Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2007, Article ID 71432, 6 pages
doi:10.1155/2007/71432

Research Article

Adaptive Resolution Upconversion for Compressed Video Using Pixel Classification

Ling Shao

Video Processing and Analysis Group, Philips Research Laboratories, High Tech Campus 36, 5656 AE Eindhoven, The Netherlands

Received 22 August 2006; Accepted 3 May 2007

Recommended by Richard R. Schultz

A novel adaptive resolution upconversion algorithm that is robust to compression artifacts is proposed. This method is based on classification of local image patterns using both structure information and activity measure to explicitly distinguish pixels into content or coding artifacts. The structure information is represented by adaptive dynamic-range coding and the activity measure is the combination of local entropy and dynamic range. For each pattern class, the weighting coefficients of upscaling are optimized by a least-mean-square (LMS) training technique, which trains on the combination of the original images and the compressed downsampled versions of the original images. Experimental results show that our proposed upconversion approach outperforms other classification-based upconversion and artifact reduction techniques in concatenation.

1. INTRODUCTION

With the continuous demand of higher picture quality, the resolution of high-end TV products is rapidly increasing. The resolution of broadcasting programs or video on storage discs is usually lower than that of high-definition (HD) TV. Therefore, those video materials have to be upconverted to fit the resolution of the HDTV. Due to the bandwidth limit of the broadcasting channels and the capacity limit of the storage media, the video materials are always compressed with various compression standards, such as MPEG1/2/4 and H.26x. These block-transform-based codecs divide the image or video frame into nonoverlapping blocks (usually with the size of 8×8 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.

Most existing resolution upconversion algorithms apply content-adaptive interpolation according to the structure or property of a region [1–7]. For compressed materials, the coding artifacts will be preserved after upscaling. These coding artifacts, for example, 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 upscaling. One solution is to reduce the coding artifacts before applying resolution upscaling. However, most coding artifact reduction algorithms [8–11] blur details while suppressing various digital artifacts. Those details lost during artifact reduction cannot be recovered during resolution upscaling. We propose to remove coding artifacts and apply resolution upconversion simultaneously in this paper. Different filter coefficients are used for different image regions based on a classification scheme that utilizes both structure and an activity metric. The optimal coefficients are obtained by making the mean square error (MSE) between the reference pixels and the processed distorted pixels minimized statistically during the training process. The distortion we use here is first downsampling then adding coding artifacts by compression.

Most superresolution algorithms [12, 13] in the literature attempt to recover high-resolution images from low-resolution images based on multiframe processing. We propose a single-frame processing solution for resolution upconversion of compressed images and video. Therefore, the proposed technique is more efficient and cost-effective.

The rest of this paper is organized as follows. Section 2 describes the classification method that determines whether a local region contains information or digital artifacts.

Table 1: Coarse classification of a region.

DR	Entropy (high)	Entropy (low)
DR (high)	Object edge/highly textured region	Strong blockiness
DR (low)	Fine texture	Mild blockiness/mosquito noise

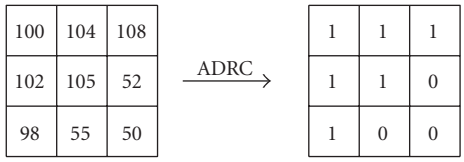

Figure 1: Illustration of ADRC coding.

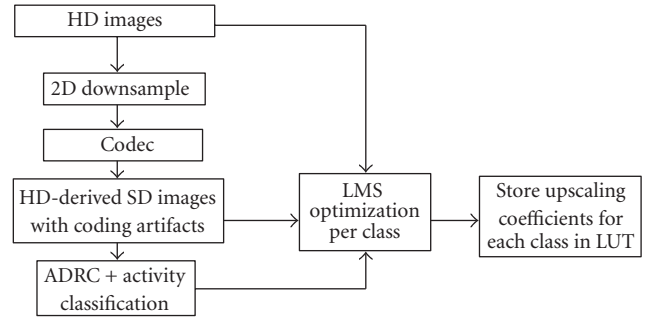

Figure 2: The training procedure of the proposed method.

In Section 3, we present the least-mean-square technique 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) [14] 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 $\text{ADRC}(x_i) = 0$ if $V(x_i) \leq V_{\text{av}}$, 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. Figure 1 shows the ADRC coding of a 3×3 aperture. ADRC has been demonstrated to be an efficient classification technique for resolution upconversion [1]. However, obviously it is not enough for compressed materials, because it cannot 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. Therefore, local activity measure should be appended to ADRC, in order to fully differentiate object details from compression artifacts.

