

# An improved mixture of Gaussians model based on color and depth data

Chao Wu<sup>1, a</sup>, HengZhu Liu<sup>1</sup>, ZhengFa Liang<sup>1</sup>

<sup>1</sup>College of Computer, National University of Defense Technology, Deya Road NO. 109, Changsha 410073, China

<sup>a</sup>xinji\_wu@163.com

**Keywords:** Background subtraction, Mixture of Gaussians model, Computer vision, Color and depth data combination

**Abstract.** In the traditional background subtraction algorithms based on color, in the special critical situation, such as face the challenges of shadows, color camouflage, illumination changes, they can't achieve good results. But depth data can solve such problems effectively, because depth sensor operating normally in the scene. We proposed an improved mixture of Gaussian model algorithm based on color and depth data in this paper, and use a new decision-making method to fuse results. In order to ensure the stability of the model, we have special treatment of the ADO area. We test our method in three datasets, the results show that proposed approach obtain good results.

## Introduction

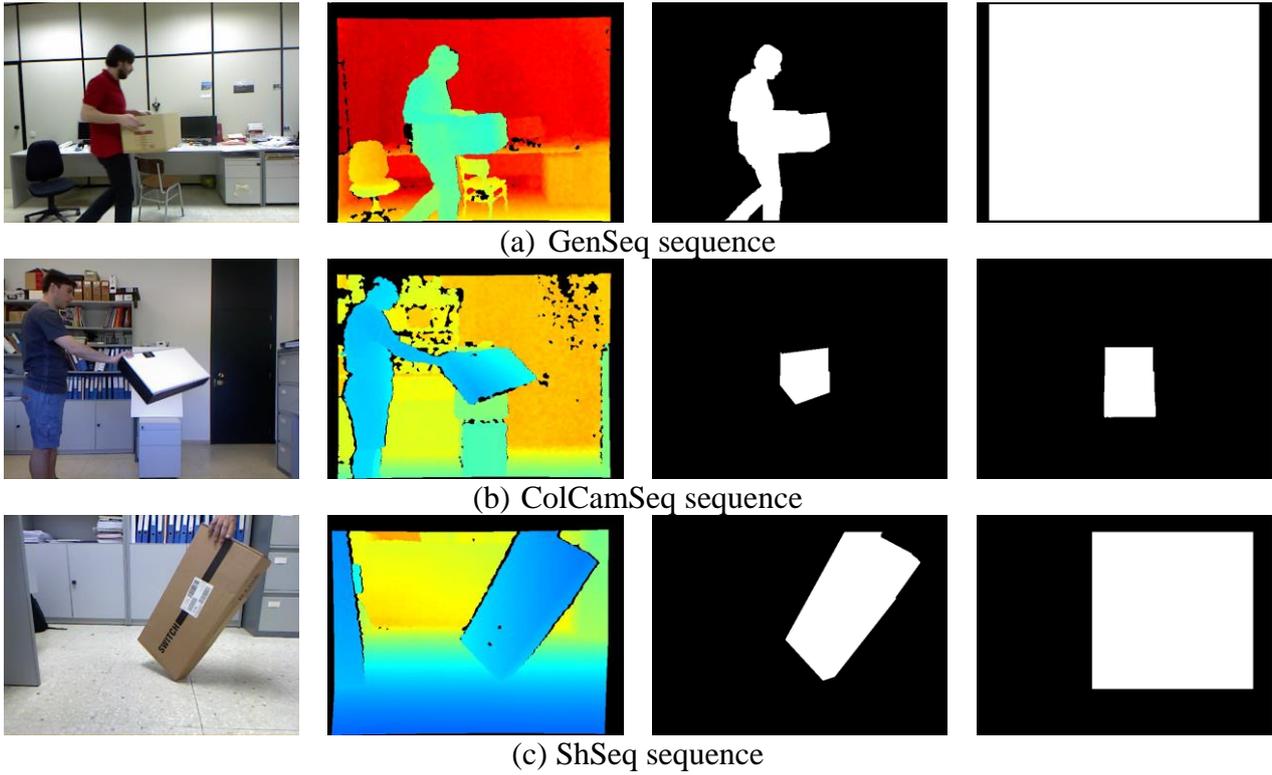
In computer vision, background subtraction is a key processing step of many computer vision systems. It aims at separating the moving objects in the scene (that constitute the foreground) from a robust model of the static environment (the background), often applied in surveillance, object detection, tracking and other applications. The traditional background subtraction [1,2,3] algorithms mainly based on RGB color data. Facing the issues of shadows, camouflage, illumination changes, traditional method can't achieve good results [4].

