MASTER

Resolution enhancement techniques for spatially layered video compression

van der Meer, P.

Award date:
2004

Link to publication

Disclaimer
This document contains a student thesis (bachelor’s or master’s), as authored by a student at Eindhoven University of Technology. Student theses are made available in the TU/e repository upon obtaining the required degree. The grade received is not published on the document as presented in the repository. The required complexity or quality of research of student theses may vary by program, and the required minimum study period may vary in duration.

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain.
RESOLUTION ENHANCEMENT TECHNIQUES FOR SPATIALLY LAYERED VIDEO COMPRESSION.

P. van der Meer
Abstract

This report describes a study on the usage of non-linear resolution enhancement for spatial scalable video coding. Spatial scalability allows reconstruction of video sequences different spatial resolutions from a single compressed bit stream. We focused on the introduction of two resolution enhancement algorithms, namely DRC by Sony and PixelPlus by Philips, into an MJPEG-based video coding scheme. The report discusses the methodology employed for the benchmarking tests, provides objective quality metrics, and gives conclusions of the subjective evaluations. The drawbacks of each of the evaluated schemes are investigated and methods to improve their efficiency are proposed and evaluated in short.
Contents

1 Introduction ............................................. 5
   1.1 Video resolutions .................................. 6
   1.2 Video compression basics ............................ 7
   1.3 Video coding standards .............................. 8
   1.4 Image scaling ....................................... 9
   1.5 Goal and outline of this thesis .................... 10

2 State of the art spatially layered video compression .... 11
   2.1 Linear scaling filters ............................... 11
      2.1.1 Frequency response .............................. 11
      2.1.2 Alias ........................................ 13
      2.1.3 Overshoot and undershoot ....................... 13
      2.1.4 Staircasing ................................... 13
      2.1.5 Further side-effects of separable filtering .... 15
      2.1.6 Simple non-separable filter .................... 15
   2.2 Spatially layered compression schemes ............... 18
      2.2.1 Spatial scalability in MPEG-2 video .......... 18
      2.2.2 DemoGraFX dual-layer codec .................. 19
      2.2.3 MPEG-based dual-layer codec ................. 19

3 Dual layer codec with non-linear resolution enhancement .. 21
   3.1 Resolution enhancement ............................. 21
      3.1.1 PixelPlus ..................................... 22
      3.1.2 Linear up-scaling as a reference ............... 23
      3.1.3 DRC .......................................... 24
      3.1.4 Separable linear up scaler and non-linear post filter 25
   3.2 Dual-layer codec framework ......................... 26
   3.3 The base layer ...................................... 29
      3.3.1 Resolution enhancement pre-filter ............. 29
      3.3.2 The base layer with separable linear up-scaling 30
### 3.3.3 The base layer with DRC

30

### 3.3.4 The base layer with FIR and non-linear post-filter

32

### 3.4 The enhancement layer

32

#### 3.4.1 Noise filtering and deflation as enhancement layer compression

33

#### 3.4.2 Block classification and jumping

34

#### 3.4.3 Modulation of high-frequent blocks

35

#### 3.4.4 Bit rate-adaptive quantization matrix

36

### 4 Evaluation of non-linear schemes

38

#### 4.1 Image quality metrics

38

##### 4.1.1 Peak Signal-to-Noise Ratio

39

##### 4.1.2 Universal Quality Metric

39

##### 4.1.3 Blockiness Impairment Metric

40

##### 4.1.4 Sharpness metric

41

#### 4.2 Rate-distortion plane measurements

42

#### 4.3 Test Sequences

44

#### 4.4 PixelPlus resolution enhancement

47

#### 4.5 Proposed dual-layer codecs with factor 2 scaling

47

##### 4.5.1 Reference codec with linear scaling

48

##### 4.5.2 The DRC-based scheme

50

#### 4.6 Proposed dual-layer codecs with rational scaling factor

55

##### 4.6.1 DRC-FIR hybrid scheme

55

#### 4.7 Proposed dual-layer codec vs. single-layer MJPEG

56

### 5 Conclusions and recommendations

60

#### 5.1 Conclusions

60

#### 5.2 Recommendations

61

### A Rate-distorsion planes

66

#### A.1 Linear up-scaling

67

#### A.2 DRC resolution enhancement

70

#### A.3 Hybrid resolution enhancement

73

### B Literature research

76

#### B.1 Project description

76

##### B.1.1 Project assignment

76

##### B.1.2 Literature research assignment

76

#### B.2 Conceptual table of contents of final report

77

#### B.3 List of search terms

78

#### B.4 Used sources

78
B.5 References .................................................. 79
  B.5.1 Selected references per source ...................... 79
  B.5.2 Selection criteria ...................................... 79
B.6 The snowball method ...................................... 79
B.7 The citation method ...................................... 80
B.8 Relation literature and contents of graduation report 82
B.9 Conclusion and recommendations ........................ 82
B.10 References .................................................... 83
Preface

This document is my Masters thesis and it is the product of a graduation project, leading to a Master of Science degree in Electrical Engineering from the Eindhoven University of Technology. This project was carried out in the Video Processing and Visual Perception (VIPs) group at Philips Research in Eindhoven.

The report assumes some knowledge of video compression. I recommend books by Mitchell [1] [2] for an introduction to this field.

The subject of this research has a very broad scope, and during my work at Philips I was able to only focus on small parts. Oleg Belik explored other related areas, and I recommend also taking a look at a summary of his work [3].

I would like to thank Philips Research for offering me the possibility to carry out the work. I must thank all of the VIPs group and especially Frits de Bruijn and Gerard de Haan for their guidance and patience, Nico Cordes, Arjan Dommisse, Michiel Klompenhouwer, Frank van Heesch, Ihor Kirenko, Franco Oberti, Michiel Oostindi, Hans Puttenstein, Jeroen Tegenbosch, Jeroen Theelen, Rene van der Vleuten and Meng Zhao for their help and input.
Chapter 1

Introduction

This report deals with digital video compression. More specifically, it is about a scalable form of video compression. In this case, scalability refers to the splitting of a video signal into various complementary video signals. Hence, various levels of quality may be provided by using just one video signal.

The applications of scalable video compression are:

- Backwards compatibility, because it allows playback on both high-end and low-end devices. The benefit is that it typically requires less bandwidth than using separate streams for each device.

- Error resilience, because transmission of scalable compressed video takes place using multiple layers. Each layer may be assigned a specific level of reliability. This can be used to guarantee the display of the most important features of the material.

In essence, the layered compression scheme encodes two types of data:

- A base layer, containing the most important features.

- Enhancement layers, containing the less important features.

There are numerous variations of this concept. For instance, temporal scalable coding where signals of different frame rates can be extracted from one stream. Another example is spatial scalable coding or spatially layered coding where one stream is used to extract signals of different spatial resolutions. The latter is the main concept discussed in this report.

Spatially layered encoding utilizes spatial scaling in order to encode the base layer at a low bit rate. One of the reasons for using image scaling is that
we reduce the amount of data in the image, while keeping the most important features. The Human Visual System (HVS) is more sensitive to low frequencies, so this is actually a reasonable way of lossy compression.

This scheme makes use of the techniques mentioned above.

- The encoder down-scales the input and compresses it to form a so-called base layer. The up-scaled base is subtracted from the input, which leaves us with the residual. This residual is then compressed into an enhancement layer.

- The decoder decompresses both layers and the up-scaled base and the enhancement are added to produce the end result.

The principle of a spatially layered coding is illustrated in Figure 1.1.

![Figure 1.1: Example of the principle of spatially layered compression](image)

### 1.1 Video resolutions

The following resolution formats are common in video compression:

- **Standard Definition (SD)** is currently the most commonly used television broadcast standard. Its resolutions are 720x576@50Hz (576 active lines) and 720x625@50Hz (576 active lines) as defined in [4]. Both are interlaced.

- **High Definition (HD)** is an emerging broadcast standard intended for high-end television. Some popular HD resolutions are 1440x1152 (high 1440 or QUAD SD, four times the size of an SD image), 1920x1080 and
1280x720. HD offers the free choice of interlaced or progressive mode for all its resolutions, however, 1920x1080 is commonly interlaced and 1280x720 is commonly progressive.

- Common Intermediate Format (CIF) has half the width and half the height of the SD format: 352x288 (progressive).
- Quarter CIF (QCIF) is again roughly half width and height compared to CIF: 176x144 (progressive).

SD is a standard defined by the International Telecommunication Union (ITU) as ITU-R Rec. BT.601-5 [4]. The most commonly used HD standard is defined by the Advanced Television Systems Committee (ATSC). CIF and QCIF are used primarily for mobile platforms and teleconferencing.

### 1.2 Video compression basics

The principle behind the most popular video compression techniques is the following: It is a scheme consisting of three basic steps.

1. Decorrelation, meaning the output lacks a form of correlation that was present in the input. This usually results in redundancies that are beneficial to compression.

2. Quantization, discarding the less important parts of the decorrelated data. This makes the compression lossy.

3. Entropy encoding, which reduces the redundancy of the quantized data.

![Figure 1.2: Blockdiagram of the principle of video coding](image)

As mentioned above this scheme is lossy, meaning it causes a loss of quality. Usually however, this results in a good compression ratio compared to lossless compression. The following techniques are based on this scheme:
• Discrete Cosine Transform (DCT) based compression. This is a scheme that first uses the DCT to achieve spatial decorrelation. The DCT domain is then quantized and is subsequently encoded with an entropy encoder. For instance a Variable Length Encoding (VLC) combined with a Run Length Encoding (RLE) is a commonly used realisation. This procedure is performed on a block wise basis.

