

Adaptive And Automatic Target Detection Without Post-Processing For Indoor Video Sequence

Charan K¹, Shreedevi B A²

Dept. of Information Science Engineering
RV College of Engineering

Abstract

The proposed work is to develop Background Subtraction methodology for indoor video sequences for color images. The indoor sequence may cause many problems in the scene due to lightning changes in it, and it may affect the target detection based on intensity of focus light. The proposed method first applies Homomorphic filtering algorithm to nullify the effect of illumination changes directly on color images, then using extended Gaussian Mixture model build a background model, using Euclidian distance as measure will detect moving foreground objects in indoor sequences. Background Subtraction is first method for any Automatic Target Detection system to work effectively. The proposed algorithm is directly works on color images and compared with existing algorithms and achieving some good improvements. Results illustrate that the proposed work achieves good results on varying illumination in indoor sequence.

1. Introduction

Background Subtraction is one of the most widely used techniques to detect moving objects from static scene [1]-[3]. Since background subtraction is a low-level task, it should consider two aspects: Accuracy and Computational resources (time and memory). First, its accuracy is critical because the output of the background subtraction is used for other high tasks,

such as tracking and recognition. Second, computational resources used for background subtraction are critical since the resources remaining after this low-level task should be used for high level tasks, and is preferable as a means of implementing this task in real-time embedded systems such as smart cameras [4], [5]. Therefore, it is important for the background subtraction method to obtain high accuracy and low resource requirements at the same time.

Background Subtraction is a typical approach in indoor sequence for foreground detection, which detects foreground regions by evaluating the difference between the input image and a reference background image. In background subtraction, tracking of objects has to be done through any of the background models. There are many background subtraction methods but many of them got failed to present the correct results for indoor sequence because of illumination changes.

Gaussian Mixture Model is considered to perform the operation but it is not able to adapt the effectiveness of illumination changes which can occur in each scene. To overcome this adaptive Gaussian Mixture Model has been proposed and its frequently adapt to the changes which is occurring in each scene. After performing improved version also there is so much of distortion and produced false results. To overcome this Homomorphic Filter is used which can easily extract

the intensity, lightness and brightness of each scene, so it helps to nullify the illumination component and produces good results for each scene. Secondly, Adaptive Gaussian Mixture Model has to be performed for Homomorphic results to get the foreground object.

The paper is organized as follows. In next section we list some related work. In section 3 the Improved GMM approach from [3] is reviewed. In sections 4 proposed method, we present to improve the algorithm. In section 5 we present some experiments.

2. Related work

It is well known that the appearance of an image can be severely affected by illumination. The problem is that the illumination causes larger variations in images. A very promising for compensating illumination changes is homomorphic filtering, based on illumination reflectance model [6], [7]. The procedure is well known in image analysis and processing but was rarely addressed.

2.1 Background Modelling

This step uses the new video frames in order to calculate and update the background model. The main aim of developing a background model is that it should be robust against environmental changes in the background, but sensitive enough to identify all moving objects. Frame difference is considered the simplest form of background subtraction. The method involves the subtraction of previous pixel frame with current pixel frame and if the value obtained from this greater than a set threshold then it is considered as the foreground. Advantages are less complex, the method can easily and quickly adapt to changes and able to distinguish the background from foreground much affectively. Disadvantages are not robust, applicable

when with fixed camera and fails to only applicable when with fixed camera.

3. ImprovedGMM

A Gaussian Mixture Model (GMM) [17] is a parametric probability density function represented as a weighted sum of Gaussian component densities. GMMs are commonly used as a parametric model of the probability distribution of continuous measurements or features in a biometric system, such as vocal-tract related spectral features in a speaker recognition system. GMM parameters are estimated from training data using the iterative Expectation Maximization (EM) algorithm or Maximum A Posteriori (MAP) estimation from a well-trained prior model.

Gaussian Mixture Model is considered in the proposed system, literature says that GMM got failed to adapt the illumination changes and also to identify the foreground object, to resolve these problems Gaussian Mixture Model was used for background subtraction[17].

