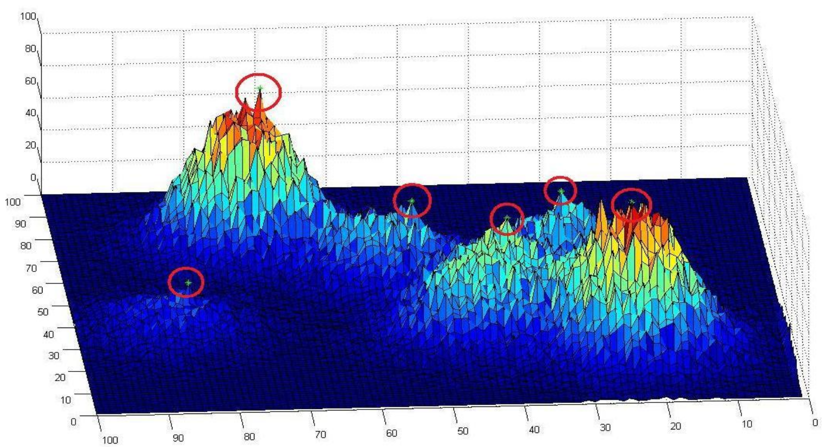
A Simple histogram based technique to obtain an initial estimate of the number of Gaussians.

- Each dimension is divided into N bins. Size of bin for dimension *i:*

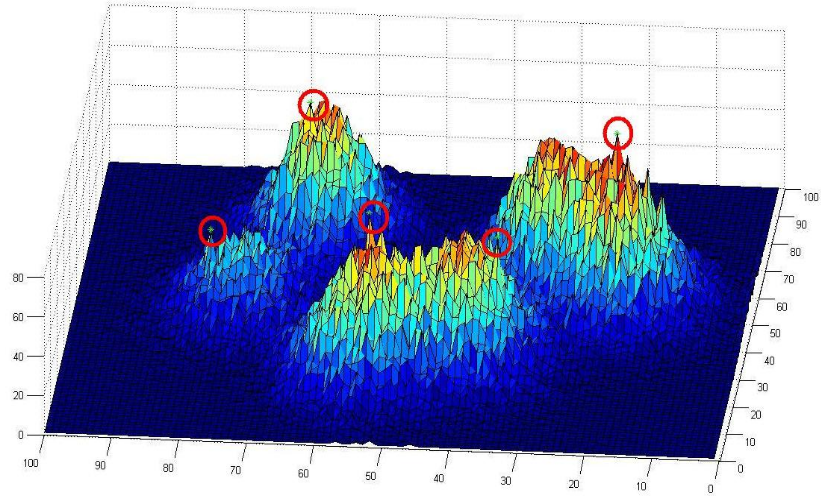
$$n_i = \frac{Max_i - Min_i}{N}$$

- Whole data space is now divided into N^D hypercuboids.
- Count the number of data points in each hypercuboid. Select the hypercuboid having maximum (locally) data points.

IMPORTANT: This would be only a crude guess. The number can be further refined using existing methods as discussed.



Actual number of clusters - 6
Number of clusters detected - 6



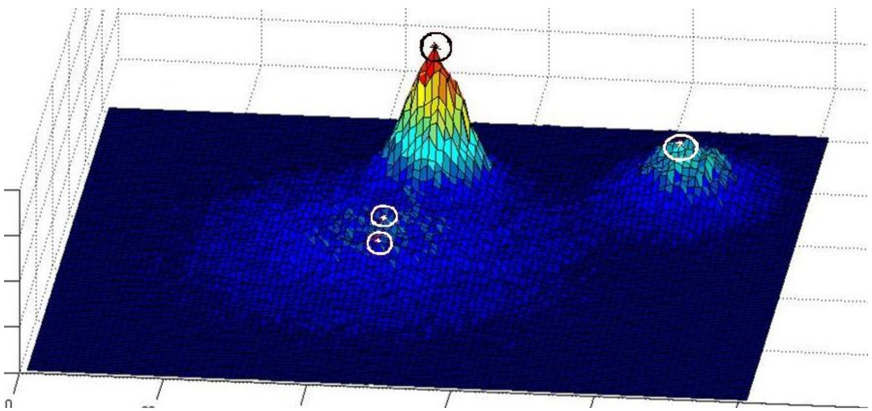
Actual number of clusters - 6
Number of clusters detected - 5

