A NEW CONSTRAINT FOR THE REGULARIZED ENHANCEMENT OF COMPRESSED VIDEO

C. Andrew Segall and Aggelos K. Katsaggelos

Department of Electrical and Computer Engineering
Northwestern University, Evanston, IL 60208, USA
{asegall, aggk}@ece.nwu.edu

ABSTRACT

A novel fidelity constraint to the image enhancement problem is presented. With this constraint, we exploit the motion vectors of a compressed video bit-stream. These vectors establish a correspondence between image pixels across a series of frames, and we guarantee that processing the decoded sequence does not violate this correspondence. We develop the constraint within the context of MPEG-2 and incorporate the constraint into a regularized enhancement algorithm. Simulations are then performed. Quantitative and qualitative results illustrate an improvement in visual quality.

1. INTRODUCTION

Image compression systems introduce a variety of artifacts into the decoded imagery. These artifacts include blocking, ringing and temporal flicker. Removing or attenuating these artifacts can significantly improve the quality of the decoded sequences, and many post-processing algorithms are proposed for this task [2]. In general, these algorithms fall into two distinct categories. Enhancement algorithms rely on a heuristic procedure to filter and improve the quality of compressed video. Recovery techniques depend on a rigorous definition of the enhanced image and utilize an optimization technique.

Here, we are concerned with recovery methods. These algorithms improve visual quality by finding a solution that balances two conflicting requirements. First, desirable characteristics of the decoded images must be present. These properties include deterministic statements such as "The image should be smooth" as well as stochastic definitions for the image distribution. Second, fidelity constraints require a solution that is consistent with any observations of the original signal. For example, we might require individual pixels to be close to the observed data. Alternatively, we could restrict the solution to a valid range of Discrete Cosine Transform (DCT) coefficients. The allowable range is extracted from the compressed bit-stream.

In this paper, we introduce a new fidelity constraint for image recovery techniques. Deterministic methods are the focus, though the constraint is also applicable to stochastic methods. The method relies on the motion vectors within a compressed bit-stream. These vectors define a relationship between blocks of pixels in different image frames, and the goal of our constraint is to maintain these temporal relationships during processing. Image sequences that do not exhibit these temporal correspondences are therefore excluded as solutions to the proposed post-processing algorithm.

The rest of the paper is organized as follows. In section 2, we provide an overview of current video compression methods. These methods motivate the definition of the new constraint, and an example shows its potential impact on a post-processing procedure. In Section 3, a method for realizing the constraint is proposed. Then, the constraint is incorporated into a regularized post-processing algorithm. Finally, simulations are presented in section 4. Visual quality metrics are used in evaluating the results.

2. MOTIVATION

Standards-based video compression systems rely on a variety of video compression algorithms. These algorithms are specified in the ITU and MPEG family of standards, and each utilizes a similar approach for video compression. In general, images are encoded as either intra- or inter-frames. In intra-mode, an image is divided into equally sized blocks. Blocks are then independently processed with the DCT, and the resulting transform coefficients are quantized. The quantized data is sent to the decoder, which computes the inverse-DCT of quantized coefficients and reassembles the image blocks into a frame.

Inter-frame compression exploits the temporal redundancies of a video sequence and is a second mode for compression. In this method, images are still divided into equally sized blocks and independently processed. However, processing begins by first finding the best estimate for each block in the previously transmitted images. Once the best match is found, its position is encoded relative to the current block and communicated to the decoder. The difference between the current block and the prediction is then calculated, transformed with the DCT, quantized and transmitted to the decoder. Finally, the current block is reconstructed by taking the inverse-DCT of the quantized information and adding it to the block referenced by the motion vectors. The image blocks are then reassembled into a complete frame.

Thinking of the compressed bit-stream in terms of fidelity constraints, we see that there are two key pieces of information that describe the original image during intra-frame compression. First, the original pixel data is approximated by the pixels of the decoded result. Thus, we are justified in saying that the original image is "close" to the decoded information, though a definition of this similarity must be provided. As a second constraint, the quantized values for the DCT coefficients convey information

This work was supported in part by a grant from the Motorola Center for Communications.

about the original image. These coefficients are transmitted to the decoder by a quantization index and scale factor. The set of coefficients that contains the original image then has a width equal to the quantization scale factor and is centered on the value of the quantizer index multiplied by the scale factor.

