

# Retinex Foreground Segmentation for Low Light Environments

Yuuki Tanaka, Yoshihiro Yamashita, Kiyoshi Nishikawa\*, Toru Yamaguchi, Takao Nishitani

Tokyo Metropolitan University, Hino-shi, Tokyo, Japan

\* E-mail: knishikawa@m.ieice.org Tel: +81-42-585-8423

**Abstract**—In this paper, we consider a method for improving the accuracy of the foreground segmentation based on the Gaussian mixture model (GMM) under low light environments. We utilize the GMM foreground segmentation in a system which enables fingertip gesture-input for a wearable device equipped with camera. In this system, one can operate the device with the fingertip by tapping the icons virtually projected on the space displayed through the glasses. However, in the low light environments, the number of segmentation errors of the GMM method would tend to increase due to the narrower range of change in the foreground region. In this paper, to reduce the segmentation errors, we consider applying the image enhancement based on the Retinex theory. Using an actually captured video sequence, we examine the effect of the proposed method.

## I. INTRODUCTION

In this paper, we consider improving the foreground segmentation based on the Gaussian mixture model (GMM) [1]–[3] in the low light environments. The proposed method uses the image enhancement based on the Retinex theory [4], [5] for the improvement.

To realize more flexible mobile computing, a wearable device controlled by the fingertip gesture is one of the possible solutions in the near future. To capture the movement of the fingertip from the video sequence taken by a mobile camera equipped to the wearable devices, foreground segmentation is required. In this paper, we consider using the GMM foreground segmentation because our target devices are wearable ones whose computational power is limited. The GMM assumes the variation of the pixel value can be modeled as a mixture of the Gaussian distributions, and it requires comparatively small amount of computation for implementation.

The problem we consider is that, when the target video is captured under low light environments, the accuracy of the foreground segmentation will be degraded. Because in those environments, variation of the pixel values will be narrower so that the variance of each Gaussian distribution will also become smaller. This makes it difficult to select the appropriate value for the threshold value used to decide the foreground or background region.

In this paper, we consider developing a method to improve the accuracy of the foreground segmentation in low light environments. For that purpose, instead of optimizing the threshold value, a method based on the image enhancement is investigated. By enhancing the image, the variance of the amplitude variation of each pixel can be controlled so as to

relax the selection of the appropriate threshold value. Besides, as the image enhancement method, we use the Retinex theory because we can expect that the color of the image is less affected at enhancement process. We consider an implementation of the image enhancement based on the Retinex with small amount of calculation. The results of applying the proposed method to a video sequence are provided and we can confirm the effectiveness.

## II. BASICS OF THE GMM FOREGROUND SEGMENTATION

Here, we briefly describe the GMM foreground segmentation [1] and shadow and reflection removal method [2], [3], [6] which are used in our system.

### A. Foreground segmentation using GMM

In the following, we assume that the image is divided into blocks of  $4 \times 4$  pixels. It is also assumed that each pixel has three components, or colors, i.e., R (red), G (green) and B (blue) of the range from 0 to 255 (i.e., 24-bit color format).

As the characteristic values, the average values of  $4 \times 4$  pixels in a block of each color are used. Those three values are applied directly to the GMM foreground segmentation instead of transforming those values.

By denoting the total number of blocks in each frame image as  $N_B$ , the characteristic values of  $i$ -th block ( $i \in \{1, 2, \dots, N_B\}$ ) at time  $t$  is expressed as  $f_{R,t}(i)$ ,  $f_{G,t}(i)$ , and  $f_{B,t}(i)$  for R, B, G components respectively. Because the same process is applied to each block, in the following description, we omit the block number  $i$ , e.g.,  $f_{R,t}$  is used for showing the R component at time  $t$ .

