

## Improved Path Detection in Tracking of Moving and Stationary objects in Moving Cameras

<sup>#</sup>Harihara Santosh Dadi, <sup>\*</sup>Gopala Krishna Mohan Pillutla, <sup>#</sup>Madhavi Latha M.

<sup>#</sup> Department Of ECE, JNT University, Hyderabad, India

<sup>\*</sup> Department Of ECE, Institute Of Aeronautical Engineering, Hyderabad, India

---

**Abstract:** Detection and tracking of moving objects is an important topic of research in the field of video surveillance. Most of the existing video surveillance cameras are stationary and the tracking algorithms are developed for the same. In recent times, the applications are not confined to stationary cameras. There is a lot of technological improvement in surveillance videos and there is a high requirement of tracking a moving object in non stationary cameras. A new calibration procedure for effectively tracking moving objects in non stationary cameras is presented in this paper. The procedure is adapted to Gaussian Mixture Model and Mean Shift Algorithms. Results show that the calibration procedure presented in this paper is best suited for all tracking algorithms.

**Index Terms:** Non stationary surveillance videos, GMM, MSA.

---

### I. Introduction

Tracking of moving objects in a sight is surely a significant subject matter in the field of computer vision. It is one of the fundamental steps in so many video-based systems. For instance, applications such as intelligent surveillance systems, robot visions and human computer interface (HCI), [1] require detection and tracking of moving objects. Numerous techniques have been proposed and are verified to be victorious for detection of moving objects in case of nonmoving cameras [2,3], but in case of pan-tilt-zoom (PTZ) or mobile cameras the existing methods do not work well because of many factors that take place when using moving cameras. Several methods have been proposed for tracking of moving object. When applied to moving cameras, the motion compensation procedure for the algorithm involves more number of computations comparative to the total number of samples employed for a pixel. This may be time-consuming for the method in significant amounts unless there is some sort of hardware support. Apart from the computations involved, when modeling the scene, it is also important that the model considers not only the errors and noises that arise in stationary cameras, but also the errors that arise when compensating for the motion of the camera. This is a serious reason that we cannot simply just adopt background subtraction algorithms for nonmoving cameras with simple motion compensation algorithms. Algorithms based on moving camera background modeling usually focus on constructing accurate model for every pixel. But for moving case, we cannot assure that the model used to estimate a pixel is in fact related to that pixel. Even the smallest amount of erroneous motion compensation could result in building the technique use wrong algorithms for few pixels. To account for such motion compensation errors, small nearby neighborhoods are considered [4]. However, considering neighborhoods increases the necessary computation, slowing down the whole algorithm. In order to take in to consideration the motion compensation errors, the nearby neighborhood pixels in small amounts are considered. But, considering these pixels increases the computations and as a result slows down the whole algorithm. Our model which is based on background subtraction is intended in a way that reduces the computational requirements and demonstrates accurate tracking of path in moving cameras. The pseudo path can be eliminated by using this method. The proposed algorithm gives an accurate trajectory of the object of interest in the case of moving camera. The algorithm assumes that the motion of the camera is either horizontal or vertical and the motion is linear. Section II gives the related work in this area. Section III describes the Gaussian Mixture Model. Section IV explains the Mean Shift Algorithm. Experimental results and conclusions are given in Section V and VI.

### II. Related Work

Roughly all existing methods for static suspicious object detection and tracking are meant for finding suspicious objects by means of a static camera in a public place, e.g., metro station, commercial center or airport hall. C. Stauffer *et.al* anticipated surveillance tracking system to detect suspicious objects automatically and draw the attention of operator to those events [5]. It involves two major parts: A cluster-based object tracking system that recognizes suspicious objects using the unsolved image parts and a Bayesian multi-person detector

