Total Variation Regularization Technique for the Improvement of Video Quality by the Reduction of Compression Artifact

Neethu Kuriakose  
Department of ECE  
Karunya University  
Coimbatore, India  
neethukuriakose12@gmail.com

Mr. Shanty Chacko  
Department of ECE  
Karunya University  
Coimbatore, India  
shantychacko@karunya.edu

Jinu Mathew  
Department of ECE  
Karunya University  
Coimbatore, India  
jinu.veliyathu@gmail.com

ABSTRACT

When we encode the data or information using lesser number of bits compared to the actual representation, it can be called as the data compression. In the case of lossy data compression, some of the media’s data is avoided to make it simplified to be stored in the available space. Compression artifact occurs as a result of the lossy data compression at the time of decompression. Many techniques are available for image denoising. In the case of good image denoising technique, one of the main advantage is the removal of noise with edge preservation. In linear denoising technique, the edges in the image are considered as the discontinuities in the image and thus they are flushed out. In non linear techniques, Total Variation (TV) filtering is one of the best approaches for reducing compression artifact. This paper proposes an efficient approach for reducing the compression artifact in videos by using Total variation regularization technique without removing the information content. We mainly focus on removing blocky noise and mosquito noise. As a result, video with improved quality is obtained with better PSNR.

Keywords- Compression artifact; Blocky noise; Mosquito noise; TV Regularization decomposition.

I. INTRODUCTION

Data compression [1] is the process in which we encode the information using lesser number of bits compared to the actual representation. There are two types of compression namely lossy compression and lossless compression [2]. In the case of lossless compression, it identifies the statistical redundancy [3] and eliminates it and thus reduces the number of bits. In the case of lossy compression, it finds out the less or marginally important data and removes it and thus reduces the number of bits. Data compression means the reduction of the size of an information file.

At low bit rate coding, the compressed signals which are block based experience many artifacts such as blocking, flickering, mosquito and ringing artifacts. When we compress each block separately, the correlation between those pixels which are at the border of neighboring blocks, are breaked and this leads to blocking artifacts. When DCT coefficients are quantized with a coarse quantization step there occurs the loss of high frequencies which leads to the ringing artifacts. Ringing artifacts are dominant along the edges and are similar to the Gibb’s phenomenon [4]. When ringing artifacts of many single compressed frames are sequencely displayed, there arise the mosquito artifacts. For the blocks on the boundary of moving object and background which have relevant interframe prediction errors in the residual signal [5], mosquito artifacts become more annoying. Due to quantizing of residual signal, temporal distortion occurs over the compressed frames and this leads to the inconsistency. For the frames at the same spatial position, sometimes there occurs the inconsistency in quality which leads to the flickering artifacts [6], [7]. These flickering artifacts are perceived more in the flat areas. They also come from different quantization levels for rate-distortion optimization.

Many denoising techniques are existing for the reduction of these artifacts. For the reduction of blocking artifacts, [8] proposes a low pass filter which remove the high frequencies which are caused by the blocky edges at borders. But, in this type of blocking artifact reduction method, since the high frequencies components of the image are removed, there comes the excessive blur. Low pass filters were applied to the DCT coefficients of the blocks which are shifted in [9]-[11]. To avoid the problem of over-blurring the images, adaptive linear filters in [10] & [11] were introduced. But, the drawback of this method is
that their computation is highly complex. The techniques in [12] & [13] uses the linear or nonlinear isotropic filters to the ringing areas for the effective reduction of ringing artifacts.

In this paper, Total Variation regularization technique is used to deal with the blocking and ringing artifacts in videos while preserving the edges in a very good manner.

Section II describes the detail of our proposed method with the detailed explanation of the Total Variation regularization method. The simulation results are shown in section III. Concluding remarks is given in section IV.

II. PROPOSED METHOD

A. Steps

1) Generated the video with artifact which is the input.
2) Video with artifact is decomposed using TV regularization decomposition method into Structure component & Texture component.
3) Structure component is passed through sobel filter to extract the edge components.
4) Edge information from the sobel filter and texture component are fed to Gaussian filter where the edge parts of texture component are filtered to remove the mosquito noise.
5) Output of Gaussian filter and structure component are added with the help of an adder.
6) Now the output of adder is passed through Deblocking Edge Filter (DEF) to remove the blocky noise to get the final artifact reduced video.

B. Total Variation Regularization

For the effective reduction of artifact, the Total Variation Regularization technique [14] is used and is explained below. This method is very effective for regularization of images. Let us consider the following function

\[ E(s) = \int [\nabla s] dx dy + \alpha \int |s|^2 dx dy \]  

(1)

The above equation is known as the ROF model for the original TV regularization. It was proposed by Rudin, Osher, and Fatemi. The TV regularization is a process which is used to minimize the function given by (1). In (1), \(|\nabla s|dx dy\) is a TV term and \(\alpha \int |s|^2dx dy\) represents the constraint condition. How much the texture component is constrained to the original input signal is represented by \(\alpha\). With the help of projected iteration technique [15] we can solve the problem of ROF minimization.