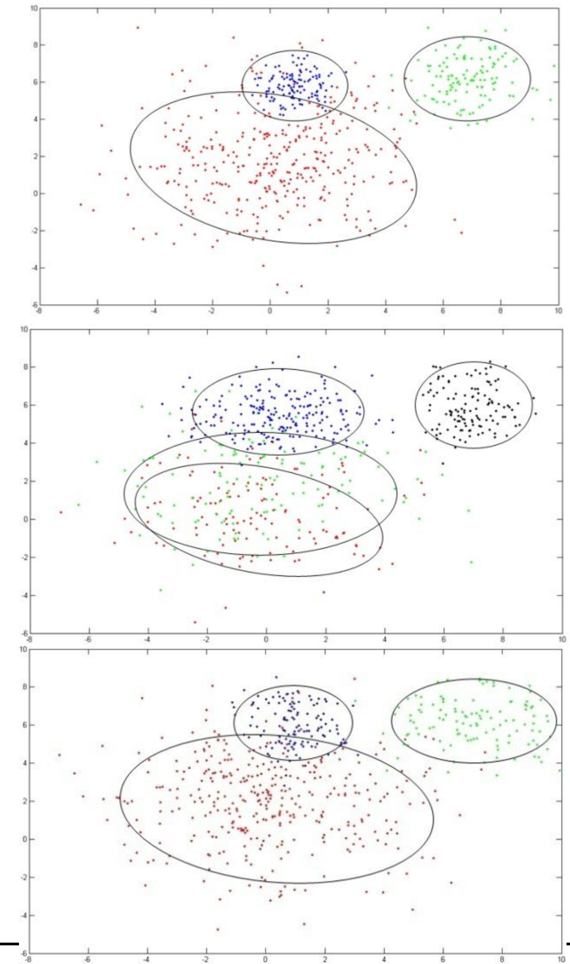
Test data set 2: $0.6N(\mu_1,\Sigma_1)+0.2N(\mu_2,\Sigma_2)+.2N(\mu_3,\Sigma_3)$

where,

$$\mu_1=\begin{bmatrix}0.2\\2\end{bmatrix},\mu_2=\begin{bmatrix}7\\6\end{bmatrix},\mu_3=\begin{bmatrix}1\\6\end{bmatrix}\qquad \Sigma_1=\begin{bmatrix}2.5^2&0\\0&2.5^2\end{bmatrix},\quad \Sigma_2=\begin{bmatrix}1^2&0\\0&1.25^2\end{bmatrix},$$

