

# Modified Unsupervised Image Segmentation based on Gaussian Mixture Model for Traffic Surveillance Applications

K.Siva Nagi Reddy  
Assoc. professor, Dept of ECE,  
MSSIST-College of  
Engineering, Mylavaram,  
Vijayawada, Krishna (D.t),  
Andhra- Pradesh, India.

Bhanu Murthy Bhaskara  
Professor and Vice Dean,  
Faculty of Integrative Sciences  
and Technology,  
Quest International University,  
Perak, MALAYSIA

B.R.Vikram  
Professor, Dept of ECE,  
Vijay Rural Engineering  
College, Nizamabad,  
Nizamabad (D.t),  
Andhra- Pradesh, India.

## ABSTRACT

This paper deals with an efficient image segmentation algorithm for video images which is quite useful for video based traffic surveillance applications. It includes video segmentations, morphological operations and labeling. In the field of surveillance system, effective video object segmentation is a conveyance for video analysis and processing. It presents a new algorithm for video object segmentation i.e. unsupervised image segmentation based on Gaussian mixture model with modified EM procedure. It uses the spatial unsupervised GMM clustering technique in which the objective function is modified or the prior term is added in the Bayesian. Firstly we use EM algorithm to estimate the distribution of input image data with which the number of mixture components is automatically determined. Secondly the segmentation is arrived at by clustering each pixel into the appropriate component according to the Minimum Message Length (MML) criterion with the help of appropriate priors like Dirichlet-Normal-Wishart (DNW) prior. The proposed technique automatically decides the best number of clusters for images. The best number of clusters is obtained by using the cluster validity criterion with the help of Gaussian distribution.

**Key words:** GMM, EM, MML, DNW, Multinomial distribution, Bayesian, Fisher Information

## 1. INTRODUCTION

Many authors have presented different models for recapturing the images based on segmentation and to bring out different features such as texture, color, shape and other aspects [1-5]. Majority of the models focus mainly on two variations: Supervised and Unsupervised. In the supervised method [6] of learning, few features about the image are known before hand and the segmentation is carried out by using these features. However in the unsupervised model [7], the features inside the image are not known beforehand. In unsupervised data it is common to cluster the images so that similar data are combined together, from which the required patterns can be identified. In order to cluster the data, a number of clustering algorithms are discussed in the literature and among all these the K-means algorithm is popularly used. Clustering based image segmentation methods have been extensively used in many applications. Unsupervised clustering has high reproducibility, as its results are based on the information of image data itself and it does not require either an assumption

of the model or the distribution of image data. However pixel based clustering algorithms cannot segment color texture images properly because they depend mainly on the intensity distribution of the pixels and ignore their geometric information. Due to the noise and intensity in homogeneities introduced in imaging process, different color textures at different locations have a similar intensity appearance. As a result the segmentation results would be totally wrong without the spatial information.

A popular method to include the spatial unsupervised GMM clustering technique is to alter the objective function or by adding the prior term in the Bayesian formulation. Bayesian method is specifically suitable for analyzing limited data as it allows updating of information by combining the current information with the prior one. Bayesian classification system was used by several researchers for different applications. A recent application of Bayesian classifier [8] is in the implementation of the Bayesian training method to construct a series of hybrid Artificial Neural Network (ANN) structures to model hot rolling force prediction of real input/output data and practical expressions. In [9], the Bayesian estimation, based on some prior assumptions on the regions for the range image segmentation has been used. Image priors were modeled by a new Markov Random Field (MRF) model.

Gaussian mixture model GMM [10, 11] has been extensively used to clustering, object detection, image segmentation, marketing analysis, speaker identification and optical character recognition etc [12 - 14]. Learning a GMM includes guideline learning for estimating all the unknown guidelines and model selection for determining the number  $k$  of Gaussian components. Guideline learning is usually started under the maximum likelihood principle by an expectation maximization EM algorithm [11], [15, 16]. A common model selection approach is featured by a two-stage implementation. With the help of suitable priors like Dirichlet-Normal-Wishart prior, efforts have been made on Minimum Message Length for learning GMM with automatic model selection. MML approach minimizes a two-part message for a statement of model and a statement of data encoded by that model involving a Fisher information matrix [17, 18]. The MML algorithm for GMM with a prior is the product of independent which is developed by Wallace et al [19].

We consider GMM with conjugate prior [20], which imposes Dirichlet prior on the mixing weights and a joint Normal Wishart prior combined as DNW. The algorithm for MML [17] is adopted here, with some changes in DNW prior. In addition to morphological operations, a labeling unit based on a contour tracing technique has been presented which extracts features at low extra costs.