• Motion compensated compression. This encodes the differences between consecutive frames to achieve temporal decorrelation. It is based on the concept of movement: a block is traced to a matching position in a reference frame, resulting in a motion vector. The motion vectors are encoded with an entropy encoder.

There are roughly two categories of frames in an encoded video stream: the intra frame and the inter frame. The intra frame uses DCT based compression exclusively, whereas the inter frame also uses motion compensation.

1.3 Video coding standards

The Joint Photographic Experts Group (JPEG) developed a standard for compression of still images using DCT compression. There is an extension of JPEG known as Motion JPEG (MJPEG), which enables the storage of multiple frames. For an introduction to JPEG and DCT based compression, we refer to Mitchell et al. [1].

The Motion Picture Experts Group (MPEG) developed a standard for compression of video material, primarily meant for storage purposes. The MPEG standards are also International Organisation for Standardization (ISO) standards, formalized by ISO International Electrotechnical Commission, Joint Technical Committee 1 (ISO/IEC JTC1). MPEG uses DCT compression, but with motion compensation. This is done by using inter frames in between the intra frames. An inter frame is typically a quarter or a third of the size of an intra frame. Because of this, MPEG can achieve a much higher compression ratio than is possible with MJPEG.

The first MPEG standard, MPEG-1, was used to encode progressive (non-interlaced) video and was limited to a maximum of SD resolution. MPEG-2 extends this to higher resolutions, such as HD mentioned above, adds scalability as used in the aforementioned schemes and supports interlaced formats. MPEG-1 and MPEG-2 and their differences are discussed by Mitchell et al. [2].
The H.261 thru H.264 standards originate from the International Telecommunication Union (ITU) Telecommunications Standardization Sector (ITU-T) and have a lot in common with the MPEG standards. Originally, the ITU and the ISO were competitors in the field of video compression and both organisations developed their own standards:

The ITU developed H.261 [5] as a standard meant primarily for teleconferencing, and appeared around the same time as the ISO standard MPEG-1 video. H.262, its follow-up, is identical to MPEG-2 video [6]. H.263 [7] is a continuation of H.262 aiming at higher compression ratios. H.264, or MPEG-4 Advanced Video Coding (AVC) [8], is currently still in development, but it is able to encode at the same quality as MPEG-2 using just half the data-rate. This is developed by the Joint Video Team (JVT): a cooperation between the ITU-T and the ISO/IEC JTC1.

1.4 Image scaling

The up-scaling and down-scaling used in spatial layering are commonly implemented using linear image scaling. This comprises of:

- up-sampling followed by interpolation filtering (up-scaling).
- low-pass filtering followed by down-sampling (down-scaling).

The low-pass filtering and interpolation are realised with a linear algorithm, typically a FIR filter.

Linear scaling introduces artifacts: unwanted artificially introduced components. With up-scaling the image loses its sharpness and there may appear 'staircases' around edges. In spatially layered compression this is a drawback. The more artifacts in the up-scaled base layer, the more enhancement layer data needs to be added in order to achieve good perceived quality.

Recently, non-linear resolution enhancement techniques have become available, significantly improving the quality of an up-scaled image. These may be used to replace the linear up-scaling in spatially layered compression.
1.5 Goal and outline of this thesis

At Philips, research in spatially layered compression has been conducted since the late 1980's. Now spatial scalability is offered as an integral part of existing video coding standards such as MPEG-2 and MPEG-4. Alternatives to these (relatively inefficient) schemes were offered by DemoGraFX [9] in the mid 1990's. The DemoGraFX dual layer codec encoded two levels of quality: HD and SD.

A similar concept is also used in the following scheme invented at Philips. Fons Bruls et al. introduced a dual layer codec that can represent HD content using less bandwidth than a single layer HD stream [10].

The goals of this project are:

- Finding options for improvement in spatial scalable compression, especially by using non-linear spatial algorithms. This means a reduction of the required bandwidth without loss of quality, or offering higher quality at the same bandwidth.

- The evaluation of these options by comparing their computational complexity with their benefits.

Chapter 2 discusses the current state of the art considering the dual layer coding scheme mentioned above.

In Chapter 3 methods are proposed for improvement in a dual layer scheme. Amongst these is the replacement of linear up-scaling with non-linear resolution enhancement. This results in various codec configurations which will be discussed in detail.

Subsequently these proposed schemes will be simulated and evaluated with respect to their costs and benefits. These benefits will be quantified with metrics that attempt to emulate aspects of the HVS. This is handled in Chapter 4.

Conclusions are drawn from the evaluations in Chapter 5. Recommendations on future improvements are provided as well.
Chapter 2

State of the art spatially layered video compression

This chapter provides an overview of currently available spatially layered video compression schemes. Additionally, a description of their backgrounds and history is presented. An overview of various linear scaling filters, some of which are used in the schemes described in this chapter, is provided as a background.

2.1 Linear scaling filters

During this project, numerous experiments with low-pass FIR filters for image scaling were conducted. Not all of these filters were eventually used in the proposed coding schemes in Chapter 3. However, the experiments have provided an overview of which filters are, or are not, suited for resolution enhancement.

2.1.1 Frequency response

The three most important FIR filters for down- and up-scaling in this report are:

- the 'Empress' filter, originating from the Bruls-method [17] for spatial layered compression, which has the following impulse response:

  \[ FIR = \{-3, 0, 19, 32, 19, 0, -3\} \]  

(2.1)
• the PixelPlus horizontal filter, which has the following impulse response:

\[ FIR = \{1, 0, -4, 0, 10, 0, -23, 0, 80, 128, 80, 0, -23, 0, 10, 0, -4, 0, 1\} \] (2.2)

• the MPEG-4 filter, which has the following impulse response:

\[ FIR = \{2, 0, -4, -3, 5, 19, 26, 19, 5, -3, -4, 0, 2\} \] (2.3)

For efficiency, all of the above filters are separable.

The low-pass filters are used for scaling by a factor of two, so logically the cutoff frequency should be at half the Nyquist frequency \(0.5\omega_0\). The frequency response of the scaling filters shows that the cut-off of the MPEG-4 filter is significantly lower than that of the other filters. This causes a more 'blurred' output. The other filters have a cutoff at half the Nyquist frequency. Additionally, the PixelPlus horizontal filter has a steeper fall-off. Note the ripples at \(\omega = 0.23\omega_0\) (resonance) and \(\omega = 0.7\omega_0\) (alias) in the frequency characteristic of the MPEG-4 filter.

![Figure 2.1: Frequency response of scaling filters](image)

Figure 2.1: Frequency response of scaling filters
2.1.2 Alias

In the frequency response, ripples beyond the Nyquist frequency may be responsible for alias. An example of extreme alias is up-scaling by means of pixel repetition:

$$FIR = \{1, 1\}$$  \hspace{1cm} (2.4)

All of the filters mentioned above suffer more or less from alias. The PixelPlus horizontal and Empress filters only have 3dB attenuation at the cut-off. The MPEG-4 filter has significant ripples beyond the Nyquist frequency, which results in more visible alias.

2.1.3 Overshoot and undershoot

These phenomena may be visible around up-scaled edges. Actually it is an aspect inherent in linear filters that have negative coefficients in their impulse response. A separable implementation will boost these artifacts. Overshoot and undershoot can be avoided by clipping against the maximum and minimum of the input. This measure has a negative effect on the rendering of texture.

The Empress filter is the only one that does not suffer from too much overshoot. This is because it has only few negative coefficients in its impulse response. However, it does not have a very steep roll-off. This is a common trade-off.

2.1.4 Staircasing

The staircasing phenomenon may be visible around up-scaled edges. Instead of looking like a straight edge, it will look jaggy, much like a staircase. It is commonly accepted that two factors contribute to this effect:

- Alias resulting from a badly conditioned frequency response, see Section 2.1.2
- A separable implementation with a strong non-isotropic 2D impulse response
All filters listed above are used in a separable implementation. The resulting 'virtual' 2D filter kernels, and their frequency responses, are non-isotropic. Using a longer 1D impulse response results in a more non-isotropic 2D kernel. In case of the PixelPlus horizontal FIR filter depicted in Figure 2.2 this is already quite visible.

![Figure 2.2: Resulting 2D impulse response and frequency response of separable implementation (non-isotropic) of the PixelPlus horizontal filter](image)

A 1D filter is suitable for separable implementation if it satisfies the following conditions. It should have little to no resonance or alias and preferably not too much overshoot or undershoot. The latter is not always possible since it imposes a trade-off (see Section 2.1.3).

The non-isotropic shape of the filter is a problem if the original 1D filter has some resonance. The resonance of the diagonal frequencies is squared com-
pared to that of the horizontal and vertical frequencies. To avoid resonance and still have a steep roll-off the order of the filter needs to be high.

The PixelPlus filter has a low amount of staircasing. Mostly, the effect is just slightly visible. It seems like it is quite well suited to a separable implementation.

For the Empress filter the same may be concluded. However, the up-scaled image is slightly less sharp.

The MPEG-4 filter really suffers from the separable implementation. Its alias and non-isotropic impulse response contribute a lot to the visibility of the staircasing effect.

2.1.5 Further side-effects of separable filtering

The 2D frequency response depicted in figures 2.2 and 2.3 show that the separable implementation has too much gain in the diagonal directions. This is visible: slightly round edges near +/− 45° are transformed into straight edges. This effect is of course strongest when the cut-off is low, since the lower frequencies contain more power. This is noticeable with the MPEG-4 filter, for instance.

