Content-adaptive up-scaling of chrominance using classification of luminance and chrominance data

M. Zhao\textsuperscript{a}, P. M. Hofman\textsuperscript{b} and G. de Haan\textsuperscript{a,\textit{b}}

\textsuperscript{a}Eindhoven University of Technology, Eindhoven, Den Dolech 2, 5600 MB Eindhoven, The Netherlands
\textsuperscript{b}Philips Research Laboratories Eindhoven, Prof. Holstlaan 4, (WO01), 5656 AA Eindhoven, The Netherlands

ABSTRACT

In a Standard Definition (SD) television system, the Y:U:V video format 4:2:2 with chrominance sub-sampling is widely used. With the advent of High Definition (HD) television, the 4:4:4 format is required for high-performance TV. High-quality up-sampling methods have been developed to perform a resolution conversion from SD signal to HD signal\textsuperscript{1}.\textsuperscript{2} Although these algorithms have been designed for spatial scaling of luminance, they may be adapted and used to up-sample the low-resolution components U,V (4:2:2) to a high-resolution UV-colour format (4:4:4). In this paper, a content-adaptive up-scaling method for chrominance is proposed, with interpolation filters that adapt to the local structure of both luminance and chrominance data. Optimal filters were computed from a large video data set in different colour formats, such that original high-resolution data in a 4:4:4 format was reconstructed from low-resolution colour data, on the basis of the Least Mean Square (LMS) criterion. By combining edge information of both luminance and chrominance, the edge in the chrominance signal can be detected more accurately, thus exploiting the wider bandwidth of the luminance signal.

Keywords: Colour format, Up-scaling, HDTV

1. INTRODUCTION

In video engineering, the chrominance signal is commonly sub-sampled compared to the luminance signal to reduce transmission bandwidth or storage space. This type of signal format makes use of the fact that the human visual system is more sensitive to spatial variations in the luminance than in the chrominance. Different Y:U:V signal formats, such as 4:1:1, 4:2:0 and 4:2:2, correspond to different levels of chrominance (UV) sub-sampling.

With the emergence of High Definition (HD) television, high-quality spatial video up-conversion techniques have been developed to portray Standard Definition (SD) signals on a HD display, with improved sharp luminance details. At present, various television manufacturers therefore incorporate a technology for up-scaling of the luminance signal in high-end television sets. These technologies focus on luminance, although additional image quality may be gained by applying similar methods to the chrominance components, U and V.

For chrominance up-scaling, we will first consider up-scaling of the sub-sampled chrominance signal to the 4:4:4 format that represents chrominance and luminance at the same resolution. Traditional linear interpolation methods that are typically used do not increase the sharpness of the edges in the chrominance signal.

The current study proposes a variant of a (non-linear) classification-based up-scaling method that has demonstrated a good price-performance ratio\textsuperscript{1}.\textsuperscript{2} The main idea of the classification based video
up-conversion technique is to define classes of local image characteristics, and to apply the same interpolation filter to all image blocks that correspond to a single class. Although similar techniques can be used for chrominance and luminance independently, it is proposed to up-scale the chrominance data using classes derived from both chrominance and luminance data. By involving the luminance component, it is expected that chrominance edges are detected more precisely, and that consistency between luminance and chrominance transients is improved.

The paper is organized as follows. In Section 2, a traditional linear up-scaling method is described. In Section 3, we will propose the new up-scaling method by classifying the luminance and chrominance content of the image. The evaluation and comparison of both up-scaling methods are presented in Section 4. Finally, in Section 5, we draw the conclusion.

2. LINEAR UP-SCALING

In this study, a linear filter was designed for the down-scaling of the 4:4:4 format to 4:2:2, according to the ITU requirements described in document Rec. ITU-R BT.601.\(^5\) The result was a center-symmetric 27-taps FIR filter. The non-zero coefficients, \(b_i\), are shown below.

\[
\begin{array}{ccccccccccccccc}
\end{array}
\]

**Figure 1.** Non-zero coefficients of the linear filter that was used in this study. This filter is compliant with ITU requirements.\(^5\)

![Figure 2](image_url)

**Figure 2.** The frequency response that corresponds to the filter coefficients in Figure 1 (thick red line). The straight thin lines mark the ITU-recommended limits in between which the filter response must pass. The connected straight lines at the left and right indicate lower and upper limits, respectively. Note, that the filter is skew-symmetrical and has a symmetrical pass-band at 3.375 MHz, at half the Nyquist frequency of 6.75 MHz.