In recent years, with the development of deep sensor technology, the RGB-D camera began to spread, such as Microsoft Kinect and stereo camera. Like Microsoft Kinect [5], the low-cost RGB-D device that are able to capture depth and color images simultaneously at frame rates up to 30 fps. The depth data is very helpful for foreground/background segmentation in interior environment, because in the process of collecting the depth data will not be affected by the color. Benefit from this advantage, there are a lot of paper focus on approaches that fuse color and depth information.

In Harville et al. [6], they proposed using 4-channel pixel vectors to build mixture of Gaussian model, the  $(Y, U, V)$  vector is extended by adding the depth data to the  $(Y, U, V, D)$  vector. Researchers often have the same idea, F. Sanchez [7] apply the codebook model algorithm to 4-channel scene that pixel value is  $(R, G, B, D)$ . Dong Tian [8] propose a depth-weighted group-wise PCA (DG-PCA) approach that is formulated as a weighted  $l_{2,1}$ -norm PCA problem with depth-based group sparse being introduced. In the recent research, tend to applied color and depth in each model not in the same model. Young-min Song [9] designed two probabilistic background models corresponding to color and depth based on GMM and denoise depth image with the model. Camplani and Salgado [10,11] proposed a foreground/background segmentation strategy based on the combination of two per-pixel statistical classifiers. Their combination is based on a weighted average considering past foreground detections and the detected edges in color and depth images. Van-Toi Nguyen [12] proposed a method for noise removal from depth data to improve the stability of background.

Fusing color and depth information can effectively solve this challenges of shadows, camouflage, illumination changes. But in special scene, some new problems lead to depth information can't be measured, such as near objects, non-reachable areas [13]. In this situation, the sensor return a zero indicating that there is no depth data, this place appear as black holes in the images is show in Fig. 1. We denote this pixels as Absent Depth Observations (ADO). In the above method, the authors didn't give the corresponding solution for this noise. In this paper, we built two mixture of Gaussians models, one based on color the other based on depth data. We improve MOG algorithm to reduce the

impact of ADO and use the new method to fusion results. Then select the strong or weak rate update model based on the results.



**Fig. 1** Samples for each sequence. The first column are color image. The second column are depth map, converts a depth grayscale to a color image for display. The black holes are ADO. The third column image are ground truth foreground. The fourth column is mask image.

### The Improved Mixture of Gaussians Model

Stauffer and Grimson [14] describes the principle of mixture of Gaussian model. The pixel-level background model is constructed using continuous  $N$  frames and the model is updated with time. At any time  $t$  the pixel values of position  $(x_0, y_0)$  can be observed. All historical values before  $t$  can be expressed as:

$$\{X_1, \dots, X_t\} = \{I(x_0, y_0, i) : 1 \leq i \leq t\} \quad (1)$$

The approach is using the mixture of Gaussian model to simulate the distribution of pixel value sequence. The probability of current pixel value  $X_t$  is:

$$P(X_t) = \sum_{i=1}^K \omega_{i,t} * \eta(X_t, \mu_{i,t}, \Sigma_{i,t}) \quad (2)$$

Where  $K$  is the number of Gaussian distributions,  $\omega_{i,t}$  is the weight estimates value of the  $i^{th}$  Gaussian in the mixture model at time  $t$  with mean value  $\mu_{i,t}$  and covariance matrix  $\Sigma_{i,t}$ , and where  $\eta$  is a Gaussian probability density function

$$\eta(X_t, \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{1}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(X_t - \mu)^T \Sigma^{-1} (X_t - \mu)} \quad (3)$$

The number of Gaussian distributions  $K$  is generally between 3 and 5, our method is 3. In order to simplify the calculation, the covariance matrix is computed as follows

$$\Sigma_{k,t} = \sigma_k^2 I \quad (4)$$

$\sigma_k^2$  is the variance of each dimension. For color image, red, green and blue three channels have the same variance value.

In this paper, we use two features. One is RGB color data, the other is depth data. As is known from the preview description, there is an ADO area in the depth map. In this area, depth data will change suddenly at some point. From zero to normal, and then mutate to zero, sometimes change frequently. This sudden change has a great impact on parameter updates. Different from depth map, the pixel value in color image has remained stable. In order to reduce the influence of the ADO region on the color space, we set up a mixture of Gaussian model for the value of two data.  $mog_{rgb}$  is a mixture model in RGB color space, and  $mog_{depth}$  is a mixture model in depth space.

**Model Initialization.** Although mixture model have  $K$  Gaussian distributions, but we only initialize one at a time. In color space, we initialize the mean value of the first Gaussian using the pixel value  $X_1$ , the other parameter use default values. When  $X_t$  not find a matching Gaussian, initialize a new distribution or replace the least weight distribution. Different color image, ADO area will resulting in performance degradation in depth space. If  $X_1 \in ADO$ , do nothing until observing normal depth information, else initialize a new Gaussian distribution.