2.1.6 Simple non-separable filter

The non-separable scaling filter proposed by Christian Hentschel has very little alias or staircasing and is very efficient. Used here is the case of scaling by a factor of 2: the 'diamond 4' kernel.

The proposed use of the diamond filter for up-scaling is for up-sampling, then using a Zero Order Hold (ZOH), finally filtering with a simple kernel. The filtering itself can be done with additions and a shift, making it about as cheap as the bilinear filter. A comparison between the two will be made and the diamond filter will also be compared to the more costly FIR filter described above.

First, the impulse response for the case of factor 2 up-scaling will be derived. The filter that is applied to the up-sampled ZOH image is as follows:
Table 2.1: Impulse response of diamond 4 when used with up-sampled ZOH image

By writing down the four cases for factor two up-scaling (for each original pixel four up-scaled ones are produced), the poly-phase impulse response can be derived. The separate impulse responses can be used on the original image.

Table 2.2: Impulse responses when using diamond 4 on original image: upper left, upper right, lower left, lower right cases

The cases are then combined into a poly-phase filter bank to obtain the final impulse response:

Table 2.3: Impulse response of diamond 4 filter

When the same procedure is repeated for the bilinear filter, the following poly-phase impulse response is found:

Table 2.4: Impulse response of bilinear filter

Both impulse responses may be used to filter an up-sampled grid, just like the FIR filters mentioned above. Now, a comparison can be made in the frequency domain as well.
The frequency response shows that the roll-off is not very steep, although a little steeper than the bilinear filter. For the horizontal and vertical directions the characteristics are a lot better. But there still is very slight alias in these cases. In the diagonal directions the roll-off is also steeper and especially the pass band is considerably narrower. This explains the high amount of blur along the diagonal direction.

Looking at the spatial domain, the following remarks can be made. Traditional bilinear interpolation can output unfiltered pixels, the diamond 4 only produces interpolated pixels. This is immediately noticeable: the interpolated pixels can be distinguished from the originals pixels, since their amplitude lies between the amplitude of the originals. Hence, a grid of original pixels is easily spotted, resulting in a slightly jaggy impression. The diamond filter doesn't suffer from this.

The diamond 4 filter is certainly the best scaling filter in terms of computational cost. It is about as cheap as bilinear scaling, but offers the smooth appearance of, for instance, bicubic scaling. This is of course still very soft and generally not suited to the requirements of resolution enhancement.

The most suited basis for resolution enhancement is probably the PixelPlus filter, since it realises a reasonable balance between a high cut-off and suppression of staircasing.
2.2 Spatially layered compression schemes

2.2.1 Spatial scalability in MPEG-2 video

Spatial scalability is offered as an integral part of existing video coding standards such as MPEG-2 [6] and MPEG-4 [11]. The scheme of the standardised algorithm is depicted in Figure 2.5. The performance of the standard solution is poor. This is largely due to the switching between the weighted locally decoded up-scaled base output and the weighted motion-compensated prediction in the enhancement encoder.

![Figure 2.5: Simplified block diagram of MPEG-2/MPEG-4 spatial scalability](image)

Figure 2.5: Simplified block diagram of MPEG-2/MPEG-4 spatial scalability, 'w1' is a weight that is transmitted along with the enhancement data.

MPEG-2 uses bilinear interpolation for half-pel motion compensation and up-scales the base layer with a bilinear filter as well. Both of these aspects cause MPEG-2 spatial scalability to perform poorly.

MPEG-4, however, uses better filtering for motion compensation, maintaining sharpness in inter frames, and uses a 7 tap FIR for up-scaling. As described in Section 2.1, this filter has a quite low cut-off frequency. This helps
it to suppress coding artifacts. Equation 2.3 presents the impulse response.

### 2.2.2 DemoGraFX dual-layer codec

A basic principle to encode an image sequence as a base layer and, separately encoded enhancement layers complementing the decoded base, was already proposed by Philips in the late eighties [12, 13]. Later, particularly driven by the development of HDTV, alternative implementations emerged of which the DemoGraFX method is often referenced [14, 9, 15, 16].

These methods are all based on linear combinations of base and enhancement layer, where the quantisation during encoding is the only non-linearity in the coding chain. Moreover, due to the extra semantic overhead, a (minor) loss in coding efficiency compared to single layer encoding is commonly accepted as the price to pay for the layered representation.

### 2.2.3 MPEG-based dual-layer codec

However recently, Bruls et al. proposed a method that is capable of achieving a dual-layer representation of HD content that requires a lower bitrate than the single layer stream [17]. This method is based on the early Philips principle, but incorporates a new content-adaptive attenuation of the enhancement signal, prior to encoding. Any combination of standard encoders can be used in this scheme, although we confine ourselves here to MPEG-2.

As can be seen in Figure 2.6 the offset is required to convert a signed signal into an unsigned signal, which is required by the MPEG encoder. This is commonly achieved by clipping the 9 bit signed residual to an 8 bit signed signal and consequently adding 128 to unsign. As mentioned above, the $\alpha$ parameter is used for content attenuation. Some successful cases are the combination of skin tone detection or edge detection with low attenuation.

As opposed to MPEG-4, the EMPRESS scaling filter has a higher cut-off, which helps to preserve sharpness.
Figure 2.6: Block diagram of the spatial scalability based on content attenuation of the enhancement data; $\alpha$ is an attenuation factor varying from '0', causing no attenuation to '1', in which case only the base signal is maintained.
Chapter 3

Dual layer codec with non-linear resolution enhancement

In this chapter, various dual-layer coding schemes are proposed. First of all, an overview of the resolution enhancement algorithms is presented. Since the proposed schemes have a lot in common, a framework is provided to start with. This framework contains abstract functions that may be implemented by the resolution enhancement algorithms.

For the enhancement layer, various options for improvement are presented. This is discussed in Section 3.4.

3.1 Resolution enhancement

Resolution enhancement aims at increasing the resolution of an image. This usually involves the application of non-linear algorithms to achieve optimal quality. Resolution enhancement algorithms can be divided into two categories:

1. **Adaptive** methods, which rely on feature detection to control the position and the strength of sharpening algorithms. PixelPlus by Philips belongs to this category.

2. **Learning** methods, which rely on neural nets or maximum likelihood operators to achieve optimal correlation between low resolution and
high resolution images. Learning may take place in advance (i.e. training) or can be done on-the-fly. Digital Reality Creation (DRC) by Sony and the proposed hybrid resolution enhancement belong to this category.

In some cases, resolution enhancement is preceded by linear up-scaling. PixelPlus and the hybrid resolution enhancement use this principle, but DRC combines these two functions in a single step.

### 3.1.1 PixelPlus

PixelPlus is a way of converting an SD signal to HD format, aiming to add sharpness. It features spatial up-scaling and frame rate up-conversion as well as some enhancement algorithms. The up-scaling doubles the horizontal resolution and increases the vertical resolution by a factor of 4/3. This is a separable linear up-scaling. Luminance Transient Improvement (LTI) [18] is used to enhance the perceived sharpness. The de-interlacing is required in order to achieve a frame rate and spatial up-conversion of good quality.

![Figure 3.1: Block schematic of PixelPlus](image)

---

22
3.1.2 Linear up-scaling as a reference

As far as separable linear up-scaling goes, the PixelPlus scaling FIR filters provide good quality. For this reason, these filters are used as the reference linear scaling scheme in the evaluation part of this report. Section 2.1 provides an analysis of the scaling filters used in PixelPlus.

For a factor of 2 up-scaling in both directions (intended for SD to QUAD SD conversion) the procedure is as follows: First, the SD image is up-sampled to the HD grid. Next, the intermediate HD pixels are interpolated by filtering the input SD pixels. This way, for each SD input pixel 4 HD pixels are output.
3.1.3 DRC

DRC or Digital Reality Creation is Sony’s commercial term for a non-linear scaler that is meant to up-scale SD images to HD (Quad SD) format. We believe it to be patented in US-patent 6,323,905 [19]. DRC must be trained in advance, it doesn’t learn on the fly. This is of course a major advantage in terms of speed. In this area a well implemented DRC can compete with a separable FIR.

DRC categorises its input in 'classes' by using an algorithm known as Adaptive Dynamic Range Coding (ADRC). For every class an optimal set of filter coefficients is derived. Output HD pixels are formed by taking a weighted sum of the twelve SD pixels using the coefficients. So basically, it makes use of non-separable linear filtering. The idea is to have an optimal filter for each class. To avoid undershoots and overshoots, the output can be clipped to the range of the twelve input pixels. Figures 3.4 and 3.5 illustrate this.

Figure 3.6 illustrates the training process. The input material is classified by ADRC and used to index an entry (class) in the coefficient data table. Care must be taken to avoid under-population of these classes, which may lead to hysteresis. In this case, an optimal filter is derived for a small sub-set of the class, which will lead to excellent results for the sub-set, but also will lead to poor results for the rest of the class. Outlying pixels and slight geometric distorsion may be visible in the output.

When the training data for one class is too varied, over-population may also occur. This results in a mediocre filter, not unlike a low-pass non-separable FIR. When all the training data is gathered, the Least Mean Squares (LMS) algorithm calculates an optimal filter for each class.

DRC produces the lowest possible squared pixel error for its training set. We want it to maintain this behaviour for other input material. As a consequence, it is recommended to supply a large and diverse training set. For more information about DRC and a comparison with other resolution enhancement algorithms, we refer to the work of Zhao et al. [20].
3.1.4 Separable linear up scaler and non-linear post filter

We developed this resolution enhancement technique as a means to combine the qualities of DRC with the option to scale to various dimensions, i.e.
the ability to use a rational scaling factor. The concept is a straightforward combination of linear scaling followed by a post-filter. The post-filter is much like DRC and is LMS trained for a given scaling factor, but it preserves the dimensions of the input image.

