# PPViBe: Privacy Preserving Background Extractor via Secret Sharing in Multiple Cloud Servers

*Xin Jin[1,3],Yaming Wu[1,2],Xiaodong Li[1,*] Yuzhen Li[1,2],Geng Zhao[1], Kui Guo[1]*
[1]Beijing Electronic Science and Technology Institute, 100070, Beijing, China
[2]Xidian University, 710071, Xi'an, China
[3]Information Technology Research Base of Civil Aviation Administration of China,
Civil Aviation University of China，Tianjin，300300
*Corresponding Author: lxd@besti.edu.cn

***Abstract*—Recently, with the maturity of cloud services and the development of distributed computing, increasing images and video data are stored in the cloud. However, cloud services are generally provided by the third party entities. In addition, the images and video which are stored in the cloud mostly depend on the security of the cloud servers, while the data are not encrypted, which is a great threat to the users' privacy. In this paper, we propose a method of privacy preserving background extraction of video surveillance in multiple cloud servers based on Chinese Remainder Theorem (CRT). The client uses the CRT to divide a video frame into several encrypted frames, which are sent to the corresponding cloud servers separately. The cloud servers conduct the algorithm of Visual Background Extractor (ViBe) in the encrypted video frames, hereafter, the results obtained are transmitted to the end server, in which the results are decrypted and combined to get the final extraction results in original videos. Finally, the results are transmitted back to the client. The proposed method has several advantages: (1) Based on our encryption method, the original extraction method in the original videos need not be changed; (2) Each cloud server learns only a portion of the video frame information, yet they cannot recover the original video even if the data is leaked; (3) Multiple cloud servers can improve the security of data and enhance the processing efficiency. The experimental results on several video surveillance datasets shows that our method can correctly extract the background as the original ViBe algorithm does while protecting the privacy of the client video data.**

***Keywords—Privacy Preserving; Video Surveillance; ViBe; CRT; Multiple Cloud Servers;***

## I. INTRODUCTION

Intelligent video surveillance has been used everywhere in our daily life and public security. Nowadays, cloud computing has changed the way of traditional video surveillance. The big data of surveillance videos are stored and automatically analyzed in the cloud server, which supports large scale video surveillance applications such as face tracking, suspect searching [3]. In addition, multiple cloud servers can provide services together for some special applications.

Background extraction has a significant application in the video surveillance system, which is an important step of motion detection and tracking. In recent years, video surveillance has been widely used in various places, yet the tremendous video data cannot be effectively processed locally. With the development of cloud services, people choose to upload videos to the cloud, using the powerful processing capability of cloud server to process and then getting the corresponding results. However, these uploaded data haven't any protection, leading to video contents directly exposed to the cloud, and the privacy of users would be seriously violated once the data are leaked. This paper combines privacy preserving and background extraction, exploiting distributed computing to ensure data security while processing video information.

### *A. Related Work.*

Over the years, increasingly privacy preserving computer vision algorithms in the cloud have been proposed [1][2][3][8][9][10][11][12][13]. In [1], a framework is proposed to carry out privacy preserving surveillance. It splits each frame into a set of random images. Each image by itself does not convey any meaningful information about the original frame, while collectively, it retains all the information. This solution is derived from a secret sharing scheme based on the Chinese Remainder Theorem (CRT), suitably adapted to image data. But the method leads to satisfactory results only in controlled environments. It can became problematic in practical applications.

Recently, Chu et al. [2] propose a method for real-time privacy preserving moving object detection in the cloud. However, it is found that this method has a main drawbacks: (1) the server can clearly see the contours of the foreground objects, which can release some privacy information of the original surveillance videos, (2) the security of their encryption method is not that well because of the less secure random scheme if the randomness functions used in their work are not based on chaotic mapping.