## 2. GAUSSIAN MIXTURE MODEL (GMM)

$$\mathcal{P}(x) = \sum_{i=1}^k \pi_i \mathcal{N}(x|\mu_i, \sigma_i) \quad (1)$$

$$\forall k: \pi_i \geq 0, \sum_{k=1}^K \pi_i = 1$$

Cost function for fitting a GMM For a point  $x_i$

$$\mathcal{P}(x_t) = \sum_{i=1}^k \pi_i \mathcal{N}(x_t|\mu_i, \sigma_i) \quad (2)$$

The likelihood of the GMM for N points (assuming independent) is

$$\prod_{t=1}^N \mathcal{P}(x_t) = \prod_{t=1}^N \sum_{i=1}^k \pi_i \mathcal{N}(x_t|\mu_i, \sigma_i) \quad (3)$$

And the (negative) log-likelihood is

$$\mathcal{L}(\theta) = -\sum_{t=1}^N \ln \sum_{i=1}^k \pi_i \mathcal{N}(x_t|\mu_i, \sigma_i) \quad (4)$$

To minimize  $\mathcal{L}(\theta)$ , differentiate first wrt  $\mu_i$

$$\frac{\partial \mathcal{L}(\theta)}{\partial \mu_i} = \sum_{t=1}^N \frac{\pi_i \mathcal{N}(x_t|\mu_i, \sigma_i)}{\sum_{j=1}^k \pi_j \mathcal{N}(x_t|\mu_j, \sigma_j)} \sigma_i^{-1} (x_t - \mu_i) \quad (5)$$

$$\text{Rearranging } \sum_{t=1}^N \gamma_{it} \mu_i = \sum_{t=1}^N \gamma_{it} x_t$$

$$\text{And hence } \mu_i = \frac{1}{r_i} \sum_{t=1}^N \gamma_{it} x_t \quad \text{weighted mean}$$

$$\text{Where } \gamma_{it} = \frac{\pi_i \mathcal{N}(x_t|\mu_i, \sigma_i)}{\sum_{j=1}^k \pi_j \mathcal{N}(x_t|\mu_j, \sigma_j)}, r_i = \sum_{t=1}^N \gamma_{it}$$

$\gamma_{it}$  Are the responsibilities of mixture component k for vector  $x_t$ . “ $r_i$ ” is the effective number of vectors assigned to component k.  $\gamma_{it}$  is not binary it lies between  $0 \leq \gamma_{it} \leq 1$

Differentiating with respect to  $\sigma_i$  gives weighted covariance

$\sigma_i = \frac{1}{N_i} \sum_{t=1}^N \gamma_{it} (x_t - \mu_i)(x_t - \mu_i)^T$  and with respect to  $\pi_i$  (enforcing the constraints that  $\sum \pi_i = 1$  with a Lagrange multiplier). Gaussian mixture model and EM algorithm with the multivariate normal wishart prior GMM [21] assumes that an observation  $X \in \mathbb{R}^d$  is distributed as a linear mixture of k multivariate Gaussian distributions, i.e.

$$\mathcal{P}(x_t|\theta_i) = \sigma_i \pi_i \mathcal{N}(x_t|\mu_i, \sigma_i) = \mathcal{N}(x_t|\mu_i, \sigma_i)$$

$$= \frac{\sigma_i^{\frac{n}{2}}}{2\pi^{\frac{nd}{2}}} \exp \left[ -\frac{1}{2} (x_t - \mu_i) \sigma_i (x_t - \mu_i)^T \right]$$

For single variants k multivariate Gaussian distributions,

$$\begin{aligned} \mathcal{P}(x_t|\theta_i) &= \sigma_i \pi_i \mathcal{N}(x_t|\mu_i, \sigma_i) = \mathcal{N}(x_t|\mu_i, \sigma_i) \\ &= \frac{\sigma_i^{\frac{1}{d}}}{2\pi^{\frac{d}{2}}} \exp \left[ -\frac{1}{2} (x_t - \mu_i) \sigma_i (x_t - \mu_i)^T \right] \end{aligned} \quad (6)$$

Where  $\sum_{i=1}^k \pi_i = 1$ , Parameters :  $\theta = \{\pi, \mu, \sigma\}$ ,

The parameters  $\sigma, \pi$  fixed, i.e., only estimate  $\mu$

### Expectation Maximization (EM) Algorithm:

**Step 1 Expectation:** Compute responsibilities using current parameters  $\pi_i, \sigma_i$  for  $i=1, 2, 3, \dots, K$  and  $t=1, 2, 3, \dots, N$ ,

$$\gamma_{it} = \frac{\pi_i^{\text{old}} \mathcal{N}(x_t|\mu_i^{\text{old}}, \sigma_i^{\text{old}})}{\sum_{j=1}^k \pi_j \mathcal{N}(x_t|\mu_j^{\text{old}}, \sigma_j^{\text{old}})}, r_i = \sum_{t=1}^N \gamma_{it} \quad (7)$$

**Step 2 Maximization:** Re-estimate parameters using computed responsibilities

$$\pi_i^{\text{new}} = \frac{r_i}{\sum_{j=1}^k r_j} \quad (8)$$

$$\mu_i^{\text{new}} = \frac{1}{r_i} \sum_{t=1}^N \gamma_{it} x_t \quad (9)$$

$$\sigma_i^{\text{new}} = \frac{1}{r_i} \sum_{t=1}^N \gamma_{it} (x_t - \mu_i^{\text{old}})(x_t - \mu_i^{\text{old}})^T \quad (10)$$

Repeat until convergence

Using EM the parameters representing the Gaussian mixture are found. In this we use fixed number of iterations .usually the large gradient in the Mixture components learning occur with in the first iterations, while as long as the EM reaches its convergence only small refinements occur. Then, it is possible to maintain a real time process without relying on the unknown convergence time of the original EM, with a negligible lack of accuracy.