Figure 3.7: Resolution enhancement using linear scaler and non-linear post-filter hybrid

As can be seen in Figure 3.8, the grid is sub-sampled by a factor of two. This significantly improves the classification process; edges and lines are more easily recognized.

3.2 Dual-layer codec framework

To prove the usefulness of non-linear resolution enhancement in spatially layered compression, it was decided to apply a bottom-up approach: starting with only a still base layer up-scaled with non-linear resolution enhancement.
If this results in a significant improvement over a linear up-scaling, the coding scheme could be expanded to support an enhancement layer and moving material. Eventually, the resolution enhancement could be tried with the an MPEG-based codec. The proposed framework uses standard MJPEG encoders and decoders and has enhancement layer capability.

![Diagram of dual-layer encoder](image)

**Figure 3.9:** Abstract framework of the proposed dual-layer encoder

Similar to the MPEG-based coding scheme proposed by Bruls et al., standard coding components are used for the enhancement layer. The same (un)signing operation is performed. Furthermore, the \( \alpha \) parameter is supplied, its main importance being blockiness reduction and 'smooth' bit rate control. There is no content-based attenuation through this parameter.

The encoder contains a switch between *pre-encoder* ("pre") and *post-decoder* ("post") modes. This represents the way in which the residual image is formed. The choice is what base layer information to subtract from the original signal:

*Pre-encoder* mode uses the up-scaled base layer before base layer coding. This exclusively stores the (inverse) scaling artifacts in the enhancement layer.
Post-decoder mode uses the up-scaled base layer after base layer coding. This causes the coding artifacts to be stored in the enhancement layer as well.

When using the post-decoder mode, it is possible to correct base layer coding artifacts. The pre-encoder mode does not allow this, but it would seem to have the advantage of not carrying extra complimentary base layer coding artifacts. In turn, this could possibly aid the coding efficiency of the enhancement layer. However, the base layer bit rate needs to be very high in order to achieve good perceived quality in this particular mode. Otherwise, even small base layer coding artifacts, enlarged by scaling, will become visible and are not corrected by addition of the enhancement layer.

It is recommended to use the post-decoder method with high bit rate budgets. This way all coding artifacts will be gone in the final enhanced image. In this mode it is also allowed to assign all bits to the enhancement layer, which is of course not useful and may be prohibited by setting a minimum base layer bit rate.

Figures 3.9 and 3.10 contain two abstract functions:

- **Down-scaler**, the down-scaling algorithm which may be implemented by any combination of linear filters and down-sampling.

- **Res enh**, the resolution enhancement function which may be implemented by the algorithms discussed in section 3.1.
3.3 The base layer

The base layer is formed by down-scaling with the appropriate algorithm and encoding with a standard MJPEG encoder. The base layer is decoded by applying a MJPEG decoder followed by resolution enhancement.

In this section the various proposed resolution enhancement techniques are introduced into the dual-layer coding framework. An additional pre-filter is introduced in order to maintain good quality resolution enhancement at all base layer bit rates.

3.3.1 Resolution enhancement pre-filter

When decoding the base layer, a filter is applied to reduce coding artifacts. This is done before up-scaling, because this stage may boost artifacts. After up-scaling, artifacts (especially ringing) cover a larger area and become hard to remove.

The artifact reduction filter that was used is a content adaptive filter. 'Adaptive' in this case means that it can be varied in strength by supplying it with the quantizer scale. The filter distinguishes pixel between values below and those above a threshold. The output pixels are calculated by the following equation:

\[ \tilde{g}_5 = \frac{\lambda g_5 + \sum_{i=1,i\neq 5}^{9} \delta_i g_i}{\lambda + \sum_{i=1,i\neq 5}^{9} \delta_i} \]  

(3.1a)

with

\[ \delta_i = \begin{cases} 
1 & \text{if } |g_i - g_5| < S \\
0 & \text{else} 
\end{cases} 
\]  

(3.1b)

and \( \lambda \) the gain for the center pixel.

The filter kernel is as follows, with \( g_5 \) the center pixel.

| \( g_1 \) | \( g_2 \) | \( g_3 \) |
| \( g_4 \) | \( g_5 \) | \( g_6 \) |
| \( g_7 \) | \( g_8 \) | \( g_9 \) |

Table 3.1: Pixel kernel for de-ringer
$S$ is a value derived from the quantization scalar. This functions like the Sigma filter [26] by Lee et al., but with a high gain ($\lambda$) for the center pixel. In other words, it is a Sigma filter directly mixed with the input center pixel. This particularly helps the filter to preserve corners.

Provided $S$ has the right value, the filter can give good results with pictures suffering from coding artifacts. The ringing almost completely disappears with minimum loss of detail. For more information see the paper by Kaup [21].

3.3.2 The base layer with separable linear up-scaling

When comparing the other base layer encoding schemes this one is used as a reference. The up-scaler is the separable FIR filter with up sampling described in Section 3.1.2. For the down-scaling, a 7 taps separable FIR filter from equation 2.1 is used.

These are optimal linear filters available with respect to sharpness, rendering of detail and their computational cost. However, there is a certain amount of staircasing present. See Section 2.1 for more information.

3.3.3 The base layer with DRC

The scaling algorithms used are:

- Down scaler: four-point-average for down-scaling, followed by the aperture correction peaking filter as illustrated in Figure 3.11
- Res enh: DRC for up-scaling, as described in Section 3.1.3

Note that the four point average centers the SD pixels in between the QUAD SD pixels. This is required, because DRC centers its output (QUAD SD) pixels in between the SD pixels. With the four point averaging, the sub-pixel shift is effectively cancelled out.

The aperture correction is required to compensate for the frequency response of the four point average, a sinc function. The four point average frequency response roll-off is not steep enough and this causes frequencies even below the Nyquist frequency to be lost. By keeping the higher frequencies, DRC is able to do a better classification and hence achieve better results.

The aperture correction has the following 2d impulse response:
From Figure 3.13, it can be seen that the pass band is widened by the aperture correction, while the alias, part of the spectrum with $|F| > 0.5$, remains about the same.

DRC may be hindered by coding artifacts in the following ways: What may be expected is edges, distorted by ringing, will be classified as texture. Another example is blocking artifacts being classified as horizontal or vertical edges. With decreasing bit rate, the ringing seemed to become visible earlier than the blocking. Compared to single-layer encoding, the blocking artifact can be easily avoided, because the uncompressed base layer already has a
much lower bit rate than in single layer encoding. Hence, the quantizer scale can be kept at a moderate level.

3.3.4 The base layer with FIR and non-linear post-filter

The scaling algorithms used are:

- Down-scaler: a linear, separable down-scaler
- Res enh: a linear, separable up scaler with the same scaling factor as the down-scaler. This up-scaler is followed by the non-linear post-filter described in 3.1.4.

The FIR down-scaling typically uses filters with a steep frequency roll-off, which has the obvious advantage of producing good down-scaled images, especially when compared to the four-point average down scaler.

3.4 The enhancement layer

The residual image is formed by subtraction of base the layer information from the input signal. The way this is performed is described in Section 3.2. Unlike the common way of directly encoding the residual image with standard video compression components to form the enhancement layer, we evaluated methods that use different compression techniques and methods that use modified video compression components.
Because the residual image contains important features like edges but also noise of various types, the idea was to use edge-preserving noise filters to gain additional compression. These noise-filters were first combined with deflate entropy encoding ¹ in order to achieve a good compression ratio. The deflate algorithm is typically able to achieve compression factors of 2 or 3 on natural images. For noisy residual data the compression factor is about the same. Without the noise the compression might be better. Eventually, it was decided to use the Bruls-method [10]: using a standard video encoder combined with the \( \alpha \) parameter, as mentioned in Section 2.2.3.

To reduce the bit rate requirements of the enhancement layer, various concepts were proposed: Block classification and block jumping, a bit rate dependent quantization matrix and modulation of high-frequent blocks. The bit rate dependent quantization matrix is used in the proposed dual-layer coding scheme.

### 3.4.1 Noise filtering and deflation as enhancement layer compression

Prior to application of MJPEG coding on the enhancement data, it was attempted to apply lossless compression using the deflate algorithm: a combination of the Lempel Ziv [23] and Huffman [24] entropy encoders. The deflate algorithm is described in the document by Deutsch [25]. Of course, lossless compression even with a relatively good entropy encoder still gives a poor compression ratio. Hence, a variety of noise filters were applied prior to coding:

- A 3x3 Sigma filter by Lee et al. [26],
- A 3x3 SUSAN filter by Smith et al. [27],
- A coring operation with a single threshold,
- A thresholding operation.

It seemed that high compression ratios could now be achieved, but both the perceived quality as well as the PSNR were poor.

¹This entropy coding method is commonly used as an efficient alternative of the patented Lempel Ziv Welch (LZW) algorithm, see the article by Welch [22] for a description. LZW is used by the Graphics Interchange Format (GIF) and by older versions of the PKZIP archiver. The deflate algorithm is used by newer PKZIP versions and the Portable Network Graphics (PNG) standard.
Typically, the noise filters did help: SUSAN seemed to respect edges and corners and aggressively filter out TV noise and film grain. Perceptually, the residual image seemed good, but the pixel error was quite high, resulting in a low quality enhanced image. This is because the SUSAN algorithm does not use the center pixel as input.

The sigma filter also filtered out TV noise and film grain and respected the edges, but the corners looked bad. The pixel error in this case was lower than with application of the SUSAN filter. Generally, the sigma filter did not increase the compression ratio enough.

