Privacy-Preserving Anomaly Detection in Bitstream Video based on Gaussian Mixture Model

Yike Chen*, Yuru Song*, Peijia Zheng*, Yusong Du* and Weiqi Luo *
* Sun Yat-sen University, Guangzhou, China
E-mail: zhpj@mail.sysu.edu.cn

***Abstract*—Advanced computer vision technologies enable nearly real-time intelligent monitoring of homes by detecting anomalies such as falls and incidents of domestic violence, thereby enhancing home safety. Leveraging affordable home cameras and cloud computing services, this technology offers significant societal benefits. However, privacy concerns present challenges for its deployment in real-world settings. This paper introduces a method for anomaly detection in encrypted bitstream videos. By analyzing the video compression standard, we incorporate new appearance information that captures variations in object sizes alongside existing motion information, enhancing anomaly detection capabilities. We also refine the feature extraction process to reduce the impact of noise. Finally, we employ Gaussian Mixture Model to model the probability distribution of the data for effective anomaly detection. Experimental results demonstrate that our proposed privacy-preserving anomaly detection method achieves a commendable balance between protecting privacy and maintaining high detection performance.**

I. Introduction

As camera prices continue to decrease, video surveillance has become increasingly prevalent, with more individuals using it to ensure their safety, thereby driving the demand and development of private smart monitoring services. However, managing the enormous data volume of surveillance videos and performing intelligent video analysis on resource-constrained local devices is challenging. Cloud servers offer powerful computing capabilities, mature video analysis models, and are willing to provide users with various video analysis and processing tasks [1] for a proper fee.

However, surveillance videos, especially home surveillance videos, contain a significant amount of privacy-sensitive information such as personal identities, family activities, and daily routines. The leakage of this privacy information poses security risks to users of cloud services, making them reluctant to upload home surveillance videos. As a result, cloud servers face the challenging dual responsibility of respecting personal privacy while enhancing the user experience.

Processing videos in the encrypted domain [2] is one of the promising solutions to the aforementioned issues. Users encrypt videos before uploading, where the cloud servers process the encrypted videos and return the processed results to users. However, directly encrypting raw videos entails significant computational costs. Surveillance videos are akin to continuous movies, generating a vast number of frames daily. Users may not have adequate computational resources to encrypt raw videos. Methods proposed by [3]–[6] advocate encrypting compressed videos, greatly improving the efficiency of video encryption and analysis. These methods compress videos first and then encrypt the compressed bitstream, reducing the amount of data that needs encryption and thereby significantly lowering encryption costs for users.

Existing encrypted domain video analytics, such as moving object detection, have been extensively researched [4]. However, there is limited research on anomaly detection in encrypted bitstream videos. Guo *et al.* [6] used motion information from bitstreams and Adaptive Kernel Density Estimation (AKDE) for anomaly detection, but their focus solely on motion information is insufficient for complex scenes.

In this paper, we build on [6] by introducing new appearance features that partially capture changes in object size, thereby complementing motion information to enhance anomaly detection. We observed significant noise in the information extracted from encrypted videos, prompting us to redesign our feature extraction method to mitigate noise impact. To address the challenges of AKDE in handling high-dimensional data within encrypted compressed videos, we employ a Multivariate Gaussian Mixture Model (GMM) [7] for anomaly detection. By inputting high-dimensional features, the model learns intricate relationships between different features, thereby achieving better anomaly detection performance. The main contributions of this paper are as follows:

1. We introduce a new encrypted video information to complement motion information in anomaly detection by partially representing changes in object appearance.
2. We propose new feature extraction methods and use the Multivariate GMM to better capture potential relationships between features.
3. Experimental results show that the proposed privacy-preserving anomaly detection method achieves higher detection accuracy than existing state-of-the-art methods on multiple datasets.

II. Related Works

A. Anomaly Detection Methods

Traditional anomaly detection methods often utilize handcrafted features, such as the hierarchical Bayesian model [8], mixture of dynamic textures [9], and sparse reconstruction [10]. Recently, Hasan *et al.* [11] used a fully convolutional autoencoder to learn normal patterns. To address the issue that both normal and abnormal events are reconstructed well, Gong *et al.* [12] introduced a memory-augmented deep autoencoder.