Let us briefly describe the GMM foreground segmentation using  $f_{R,t}$ , and the same process will be applied to G and B components. It is assumed that the probability density function (PDF) of  $f_{R,t}$  is expressed as a weighted sum of multiple Gaussian distributions, i.e.,

$$p(f_{R,t}) = \sum_{k=1}^K \omega_{k,t} \times N(\mu_{k,t}, \sigma_{k,t}) \quad (1)$$

where  $p(\cdot)$  shows the PDF of  $f_{R,t}$ , and  $N(\mu_{k,t}, \sigma_{k,t})$  the Gaussian distribution of the mean  $\mu_{k,t}$  and the standard deviation  $\sigma_{k,t}$ . In this model,  $K$  weighted Gaussian distributions are summed, and  $\omega_{k,t}$  shows the weight for the  $k$ -th distribution.

For  $p(f_{R,t})$  to be a valid PDF, the weights  $\omega_{k,t}$  must satisfy

$$\sum_{k=1}^K \omega_{k,t} = 1 \quad (2)$$

to ensure that the total probability is one. In the following, it is assumed that  $\omega_{k,t}$  and corresponding distributions are sorted according to their amplitudes, i.e.,

$$\omega_{1,t} \geq \omega_{2,t} \geq \dots \geq \omega_{K,t} \quad (3)$$

and the first  $B$  ( $B < K$ ) distributions will be used as the background model.

At time  $t + 1$ ,  $f_{R,t+1}$  is calculated, and we decide whether it belongs to the background model. For that, the following equation will be evaluated,

$$|f_{R,t+1} - \mu_{k,t}| < T \quad k = 1, \dots, B \quad (4)$$

where  $T$  is a predefined threshold value.

We use the following equation as  $T$  [7], [8]:

$$T = \delta \times \mu_{k,t} + 2.5 \times \sigma_{k,t} \quad (5)$$

where  $\delta$  is an adjustable parameter. When  $f_{R,t+1}$  does not satisfy the relation (4), the block will be classified as the foreground. The same procedure will be repeated for  $f_{G,t}$ , and  $f_{B,t}$ .

Then, Gaussian models will be updated as below. When  $f_{R,t+1}$  is classified as the background,  $\mu_{k,t}$ ,  $\sigma_{k,t}$ , and  $\omega_{k,t}$  are updated by

$$\mu_{k,t+1} = (1 - \gamma) + \gamma f_{R,t+1} \quad (6)$$

$$\sigma_{k,t+1}^2 = (1 - \gamma)\sigma_{k,t}^2 + \gamma(f_{R,t+1} - \mu_{k,t})^2 \quad (7)$$

where  $\gamma$  is a forgetting factor ( $\{0 < \gamma \leq 1\}$ ). Also, the weight will be slightly increased by

$$\omega_{k,t+1} = (1 - \zeta)\omega_{k,t} + \zeta \quad (8)$$

where  $\zeta$  is called the learning rate [1].

When  $f_{R,t+1}$  is not included in the background model, only the weight  $\omega_{k,t}$  is updated by

$$\omega_{k,t+1} = (1 - \zeta)\omega_{k,t} \quad (9)$$

and  $\mu_{k,t}$  and  $\sigma_{k,t}$  are unchanged. Note that we need to adjust the values of  $\omega_{k,t}$  ( $k = 1, \dots, K$ ) to satisfy the relation (2).

### B. shadow and reflection removal

When the GMM foreground segmentation is used, we consider the effect of the shadow casted by the hand, or reflections from it on the accuracy of the detection of the foreground objects. To reduce the detection errors due to the shadows or reflections, we proposed a method to remove them [7]. The method uses the similar formula as those of the GMM segmentation.

