The CSI Journal on
Computer Science and Engineering
Vol. 19, No. 2, 2025
Pages 27-32
doi.org/10.22034/jcse.2025.181902
Regular Paper

Enhance the efficiency of rate-control in VVC standard using the characteristics of human visual system

Maryam Ranjbar[1] Hoda Roodaki[1]

[1] Faculty of Computer Engineering, K. N. Toosi University of Technology, Tehran, Iran

Abstract

Rate control (RC) plays an essential role in video coding systems and makes the bitrate of an encoded video equal to the target bitrate while minimizing the distortion of the compressed video. Incorporating the characteristics of the human visual system (HVS) into RC makes it more efficient since the human eyes are the ultimate recipient of a video. In this paper, we rectified the rate-distortion algorithm of versatile video coding (VVC) standard to be more correlated with the perceptual video quality. For this purpose, the structural similarity index (SSIM) and the PSNR-HVS-M quality metrics are used as criteria that more closely resemble the way humans perceive structural distortions. Experimental results show that the proposed rate-distortion model can attain at most 1.77% and 1.8% bitrate saving in low delay (LD) and random access (RA) configurations, respectively, compared to the VVC standard as the baseline.
Keywords: Rate control, versatile video coding standard, human visual system.

1. Introduction

Numerous compression standards have been created to meet the diverse requirements of video applications. VVC, the latest video coding standard, was finalized in July 2020 [1]. It succeeds High-Efficiency Video Coding (HEVC) [2], as it was also developed by the Joint Video Experts Team (JVET) of ITU-T VCEG and ISO/IEC MPEG. When compressing video, various application constraints need to be taken into account, including latency and bandwidth. To achieve optimal visual quality for video applications under real-world constraints, Rate Control (RC) is used in video compression [3]. RC aims to allocate bits wisely and maintain constant video quality [4].

VVC uses the R-λ rate control algorithm [5], which assigns the bit rate based on the resolution of encoded video, frame rate, and current channel bandwidth. However, the R-λ model's target bit allocation is based on the Mean Absolute Difference (MAD), which does not always reflect all the features of the image content and ignores subjective visual experience [1]. Therefore, incorporating the characteristics of the Human Visual System (HVS) into RC is desirable to improve the visual quality.

The Structural Similarity Index (SSIM), PSNR- HVS, PSNR_HVS_M, and VIFP are four famause quality metrics based on the human visual system.

The Structural Similarity Index (SSIM) is a perceptual metric that quantifies image quality degradation caused by processing such as data compression. It is a full reference metric that requires two images – a reference image and a processed image which is typically compressed. The measure between two images x and y is [6]:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y+C_1)(2\sigma_{xy}+C_2)}{(\mu_x^2+\mu_y^2+C_1)(\sigma_x^2+\sigma_y^2+C_2)} \quad (1)$$

The μx, μy are the pixel sample mean of x and y and σ^2, σ^2 are the variance of x and y, respectively. C_1 and C_2 are two variables to stabilize the division with weak denominator [6]. Since, the evaluation of image quality using MAD and PSNR does not take into account the characteristics of the HVS and image content information, their outcomes are not always indicative of subjective quality scores. PSNR only considers the absolute difference between the distorted and original images, without accounting for factors that influence human visual perception [7]. To enhance the effectiveness of these metrics, the PSN-HVS and PSNR-HVS-M models have been

developed. PSNR-HVS is an extension of PSNR that incorporates properties of the human visual system such as contrast perception and PSNR-HVS-M improves on PSNR-HVS by additionally taking into account visual masking [8]. The visual information fidelity in pixel domain (VIFP) is used as a measure. VIFP is obtained by quantifying two mutual information quantities: the mutual information between the input and output of the HVS channel in the absence of distortion, and the mutual information between the input of the distortion channel and the output of the HVS channel for the test image. The assumption is made that in the absence of distortion, the signal passes through the HVS channel of a human observer before entering the brain, which extracts cognitive information. For distorted images, it is assumed that the reference signal has passed through another distortion channel before entering the HVS. By combining these two quantities, a visual information fidelity measure is derived [9]. In this paper, we rectify the rate-distortion model in VVC standard using SSIM and PSNR-HVS-M metrics to measure distortion most correlated to perceptual video quality. The rectified scheme can better maintain structural information and improve the visual quality of encoded videos.
The remainder of this paper is arranged as follows. Section 2 reviews the Rate control in video coding standards. Section 3 reviews the related work. Section 4 and 5 describe the proposed method and the experimental results, respectively. Finally, section 6 concludes the paper.

