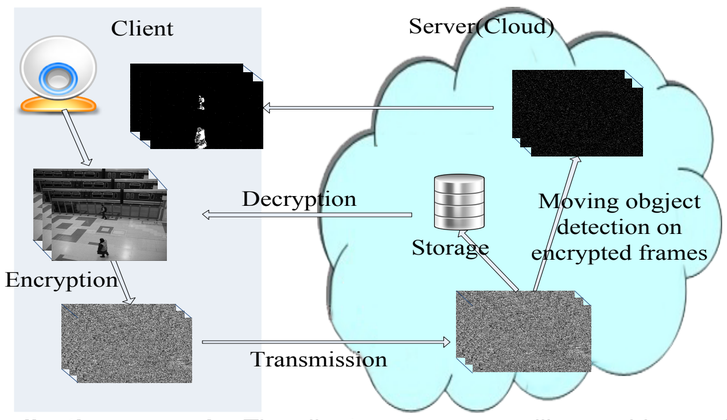
Private Video Foreground Extraction through Chaotic Mapping based Encryption in the Cloud

Xin Jin[1,3,*], Kui Guo[1], Chenggen Song[1], Xiaodong Li[1,*], Geng Zhao[1], Jing Luo[1,2], Yuzhen Li[1,2], Yingya Chen[1], Yan Liu[1,2], and Huaichao Wang[3]

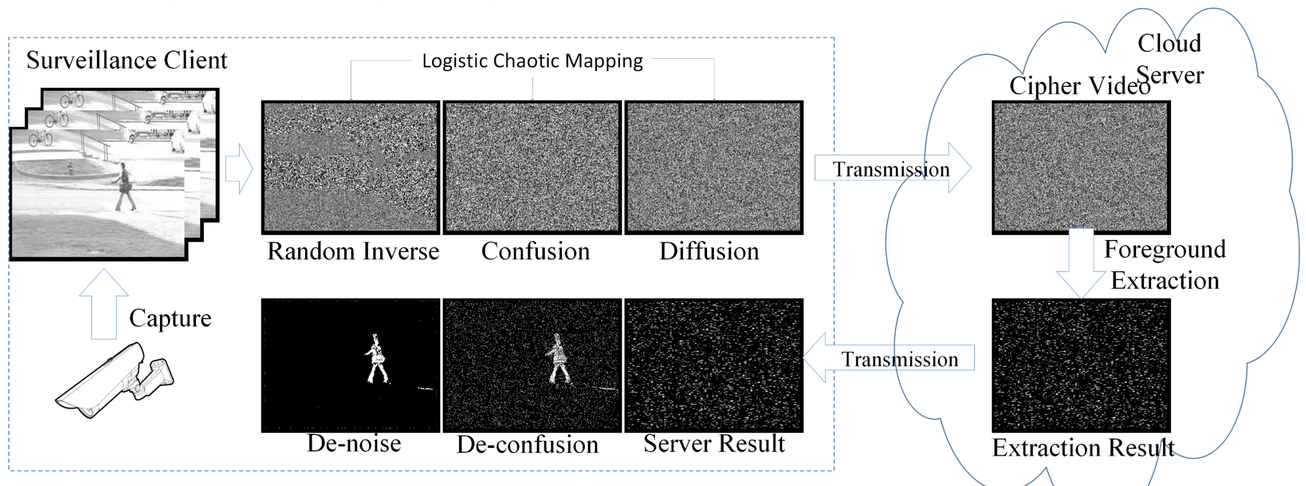
BestiVictory

[1]**Beijing Electronic Science and Technology Institute, Beijing 100070, China**
GOCPCCC Key Laboratory of Information Security, Beijing 100070, China
[2]Xidian University, Xi'an, 710071, China
[3]Information Technology Research Base of Civil Aviation Administration of China, Civil Aviation University of China, Tianjin 300300, China
*{jinxin,lxd}@besti.edu.cn