Coring had the advantage of a reasonably low pixel error and high perceptual quality (similar to sigma), but only when using low thresholds. For a good compression ratio, the threshold typically needed to be a lot higher, leading to a low quality enhancement layer. The same applies to thresholding. But in this case the final enhanced image was of even lower quality.

3.4.2 Block classification and jumping

Jumping, or skipping, of blocks, for instance macro blocks in MPEG coding, is known to give a substantial increase in compression [28].

Block classification is required as a criterion for block jumping. Classification in the DCT domain is a good choice, because of the low computational cost. The algorithm can easily separate edges from texture and from plain areas. When reducing the bit rate of the enhancement layer the obvious choice is to keep the edges and jump over the rest. When the allocated bit rate is somewhat higher only the plain areas should be skipped. For more information about classification in the DCT domain, see the report by Yang [29]. The actual algorithm originates from Tong [30].

The algorithm distinguishes 3 cases: texture, edge and plain. It distinguished between these classes by specifying disjoint areas in the 8x8 DCT block: DC, L (Low-frequency), E (Edge) and H (High-frequency). Various conditions involving the absolute pixel sums over these areas lead to the final classification.

The original algorithm was apparently not intended for application on residual images. Especially the classification criteria for the plain class are ill-suited. Since plain areas in enhancement layers have mostly a DC component close to zero, the plain area classification could be implemented much simpler, for instance with a threshold of the block’s absolute sum.
The noisy character of the residual from DRC causes a lot of 'plain' blocks to be classified as texture, whereas these blocks do not contribute to the perceived quality. The classification algorithm’s parameters may be adjusted to compensate for this effect. This is achieved by lowering a threshold.

3.4.3 Modulation of high-frequent blocks

Modulation of high-frequent blocks to low-frequent blocks was attempted to try to reduce the bit rate while keeping the standard zigzag function to remain compliant to MJPEG or MPEG standards. The standard zigzag function works very well on natural signals, since most energy is contained in the low coefficients, but on residuals with a lot of high-frequent data, it has no advantage. The array of coefficients needs to be short without losing the important frequency bands.

![Figure 3.14: Signal spectrum before (left) and after (right) modulation.](image)

The modulation is performed by multiplication of an 8x8 block with the following scalar matrix:

\[
M = \begin{pmatrix}
-1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\
1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\
-1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\
1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\
-1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\
1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\
-1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\
1 & -1 & 1 & -1 & 1 & -1 & 1 & -1
\end{pmatrix}
\]
The idea was implemented with a classification process to mark 'high-frequent' blocks. This resulted in an additional bitmap \( w/8 \times h/8 \) bits long, a relatively small overhead. High-frequent blocks were modulated at the encoder side and demodulated at the decoder side. Other blocks are processed like before. The decoder checks a block's mark bit for a zero and skips demodulation.

Experiments revealed this technique did not lead to great improvements in bit rate, because commonly important blocks in the enhancement layer, for instance containing edges, are not completely high-frequent. Hence, with modulation prior to encoding, the amount of stored coefficients per block still is quite high. An option would be to quantize the low frequencies harder, but this would lead to significant perceptual degradation.

### 3.4.4 Bit rate-adaptive quantization matrix

A bit rate-dependent quantization matrix was realised by mixing of weighted and unweighted matrices. The balance equation between these matrices was derived from tests to gain optimal perceived image quality, see Section 4.1. This was performed for a multitude of bit rates.

The end result was the following equation:

\[
Q = a \cdot U + (1 - a)W
\]

(3.2a)

\[
U = \begin{pmatrix}
16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\
16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\
16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\
16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\
16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\
16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\
16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\
16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\
\end{pmatrix}
\]

(3.2b)
\[ W = \begin{pmatrix} 
8 & 4 & 4 & 4 & 4 & 16 & 16 \\
4 & 4 & 4 & 4 & 16 & 16 & 16 \\
4 & 4 & 4 & 4 & 16 & 16 & 16 \\
4 & 4 & 4 & 16 & 16 & 16 & 32 \\
4 & 4 & 16 & 16 & 16 & 32 & 32 \\
4 & 16 & 16 & 16 & 32 & 32 & 32 \\
16 & 16 & 32 & 32 & 32 & 32 & 32 \\
16 & 16 & 32 & 32 & 32 & 32 & 32 
\end{pmatrix} \] (3.2c)

\( U \) is the unweighted matrix. Such a quantization matrix is required to preserve detail and texture, but it easily leads to the blocking artifacts at a low bit rate. \( W \) is the weighted matrix. This is similar to the standard quantization matrix of JPEG or MPEG. Low frequencies are better preserved than high ones. \( Q \) is the resulting quantization matrix after mixing.

\[ a = \sqrt{1 - q/100} \] (3.3a)

\[ 0 \leq q \leq 100 \] (3.3b)

\( a \) is the mix variable. \( q \) represents the quantizer scale (0: worst quality, 100: best quality).

This kind of mixing is a method to combine advantages of both weighted and unweighted quantization. For all enhancement layer bit rates a good setting is obtained. At low bit rates there is little blocking, and at high bit rates all details are present. Subsequently, this technique was incorporated in the proposed dual-layer coding framework.
Chapter 4

Evaluation of non-linear schemes

In this chapter, the proposed schemes, equipped with various types of DRC (and PixelPlus) non-linear enhancement, are evaluated. This is done by comparing the results of these schemes with those of a reference scheme that exclusively uses linear scaling. A test set of sequences and images is compiled and used as input for each codec. The comparison is made by using objective quality metrics and is accompanied by a subjective evaluation where necessary.

First, the objective metrics, and the methods of measurement for which they are used, are explained. Thereafter, we present the test set. Subsequently, the resulting scores from the metrics along with frames of the output are visualized. A discussion follows, explaining the weaknesses and strengths of the non-linear enhancement techniques, with respect to the applications of dual layer compression. For completeness, the efficiency of the proposed non-linear dual layer codec is compared to that of a single layer codec.

4.1 Image quality metrics

Objective quality metrics are used to quantify an estimate of the perceived quality. Their importance in this case is, that a large number of measurements are needed per frame. The reason for the rate-distortion plane measurements is discussed in section 4.2. Using subjective measurements, this would simply cost too much time. Generally, objective metrics simulate the
HVS with respect to local features (BIM, sharpness) or globally (PSNR, Q-metric). These local metrics focus on the HVS' ability to see certain aspects that are beneficial to the perceived quality, such as sharpness or contrast. More specifically, such a metric can also simulate the sensitivity of the HVS to artifacts arising from video coding or video processing.

Metrics may use a reference (i.e. have access to the original material), which helps to evaluate the quality degradation. Having access to the complete original material is called full-reference. Most full-reference metrics simulate the HVS as a whole. Another option is not using any reference, no-reference.

4.1.1 Peak Signal-to-Noise Ratio

A well-known reference is the Peak Signal-to-Noise Ratio (PSNR), based on the Mean Squared Error (MSE), which is defined as follows:

\[ MSE = \frac{1}{N} \sum_{i=1}^{N} (x_i - y_i)^2 \] (4.1)

Where \( x \) and \( y \) denote pixels in the input image and reference image respectively. \( N \) denotes the number of pixels.

The PSNR is defined as the ratio between the peak (maximum possible) squared pixel error and the \( MSE \). A logarithmic scale is then applied, resulting in the following equation (for 8bit images):

\[ PSNR = 10 \log_{10} \frac{255^2}{MSE} \] (4.2)

The use of the PSNR as a quality measure in video coding is commonly known to be a bad predictor for the perceived visual quality.

4.1.2 Universal Quality Metric

The Q-metric proposed by Wang \textit{et al.} is an alternative measure that showed to represent the quality degradation in a way that was consistent with our own observation. The definition of the Q-metric can be written as

\[ Q = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \frac{2\bar{x}\bar{y}}{\bar{x}^2 + \bar{y}^2} \frac{2\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2} \] (4.3)
where \( \vec{x} \) and \( \vec{y} \) are vectors with original and distorted data, respectively. The symbols \( \bar{x} \), \( \sigma_x \), and \( \sigma_{xy} \) denote the estimated mean value, standard deviation and cross correlation in the traditional sense. The value of \( Q \) theoretically ranges from '-1' to a maximum value of '1', meaning that \( \vec{x} \) and \( \vec{y} \) are identical. The three terms in (4.3) respectively model the loss of correlation, the distortion between the mean signals and the distortion between the variances.

The Q-metric is known to effectively estimate the perceived quality degradation for a wide range of distortions such as Gaussian noise and blurring, but also for blocking and ringing. This makes it suited to measurements on intra coded material, which is used in this chapter.

4.1.3 Blockiness Impairment Metric

Another quality metric, the Blockiness Impairment Metric (BIM), uses a more specific approach. It exclusively tries to quantify the blockiness artifacts. As is well known, these are the most annoying of all coding artifacts.

This no-reference algorithm is based on the inter pixel difference between each of the block boundaries. It takes into account various effects that can mask block artifacts. In bright, dark or spatially busy areas this is often the case. Also, it tries to compensate for the case that highly geometric content may be seen as artifacts. This is done by evaluating the pixels not on the block borders, which results in the non-block-edge blockiness. Subsequently, the ratio between this and the block-edge blockiness and the non-block-edge blockiness forms the BIM.

The BIM is explained in detail in the article by Wu and Yuen [31].

The BIM is composed of two individual metrics:

- the \( H_{BLM} \) for the horizontal block edges,
- the \( V_{BLM} \) for the vertical block edges.