2. Rate control in video coding standards

In video coding standards, rate control comprises buffer control, rate-distortion modeling, and target bit allocation as its core elements [2]. The bit allocation is divided into three layers, namely the basic unit (BU), frame, and Group of Picture (GOP) layers [3].
In GOP layer, the following equation is used to assign propoer bitrate to each GOP, where, NGOP, N_{coded}, SW, and R_{coded} represent the number of frames in the current GOP, the current video sequence's consumed bits, the size of the smooth window, and the number of frames encoded, respectively. SW is set to 40 in the actual encoding [2].

$$T_{GOP} = \frac{R_{PicAvg} \times (N_{coded} + SW) - R_{coded}}{SW} \times N_{GOP} \quad (2)$$

In frame layer, strategy for bit allocation is as follows [2]:

$$T_{pic} = \frac{\omega_{picCur} \times (T_{GOP} - R_{coded})}{\sum_{picNotCode} \omega_{pic}} \quad (3)$$

Where ω_{PicCur} and $\sum_{PicNotCode} \omega_{Pic}$ represent the weights of the current frame and the sum of the whight of all the uncoded frames in the current GOP, respectively.
When it comes to the first sequence, the allocation strategy for the first frame is determined based on the lack of prior knowledge about its content [2].

$$R_{PicAvg} = \frac{R_{target}}{F} \quad (4)$$

$$T_{Intra} = \omega_I \times R_{PicAvg} = \omega_I \times \frac{R_{target}}{F} \quad (5)$$

$$bpp = \frac{R_{PicAvg}}{W \cdot H} \quad (6)$$

Where, R_{target}, R_{PicAvg}, F, T_{Intra}, bpp, W, and H represent the current channel bandwidth, the average code rate, the frame rate of the current frame, the number of bits allocated for the first frame, the number of allocated bits per pixel, the width and height of the current frame, respectively. The empirical weight ω_I is calculated according to the range of bpp.
Largest coding-unit (LCU) is the basic unit in VVC, so similar LCUs have the same QPs. The LCU bit allocation strategy is the same as the frame-level strategy.

$$T_{BU} = \frac{T_{PIC} - R_{CodPic} - R_{Head}}{\sum_{BUNotCoded} \omega_{BU}} \times \omega_{BUcur} \quad (7)$$

Where, RHead, $R_{CodedPic}$, ω_{BUCur}, and $\sum_{BUNotCoded} \omega_{BU}$ represent the number of bits of information in the current frame header, the number of bits consumed by the coded BU, the weight of the current BU, and the sum of the not coded BU weights, respectively. The bit allocation weight is based on the MAD value. The MAD value describes the previously encoded prediction block and the error of the current block.
In the R-λ model, QP has less influence on the output bit stream than λ. First, conforming to the number of allocated bits, the Lagrange coefficient λ is calculated, afterward based on the relationship between QP and k, QP is calculated as follows:

$$\lambda = \alpha \times bpp^{\beta} \quad (8)$$

$$QP = 4.2005 \ln \lambda + 13.7122 \quad (9)$$

In these equations, bpp, β, and α represent the mean number of bits assigned to the pixel point and coding parameters with primary values of -1.367 and 3.2003. It should be noted that bpp can be obtained from the LCU level or the frame level.

3. Related work

In this section, various rate control algorithm is reviewed. The method in [10] propose a novel λ-domain rate control algorithm based on the R - λ model, and implement it in high efficiency video coding (HEVC). The method proposed in [11] uses an optimal bit allocation scheme for coding tree unit level rate control in HEVC. For this purpose, first a novel R-D estimation instead of the existing R– λ estimation is proposed. Obtaining a closed-form solution to the optimization formulation is not feasible, which is why a Recursive Taylor Expansion (RTE) method is suggested to solve it iteratively. This approach allows for an approximate closed-form solution, resulting in successful optimal bit allocation and bit reallocation. In [12], a new rate control framework is introduced for High Efficiency Video Coding (HEVC) that utilizes the Lagrange multiplier. The proposed approach assumes constant quality control and establishes a novel relationship between the Lagrange multiplier and distortion. By incorporating the proposed distortion model and buffer status, It is possible to obtain a practical solution to minimize distortion variation at the coding tree unit level across video frames. A novel approach to optimize Rate-Distortion (R-D) performance for High Efficiency Video Coding (HEVC) is presented in [13]. This method focuses on frame-level bit allocation optimization. To overcome the