Most recently, Jin el al. propose an improved method [3] of chaotic mapping for privacy preserving, which outperform the method of [2] in the security issue. However, we found that when the only cloud server is attacked and the key lost, the attacker can get all the information from cloud server. This method has high dependence on encryption key. In addition, both [2] and [3] use the mixture of Gaussian model to extract the background. They use different random schemes for the frame confusion and diffusion. However, the correctness is slightly reduced because of the decryption error.

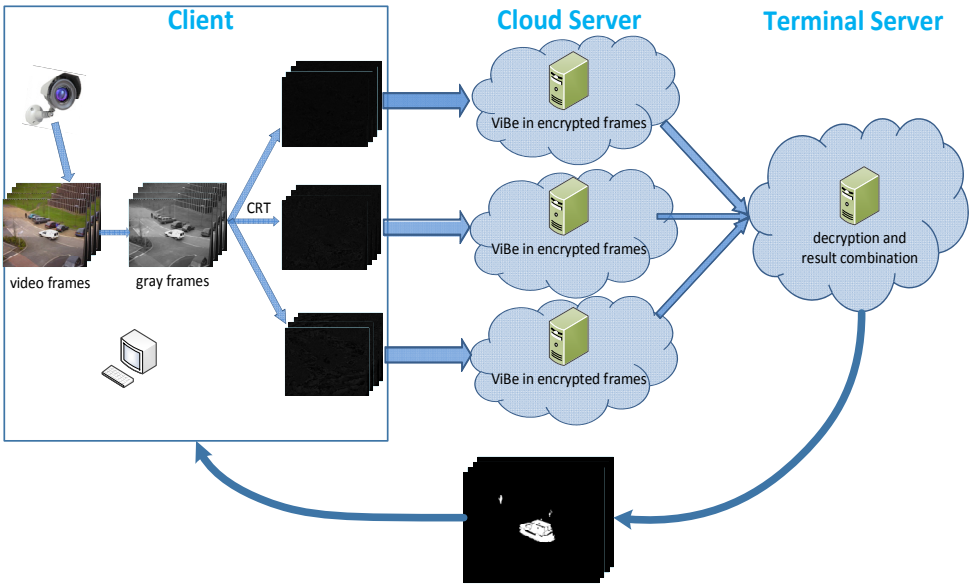


Fig. 1.Method flow char

### B. *Our Approach.*

In order to protect the security of user video data, we propose a method of privacy preserving background extraction of video surveillance, which combine the ViBe detection algorithm and encryption scheme based on Chinese Remainder Theorem (CRT). The computation operations of the mixture of Gaussian method for background extraction are too complicated to integrated into the CRT based secret sharing method. While the ViBe contains only simple operations and is suitable for the CRT framework with nearly homomorphic encryption properties.

As shown in Fig.1, the client does the segmentation process for the video frame .The results are separately sent to different cloud servers. The cloud servers receive the encrypted video frames. In addition, each server learns only a portion of information of the original video frame. The cloud servers use ViBe background detection algorithm [4][5] to process encrypted video frames, following, the processing results in each cloud server are sent to the end server. It is in that place where we judge whether the pixel is in the background or not. Finally, the result is transmitted back to the client. Owing to each server learning only a portion of encrypted information of the video frame, they cannot directly get the information of original video frame. Furthermore, a single server cannot recover the original video frames. It is proved to be a very valid method of privacy preserving background extraction for intelligent video surveillance. The proposed method has several advantages:

*1)* Based on our encryption method, the original extraction method in the original videos need not be changed;

*2)* Each cloud server learns only a portion of the video frame information, yet they cannot recover the original video even if the data is leaked;

*3)* Multiple cloud servers can improve the security of data and enhance the processing efficiency.

This paper is organized as follows. In section 2, we introduce basic theory of the proposed algorithm. The design of the algorithm in detail is described in the section 3. Simulation results put in the section 4.The security analysis is discussed in section 5. Section 6 gives the conclusion.

