

Background Subtraction with Dirichlet process Gaussian Mixture Model (DP-GMM) for Motion Detection

Himani K. Borse¹, Bharati Patil²

^{1,2}G.H. Raisoni College of Engineering and Management, Wagholi, Pume, India

Abstract: Video analysis often starts with background subtraction. This problem is often loomed in two steps: Per-pixel background model followed by regulation scheme. A background model allows it to distinguished on Per-pixel basis from foreground, though the regularization combines information from adjacent pixels. Dirichlet process Gaussian mixture models is a method, which are used to approximate per-pixel background distributions followed by probabilistic regularization. Per pixel modes are automatically count by using non-parametric Bayesian method, avoiding over-/under- fitting. We implemented this method using FPGA and also compare the results with different methods like Background subtraction; Frame difference and Neural map and shows how this method is superior then previous methods.

Keywords: Background subtraction, Dirichlet processes, video analysis.

1. Introduction

This method is a non-parametric Bayesian method that spontaneously estimates the Number of mixture components is automatically estimate by this method to model the pixels background color distribution, e.g. Single mode pixel generate at the trunk and in the sky, when the tree is waving forward and backward in front of sky creates two modes pixels in the area where braches wave, i.e. pixel transition among leaf and sky regularly. If it requires more modes to denote multiple leaf color this will take place automatically, and, of excessive significance for term surveillance, this model will update with time. e.g., on a calm day if tree is not moving it will return to single model pixel throughout. it avoids over-under fitting as it is a fully Baysian. However, two issues of standard DP-GMM model: 1) update techniques of the existing model cannot cope with the scene changes common in real-world applications; 2) if we used this model for continues video then more computation and memory is required.

This model usages a Gaussian mixture model (GMM) for a Per pixel density estimation is carried by Gaussian mixture model and this is followed by connected component of regulation. Its mixture model has two components foreground and background. -This model categorizes values based on their mixture components, which is allocated to the foreground or the background larger components belongs to background and remaining belongs to foreground If object hangs time then this assumption is not good.it takes a fixed component count, which does not equal actual validity, where number of modes varies between pixel

2. Block Diagram

Video this is in AVI format is taken because processing take place on AVI format. Then this video is converted into frames and then color conversion is take place that convert the color (RGB) images into gray form. And reduces lighting effect occurs at the time of capturing of video.The proposed

method normally splits into two parts—a per-pixel background model and a regularisation step

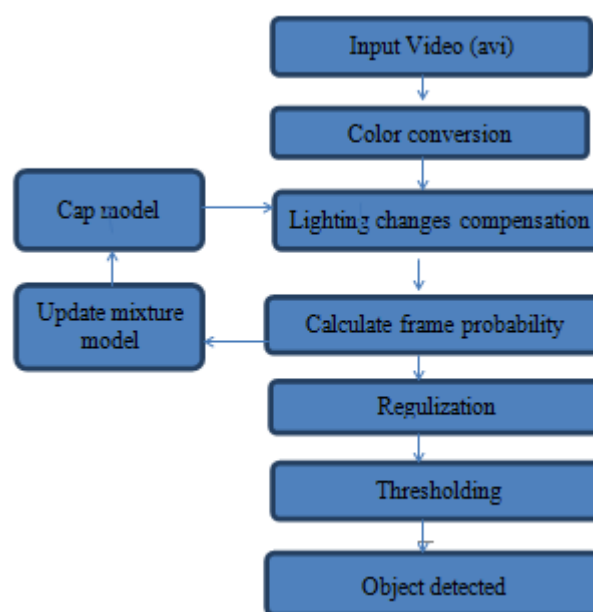


Figure 1: Block diagram of the approach

2.1 Per-Pixel Background Model

Every single pixel has its multi-model density estimate, used to model $P(x/bg)$ where x is the pixel color channels vector. The Dirichlet process (DP) Gaussian mixture model is used it can be observed as the Dirichlet distribution prolonged to an infinite number of components, which permits it to obtain the true number of mixtures essential to signify the data. Each pixel a value of values reaches, one with every frame-the model has to be constantly using incremental learning. Dirichlet process, first using the stick breaking construction then secondly using the Chinese restaurant process (CRP). Gives clean description of concept is provided by stick breaking, whereas the Chinese restaurant process integrates out inappropriate variables and offers the formulation we actually solve.