The EM present several draw backs .particularly, methods that require mixture estimates for various numbers of components may converge to the boundary of the parameters space. This means that, e.g. for Gaussian mixture, the covariance matrix’s may become singular. A way to address this problem is to use adequate priors for the parameters. We use the dirchilet normal wishart prior, because it promotes configurations where the prior probability is either 0 or 1. .besides, it also speeds-up the components with respect to the other priors like Jeffrey’s prior or minimum entropy prior, therefore to a faster convergence

### 3. DIRICHLET-NORMAL-WISHART PRIOR

Consider GMM in Eq. (6) with the prior  $p(\theta)$  on its parameters  $\theta = \{\pi_j, \mu_j, \sigma_j\}$ . The conjugate prior [20], which imposes a Dirichlet prior on the mixing weights

$\pi_j$ , a joint Normal-Wishart prior that consists of a Normal distribution on  $\mu_j$  conditional on the Wishart distributed  $\sigma_j$  denote this conjugate prior as DNW. The DNW prior is expressed by

$$P(\theta) = DP(\pi|\alpha, \beta) \prod_{i=1}^k \mathcal{N}(x_i|\mu_i, \sigma_i/\epsilon)W(\sigma_i|\varphi, \vartheta) \quad (11)$$

#### 3.1 Fisher information matrix

Fisher information [22, 23] is a main concept in the theory of statistical inference and it is defined as: Let  $X = (X_1, \dots, X_N)$  be a random sample, and let  $P(X_N/\theta)$  denote the probability density function for some model of the data, which has parameter vector  $\theta = (\theta_1, \theta_2, \dots, \theta_k)$ . Then the Fisher information matrix in  $F(\theta)$  of sample size  $n$  is given by the  $k \times k$  symmetric matrix whose  $i, i^T$  element is given by the covariance between first partial derivatives of the log-likelihood,  $F(\theta)$  is the Fisher information matrix

$$F(\theta) = -E \left[ \frac{\partial^2 \ln p(X_N|\theta)}{\partial(\theta) \partial(\theta)^T} \right] \quad (12)$$

Due to the difficulty of getting an exact analytical expression of  $F(\theta)$ . By following [17] to approximate  $F(\theta)$  by the following block-diagonal complete data Fisher information matrix  $F_c(\theta)$  such that

$$F_c(\theta) - F(\theta) \text{ is positive definite,} \\ F_c(\theta) = N \times \text{blockdiag}[\pi_1 F_c(\theta_1), \pi_2 F_c(\theta_2) \dots \pi_k F_c(\theta_k), F_c(\pi)] \quad (13)$$

$$F_c(\theta_i) \propto |\sigma_i|^{-d}, |F_c(\pi)| \propto \prod_{i=1}^k \pi_i^{-1} \quad (14)$$

Where  $F_c(\theta_i)$  is the Fisher information of the  $k^{\text{th}}$  Gaussian component and  $F_c(\pi)$  is the Fisher information of the mixing weights. It follows from Eqs. (12)– (14) that is given as

$$|F(\theta)| \propto N^{k(\rho+1)} \prod_{i=1}^k \pi_i^{\rho-1} \prod_{i=1}^k |\sigma_i|^{-d} \\ P(\theta) \propto N^{\frac{k(\rho+1)}{2}} \prod_{i=1}^k \pi_i^{\frac{\rho-1}{2}} \prod_{i=1}^k |\sigma_i|^{-\frac{d}{2}} \quad (15)$$

Where  $\rho = d + d(d+1)/2$  is the number of free parameters in each Gaussian component.

#### 3.2 Dirichlet distribution

The Dirichlet distribution [24] forms our first step toward understanding the DPM model. The Dirichlet distribution is a multi-parameter generalization of the Beta distribution and defines a distribution over distributions, i.e. the result of sampling a Dirichlet is a distribution on some discrete probability space. Let  $\pi = (\pi_1, \pi_2, \dots, \pi_n)$  be a probability distribution on the discrete space  $X = (x_1, x_2, x_3, \dots, x_n)$ .  $P(X = x_i) = \pi_i$  where  $X$  is a random variable in the space. The Dirichlet distribution on  $\pi$  is given by the formula

$$DP(\pi|\beta, \alpha) = \frac{\Gamma(\alpha)}{\prod_{i=1}^k \Gamma(\alpha\beta_i)} \left( \prod_{i=1}^k \pi_i^{\alpha\beta_i-1} \right), \quad (16)$$

$$\alpha > 0, \sum_{i=1}^k \beta_i = 1, \beta_i \geq 0, \beta = [\beta_1, \beta_2, \dots, \beta_k]^T$$

Where  $\beta_i = (\beta_1; \beta_2, \dots, \beta_n)$  is the base measure defined on and is the mean value of  $\pi$ , and  $\alpha$  is precision parameter that say show concentrated the distribution is around  $\beta$ . Both  $\pi$  and  $\beta$  are normalized, i.e. sum to unity, since they are proper probability distributions.  $\alpha$  can be regarded as the number of pseudo-measurements observed to obtain  $\beta$ , i.e. the number of events relating to the random variable  $X$  observed a priori. The greater the number of pseudo-measurements the more our confidence in  $\beta$ , and hence, the more the distribution is concentrated around  $\beta$ .