$$\Sigma_3=\begin{bmatrix}0.7^2&0\\0&0.7^2\end{bmatrix}$$



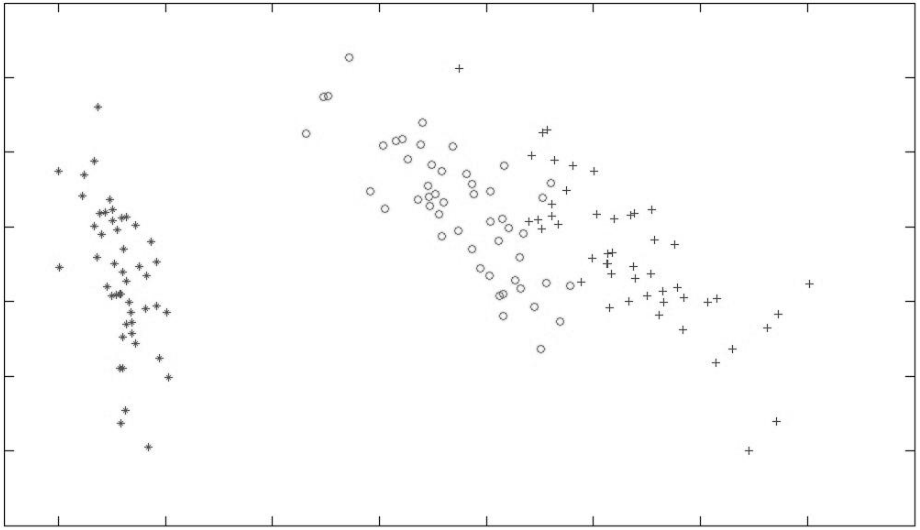


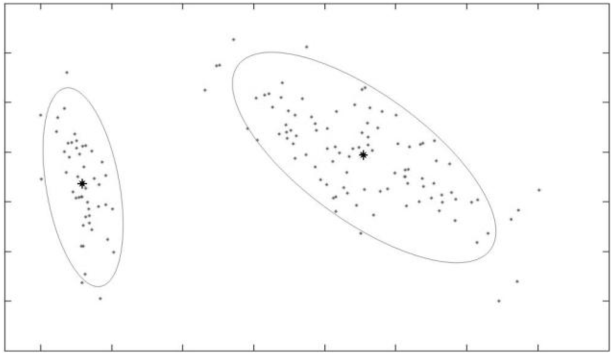
Number of Components	System Mutual Information $I(\Theta)$	Component	Component Mutual Information $I(\mathbf{u}_i,\Theta^{-i})$
4	-0.0203	1	0.0110
		2	0.0073
		3	-0.0596
		4	-0.0177
3*	-0.0570	1	-0.0663
		2	-0.0656
		3	-0.0200

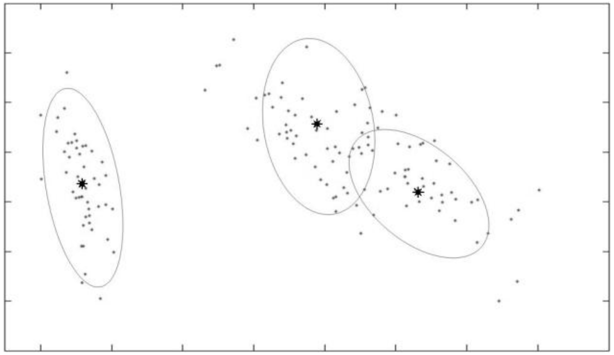
Test data set 3: Iris Data set*

The Iris Data Set is perhaps the best known database to be found in pattern recognition literature. The dataset contains 3 classes of 50 instances each, where each class refers to a type of Iris plant. One class is linearly separable from the other two; the latter are NOT linearly separable from each other.

A. Asuncion and D. Newman, “UCI Machine Learning Repository,” 2007. [online]. Available: http://www/ics.uci.edu/~mlearn/MLRepository.html







Number of Components	System Mutual Information $I(\Theta)$	Component	Component Mutual Information $I(\mathbf{u}_i, \Theta^{-i})$
2	-0.0013	1	-0.0013
		2	-0.0013
3*	-0.0269	1	-0.0004
		2	-0.0404
		3	-0.0404
4	-0.0177	1	0.0002
		2	-0.0214
		3	-0.0137
		4	-0.0343

Bayesian Learning: Other Applications

There are other applications possible if we integrate

- Peak detection technique
- Bayesian learning approach

Applications

- Clustering (unsupervised pattern classification)
- Image segmentation
- Satellite image classification