Yang *et al.* [13] used keyframes to restore video events and explore temporal relationships within the video. Furthermore, to enhance the speed of anomaly detection, some methods operate directly in the compressed domain [14], [15]. However, these methods neglect the need for privacy preservation in anomaly detection.

B. Privacy-Preserving Anomaly Detection Methods

Cheng *et al.* [16] proposed a secure video anomaly detection scheme using secret sharing and Bloom filter, but its generalizability is limited. Yan *et al.* [17] masked privacy-sensitive regions in decompressed video data to protect privacy, but their approach is computationally complex and offers limited privacy protection. Allahham *et al.* [18] utilized federated learning techniques, but the method incurs high computational costs, making it unsuitable for real-time processing. Guo *et al.* [6] proposed an anomaly detection scheme for encrypted compressed videos, but their scheme lacks robustness and shows a significant performance drop in complex scenarios. Therefore, designing a real-time and robust anomaly detection scheme on encrypted video remains a challenging task.

III. Problem Statement

A. System Model

As illustrated in fig. 1, our encrypted video anomaly detection system involves two main entities: The data owner, $\mathcal{O}$, possesses the original video data but has limited storage and computational capabilities. The cloud computing service provider, $\mathcal{S}$, acts as a third-party entity offering robust storage and computational resources at a reasonable cost to $\mathcal{O}$ for video anomaly detection cloud services. By outsourcing the video data, $\mathcal{O}$ can access the video data anytime, anywhere using terminal devices such as smartphones and benefit from the video anomaly detection cloud services. The cloud server $\mathcal{S}$ performs anomaly detection tasks on the encrypted bitstream video and returns the detection results to $\mathcal{O}$. $\mathcal{O}$ decrypts the results using a key to obtain the anomaly detection outcomes.

B. Threat Model

We adopt the setting established in similar literature [19], assuming $\mathcal{S}$ to be a semi-honest adversary. This means $\mathcal{S}$ will follow the protocol but may attempt additional computations to access private information from the video. We employ the video encryption scheme from [5], [20], which has been extensively validated for its ability to protect video privacy. Only $\mathcal{O}$ possesses the decryption key; $\mathcal{S}$ does not have access to the key. During the execution of anomaly detection, $\mathcal{O}$ does not interact with $\mathcal{S}$, preventing leakage of the video's key information. Therefore, $\mathcal{S}$ cannot obtain sensitive information from the encrypted video. Our encrypted video anomaly detection system ensures security.

IV. Proposed Privacy-Preserving Anomaly Detection Scheme over Encrypted Video

The proposed privacy-preserving video anomaly detection framework is illustrated in fig. 2. To facilitate our discussion, we focus on H.264 videos and use a selective encryption scheme to encrypt the compressed bitstreams, as detailed in [5]. It should be noted that video encryption algorithms compatible with other formats [21], [22] are also applicable to our framework. Our framework directly analyzes encrypted video bitstreams using compression standards, extracting entropy-coded data related to anomalies. From these entropy-coded data, we employ statistical analysis methods to extract abnormal behavioral characteristics for each frame in the compressed domain. These extracted compressed domain features serve as representations of each frame. We utilize a Multivariate GMM to construct a probability distribution of abnormal events and employ this distribution to detect abnormal frames. We will now elaborate on each specific part.

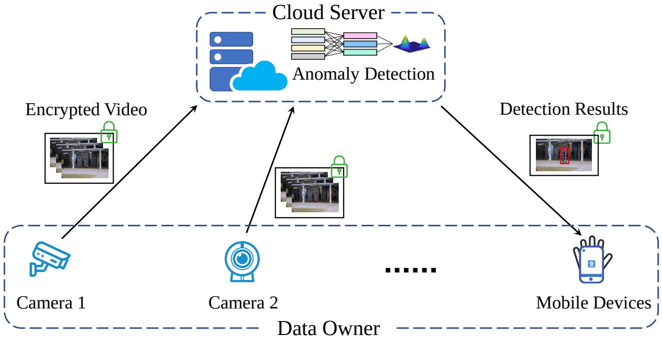


Fig. 1. System Model

A. Abnormal Information Estimation on Encrypted Video

By analyzing video compression standards, we directly extract syntax elements from compressed entropy-coded data to reveal their correlations with abnormal behaviors. Specifically, in addition to the three types of compressed domain abnormal behavior information identified in [6], we introduce a new type: residual density, which effectively captures appearance changes associated with object size variations. Next, we will provide detailed explanations of these four types of compressed domain abnormal behavior information.

