Turkish Journal of Computer and Mathematics Education *Vol.10 No.02 (2019), 975-981*
DOI: https://doi.org/10.17762/turcomat.v10i2.13578

Research Article

Versatile Video Coding (VVC) Standard: Overview and Applications

Amit Kumar Mishra
Department of Computer Science & Information Technology, Graphic Era Hill University, Dehradun Uttarakhand India 248002

Abstract
Information security includes picture and video compression and encryption since compressed data is more secure than uncompressed imagery. Another point is that handling data of smaller sizes is simple. Therefore, efficient, secure, and simple data transport methods are created through effective data compression technology. Consequently, there are two different sorts of compression algorithm techniques: lossy compressions and lossless compressions. Any type of data format, including text, audio, video, and picture files, may leverage these technologies. In this procedure, the Least Significant Bit technique is used to encrypt each frame of the video file format to be able to increase security. The primary goals of this procedure are to safeguard the data by encrypting the frames and compressing the video file. Using PSNR to enhance process throughput would also enhance data transmission security while reducing data loss.
Keywords: Versatile Video Coding, Image and Video Compression, Data Encryption, Least Significant Bit Technique.

1. Introduction

The main objectives of this process are to protect the data by compressing the video file and encrypting the frame data. Data transmission security will be improved while data loss was decreased by using PSNR to speed up processes. Picture compression uses data compression to encrypt the original image using a minimal number of bits. Reducing picture redundancy and storing or transmitting data in an effective manner are the goals of image compression.

Reduced accuracy and a greater compression ratio are the two goals of quantization. As an illustration, the original image requires 8 bits to store one element per pixel. If we choose less bits, like 6, to preserve the image's information, the storage required will be reduced and the image may be compressed. The drawback of quantization is that it is a lossy process, leading to accuracy loss and irrecoverable distortion.

Entropy coding's primary goal is to reduce the average length of the picture. In entropy coding, codewords are assigned to related symbols in accordance with the likelihood of the symbols. Entropy encoders, which replace symbols denoted by equal-length codes with codewords whose length is inversely proportionate to the associated probability, are typically used to compress data. Lossy or lossless image compression is possible. For archiving purposes, lossless compression is recommended and is frequently used for medical imaging, technical drawings, clip art, and comics. Lossy compression techniques generate compression artefacts, especially when utilised at low bit rates. When a significant drop in bit rate is desired but a little (and perhaps undetectable) loss of quality is acceptable, lossy algorithms are particularly well suited for natural pictures like photography. Visually lossless compression is the lossy compression that results in unnoticeable changes.

Encoding communications (or information) so that only authorised parties may read it is known as encryption in cryptography. Although encryption doesn't stop hacking, it helps lessen the chance that the data will be read by the hacker. An encryption method encrypts the message or data (known as plaintext) using an encryption algorithm to produce unintelligible ciphertext (ibid.). Typically, a key for encryption is used, which defines how the message should be encoded. Any opponent who has access to the ciphertext should be unable to decipher the original message in any way. However, an authorised person can use a decryption technique to decipher the ciphertext, which often necessitates a secret decryption key that adversaries are unable to get. An encryption technique often requires a key-generation mechanism to generate keys at random for technical reasons.

Since the early 20th century, the names "biometrics" and "biometry" have been used to denote the area of study and development of mathematical and statistical methods pertinent to problems with data analysis in the biological sciences. These terms are now also used to describe the emerging field of information technology that focuses on the automated identification of individuals using biological characteristics, including those predicated upon speech patterns, dynamic signatures, fingerprints, facial recognition, or hand measurements, particularly for authentication purposes.

The science of assessing bodily traits that are particular to each person, known as biometrics, enables us to confirm that a person is who they say they are. The storing and transmission of enormous amounts of data is another major issue with biometrics. The biometric signature is "anormalized" using a computer programme to take the same shape as signatures stored in the system's database. Biometric technologies may be able to prevent unwanted access to or fraudulent usage of ATMs, mobile devices, smart cards, desktop PCs, and computer networks by replacing PINs. The majority of the top banks are thinking about implementing biometrics for ATMs and as a broad strategy to thwart card theft. In the future, it will affect our daily life more.

2. Literature Survey

In many surveillance situations, when there is consistently a great distance among the camera as well as the objects (people) of interest, face image resolution augmentation is typically desirable. Face photos have a more regular structure than the generic images outlined above, making them easier to manipulate. In fact, we can work with input photographs of lesser quality for face super-resolution. To recover details, the fundamental strategy is to first utilise the face prior to magnify the input to a respectable medium resolution. Next, the local sparsity prior model is used. In fact, the answer is also handled in two phases as follows: 1) Global model: recover a medium-high resolution face picture using reconstruction constraints; nevertheless, the face subspace is the only place where the solution is explored; and 2) Local model: recover image features using the local sparse model. To save a picture in the dictionary using this way would take up too much RAM is major drawback of the system proposed [1].