drawbacks of the Mixture Laplacian Distribution (MLD) model, which is complex, a new Synthesized Laplacian Distribution (SynLD) model is introduced. This model describes the DCT transformed coefficients using Kullback-Leibler (KL) divergence analysis. Additionally, the study investigates quality dependencies among frames and proposes a linear relationship between the Quality Dependency Factor (QDF) and skip mode percentage for QDF prediction. Based on the proposed SynLD model and QDF prediction method, a ρ-domain based frame-level bit allocation method is proposed.
The propose approach in [14] involves incorporating a perceptual video quality metric into the rate distortion optimization process for 3D-HEVC. This is achieved by utilizing PSNR-HVS as a distortion measure in the coding unit (CU) mode selection process. PSNR-HVS takes into account the unique characteristics of the human visual system.
Previous research has shown that the subjective quality of a video is not always consistent with its image fidelity. To address this issue, [15] have developed a video coding method that improves subjective quality while minimizing fidelity degradation. The approach proposes in [4] utilizes a just noticeable distortion (JND)-based perceptual rate control method for high efficiency video coding. The JND factor of a coding unit is used as a weight for bitrate allocation and R-D modelling is conducted based on this factor. The resulting R-D model is then integrated into an existing rate control framework to improve coding efficiency. The algorithm has been implemented in the newest video coding standard.
The aim of the study in [16] is to improve the visual quality of 3D video by proposing a joint bit allocation scheme based on structural similarity (SSIM). The perceptual quality of a synthesized view is influenced by both texture and depth map quality, therefore SSIM-based rate-distortion optimization is applied to both texture and depth map. SSIM is utilized as a distortion metric in mode decision and motion estimation. An experimental approach is used to establish an SSIM-based distortion model for the synthesized view. As SSIM cannot be directly related to quantization step, the distortion model is converted into mean squared error (MSE) using an SSIM-MSE relation. The bit allocation problem is solved using the Lagrange multiplier method. The objective of the study in [4] is to enhance the visual quality of HEVC by introducing a rate control algorithm that is based on visual perception. The proposed algorithm optimizes the bit allocation weight of the LCU level by taking into account the visual perception of both luminance and motion. This helps to improve the subjective quality of the video. Additionally, λ and QP are adjusted in conjunction with the bit allocation weight to further enhance the rate distortion performance. In [5] a video bit-rate controller that is fully compatible with the requirements of real-time applications is presented. The controller utilizes a multi-layer perceptron (MLP) neural network to determine the appropriate quantization parameter (QP) modification at the frame level. The QP derivation process takes into account the buffer occupancy to ensure reliable buffer control. An enhanced R-λ rate control model, referred to as IRLRC, is proposed in [17] which leverages joint spatial-temporal domain information and human visual system characteristics. The model utilizes gradient information in the joint spatial-temporal domain to guide bit allocation at both the frame and coding tree unit (CTU) levels. Additionally, the temporal coefficient is adaptively corrected to improve the accuracy of the model.
Both proposed solutions in [18] aim to improve the rate control performance of VVC low-delay coding by taking into account the dependencies between frames. The first solution utilizes a distortion model to allocate bits based on the correlation between frames, while the second solution uses the differences between frames to adaptively allocate bits with lower complexity. Overally, these solutions can lead to better video quality and more efficient use of bitrate in VVC low-delay coding scenarios.

4. Proposed method

This section details the proposed approach to rate control algorithm for VVC standard, which takes into account the visual characteristics of the human visual system. The method involves selecting a quality evaluation criterion that is appropriate for human eye perception and modifying the bit rate control algorithm in VVC standard accordingly.