that elucidates as much of the scene as possible. If a probable suspicious object is identified, the operator is notified, and the method presents the operator the suitable key frames for inferring the scene. D. Harihara Santosh et al, proposed cluster analysis [6] for suspicious object detection. First, the short and long-term backgrounds are separately constructed. After that, based on the two background models, two foreground models are acquired. The suspicious object can be noticed by four hypotheses based on the two background models and two foreground models. D. Harihara Santosh et al., also anticipated to use a tracker based on background-subtraction and the projection of the center of mass of each object on the ground is marked [7]. The tracker recognizes object segmentations and eligible them for chance of a “drop-off” event. The nonmoving object detector running in parallel with the tracker quickly identifies probable stationary foreground objects by watching for pixel regions that consistently deviate from the background for a set duration of time. The nonmoving object tracker running in line with the detector identifies quickly the most probable nonmoving foreground objects by monitoring for cluster regions that repeatedly for set duration of time deviate from the background. Using location information, the suspicious objects are linked within the views of multiple cameras and a time-weighted voting scheme among the camera views is used to tell the final notice and eliminate the view dependencies effects. Chris Stauffer et al. recommend adapting parallel approach [8]. The primary step is to track objects in the view using a trans-dimensional MCMC tracking model apt for general cluster tracking tasks. The tracker adapts a single view camera and it won’t not discriminate between luggage and people. By analyzing the response of the tracking system in a tracking and detection process, the lasting problem of identifying whether a suitcase is left unattended is solved. All the above mentioned techniques use the nonmoving camera mounted in few open places, where stationary background is used. For some application scenarios however, to keep monitor on is too huge to use nonmoving cameras. Therefore, it is essential to employ moving camera to search these places. We use in this paper a camera mounted on a moving platform to perfectly track the object. The tracking results of the moving camera of both the existing and proposed techniques are compared.

### **III. Object Detection Using Gaussian Mixture Model**

There is a parametric model of probability density function for object detection. That model is coined as Gaussian Mixture Model (GMM). This model is represented in following way: GMM is equal to the weighted summation of the so called Gaussian component densities.

This Gaussian Mixture Model can be used as background model. To get the anticipated outcomes pixels of frame are removed from necessary video [9]. This background subtraction includes different issues that implicit emerging an algorithm that can used in recognizing the object. Moreover, it could be capable in responding to different alterations like moving and halt of motion objects and illumination.

#### **A. Background Subtraction**

Background subtraction is relied on below mentioned four critical steps:

##### **1) Preprocessing**

Temporal or spatial smoothing will be applicable at early stages of preprocessing to eradicate noise in the device. This noise could be an issue under the various intensities of light. Smoothing technique involves deleting different elements of environment such as rain and snow. In real time systems, to minimize the data rates for processing, frame rate and frame size are generally used. The important factor while preprocessing the technique that can be used for data format using background subtraction model. Various algorithms are there to process intensity of luminance that is single valued scalar per every pixel.

The below two figures shown among which left one shows snow and on right side shows spatial and temporal smoothing application. This outcomes removal of snow in more effective way to get clear image.



**Figure 1:** Image on the left shows snowing and image on the right is a resultant of smoothing effect

## 2) **Background Modeling**

Using video frame this step computes background model. The important motto of this emergence of background model is to withstand the alterations in environment in background. However, it is complicate to recognize the motion objects.

## 3) **Foreground Detection**

This step recognizes the frame pixels. This foreground detection will differentiate video frame and background model. General way for foreground detection will verify that pixel is distinguished from background.

## 4) **Data Validation**

Ultimately, the final step eradicates the pixels which are not used in image.

It includes enlighten of process for foreground mask which is relied on data that is actually extracted from background model.

There are efficient key points where it lags. These are mentioned below:

1. Disregarding any correlation among adjacent pixels.
2. The rate of amendment may not be compatible with the motion velocity of the so called foreground object.
3. Pixels which are Non-stationary, from casting shadow or moving leavers by motion objects that actually violates from the true one.

## **B. Algorithm of Gaussian Mixture Model:**

To get well aware of algorithm that applied at background subtraction, below steps are studied.