Due to the fact that the majority of feasible sparsity measures are not convex, sparse approximation issues are computationally difficult. Numerous heuristic techniques have been developed for generating sparse approximations, however there are few assurances of their effectiveness in the literature. By substituting a corresponding convex function for the nonconvex sparsity measure, the convex relaxation method creates a convex programming issue. However, these methods can only be relied upon to find a locally optimum solution. A Gaussian sequence that was unrelated to the data held the secret. Their encoder reduces the encrypted data's size to 1 bit per sample. Carrier images and SCAN patterns produced by SCAN technique are both used in hybrid image encryption. After using the scan approach on the original picture or the carrier image, the original image and carrier image are combined to produce a severely deformed encrypted image. We obtain the encrypted picture by using the reverse procedure. It is possible to create new physical sampling systems that capture discrete, low-rate, incoherent samples of the analogue signal directly. However, one drawback is that the sparse basis on which the signal is to be represented could not be known or might not be accessible [4].

Other academics have lately started researching on alternate algorithms that could improve ISTA's performance. These techniques depends upon computing the subsequent iterate relied on not just the preceding one but also on two or more previously calculated iterates, just like FISTA proposes in this procedure. FISTA, where k is the iteration counter, has been demonstrated to converge in function values as O(1/k2). Theoretically, the global rate of convergence for both methods is the same, but conceptually and computationally, the two approaches are very different. The primary differences between FISTA and the suggested technique presented include the following: (a) on the algorithmic building blocks, the latter builds recursively a sequence of estimate functions k(•) that approximates F(•), whereas FISTA employs merely the standard projection-like step, assessed at an auxiliary point very particularly designed in terms of the two preceding iterates and an explicit dynamically updated step. Last but not least, we note that FISTA already had the values produced by ISTA and MTWIST at iterations 275 and 468, respectively, before they were obtained by those two methods at iteration 10,000 [5].

a novel biometric method employing palm print technology for online personal identification. A cutting-edge gadget for online palm print picture collecting and a powerful algorithm for quick palm print identification make up the system's two components. In order to make picture alignment for feature extraction easier, a reliable image coordinate system is created. The experimental findings show that the suggested solution is workable. It is possible to extract the main lines using algorithms like the stack filter. Because different people's palm prints may have identical primary lines, these principal lines may not adequately capture the distinctiveness of each individual's palm print. Author uses a feature vector, which is made up of two feature matrices—real and imaginary—to represent picture data in order to properly explain the matching process [6].

The feature level fusion method is used to increase the effectiveness of palm print identification. On a palm print picture, many elliptical Gabor filters are used, each with a distinct orientation, to extract the phase information. These filters are then combined using a fusion rule to create a single feature known as the Fusion Code. The suggested system's four key phases are as follows: Send a palm print image from our palm print scanner to a computer. Establish a coordinate system using the two key points that are located between the fingers, and then use that system to extract the centre portions. Different palm print pictures are positioned as a consequence. Use a variety of Gabor filters to convolute the core regions. Fusion Code is a feature vector created by coding the phases after merging the filter outputs. To compare the similarity of two Fusion Codes, use the normalised hamming distance; then, apply a dynamic threshold to make your choice [7].

According to this procedure, the main lines on the palm were used to verify the palm print. The modified finite Radon transform, which can successfully and efficiently extract primary lines even when the palm print pictures contain numerous long and severe wrinkles, is presented for the feature extraction step. Automatic personal verification is a critical issue that must be appropriately addressed in today's networked world. And one of the most significant and practical answers in this area is biometrics. Researcher interest in palm print-based verification systems (PVS) has increased recently. The PVS offers a number of unique benefits, including easy self-positioning, cheap cost capturing devices, low-resolution images, steady lines, and rich texture features. The most common methods for using texture involve extracting texture features using the 2-D Gabor filter, which has been found to function satisfactorily in regards of recognition rate and processing speed. Two significant flaws in the proposed system are the complexity of the method to construct and the fact that the encrypted picture is compressed, meaning that the technique cannot compress well [8].