#### 3.3 Wishart distribution

The Wishart distribution [25] can be characterized by its probability density function, as follows. Let  $\varphi$  be a  $d \times d$  symmetric matrix of random variables that is positive definite. Let  $\vartheta$  be a (fixed) positive definite matrix of size  $d \times d$ . Then, if  $\vartheta \geq d$ ,  $\sigma$  has a Wishart distribution with  $d$  degrees of freedom if it has a probability density function given by

$$W(\sigma_i|\varphi, \vartheta) = \frac{|\varphi|^{\frac{\vartheta}{2}} |\sigma_i|^{\frac{\vartheta-d-1}{2}} 2^{-\frac{d\vartheta}{2}}}{\pi^{\frac{d(d-1)}{4}} \prod_{j=1}^d \Gamma(\frac{\vartheta-j+1}{2})} \exp \left\{ -\frac{1}{2} \text{Tr}(\sigma_i \varphi) \right\} \\ \varphi > 0, \vartheta > d - 1 \quad (17)$$

Where  $\Gamma(\cdot)$  the multivariate gamma is function,

#### 4. MINIMUM MESSAGE LENGTH

MML is an information theoretic restatement of Occam's razor. Among different models, the one generating the shortest overall code length is regarded most likely to be the best, Where the code length is

$$\text{Length}(X, \theta) = \text{Length}(X|\theta) + \text{Length}(\theta),$$

That consists of a statement of the model. As introduced in Refs. [17, 18], MML seeks a model  $\theta$  by maximizing the objective function:

$$J_{\text{MML}}(\theta) = \ln p(X_N|\theta) + \ln p(\theta) - \frac{1}{2} \ln |I(\theta)| \quad (18)$$

In the above equation  $I(\theta)$  is Fisher information matrix given in Equation (12).

From equations (16)(17) & (12) to consider the following  $J_{\text{MML}}(\theta)$  with the DNW prior given in Eq. (11) will gives as follows:

$$J_{\text{MML}}^{\text{DNW}}(\theta, \zeta) = \ln p(X_N|\theta) + \ln DP(\pi|\alpha, \beta) + \mathcal{N}(x_t|\mu_t, \sigma_t/\epsilon) \\ + \sum_{i=1}^k \ln W(\sigma_i|\varphi, \vartheta) - \frac{1}{2} \ln |F(\theta)| \\ = \ln P(X_N|\theta) + \sum_{i=1}^k \left( \alpha\beta_i - \frac{\rho+1}{2} \right) \ln \pi_i + \frac{\vartheta}{2} \sum_{i=1}^k \ln |\sigma_i| - \\ \frac{\epsilon}{2} \sum_{i=1}^k (\mu_i - m_i)^T \sigma_i (\mu_i - m_i) - \frac{1}{2} \sum_{i=1}^k \text{Tr}(\sigma_i \varphi) + \ln \Gamma(\alpha) - \\ \sum_{i=1}^k \ln \Gamma(\beta_i) + \frac{kd}{2} \ln(2\pi) + \frac{kd}{2} - \frac{kd\vartheta}{2} \ln 2 - k \ln \Gamma\left(\frac{\vartheta}{2}\right) - \\ \frac{k\rho}{2} \ln N \quad (19)$$

Fix the hyper parameters as  $\alpha = k/2$ , each  $\beta_i = 1/k$ ,  $\beta = 0$ ,  $\varphi = 0$  and  $\vartheta = 0$

## 5. SEGMENTATION ALGORITHM OF GMM WITH MML- DNW

The segmentation consists of two steps. The first step is mixture estimation. The method is used to estimate the mixture parameters and determine the number of Gaussian components. The second step is pixel clustering. The segmentation is carried out by assigning each pixel into a proper cluster according to the Maximum likelihood (ML) estimation is:

**E-step:** for  $i=1, 2, 3 \dots k$ , and  $\gamma_{it}$  and  $r_i$  by equation (7)

**M-step:** update the parameters

$$\begin{aligned} \theta &= \{\pi_i, \mu_i, \sigma_i\}_{i=1}^k \\ \pi_i^{\text{new}} &= \frac{r_i}{\sum_{j=1}^k r_j} \\ r_i &= \sum_{t=1}^N \gamma_{it} + \alpha^{\text{old}} \beta_i^{\text{old}} - \frac{\rho+1}{2}, \\ \text{Where } \rho &= d + \frac{d(d+1)}{2}, \end{aligned} \quad (20)$$

$$\begin{aligned} \mu_i^{\text{new}} &= \frac{1}{r_i + \varepsilon^{\text{old}}} [\sum_{t=1}^N \gamma_{it} x_t + \varepsilon^{\text{old}} m_i^{\text{old}}], \\ o_{it} &= x_t - \mu_i^{\text{old}}, \end{aligned} \quad (21)$$

$$\sigma_i^{\text{new}} = \frac{1}{r_i + \vartheta^{\text{old}}} [\sum_{t=1}^N \gamma_{it} o_{it} o_{it}^T + \beta^{\text{old}} (\mu_i^{\text{old}} - m_i^{\text{old}})(\mu_i^{\text{old}} - m_i^{\text{old}}) + \varphi^{\text{old}}] \quad (22)$$