1) Consumed Bits: In common video compression standards, the basic coding unit is the macroblock. Video compression aims to represent more content with less data, so repetitive content is predicted to reduce the number of bits per macroblock. In contrast, during abnormal events, affected macroblocks often consume more bits due to unusual object motion. Thus, the bit consumption of macroblocks implicitly reveals information about abnormal behaviors.

Taking the example of the H.264 encoding scheme, its slice data syntax structure is represented as:

$$\left\{\mathbf{SH}, \mathbf{MB}'_1, \mathbf{MB}'_2, \mathbf{SR}_3, \cdots, \mathbf{MB}'_i, \cdots, \mathbf{MB}'_n\right\},$$

where $\mathbf{SH}$ denotes slice header, $\mathbf{MB}'_i$ represents the i-th encrypted macroblock in the slice, and $\mathbf{SR}_3$ indicates that the 3rd macroblock is a skipped macroblock.

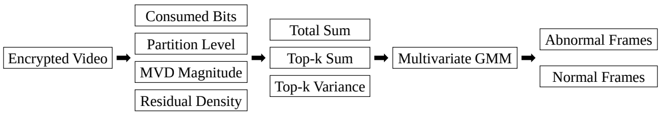


Fig. 2. Overview of proposed framework

The number of bits consumed $b_i(t)$ by the i-th macroblock in frame t is calculated as:

$$b_i(t) = \begin{cases} 0, & \text{skipping } \mathbf{MB}_i, \\ \mathcal{S}(\mathbf{SR}_{i+1}) - \mathcal{S}(\mathbf{MB}'_i), & \text{skipping } \mathbf{MB}_{i+1}, \\ \mathcal{S}(\mathbf{MB}'_{i+1}) - \mathcal{S}(\mathbf{MB}'_i), & \text{otherwise}, \end{cases} \quad (1)$$

where $\mathcal{S}(\cdot)$ denotes the starting address in the bitstream.

2) Partition Level: In compressed videos, b_i in I-frames is generally significantly higher than in P-frames. Additionally, the aggregation or dispersion of normal objects affects b_i. Consequently, relying solely on b_i can result in a high false positive rate in anomaly detection. Typically, 16×16 macroblocks can be subdivided into smaller sub-MBs. To better capture motion during decoding, macroblocks in abnormal areas of P-frames may be divided into more smaller sub-MBs. For instance, a regular macroblock might be divided into two 8×16 sub-MBs, whereas a macroblock in an abnormal area may be divided into sixteen 4×4 sub-MBs. The partition level $p_i(t)$ of a macroblock is calculated as:

$$p_i(t) = \mathcal{P}\left(\mathbf{ST}_i,\, \mathbf{MT}_i\right), \quad (2)$$

where $\mathbf{ST}_i$ denotes the slice type, $\mathbf{MT}_i$ denotes the MB type, and $\mathcal{P}(\cdot,\cdot)$ calculates the number of its sub-MBs.

3) MVD Magnitude: Motion vectors (MVs) are widely used features in bitstream-based video anomaly detection [23], [24]. However, MVs in encrypted bitstream videos are unavailable. This paper selects MVD magnitude as a substitute for MV. In bitstream videos, MVs are not directly encoded but indirectly encoded by predictive motion vectors and actual motion vector differences (MVDs). MVD consists of horizontal $\mathrm{MVD_x}$ and vertical $\mathrm{MVD_y}$ components. We use the magnitude of MVDs of a macroblock as an estimate for MVs to provide richer motion information for anomaly detection. The MVD magnitude $m_i(t)$ of a macroblock is calculated as

$$m_i(t) = \sum_{k \in \mathbb{I}_{mv}(t)} \lfloor \sqrt{L_{\mathrm{x}}^k(t)^2 + L_{\mathrm{y}}^k(t)^2} \rfloor, \quad (3)$$

where $\mathbb{I}_{mv}(t)$ represents the set of addresses for 4×4 blocks within the macroblock, and $L_{\mathrm{x}}^k(t)$ and $L_{\mathrm{y}}^k(t)$ represent the code word lengths of $\mathrm{MVD_x}$ and $\mathrm{MVD_y}$, respectively.