Foremost thing to be done is to distinguish every input pixel for mean ' $\mu$ ' for components. If pixel value nearer to mean of selected component, that particular component is taken as compatible component. To be a compatible component, the difference of pixel and mean obtained should be less. This is matched with standard deviation with scaling factor of D. Secondly, Gaussian weight, mean and standard deviation (variance) are updated to replicate the obtained new pixel value. Further components which are non-matched decreases to weight ' $w$ ' and the mean and standard deviation won't change. This relied on learning component ' $p$ ' to state the instant alterations.

Thirdly, categorize the components are which portions of background model. To accomplish this task a threshold value is used as component weights ' $w$ '.

Fourthly, we regulate the pixels of foreground. Here the recognized pixels as foreground will not be suitable with any other components that are firm to the background.

## **C. Methodology**

The GMM is defined as a Combination of Gaussian distributions of K that analyze the alterations of state for equivalent pixels of frames. Hence algorithm established put on Gaussian mixtures to individual frame and alters images from colorful images to binary. The pixels which has no change in its state is represented by value 1 that is black color and the pixels which change their state is represented as value 0 that is white color. Hence, creating positions for motion objects of video is made possible.

Below shown a figure 2 that represents the pixels equivalent to road of the background image that doesn't suffer changes in state, consequently the black color that is value 0 is represented and seems black in color as revealed in below figure 2.a. The pixels of image equivalent to cars experience radical state alteration, Hence the white color which is represented as value 1 is represented and cars seems as white in color which is shown in below figure 2.b



Figure 2.a

Figure 2.b

## **IV. MEAN SHIFT ALGORITHM**

Fakunaga and Hostetter proposed an algorithm called as Mean-shift (MS) object-Tracking Approach. It is basically iterative expectation maximization – clustering algorithm executed within local search regions [10].

In other words, it is a type of non-parametric clustering algorithm that does not need prior information of the number of clusters and also does not constrain the shape of the clusters. That is, the Mean-shift algorithm is a nonparametric density gradient estimator. The following steps are iterated in order to track the object by using the Mean-Shift algorithm: A. Select a search window size and the initial position of the search window. B. Estimate the mean position in the search window. C. Center the search window at the mean position estimated in Step B. D. Repeat Steps B and C until the mean position moves less than a preset threshold. That is, until convergence is achieved.

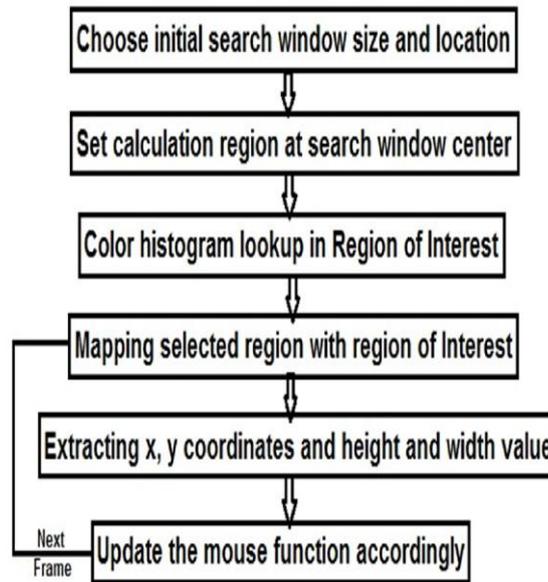


Fig. 3: Procedure of Object Tracking

After the Mean-Shift Algorithm is executed on various videos it is concluded that when the target moves so fast that the target area in the two neighboring frame will not overlap, tracking object often converges to a wrong object. Because of this issue, traditional Mean-Shift Algorithm gets failed to track fast moving object. But there are some solutions like combining Kalman filter or particle filter with Mean-shift Algorithm. The center of convergence will become more accurate as at first it predicts the direction and speed of the object and then adjusting the search window center of the mean shift convergence. However, such methods are unsuited to be using in real time tracking systems as these require high CPU costing to estimate the moving object location. The kernel-based tracking algorithm, when incorporated with prior task-specific information, can gain adequate results. This enhanced model could profitably detect and track a human subject in arbitrary motion and in a situation where there is a certain alter in radiance. It was constituted that a well built alteration in radiance induced the procedure to experience quite major deformation in the probability distribution image. Hence, the non-adaptive behavior of the mean shift algorithm may conduct to a wrong tracking conclusion.