Update the prior hyper-parameters  $\{m_i\}$  in the following cases:

General case  $\mu_i^{\text{old}} = m_i^{\text{old}}$

Special case (constrain each  $m_i = m$ ):  $\forall i$ ,

$$m_i^{\text{new}} = (\sum_{i=1}^k \sigma_i^{\text{old}})^{-1} (\sum_{i=1}^k \sigma_i^{\text{old}} \mu_i^{\text{old}}),$$

Then update the prior hyper-parameters  $\{\alpha, \beta, \varepsilon, \varphi, \vartheta\}$ :

$$\begin{aligned} \alpha_i^{\text{new}} &= \frac{\alpha_i^{\text{old}} \eta \delta \alpha_i}{\sum_{j=1}^k (\alpha_j^{\text{old}} + \eta \delta \alpha_j)}, \\ \delta \alpha_i &= \ln \pi_i^{\text{old}} - \psi(\beta^{\text{old}} \alpha_i^{\text{old}}) + \psi(\beta^{\text{old}}) \end{aligned} \quad (23)$$

$$\beta^{\text{new}} = \beta^{\text{old}} + \eta \delta \beta, \quad \delta \beta = \sum_{i=1}^k \alpha_i^{\text{old}} \delta \alpha_i \quad (24)$$

$$\varepsilon^{\text{new}} = \frac{kd}{\sum_{i=1}^k (\mu_i^{\text{old}} - m_i^{\text{old}})^T \sigma_i^{\text{old}} (\mu_i^{\text{old}} - m_i^{\text{old}})} \quad (25)$$

$$\begin{aligned} \varphi^{\text{new}} &= k \vartheta^{\text{old}} \left( \sum_{i=1}^k \sigma_i^{\text{old}} \right), \\ \vartheta^{\text{new}} &= \vartheta^{\text{old}} + \eta \delta \vartheta (\vartheta^{\text{old}}, \zeta^{\text{old}}), \end{aligned} \quad (26)$$

$$\begin{aligned} \delta \vartheta (\theta, \zeta) &= \\ \sum_{i=1}^k \ln |\sigma_i| + k \ln |\varphi| - k_d \ln 2 - k \sum_{j=1}^d \psi \left( \frac{\vartheta+1-j}{2} \right) \end{aligned} \quad (27)$$

Where  $\zeta = \{\alpha, \beta, \varepsilon, \varphi, \vartheta\}$ ,  $i=1, 2 \dots k$  do if  $\pi_i \rightarrow 0$  then discard component  $i$ , let  $k = k - 1$  and  $\tau = \tau + 1$ ; The MML objective function as  $U_{\text{MML}}(\tau)$  by Eq. (19).

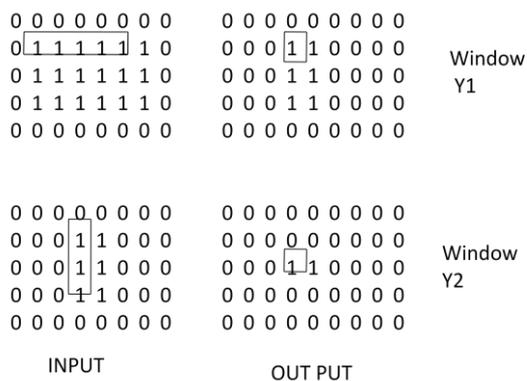
## 6. MORPHOLOGY

Erosion and Dilation (E&D) are the two foundations in mathematical morphology, since most morphologic operations can be broken down into these two basic operations [26]. For example, operations such as opening, closing, gradient, and skeletonization are performed with these two base functions. To increase the overall performance of the process, it is also desirable that the size of the Structuring Element (SE) can be changed, With a flexible SE size comes the ability to compensate for different types of noise and to sort out certain types of objects in the mask, e.g., high and thin objects (standing humans) or wide and low objects (side view of cars). In this paper  $X$  will represent the binary input image and  $Y$  the structuring element. If the SE is both reflection invariant, i.e.  $Y = \hat{Y}$ , and decomposable, i.e. Then the following two equations can be derived

$$\begin{aligned} Y &= Y_1 \oplus Y_2 \\ X \oplus Y &= (X \oplus Y_1) \oplus Y_2 = ((X' \ominus Y_1) \ominus Y_2)^1 \\ X \ominus Y &= X \ominus (Y_1 \oplus Y_2) = (X \ominus Y_1) \ominus Y_2 \end{aligned} \quad (28)$$

Where  $\hat{\cdot}$  is bit inversion,  $\oplus$  is dilation, and  $\ominus$  is erosion. In addition, for a SE to be reflection invariant it has to be symmetric both in respect to the  $x$  and  $y$  direction, e.g., a square or a circle. However, one common class of SEs that is both decomposable and reflection invariant is rectangles of ones. This type of SE is well suited for the opening and closing operations that are needed in this system.