The activity measure we employ is the local entropy coupled by dynamic range of a region. Local entropy has been shown to be a good measure for distinguishing information from digital noise [8]. The local entropy is calculated on the probability density functions (PDFs) of some descriptors inside a region. The PDFs are approximated by the histogram of a descriptor. Considering the context of video processing, we employ luminance intensity as the descriptor. Therefore, the entropy calculation can be defined as

$$H = -\sum_{i=1}^{N} P_R(i) \log_2 P_R(i), \tag{1}$$

where i indicates the bin index in the histogram, N is the total number of bins, and R is a local region around the central pixel over which the entropy is calculated. A region with high activity has a distributed histogram, while the histogram of a region with low activity usually only contains a few peaks. Note that the distribution of the histogram is dominated by the local structure of the region, such that noise and coding artifacts will not affect the overall distribution of the histogram.

According to the information theory, H has a higher value for a spread-out histogram than a peaked one [8], that is, the entropy value of a complex region tends to be larger than a smooth region. Entropy H can be also used as a local blockiness metric, because blocking artifacts reduce the variation of intensities, thus decrease the entropy value. Typically, the entropy value of a region decreases when increasing the compression rate.

To further quantize a region's activity or coding artifacts, entropy should be coupled with dynamic range (DR). DR is defined as the absolute difference between the maximum and minimum pixel values of a region. Table 1 depicts a coarse classification of a region based on the combination of entropy and dynamic range. Here, each 1 bit is used for both entropy and DR. Ringing artifact can be also differentiated, because it usually has a medium-valued entropy and a relatively low DR. For more detailed description of the classification method based on entropy and DR, please refer to [9].

Accordingly, a pixel and its surrounding region can be classified based on the structure, which is represented by ADRC, and the activity measure, which is the local entropy plus dynamic range.

3. LEAST-MEAN-SQUARE OPTIMIZATION

In this section, the least-mean-square (LMS) optimization technique is described to produce optimal coefficients for

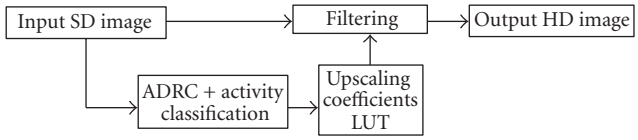

Figure 3: The filtering procedure of the proposed method.

each class based on the pixel classification of the previous section. Figure 2 shows the proposed optimization procedure. Uncompressed HD reference images are first downsampled using bilinear interpolation. The downsampled images are then compressed to introduce coding artifacts. We refer to the compressed downsampled 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.

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:

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

where $w_c(i), i \in [1 \cdots n]$, are the desired coefficients, n is the number of pixels in the aperture, and j indicates a particular aperture belonging to class c.

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

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

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

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

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

$$\begin{aligned} \begin{bmatrix} w_c(1) \\ w_c(2) \\ \vdots \\ w_c(n) \end{bmatrix} &= \begin{bmatrix} \sum_{j=1}^{N_c} F_{D,c}(1,j)F_{D,c}(1,j) & \vdots & \sum_{j=1}^{N_c} F_{D,c}(1,j)F_{D,c}(n,j) \\ \sum_{j=1}^{N_c} F_{D,c}(2,j)F_{D,c}(1,j) & \cdots & \sum_{j=1}^{N_c} F_{D,c}(2,j)F_{D,c}(n,j) \\ \vdots & \vdots & \vdots \\ \sum_{j=1}^{N_c} F_{D,c}(n,j)F_{D,c}(1,j) & \cdots & \sum_{j=1}^{N_c} F_{D,c}(n,j)F_{D,c}(n,j) \end{bmatrix}^{-1} \\ &\times \begin{bmatrix} \sum_{j=1}^{N_c} F_{D,c}(1,j)F_{R,c}(j) \\ \sum_{j=1}^{N_c} F_{D,c}(2,j)F_{R,c}(j) \\ \vdots \\ \sum_{j=1}^{N_c} F_{D,c}(n,j)F_{R,c}(j) \end{bmatrix}. \end{aligned} \tag{5}$$

The LMS-optimized coefficients for each class are then stored in a lookup table (LUT) for future use. Figure 3 shows the filtering procedure of resolution upconversion for compressed materials using the optimized coefficients retrieved from the LUT. A more comprehensive explanation of the LMS optimization technique can be found in [1].