In 4.4 the equation for the HBIM is presented. We provide the HBIM, since the HBIM and the VBIM are almost identical.

\[
HBIM = \frac{7M_{h,0}}{\sum_{n=1}^{7} M_{h,n}} \quad (4.4a)
\]
\[ M_{h,n} = \sqrt{\frac{1}{Nc/8-1} \sum_{k=1}^{Nc/8-1} (w_k(f_{3k+n} - f_{3k+n+1}))^2} \]  

(4.4b)

Where \( f_i \) denotes the pixel value at column \( i \). The row length or number of columns is denoted by \( N_c \). \( w_i \) denotes the weight, for luminance masking amongst others, of the pixel at column \( 8i \). This weight is calculated from the mean and the standard deviation of the left and right block rows.

Usually, it is sufficient to combine both into one score by averaging:

\[ BIM = \frac{VBIM + HBIM}{2} \]  

(4.5)

The BIM score ranges from zero to infinity (in theory). A \( BIM < 1.5 \) often indicates that the image is free from blocking artifacts. A score close to 2 indicates some minor blocking. Anything with a higher score is either very hindered by blocking artifacts or contains highly geometric shapes.

### 4.1.4 Sharpness metric

The no-reference sharpness metric introduced by Caviedes et al. [32] enables us to measure the degradation in sharpness caused by coding. Furthermore, it can also quantify the sharpness in enhanced material. The PSNR, a full-reference metric, fails in doing the latter. This is because enhanced material can be perceived as very sharp even when having a high pixel error.

The sharpness metric exclusively measures the sharpness of edges. This helps against the influence of for instance noise. Each edge pixel is assigned a local sharpness. The average sharpness over all of these edge pixels forms the final score of the metric. The local sharpness is calculated by taking a 2D DCT of the local neighbourhood, normalising it and applying a 2D kurtosis calculation.

This sharpness metric is known to sometimes evaluate material suffering from blocking artifacts as very sharp. To compensate for this side-effect, it is recommended to additionally use a blockiness metric, for instance, the BIM. This will expose the cases where sharpness score should be neglected.
4.2 Rate-distortion plane measurements

In case of just one varying bit rate, a rate-distorsion curve is a good way of illustrating the effect of the bit rate on the perceived quality. When using multiple bit rates however, two in this case, this will not suffice. A solution is presenting multiple curves, but these will take up alot of space. Alternatively, the multiple curves can be incorporated into one surface: a rate-distortion plane.

For dual layer compression there are two bit rates: the base layer bit rate (base bit rate) and the enhancement layer bit rate (enhancement bit rate). The total bit rate is the sum of these two. We are interested in finding the optimal quality as a function of the two bit rates, so we can proceed as follows: The horizontal axis represents the base bit rate and the vertical axis represents the enhancement bit rate. Now, on a diagonal the total bit rate is constant and along this line we expect to find an optimal quality.

The plane is best presented by a contour plot, because 3d visualisations are often harder to interpret. The example below depicts a simple (and highly unlikely) case where the base bit rate and enhancement bit rate both contribute equally to the quality. As a consequence, the optimal bit rate allocation (the optimal combination of base and enhancement bit rate) can just be any random bit rate allocation.
Figure 4.1: Example of a fictive rate-distortion plane using the Q-metric quality scale. The contours are completely diagonal, meaning any bit rate allocation is optimal.

We mark the optimal quality on each diagonal with a cross and connect it with the cross on the next diagonal (of higher total bit rate). The grid has a precision of 0.1 bpp, so the step between consecutive optima is 0.1bpp as well. This is the finest resolution possible, because of the low precision of the bit rate control: the base bit rate is controlled exclusively by variation the quantizer scale. This quantizer scale is then kept constant over a frame.

The enhancement bit rate may be varied by variation of both the quantizer scale as well as the $\alpha$ parameter, as defined by Bruls et al. [10]. This however, still doesn't allow a resolution of say 0.05 bpp. For all the measurements in this chapter, the bits per pixel are normalised to the dimensions of the input material.

The example below illustrates a more realistic case. The optimal bit rate allocation is strongly biased towards a high base bit rate.
Using the optimal bit rate trajectory as for instance shown above it becomes possible to present an optimal rate-distortion curve.

4.3 Test Sequences

These sequences are grabbed from progressive material. This is important, because de-interlaced video may contain artifacts such as blur in the vertical direction or jaggies, which can severely influence the objective metric scores.

- **Music** This is a high definition QUAD SD sequence. The content is rich in contrast and there are a lot of diagonal lines as well as text.
- **Bicycle** This is a standard definition sequence at 50Hz. The content is highly geometric with a lot of diagonal edges and is rich in contrast.
- **Kiel harbour** This is a standard definition sequence with a lot of straight lines. It also contains a big low frequent area.
• **Market** This SD sequence contains curved and straight edges and numerous textured areas.

• **Tram** This is a QUAD SD sequence which contains numerous edges and lines both curved and straight. It also contains a lot of natural components however.

• **Nets** This is a QUAD SD sequence which contains a huge amount of curved and straight lines of varying thickness.
Figure 4.3: frames from original test sequences. from left to right and from top to bottom: a) Music, b) Bicycle, c) Kiel harbour, d) Market, e) Tram, f) Nets
4.4 PixelPlus resolution enhancement

The enhancement algorithms of PixelPlus are meant to improve the perceived sharpness, but they do not improve the pixel error much. In turn, this has little impact on the PSNR, which is typically just 0.2 to 0.3dB higher. This does not aid the usage of an enhancement layer. Typically, a larger part of the bit rate needs to be devoted to this layer than when using DRC, for instance.

The evaluated LTI is proposed for implementation in Pixel++ by Tegenbosch et al.. This LTI version is to be the successor of the LTI in PixelPlus2 by Bellers et al. [18].

![Figure 4.4: The bicycle sequence up-scaled and down-scaled with PixelPlus scaling filters (left) and consequently enhanced with LTI (right). The PSNR of the left and right image is 28.1dB and 28.3dB respectively.](image)

4.5 Proposed dual-layer codecs with factor 2 scaling

The original DRC algorithm up-scales by a factor of two in both the horizontal and the vertical direction. For this reason, the comparison with codecs exclusively with a scaling factor of two is made here. In Section 4.6 the results from schemes with a variable (rational) scaling factor are presented.
4.5.1 Reference codec with linear scaling

In Appendix A.1 rate-distortion planes are presented. These are taken from the codec with linear scaling. Figures A.1, A.2 and A.3 present the Q metric, PSNR and BIM planes respectively.

Logically, it seems that in order to achieve optimal BIM, a high bit rate should be allocated to the base layer. This is apparent from all measurements. For an optimal PSNR, the choice is to have a much lower bit rate allocated to the base layer. The Q metric seems somewhere in between these trends.

The Q metric rate-distortion planes indicate that nets, tram and especially the kielp sequence have a low perceived base layer quality. Subsequently, these sequences require a high enhancement bit rate. This is because of the large amount of sharp edges and lines and the overall spatial business of these sequences. Music and bicycle contain large flat (spatially inactive) areas and hence don’t require an enhancement bit rate as high as for instance kielp does.

The separable, linear up-scaling causes artifacts which are responsible for the low perceived quality of the base layer:

- **Staircasing**, frequently found on diagonal lines and edges. The reasons for occurrence of this particular artifact are discussed in Section 2.1.

- **Blurring**, most noticeable on horizontal and vertical lines and edges and on texture. Blurring is caused by the low-pass behaviour of linear scaling filters.

![Figure 4.5: Region of 'bicycle' compressed to 1.0 bpp and subsequently up-scaled linearly](image)
As can be seen in the *Bicycle* grab, the diagonal edges are staircased and blurred. *Tram* suffers from staircasing around the right edge of the wind shield and around the lines on the side of the tram and on the road. *Market* has loss of texture on the baskets at the top and blurring on the hat in the center.
4.5.2 The DRC-based scheme

This part contains results from the proposed DRC-based codec as specified in Chapter 3. The DRC uses a LUT generated by a training session of 800 frames taken from various QUAD SD sequences. The training set and the test set did not overlap. The down-scaled (SD) input was provided to the training process by application of the four point average algorithm without aperture correction.

In Appendix A.2 rate-distorsion planes are presented. These are taken from the codec with DRC resolution enhancement. Figures A.4, A.5 and A.6 present the Q metric, PSNR and BIM planes respectively.

The BIM may suffer from DRC resolution enhancement. The low BIM score is caused by the increased sharpness and by the four-point average down-scaling. Using an alternate down-scaler, such as the FIR in Equation 4.6, and training for this down-scaler, yields a significantly lower BIM.

From the PSNR planes, it may be concluded that it is preferred to allocate slightly more bits to the base layer than with linear scaling. The Q metric indicates an even stronger preference for a higher base bit rate.

A quick look at the output of DRC shows that up-scaled edges are very sharp. Much more so than the output of a the linear up-scaler. The classes for edges are apparently easily recognized. In areas with a lot of texture, the results are closer to that of traditional scalers.

The bicycle sequence clearly illustrates the effect of DRC. The sharp diagonal edges are recognized and appropriate filters are selected for up-scaling. The perceived sharpness of the DRC up-scaling is clearly higher than that of the linear up-scaling. Figure 4.8 shows a zoomed region of the image. Notice the blurring and peaking in the left image. The DRC up-scaled image does not suffer from these artifacts.
The residual image of DRC shows that the residual edges are smaller or have lower amplitude. In some cases the residual edges disappear. Perceptually, the residual of DRC might seem to give a noisier impression. However, this seemed not to influence the coding efficiency.