The illumination in the scene could change suddenly (switching on/off lights). A new object could be brought into the scene or a present object removed from it. In order to adapt to changes update the training set by adding new samples and discarding the old ones[17]. Choose a reasonable time period T and at time t ,

$$X_T = \{x^{(t)}, \dots, x^{(t-T)}\} \quad (1)$$

For each new sample update the training data set X_T and re-estimate

$$\hat{P}(\vec{x}|X_T, BG) \quad (2)$$

Among the samples some values that belong to the foreground objects and should denote this estimate as

$$P(\vec{x}^{(t)}|X_T, BG + FG) \quad (3)$$

GMM with M components

$$P(\vec{x}^{(t)}|X_T, BG + FG) = \sum_{m=1}^M \hat{\pi}_m N(\hat{x}; \hat{\mu}_m, \hat{\sigma}_m^2 I) \quad (4)$$

Where $\hat{\mu}_1, \dots, \hat{\mu}_M$ are the estimates of the means and $\hat{\sigma}_1, \dots, \hat{\sigma}_M$ are the estimates of the variances that describe the Gaussian components. The covariance matrices are assumed to be diagonal and the identity matrix I has proper dimensions. The mixing weights denoted by $\hat{\pi}_m$ are non-negative and add up to one. Given a new data sample $\vec{x}^{(t)}$ at time t the recursive update equations are.

$$\hat{\pi}_m \leftarrow \hat{\pi}_m + \alpha (o^{(t)}_m - \hat{\pi}_m) \quad (5)$$

$$\hat{\mu}_m \leftarrow \hat{\mu}_m + o^{(t)}_m \left(\frac{\alpha}{\hat{\pi}_m} \right) \vec{\delta}_m \quad (6)$$

$$\hat{\sigma}_m^2 \leftarrow \hat{\sigma}_m^2 + o^{(t)}_m \left(\frac{\alpha}{\hat{\pi}_m} \right) \vec{\delta}_m^T \vec{\delta}_m - \hat{\sigma}_m^2 \quad (7)$$

Where $\vec{\delta}_m = \vec{x}^{(t)} - \hat{\mu}_m$ instead of the time interval T that was mentioned above, here constant α describes an exponentially decaying envelope that is used to limit the influence of the old data. The same notation having the mind that approximately $\alpha=1/T$. For a new sample the ownership $o^{(t)}_m$ is set to 1 for the 'close' zero component with largest $\hat{\pi}_m$ and the others are set to 0. The squared distance from the m-th component is calculated as

$$D_m^2(\vec{x}^{(t)}) = \vec{\delta}_m^T \vec{\delta}_m / \sigma_m^2 \quad (8)$$

If there are no 'close' components a new component is generated with

$$\hat{\pi}_{M+1} = \alpha, \hat{\mu}_{M+1} = \vec{x}^{(t)} \quad (9)$$

And

$$\hat{\sigma}_{M+1} = \sigma_0 \quad (10)$$

Where σ_0 is some appropriate initial variance. If the maximum number of component is reached discard the component with smallest $\hat{\pi}_m$

The foreground object will be represented by some additional clusters with small weights $\hat{\pi}_m$. Therefore approximate the background model by the first B largest clusters

$$P(\vec{x}^{(t)}|X_T, BG) = \sum_{m=1}^M \hat{\pi}_m N(\hat{x}; \hat{\mu}_m, \sigma_m^2 I) \quad (11)$$

If the components are stored to have descending weights $\hat{\pi}_m$

$$B = \text{argmin}_b (\sum_{m=1}^b \hat{\pi}_m > (1 - C_f)) \quad (12)$$

Where C_f is a measure of maximum portion of the data that can belong to foreground objects without influencing the background model. For example, if a new object comes into a scene and remains static for some time it will probably generate an additional stable cluster. Since the old background is occluded the weight π_{B+1} of the new cluster will be constantly increasing. If the object remains static long enough, its weight becomes larger than C_f and it can be considered to be part of the background. By looking this equation

$$\hat{\pi}_m \leftarrow \hat{\pi}_m + \alpha (o^{(t)}_m - \hat{\pi}_m)$$

conclude that the object should be static for approximately $\log(1 - C_f) = \log(1 - \alpha)$ frames.