The intra-mode constraints are often utilized in post-processing algorithms [4,7]. In both cases, the goal is to restrict any modifications to the decoded image data so that the constraints are not violated. (Violating the constraints guarantees a solution unequal to the original image.) In inter-mode encoding, an additional description of the original image appears in the compressed bit-stream. This suggests another constraint on the solution. In this mode, motion vectors identify the best match between each block in the original image and previously encoded data. With knowledge of this relationship, it is reasonable to constrain a post-processed solution to be "close" to these estimated values.

Satisfying the proposed *motion vector constraint* requires that a post-processing algorithm define a measure of similarity between a solution and the predicted blocks. Within our framework, any definition is allowable. However, the constraint should utilize the same definition as enforced by the encoder. In typical ITU or MPEG compression systems, the purpose of the motion vectors is to reduce the bit-rate requirement for video transmission. While many different measures are conceivable, most procedures rely on the *sum of the absolute difference* (SAD) as the similarity test. This definition is well-suited for video coding, as it reduces the magnitude of the error residual and simplifies coding.

Independent of the similarity measure, the motion vector constraint requires that motion vectors calculated for the post-processed image correspond to the information transmitted in the bit-stream. Using the SAD metric as an example, we should then require that the minimum SAD between each post-processed block corresponds to the block of pixels referenced by the motion vector, when compared to all other pixel locations within a search region. This is stated as

$$\mathrm{P}_{\mathrm{MV}}\left(\mathbf{b}_i\right) \in \left\{ \mathbf{b}_i : \left| \mathbf{b}_i - \hat{\mathbf{b}}_{i,i} \right| \leq \left| \mathbf{b}_i - \hat{\mathbf{b}}_{i,j} \right|, \ \forall j \neq i \right\}, \tag{1}$$

where P_{MV} is the motion vector constraint operator, $\mathbf{b}_i$ is the i^{th} block in the current frame, $\hat{\mathbf{b}}_{i,i}$ is the estimate for the i^{th} block as denoted by the motion vector and $\hat{\mathbf{b}}_{i,j}$ are all of the blocks that were considered by the encoder as best matches.

Defining the blocks that were not chosen as the best match is an important component of the constraint, as it implicitly determines the amount of similarity between the original image and the estimate. In the standards, the specific region and pattern of the motion vector search is undefined. Instead, a maximum search range is identified as well as the precision of the motion vectors. Many applications do not have the resources for an exhaustive search over all potential matches. Instead, a reduced search or imperfect strategy must be considered. Nevertheless, rejected candidates express additional knowledge of the original image and should benefit the post-processing procedure.

The maximum number of search locations is defined in all of the standards, and we utilize this information to illustrate the potential influence of the motion vector constraint. Consider an interlaced compression system based on MPEG-2 [1]. In this scenario, motion vectors are calculated for each 16x8 block of the current frame and are encoded at half-pixel resolution. The maximum extent of the search is ±64 pixels, and two modes of inter-frame compression are available. When encoding using the first mode, a P-field is generated and relies on two previous fields as temporal references. A full search of the candidate motion vectors produces 107,519 rejected search locations. More constraints are introduced when encoding a B-field. In this situation, two motion vectors are transmitted for each 16x8 block and averaging the two references is an acceptable way to form the estimate. In the case of the maximum search window, there are 1,156,055,038 rejected search locations. This effectively provides $1\mathrm{x}10^9$ examples of what each block in the original image does not look like, which is a significant amount of information about the original image.

3. PROPOSED ALGORITHM

To realize the motion vector constraint, we propose the following algorithm

$$\mathrm{P}_{\mathrm{MV}}\left(\mathbf{b}_i\right) = \mathbf{b}_i + \alpha\left(\hat{\mathbf{b}}_{i,i} - \mathbf{b}_i\right), \tag{2}$$

where α is the minimum positive value that makes the statement

$$\mathrm{M}_{Similarity}\left(\mathbf{b}_i + \alpha\left(\hat{\mathbf{b}}_{i,i} - \mathbf{b}_i\right) - \hat{\mathbf{b}}_{i,i}\right) \leq \mathrm{M}_{Similarity}\left(\mathbf{b}_i + \alpha\left(\hat{\mathbf{b}}_{i,i} - \mathbf{b}_i\right) - \hat{\mathbf{b}}_{i,j}\right) \tag{3}$$

true. $\mathrm{M}_{Similarity}$ is the similarity measuring operator. In the case of the SAD test for similarity, we would restate (3) as

$$\left| \mathbf{b}_i + \alpha\left(\hat{\mathbf{b}}_{i,i} - \mathbf{b}_i\right) - \hat{\mathbf{b}}_{i,i} \right| \leq \left| \mathbf{b}_i + \alpha\left(\hat{\mathbf{b}}_{i,i} - \mathbf{b}_i\right) - \hat{\mathbf{b}}_{i,j} \right|. \tag{4}$$

With the procedure, a post-processing algorithm takes each image block and moves it closer to its predicted value. Blocks are only modified if the similarity between the predicted block and the current solution is greater than the similarity between a rejected block and the current solution. Under this condition, the solution block is modified until the constraint is satisfied. A visual example of the procedure is shown in Figure 1.