4.1. Selecting the appropriate quality assessment metric

In odre to find the proper quality assessment metric for rate control algorithm, four various metrics that described in section 1, SSIM, PSNR-HVS, PSNR-HVS-M, and VIFP are considered. Then, eight video sequences encoded by four fixed QPs (22, 28, 32, and 38) in random access configuration and these four-quality metrics are used to evaluate the quality of decoded sequences. TABLE I shows a summary of the information on these sequences.
Then we have used subjective tests to compare the correlation of these objective metrics with visual perception. The correlation coefficients between the overall perception quality and these objective quality metrics are used for this issue. The DSCQS method, as described in ITU-R Recommendation 500 [19] was utilized to assess the subjective quality of the decoded sequences. The experiment involved forty-five non-expert viewers with no background in video processing or quality assessment. A 19-inch monitor was used to display the material at its main resolution on a 15.6-inch monitor, with a viewing distance set to four times the screen height in accordance with Rec. ITU-R 812.

TABLE I. Test sequences information

Sequence	Frame size	Frame rate (fps)
BasketballDrive	1920 × 1080	50
Cactus	1920 × 1080	50
BQTerrace	1920 × 1080	60
KristenAndSara	1280 × 720	60
FourPeople	1280 × 720	60
BasketballDrill	832 × 480	50
PartyScene	832 × 480	50
BlowingBubbles	416 × 240	50

The viewers were presented with the original and decoded sequences randomly, with a 3-second gray display between them. They then rated the subjective quality of both sequences on a scale from 1 to 5, corresponding to "Bad", "Poor", "Fair", "Good", and "Excellent". The subjective quality was expressed as the difference between the ratings for the source

and decoded sequence. The total MOS of the sequences was calculated by averaging the numerical values obtained from different viewers as shown in TABLE II.

TABLE II. The total MOS of the sequences that calculated by averaging the numerical values obtained from different viewers

Sequence	QP			
	22	28	32	38
BasketballDrive	4.88	4.68	4.51	4.44
Cactus	4.93	4.82	4.68	4.55
BQTerrace	4.95	4.82	4.71	4.67
KristenAndSara	4.82	4.77	4.22	3.95
FourPeople	4.75	4.64	4.4	4.04
BasketballDrill	4.75	4.68	4.24	4.15
PartyScene	4.91	4.86	4.53	4.37
BlowingBubbles	4.88	4.68	4.44	4.06

The correlation coefficients between the objective quality metrics and MOS are shown in Figure 1. The results indicated that SSIM and PSNR-HVS-M criteria have the highest correlation with the visual system of the human eye.

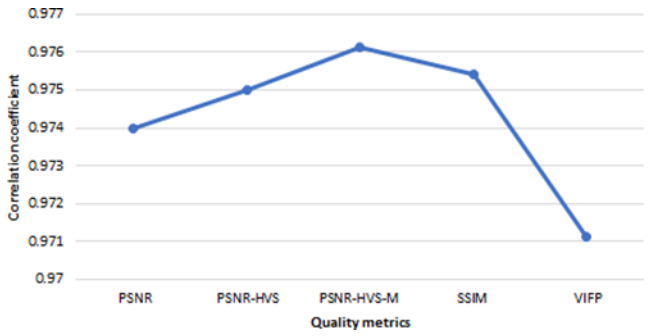

Figure 1. The correlation coefficient between MOS and various objective quality metrics

4.2 Proposed method rate control algorithm in VVC standard

To incorporate human eye characteristics into the rate control algorithm, adjustments were made to the weight value in equation (3). The VVC standard calculates the weight value based on the MAD value of the previously encoded prediction block and the error of the current block. In order to apply the characteristics of the human eye in the bit rate control algorithm, since SSIM measures the similarity between two images and ranges from 0 to 1, (1-SSIM) was utilized to modify the proposed weight in equation (3). The reson is that, as SSIM increases, the similarity between two images increases, so (1-SSIM) can be used to indicate the distortion or difference between them. Similarly, the PSNR-HVS-M criterion can be used to rectify the weight value used in equation (3).

In a similar way, we have used SSIM and PSNR-HVS-M in order to find the proper weight values at LCU level as shown in equation (7). Given the bits allocated to the current frame, bpp is updated using equation (6) for the next frame. Finally, λ and QP are calculated using equation (8) and (9).

5. Experimental results

In this part, we will evaluate the performance of our rectified rate control algorithm based on human visual system.

We compare the results of rectified rate control with the results of VTM reference software [20] which implemented the original rate control algorithm of VVC standard. Random access (RA) and low delay (LD) encoding configurations of VTM are used to extract our results. 33 and 10 frames are encoded for each sequence in RA and LD encoding configurations, respectively.

