

Shadow Suppression using RGB and HSV Color Space in Moving Object Detection

Shailaja Surkutlawar

EXTC

Vivekananda education society of information
technology
Mumbai, India

Prof. Ramesh K Kulkarni

EXTC

Vivekananda education society of information
technology
Mumbai, India

Abstract— Video-surveillance and traffic analysis systems can be heavily improved using vision-based techniques to extract, manage and track objects in the scene. However, problems arise due to shadows. In particular, moving shadows can affect the correct localization, measurements and detection of moving objects. This work aims to present a technique for shadow detection and suppression used in a system for moving visual object detection and tracking. The major novelty of the shadow detection technique is the analysis carried out in the HSV color space to improve the accuracy in detecting shadows. This paper exploits comparison of shadow suppression using RGB and HSV color space in moving object detection and results in this paper are more encouraging using HSV colour space over RGB colour space.

Keywords— Shadow detection; HSV color space; RGB color space.

I. INTRODUCTION

Surveillance systems have wide demand in public areas, such as airports, subways, entrance to buildings. In this context, reliable detection of moving objects is most critical requirement for the surveillance systems. To detect a moving object, a surveillance system usually utilizes background subtraction. The key of background subtraction is the background model. In the moving object detection process, one of the main challenges is to differentiate moving objects from their cast shadows.

Moving cast shadows are usually misclassified as part of the moving object making the following stages, such as object classification or tracking, to perform inaccurate. The Gaussian mixture model (GMM) [1] represented the statistics of one pixel over time can cope with multi-modal background distributions. However, a common problem for this approach is to find the right balance between the speed at which the model adapts to changing background, and the stability.

The shadow points and the object points share two important visual features: motion model and detectability. Since the most common techniques for foreground object detection in dynamic scene are inter-frame difference or background suppression, all the moving points of both objects and shadows are detected at the same time.

Moreover, shadow points are usually adjacent to object points and with the more commonly used segmentation techniques shadows and objects are merged in a single blob. These aspects cause two important drawbacks: The former is that the object shape is falsified by shadows and all the measured geometrical properties are affected by an error (that varies during the day and when the luminance changes). This affects both the classification and the assessment of moving object position (normally given by the shape centroid), as, for instance, in traffic control systems that must evaluate the trajectories of vehicles and people on a road. The second problem is that the shadows of two or more objects can create a false adjacency between one another, thus detecting them as merged in a single affects many higher level surveillance tasks such as counting and classifying individual objects in the scene. In order to avoid the drawbacks due to shadows, a new technique of shadow suppression using HSV colour space is proposed.

The paper is organized as follows: Section II deals with the background subtraction using Gaussian Mixture Model to classify the pixels as background or foreground by thresholding the difference between the background image and the current image, Section III deals with Post processing techniques for suppressing shadow using HSV and RGB colour space, Section IV discusses the experimental results of shadow suppression techniques. Finally, the conclusion is given in Section V.

II. BACKGROUND SUBTRACTION

In the model of Mixture of Gauss [1] [4] [5], the background is not a single frame without any moving objects. Gaussian Mixture Model (GMM) is thought to be one of the best background modeling methods and works well when gradual changes appear in the scene [2] [3]. The GMM method models the intensity of each pixel with a mixture of k Gaussian distributions. The probability that a certain pixel has a value X_t at time can be written as

$$P(X_t) = \sum_{i=1}^k \omega_{i,t} \cdot \eta(X_t, \mu_{i,t}, \Sigma_{i,t}) \quad (1)$$

Where k is the number of distributions (currently, 3 to 5 is used), $\omega_{i,t}$ is the weight of the k^{th} Gaussian in the mixture at

time t and $\eta(X_t, \mu_{i,t}, \Sigma_{i,t})$ the Gaussian probability density function.

$$\eta(X_t, \mu_{i,t}, \Sigma_{i,t}) = \frac{1}{(2\pi)^3/2 |\Sigma_{i,t}|^{1/2}} e^{\left\{ \frac{-1(X_t - \mu_{i,t})^T \Sigma_{i,t}^{-1} (X_t - \mu_{i,t})}{2} \right\}} \quad (2)$$

Where, $\mu_{i,t}$ is the mean value and $\Sigma_{i,t}$ is the covariance of the k^{th} Gaussian at time t . For computational reasons, the covariance matrix is assumed to be of the form

$$\Sigma_{k,t} = \sigma^2 I \quad (3)$$

Where σ is the standard deviation. This assumes that the red, green, and blue pixel values are independent and have the same variance, allowing us to avoid a costly matrix inversion at the expense of some accuracy.