Incorporating the motion vector constraint into a post-processing algorithm leads to a novel procedure. Consider the deterministic framework presented in [3, 5]. In this work, the post-processing algorithm is posed as the minimization of

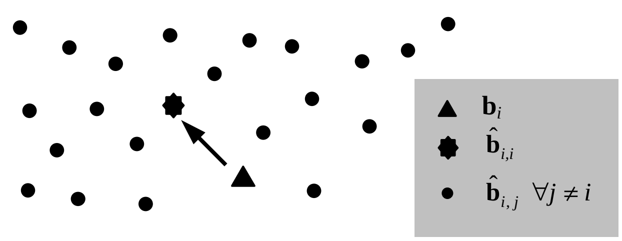


Figure 1. Simplified representation of the motion vector constraint. The current block is moved towards the estimated block until the sum of absolute difference criterion is satisfied.

$$J(\mathbf{f}) = \mu_1 \| \mathbf{g} - \mathbf{f} \|^2 + \lambda_1 \| \mathbf{Q_1 f} \|^2 + \lambda_2 \| \mathbf{W_2 Q_2 f} \|^2 + \lambda_3 \| \mathbf{W_3 Q_3 f} \|^2 + \lambda_4 \| \mathbf{f} - \mathbf{f_{MC}} \|^2$$

$$s.t.\ \mathrm{P_{DCT}}(\mathbf{f}) \in DCT_{Transmitted} \qquad (5)$$

where **g** is a decompressed image, **f** is the sought after enhanced image, $\mathbf{Q_1}$ is a high-pass operator that measures smoothness across the entire image, $\mathbf{Q_2}$ is a horizontally oriented high-pass operator, $\mathbf{W_2}$ restricts the influence of the operator to the horizontal boundaries of the blocks, $\mathbf{Q_3}$ is a vertically oriented high-pass operator, $\mathbf{W_3}$ restricts the influence of the operator to the vertical boundaries of the blocks, $\mathbf{f_{MC}}$ are previously enhanced images projected through a motion field to the current frame location, $DCT_{Transmitted}$ represents the allowable range of the transmitted DCT coefficients and $\mathrm{P_{DCT}}$ is the projection operator onto this range. The importance of the fidelity term $\| \mathbf{g} - \mathbf{f} \|^2$ is controlled by the parameter μ_1 and is traditionally equal to one. The relative influence of the within image, between block and inter-frame smoothness measures are controlled by λ_1, λ_2, λ_3 and λ_4, respectively.

Incorporating the motion vector constraint and utilizing the method of successive approximations to minimize (5), an iterative solution evolves. The procedure becomes

$$\begin{aligned} \mathbf{f_{k+1}} = \mathrm{P_{MV}}\,\mathrm{P_{DCT}}\Big(\mathbf{f_k} - \gamma \big\{ & \mathbf{f_k} - \mathbf{g} + \lambda_1 \mathbf{Q_1^T Q_1 f_k} \\ & + \lambda_2 \mathbf{Q_2^T W_2^T W_2 Q_2 f_k} \\ & + \lambda_3 \mathbf{Q_3^T W_3^T W_3 Q_3 f_k} \\ & + \lambda_4 (\mathbf{f_k} - \mathbf{f_{MC}}) \big\}\Big) \end{aligned} \qquad (6)$$

where $\mathrm{P_{MV}}$ is the motion vector constraint in (2) and γ determines the convergence and the rate of convergence of the algorithm.

4. EXPERIMENTS

Incorporating constraints into a post-processing algorithm is only worthwhile if it improves the visual quality of the decoded sequence. In this section, we present experimental results utilizing the proposed motion vector constraint. Results consist of visual quality measurements and visual examples. For the simulations, we process 100 frames of an MPEG-2 video sequence. The sequence consists of interlaced CCIR601 video data with a spatial resolution of 720x240 pixels and a frame rate of 60 fields per second. The bit-stream is generated with the TMN5 rate controller operating at MAIN profile and MAIN level and a target bit-rate of 4.25Mbps. When finding the motion vectors, the encoder searches an 8x8 region for a P-field and a 3x3 region for a B-field. In both instances, results utilize the same parameters when defining the rejected matches for the motion vector constraint.