The work done by the Meyer about the Rudin-Osher-Fatemi (ROF) algorithm is the first work of the image decomposition model. We can now discuss about the ROF model as in [16].

Let the observed image be \(i\). Let \(s\) be the ideal scene or it is the image which we want to retrieve and \(n\) be the noise. \(i\) is the addition of \(s\) and \(n\). To get \(s\), the solution is to minimize the following functional.

\[ E_{\alpha}^{\text{ROF}}(s) = K(s) + \alpha \int |s|^2 dx dy \]  

(2)

Consider that the image is decomposing into structural and textural part. That is, \(i = s + t\). Then, (2) can be modified as

\[ E_{\alpha}^{\text{ROF}}(s,t) = K(s) + \alpha \int |t|^2 dx dy \]  

(3)

where \(s\) denotes the structural part and \(t\) denotes textural part. Inorder to solve the Meyer model, consider a dual method approach. This approach arises because of the dual relation between \(H\) and Bounded Variation (BV) spaces. BV is the Bounded Variation spaces. It is one of the spaces of interest. This space is frequently used in image processing since it is very good to modelize structures in images. A modified functional is proposed to minimize

\[ E_{\alpha,\beta}^{\text{AU}}(s,t) = K(s) + K'(\frac{t}{\beta}) + (2\alpha)^{-1} \|i - s - t\|^2_{M^2} \]  

(4)

and

\( (s,t) \in BV(\emptyset) \times H^{\beta} \)  

(5)

The set \(H^{\beta}\) can be considered as the subset in \(H\) where \(\forall s \in H^{\beta}, \|s\|_H \leq \beta\). Also, \(K^*\) can be called as the dual operator of \(K\) (\(K^* = K\)).

\[ K^*(t) = \begin{cases} 0 & \text{if } t \in H^1 \\ +\lambda & \text{else} \end{cases} \]  

(6)

Here, the projector over the sets \(H^{\beta}\) \(\forall \beta\) are the Chambolle’s projectors. These projectors are denoted as \(P_{H^{\beta}}\).

Now, consider the algorithm that gives the minimizers \((s,t)\) of \(E_{\alpha,\beta}^{\text{AU}}(s,t)\).

- Fix \(t\), now search for the minimizer \(s\) of

\[ \inf(K(s) + (2\alpha)^{-1} \|i - s - t\|^2_{M^2}) \]  

(7)
Fix s, now search for the minimizer t of

\[ E_{\alpha,\beta}^{AU}(s, t) = K(s) + K^{\gamma}(\frac{t}{\beta}) + (2\alpha)^{-1}||s - t||_{L^2}^2. \tag{8} \]

We can find the solutions of (7) and (8) with the help of Chambolle’s results.

Solution of (7) is

\[ s^* = i - t^* - P_{H_{\alpha}}(i - t^*) \tag{9} \]

Solution of (8) is

\[ t^* = P_{H_{\beta}}(i - s^*) \tag{10} \]

So, we can write the algorithm for image decomposition as

1) First we have to initialize the initial values of structural and textural component as 0. That is, \( s_0 = t_0 = 0 \).

2) Now we have to perform the iteration up to \( n+1 \) times.

- Texture component

\[ t_{n+1} = P_{H_{\beta}}(i - s_n) \tag{11} \]

- Structure component

\[ s_{n+1} = i - t_{n+1} - P_{H_{\alpha}}(i - t_{n+1}) \tag{12} \]

3) Continue this algorithm until

\[ \max[|P_{H_{\alpha}} - s_n||t_{n+1} - t_n|] \leq \epsilon \tag{13} \]

or a prescribed maximal number of iterations is reached. For this algorithm, the value of the parameter \( \lambda \) can be taken as 50 and \( \mu \) can be taken as 100. For solving the above mentioned algorithm, we want the following equations.

Let the size of the processed image be \( P \times Q \). Let \( X = \mathbb{R}^{P \times Q} \) and \( Y = X \times X \). Assume \( s \in X \). Then, \( \nabla s \), which is the discrete gradient of \( s \) can be defined as

\[ (\nabla s)_{i,j} = \left( \begin{array}{c} s_{i+1,j} - s_{i,j} \\ 0 \end{array} \right) \text{ if } i < P - 1 \\
\left( \begin{array}{c} 0 \\ s_{j+1,i} - s_{j,i} \end{array} \right) \text{ if } j < Q - 1 \tag{14} \]

where \( \nabla s \in Y = X \times X \)

\( (\nabla s)_{i,j}^{1} = \left( \begin{array}{c} \nabla_{1}s_{i,j} \end{array} \right) \)

and

\( (\nabla s)_{i,j}^{2} = \left( \begin{array}{c} \nabla_{2}s_{i,j} \end{array} \right) \)

