Image representation and compression using steered hermite transforms

DOI:

[10.6100/IR502348](https://doi.org/10.6100/IR502348)

Document status and date:
Published: 01/01/1997

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

Download date: 16. Nov. 2023

Image Representation and
Compression Using
Steered Hermite Transforms

Toon van Dijk

Image Representation and

Compression Using

Steered Hermite Transforms

cover: Celsus-bibliotheek in Ephesos, symbool voor wijsheid, daadkracht, inzicht en wetenschap (Selçuk, 1994).

Image Representation and

Compression Using

Steered Hermite Transforms

door

Antoon Meindert van Dijk

geboren te Wijchen

Dit proefschrift is goedgekeurd door de promotoren:

prof.dr.ir. J.A.J. Roufs
prof.dr.ir. J.B.H. Peek

en de copromotor

dr.ir. J.B.O.S. Martens

voor Marjolein
en mijn ouders

Acknowledgements

Contents

Chapter 1

Introduction

1.1 Perception and Image Processing

The introduction of multimedia applications in our daily life has made high demands upon the performance and the quality of technical systems that integrate information sources such as text, sound, images and video. In order to establish high quality services, optimal use of available resources is essential. For digital image processing, the optimization with respect to image quality makes use of our knowledge of human visual perception. How to optimally use our knowledge of perception in image processing, however, has not yet been fully determined and more research is needed to link perception and image processing in a more efficient way.

Detailed knowledge of the visual system can, first of all, be used to predict visual behavior and calculate the effect of certain stimuli. Second, characteristics of the visual system can be used effectively to improve image processing algorithms. In the past, there has been a particular interest in measuring both spatial and temporal properties of the visual system, and a considerable amount of research has been dedicated to model threshold perception and luminance-brightness relations. The assumption that a better insight in visual functions could be obtained by analyzing underlying problems from an information processing point of view [86], led to the development of models that closely follow the processing steps of the early visual system (i.e. at the level of the retinal and cortical receptive fields). From an image processing point of view, mathematical descriptions of the information analysis in the visual system enable algorithms, for instance, to operate in a representation domain that is similar to the domain used by the visual system. An example of this approach is found in color processing applications that use a three-dimensional representation of the color space.

Also in the field of image compression, more attention should be paid to perception processes. In the first place, knowledge of perception leads the development of efficient image compression techniques, because visual properties can be exploited to remove image information that can not be perceived by the

human eye. Second, for lossy compression techniques, the notion of perceptual image quality [120] and image quality understanding has become increasingly important. In particular there is a need for techniques that measure the perceptual distortion between the coded image and the original image, either by calculating some objective error measure or by carrying out subjective experiments.

In this thesis, image compression techniques are presented that are based on image analysis with a Hermite transform, an image description technique that uses analysis filters that model the information analysis carried out by the retinal and cortical visual receptive fields [90, 155]. The image compression research in this thesis, however, is motivated by the assumption that optimal image quality for a given bit rate can only be achieved if the representation of images focuses at the explicit description of those structures in the image that contribute most to perceived image quality. Further, this thesis discusses a number of experimental methods and data analysis techniques that can be used to evaluate perceived image quality of compressed images.

1.2   Image Compression

Over the last decades, ground-breaking work has been done in the area of digital image compression, starting with simple techniques of pulse code modulation (PCM), delta modulation and various prediction techniques (e.g., differential PCM). With these basic techniques, supported by quantization techniques [60] and lossless entropy coding of the resulting source symbols [111], the first compression algorithms were developed. An example of an entropy coding technique is Huffmann coding that assigns, depending on the frequency histogram of the source symbols, fixed sequences of bits (codewords) of different length to the source symbols in order to reduce the average codeword length. In other techniques, such as Lempel-Ziv coding and arithmetic coding, there is no one-to-one correspondence between codewords and source symbols. In arithmetic coding, a codeword is assigned to an entire sequence of source symbols, and slightly different source sequences may result in significantly different code sequences. An introduction to entropy coding techniques can be found in [111].

With transform coding or decomposition techniques [4] such as the Karhunen-Loève transform (KLT), the discrete cosine transform (DCT), subband coding and wavelet coding [79, 138], a huge step forward was taken. By decomposing signals into different localized patterns, exploitation of the uneven distribution of signal energy in "natural" real-world images became possible. Also, frequency decomposition techniques facilitated incorporation of properties of the human visual system such as frequency sensitivity and contrast masking [123, 146]. Furthermore, transform coding techniques based on models of spa-

tial processing in the early stages of human vision were developed. Amongst these, the Cortex transform [142, 143], the Gabor transform [30] and the Hermite transform [90], all of them imitating specific information processing stages in the early visual system.

The need for international standards for (continuous-tone) image compression resulted, in the eighties, in the definition of the DCT-based JPEG standard [108] for still-color-image coding, the H.261 standard for low-bit-rate ISDN video services and, in the nineties, the MPEG first and second phase (MPEG-1 and MPEG-2) standards for both high quality audio and video compression [7] (JPEG and MPEG are acronyms for the Joint Photographic Experts Group and the Motion Picture Experts Group, respectively). The basis of all these techniques is the DCT, in combination with standard tools for image compression such as motion compensation and entropy coding. Transformation techniques other than the DCT, such as the lapped orthogonal transform (LOT) [56, 84] or wavelet compression remain, of course, under investigation and will be used whenever applicable. At this moment, the highest performance image coders are advanced waveform coders based on wavelet decompositions [138, 140]. An example of such a coder is Shapiro's Zerotree algorithm [11, 128], which will be used in Chapter 3 as a reference coder.

The compression research in this thesis is concerned with the concept of *second-generation coding techniques*. This concept was introduced in the eighties by Kunt et al. [74] and refers to compression techniques that make use of our knowledge on *how and what* the eye sees. These techniques should thus account for properties of vision mechanisms of the brain by analyzing images more closely to the way the visual system analyzes information and by describing images in terms of more nearly physical entities such as contours, textures and uniform regions. These content-based approaches are usually referred to as segmentation or region-based coding (Kunt [74]), contour/texture-based coding [20] or edge-based coding [80, 14]. Such an approach would then lead to more compact image representations and hence to higher compressions. Typically, at low bit rates, such compression techniques perform quite well compared to transform coding techniques such as the DCT or subband coding. For high quality image reconstruction, however, transform coding techniques outperform second-generation coders. Apparently, the exploitation of second-generation coding concepts for high quality image compression is not easy. This thesis attempts to bridge the gap between conventional transform coding techniques and second-generation coders. The research in this thesis will show that on the basis of the coefficients of the Hermite transform, an image representation can be obtained that describes local image structures (e.g., edge and line segments) in terms of primitives such as orientation, position, height, etc. In Section 1.4, second-generation coding methods are discussed in more detail.

At the present time, there is a widespread interest in model-based coding, a promising technique in the area of very low-bit-rate compression. Model-based coding achieves high compression by modeling and analyzing the objects in a (moving) scene. The parameters needed to animate the model are then used for object reconstruction. Because of applications in video-telephony and video-conferencing, up to now, most work in model-based coding has been concerned with modeling and 3D tracking of the human head and of facial expressions. For broader applications such as television, model-based coding still faces a lot of problems. In future generation model-based coders, increasing amounts of knowledge and intelligence will have to be available for the coding process [107].

1.3   Image Quality Evaluation

With the development of new digital techniques for image coding and image processing, the need for reliable methods to measure image quality has increased accordingly. Evaluation of image quality is not only necessary to select the best algorithms amongst a set of possible candidates, but also to optimize parameter settings. Although a variety of instrumental or objective image quality measures has been developed [77, 85, 12], a measure showing a high correlation with subjective scores is not yet available. Therefore, experiments with subjects are still the only reliable way of determining the perceived quality of images. In this thesis, a range of experiments for evaluating the quality of compressed images will be discussed.

Reliable and valid techniques for assessing perceptual image quality require some level of standardization. Therefore, the International Telecommunication Union (ITU), formerly CCIR, has recommended a number of scaling techniques: single stimulus scaling, double stimulus scaling and a scaling procedure based on difference judgments [21]. For these techniques, different response scales can be used, including both adjective scales using quality indications such as *excellent, good* and *bad*, and numerical scales indicating image quality or differences in image quality (e.g., *1,...,10* or *-5,...,5*). Also, experimental conditions and display characteristics are specified.

Past IPO research projects have indicated that numerical category scaling provides an efficient and valid way to assess image quality in both appreciation-oriented and performance-oriented settings [120]. The choice of an appropriate method, however, is not always trivial, since the outcome of a subjective experiment can be susceptible to effects caused by the nature and the statistics of the stimulus set [36, 117]. Today, clear guidelines on how to select the stimulus set and when to select a specific experimental set-up are still missing. Also, no rules of thumb exist for selecting appropriate data analysis techniques.

Amongst others, this thesis aims at providing the reader with some feeling for setting-up quality evaluation experiments with compressed images.

For moving video, the standards of the ITU-R can be applied to 10-second sequences [21]. In the nearby future, more sophisticated methods for measuring quality of image sequences or video will have to be developed. An example is the single stimulus continuous quality evaluation method presented in [53], which is developed for measuring the time-varying perceptual image quality of a sequence of images. In this method, subjects continuously indicate the perceived strength of image quality by moving a slider along a graphical scale.

1.4   Image Representation Based on Human Vision

The conventional way of describing an image is in terms of its canonical pixel-based representation. Other image description techniques are based on *image transformations* [111]. Such a transformation converts a canonical image representation into a representation in which specific properties of an image are described more explicitly. Image transformations are applied in order to simplify image manipulations (e.g. image enhancement) or to make image descriptions more suited for compression. In most transformations, images are locally approximated within a window by a linear combination of a number of a priori selected patterns. The coefficients of such a decomposition then provide the desired image representation.

A commonly used transformation for image compression is the discrete cosine transform (DCT). The standardized JPEG system, for example, first analyzes the original image by partitioning it into 8x8 pixel blocks and decomposing each block into 64 2D DCT basis functions. The advantage of this decomposition is that signal energy is concentrated in a small number of coefficients (frequency components). Actual image compression can be achieved by quantizing and coding the obtained DCT coefficients. After decoding, the reconstructed image is synthesized by applying the inverse DCT. This classic compression scheme, based on *image analysis* followed by *image synthesis* can, of course, also be applied using other transformations than the DCT. Schematically, this scheme is shown in Fig. 1.1.

Although conventional 2D image transformations have proved to result in useful image descriptions, especially for image compression applications, a few remarks on these transformations are in place. First, the analysis of most conventional transformations is poorly matched to the analysis carried out by the visual system in the early visual pathway. Second, classic analysis schemes are motivated by frequency separation and energy compaction, thus resulting in a (local) representation of images consisting of a DC component and a number of
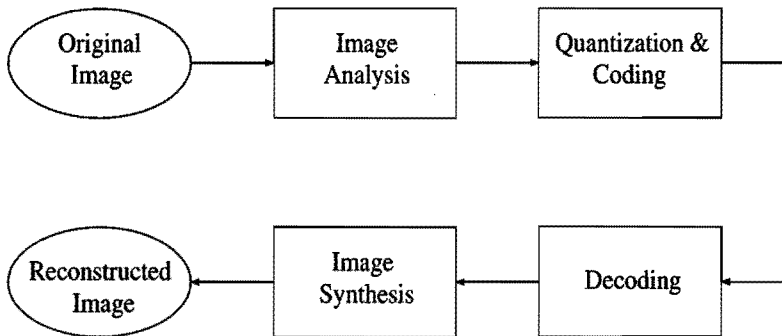
Figure 1.1: Classic transform coding scheme for image compression, based on image analysis by an energy compacting transformation and image synthesis by the inverse transformation.

2D frequency components. Natural images, however, are typically perceived as a collection of real world objects built of lines, edges and texture, rather than a collection of frequency patterns. Also, conventional transformations do not explicitly account for natural image structures that are of interest in visual perception. In a DCT representation, for example, line segments oriented along a certain angle are expected to have energy contributions in many different frequency components.

One may conclude that conventional transforms usually do not fit Kunt's second-generation concept (*how and what*). In the past, though, there have been a number of techniques that can be regarded upon as second-generation. In the late sixties, Graham [51] already constructed a coding scheme in which images were represented by a low-pass component and a "synthetic highs" component that resulted from coding explicit edge information and filtering with 2-dimensional reconstruction filters. Graham's work was mainly based on the synthetic highs coder developed by Schreiber et al. [127]. Although the link between this coding method and the human visual system is restricted to the exploitation of the lateral inhibition phenomenon, Kunt [74] considered this technique as the first example of second-generation image coding. A method that uses the idea of lateral inhibition at multiple scales is the Laplacian pyramid [19], since it splits the frequency spectrum in a number of components which are approximately one octave apart.

Directional decomposition based coding was introduced by Kunt, motivated by neurophysiological evidence of directionally sensitive cells in the human visual

system [74, 75]. In this approach, the image decomposition is performed with a 2-D filter bank producing a low-pass image and a number of directional images containing high frequency components. As the information present in directional images is concentrated in edges, these images are coded using an edge-based coding strategy.

Another coding technique that makes use of the explicit representation of features of real-world objects has been developed by Carlsson [20]. In Carlsson's approach information is coded about geometric structure and contour grey level intensities alone. Between contours, images are reconstructed by smooth interpolation of the grey level data. Mallat and Zhong developed a second generation coding algorithm based on image representation by multiscale edges [80]. In [79] they showed that, using the wavelet transform, one can represent an image completely with multiscale edges. More recently, van Beek [14] proposed a low-bit-rate compression scheme based on full image reconstruction from contour information only.

In general, only at low bit rates, the performance of second-generation coders is better than the performance of conventional transform coders. The contour coding concepts used by Carlsson [20] and van Beek [14], for instance, always result in low quality images, because no texture information can be coded. At low bit rates, however, these algorithms perform very well, because reconstructed image quality is relatively high. In Kunt's paper [74], no indications
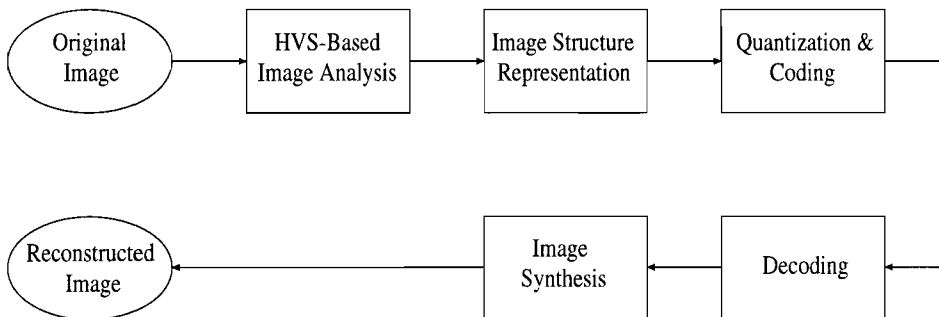
Figure 1.2: Image compression scheme based on HVS analysis. In this scheme image analysis corresponds to the measurement of visual information carried out by the early visual system. In this thesis the Hermite transform is used for image analysis. On the basis of the results of these measurements, an explicit representation of perceptually relevant image structures is obtained.

on the relative performance of the proposed coders are given at all. Also, in most of the papers, only images with pronounced edge structures are used for demonstration purposes. In this thesis, coding concepts are presented that may bridge the gap between conventional (high quality) transform coding methods and second-generation coders.

A typical second-generation coding scheme is proposed in Fig. 1.2. It incorporates a (local) scene analysis that corresponds, to some extent, to the information analysis as carried out by the visual system. It is not sufficient, however, to base coding techniques on human visual system (HVS) information analysis only. The analysis stage should be followed by representation of perceptual relevant scene elements. This means that image representations should be able to represent the specific nature of these structures (orientation, profile, scale) in a more explicit way. Classification and grouping of detected local image elements is necessary to obtain image representations that are more related to real world structures. After HVS image analysis the image should contain components that can be processed separately, depending on their characteristics and perceptual relevance (e.g. texture, contours, monotone areas). In this coding diagram, the synthesis stage does no longer have to be directly related to the analysis stage (in the sense of a forward and an inverse transformation).

1.5   Image Analysis with Hermite Transforms

The compression research in this thesis is based on the *Hermite transform*, an image transformation technique introduced by Martens in [90]. The Hermite transform uses overlapping Gaussian windows and projects images locally onto a basis of polynomials. In Fig. 1.3 Hermite filter functions are shown for a one-dimensional Hermite transform. Because the analysis filters are derivatives of Gaussians, Hermite analysis is in close agreement with the information analysis carried out by the human visual system. It has been shown by Hubert and Wiesel [58] that the visual cortex contains receptive fields with different profiles and orientations. According to the Gaussian derivative theory [155], the shapes of these fields can be approximated quite well by the derivatives of 2-dimensional Gaussians. The Hermite transform thus matches the Gaussian derivative theory [90, 155], since it models the information analysis carried out by the retinal and cortical visual receptive fields. Because receptive fields occur in varying size, each field is suited for detecting the presence of a specific spatial frequency. With a two-dimensional Hermite transform, field sizes can be modeled by varying the standard deviation σ of the Gaussian envelope, while orientation selectivity can be obtained by rotating the Hermite filters.

The principle of receptive fields with different sizes has, amongst others, been mathematically described by Witkin's *scale-space* formalism [151]. An applic-
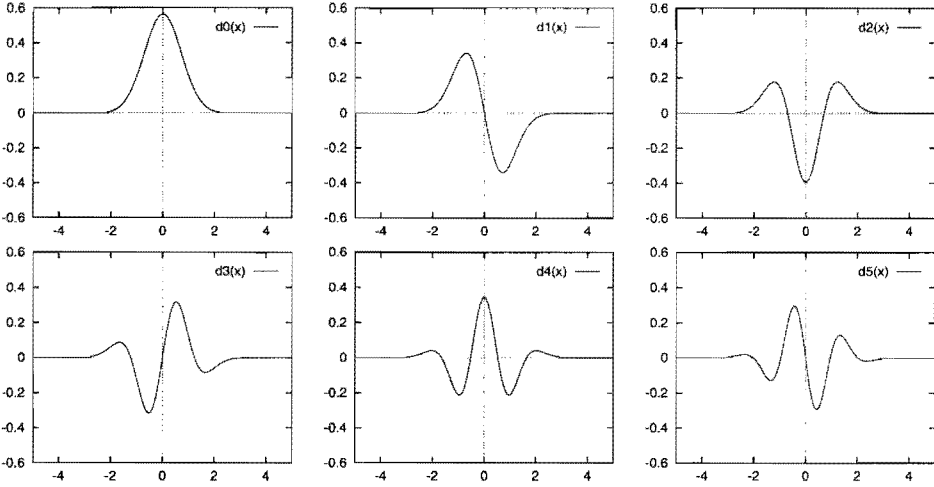
Figure 1.3:   Hermite filter functions $d_n(x)$ in the spatial domain for order $n = 0, ..., 5$ and $\sigma = 1$. Details of the Hermite transform are discussed in Chapter 3.

ation of this approach in 2D image coding can be found in [92] and is based on the observation that images typically contain structures at many different scales. For the approximation of local signals, a scale-space coder therefore uses pattern sets that differ in spatial scale and sampling distance. Of course, the scale-space approach is closely related to the Laplacian pyramid of Burt & Adelson [19]. Although the Hermite transform can very well be applied in a multi-scale coding scheme, this thesis does not address any scale-space aspects of image representations. Applications of the Hermite transform in pyramidal coding schemes can be found in [89] and [91].

This thesis shows that the Hermite transform can be applied in a straightforward way in the analysis-synthesis compression scheme of Fig. 1.1. It is also suited, however, to be used in the second-generation compression scheme of Fig. 1.2, because on the basis of the transform coefficients local primitives of perceptually pronounced image structures such as edges and lines can be extracted. Primitives of edge segments, for example, are orientation, position and height. By additional processing, these local image descriptions can be grouped and compression schemes can be enabled to exploit redundancies between adjacent local image descriptions. In this thesis, amongst others, a Hermite-based contour-coder will be presented.

1.6   Scope of this Thesis

Because of the evolution of techniques such as desktop imaging, video-phone and multimedia information retrieval, the role of visual perception in image processing has become more and more important. For data reduction, one possible approach is to study image compression techniques that make use of perceptually-motivated image representations.

This thesis aims at developing an image description technique that matches the image compression scheme proposed in Fig. 1.2. On the basis of image analysis with a Hermite transform, a content-based image description will be constructed that represents perceptually important image structures such as contours and lines, explicitly. The proposed image representation will also be included in an image compression algorithm. In this way, the efficiency and the practical use of the image representation can be evaluated. The research in this thesis is restricted to the coding of static or still greyscale images.

Furthermore, specific attention is paid to finding simple valid methods on how to measure image quality in order to evaluate the performance of a specific compression scheme. In **Chapter 2** a number of scaling methods and data analysis techniques for the subjective evaluation of the perceived image quality of compressed images are discussed. Amongst these, both direct quality scaling experiments and comparison-based experiments. Subjective evaluations are also carried out in order to compare the image compression techniques introduced in this thesis with an existing reference coder. Because it is important to use a well-known reference coder, a JPEG coder was selected for reference purposes.

Chapter 2 shows that numerical category scaling provides an efficient way to measure image quality, even if the quality range of the stimulus set is very small. It also demonstrates, however, that the outcome of a subjective experiment can be susceptible to effects caused by the characteristics of the stimulus set. Therefore, it is sometimes necessary to make use of alternative scaling methods. One of those methods is based on Anderson's functional measurement theory, and requires observers to determine quality differences between all possible combinations of coded images [118].

Chapter 3 discusses the Hermite transform and introduces a new orientation-adaptive form of this transform that is based on the steering properties of the Hermite filters. This *steered Hermite transform* enables rotation of the Hermite frequency patterns that are used for image description and integrates explicit coding of local pattern orientations. Chapter 4 will show that the steered Hermite transform constitutes a basis for bridging the gap between conventional transform coding techniques and second-generation coding techniques, because it also enables detection and explicit representation of perceptually important

image structures such as contours and lines.

In order to demonstrate the efficiency of the Hermite transform in an image compression algorithm, a compression scheme based on an orientation-adaptive steered Hermite transform is presented. Comparisons with other compression techniques such as JPEG show that the proposed scheme performs very well at high compression ratios, not only in terms of peak-signal-to-noise ratio but also in terms of perceptual image quality.

In **Chapter 4** explicit coding of contours is integrated in the Hermite coding scheme of Chapter 3. It demonstrates that at edge locations the local Hermite frequency description can be transformed into an edge parameter description, making use of (Gaussian) edge primitives such as orientation, position, edge height and edge blur. In this chapter, images are represented by contour information and low-frequency information alone. Redundancies along image contours are exploited in a contour-based image compression scheme. Compression results show, however, that the overall performance of the Hermite coding scheme presented in Chapter 3, is better than the contour-based scheme of Chapter 4. One may conclude that more research effort is still needed to improve the Hermite contour-coding scheme.

Chapter 5 focuses on the representation and reconstruction of uniform regions and demonstrates that image quality of uniform regions can be improved by spending more effort on the design of the decoder. An algorithm using adaptive DC filtering is proposed that reduces segmentation artifacts and blocking artifacts in block transform coding schemes. The proposed algorithm is particularly effective for low-bit-rate coding. The research in Chapter 5 demonstrates that for JPEG-coded images the DC filtering algorithm in combination with the JPEG AC-prediction method [108], reduces segmentation artifacts and blocking artifacts in a JPEG coding scheme. Results of DC filtering on Hermite-coded images are also given.

Chapter 6 discusses the philosophy used in this thesis and summarizes the main results. Also, suggestions for future research are indicated.

