An Overview and Performance Evaluation of Classification-Based Least Squares Trained Filters

Ling Shao, Hui Zhang, and Gerard de Haan, *Senior Member, IEEE*

Abstract—An overview of the classification-based least squares trained filters on picture quality improvement algorithms is presented. For each algorithm, the training process is unique and individually selected classification methods are proposed. Objective evaluation is carried out to single out the optimal classification method for each application. To optimize combined video processing algorithms, integrated solutions are benchmarked against cascaded filters. The results show that the performance of integrated designs is superior to that of cascaded filters when the combined applications have conflicting demands in the frequency spectrum.

Index Terms—Adaptive filters, classification, integrated processing, least squares optimization, performance evaluation, trained filters, video enhancement.

I. INTRODUCTION

DISPLAYS based on liquid crystal (LCD) and plasma (PDP) technologies are rapidly replacing cathode ray tubes (CRTs) in the consumer markets. Those displays offer a higher picture resolution (multimillion pixels) with very good contrast. However, the source video materials have not all improved to the same extent. In order to improve the picture quality of relatively poor source materials, high-end display manufacturers are putting much effort on video enhancement. Typically, digital filtering algorithms are designed for sharpness enhancement, noise/coding artifacts reduction, and resolution up-conversion, etc. The filter coefficients of video enhancement are usually heuristically optimized [1]–[5], which often involves tedious tuning and testing. Recently, classification-based least squares (LS) filters have been proposed for video enhancement applications including resolution upscaling [6], [7] and coding artifacts reduction [8], [9], which yield promising results. The experimental results in our previously published papers [7]–[9] show the superiority of the classification-based LS filters over other heuristically designed adaptive filters. The main idea is that unclassified LS filters, i.e., the LS optimization is done on the image as a whole, may perform poorly on individual image regions, since a unique LS filter is designed for all pixels in an image. By distinguishing relevant local image characteristics,

the LS filters optimized for separate classes are far superior. The local image characteristics can be classified using local structure or activity information. For each class, the optimized filter coefficients are obtained by an off-line training process, which trains on the combination of targeted images and the degraded versions thereof that act as source. Hence, we shall call the LS optimization design “trained filters.”

In Section II, the trained filters will be discussed in a general framework, which consists of the off-line training process and the run-time filtering process. During the training process, the degradation of targeted images can be specified to suit various applications. Two components, namely classification and degradation, are the most important for the training process.

The classification methods used for trained filters are crucial to ensure best adaptation for relevant local image patterns. Different filter coefficients should be used for different local image content based on the classification. For different applications, the classification usually has to be individually designed. For example, structure information may be the most critical for resolution up-conversion, but local activity measures can be more important for coding artifact reduction. Therefore, the investigation of different classification methods for various video enhancement algorithms will be presented in Section III.

Moreover, different video processing algorithms are often applied in cascade. Those algorithms may perform very well separately, but problems may occur due to the cascading. For example, in the video chain, sharpness enhancement and noise reduction are usually designed separately. The essence of sharpness enhancement is that high frequencies of the spatial image spectrum are amplified compared to the low spatial frequencies, while a noise reduction filter tends to do the opposite, i.e., suppress high frequencies relative to the low frequencies. Hence, there is a conflicting spectral demand on the two procedures, and the utilization of one leads to the deterioration of the other. If noise reduction occurs after sharpness enhancement, the low-pass filter will suppress the enhanced details created by the sharpness enhancement procedure. Usually, sharpness enhancement is applied after noise reduction as this leads to a more acceptable behavior [10]. However, the sharpness enhancement procedure also tends to boost the remaining noise. In Section IV, combined video enhancement algorithms including sharpness enhancement with coding artifact reduction, resolution up-conversion with coding artifact reduction, and sharpness enhancement with resolution up-conversion will be presented. An evaluation of the integrated algorithms and the cascaded filters will also be given to show whether or not the integrated solutions outperform the cascaded techniques, either in performance, or in cost.

In Section V, we conclude this paper and make some suggestions for implementation.

Manuscript received October 5, 2007; revised May 20, 2008. Current version published September 10, 2008. The associate editor coordinating the review of this manuscript and approving it for publication was Vicent Caselles.

L. Shao and G. de Haan are with the Video Processing and Analysis Group, Philips Research Laboratories, High Tech Campus 36, Eindhoven 5656 AE, The Netherlands (e-mail: l.shao@philips.com; g.de.haan@philips.com).

H. Zhang is with the Department of Computer Science and Technology, United International College, Zhuhai, China (e-mail: amyzhang@uic.edu.hk).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIP.2008.2002162

Fig. 1. Training process of the trained filters.

The contributions of this paper, compared to previous works [7]–[9], include that we analyze the classification-based LS algorithm on different video enhancement scenarios and the most appropriate solutions are proposed for each application, which make the trained filters an excellent choice for picture quality improvement in modern display products. The previous works all attempt to tackle individual problems but have not put much effort on making this algorithm a generalized solution for video restoration, which is exactly the objective of this paper. The comparison between integrated methods and cascaded methods gives the future design of combined video processing systems a guide. The comprehensive description and evaluation of the trained filters also serve as a reference for future research on video enhancement.

