Adaptive GOF residual operation algorithm in video compression

Jiyan Pan, Bo Hu

Digital Signal Processing and Transmission Laboratory, E.E. Dept., Fudan University
No.220 Handan Road, Shanghai, 200433, P.R.China

ABSTRACT

Residual operations are introduced in many video compression algorithms to exploit temporal redundancy in video sequences. One of the most effective and computationally simple algorithms is 3D-DWT-SPIHT algorithm. Nevertheless, it only utilizes intra-GOF temporal redundancy. In order to eliminate inter-GOF redundancy, the Simple GOF Residual Operation (SGRO) algorithm was introduced, where residual operations are performed every two GOFs. However, when background mutations occur, the PSNR of reconstructed target GOF significantly drops; on the other hand, when video sequences are temporally stationary, SGRO algorithm fails to fully utilize inter-GOF temporal redundancy due to its imperative insertion of no-residual-operation GOFs. In this paper, we propose a new Adaptive GOF Residual Operation (AGRO) algorithm, based upon a criterion on residual operations, which is derived from an empirical formula of image complexity established by us. AGRO algorithm always selects the best residual operation manner according to the contents of video sequences: by detecting contents of video sequences, it cancels residual operations where background mutations happen, while encouraging residual operations where video sequences are temporally stationary. Therefore, AGRO algorithm prevents the significant drop in compression effects resulted from background mutations, and in the meantime, fully utilizes inter-GOF temporal redundancy. In addition, AGRO algorithm demonstrates an innate error-propagation-resistant property. Numerical results show that AGRO algorithm renders a significant PSNR increase over SGRO algorithm whenever background mutations occur or temporally stationary sequences dominate.

Keywords: Residual operation, wavelet, SPIHT, video compression, image complexity, background mutation

1. INTRODUCTION

One of the major parts of video compression is to exploit redundancy as much as possible in video sequences. The redundancy in video sequences includes spatial redundancy and temporal redundancy. The former one is exploited by using various kinds of transforms, such as DCT[1] or DWT[2]. As DWT does not require dividing an image into blocks as DCT does, the reconstructed images do not suffer from block effect that DCT-based video compression usually has. The elimination of temporal redundancy can be achieved by using motion estimation[3] or 3D-DWT[4]. Generally, if 3D-DWT is used, combined with 3D-SPIHT[5, 6], it is not necessary to perform the somewhat complicated and resource-consuming motion estimation, and the combined algorithm is called 3D-DWT-SPIHT algorithm[7].

The mechanism that 3D-DWT-SPIHT algorithm achieves good compression performance can be explained as follows. DWT demonstrates a property of energy concentration, which concentrates most of the image energy into relatively few low-frequency-subband DWT coefficients, leaving the majority of DWT coefficients (located in high frequency subbands) rather small[2, 8]. Unlike DFT, spatial and temporal formation is not lost during DWT process, making it possible to know the exact location of image details at various scales (in different resolutions). Generally speaking, in the case of multilevel DWT, the DWT coefficients located in the low frequency subband are much larger than those located in high frequency subbands, and the DWT coefficients located in higher levels of high frequency subbands are larger than those located in lower levels of high frequency subbands. In the perspective of raising PSNR during the process of sequence reconstruction, the larger a DWT coefficient is, the more contribution the coefficient will make in increasing PSNR, thus the more important the coefficient is. Therefore, at the side of an encoder, larger coefficients should enjoy a higher priority to be transmitted. Such a principle is embodied in SPIHT algorithm, which makes use of the properties of DWT coefficients mentioned above, and the property of self-similarity of DWT coefficients along the same spatial orientation[5].

Visual Communications and Image Processing 2005, edited by Shipeng Li,
Fernando Pereira, Heung-Yeung Shum, Andrew G. Tescher, Proc. of SPIE Vol. 5960
(SPIE, Bellingham, WA, 2005) · 0277-786X/05/$15 · doi: 10.1117/12.632503

