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—Video enhancement, Least Squares optimization, trained filters, adaptive filters, classification, integrated processing, performance evaluation.

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 (multi-million 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 Least Squares filters, i.e. the Least Squares optimization is done on the image as a whole, may perform poorly on individual image regions, since a unique Least Squares 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 2, 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 3.

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 4, 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 5, we conclude this paper and make some
suggestions for implementation.
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. Figure 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 Least Squares (LS) minimization.

![Diagram of training process](image)

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:

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

where \( w_c(i), i \in [1...n] \) 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:

\[
e^2 = \sum_{j=1}^{N_c} (F_{R,c}(j) - F_{F,c}(j))^2
\]

\[
= \sum_{j=1}^{N_c} [F_{R,c}(j) - \sum_{i=1}^{n} w_c(i) F_{D,c}(i,j)]^2
\]  

(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 \( w_c(k), k \in [1...n] \) should be equal to zero.

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

(3)

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

\[
\begin{bmatrix}
w_c(1) \\
w_c(2) \\
\vdots \\
w_c(n)
\end{bmatrix} = \frac{1}{\sum_{j=1}^{N_c} F_{R,c}(j)} \begin{bmatrix}
\sum_{j=1}^{N_c} F_{R,c}(1,j) F_{D,c}(1,j) \\
\sum_{j=1}^{N_c} F_{R,c}(2,j) F_{D,c}(2,j) \\
\vdots \\
\sum_{j=1}^{N_c} F_{R,c}(n,j) F_{D,c}(n,j)
\end{bmatrix}
\]

(4)

The LS optimized coefficients for each class are then stored in a
look-up table (LUT) for future use.

![Diagram of filtering process](image)

Fig. 2. The filtering process of the trained filters.

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 inter-class differences are
ideally much larger than the intra-class variations. In theory,
the class number can be huge, considering \( N \) number of pixels
that 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:

\[
ADRC(x_i) = \begin{cases} 
0, & \text{if } V(x_i) \leq V_{av} \\
1, & \text{otherwise,}
\end{cases}
\]

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. Figure 3 shows a diagram of the ADRC code on a 3x3 block.

![Fig. 3. The ADRC code of a 3x3 block.](image)

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:

\[
H = -\sum_{i=1}^{N} P_R(i) \log_2 P_R(i)
\]

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:

\[
MAG = \frac{1}{N-1} \sum_{i=1}^{N-1} |F(0) - F(i)|
\]

where \( F(i) \) denotes the intensity value of a pixel in a region, \( F(0) \) is the intensity of the pixel in the centre, and \( N \) is the number of pixels in the region. Standard Deviation (STD) is employed as another complexity metric for a local region in [15].

Figure 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.

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

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 Figure 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, 2 more bits are used for representing the complexity of a region. So, in total 14 bits are used for classification. The 4 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, intra-frame processing is more cost-effective than inter-frame 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 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 5x5. 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 the next section.

For objective evaluation, we calculate the mean square error (MSE) between the target sequences and the result sequences processed on the source sequences. Figure 6 depicts the snapshots of the six test sequences we used for experiments. All the test sequences are excluded from the training set. Table 1 shows the MSE scores of different classification methods on different sequences.

From Table 1, 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 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. Figure 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 2 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 increases 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.
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.ijg.org), because sufficient coding artifacts are visible and image details are not completely lost using that quality level.

Table 3 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.

For testing the usefulness of structure information, Table 4 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.

Table 2
MSE scores for sharpness enhancement for sequences blurred with different Gaussian filters

<table>
<thead>
<tr>
<th>Sequence/Blur</th>
<th>Trained Filters</th>
<th>LT1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bicycle/σ=0.8</td>
<td>50.73</td>
<td>27.25</td>
</tr>
<tr>
<td>Bicycle/σ=0.9</td>
<td>15.99</td>
<td>33.71</td>
</tr>
<tr>
<td>Bicycle/σ=1.0</td>
<td>13.31</td>
<td>43.45</td>
</tr>
<tr>
<td>Bicycle/σ=1.1</td>
<td>24.20</td>
<td>55.30</td>
</tr>
<tr>
<td>Bicycle/σ=1.2</td>
<td>42.96</td>
<td>68.16</td>
</tr>
<tr>
<td>Football/σ=0.8</td>
<td>44.58</td>
<td>30.96</td>
</tr>
<tr>
<td>Football/σ=0.9</td>
<td>17.60</td>
<td>37.74</td>
</tr>
<tr>
<td>Football/σ=1.0</td>
<td>14.49</td>
<td>46.67</td>
</tr>
<tr>
<td>Football/σ=1.1</td>
<td>22.55</td>
<td>56.96</td>
</tr>
<tr>
<td>Football/σ=1.2</td>
<td>37.48</td>
<td>67.85</td>
</tr>
<tr>
<td>Girlsea/σ=0.8</td>
<td>10.08</td>
<td>7.40</td>
</tr>
<tr>
<td>Girlsea/σ=0.9</td>
<td>5.02</td>
<td>8.17</td>
</tr>
<tr>
<td>Girlsea/σ=1.0</td>
<td>4.40</td>
<td>9.76</td>
</tr>
<tr>
<td>Girlsea/σ=1.1</td>
<td>5.94</td>
<td>11.91</td>
</tr>
<tr>
<td>Girlsea/σ=1.2</td>
<td>8.81</td>
<td>14.32</td>
</tr>
</tbody>
</table>