To demonstrate that this filter meets the ITU requirements, the frequency response of this FIR filter is shown in Figure 2. This filter was also used for 4:2:2 to 4:4:4 up-sampling, as a reference to the new non-linear up-sampling method as proposed in this paper.
3. CLASSIFICATION BASED CHROMINANCE UP-SCALING

In a classification based up-scaling method, the momentary filter coefficients, during interpolation, depend on the local content of the image, which can be classified into classes based on the pattern of the block. Good examples of such methods are commercially named DRC\textsuperscript{1} and resolution synthesis.\textsuperscript{2} We choose DRC for its easy implementation and good performance. For classification, various data compression techniques can be used. In this paper, we applied ADRC (Adaptive Dynamic Range Coding),\textsuperscript{3} which reduces the class data to just a single bit per pixel in the aperture. To obtain the filter coefficients, a learning process is performed in advance. The learning process employs both the 4:4:4 (YUV) format video signal and the its sub-sampled version as the training material and uses the Least Mean Square (LMS) algorithm to obtain the optimal filter coefficients for each class. The training process is computational intensive due to the large number of classes. The adaptive interpolation coefficients will be obtained after the training and stored in a Look-Up-Table (LUT). Fortunately it can be done off-line and needs to be performed only once.

3.1. Aperture Definition

The proposed aperture of the filter depends on the input video format. For a 4:2:2, or a 4:1:1, (YUV) format it makes sense to use a horizontal up-scaling filter, since the horizontal resolution of the luminance is higher than that of the chrominance. For a 4:2:0 (YUV) signal, a 2-D aperture seems appropriate as also the vertical luminance resolution is higher than that of the chrominance. In all cases, a classical DRC-like up-scaling of the chrominance using chrominance data only can follow the proposed processing. Thus, for a 2-D up-scaling to HDTV 4:4:4 format of an SD 4:2:2 format, we would propose to perform a purely horizontal up-scaling, resulting in a 4:4:4 SD format followed by a classical DRC-like up-scaling to arrive at the HD 4:4:4 format.

![Figure 3](image)

**Figure 3.** (A). The ten chrominance-only 1D aperture, named C1. (B). The six chrominance-1D and six luminance-1D aperture, named C1Y1. (C). The ten chrominance-2D and four luminance-1D aperture, named C2Y1. (D). The six chrominance-1D and eight luminance-2D aperture, named C1Y2. For example, C1Y2 indicates a one-dimensional chrominance and two-dimensional luminance classification aperture. In contrast to the classification aperture which can depend on luminance also, the filter aperture is always composed of the chrominance samples only. The center pixels, indicated by a box, represent pixel positions of the up-converted chrominance data.

For up-scaling, we first define a filter aperture and a content classification aperture. Those two apertures can be the same or can be different from each other in the sense that the filter aperture acts on pixel samples used for interpolation only, whereas the content aperture may also contain other data.
for classification. Because the number of classes increases exponentially with the number of pixels in the classification aperture, there is always a tradeoff between aperture size and number of classes.

In this paper, we will show our investigation on 4:2:2 to 4:4:4 up-conversion. As a reference, we start with an aperture with pure chrominance data. Taking into account the number of classes, we first chose ten horizontal pixels that compose of five U and five V data (Aperture A in Figure 3). After this, the luminance data is included. Various combinations of luminance and chrominance data are shown in Figure 3. With the detailed edge information in the luminance data, we can improve the classification of the image content which will lead to more accurate interpolation. From a theoretical point of view, it makes no difference to include the luminance data for filtering or not. However, to reduce the complexity of the interpolation, it is straightforward to include the luminance data only in the classification and use chrominance data for both classification and interpolation.

3.2. ADRC Coding

In this study, filters are computed for each category of image blocks, rather than for each image block individually. Categories were created by 1-bit quantization, which creates a limited number of categories compared to 8-bit quantization, for example. With Adaptive Dynamic Range Coding (ADRC)\(^3\) each pixel can be encoded into one bit, using:

\[
Q = \begin{cases} 
0 & : F_{SD} < F_{AV} \\
1 & : F_{SD} \geq F_{AV}
\end{cases}
\]

Here \(F_{SD}\) is the luminance or chrominance value of the SD pixel and \(F_{AV}\) is the average luminance/chrominance value of the pixels in the current aperture. \(Q\) is the ADRC encoding result. This encoding scheme simply encode current pixel into 1 if its value is above the average value of the pixels in the classification aperture otherwise it will be encoded to 0.