4) Residual Density: Residual density has been proposed for detecting moving objects in videos [5]. We introduce residual density as appearance information to complement motion-based anomaly detection. During video compression, residual data undergoes DCT transformation, quantization, and then entropy coding. Given a macroblock, if its DCT block $U_{dct}^k(t)$ contains δ_k 4×4 blocks, the density of non-zero coefficients $d_i(t)$ is calculated as:

$$d_i(t) = \sum_{k \in \mathbb{I}_{dct}(t)} \delta_k \cdot \lfloor \frac{\rho\left(U_{dct}^k(t)\right)}{\sqrt{\delta_k}} \rfloor, \quad (4)$$

where $\rho(\cdot)$ denotes the number of non-zero coefficients in a DCT block, and $\mathbb{I}_{dct}(t)$ denotes the set of DCT block addresses within the macroblock.

B. Feature Extraction from Encrypted Compressed Videos

We extracted four types of estimated information from encrypted compressed videos: consumed bits $b_i(t)$, partition level $p_i(t)$, MVD magnitude $m_i(t)$, and residual density $d_i(t)$. We use $\mathbf{f}_i(t)$ to denote any one of these four types of information for the i-th macroblock in frame t, and $\mathbb{I}(t)$ denotes the set of macroblock addresses in frame t. Next, we consider these informational features at the frame level, including the sum of all macroblocks, the sum of representative macroblocks, and the variance of representative macroblocks.

1) Total Sum: For the estimated information $\mathbf{f}(t)$, we compute the total sum of all macroblocks in a frame as a rough feature of that frame, which is given by:

$$\lambda_t^{\mathbf{f}} = \sum_{i \in \mathbb{I}(t)} \mathbf{f}_i(t), \quad (5)$$

2) Top-k Sum: During abnormal events, not all macroblocks in a frame are related to anomalies. In fact, abnormal events only involve a small number of macroblocks with prominent features. Based on this observation, we isolate the top α ranked macroblocks in a frame, termed representative macroblocks. Considering only the information from representative macroblocks can mitigate noise interference from non-representative macroblocks. The sum of information extracted from representative macroblocks in a frame is considered as a refined feature, which is computed by:

$$\delta_t^{\mathbf{f}} = \sum_{i \in \mathbb{I}_\alpha(t)} \mathbf{f}_i(t), \quad (6)$$

where $\mathbb{I}_\alpha(t)$ denotes the set of addresses of the top α ranked macroblocks in a frame.

3) Top-k Variance: Based on the identified representative macroblocks, in contrast to [6], we consider the variance among representative macroblocks. Since anomalies result in increased variance among representative macroblocks, this variance can be utilized as an additional refined feature for anomaly detection, which is computed by:

$$\phi_t^{\mathbf{f}} = \frac{1}{|\mathbb{I}_\alpha(t)| - 1} \sum_{i \in \mathbb{I}_\alpha(t)} \left(\mathbf{f}_i(t) - \bar{\mathbf{f}}_i(t)\right)^2, \quad (7)$$

where $\bar{\mathbf{f}}_i(t)$ denotes the mean value of $\mathbf{f}_i(t)$ over all representative macroblocks.

Thus, we obtain a total of 12 features:

$$[\lambda_t^c, \lambda_t^p, \lambda_t^m, \lambda_t^d, \delta_t^c, \delta_t^p, \delta_t^m, \delta_t^d, \phi_t^c, \phi_t^p, \phi_t^m, \phi_t^d],$$

which is represented collectively as $[\Lambda_t, \Delta_t, \Phi_t]$. All features together form a high-dimensional feature that represents the frame and is used for subsequent anomaly detection.