3. Proposed System

A modified or more sophisticated variation of the VQ approach is the mean-removed vector quantization (MRVQ). The objective of MRVQ is to offer images with a higher quality than genuine VQ. The codebook creation, picture encoding, and image decoding are its three main steps. To accomplish the objective, each picture block is encoded using the block mean value and the residual vector index. For photos of great quality, MRVQue requires a bit rate that is significantly higher than VQ's. Mean-removed VQ codebook design is done using the training pictures. Multistage vector quantization is a quantization procedure that has a number of steps. Using a tiny codebook relative vector quantizer, the input vector is first quantized. A separate codebook is used to quantize the mistake once more in the second stage. The error between the prior stage's input vector and the stage's quantized outputs is quantized in the following stage.

The above-mentioned approach has a number of drawbacks, such as the fact that the scale of the final image is different from the original image, that it is difficult and expensive to compute, and that it cannot achieve the greatest spectral compression.

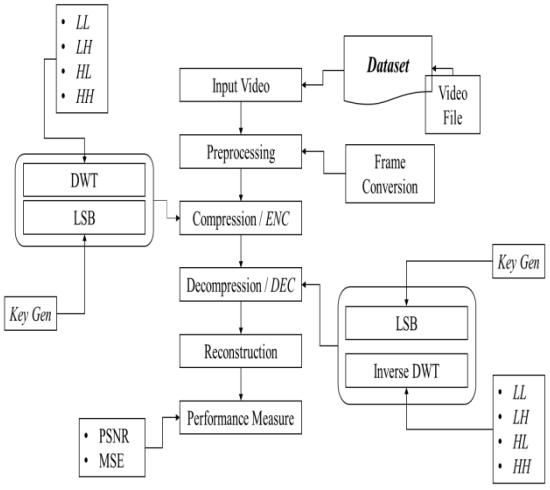

Fig 1: Flow Diagram

Most recent global video coding standard, recognized as Versatile Video Coding (VVC), was published in July 2020. It was created by the Joint Video Experts Team (JVET) of the ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG) to enable a larger range of current media material and new applications as well as to address the always expanding need for greater video compression.
The High Efficiency Video Coding (HEVC) standard, which achieved considerable bit rate reductions of over 50% over its predecessor for equivalent video quality, and around 75% over the presently most popular format, the Advanced Video Coding (AVC) standard, are described in this work along with the underlying compression methods. A key benefit of the suggested methodology is that the recommended feature extraction method performs well when dealing with noisy data sets.

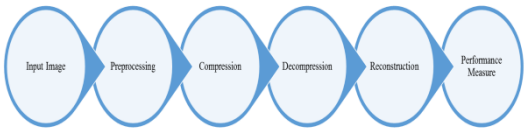

Fig 2: System Architecture

Various phases that are involved during implementation of proposed approach are being explained under following section:

1. **Image Acquisition**
   Every vision system begins with the picture acquisition phase. Picture acquisition is the process of digitising and archiving an image. To perform the numerous visual jobs required today, the image may be processed through a number of approaches after it has been taken. If the image was not gathered properly, the anticipated tasks would not be feasible even with the aid of any form of image improvement. Utilize the uigetfile and imread functions to first capture the input image from the source file. With the aid of the imresize feature, the user is able to alter the size of the picture by giving Pixel Height and Width.
2. **Preprocessing**
   A digital picture's resizing is referred to as image scaling in computer graphics and digital imaging. Upscaling or resolution improvement phrasing utilised in video technology to explain how digital content has expanded. The visual primitives that make up a vector graphic picture may be resized via geometric transformations without compromising the image's quality. Raster graphics images must be scaled by creating a new picture with more or less pixels. When the number of pixels is reduced (scaling down), there is typically a noticeable quality reduction.
3. **Compression / ENC**
   Any wavelet transform for which the wavelets are discretely sampled is referred to as a discrete wavelet transform (DWT) in numerical analysis and functional analysis. It incorporates both frequency and position information, giving it, like other wavelet transforms, a significant advantage over Fourier transforms in terms of temporal precision (location in time).
4. **Reconstruction**
   The picture was first encrypted by the user and then decrypted using the generated key. It is the picture encryption process done backwards. The repositioned column and rows carry over the original picture retrieval and location. The information was then kept in the designated place. The decryption procedure won't be carried out if the user provides the incorrect key, making it impossible to recreate the data that was encrypted into the image.
5. **Performance Measure**
   To analyse performance, the PSNR value is calculated. Maximum Signal-to-Noise Ratio An approximate measure of how well a reconstruction is perceived by people is PSNR. It is determined using the Mean Square Error (MSE). For an improved image, the PSNR value is high. When comparing a processed picture to the original (non-compressed) image, PSNR is a regularly used indicator of how closely two images resemble one another. When comparing compression codecs, one reconstruction may occasionally seem to be more similar to the original than another even if it has a lower PSNR because it uses an estimate for human perception of reconstruction quality (A greater PSNR would typically suggest a higher-quality reconstruction). This metric's range of validity must be

carefully considered because it can only be utilised to evaluate outcomes from the same codec (or codec type) and similar material in a way that is indisputably valid.