II. TRAINED FILTERS

The trained filters are composed of two parts: the off-line training process and the run-time filtering process. Fig. 1 shows the training process of the trained filters. Target images are first degraded according to the specification of the application. The degradation during the training process is the inverse of the desired enhancement in the filtering process. For example, the degradation during the training process for sharpness enhancement is blurring. We shall refer to the images after degradation as (simulated) source images. In the source images, each pixel is classified according to the local properties of the image using an applicable specific pixel classification method. All the pixels and their neighborhoods belonging to a specific class and their corresponding pixels in the target (original) images are accumulated, and the optimal coefficients are obtained from a LS minimization.

Let $F_{D,c}(i, j)$ be the apertures of the source images and $F_{R,c}(j)$ be the corresponding pixels in the target images for a particular class c , respectively. Then the filtered pixels $F_{F,c}(j)$ can be obtained by the desired optimal coefficients as follows:

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

Fig. 2. Filtering process of the trained filters.

where $w_c(i), i \in [1 \dots n]$ 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 target pixels is

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

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

$$\begin{aligned} \frac{\partial e^2}{\partial w_c(k)} &= \sum_{j=1}^{N_c} 2 F_{D,c}(k, j) \\ &\quad \times \left[F_{R,c}(j) - \sum_{i=1}^n w_c(i) F_{D,c}(i, j) \right] \\ &= 0. \end{aligned} \quad (3)$$

By solving the above equation using Gaussian elimination, we will get the optimal coefficients as follows [see (4), shown at the bottom of the page]. The LS optimized coefficients for each class are then stored in a look-up table (LUT) for future use.

Fig. 2 shows the filtering process of the algorithm using the optimized coefficients. Each pixel to be filtered is classified according to its local image properties using exactly the same classification method as during training. The coefficients are retrieved from the LUT using the classification to index the LUT. The pixels are then filtered using the optimized coefficients.

III. PIXEL CLASSIFICATION

In this section, some classification methods will be evaluated for sharpness enhancement, coding artifacts reduction and resolution up-conversion. Classification is used to distinguish local image characteristics so that the interclass differences are ideally much larger than the intraclass variations. In theory, the class number can be huge, considering N number of pixels that

$$\begin{bmatrix} w_c(1) \\ w_c(2) \\ \vdots \\ w_c(n) \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{N_c} F_{D,c}(1, j) F_{D,c}(1, j) & \dots & \sum_{j=1}^{N_c} F_{D,c}(1, j) F_{D,c}(n, j) \\ \sum_{j=1}^{N_c} F_{D,c}(2, j) F_{D,c}(1, j) & \dots & \sum_{j=1}^{N_c} F_{D,c}(2, j) F_{D,c}(n, j) \\ \dots & \dots & \dots \\ \sum_{j=1}^{N_c} F_{D,c}(n, j) F_{D,c}(1, j) & \dots & \sum_{j=1}^{N_c} F_{D,c}(n, j) F_{D,c}(n, j) \end{bmatrix}^{-1} \begin{bmatrix} \sum_{j=1}^{N_c} F_{D,c}(1, j) F_{R,c}(j) \\ \sum_{j=1}^{N_c} F_{D,c}(2, j) F_{R,c}(j) \\ \dots \\ \sum_{j=1}^{N_c} F_{D,c}(n, j) F_{R,c}(j) \end{bmatrix} \quad (4)$$

Fig. 3. ADRC code of a 3×3 block.

are valued between 0 and 255. In practice, various techniques can be employed to compress the classification. The purpose is to classify the most critical characteristics specific for the application. For most video enhancement applications, local structure and local complexity are relevant features. Straightforwardly, Principal Component Analysis (PCA) and Discrete Cosine Transform (DCT) could be used to classify the frequency information of a local region. However, in the scenario of real-time processing, simplified and efficient classification methods are preferred.

Adaptive Dynamic Range Coding (ADRC) [11] has been proposed for representing the structure of a region because of its high efficiency and simplicity. The ADRC code of each pixel x_i in an observation aperture is defined as: $ADRC(x_i) = 0$, if $V(x_i) \leq V_{av}$; 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. The ADRC code of an image kernel is the concatenation of the ADRC codes of all the pixels in that kernel. Fig. 3 shows a diagram of the ADRC code on a 3×3 block.

Obviously, only using structure for classification is not sufficient, because the structure of noise or coding artifacts could be exactly the same as that of object details, and because high contrast structures and low contrast structures should be treated differently. Hu and de Haan [12] attempted to further differentiate coding artifacts from object details by adding some contrast information to ADRC, by using the dynamic range (DR), to ADRC. DR is simply the absolute difference between the maximum and minimum pixel values of the filter kernel. Shao [13] proposed to use local entropy as an activity measure for distinguishing complex regions from uniform regions. The entropy value is calculated on the probability density functions of the pixel intensity distribution. The local entropy of a region can be defined as follows:

$$H = - \sum_{i=1}^N P_R(i) \log_2 P_R(i) \quad (5)$$

where i indicates the bin index, $P_R(i)$ is the probability of pixels having a value in the range of bin i and R is a local region inside which the entropy is calculated. Another activity measure called Mean Absolute Difference (MAG) was presented by Shao [14] for determining the complexity of a region. MAG is defined as follows:

$$MAG = \frac{1}{N-1} \sum_{i=1}^{N-1} |F(0) - F(i)| \quad (6)$$

where $F(i)$ denotes the intensity value of a pixel in a region, $F(0)$ is the intensity of the pixel in the center, and N is the

Fig. 4. Classification based on the complexity measures: (a) DR; (b) MAG; (c) STD; (d) Entropy.

number of pixels in the region. Standard Deviation (STD) is employed as another complexity metric for a local region in [15].