2.1.1 Dirichlet process

In probability theory, are family of stochastic method whose realization are probability distribution. In other ways a Dirichlet process is a probability distribution whose domain is itself a group of probability distribution. It is frequently used in Bayesian interference to define the preceding knowledge nearly the distribution of random variables, that is, how possibly it is that the random variable are distributed allowing to one or another specific distribution.

The Dirichlet process is presented by a base distribution H and a positive real number α called the concentration parameter. The base distribution is the probable the process value, that is, the dirichlet process perform distribution "around" the base distribution in the manner that a normal distribution draws real numbers around it means. Though, even if the base distribution is constant the distributions drawn from the Dirichlet process remainsurely discrete. The concentration parameter identifies how robust this discretization: is in the limit of $\alpha \rightarrow 0$, the realization are all concentrated on a particular value, whereas in the limit of $\alpha \rightarrow \infty$ the realization becomes between the two extreme the realization are discrete distribution with fewer and fewer concentration as increases.

The Dirichlet process that may also be understood as the infinite-dimensional generalization of the Dirichlet distribution. In the related manner as the dirichlet distribution is the infinite conjugate prior., In the same manner as the Dirichlet distribution is the categorical distribution conjugate prior, the dirichlet process is the infinite conjugate prior, non-parametric discrete distribution, A particularly important application of the dirichlet process is the prior probability distribution in infinite mixture model. Assume that the production of values X_1, X_2, \dots can be described with the following algorithm.

Input: H (Probability distribution called Base distribution), α (Positive real number called Concentration parameter)

1. Draw X_1 from the distribution H .
2. Forn: $n > 0$
1. With probability $\frac{\alpha}{\alpha+n-1}$ draw X from H .
2. With probability $\frac{n_x}{\alpha+n-1}$ set $x_n = x$, where n_x is the number of previous observations X_j , $j < n$ such that $X_j = x$.

The X_1, X_2, \dots the observations are dependent, then we have to think through the previous results when producing value. They are still, replaceable. This fact may be displayed by computing joint probability distribution of the observations and seeing that the resultant formula only based on which value X occur among the observations and how many duplications they each have. Procedure of the above algorithm:

1. Draw a distribution P from $DP(H, \alpha)$
2. Draw observations independently X_1, X_2, \dots from P .

2.1.2 The Chinese restaurant process

The "Chinese restaurant process" name is stated from the following analogy: imagine an infinitely big restaurant having an infinite tables, and capable to serve an infinite dishes. The restaurant in question works with a slightly unusual seating policy whereby new dinners are seated either at presently working table with probability proportional to the number of guests at present seated there, or at an unfilled

table by means of probability proportional to a constant. Guests who sit at engaged table essential order the identical dish as those presently seated, whereas guests assigned a new table are served a different dish at random.

The dishes distribution after J guests are served is sample drawn as described below.

Suppose that J samples, $\{\theta_j\}_j^J = 1$ samples must before been got according to Chinese restaurant process, the $(J + 1)^{th}$ sample should be drawn from

$$\theta_{(J+1)} \sim \frac{1}{(H(S)+1)} (H + \sum_{j=1}^J \delta_{\theta_j})$$

Where δ_{θ} is atomic distribution. Understanding this, two properties are clear:

1. Even if S is uncountable set, there is finite (i.e. non zero) probability that two samples will have nearly the similar value. A dirichlet process samples are discrete.
2. The dirichlet process shows a self-reinforcing property the further every so often a identified value has been sampled in the previous, the best probable it is to be sampled again

2.1.3 Stick breaking

A third approach to the Dirichlet process is so-called stick-breaking process view, a dirichlet process are distributions over a set S . As the distribution drawn is discrete with probability 1. In the sticking-breaking procedure opinion, we clearly use the discreteness and provide the probability mass function of this random discrete distribution as:

$$f(\theta) = \sum_{k=1}^{\infty} \beta_k \cdot \delta_{\theta_k}(\theta)$$

Where δ_{θ_k} is indicator function which estimates to zero all over, excepts for. $\delta_{\theta_k}(\theta_k) = 1$. Then this distribution is itself, its mass function is parameterized through two of random variable: the locations $\{\theta_k\}_{k=1}^{\infty}$ and the corresponding probabilities $\{\beta_k\}_{k=1}^{\infty}$. In the current deprived of proof what these random variables are.