Chapter 2

Subjective Quality Evaluation of Compressed Images

Abstract

With limited effort, one can obtain valuable information on the subjective performance of image compression schemes. The choice of an appropriate experimental technique, however, is not always trivial, since both the subject's task and the experimental method itself should depend on the properties of the stimulus set (quality range, kind of artifacts). The goal of this chapter is to discuss a number of numerical scaling techniques that can be used to assess perceived image quality. It is shown how to analyze data obtained in numerical category scaling experiments and how to set up such experiments. The results of several subjective experiments illustrate that numerical category scaling techniques provide an efficient means not only for obtaining compression ratio versus quality curves that characterize coder performance over a broad range of compression ratios, but also for assessing perceived image quality in a much smaller range (e.g., close to threshold level). However, the nature of the artifacts introduced by different coders can cause problems when evaluations are carried out using direct numerical category scaling. The latter is demonstrated by comparing the results of a direct scaling method and a scaling technique in which subjects have to determine quality differences between all possible combinations of coded images.

---

[1] The research described in this chapter has been published as a paper in Signal Processing [36]. Some results were also presented at the International Conference for Image Processing ICIP-96, Lausanne [34]

2.1 Introduction

Today's large variety of algorithms for data compression has created a growing need for methods to judge (new) compression algorithms. It is generally agreed, however, that objective error measures such as the commonly used root-mean-squared error (RMSE), defined by

$$\text{RMSE} = \sqrt{\frac{1}{N^2} \sum_{x=1}^{N} \sum_{y=1}^{N} \left[f(x,y) - \hat{f}(x,y) \right]^2}, \qquad (2.1)$$

where f denotes the original NxN image and $\hat{f}$ the reconstructed image, do not always correlate well with subjective quality ratings. Although more advanced measures are being developed, experiments involving subjects are still the most reliable way of determining the perceived quality of coded images.

A well known problem in papers on image compression, especially when very high compression ratios are involved, is the lack of quality assessment. Usually, bit rates and signal-to-noise ratios are specified, but subjective quality evaluations are not carried out. Probably, researchers are unfamiliar with quality assessment techniques or they consider them to be too time-consuming. This chapter demonstrates that, with limited additional effort, one can obtain valuable information on the subjective performance of image compression schemes.

A number of evaluation techniques which can be used to measure perceived image quality is discussed. It is shown how to analyze data obtained in such experiments and how to set up the experiments. In all experiments techniques that make use of numerical category scaling are used. Although past IPO research projects [120] have indicated that numerical category scaling techniques provide an efficient and valid mean for assessing image quality, it will be shown that the choice of an appropriate experimental technique is not always trivial.

For a subjective evaluation of a compression algorithm, this algorithm must be compared with a reference coder. In this chapter a JPEG-based interpolative coding scheme is compared with a standard JPEG baseline sequential coder (the JPEG compression algorithm is based on the discrete cosine transform which is usually indicated by the acronym DCT). The different nature of the artifacts introduced by these coders causes problems when evaluations are carried out using direct numerical category scaling. This will be demonstrated by using both a direct scaling method and a scaling technique in which subjects have to determine quality differences between all possible combinations of coded images. The latter technique (based on Anderson's functional measurement theory) should be preferred in this case.

All test scenes used in the experiments were acquired from a Kodak PhotoCD demonstration disc. The advantage of using these images is that they are varied

and have high image quality. The original images with a resolution of 512x768 pixels were cut to 480x240 pixels in order to enable us to display two images simultaneously. For all the experiments we used 8-bit grayscale versions only. Fig. 2.1 shows all the original test scenes.

Section 2.2, presents a JPEG-based interpolative image compression scheme. Subjective evaluations of this scheme are discussed in Sections 2.3, 2.4, and 2.5. Section 2.3 discusses the results of coder comparison experiments carried out by means of direct numerical category scaling. In Section 2.4 an assessment technique in accordance with Anderson's functional measurement theory is used. Both methods are discussed in some detail. In particular this chapter shows how to average results over different subjects and how to interpret differences between subjects. Section 2.6 demonstrates the performance of a Hermite coder [35] (see chapter 3) by applying a multi-dimensional scaling technique. In Section 2.7, finally, some results of subjective evaluations of a standard JPEG baseline sequential coder using scene-optimized DCTune [146] quantization matrices are shown.

2.2 JPEG-DCT versus Interpolative DCT Coding

The first compression algorithm evaluated is the JPEG-based interpolative coding scheme proposed by Zeng and Venetsanopoulos in [156]. The principle of this scheme is as follows: first, the original image is low-pass filtered and subsampled. This subsampled image is then DCT-coded using a standard JPEG baseline sequential coder. The reconstructed image is obtained via DCT decoding and interpolation (upsampling and low-pass filtering). In this case a 2x2 average was used as the low-pass filter and a down-sampling factor of 2 was applied in each direction to obtain the decimated image. A 4x4 binomial filter was used as the interpolator.

In order to test the interpolative scheme a standard JPEG baseline sequential coder was used as a reference [108]. Because the interpolative coding scheme was designed to reduce blocking artifacts, the comparison was carried out with low quality images only (compression ratios approximately between 10 and 60). Fig. 2.2 presents reconstructed fragments of the "children" test image showing both the typical coding artifacts introduced by the reference coder (left image) and the impairments caused by the interpolative coder (right image). Note that JPEG-DCT coding typically introduces blocking artifacts, whereas the interpolative coding typically introduces blur.

By scaling the standard DCT quantization matrix, a set of coded images was obtained for each coder. For each test scene this resulted in twelve (2x6) coded images at various compression ratios. Fig. 2.3 shows the peak-signal-to-noise

Figure 2.1: The original Kodak images "children", "doll", "motor-bike", "sails", and "girls". The bottom right image is a cut (480x240 pixels) from the famous "lena" image.
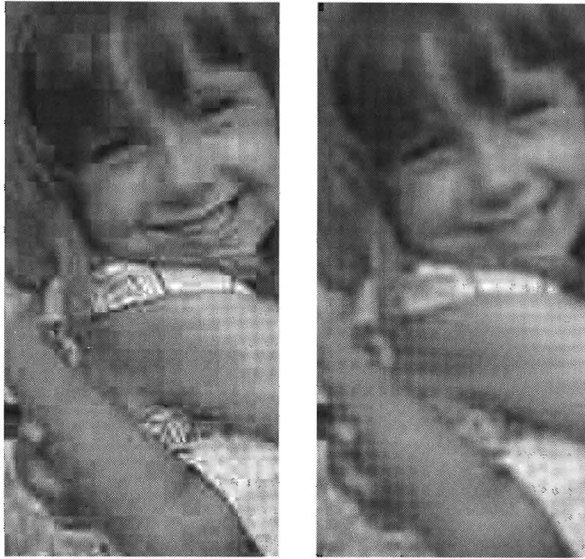
Figure 2.2: Fragments of "children" showing typical coding artifacts due to JPEG-DCT coding (C=28) and interpolative DCT coding (C=43).

ratio (PSNR) values for two test images as a function of the compression ratio C, which is defined by the number of bytes needed to represent the original image divided by the number of bytes needed to represent the compressed image. The relation between the PSNR and the root-mean-squared error (RMSE) is given in Section 3.2, Eq. (3.11). On the basis of the PSNR performance curves of a representative set of test images, it was found that the interpolative scheme typically outperforms the standard coder at compression ratios higher than 30. The same results were found in [156]. In order to get a more reliable impression of the coder performances, however, a comparison was carried out based on subjective quality judgements. This subjective evaluation, using two different category scaling techniques, is discussed in the next three sections.

2.3 Quality Assessment I (Direct Numerical Category Scaling)

Perceived image quality is often assessed by means of direct category scaling. In a category scaling experiment, subjects are asked to classify images into a number of categories. For the measurement of perceived image quality, for
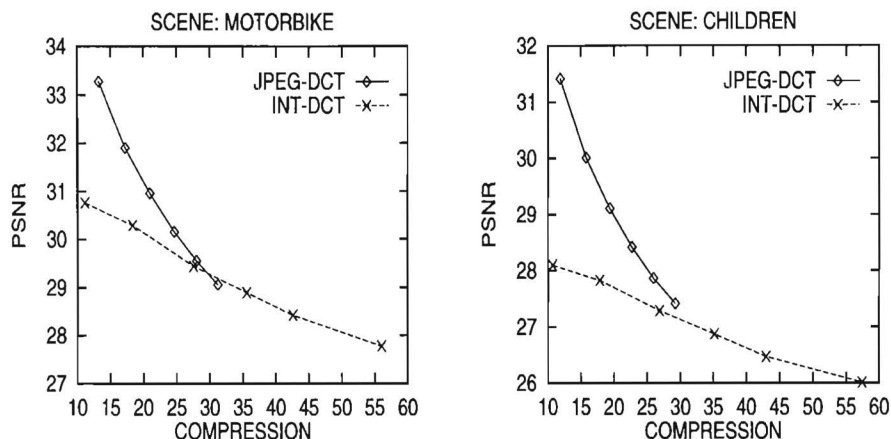
Figure 2.3: PSNR versus compression ratio curves for both compression algorithms. Results are shown for the test scenes "children" and "motorbike".

example, the CCIR [21] has recommended a 5-point scale using the adjectives *bad*, *poor*, *fair*, *good* and *excellent* . In a similar way, descriptive impairment and comparison scales have been defined.

An alternative way of identifying the points on a category scale is to label each category with a numerical value. It has been argued [116, 133] that the validity of numerical category scales is higher, because numbers divide the scale in intervals of equal size. Also, numerical scales can easily be adapted to specific ranges of image qualities. This does not mean, however, that the sensations belonging to adjacent categories are equidistant [120]. This section discusses the results of a direct numerical category scaling experiment in which the performance of the image coders mentioned in Section 2.2 is compared. It is shown how to set up the experiments and how to analyze the data obtained in such experiments.

2.3.1 Method

Subjective quality evaluations were carried out for the test scenes "children" and "motorbike". For stimulus display, a 50 Hz non-interlaced BARCO CCID7351B monitor was placed in a dark room against a white, dimly lit 2.5 cd/m2 background. The subjects observed the coded images from a dis-

tance of 160 cm, corresponding to a ratio of 6:1 between viewing distance and monitor height. All viewing conditions satisfied CCIR recommendation 500-3 [21]. On the monitor, each stimulus occupied an area of 25 x 12.5 cm, resulting in a horizontal size of 4.5 degrees of visual angle and a resolution of 53 pixels per degree of visual angle.

Two male and two female subjects between 24 and 28 years of age participated in the experiments. All subjects had experience with quality evaluations of processed natural images. They had normal or corrected-to-normal vision and a visual acuity, measured on a Landolt chart, between 1.5 and 2.0. The subjects did not have any knowledge of the way in which the presented images had been processed.

In one session, all 12 reconstructed images of one test scene were presented 4 times in a random sequence. Separate scenes were handled in separate sessions. All 48 stimuli were displayed on the monitor for 5 seconds, followed by an homogeneous 11 cd/m2 adaptation field which lasted at least 2 seconds (the luminace of this field matched the average luminance of the test scenes). The subjects were asked to rate the overall perceived quality of each image using a numerical quality scale ranging from 1 to 10. Before starting a session, subjects had to judge a training sequence in order to get an idea of the quality range of the images. This randomly displayed training sequence contained all 12 coded images.

In a quality evaluation experiment, the duration of the stimulus presentation should provide subjects enough time to perform their task. A too long or unlimited presentation time might bias the judgment of the subjects and results in unnecessary long experimental sessions and different experimental conditions across subjects. Since the optimum presentation time depends on the quality range of the stimuli and the number of stimuli displayed at the same time, all presentation times mentioned in this thesis were determined by carrying out some short pilot experiments.

2.3.2   Z-score Results

Quality scores for each image were obtained by taking the mean of the four scores corresponding to that particular scene. Before averaging over all subjects, however, it is common to minimize the variation between the individual quality scores. This variation is caused by the fact that not all subjects use the full range of the numerical scale in classifying images. A simple method that normalizes a set of individual scores on the mean and standard deviation of those scores, is the z-score transform described by Hays [55]. This transform thus converts each score into a z-score that indicates the deviation from the mean score in standard deviation units (see Appendix 2.A for the exact
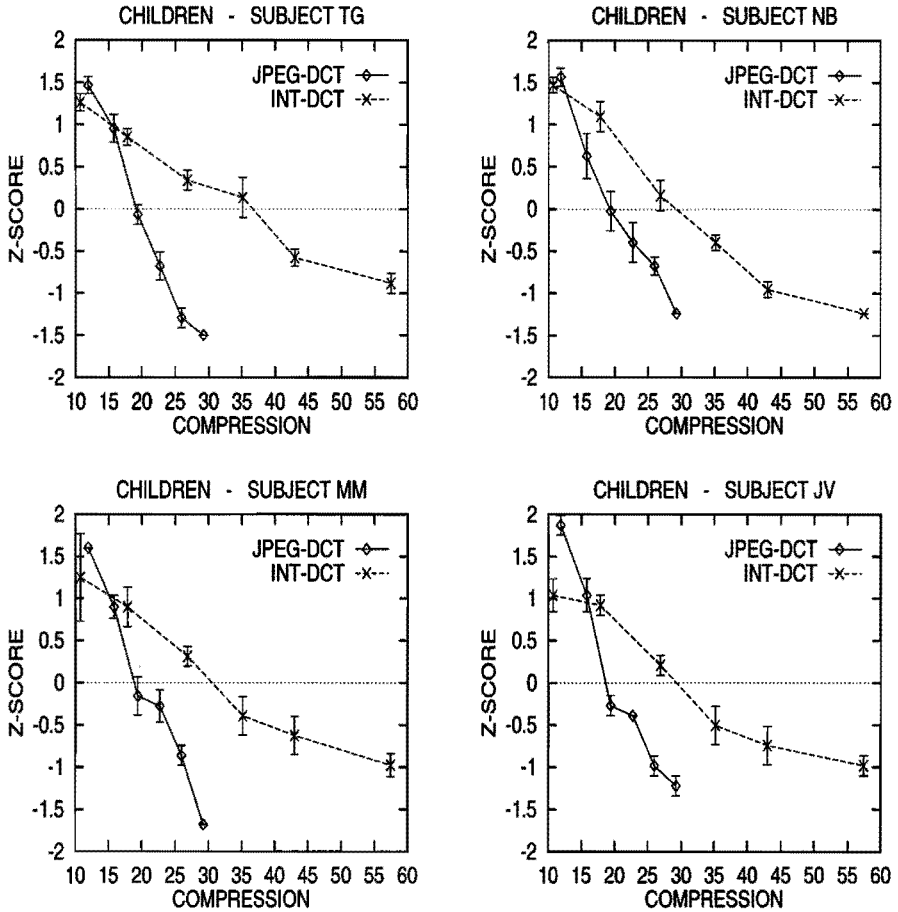
Figure 2.4: Individual z-scores obtained via direct numerical scaling. The plots indicate the perceived image quality of the "children" test scene as a function of the compression ratio. The lengths of the error bars are equal to twice the standard error of the mean.

definition of z-scores). It should be noted that the name z-score transform is somewhat confusing, because this transform has no relation with the well-known z-scores used in statistics). In Fig. 2.4 the z-scores of all individual subjects are shown for the "children" scene. Similar results were obtained for the "motorbike" scene.

In Fig. 2.5 the averaged z-scores have been plotted for both test scenes. The
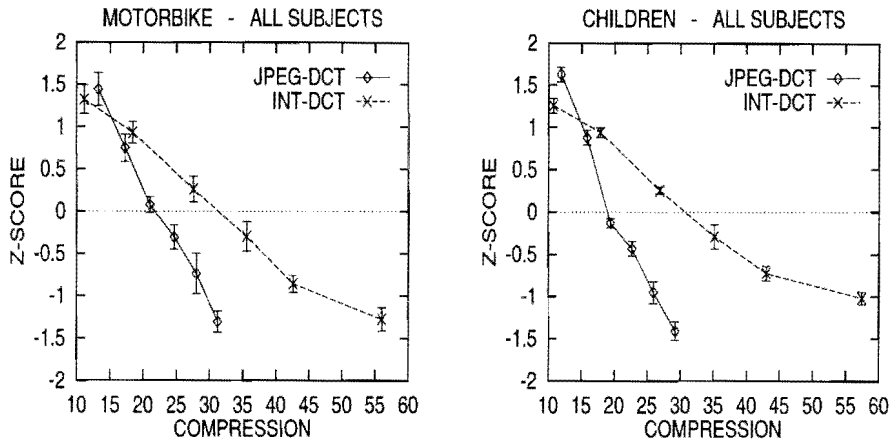
Figure 2.5: Perceived quality versus compression ratio curves for the two coders. The results were obtained by averaging over all individual z-scores. The error bars indicate the standard error of the mean calculated between the subjects.

error bars are equal to the standard error of the mean calculated over the mean scores of each subject. Since the lengths of the error bars are very small, the differences between the subjects are not very pronounced. Comparison of these results with the PSNR curves of Fig. 2.3 shows that for both scenes the equal-performance points of the algorithms with respect to the perceptual quality (C=16 and C=15 for the scenes "motorbike" and "children", respectively) have shifted considerably. From these plots, one would conclude that the interpolative coding scheme performs even better than could be expected on the basis of the PSNR curves. Section 2.4, however, demonstrates that these results are biased because of the difference in nature of the artifacts present in the reconstructed images.

Although averaging over individual z-scores does make sense, it implies that differences between individual subjects are not explicitly taken into account. Not only are all subjects given the same "weight", but averaging over subjects also removes information about differences that may exist between subjects. Section 2.4 shows that these differences can be significant and meaningful. Therefore, analysis and interpretation of individual differences should always be part of a subjective evaluation experiment. In the next paragraph a method for data analysis based on Thurstone's category scaling model is discussed. This Thurstone scaling method does not only facilitate the interpretation of

individual differences, but also provides a more perceptual view on the process of category scaling.

2.3.3 Thurstone's Law of Categorical Judgement

In literature, results of numerical category scaling are not always presented in a raw (or normalized) form. An alternative way of analyzing and presenting data obtained in scaling experiments is to transform the data into an interval scale assuming a psychologically linear continuum. A technique for such a transformation, based on Thurstone's law of categorical judgement is, for instance, described by Torgerson in [135].

The ideas underlying Thurstone's law of categorical judgement can be described as follows. Thurstone assumes that every stimulus generates some momentary impression on a psychological continuum, a psychological attribute, of which the strength is characterized by a specified number of ordered categories. The response to a given stimulus however is stochastic, as both the position of a stimulus impression and the positions of the category boundaries are Gaussian distributed. In Fig. 2.6 this concept of stimulus distributions and boundary distributions is visualized. Thurstone's law of categorical judgement now consists of a set of equations that describe the judgement of an observer when asked to place stimuli into a number of ordered categories. The complete form of the law of categorical judgment [135]:

$$t_g - s_j = x_{jg}\sqrt{(\sigma_j^2 + \sigma_g^2 - 2r_{jg}\sigma_j\sigma_g} \qquad (2.2)$$

with $j = 1, ..., n,$; $g = 1, ..., m$; $m + 1$ the number of categories and

$$
\begin{aligned}
s_j &= \text{scale value of stimulus } j \\
t_g &= \text{mean location of the } g\text{-th category boundary} \\
\sigma_j &= \text{standard deviation of the } j\text{-th stimulus sensation} \\
\sigma_g &= \text{standard deviation of the } g\text{-th category boundary} \\
r_{jg} &= \text{correlation between momentary positions of stimulus } j \text{ and} \\
    &\quad \text{category boundary } g \\
x_{jg} &= \text{unit normal deviate corresponding to the proportion of times} \\
    &\quad \text{stimulus } j \text{ is sorted below boundary } g.
\end{aligned}
$$

Unfortunately, the general law of categorical judgement is not solvable in its complete form, so when this law is applied, additional constraints must always be set. In this chapter, a model (Thurstone's model I) is applied that involves
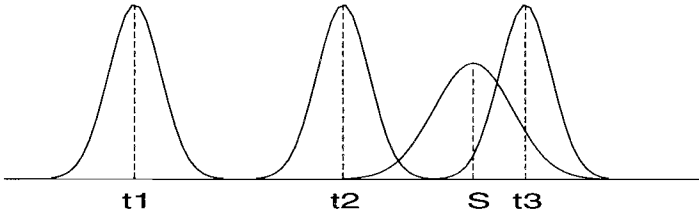
Figure 2.6: Position distributions on a psychological continuum for a stimulus S and category boundaries t1, t2 and t3.

replications over trials within one subject with constraints of condition D (for details see [135]). In condition D, the number of parameters is limited by assuming that 1) the standard deviations of the category boundaries σ_g as well as σ_j of the stimulus locations are constant, and 2) the correlation r_{jg} between the momentary positions of a stimulus and a category boundary is also constant for all values of j and g. Equation 2.2 then reduces to

$$t_g - s_j = x_{jg}c, \qquad (2.3)$$

with $j = 1, ..., n$ and $g = 1, ..., m$.

For all subjects, the raw scaling data obtained in the previously described experiment were transformed to a Thurstone scale using the in-house software package TurcatD. The output of this program includes maximum likelihood (ML) estimates of the mean stimulus positions, the standard deviation of the stimulus distributions and some stress value indicating the goodness of fit of the scaling data (the standard deviation of the category boundaries is set to zero). Since the experiment involved replications of stimuli over one subject (Torgerson calls this a model I experiment), each psychological scale calculated reflects the behaviour of one particular subject for one particular test scene. The stimulus standard deviation estimated for a subject can be regarded as a measure of the consistency of that subject. It should be noted, however, that this standard deviation is also influenced by the range of the numerical scale values used by that subject (some subjects do not use the full range of the category scale).

Because the psychological scale constructed in Thurstone scaling is considered to be a true interval scale, equal differences in the percept judged are reflected by equal distances on the scale. Direct interpretation of individual results, however, is difficult since it should be noted that a Thurstone scale is determined but for a linear transformation, and thus offset and scale may be chosen

freely. In order to ease interpretation of individual subject behaviour, we normalized all Thurstone scales by subtracting the mean value of the Thurstone scores and rescaling so that one unit on the quality axis corresponded to the estimated stimulus standard deviation of a subject.
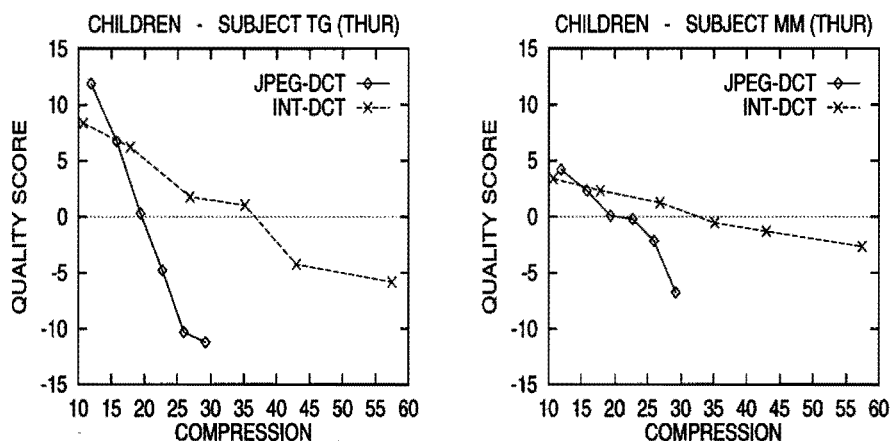
Figure 2.7:   Normalized Thurstone scores expressed in stimulus standard deviation units for the subjects TG and MM. Results are shown for the "children" test scene.

On such a normalized psychological scale, differences between observers become clear. Fig. 2.7 shows the normalized individual results of subjects TG and MM for the "children" scene. The difference between the stimulus standard deviation at the psychological scale estimated for the subjects TG and MM (0.24 and 0.38, respectively) is reflected by the larger range of quality scores calculated for subject TG. These standard deviations are influenced by the scaling range and the consistency of the subject. Also, it can be seen that there is some difference between the specific shapes of the curves obtained for both subjects.

