Coding Artifacts Robust Resolution Up-conversion

Ling Shao, Hao Hu, Gerard de Haan

Philips Research Europe, High Tech Campus 36, Eindhoven, The Netherlands

ABSTRACT

In this paper, an integrated resolution up-conversion and compression artifacts removal algorithm is proposed. Local image patterns are classified into object details or coding artifacts using the combination of structure information and activity measure. For each pattern class, the weighting coefficients for up-scaling and artifact reduction are optimized by a Least Mean Square (LMS) training technique, which trains on the combination of the original images and the compressed down-sampled versions of the original images. The proposed combined algorithm is proven to be more effective than previous classification based techniques in concatenation.

Index Terms— Image up-scaling, Compression artifacts removal, Trained filter, Classification, Adaptive dynamic range coding, Standard deviation

1. INTRODUCTION

High definition television (HDTV) is becoming the standard appliance of each modern household. The resolution of HDTV is usually higher than that of legacy videos, which are still ubiquitous in broadcasting programs or recorded media. Those low-resolution video materials have to be up-converted to fit the resolution of HDTV. Moreover, video materials are always compressed with various compression standards, such as MPEG-4 and H.264. These block transform based codecs divide the image or video frame into non-overlapping blocks (usually with the size of 8 x 8 pixels), and apply discrete cosine transform (DCT) on them. The DCT coefficients of neighboring blocks are thus quantized independently. At high or medium compression rates, the coarse quantization will result in various noticeable coding artifacts, such as blocking, ringing and mosquito artifacts.

Linear resolution up-scaling techniques, such as bi-linear and bi-cubic interpolations, usually blurred image details. Advanced resolution up-conversion algorithms [1-6] are designed to be adaptive to local structure or edge orientation, which makes them capable of preserving edges and fine details in the image content. Zhao et al. [7] compared the state-of-the-art up-scaling techniques both objectively and subjectively, and concluded that the structure-adaptive LMS training technique, proposed by Kondo et al. [1], performs the best. The training algorithm can preserve structures and fine textures perfectly, when the image is clean and noise free. However, when the image is compressed, coding artifacts will be preserved and enlarged after up-scaling. These coding artifacts, e.g. blocking artifacts, will be even more difficult to remove than those in the original low-resolution image, because the coding artifacts will spread among more pixels and become not trivial to detect after up-scaling. In the video chain, coding artifacts are usually suppressed using certain low-pass filtering techniques before applying resolution up-scaling. However, most coding artifact reduction algorithms blur image details while removing various digital artifacts. Those details lost during artifact reduction cannot be recovered during resolution up-scaling. We propose a combined artifacts reduction and resolution up-scaling approach in this paper. Optimized filter coefficients are used for different image regions based on a classification method that utilizes both structure and activity information. The optimal coefficients are obtained by making the mean square error (MSE) between the reference pixels and the processed degraded pixels minimized statistically during the training process. The degradation we use here is first down-sampling then adding coding artifacts by compression.

Section 2 describes the classification method that attempts to classify a local region into various object details or coding artifacts. In Section 3, we present the LMS training process to obtain the optimized coefficients for each class. Experimental results and performance evaluation are given in Section 4. Finally, we conclude our paper in Section 5.

2. PIXEL CLASSIFICATION

Adaptive Dynamic Range Coding (ADRC) [8] has been successfully used for representing the structure of a region. The ADRC code of each pixel x_i in an observation aperture is defined as: $ADRC(x_i)=0$, *if* $V(x_i)<V_{av}$; 1, *otherwise*, where $V(x_i)$ is the value of pixel x_i, and V_{av} is the average of all the pixel values in the aperture. ADRC has been demonstrated to be an efficient classification technique for resolution up-conversion [1]. However, it obviously is not enough for compressed materials, because it can not distinguish object details from coding artifacts.

For example, the ADRC codes of an object edge could be exactly the same as that of a blocking artifact. Hao and de Haan [9] proposed to use dynamic range to further differentiate coding artifacts from object details. Dynamic range is simply the absolute difference of the maximum and minimum pixel values of a region. Since dynamic range is not robust to shot noise, we propose to use local activity measure to be appended to ADRC. The activity measure we employ here is standard deviation (*STD*), because the *STD* of object details tends to higher than that of coding artifacts. The STD of a region is defined as follows:

$$STD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(v_i - \bar{v})^2} \qquad (1)$$

where v_i indicates the pixel value of the *i*th pixel, $\bar{v}$ is the average of the pixel values, and *N* is the pixel number of the region around the central pixel over which the *STD* is calculated. A region with high activity has a high *STD* value, while the *STD* value of a region with low activity usually only contains coding artifacts.