Here, let us describe the idea of the the shadow removal. When a block is decided as the foreground by the GMM, then it is checked that whether the block is shadow or not

by the following steps. At first, we denote the maximum characteristic value of the block as  $f_{c1}$ , namely,

$$f_{c1} = \max f_C \quad (10)$$

where  $C$  is one of R, G, or B, i.e.,  $C \in \{R, G, B\}$ . Besides,  $f_{c2}$ , and  $f_{c3}$  denote the other ones. Then, we calculate the following value  $\beta_{c1}$ , i.e.,

$$\beta_{c1} = f_{c1}/\mu_{c1} \quad (11)$$

where  $\mu_{c1}$  is the mean value of one of the Gaussian distribution which is nearest to  $f_{c1}$ . Hereafter, we omit to denote  $k$  to show the  $k$ -th distribution of the GMM model for the mean and the standard deviation by always assuming that the appropriate distribution is selected. Similarly, we can define  $\beta_{c2}$ , and  $\beta_{c3}$  as

$$\beta_{c2} = f_{c2}/\mu_{c2}, \quad \beta_{c3} = f_{c3}/\mu_{c3}. \quad (12)$$

In the conventional method, it is assumed that, when the block is in the shadow, characteristic value  $f_C$  are attenuated at the same ratio in RGB components. In other words, if  $\beta_{c2}$  and  $\beta_{c3}$  are close to  $\beta_{c1}$ , then the method decide the block is shadow.

Instead of directly calculating (12), we calculate the following for  $f_{c2}$

$$f'_{c2} = f_{c2} \times \beta_{c1}. \quad (13)$$

Then, we evaluate the condition similar to the GMM as

$$|f'_{c2} - \mu_{c2}| < T_C \quad (14)$$

where  $T_C$  is the threshold value. If the condition (14) is satisfied then the block is decided to be shadow and it will be classified as the background. The same process will be repeated for  $c3$ .

To determine the threshold  $T_C$ , originally in [3], the following equation simmlar to equation (5) is proposed, namely,

$$T_C = 2.5 \times \sigma_{c2} + \mu_{c2} \times \delta. \quad (15)$$

However, we found that there are environmetns where this setting of  $T_C$  does not work, so that, in [3], we considered and proposed four equations to determine  $T_C$ . For details of methods for detemining  $T_C$ , please refer to the reference [3].

### C. Problem of the conventional method

Although the conventional method achieves good separation performance in most cases, we can observe the increase in false detection of a foreground region as the background under some specific conditions. Especially, we confirmed that the performance of the method decreases for the videos captured under low light environments. Here, we show an example of such situations.

In Fig. 1 and Fig. 2, we show frame images taken from video sequences captured at the same place under different light conditions. Namely, Fig. 1 is captured under brighter condition, and, on the other hand, Fig. 2 under darker condition.

In Figs 3 and 4, we show the result of the foreground segmentation applied to the video sequences of Figs 1 and

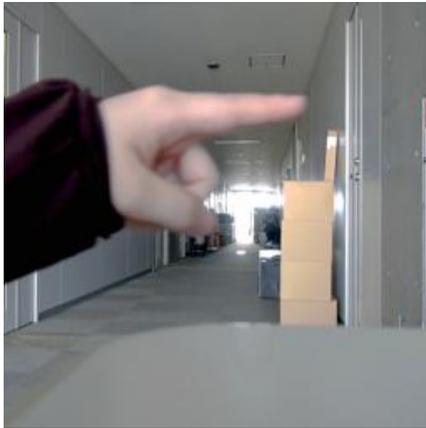


Fig. 1. Frame image captured under the brighter light condition.



Fig. 3. Result of the GMM foreground segmentation applied to Fig. 1.

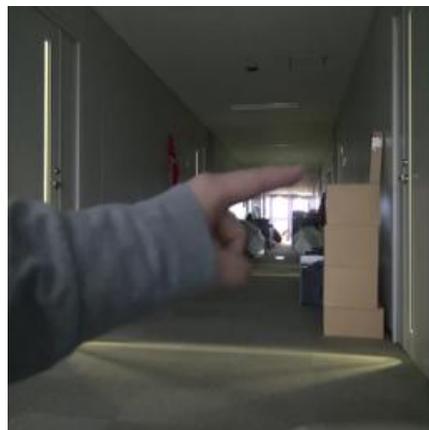


Fig. 2. Frame image captured under the darker light condition.