Averaging over quality scores expressed in units of standard deviation implies that each subject is assigned a different weighting factor. The higher a subject's standard deviation, the smaller the weighting factor assigned. This corresponds to the idea that more consistent subjects are more reliable and should be assigned a higher weighting factor. One may compare the normalized individual Thurstone scores of the subjects TG and MM in Fig. 2.7 to the individual z-scores in Fig 2.4. The shapes of the curves only show small differences. Fig. 2.8 shows this "weighted" average of Thurstone scaling results for both test
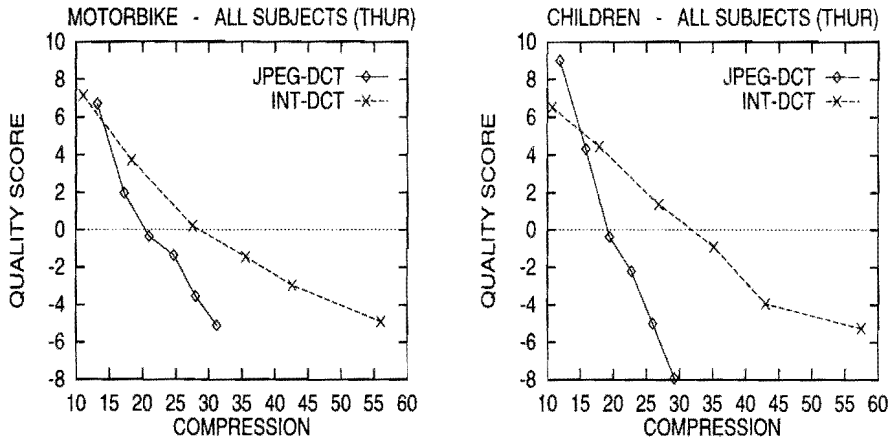
Figure 2.8:   Averaged Thurstone scaling results indicating overall perceived quality of the test scenes "motorbike" and "children".

scenes. It can be seen that compared to the performance curves in Fig. 2.5, only small changes have occurred (especially for the "motorbike" scene). In the next section, instead of using direct numerical category scaling, a second quality evaluation method, based on functional measurement theory, is demonstrated.

2.4   Quality Assessment II (Functional Measurement Theory)

The method of direct numerical category scaling has proved to be an efficient tool for image quality assessment [120]. One may doubt, however, the validity of quality scores derived from a direct scaling experiment when comparing the performance of two coders that cause artifacts of a very different nature. In image quality research, it is known that if multiple scenes have to be judged in a session of a direct category scaling experiment, subjects may use a separate scale for each of those scenes. It is possible, however, that in the case where only one scene, but multiple types of distortion are presented, subjects will also build up separate internal quality scales. If that were indeed to be the case, it would not be possible to compare direct quality ratings obtained for two different coders.

A possible solution to this problem would be to force subjects to link the quality ratings for both coders. An approach that achieves this is offered by Anderson's Functional Measurement Theory (FMT) [5]. In this approach, image qualities are compared rather than separately evaluated. Although functional measure-

ment theory was originally developed as a broad framework for the efficient description of non-observable psychological processes underlying the comparison of stimuli, it can be used quite efficiently for quality assessment of coded images [116]. An essential assumption in functional measurement theory is that when two different independent stimuli Φ_i and Φ_j which evoke sensations s_i and s_j, respectively, these sensations are combined to form some internal or psychological response Ψ_{ij}. It is also assumed that subjects use simple rules such as addition, multiplication and subtraction when combining the sensations s_i and s_j into Ψ_{ij}. Subsequently, this psychological response is mapped onto an overt response R_{ij}, which is usually a number on a category scale. Anderson argued that the mapping from the stimuli to the evoked sensations, the combination function from the sensations into Ψ_{ij} and the judgement mapping into R_{ij} can be determined simultaneously if all stimuli are presented in a factorial design. Under the assumption that a subtractive combination rule is used when subjects are asked to judge differences in quality and that a linear judgement function is used, then the factorial plot of the overt responses will consist of a set of parallel curves and the marginal means of the factorial design represent the sensation strengths on an interval scale.

2.4.1   Method

In the FMT experiment the test scenes "motorbike" and "children" were again used for evaluation. The subjects were the same as those who participated in the direct scaling experiment. No changes were made in the viewing conditions, except for the fact that now each stimulus contained two images, simultaneously displayed on the left and on the right hand sides of the screen, with a spacing of 30 pixels (about 1.5 cm). The stimulus presentation time was extended to 6 seconds in order to give the subjects enough time to observe both images.

All 12 stimuli (6 for each coder) were factorially combined to form 144 stimulus pairs. The subjects were asked to rate the difference in quality between the two images using a scale ranging from -10 to +10. The plus and the minus signs were used to indicate whether the left or the right image was preferred. Again, a training session of 12 randomly selected image pairs was presented to each subject before the start of the actual experiment.

2.4.2   Experimental Results

For each subject, one 12x12-element matrix was obtained for each test scene (one row and one column per stimulus), with elements (i, j) representing the score given by the subject for the difference in quality between the pair of
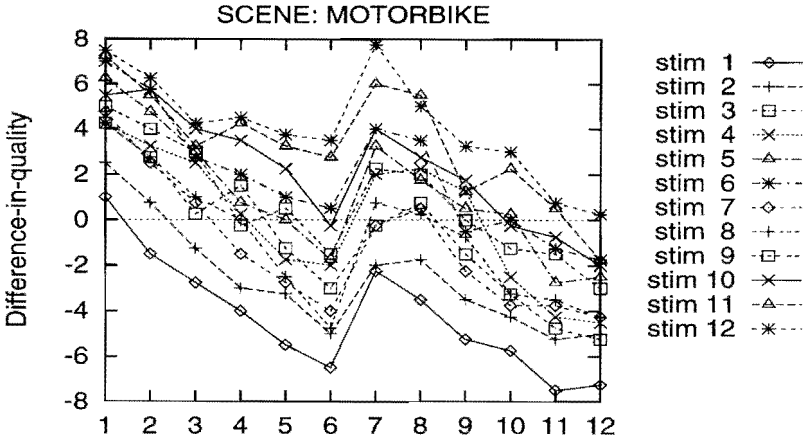
Figure 2.9:   Difference-in-quality scores, averaged over all the subjects. The numbers 1-6 and 7-12 on the x-axis correspond to the JPEG-DCT and the interpolative-DCT images, respectively. Each point on the curves indicates the score for the difference-in-quality between the corresponding row and column stimuli in the 12x12 stimulus matrix. Note the parallelism of the curves.

stimuli, stimuli i and j being displayed on the right and the left hand sides of the screen, respectively. In Fig. 2.9 the scores in this 12x12 matrix averaged over all subjects are plotted. In this plot, the lowest curve corresponds to stimulus 1, which is the highest quality JPEG-coded image. Note that the curve shows that the other 11 stimuli are all rated as having lower quality in each comparison. The scores within the different rows and columns form approximately parallel curves, which is a necessary condition to apply the FMT method.

Although parallelism can be verified graphically, it was also tested statistically by calculating the interaction between rows and columns by means of a standard two-way analysis of variance method [22]. Since no significant interaction was found, it is now possible, according to FMT, to determine a quality score for each stimulus by averaging (with opposite sign) the row and column means of the matrix that correspond to that stimulus. This was done for each stimulus. A general quality score for all subjects can now be obtained by averaging the individual quality scores. Before being averaged over the subjects, the individual scores were normalized using a z-score transform [55] (see Appendix 2.A). Fig. 2.10 shows the averaged FMT quality curves for both test

scenes. The FMT curves differ considerably from the quality curves obtained using the direct scaling method (Fig. 2.5). Individual FMT curves for the test scene "children" are given in Fig. 2.11.
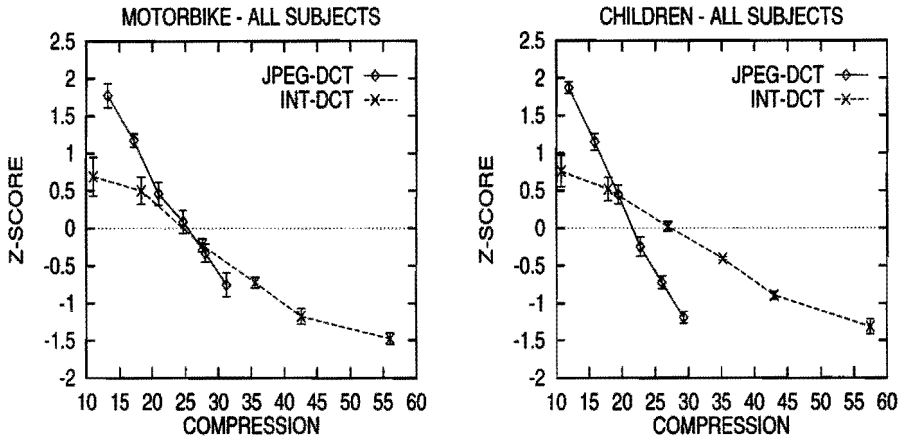
Figure 2.10: Subjective quality curves obtained by using Anderson's FMT.

In comparison with the PSNR curves, the equal-quality point has shifted slightly towards the lower compression ratios. This shift, however, is much smaller than the effect that was found in the direct category scaling experiment.

In Table 2.I, for each individual subject, the equal-performance points of the two coding algorithms are given for both the direct scaling experiment and the comparison-based experiment. Although there are some differences between the subjects, it can be observed that these individual data all show the same effect: equal-performance points found for the two methods differ considerably. Also, note the similarity between the PSNR curve calculated for the "motor-bike" scene (see Fig. 2.3) and the quality curve obtained using the functional measurement method (Fig. 2.10). In the case of this image, the objective PSNR measure performs quite well. For the "children" image, however, the PSNR curves of Fig. 2.3 and the subjective quality curves differ significantly.
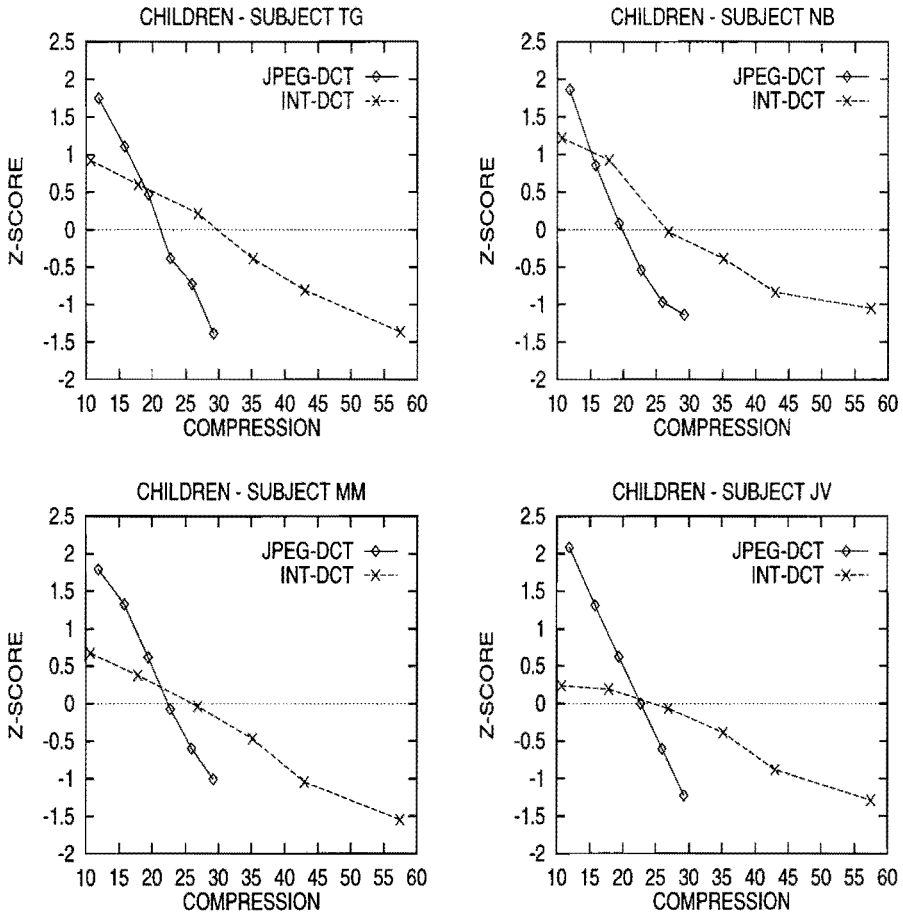
Figure 2.11: Individual FMT quality scores of all subjects for the test image "children".

	Compression at equal-quality point			
	Motorbike		Children	
Observer	Direct Scaling	FMT	Direct Scaling	FMT
mm	C=12	C=18	C=16	C=22
tg	C=12	C=19	C=16	C=19
nb	C=19	C=35	C=13	C=15
jv	C=18	C=28	C=17	C=22
average	C=16	C=27	C=15	C=20

Table 2.I: Compression ratios at the equal-quality points resulting from both the direct category scaling method and FMT. Results are shown for the test scenes "children" and "motorbike".

2.5   Functional Measurement Theory versus Direct Category Scaling

The experimental results of Sections 2.3 and 2.4 clearly show that direct numerical scaling and functional measurement theory (FMT) do not necessarily lead to identical results. What could be a problem is that reconstructed images from the two coders can be easily identified. This stimulus identification leads directly to problems in the sense that observers may use separate quality scales for the two coders. The nature of the artifacts is also an important factor. Blocking, for instance, is always perceived by a subject as annoying. Blur, however, is a more "natural" artifact, which is only recognized when a blurred image is compared with an image with less blur. Comparison of Figures 2.5 and 2.10 shows that in Fig. 2.5 the quality curves of both test scenes have been "stretched" to fit into the full quality range. An hypothesis is that instead of judging the images at the level of overall image quality, subjects judge the images at the level of a specific artifact. With Anderson's functional measurement theory, however, subjects are forced to link the quality scales for both coders. Therefore, direct scaling should not be used in a coder comparison experiment if the reconstructed images from the various coders are characterized by very different artifacts.

Of course, this does not mean that the method based on functional measurement should always be preferred. De Ridder and Majoor [116], for instance, carried out some experiments to measure impairment of perceptual image quality due to quantization errors in a scale-space coding algorithm. They found that simple direct category scaling and the scaling procedure in accordance with functional measurement theory gave rise to the same functional relation-

ship between perceptual image quality and quantization error. This means that direct category scaling is very well applicable, for instance, in experiments in which different parameters of one specific coder are varied. Also, in some cases it is physically impossible to use functional measurement theory, because it requires two stimuli to be displayed at the same time.

It was already mentioned that direct scaling experiments are much less time-consuming than complete comparison experiments. An intermediate solution when comparing different coders would be to perform a direct scaling experiment to obtain several "separate" scales, and to link these scales by carrying out an additional comparison experiment using a small subset of the stimuli. All scales previously obtained by direct category scaling can then be linearly transformed so that they can be compared with each other. Another solution would be the use of metric multi-dimensional scaling techniques that allow comparison experiments in which only a subset of all image pairs has to be judged. Multi-dimensional scaling is discussed in the next section.

2.6 Multi-dimensional Scaling

In the FMT experiment of Section 2.4, the subjects were asked to indicate whether they preferred the left or the right image. By applying metric multi-dimensional scaling (MDS) techniques [126], however, preference information can be omitted, and quality scales are constructed using dissimilarity data only. Originally, multi-dimensional scaling methods were developed based on non-metric techniques using ordinal relations obtained via comparison of pairs [119]. The problem solved by MDS techniques is similar to constructing the map of a country, given the distances between all the cities. Usually, in MDS analysis, an n-dimensional configuration (n being specified by the user) is calculated through statistical maximum-likelihood estimation.

The method used in this thesis is a so called metric MDS technique, which assumes that the perceived differences in quality can be represented on a ratio scale, and hence that the distances between the stimuli in the MDS space are proportional to the difference in quality judgments. Non-metric techniques, on the other hand, assume ordinal scaled data, which implies that there is only a monotonic relation between the distances in the MDS space and the difference judgments. The non-metric MDS techniques are very time-consuming, because they require a large number of observations. Also, the comparison of stimulus pairs requires four images to be displayed at the same time. It is beyond the scope of this thesis, however, to explain in detail how MDS works. A detailed overview of MDS techniques can be found in [126].

The power of MDS techniques is demonstrated by using the difference-in-

Figure 2.12:   Stimulus positions in a two-dimensional space calcu-
lated by Multiscale from the difference-in-quality scores for the twelve
reconstructed images of the "motorbike" test scene. The images of
each coder are indicated by separate symbols. Since the x-axis corres-
ponds to the principal component, the position on this axis indicates
the perceived image quality of each image (in arbitrary units).

quality data obtained in the FMT experiment of Section 2.4 as input for the
MDS "Multiscale" software package. In the "Multiscale" program, a Euclidean
distance measure is used. Based on the difference-in-quality scores given by the
subjects, a two-dimensional MDS space which was calculated by "Multiscale"
for the "motorbike" image. The result is shown in Fig. 2.12. Quality scores
are obtained by taking the principal component (dimension 1) of the two-
dimensional MDS space (in the case of ideal subjects giving perfectly consistent
scores, all stimuli should have been positioned on a straight line). After nor-
malization by a z-score transform, these scores were plotted against the FMT
quality scores in Fig. 2.13. In this figure it can be seen that both data ana-
lysis methods yield similar results (a correlation of 0.993 was found). In [119],
coinciding data are also found for (fast) direct category scaling and a method
of non-metric scaling.

Although MDS analysis can, of course, be used to determine quality scores, it
also provides a way of getting a better grip on image quality understanding.
In [66], for example, MDS techniques are used to characterize images degraded
by blur and noise by locating them in a multi-dimensional space. This section
presents the results of an experiment using MDS to determine the perform-

Figure 2.13:   Comparison between quality scores obtained by FMT and multi-dimension scaling. Results are shown for the "motorbike" test scene.

ance of the orientation-adaptive 2D Hermite coders discussed in [35]. Again, a JPEG-DCT coder was used as a reference. For each test scene, 12 coded images were generated: 6 JPEG-coded images and 6 Hermite-coded images. Fig. 2.14 shows the PSNR curves obtained for the 4 test scenes used in the experiment.

2.6.1   Method

Subjective quality evaluations were carried out for 4 test scenes ("girls", "motorbike", "doll", and "sails"). For stimulus display, a 50 Hz non-interlaced BARCO CCID7351B monitor was placed in a dark room against a white, dimly lit 2.5 cd/m2 background. The subjects observed the coded images from a distance of 100 cm. Since each stimulus occupied an area of 25 x 26.5 cm on the monitor, the ratio between viewing distance and stimulus height was 4:1, resulting in a horizontal size of 15 degrees of visual angle and a resolution of 34 pixels per degree of visual angle. All further viewing conditions satisfied CCIR recommendation 500-3 [21].

Four subjects between 26 and 29 years of age (3 male, 1 female) participated in the experiments. All subjects had experience with quality evaluations of processed natural images. They had normal or corrected-to-normal vision and a visual acuity, measured on a Landolt chart, between 1.5 and 2.0. Only one of
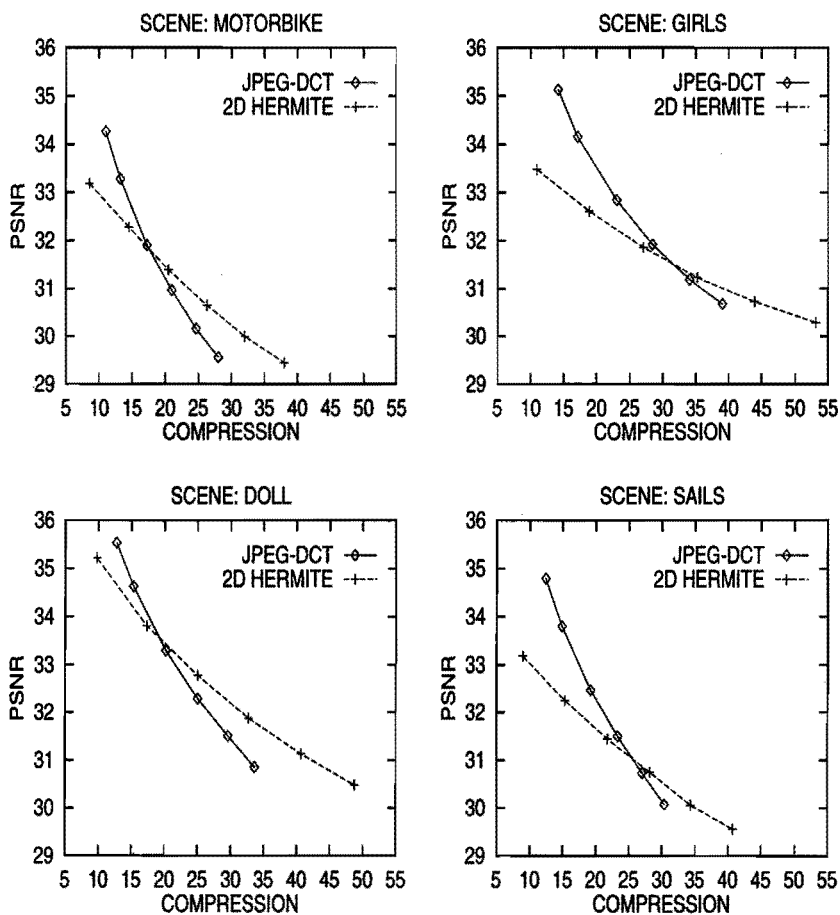
Figure 2.14:   PSNR versus compression ratio curves for a Hermite coder. JPEG-DCT was used as a reference coder. Results are shown for 4 test scenes.
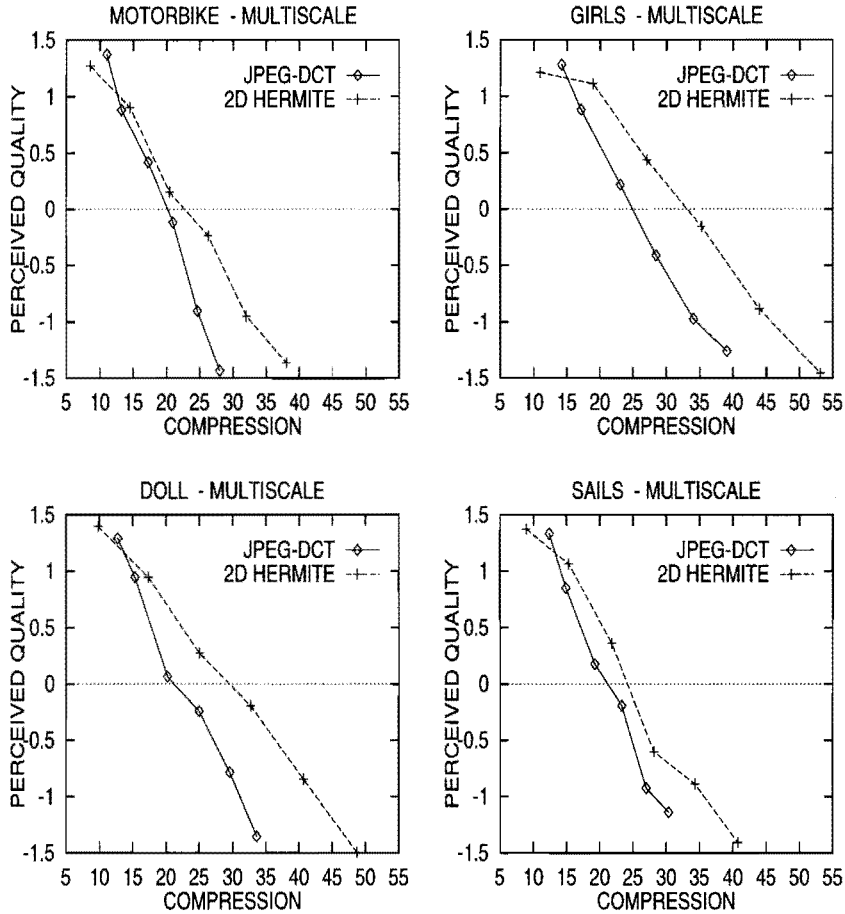
Figure 2.15:   Perceived quality curves obtained with "multiscale" for both the Hermite-coded images and the JPEG-DCT-coded images.

the subjects had detailed knowledge of the way in which the presented images had been processed.

All 12 stimuli (6 for each coder) were factorially combined to form 36 different image pairs, each containing one JPEG-coded image and one Hermite-coded image. Since all pairs were displayed twice, 72 stimuli were presented per session. The subjects were asked to rate the difference in quality between the two images using a scale ranging from 0 to 10. A training session of 12 randomly selected image pairs was presented to each subject before the start of the actual experiment. The stimulus presentation time was 8 seconds.

