Transform Domain Restoration and Assessment for Compressed Videos

Xiao-hua Li*

The Information Center of Zhejiang University, China

* Corresponding author

Keywords: Video restoration, Wavelet, Transform domain, PVA.

Abstract. We test a sparsity-based soft decoding approach to restore compressed videos directly in the transform domain of compression (wavelet domain specifically examined in this paper) for non-material cultural heritage. Restoring transform coefficients rather than pixel values prevents the propagation of quantization errors in the spatial domain. A quality assessment model is employed for perceptual video compression applications to assess the quality of the restored results.

Introduction

The past years have witnessed a rapidly growth of research works on sparsity-based video analysis and processing. However, to the best of our knowledge, there is few works that applied the sparsity-based restoration approach to restore compressed video. Ironically, the most common cause of image degradation in practice is compression. Modern sensors, even the consumer grade digital cameras, offer sufficiently high spatial and spectral resolutions and high signal-to-noise ratio to meet the quality requirements of most users without any further processing of the raw data. But compression is and will continue to be a vital component of almost all visual communication and computing systems and products, because the sheer volume of video data can easily overwhelm the communication bandwidth and in-device storage.[1] The relative lack of advance in sparsity-based restoration of compressed videos is perhaps due to the fact that the compression noises are much more difficult to model than other degradation sources such as blurring and imaging device noises. The non-linearity of quantization operations in image compression systems makes quantization noises signal dependent, far from being white and independent as commonly assumed by works on other video restoration tasks.[2]

Inspired by the work proposed in [1] which describes transform domain based restoration for images, we focus on soft decoding of wavelet domain compressed videos for the reason that the code streams of wavelet-based compression methods have certain amount of residual redundancy in the form of inter-block correlation. Unlike most existing video restoration methods, we design soft decoding algorithm to work directly in the transform domain instead of spatial domain. This is because inverse wavelet transform is required if the restoration is carried out in pixel domain, and it will propagate an isolated quantization error, which is originally confined to a wavelet coefficient, to all pixels of the block being restored.

It is widely known that our perceptual sensitivity to picture distortions varies with luminance level, the existence of textures, and the level of artifacts. This is often described in terms of visual masking and contrast sensitivity characteristics, which have often been utilized in quality metrics such as just noticeable distortion (JND) [3, 4] and visual signal-tonoise ratio (VSNR) [5]. This has provided a basis for many perceptual video compression algorithms including [6–8]. Moreover, visual statistics and features are also employed for developing quality assessment methods [9]. Recently, the near-threshold and supra-threshold perception strategy has been successfully exploited in metrics such as VSNR [5] and MAD [9], where human perception is modeled as two distinct processes.

Finally, Zhang and Bull proposed an artifact-based video metric (AVM) [10] for their analysis-synthesis video coding framework, which is used both in-loop for RQO as well as outside the
coding framework for performance evaluation. Based on the subjective experiment results reported, this performs well for both conventionally compressed and synthesized content.

In order to assess the restoration of the results, we also employ a quality assessment model for perceptual video compression applications. This approach non-linearly combines noticeable distortion and blurring artifacts based on a complex wavelet transform decomposition and motion analysis. PVM and its later version, PIM, provide competitive correlation performance with subjective judgments and offer efficiency and flexibility for in-loop processing [4].

Proposed Algorithm

Sparsity Model

In block-wise wavelet coding, a video is partitioned into non-overlapping blocks for each frame. Wavelet is performed on pixel block independently; coefficients of wavelet transformation are scalar quantized according to a quantization table. Research on image and video statistics reveals that pixel patches can be well approximated by a sparse linear combination of elements from an appropriately chosen dictionary. Using this observation as a prior for soft decoding, we seek a sparse representation of each block of wavelet coefficients. Let \( D = [d_1, d_2, ..., d_k] \) be the dictionary matrix, where each \( d_i \) represents a basis vector in the dictionary. A wavelet patch \( y \) can be represented as a linear combination of atoms in the dictionary \( D \) plus some perturbation \( \varepsilon \), that is, \( y = Da + \varepsilon, a \in \mathbb{R}^{k \times 1} \). The model is sparse if \( \|e\| \ll \|y\| \) and \( \|a\|_0 \ll k \).

Bases for each model are adaptive to the particular homogeneous subset. For this reason, we divide the hard decoded image into a set of overlapped blocks of size 8x8, and perform wavelet on these blocks; the wavelet coefficient vectors constitute a training data set. In the construction of the sparsity dictionary for restoring wavelet coding block \( y_i \), we take advantage of the non-local self-similarity of natural images in learning, and collect similar patches by non-local patch grouping (NLPG) in the training data set. The NLPG procedure guarantees that only the similar sample blocks are used in dictionary learning. [1]

Soft Decoding

As analysed in [1], all blockwise WAVELET-based image/video compression methods suffer from a common problem, that is, sample blocks are encoded independent of each other. Inter-block correlations are totally ignored. This not only reduces the coding efficiency in the first place, but also limits the modeling capability of sparsity-based image prior. The problem is aggravated for low bit rates as vital structural information of the source image is lost or distorted due to the quantization process. One way to solve the problem is to impose structural sparsity constraints in soft decoding. Specifically, we explicitly introduce a regularization term into the following optimization problem to preserve the consistency of sparse codes for similar local patches:

\[
\min_{i \in [n], \langle a, \tilde{a} \rangle} \left\{ \|y_i - Da\| + \lambda \sum_{i \neq j} \|a_i - a_j\| + \gamma \sum_{i \neq j} W_{ij} \right\},
\]

where \( n \) is the number of sample blocks in the video, \( W_{ij} \) measures the similarity between two patches. Here we employ Euclidean distance for simplicity. In addition to the sparsity image prior, the wavelet video code stream contains strong pieces of side information on the original video that should be exploited to improve restoration performance. Finally, we formulate our problem of soft decoding as the following constrained convex optimization problem:
Perception-based Video Metric