Fig. 4. Result of the GMM foreground segmentation applied to Fig. 2

2 respectively. From these figures, we can confirm that, for the image taken under brighter condition, the foreground segmentation can separate the hand region. On the other hand, the large part of hand and finger region were lost in Fig. 4.

The accuracy of the GMM foreground segmentation depends on the selection of the threshold values in the equations (4) or (15) as we considered in [7]. However, we cannot know the environment where the device is used in advance, so that it is rather difficult to assign the appropriate values to the thresholds. Therefore, in this paper, we consider another approach, i.e., the application of image enhancement before the GMM segmentation for relaxing the selection of the threshold values.

### III. RETINEX FOREGROUND SEGMENTATION

Let us consider applying the image enhancement for relaxing the selection of the threshold values of the GMM foreground segmentation.

#### A. Retinex image enhancement

For processing the images captured in low light environments, we consider applying the image enhancement before the GMM foreground segmentation.

In that process, the following points are required

- 1) Characteristics values  $f_{C,t}$  of background region of the same block shall be within a Gaussian distribution of the background model. In other words, if  $f_{C,t}$  largely changes from time to time due to the image enhancement, the block may be classified as foreground.
- 2) For real time processing, additional computational cost shall be small.

Based on these requirements, we consider applying the image enhancement based on the Retinex theory. The Retinex theory is originally considered by Land for the treatment of color constancy [4]. The theory of Retinex has been continuously developed [9], [10] and one of its applications to image processing is the image enhancement [11]. Although there are several model of the Retinex theory, in this paper,

we use a formula proposed in [5] which is derived based on the expression of [11] for ease of implementation.

The formula of the Retinex used in the proposed method is given as

$$R_i(x, y) = \log I_i(x, y) - \log [F(x, y) * I_i(x, y)] \quad (16)$$

where  $I_i$  is the image distribution in the  $i$ -th spectral band. Besides,  $F(x, y)$  is called the surround function, and “\*” shows the convolution operation. As the surround function  $F(x, y)$ , the Gaussian function is suggested in [11].

As an extension of (16), [5] proposes an implementation method of the Retinex based on the multi-rate signal processing. The method uses multiple low pass filters as  $F(x, y)$  instead of the Gaussian function and, so that, it can be implemented with reduced amount of computation.

In the method of [5], for each pixel, the following equation is calculated

$$I_{R,i}(x, y) = \begin{cases} \log I_i(x, y) - \log \bar{I}_a(x, y) & (I_i(x, y) < 127) \\ -\log(255 - I_i(x, y)) + \log(255 - \bar{I}_a(x, y)) & (127 < I_i(x, y)) \end{cases} \quad (17)$$

where  $\bar{I}_a$  is the output of the low pass filter. After the calculation of the equation (17), the output signal  $I_{R,i}(x, y)$  is multiplied and truncated to be fit to the 8-bit color representation [5].

### B. Proposed Retinex implementation

Our objective is to apply the image enhancement based on the Retinex theory into real-time finger gesture input system. For that purpose, the amount of calculation should be reduced as possible.

As compared in [5], the implementation using low pass filters instead of the Gaussian filter can reduce the number of required calculation. However, it still needs filtering for each pixel adding to the Retinex operation of the equation (17). Besides, the method implements the multi scale Retinex (MSR) using the multi rate processing and it requires additional computation.

Hence, we think of using the single scale Retinex (SSR) as the basic in the proposed method for reducing the computation. Although the MSR can achieve better color constancy than SSR, this selection is done because our purpose is the segmentation of the hand and fingertip from the background, and exact color of objects are not required.