Depicted in Figure 4.9 is the same region encoded at 0.5 bits per pixel (bpp). Both up-scaled images appear to suffer from coding artifacts. The image up-scaled with DRC, however, still has sharper edges.

To illustrate the effect of coding artifacts on resolution enhancement up-scaled images, the rate-distortion planes for the bicycle sequence are presented in Figure 4.10.

Clearly bicycle is an optimal case for DRC: the sequence contains a lot of sharp edges. The improvement is about 2 dB in the best case. The increase in Q metric score is about 0.02. The rate-distortion curve shows a clear improvement for almost all bit rates as can be seen in Figure 4.12.

Note that the Q metric curves in the right of Figure 4.12 provide a good overview of the perceived quality increase. The 'chair' shape is very typical
Figure 4.9: Region of 'bicycle' compressed to 0.5 bpp and subsequently up-scaled linearly (left) and with DRC (right).

Figure 4.10: Rate-distortion planes of bicycle sequence coded using reference scheme (left) and DRC-based scheme (right). Values along the height lines denote the PSNR (dB).

of dual-layer coding. Increasing the base layer yields a vast improvement of quality at first, later on the enhancement layer bit rate needs to be increased, but this only really starts to pay off when a very high amount of bits are devoted to the enhancement layer. The PSNR curve doesn't show this.

When analysing the artifacts present in the up-scaled image by DRC, it seems that only with extremely low base bit rate, the perceived quality is inferior to that of the scheme with linear up-scaling, i.e. base bit rates around 0.1 bpp. This is due especially to the high blockiness. The linear up-scaling blurs this artifact more than DRC does. The pre-filter used in the proposed scheme helps to reduce this effect.

Figure 4.13 shows a zoomed area of the 'tram' sequence processed by the linear scheme and the DRC-based scheme, respectively. When comparing the two zooms, notice that on right edge of the wind shield the DRC output
Figure 4.11: Rate-distortion planes of bicycle sequence coded using reference scheme (left) and DRC-based scheme (right). Values along the height lines denote the Q metric score.

Figure 4.12: Rate-distortion curves of the bicycle sequence. The PSNR is presented left and the Q-metric score is presented right.

does not show the staircasing artifact. Furthermore, from Figure 4.14 it is apparent that the DRC rendering of the basket texture and the hat edges is clearly superior.

Unfortunately, in a late stage of this project it was discovered that the four point average is an insufficient means of down-scaling when using DRC. Various lines with a small slope show staircasing. The BIM also clearly indicates this. Sequences from the test set that suffer from this kind of artifact are kielp, tram and nets. Figure 4.15 illustrates this phenomenon.
Figure 4.13: Region of 'tram' compressed to 1.0 bpp and subsequently up-scaled linearly (left), respectively up-scaled using DRC (right).

Figure 4.15: Region of the kielp sequence down-scaled with four point average and aperture correction (left) and an eight tap separable FIR respectively, prior to resolution enhancement with DRC.

\[ FIR = \{-2, -2, 8, 28, 28, 8, -2, -2\} \]  \hspace{1cm} (4.6)

An alternative down-scaler was used to limit the staircasing artifact. DRC was trained again for this down-scaler. It uses a separable implementation of the FIR filter of Equation 4.6. Note that, like the four point average, it
also shifts by a half pixel.

4.6 Proposed dual-layer codecs with rational scaling factor

All coding schemes evaluated in this section are able to scale by rational factors. This is for instance required to scale SD material to ATSC resolutions. The scaling factors used for conversion from SD to ATSC are relatively close to 2.0. This formed the motivation to simply apply a scaling factor of two to enable the use of the test set defined in 4.3. Hence, the measurement data of the reference coding scheme as described in 4.5.1 may be used.

4.6.1 DRC-FIR hybrid scheme

The non-linear filter was trained with a set of twenty images: ten at SD resolution and ten at QUAD SD resolution. The down-scaled input was provided to the training process by application of the same down scaler as used in the dual-layer hybrid scheme.

Rate-distortion planes are presented by Figures A.7, A.8 and A.9, which represent the Q metric, PSNR and BIM planes respectively.
In terms of Q and PSNR, the results are comparable with those of the DRC-based scheme. The BIM is generally lower than that of the DRC-based scheme. Subjectively speaking, this is reflected by alias-free edges generated by the hybrid resolution enhancement. However, DRC can produce similar results with the appropriate training.

The consequences for the preferred bit rate allocation are similar to those of the DRC-based scheme: a slightly higher bit rate should be allocated to the base layer.

Subjective evaluation reveals that edges are significantly sharper. Also, the staircasing produced by the separable FIR is strongly suppressed. This is illustrated by the music sequence.

As mentioned above, the results of this resolution enhancement are close to those of DRC. However, the post-filter needs to suppress the artifacts generated by the linear up-scaling. If the linear up-scaler has a low cut-off frequency, this may result in hysteresis of the post-filter. Providing the post filter is well-trained, the PSNR can increase by 0.3dB up to 2.0dB.

Figure 4.16: A frame from the music sequence coded at 0.75bpp and consequently scaled linearly (left) and using the FIR-DRC hybrid (right) respectively.

4.7 Proposed dual-layer codec vs. single-layer MJPEG

These measurements were conducted to verify if the proposed scheme was able to match the quality of a single-layer system. This would also show that the scheme was properly implemented. The results from the Empress codec
by Bruls et al. [17] formed the motivation to improve the proposed system and to match these figures. The BIM showed that the blockiness at low bit rates was significantly lower than that of a single-layer codec.

To get a good impression, the goal was set to achieve the same PSNR and BIM scores as Empress in its comparison with the performance of a single-layer codec. This lead to the two rate-distortion curves displayed further on in this section.

The comparison is made between the following schemes:

- The evaluated dual-layer scheme is the proposed DRC-based codec with factor 2 scaling. The DRC training data did not contain images from the test set used in this section.

- The evaluated single-layer scheme is a standard MJPEG codec.

Measurements were performed on a range of still material, including frames from the test set of 4.3: eight pictures in total, both standard definition and high definition. The results are averaged measurements of this material. Each measurement is done at 19 bit rates, varying from 0.2 bpp up to 2.0 bpp. 0.2 bpp is the lowest bit rate single-layer JPEG can handle. 2.0 bpp is considered high quality.

For each measurement, optimal codec settings are used. Meaning the settings for the BIM measurements are largely different from the settings used for the PSNR measurements.
Clearly, in terms of PSNR, the proposed codec is at least as good as the reference single-layer MJPEG codec. Of course, this definitely doesn’t say everything about the perceived quality, because the PSNR is largely blind to blocking artifacts. Hence, the BIM was chosen to complement the PSNR measurements.
In terms of BIM the proposed codec is clearly superior at any bit rate, even at the very high ones. Typically for this test set a BIM below 1.5 indicates good quality. When looking at the rate-distorsion BIM curve in Figure 4.18, the BIM goes down with increasing bit rate. This in contrast to the Bruls-method, where the BIM goes up with increasing bit rate.

Additionally, sharpness measurements were conducted. In most cases the single-layer reference shows a higher score, but just by a slight margin.
Chapter 5

Conclusions and recommendations

Several options for resolution enhancement have been compared in a motion JPEG (MJPEG) based dual-layered coding scheme. The aim of the comparison was to provide an overview of the use of non-linear resolution enhancement as a replacement of traditional linear up-scaling. Several versions of the resolution enhancement algorithms Digital Reality Creation (DRC) and Luminance Transient Improvement (LTI) were evaluated.

Additionally, various methods for improvement of the enhancement layer compression were proposed. This involved the application of various spatial classification and modification algorithms on the residual image and the use of different coding parameters and algorithms to form the enhancement layer.

5.1 Conclusions

The following conclusion can be drawn. The proposed DRC-based scheme can decrease the bit rate, while keeping the same image quality. This is true especially for material with sharp diagonal edges or curves. In this case, DRC produces a sharper up-scaled image, which leaves less residual data to compress. In other cases, the proposed scheme performs similar to other dual-layer codecs.

It is recommended to use the enhancement layer only if its assigned bit rate is high enough. This is also the case with codecs that use linear up-scaling. However, with non-linear resolution enhancement the threshold is
even higher. Only if a really large bit rate is available, usage of the enhancement layer is recommended.

It needs to be said that using DRC in a codec is not straightforward. Sometimes rational scaling factors are required for conversion of 4:3 standard definition (SD) to wide screen high definition (HD) formats. DRC was designed exclusively for conversion from SD to QUAD SD. Also, additional measures are required to avoid amplification of coding artifacts. These artifacts may be suppressed by application of a de-ringer prior to resolution enhancement.

An additional hybrid resolution enhancement scheme was proposed to combine the advantages of the DRC algorithm with the flexibility of linear scaling methods. This has proven to work quite elegantly. The requirement is a good down-scaler which gives this method the advantage of alias-free upscaled results. The results are very similar to those of DRC and in some cases superior.

The proposed improvements for the enhancement layer compression, as described in Section 3.4, were mainly without success, except for the option of a bit rate adaptive quantization matrix. This keeps the most important features of the residual image at low bit rate and provides the lowest pixel error at high bit rate.

Recently, experiments with the integration of the hybrid resolution enhancement algorithm into the EMPRESS coding scheme were conducted. It seems it does not aid the perceived quality, nor the score of objective metrics. We currently believe this to be caused by additional artifacts resulting from interframe coding.

5.2 Recommendations

Recently, there have been some advances in the optimisation of DRC. A 3 x 3 kernel can be used instead of the 12-pixel kernel that was used at the time of evaluation. Tests have shown that this version produces up-scaled images of comparable quality. This new DRC version especially reduces the size of the LUT.