4. Results

The major goal of this procedure is to protect the data by compressing the video file and encrypting the frames. By using PSNR, this method also aims to increase performance while reducing data loss. Because compressed data is safer than uncompressed pictures, image and video compression and encryption are crucial components of information security. Additionally, handling data with small numbers of elements is simple. Because of this, data may be sent in an effective, secure, and simple manner thanks to effective data compression technologies. The Least Significant Bit technique is used to encrypt each frame in the video file format as part of this procedure to increase security.

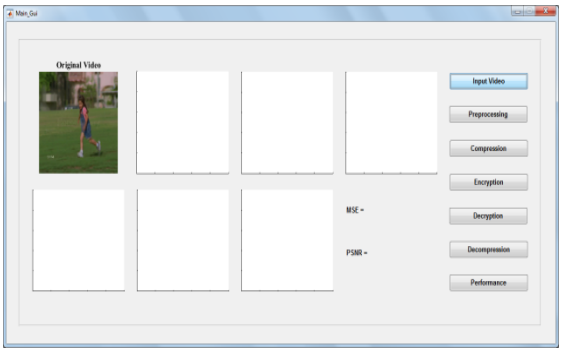

Fig 3: Input Image

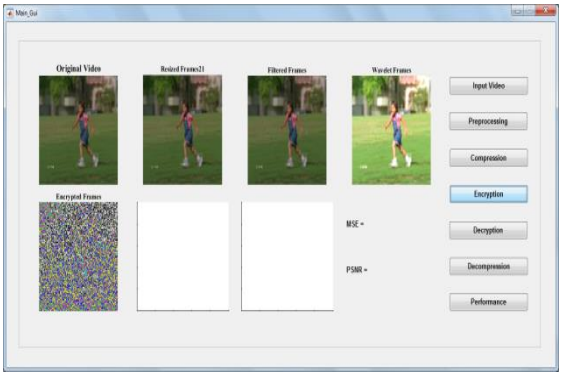

Fig 4: Encryption

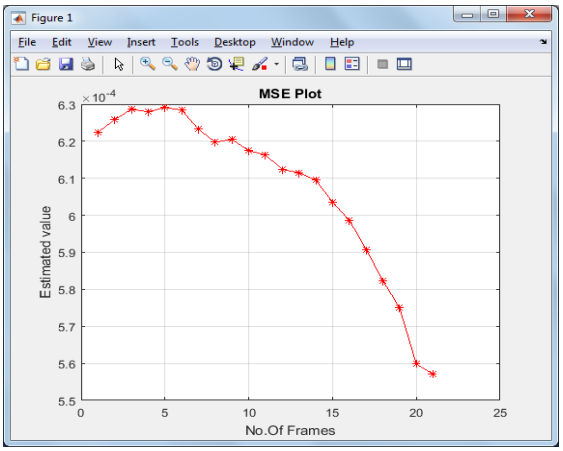

Fig 5: MSE Plot

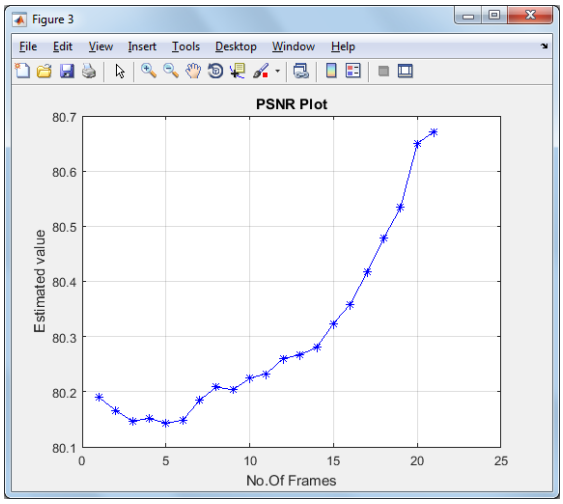

Fig 6: PSNR Plot

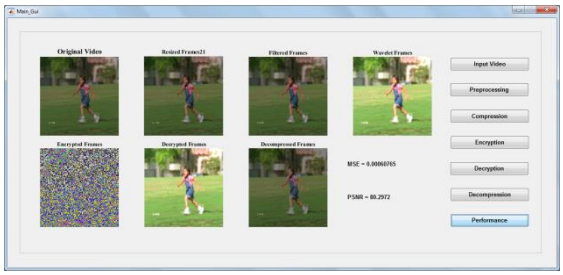

Fig 7: Performance Analysis

5. Conclusion

This essay is a project that effectively used JPEG picture compression. Software called MATLAB is used to develop the system. In the end, an 8x8 Compressed DCT picture was produced after this project underwent exhaustive testing in the MATLAB environment on Windows XP. The blocking effect is one of the DCT's key issues and a source of criticism. Images in DCT are divided into blocks of 8x8 or 16x16 or larger. These blocks provide a difficulty since they become evident when the image is compressed at higher levels. The blocking effect is what is known as this. Only 4 coefficients are kept after this picture has been compressed using 8x8 blocks. This image clearly shows the blocking effect.

6. Future Enhancement

The two lossless data compression methods LZ77 and LZ78 will be employed in future study to compress the image.

Reference

[1] L. Guan, S. W. Perry, and H.-S. Wong, *Adaptive Image Processing: A Computational Intelligence Perspective*. Boca Raton, FL, USA: CRC Press, 2001.

[2] G. J. Sullivan, J.-R. Ohm, W.-J. Han and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard", IEEE Trans. Circuits Syst. Video Technol., vol. 22, no. 12, pp. 1649-1668, Dec. 2012.

[3] Halem, M., Chukkapalli, S. S. L., Shivadekar, S., & Nguyen, P. Satellite Data Fusion of Multiple Observed XCO2 using Compressive Sensing and Deep Learning. In AGU Fall Meeting Abstracts (Vol. 2019, pp. B11F-2400).

