Shahid Chamran University of Ahvaz

Iranian Association of Electrical and Electronics Engineers

Journal of Applied Research in Electrical Engineering

E-ISSN: 2783-2864
P-ISSN: 2717-414X
Homepage: https://jaree.scu.ac.ir/

Research Article

Smart AI-based Video Encoding for Fixed Background Video Streaming Applications

Mohammadreza Ghafari[1], Abdollah Amirkhani[2,*], Elyas Rashno[3], and Shirin Ghanbari[4]

[1]*Department of Electrical Engineering, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran*

[2]*School of Automotive Engineering, Iran University of Science and Technology, Tehran, Iran*

[3]*Department of Computer Engineering, Iran University of Science and Technology, Tehran, Iran*

[4]*Department of Computer Science and Electronic Engineering, University of Essex, Essex, UK*

* *Corresponding Author: amirkhani@ieee.org*

Abstract: This paper is an extension of our previous research on presenting a novel Gaussian Mixture-based (MOG2) Video Coding for CCTVs. The aim of this paper is to optimize the MOG2 algorithm used for foreground-background separation in video streaming. In fact, our previous study showed that traditional video encoding with the help of MOG2 has a negative effect on visual quality. Therefore, this study is our main motivation for improving visual quality by combining the previously proposed algorithm and color optimization method to achieve better visual quality. In this regard, we introduce Artificial Intelligence (AI) video encoding using Color Clustering (CC), which is used before the MOG2 process to optimize color and make a less noisy mask. The results of our experiments show that with this method the visual quality is significantly increased, while the latency remains almost the same. Consequently, instead of using morphological transformation which has been used in our past study, CC achieves better results such that PSNR and SSIM values have been shown to rise by approximately 1dB and 1 unit respectively.

Keywords: Artificial Intelligence, video coding, background subtraction, color clustering, mixture of gaussian model.

Article history
Received 14 March 2022; Revised 28 June 2022; Accepted 06 August 2022; Published online 28 February 2023.

How to cite this article
M. Ghafari, A. Amirkhani, E. Rashno, and S. Ghanbari, "Smart AI-based video encoding for fixed background video streaming applications," *J. Appl. Res. Electr. Eng.*, vol. 2, no. 1, pp. 37-44, 2023. DOI: 10.22055/jaree.2022.40295.1051

1. Introduction

Artificial Intelligence (AI), as a branch of computer science, has been able to affect people's lives. More specifically, AI has provided algorithms for computer professionals that were previously difficult to implement. Nowadays, AI is widely used within image processing applications including face recognition or object detection. However, due to the complexities and challenges of video compression, the feasibility of using these algorithms in video compression has been less used. One of the main applications of video compression is in traffic cameras and video surveillance (CCTV) that are continuously streaming data to their users. Since these CCTV cameras are streaming video to their destinations, day and night, it seems necessary to use a method that can take up less bandwidth. Practically, CCTV is one of the most important technologies in the security field.

Today, due to network limitations, CCTVs are applied at variant private and public places [1]. CCTV is used within real applications including identity offenders and prevent crime purposes [2]. An efficient CCTV technology was proposed by Harikrishnan [3] for business activities. In this paper, the author studied the impact of area condition on the CCTV technology and its utilization, and by using data analysis statistics, they proposed practical ideas in improving CCTV cameras. To address its challenges, Carli [4] tried to introduce CCTVs as a tool for crime prevention. Also, Goold [5] tried to address challenges in regards to CCTV responsibilities. On the other hand, Welsh and Farrington [6]