Table III compare the performance of the rectified and original rate control algorithms using RA and LD encoding configurations. The negative numbers indicate the improvement of the compression efficiency of the proposed method compared to the original method. The results of the decoded video sequences usig original and rectified rate control algorithm are compared subjectively in Figure 2 and Figure 3 for more clarification.

TABLE III. Performance comparison of rectified rate control based on SSIM and PSNR-HVS-M over original rate control in VVC for RA and LD encoding configurations

Sequence	SSIM		PSNR-HVS-M	
	RA	LD	RA	LD
	BD-rate (%)	BD-rate (%)	BD-rate (%)	BD-rate (%)
BasketballDrive	-0.42	-2.91	-0.31	-2.14
Cactus	-0.96	-0.66	-0.77	-0.54
BQTerrace	-3.86	-0.31	-3.86	-2.06
KristenAndSara	-0.20	-0.23	-0.12	-0.80
FourPeople	-2.91	-2.31	-4.27	-2.41
BasketballDrill	-4.59	-1.91	-4.59	-0.34
PartyScene	-0.04	-3.09	-0.29	-0.80
BlowingBubbles	-0.51	-2.76	-0.16	-0.10
Average	-1.69	-1.77	-1.8	-1.15

Figure 2. Comparison of the visual quality of BasketballDrive video using using original and rectified rate control algorithm, (a) video compressed using rectified rate control, (b) video compressed using VVC rate control (c) original video.

Figure 3. Comparison of the visual quality of Cactus video using using original and rectified rate control algorithm

(a) video compressed using rectified rate control, (b) video compressed using VVC rate control (c) original video.

Finally, we compared the results of our rectified rate control algorithm with MORC method [21]. This approach aims to achieve optimal solutions by simultaneously minimizing average distortion and quality fluctuation. The multi-objective problem is converted into a single-objective problem using a weighting method. A convex optimization problem is formulated to allocate rates while achieving better trade-offs among three objectives: minimizing average video distortion, meeting rate constraints, and minimizing video quality fluctuation. The D-λ model is used to propose a two-stage method to solve the optimization problem.
The results of this comparison are shown in TABLE IV that shows the better performance of our proposed approach against VVC rate control algorithm.

TABLE IV. Comparing the performance of our proposed methods and the MORC method [21] against the VVC rate control method

Sequence	Rectified rate control algorithm using SSIM against VVC	Rectified rate control algorithm using PSNR-HVS-M against VVC	MORC method against VVC
	BD-rate (%)	BD-rate (%)	BD-rate (%)
BasketballDrive	-0.42	-0.31	-2.90
Cactus	-0.96	-0.77	0.40
BQTerrace	-3.86	-3.86	2.90
PartyScene	-0.04	-0.29	-0.3
BasketballDrill	-4.59	-4.59	-2.3
BlowingBubbles	-0.51	-0.16	-3.80
FourPeople	-2.91	-4.27	-2.40
KristenAndSara	-0.20	-0.12	-4.90
Average	-1.68	-1.79	-1.66

6. Conclusion

In this paper, a rectified rate control algorithm for VVC standard is proposed that uses the SSIM and PSNR-HVS-M metrics to assign appropriate rate to frames and LCUs. The VVC standard uses the MAD metric in its rate control and bit allocation algorithms that is not properly match with human visual system. Experimental results show that the proposed rate control algorithm can attain better performance over VVC standard using low delay (LD) and random access (RA) encoding configurations.