- 1) The locations θ_k are identically and independent distributed according to H , base distribution of the dirichlet process. The probabilities β_k are specified by a procedure approximating the breaking of a unit-length stick:

$$\beta_k = \beta'_k \cdot \prod_{i=1}^{k-1} (1 - \beta'_i)$$

Where β'_k are independent random variables with the beta distribution $Beta(1, \alpha)$.

The correspondence to 'stick-breaking' can be realized through seeing as β_k the length of a part of a stick. We begin with a unit-length stick and every step we halt a portion of the remaining stick according to β'_k and assign this broken-off piece to β_k . The formula can be understood by observing that subsequently the first $k - 1$ values have their portions allocated, the length of the rest of the stick is $\prod_{i=1}^{k-1} (1 - \beta'_i)$ and this portion is broken according to β'_k and becomes assigned to β_k .

- 2) The smaller α is, the fewer of the stick will be left consequent values (on average), and resulting further concentrated distributions.

In Stick breaking, stick is continuously break infinite times and divides the samples into different Chinese restaurant sets. Integrating out the draw from the DP indications to better convergence, but more significantly replaces the

infinite set of stick with the computationally controllable finite set of tables.

2.2 Probabilistic Regularisation

Per-pixel background model does not take information from the adjacent or neighboring pixel so causes it susceptible to noise and camouflage. Additionally, Gibbs sampling introduces certain amount of noise i.e. dithering effect at the boundary between foreground and background. This issues is resolved by using markov random field, with node of each pixel connected to four way neighborhoods. It is a binary labeling problem where every single pixel corresponds to the either foreground or to background.

The key is to consider the best probable solution, where probability can be separated into two terms. First, every single pixel has a probability of going to the background or foreground, straightacquired from the model. Threshold value is set to 45, the pixel above threshold is considered as foreground and below background. Cap model is used to update the model Instead of repeating the same calculation. Whatever output is obtained that is updated so when next same output is generated then the previous outputs are taken that are stored so reduces same calculations.

3. Hardware

This method is implemented in FPGA.

3.1 FPGA

FPGAs contain an array of programmable logic blocks, and reconfigurable interconnects the logic blocks together, like various logic gates that can be present inter wired in different configuration. Logic blocks are configured to achievedifficult combinational functions or simple logic gates such as the AND or XOR. In most FPGAs, logic blocks also contain memory elements, which can be simple flip-flops or additionalwhole blocks of memory.

Board Features

- FPGA Spartan XC3S50A in TQG 144 package
- USB 2.0 interface for on-board flash programming.
- Flash memory: 16 Mb SPI flash memories (M25P16).
- FPGA configuration via JTAG and USB
- 39 IOs for user defined purposes
- Six Push buttons ,8 LEDs and 8 way IP switch for user defined purposes
- One VGA Connector One Stereo Micro SD Card Adapter Three Seven Segment Displays
- On-board voltage regulators for single power rail operation

FPGA receives the current and reference frame data through serial communication bus UART. and stored data in RAM and Perform the and generates the output. This output is given to matlab through UART3.

3.2 RAM

A typical RAM cell has only four connections: Data in (the D pin on the D flip-flop), data out (the Q pin on the D flip-

flop), Write Enable (often abbreviated WE; The C pin on the D flip-flop), and Output Enable (the Enable pin which we added).For our board, ram is 16 Mb. Ram stored current frame, reference frame, intermediately generated result (processing) and output.

3.3 UART

The universal asynchronous receiver/transmitter (UART) receives bytes of data and sends the individual bits in sequential manner.at the destination, a second UART re-assembles the bits into whole bytes Each UART holds a shift register which convert serial data into parallel form and vice versa. Transmission of single bit of bye in single wire is less costly than transmission of multiple data in parallel form in multiple wires.

3.3.1 Asynchronous serial communication terms

In asynchronous transmitter and receiver cannot share common clock.

- Clock Start bit - shows the initiation of the data word. When detected, the synchronizes with the new data stream. Check if it is one then transmission starts.
- Stop bit-shows the end of the data word. The stop bit symbols the end of transmission. Check low transition indicate stop bit
- Parity bit-inserted for error detection (optional). The parity bit is inserted to make the number of 1's even (even parity) or odd (odd parity). This bit is used by the receiver to detect for transmission errors. In our system, even parity is used. So no of ones are even then parity bit set else reset (low)
- Ack bit: if parity of matlab and FPGA matches then this bit is high. Data bits-the actual data to be transmitted bit data transmitted.
- Baud rate-the bit rate of the serial port. 9600 boud rate is set.