Thus, the distribution of recently observed values of each pixel in the scene is characterized by a mixture of Gaussians. A new pixel value will, in general, be represented by one of the major components of the mixture model and used to update the model.

However, it fails when there are sharp changes, such as sudden illumination changes or sudden partial changes in the background. To tackle this problem, some improvement has been made in recent researches. In [6], every frame is processed on pixel level, region level and frame level with color and gradient information to overcome the problem caused by sudden illumination changes based on GMM. In [7], a hierarchical GMM using state models without temporal correlation on different scales is proposed to handle sharp changes. Zivkovic presented an improved GMM algorithm automatically fully adapting to the scene, by choosing the number of components for each pixel in an online procedure [8] [9], which leads to big improvement in reduced processing time and slight improvement in segmentation result.

If the pixel process could be considered a stationary process, a standard method for maximizing the likelihood of the observed data is expectation maximization. Unfortunately, each pixel process varies over time as the state of the world changes, therefore an approximate method which essentially treats each new observation as a sample set of size 1 and uses standard learning rules to integrate the new data.

If lighting changes occurred in a static scene, it would be necessary for the Gaussian to track those changes. If a static object was added to the scene and was not incorporated in to the background until it had been there longer than the previous object, the corresponding pixels could be considered foreground for arbitrarily long periods. This would lead to accumulated errors in the foreground estimation, resulting in poor tracking behavior. These factors suggest that more recent observations may be more important in determining the Gaussian parameter estimates. Since there is a mixture model for every pixel in the image, implementing an exact Expectation maximization algorithm on a window of recent data would be costly.

Instead, we implement an on-line K-means approximation. Every new pixel value X_t , is checked against the existing k Gaussian distributions, until a match is found. A match is defined as a pixel value within 2.5 standard deviations of a distribution. GMM algorithm can be summarized as:

- Initialize each pixel of the scene with k Gaussian distributions
- Every new pixel value X_t , is checked against the existing Gaussian distributions until a match is found.
- A match is defined as a pixel value within 2.5 standard deviations of a distribution.
- If none of k -distributions match current pixel value, least probable distribution is go out.
- A new distribution with current value as mean value, an initially high variance, and low prior weight, is entered.
- Prior weights of k distributions at time adjusted as

$$\omega_{k,t} = (1 - \alpha)\omega_{k,t-1} + \alpha(M_{k,t}) \quad (4)$$

- μ_t and σ for unmatched distributions remain the same.
- Parameters of distribution matching new observation are updated as :

$$\mu_t = (1 - \rho)\mu_{t-1} + \rho(X_t) \quad (5)$$

$$\sigma_t^2 = (1 - \rho)\sigma_{t-1}^2 + \rho(X_t - \mu_t)^T (X_t - \mu_t) \quad (6)$$

$$\rho = \alpha \eta(X_t | \mu_{t-1}, \sigma_{t-1}) \quad (7)$$

- Gaussians are ordered by the value of ω/σ
- 1st B distributions are chosen as background model, where

$$B = \text{arg min}_b (\sum_{k=1}^b \omega_k > T) \quad (8)$$

III. SHADOW SUPPRESSION TECHNIQUE

Shadows are due to the occlusion of light source by an object in the scene. In particular, that part of the object not illuminated is called self-shadow, while the area projected on the scene by the object is called cast shadow [10]. This last one is more properly called moving cast shadow if the object is moving.

A. Normalized RGB color space

The Normalized RGB space aims to separate the chromatic components from the brightness component. The red, green and blue channel can be transformed to their normalized counterpart by using the formulae

$$l = R + G + B, r = R/l, g = G/l, b = B/l \quad (9)$$

When $l \neq 0$ and $r = g = b = 0$ otherwise.

One of these normalized channels is redundant, since by definition r , g , and b sum up to 1. Therefore, the Normalized RGB space is sufficiently represented by two chromatic components r and g and a brightness component 1. Normalized RGB suffers from a problem inherent to the normalization namely noise sensor or compression noise at low intensities results in unstable chromatic components.