An example of erosion with a decomposed SE is shown in Fig. 1. were the SE is decomposed into  $Y_1$  and  $Y_2$ . The input is first eroded with  $Y_1$  and then  $Y_2$ . The first position of  $Y_1$  and  $Y_2$  that produce a one is shown in the Figure, together with location in the output of this one. With a decomposed SE, the number of comparisons per output is decreased from the number of ones in  $Y$  to the number of ones in  $Y_1$  plus  $Y_2$ . To perform dilation, the input  $X$  and the result is inverted. Hence, the same inner kernel can be used for both operations. With rectangular Structuring Element of ones, erosion can be performed as a summation followed by a comparison. To perform binary erosion, bits in  $X$  that lies directly below the current position of  $Y$  are added and compared to the size of  $Y$ . If the sum is equal to the size of  $Y$  the result is one otherwise zero. When combining this with decomposition, the summation can be broken up into two stages, where the first stage compares the number of one's under  $Y_1$  to the width of  $Y_1$  and the second stage compares the number of one's under  $Y_2$  in the result from the first stage to the height of  $Y_2$ .



**Fig 1: Input and output to an erosion operation were the SE is decomposed into B1 and B2.**

### 7. LABELING

After segmentation, a binary frame is produced containing connected clusters of pixels that represent different objects. Assuming that noise has been removed by the morphology unit, the frame now only contains objects of interest that can be tracked and classified. To be able to separate and distinguish between these clusters, they have to be identified, i.e. labeled. Various labeling algorithms have been proposed and a survey of various algorithms can be found in [27]. A common property for all these algorithms are that they are memory access intense. Furthermore, all algorithms have to handle the same obstacle, i.e. label collisions. In a binary image, a typical label collision occurs when a u shaped object is encountered. Since an image typically is scanned from top to bottom and from left to right, it is not possible to know that the two pillars in the u is part of the same object until the bottom part of the u is encountered. Two common methods to handle this problem are:

- **Equivalence Table Based Algorithms** – Two scans with a corresponding equivalence table.
- **Contour Tracing Based Algorithms** – A single scan with contour tracing. Equivalence table based algorithms [28] scan through the memory writing every label collision into an equivalence table. The first label scan is completed on the fly as the frame is written into the memory by comparing each pixel with its neighbors to the left and above. After the first scan all pixels are assigned a label and all collisions have been detected. The second scan resolves all collisions and reduces the number of labels per cluster to one. Contour tracing based algorithms [29] is a technique that requires one global scan together with some additional random memory accesses for the contour tracing procedure. The major advantage is that label collisions will never occur since when an unlabeled cluster is encountered, the contour of that cluster is labeled immediately and every pixel within a label is regarded as part of the same cluster. If a cluster with a labeled contour is encountered, the scan proceeds without modification continuing until an unlabeled pixel is reached, restarting the contour tracing procedure, or the last pixel is reached. If a cluster has a hole inside its contour, this hole will not be traced. Every pixel between two labels can therefore be

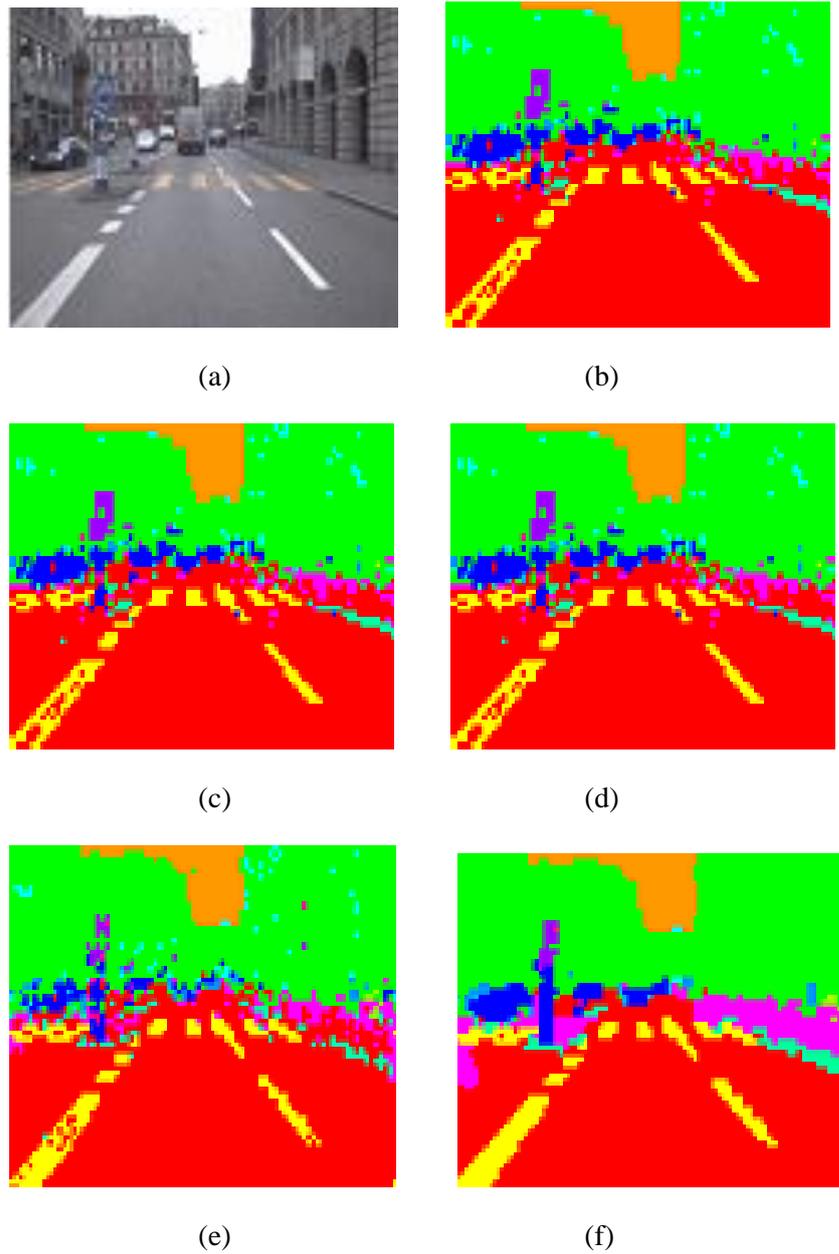
considered a part of an object, cluster holes are filled. Extracting properties by post processing in the tracking stage or by a general purpose processor can be time consuming, thus every property that can be extracted by an algorithm without inferring additional hardware complexity should be seen as an advantage.