Nevertheless, 3D-DWT only utilizes temporal redundancy within a group of frames (GOF). In order to achieve even better compression effects, the temporal redundancy between adjacent GOFs should also be eliminated. Moreover, in 3D-DWT-SPIHT algorithm, the ideal situation is that almost all the energy is concentrated into the low-frequency-subband coefficients, and in such a situation, 3D-SPIHT can achieve very high efficiency. However, if there are many edges, strips, and other high frequency components in images, the DWT coefficients located in high frequency subbands are relatively large, and under a given compression rate, some of those high-frequency coefficients are not transmitted during the process of SPIHT encoding, thus making PSNR drops significantly. In order to decrease the amount of high-frequency components in images, residual operation should be introduced to reduce the complex still structures of images. A GOF residual operation algorithm is developed to serve this purpose[9]. In this algorithm, video sequences are divided into 8-frame GOFs, and residual operations are performed on these GOFs alternatively; that is, two GOFs make up a GOF group, of which the first GOF is processed without residual operation and the last image of it serves as a reference image during the process of residual operation performed on the second GOF. As the GOF residual operation is non-adaptive, we refer to this algorithm as simple GOF residual operation (SGRO) algorithm in this paper. When video sequences are temporally stationary in a wide range, SGRO algorithm provides better compression effects than the algorithm without residual operations.

However, SGRO algorithm performs residual operations in a fixed manner, with no regard for the contents of video sequences. As a result, in many cases where scene changes or background mutations occur (which are quite normal in actual video sequences), the residual operation surprisingly leads to a significant drop in peak signal to noise ratio (PSNR), because in these cases, residual images become more complex ones, with more edges, strips and details, which are harmful to efficient compression. So, when a major scene change or background mutation happens to occur between target images (the images that are currently being compressed) and the reference image, the initial intention of reducing complex structures will not be realized by residual operation at all, and such operation will even make things worse.

On the other hand, when video sequences are temporally stationary for a relatively long period of time, SGRO algorithm fails to fully utilize temporal redundancy between GOFs because of its imperative insertion of a no-residual-operation processed GOF every two GOFs to prevent error propagation[9]. As a matter of fact, much more residual operations could be done in a temporally stationary sequence, as long as a method to effectively control error propagation is figured out.

In sum, SGRO algorithm does not provide an optimal method for GOF residual operations. In this paper, a criterion is proposed that helps make judgment on whether to perform residual operation or not in a certain location of a video sequence, preceded by the definition of the “complexity” of a GOF. Based on such a criterion, a new algorithm called “adaptive GOF residual operation algorithm” (AGRO algorithm) is derived. Simulation results show that AGRO algorithm renders a significant PSNR increase over SGRO algorithm whenever background mutations occur or temporally stationary sequences dominate.

This paper is organized as follows. Section 2 introduces and defines a quantitative measure of the complexity of images. Section 3 gives the criterion on residual operation. AGRO algorithm is detailed in Section 4. Simulation results are given in Section 5. The conclusion of this paper is in Section 6.

2. DEFINITION OF *COMPLEXITY* OF IMAGES

Intuitively, the complexity of an image is related with the amount of fine structures contained in the image. Although for human eyes, it is not difficult to tell whether or not an image is more complex than another one, machines do not have such a capability. Therefore, a quantitative measure that is appropriate to reflect the complexity of images has to be found for machine processing.

Take a look at Fig. 1 and Fig. 2. We could tell immediately that Fig. 1 is more complex than Fig. 2, because there are more details in Fig. 1, such as the shelves, books, trees, and the stuffs on the man’s desk. In these areas, grey levels fluctuate frequently and drastically. By contrast, a large area of Fig. 2 is covered by a flat white wall, where few changes in gray levels exist. Therefore, image complexity (Precisely, the image complexity of Y component, and so for the other color components) is actually reflected by grey level fluctuations in the image. Firstly, the amount of such fluctuations plays a major role in determining image complexity, because with the increase of the number of grey level fluctuations, more and more high-frequency subband DWT coefficients in different spatial orientations will become greater, leading

to a reduce in compression efficiency. Secondly, the intensity of such fluctuations also makes contributions to image complexity, for it determines to what extent high-frequency subband coefficients get larger in value.