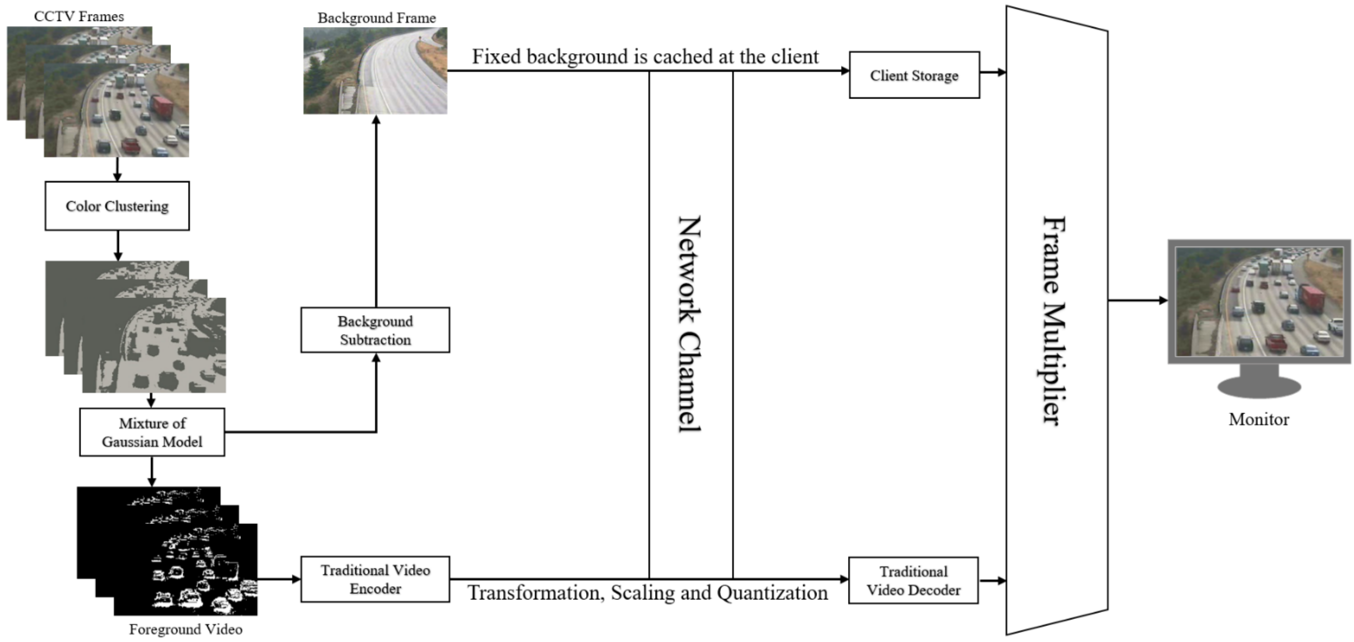

Fig. 1: Our proposed structure for AI-based video coding.

suggested a new method to improve the quality of CCTV and its efficiency. In order to use the CCTV camera as an analytical tool with image processing, the important features of 3D surveillance are extracted by the OpenCV library to introduce a new method of video analytics [7]. Another use of CCTV is as an analytical tool as proposed by Kumar et al. [8], whereby they define a new model for shot detection in MATLAB. The use of Digital Video Recorder (DVR) and Network Video Recorder (NVR), both are major components of CCTV cameras that perform compression and video capturing. In this way, a video compression method specifically for CCTVs can reduce the bandwidth but has an impact on image quality [9]. Another challenging issue with CCTV cameras is the limitation of its storage, which has been the subject of several studies. The main reason of this challenge is due to the large video files which are constantly recorded all the time. To overcome this challenge, cloud-based services are now available that can provide unlimited storage space [10]. Also, there are methods relating to storage optimization as to optimize CCTV storage and remove similar frame footage. This process may lead to the reduction of video size [11].

Today, video image enrichment in CCTV cameras is performed by artificial intelligence algorithms. This enrichment is achieved through visual models, the most well-known being object tracking, pattern recognizing and face detection [12]-[14]. CCTV image transfer optimization depends on video compression algorithms. The main reason for this claim is that by optimizing the video compression, the volume of the transmitted video is reduced and in addition to requiring less storage space, it also occupies less bandwidth.

Among conventional video coding techniques, block-based estimation is a practical method to reduce temporal redundancy in video streaming. In this way, a background-base model overlaying on High Efficiency Video Coding (HEVC) in a surveillance video streaming can reduce the network bandwidth occupation [15]. In a similar work, Zhang et al. [16] proposed a background-based method that tries to separate the background from the foreground for achieving better accuracy in surveillance video coding. In this paper, the authors have used two methods namely, Background Reference Prediction (BRP) and Background Differences Prediction (BDP) to encode the video frames. In line with image transfer problems, Guo et al. [17] proposed a novel algorithm to overcome the high-quality of large videos files during the transformation process by background removal. Predicting adaptive motion units, namely, PA-search is a novel predictive method that has been proposed for motion estimation [18]. In this approach the author defined a smart algorithm for searching the motions of video frames while keeping algorithm compression complexities at a minimum. In another similar work, Kim and Lee [19] have proposed an efficient method to separate the background from foreground based on the integer motion estimation method for video compression. This video coding system helps to reduce the external power consumption to the encoder side and optimize its efficiency. Double background-based coding, as proposed by Li et al. [20] is defined by two different background frames based on reconstructed frames and the original frames, simultaneously. This method, tested by HM14.0, helped to improve the compression performance by saving the average of 17.32 bit rate. The distance parameter is one of the challenges of CCTVs, which makes increases the complexity of face identification. To overcome this challenge, a two phased object recognition is defined by Celine and Agustin [21] that enhances the facial quality. At the first phase the dataset including faces are created, then at the second phase, good quality images are produced. Moreover, automatically detecting illegal activities such as weapon detection can be critical for security applications. However, as the detection of such objects is time consuming, it practically affects the efficiency of such systems. In this regard, the weapon detection system has been optimized using Yolov4 with an accuracy of 90% in detection [22]. Another detection system has been investigated by Powale et al. [23] that identifies people even in low resolution images. The results of the study show that the proposed detection system based on