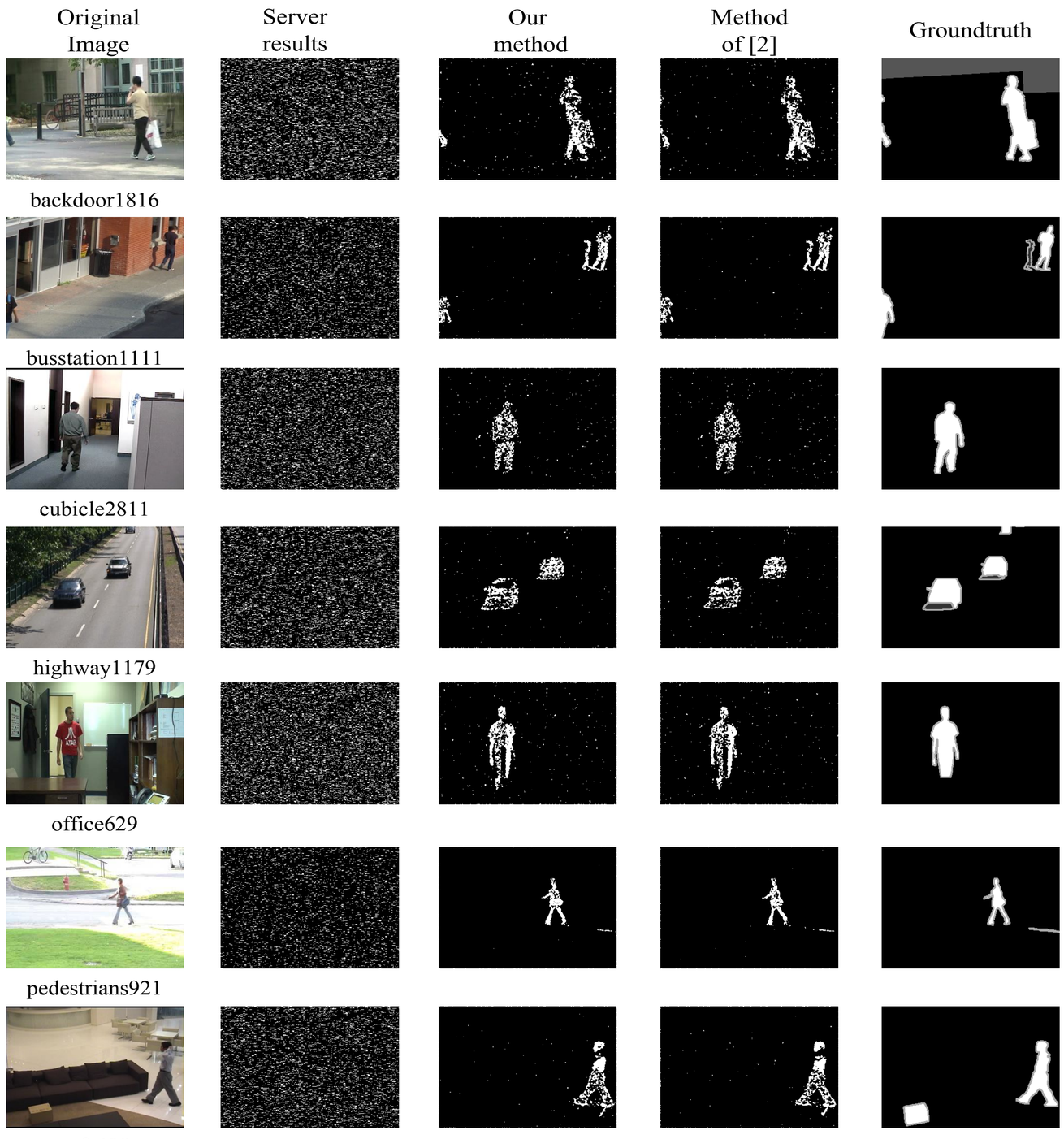
Background. Recently, storage and processing large-scale visual media data are being outsourced to Cloud Data Centres (CDCs). However, the CDCs are always third party entities. Thus the privacy of the users' visual media data may be leaked to the public or unauthorized parties.In this paper we propose a method of privacy preserving foreground extraction of video surveillance through chaotic mapping based encryption in the cloud.



The application scenario. The client captures surveillance videos, which are then encrypted by our proposed chaotic mapping based encryption method. The encrypted surveillance videos are transmitted to the cloud server, in which the foreground extraction algorithm is running on the encrypted videos. The results are transmitted back to the client, in which the extraction results are decrypted to get the extraction results in plain videos. The server learns nothing and the client knows where the foreground objects are.



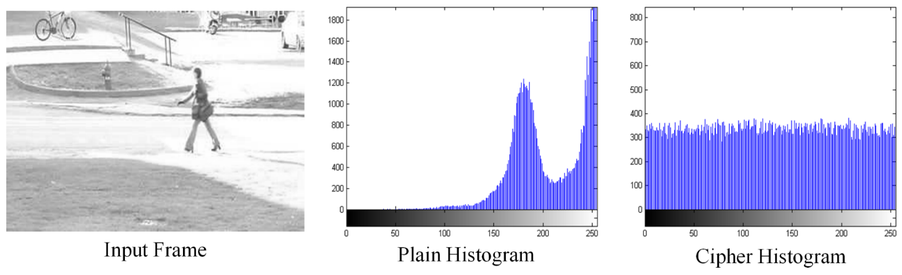
The overview of our method. In the client, each frame of the captured videos is first been randomly inversed pixel by pixel using a random matrix generated by the logistic mapping. Then we use the logistic mapping to confuse the randomly inversed frame, followed by a diffuse operation to get the final cipher frame. Each are transmitted to the cloud server. The cloud server use the standard mixture Gaussian model to extract foreground directly in each cipher frame. The server can only learn the cipher result and can not recognize anything in the server result. This result is transmitted back to the client. The client de-confuse the server result to get the intermediate extraction result, followed by the median filter to de-noise so as to obtain the final foreground extraction result.