Under the consideration of saving computational cost, RGB space based method proposed by Horprasert in [4] is adopted. The basic idea in [4] is that shadow has similar chromaticity but lower brightness. For a given observed pixel value I_i , a brightness distortion, α_i , and a color distortion CD_i , is calculated by,

$$\alpha_i = \arg_{\alpha_i} \min(I_i - \alpha_i E_i)^2 \quad (10)$$

$$CD_i = \|I_i - \alpha_i E_i\| \quad (11)$$

Where E is the expected chromaticity line, α_i equals 1 if the brightness of the given pixel in the current frame is the same as in the background image. α_i , is less than 1 if it is darker and greater than 1 if it becomes brighter than the expected brightness. Then, the criteria for shadow pixels simply becomes,

$$\begin{cases} \tau_a < \alpha_i < 1 \\ CD_i < \tau_{CD} \end{cases} \quad (12)$$

In [11], τ_a and τ_{CD} are predefined thresholds $\tau_a = 0.7$ and $\tau_{CD} = 5$, in our experiments.

A. HSV color space

In literature, many works have been published on shadow detection topic. Jiang and Ward [10] extract both self-shadows and cast shadows from a static image. They use a three level processes approach:

1. The low level process extracts dark regions by thresholding input image.
2. The middle level process detects features in dark regions, such as the vertexes and the gradient of the outline of the dark regions and uses them to further classify the region as penumbra (part of the shadow where the direct light is only partially blocked by the object), self-shadow or cast shadow.
3. The high level process integrates these features and confirms the consistency along the light directions estimated from the lower levels.

It addresses the problem of segmentation of moving objects, hence an approach for detecting moving cast shadows on the background, without computing static shadows is defined. In [12], the authors detail the shadow handling system using signal processing theory. Thus, the appearance of a point belonging to a cast shadow can be described as:

$$S_k(x, y) = E_k(x, y) \rho_k(x, y) \quad (13)$$

Where S_k is the image luminance of the point of coordinate (x, y) at time instant t . $E_k(x, y)$ is the irradiance and it is computed as follows:

$$E_x(x, y) = \begin{cases} C_A + C_P \cos \angle(N(x, y), L) & \text{illuminate} \\ C_A & \text{shadowed} \end{cases} \quad (14)$$

Where C_A and C_P are the intensity of the ambient light and of the light source, respectively, L the direction of the light source and $N(x, y)$ is object surface normal. $\rho_k(x, y)$ is the reflectance of the object surface. In [12], some hypotheses on the environment are outlined:

- I. strong light source
- II. static background (and camera)
- III. planar background

Most of the papers take implicitly into account these hypotheses. In fact, typically the first step computed for shadow detection is the difference between the current frame and a reference image, as in [12], or a reference frame, typically named background model [13][14][15][16]. Difference $D_k(x, y)$ can be written as:

$$D_k(x, y) = S_{k+1}(x, y) - S_k(x, y) \quad (15)$$

Let us consider that a previously illuminated point is covered by a cast shadow at frame $k + 1$. According to the hypothesis 2 in [12] of a static background, reflectance $\rho_k(x, y)$ of the background does not change with time, thus we can assume that

$$\rho_{k+1}(x, y) = \rho_k(x, y) = \rho(x, y) \quad (16)$$

$$D_k(x, y) = \rho(x, y) C_P \cos \angle(N(x, y), L) \quad (17)$$

Thus, if hypothesis 1 in [12] holds, C_P in eq.17 is high. Summarizing, if hypotheses 1 and 2 in [12] hold, difference in eq. 6 is high in presence of cast shadows covering a static background. This implies that shadow points can be obtained by thresholding the frame difference image eq. 17 detects not only shadows, but also foreground points. In [13] Kilger uses a background suppression technique to find the moving objects and moving cast shadows in the scene. Then, for each object, it exploits the information on date, time and heading of the road computed by its system to choose whether to look for vertical or horizontal edges to separate shadows from objects.