4. EXPERIMENTS AND EVALUATION

In this section, the experimental results of the proposed algorithm are presented. For the optimization procedure, a set of 500 images is used for training. We demonstrate the algorithm with the upscaling factor of 2×2. Therefore, the bilinear interpolation with the scaling factor of 2×2 is used for downsampling during training. Obviously, other upconversion factors can also be achieved. The baseline JPEG software from the Independent JPEG Group website (http://www.ijg.org) is adopted to be the codec 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. An aperture of 3×3 pixels, as depicted in Figure 4, is used for classification in our implementation. Therefore, 8 bits are needed for ADRC coding, since 1 bit can be saved by bitinversion [15]. For the activity measure, we use 2 bits for local entropy and 2 bits for dynamic range. Totally, 12 bits are used for classification.

Table 2: Comparison of numbers of coefficients in the LUT of the three algorithms.

Algorithm	Reference [15] + reference [1]	Reference [1] + reference [15]	Proposed
No. coefficients	$4096 \times 16 \times 13 + 256 \times 9$	$256 \times 9 + 4096 \times 16 \times 13$	$256 \times 16 \times 9$

Table 3: MSE comparison of different algorithms.

Sequence	Reference [1]	Reference [15] + reference [1]	Reference [1] + reference [15]	Proposed
Hotel	116.28	113.40	108.53	104.92
Parrot	36.13	32.15	35.05	31.92
Girl	66.93	59.85	63.72	59.42
Bicycle	183.48	164.25	170.19	161.43
Helicopter	89.01	89.81	83.07	82.85
Game	208.80	209.74	198.27	192.82

For benchmarking, we compare our algorithm with two state-of-the-art classification-based resolution upconversions [1] and artifact reduction [15] methods in concatenation. ADRC is used for classification in the resolution upconversion algorithm. Same as our proposed approach, a 3×3 aperture is used for classification and interpolation. The coding artifact reduction method is based on the classification of structure by adaptive dynamic-range coding (ADRC) and relative position of a pixel in the coding block grid. A diamond-shape 13- 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 upconversion then doing coding artifact reduction, the classification of coding artifact reduction is carried out on the upscaled HD signal and the relative position of a pixel in the block grid is also upscaled 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 analysis-based filtering techniques. For cost comparison, Table 2 shows the numbers of coefficients that need to be stored in lookup 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 downsampled then compressed using the same setting used during the training. Figure 5 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), that is, we calculate the MSE between the original HD sequences and the result sequences processed on the compressed downsampled versions of the original sequences. Table 3 shows the results of the proposed algorithm in comparison to the results of first applying coding artifact reduction then upconversion and first applying upconversion then artifact reduction. The result of resolution upconversion 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 upconversion and artifact reduction affects the performance of the concatenated method. For some

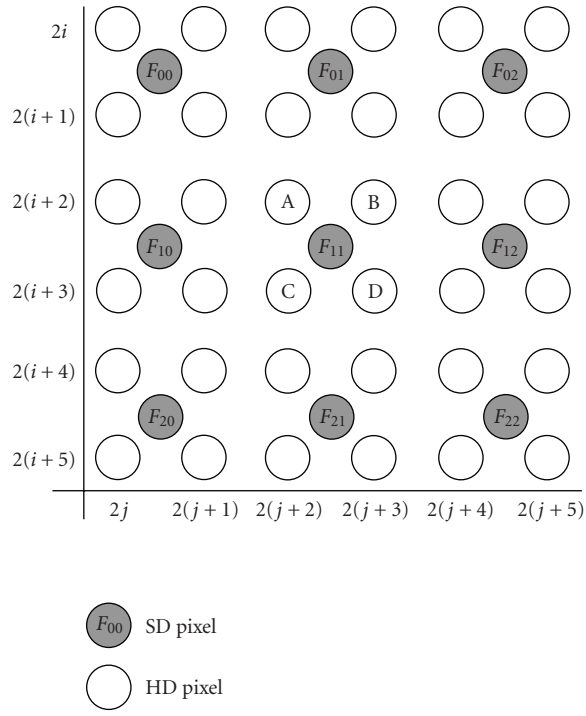

Figure 4: Aperture used in the proposed method. The white pixels are interpolated HD pixels (F_{HD}). The black pixels are SD pixels (F_{SD}), with F_{12} as a shorthand notation for $F_{SD}(1, 2)$ and so forth. The HD pixel A that corresponds to $F_{HD}(2(i+2), 2(j+2))$, is interpolated using nine SD pixels (F_{00} up to F_{22}).

(a) Hotel (b) Parrot (c) Girl

(d) Bicycle (e) Helicopter (f) Game

FIGURE 5: Snapshots of test sequences for experiments.

(a) (b) (c)

FIGURE 6: The cutouts of the girl sequence processed using the three methods: (a) first artifact reduction then resolution upconversion; (b) first resolution upconversion then artifact reduction; (c) the proposed method.

sequences, applying artifact reduction first gives better results; for other sequences, vice verse.