2.6.2   Results

Fig. 2.15 shows the quality curves produced by "Multiscale". For every scene, the output of "Multiscale" was z-score transformed and plotted against the compression ratio. It can be observed that the curves representing perceived quality differ considerably from the corresponding PSNR curves. In terms of coding artifacts, part of this result can be explained by the fact that, especially at compression ratios up to 30, the visibility of blocking artifacts is more pronounced in DCT-coded images than in Hermite-coded images. Hermite-coded images, however, are less sharp than DCT-coded images. Apparently, the introduction of artificial image structures greatly affects perceived image quality. A similar effect is found in the experiment described in Section 2.7, in which DCTune-coded images are clearly preferred over standard JPEG-DCT-coded images.

In the previous sections, a range of quality evaluation methods based on numerical scaling have been presented in combination with appropriate data processing techniques. In the examples discussed in these sections, however, the quality differences between the images used for evaluation were rather pronounced. Section 2.7 demonstrates that numerical scaling can also be used in order to measure very small differences in quality.

2.7   JPEG Coding Using Optimized Quantization Matrices

In JPEG-DCT coding, the quality and bit rate of a compressed image are determined by the quantization matrix which defines the reconstruction levels for all transform coefficients. Usually, quality and bit rate are varied through scalar multiplication of some "standard" JPEG quantization matrix. Proper design of the quantization matrix, however, can improve the performance of a DCT coder significantly. This section shows some results of an experiment in which the effect of scene-optimized quantization matrices on perceived image

quality is evaluated. The main goal is to demonstrate that scaling experiments can also be carried out when differences between reconstructed images are fairly small. All scene-optimized matrices we used were produced using Watson's "DCTune" technique [146, 147].

Optimized quantization matrices can be constructed by applying models that describe the visibility of quantization artifacts. The DCTune matrices calculated for our experiments were designed using a model which incorporated the visibility of DCT frequencies, the effects of display resolution and luminance characteristics, frequency summation, spatial summation, and contrast masking. Although optimized matrices were calculated for grayscale images only, it should be noted that the DCTune technique has also been extended to color images [147]. It is not within the scope of this thesis, however, to discuss the DCTune technique in detail. Fig. 2.16 shows both the standard JPEG quantization matrix and a scene-optimized DCTune matrix for the "children" test scene. The latter matrix was designed for a bit rate of approximately 1 bit/pixel, a display resolution of 34 pixels per degree of visual angle, and a mean luminance of 17 cd/m2.

$$
\begin{array}{|rrrrrrrr|}
16 & 11 & 10 & 16 & 24 & 40 & 51 & 61 \\
12 & 12 & 14 & 19 & 26 & 58 & 60 & 55 \\
14 & 13 & 16 & 24 & 40 & 57 & 69 & 56 \\
14 & 17 & 22 & 29 & 51 & 87 & 80 & 62 \\
18 & 22 & 37 & 56 & 68 & 109 & 103 & 77 \\
24 & 35 & 55 & 64 & 81 & 104 & 113 & 92 \\
49 & 64 & 78 & 87 & 103 & 121 & 120 & 101 \\
72 & 92 & 95 & 98 & 112 & 100 & 103 & 99
\end{array}
\qquad
\begin{array}{|rrrrrrrr|}
11 & 10 & 11 & 17 & 28 & 63 & 143 & 255 \\
10 & 13 & 12 & 15 & 22 & 41 & 93 & 255 \\
11 & 12 & 24 & 30 & 41 & 69 & 178 & 255 \\
17 & 16 & 31 & 60 & 85 & 145 & 255 & 255 \\
31 & 24 & 46 & 103 & 255 & 255 & 255 & 255 \\
65 & 47 & 78 & 163 & 255 & 255 & 255 & 255 \\
169 & 95 & 138 & 255 & 255 & 255 & 255 & 255 \\
255 & 255 & 255 & 255 & 255 & 255 & 255 & 255
\end{array}
$$

Figure 2.16: Standard JPEG quantization matrix (left) and a DCTune matrix (right), designed for the "children" test scene. Note that the DCTune matrix typically compresses the high frequency coefficients, while the DC coefficient is less severely quantized.

DCTune matrices were calculated for two different bit rates for four test scenes. One matrix for high-quality compression (about 0.8 - 1.0 bits/pixel) and one matrix for lower-quality compression (about 0.4 - 0.5 bits/pixel). All test scenes were then compressed with a standard JPEG baseline sequential coder using these DCTune matrices. Subsequently, for each of the eight DCTune-coded images obtained, five JPEG comparison images were generated using quantization matrices, obtained through scalar multiplication of the standard JPEG quantization matrix. Each set of five JPEG comparison images was generated

so that one of these images matched the compression ratio of the corresponding DCTune-coded image. In Fig. 2.17 PSNR versus compression ratio curves in the lower quality range have been plotted for all test scenes. It can be seen that for all these scenes, according to the PSNR measure, the standard JPEG matrix outperforms the DCTune matrix.

2.7.1   Method

In order to be able to evaluate experimentally the perceptual quality of each DCTune-coded image compared to the quality of the five corresponding JPEG images coded with the standard quantization matrix, a so-called "constant stimulus" method was used. In this technique, stimuli are used that consist of two images: one "constant" image that occurs in every stimulus (in this case the DCTune-coded image), and one comparison image which is varied (the five corresponding JPEG-coded images were used as comparison images). The first experiment that was carried out involved the performance assessment of the four DCTune matrices designed for the higher compression range. Although each DCTune-coded image already functions as the "constant" image, it was used as an additional sixth comparison image.

Except for the viewing distance and the stimulus presentation time, which were set to 100 cm and 9 seconds, respectively, no changes were made in the viewing conditions. Since each stimulus occupied an area of 25x26.5 cm, this resulted in 15 degrees of visual angle, and a ratio of 4:1 between viewing distance and stimulus height. The display resolution was 34 pixels per degree of visual angle. The mean luminance of the images displayed on the monitor, having a peak luminance of 60 cd/m2 and a gamma of 2.5, varied between 5 and 17 cd/m2.

Two male and two female subjects between 22 and 28 years of age participated in the experiments. All subjects had experience with quality evaluations of processed natural images. They had normal or corrected-to-normal vision, and a visual acuity, measured on a Landolt chart, between 1.5 and 2.0. Only one of the subjects had detailed knowledge of the set-up of the experiment and the kind of stimuli that were used.

The subjects' task was to rate the difference in quality between the two displayed images, using a numerical category scale ranging from -5 to +5. The plus and minus sign indicated whether the left or the right image was preferred. The stimuli were presented in a random sequence, containing stimuli of all four test scenes. Because each DCTune-coded image was combined with 6 comparison images, and each stimulus combination was repeated 6 times, the number of image pairs to be compared in one session was 144. The experiment was balanced in the sense that each DCTune-coded image occurred an equal number of times on the left and on the right side of the screen. A training set
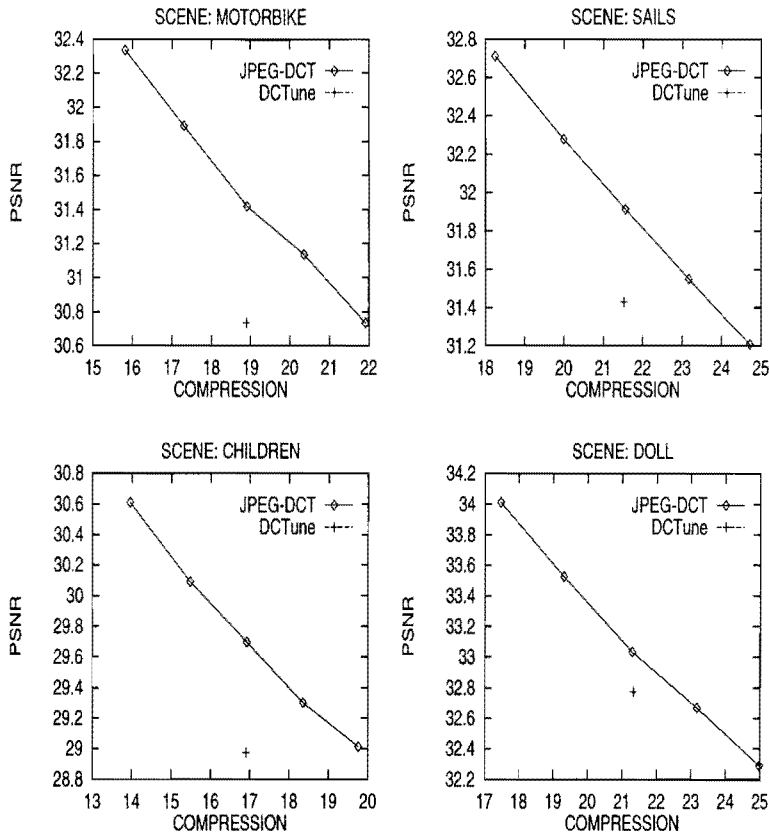
Figure 2.17: Compression ratio versus PSNR plots for all comparison images.

of 16 randomly selected stimuli containing all test scenes was presented to the subjects before the start of a session.

2.7.2 Results

For each of the test scenes, quality scores for the 6 comparison images were obtained by taking the mean of the 6 scores indicating the distance between that particular image and the corresponding "constant" DCTune-coded image. Before being averaged over all subjects, these scores were z-score transformed to reduce the variation between the subjects. In Fig. 2.18 the averaged z-scores of all comparison images have been plotted. In each plot, the dotted
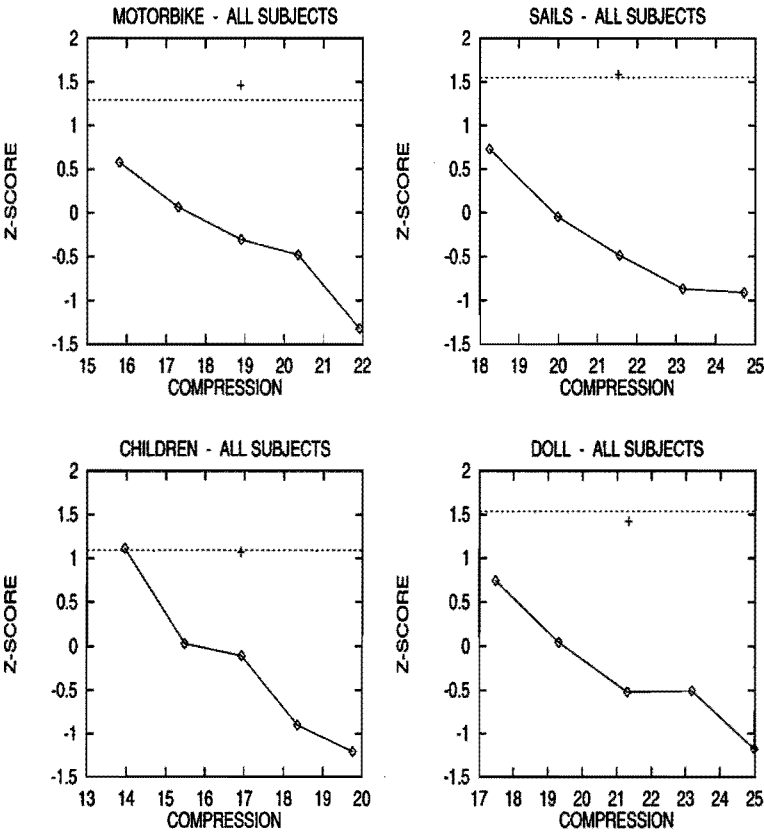
Figure 2.18:   Averaged subjective quality ratings for all comparison images expressed in z-scores. Each dotted line indicates the quality of the corresponding DCTune-coded scene with which all six images were compared.

horizontal line indicates the quality of the corresponding DCTune-coded image that was used as the "constant" image. The qualities rated for the DCTune-coded comparison images (almost) coincide with these lines, because these quality scores were obtained by comparing pairs of identical DCTune-coded images.

The z-score plots clearly indicate that the rate/distortion performance of a DCT coder can indeed be improved by designing proper quantization matrices. In the case of the "children" test scene, the perceived quality of the DCTune-coded image exactly matches the quality of the best comparison image. For the other test scenes, the perceived image quality of the DCTune-coded image even exceeds the quality of the best comparison image. This means that it would have been better to select comparison images in an even higher quality range. From these plots one may additionally conclude that subjects are very well capable of rating images in a small quality range.

A similar comparison experiment was carried out using the DCTune matrices calculated for the high quality range. Figure 2.19 shows both the PSNR's (left) and the subjective quality ratings (right) obtained for the "doll" test scene. The perceived image quality - expressed in z-scores - was averaged over 3 subjects. It can be observed that for these subjects, the effect of using a scene-optimized quantization matrix was found to be very small. One of the subjects clearly preferred the DCTune-coded image. Similar results were obtained for the other three test scenes.
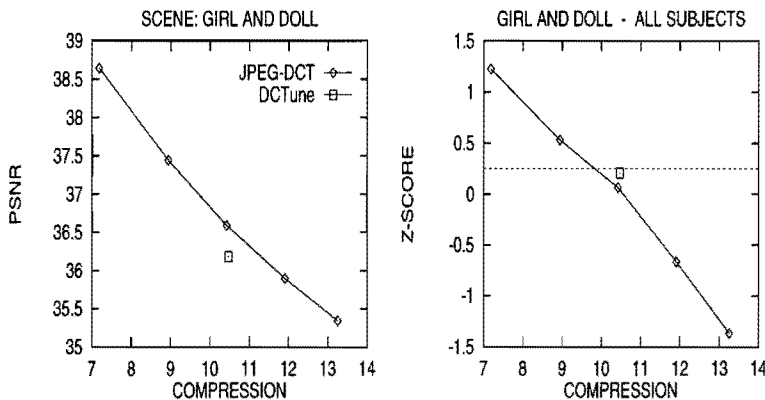
Figure 2.19: Left: PSNR plot of the comparison images. Right: averaged subjective quality ratings of the comparison images expressed in z-scores. The dotted line indicates the quality of the "constant" DCTune-coded image. Results are shown for the "doll" test scene.

2.8   Conclusions

In this chapter, the set-up has been described of a number of experiments in which subjects had to judge image quality using numerical category scaling. Numerical scaling in itself is a simple technique, although it can be applied in many different ways. It can be used in both direct rating experiments, and in experiments in which subjects are asked to rate quality differences between two stimuli rather than scaling quality directly. In order to give the reader some feeling for numerical scaling, a number of scaling experiments have been discussed in which various techniques were applied in varying situations. From these experiments it has become clear that numerical scaling techniques can and should be adapted to the typical characteristics of the stimulus set. The necessity of carrying out subjective experiments should also be stressed. Several experiments described in this chapter demonstrate that objective measures such as the peak-signal-to-noise ratio (PSNR) do not give any information about the annoyance of coding artifacts. It was found that subjective image quality scores may differ substantially from PSNR values.

The choice for an appropriate scaling technique is not always trivial. The single stimulus method used in this chapter, for example, is one of the evaluation methods recommended by the International Telecommunication Union (ITU). However, one should be very careful applying the method because in some cases it will bias the final outcome of an experiment. This was demonstrated by evaluating the performance of an image coder using both a (single stimulus) direct category scaling method and a scaling method based on Anderson's functional measurement theory (FMT). It appeared that results obtained in the direct category scaling experiments were biased because the image coders that were evaluated caused reconstruction artifacts of a very different nature. Only by using difference-in-quality ratings, valid quality ratings could be produced. Other research [117] has indicated that the single stimulus method is also susceptible to contextual effects due to stimulus spacing or frequency of occurrence of stimuli. It was demonstrated that these contextual effects even occur in the double stimulus method using separate scaling rates for both images. Also in this case it was shown that by using difference ratings, the bias could be reduced significantly.

It was shown that, in general, results from scaling experiments can be obtained by normalizing individual data with a z-score transform and averaging over different subjects, although more sophisticated methods can be applied. By assuming Thurstone's model for category scaling, for example, category scaling data can be transformed into an interval scale on a psychologically linear continuum. Thurston's model, however, requires a significant amount of stimulus observations in order to generate reliable results. Further, metric multi-dimensional scaling (MDS) was introduced as a valuable tool in analyzing and

interpreting experimental difference-in-quality data, and it was demonstrated for one set of experimental difference-in-quality data that both Anderson's FMT method and MDS resulted in the same quality scores. The suitability of multi-dimensional scaling techniques for the quality assessment of coded images was also demonstrated by evaluating the performance of a 2D Hermite coder. Disadvantages of both FMT and MDS are, of course, that multiple images have to be displayed at the same time and more observations are needed.

Finally, by assessing the effect of using scene-optimized quantization matrices on the quality of JPEG-coded images, it was demonstrated that category scaling can also be used if differences in image quality are fairly small. This experiment is also an example of how to adapt the range of the rating scale to the image quality range of a stimulus set. In general, when differences in the stimulus set are very small, a narrow rating scale (e.g. -5 to +5) should be preferred over a broad rating scale (e.g. -10 to +10), in order to enable subjects to feel comfortable when performing their task.

Acknowledgement

Appendix 2.A
Z-score Transform

Each score obtained in a category scaling experiment can be converted into a standardized score, or z-score, expressing the deviation from the mean score in standard deviation units [55]. The z-score z_j for a score x_j is given by

$$z_j = \frac{x_j - \bar{x}}{s}.$$

The z-score thus tells how many standard deviations x_j is removed from the mean. The mean and standard deviation of the set of observations are given by [22] :

$$\bar{x} = \frac{1}{N} \sum_{j=1}^{N} x_j \qquad \text{and} \qquad s^2 = \frac{1}{N-1} \sum_{j=1}^{N} (x_j - \bar{x})^2$$

respectively. The z-scores defined here have no relation with the well-known z-scores used in statistics. Therefore, the name z-score is somewhat confusing. Note that z-scores are normalized according to the mean and standard deviation of the observations of one particular subject.

Chapter 3

Image Representation Using Steered Hermite Transforms

Abstract

This chapter is concerned with the possibility of whether the Hermite transform can bridge the gap between conventional transform coding techniques and second-generation coding techniques that make use of explicit descriptions of perceptually important image structures (e.g., edge contours and lines). For image analysis the Hermite transform is selected, because this transform can be used efficiently for extracting (local) image primitives such as pattern orientation and position. In order to demonstrate the efficiency of the Hermite transform in image compression, an image compression scheme is presented on the basis of a new orientation-adaptive form of the Hermite transform (a steered Hermite transform). Comparisons with other compression techniques such as JPEG show that the proposed scheme performs very well at high compression ratios, not only in terms of peak-signal-to-noise ratio but also in terms of perceptual image quality. In Chapter 4, the steered Hermite transform will be used to construct contour-based image representations.

3.1 Introduction

Image compression is usually achieved by applying some fixed information-preserving transformation. This transformation must be followed by quantization and entropy coding of the transform coefficients in order to achieve actual

---

[1]The research described in this chapter has been published as a paper in Signal Processing [35]. Some results were also presented at the International Conference for Image Processing ICIP-96, Lausanne [33]

compression. The advantages of applying a transformation are twofold. First, it removes statistical redundancies by decorrelating the image data. Second, by making certain aspects of the information content explicit (e.g., frequency information), it becomes easier to take advantage of subjective irrelevance of information. For example, human visual characteristics such as spatial frequency selectivity and spatial masking properties can be incorporated into compression schemes that decompose signals into different frequency components [124, 2]. In the case of still images, however, one believes [157] that (1) no substantial further improvements in reducing the amount of irrelevant information can be expected, and (2) most conventional image decomposition techniques are limited in their redundancy reduction properties due to their inability to detect essential structural aspects in images. Future image compression techniques thus require the development of new image coding philosophies.

In the eighties, Kunt et al. [74] introduced the concept of *second-generation coding techniques.* Although this concept has led to a plethora of methods, the basic principles have remained the same: human visual perception should be incorporated not only in the process of determining irrelevant information, but also in the stages of information extraction and representation. Image analysis should thus be related more closely to the way in which the visual system analyzes visual information, and image representations should be tuned to the explicit description of perceptually important image structures (e.g., edge contours and textures). Second-generation coding techniques thus require image analysis tools that are capable of detecting and extracting visually important image structures. The purpose of this thesis is to demonstrate that the Hermite transform can bridge the gap between conventional transform coding techniques and second-generation coding methods which represent visually important image structures more explicitly.

The Hermite transform was introduced in [90]. It is a local image decomposition technique that uses overlapping Gaussian windows and projects images locally onto a basis of orthogonal polynomials. Since the analysis filters needed for the Hermite transform are derivatives of Gaussians (see Section 1.5), the transform matches– according to the Gaussian derivative theory [154] –the analysis carried out by the cortical receptive fields. Section 3.3 shows that the Hermite filters also provide the information necessary to calculate Hermite filter responses in arbitrary orientations. Orientational selectivity of receptive fields can thus be imitated, while different receptive field sizes can be modeled by varying the standard deviation of the Gaussian envelope.

Up to now, Hermite transforms have been used in applications such as image compression [89, 91], image deblurring [92], noise reduction [41], and estimation of perceived noise and blur [64]. In particular, this work has shown that the Hermite transform is very efficient in image analysis (detection and ex-

traction of image features). In this chapter the transform is used in an image coding algorithm that adapts to local orientations in an image. In this way, an image representation which efficiently represents one-dimensional structures is obtained. Chapter 4 will show that it provides a basis for image representations in which global image structures such as edges and lines can be described efficiently and explicitly. This chapter, however, is limited to the representation of local image features.

The organization of this chapter is as follows. Section 3.2 discusses the basic principles of the Hermite transform, while Section 3.3 focuses on its steering property. Section 3.4 demonstrates the energy compaction properties of a steered Hermite transform. Section 3.5 presents an image compression scheme based on a Hermite transform that adapts to local image orientation (an extended compression scheme using local edge parameter descriptions will be presented in Chapter 4). Results and comparisons with other coding schemes are dealt with in Section 3.6. Section 3.7 discusses the coding philosophy used and presents suggestions for future work. An extended compression scheme using local edge parameter descriptions will be presented in Chapter 4.

3.2   Hermite Transform

The Hermite transform was introduced in [90] as a signal expansion technique in which a signal is windowed at equidistant positions and locally described as a weighted sum of polynomials. This section briefly discusses the basic principles and main characteristics of the transform from an image coding point of view. Section 3.3 concentrates on the kind of information extracted by the transform and its steering property.

In the case of two-dimensional signals, the orthonormal Hermite basis functions used for signal decomposition are constructed by taking the Gaussian window function

$$v(x,y) = \frac{1}{\sigma\sqrt{\pi}}\exp\left[-\frac{x^2+y^2}{2\sigma^2}\right] \tag{3.1}$$

and the polynomials

$$g_{n-m,m}(x,y) = \frac{1}{\sqrt{2^n(n-m)!m!}}H_{n-m}(x/\sigma)H_m(y/\sigma), \tag{3.2}$$

which are orthonormal with respect to $v^2(x,y)$, i.e.

$$\int_{-\infty}^{\infty}dx\int_{-\infty}^{\infty}dy\,v^2(x,y)g_{n-m,m}(x,y)g_{l-k,k}(x,y) = \delta_{mk}\delta_{nl} \tag{3.3}$$

for $n, l = 0, ..., \infty$; $m = 0, ..., n$; $k = 0, ..., l$. In this notation, H_n is the standard Hermite polynomial of order n $(n=0,...,\infty)$, defined by [1]

$$H_n(x) = (-1)^n \exp(x^2) \frac{d^n}{dx^n} \left[\exp(-x^2)\right].$$ (3.4)

The orthonormal Hermite basis functions of order n expressed in cartesian coordinates are thus given by

$$h_{n-m,m}(x, y) = v(x, y) g_{n-m,m}(x, y),$$ (3.5)

where $n - m$ and m are the orders with respect to x and y, respectively. The Hermite transform uses this basis for representing a signal $f(x, y)$ localized by the Gaussian window function $v(x, y)$. Decomposition of such a localized signal thus results in a signal expansion

$$v(x, y) \cdot f(x, y) = \sum_{n=0}^{\infty} \sum_{m=0}^{n} f_{n-m,m} \cdot h_{n-m,m}(x, y)$$ (3.6)

with $f_{n-m,m}$ the cartesian Hermite coefficients. Since the Hermite transform decomposes images within Gaussian windows located on a square sampling grid $\mathcal{P}$, the coefficients $f_{n-m,m}$ are obtained by filtering the image using the (separable) Hermite filters with impulse responses

$$d_{n-m,m}(x, y) = v^2(-x, -y) g_{n-m,m}(-x, -y),$$ (3.7)

followed by sampling of the filtered output at the selected signal expansion positions. Plots of the filter functions $d_n(x)$ for a one-dimensional Hermite transform can be found in Chapter 1, Fig. 1.3. Because the Hermite transform is a continuous signal expansion, each local signal is represented by an infinite number of coefficients. For practical applications, however, the transform coefficients can only be calculated up to a finite order N, which implies that images cannot be reconstructed perfectly. Further, in this thesis, the lengths L of the Hermite filters are limited to $L = 10$ or $L = 16$. Filter taps of typical Hermite filters used are given in Appendix 3.A. An example Hermite decomposition is shown in Fig. 3.1.