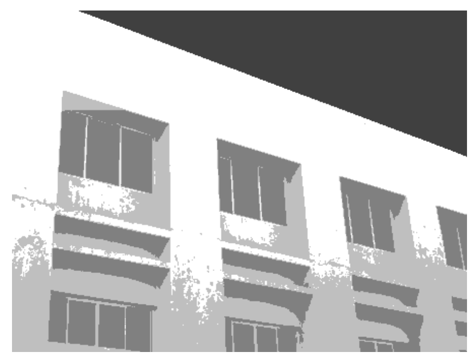
Results of peak detection followed by Bayesian learning to determine the number of clusters in images and the cluster regions

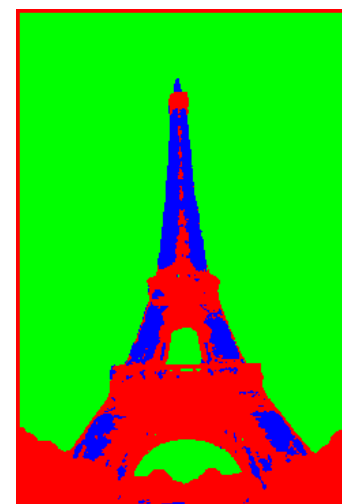
Number of clusters: 5

Number of clusters: 7

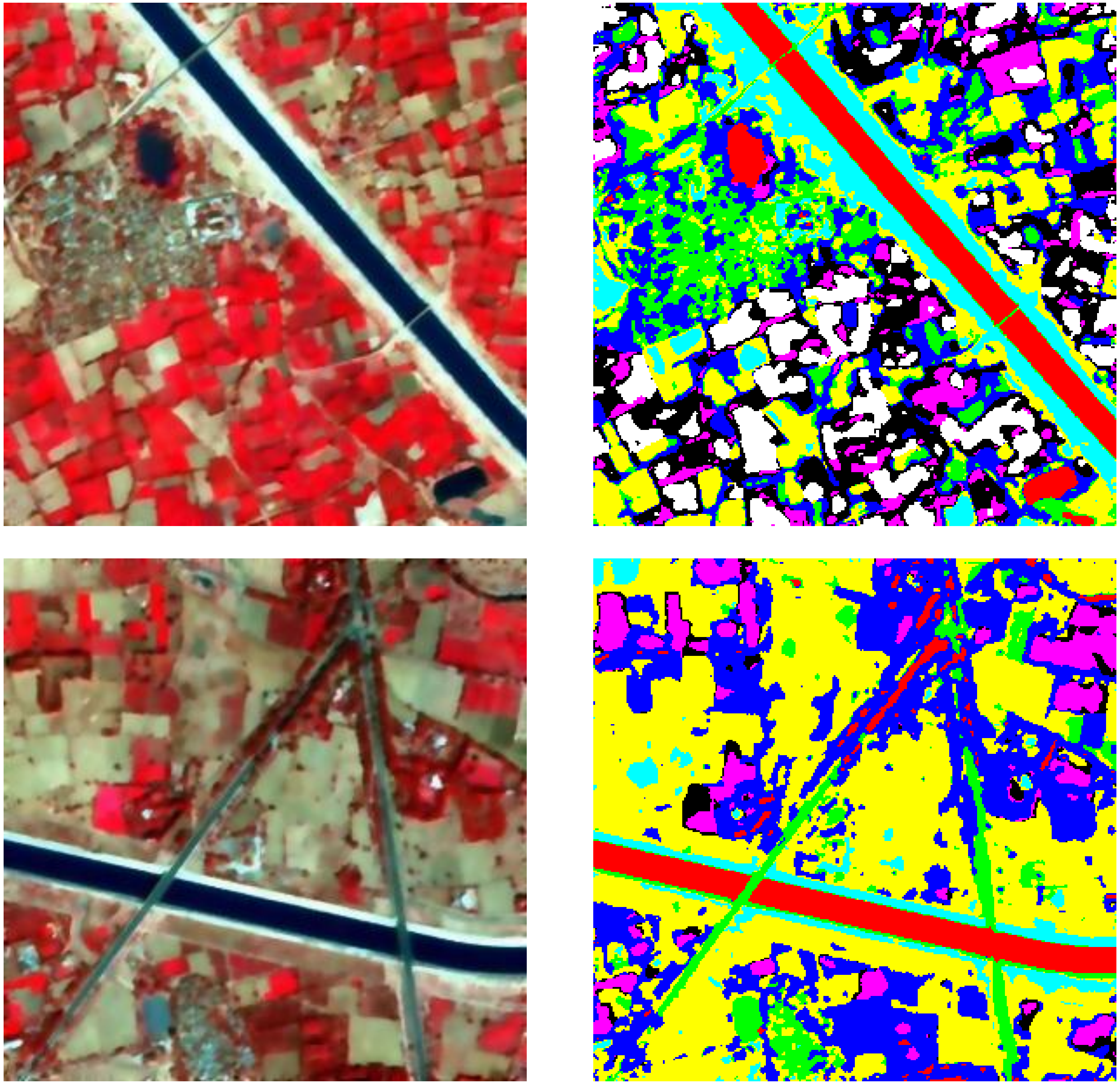
Number of clusters: 5

Number of clusters: 3





Peak detection followed by Bayesian learning for Satellite image classification.



*Satellite Images: courtesy BISAG, Gandhinagar

Acknowledgement

- Abhishek Singh (BTech student)
- Padmini Jaikumar (BTech student)

References

[1] J. Bilmes. A gentle tutorial of the EM algorithm and its application to parametric estimation for Gaussian mixture and hidden markov models. Technical Report, Univ. Calif. Berkeley, 1997.

[2] D.L Davies and D.W Bouldin. A cluster separation measure. *IEEE Trans. PAMI,* 1:224-227, 1979.

[3] A.Dempster, N.Laird, and D.Rubin. Maximum likelihood from incomplete data via the EM algorithm. *JRSS*, 39(1):1-38, 1977.

[4] Y Lee, K.Y Lee, and J Lee. The estimating optimal number of gaussian mixtures based on incremental k-means for speaker identification. *International Journal of Information Technology*, 12(7):13-21, 2006.

[5] A Likas, N Vlassis, and J.J Verbeek. The global k-means clustering algorithm,*The Journal of the Pattern Recognition Society,* 36:451-461, 2003.