Since the average luminance value and the average chrominance value are not correlated in the same sense as neighbouring luminance pixels, the calculation of the classes needs to be modified. We concatenate the individual class bits of all components (Y/U/V) to form a single class-index that is the entry to the filter coefficients Look-Up Table (LUT). Taking aperture B in Figure 3 as an example, the class index is generated as follows shown in Figure 4.

![Figure 4. Class index is a concatenation of the Q values of Y, U and V components in the current classification aperture. In this example image block, it is 001100 − 001 = 011.](image)

ADRC encoding with more bits per pixel can be thought of in a similar way. But with more bits ADRC encoding, the number of pixels in the aperture has to be constrained to make the system...
affordable. With the classification scheme defined above, there is no chance to further reduce pixels in the classification aperture.

3.3. Class Inversion

It has been shown before\(^4\) that for luminance up-scaling, if we invert the picture data, the coefficients in the LUT should remain the same. Consequently, any binary class (with class code CLSS) and its inverted version (with class code CLMAX-CLSS, CLMAX indicates the total number of classes) should yield the same interpolation coefficients. By combining the two complementary classes, the size of the LUT reduces with a factor of two without any loss of image quality. It seems reasonable that this holds equally well for the chrominance data U and V. Therefore, calculating the ADRC-coefficients independently for the luminance and chrominance values has an attractive side-effect. Since we propose to apply ADRC coding to Y, U and V data independently, we end up with half the number of classes for each component, i.e. we save 3-bits, or a factor of 8, in the address-space of the LUT.

3.4. Least Mean Square Algorithm

To clarify the use of the LMS algorithm used in the training process within each class, let \(F_{HD}\) be the chrominance value of the original (not the up-converted) HD pixels and \(F_{HI}\) be the value of the interpolated ones, which is a weighted sum of the SD pixels \(F_{SD}\) in the interpolation aperture. Here, \(F_{SD}\) can be luminance or chrominance values. Suppose we choose aperture A in Figure 3 as the interpolation aperture, the equation to interpolate a chrominance value of the pixel is:

\[
F_{HI} = \sum_{k=0}^{9} w_{k,c} F_{SD}(k)
\]

where \(w_{k,c}\) are weights for class \(c\). The squared error for one class over a large number of images is:

\[
e^2 = \sum_i (F_{HD}(i) - F_{HI}(i))^2
\]

In the training process, one class contains in total a number of \(t\) samples. The \(p^{th}\) HD sample in class \(c\) of sample U is interpolated as:

\[
F_{HI,p} = \sum_{k=0}^{9} w_{k,c} F_{SD,p}(k) \quad (p = 1, 2, ..., t)
\]

The interpolation error of this \(p^{th}\) interpolation sample is:

\[
e_{p,c} = F_{HD,p} - F_{HI,p} = F_{HD,p} - \sum_{k=0}^{9} w_{k,c} F_{SD,p}(k) \quad (p = 1, 2, ..., t)
\]

Consequently, the total squared error of this class can be expressed as:

\[
e^2_c = \sum_{p=1}^{t} e^2_{p,c}
\]

To find the minimum, we calculate the first derivative of \(e^2_c\) to each \(w_{k,c}\):

\[
\frac{\partial e^2_c}{\partial w_{k,c}} = \sum_{p=1}^{t} 2(e_{p,c}) \frac{\partial e_{p,c}}{\partial w_{k,c}} = -\sum_{p=1}^{t} 2F_{SD,p}(k)e_{p,c} \quad (k = 0, 1, ..., 9)
\]
The minimum occurs when the first derivative is zero.