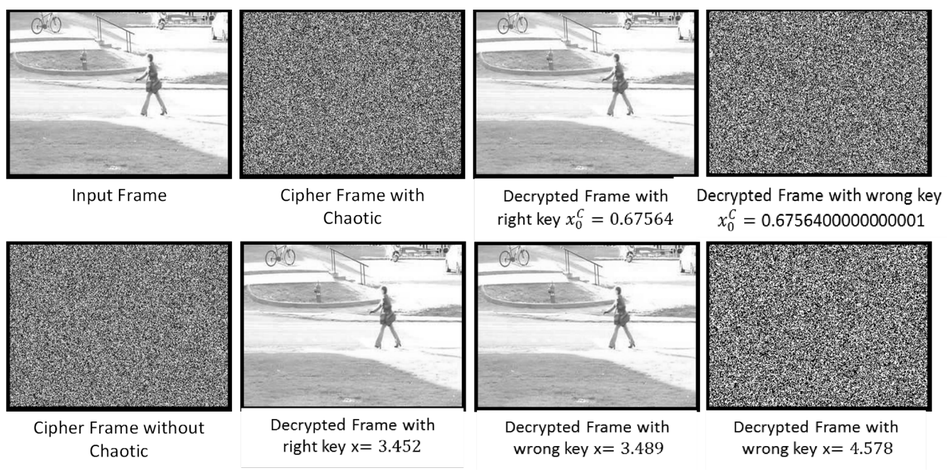
Parts of the foreground extraction results. The input plain video frames are shown in the first column. The extraction results in the server using our method are shown in the second column. The extraction results after the encryption and median filter in the client are shown in the third column. The extraction results in the server using the method in [2] are shown in the fourth column. The serve can clearly observe the contours of the foreground objects and the privacy of the client video is leaked. The ground truth manually segmented and annotated in [12] are shown in the last column.

The proposed method has several **advantages**:
(1) The server only learns the obfuscated extraction results and can not recognize anything from the results.
(2) Based on our encryption method, the original extraction method in the plain videos need not be changed.
(3) The chaotic mapping ensure high level security and the ability to resistant several attacks.

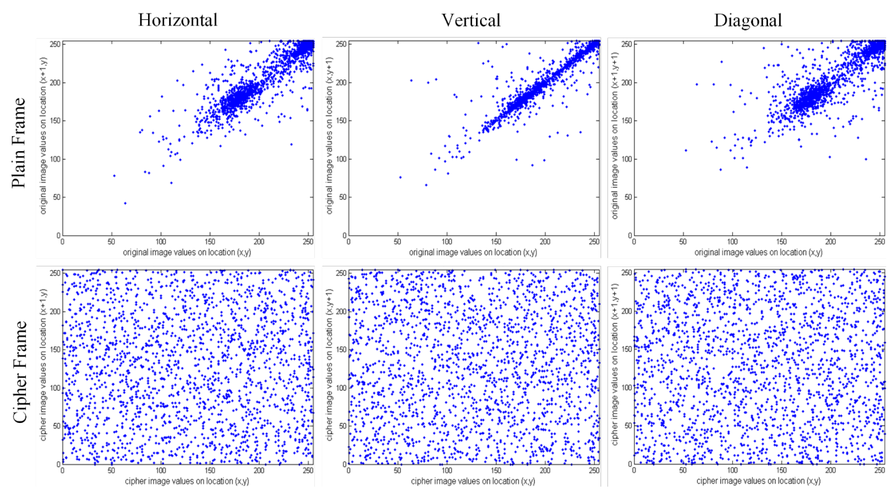
Main References:
[1] Upmanyu, M., Namboodiri, A.M., Srinathan, K., Jawahar.C.V. Ecient Privacy Preserving Video Surveillance. IEEE 12th International Conference on Computer Vision (ICCV), 1639-1646 (2009)
[2] Chu, K.Y., Kuo Y.H., Hsu W.H. Real-Time Privacy-Preserving Moving Object Detection in the Cloud. ACM Multimedia, 597-600 (2013)



The **histogram** of the input frame before and after encryption.



Decrypted with wrong key. First line: using our method we slightly change the key and get the wrong decrypted result. Second line: using a standard random function,the sensitivity of the secret key x in the confusion step as a example is not well, and is much easier be attacked by the bruteforce attack.



The **correlation** of the input frame before and after encryption in 3 directions (horizontal, vertical and diagonal).

videos/frames	Our Method	Method of [2]
backdoor/1816	0.795028	**0.807418**
busstation/1111	**0.94699**	0.939721
cubicle/2811	**0.974191**	0.973237
highway/1179	**0.936468**	0.929562
office/629	**0.901332**	0.897626
pedestrians/921	**0.988313**	0.984055
sofa/628	0.944334	**0.945863**

The **comparison** of the correctness rate with [2] in seven video sequences. The overall correctness rate in the cipher videos of our method is only slightly higher than that of [2] to some extend.