Since the Hermite transform results in an overcomplete signal description, image reconstruction from the transform coefficients is not straightforward. The inverse Hermite transform does not require all positions of a local signal to contribute equally in the reconstruction process. By applying a synthesis window $s(x, y)$, the contribution of all positions within a local signal description can be set. In this thesis, a synthesis window $s(x, y) = v(x, y)$ is used. For reconstruction, we thus have local signals of the form $v_p^2(x, y) f(x, y)$ for all positions $p \in \mathcal{P}$, were $v_p(x, y)$ denotes $v(x - p_x, y - p_y)$ for p with coordinates (p_x, p_y),

Figure 3.1:   Cartesian Hermite coefficients up to order 2 for a Gaussian window with spread $\sigma = 1.3$ and a window spacing $d = 4$. For display purposes, the $n + 1$ coefficients of order n, located in the nth diagonal, are shown at the same, maximum contrast.

and $\mathcal{P}$ represents the set of all window positions. Synthesis is now carried out by adding all local signals and correcting for the modulation introduced by the window $v^2(x, y)$. For the reconstructed signal $\hat{f}(x, y)$, we thus have

$$\hat{f}(x, y) = \frac{\sum_p v_p^2(x, y) f(x, y)}{\sum_p v_p^2(x, y)} \tag{3.8}$$

which equals interpolation of the Hermite coefficients with the (separable) synthesis functions

$$s_{n-m,m}(x, y) = \frac{g_{n-m,m}(x, y) \cdot v^2(x, y)}{\sum_{p \in \mathcal{P}} v_p^2(x, y)}. \tag{3.9}$$

Note that synthesis requires complete coverage of the original signal, that is,

$$\sum_{p \in \mathcal{P}} v_p^2(x, y) \geq 0. \tag{3.10}$$

The free parameters in a Hermite transform decomposition are the Gaussian window spread σ, which defines the size of the Hermite basis functions, and the distance d between adjacent window positions. Given a fixed sampling distance d, one has to choose the appropriate value for σ. In Fig. 3.2, the effect of σ on the peak-signal-to-noise ratio (PSNR) for sampling distances 4 and 8 is shown for the image "Lena" (the original image is given in Fig 4.2). The different curves in each plot correspond to different values of the order N of the Hermite transform that was used in reconstructing the image. It can be observed that the PSNR, which is related to the root-mean-squared error (defined in Section 2.1) by

$$\text{PSNR} = 20 \cdot \log_{10} \left(\frac{255}{\text{RMSE}} \right), \tag{3.11}$$

varies considerably as a function of the window spread. Similar results were obtained for other test images. The value of σ is thus essential when applying a Hermite transform. Of course, the higher the order of the transform, the higher the PSNR of the reconstructed image. Note that in Eq. (3.11) the value 255 equals the peak-luminance of an 8-bit grayscale image.

In terms of coding artifacts, different values of σ result in different types of artifacts in the reconstructed image. A small σ causes "speckles", while a high σ results in Gibbs-phenomenon-like artifacts such as ringing and blur. These coding artifacts are demonstrated in Fig. 3.3. The "speckles" are visible mainly along high contrast edges. In coding applications, appropriate values for σ were found to be approximately 1.3 and 2.5 for sampling distances of 4 and 8 pixels, respectively. These values do not exactly correspond to the optimum PSNR (see Fig. 3.2) due to the effects of quantization of coefficients in coding applications. A rule of thumb for the selection of d and σ is given by the ratio $\tau = d/\sigma$ (τ should have an approximate value of 3). Further, a value of $d = 4$ is typically used for high quality coding at moderate bit rates, while a window spacing of $d = 8$ should be used for low-bit-rate coding (see also Section 3.6).

Because the cartesian Hermite transform is (locally) a unitary transform, the signal energy of each local signal can be obtained by calculating

$$E_\infty = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy [v(x, y) \cdot f(x, y)]^2 \tag{3.12}$$
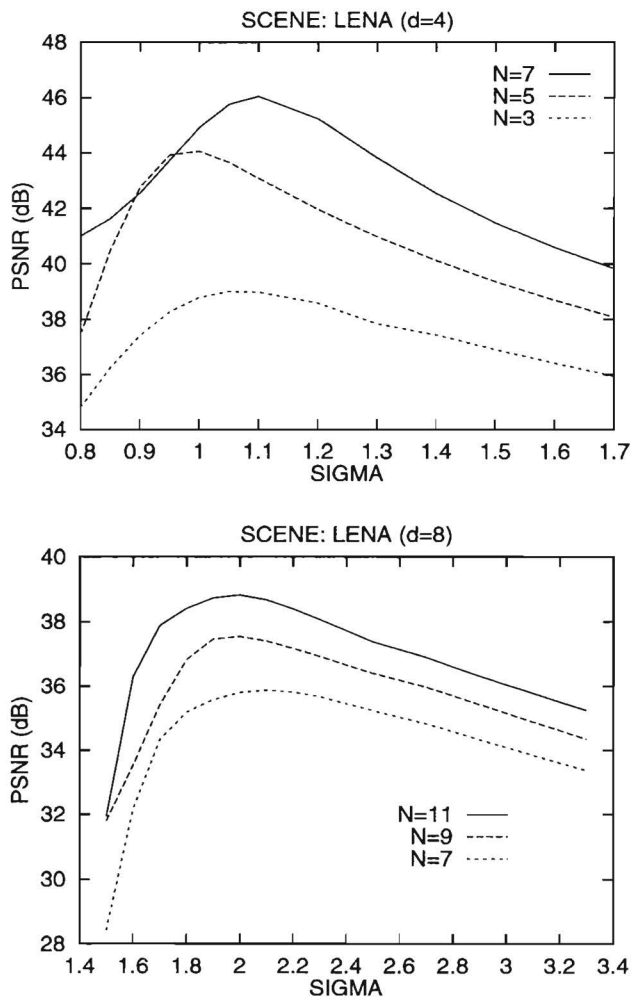
Figure 3.2: Effect of the Gaussian window spread σ on the PSNR of synthesized images. The figure at the top shows curves for different values of N and a window spacing of 4. The figure at the bottom shows PSNR curves obtained for a window spacing of 8. Results are given for the 512x512 8-bit grayscale image "Lena".
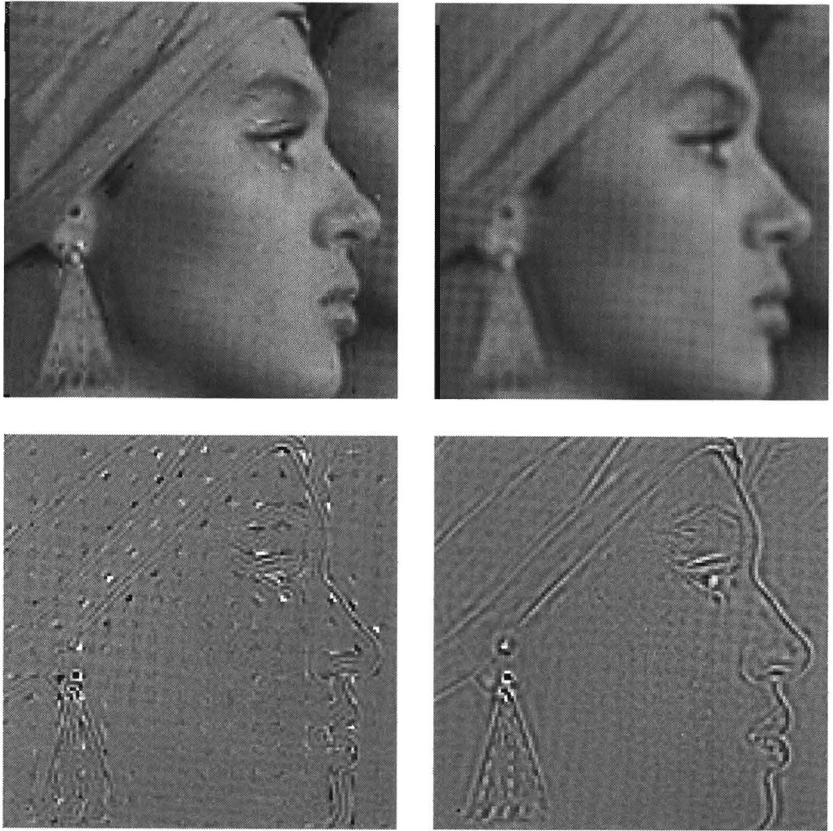
Figure 3.3:   Top: Details of an image synthesized from Hermite coefficients up to order $N = 7$. For both images a window spacing of 8 pixels was used. The spread σ of the Gaussian window was set to 1.75 pixels for the left image and 2.9 pixels for the right image, so that the PSNR was the same for both images. Bottom: corresponding error images.

or, in terms of its Hermite coefficients (Parseval's theorem)

$$E_\infty = \sum_{n=0}^{\infty} \sum_{m=0}^{n} [f_{n-m,m}]^2. \qquad (3.13)$$

Fig. 3.4 gives an impression of the cumulative signal energy contained in the coefficients of a Hermite decomposition of the image "Lena" as a function of the order N, expressed in a percentage of the total AC energy (equal to $E_\infty - f_{0,0}^2$). Curves are drawn for two different parameters setting of the Hermite transform. It can be seen that in both cases, at reasonable values of N, about 80 % of the total amount of AC energy is captured by the Hermite coefficients. In the lower part of Fig. 3.4, the effect of the order N on the PSNR of the reconstructed images is plotted. In this thesis, typical orders N used in coding are between 3 and 7 for a window spacing of 4, and between 7 and 11 for a window spacing of 8.

3.3   Steered Hermite Transform

In [46] Freeman and Adelson describe the problem of finding the response of a filter at arbitrary orientations. They use the term *steerable filter* to describe a class of filters in which a filter of arbitrary orientation is synthesized as a linear combination of a set of *basis filters*. Since all Hermite filters are polynomials times a radially symmetric window function, it is easy to prove that the $n + 1$ Hermite filters of order n form a steerable basis for every individual filter of order n. More specifically, rotated versions of a filter of order n can be constructed by taking linear combinations of the filters of order n. This property of the Hermite filters, which results in a new form of the Hermite transform, will be explained in the next paragraphs.

The Fourier transform of the 2D filter functions $d_{n-m,m}(x, y)$ is given by

$$\hat{d}_{n-m,m}(\omega_x, \omega_y) = \alpha_{n-m,m}(\theta)\hat{d}_n(\omega) \qquad (3.14)$$

where $\omega_x = \omega \cos(\theta)$ and $\omega_y = \omega \sin(\theta)$. The Fourier transform of each filter function can thus be written as a product that separately expresses orientation selectivity and radial frequency selectivity. The radial frequency of the 2D filter functions of order n equals the nth order derivative of a Gaussian and is given by

$$\hat{d}_n(\omega) = \frac{1}{\sqrt{2^n n!}} \cdot (-j\omega\sigma)^n \cdot \exp\left[-(\omega\sigma)^2/4\right], \qquad (3.15)$$
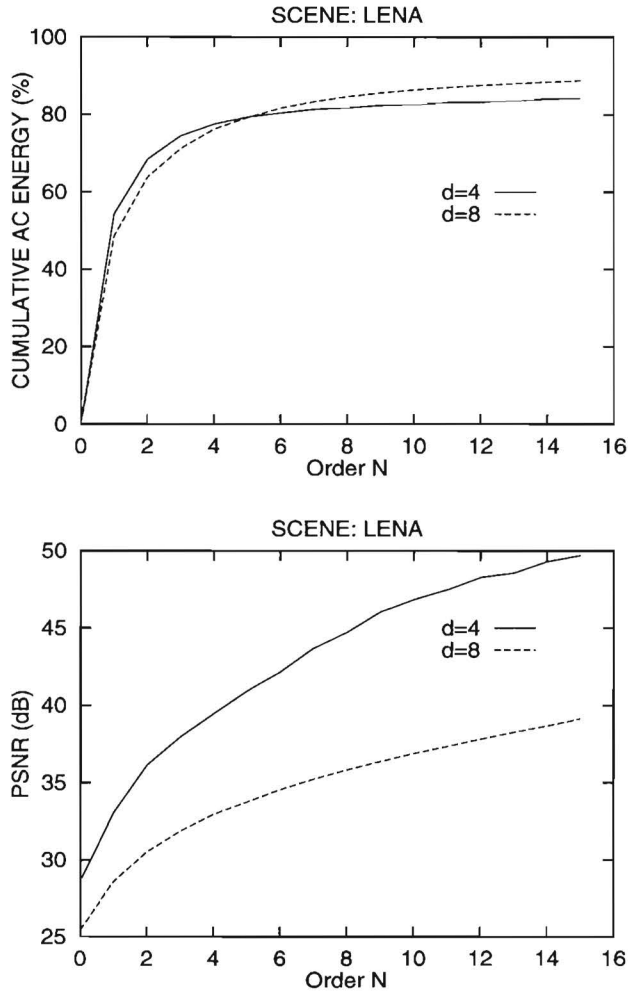
Figure 3.4:   Top: Cumulative contribution of AC signal energy contained in the coefficients of the Hermite transform as a function of order N. The curves were obtained for windows spacings of $d = 4$ and $d = 8$ and a Gaussian window spread σ of 1.3 and 2.5, respectively. The length of the Hermite filters was set to $L = 16$. Results are calculated for the 512x512 8-bit grayscale image "Lena". Bottom: Calculated PSNR values of the reconstructed image "Lena" as a function of order N.

while the orientation selectivity is expressed by

$$\alpha_{n-m,m}(\theta) = \sqrt{\binom{n}{m}} \cdot \cos^{n-m}\theta \cdot \sin^{m}\theta. \tag{3.16}$$

Note that the radial frequency selectivity $\hat{d}_n(\omega)$ is the same for all $n+1$ filters of order n and that these filters differ only in their orientation selectivity. In terms of orientation frequency functions, the steering property of the Hermite filters can be expressed by

$$\alpha_{n-m,m}(\theta - \theta_0) = \sum_{k=0}^{n} c_{m,k}^{(n)}(\theta_0)\alpha_{n-k,k}(\theta) \tag{3.17}$$

with $c_{m,k}^{(n)}(\theta_0)$ being the *steering coefficients*. By properly "steering" the Hermite filters at each position, image representations adapted to local orientation content can be obtained. Section 3.4 shows that such adaptivity results in significant energy compaction.

Fig. 3.5 shows the algorithmic structure for deriving a steered Hermite transform. First, a cartesian Hermite transform is applied to the original image. On the basis of the coefficients of this transform, an orientation image indicating the direction of maximum energy for each window position is determined. Next, this orientation image is used to "steer" the cartesian Hermite decomposition. In Fig. 3.5 the orientation image is displayed in the lower right corner of the steered Hermite decomposition.

Especially for patterns that are locally 1D, the Hermite transform provides a very efficient representation. This representation consists of a parameter θ, indicating the orientation of the pattern, and a small number of coefficients, indicating the profile of the pattern perpendicular to its orientation. For a 1D pattern with orientation θ, the following relation holds:

$$f_{n-m,m}^{\theta} = \begin{cases} \sum_{k=0}^{n} \alpha_{n-k,k}(\theta) \cdot f_{n-k,k}, & m = 0 \\ 0, & m > 0. \end{cases} \tag{3.18}$$

For such a pattern, steering over θ thus results in a compaction of energy into the coefficients $f_{n,0}^{\theta}$, while all other coefficients are set to zero. This means that a steered Hermite transform offers a way to describe 1D patterns explicitly on the basis of their orientation and profile. Fig. 3.6 shows a detail of an image that has been reconstructed by using local 1D approximations only (that is, only the "1D-coefficients" $f_{n,0}^{\theta}$ of the steered transform have been used in the reconstruction). As expected, artifacts are visible mainly in areas that are typically 2D (e.g., high curvature edges). In [91] a similar approach was applied
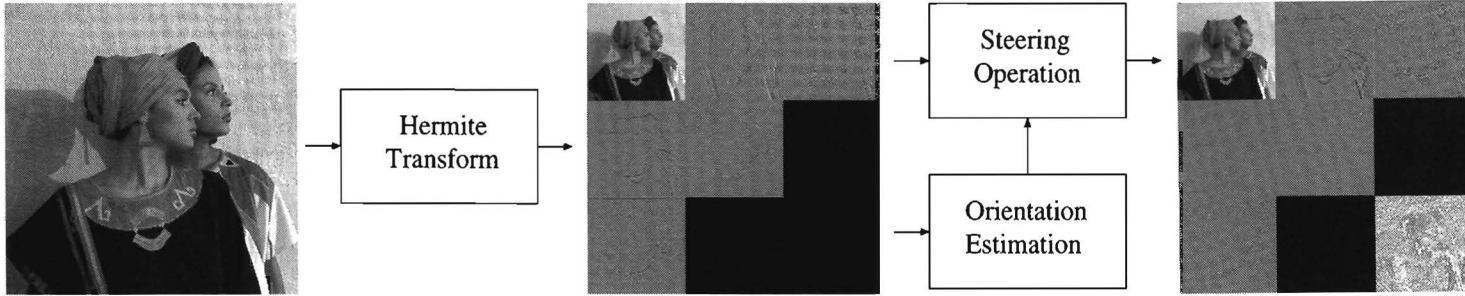
Figure 3.5: Algorithmic structure for deriving a steered image decomposition. First, the original image is transformed into a cartesian Hermite decomposition. This decomposition is then used to determine an orientation image to control the steering operation.

Figure 3.6: Details of the "doll" image (128x128 pixels). Left: original image. Right: full 1D reconstruction from an adaptive Hermite transform with parameter setting $d = 4$, $\sigma = 1.3$, and $N = 5$.

in approximating 1D local image structures using a discrete (binomial) Hermite transform. This transform, however, does not support smooth rotation of its basis functions and, also, no information other than 1D is retained.

3.4 Energy Compaction

For image compression purposes, it is important that signal energy is compacted into a small number of coefficients. The Hermite transform has the advantage that a high degree of energy compaction can be achieved by adaptively steering the transform. In Section 3.2 expressions were given to calculate the signal energy E_∞ of each local signal. In a similar way, energy up to order N, E_N can defined by

$$E_N \;=\; \sum_{n=0}^{N} \sum_{m=0}^{n} [f_{n-m,m}]^2, \tag{3.19}$$

which adds over all squared coefficients up to order N only. Due to the perfect energy compaction property of the (steered) cartesian Hermite transform for 1D patterns, we can distinguish 1D energy terms and 2D energy terms. That is, for each local signal we have

$$E_N^{1D}(\theta) = \sum_{n=1}^{N} [f_{n,0}^\theta]^2, \tag{3.20}$$

$$E_N^{2D}(\theta) = \sum_{n=1}^{N} \sum_{m=1}^{n} [f_{n-m,m}^{\theta}]^2, \tag{3.21}$$

and thus

$$E_N = E_0 + E_N^{1D}(\theta) + E_N^{2D}(\theta) \tag{3.22}$$

where $E_0 = f_{0,0}^2$ indicates DC energy.

Since most images contain a large number of structures which are (approximately) 1D, significant energy compaction can be obtained by properly steering the Hermite transform into the directions of maximum energy $E_N^{1D}(\theta)$. Table 3.I demonstrates the effect of steering on the energy contained in the higher-order (AC) Hermite coefficients. The upper part of Table 3.I shows the energy distribution for the cartesian Hermite transform, while the lower part shows the energy distribution after a steering operation. In the steered decomposition, more than 91 % of the AC energy is compacted into the 1D coefficients, while the 2D energy components contain less than 9 % of the energy. In this chapter, all local orientations were arrived at by maximizing $E_N^{1D}(\theta)$ in 16 discrete-valued orientations between 0 and 180 degrees. With much less complexity, results close to optimum can be obtained for most window positions by using $\tan(\theta) = f_{0,1}/f_{1,0}$ as an estimator for θ.

In the case of the Hermite transform, however, there is still a considerable amount of correlation between the transform coefficients. Therefore, by de-correlating the AC coefficients, the energy packing can be further optimized. Optimum decorrelation can be obtained by applying a decorrelating Karhunen-Loève transform (KLT) [23] to the AC coefficients. In the rest of this section we consider transform coefficients to be random variables (denoted in bold-face). The KLT of a vector $\mathbf{f}$ is obtained by calculating the covariance matrix $COV(\mathbf{f})$, of which the elements are given by

$$COV(\mathbf{f})(i,j) = \sigma_{ij}^2 = \mathcal{E}[(\mathbf{f}_i - \bar{\mathbf{f}}_i)(\mathbf{f}_j - \bar{\mathbf{f}}_j)^T]. \tag{3.23}$$

Now, by calculating the eigenvectors of this matrix, a transformation matrix $\mathbf{T}$ is obtained that completely decorrelates the elements of $\mathbf{f}$. In other words, we have

$$\mathbf{T} \cdot [COV(\mathbf{f})] \cdot \mathbf{T}^T = \mathbf{\Lambda} \tag{3.24}$$

where the rows of $\mathbf{T}$ are the eigenvectors and $\mathbf{\Lambda}$ contains the corresponding eigenvalues.

The gain of an adaptive cartesian Hermite transform over a common cartesian Hermite transform can be demonstrated by calculating the energy compaction

-	39.33	10.23	4.12	2.57	1.42	1.29	0.75
16.97	4.40	1.94	1.23	0.65	0.52	0.33	
3.67	1.57	0.98	0.55	0.44	0.27		
1.62	0.77	0.47	0.33	0.23			
0.77	0.42	0.28	0.20				
0.48	0.25	0.17					
0.40	0.16						
0.21							

—	54.69	16.32	7.75	5.16	3.08	2.53	1.64
1.61	0.84	0.47	0.30	0.23	0.20	0.16	
1.15	0.54	0.31	0.20	0.20	0.12		
0.48	0.20	0.12	0.09	0.07			
0.35	0.15	0.13	0.07				
0.21	0.12	0.07					
0.23	0.07						
0.12							

Table 3.I: Distribution of the AC energy up to order 7 over the individual coefficient bands of a Hermite transform on "Lena" with $(d, N, \sigma) = (8, 7, 2.7)$. The energy in each coefficient band is given as the percentage of the total amount of energy (total 100 percent). The energies in the coefficient bands of order n are displayed in the n-th diagonal. Top: no orientation adaptivity; bottom: orientation adaptive transform with 16 possible orientations, optimized on $E_7^{1D}(\theta)$

characteristics of both transforms [23]. Before calculating the transform efficiencies, however, we first decorrelate the transform coefficients by applying a Karhunen-Loève transform. Maximum energy compaction is thus guaranteed for both transforms.