[6] S Ray and R Turi. Determination of number of clusters in k-means clustering and application in color image segmentation. In Proceedings of *ICAPRDT'99, 1999.*

[7] G Schwarz. Estimating the dimension of a model. *Annals of Statistics,* 6:461-464, 1978.

8] A. Singh, P. Jaikumar, S.K Mitra, and M.V Joshi. Low contrast object detection and tracking using gaussian mixture model with split-and-merge operation, *International Journal of Image and Graphics (Submitted)* 2008.

[9] A. Singh, P. Jaikumar, S.K Mitra, M.V Joshi, and A. Banerjee. Detection and tracking of objects in low contrast conditions. In Proceedings of *NCVPRIPG 2008,* pp. 98-103, 2008.

[10] A.F.M Smith and A.E Gelfand. Bayesian statistics without tears: A sampling-resampling perspective. *The American Statistician,* 46(2): 84-88, May 1992.

[11] C Stauffer and W.E.L Grimson. Adaptive background mixture models for real-time tracking. In *Proc. CVPR,* pp. 599-608, 1999.

[12] N Ueda, R Nakano, Z Ghahramani, and G.E Hinton. SMEM algorithm for mixture models. *Neural Computation,* 12(9): 2109-2128, 2000.

[13] L Xu and M.Jordan.On Convergence properties of the em algorithm for gaussian mixtures.*Neural Computation,* 8:129-151, 1996.

[14] Z.R Yang and M. Zwolinski.Mutual information theory for adaptive mixture models. *IEEE Trans. PAMI,* 23(4): 396-403,2001.

[15] Z Zhang, C Chen, J Sun, and K.L Chan. Em algorithms for gaussian mixtures with split-and-merge