To simplify the expression to calculate the minimum, let:

\[ X_{k,l} = \sum_{p=1}^{t} F_{SD,p}(k) \cdot F_{SD,p}(l) \quad (k, l = 0, 1, \ldots, 9) \]  

and:

\[ Y_{k,l} = \sum_{p=1}^{t} F_{SD,p}(k) \cdot F_{HD,p} \quad (k = 0, 1, \ldots, 9) \]  

Now the coefficients \( w_{kl,c} \) for class \( c \) are obtained by solving:

\[
\begin{bmatrix}
X_{00} & X_{01} & \cdots & X_{09} \\
X_{10} & X_{11} & \cdots & X_{19} \\
X_{20} & X_{21} & \cdots & X_{29} \\
\vdots & \vdots & \ddots & \vdots \\
X_{90} & X_{91} & \cdots & X_{99}
\end{bmatrix}
\begin{bmatrix}
w_{0,c} \\
w_{1,c} \\
w_{2,c} \\
\vdots \\
w_{9,c}
\end{bmatrix}
= 
\begin{bmatrix}
Y_{0} \\
Y_{1} \\
Y_{2} \\
\vdots \\
Y_{9}
\end{bmatrix}
\]  

(10)

With all the filter coefficients calculated and stored in a Look-Up Table (LUT) in advance during training, the interpolation becomes very simple, using equation (2). \( w_{k,c} \) is looked up in the LUT, that is addressed by the ADRC coded representation of the pixels in the classification aperture.

4. EVALUATION

We first down-sample a 4:4:4 format video signal to a 4:2:2 format and then up-convert it to 4:4:4 format using our proposed algorithm and linear up-conversion respectively. Then the MSE between chrominance signal of the original images and up-converted ones was calculated. The procedure for this evaluation is shown in Figure 5. We could compare the result of each aperture using the combination of objective MSE criterion and subjective evaluation.

![Figure 5. Evaluation Flowchart](image)

4.1. Test images

Test images used in the evaluation include natural images and synthetic images. The natural images that are used show the effectiveness of the algorithm in normal applications. However, we should bear in mind that the improvement could only be achieved at edges in chrominance, which are not so dominant in natural images than we should expect and the advantage clearly shows in an MSE-score.
Synthetic images with abundant chrominance transitions are, therefore, made to more convincingly show the advantages of the algorithm. Shown in Figure 6. Image A and B are natural images with lots of details and vivid colour. Image A has a resolution of $2048 \times 1536$ and B is $1920 \times 1080$. Images C and D test images, both with a resolution of $720 \times 576$. The idea behind C is to create all possible types of vertical edges, including simple edges, double-step edges and thin lines of different width, with different combination of colours (R,G,B,C,M,Y). Image D is a colour zone plate with edges at different orientations and with different width. The equations below specify how the image has been computed. In short, a gray-scale zone plate with values between 0.0 and 1.0 was used as a weighting factor to interpolate between two RGB-triplets, $\vec{A} = (r_A, g_A, b_A)$ and $\vec{B} = (r_B, g_B, b_B)$, thus creating the colour zone plate, $\vec{C}(i, j)$:

$$\vec{C}(i, j) = w(i, j) \cdot \vec{A} + (1 - w(i, j)) \cdot \vec{B},$$

with

$$w(i, j) = \frac{1}{2} + \frac{1}{2K} \text{median} \left[ \sin \left( \frac{\pi \Delta}{2D} (i^2 + j^2) \right), -d, +d \right]$$

with $(i, j)$ the image indices, $W(i, j)$ the local weight at $(i, j)$, $\Delta$ the sampling interval of the zone plate, $D = 500/\Delta$, and a scaling factor. The factor $A$ defines a clipping amplitude for the sine-based zone plate, thus creating sharp edges of varying widths, where clipping values $-d$ and $+d$ correspond
to $\vec{B}$ and $\vec{A}$, respectively. Typical values were $K = 0.5$ and $d = 0.25$. An example of the test image is shown in Figure 6 D.