Fig. 3.7 demonstrates the energy compaction effect resulting from a steering operation. For all $i\epsilon1...K$ (K being the number of transform coefficients given by $(N + 1)(N + 2)/2$), the sum of the estimated entropies of the first i KLT transform coefficients has been plotted against the sum of the variances of the omitted $K - i$ higher coefficients. The entropy of each individual transform coefficient expressed in units of bits per pixel of the original image is defined as

$$H = \frac{1}{d^2} \sum_s -P(s) \log_2 P(s) \qquad (3.25)$$

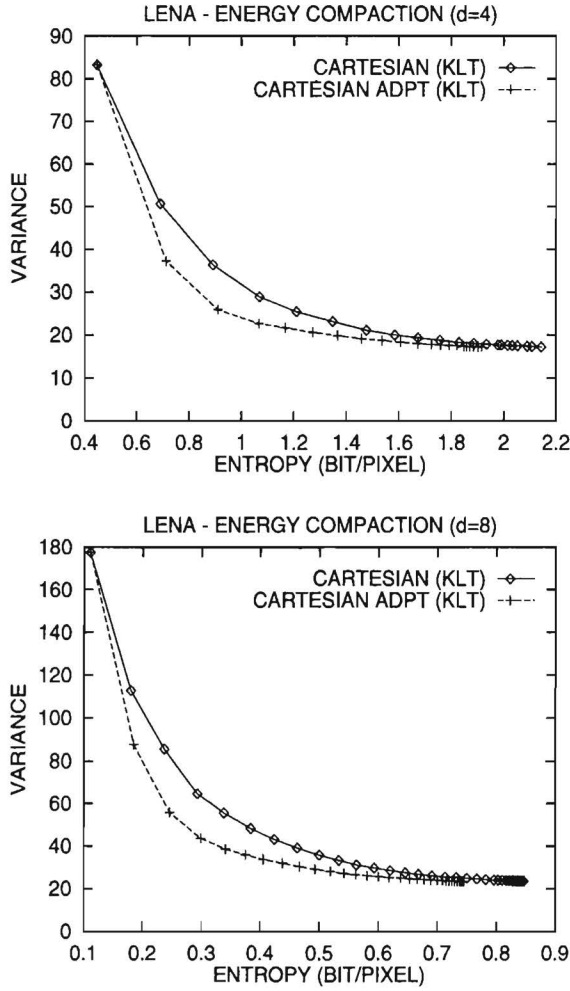where $P(s)$ is the frequency histogram of coefficient values s and d is the window

Figure 3.7: Energy compaction due to orientation adaptivity for two different settings of the Hermite transform. The plots show the variance of omitted coefficients versus estimated cumulative entropy of the KLT transform coefficients. Top: $(d, N, \sigma) = (4, 5, 1.3)$; bottom: $(d, N, \sigma) = (8, 7, 2.7)$.

spacing in pixels. Before the entropy calculation, all coefficients were uniformly quantized using quantization steps of 1. The total entropy of the image in bits per pixel is given by the sum of the individual coefficient entropies. The variance of each KLT coefficient $\mathbf{f}_i^{klt}$ $i \geq 1$ is expressed by

$$\sigma_i^2 = \mathcal{E}[(f_i^{klt} - \bar{f}_i^{klt})^2] = \mathcal{E}[(f_i^{klt})^2] \tag{3.26}$$

since the higher-order Hermite coefficients can be considered to be a zero-mean process. Curves are shown for the image Lena for window spacings of 4 and 8 pixels, respectively. Steering of the Hermite transform thus results in a significant gain in coding efficiency. Although curves and data have been plotted for "Lena" only, the same results are found for other test images. Typically, for the AC coefficients, an entropy reduction of about 20 % can be achieved (depending on the number of coefficients used in the image reconstruction). As can be seen in these plots, a considerable amount of energy is still contained in the Hermite coefficients of orders higher than N. This variance equals $\mathcal{E}[E_\infty - E_N]$.

A Karhunen-Loève transform is image dependent. However, coder complexity can be reduced by applying a fixed transformation to the Hermite coefficients. All compression results given in sections 5 and 6 were obtained by applying a fixed KLT transformation matrix derived from eight 512x768 test images. This test set did not include any of the images used in the experiments. Simulations showed that the fixed matrix resulted in energy compaction very close to the optimum.

An impression of the computational complexity of a Hermite coder: for an 8x8 DCT coder, 16 1D DCTs of length 8 are needed, each containing 64 multiplications in a straightforward implementation, adding up to a total of 16 multiplications per pixel. For a Cartesian Hermite transform with window spacing d, order N and filter length L, on the average, without any computational optimizations, a total of $d(N + 1)$ filter operations of length L are needed per window position in the horizontal direction and $(N + 1)(N + 2)/2$ filter operations of length L in the vertical direction, adding up to a total of $(d + (N + 2)/2)(N + 1)L$ multiplications for every window. Per window position, the steering operation takes $\sum_{n=2}^{N+1} n^2$ multiplications, while the KLT with a fixed matrix takes $((N + 1)(N + 2)/2)^2$ multiplications. For a typical orientation adaptive Hermite transform with parameters $d = 8$, $N = 7$ and $L = 16$, followed by a KLT transform, the total computational complexity is about 3 times the complexity of a separable 8x8 DCT transform.

3.5   Image Compression Scheme

A compression schemes incorporating an orientation adaptive Hermite transform is presented. The general framework for a Hermite compression algorithm

includes three components: (1) image analysis and representation, (2) quantization of all coefficients, and (3) ordering of the quantized coefficients into a bytestream, followed by entropy encoding.

3.5.1   Image Analysis and Representation

Image analysis is done by performing a cartesian Hermite transform, followed by extraction of local orientations and adaptive steering of the transform coefficients (Fig. 3.5). Orientations are detected by maximizing the energy measure $E_N^{1D}(\theta)$ for each window position (typically, 8 or 16 possible orientations are distinguished). Good results, however, can already be obtained by maximizing $E_1^{1D}(\theta)$ (i.e. $\tan(\theta) = f_{0,1}/f_{1,0}$ is used as an estimator). After the steering operation, extra energy compaction is obtained by applying a Karhunen-Loève transformation to the AC coefficients. An KLT, however, is image-dependent and requires transmission of the KLT matrix. In our implementation, coding of this information is avoided by applying a fixed transformation derived from eight 512x768 grayscale test images, which resulted in almost optimum energy compaction.

In the case of restricting to 1D polynomial descriptions, only the 1D coefficients of the steered transform have to be calculated. For such a 1D Hermite coder, a fixed KLT is applied on the 1D Hermite coefficients only. The difference between a 1D coder and a 2D coder is that in the case of a 1D coder, only the 1D AC coefficients are retained. In the case of a 2D coder, all AC coefficients are retained.

3.5.2   Quantization

All coefficients are uniformly quantized and mapped onto a set of integer values. By scaling the quantization stepsizes, different compression ratios are obtained. The results in this chapter were obtained by taking all stepsizes equal to 1, although different stepsizes can be selected for each coefficient band. Since the orientation image is already discrete-valued, this image need only be mapped to integer values.

3.5.3   Bytestream Conversion & Entropy Coding

Before the actual entropy coding is performed, all quantized coefficients are rearranged into a bytestream to take advantage of the many zeros produced by the quantization (every coefficient is coded in one byte). The bytestream is built up as follows :
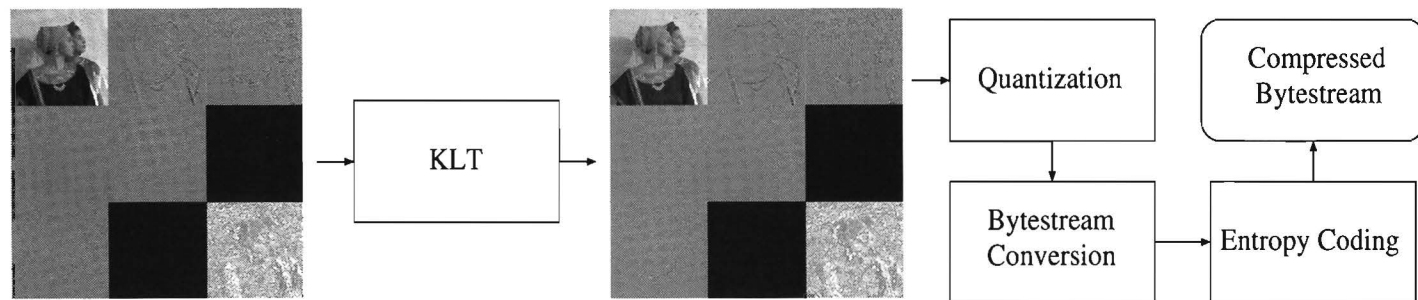
Figure 3.8: Block diagram of a 2D Hermite coder based on a steered Hermite transform. A KLT is applied to decorrelate the AC Hermite coefficients. After quantization, all coefficients are reordered into a bytestream and entropy coded.

- All DC coefficients. The DC values (except the first) are coded with a lossless differential pulse code modulation (DPCM) scheme using the previous DC coefficient as a predictor.

- Then, for each individual window position, the orientation is coded, followed by all KLT coefficients for that position. An end-of-block symbol is used to signal that all remaining KLT coefficients for that window position are zero. The orientation is only coded if one or more KLT coefficients are non-zero.

After rearrangement of the coefficients, an arithmetic coder is used to further compress the coefficients in the bytestream. The arithmetic coder used was obtained from a public domain entropy coding toolbox. A block diagram of a 2D Hermite coder is given in Fig. 3.8. Compression results obtained with 1D and 2D Hermite coders are given in Section 3.6.

3.6   Compression Results

In Fig. 3.9 PSNR performance curves obtained for the 512x512 8-bpp image "Lena" have been plotted for different types of Hermite coders. For comparison, performance curves are also shown for two other compression algorithms: JPEG using a Huffman-based entropy coding of the transform coefficients [108] and Shapiro's Zerotree coder [128], based on a discrete wavelet decomposition and the specific "Zerotree approach" in combination with arithmetic coding to code coefficients across different scales. The Zerotree algorithm was selected as a reference because it performs extremely well over a large range of compression ratios. It also demonstrates that the performance of JPEG, which is based on the discrete cosine transform (DCT), deteriorates rapidly at compression ratios above 30. In Shapiro's paper, all wavelet coefficients are ordered in a tree-like data structure. The name Zerotree refers to the zerotree symbol that is used to indicate that all coefficients in a subtree are insignificant.

The plots on the left give results for a 1D Hermite coder in which only the coefficients $f_{n,0}^{\theta}$ are used for reconstruction. The plots on the right give results for a 2D Hermite coder in which all Hermite coefficients are retained. All coded images were obtained by applying an adaptive cartesian Hermite transform with 16 possible orientations, followed by a KLT to decorrelate the AC coefficients. From these plots we can conclude the following. First, it can be seen that there is not much difference in performance between 1D and 2D coders, although 1D coders saturate at high bit rates. At low bit rates, performance is almost identical. Second, for low-bit-rate coding, a large window spacing (e.g., 8) is clearly preferred, while high-bit-rate coding requires a smaller window
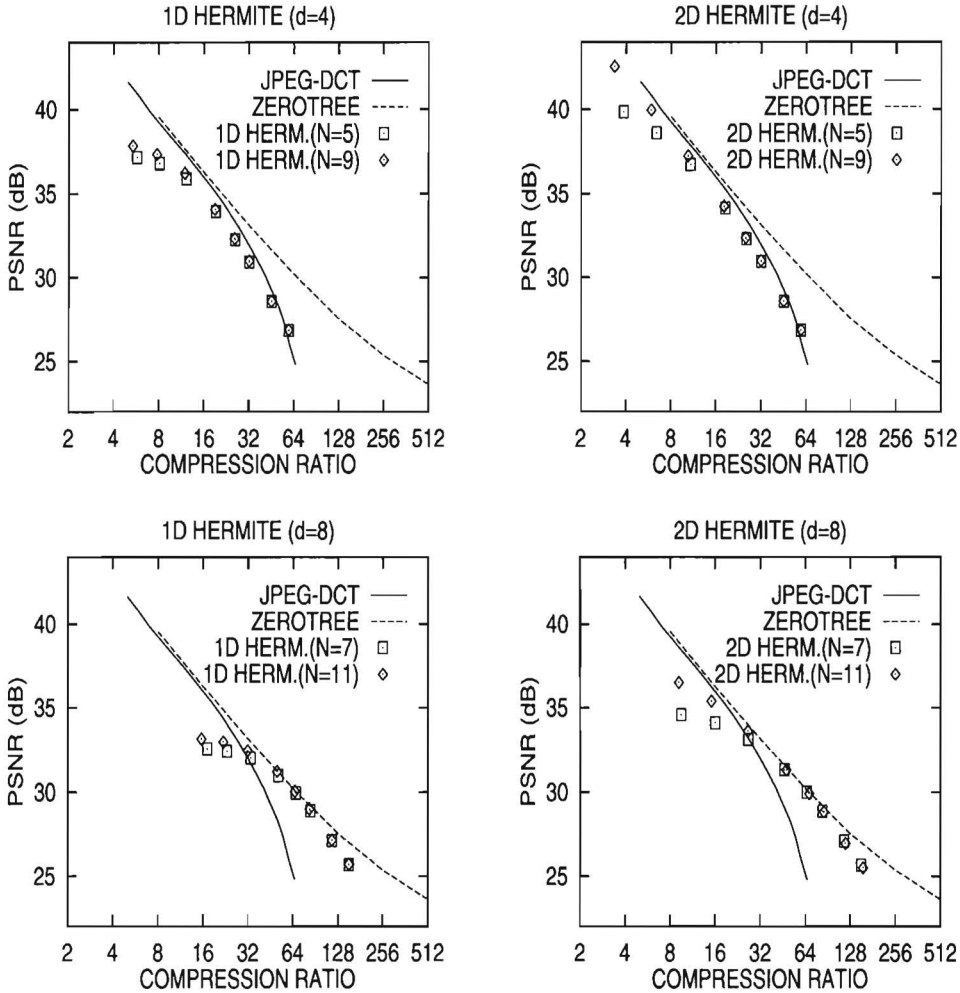
Figure 3.9: Compression results of 1D Hermite coders (left column) and 2D Hermite coders (right column). Parameter settings of the Hermite transform used: $(d, \sigma) = (4, 1.3)$ for the plots at the top and $(d, \sigma) = (8, 2.7)$ for the plots at the bottom. The curves were obtained by varying the quantization steps for the Hermite coefficients between 1 and 32. For all coders, 16 different orientations were used.

spacing. Furthermore, it is clear that Hermite coders perform quite well at compression ratios between 32 and 64 and approximate the performance of the Zerotree coder. Note that at these high ratios, taking a higher order N does not significantly affect coder performance.
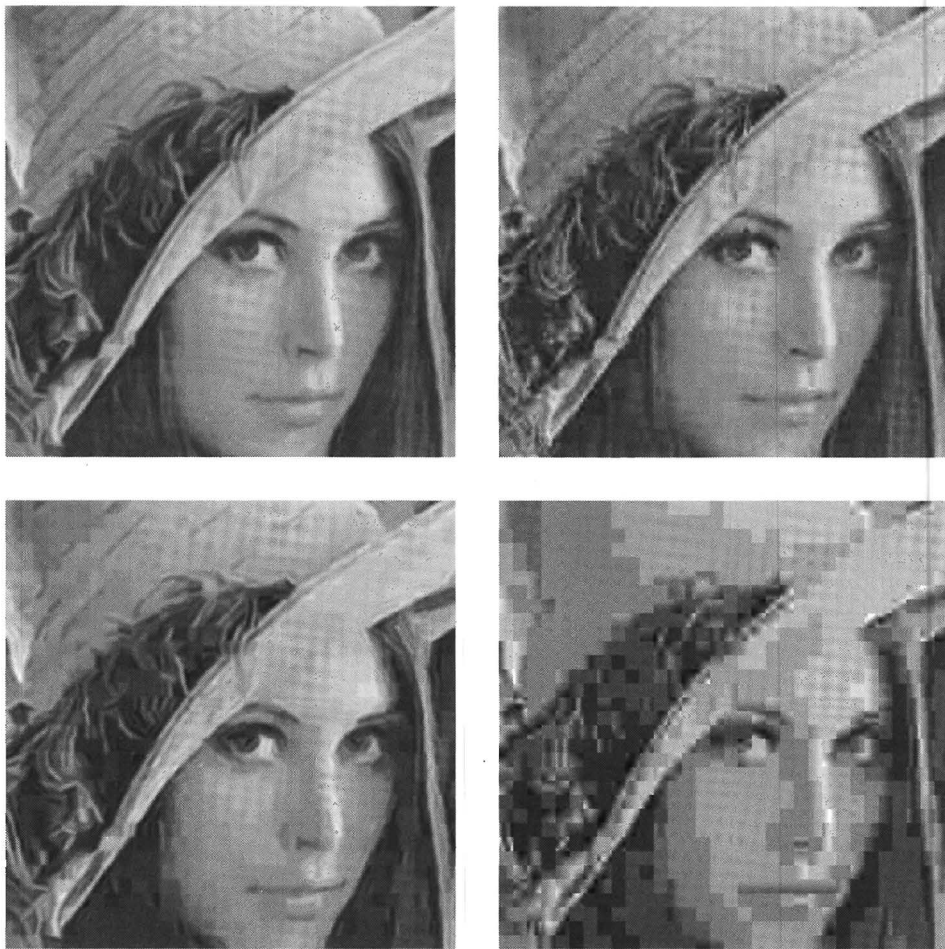
Figure 3.10:   Fragments of "Lena". Left: 1D Hermite transform with compression ratio C=31.86, PSNR=32.46 (top) and C=66.11, PSNR=30.05 (bottom). Right: JPEG-DCT C=32.67, PSNR=31.86 (top) and C=60.07, PSNR=26.02 (bottom).

Fig. 3.10 shows a number of reconstructions of "Lena". It can be seen that the steered Hermite transform reproduces edge-like structures very well. For example, at a compression ratio of 32 the edge-like structures in the 1D reconstructed image show hardly any distortion. At the same bit rate, the quality of DCT-coded edges is much less. This difference is even more apparent in the images compressed at a ratio of about 64. Since the Hermite transform is an (overlapped) block transform, blocking artifacts become visible only at high compression ratios. At equal bit rates, however, blocking artifacts are much more prominent in JPEG-DCT-coded images than in Hermite-coded images.

We found that on the basis of a PSNR criterion, for most test images, JPEG should be preferred over a Hermite coder at compression ratios lower than 25. However, results of quality assessment experiments using subjects indicate that the perceptual quality of Hermite-coded images is much better than would be expected on the basis of the PSNR. Figure 3.11 shows the results of the subjective quality evaluation of a 2D Hermite coder for the test scenes "doll" and "girls". PSNR curves and subjective quality ratings have been plotted for a 2D Hermite coder and a standard JPEG baseline coder. For the Hermite coder used in the experiment, a parameter setting $(d, N, \sigma) = (8, 11, 2.5)$ was applied. Perceived quality was measured via metric multi-dimensional scaling, a method in which subjects are asked to rate quality differences between pairs of coded images (details can be found in Section 2.6). It can be observed that the subjective quality ratings differ significantly from the PSNR curves. The difference is due mainly to less severe blocking artifacts in the Hermite-coded images.

The typical artifacts caused by Shapiro's Zerotree coder differ from the artifacts introduced by a Hermite coder. Therefore, it would also be interesting to evaluate perceived image quality of this Zerotree coder. Shapiro's paper [128] shows that, at high compression ratios, the structures in the Zerotree-reconstructed images suffer from blur and ringing artifacts due to the wavelet scheme used. At extremely low bit rates ($C > 100$), the images fade instead of showing blocking and segmentation artifacts, as is the case with a Hermite coder. In this thesis, no subjective evaluation of the Zerotree coder is presented.

3.7 Discussion and Conclusions

In this chapter, a new form of the Hermite transform is introduced that enables (local) rotation of the Hermite frequency patterns. From an image compression point of view, it is important that the orientation adaptivity of the steered Hermite transform results in a significant energy compaction. Especially for patterns that are locally one-dimensional, this energy compaction provides an efficient representation, consisting of the orientation of the pattern and
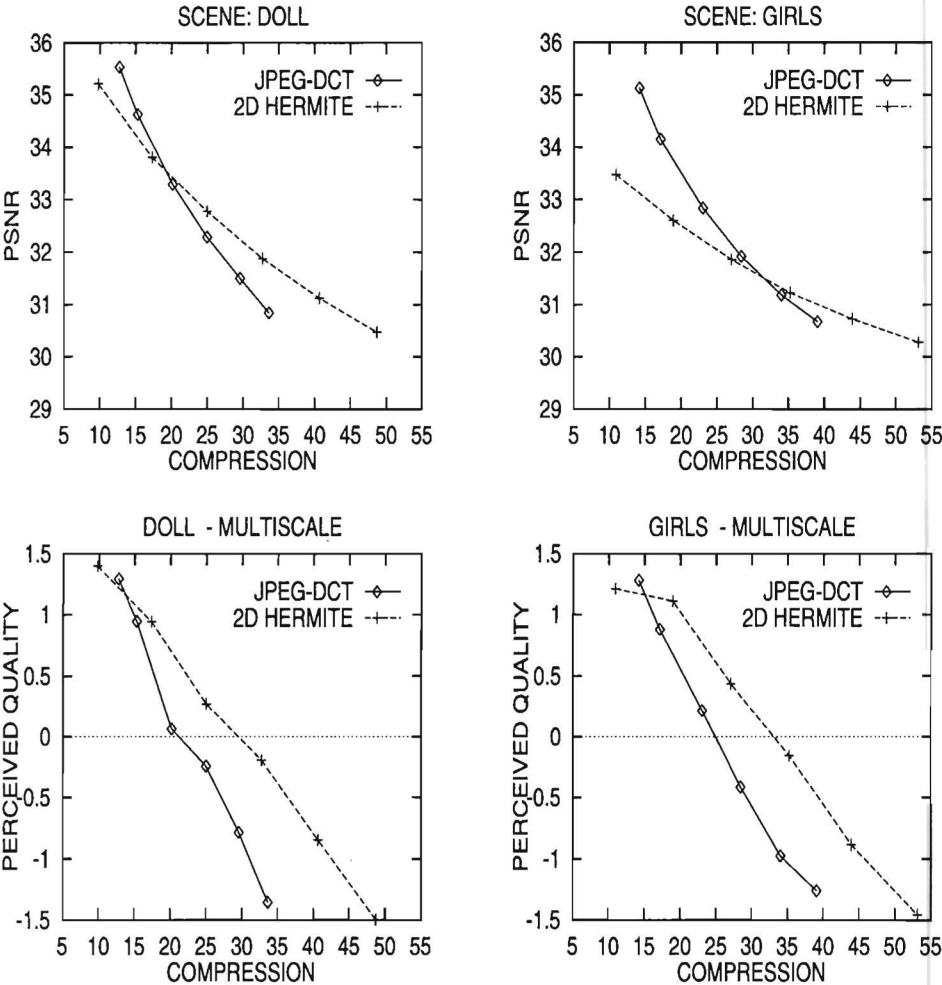
Figure 3.11: Subjective quality ratings. Top: PSNR curves for a 2D Hermite coder and JPEG; bottom: corresponding subjective quality ratings. Results are shown for the test scenes "doll" and "girls".

a number of coefficients describing the profile of the pattern. One-dimensional approximations of two-dimensional elements are obtained by discarding the residue energy in the 2D coefficients of the steered Hermite transform. An extended mathematical framework on the theory of the steered Hermite transform as described in this chapter, will be published in the near future [95].

Using the steered Hermite transform in an image compression scheme, shows that an image coder can be constructed that performs well compared to other compression algorithms. In order to benchmark the performance of this image coder, both 1D and 2D Hermite coders are compared to a standard JPEG DCT-based reference coder. For high compression ratios, Shapiro's Zerotree wavelet coder is used as a second reference. At low bit rates, the Hermite transform should be preferred over JPEG, and it may even compete with the Zerotree coder. Since the objective peak-signal-to-noise ration (PSNR), however, does not correlate well with perceived image quality more extensive evaluations of coder performances are needed. In particular, it would be nice to compare a Hermite coder with a Zerotree wavelet coder. Experiments with subjects do indicate that for equal PSNR's, the perceptual quality of Hermite-coded images is better than that of JPEG-coded images.