## II. BASIC THEORY.

### A. *Visual Background Extractor*

Visual Background Extractor (ViBe) [4][5] is an algorithm of pixel-level background modeling, occupying less memory and having high processing efficiency. Generally speaking, the values of pixel in background and pixels surrounding it possess the characteristics of having a small change within a certain time. Making use of the above characteristic, the sample model, which is used to judge whether the background or not, is established for each pixel.

### B. *The Chinese Remainder Theorem*

The Chinese Remainder Theorem (CRT) is a result about congruence in number theory and its generalizations in abstract algebra. According to the CRT, we construct $i$ congruencies for one pixel by $i$ different prime numbers. In that way, the video frame is divided into $i$ encryption frames. On the basis of the property of CRT, it does not change the solution of congruence while doing the same operation for the remainder. And the pixel value after taking the remainder will not leak any information of the video frame.

Set prime number to $p$ and one pixel value as $d$. Using CRT directly, $d^*$ is obtained result.

$$d^* = d \bmod p \tag{1}$$

But taking into account of the correlation between the pixels, it would leak out a portion of the image information. Thence, it is imperative to make the following transformation:

$$\begin{aligned} d^\dagger &= (d * s + \eta) \bmod p \\ \eta &\in (0, r_{\max}) \\ r_{\max} &< s \end{aligned} \tag{2}$$

Where $s$ is a constant, $\eta$ is a random number, $d^\dagger$ is obtained result. It reputes that d turns into a large number. Through the above transformation, we can reduce the correlation between the image pixels and reduce the leaking of image information. It is assured that after d becoming large number, the random number $\eta$ has a very small effect on the accuracy overall. This paper uses the approach as the (2).

## III. Privacy Preserving ViBe

### A. The Client

The client transforms the captured video frames into single-channel gray frames. Then we use CRT to process the obtained gray frames. Selecting *m* prime numbers: $\{p_1, p_2 ... p_i ... p_m\}$ (m takes 3 in this paper). Then we divided every pixel into three pieces by using the (2). For example, set a pixel' value is *α*, and we can get three results by (3):

$$\begin{aligned} \alpha_1 &= (\alpha * s + \eta) \bmod p_1 \\ \alpha_2 &= (\alpha * s + \eta) \bmod p_2 \\ \alpha_3 &= (\alpha * s + \eta) \bmod p_3 \end{aligned} \tag{3}$$

By the above process, as the (3), a video frame can be divided into three encrypted frames. After that, the three encrypted frames are transmitted to three corresponding server, respectively.

### B. The Cloud Servers

*1) Background modeling and update.*

After the corresponding cloud server receiving the encrypted frame, the image pixels of the first frame adopt ViBe modeling. For each pixel, the neighboring pixel values are selected to get its sample model. Formally, let us denote by x one pixel of the image, and by n the number of values in the 8-connected regions centered on x. The n values are denoted as (4):

$$V(n) = \{x_1, x_2, ..., x_i, ...x_n\} \tag{4}$$

Where $V(x)$ is the sample model of pixel *x*, $x_i$ is the *i*-th randomly achieved value. In order to make the above background model able to adjust to the altering background, such as the changing illumination, background objects and so on, each sample model of one pixel has a certain probability to be updated over time.

*2) Model application.*

After modeling using the first frame, we successively start to process the other frames. The value of new pixel y in the second frame is made the difference with the pixel values within the sample model, respectively. Then, we do modulo for each difference using a prime number which corresponding to the one in client. As shown in (4)-(6).

$$C(y) = \{y - x_1, y - x_2, ..., y - x_i, ...y - x_n\} \tag{5}$$

$$c_i = (y - x_i) \bmod p_t \tag{6}$$

$$C(y) = \{c_1, c_2, ..., c_i, ..., c_n\} \tag{7}$$

Thus, we get a set for every pixel. Certainly, the above operations are carried out on each frame and each server. Afterwards, the difference sets are all sent to the terminal server.