Two characteristics of the Human Visual System (HVS) are commonly exploited in objective quality metric development: visual masking and a two-stage perception strategy. Texture masking implies that greater distortion can be tolerated by the HVS in textured regions (static and dynamic) than in plain luminance areas [4]. Two-stage perception theory observes that human perception tends to estimate distortion for high quality videos, while it detects artifacts in low quality cases. These two phenomena provide a basis for the proposed methods.

Here we adopt the architecture of the perception-based video metric (PVM) in our algorithm, which is shown in Figure 1 [4]. This approach is an enhanced version of AVM [8] that combines noticeable distortions and blurring artifacts using a modified geometric mean model [11]. This emulates the distortion-artifact perception process.

Interblock Coding Dependency

Block-based predictive coding is employed in AVS intraframe coding. For each coding block, the neighboring reconstructed pixels are used for intra-prediction. To accommodate various video...
contents and further enhance the prediction performance, 5 different prediction modes are defined for each 8*8 block in AVS profile, including a DC prediction mode and 4 directional prediction modes. As shown in Figure 2. For each prediction mode, the neighboring reconstructed pixels will be copied along the corresponding direction to form the prediction result.[9]

After the intra-prediction, the residue block can be calculated by subtracting the prediction block from the original block in pixel domain. Then, the wavelet transform is performed to obtain the wavelet coefficient block. Let us assume the wavelet coefficients at different frequency positions are independent and follow Gaussian distribution. Based on the classic R-D theory, the total bitrate can be estimated as:

\[ R = \sum_{i=0}^{n} \sum_{j=0}^{m} \log \frac{\omega_{ij}}{D_{ij}} \]

In this paper, an empirical approach is employed to estimate \( \omega_{ij} \). From extensive experiments, we found that \( \omega_{ij} \) can be calculated as:

\[ \omega_{ij} = \alpha_{ij} d + \beta_{ij} \]

where \( \omega_{ij} \) and \( \beta_{ij} \) are frequency position dependent parameters. \( d \) is the average distortion of the neighboring reconstructed pixels. For the wavelet coefficients at each frequency position, \( \omega_{ij} \) increases monotonically with \( d \), which is consistent with our expectation. If \( d \) is large, the neighboring pixels used for intra-prediction tend to have large quantization distortion. This will lead to worse prediction and more residue. For \( \beta_{ij} \), it can be viewed as a measure of the intrinsic difference, the so-called innovation signal, between the original neighboring pixels and current block.[12,13] In other words, even the neighboring pixels are free of quantization distortion, the residue signals are often not zero. Typically the value of the innovation signal depends on the characteristics of the image content itself. It should be noted that the wavelet coefficients at different frequency positions correspond to different values of \( \omega_{ij} \) and \( \beta_{ij} \). (2) is an interesting finding, as it provides us a quantitative, instead of conventionally conceptual, measure of the coding dependency inhabits in the intra-prediction based image coding.[9]

**Termination Strategy**

Errors in the residual information leading to a syntactically valid sequence of codewords are problematic, since the prediction information is interlaced with the residual information.[14] Upon desynchronization, the correction process will force the use of a valid codeword in any modeled syntax element, no matter how unlikely. To address this problem, we use a threshold to stop the correction process when the selected codeword seems too unlikely. Exploiting the fact that most codewords are short, it becomes highly improbable that multiple bits inside the same codeword are erroneous, even at very high bit error rates. Thus, we propose to stop the correction process when the Hamming distance between the likeliest codeword and the received bits is greater than 1.

Furthermore, our implementation uses a greedy approach to find the series of likeliest codewords where a single codeword is retained at each step, and the remaining outcomes are discarded. Type I errors (i.e. changing intact bits) and type II errors (i.e. keeping corrupted bits) may desynchronize the bitstream, leading to an unlikely path requiring that multiple bits be flipped. To avoid following such a path, we stop the correction process when the ratio of flipped bits over interpreted bits exceeds \( 10^{-2} \). Once the decoding process stops, either because the slice was entirely decoded, an error was detected, or a threshold was reached, the extracted macroblocks are reconstructed and the missing macroblocks, if any, are marked for concealment.[14]

**Conclusion**

We proposed a new sparsity-based soft decoding approach for the restoration of compressed videos in wavelet domain. The main contribution of this work is the exploitation of inter block correlations by a
technique of collaborative sparse coding. The experimental results are encouraging, opening up the possibility of significantly improving the quality of wavelet compressed images in a postprocess. The employed PVM and its later version PIM provide competitive performance in contrast to existing video quality assessment approaches. It also offers the benefit of low complexity and latency and is therefore suitable for in-loop RQO. Future research should focus on the RQO application using the proposed metrics. This work is supported in part by the National Key Technology R&D Program projects (2012BAH27F02, 2013BAH27F03).

References