C. Anomaly Detection Algorithm

We use the Multivariate GMM for its ability to capture normal patterns in the data, thereby fully exploiting the potential relationships between different features. Specifically, during the training phase, we use the extracted features mentioned above to construct a probability density function using Multivariate GMM based on the training dataset, i.e.,

$$\begin{aligned} \mathbf{x}_t &= [\Lambda_t, \Delta_t, \Phi_t]^T, \\ p_{\mathcal{M}}(\mathbf{x}_t) &= \sum_{k=1}^{K} \pi_k \cdot \mathcal{N}\left(\mathbf{x}_t \mid \boldsymbol{\mu_k}, \Gamma_k\right), \\ \sum_{k=1}^{K} \pi_k &= 1, \quad \pi_k > 0, \end{aligned} \tag{8}$$

where $\mathcal{N}\left(\mathbf{x}_t \mid \boldsymbol{\mu_k}, \Gamma_k\right)$ represents the probability density function of the k-th Gaussian component with mean $\boldsymbol{\mu_k}$ and covariance matrix Γ_k. K denotes the number of Gaussian components, and π_k represents the weight for the k-th Gaussian component.

The obtained GMM provides a representation of the probability of normal behaviors in the video data. During the testing phase, using the high-dimensional features $[\lambda_t, \Delta_t, \Phi_t]$ related to anomalies in each frame of the test video and the obtained GMM, we can compute the probability p_t of an event occurring in that frame. Frames with low probabilities are identified as potential anomalies, indicating deviations from the learned normal patterns. Based on the probability of event occurrence, the confidence s_t that a frame is an anomaly is computed as follows:

$$s_t = 1 - p_{\mathcal{M}}(\mathbf{x}_t). \tag{9}$$

After obtaining the anomaly confidence for each frame, and considering that abnormal events in videos typically develop gradually, we apply median filtering to the confidence values. Finally, by setting an appropriate threshold θ, frames with confidence $s_t \geq \theta$ are classified as anomalous frames, while frames with $s_t < \theta$ are classified as normal frames.

V. Experimental Results

A. Settings

The proposed privacy-preserving anomaly detection scheme was tested on the UCSD Ped1 [9], UCSD Ped2 [9], and CUHK Avenue [25] datasets. For convenience in processing, we resized the video resolution to 1280×720 and set the frame rate to 25 fps. The feature extraction was performed with $\alpha = 4\%$. All experimental results were obtained using MATLAB on a 64-bit Windows 10 PC with Intel(R) Core(TM) i5-10400 CPU @ 2.90GHz and 16GB memory.

B. Datasets

1) UCSD Peds: The UCSD Peds [9] dataset is captured on a campus, comprising two sub-datasets: UCSD Ped1 and UCSD Ped2. UCSD Ped1 includes 70 videos of size 238×158, captured from a camera with a perpendicular viewpoint to the road, resulting in significant variations in the sizes of moving foreground objects. UCSD Ped2 includes 28 videos of size 360×240, captured from a camera with a parallel viewpoint to the road, resulting in smaller variations in the sizes of moving foreground objects. Anomalies in these datasets are related to unexpected objects and pedestrian behaviors.

TABLE I
AUC and EER comparison with different methods.

Dataset	*UCSD Ped1*		*UCSD Ped2*		*CUHK Avenue*	
Method	AUC↑	EER↓	AUC↑	EER↓	AUC↑	EER↓
Kmeans [26]	0.59	0.44	0.57	0.53	0.70	0.36
iForest [27]	0.70	0.37	0.76	0.29	0.73	0.32
Guo [6]	0.69	0.38	-	-	0.79	0.29
Proposed	**0.72**	**0.35**	**0.81**	**0.24**	**0.80**	**0.26**

TABLE II
Performance comparison between plaintext and encrypted domains. CB and FS represent inputs of compressed bitstream and frame sequence, respectively.