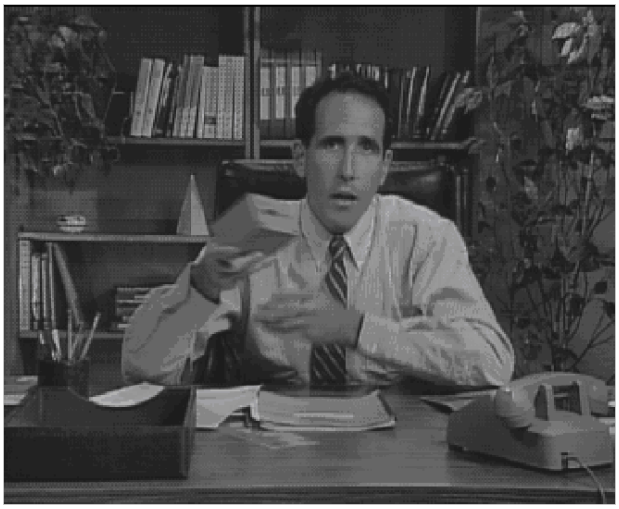

Figure 1: An image taken from “salesman”

Figure 2: An image taken from “children”

From what has been discussed above, it is reasonable to define image complexity as a combination of the amount of grey level changes contained in the image, and the intensity of such changes. The next problem is how to acquire the two quantities mentioned above. In order to get an image that only includes grey level fluctuation information of the original image, we could subtract the original image by a shifted version of it. That is, the difference of the original image by the shifted version of it is calculated. In the areas where grey level keeps constant, the difference is zero; in other areas where fluctuations or mutations of grey levels occur, for example, edges, the difference is nonzero, and is proportional to the intensity of the fluctuations or mutations. The shifted version of the original image is generated by subsequently shifting the original image one pixel horizontally and vertically, so that both the vertical and horizontal changes of gray levels can be detected. If *I* represents for the original image, and *D* represents for the difference image, the difference operation can be expressed as follows,

$$D(row, col) = I(row, col) - I(row+1, col+1)\,. \tag{1}$$

For Fig. 1 and Fig. 2, the effect of image difference is illustrated in Fig. 3 and Fig. 4.

In Fig. 3 and Fig. 4, color grey represents zero value. We could see that Fig. 4 is covered by a large area of zero values, which corresponds with the areas with constant grey levels in Fig. 2. Yet, nonzero values almost dominate Fig. 3, indicating that there are quite a lot of grey level changes in the original image, i.e., Fig. 1.

After the difference image is acquired, the number of pixels with nonzero values is appropriate to reflect the amount of grey level fluctuations in the original images, and the root mean square (RMS) of all those nonzero values serves as a proper measure for the intensity of grey level changes. Let’s denote the number of nonzero pixels and the root mean square of the nonzero values by *Cnt* and *RMS*, respectively. Experiments show that *Cnt* affects compression efficiency more than *RMS* does, and with the rise of *RMS*, compression efficiency becomes increasingly insensitive to the change of *RMS*. Based on these facts, and a great number of experimental data, the following empirical formula of image complexity is proposed:

$$Comp = Cnt \cdot \ln(RMS)/100\,, \tag{2}$$

where *Comp* stands for image complexity. That is, image complexity is proportional to the product of the number of nonzero pixels in the difference image and the logarithm of the root mean square of those nonzero values. The constant in the denominator serves to adjust image complexity within a proper value range. For example, the complexity of Fig. 1 is 2414, and the complexity of Fig. 2 is 1889.

This empirical formula provides an effective measure of image complexity in the respect of its influence on compression efficiency, and hence the PSNR of reconstructed GOF made up with such images. The increase in image complexity leads to the decrease in compression efficiency and a drop in the PSNR of recovered GOF, which is confirmed by experiment results given in the next section.