Convolutional Neural Networks (CNNs), has been able to reach an accuracy of 94.55% for the TinyFace [24] dataset. For object detection, between Deep Learning models, YOLO has shown to outperform. Pillai et al. [25] have proposed a Mini-YOLO method that has comparable accuracy with YOLO, however, the model size and computational cost were reduced significantly. In another application, YOLO has been used for real-time applications due to its high speed in object detection. In this regard, a novel real-time approach based on machine learning and deep learning methods to detect human faces in CCTV images has been investigated by Rehmat Ullah et al. M. S. Pillai, G. Chaudhary, M. Khari, and R. G. Crespo, "Real-time image enhancement for an automatic automobile accident detection through CCTV using deep learning", *Soft Computing*, vol. 25, no. 18, pp. 11929-11940, 2021.

[26]. The novel part of this research refers to the pre-processing part that enhances the image quality for better detection. In a similar work, Pan et al. [27], use YOLOv3 with the help of the COCO dataset to achieve better detection from CCTV video streaming. The experimental results of this method showed that YOLOv3 has performed 44% accuracy during the day and 41% accuracy during the night.

Reviewing these works, have motivated us to define a novel video coding system, which can efficiently reduce the network bandwidth. The fundamental concept of this coding is referred to the background subtraction process, whereby the fixed background is sent through the channel only once. In this regard, it is cached at the client side and is multiplied to the moving objects.

The rest of the paper is arranged as follows. Section 2 describes the proposed scientific methodology and its experimental results are presented in Section 3. Finally, Section 4 concludes this paper and discusses future work.

2. Proposed Method

The proposed compression technique extends the authors previous work [28] based upon an improved Mixture of Gaussian Model, MOG2 algorithm to separate the background image from moving objects in the video, while for getting better visual quality, Color Clustering (CC) method is added before the MOG2 process. In fact, instead of compressing the entire video frame, this technique causes the video frames to be separated into moving objects and the background image and whereby the background image occupies the network bandwidth only once. In other words, the moving objects frames create the foreground video while the background image creates the background video. Our previous results showed that for using background subtraction, the produced mask from MOG2 is very noisy and it should be denoised using Morphological Transformation (MT). As the visual quality of this transformation is not satisfactory, the first step of the proposed architecture is to use CC as a clean mask.

Fig. 1 shows the overall process of compressing and retrieving video images on a Network Video Recorder (NVR). In this process, video frames are captured by the CC block for clustering the colors. Then it passes to MOG2 block to estimate the fixed background and separate foreground from background parts. At the end of this block, the background subtraction operation is started to detect and extract moving objects. It is necessary to pass one full cycle for achieving background and after the CC and MOG2 process, the fixed background image is cached at the client-side. At the server's side, the foreground video is going to be compressed by conventional video codecs. In our approach, we have used H.265, namely the HEVC video codec to compressed moving objects. At the end, to the decoder side, the stored background image and encoded foreground video are multiplied to achieve the full video frames including foreground and background. With these explanations, our proposed architecture consists of three main parts, each of which is discussed below.

2.1. Color Clustering

The idea of using CC is due to its capability to better separate the components of objects. In our study, as the foreground objects includes the moving objects, it would be more efficient to separate colors then applying the background subtraction method. In this regard, as we wanted to separate colors in order for having better masks, and we have just two kinds of objects, foreground and background objects, we have set CC=2 in order for getting a denoised mask. Clustering method has been done by k-means color quantization which in RGB channel we have used Euclidean distance for comparing and clustering the colors. For every colorful image we have three channels including red, green and blue which creates the color-based image. Since we are looking for the distance of each color, the distance between two colors calculates by (1):

$$d = \sqrt{(R_2 - R_1)^2 - (G_2 - G_1)^2 - (B_2 - B_1)^2} \tag{1}$$

where *R*,*G*,*B* are the representatives of the red, green and blue channels.

2.2. Gaussian Mixture Model

The basis of the MOG2 method is a pixel-by-pixel review in a video sequence in which motion information is obtained from differences between the frames. Using this method, moving objects are detected that can be a tool for creating a mask. Multiplying this mask throughout the video will separate the background image from moving objects. The process of creating the mask is relevant to Gaussian distribution.