**D. The kernel mask:**

The objects density estimates were weighted by a monotonically decreasing Epanechnikov kernel given by:

$$K_E(x) = \frac{1}{2} c_d^{-1} (d + 2) (1 - \|x\|^2) \quad (1)$$

If  $\|x\| < 1$  and 0 otherwise

Where  $c_d$  is the volume of the unit d-dimensional sphere and  $x$  are the normalized pixel coordinates within the target, relative to the center (i.e.  $\|x\|^2$  is a squared Euclidean distance of each pixel from the center of the target. Since we were dealing with a two dimensional image space, our kernel function was of the form:

The rational for using a kernel to assign smaller weights to pixels farther from the center is that those pixels are the least reliable, since they are the ones most affected by occlusion or interference from the background and noise. A kernel with Epanechnikov profile was essential for the derivation of the smooth similarity function between the distributions, since its derivative is constant; thus the kernel masking lead to a function suitable for gradient optimization, which gave us the direction of the target’s movement. The search for the matching target candidate in that case was restricted to a much smaller area and therefore much faster than the exhaustive search.

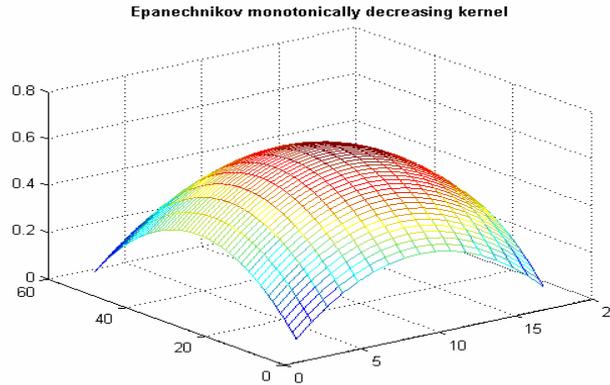


Figure 4: kernel mask

**E. Distance minimization:**

Based on the fact that the probability of classification error is directly related to the similarity of the two distributions, the choice of the similarity measure in was such that it was supposed to maximize the Bayes error arising from the comparison of target and candidate pdf's. Being a closely related entity to the Bayes error, a Bhattacharyya coefficient was chosen and its maximum searched for to estimate the target localization.

Bhattacharyya coefficient of two statistical distributions is defined as:

$$\rho[p(y),q] = \int \sqrt{p_z(y)q_z} dz \tag{2}$$

and in our case distributions  $p_z$  and  $q_z$  are the histograms of the target and the candidate, respectively.

**F. Mean shift:**

In order to find the best match of our target in the sequential frames, we needed to maximize the Bhattacharyya coefficient, which means that we needed to maximize the term

$$\sum_{i=1}^n \omega_i k\left(\left\|\frac{y-x_i}{h}\right\|^2\right) \tag{3}$$

where  $h$  is the kernel's smoothing parameter, or bandwidth, and  $-i$  is given by

$$\omega_i = \sum_{u=1}^m \sqrt{\frac{\hat{q}_u}{\hat{p}_u(\hat{y}_0)}} \delta[b(x_i) - u] \tag{4}$$

and  $\delta$  is the Kronecker delta function, equal to 1 only at  $u$  and 0 otherwise (i.e. only equal to 1 at the particular binu). The terms  $u \hat{q}$  and  $u \hat{p}$  are the values of the target and candidate histograms corresponding to pixel  $x_i$  of the candidate object. This mapping of color values given by  $u \hat{q}$  and  $u \hat{p}$  can be visualized to demonstrate how the target object changes over time and what the corresponding distribution of weights is the figures below show the gray scale images of some of these mappings, taken from the football sequence:

Note that the sum is actually a density estimate (i.e. a histogram) of the object centered at  $y$  in the current frame, computed with a kernel profile  $k(x)$  and weighted by  $-i$ . The maximum of this density in the local neighborhood (starting from the last known position of the target) gives us the most probable target position in the current frame, and it can be found by employing a mean shift procedure. During this procedure, the center of the target candidate is successively

$$\hat{y} = \frac{\sum_{i=1}^n x_i \omega_i g\left(\left\|\frac{\hat{y}_0 - x_i}{h}\right\|^2\right)}{\sum_{i=1}^n \omega_i g\left(\left\|\frac{\hat{y}_0 - x_i}{h}\right\|^2\right)} \tag{5a}$$

Where  $\hat{y}_0$  is the current location of the candidate center and  $g(x)$  is the derivative function. Since the derivative of the Epanechnikov kernel profile is constant, the above expression reduces to a weighted distance average

$$\hat{y} = \frac{\sum_{i=1}^n x_i \omega_i}{\sum_{i=1}^n \omega_i} \quad (5b)$$

The details of the mean-shift procedure are outlined .Here we will only refer to the implementation choices that we made and the results that we obtained.

### V. Data Sets

Soccer Data set 1 and 2 are used in this paper. The Camera used is a stationary Camera for Soccer Data Set 1 and in the Data Set 2 the camera is a non stationary. The data set 2 taken in which the camera is moving only in horizontal motion. Sample frames are shown in the following figures. The players in the video are moving and some are stationary. There are full occlusions and partial occlusions among the players. In Data Set 1, there are 20 players in the video. The video is of 25 frames per second, the duration is 1 minute 58 seconds and the number of frames are 2939. In the Data Set 2, there are 15 players; the video is of 7 frames per second. The duration is of 8 seconds. The camera used for Data Set 1 is a stationary camera.



**Figure 5:** Sample frame from the data set 1 (Stationary Camera)



**Figure 6:** Sample frame from the data set 2 (Non-stationary Camera)

## VI. EXPERIMENTAL RESULTS

In order to see the improvement of path detection in various algorithms namely, Mean Shift Algorithm and Gaussian Mixture Model, first the path has been detected. Figure 7 show that the template formed on the player in Dataset 1. Figure 8 show that the template formed on the player in Data set 2. For MSA algorithm ellipse template has been used and for GMM rectangle template has been used.



**Figure 7:** Template used in Data Set 1 for MSA Algorithm



**Figure 8:** Template used in Data Set 1 for GMM Algorithm

Figure 9 shows the template used in the Dataset 2 in MSA Algorithm. Figure 10 shows the template used in the Dataset 2 in GMM and also in GMM with Kalman Filter algorithms.



**Figure 9:** Template used in Data Set 2 for MSA Algorithm (Non Stationary Camera)



Figure 10: Template used in Data Set 2 for GMM Algorithm

Figure 11 and 12 shows the path formed in Data set 1 with MSA and GMM respectively. Since the data set 1 is formed using Stationary camera, the path shows is exactly the path of the player.



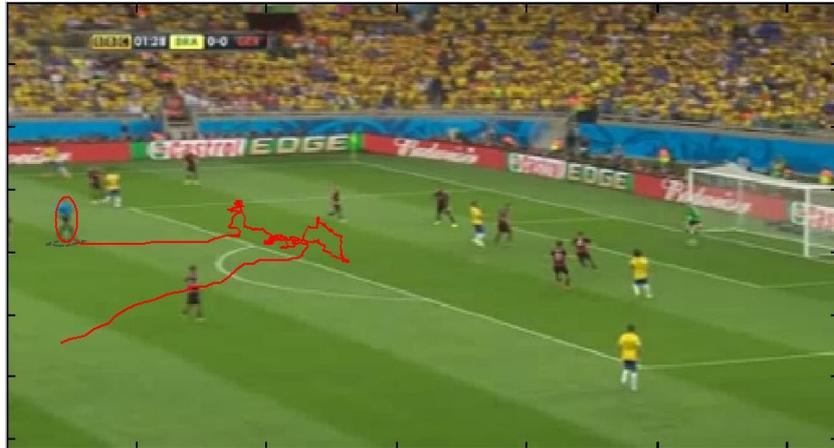
Figure 11: Path Detection in Stationary Camera (Data Set 1) using MSA Algorithm