Parameters for the algorithm are chosen to provide a reasonable post-processing method. At the same time, we present an algorithm that relies heavily on the motion vector constraint. Towards this goal, experiments utilize the method of (6) but disable any smoothing between temporal frames by setting λ_4=0. Thus, each frame is processed independently. Also, we minimize the impact of the linear constraints with the parameter choice λ_1=0, λ_2=λ_3=1 and μ_1=0. This results in an algorithm that attempts to smooth across block boundaries, with no penalty for deviating from the decoded intensities. The DCT constraint is realized by calculating the DCT at each iteration, projecting the invalid coefficients to the nearest valid value and calculating the inverse DCT. The entire algorithm terminates when $\|\mathbf{f_{k+1}} - \mathbf{f_k}\|^2 / \|\mathbf{f_k}\|^2 \leq 10^{-6}$.

To quantify the improvement in visual quality, we utilize the metric discussed in [6]. Effectively, this metric incorporates a model of the human visual system into a weighted error calculation. These weights are dependent on several characteristics of human perception, including a variety of luminance and spatial masking properties. With this model, visibility thresholds for each DCT coefficient of a block are calculated. These thresholds define the errors required for the perception of a visual artifact. Then, the metric is computed by finding the quantization error for each coefficient and dividing by the corresponding threshold. Within this framework, measurements less than one denote an invisible error. Values greater than one identify visible artifacts.

Visual quality measurements for each of the 200 fields are provided in Figure 2. Care must be taken in interpreting the values, as the scale only provides relative measurements. However, it is reasonable to expect images with higher values for the measurement to contain more visual errors. In the figure, (a) displays the visual quality of the decoded image sequence (without processing) for each field, while (b) plots the visual quality of the post-processing image sequence for each field. The difference between these two sequences is shown in (c), where positive numbers represent an improvement in visual quality after post-processing. From the figure, we see that the post-processing algorithm improves the visual quality metric of every frame of the sequence. The average improvement is 4.04, which is expressed in terms of a spatially pooled, just-noticeable difference. The reduction in the visual quality metric suggests that the algorithm improves visual quality.

Perhaps the most important evidence in evaluating visual quality is the inspection of actual images. Two small portions of a CCIR field are shown in Figure 3. The decoded images (without post-processing) are displayed in (a) and (c), while the post-processed results are shown in (b) and (d). From the figure, we see that the post-processing algorithm is able to reduce the ringing and blocking artifacts appearing around the statue in the top set of images. Also, the blocking artifact in the second set of images is reduced. Perhaps a more interesting observation is that de-blocking and de-ringing is accomplished without blurring significant features. Notice the preservation of the monument feature at the lower left of the top set of images. Also, observe the strong vertical feature at the right of the bottom set of images. The motion vector constraint facilitates this adaptive smoothing, as the significant features appear in previous images and are maintained by the constraint.

5. CONCLUSIONS

In this paper, we introduce a new fidelity constraint for regularized video enhancement. The constraint relies on the motion vectors in the compressed bit-stream and guarantees that relationships defined by the motion vectors are maintained during processing. The motion vector constraint is incorporated into a deterministic post-processing algorithm, and the novel technique is applied to an MPEG-2 coding scenario. Quality is measured with a visual quality metric and visual examples.

6. REFERENCES

[1] ISO/IEC JTC1/SC29 International Standard 13818-2, *Information technology – Generic coding of moving pictures and associated audio information: Video,* 1995.

[2] C.A. Segall and A.K. Katsaggelos, "Pre- and Post-Processing Algorithms for Compressed Video Enhancement," *Proceedings of the Thirty-Fourth Asilomar Conference on Signals, Systems and Computer*, Oct. 29-Nov. 1, 2000.

[3] C.A. Segall and A.K. Katsaggelos, "Enhancement of Compressed Video using Visual Quality Metrics," *Proceedings of the IEEE International Conference on Image Processing*, Vancouver, BC, Canada, Sept. 10-13, 2000.

[4] *Signal Recovery Techniques for Image and Video Compression and Transmission*, A. K. Katsaggelos and N. P. Galatsanos, editors, Kluwer Academic Publishers, 1998.

[5] C.-J. Tsai, P. Karunaratne, N.P. Galatsanos and A.K. Katsaggelos, "A Compressed Video Enhancement Algorithm," *Proceedings. of the International Conference on Image Processing,* Kobe, Japan, Oct. 25-28, 1999.

[6] A.B.Watson, "Image Data Compression having Minimum Perceptual Error," U.S. Patent 5,426,512, 1995.