For subjective comparison, Figure 6 shows the results of the three methods on the girl sequence. It is easy to see that the result of first applying upconversion then artifact reduction contains more residual artifacts than the proposed algorithm, because upscaling makes coding artifacts spread out in more pixels and the enlarged coding artifacts are more difficult to remove. The result of first applying artifact reduction then resolution upconversion is blurrier than our proposed algorithm, because the artifact reduction step blurs some details, which cannot be recovered by the upscaling step.

5. CONCLUSION

In this paper, a compression artifacts robust resolution upconversion approach is proposed. Structure and activity information are employed to classify an aperture into object details or coding artifacts. Based on the classification, a least-mean-square optimization technique is used to obtain the

optimized weighting coefficients for upscaling. The optimization is done using a training set composed of the original HD images and the compressed downsampled versions of the original images. The experimental results are compared to two classification-based artifact reduction and resolution upconversion algorithms in concatenation. Our proposed approach outperforms the other two both objectively and subjectively.

REFERENCES

[1] T. Kondo, Y. Node, T. Fujiwara, and Y. Okumura, "Picture conversion apparatus, picture conversion method, learning apparatus and learning method," US patent: no. 6,323,905, November 2001.

[2] C. B. Atkins, C. A. Bouman, and J. P. Allebach, "Optimal image scaling using pixel classification," in *Proceedings of IEEE International Conference on Image Processing (ICIP '01)*, vol. 3, pp. 864–867, Thessaloniki, Greece, October 2001.

[3] X. Li and M. T. Orchard, "New edge-directed interpolation," *IEEE Transactions on Image Processing*, vol. 10, no. 10, pp. 1521–1527, 2001.

[4] J. A. P. Tegenbosch, P. M. Hofman, and M. K. Bosma, "Improving non-linear up-scaling by adapting to the local edge orientation," in *Visual Communications and Image Processing*, vol. 5308 of *Proceedings of SPIE*, pp. 1181–1190, San Jose, Calif, USA, January 2004.

[5] N. Plaziac, "Image interpolation using neural networks," *IEEE Transactions on Image Processing*, vol. 8, no. 11, pp. 1647–1651, 1999.

[6] R. G. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 29, no. 6, pp. 1153–1160, 1981.

[7] H. Greenspan, C. H. Anderson, and S. Akber, "Image enhancement by nonlinear extrapolation in frequency space," *IEEE Transactions on Image Processing*, vol. 9, no. 6, pp. 1035–1048, 2000.

[8] L. Shao and I. Kirenko, "Content adaptive coding artifact reduction for decompressed video and Images," in *Proceedings of International Conference on Consumer Electronics (ICCE '07)*, pp. 1–2, Las Vegas, Nev, USA, January 2007.

[9] L. Shao, "Unified compression artifacts removal based on adaptive learning on activity measure," to appear in *Digital Signal Processing*.

[10] I. Kirenko, R. Muijs, and L. Shao, "Coding artifact reduction using non-reference block grid visibility measure," in *Proceedings of IEEE International Conference on Multimedia and Expo (ICME '06)*, pp. 469–472, Toronto, Ontario, Canada, July 2006.

[11] M. Yuen and H. R. Wu, "Reconstruction artifacts in digital video compression," in *Digital Video Compression: Algorithms and Technologies*, vol. 2419 of *Proceedings of SPIE*, pp. 455–465, San Jose, Calif, USA, February 1995.

[12] W. T. Freeman and E. C. Pasztor, "Markov networks for super-resolution," in *Proceedings of the 34th Annual Conference on Information Sciences and Systems (CISS '00)*, Princeton, NJ, USA, March 2000.

[13] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '00)*, vol. 2, pp. 372–379, Hilton Head Island, SC, USA, June 2000.

[14] T. Kondo, Y. Fujimori, S. Ghosal, and J. J. Carrig, "Method and apparatus for adaptive filter tap selection according to a class," US patent: no. 6,192,161 B1, February 2001.

[15] M. Zhao, R. E. J. Kneepkens, P. M. Hofman, and G. de Haan, "Content adaptive image de-blocking," in *Proceedings of IEEE International Symposium on Consumer Electronics (ISCE '04)*, pp. 299–304, Reading, Mass, USA, September 2004.

Ling Shao is a Research Scientist at the Video Processing and Analysis Group, Philips Research Laboratories, Eindhoven, The Netherlands. He did his B.Eng. degree in electronics engineering at the University of Science and Technology of China, and his M.S. degree in medical imaging and Ph.D. degree in computer vision at Oxford University in the UK. From March to July 2005, he worked as a Senior Research Engineer at Queen's University of Belfast. His research interests include image/video processing, computer vision, and medical imaging.