The weight π_m describes how much of the data belongs to the m-th component of the GMM. It can be regarded as the probability that a sample comes from the m-th component and in this way $\pi_m - s$ define an underlying multinomial distribution. Let us assume that, t data samples and each of them belongs to one of the components of the GMM. Let us also assume that the number of samples that belong to the m-th component is

$$n_m = \sum_{i=1}^t o^{(i)}_m \quad (13)$$

Where $o_m^{(i)}$ -s are defined in the previous section. The maximum weights are constrained to sum up to one. Take this into account by introducing the lagrange multiplier λ . The maximum likelihood estimate follows from:

$$\frac{\theta}{\theta \hat{\pi}_m} (\log L + \lambda (\sum_{m=1}^M (\hat{\pi}_m - 1))) = 0 \quad (14)$$

After getting rid of

$$\lambda: \hat{\pi}_m^{(t)} = \frac{n_m}{t} = \frac{1}{t} \sum_{i=1}^t o_m^{(i)} \quad (15)$$

The estimate from t samples is denoted as $\hat{\pi}_m^{(t)}$ and it can be rewritten in recursive form as a function of the estimate $\hat{\pi}_m^{(t-1)}$ for t-1 samples and the ownership $o_m^{(t)}$ of the last sample.

$$\hat{\pi}_m^{(t)} = \hat{\pi}_m^{(t-1)} + 1/t(o_m^{(i)} - \hat{\pi}_m^{(t-1)}) \quad (16)$$

If the influence of the new samples is fix by fixing 1/t to $\alpha = 1/T$ get the update equation. This fixed influence of the new samples means that rely more on the new samples and the contribution from the old samples is downweighted in an exponentially decaying manner as mentioned before.

Prior knowledge for multinomial distribution can be introduced by using its conjugate prior, the Dirichlet prior

$$P = \prod_{m=1}^M \pi_m^{c_m} \quad (17)$$

The coefficients C_m have a meaningful interpretation. For the multinomial distribution, the C_m presents the prior evidence for the class m the number of samples that belong to that class a priori. Negative prior evidence means, accept that the class m exists only if there is enough evidence from the data for the existence of this class. This type of prior is also related to minimum message length criterion that is used for selecting proper models for given data. The MAP solution that includes the mentioned prior follows from

$$\frac{\theta}{\theta \hat{\pi}_m} (\log L + \log P + \lambda (\sum_{m=1}^M (\hat{\pi}_m - 1))) = 0 \quad (18)$$

Where

$$P = \sum_{m=1}^M \hat{\pi}_m^{-c} \quad (19)$$

$$\hat{\pi}_m^{(t)} = \frac{1}{K} (\sum_{i=1}^t o_m^{(i)} - c) \quad (20)$$

Where

$$K = \sum_{m=1}^M (\sum_{i=1}^t o_m^{(i)} - c) = t - M_c \quad (3.30)$$

Rewrite as

$$\hat{\pi}_m^{(t)} = \frac{\hat{\pi}_m - c/t}{1 - M_c/t} \quad (21)$$

Where $\hat{\pi}_m = \frac{1}{t} \sum_{i=1}^t o_m^{(i)}$ are the ML estimate and the bias from the prior is introduced through c/t. The bias decreases for larger data sets. However, if a small bias is acceptable then keep it constant by fixing c/t to $C_T = c/T$ with some large T. This means that the bias will always be the same as if would have been for a data set with T samples. It is easy to show that the recursive version with fixed $c/t = C_T$ is given by

$$\begin{aligned} \hat{\pi}_m^{(t)} = & \\ & \hat{\pi}_m^{(t-1)} + 1/t \left(\frac{o_m^{(t)}}{1 - M C_T} - \hat{\pi}_m^{(t-1)} \right) - \\ & 1/t \frac{C_T}{1 - M C_T} \end{aligned} \quad (22)$$

Since expect usually a few components M and C_T is small let assume $1 - M C_T \approx 1$. As mentioned set 1/t to α and get the final modified adaptive update equation

$$\hat{\pi}_m \leftarrow \hat{\pi}_m + \alpha (o_m^{(i)} - \hat{\pi}_m) - \alpha C_T \quad (23)$$

This equation is used instead of this equation $\hat{\pi}_M \leftarrow \hat{\pi}_M + \alpha (o_M^{(t)} - \hat{\pi}_M)$. After each update need to be normalize π_m -s so that it is up to one. Then start with GMM with one component centered on the first sample and new components are added as mentioned in the previous. The Dirichlet prior with negative weights will suppress the components that are not supported by the data and discard the component m when its weight π_m becomes negative. This also ensures that the mixing weights stay non-negative. For a chosen $\alpha = 1/T$ we

could require that at least $c = 0.01 * T$ samples support a component and we get $C_T = 0.01$.