Fig. 4 shows the classification of pixels on Lenna using the four complexity measures discussed above. The superimposed dots indicate pixels with the value of a particular complexity measure above a certain threshold. We can see that all the four complexity measures have similar behavior with some subtle difference on detailed regions.

In the following, structure information using ADRC coupled with one of the complexity measures described above will be used for classification, and the performance of those complexity measures will be evaluated. For all the three applications, a 13 pixel diamond-shaped aperture, as depicted in Fig. 5, is used for both classification and filtering. Therefore, 12 bits are needed for the ADRC code, because 1 bit can be saved using bit-inversion [16], and two more bits are used for representing the complexity of a region. So, in total, 14 bits are used for classification. The four quantization levels of each complexity measure are manually adjusted to make the classifications more appropriate for individual applications. It would be beneficial to use more bits for classifying complexity, but for the consideration of cost fewer bits are preferred. For TV systems, intraframe processing is more cost-effective than interframe processing. So, no temporal filtering will be discussed here. In order to guarantee that enough instances of each class occur in the training set, a large number of images (more than 10000 in HD resolution) with a variety of image content are used for training. Besides, a fall-back mechanism is adopted when there is not enough occurrence of a particular class in the training data, i.e., default filters will be used for image patterns that seldom occur.

A. Sharpness Enhancement

For sharpness enhancement, filter coefficients should be adaptive to structure or edge orientation, and edges of different magnitudes should be enhanced differently. Therefore, the combination of structure and some complexity measure is

Fig. 5. Diamond-shaped aperture; the black pixel indicates the central pixel.

Fig. 6. Snapshots of the test sequences.

used for classification. During the training process, the applied degradation to simulate the source images is Gaussian blur. The Gaussian filter we use represents a standard normal distribution, i.e., $\mu = 0$ and $\sigma = 1$, with a filter footprint of 5×5 . In principle, some noise should also be added for degradation to test the robustness of the sharpness enhancement to noise, but it can be considered as an integrated filter, which will be discussed in Section III-B.

For objective evaluation, we calculate the mean square error (MSE) between the target sequences and the result sequences processed on the source sequences. Fig. 6 depicts the snapshots of the six test sequences we used for experiments. All the test sequences are excluded from the training set. Table I shows the MSE scores of different classification methods on different sequences.

From Table I, we can see that *ADRC* coupled with a complexity measure always performs better than just using *ADRC*. The *ADRC* code of an object edge could be exactly the same as that of a flat region, when there is small variation of pixel values due to noise in the flat region. By distinguishing object

TABLE I
MSE SCORES FOR SHARPNESS ENHANCEMENT

	<i>ADRC</i>	<i>ADRC+DR</i>	<i>ADRC+MAG</i>	<i>ADRC+STD</i>	<i>ADRC+Entropy</i>
Bicycle	16.41	13.48	13.12	13.31	14.11
Football	15.12	14.49	14.42	14.49	14.78
Girlsea	4.73	4.48	4.21	4.4	4.57
Teen	5.76	4.84	4.79	4.94	5.41
Wheel	10.43	7.67	7.57	7.61	8.67
Yvonne	10.22	8.83	8.76	8.84	9.4

edges from uniform regions using a complexity measure during training, the trained filters tend to be more optimal for sharpening object details. Among the complexity measures, *MAG* performs the best. Fig. 7 shows the result of sharpness enhancement using *ADRC* and *MAG* for classification on the Bicycle sequence.

To test the robustness of the trained filters, Table II shows the results when input sequences are blurred slightly differently to that used during training. *ADRC* plus *STD* is used for classification. The MSE scores increase with the enlargement of the difference between the Gaussian blurs used during training and during test sequence simulation. For benchmarking, the results of Luminance Transient Improvement (LTI) [17] are also shown. For most cases, the trained filters outperform LTI except when the test sequences are blurred much milder than the simulated degraded sequences during the training process. For those cases, the trained filters tend to over sharpen the signals.

B. Coding Artifact Reduction

The aim of coding artifact reduction is to suppress visible coding artifacts but at the same time preserve object details. Here, structure information is not only important for determining the edge orientation of object details, but also useful for distinguishing blocking artifacts from object edges. Since coding artifacts could have the similar structure pattern as object details, complexity measures are used to further differentiate them, since we expect coding artifacts to be less visible in complex areas and coding artifacts tend to have a lower amplitude than object details. Compression standards, such as JPEG, MPEG-2, and H.264, can be used for degrading the target sequences. Since images compressed by JPEG with the same compression rate have more consistent artifacts level and for the purpose of simplicity, JPEG is adopted as the codec for degrading the target images during training. For motion compensated coding standards, e.g., MPEG, we expect it would be beneficial to make separate training processes for I-frames, B-frames and P-frames. The JPEG quality in our implementation is set to be 20 (<http://www.iijg.org>), because sufficient coding artifacts are visible and image details are not completely lost using that quality level.