Let us show the expression of the Retinex in the proposed method. At first, because we use SSR, there is only one  $I(x, y)$  instead of multiple  $I_i(x, y)$  in (17). Then, we slightly change the equation (17) as

$$I_R(x, y) = \begin{cases} \log I(x, y) - \log I_a & (I(x, y) < 127) \\ -\log(255 - I(x, y)) + \log(255 - I_a) & (127 < I(x, y)) \end{cases} \quad (18)$$

where we use  $I_a$  instead of  $\bar{I}_a(x, y)$  in the equation (17).  $I_a$  is the average of a frame image calculated as

$$I_a = \frac{1}{WH} \sum_{x=1}^W \sum_{y=1}^H I(x, y) \quad (19)$$

where  $W$  and  $H$  are the width and the height of the image. We note that calculation of  $I_a$  corresponds to use the low pass filter of the size of the image size and whose coefficients are all one. From this point of view, we can regard the equation (18) as a special case of (17). The process will be repeated for three color components, i.e., R, B, and G.

We note that the proposed approximation of the Retinex calculation may change the color of the objects. However, we consider that this may not affect the accuracy of the foreground segmentation in the proposed method. Because in our problem, the purpose is to extract the hand or finger region, and exact color constancy is not required.

### C. Gain normalization

As the final step, the pixel values  $I_R(x, y)$  are magnified and then clipped into the range 0 to 255.

We use the following equation for magnification, i.e.,

$$I_o(x, y) = \alpha I_R(x, y) + \beta \quad (20)$$

where  $\alpha$  and  $\beta$  are predefined constants. After some trials with varying the values, we set  $\alpha = 64$ , and  $\beta = 128$ , in the simulation shown in the next section. Note that these values shall be adjusted according to the lightning conditions.

Finally, the pixel value is clipped using the following equation:

$$\bar{I}(x, y) = \begin{cases} 255 & I_o(x, y) > 255 \\ I_o(x, y) & 0 \leq I_o(x, y) \leq 255 \\ 0 & I_o(x, y) < 0 \end{cases} \quad (21)$$

In Fig. 5, a diagram of the proposed procedure for calculating the Retinex is shown.

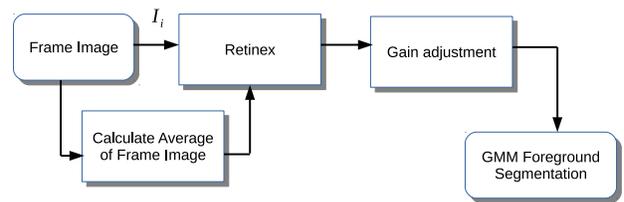


Fig. 5. Flow of the proposed Retinex implementation

## IV. SIMULATION RESULTS

Here, we provides results of simulations to show the effectiveness of the proposed method. We applied the proposed method to the video sequence whose one frame is shown in Fig. 6 which was captured under the same conditions as Fig. 2.



Fig. 6. A frame of the original video sequence for processing

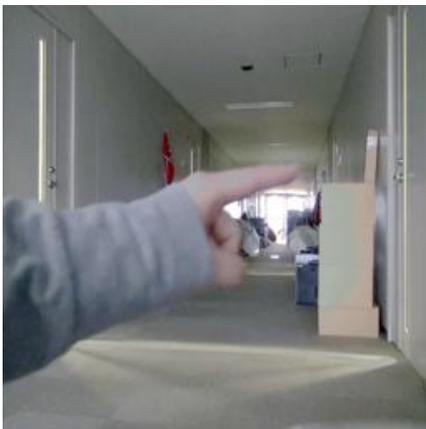


Fig. 7. Result of image enhancement by applying the proposed Retinex processing to the image in Fig. 6.

*A. Effect of the Retinex*

First, we confirm the image enhancement ability of the proposed Retinex processing by showing a result of applying it. In Fig. 7, the enhanced image is shown. By comparing Fig. 6 and Fig. 7, we can confirm that the method enhances the image.

At the same time, we notice that the colors of the some parts changes unnaturally, e.g., part of finger, or objects on the right side. These changes of colors are due to the approximation in the proposed calculation of the Retinex.

*B. Results of foreground segmentation*

In Fig. 8 and Fig. 9, the results of applying the GMM foreground segmentation are shown. As the threshold values for the GMM and the shadow/reflection removal, we used the type IV equation proposed in [7].