Accordingly, a pixel and its surrounding region can be classified based on the structure, which is represented by ADRC, and the activity measure, which we use a simple standard deviation.

3. LEAST MEAN SQUARE OPTIMIZATION

In this section, the Least Mean Square (LMS) optimization technique is described to produce optimal coefficients for each class based on the pixel classification of the previous section. Fig. 1 shows the proposed optimization process. Uncompressed HD reference images are first down-sampled using bi-linear interpolation. The down-sampled images are then compressed to introduce coding artifacts. We refer the compressed down-sampled images as corrupted images. Each pixel in the corrupted images is then classified on that pixel's neighborhood using the classification method described in the previous section. All the pixels and their neighborhoods belonging to a specific class and their corresponding pixels in the reference images are accumulated, and the optimal coefficients are obtained by making the Mean Square Error (MSE) minimized statistically.

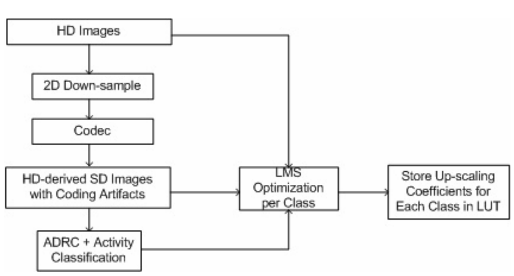

Fig. 1: The training process of the proposed method.

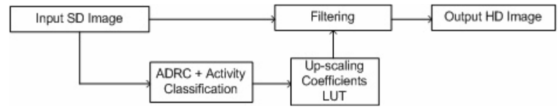

Fig. 2: The filtering process of the proposed method.

Let $F_{D,c}$, $F_{R,c}$ be the apertures of the distorted images and the reference images for a particular class *c*, respectively. Then the filtered pixel $F_{F,c}$ can be obtained by the desired optimal coefficients as follows:

$$F_{F,c} = \sum_{i=1}^{n} w_c(i) F_{D,c}(i,j) \qquad (2)$$

where $w_c(i), i \in [1...n]$ are the desired coefficients, and *n* is the number of pixels in the aperture.

The summed square error between the filtered pixels and the reference pixels is:

$$\begin{aligned} e^2 &= \sum_{j=1}^{N_c} (F_{R,c} - F_{F,c})^2 \\ &= \sum_{j=1}^{N_c} [F_{R,c}(j) - \sum_{i=1}^{n} w_c(i) F_{D,c}(i,j)]^2 \end{aligned} \qquad (3)$$

where N_c represents the number of pixels belonging to class *c*. To minimize e^2, the first derivative of e^2 to $w_c(k), k \in [1...n]$ should be equal to zero.

$$\frac{\partial e^2}{\partial w_c(k)} = \sum_{j=1}^{N_c} 2F_{D,c}(k,j)[F_{R,c}(j) - \sum_{i=1}^{n} w_c(i) F_{D,c}(i,j)] = 0 \qquad (4)$$

By solving the above equation using Gaussian elimination, we will get the optimal coefficients as follows:

$$W = X^{-1}Y \qquad (5)$$

where

$$W = \left[w_c(1), w_c(2), \cdots, w_c(n)\right]^T$$

$$X = \begin{bmatrix} \sum_{j=1}^{N_c} F_{D,c}(1,j)F_{D,c}(1,j) & \cdots & \sum_{j=1}^{N_c} F_{D,c}(1,j)F_{D,c}(n,j) \\ \sum_{j=1}^{N_c} F_{D,c}(2,j)F_{D,c}(1,j) & \cdots & \sum_{j=1}^{N_c} F_{D,c}(2,j)F_{D,c}(n,j) \\ \cdots & \cdots & \cdots \\ \sum_{j=1}^{N_c} F_{D,c}(n,j)F_{D,c}(1,j) & \cdots & \sum_{j=1}^{N_c} F_{D,c}(n,j)F_{D,c}(n,j) \end{bmatrix}$$

$$Y = \left[\sum_{j=1}^{N_c} F_{D,c}(1,j)F_{R,c}(j), \sum_{j=1}^{N_c} F_{D,c}(2,j)F_{R,c}(j), \cdots, \sum_{j=1}^{N_c} F_{D,c}(n,j)F_{R,c}(j)\right]^T$$

The LMS optimized coefficients for each class are then stored in a look-up table (LUT) for future use. Fig. 2 shows the filtering process of resolution up-conversion for compressed materials using the optimized coefficients retrieved from the LUT.