Table III shows the MSE scores between the target sequences and the result sequences processed on the source sequences. The combination of *ADRC* and a complexity measure always outperforms only using *ADRC*, and *STD* appears to be the most effective complexity measure for coding artifact reduction.

(a)

(b)

Fig. 7. (a) Gaussian blurred image. (b) Output after sharpness enhancement.

TABLE II
MSE SCORES FOR SHARPNESS ENHANCEMENT FOR SEQUENCES BLURRED
WITH DIFFERENT GAUSSIAN FILTERS

Sequence/Blur	Trained Filters	LTI
Bicycle/ $\sigma=0.8$	50.73	27.25
Bicycle/ $\sigma=0.9$	15.99	33.71
Bicycle/ $\sigma=1.0$	13.31	43.45
Bicycle/ $\sigma=1.1$	24.20	55.30
Bicycle/ $\sigma=1.2$	42.96	68.16
Football/ $\sigma=0.8$	44.58	30.96
Football/ $\sigma=0.9$	17.60	37.74
Football/ $\sigma=1.0$	14.49	46.67
Football/ $\sigma=1.1$	22.55	56.96
Football/ $\sigma=1.2$	37.48	67.85
Girlsea/ $\sigma=0.8$	10.08	7.40
Girlsea/ $\sigma=0.9$	5.02	8.17
Girlsea/ $\sigma=1.0$	4.40	9.76
Girlsea/ $\sigma=1.1$	5.94	11.91
Girlsea/ $\sigma=1.2$	8.81	14.32

TABLE III
MSE SCORES FOR CODING ARTIFACT REDUCTION
USING ADRC AND COMPLEXITY

	ADRC	ADRC+ DR	ADRC+ MAG	ADRC+ STD	ADRC+ Entropy
Bicycle	44.27	43.02	42.01	41.48	43.58
Football	45.75	44.88	44.49	44.30	45.24
Girlsea	18.47	18.03	17.93	17.81	18.22
Teeny	19.04	18.03	17.93	17.78	18.42
Wheel	36.79	35.15	34.04	33.37	35.92
Yvonne	37.06	35.76	35.39	34.93	36.37

TABLE IV
MSE SCORES FOR CODING ARTIFACT REDUCTION ONLY USING COMPLEXITY

	DR	MAG	STD	Entropy
Bicycle	53.86	53.38	52.26	54.41
Football	49.15	48.93	48.34	49.64
Girlsea	19.18	19.18	18.84	19.60
Teeny	19.48	19.50	19.18	20.06
Wheel	44.04	43.54	41.82	45.03
Yvonne	39.27	39.28	38.36	40.09

For testing the usefulness of structure information, Table IV illustrates the MSE results just using complexity for classification. It is easy to see that the MSE scores become worse especially for sequences dominated by structures, e.g., Bicycle and Wheel.

Fig. 8 shows the results on the Teeny sequence using the combination of ADRC and STD for classification. The combination of ADRC and STD can remove severe coding artifacts and at the same time preserve object edges.

The robustness of trained filters for coding artifact reduction is evaluated by testing the algorithm on sequences compressed at different quality levels. The classification method is the combination of ADRC and STD. From Table V, we can see that the MSE scores decrease when the quality of the test sequences increases. For benchmarking, the results of two content adaptive algorithms for compression artifacts removal [1], [18] are also depicted. Trained filters produce significantly better results for all the sequences with different quality levels, which confirms the robustness of trained filters for artifact reduction.

C. Resolution Up-Conversion

The purpose of image upscaling is to increase the number of pixels and preserve all the details in the original image. Resolution upscaling techniques usually result in blurred images, because the scaling process does not add new frequency components. Our goal is to actually *extend the spatial video spectrum*, i.e., create frequency components that are not present in the SD-signal, which could not even be represented on an SD-display, but contribute to an increased picture quality when shown on an HDTV-screen. Fig. 9 illustrates the concept. The trained filters function as nonlinear filters due to the pixel classification mechanism, which enables the filters to create new frequency components. To distinguish this technology from scaling, we shall call it *Resolution Up-conversion*. Structure information is the most important for resolution up-conversion. We shall also verify if complex regions and flat regions could profit from being up-converted differently. We demonstrate the resolution

(a)

(b)

Fig. 8. (a) Image with coding artifacts. (b) Output after coding artifact reduction.

up-conversion algorithm with the scaling factor of 2 both horizontally and vertically. Therefore, the “degradation” during the training process is a downscaling with the factor of 2 both horizontally and vertically. The downscaling method used is simply a bi-linear filter.

Table VI shows the MSE results of up-conversion between the target sequences and the up-converted source sequences. We can see that the complexity measures do not contribute to the MSE scores for resolution up-conversion. This may be because contrast information is irrelevant for interpolation. However, if there is noise in the images being up-converted, which is normally the case for broadcasted materials, adding a complexity measure will still be beneficial. We will prove the advantage of including a complexity measure for up-converting images with coding artifacts in Section IV-B.