For every Gaussian distribution η we have

$$\eta(t) = \frac{1}{\sqrt{2\pi\omega}} exp\left(-(t-\mu)^T \omega^{-1} (t-\mu)\right) \tag{2}$$

where μ is the mean vector and ω is the covariance matrix. If we assume that there are k Gaussian distributions according to each t variable, which refers to the frame sequences in the time domain, the Gaussian Mixture Model (GMM) is a mixture of k Gaussians that describes the random variable t.

This assumption leads to (3):

$$Pr(t) = \sum_{k=1}^{k=K} W_k * \eta(t) s.t{:} \sum_{k=1}^{k=K} W_k = 1 \tag{3}$$

where W_k is the k^{th} weight of the Gaussian model.

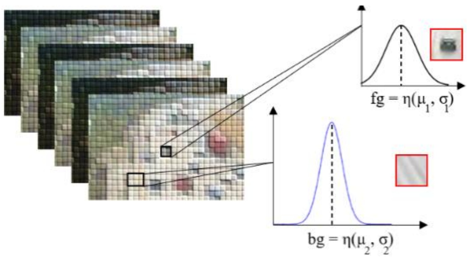

Fig. 2: Foreground and background Gaussian modelling.

Since the background frames appear frequently during video streaming, the ideal background Gaussian model should have a high weight with low variance. To select the first *n* Gaussians as Background Models (BM) we have:

$$BM = argmin_n\left(\sum_{k=1}^{n} W_k > T_n\right) \quad (4)$$

where T_n is the model selection threshold. The range of this threshold for MOG2 model is from 1 to 255. Based on Monte Carlo analysis as performed by Matczak et al. M. Ghafari, A. Amirkhani, E. Rashno, S. Ghanbariet, "Novel gaussian mixture-based video coding for fixed background video streaming," in *2022 12th Iranian/Second International Conference on Machine Vision and Image Processing (MVIP)*, Ahvaz, Iran, 2022.

[29] we set this variable threshold to 16, which exactly fits our requirements. During the subtracting process, if block *B* is matched with BM, it would be labelled as background block, otherwise it is selected as the foreground block. Fig. 2 depicts the subtraction processing phase for creating the foreground and background model.

2.3. Video Encoder

The video compression technique has been used for many years. Due to the H.264 codec, an evolution was made in video compression techniques. Recently, with the standardization of the H.265 codec, many devices will support this codec, which can significantly decrease the bandwidth being used. Due to the popularity and advantages of this codec, this research has also used this codec.

2.4. Frame Multiplier

By separating the background image from the original video frames, on the decoder side, the image must be multiplied by the foreground frames that have been compressed by the video encoder to restore the original frames. In fact, at this point, it is enough to multiply the moving objects on the pre-cached background image to restore the same original frames. Since the detection of moving objects by the MOG2 method is faced with a small error, in some parts of the frame noise is visible and whereby causes this method not to be an ideal filter to separate the background image from moving objects. In this regard, the proposed methodology used MT to remove noises and to create an ideal mask.

3. Experimental Results

In applying the proposed methodology, a highway video streaming dataset [30] that completely covers the mentioned requirement is used. The camera position is fixed such that the background can be extracted from the main frames. By applying this method, the experimental results of this article can be examined in four areas.

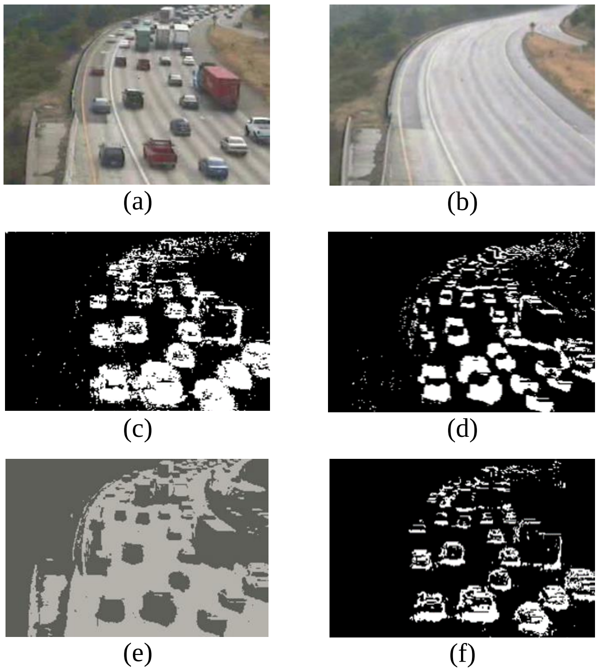

Fig. 3: Procedure of creating an ideal mask from the original frame, (a) Original frame, (b) Subtracted background, (c) MOG2 mask, (d) MOG2 mask after MT, (e) CC frame, (f) Ideal mask (CC+MOG2).