### C. The Terminal Server

$$\begin{aligned} C_{p_1}(y) &= \{c_{11}, c_{12}, ..., c_{1i}, ..., c_{1n}\} \\ C_{p_2}(y) &= \{c_{21}, c_{22}, ..., c_{2i}, ..., c_{2n}\} \\ C_{p_3}(y) &= \{c_{31}, c_{32}, ..., c_{3i}, ..., c_{3n}\} \end{aligned} \tag{8}$$

After received all the difference sets (8) from all the three cloud servers, the terminal server will construct congruence equations as (9).

$$\begin{aligned} y_1 &\equiv c_{11} \bmod p_1 \\ y_1 &\equiv c_{21} \bmod p_2 \\ y_1 &\equiv c_{31} \bmod p_3 \end{aligned} \tag{9}$$

Then we can solve the equations by CRT to get decrypted difference set $D(y) = \{y_1, y_2, ..., y_i, ..., y_n\}$ for every pixel. Compares each value in $D(y)$ to the threshold R, statistic the number of less than R，denoted by #. #min represents the minimum matching value. When # is more than #min, the pixel is determined to be the background one. At last, all the results are transmitted to the client.

## IV. Experimental Results

We test the proposed method in various categories of surveillance videos from a large public dataset [6], which contains nearly 16000 manually annotated surveillance video frames and several subset from other public datasets. The tested surveillance videos contains categories of baseline, shadow, night videos, and intermittent object motion. We test the correctness rate of the background extraction, the visual results and the security analysis. In addition we compare our method with that of Jin et al. [3] and Chu et al [2]. Notice that, all the background extraction experiments run in cipher video frames. The experimental results reveal that:

*1)* The correctness rate of the background extraction of our PPViBe is the same as that of the original ViBe.

*2)* The encryption results in multiple cloud servers can protect the contents of the client video and cannot be recognized.

In our experiments, we chose the model containing 20 sample values, set s to 33, a sphere of threshold 20 and the # min was set to 2. Consider for the update of the sample, every pixel model would update the sample values in probability of 1/16. As for CRT encryption, we picked 3 prime numbers 19, 29 and 31.

### *A. The Correctness Rate*

The correctness rate of the background extraction is defined as the number of pixels correctly labelled as foreground or background against the total pixels in the video sequences. Both [2] and [3] use the mixture of Gaussian model to extract the background. The correctness is slightly reduced because of the decryption error. While our PPViBe method has not decryption error and obtain the same correctness rate as that of the ViBe, as shown in Table 1.

TABLE I. THE CORRECTNESS RATE OF EXTRACTION USING VIBE, PPVIBE, METHOD OF [2] AND [3]

| Videos/ Frames | ViBe | PPViBe | Method of [2] | Method of [3] |
|---|---|---|---|---|
| backdoor/1186 | 0.822044 | 0.822044 | 0.807418 | 0.795028 |
| bus station/1111 | 0.943676 | 0.943676 | 0.939721 | 0.946990 |
| cubicle/2811 | 0.983418 | 0.983418 | 0.973237 | 0.974191 |
| highway/1179 | 0.951065 | 0.951065 | 0.929562 | 0.936468 |
| office/629 | 0.905409 | 0.905409 | 0.897626 | 0.901332 |
| pedestrians/921 | 0.983189 | 0.983189 | 0.984055 | 0.988313 |
| sofa/628 | 0.946832 | 0.946832 | 0.945863 | 0.944334 |

### *B. Encryption Results*

At client, we encrypt the gray frames by CRT. According to the selected prime numbers, we can get corresponding encryption results. As shown in Fig.2, these are three encrypted video frames that the three cloud server received. Visually, the frames after encryption didn't contain any of original frame information. That means cloud servers can barely learn any message about plain frame from encrypted frame. Even if encryption frames leak out from cloud server, it cannot recover pain frame through leakage, unless you have got encrypted frames from all cloud servers.