TABLE V
MSE SCORES FOR CODING ARTIFACT REDUCTION FOR SEQUENCES COMPRESSED WITH DIFFERENT QUALITY LEVELS

Sequence/JPEG Quality	Trained Filters	Ref [18]	Ref [1]
Bicycle/10	80.35	102.43	110.67
Bicycle/15	50.74	73.58	75.28
Bicycle/20	41.48	59.82	58.29
Bicycle/25	29.96	51.26	47.68
Bicycle/30	25.14	45.36	40.36
Football/10	78.12	90.70	97.22
Football/15	53.58	66.96	68.04
Football/20	44.30	55.11	53.42
Football/25	35.41	47.93	44.34
Football/30	31.01	43.01	38.20
Girlsea/10	38.35	40.08	44.87
Girlsea/15	23.19	25.58	28.15
Girlsea/20	17.81	19.08	20.46
Girlsea/25	13.57	15.74	16.24
Girlsea/30	11.55	13.55	13.55

Fig. 9. Illustration showing the difference between spatial scaling (top) and resolution up-conversion (bottom) of an SD-image to an HD-grid. The bottom picture is sharper, and its spectrum shows many new high-frequency components. (Reproduced from Zhao *et al.* [21]).

TABLE VI
MSE SCORES FOR RESOLUTION UPSCALING USING ADRC AND COMPLEXITY

	ADRC	ADRC+ DR	ADRC+ MAG	ADRC+ STD	ADRC+ Entropy
Bicycle	67.98	67.96	67.97	67.91	71.76
Football	68.22	68.13	68.53	68.51	70.11
Girlsea	14.26	14.29	14.71	14.69	14.23
Teeny	17.66	17.64	18.14	18.13	17.49
Wheel	43.98	43.95	43.90	43.82	47.90
Yvonne	47.25	47.27	47.40	47.39	49.00

Fig. 10 depicts the result of resolution up-conversion using ADRC for classification on the Bicycle sequence. Both edges and fine details are preserved very well.

Since the alignment of pixels in a filter aperture used for up-scaling is critical for trained filters, the scaling factor should be predefined during the training process. Fortunately, only a certain number of scaling factors is required for TV display systems. For those scaling factors that are not trained beforehand,

(a)

(b)

Fig. 10. (a) Downscaled image. (b) Output after resolution up-conversion.

TABLE VII
MSE SCORES FOR RESOLUTION UPSCALING OF DIFFERENT METHODS

	Trained Filters	Bi-cubic	NEDI
Bicycle	67.98	76.54	72.30
Football	68.22	75.69	73.05
Girlsea	14.26	17.45	18.02
Teeny	17.66	21.50	22.88
Wheel	43.98	53.83	48.61
Yvonne	47.25	56.20	52.86

the trained filters can be combined with other scaling methods, such as bi-linear interpolation, to make it more flexible.

To evaluate the performance of trained filters for resolution up-conversion, we compare its MSE scores with two other scaling techniques, namely Bi-cubic interpolation and New Edge Directed Interpolation (NEDI) [19]. *ADRC* is adopted as the classification method for trained filters here. Table VII shows that trained filters outperform the other two methods.

D. Discussion

We evaluated the classification methods for sharpness enhancement, coding artifact reduction, and resolution up-conversion in this section. Structure information using *ADRC* is proven to be relevant for all three applications. Complexity measures are useful for sharpening and coding artifact reduction, but can be discarded for resolution up-conversion. The underlying reason might be that both sharpening and coding artifact reduction tend to do filtering perpendicular to edges: object edges for

sharpening and block edges for artifact reduction, whereas resolution up-conversion does filtering along edges. Therefore, the differentiation between high amplitude edges and low amplitude edges is more relevant for sharpening and coding artifact reduction. For sharpness enhancement, *MAG* produces the best results as a complexity measure; and *STD* is the most effective for coding artifact reduction.

IV. INTEGRATED PROCESSING VERSUS CASCADED PROCESSING

In the video processing chain, various video enhancement algorithms are applied in cascade. The optimization of one may lead to the deterioration of the other, especially when there is a conflict in spectral demands. In this section, we evaluate the performance of integrated processing in comparison to cascaded processing. For the integrated method, one filtering process is used to meet the needs of multiple requirements, e.g., sharpness enhancement and compression artifacts removal. On the contrary, the cascaded method is composed of two filtering processes and the output of the first filtering process is used as the input for the second filtering process. The best classification method for each application from the previous section will be utilized to construct the cascaded methods. The training processes of individual filters we employ to build the cascaded systems are exactly the same as described in the previous section. During the filtering process of the cascaded methods, the pixels are first classified and the first transformation is applied, then we classify again the pixels of this 'intermediate' image, and apply the second transformation according to the new classification. In theory, the performance of the cascaded approach would be improved, if the sequential filtering process was taken into account during the training, i.e., if we had the knowledge of how the image is previously filtered when designing the second part of the cascaded filters. However, in practice, the prior knowledge of how an image is previously processed is not always available or cannot be easily obtained or detected. Therefore, we currently do not take the prior knowledge into account when constructing the cascaded filters. During the training process of the integrated solutions, target images are first degraded according to the requirements of the desired integrated filter, e.g., first blur then compression. Then, classification-based LS optimization is carried out to output optimized filter coefficients for each class. The filtering process of the integrated methods is the same as that of an individual filter in Section III, i.e., only one pass of filtering is needed. We employ *ADRC* plus *STD* to be the classification method for the integrated methods, because *STD* as a complexity measure gives one of the best MSE scores for both sharpness enhancement and coding artifact reduction in the previous section. As in Section III, a 13 pixel diamond-shaped aperture is used for both classification and filtering, and 12 bits are used for *ADRC*, 2 bits are used for *STD*, respectively. The settings for degradation during the training process for the integrated methods are exactly the same as for the separate methods discussed in Section III.