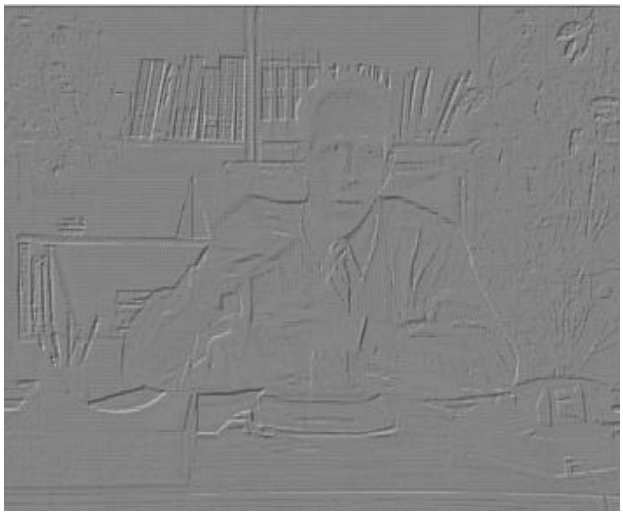

Figure 3: Difference image of Figure 1

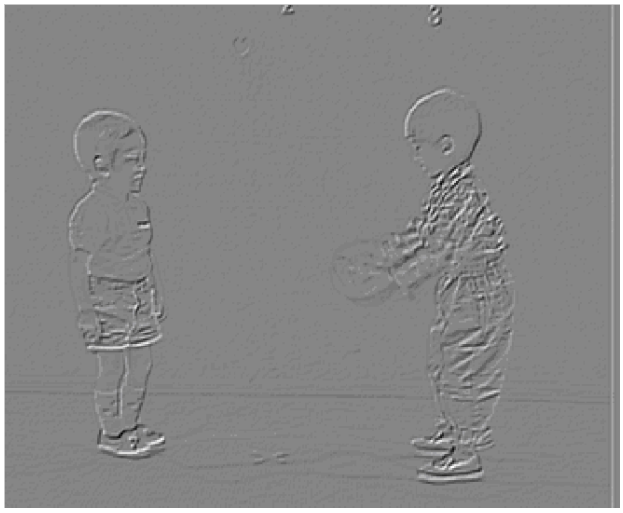

Figure 4: Difference image of Figure 2

3. CRITERION ON RESIDUAL OPERATION

In this section, the role that residual operation plays in the process of video compression is examined from the perspective of image complexity defined in the previous section and the PSNR of reconstructed GOF. Suppose the GOF currently under compression (referred to as target GOF) is the one composed of images similar to Fig. 1, that is, Fig.1 is typical of all the images in the GOF and its complexity can represent the complexity of the entire GOF. In the first case, the reference image is the one displayed in Fig. 5, which is only slightly different from Fig. 1. The residual image of Fig.1 by Fig. 5, which is acquired by subtracting Fig.5 from Fig. 1, is shown in Fig. 6. As we could see in Fig. 6, the complex still structures behind the man (e.g., shelves, books, trees, and the stuffs on his desk) disappear in the residual image, which is beneficial for further compression operation.

However, if the reference image is an image taken from another sequence “Children” (this is possible when the story about children are over and a salesman begins to tout about their toys.), as is shown in Figure 7, the residual image will be extremely complex, more than both images, as is shown in Figure 8.

In Fig. 8, the complex details of both Fig. 1 and Fig. 7 are present due to the residual operation. Now let’s take a look at the complexity of Fig. 1, Fig. 6 and Fig. 8: the complexity of Fig. 1 is 2414; the complexity of Fig. 6 is 2145; and the complexity of Fig. 8 is 2883. From the comparison of the complexity of the three figures, it is evident that when the reference image is Fig. 5, the residual image will have a lower complexity than the original one, and, the residual image is even more complex than the original one if the reference image is Fig. 7. Considering the relationship between PSNR of the reconstructed GOF and the complexity of its typical image, we might expect that residual operation in the first case increases PSNR of the recovered target GOF, while lower PSNR is yielded by the residual operation in the second case, where Fig. 7 serves as the reference image. Simulation results have confirmed our expectation: under the same compression rate, the PSNR without residual operation is 31.4dB, the PSNR in the first case is 35.5dB, and the PSNR in the second case is 28.6dB.

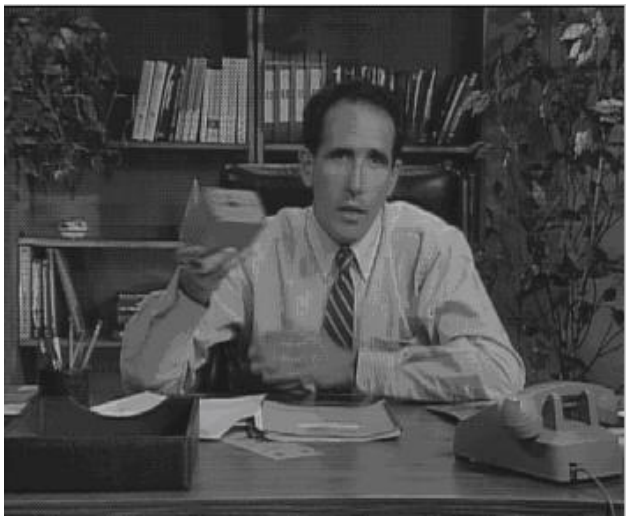

Figure 5: A reference image taken from “Salesman”

Figure 6: Residual image of Figure 1 by Figure 5

Figure 7: A reference image taken from “Children”

Figure 8: Residual image of Figure 1 by Figure 7

More experiment results are plotted in Fig. 9 through Fig. 12, each of which demonstrates the relationship of PSNR vs. complexity on a target GOF taken from the standard video sequences “Salesman”, “Miss American”, “Children”, and “Claire”, respectively. The reference image for each target GOF is taken from a wide range: from images of the same sequence, images of other sequences, rotated images, clipped images, random images, etc.