In comparison, both types of labeling algorithms can extract maximum and minimum coordinates and the number of pixels in each cluster. In addition, the contour tracing algorithm can calculate moments (center of gravity) and fill holes inside the clusters. Thus, the contour tracing label algorithm is more suitable to be used in a automated surveillance system.

### 8. RESULTS & CONCLUSIONS

A visual traffic surveillance application oriented, probabilistic approach based large scale moving objects strategy has been presented in this paper. The modified proposal of an unsupervised color image segmentation method based on its Gaussian mixture model with DNW prior is presented. The color image that has been observed was considered as a mixture of multivariate normal densities. The segmentation is processed by ML estimation with MML Bayesian and DNW prior. The advantage of our method lies in two features, first by using the K means algorithm we successfully initialized the problem of EM algorithm. Second advantage of the algorithm is that the amount of segmentation region is automatically determined by the MML Bayesian prior with DNW, so the algorithm is called unsupervised. This approach shows good performance in terms odd adaptability, accuracy and robustness. Experimental results show that this method has better segmentation results when compared to other unsupervised methods.

The segmentation results in Fig. 2 shows that the proposed system determine accurately the object borders between the regions that are characterized by similar color compositions and it is compared with MML-Jeff, FCR, and CTM. The comparison is based on four performance measures namely, Probabilistic Rand Index (PRI), the Variation of Information (VOI), Global Consistency measure (GCE) and the boundary displacement error,(BDE) following [30,31]. PRI calculates the consistency between the computed segmentation and the ground truth. The VOI measures the amount of randomness in segmentation which cannot be explained by other. The GCE measures the extent to which one segmentation map can be viewed as are refinement of segmentation. BDE measures the aver-age displacement error of one boundary pixel and the closest boundary pixels in the other segmentation. Table .1 shows the performance measures for the obtained segmentation results by MML-Jeff, FCR, and CTM and the MML-DNW on the traffic images presented in Figures 1. Higher is better for PRI and lower is better for VOI, GCE and BDE.



**Fig 2:** a) original image b) GMM with MML with Jeffries prior Method c) FCR Method d) CTM Method e) GMM with MML with DNW f) after morphological operation.

**Table 1.** Performance of proposed algorithm with existing algorithms for the given images

METHOD	PRI	GCE	VOI	BDI
MML-DNW	0.892	0.156	1.45	5.32
MML-JEF	0.885	0.173	1.42	5.27
FCR	0.782	0.211	2.10	7.54
CTM	0.76	0.178	1.93	7.62