In [17], a statistical posterior estimation of the pixel probabilities of membership to the class of background, foreground or shadow points, authors use three sources of information: local, based on the assumption that the appearance of a shadowed pixel can be approximated using a linear transformation of the underlying pixel appearance, according with the fact that the difference of eq. 17 should be positive; spatial, which iterates the local computation by re-computing the a-priori probabilities using the a-posteriori probabilities of the neighborhood; temporal, which predicts the position of shadows and objects from previous frames, therefore adapting the a-priori probabilities. The approach in [12] exploits the local appearance change due to shadow by computing the ratio $R_k(x, y)$ between the appearance of the

pixel in the actual frame and the appearance in a reference frame:

$$R_k(x, y) = \frac{S_{k+1}(x,y)}{S_k(x,y)} \quad (18)$$

That can be rewritten as ratio between irradiance and reflectance by using eq. 13 and eq. 16 as

$$R_k(x, y) = \frac{E_{k+1}(x,y)}{E(x,y)} \quad (19)$$

If static background point is covered by a shadow, we have:

$$R_k(x, y) = \frac{C_A}{C_A + C_P \cos \angle (N(x,y), L)} \quad (20)$$

This ratio is less than 1. In fact, the angle between $N(x, y)$ and L is in range between $-\pi/2$ to $\pi/2$ therefore the Cos function is always positive. Moreover, due to hypothesis 3, we can assume $N(x, y)$ as spatially constant in a neighborhood of the point, as background is supposed planar in neighborhood.

In [12], authors exploit the spatial constancy of N to detect shadows by computing the variance in a neighborhood of the pixel of the ratio $R_k(x, y)$: a low variance means that assumption 3 holds, then they mark that pixel as "possible shadow", eq. 20 can be seen as the ratio between the luminance after and before shadow appears. In a similar way, Davis et al. [11] [14] define a local assumption on the ratio between shadow and shadowed point luminance. This is based on the hypothesis that shadows darken the covered point, as eq. 20 and the considerations above confirm.

IV. EXPERIMENTAL RESULTS

Original frames for the experiments showed in Fig. a, Fig. b, Fig. c. On this original frames Background subtraction using GMM is applied, as a result background pixels, foreground pixels and some shadow pixels (falsely segmented as foreground pixels) shown as black and white respectively in Fig. d, Fig. e, Fig. f. Post processing techniques for shadow suppression using HSV and RGB color space applied on Fig. d, Fig. e, Fig. f, results shown in Fig. g, Fig. h, Fig. i using RGB color space and Fig. j, Fig. k, Fig. l using HSV color space. Results show that shadow suppression is better using HSV as compared to RGB color space.

V. CONCLUSION

Moving objects detection and segmentation is a fundamental step in many applications based on vision. Mixture of Gaussians is the frequently used method to subtracting moving objects from background. But its results are not good enough in some cases. In this paper, a post-processing method is proposed to solve this problem. The results with more complete boundaries provided by the color clustering is used to verify the outputs of mixture of Gaussians, and thus two possible false segmentations can be corrected effectively. Moving shadow suppression using RGB

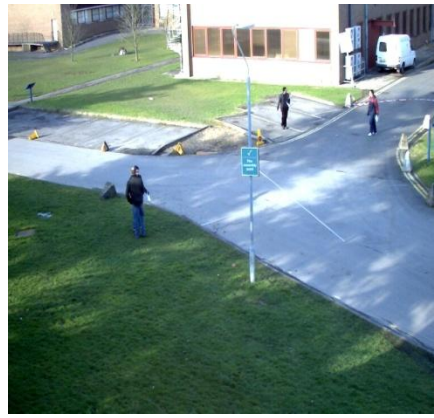
and HSV colour spaces and small region median filter are also adopted. This paper compare shadow suppression results using RGB and HSV colour space and found that results of HSV are good over RGB colour space.