[7] Y. Yang, N.P. Galatsanos and A.K. Katsaggelos, "Regularized Reconstruction to Reduce Blocking Artifacts of Block Discrete Cosine Transform Compressed Images," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.3, no.6, pp.421-432, Dec. 1993.

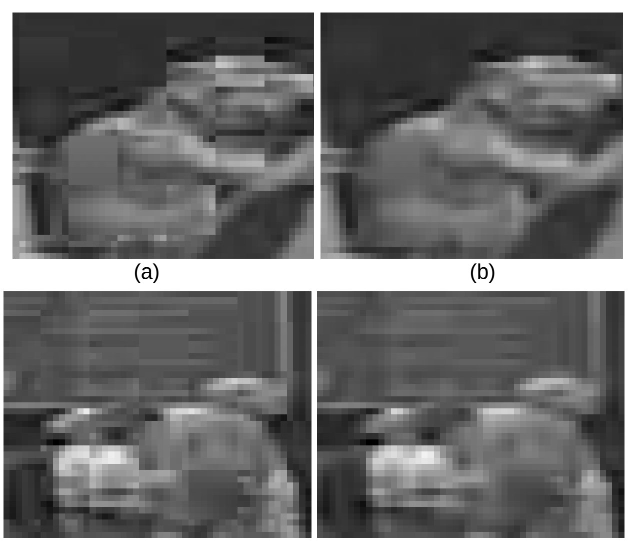
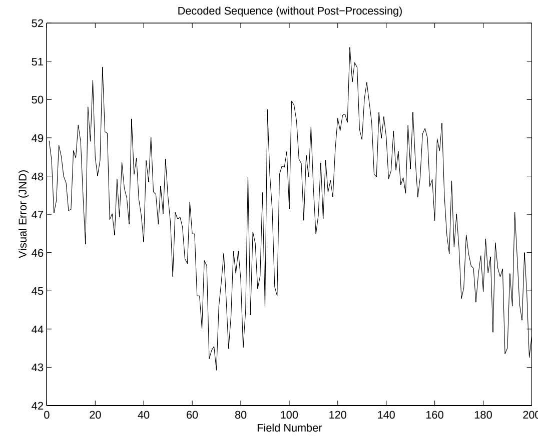
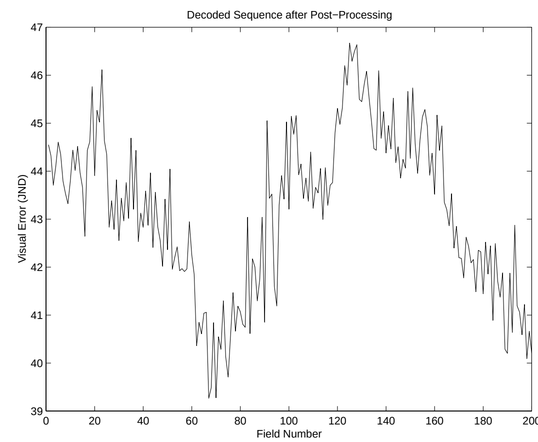


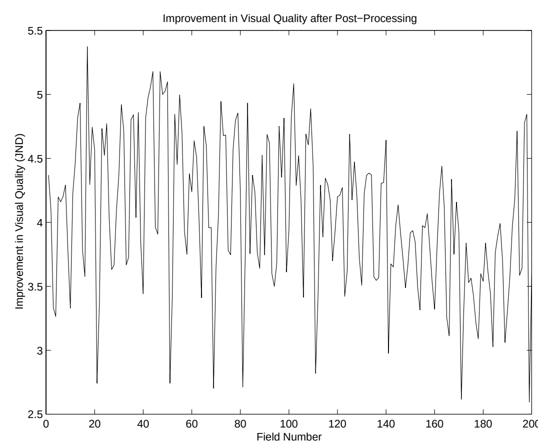
Figure 3. Two examples of the motion vector constraint: (a) and (c) are the decoded images without processing; (b) and (d) are the images after post-processing. The motion vector constraint facilitates removal of blocking and ringing artifacts but preserves significant image content.



(a)



(b)



(c)

Figure 2. Visual quality measurements versus field number. Visual quality is a pooled just-noticeable difference calculation and suggests that the motion vector constraint improves visual quality: (a) quality of decoded sequence; (b) quality of sequence after post-processing with (6), and (c) improvement in visual quality resulting from the motion vector constraint. In (a) and (b), higher values for the visual quality measurement reflect more severe degradations. In (c), positive values for the metric suggests that the post-processing algorithm improves the visual results.