### *C. Background Extraction Results*

As shown in the Fig.3, we get the same background using ViBe by multiple cloud servers from encrypted frames or unencrypted frames. Thus the results from ViBe background extraction are the same, whether video frames have been encrypted or not. Furthermore, the information of video frames will not directly exposed in cloud servers after be encrypted. This strategy is an efficient way to protect the privacy of user videos. The results of [2] and [3] are not that satisfied because of (1) the decryption error introduced by the chaotic mapping, (2) the performance limitation of the mixture of Gaussian methods.


Fig. 2.Encrypted video frames. The original frames are split into encrypted shares. The video contents are protected and nearly nothing can be recognized in the cloud servers.

## V. CONCLUSION AND DISCUSSION

In this paper, we propose a method of privacy preserving background extractor via secret sharing in multiple cloud servers. The client separates video frames encrypted using the CRT and sends to several cloud servers. Then cloud severs and terminal server will work cooperatively to extract background by ViBe. All the test results show that our method could get precisely background while video frames are safely protected. This is the first time that the ViBe is integrated into the CRT based secret sharing framework. In the future work, we will integrate more video surveillance algorithms to Secure Multi-party Computation (SMC) framework and make the computer vision in the cloud more secure.

## VI. ACKNOWLEDGEMENTS

This work is partially supported by the National Natural Science Foundation of China (Grant NO.61402021, 61402023), the Science and Technology Project of the State Archives Administrator (Grant NO.2015-B-10), the open funding project of State Key Laboratory of Virtual Reality Technology and Systems, Beihang University (Grant NO. BUAA-VR-16KF-09), the Fundamental Research Funds for the Central Universities (NO. 2014GCYY02, 2014GCYY04, 2016LG03, 2016LG04), and the Open Project Foundation of Information Technology Research Base of Civil Aviation Administration of China (NO. CAAC-ITRB-201403)