4. Algorithm of Overall system flow:

- First the predefined input which is in AVI format is taken.
- Then the video is converted into frames and then binary images (contains only 0 and generated in matlab.
- Then bit of binary image is transmitted to FPGA through serial communication in our case we used UART serial communication. And stored bits in RAM.
- FPGA do all the processing and generate the output. Then 8 bit of output is transmitted to Matlab through UART and
- then display the out on Matlab GUI.

4.1 Overall System Flowchart

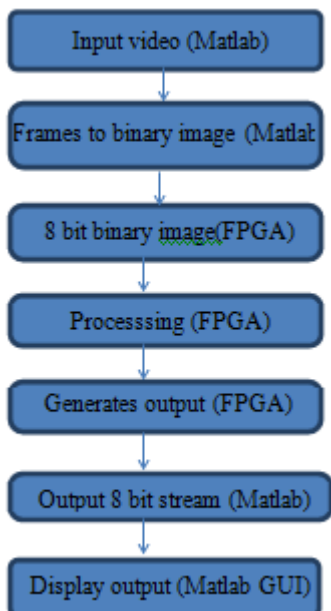


Figure 4.1: Overall system flowchart

5. Performance and Results

5.1 Output Parameters

5.1.1 MSE: MSE as a measure signal fidelity. The signal fidelity compare the quantitative score of two signal so we can described how to signals are similar and level of distortion (noise) between them. Typically, it is considered that onesignal is an original signal, while the other signal is distorted or contaminated with errors.

Assume that $\mathbf{x} = \{x_i | i= 1, 2, \dots, N\}$ and $\mathbf{y} = \{y_i | i= 1, 2, \dots, N\}$ are two finite-length, discrete signals (e.g., visual images), where N is the number of signal samples (pixels, if the signals are images) and x_i and y_i are the i th samples in \mathbf{x} and \mathbf{y} , respectively.

The MSE is defined as:

$$MSE(x,y) = \frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2$$

The MSE, also denote to the error signal $e_i = x_i - y_i$, which is the difference of the original and contaminated signals. If one of the signals is an original signal of suitable (or possibly original) value and the other is a distorted form of it whose value is being calculated, then this may also be described as a measure of signal quality

5.1.2 PSNR: Peak signal-to-noise Ratio

Abbreviated as PSNR. It is the ratio of the maximum possible signal power to the corrupting noise power that distracts the fidelity of the representation. Because many signals have several varied dynamic range, PSNR is usually stated with the help of the logarithmic decibel scale. It is most frequently used for measuring the superiority of reconstruction of lossy compression codecs (e.g. for image compression). The signal here is an original data, and noise is error introduced by the compression. When relating compression codecs PSNR is an estimate to human perception of reconstruction value. Even though higher PSNR value shows the reconstruction is of higher quality.

PSNR is most easily well-defined via mean squared error (MSE).

The PSNR (in dB) is defined as:

$$\begin{aligned} PSNR &= 10 \log_{10} \left(\frac{MAX^2}{MSE} \right) \\ &= 20 \cdot \log_{10} \left(\frac{MAX_1}{\sqrt{MSE}} \right) \\ &= 20 \cdot \log_{10}(MAX_1) - 20 \cdot \log_{10}(MSE) \end{aligned}$$

Here, MAX₁ is the extreme possible pixel image value. When the pixel are denoted with the help of bits per sample, this value is 255. Mostly, when denoted using linear PCM with B bits per sample, MAX₁ is $2^B - 1$.

5.1.3 Entropy

Entropy is a statistical degree of uncertainty that can be used to describe the texture of the input image. Entropy is an index to evaluate the how much information (quantity) contained in an image. Entropy is defined as

$$E = - \sum_{i=0}^{L-1} p_i \log_2 p_i$$

Where L is the total grey levels, $p = \{p_0, p_1, \dots, p_{L-1}\}$ is the probability distribution of each level

5.1.4 Correlation

Normalized cross correlation are used to find out likenesses between current and reference image is given by the following equation

$$NCC = \frac{\sum_{i=1}^m \sum_{j=1}^n (A_{ij} * B_{ij})}{\sum_{i=1}^m \sum_{j=1}^n (A_{ij})^2}$$

5.2 Results

Output of different motion detection a technique are calculated and shows that how this method is superior than all other methods.