3.1. Numerical Calculations

Due to the removal of the background image from the main video frames and the compression of only foreground frames, pixels per frame will be significantly reduced. In fact, by reducing the number of pixels per frame, the number of numerical calculations for the compression operation at the encoder side will be reduced by the same amount. Fig. 3 illustrates the number of pixels reduced by comparing the 100 original frames for both MT+MOG2 and CC+MOG2 methods. Analyzing this figure shows that in the initial 100 frames, the total number of pixels per frame will be reduced by an average of 35000 pixels for MT+MOG2 method while the CC+MOG2 gives a better reduction, 38000 pixels reduction per frame. In other words, 45.57% of the processing load is reduced by MT+MOG2 method and 49.47% by CC+MOG2 method. Henceforth, the CC method will result in better reductions as its associated mask is cleaner as compared to the MT method.

3.2. Bandwidth Transmission

The occupied network bandwidth will decrease dramatically for two reasons. The first reason is that due to the subtraction process, the background image is occupied the bandwidth only once. The second reason is that the bandwidth occupation rate is directly dependent on the content within the frames and since the content of each frame is reduced, the transmitted bandwidth will be saved.

Fig. 6 illustrate the I/P/B frames bandwidth occupation rate for three methods. Since the I-frame occupies a very large amount of bandwidth, it will be very important to check this frame rate. This research illustrates that due to the removal of large quantities of pixels, which is referred to the separation of foreground and background frames, it has been able to

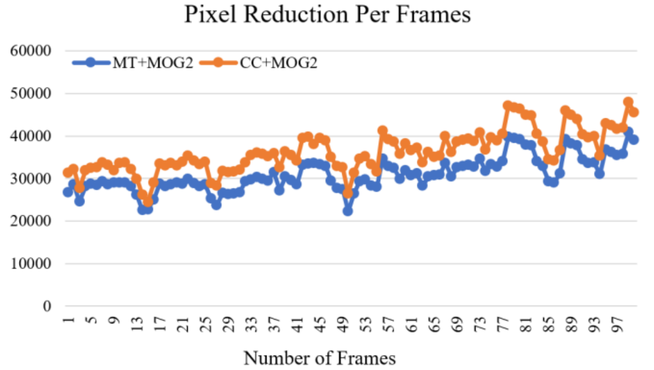

Fig. 4: Pixel reduction per frames to reduce the total calculation.

reduce the bandwidth occupation by 48% in the I-frame using MT+MOG2 and 67.4% through CC+MOG2. In return, for this significant reduction in the reference frames, P/B-frames have been increased to a very small extent, which will take up very little bandwidth compared to reference frames.

3.3. Visual Quality

Due to the rapid changes of moving objects in the real-time video streaming, noises will be present in the Gaussian filter. To optimize this noisy mask to an ideal mask, our proposed last methodology used MT to reduce noises. Due to applying several filters during the AI process, there will have a slight reduction in visual quality compared to the original video. However, in our newer method, instead of using MT, we have used CC to achieve better results.

Structural Similarity Index Measurement (SSIM) and Peak Signal to Noise Ratio (PSNR) algorithms are useful visual quality assessment algorithms in which the aim is to examine the qualitative changes compared to the original CCTV video.

Fig. 8 and Fig. 9 compare the visual quality of three compressed videos using the CC+MOG2, MT+MOG2 and, traditional solutions. The investigation of the above compressed videos for PSNR and SSIM values illustrates an approximate 1 dB and 1 unit improvement by using CC+MOG2 in comparison to only using MT+MOG2 solution.

3.4. Latency

As the proposed coding system uses a background subtraction process, it is expected that latency is slightly increased. The main part of the process which pushes delays in an end-to-end cycle refers to MT process, which needs to scan whole the frame then creates a noise free mask. In our newer strategy, as we have used CC instead of MT, it is expected to perform better in reducing the total latency. The FFMPEG encoding process depicts that approximately twice the time is required for MT+MOG2 encoding, and the use of the CC+MOG2 method has resulted in very little reduction in latency. This result shows that the majority of latency is due to using MOG2 and not MT or CC processes. Fig. 10 illustrates the encoding latency for the three above methodologies.

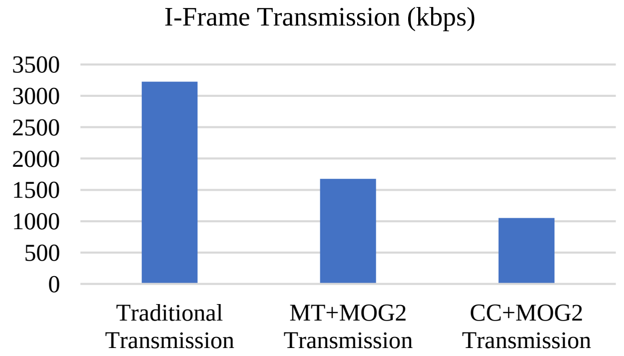

Fig. 5: Comparison of I-Frame transmission reductions.