## 9. REFERENCES

- [1] Reed T. and du Buf J.M.H. (1993) CVGIP Image Understanding, vol. 57, pp. 359–372.
- [2] Li J., Najmi A., and Gray R.M. (2000) IEEE Trans.Sig. Proc., vol. 48, no. 2, pp. 517–533.
- [3] Povlow B.R. And Dunn S.M. (1995) IEEE Trans. Patt. Anal. Mach. Intell., vol. 17, no. 10, pp. 1010–1014.
- [4] G. Fan and X.-G. Xia (2000) Proc. 34th Asilomar Conf. Signals, Systems and Computers.
- [5] Choi H. And Baraniuk R.G. (1999) Proc. SPIE Tech.Conf. Mathematical Modeling, Bayesian Estimation and Inverse Problems, pp. 306–320.
- [6] Prasad Reddy PVGD et al (2007) International Journal of Computer Science and Network Security, VOL.7 No.4.
- [7] Safarinejadian, B., M.B. Menhaj and M. Karrari, 2009. Distributed data clustering using expectation Maximization algorithm. J. Applied Sci., 9: 854-864. DOI: 10.3923/jas.2009.854.864
- [8] Moussaoui, A., Y. Selaimia and H.A. Abbassi, 2006. Hybrid hot strip rolling force prediction using a Bayesian trained artificial neural network and analytical models. Am. J. Applied Sci., 3: 1885-1889. DOI: 10.3844/ajassp.2006.1885.1889
- [9] Mazouzi, S. and M. Batouche, 2007. Range image segmentation by randomized region growing and Bayesian edge regularization. J. Comput. Sci., 3: 310-317. DOI: 10.3844/jcssp.2007.310.317.
- [10] Constantinopoulos C, Titsias M K. Bayesian feature and model selection for Gaussian mixture models. IEEE Trans-actions on Pattern Analysis and Machine Intelligence, 2006, 28(6): 1013–1018
- [11] Redner R, Walker H. Mixture densities, maximum likelihood and the EM algorithm. SIAM Review, 1984, 26(2): 195–239
- [12] Engel A, den Broeck C P L V. Statistical Mechanics of Learning. New York: Cambridge University Press, 2001
- [13] Constantinopoulos C, Likas A. Unsupervised learning of Gaussian mixtures based on variational component splitting. IEEE Transactions on Neural Networks, 2007, 18(3): 745–755
- [14] Verbeek J, Vlassis N, Krose B. Efficient greedy learning of Gaussian mixture models. Neural Computation, 2003, 15(2):469–485
- [15] Xu L, Jordan M I. On convergence properties of the EM algorithm for Gaussian mixtures. Neural Computation, 1996, 8(1):129–151
- [16] McLachlan G J, Krishnan T. The EM Algorithm and Extensions (Wiley Series in Probability and Statistics). New York: Wiley-Interscience, 2007
- [17] Wallace C, Boulton D. An information measure for classification. The Computer Journal, 1968, 11(2): 185–194
- [18] Wallace C S, Dowe D L. Minimum message length and Kolmogorov complexity. The Computer Journal, 1999, 42(4):270–283
- [19] Figueiredo M A F, Jain A. K. Unsupervised learning of finite mixture models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(3): 381–396
- [20] Gelman A, Carlin J B, Stern H S, Rubin D B. Bayesian Data Analysis. 2nd ed. Texts in Statistical Science. Boca Raton: Chapman & Hall/CRC, 2003
- [21] Constantinopoulos C, Titsias M K. Bayesian feature and model selection for Gaussian mixture models. IEEE Trans-actions on Pattern Analysis and Machine Intelligence, 2006, 28(6): 1013–1018
- [22] Lehman, E. L., & Casella, G. (1998). Theory of Point Estimation (2nd edition). New York, NY: Springer.
- [23] Schervish, M. J. (1995). Theory of Statistics. New York: Springer.
- [24] D. Blei, A. Ng, and M. Jordan, Latent Dirichlet Allocation, Journal of Machine Learning Research, 3:993-1022, (2003).
- [25] Anderson, T. W. (2003) an introduction to multivariate statistical analysis, 3rd ed. John Wiley and Sons, New York
- [26] J. Serra, Image Analysis and Mathematical Morphology, Vol 1. Academic Press, 1982.
- [27] W. Kesheng, O. Ekow, and S. Arie, “Optimizing connected components labeling algorithms,” in SPIE Int. Symposium on Medical Imaging, San Diego, CA, USA, Feb 2005.
- [28] K. Suzuki, H. Isao, and S. Noboru, “Linear-time connected-component labeling based on sequential local operations,” Journal of CVIU, vol. 89, pp. 1–23, jan 2003.
- [29] F. Chang, C. J. Chen, and C. J. Lu, “A linear-time component-labeling algorithm using contour tracing technique,” Journal of CVIU, vol. 93, pp.206–220, feb 2004.
- [30] Ma, Y., H. Derksen, W. Hong and J. Wright, 2007. Segmentation of multivariate mixed data via lossy data coding and compression. IEEE Trans. PAMI., 29: 1546-1562. DOI: 10.1109/TPAMI.2007.1085
- [31] Mignotte, M., 2008. Segmentation by fusion of histogram based k-means clusters in different color spaces. IEEE Trans. Image Process. 17: 780-787. DOI: 10.1109/TIP.2008.920761

## 10. AUTHERS PROFILE

**SIVANAGIREDDY.KALLI** received his B.E. degree in ECE the M.E. degree in VLSI DESIGN, and pursuing the Ph.D. in image processing algorithms implementing on FPGA. He had 9 years teaching experience .He was a, Assistant professor, with Department of ECE in repudiated institutions from 2003to2006 and Associate professor, in ARJUN College of Technology Science, HYDERABAD .At present working as Associate professor and In-charge Principal in MSSIST-College of Engineering in Mylavaram, Vijayawada, A.P. He was published 6 International papers and his research interests include Digital image processing VLSI Design, FPGA, Analog and digital communications.

**BHANU BHASKARA** received his Ph.D degree in Digital Signal Processing. He has over three decades of teaching, research and industry experience. Currently he is serving as the Professor and Vice Dean of the Faculty of Integrative

Sciences and Technology at Quest International University Perak, Malaysia.