Table 4
MSE scores for coding artifact reduction only using complexity

<table>
<thead>
<tr>
<th></th>
<th>DR</th>
<th>MAG</th>
<th>STD</th>
<th>Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bicycle</td>
<td>53.86</td>
<td>53.38</td>
<td>52.26</td>
<td>54.41</td>
</tr>
<tr>
<td>Football</td>
<td>49.15</td>
<td>48.93</td>
<td>48.34</td>
<td>49.64</td>
</tr>
<tr>
<td>Girlsea</td>
<td>19.18</td>
<td>19.18</td>
<td>18.84</td>
<td>19.60</td>
</tr>
<tr>
<td>Teeny</td>
<td>19.48</td>
<td>19.50</td>
<td>19.18</td>
<td>20.06</td>
</tr>
<tr>
<td>Wheel</td>
<td>44.04</td>
<td>43.54</td>
<td>41.82</td>
<td>45.03</td>
</tr>
<tr>
<td>Yvonne</td>
<td>39.27</td>
<td>39.28</td>
<td>38.36</td>
<td>40.09</td>
</tr>
</tbody>
</table>

Fig. 8. (a) Image with coding artifacts, (b) output after coding artifact reduction.
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 5, 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 [18, 1] 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.

<table>
<thead>
<tr>
<th>Sequence/JPEG Quality</th>
<th>Trained Filters</th>
<th>Ref [18]</th>
<th>Ref [1]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bicycle/10</td>
<td>80.35</td>
<td>102.43</td>
<td>110.67</td>
</tr>
<tr>
<td>Bicycle/15</td>
<td>50.74</td>
<td>73.58</td>
<td>75.28</td>
</tr>
<tr>
<td>Bicycle/20</td>
<td>41.48</td>
<td>59.82</td>
<td>58.29</td>
</tr>
<tr>
<td>Bicycle/25</td>
<td>29.96</td>
<td>51.26</td>
<td>47.68</td>
</tr>
<tr>
<td>Bicycle/30</td>
<td>25.14</td>
<td>45.36</td>
<td>40.36</td>
</tr>
<tr>
<td>Football/10</td>
<td>78.12</td>
<td>90.70</td>
<td>97.22</td>
</tr>
<tr>
<td>Football/15</td>
<td>53.58</td>
<td>66.96</td>
<td>68.04</td>
</tr>
<tr>
<td>Football/20</td>
<td>44.30</td>
<td>55.11</td>
<td>53.42</td>
</tr>
<tr>
<td>Football/25</td>
<td>35.61</td>
<td>47.93</td>
<td>44.34</td>
</tr>
<tr>
<td>Football/30</td>
<td>31.01</td>
<td>43.01</td>
<td>38.20</td>
</tr>
<tr>
<td>Girlsea/10</td>
<td>38.35</td>
<td>40.08</td>
<td>44.87</td>
</tr>
<tr>
<td>Girlsea/15</td>
<td>23.19</td>
<td>25.58</td>
<td>28.15</td>
</tr>
<tr>
<td>Girlsea/20</td>
<td>17.81</td>
<td>19.08</td>
<td>20.46</td>
</tr>
<tr>
<td>Girlsea/25</td>
<td>13.57</td>
<td>15.74</td>
<td>16.24</td>
</tr>
<tr>
<td>Girlsea/30</td>
<td>11.55</td>
<td>13.55</td>
<td>13.55</td>
</tr>
</tbody>
</table>

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. Figure 9 illustrates the concept. The trained filters function as non-linear 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 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 6 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 4.2.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>MSE Scores for Resolution Up-scaling Using ADRC and Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bicycle</td>
<td>ADRC</td>
</tr>
<tr>
<td>Bicycle</td>
<td>67.98</td>
</tr>
<tr>
<td>Football</td>
<td>68.22</td>
</tr>
<tr>
<td>Teeny</td>
<td>17.66</td>
</tr>
<tr>
<td>Wheel</td>
<td>43.98</td>
</tr>
<tr>
<td>Yvonne</td>
<td>47.25</td>
</tr>
</tbody>
</table>

Figure 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 upscaling is critical for trained filters, the scaling factor should be pre-defined 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, the trained filters can be combined with other scaling methods, such as bi-linear interpolation, to make it more flexible.