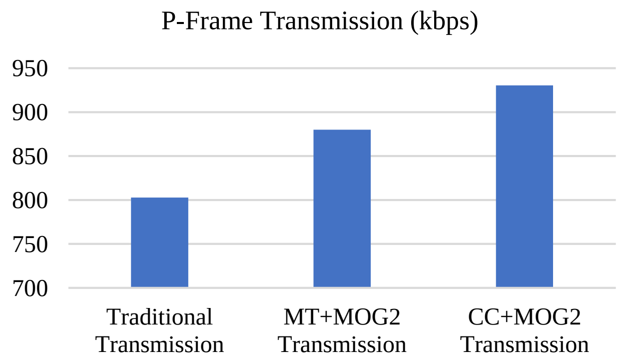

Fig. 6: Comparison of P-Frame transmission reductions.

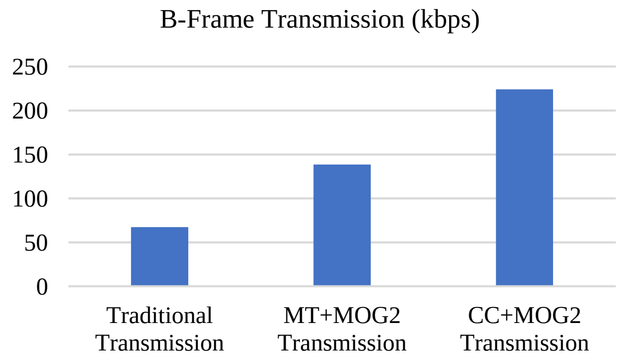

Fig. 7: Comparison of B-Frame transmission reductions.

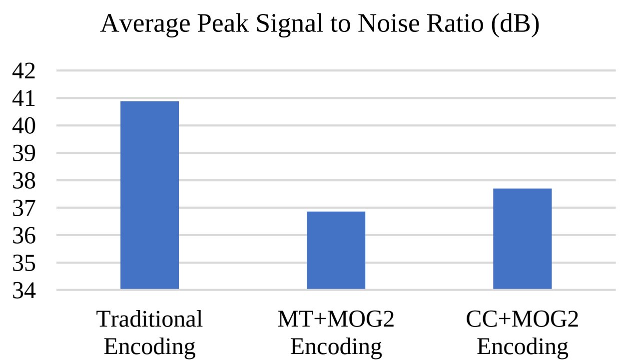

Fig. 8: The PSNR comparison of traditional, MT+MOG2 and CC+MOG2 encoding.

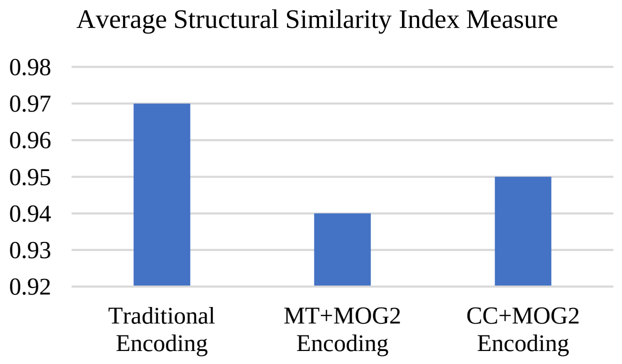

Fig. 9: The SSIM comparison of traditional, MT+MOG2 and CC+MOG2 encoding.

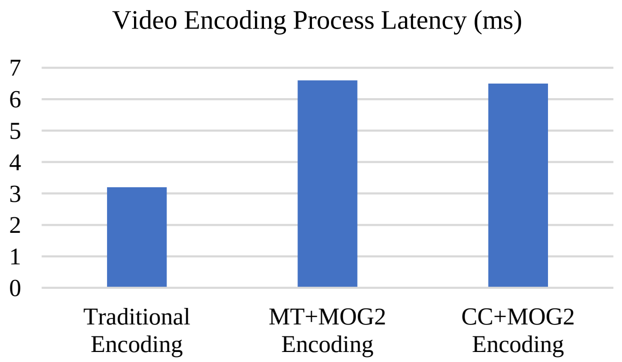

Fig. 10: Comparison of video encoding processes.