REFERENCES

- [1] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Ft. Collins, 1999: pp.246-252.
- [2] C. Stauffer, and W. E. L. Grimson, "Learning patterns of activity using real-time tracking," IEEE Trans. on PAMI, Vol. 22, No. 8,2000 pp. 747-757.
- [3] P. Kumar, and A. Mittal, "Study of robust and intelligent surveillance in visible and multimodal framework," informatica, Vol. 31, 2007 pp. 447-461.
- [4] Horprasert T, Harwood D, Davis L S. "A Statistical Approach for Real-time Robust Background Subtraction and Shadow Detection," in Proceedings of IEEE ICCV'99 Frame-Rate Workshop, 1999, pp.1-19.
- [5] KaewTraKulPong P., Bowden R, "An Improved Adaptive Background Mixture Model for Real time Tracking with Shadow Detection," in proceedings of 2nd European Workshop on Advanced Video Based Surveillance Systems, Sept 2001, Pages:1-5.
- [6] O.Javed, K.Shafique, and M .Shah., "A hierarchical approach to robust background subtraction using color and gradient information," In Workshop on Motion and Video Computing, Dec. 2002, pp. 22–27.
- [7] Y.D.Sun, and B. Z. Yuan, "Hierarchical GMM to handle sharp changes in moving object detection," in Electronics Letters, Vol. 40, No. 13, 2004, pp. 801-802.
- [8] Z.Zivkovic, "Improved adaptive gaussian mixture model for background subtraction," in proceedings ICPR, 2004.
- [9] Z. Zivkovic and F.vander Heijden, "Efficient adaptive density estimation per image pixel for the task of background subtraction," Pattern Recognition Letters, Vol. 27, No. 7, 2006, pp. 773-780.
- [10] C. Jiang and M.O. Ward, "Shadow identification," in Proceedings of IEEE Int'l Conference on Computer Vision and Pattern Recognition, , 1992, pp. 606–612.
- [11] I. Haritaoglu, D. Harwood, and L.S. Davis, "W4: Real-time surveillance of people and their activities," in proceedings of IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 8, Aug. 2000, pp. 809–830.
- [12] J. Stauder, R. Mech, and J. Ostermann, "Detection of moving cast shadows for object segmentation," in proceedings of IEEE Transactions on Multimedia, vol. 1, no. 1, Mar. 1999, pp. 65–76.
- [13] M. Kilger, "A shadow handler in a video-based real-time traffic monitoring system," in Proceedings of IEEE Workshop on Applications of Computer Vision, 1992, pp. 11–18.
- [14] A. Elgammal, D. Harwood, and L.S. Davis, "Non-parametric model for background subtraction," in Proceedings of IEEE ICCV'99 FRAME-RATE Workshop, 1999.
- [15] T. Horprasert, D. Harwood, and L.S. Davis, "A Statistical approach for real-time robust background subtraction and shadow detection," in Proceedings of IEEE ICCV'99 FRAME-RATE Workshop, 1999.
- [16] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, "Statistical and knowledge-based moving object detection in traffic scene," in Proceedings of IEEE Int'l Conference on Intelligent Transportation Systems, Oct. 2000, pp. 27–32.
- [17] I. Mikic, P. Cosman, G. Kogut, and M.M. Trivedi, "Moving shadow and object detection in traffic scenes," in Proceedings of Int'l Conference on Pattern Recognition, Sept. 2000.
- [18] N. Herodotou, K.N. Plataniotis, and A.N. Venetsanopoulos, "A color segmentation scheme for object-based video coding," in Proceedings of the IEEE Symposium on Advances in Digital Filtering and Signal Processing, 1998, pp. 25–29.