**Background Moving Object Detection.** For new value  $X_t$ , we need to determine whether it is a background or a foreground. In Stauffer and Grimson, the author's method is: first find the matching distribution and update parameter, then distribution is sorted by weight, finally make a decision. For weaken the impact of ADO, we improve their method, first find the matching distribution, sort the distribution, then classify  $X_t$ , finally update the model based on the results.

In color image, we have got the model of the frame sequence  $\{X_{1,rgb}, \dots, X_{t-1,rgb}\}$  before time  $t$ . For value  $X_{t,rgb}$ , calculate its Mahalanobis distance to each Gaussian distribution select the best match distribution  $k_{hit}$  by threshold. The weight of each Gaussian distribution in the model is recalculated as follows

$$w_{k,t-1} = \frac{1}{\sum_{i=1}^K w_{k,t-1}} w_{k,t-1} \quad (5)$$

Then sort the Gaussian distribution according to the  $w_{k,t-1}/\sigma$  value. After sorting, select the first Gaussian distribution  $B$  that satisfies the condition, the expression in Eq.6. If  $k_{hit} > B$ ,  $X_{t,rgb}$  is foreground, else if background. In depth map, if  $X_{t,depth} \in ADO$ , the result depends on  $mog_{rgb}$ .

$$B = argmin_b (\sum_{k=1}^b \omega_k > T) \quad (6)$$

For color image, the result of mixture models is  $RES(mog_{rgb}) = \{f_{rgb}, b_{rgb}\}$  ( $f$  is foreground,  $b$  is background), for depth map, the result in is  $RES(mog_{depth}) = \{f_{depth}, b_{depth}, ADO\}$ . The last fusion result are computed as follows

$$\begin{aligned} & \text{if } X_{t,depth} \in ADO : res = RES(mog_{rgb}) \\ & \text{else :} \\ & \quad \text{if } f_{rgb} \text{ and } f_{depth} : res = fg \\ & \quad \text{else if } f_{rgb} \text{ and } b_d : res = bg \quad (7) \\ & \quad \text{else if } b_{rgb} \text{ and } f_d : \\ & \quad \quad \text{if } X_{t,depth} < X_{t-1,depth} : res = fg, \text{ else : } res = bg \\ & \quad \text{else : } res = bg \end{aligned}$$

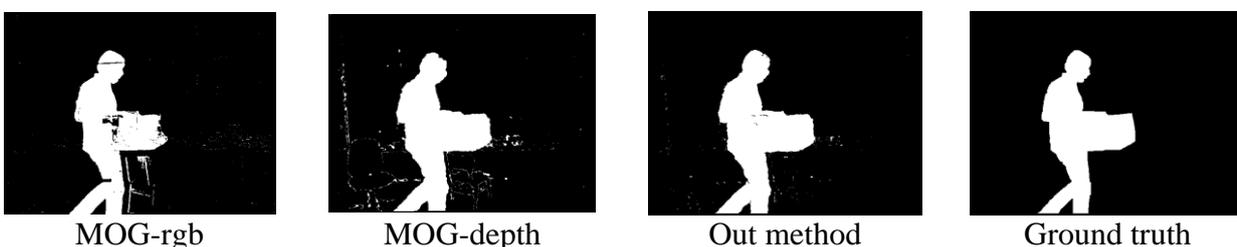


Fig.2 The results in GenSeq

**Model Update.** In order to ensure the stability of the model, we use two update rates: strong and weak value. Use strong value when the result is background, mixture model will be better adapted to light, shadow effects. If the result is foreground, use weak update rate. This will enhance the stability of the mixture model, because the change of foreground value is frequently.

The weight  $\omega_{k,t}$  of the  $k$  distribution at time  $t$ , are adjusted as follows:

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

Where  $\alpha$  is the learning rate<sup>2</sup>, select the strong or weak rate based on the results. And  $M_{k,t}$  is 1 for the model which matched and 0 for the remaining models. The  $\mu$  and  $\sigma$  parameters for unmatched distributions remain the same. The parameters of the distribution which matches the new observation are updated as follows

$$\begin{aligned} \mu_t &= (1 - \rho)\mu_{t-1} + \rho X_t \\ \sigma_t^2 &= (1 - \rho)\sigma_{t-1}^2 + \rho(X_t - \mu_t)^T(X_t - \mu_t) \\ \rho &= \alpha\eta(X_t | \mu_k, \sigma_k) \end{aligned} \quad (9)$$