The simulation results demonstrate a monotone decreasing relationship between GOF complexity (represented by the complexity of its typical image) and the PSNR of the reconstructed GOF under all compression rates, laying a foundation for the criterion on residual operation described below.

The criterion on residual operation: Residual operation is performed only under the condition that the complexity of the residual target GOF is lower than that of the original target GOF. The complexity of a GOF is represented by the complexity of its typical image which is defined as the image in the middle of the GOF.

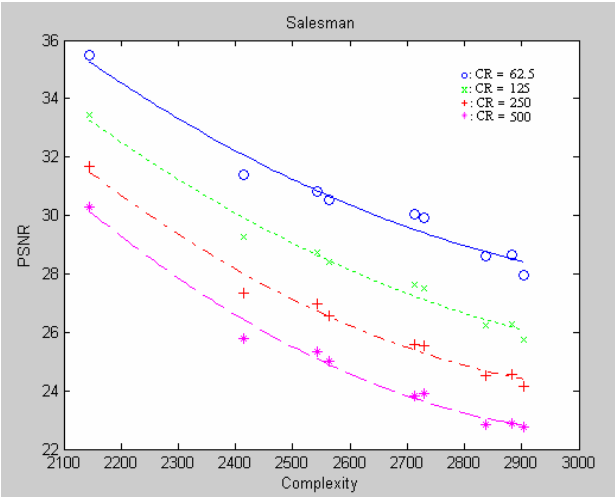

Figure 9: PSNR vs. Complexity on “Salesman”

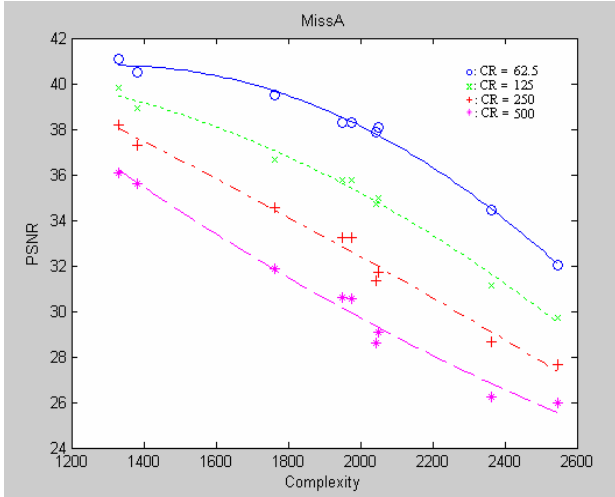

Figure 10: PSNR vs. Complexity on “MissA”

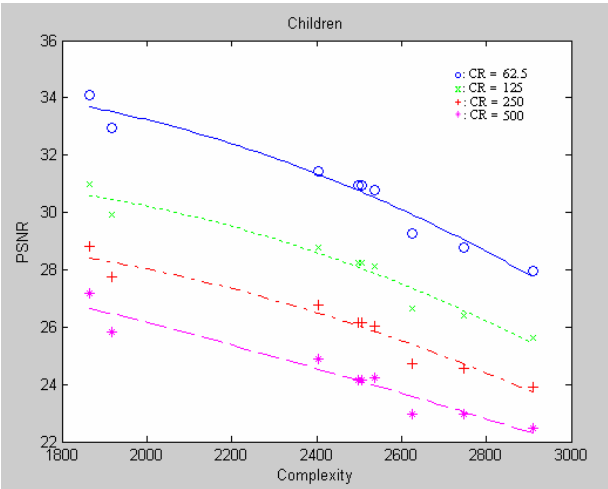

Figure 11: PSNR vs. Complexity on “Children”