The result of applying the image in Fig. 6 is shown in Fig. 8. From this figure, we can see that almost all of the hand and

finger region were deleted by the foreground segmentation, and the shadow/reflection removal.

On the other hand, Fig. 9 shows the result of applying the proposed method to Fig. 7. We can confirm that, in this case, the shape of hand and figure are clearly segmented. Besides, the color change mentioned in the previous subsection do not affect the results of the segmentation.

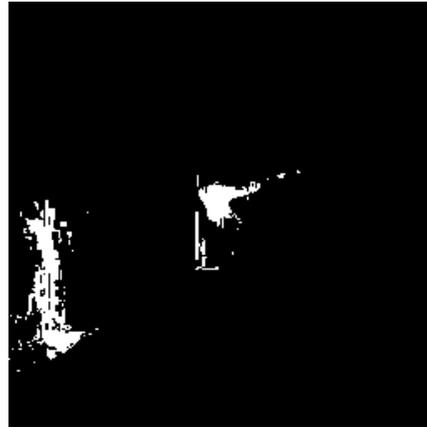


Fig. 8. Result of the GMM foreground segmentation.



Fig. 9. Results of the GMM foreground segmentation

V. CONCLUSION

In this paper, we considered a method to improve the GMM foreground segmentation under low light environments. As a method for reducing the false background decision, we proposed to use image enhancement based on the Retinex theory. The proposed method uses an approximation of the exact Retinex operation to reduce the required amount of calculation. The possibility of reducing the decision errors by the proposed method was confirmed by the results of simulation.

As a future work, the theoretical analysis of the effectiveness of the proposed Retinex implementation shall be considered. Based on the results of the analysis, conditions for the proposed method to reduce the false decision shall be shown.

## REFERENCES

- [1] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, 1999, vol. 2, p. 252 Vol. 2.
- [2] Y. Yamashita, T. Nishitani, T. Yamaguchi, and B. Oh, "Software implementation approach for fingertip detection based on color multi-layer gmm," in *The 18th IEEE International Symposium on Consumer Electronics (ISCE 2014)*, June 2014, pp. 1–2.
- [3] K. Nakagami, T. Shiota, and T. Nishitani, "Low complexity shadow removal on foreground segmentation," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2011, pp. 1081–1084.
- [4] E. H. Land, "An alternative technique for the computation of the designator in the retinex theory of color vision," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 83, no. 10, pp. 3078–3080, May 1986.
- [5] T. Okuno and T. Nishitani, "Efficient multi-scale retinex algorithm using multi-rate image processing," in *2009 16th IEEE International Conference on Image Processing (ICIP)*, Nov 2009, pp. 3145–3148.
- [6] T. Shiota, K. Nakagami, and T. Nishitani, "Transform domain shadow removal for foreground silhouette," *IEICE Trans. on Fundamentals*, vol. E96-A, no. 3, pp. 667–674, March 2013.
- [7] K. Nishikawa, Y. Yamashita, T. Yamaguchi, and T. Nishitani, "Consideration on performance improvement of shadow and reflection removal based on gmm," in *2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, Dec 2016, pp. 1–7.
- [8] J. Zhu, S. C. Schwartz, and B. Liu, "A transform domain approach to real-time foreground segmentation in video sequences," in *Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, March 2005, vol. 2, pp. 685–688.
- [9] D. Zosso, G. Tran, and S. J. Osher, "Non-local retinex—a unifying framework and beyond," *SIAM Journal on Imaging Sciences*, vol. 8, no. 2, pp. 787–826, 2015.
- [10] S. O. Dominique Zosso, Giang Tran, "A unifying retinex model based on non-local differential operators," 2013.
- [11] D. J. Jobson, Z. Rahman, and G. A. Woodell, "Properties and performance of a center/surround retinex," *IEEE Transactions on Image Processing*, vol. 6, no. 3, pp. 451–462, Mar 1997.