Plaintext Domain						
Method	Input	*UCSD Ped1*		*CUHK Avenue*		
		AUC↑	EER↓	AUC↑	EER↓	
USTN-DSC [13]	FS	-	-	**0.90**	-	
Kiryati [14]	CB	-	-	0.72	0.34	
Biswas [15]	CB	**0.79**	**0.24**	0.69	0.34	
Guo [6]	CB	0.69	0.38	0.79	0.29	
Proposed	CB	0.72	0.35	0.80	0.26	

Encrypted Domain					
Method	Input	*UCSD Ped1*		*CUHK Avenue*	
		AUC↑	EER↓	AUC↑	EER↓
MVDM [6]	CB	0.62	0.43	0.65	0.40
MVDF [6]	CB	0.58	0.45	0.53	0.47
Guo [6]	CB	0.69	0.38	0.79	0.29
Proposed	CB	**0.72**	**0.35**	**0.80**	**0.26**

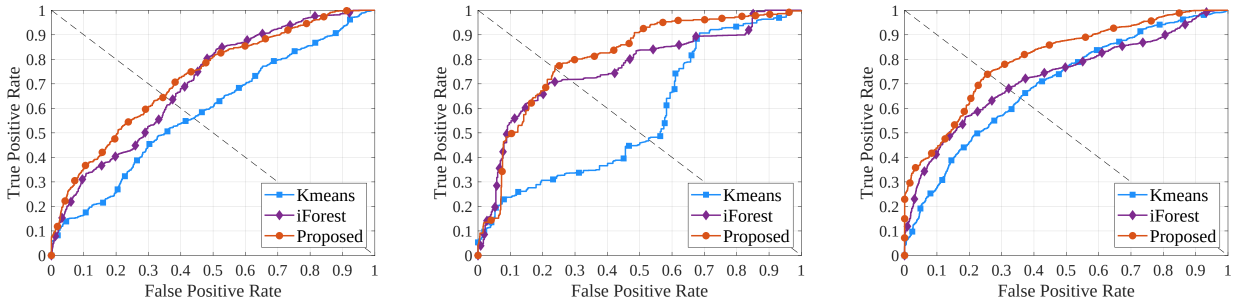
2) CUHK Avenue: The CUHK Avenue [25] dataset is captured on a campus avenue using a camera with slight jitter. It comprises 28 videos with a resolution of 640×320. Anomalies in this dataset mainly involve abnormal movements such as sudden running and littering.

C. Comparison results

We quantitatively evaluate the video results using widely used metrics in video anomaly detection: Area Under the Curve (AUC) and Equal Error Rate (EER), which can be obtained from receiver operating characteristic (ROC) curves. A higher AUC indicates better anomaly detection performance, whereas a lower EER indicates the opposite. Quantitative comparisons of the proposed method's detection performance in the encrypted domain with other methods are shown in Table I. Corresponding ROC curves are depicted in Figure 3.

The proposed method achieves AUC values of 0.72, 0.81, and 0.80, and EER values of 0.35, 0.24, and 0.26 on Ped1, Ped2, and Avenue datasets, respectively. Comparisons with Guo [6] show improvements in AUC by 3% and 1% on UCSD Ped1 and CUHK Avenue datasets, respectively. Quantitative comparisons demonstrate that the proposed method surpasses Guo's method, achieving more robust anomaly detection performance.

In Table I, we also compare our method with other methods such as Kmeans [26] and iForest [27], using features extracted



(a) UCSD Ped1

(b) UCSD Ped2

(c) CUHK Avenue

Fig. 3. Comparison of ROC curves for different methods.

in this paper for anomaly detection. On the UCSD Ped2 dataset, our method achieves an AUC that is 5% higher than that of the second-best method. Experimental results indicate that the proposed method achieves state-of-the-art performance on all three anomaly detection datasets.

In Figure 3, proposed method consistently positions above and to the left of the ROC curves, indicating better representation of data probability distributions and thorough exploration of potential relationships between different features, resulting in robust anomaly detection capabilities.

We also compare our method with methods in plaintext domain such as Kiryati [14], Bitwas [15], and USTN-DSC [13]. Experimental results, as shown in Table II, indicate that USTN-DSC [13], a deep learning-based method on frame sequences, achieved the best anomaly detection performance on the CUHK Avenue dataset due to its powerful representation capabilities using neural networks. However, our method performs the best among methods in the compressed domain and shows comparable performance with deep learning-based methods. On UCSD Ped1, our method also demonstrates performance comparable to similar methods in the compression domain.