With the Hermite transform, perfect image reconstruction is not possible due to the continuous character of the signal expansion. Also, the computational complexity of the transform increases when calculating the transform coefficients up to a high order N. In the first years of the research project that resulted in the writing of this thesis -in order to avoid the disadvantages of the continuous Hermite transform-, a lot of effort was spent using a discrete equivalent of the Hermite transform [90] in an orientation adaptive image representation technique. Unfortunately, the discrete character of this (binomial) Hermite transform caused too many distortions when approximating filter rotations. The performance of a non-adaptive discrete Hermite transform in a compression scheme was also found to be insufficient.

Although the image compression scheme presented in this chapter covers only local image descriptions, the Hermite transform provides an excellent basis for image representations containing explicit descriptions of global image structures such as edge contours, lines and surfaces. In Chapter 4, explicit coding of other local image features and representation of global image structures will be addressed. Furthermore, the perceptual quality of an image is significantly affected by artificial, unnatural structures. Therefore, it is worthwhile to compress areas in which blocking artifacts are mostly visible (e.g., slope areas) less severely. At high compression ratios ($C > 32$), however, block coders such as JPEG and the Hermite transform spend a relatively large number of bytes on the representation of the DC coefficient. Since quantization of the DC coeffcients causes blocking and segmentation artifacts, performance of block coders

could be improved by developing alternative coding methods for the DC coeff-cient. In Chapter 5 of this thesis, an algorithm for reconstructing smooth areas in the image wil be discussed.

n=0	n=1	n=2	n=3	n=4	n=5	n=6	n=7
0.000007	0.000032	0.000097	0.000230	0.000442	0.000700	0.000914	0.000951
0.000555	0.001945	0.004448	0.007496	0.009502	0.008590	0.004174	-0.001854
0.014381	0.035699	0.053198	0.049208	0.018545	-0.019730	-0.035065	-0.015654
0.123841	0.183165	0.111492	-0.040862	-0.117121	-0.043267	0.070682	0.074481
0.361215	0.177346	-0.169235	-0.178019	0.088629	0.161924	-0.040723	-0.141137
0.361215	-0.177346	-0.169235	0.178019	0.088629	-0.161924	-0.040723	0.141137
0.123841	-0.183165	0.111492	0.040862	-0.117121	0.043267	0.070682	-0.074481
0.014381	-0.035699	0.053198	-0.049208	0.018545	0.019730	-0.035065	0.015654
0.000555	-0.001945	0.004448	-0.007496	0.009502	-0.008590	0.004174	0.001854
0.000007	-0.000032	0.000097	-0.000230	0.000442	-0.000700	0.000914	-0.000951

Table 3.II: Hermite analysis filters of length $L = 10$ up to order $N = 7$ for $\sigma = 1.3$.

n=0	n=1	n=2	n=3	n=4	n=5	n=6	n=7
0.000007	-0.000035	0.000118	-0.000304	0.000641	-0.001132	0.001677	-0.002055
0.000824	-0.003137	0.007863	-0.014725	0.021222	-0.022966	0.016325	-0.002232
0.085735	-0.233167	0.387775	-0.418499	0.233260	0.090612	-0.313542	0.238408
0.914265	-1.491884	1.074922	0.205422	-1.098512	0.617911	0.591164	-0.936679
0.999169	-0.543476	-0.497490	0.599977	0.267667	-0.601746	-0.110723	0.579871
0.999169	0.543476	-0.497490	-0.599977	0.267667	0.601746	-0.110723	-0.579871
0.914265	1.491884	1.074922	-0.205422	-1.098512	-0.617911	0.591164	0.936679
0.085735	0.233167	0.387775	0.418499	0.233260	-0.090612	-0.313542	-0.238408
0.000824	0.003137	0.007863	0.014725	0.021222	0.022966	0.016325	0.002232
0.000007	0.000035	0.000118	0.000304	0.000641	0.001132	0.001677	0.002055

Table 3.III: Hermite synthesis filters of length $L = 10$ up to order $N = 7$ for $\sigma = 1.3$.

n=0	n=1	n=2	n=3	n=4	n=5	n=6	n=7	n=8	n=9
0.000035	0.000144	0.000397	0.000832	0.001385	0.001834	0.001861	0.001250	0.000129	-0.000970
0.000309	0.001106	0.002588	0.004470	0.005819	0.005419	0.002742	-0.001178	-0.003987	-0.003670
0.002003	0.006069	0.011621	0.015494	0.013656	0.004994	-0.005966	-0.011321	-0.006675	0.003659
0.009512	0.023572	0.034747	0.030952	0.009034	-0.016940	-0.025139	-0.008249	0.015661	0.020472
0.033064	0.063706	0.064020	0.020564	-0.034147	-0.047283	-0.006821	0.037613	0.031775	-0.014118
0.084146	0.115778	0.054721	-0.048514	-0.079441	-0.006626	0.066799	0.040641	-0.041012	-0.056006
0.156831	0.129448	-0.032343	-0.118203	-0.021501	0.094871	0.051002	-0.069494	-0.066603	0.045388
0.214097	0.058900	-0.135795	-0.068326	0.104963	0.072314	-0.085033	-0.073897	0.070125	0.074108
0.214097	-0.058900	-0.135795	0.068326	0.104963	-0.072314	-0.085033	0.073897	0.070125	-0.074108
0.156831	-0.129448	-0.032343	0.118203	-0.021501	-0.094871	0.051002	0.069494	-0.066603	-0.045388
0.084146	-0.115778	0.054721	0.048514	-0.079441	0.006626	0.066799	-0.040641	-0.041012	0.056006
0.033064	-0.063706	0.064020	-0.020564	-0.034147	0.047283	-0.006821	-0.037613	0.031775	0.014118
0.009512	-0.023572	0.034747	-0.030952	0.009034	0.016940	-0.025139	0.008249	0.015661	-0.020472
0.002003	-0.006069	0.011621	-0.015494	0.013656	-0.004994	-0.005966	0.011321	-0.006675	-0.003659
0.000309	-0.001106	0.002588	-0.004470	0.005819	-0.005419	0.002742	0.001178	-0.003987	0.003670
0.000035	-0.000144	0.000397	-0.000832	0.001385	-0.001834	0.001861	-0.001250	0.000129	0.000970

Table 3.IV: Hermite analysis filters of length $L = 16$ up to order $N = 9$ for $\sigma = 2.5$.

n=0	n=1	n=2	n=3	n=4	n=5	n=6	n=7	n=8	n=9
0.000128	-0.000545	0.001544	-0.003337	0.005741	-0.007909	0.008457	-0.006240	0.001449	0.003834
0.001659	-0.006099	0.014685	-0.026195	0.035442	-0.034850	0.019960	0.004525	-0.024554	0.025828
0.021041	-0.065465	0.129145	-0.178530	0.165885	-0.071130	-0.061084	0.137686	-0.094315	-0.031998
0.217550	-0.553792	0.842995	-0.786775	0.271346	0.394807	-0.658000	0.267568	0.374692	-0.570202
0.782450	-1.549172	1.615565	-0.581851	-0.823117	1.249242	-0.258351	-0.963241	0.915934	0.303667
0.978959	-1.384457	0.692228	0.565202	-0.999146	0.126383	0.839124	-0.565538	-0.502159	0.769915
0.998341	-0.847121	-0.197661	0.788505	-0.163354	-0.643272	0.371957	0.476262	-0.490813	-0.310201
0.999872	-0.282806	-0.650455	0.337129	0.515633	-0.366761	-0.428357	0.385348	0.362156	-0.397454
0.999872	0.282806	-0.650455	-0.337129	0.515633	0.366761	-0.428357	-0.385348	0.362156	0.397454
0.998341	0.847121	-0.197661	-0.788505	-0.163354	0.643272	0.371957	-0.476262	-0.490813	0.310201
0.978959	1.384457	0.692228	-0.565202	-0.999146	-0.126383	0.839124	0.565538	-0.502159	-0.769915
0.782450	1.549172	1.615565	0.581851	-0.823117	-1.249242	-0.258351	0.963241	0.915934	-0.303667
0.217550	0.553792	0.842995	0.786775	0.271346	-0.394807	-0.658000	-0.267568	0.374692	0.570202
0.021041	0.065465	0.129145	0.178530	0.165885	0.071130	-0.061084	-0.137686	-0.094315	0.031998
0.001659	0.006099	0.014685	0.026195	0.035442	0.034850	0.019960	-0.004525	-0.024554	-0.025828
0.000128	0.000545	0.001544	0.003337	0.005741	0.007909	0.008457	0.006240	0.001449	-0.003834

Table 3.V: Hermite synthesis filters of length $L = 16$ up to order $N = 9$ for $\sigma = 2.5$.

Chapter 4

Contour-Based Image Representation and Compression

Abstract

The steered Hermite transform introduced in Chapter 3 can be used as a basis for deriving an edge-parameter-based image representation. This chapter describes how to calculate edge parameters such as position, height and blur from the frequency coefficients of the steered Hermite transform. In particular, the research in this chapter demonstrates that edge descriptions in terms of edge parameters may decrease the entropy of edge representations. By chaining local edge points into edge contours, images are reconstructed based on contour information supplemented with DC components. Also, an impression is given of the variation of edge parameters along image contours. Although the proposed edge parameter descriptions are integrated into an image compression scheme that exploits parameter redundancies along contours, the performance of this compression scheme will have to be improved in order to obtain optimal results.

4.1 Introduction

A significant part of the information content of an image is contained in edges. In this thesis, the assumption is made that the perceptual quality of an image is largely determined by the accuracy with which edge information can be reconstructed. Conventional transform coding techniques, however, have a weakness in coding edges. In a DCT (discrete cosine transform) representation, for example, edge segments oriented along a certain angle are expected to have energy distributions in many frequency components. The orientation adaptive Hermite transform presented in Chapter 3 significantly reduces this problem. However, from a perceptual point of view, it still seems unnatural to decompose

edge structures into polynomial or frequency components since edges are not perceived as a sum of frequency patterns. Another disadvantage of transform coding techniques, is that they do not efficiently exploit redundancies along edge-like structures.

The use of explicit contour coding represents a break with conventional waveform coding methods. The importance of edge coding, however, has been recognized for a long time. Already in 1967, Graham [51] constructed a compression algorithm that was based on separate coding of low frequency information and edge-based coding of high frequency information by describing contour direction, gradient direction and gradient magnitude. Carlsson [20] presented an image compression algorithm based on coding information about the geometric structure and contour grey level intensities. A similar approach is used by van Beek [14]. Image reconstruction from contour information only, however, usually results in images that look somewhat unnatural due to smooth interpolation between reconstructed contours. In [74, 75] Kunt describes coding methods based on approximation of contours obtained from region growing, and methods that use a directional decomposition of the image. Other research using edge-based coding has been performed by Mallat [80, 81] who describes a second-generation coding algorithm based on multiscale edge representation. In his approach images are reconstructed using wavelet coefficients at edge locations that are important for image visualization.

This chapter integrates explicit coding of edges in a (Hermite) transform coding scheme. The steered Hermite transform presented in Chapter 3 will be used as a basis for developing a contour-based image representation. Section 4.2 demonstrates that at edge locations the local Hermite frequency description can be transformed into a description of the edge in terms of location, height and blur. Section 4.3 concentrates on linking local edge descriptions along the contours in the image. In this way, redundancies along edge contours can be exploited since the edge primitives along an edge are expected to vary only slightly. The contour-based representation is integrated into an image compression scheme in Section 4.4. Results are shown in Section 4.5.

4.2   Edge Parameter Coding

The orientation adaptive Hermite transform provides a basis for describing local image structures more explicitly. This section will demonstrate that edge primitives, such as edge height and edge position, can be deduced from the coefficients of an adaptive Hermite transform. The underlying idea is that by distinguishing different types of local image structures and by representing these structures by their specific parameters, a more efficient and more flexible image representation can be constructed. Also, separate coding of classes of

image structures facilitates independent quality control of these structures.
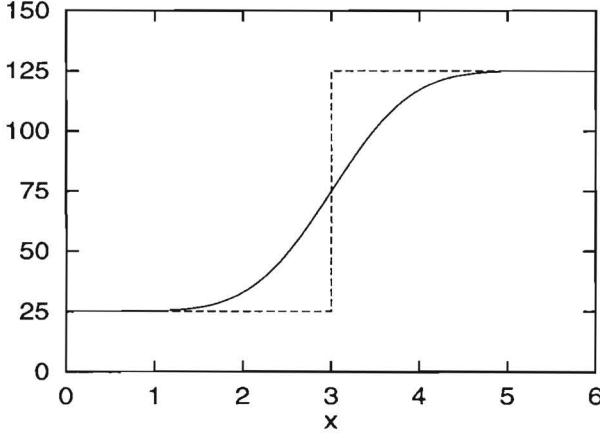
Figure 4.1:   The cross-section of a blurred edge along the direction of the edge (drawn curve), for an edge with height $\Delta V = 100$, mean value $V_e = 75$, located at a distance $d = 3$ from the origin and blurred with a Gaussian blurring kernel with $b = 1$.

We use the Gaussian edge model

$$\text{edge}(x; V_e, \Delta V, b, d) = V_e + \frac{\Delta V}{2} \text{erf}\left[\frac{x - d}{b}\right] \qquad (4.1)$$

to describe local edge profiles with mean signal value V_e, edge height ΔV, blur parameter b and distance to the origin d (see Fig. 4.1, obtained from [64]). At edge locations, the Hermite frequency coefficients can be expressed in terms of local edge parameters. More specifically, for a Gaussian window of spread σ, centered at the origin, it can be shown [64] that for a Gaussian edge the Hermite coefficients ($f_{n,0}^{\theta}$ is denoted by f_n) are given by

$$f_n = \left(\frac{\sigma}{\sqrt{\sigma^2 + b^2}}\right)^{n-1} \cdot \frac{\Delta V \cdot \sigma}{\sqrt{2n}} \cdot d_{n-1}(-d; \sqrt{\sigma^2 + b^2}) \qquad (4.2)$$

for $n = 1, ..., \infty$, where $d_n(x; \sigma)$ is the nth order derivative of a Gaussian [90] i.e.,

$$d_n(x; \sigma) = \frac{1}{\sqrt{2^n n!}} \cdot \frac{d^n}{d\left(\frac{x}{\sigma}\right)^n} \left[\frac{1}{\sigma\sqrt{\pi}} e^{-x^2/\sigma^2}\right]. \qquad (4.3)$$

Note that the coefficients f_n do not depend on the orientation θ of the edge, but only describe the profile of the edge. By solving Eq. (4.2) for f_n up to order three [64], an estimation for the edge blur

$$b = \sigma \cdot \left[\left(\frac{2f_2^2}{f_1^2} - \frac{\sqrt{6}f_3}{f_1} \right)^{-1} - 1 \right]^{1/2} \tag{4.4}$$

is obtained, while the edge distance and the edge height are given by

$$d = \sigma \cdot \frac{f_2}{f_1} \cdot \left(\frac{2f_2^2}{f_1^2} - \frac{\sqrt{6}f_3}{f_1} \right)^{-1} \tag{4.5}$$

and

$$\Delta V = f_1 \sqrt{2\pi} \left(\frac{2f_2^2}{f_1^2} - \frac{\sqrt{6}f_3}{f_1} \right)^{-(1/2)} \exp \left[\frac{f_2^2}{f_1^2} \left(\frac{2f_2^2}{f_1^2} - \frac{\sqrt{6}f_3}{f_1} \right)^{-1} \right] \tag{4.6}$$

respectively. The mean value V_e can be calculated by substituting the above parameter values in

$$f_0 = V_e - \frac{\Delta V}{2} \mathrm{erf} \left[\frac{d}{\sqrt{\sigma^2 + b^2}} \right]. \tag{4.7}$$

In this chapter, however, Eq. (4.7) is not solved for V_e, but the original DC value f_0 is used. By adopting an edge model, edges can thus be described explicitly in terms of local edge parameters (only edge parameters have to be retained). The Hermite coefficients can directly be reconstructed from the edge parameters by applying Eq. (4.2).

Edge parameter descriptions can be included in an image compression scheme by detecting edge locations and replacing the Hermite coefficients by edge parameters. In this section edge structures are detected by selecting local 1D image structures by thresholding the relative amount of 1D energy in the coefficients of order $n > 1$,

$$\frac{E_N^{1D}(\theta) - E_1^{1D}(\theta)}{E_N^{2D}(\theta) - E_1^{2D}(\theta)} \geq th1, \tag{4.8}$$

followed by thresholding $E_{1D}(\theta) \geq th2$ in order to discard low energy 1D structures. Subsequently, edge parameters are estimated for each selected window position, and edge positions are separated from non-edge positions by thresholding the relative error energy in the 1D coefficients, which is given by

$$\frac{\sum_{n=1}^{N} [f_n - \hat{f}_n]^2}{\sum_{n=1}^{N} [f_n]^2} \geq th3 \tag{4.9}$$

Figure 4.2:   Top: Detected edges in "Mondrian" and "Lena". The parameters of the Hermite transform were set to $(d, \sigma, N) = (4, 1.3, 5)$. Bottom: The original 512x512 8-bit greyscale images "Mondrian" and "Lena".

with $\hat{f}_n$ the 1D Hermite coefficients estimated from the edge parameters. Fig. 4.2 shows edge detection results based on a Hermite transform with parameter settings $(d, \sigma, N) = (4, 1.3, 5)$ and $(th1, th2, th3) = (0.8, 5, 0.1)$ for the images "Mondrian" and "Lena". To be able to extend the compression scheme of Chapter 3 with edge parameter coding, the coder must, of course, include the binary classification image indicating whether a window position is described by its Hermite coefficients or by a set of edge parameters.

The entropy reduction effect of edge parameter coding is demonstrated in Fig. 4.3. Peak-signal-to-noise ratio (PSNR) performance curves are shown for two images: "Mondrian", containing many step edges, and "Lena". The data were obtained by varying the quantization stepsizes for both DC and AC coefficients between 1 and 16. For the orientation adaptive cartesian Hermite decomposition, total entropy was calculated by adding the estimated entropy needed to
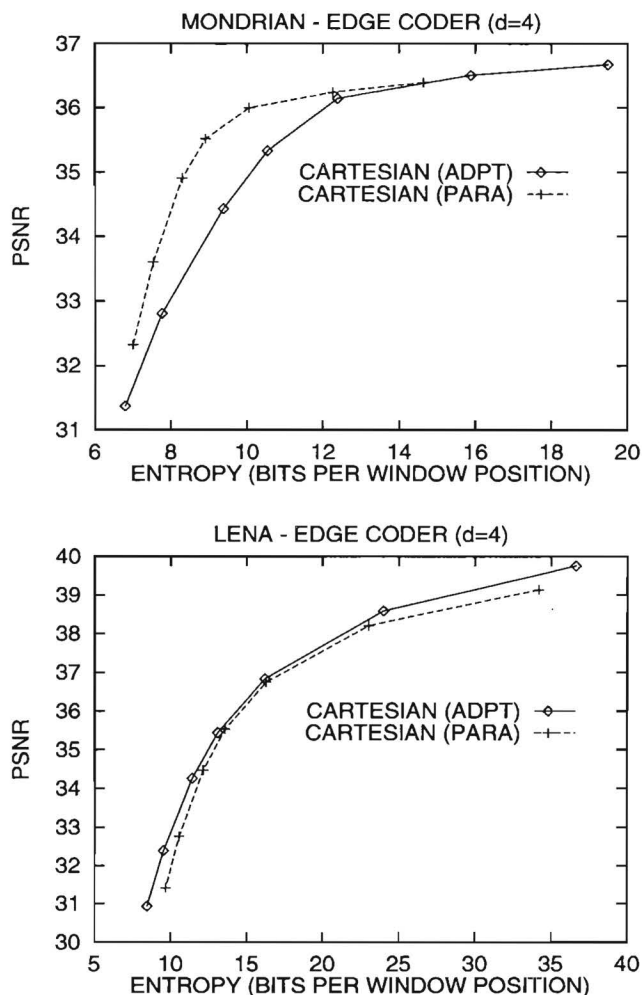
Figure 4.3: Effect of edge parameter coding on the estimated entropy needed to code both DC and AC Hermite coefficients. Results are shown for the images "Mondrian" (top) and "Lena" (bottom). Curves were obtained by varying the quantization stepsizes for the AC and DC coefficients (stepsizes were set to 1, 2, 4, 6, 8, 12 and 16). The quantization stepsizes for the parameters d, ΔV and b were set to 0.5, 16 and 2, respectively. Sixteen different orientations were used for all coders.

code the DC, the AC and the orientation parameter image using Eq. (3.25). The entropy for the edge parameter coder was calculated by adding the estimated entropies of the DC and orientation image, the edge position image, and the (conditional) entropies of the AC coefficients and the edge parameters d, ΔV, and b. In the case of the "Mondrian" image, the entropy is reduced significantly. For the "Lena" image, however, the orientation adaptive Hermite coder performs slightly better.

Feature-based coding can be extended by also representing other image structures explicitly. Line segments, for example, can be modeled by adding two (step) edges, resulting in line primitives such as position, line height, line width and mean value. For compression purposes, however, explicit coding of line segments and other structures such as corners and surfaces is still under investigation.

4.3 Contour Coding

Section 4.2 showed how to calculate local edge primitives from the coefficients of a steered Hermite transform. The rest of this chapter focuses on image representations containing only contour information and DC information. In this section, edges are detected by linking local maxima in the total energy image, which results in chains of local edge points. Weak edges are discarded by thresholding both the average energy of the edge points in the chain and the length of the chain. Up to now, only image descriptions within local windows have been considered. By linking local edge descriptions along contour-like structures it should be possible to exploit redundancies between adjacent edge positions. It can be expected that the edge parameters along a contour show only moderate variation. An example of edge chains detected in the test images "sails" and "motorbike" are given in Fig. 4.4. The original 480x240 8-bit greyscale images are given in Fig. 2.1. Note the difference between Fig. 4.4 and Fig. 4.2, in which the individual edge points are not linked into contours. By increasing the contour length threshold from 1 to 5, however, is becomes clear that the contour detection still results in many small contours.

When using image representations based on edge parameters only, a correct setting of the parameters of the Hermite transform is important. First of all, the sampling distance d should be small in order to enable accurate positioning of analysis windows. In this chapter, a typical sampling distance of $d = 2$ is used, since unacceptable results were obtained by taking $d = 4$ or $d = 8$. Also, the size of the analysis window, indicated by σ, determines the accuracy with which edges can be represented. A typical value for σ was found to be smaller than 2, although higher values can be selected for very low bit rate coding. A small d and a small σ, however, affect the efficiency with which low-

Figure 4.4:   Contour detection results in "sails" and "motorbike" from a steered Hermite transform with parameter setting (d, σ, N) = $(2, 1.3, 5)$. Results were obtained for an average energy threshold of 20, while the contour length thresholds were set to 1 and 5 edge points for the left and right images, respectively.

frequency information can be coded. A small d results in a large number of DC coefficients. This can be solved by using only those DC coefficients positioned on a sublattice in the DC image (e.g., a 2x2 or 4x4 sublattice can be used). In the case of a small σ, however, only a small subsampling lattice can be applied. Efficient coding of DC coefficients will be discussed in more detail in Section 4.4.

An example of an image reconstruction based on edge parameters only is given in Fig. 4.5. Image analysis was carried out using a steered Hermite transform with parameters (d, σ, N) = $(2, 1.3, 5)$ and 16 possible orientations, while the analysis filter length was set to $L = 10$. The result of the contour detection is shown in Fig. 4.5 (a). After contour detection, edge parameters were calculated for all contour positions. Since only edge parameter descriptions at contour positions were used in the reconstruction process, an additional modulation correction was applied to the reconstructed image after image synthesis. This is necessary because the modulation correction $\sum_{p \in \mathcal{P}} v_p^2(x, y)$ of the Hermite synthesis filters in Eq. (3.9) adds up over all window positions. The result is shown in image (b). Image (c) was reconstructed by using both the edge parameters along the contours and a subsampled version of the DC image (a sampling factor of 2 was used). In the contour coding approach in this chapter, DC information is not derived from the edge information (as is done in [14]), but is coded explicitly by retaining DC coefficients.