Figure 12: Path Detection in Stationary Camera (Data Set 2) using GMM Algorithm

Figure 13 and 14 shows the path formed in Data set 2 with MSA and GMM respectively. Since the data set 1 is formed using Stationary camera, the path shows is not exactly the path of the player. Because the player of interest is stable and the camera is moving and as a consequence the player appears to be moving. This is called pseudo path detection.

In order to solve the problem of pseudo path detection in non stationary cameras, the pole in the Soccer video is taken as the reference and the difference is considered in solving the problem.



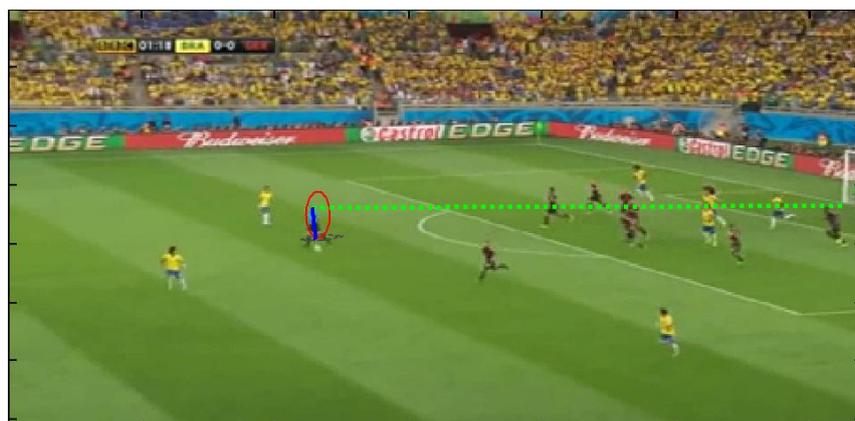
**Figure 13:** Path Detection in Non stationary Camera (Data Set 2) using MSA Algorithm



**Figure 14:** Path Detection in Non stationary Camera (Data Set 2) using GMM Algorithm

By using the new path detection technique, the exact movement of the player has been detected. Figure 15 and 16 shows the path detection using MSA and GMM. The Soccer pole shown in the frames has been taken as the reference and is shown in green dotted line. Because the distance between the pole and the player is considered in path detection, therefore the path shown is exact.

In this technique, the assumption made is that there is only a horizontal movement of camera.



**Figure 15:** Path Detection in Non stationary Camera (Data Set 2) using MSA Algorithm with Path Detection Technique



Figure 16: Path Detection in Non stationary Camera (Data Set 2) using GMM Algorithm with Path Detection Technique

## VII. Conclusions

From the results we concluded that for non stationary cameras the new technique we introduced in this paper is best suited when the objects are not moving. The new technique has been verified on MSA and GMM. The three algorithms showed excellent path detection results when the objects are not moving. Both the stationary camera and non stationary camera datasets have been taken and both are soccer videos. Here we assumed that the camera is moving only in the horizontal direction. The future scope could be when the camera is moving in both the directions. Here we have used only the tracking of on object. We will work out multiple object tracking in the future.