Also, the numerical divergence operator \( \text{div}: Y \rightarrow X \) can be defined by the following:

\[ \text{div}(q)_{i,j} = \begin{cases} q_{i,j} - q_{i,j+1} & \text{if } i < P - 1 \\ q_{i,j} & \text{if } i = 0 \\ -q_{i+1,j} & \text{if } i = P - 1 \\ q_{i,j+1} - q_{i,j} & \text{if } j < Q - 1 \\ q_{i,j} & \text{if } j = 0 \\ -q_{i,j+1} & \text{if } j = Q - 1 \end{cases} \tag{17} \]

where \( \text{div} = -\nabla^* \) (adjoint of \( \nabla \) is \( \nabla^* \)), and \( q \in Y(q = (q^1, q^2)) \).

Also, for every \( q \in Y \) and \( s \in X \), \( \langle -\text{div}q, s \rangle = \langle q, \nabla s \rangle \).

The equation for Chambolle’s Projector can be defined as

\[ q_{i,j}^{n+1} = \frac{q_{i,j}^{n} - \tau \left( \nabla^2 \text{div}^*(q^{n} - \frac{h}{\alpha})_{i,j} \right)}{1 + \tau |\nabla^2 \text{div}^*(q^{n} - \frac{h}{\alpha})_{i,j}|} \tag{18} \]

where \( h \in X, \alpha > 0, n \geq 0, \tau > 0 \).

Here, by using TV regularization decomposition method, the input video \( i \), is decomposed into two sub components, \( s \) and \( t \). The first component, \( s \), is known as the structure component and the second component, \( t \), is known as the texture component.

Structure component, which is obtained by the decomposition of the input video using the TV regularization method, is well-structured and it models the homogeneous objects which are contained in the video. It comprises of smooth signals and the edge with only very little amount of noise.

Texture component, which is obtained by the decomposition of the input video using the TV regularization method, contains both the textures and noise. All the mosquito noise and blocky noise are isolated into the texture component. Selective filtering is implemented in texture component since simple filtering leads to the loss of small texture components which leads to the degradation of picture.

Structure component is passed through the sobel filter to extract the edge components. The Gaussian and the Deblocking Edge Filter are controlled by using the edge information from this filter.
The texture component and edge information from the sobel filter are fed to the Gaussian filter. Gaussian filter process only the edge components of the texture component thus removing the mosquito noise.

Now, the output of Gaussian filter and structure component are added with the help of an adder. Finally, output of adder is passed through Deblocking Edge Filter [17] for removing blocky noise.

III. SIMULATION RESULTS

To demonstrate our proposed method, which reduces the compression artifact, simulations have been performed for many moving images. We apply our proposed method to moving images such as Car traffic, Goldfish, Stefan and Horse racing. In this paper, comparison results are presented only for the Car traffic and Goldfish since the simulation results for the other moving images are similar to those of these ones. The experimental moving images are the Car traffic, which is 160×120 pixels per frame with total 119 frames and the Goldfish, which is 320×240 pixels per frame with total 1542 frames as shown in Fig. 1 and Fig. 2.

In the proposed method, the input video with artifact is decomposed into texture and structure component using the TV regularization decomposition method. The structure component gives the details of the edge components and comprises of smooth signals and edges and is shown in Fig. 5 and Fig. 6.
The mosquito noise, which occurs due to the quantization of high frequency coefficients and the blocky noise, which occurs due to the quantization of low frequency coefficients, is separated into the texture component.

The structure component is passed through sobel filter to extract only the edge components. The edge information from the sobel filter and the texture component are fed to the Gaussian filter. Gaussian filter process only the edge components according to the information about the edges of the texture component thus removing the ringing noise. With the help of an adder, the structure component and gaussian filter output are added. Thus artifact reduced video is obtained at the output of adder and is shown in Fig. 7 and Fig. 8.

Finally, the adder output is passed through DEF which is used for the reduction of the blocky noise. Thus the artifact reduced video is obtained at the DEF output which is given in Fig. 9 and Fig. 10.

The relationship between PSNR and frame number is shown in Fig. 11.
IV. CONCLUSION

This paper proposes an effective method for the reduction of compression artifacts in video. In the proposed method, with the help of Total Variation regularization decomposition method the input video with artifact is decomposed into texture and structure component. After the decomposition, all the compression artifacts are accumulated in the texture component. Then with the help of gaussian filter ringing noise is decreased which were detected with the help of sobel filter. Finally by using the DEF method blocky noise is decreased without any loss in small texture components.

In the experimental results it is clear that there is only less blocky noise and mosquito noise in all the frames of the output video reconstructed using the proposed method. Also, the PSNR in all the frames is also higher.

REFERENCES