a) Original Frame 1



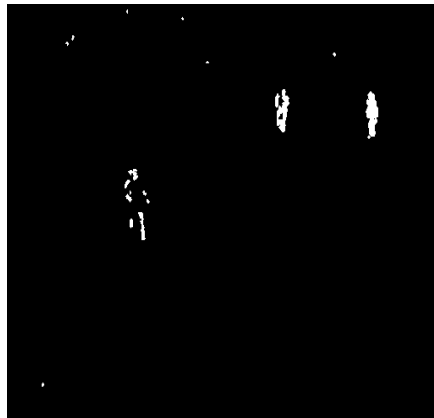
b) Original Frame 2



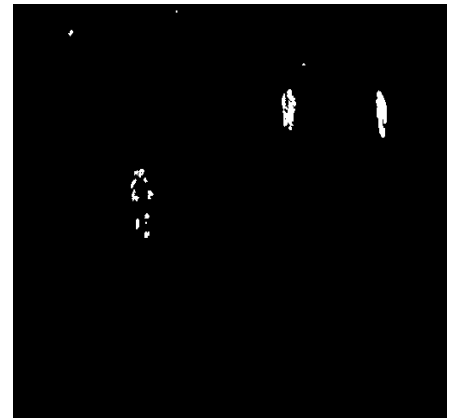
c) Original Frame 3



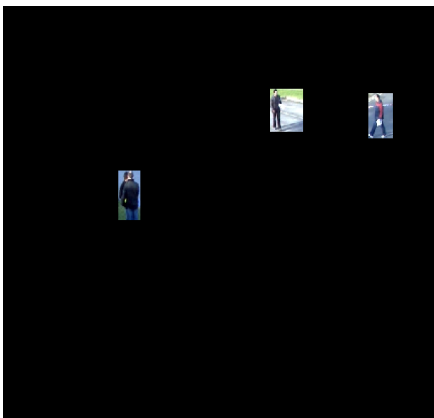
d) Shadow Detection of Frame 1



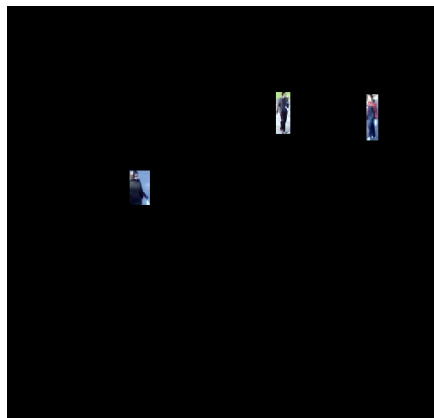
e) Shadow Detection of Frame 2



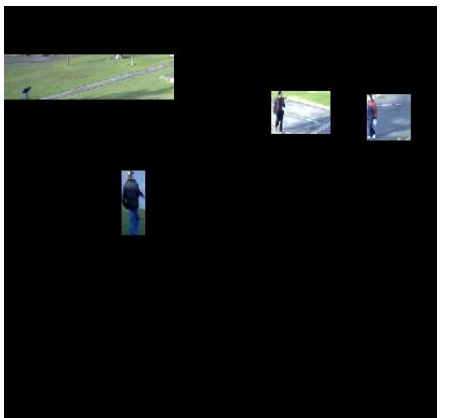
f) Shadow Detection of Frame 3



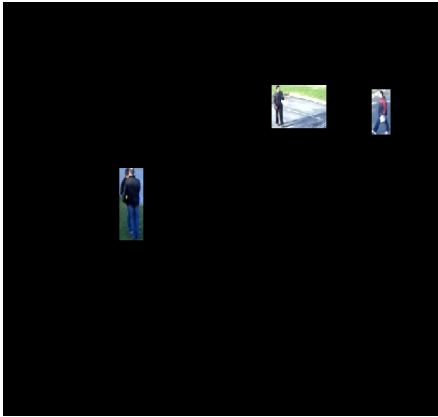
j) HSV Result Of Frame 1



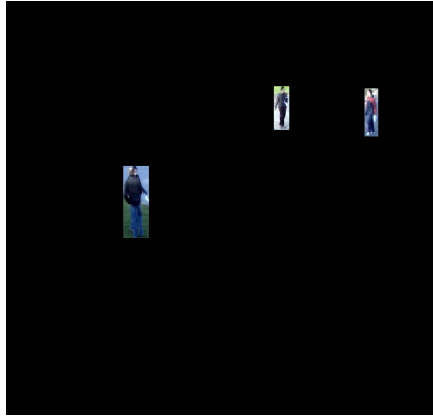
k) HSV Result Of Frame 2



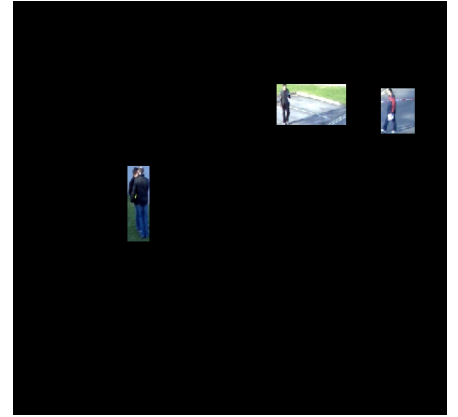
l) HSV Result Of Frame 3



g) RGB Result Of Frame 1



h) RGB Result Of Frame 2



i) RGB Result Of Frame 3