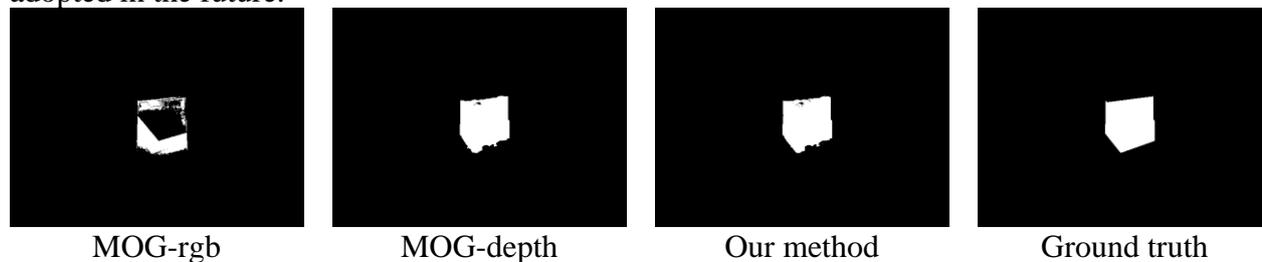
## Experimental Result

We evaluating our background subtraction algorithm by use three datasets: GenSeq, ColCamSeq, ShSeq [10], and is show in Fig. 1. GenSeq sequence used to test the overall performance of the algorithm in case of complex scenarios. ColCamSeq sequence aims at testing the performance of the algorithms when the color camouflage problem occurs. ShSeq is to highlight the impact that shadows projected moving object have on the foreground/background segmentation algorithm.

At the same time we tested several other algorithms. One is MOG based on color, the second is MOG based on depth data. The results of the various methods is show in Fig. 2, Fig. 3, Fig. 4. As the Fig. 2 show,  $MOG_{depth}$  method have a lot of noise because ADO, but it is reduced in our method. In Fig. 3,  $MOG_{rgb}$  method didn't have detect the foreground because camouflage, our method can solve it. In Fig. 4, our method can eliminate shadows and detect the contour of box.

## Conclusions

In this paper, we propose an improved mixture of Gaussian model, we built two independent model for color and depth data. Our method can effectively eliminate shadows, and solve the color camouflage. Because of the sensor, there are ADO in depth map. Aim to reduce the impact of ADO on the model, we use new decision-making methods to fuse the color and depth results. As our future work, we expect to use the new model based on color and depth information, because our method has a high computational complexity. Large datasets include more challenges of depth data will be adopted in the future.



**Fig.3** The results in ColCamSeq.



MOG-rgb

MOG-depth

Our method

Ground truth

**Fig.4** The results in ShSeq.

## References

- [1] A. Sobral, A. Vacavant: *Computer Vision and Image Understanding*, 2014, 122: 4-21
- [2] T. Bouwmans: *Recent Patents on Computer Science*, 2011, 4(3): 147-176
- [3] N. Goyette, P.-M. Jodoin, F. Porikli, J. Konrad, P. Ishwar, in: *Workshop on Change Detection at CVPR, IEEE*, 2012
- [4] T. Bouwmans: *Computer Science Review*, vol. s 1112, pp. 31–66, 2014
- [5] Z. Zhang: *IEEE Multimedia* 19.2(2012):4-10
- [6] M. Harville, G. Gordon, J. Woodfill in: *Proceedings of the IEEE Workshop on Detection and Recognition of Events in Video, IEEE Computer Society, Los Alamitos, CA, USA*, pp. 311, 2001
- [7] E.J. Fernandez-Sanchez, L. Rubio, J. Diaz, E. Ros: *Machine Vision and Applications*, 2014, 25(5):1211-1225
- [8] D. Tian, H. Mansour, A. Vetro in: *Image Processing (ICIP), 2015 IEEE International Conference on*, Sept 2015, pp. 3230–3234
- [9] YM. Song, S.J Noh, J. Yu, CW, Park in: *Control, Automation and Information Sciences (ICCAIS), 2014 International Conference on. IEEE*, 2014, pp. 132–135
- [10] M. Camplani, L. Salgado: *Journal of Visual Communication & Image Representation*, 2014, 25(1):122–136
- [11] M. Camplani, C.R.D. Blanco, L. Salgado: *Pattern Recognition Letters*, 2014, 50:23-33
- [12] V.T. Nguyen, V. Hai, T.H. Tran: *Some Current Advanced Researches on Information and Computer Science in Vietnam*. 2014:49-63
- [13] G. Moyà-Alcover, A. Elgammal, A. Jaume-I-Capó, J. Varona,: *Pattern Recognition Letters*, 2016
- [14] C. Stauffer, W.E.L Grimson: *Computer Vision and Pattern Recognition*, 1999. *IEEE Computer Society Conference on. IEEE Xplore*, 1999:252 Vol. 2