<table>
<thead>
<tr>
<th></th>
<th>MSE</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>27-tap filter</td>
<td>[1 2 1] filter</td>
<td>C1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Image A</td>
<td>0.79</td>
<td>0.77</td>
<td>0.88</td>
<td>0.85</td>
<td>1.53</td>
</tr>
<tr>
<td>Image B</td>
<td>4.69</td>
<td>1.85</td>
<td>5.61</td>
<td>2.16</td>
<td>6.05</td>
</tr>
<tr>
<td>Image C</td>
<td>116.48</td>
<td>109.46</td>
<td>154.95</td>
<td>146.63</td>
<td>102.94</td>
</tr>
<tr>
<td>Image D</td>
<td>38.49</td>
<td>47.79</td>
<td>62.35</td>
<td>80.67</td>
<td>33.12</td>
</tr>
<tr>
<td>Average</td>
<td>40.11</td>
<td>39.96</td>
<td>55.95</td>
<td>57.58</td>
<td>35.91</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>MSE</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C1Y1</td>
<td>C2Y1</td>
<td>C1Y2</td>
<td></td>
</tr>
<tr>
<td>Image A</td>
<td>0.67</td>
<td>0.67</td>
<td>0.71</td>
<td>0.71</td>
</tr>
<tr>
<td>Image B</td>
<td>5.27</td>
<td>1.89</td>
<td>5.68</td>
<td>2.04</td>
</tr>
<tr>
<td>Image C</td>
<td>81.32</td>
<td>78.18</td>
<td>77.68</td>
<td>70.00</td>
</tr>
<tr>
<td>Image D</td>
<td>30.72</td>
<td>38.56</td>
<td>30.07</td>
<td>38.54</td>
</tr>
<tr>
<td>Average</td>
<td>29.50</td>
<td>29.83</td>
<td>28.53</td>
<td>27.82</td>
</tr>
</tbody>
</table>

Table 1. MSE analysis result

4.2. Objective comparison – $MSE$ analysis

A $MSE$ analysis based on the four test images with different up-scaling methods was performed. The methods we evaluated include the 27-tap interpolation FIR filter, the simple [1 2 1] FIR interpolation filter and our proposed classification-based method with various apertures shown in Figure 3.

From Table 1, we can clearly see that the introduction of the luminance signal in the classification improves the up-scaling quality compared to the traditional linear up-scaling method, especially for our test images. For natural images, the difference is rather small. This is due to the fact that the differences only occur at transitions in edges, which are a minor part of the natural image. In average, the $MSE$ dropped about 30% in both U and V components, taking the traditional linear method as the reference.

4.3. Subjective comparison

Subjective comparison for reader is important in video engineering because the objective metric $MSE$ is not always reliable. To better depict the differences between all methods, we show the chrominance components in the luminance channel, which will help to perceive the detailed differences.

As we can see in Figure 7, the 27-tap low-pass interpolation FIR filter performs well in most areas and gives clear details. However, it brings strong ringing at the sharp edges, which can be very annoying for the perceived image quality. The 3-tap FIR filter corrected this problem to some extent, but makes the image staircases and blurred in the detailed areas like the parallel lines in the bottom-left part of the image blocks (area $N$) in Figure 7. With our proposed classification-based method, the up-converted image keeps the sharpness without generating any ringing. This can be observed mainly in the edge transition area in the top part of the image blocks (area $M$) in Figure 7. To further illustrate the advantage of using luminance data during classification, we show another detailed area in our test image in Figure 8. Within this image, it is clear that with the help of luminance data in the classification, we will obtain sharper chrominance edges in the up-converted image.
Figure 7. Comparison of different up-scaling method. Here we only show the U component, the V component performs in the same way. Top-left is the Original, (A). up-converted using 27-tap FIR filter, (B). 3-tap FIR filter, (C). classification-based method using aperture A shown in Figure 3, (D). classification-based method using aperture D shown in Figure 3.

Figure 8. Comparison of using luminance data in classification. TOP: U component. BOTTOM: Waveform of cross-section in U component at the position indicated by the white line. Left is the Original, (A) up-converted with 27-tap FIR filter, (B) up-converted with classification-based method using aperture A shown in Figure 3, (C) classification-based method using aperture D shown in Figure 3. The V component performs in the same way.
5. CONCLUSION

We conclude that a classification based chrominance up-conversion using luminance information is a very effective method for video chrominance up-conversion. With this method, we can up-convert the chrominance components in the video signal with clear and distinct colour at the edges. Classification using a 2-D luminance aperture will make better use of edge information in the luminance signal. Compared with the traditional linear method, this new algorithm has a very good performance. It gives distinct colour transitions without introducing any overshoots. The algorithm uses a Look-Up-Table (LUT) for obtaining interpolation coefficients. A simple threshold operation generates the address for the LUT, and the pixels involved in interpolation is limited to 6 chrominance data. All this will yield an easy hardware implementation with low cost.

REFERENCES