References

[1] B. Bross, J. Chen, J.-R. Ohm, G. J. Sullivan, and Y.-K. Wang, “Developments in international video coding standardization after AVC, with an overview of versatile video coding (VVC),” *Proc. IEEE*, vol. 109, no. 9, pp. 1463–1493, 2021.
[2] G. Marquant, C. Salmon-Legagneur, F. Urban, and P. d. Lagrange, “Spatial scalability with VVC: Coding performance and complexity,” in *Proc. SPIE Opt. Eng. + Appl.*, 2022.
[3] A. A. Ramanand, I. Ahmad, and V. Swaminathan, “A study of rate control for H.265/HEVC video compression,” Univ. Central Arkansas, 2020.
[4] Z. Feng, P. Liu, and K. Jia, “Visual perception based rate control algorithm for HEVC,” *J. Phys.: Conf. Ser.*, vol. 960, no. 1, p. 012041, 2018.
[5] S. Kim, D. Pak, and S. Lee, “SSIM-based distortion metric for film grain noise in HEVC,” *Signal, Image Video Process.*, vol. 12, pp. 489–496, 2018.
[6] M. Zhou, X. Wei, S. Kwong, W. Jia, and B. Fang, “Just noticeable distortion-based perceptual rate control in HEVC,” *IEEE Trans. Image Process.*, vol. 29, pp. 7603–7614, 2020.
[7] F. Raufmehr, M. R. Salehi, and E. Abiri, “A frame-level MLP-based bit-rate controller for real-time video transmission using VVC standard,” *J. Real-Time Image Process.*, vol. 18, pp. 751–763, 2021.
[8] F. Li, S. Krivenko, and V. Lukin, “A two-step approach to providing a desired visual quality in image lossy compression,” in *Proc. 2020 IEEE 15th Int. Conf. Adv. Trends Radioelectron., Telecommun. Comput. Eng. (TCSET)*, Lviv-Slavske, Ukraine, 2020.
[9] A. Banitalebi-Dehkordi, M. T. Pourazad, and P. Nasiopoulos, “3D video quality metric for 3D video compression,” in *Proc. IVMSP 2013 IEEE*, 2013, pp. 1–4.
[10] B. Li, H. Li, L. Li, and J. Zhang, “λ domain rate control algorithm for high efficiency video coding,” *IEEE Trans. Image Process.*, vol. 23, pp. 3841–3854, 2014.
[11] S. Li, M. Xu, Z. Wang, and X. Sun, “Optimal bit allocation for CTU level rate control in HEVC,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 11, pp. 2409–2424.
[12] M. Wang, K. N. Ngan, and H. Li, “Low-delay rate control for consistent quality using distortion-based Lagrange multiplier,” *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 2943–2955.
[13] W. Gao, S. Kwong, H. Yuan, and X. Wang, “DCT coefficient distribution modeling and quality dependency analysis based frame-level bit allocation for HEVC,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 1, pp. 139–153, 2015.
[14] S. Valizadeh, P. Nasiopoulos, and R. Ward, “Perceptual rate distortion optimization of 3D–HEVC using PSNR-HVS,” *Multimed. Tools Appl.*, vol. 77, pp. 22985–23008, 2018.
[15] Q. Wang, H. Yuan, J. Huo, and P. Li, “A fidelity-assured rate distortion optimization method for perceptual-based video coding,” in *Proc. 2019 IEEE Int. Conf. Image Process. (ICIP)*, 2019, pp. 4135–4139.
[16] P. K. Biswas, “SSIM-based joint-bit allocation for 3D video coding,” *Multimed. Tools Appl.*, vol. 77, pp. 19051–19069, 2018.
[17] Z. Zhao, S. Xiong, W. Sun, X. He, and F. Zhang, “An improved R-λ rate control model based on joint spatial-temporal domain information and HVS characteristics,”

Multimed. Tools Appl., vol. 80, pp. 345–366, 2021.
[18] H. Liu, S. Zhu, and B. Zeng, "Inter-frame dependency-based rate control for VVC low-delay coding," *IEEE Signal Process. Lett.*, vol. 29, pp. 2727–2731, 2021.
[19] ITU-R, *Recommendation BT.500-10: Methodology for the subjective assessment of the quality of television pictures*, 2000.
[20] Fraunhofer Heinrich Hertz Institute, *VVCSoftware VTM*, accessed June 2023. [Online]. Available: https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM
[21] F. Liu and Z. Chen, "Multi-objective optimization of quality in VVC rate control for low-delay video coding," *IEEE Trans. Image Process.*, vol. 30, pp. 4706–4718, 2021.

Maryam Ranjbar received received B.S. and M.S. degrees in Computer engineering from the University of Lurestan and K. N. Toosi University of Technology.
Email: ranjbarmaryam@email.kntu.ac.ir

Hoda Roodaki Hoda Roodaki received B.S. and M.S. degrees in Computer engineering from the University of Tehran, and Sharif University of Technology, in 2005 and 2007, and a Ph.D. degree in Computer Architecture from the University of Tehran, Iran in 2014. Since 2015, she has been an Assistant Professor at the Computer Engineering Department, K. N. Toosi University, Tehran, Iran.
Email: hroodaki@kntu.ac.ir