For the encrypted video, we compare our results with those reported in [6], demonstrating that proposed method achieves the best anomaly detection performance compared to other methods.

D. Ablation Study

We conduct ablation experiments to evaluate the effectiveness of residual density information for representing appearance and top-k variance features. As shown in Figure III, removing variance features and appearance information results in a reduced AUC of 0.76 and an increased EER of 0.30 on UCSD Ped2, highlighting the importance of these features. Incorporating top-k variance with the three types of estimated information improves the AUC to 0.77 and reduces the EER to 0.28 on UCSD Ped2, demonstrating the efficacy of including variance features.

Adding appearance information alone increases the AUC to 0.78 and decreases the EER to 0.28, highlighting the necessity of including appearance information. By introducing appearance information and redesigning feature extraction, the model achieves an AUC of 0.81 and an EER of 0.24 on UCSD Ped2, demonstrating that both enhancements contribute to improving model performance.

TABLE III
ABLATION STUDIES ON UCSD PED2 DATASET

Method		UCSD Ped2	
Top-k Variance	Residual Density	AUC↑	EER↓
✘	✘	0.76	0.30
✔	✘	0.77	0.28
✘	✔	0.78	0.28
✔	✔	**0.81**	**0.24**

TABLE IV
INFLUENCE OF DIFFERENT α VALUES ON AUC AND EER ON UCSD PED2 DATASET

α	1%	2%	3%	4%	5%	6%	10%
AUC↑	0.74	0.78	0.79	**0.81**	0.80	0.79	0.73
EER↓	0.33	0.26	0.30	**0.24**	**0.24**	0.27	0.36

Furthermore, we examine the impact of different values of α on the performance of representative macroblocks in anomaly detection, as shown in Figure IV. We observe that the choice of α significantly affects model performance. Increasing α from a small value of 1% to 4% improves the AUC from 0.74 to 0.81. However, further increasing α results in decreased model performance because macroblocks selected by a too large α being influenced more by noise from regions unrelated to anomalies. Therefore, selecting an appropriate α value is crucial for achieving effective anomaly detection performance.

VI. CONCLUSIONS

We propose an anomaly detection algorithm for encrypted bitstream videos based on the Multivariate GMM. To address the limitation of existing methods, where motion information alone is insufficient for representing anomalies, we introduce complementary appearance information to enhance detection capabilities. We have redesigned the feature extraction method to reduce the impact of noise in the encrypted domain on feature extraction. Additionally, by integrating the extracted features with the Multivariate GMM, we develop a more robust anomaly detection model for encrypted bitstream videos, particularly in complex scenarios. In future work, we plan to explore end-to-end anomaly detection methods based on deep learning for encrypted bitstream videos, aiming to reduce reliance on manual feature design.

References

[1] B. Alouffi, M. Hasnain, A. Alharbi, W. Alosaimi, H. Alyami, and M. Ayaz, "A systematic literature review on cloud computing security: Threats and mitigation strategies," *IEEE Access*, vol. 9, pp. 57 792–57 807, 2021.

[2] C. Lin, K. Muchtar, J. Lin, Y. Sung, and C. Yeh, "Moving object detection in the encrypted domain," *Multim. Tools Appl.*, vol. 76, no. 7, pp. 9759–9783, 2017.

[3] X. Tian, P. Zheng, and J. Huang, "Robust privacy-preserving motion detection and object tracking in encrypted streaming video," *IEEE Trans. Inf. Forensics Secur.*, vol. 16, pp. 5381–5396, 2021.

[4] C. Liu, X. Ma, S. Cao, J. Fu, and B. B. Zhu, "Privacy-preserving motion detection for HEVC-compressed surveillance video," *ACM Trans. Multim. Comput. Commun. Appl.*, vol. 18, no. 1, pp. 1–27, 2022.

[5] X. Tian, P. Zheng, and J. Huang, "Secure deep learning framework for moving object detection in compressed video," *IEEE Trans. Dependable Secur. Comput.*, vol. 21, no. 4, pp. 2836–2851, 2024.