3.3 Foreground detection

This method is used for detecting foreground object is highly dependent on background model. It compares the video frame with the background frame, and identifies the foreground object from the frame.

3.4 Data Validation

Finally, this step eliminated any pixels which are not connected to the image. It involves the process of improving the foreground mask based on the information obtained from the outside background model.

Most background models lack three main points which are ignoring any correlation between the neighboring pixels, the rate of adaption may not match the moving speed of the foreground object, non-stationary pixels, from moving leavers or shadow cast by moving objects are at times mistaken for true foreground objects.

4. Proposed System

In background subtraction, tracking of objects has to be done through any of the background models. There are many background subtraction methods but many of them got failed to present the correct results for indoor sequence because of illumination changes. To overcome this adaptive Gaussian Mixture Model has been proposed and its frequently adapt to the changes which is occurring in each scene. After performing improved version also there is so much of distortion and produced false results. To overcome this Homomorphic Filter is used which can easily extract the intensity, lightness and brightness of each scene, so it helps to nullify the illumination component and produces good results for each scene. Secondly, Adaptive Gaussian Mixture Model has to be performed for Homomorphic results to get the foreground object

5. Experiments

The Homomorphic Filter enhances the original color image by filtering the lightness or any darkness in an image. The filtered image is transformed back to RGB but the resulting RGB image has identical matrices for the Red, Green and Blue planes, so the image displays as shades of gray.

4.1 Background Subtraction

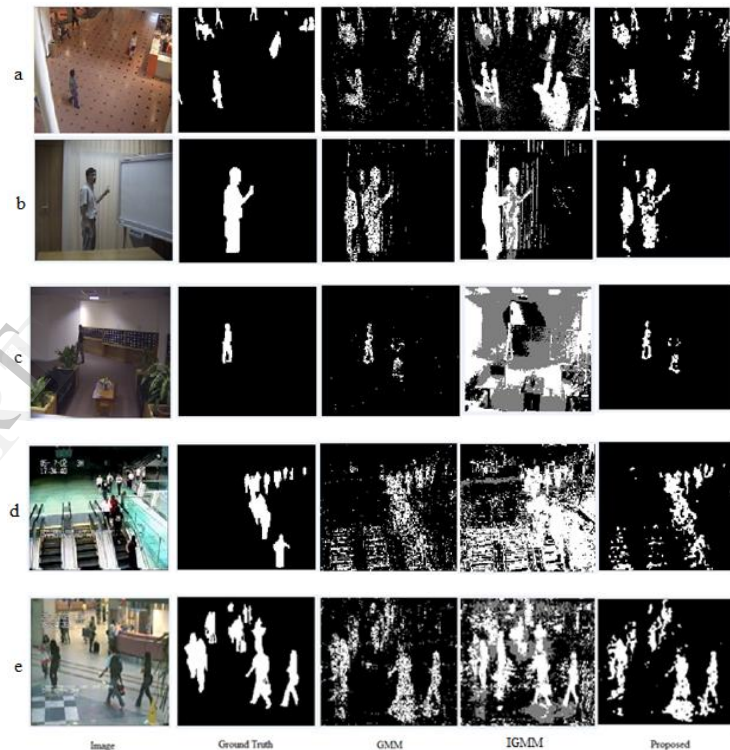


Fig 1: Foreground segmentation results

In figure 1(a), the frame size is handled as 320*256 and some peoples are walking in it in the mall as taken as input with some lightning changes in it.

In figure 1(b), the frame size is handled as 160*128 and one person is teaching in the class room with more focus of light on him taken as input.

In figure 1(c), the frame size is handled as 160*128 and one person is moving towards light from dark so it is highly illuminated frame taken as input.