Table 7 shows the MSE scores for resolution up-scaling of different methods.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Trained Filters</th>
<th>Bi-cubic</th>
<th>NEDI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bicycle</td>
<td>67.98</td>
<td>76.54</td>
<td>72.30</td>
</tr>
<tr>
<td>Football</td>
<td>68.22</td>
<td>75.69</td>
<td>73.05</td>
</tr>
<tr>
<td>Girlsea</td>
<td>14.26</td>
<td>17.45</td>
<td>18.02</td>
</tr>
<tr>
<td>Teeny</td>
<td>17.66</td>
<td>21.50</td>
<td>22.88</td>
</tr>
<tr>
<td>Wheel</td>
<td>43.98</td>
<td>53.83</td>
<td>48.61</td>
</tr>
<tr>
<td>Yvonne</td>
<td>47.25</td>
<td>56.20</td>
<td>52.86</td>
</tr>
</tbody>
</table>

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 7 shows that trained filters outperform the other two methods.

![Fig. 10. (a) Downscaled image, (b) Output after resolution up-conversion.](image)

**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 Least Squares 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 3, 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 3, 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 3.

**A. Integrated Sharpening and Coding Artifact Reduction**

In the video chain, sharpness enhancement is usually applied after coding artifact reduction, because the coding artifacts become 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 8 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].

### Table 8

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

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 upscaling. During the training process of the integrated coding artifact reduction and resolution up-conversion, the degradation is first downscaling then compression.

Table 9 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 6, one can see that the 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.

### Table 9

<table>
<thead>
<tr>
<th></th>
<th>Cascaded</th>
<th>Integrated (ADRC+STD)</th>
<th>Integrated (STD)</th>
<th>Ref [20]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bicycle</td>
<td>175.79</td>
<td>173.74</td>
<td>173.74</td>
<td></td>
</tr>
<tr>
<td>Football</td>
<td>152.42</td>
<td>152.17</td>
<td>152.17</td>
<td></td>
</tr>
<tr>
<td>Girlsea</td>
<td>56.24</td>
<td>56.72</td>
<td>56.72</td>
<td></td>
</tr>
<tr>
<td>Teeny</td>
<td>58.35</td>
<td>58.44</td>
<td>58.44</td>
<td></td>
</tr>
<tr>
<td>Wheel</td>
<td>144.57</td>
<td>141.27</td>
<td>141.27</td>
<td></td>
</tr>
<tr>
<td>Yvonne</td>
<td>130.33</td>
<td>130.49</td>
<td>130.49</td>
<td></td>
</tr>
</tbody>
</table>

Figure 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.

### C. Integrated Sharpening and Resolution Up-conversion

In the video processing chain, sharpness enhancement is usually applied after resolution up-conversion. During the training process of the integrated sharpening and resolution up-conversion, the degradation is first Gaussian blur then linear downscaling.

Table 10 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.

Figure 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.

### TABLE 10

<table>
<thead>
<tr>
<th></th>
<th>Cascaded</th>
<th>Integrated (ADRC+STD)</th>
<th>Ref[4]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bicycle</td>
<td>425.77</td>
<td>69.22</td>
<td>81.60</td>
</tr>
<tr>
<td>Football</td>
<td>362.50</td>
<td>69.83</td>
<td>78.69</td>
</tr>
<tr>
<td>Girlsea</td>
<td>115.23</td>
<td>15.32</td>
<td>19.59</td>
</tr>
<tr>
<td>Teeny</td>
<td>140.35</td>
<td>19.24</td>
<td>25.51</td>
</tr>
<tr>
<td>Wheel</td>
<td>229.33</td>
<td>40.85</td>
<td>51.13</td>
</tr>
<tr>
<td>Yvonne</td>
<td>244.59</td>
<td>45.24</td>
<td>60.69</td>
</tr>
</tbody>
</table>

**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 3.3. 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 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 Least Squares 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