[4] Y. Zhang, X. Ji, H. Wang, and Q. Dai, "Stereo interleaving video coding with content adaptive image subsampling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 7, pp. 1097–1108, Jul. 2013.

[5] C. Liu, A. Zhou, Q. Zhang, and G. Zhang, "Adaptive image segmentation by using mean-shift and evolutionary optimisation," *IET Image Process.*, vol. 8, no. 6, pp. 327–333, 2014.

[6] Jeysudha, A., Muthukutty, L., Krishnan, A., & Shivadekar, S. (2017). Real Time Video Copy Detection using Hadoop. International Journal of Computer Applications, 162(9), 42-45.
[7] Sayed, A., Sardeshmukh, M., & Limkar, S. (2014). Improved Iris Recognition Using Eigen Values for Feature Extraction for Off Gaze Images. In ICT and Critical Infrastructure: Proceedings of the 48th Annual Convention of Computer Society of India-Vol II: Hosted by CSI Vishakapatnam Chapter (pp. 181-189). Springer International Publishing.
[8] R. Ferzli and L. J. Karam, "A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB)," *IEEE Trans. Image Process.*, vol. 18, no. 4, pp. 717–728, Apr. 2009.
[9] H. H. Tong and A. N. Venetsanopoulos, "A perceptual model for JPEG applications based on block classification, texture masking, and luminance masking," in *Proc. Int. Conf. Image Process. (ICIP)*, 1998, pp. 428–432.
[10] G. Yammine, E. Wige, and A. Kaup, "A no-reference blocking artifacts visibility estimator in images," in *Proc. Int. Conf. Image Process. (ICIP)*, Sep. 2010, pp. 2497–2500.
[11] K.-H. Thung, R. Paramesran, and C.-L. Lim, "Content-based image quality metric using similarity measure of moment vectors," *Pattern Recognit.*, vol. 45, no. 6, pp. 2193–2204, 2012.
[12] L. Li, H. Zhu, G. Yang, and J. Qian, "Referenceless measure of blocking artifacts by Tchebichef kernel analysis," *IEEE Signal Process. Lett.*, vol. 21, no. 1, pp. 122–125, Jan. 2014.
[13] J. S. Lee, "Digital image enhancement and noise filtering by use of local statistics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-2, no. 2, pp. 165– 68, Mar. 1980.
[14] K. Lee, D. S. Kim, and T. Kim, "Regression-based prediction for blocking artifact reduction in JPEG-compressed images," *IEEE Trans. Image Process.*, vol. 14, no. 1, pp. 36–48, Jan. 2005.
[15] E. Nadernejad, N. Burini, and S. Forchhammer, "Adaptive deblocking and deringing of H.264/AVC video sequences," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, May 2013, pp. 2508–2512.
[16] B. Gatos, I. Pratikakis, and S. J. Perantonis, "Adaptive degraded document image binarization," *Pattern Recognit.*, vol. 39, no. 3, pp. 317–327, 2006.