In figure 1(d), the frame size is handled as 160*130 and more peoples are moving in elevator with some lightning changes on it taken as input.

In figure 1(e), the frame size is handled as 176*144 and more peoples are moving in mall with more focus of light on them taken as input.

The above mentioned frames are taken as input for background subtraction with some lightning changes on it. The Homomorphic Filter is applied to eliminate the lightning changes in it and later these enhanced frames will be given as input for background subtraction.

Image: Sample images to perform Background Subtraction.

Ground truth: Estimated Ground truth for given images.

GMM: Foreground results for Gaussian Mixture Model.

Improved GMM: Foreground results for extended GMM.

Proposed: Foreground results for proposed approach.

The proposed method was evaluated and compared with GMM and Improved GMM using different type of illumination images. The experimental results are presented to show that the proposed methods can achieve promising performance for illumination changes in background subtraction and foreground object extraction. The proposed system detects and tracks the moving objects exactly. The background scene is modelled using a set of consecutive frames. The object pixels are segmented out from its background followed by morphological process to eliminate noisy pixels for producing better results.

4.2 Performance Evaluation

Three criteria are used for performance evaluation, such as

- False Positive Ratio

- False Negative Ratio
- True Positive Ratio

	FPR			FNR			TPR		
	GMM	IGMM	Proposed	GMM	IGMM	Proposed	GMM	IGMM	Proposed
1	1913	1628	1496	3292	13442	2732	2335	2620	2752
2	580	755	568	888	2489	686	1032	857	1044
3	117	127	100	161	5897	144	177	167	194
4	611	233	502	1764	7693	947	989	1367	1098
5	1783	928	1590	1644	3586	1339	2546	3401	2739

Table 1: Resulted values for performance evaluation.

In the above table the results are tabulated for the performance of resulting images. Where FP is the number of false positives, FN is the number of false negatives and TP is the number of true positives.

Recall is defined by:

$$\text{Recall} = \frac{tp}{tp + fn} \tag{24}$$

	Recall			Precision		
	GMM	IGMM	Proposed	GMM	IGMM	Proposed
1	41.49	16.31	50.182	54.96	61.67	64.78
2	53.75	25.61	60.34	64.01	53.16	64.76
3	52.36	2.75	57.39	60.204	56.802	65.986
4	35.92	15.09	53.69	61.81	85.43	68.625
5	60.76	48.67	67.16	58.812	78.56	63.27

Table 2: Tabulated results for precision and recall

In the above table the results are tabulated for recall and precision. Recall gives the percentage of detected true positives as compared to the total number of true positives in the ground truth, while precision gives the percentage of detected true positives as compared to the total number of items detected.

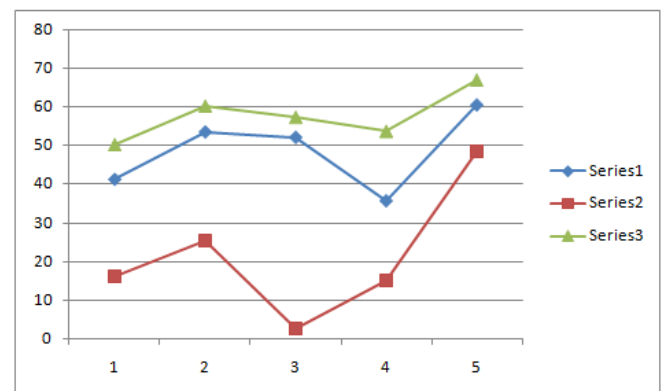
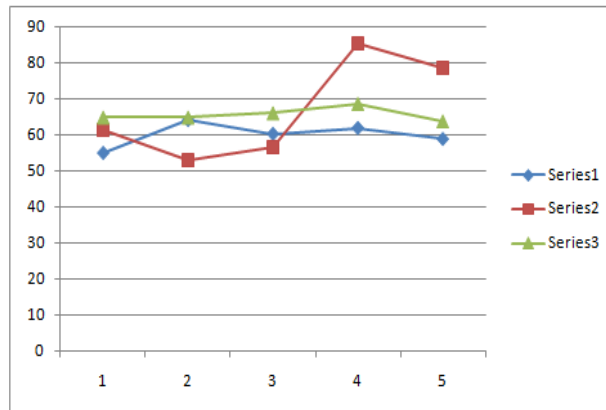


Fig 2:Graph for Recall

The above graph is plot through resulting values of recall, series 1 is for GMM, series 2 is for IGMM and series 3 is for proposed method.