4. EXPERIMENTS AND EVALUATION

In this section, the experimental results of the proposed algorithm are presented. For the optimization process, a set of 500 images with HD resolution is used for training. We demonstrate the algorithm with the up-scaling factor of 2 both horizontally and vertically. Therefore, bi-linear interpolation with the scaling factor of 2 both horizontally and vertically is used for down-sampling during training. Obviously, other up-conversion factors can also be achieved. Since we do not take temporal information into account, JPEG is adopted for introducing coding artifacts. The quality factor of JPEG is set to be 20. Obviously, other codecs, such as MPEG or H.264, can also be used, but I-frames, P-frames and B-frames have to be treated separately. A region of 3x3 pixels, as depicted in Fig. 3, is used for classification in our implementation. Therefore, 8 bits are needed for ADRC coding, since 1 bit can be saved by bit-inversion [10]. For the activity measure, we use 2 bits for standard deviation. Totally, 10 bits are used for classification.

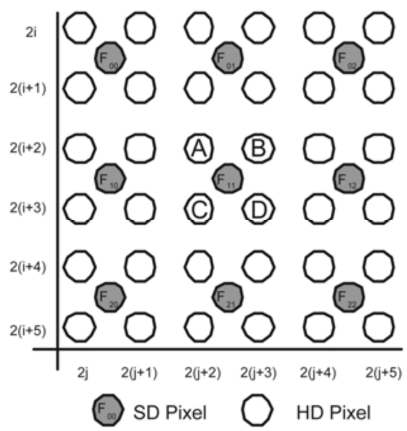

Fig. 3: Aperture used in the proposed method. The white pixels are interpolated HD pixels (F_{HI}). The black pixels are SD pixels (F_{SD}), with F_{12} a shorthand notation for $F_{SD}(1,2)$, etc. The HD pixel A that corresponds to $F_{HI}(2(i+2),2(j+2))$, is interpolated using nine SD pixels (F_{00} up to F_{22}). (Courtesy of M. Zhao, Philips Research Eindhoven.)

Table 1: The numbers of coefficients stored in the LUT of the three algorithms.

Algorithm	Ref [10]+Ref [1]	Ref [1]+Ref [10]	Proposed
# coefficients	4096x16x13 + 256x9	256x9 + 4096x16x13	256x4x9

Table 2: MSE scores of different algorithms.

Sequence	Ref [1]	Ref [10] + Ref [1]	Ref [1] + Ref [10]	Proposed
Hotel	116.3	113.4	108.5	104.1
Parrot	36.1	32.2	35.1	31.3
Teeny	66.9	59.9	63.7	58.7
Bicycle	183.5	164.3	170.2	159.8

Table 3: Comparison of the combined methods.

Sequence	ADRC+Position	ADRC+STD
Hotel	107.7	104.1
Parrot	32.6	31.3
Teeny	60.0	58.7
Bicycle	165.8	159.8

For evaluation, the proposed algorithm is benchmarked against two classification based resolution up-conversion [1] and artifact reduction [10] methods in concatenation. Same as our proposed approach, an ADRC code of a 3x3 aperture is used for classification in up-scaling. The classification method used in coding artifact reduction is the combination of structure by Adaptive Dynamic Range Coding (ADRC) and relative position of a pixel in the coding block grid. A diamond shape 13 pixel aperture is used, which requires 12 bits for ADRC and 4 bits for relative position coding. The drawback of this method is that block grid positions are not always available, especially for scaled material. For the cascaded method of first applying resolution up-conversion then doing coding artifact reduction, the classification of coding artifact reduction is carried out on the up-scaled HD signal and the relative position of a pixel in the block grid is also up-scaled accordingly to suit the HD signal. The coefficients of both methods are obtained by the LMS technique. These two methods have significant advantages over other heuristically designed filtering techniques. For cost comparison, Table 1 shows the numbers of coefficients that need to be stored in look-up tables (LUT) for each of the three algorithms. The proposed algorithm is much more economical than the other two in terms of LUT size. Since the training process is done offline and only needs to be done once, thus the computational cost is limited for all the three methods.

We test the algorithms on a variety of sequences first down-sampled then compressed using the same setting as during the training. Fig. 4 shows the snapshots of the sequences we use. All the test sequences are excluded from the training set. The objective metric we use is mean square error (MSE), i.e. we calculate the MSE between the original HD sequences and the result sequences processed on the compressed down-sampled versions of the original sequences. Table 2 shows the results of the proposed algorithm in comparison to the results of first applying coding artifact reduction then up-conversion and first applying up-conversion then artifact reduction. The result of resolution up-conversion using the method in [1] without applying artifact reduction is also shown for reference.