4. Conclusion

In this paper, an end-to-end AI-based video coding for separating the foreground and background frames with the help of using CC method has been improved. Through this procedure the total number of encoder computations was reduced while the visual quality reduction was improved by the CC process. Also, as the background is sent just one at a time, the network bandwidth occupation is reduced by 48% by MT+MOG2 and 67.4% by CC+MOG2. In visual quality assessment, 1dB and 1 unit improvement in PSNR and SSIM were achieved through the proposed method. For future works, the CC process can be changed by investigation on adaptive thresholding method to check if the optimization and improvement would be increased.

CRediT Authorship Contribution Statement

Mohammadreza Ghafari: Conceptualization, Data curation, Formal analysis, Methodology, Project administration. **Abdollah Amirkhani:** Project administration, Supervision, Validation. **Elyas Rashno:** Investigation, Methodology. **Shirin Ghanbari:** Validation, Writing - original draft, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. The ethical issues; including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, redundancy has been completely observed by the authors.

References

[1] S. Arora, K. Bhatia, and V. Amit, "Storage optimization of video surveillance from CCTV camera," in *2016 2nd International Conference on Next Generation Computing Technologies (NGCT)*, 2016, pp. 710-713.

[2] S. Vítek, M. Klíma, and L. Krasula, "Video compression technique impact on efficiency of person identification in CCTV systems," in *2014 International Carnahan Conference on Security Technology (ICCST)*, 2014, pp. 1-5.

[3] G. Harikrishnan, "Role of CCTV in business organization: A case study," *International Journal of Commerce, Business and Management*, vol. 3, no. 3, pp. 466-470, 2014.

[4] V.Carli, "Assessing CCTV as an effective safety and management tool for crime-solving, prevention and reduction," Montreal, 2008. [Online]. Available: https://policycommons.net/artifacts/1217708/assessing-cctv-as-an-effective-safety-and-mangement-tool-for-crime-solving-prevention-and-reduction/1770796

[5] B. Goold, "Public area surveillance and police work: The impact of CCTV on police behaviour and autonomy," *Journal of Surveillance and Society*, vol. 1, no. 2, pp. 191-203, 2003.

[6] B. C. Welsh, and D.P. Farrington, "Effects of closed circuit television surveillance in reducing crime," *Campbell Systematic Reviews*, vol. 4, no. 1, pp. 1-73, 2008.

[7] B. Srilaya, LV. Kumar, and LNP. Boggavarapu, "Surveillance using video analytics," in *Proc. of Int. Conf. on Computational Intelligence and Information Technology*, 2013.

[8] Kumar, M. B. Punith, and P. S. Puttaswamy, "Video to frame conversion of TV news video by using MATLAB," *International Journal of Advance Research in Science and Engineering*, vol. 3, no.3, pp. 95-101, 2014.

[9] D. R. Zaghar, and TE. Abdulabbas, "A Compression Algorithm for Video Surveillance System," *International Journal of Computer Applications*, vol. 118, no.24, pp. 19-22, 2015.

[10] CCTV Services (Jan. 6, 2023). Unlimited Video Storage. [Online]. Available: http://www.cctvservices.net/unlimited-video-storage/

[11] S. Arora, K. Bhatia, and V. Amit, "Storage optimization of video surveillance from CCTV camera," in *2016 2nd International Conference on Next Generation Computing Technologies (NGCT)*, 2016, pp. 710-713.

[12] C. Held, J. Krumm, P. Markel, and R. P. Schenke, "Intelligent Video Surveillance," *Computer*, vol. 45, no. 3, pp. 83-84, 2012.

[13] F. Lv, J. Kang, R. Nevatia, I. Cohen, and G. Medioni, "Automatic tracking and labeling of human activities in a video sequence," in *Proc. the 6th IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS04)*, 2004.

[14] P. C. Ribeiro, J. Santos-Victor, and P. Lisboa, "Human activity recognition from video: Modeling feature selection and classification architecture," in *International Workshop on Human Activity Recognition and Modeling*, pp. 61-70, 2005.

[15] L. Ma, H. Qi, S. Zhu, and S. Ma, "A fast background model based surveillance video coding in HEVC," in *IEEE Visual Communications and Image Processing Conference*, Valletta, pp. 237-240, 2014.

[16] X. Zhang, T. Huang, Y. Tian, and W. Gao, "Background-modeling-based adaptive prediction for surveillance video coding," *IEEE Transactions on Image Processing*, vol. 23, no. 2, pp. 769-784, 2014.

[17] S. Guo, Y. Wang, Y. Tian, P. Xing, and W. Gao, "Quality-progressive coding for high bit-rate background frames on surveillance videos," in *2015 IEEE International Symposium on Circuits and Systems*, 2015, pp. 2764-2767.