## References

- [1]. Rogers, Y., Sharp, H. and Preece, J.J. (2011) Interaction Design: Beyond Human-Computer Interaction (3rd ed.). John Wiley and Sons
- [2]. G. Stein, "Tracking from Multiple View Points: Self-calibration of Space and Time", IEEE Proc. CVPR, pp. 521-527, 1999.
- [3]. J. Kang, I. Cohen and G. Medioni, "Continuous Multi-Views Tracking using Tensor Voting", In Proc. of IEEE MCV, 2002.
- [4]. "Digital image processing using MATLAB" 2nd edition by Rafael C Gonzalez, Richard E. Woods and Steven L. Eddins, TMH publications, New Delhi ; 2011
- [5]. C. Stauffer, W. E. L. Grimson, "Learning Patterns of Activity Using Real-time Tracking", IEEE Trans. Pattern Analysis and Machine Intelligence, 22(8), 747-757, 2000
- [6]. D.Harihara Santosh and P.G. Krishna Mohan, "Tracking Manifold Objects in Motion Using Gaussian Mixture Model And Blob Analysis" at IEEE International Conference on Convergence of Technology – 2014 conducted at Pune during the period 5th to 7th April 2014, pp. 1-7.
- [7]. D.Harihara Santosh and P.G. Krishna Mohan, "Multiple Objects Tracking Using Extended Kalman Filter, GMM and Mean Shift Algorithm – A comparative Study", at IEEE International Conference on Advanced Communication Control and Computing Technologies (ICACCCT) at Ramanathapuram during the period 8th to 10th May 2014, pp. 1484 – 1488.
- [8]. Chris Stauffer, W Eric L Grimson, "Adaptive background mixture models for real-time tracking", Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference, vol. 2, pp. 246-252.
- [9]. D.Harihara Santosh and P.G. Krishna Mohan, "Multiple Human Tracking and Prediction under severe Occlusions using GMM and Kalman Filter" IJAER, Volume 10, Number 11 (2015) pp. 29385-29404.
- [10]. Georgescu B, Shimshoni I, and Meer P, "Mean Shift Based Clustering in High Dimensions: A Texture Classification Example," In: Proc. ICCV, Oct. pp. 456-463, 2003.



**Mr. D. Harihara Santosh** obtained his B. Tech. and M. Tech degrees from JNT University, Hyderabad in the year 2005 and 2010. Presently he is pursuing Ph.D. in Video Processing at JNTU, Hyderabad. He is presently pursuing his Ph.D. under the Guidance of Dr. P.G. Krishna He has 9 publications in both International and National Journals and presented 20 papers at various International and National Conferences. His areas of interests are Image and Video Processing.



**Dr. Gopala Krishna Mohan Pillutla** presently working as Professor in Institute of Aeronautical College of Engineering, Hyderabad. He Worked as Head of ECE Dept. , Member of BOS for ECE faculty at University Level, Chairman of BOS of EIE group at University level, Chairman of BOS of ECE faculty for JNTUCEH, Member of selection committees for Kakitaya, Nagarjuna University, DRDL and convener for Universite a Hidian committees. He has more than 50 papers in various International and National Journals and Conferences. His areas of interests are Signal Processing, Communications.



**Madhavi Latha Makkena** obtained her B.Tech, from Nagarjuna University in the year 1986. M. Tech. and Ph.D, degrees from the University of JNTU in the year 1993 and 2002 respectively. India. She is presently working as Professor in the Department of ECE, JNT University, Telangana, She is having 24 years of teaching experience, 3 years of industry experience and 14 years of research experience. She guided 16 Ph.Ds. she served as Board Of Studies for ECE & ETM board and also the board of Studies for BME\EIE\ICE\ECM board JNTUH in the year 2012 – 2014. She Convened for more than seven number of various seminars, conferences and workshop both nationally and internationally. She is a Life Member MISTE, Fellow of IETE – F169872, Member of IEEE and Chairperson for IEEE Women in Engineering Society, Hyderabad Section, 2010. She had 175 publications in various journals and conferences both nationally and internationally. She authored a book titled “Signals, Systems and Communications”. Her areas of interests are Signal, Image & Speech Processing, Biomedical Signal Processing, Wavelet Transform Applications, Mojette Transform Applications, VLSI, Analog & Mixed Signal Design Low Power VLSI Embedded Systems, DSP Processors & Controllers and Communications.