Background Subtraction: Ddifference between current and reference frame

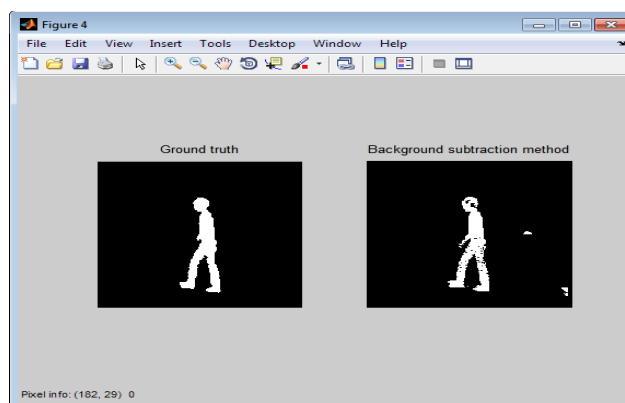


Figure 5.2.1: Output of background subtraction method

Frame difference: difference between two consecutive frames

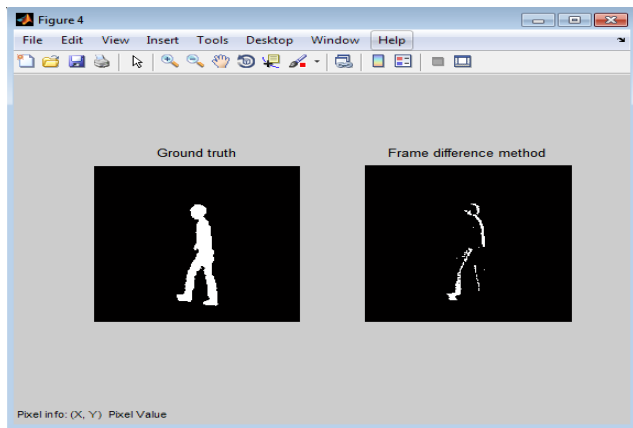


Figure 5.2.2: Output of Frame Difference method

Neural Map method: pixel mapped into 3*3 mapping of current and reference frame.

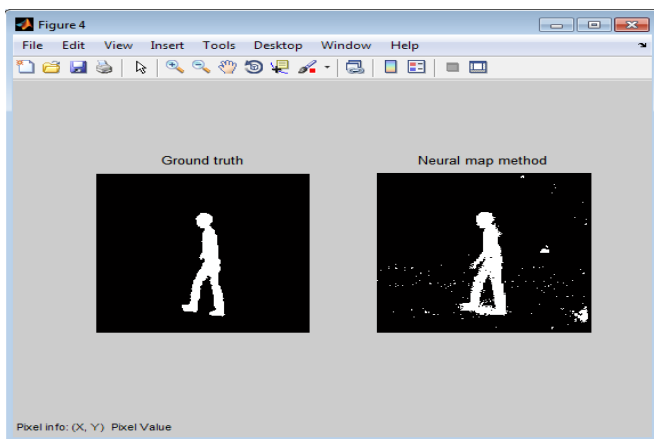


Figure 5.2.3: Output of Neural Map method

Background Subtraction with Dirichlet process Gaussian mixture model (DP-GMM)

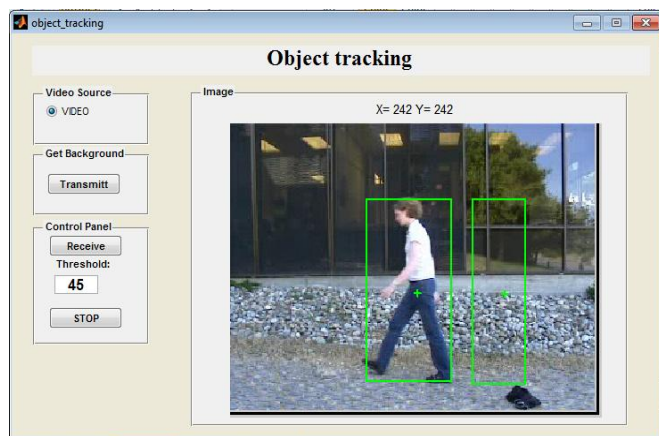


Figure 5.2.4: Output of Neural Map method

6. Comparison of Different Method Parameters

Method /parameter	Background subtraction method	Frame Difference method	Neural Map method	DP-GMM
MSE	0.0115	0.0487	0.0245	0.0102
PSNR	67.5392	61.2563	64.243	68.673
Entropy	0.2950	0.0917	0.3766	0.2870
Correlation	0.8832	0.2783	0.8041	0.8812