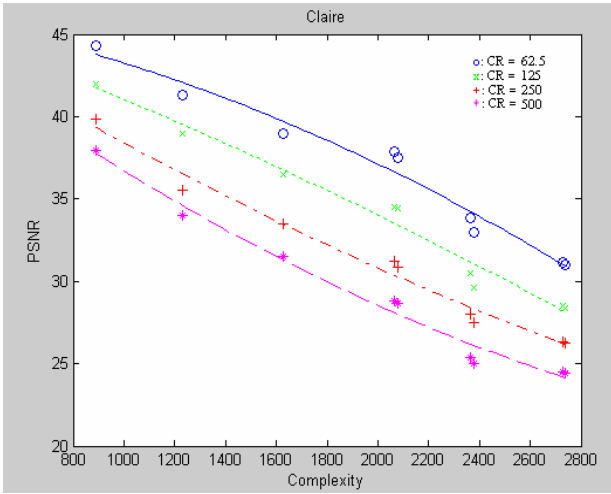

Figure 12: PSNR vs. Complexity on “Claire”

One thing to be noted is that since 3D-DWT is employed, motion within the GOF—the temporal complexity of the GOF—also has a significant influence on the PSNR of the GOF, thus the relationship between GOF (spatial) complexity and reconstructed GOF PSNR is derived under the assumption that temporal complexity of the GOF is constant. However, such an assumption imposes no restriction on the application of the proposed judgment criterion, because the only difference between whether to perform residual operation or not lies in the reference image alone (processing without residual operation amounts to performing residual operation with a zero-valued reference image), and no matter whether to perform residual operation, the motion information within the target GOF remains unchanged, that is, the temporal complexity of the target GOF is not affected by the judgment made on whether to perform residual operation. Therefore, we can simply use the criterion to make judgments on residual operation without having to concern about the influence of temporal GOF motion imposed on the PSNR of the recovered GOF.

4. ADAPTIVE GOF RESIDUAL OPERATION ALGORITHM

Based upon the criterion on residual operation mentioned in the previous section, a new residual operation algorithm called “adaptive GOF residual operation (AGRO) algorithm” is proposed in this paper.

In AGRO algorithm, the entire video sequence is divided into GOFs consisting of a certain number of images. Here, a GOF contains 8 images, so the typical image of a GOF is the 4th or the 5th frame of the GOF. There is no perceptible difference whether the 4th or the 5th frame is taken as the typical frame, so we choose the 5th frame here. In the simple GOF residual operation (SGRO) algorithm, residual operation is performed on GOFs alternately and fixedly, and those GOFs are called original (or reference) GOFs and residual GOFs, respectively[9]. In the new algorithm, no fixed reference GOFs or residual GOFs exist; actually a GOF is always a reference GOF, the last image of which serves as the possible reference image for the next target GOF. What matters in the new algorithm is whether residual operation is to perform on the current target GOF, depending on the principle of maximizing the PSNR when the GOF is reconstructed. Before processing any target GOF, judgment on whether to perform residual operation is made:

4.1 calculate the complexity of the typical image of the target GOF (Comp_Org);
4.2 calculate the residual image between the typical image of the target GOF and the last image of the previous recovered GOF (reference image);
4.3 calculate the complexity of the residual typical image (Comp_Res);
4.4 compare Comp_Org and Comp_Res, if the former is larger than the latter, residual operation is adopted; if not, process the target GOF without residual operation. The last image of the reconstructed target GOF becomes the possible reference image of the next GOF.

The flowchart of AGRO algorithm is illustrated in Figure. 13.