From the results, one can see that the proposed algorithm outperforms the other two concatenated methods for all sequences. The results also reveal that the order of applying up-conversion and artifact reduction affects the performance of the concatenated method. For some sequences, applying artifact reduction first gives better results; for other sequences, vice verse. For testing the effectiveness of the proposed classification method, the results of the combined algorithm with the classification of ADRC and the relative position of the pixel in the block grid are shown in Table 3. We can see that ADRC plus standard deviation performs better for all sequences.

Fig. 4: Snapshots of test sequences for experiments.

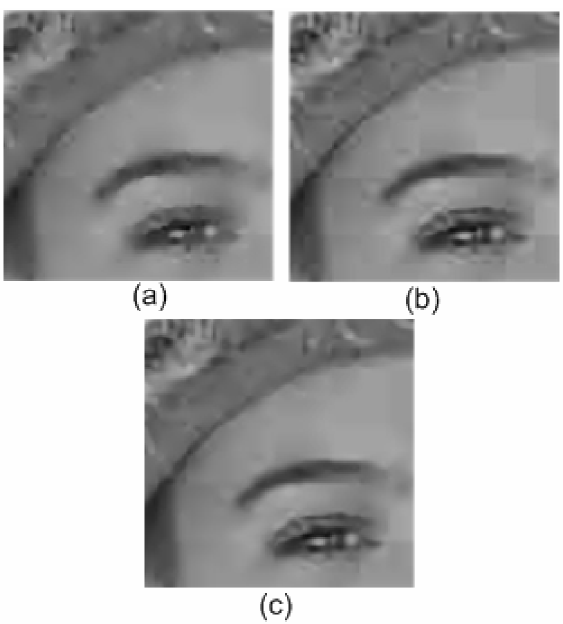

Fig. 5: The results of the Teeny sequence processed using the three methods. (a) First artifact reduction then resolution up-conversion; (b) First resolution up-conversion then artifact reduction; (c) The proposed method.

For subjective comparison, Fig. 5 shows the results of the three methods on the Girl sequence. It is easy to see that the result of first applying up-conversion then artifact reduction contains more residual artifacts than the proposed algorithm, because up-scaling makes coding artifacts spread out in more pixels and more difficult to remove. The result of first applying artifact reduction then resolution up-conversion is blurrier than our proposed algorithm, because the artifact reduction step blurs some details, which cannot be recovered by the up-scaling step.

5. CONCLUSION

In this paper, an integrated coding artifact reduction and resolution up-conversion approach is proposed. Structure information based on ADRC and activity measure based on standard deviation are employed to classify an image region into object details or coding artifacts. Based on the classification, a least mean square optimization technique is used to obtain the optimized weighting coefficients. The optimization is done using a training set composing of the original HD images and the compressed down-sampled versions of the original images. The experimental results are compared to two classification based artifact reduction and resolution up-conversion algorithms in concatenation. Our proposed approach outperforms the other two both objectively and subjectively.

REFERENCES:

[1] T. Kondo *et al.*, “Picture conversion apparatus, picture conversion method, learning apparatus and learning method”, US-patent 6,323,905, Nov. 2001.

[2] C. B. Atkins *et al.*, “Optimal image scaling using pixel classification”, *Proc. ICIP 2001*, Vol. 3, pp. 864-867, 2001.

[3] X. Li and M. T. Orchard, “New edge-directed interpolation”. *IEEE Trans. Image Process.*, Vol. 10, No 10, Oct. 2001, pp. 1521-1527.

[4] J. Tegenbosch *et al.*, “Improving nonlinear up-scaling by adapting to the local edge orientation”, *Proc. SPIE*, Vol. 5308, pp. 1181-1190, Jan. 2004.

[5] N. Plaziac, “Image interpolation using neural networks”, *IEEE Trans. Image Process.*, Vol. 8, No. 11, pp. 1647-1651, Nov. 1999.

[6] R. G. Keys, “Cubic convolution interpolation for digital image processing”, *IEEE Trans. Acous., Speech and Signal Process.*, Vol. ASSP-29, No. 6, pp. 1153-1160, Dec. 1981.

[7] M. Zhao et al., “Making the best of legacy video on modern displays”, *Journal of the Society for Information Display*, Vol. 15(1), pp. 49-60.

[8] T. Kondo *et al.*, “Method and apparatus for adaptive filter tap selection according to a class”, US-Patent: US 6, 192, 161 B1, Feb. 20, 2001.

[9] H. Hu and G. De Haan, “Simultaneous coding artifact reduction and sharpness enhancement”, Proc. *IEEE International Conference on Consumer Electronics*, Jan. 2007.

[10] M. Zhao, *et al.*, “Content adaptive image de-blocking”, *Proc. ISCE*, Sep. 2004.