[18] Y. Tian, J. Yan, S Dong, and T. Huang, "PA-Search: predicting units adaptive motion search for surveillance video coding," *Computer Vision and Image Understanding*, vol. 170, pp. 14-27, 2018.

[19] H. Kim, and H. Lee, "A low-power surveillance video coding system with early background subtraction and adaptive frame memory compression," *IEEE Transactions on Consumer Electronics*, vol. 63, no. 4, pp. 359-367, 2017.

[20] H. Li, W. Ding, Y. Shi, and W. Yin, "A double background based coding scheme for surveillance videos," in *Data Compression Conference*, Snowbird, UT, 2018, pp. 420-420.

[21] J. Celine and S. A. A, "Face Recognition in CCTV Systems," in *2019 International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 2019, pp. 111-116.

[22] M. T. Bhatti, M. G. Khan, M. Aslam, and M. J. Fiaz, "Weapon Detection in Real-Time CCTV Videos Using Deep Learning," *IEEE Access*, vol. 9, pp. 34366-34382, 2021.

[23] S. Powale, A. Dhanawade, S. Bagwe, S. Kawale, N. L. Chutke, and S. Chavan, "Person identification in low resolution CCTV footage using deep learning," in *2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, 2020, pp. 236-240.

[24] Computer Vision Group, School of Electronic Engineering and Computer Science, Queen Mary University of London. (Jan. 6, 2023). TinyFace: Face Recognition in Native Low-resolution Imagery. [Online]. Available: https://qmul-tinyface.github.io/

[25] M. S. Pillai, G. Chaudhary, M. Khari, and R. G. Crespo, "Real-time image enhancement for an automatic automobile accident detection through CCTV using deep learning", *Soft Computing*, vol. 25, no. 18, pp. 11929-11940, 2021.

[26] R. Ullah et al., "A real-time framework for human face detection and recognition in CCTV images," *Mathematical Problems in Engineering*, article 3276704, 2022.

[27] Pan, S.-H., S.-C.J.S. Wang, and Materials, "Identifying vehicles dynamically on freeway CCTV images through the Yolo deep learning model," *Sensors and Materials*, vol. 33, no. 5, pp. 1517-1530, 2021.

[28] M. Ghafari, A. Amirkhani, E. Rashno, S. Ghanbariet, "Novel gaussian mixture-based video coding for fixed background video streaming," in *2022 12th Iranian/Second International Conference on Machine Vision and Image Processing (MVIP)*, Ahvaz, Iran, 2022.

[29] G. Matczak, and P. Mazurek, "Comparative Monte Carlo Analysis of Background Estimation Algorithms for Unmanned Aerial Vehicle Detection," *Remote Sensing*, vol. 13, no. 5, 2021.

[30] Kaggle. (Jan. 6, 2023). *Highway Traffic Videos Dataset*. [Online]. Available: https://www.kaggle.com/aryashah2k/highway-traffic-videos-dataset

BIOGRAPHY

Mohammadreza Ghafari received his Telecommunication master's degree from Amirkabir University of Technology. He has graduated with his B.Sc. degree, Electrical Engineering and was honoured as the best student of the year at IRIB university. His background is mainly covering computer networks and video processing which he also participated in many national and academic projects in these fields. He has been working on Software Defined Network, Cloud Computing and Video Compression methods for many years, while his main interest is Cloud Gaming.

Abdollah Amirkhani received the M.Sc. and Ph.D. degrees (with honors) in electrical engineering from Iran University of Science and Technology (IUST), Tehran, in 2012 and 2017, respectively. He earned the Outstanding Student Award (2015) form the First Vice President of Iran. In 2016, he was awarded by the Ministry of Science, Research and Technology. He is an Assistant Professor in the school of automotive engineering at IUST. He is the Associate Editor of the "Engineering Science and Technology, an International Journal". He has been actively involved in several National R&D projects, related to the development of new methodologies and learning algorithms based on AI techniques. His research interests are in machine vision, fuzzy cognitive maps, data mining and machine learning.

Elyas Rashno received the B.Sc. degree in computer engineering from the Isfahan University of Technology, Isfahan, Iran, in 2015, the M.Sc. degree in artificial intelligence from the Iran University of Science and Technology, Tehran, Iran, in 2018. Since 2018, he has been an R&D member of a company that works on deep learning models. His research interests include machine learning, deep learning, image processing, biometrics, content-based image retrieval, and evolutionary computing.

Shirin Ghanbari received her M.Sc. in E-Commerce and Ph.D. in object segmentation and video tracking from the Computer Science and Electrical Engineering department from the University of Essex, United Kingdom, in 2005 and 2010. Recently, she leads a team for data analysis using the latest AI technologies for both text and video files.