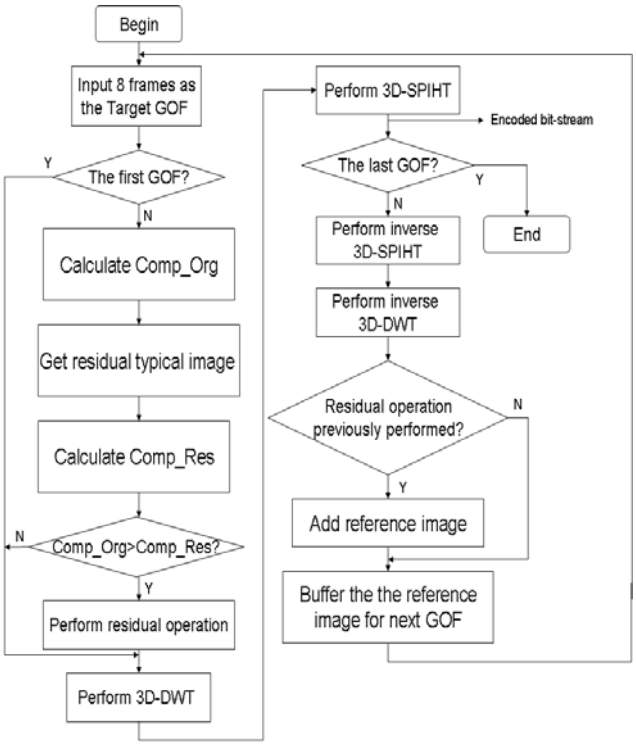

Figure 13: Flowchart of AGRO algorithm

AGRO algorithm always selects the best residual operation manner according to the contents of actual video sequences: by detecting contents of video sequences, it cancels residual operations where background mutations happen, while encouraging residual operations where video sequences are stationary in the temporal direction. As a result, PSNR will not suffer from background mutations any more. By contrast, SGRO algorithm does not care whether background mutations happen to take place when it performs residual operations[9], hence, great loss of compression performance is incurred.

Since in AGRO algorithm, the judgment on whether to perform residual operation is optimal in the sense of maximizing PSNR, there is no need to imperatively insert a GOF on which no residual operation is conducted every two GOFs, as is performed in SGRO algorithm. The imperative insertion of such GOFs in SGRO algorithm is used to prevent error propagation. However, AGRO algorithm exhibits an innate error-propagation-resistant property, because if error ever accumulates to such an extent that residual operation would result in lower PSNR, no residual operation would be performed on the current GOF, thus still maximizing the PSNR of any GOF. Therefore, no error propagation will occur in AGRO algorithm. On the other hand, when video sequences are stationary along temporal direction for quite a long period, frequent insertion of no-residual-operation GOFs, as is done in SGRO algorithm, will significantly reduce the exploitation of temporal redundancy between GOFs, which also leads to a considerable decrease in PSNR, compared with the compression performance of AGRO algorithm, where such temporal redundancy is always fully exploited.
PSNR loss due to background mutations occurring at the edge of two adjacent GOFs can be completely avoided by AGRO algorithm. However, if a background mutation occurs within the target GOF, complete prevention of PSNR loss can not be achieved; yet, AGRO algorithm still ensures that such PSNR loss gets minimized, because if AGRO algorithm decides to perform residual operation on the current GOF, the complexity of at least more than half of the frames in the current GOF can be reduced by residual operation, and residual operation will still increase the PSNR of the current GOF compared with processing the current GOF without residual operation, and vice versa.

5. NUMERICAL RESULTS

Simulation is conducted on a 48-frame video sequence composed of images from standard video sequences in order to examine the compression performance of AGRO algorithm compared with that of SGRO algorithm. The compression effects when no residual operation is employed are also included for comparison. Numerical results are listed in Table 1.

From the simulation results, the following conclusions could be drawn:

A. When no background mutation occurs, AGRO algorithm outperforms SGRO algorithm as a result of the greater exploitation of temporal redundancy. And, in this case, the algorithm without residual operation fails to achieve compression effects as good as those of the algorithms with residual operations (both adaptive and simple), because it does not utilize inter-GOF redundancy at all.
B. An average PSNR increase of over 2dB is achieved by AGRO algorithm when the sequence includes 5 background mutations. Since a background mutation occurs every 8 frames in the video sequence, residual operations are always cancelled. As a result, AGRO algorithm is equivalent to the algorithm without residual operation. SGRO algorithm, which runs with a fixed residual operation manner, fails to gain any improvement from residual operations at all; on the contrary, such residual operations significantly lower the PSNR.
C. Even if only one background mutation occurs in the sequence, experiment results show that the PSNR increase over SGRO algorithm can still be as high as 0.8dB~1.7dB when adaptive residual operation is performed. AGRO algorithm also achieves better compression effects than the algorithm without residual operation.
   It deserves to be noted that due to the only one background mutation, the PSNR of SGRO algorithm falls below that of the algorithm without residual operation. As no background mutation occurs in the first forty frames, the PSNR drop has been somewhat compensated by residual operations on those frames in SGRO algorithm; otherwise, the PSNR drop due to the background mutation occurring at the 41st frame would be extremely significant.
   When the number of background mutations is between 1 and 5, the PSNR gain by AGRO algorithm over SGRO algorithm ranges from 0.8dB to 2.3dB.
D. If a background mutation occurs within a GOF, rather than between two GOFs, although PSNR suffers a lot in this case, AGRO algorithm still demonstrates its superiority over SGRO algorithm and the algorithm without residual operation, no matter whether in the case where AGRO algorithm cancels residual operations or in the case where residual operations are adopted by AGRO algorithm.