**Fig 3:** Graph for Precision

The above graph is plot through resulting values of Precision, series 1 is for GMM, series 2 is for IGMM and series 3 is for proposed method.

6. Conclusion

The proposed approach is capable of detecting motion and extracting object information which involves human as object has been described. The strength of the proposed technique is robust against illumination changes in indoor sequence. Main contribution towards the success of the detection in proposed method is because of nullifying the effectiveness of illumination later use the enhanced images to extract the foreground object through the proposed background model.

In future we can consider the outdoor environment also to nullify the effectiveness of distortion in the scene using background subtraction. We need to add many more features to the proposed system for outdoor environment because it is more complicated than indoor environment so it is necessary to find out the best way to get the solution for outdoor environment.

References

- [1]M. Piccardi, "Background subtraction techniques: A review," in Proc. IEEE Int. Conf. Syst., Man Cybern., Oct. 2004, pp. 3099–3104.
- [2]R. J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam, "Image change detection algorithms: A systematic survey," IEEE Trans. Image Process., vol. 14, no. 3, pp. 294–307, Mar. 2005.
- [3]Y. Benezeth, P. M. Jodoin, B. Emile, H. Laurent, and C. Rosenberger, "Review and evaluation of commonly-implemented background subtraction algorithms," in Proc. IEEE Int. Conf. Patt. Recog., Dec. 2008, pp. 1–4.
- [4]B. Rinner and W. Wolf, "An introduction to distributed smart cameras," Proc. IEEE, vol. 96, no. 10, pp. 1565–1575, Oct. 2008.
- [5]A. N. Belbachir, Smart Cameras. New York: Springer, 2009.
- [6]R.C. Gonzalez, R.E. Woods, "Digital Image Processing", Prentice Hall, Upper Saddle River, NJ, 2002
- [7]S.E. Umbaugh, "Computer Imaging: Digital Image Analysis and Processing", CRC Press, Florida, 2005
- [8]C. Stauffer and W. E. L. Grimson. Adaptive background mixture models for real-time tracking. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1999, 2:252, 1999. URL: <http://doi.ieeecomputersociety.org/10.1109/CVPR.1999.784637>.
- [9] A.M. Elgammal, D. Harwood, and L.S. Davis, "Non-Parametric Model for Background Subtraction". ECCV, 2000, pp. 751–767.

- [10] P. Noriega, and O. Bernier, "Real Time Illumination Invariant Background Subtraction Using Local Kernel Histograms". *BMVC*, 2006, volume 3, pp. 979–988..
- [11] D. J. Ketcham, R. Lowe, and W. Weber, "Seminar on image processing," in *Real-Time Enhancement Techniques*, 1976, pp. 1–6. Hughes Aircraft.
- [12] R. Hummel, "Image enhancement by histogram transformation," *Comp. Graph. Image Process.*, vol. 6, pp. 184–195, 1977.
- [13] V. T. Tom and G. J. Wolfe, "Adaptive histogram equalization and its applications," *SPIE Applicat. Dig. Image Process. IV*, vol. 359, pp. 204–209, 1982.
- [14] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. H. Romeny, J. B. Zimmerman, and K. Zuiderveld, "Adaptive histogram equalization and its variations," *Comp. Vis. Graph. Image Process.*, vol. 39, no. 3, pp. 355–368, 1987.
- [15] Russ, J., "The Image Processing Handbook," 4th ed., CRC Press, 2002
- [16] R.C. Gonzalez, R.E. Woods, "Digital Image Processing", Prentice Hall, Upper Saddle River, NJ, 2002
- [17] S.E. Umbaugh, "Computer Imaging: Digital Image Analysis and Processing", CRC Press, Florida, 2005
- [17] Zoran Zivkovic, "Improved Adaptive Gaussian Mixture Model for Background Subtraction", In Proc. ICPR, 2004