A. Integrated Sharpening and Coding Artifact Reduction

In the video chain, sharpness enhancement is usually applied after coding artifact reduction, because the coding artifacts be-

TABLE VIII
MSE SCORES OF SHARPENING WITH CODING ARTIFACT REDUCTION

	Cascaded	Integrated (<i>ADRC</i> + <i>STD</i>)	Integrated (<i>STD</i>)	Ref [20]
Bicycle	214.13	95.01	113.57	162.05
Football	176.46	88.14	93.68	130.02
Girlsea	60.92	30.84	31.80	47.74
Teeny	68.86	31.53	32.97	53.35
Wheel	163.10	74.68	85.49	117.20
Yvonne	134.36	72.51	76.04	99.60

come more difficult to remove if they are enhanced first. During the training process of the integrated sharpening and coding artifact reduction, the degradation is first applying Gaussian blur then adding coding artifacts using JPEG.

Table VIII shows the MSE scores of the integrated sharpening and coding artifact reduction between the target sequences and the filtered outputs of the decompressed blurred versions of the target sequences in comparison to the cascaded method. The results of the integrated method just using *STD* for classification are also shown. The integrated methods, especially the one based on the classification of *ADRC* and *STD*, outperform the cascaded method significantly. For benchmarking, the results of another combined method for sharpness enhancement and coding artifact reduction [20] are also shown. The integrated trained filters perform much better than the approach in [20].

Fig. 11 shows the output of the integrated method using *ADRC* and *STD* in comparison to that of the cascaded method on the Teeny sequence. It is easy to see that the integrated method can simultaneously sharpen object details and remove coding artifacts. The cascaded method creates overshoots and enhances some remaining artifacts from the artifact reduction step.

B. Integrated Coding Artifact Reduction and Resolution Up-Conversion

The purpose of resolution up-conversion is to preserve the object details and create new frequency components from the original low-resolution images. If there are coding artifacts in the low-resolution images, they will also be preserved and even enhanced, which makes them even more difficult to remove, because the coding artifacts will be more spread-out and harder to distinguish from object details. In the video processing chain, coding artifacts are usually first reduced before resolution up-scaling. During the training process of the integrated coding artifact reduction and resolution up-conversion, the degradation is first downscaling then compression.

Table IX shows the MSE results of the integrated and cascaded methods between the target sequences and the filtered source sequences. The difference between the integrated method and the cascaded method is subtle. In [8], the integrated coding artifact reduction and resolution up-conversion is shown to be more effective than the cascaded method in terms of MSE. However, a different classification technique, which is the combination of *ADRC* and the relative position of the pixel in the coding block, was used for coding artifact reduction for the cascaded method. From Table VI, one can see that the

(a)

(b)

(c)

Fig. 11. (a) Blurred image with coding artifacts. (b) Output of the cascaded method. (c) Output of the integrated method.

cascaded method can perform as well as the integrated method for coding artifact reduction and resolution up-conversion, if a better classification technique is utilized.

Fig. 12 depicts the results of both the integrated method and the cascaded method on the Bicycle sequence. No noticeable difference can be perceived between the results, which is also the case for a large number of sequences we test on.

TABLE IX
MSE SCORES OF CODING ARTIFACT REDUCTION
WITH RESOLUTION UP-CONVERSION

	Cascaded	Integrated (ADRC+STD)
Bicycle	175.79	173.74
Football	152.42	152.17
Girlsea	56.24	56.72
Teeny	58.35	58.44
Wheel	144.57	141.27
Yvonne	130.33	130.49

(a)

(b)

(c)

Fig. 12. (a) Downscaled image with coding artifacts. (b) Output of the cascaded method. (c) Output of the integrated method.

C. Integrated Sharpening and Resolution Up-Conversion

In the video processing chain, sharpness enhancement is usually applied after resolution up-conversion. During the training

TABLE X
MSE SCORES OF SHARPENING WITH RESOLUTION UP-CONVERSION

	Cascaded	Integrated (ADRC+STD)	Ref [4]
Bicycle	425.77	69.22	81.60
Football	362.50	69.83	78.69
Girlsea	115.23	15.32	19.59
Teeny	140.35	19.24	25.51
Wheel	229.33	40.85	51.13
Yvonne	244.59	45.24	60.69

process of the integrated sharpening and resolution up-conversion, the degradation is first Gaussian blur then linear down-scaling.