DRC has recently been tried in a H.264 based dual layer scheme, using interlaced test sequences. For more information see [33]. It seems the results are satisfactory and comparable to using MPEG-4 filters, but a better de-interlacer could improve the PSNR.
For this project, DRC was trained with HD images only. Due to the differences in the SD and HD spectra, the CIF to SD up-scaling, as opposed to the SD to HD up-scaling, might have been sub-optimal. It is recommended to try and train DRC to obtain a specialized LUT for CIF to SD up-scaling.

It was proven that a DRC-based dual-layer MJPEG coding scheme may provide at least the compression ratio of a single-layer MJPEG scheme. This could mean DRC could aid coding efficiency in wavelet compression or possibly in a triple-layer scheme MJPEG scheme.
Bibliography


Appendix A

Rate-distortion planes

This appendix contains all rate-distortion plane measurements taken from the three proposed dual-layer coding schemes:

- with linear up-scaling
- with DRC resolution enhancement
- with hybrid non-linear resolution enhancement

The means by which the measurements are conducted are defined in section 4.2. The test set is specified in section 4.3. The metrics used for these measurements are listed in 4.1.

The test set contains 3 SD sequences and 3 HD sequences. The rate-distortion planes for the SD and HD sequences are depicted on the top and bottom rows respectively.
A.1 Linear up-scaling

Figure A.1: Q-metric rate-distorsion planes from the dual-layer coding scheme with linear scaling. The top row from left to right: Bicycle, Market, Kiel Harbour. The bottom row from left to right: Music, Tram, Nets.
Figure A.2: PSNR rate-distorsion planes from the dual-layer coding scheme with linear scaling. The top row from left to right: Bicycle, Market, Kiel Harbour. The bottom row from left to right: Music, Tram, Nets.
Figure A.3: BIM rate-distortion planes from the dual-layer coding scheme with linear scaling. The top row from left to right: Bicycle, Market, Kiel Harbour. The bottom row from left to right: Music, Tram, Nets.
A.2 DRC resolution enhancement

Figure A.4: Q-metric rate-distortion planes from the dual-layer coding scheme with DRC resolution enhancement. The top row from left to right: Bicycle, Market, Kiel Harbour. The bottom row from left to right: Music, Tram, Nets.
Figure A.5: PSNR rate-distortion planes from the dual-layer coding scheme with DRC resolution enhancement. The top row from left to right: Bicycle, Market, Kiel Harbour. The bottom row from left to right: Music, Tram, Nets.
Figure A.6: BIM rate-distorsion planes from the dual-layer coding scheme with DRC resolution enhancement. The top row from left to right: Bicycle, Market, Kiel Harbour. The bottom row from left to right: Music, Tram, Nets.
A.3 Hybrid resolution enhancement

Figure A.7: Q-metric rate-distortion planes from the dual-layer coding scheme with hybrid resolution enhancement. The top row from left to right: Bicycle, Market, Kiel Harbour. The bottom row from left to right: Music, Tram, Nets.
Figure A.8: PSNR rate-distortion planes from the dual-layer coding scheme with hybrid resolution enhancement. The top row from left to right: Bicycle, Market, Kiel Harbour. The bottom row from left to right: Music, Tram, Nets.
Figure A.9: BIM rate-distortion planes from the dual-layer coding scheme with hybrid resolution enhancement. The top row from left to right: Bicycle, Market, Kiel Harbour. The bottom row from left to right: Music, Tram, Nets.
Appendix B

Literature research

B.1 Project description

This chapter gives a description of the assignment for the graduation project.

B.1.1 Project assignment

The Video Processing and Visual Perception group from Philips primarily researches displays and video compression. Algorithms play an important part in this process. The assignment focuses on the application of algorithms in video compression schemes.

My assignment involves researching the application of potentially useful algorithms and ideas in video compression. The type of video compression we are interested in is the so-called two-layer compression scheme.

B.1.2 Literature research assignment

The goal of this literature research is to obtain an overview of prior art in the field of layered video encoding. Additionally an overview of algorithms suited for possible future application in this field is desired.

Industry standards, articles from journals and conferences as well as patents should contain valuable information. A worldwide search is a good idea, since a lot of research in this field is done in the USA, Europe and Japan.
Both background literature research as well as reviewing some of the recent publications has to be done. The background research is especially important because even a small detail in a dual layer codec from years back can influence a future patent. The recent publications are important for obtaining an overview of potentially useful techniques. These techniques are not directly related to dual layer encoding. However they may be of use in the future.

The aspects of recent techniques I am interested in are: the configuration of image scalers, enhancement algorithms (possibly non-linear) and artifact reduction algorithms. I think a lot of information about this could be found from the mid nineties until now.

I estimate some of the earliest background literature will be from the late eighties. So it seems like an appropriate choice to search from this time until recent years.

Documents of interest are: patents (see above), articles and industry standards. The industry standards are important for the practical part of the graduation project: how to implement various techniques for instance. Also, they are important for the same reasons why patents are. Articles are important because they tend to give a good overview, whereas patents involve a lot of legalese and standards often contain too much background data.

The literature and patents will probably originate from electronics corporations. Sony and Philips have a substantial amount of patents in the field of video coding for instance. Some of them might deal with layered coding.

**B.2 Conceptual table of contents of final report**

The concept index of the graduation report consists of the following chapters:

- Contents
  1. Introduction
  2. Currently available techniques
  3. Dual layer codec with drc
  4. Comparison
  5. Conclusions and recommendations
References

A. Literature research

B.3 List of search terms

English:
- video
- coding
- layered video compression
- spatial scalability
- spatial scalable
- layered
- layer
- image scaling
- training
- deblocking

B.4 Used sources

- INSPEC, all databases 1987-2003
- IEEE-IEE Electronic Library (IEL) Online, all databases 1988-2003
- Philips library 1980-2003
- Citeceer

The used sources are all online. Citeceer was used autonomously and not for the citation method. MicroPatent was used since Philips is subscribed to it. The IBM patent search was avoided on purpose, because the register queries to use to their advantage. I.e. using the IBM database, risks a potential patent application by IBM, resulting in a loss for Philips.
B.5 References

B.5.1 Selected references per source

Note that this is the final selection of references, i.e. the irrelevant titles are removed. Also the titles are de duplicated, meaning titles from one library don’t overlap with titles from another.

<table>
<thead>
<tr>
<th>source</th>
<th>period</th>
<th>hits</th>
</tr>
</thead>
<tbody>
<tr>
<td>IEEE-IEE Electronic Library (IEL) Online</td>
<td>1987-2003</td>
<td>14</td>
</tr>
<tr>
<td>INSPEC</td>
<td>1988-2003</td>
<td>5</td>
</tr>
<tr>
<td>Philips library</td>
<td>1980-2003</td>
<td>4</td>
</tr>
<tr>
<td>Citeceer</td>
<td></td>
<td>0</td>
</tr>
</tbody>
</table>

B.5.2 Selection criteria

Reasons for selecting references for the literature list are:

- literature containing general information on spatial scalable video compression
- literature from Philips giving an overview of the available techniques within the company
- literature that deals with artifact reduction in compressed material
- literature about up-scaling and down-scaling algorithms
- Philips patents on spatial scalable compression
- Patents from other companies on spatial scalable compression

The moment of selection was during the writing of the graduation report.

B.6 The snowball method

The snowball method uses references to search for related literature. As starting point references [1] and [2] were chosen. Both are recent publications discussing spatially scalable compression.
B.7 The citation method

The journal article by Ghanbari [1] is the first one to give a two-layer encoding scheme based on spatial scalability. Hence this article is chosen as the starting point. For this diagram the cite function of IEL online was used.
Figure B.2: The citation method diagram.
B.8 Relation literature and contents of graduation report

<table>
<thead>
<tr>
<th>Chapter</th>
<th>reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Introduction</td>
<td>2</td>
</tr>
<tr>
<td>2. Currently available techniques</td>
<td>2 3 25</td>
</tr>
<tr>
<td>3. Dual layer codec with DRC</td>
<td>1 4 8 11 18 25</td>
</tr>
<tr>
<td>4. Comparison</td>
<td>4 5 6 12</td>
</tr>
<tr>
<td>5. Conclusions and recommendations</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subject</th>
<th>reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatially scalable coding</td>
<td>12, 25</td>
</tr>
<tr>
<td>Image scaling</td>
<td>8 6</td>
</tr>
<tr>
<td>Artifact reduction</td>
<td>1, 4 18</td>
</tr>
<tr>
<td>Coding basics</td>
<td>2 3</td>
</tr>
<tr>
<td>Year</td>
<td>93 96 97 98 01 02 03</td>
</tr>
</tbody>
</table>

B.9 Conclusion and recommendations

Large parts of the literature found are about error resilience. It is an important characteristic of layered coding, but my research was directed primarily at compression and image quality. Hence error resilience is a subject that I do not deal with in my report. Only a part of the results from the citation method are used as references in my graduation report.

Useful information was found in all types of publications as well as standards and patents. However, most of the information was concentrated in articles. Searching from the late eighties until 2003 proved wise. No documents on layered coding from before 1989 seemed to exist. The early documents however did not really add up to my insights and so most will not be referenced in the final graduation report.

The most important authors and their institutes are:

De Bruijn (Royal Philips Netherlands, Eindhoven, Netherlands) Zhao Eindhoven University of technology, Eindhoven, The Netherlands) Bruls (Royal Philips Netherlands, Eindhoven, Netherlands) Vervoort (Royal Philips Netherlands, Eindhoven, Netherlands) Kaup (Siemens Corporate Technology, Munich, Germany)
B.10 References