Table 1: Comparison of PSNR (dB) achieved by AGRO algorithm, SGRO algorithm, and the algorithm without residual operations

CR	AGRO algorithm	SGRO algorithm	Algorithm without residual operation
No background mutation (1~48: Salesman)			
50	34.7589	33.7980	33.0904
100	31.8451	30.9602	30.3929
200	29.5058	28.8493	28.2788
400	27.6309	27.1547	26.7327
5 background mutations (1~8:Salesman, 9~16:Children, 17~24:Salesman, 25~32:Children, 33~40:Salesman, 41~48:Children)			
50	33.3057	31.0055	33.3057
100	30.3207	28.6575	30.3207
200	28.0992	27.0075	28.0992
400	26.4730	25.3706	26.4730
1 background mutation (1~40: Salesman, 41~48: Children)			
50	34.3538	32.6324	33.1654
100	31.3648	30.0117	30.3678
200	29.0947	28.0793	28.2153
400	27.2843	26.4373	26.6289
1 background mutation within a GOF and the corresponding residual operation is cancelled in AGRO algorithm (1~28: Salesman, 29~48: Children)			
50	31.2999	31.1240	31.0405
100	28.4004	28.2956	28.2157
200	26.8006	26.6778	26.5753
400	25.5159	25.3842	25.2757
1 background mutation within a GOF and the corresponding residual operation is adopted in AGRO algorithm (1~29: Salesman, 30~48: Children)			
50	31.0936	30.9822	30.5469
100	28.1206	28.0577	27.8818
200	26.2127	26.1017	25.8297
400	24.5466	24.4434	24.1460

6. CONCLUSION

In this paper, we have proposed a new algorithm, named adaptive GOF residual operation (AGRO) algorithm, which adaptively perform residual operations between GOFs. Such an algorithm is based upon an experiment-confirmed criterion on whether or not to perform residual operations between GOFs in order to achieve optimal compression effects. AGRO algorithm prevents PSNR from dropping by canceling residual operation when background mutations occur, and enhance compression effects by performing more residual operations when the sequence is temporally stationary. In the latter circumstance, error propagation is naturally under control due to the PSNR-optimized residual operation judgment in AGRO algorithm. Simulation results show that AGRO algorithm considerably raises the PSNR of reconstructed video sequence under all compression rates and all sequence circumstances.

Further research work can be conducted in the aspect of tackling with the problem of significant PSNR drop when background mutations occur *within* a GOF.

REFERENCES

1. Rao, K. R. and Yip, P., "Discrete Cosine Transforms - Algorithms, Advantages, Applications," Academic Press, 1990.
2. M. Antonini, Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," IEEE Trans. Image Processing, pp. 205-220, Apr. 1992.
3. Didier J. LeGall., "The MPEG video compression algorithm," in Signal Processing: Image Communication, **vol. 4**, No. 2, pp. 129-140, Apr. 1992.
4. Weeks, M. and Bayoumi, M., "3D discrete wavelet transform architectures," in *Proc. of the 1998 IEEE International Symp. on Circuits and Systems*, vol. 4, pp. 57-60, June 1998.
5. A. Said and W. Pearlman, "A new, fast and efficient image codec based on set partitioning in hierarchical trees," IEEE Trans, Circuits Syst. Video Technol., **vol. 6**, pp. 243-250, June 1996.
6. Y. W. Chen and W. A. Pearlman, "Three-dimensional subband coding of video using the zerotree method," in *Symp. on Vis. Comm. and Image Proc.*, SPIE vol. 2727, Mar. 1996.
7. B.-J. Kim and W. Pearlman, "An embedded wavelet video coder using three-dimensional set partitioning in hierarchical trees (SPIHT)," in *IEEE Data Compression Conference*, Mar. 1997.
8. S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," IEEE Trans. Pattern Anal. Machine Intell., **vol. 11**, pp. 674-693, July 1989.
9. Jia Hu, Wen-Qi Ding, Li-Ming Zhang, and Bo Hu, "Research on a novel 3D-DWT-SPIHT coding by alternate groups of original and residual error frame," in J. Infrared Millim. Waves, **vol. 23**, No. 4, pp. 265-270, Aug. 2004.