Table X shows the MSE scores of the two methods between the target sequences and filtered source sequences. The integrated method outperforms the cascaded method tremendously for all the test sequences. For benchmarking, the results of another combined method for sharpness enhancement and resolution up-conversion [4] are also shown.

Fig. 13 illustrates the results of both methods on the Bicycle sequence. Lots of overshoots and distortion are noticeable in the output of the cascaded method; while the integrated method produces a sharp and favorable output.

D. Discussion

In this section, integrated methods are evaluated and compared to the cascaded methods. The integrated sharpening and coding artifact reduction and the integrated sharpening and resolution up-conversion both outperform their cascaded counterparts significantly, whereas the integrated coding artifact reduction and resolution up-conversion gives a similar performance as the cascaded method. The results comply very well with the spectral behavior of each filtering process. Obviously, sharpness enhancement amplifies high frequencies of the image spectrum. Resolution up-conversion is essentially an all-pass filtering process, but it also creates new higher frequencies as we discussed in Section III-C. Unclassified coding artifact reduction approaches use low-pass filters, because they smooth both coding artifacts and object details. Coding artifact reduction based on the proposed trained filters reduces coding artifacts as well as reconstructs object edges and structures according to classification. Generally, coding artifact reduction based on trained filters can be considered as an all-pass filtering process, because for certain regions containing coding artifacts it serves as a low-pass filter and for other regions consisting of object details it enhances high frequencies of the signal.

There is a conflicting spectral demand between sharpness enhancement and coding artifact reduction, because sharpening boosts high frequencies and coding artifact reduction tends to suppress certain high frequencies on artifacts. For sharpening and resolution up-conversion, sharpness enhancement tends to over-enhance the new high frequencies created by resolution up-conversion in the cascaded method, which results in overshoots and downshoots. Therefore, the integrated sharpening and coding artifact reduction and the integrated sharpening and resolution up-conversion have advantages over the cascaded methods. However, if the spectral conflict is subtle, which is the case for the integrated coding artifact reduction and resolution

Fig. 13. (a) Downscaled blurred image. (b) Output of the cascaded method. (c) Output of the integrated method.

up-conversion, the advantages of the integrated method become small.

In terms of cost of the look-up table (LUT) for storing filter coefficients, the integrated methods always only consume half of the size of that of the cascaded methods.

V. CONCLUSION

In this paper, we have presented an overview of the LS trained filters. The performance of the trained filters is evaluated on a number of video enhancement scenarios, including sharpness enhancement, coding artifact reduction and resolution up-conversion. Different classification methods, such as structure based on *ADRC* and complexity measure based on *STD*, are adopted and compared. The combination of structure

information and a complexity measure is proven to be effective for coding artifact reduction and sharpness enhancement, while for resolution up-conversion complexity measures can be discarded.

Integrated processing methods using trained filters are also investigated and compared to the cascaded methods. When there is a conflicting spectral demand between the processing procedures, the integrated methods outperform the cascaded methods significantly; otherwise, the difference tends to be subtle.

In this paper, the trained filters are only applied in a pre-defined scenario, e.g., the coding artifact reduction algorithm is only used for a particular compression rate. For real-life implementation, some quality metrics should be employed to classify the source materials into different quality levels, such as the sharpness level and the artifact level. The quality metrics can be integrated into the classification. During the training process, the target images can be degraded using different levels of degradation functions. Classification methods are designed to distinguish both the quality and the content for each quality level. Quality metrics are considered to be very important for both heuristically designed filters and the proposed trained filters. In future work, quality metrics will be integrated into trained filters to make them more flexible for different source materials.

For some applications, such as compression artifacts removal, temporal information may be beneficial for reducing the flicking effect between video frames. Either motion adaptive or motion compensated methods can be adopted. Both methods need to rely on motion information, which makes the methods too expensive for implementation.

REFERENCES

- [1] L. Shao and I. Kirenko, "Coding artifact reduction based on local entropy analysis," *IEEE Trans. Consum. Electron.*, vol. 53, no. 2, pp. 691–696, May 2007.
- [2] I. Kirenko, R. Muijs, and L. Shao, "Coding artifact reduction using non-reference block grid visibility measure," presented at the IEEE Int. Conf. Multimedia and Expo, Toronto, ON, Canada, Jul. 2006.
- [3] H. Greenspan *et al.*, "Image enhancement by nonlinear extrapolation in frequency space," *IEEE Trans. Image Process.*, vol. 9, no. 6, pp. 1035–1048, Jun. 2000.
- [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] Y. Yang and L. Boroczky, "A new enhancement method for digital video applications," *IEEE Trans. Consum. Electron.*, vol. 48, no. 3, pp. 435–443, Aug. 2002.
- [6] T. Kondo, Y. Node, T. Fujiwara, and Y. Okumura, "Picture conversion apparatus, picture conversion method, learning apparatus and learning method," U.S. patent 6,323,905, Nov. 2001.
- [7] L. Shao, "Adaptive resolution upconversion for compressed video using pixel classification," *EURASIP J. Adv. Signal Process.*, vol. 2007, 2007.
- [8] M. Zhao, R. E. J. Kneepkens, P. M. Hofman, and G. d. Haan, "Content adaptive image de-blocking," presented at the IEEE Int. Symp. Consumer Electronics, Sep. 2004.
- [9] L. Shao, "Unified compression artifacts removal based on adaptive learning on activity measure," *Digital Signal Process.*, vol. 17, no. 6, pp. 1065–1070, Nov. 2007.
- [10] O. A. Ojo and T. G. Kwaiitaal-Spassova, "An algorithm for integrated noise reduction and sharpness enhancement," *IEEE Trans. Consum. Electron.*, vol. 46, no. 3, pp. 474–480, Aug. 2000.
- [11] T. Kondo and K. Kawaguchi, "Adaptive dynamic range encoding method and apparatus," U.S. patent 5,444,487, Aug. 1995.
- [12] H. Hu and G. d. Haan, "Simultaneous coding artifact reduction and sharpness enhancement," presented at the IEEE Int. Conf. Consumer Electronics, Jan. 2007.