7. Conclusion

The method is based on an present model, that is DP-GMMs, with different model learning algorithms developed to make it both suitable for background modeling, and computationally scalable. This method handles the dynamic background. Infinite no of mixture components are used so whatever object is detected is more accurate than the other methods. And also handles the scene changes. It also shows this one to have decent performance in various other parts mainly on dealing with densenoise. It handles camouflage and shadow effect. From the above table, we can conclude that the output or human body detection done by Background subtraction with Dirichlet Process Gaussian Mixture model method is more accurate than the another methods

8. Application

- Object detection
- Video surveillance
- Object tracking
- Traffic monitoring

9. Future Scope

Combining information from adjacent pixel in regularization does not completely achieve the information available. A more challenging method of spatial information transmission would be desirable-a reliant Dirichlet process might provide this. Sudden complex lighting fluctuations are not controlled by this method, which means it fails to handle certain indoor lighting changes. Still, a more sophisticated typical of the foreground and a clear model of left object could additional improve our method.

References

- [1] Weiming Hu, TieniuTan, "A Survey on Visual Surveillance of Object Motion and Behaviors" iee transaction on systems, and cybernetics-Part C: application and reviews, vol. 34, no. 3, august 2004
- [2] QiZang and ReinhardKlette, "Object Classification and Tracking in Video Surveillance", unpublished
- [3] Kinjal A Joshi, Darshak G. Thakore, "A Survey on Moving Object Detection and Tracking in Video Surveillance System", International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-3, July 2012
- [4] Zhen Tang, Zhenjiang Miao, "Fast Background Subtraction and Shadow Elimination Using Improved Gaussian Mixture Model", HAVE 2007 - IEEE International Workshop on Haptic Audio Visual Environments and their Applications, Ottawa - Canada, 12-14 October 2007
- [5] NishuSingla, "Motion Detection Based on Frame Difference Method" International Journal of Information & Computation Technology. ISSN 0974-2239 Volume 4, Number 15 (2014), pp. 1559-1565
- [6] M. Julius Hossain, M. Ali AkberDewan, and OksamChae, "Edge Segment-Based Automatic Video Surveillance", Journal on Advances in Signal Processing Volume 2008,

- [7] Tom S.F. Haines and Tao Xiang ,”*Background Subtraction with Dirichlet Process Mixture Models*”ieee transaction on pattern analysis and machine intelligence,vol. 36,no 4,April 2014
- [8] Lucia Maddalena and Alfredo Petrosino,,” *A Self-Organizing Approach to Background Subtraction for Visual Surveillance Applications*”,ieee transaction on image processing vol. 17, No. 7, July 2008
- [9] BabakShahbaba and Radford Neal ,”*Nonlinear Models Using Dirichlet Process Mixtures*”,Journal of Machine Learning Research 10 (2009) 1829-1850
- [10] Larissa ValmyAnd Jean Vaillant,”*Bayesian Inference on a Cox Process Associated with aDirichlet Process*”, International Journal of Computer Applications (0975 8887) Volume 95 - No. 18, June 2014
- [11] Ibrahim SayginTopkaya, HakanErdogan and FatihPorikli” *Counting People by Clustering Person Detector Outputs*”, 2014 11th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)
- [12] Radford M. Neal”*Markov Chain Sampling Methods for Dirichlet Process Mixture Models*”, *Journal of Computational and Graphical Statistics*, Vol. 9, No. 2. (Jun., 2000), pp. 249-265.
- [13] Daniel J. Navarro, Thomas L. Griffithsb, Mark Steyvers, Michael D. Lee,” *Modeling individual differences using Dirichlet processes*”, *Journal of Mathematical Psychology* 50 (2006) 101–122
- [14] Yee WhyeTeh, Michael I. Jordan, Matthew J. Beal and David M. Blei, “*Hierarchical Dirichlet Processes*”, *Journal of the American Statistical Association*, Vol. 101, No. 476 (Dec., 2006), pp.1566-1581
- [15] DilanGorur and Carl Edward Rasmussen,”*Dirichlet Process Gaussian Mixture Models: Choice of the Base Distribution*”,*Journal of computer science and technology* 25(4): 653–664 July 2010.