[6] J. Guo, P. Zheng, and J. Huang, "Efficient privacy-preserving anomaly detection and localization in bitstream video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 9, pp. 3268–3281, 2020.

[7] M. Varma and D. Ray, "Learning the discriminative power-invariance trade-off," in *ICCV*, 2007.

[8] X. Wang, X. Ma, and E. Grimson, "Unsupervised activity perception by hierarchical bayesian models," in *CVPR*, IEEE, 2007.

[9] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos, "Anomaly detection in crowded scenes," in *CVPR*, 2010.

[10] Y. Cong, J. Yuan, and J. Liu, "Sparse reconstruction cost for abnormal event detection," in *CVPR*, 2011.

[11] M. Hasan, J. Choi, J. Neumann, A. K. Roy-Chowdhury, and L. S. Davis, "Learning temporal regularity in video sequences," in *CVPR*, 2016.

[12] D. Gong, L. Liu, V. Le, *et al.*, "Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection," in *ICCV*, 2019.

[13] Z. Yang, J. Liu, Z. Wu, P. Wu, and X. Liu, "Video event restoration based on keyframes for video anomaly detection," in *CVPR*, 2023.

[14] N. Kiryati, T. Riklin-Raviv, Y. Ivanchenko, and S. Rochel, "Real-time abnormal motion detection in surveillance video," in *ICPR*, 2008.

[15] S. Biswas and R. V. Babu, "Anomaly detection in compressed H.264/AVC video," *Multim. Tools Appl.*, vol. 74, no. 24, pp. 11 099–11 115, 2015.

[16] H. Cheng, X. Liu, H. Wang, Y. Fang, M. Wang, and X. Zhao, "Securead: A secure video anomaly detection framework on convolutional neural network in edge computing environment," *IEEE Trans. Cloud Comput.*, vol. 10, no. 2, pp. 1413–1427, 2022.

[17] J. Yan, Y. Yang, and S. M. Naqvi, "Object detection oriented privacy-preserving frame-level video anomaly detection," in *ICASSP*, 2024.

[18] A. Al-lahham, M. Z. Zaheer, N. Tastan, and K. Nandakumar, "Collaborative learning of anomalies with privacy (clap) for unsupervised video anomaly detection: A new baseline," in *CVPR*, 2024.

[19] S. Hu, Q. Wang, J. Wang, Z. Qin, and K. Ren, "Securing SIFT: privacy-preserving outsourcing computation of feature extractions over encrypted image data," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 3411–3425, 2016.

[20] F. Peng, X. Gong, M. Long, and X. Sun, "A selective encryption scheme for protecting H.264/AVC video in multimedia social network," *Multim. Tools Appl.*, vol. 76, no. 3, pp. 3235–3253, 2017.

[21] Y. Wang, M. O'Neill, and F. Kurugollu, "A tunable encryption scheme and analysis of fast selective encryption for CAVLC and CABAC in H.264/AVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 9, pp. 1476–1490, 2013.

[22] D. Xu, R. Wang, and Y. Q. Shi, "Data hiding in encrypted H.264/AVC video streams by codeword substitution," *IEEE Trans. Inf. Forensics Secur.*, vol. 9, no. 4, pp. 596–606, 2014.

[23] S. Biswas and R. V. Babu, "Real time anomaly detection in H.264 compressed videos," in *NCVPRIPG*, 2013.

[24] S. Biswas and R. V. Babu, "Anomaly detection in compressed H.264/AVC video," *Multim. Tools Appl.*, vol. 74, no. 24, pp. 11 099–11 115, 2015.

[25] C. Lu, J. Shi, and J. Jia, "Abnormal event detection at 150 FPS in MATLAB," in *ICCV*, 2013.

[26] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *5th Berkeley Symp. Math. Stat. Prob.*, 1967.

[27] F. T. Liu, K. M. Ting, and Z. Zhou, "Isolation-based anomaly detection," *ACM Trans. Knowl. Discov. Data*, vol. 6, no. 1, pp. 1–39, 2012.