- [13] L. Shao and I. Kirenko, "Content adaptive coding artifact reduction for decompressed video and images," presented at the IEEE Int. Conf. Consumer Electronics, Jan. 2007.
- [14] L. Shao, "Simultaneous coding artifact reduction and sharpness enhancement for block-based compressed images and videos," *Signal Process.: Image Commun.*, vol. 23, no. 6, pp. 463–470, Jul. 2008.
- [15] L. Shao, H. Hu, and G. de Haan, "Coding artifacts robust resolution up-conversion," presented at the IEEE Int. Conf. Image Processing, San Antonio, TX, Sep. 2007.
- [16] T. Kondo et al., "Method and apparatus for adaptive filter tap selection according to a class," U.S. patent 6,192,161, Feb. 2001.
- [17] P. Rieder and G. Scheffler, "New concept on denoising and sharpening of video signals," *IEEE Trans. Consum. Electron.*, vol. 47, no. 3, pp. 666–671, Aug. 2001.
- [18] S. Kim, J. Yi, H. Kim, and J. Ra, "A deblocking filter with two separate modes in block-based video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 1, pp. 156–160, Feb. 1999.
- [19] X. Li and M. T. Orchard, "New edge-directed interpolation," *IEEE Trans. Image Process.*, vol. 10, pp. 1521–1527, Oct. 2001.
- [20] I. Kirenko, L. Shao, and R. Muijs, "Enhancement of compressed video signals using a local blockiness metric," presented at the IEEE Int. Conf. Acoustics, Speech, and Signal Processing, Las Vegas, NV, Mar.–Apr. 2008.
- [21] M. Zhao, M. Bosma, and G. d. Haan, "Making the best of legacy video on modern displays," *J. Soc. Inf. Process.*, vol. 15, no. 1, pp. 49–60, Jan. 2007.

Ling Shao received the B.Eng. degree in electronic engineering from the University of Science and Technology of China and the M.Sc. degree in medical imaging and the Ph.D. (D.Phil.) degree in computer vision from the Robotics Research Group, University of Oxford, Oxford, U.K.

He is a Senior Research Scientist in the Video Processing and Analysis Group, Philips Research Laboratories, Eindhoven, The Netherlands. From March to July 2005, he was a Senior Research Engineer with the Institute of Electronics, Communications and In-

formation Technology, Queen's University of Belfast, U.K. He has published over 30 journal and conference papers in image/video processing, computer vision, and medical imaging.

image sequences.

Hui Zhang received the B.Eng. and M.Eng. degrees in computer science and engineering, with honors, from Zhejiang University, China, in 1999 and 2002, respectively, and the Ph.D. degree in computer vision from the University of Hong Kong in 2007.

She is now an Assistant Professor with the Department of Computer Science and Technology, United International College, Zhuhai, China. Her research interests are in image processing and computer vision, including camera calibration, model reconstruction and representation, and motion estimation from

Gerard de Haan (SM'97) received the B.Sc., M.Sc., and Ph.D. degrees from the Delft University of Technology, Delft, The Netherlands, 1977, 1979, and 1992, respectively.

He joined Philips Research in 1979. He has led research projects in the area of video processing, and participated in European projects. He has coached students from various universities and has been teaching for the Philips Centre for Technical Training since 1998. Since 2000, he has been a Research Fellow in the Video Processing and Analysis

group of Philips Research, Eindhoven, The Netherlands, and a part-time full Professor at the Eindhoven University of Technology teaching video processing for multimedia systems. He has a particular interest in algorithms for motion estimation, video format conversion, and image enhancement. His work in these areas has resulted in several books, more than 130 papers, about 100 patents and patent applications, and various commercially available ICs.

Dr. de Haan was a winner in the 1995, 1997, 1998, 2002, and 2003 ICCE Outstanding Paper Awards program, the 1998 recipient of the Gilles Holst Award, and the Chester Sall Award from the IEEE Consumer Electronics Society in 2002. The Philips "Natural Motion Television" concept, based on his Ph.D. studies, received the European Video Innovation Award in 1995 from the European Imaging and Sound Association. In 2001, the successor of this concept "Digital Natural Motion Television," received a "Business Innovation Award" from the Wall Street Journal Europe. He serves on the program committees of various international conferences on image and video processing.