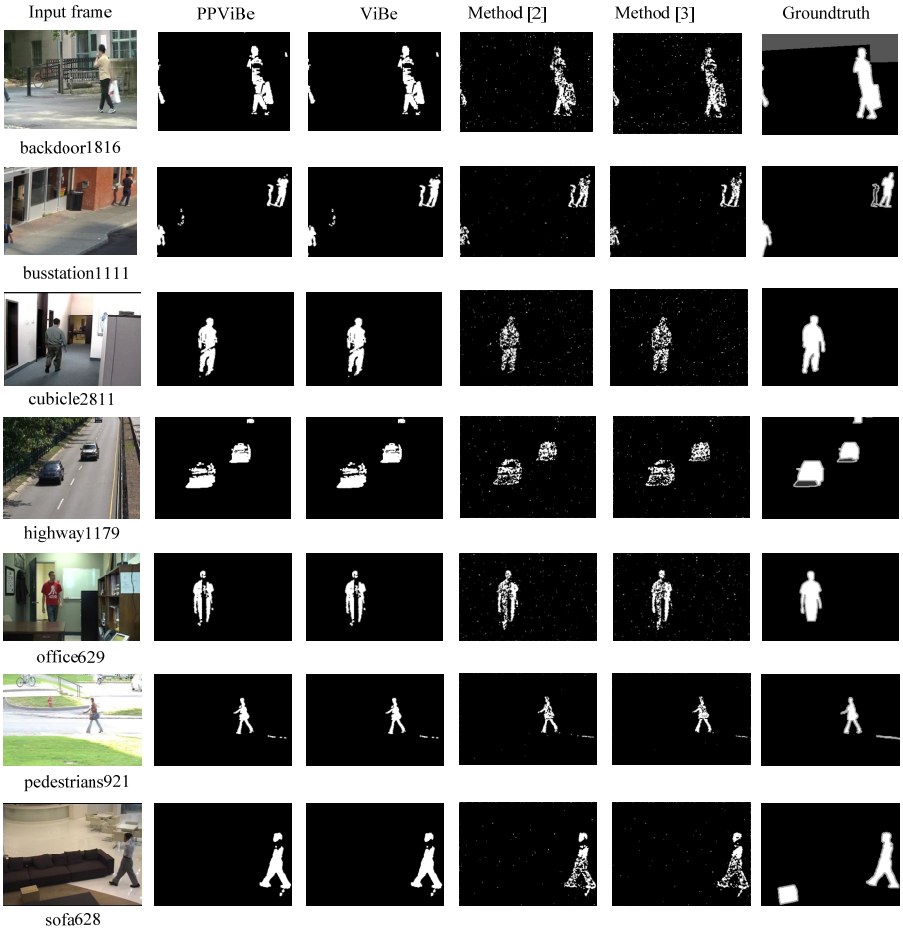


Fig. 3.Background extraction results

## REFERENCES


[1] Upmanyu,M., Namboodiri,A.M., Srinathan,K., Jawahar.C.V. Ecient Privacy Preserving Video Surveillance. IEEE 12th International Conference on Computer Vision (ICCV), 1639-1646 (2009).

[2] Chu, K.Y., Kuo Y.H., Hsu W.H. Real-Time Privacy-Preserving Moving Object Detection in the Cloud. ACM Multimedia, 597-600 (2013).

[3] . Xin Jin, Kui Guo, Chenggen Song, Xiaodong Li, et al. Private Video Foreground Extraction through Chaotic Mapping based Encryption in the Cloud. The 22nd International Conference On Multimedia Modelling (MMM), 562-573 (2016).

[4] Olivier Barnich, Marc Van Droogenbroeck. ViBe: A Universal Background Subtraction Algorithm for Video Sequences. IEEE Transactions on Image Processing. 20(6): 1709-1724, June (2011).

[5] Olivier Barnich, Marc Van Droogenbroeck. VIBE: A Powerful Random Technique to Estimate the Background in Video Sequences. IEEE.978-1-4244-2354-5 (2009).

[6] Wang, Y., Jodoin, P.M., Porikli, F., Konrad, J., Benezeth, Y., Ishwar, P. CDnet. An Expanded Change Detection Benchmark Dataset, in Proc. IEEE Workshop on Change Detection (CDW-2014) at CVPR-2014, pp. 387-394. (2014)

[7] Osadchy, M., Pinkas, B., Jarrous, A., Moskovich, B.: Sci_ - A system for secure face identification. In: 31st IEEE Symposium on Security and Privacy, S&P 2010, 16-19 May 2010, Berleley/Oakland, California, USA. (2010) 239-254.

[8] Sohn, H., Plataniotis, K.N., Ro, Y.M.: Privacy-preserving watch list screening in video surveillance system. In: Advances in Multimedia Information Processing - PCM 2010 - 11th Pacific Rim Conference on Multimedia, Shanghai, China, September 21-24, 2010, Proceedings, Part I. (2010) 622-632.

[9] Chu, C., Jung, J., Liu, Z., Mahajan, R.: strack: Secure tracking in community surveillance. In: Proceedings of the ACM International Conference on Multimedia, MM '14, Orlando, FL, USA, November 03 - 07, 2014. (2014) 837-840.

[10] Avidan, S., Butman, M.: E_cient methods for privacy preserving face detection. In: Advances in Neural Information Processing Systems 19, Proceedings of the Twentieth Annual Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada, December 4-7, 2006. (2006) 57-64.

[11] Shashank, J., Kowshik, P., Srinathan, K., Jawahar, C.V.: Private content based image re-trieval. In: 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2008), 24-26 June 2008, Anchorage, Alaska, USA. (2008).

[12] Fanti, G.C., Finiasz, M., Ramchandran, K.: One-way private media search on public databases: The role of signal processing. IEEE Signal Process. Mag. 30(2) (2013) 53-61.

[13] . Bost, R., Popa, R.A., Tu, S., Goldwasser, S.: Machine learning classification over encrypted data. In: 22nd Annual Network and Distributed System Security Symposium, NDSS 2015, San Diego, California, USA, February 8-11, 2014. (2014).