In Figs. 4.6 and 4.7 impressions of the variation in the edge parameters along
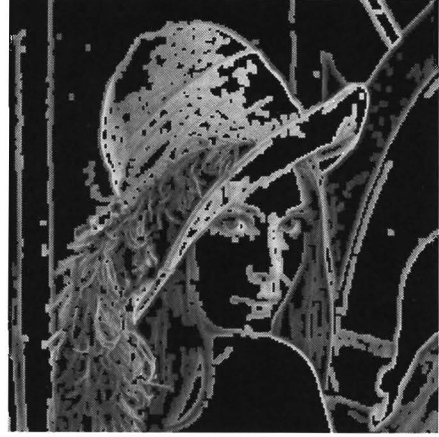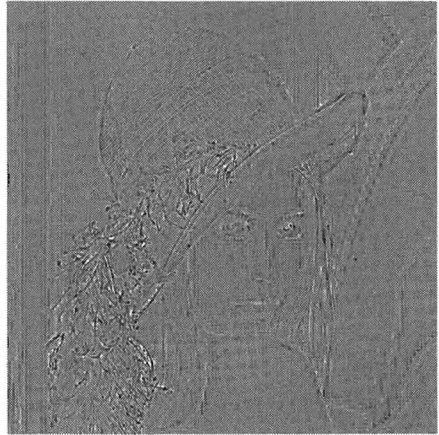
Figure 4.5:   Edge reconstruction using a Hermite transform with parameters $(d, \sigma, N) = (2, 1.3, 5)$ with 16 different orientations and filter length $L = 10$. (a) Edge positions used for reconstruction; (b) Edge parameter reconstruction using a synthesis filter with length $L = 6$; (c) Edge parameter reconstruction using DC coefficients on a 2x2 sublattice; (d) Error image, scaled by a factor of 4 and 128 added for display.

two pronounced contours in the "Lena" image are given. For each edge point in these contours, the DC value is shown along with the edge parameters d, ΔV and θ. Although the orientation parameter θ and the edge height ΔV only show moderate variation, the sampling effect is clearly visible in the edge position parameter d and the variation in the DC values along the contour. Due to the near vertical orientation of the contour along Lena's hair, a number of sign changes occur in the orientation parameter θ. These sign changes are also reflected in the value of the edge height ΔV. Further, notice the cross-correlation between the DC value and distance parameter d.

4.4   Image Compression Scheme

In Section 4.3 a qualitative impression of contour coding using edge parameters derived from a steered Hermite transform was given. In this section a more quantative approach is presented by including contour-based representations into an image compression scheme. Next to the efficient representation of edge information, the low-frequency information present in the image has to be coded, since this information is not derived from the edge representation. Before compression of edge parameters is discussed, a method for coding the DC values is presented.

4.4.1   Coding of DC Values

When applying a contour representation based on a Hermite transform with only a small sampling distance (e.g. $d = 2$), coding of the remaining DC component requires a relative large number of bytes. In the compression scheme presented in this chapter, the DC image is first subsampled in order to decrease the number of DC coefficients. Second, the remaining DC coefficients are heavily quantized (typically, quantization stepsizes between 16 and 32 are used), causing reconstructed images to show serious segmentation artifacts. In order to suppress these artifacts, an adaptive DC filtering algorithm is applied to the subsampled DC image at non-edge positions.

The filtering algorithm applied is similar to the one that is proposed in Chapter 5, and will not be discussed in detail here. The effect of DC filtering on reconstructed image quality, however, is demonstrated in Fig. 4.8. Both images in Fig. 4.8 are reconstructed by using quantized DC coefficients positioned on 4x4 sublattice. The edge parameters used for reconstructing all image contours were only slightly quantized. Since the distance d between the Hermite analysis windows was set to 2, the effective sampling distance of the DC coefficients was 8. The left image was reconstructed without applying the DC filtering algorithm. In this image, segmentation artifacts are clearly visible.

Figure 4.6: Edge parameters along one of the contours of Lena's hair. Plotted are the DC coefficient, edge distance d, edge height ΔV and edge orientation θ. Edge parameters were obtained for a Hermite transform with parameters $(d, \sigma, N) = (2, 1.3, 5)$ using 16 different orientations. Edges were constructed along maximum energy ridges.
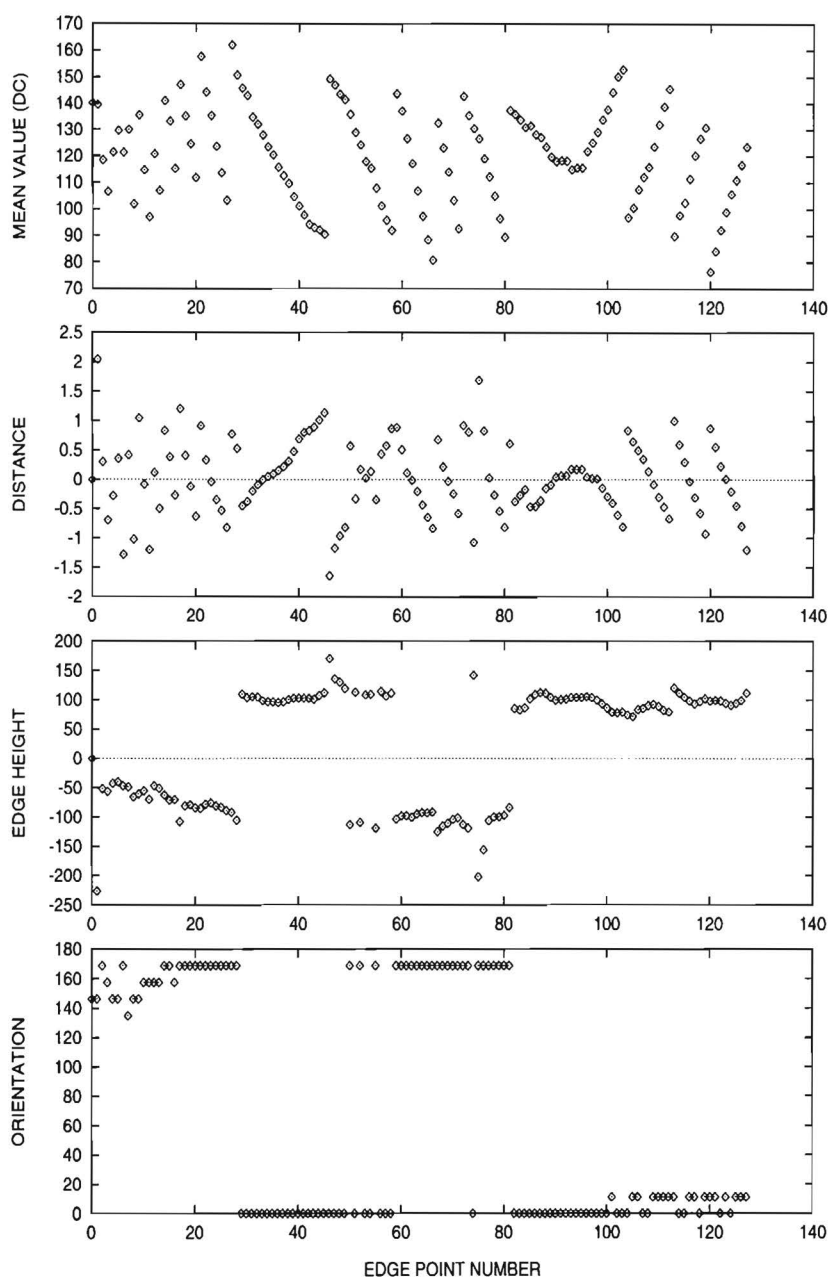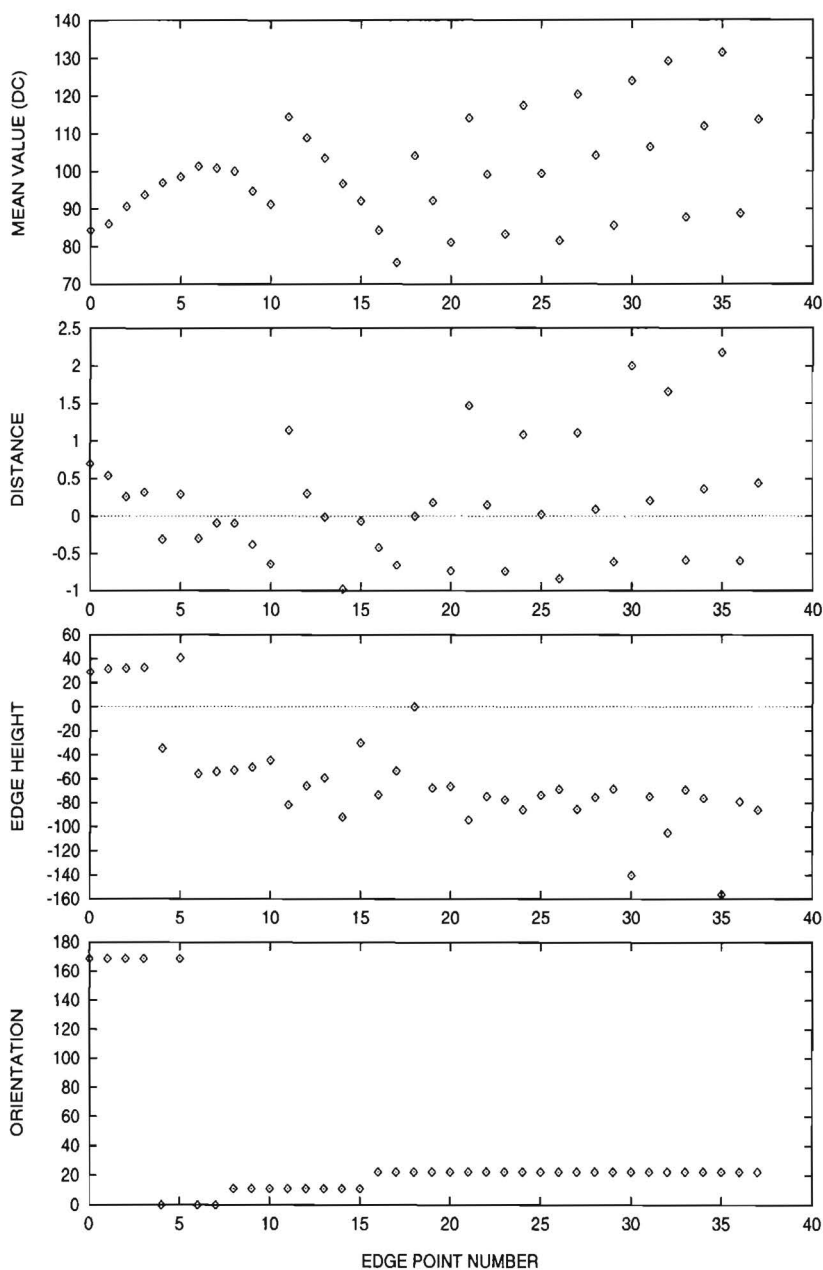
Figure 4.7: Edge parameters along one of the contours of Lena's face. Plotted are the DC coefficient, edge distance d, edge height ΔV and edge orientation θ. Edge parameters were obtained for a Hermite transform with parameters $(d, \sigma, N) = (2, 1.3, 5)$ using 16 different orientations. Edges were constructed along maximum energy ridges.

In the right image these artifacts are suppressed due to the effect of the DC filtering algorithm.

Figure 4.8:   Contour-coding using DC coefficients on a 4x4 sublattice (the DC quantization stepsize was set to 32). Left image: reconstruction without DC filtering; Right image: reconstruction with DC filtering. The image was coded using a Hermite transform with parameters $(d, \sigma, N) = (2, 2.5, 5)$ with 16 different orientations and filter length $L=16$.

4.4.2   Information to be Coded

The Hermite contour-coding process starts with selecting the settings (d, σ, N) of the Hermite transform, and the number of orientations to be distinguished (in this thesis, this number is usually 16). Next, based on the results of the steered Hermite transform, image contours are detected and edge parameter descriptions of all contour positions are calculated. The removal of weak contours requires the setting of two additional parameters: threshold $T1$ on the contour length and threshold $T2$ on the average contour energy. At non-contour positions, only DC information is retained. All higher-order Hermite frequency coefficients are set to zero. Next, one has to specify the following quantization steps: DC quantization for both DC values at contour positions and DC values at non-contour positions, and quantization for the edge parameters d, ΔV, and b. After quantization, the Hermite-based contour coding scheme retains the following information:

- Information on all contour positions

- Edge parameters at all contour positions. These parameters are: DC value, edge orientation θ, edge position d, edge height ΔV and edge blur parameter b.

- DC coefficients at non-contour positions. Only DC coefficients positioned on a sublattice of the DC image are retained.

The parameter settings of the Hermite transform are not coded. Before entropy coding is applied, all information is stored in an array of bytes. The structure of this array is depicted in Table 4.I. The first two bytes are used for coding the total number of contour locations. Next, the position of each contour is described by the x and y coordinates of its starting point and a lossless chain coding of all other points. For the chain coding, codeword values 0-7 are used in order to code the direction of the next point in the chain. The end of each contour specification is marked with an end-of-chain (EOC) codeword with a value of 8. Next, all DC values at the contour locations are coded, followed by coding of the other parameters. This means that the total number of bytes required for storing the edge locations and the edge parameters equals $2 + 6 * nr\_chain\_points + 2 * nr\_chains$. This number does not include the number of bytes needed to code the DC coefficients at non-edge positions.

4.4.3   Entropy Coding

Before the byte stream described in the previous paragraph is entropy coded using an arithmetic coder, a prediction scheme is applied in order to decrease the overall entropy. For coding that part of the byte stream which specifies all contour locations, a first order differential pulse code modulation (DPCM) scheme is applied to the y-coordinates of the starting points. No prediction is applied to the x coordinates. The chain coding symbols (range $[0, ..., 7]$) are also DPCM coded. After prediction, however, the range of all possible chain coding symbols will extend to $[-7, ..., 7]$. This can be solved by adding or subtracting a value of 8, which will map all values onto the range $[-3, ..., 3]$. The EOC value of 8 is not changed.

A first order DPCM scheme is also applied to each block of bytes containing one of the edge parameters DC, θ ΔV, and b. The edge distance d is not DPCM coded, since the prediction scheme increased the entropy of this parameter for most test images. Since the DPCM coding increases the range of values used to represent the orientation parameter θ, these values are mapped onto the range $[-\#\theta/2, ..., \#\theta/2 - 1]$, with $\#\theta$ the number of possible orientations.

Information to be coded	# bytes

# chain points						2

x-pos	y-pos	dir	.....	dir	EOC	2 + chain length
.....						
x-pos	y-pos	dir	.....	dir	EOC	2 + chain length

DC	.....	DC	# chain points
θ	.....	θ	# chain points
d	.....	d	# chain points
ΔV	.....	ΔV	# chain points
b	.....	b	# chain points

DC	.....	DC	—

Table 4.I: Structure of the byte stream in the Hermite contour-coding scheme. The table shows the number of bytes needed to store both edge positions and edge parameters. The number of bytes needed to store the remaining DC coefficients depends on the image content. See text in Section 4.4.2 for a detailed explanation.

4.5 Compression Results

In Table 4.II and Table 4.III contour coding results obtained for the test scenes "Lena" and "Sails" are indicated. In these tables, detailed information is given on the amount of information needed to code the different parts of the total byte stream. The tables divide the total byte stream into separate parts: First, that block of the byte stream needed to code all contour position information. Second, five equally-sized blocks corresponding to the edge parameters DC, θ, d, ΔV and b. Finally, the last block of the byte stream containing the remaining DC coefficients.

The initial size of each block is indicated in the column labeled *#bytes(1)*. The entropy of the codewords (expressed in number of bits) is given in the column *entropy(1)*. The next column, labeled *entropy(2)* indicates the entropy after DPCM coding, and in column *#bytes(2)* the total number of bytes needed for entropy coding the entire block is calculated (based on the entropy after DPCM coding). The stepsizes used for quantization are indicated in the column labeled *Q*. Finally, the tables indicate for each block the percentage of the

(a)

Lena (PSNR $= 30.63$, $C = 15.1$)				$(d, \sigma, N, \#\theta) = (2, 2.5, 5, 16)$		
No_of_chains $= 694$			Thresholds: $T1 = 2$ (length) $T2 = 20$ (energy)			
	Q	# bytes (1)	entropy (1)	entropy (2)	# bytes (2)	%
Contours	—	8788	3.418	3.418	3755	21.6
DC	8	7396	4.028	2.688	2485	14.3
θ	—	7396	3.711	2.317	2142	12.3
d	0.5	7396	2.939	2.939	2717	15.6
ΔV	8	7396	4.934	3.295	3046	17.5
b	1	7396	2.212	1.844	1705	9.8
DC	8	3610	4.140	3.406	1537	8.8
		49378	4.038	3.249	17387	100.0

(b)

Lena (PSNR $= 29.57$, $C = 13.8$)				$(d, \sigma, N, \#\theta) = (2, 1.3, 5, 8)$		
No_of_chains $= 912$			Thresholds: $T1 = 2$ (length) $T2 = 20$ (energy)			
	Q	# bytes (1)	entropy (1)	entropy (2)	# bytes (2)	%
Contours	—	9827	3.555	3.555	4367	23.0
DC	16	7999	3.093	2.334	2334	12.3
θ	—	7999	2.769	2.093	2093	11.0
d	1	7999	2.483	2.483	2483	13.1
ΔV	16	7999	4.007	3.182	3182	16.8
b	2	7999	1.972	1.987	1987	10.5
DC	32	14303	2.294	1.414	2528	13.3
		64125	2.882	2.435	18974	100.0

(c)

Lena (PSNR $= 28.70$, $C = 21.8$)				$(d, \sigma, N, \#\theta) = (2, 1.3, 5, 4)$		
No_of_chains $= 657$			Thresholds: $T1 = 2$ (length) $T2 = 50$ (energy)			
	Q	# bytes (1)	entropy (1)	entropy (2)	# bytes (2)	%
Contours	—	7690	3.458	3.458	3324	27.6
DC	32	6372	2.109	1.654	1318	11.0
θ	—	6372	1.791	1.236	984	8.2
d	2	6372	1.614	1.614	1285	10.7
ΔV	32	6372	3.145	2.416	1924	16.0
b	8	6372	0.713	0.796	634	5.3
DC	32	14696	2.297	1.390	2553	21.2
		54246	2.161	1.795	12022	100.0

Table 4.II: Contour coding results for the 512x512 8-bit greyscale
"Lena" test image. See text for a more detailed explanation.

(a)

Sails (PSNR = 26.34, C = 15.3)				$(d, \sigma, N, \#\theta) = (2, 2.5, 5, 16)$		
No_of_chains = 742		Thresholds: $T1 = 2$ (length) $T2 = 20$ (energy)				
	Q	# bytes (1)	entropy (1)	entropy (2)	# bytes (2)	%
Contours	—	9422	3.090	3.090	3639	21.3
DC	8	7934	4.076	2.479	2458	14.4
θ	—	7934	3.626	2.348	2329	13.6
d	0.5	7934	3.033	3.033	3008	17.6
ΔV	8	7934	4.945	3.161	3135	18.3
b	1	7934	1.250	1.236	1225	7.2
DC	8	3614	4.315	2.899	1309	7.7
		52706	3.476	2.607	17103	100.0

(b)

Sails (PSNR = 27.62, C = 12.8)				$(d, \sigma, N, \#\theta) = (2, 1.3, 5, 8)$		
No_of_chains = 1075		Thresholds: $T1 = 2$ (length) $T2 = 20$ (energy)				
	Q	# bytes (1)	entropy (1)	entropy (2)	# bytes (2)	%
Contours	—	12451	3.232	3.232	5030	23.8
DC	16	10297	3.225	2.161	2782	13.1
θ	—	10297	2.694	2.022	2603	12.3
d	1	10297	2.386	2.386	3071	14.5
ΔV	16	10297	4.011	2.954	3863	18.2
b	2	10297	1.276	1.324	1704	8.0
DC	32	13806	2.488	1.226	2115	10.0
		77742	2.759	2.186	21168	100.0

(c)

Sails (PSNR = 26.84, C = 18.6)				$(d, \sigma, N, \#\theta) = (2, 1.3, 5, 4)$		
No_of_chains = 890		Thresholds: $T1 = 2$ (length) $T2 = 50$ (energy)				
	Q	# bytes (1)	entropy (1)	entropy (2)	# bytes (2)	%
Contours	—	10918	3.151	3.151	4300	30.5
DC	32	9134	2.215	1.504	1717	12.2
θ	—	9134	1.735	1.194	1364	9.7
d	2	9134	1.522	1.522	1738	12.3
ΔV	32	9134	3.094	2.163	2470	17.5
b	8	9134	0.190	0.302	345	2.5
DC	32	14087	2.495	1.218	2145	15.2
		70675	2.057	1.579	14079	100.0

Table 4.III: Contour coding results for the 512x512 8-bit greyscale "Sails" test image. See text for a more detailed explanation.

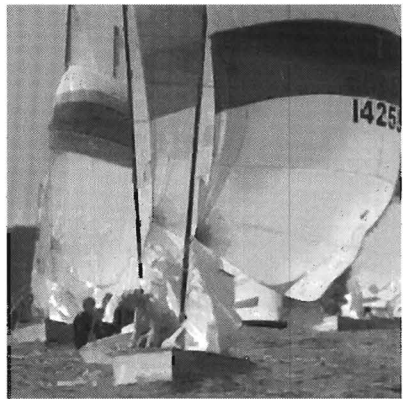4.II (a)   4.III (a)
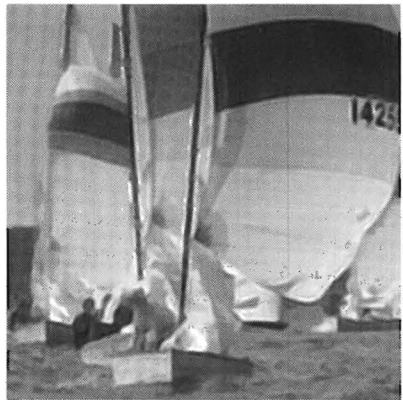
4.II (b)   4.III (b)

4.II (c)   4.III (c)

Figure 4.9: Results of contour coding on "Lena" and "Sails".

total bytestream (after entropy coding) corresponding to that particular block. The compression ratios indicated in the examples are based on the entropy calculations. Actual compression ratios obtained via arithmetic coding of the byte stream typically deviate less than a few percent.

Fig. 4.9 shows fragments (size 360x360 pixels) of the reconstructed images that correspond to the compression results given in Tables 4.II and 4.III. In the examples 4.II and 4.III (a) contour coding results are given for the test scenes "Lena" and "Sails" at a compression ratio of $C = 15$ for both images. In example (a), the Gaussian window spread σ used in the Hermite transform was set to 2.5 and the number of possible orientations was set to 16. In examples (b) and (c) the spread parameter σ was set to 1.3, while the number of orientations was set to 8 and 4, respectively. The results given in part (b) correspond to compression ratios $C = 14$ for "Lena" and $C = 13$ for "Sails". The parameter settings of the Hermite transform can be found in the headings of each table. The values of the thresholds $T1$ on the contour length and $T2$ on the average contour energy are also indicated.

The effect of σ on reconstructed image quality is clearly visible. Small values of σ result in sharp edge reconstructions, while high values cause edge blur. Compare, for example, the black digits in the reconstructions (a) and (b) of the "Sails" image. In general the Hermite contour-coded images are fairly well reconstructed at moderate compression rates ($C \approx 16$). At lower bit rates the contours rapidly deteriorate and image quality becomes unacceptable. When increasing the thresholds on both contour length and average contour energy, too few image contours remain. Also, course quantization of the edge parameters results in poor image quality. This means that compared to the compression algorithm based on the steered Hermite transform (see Chapter 3), the performance of the contour-based Hermite coder does not yet meet the expectations (both in terms of PSNR performance and in terms of perceptual image quality).

The figures in Tables 4.II and 4.III show that only a small percentage of the total compressed bytestream is needed to code the low-frequency uniform regions. Due to the DC filtering algorithm applied in the uniform regions, the quality of the DC component remains fair, even for quantization steps of 32. The contribution of the contour positions in the total bytestream, however, is rather high. For the "Sails" image, example (c), this contribution is about 30 percent, with a bit rate of 3.8 bits per edge point (this corresponds to 0.5 bits per window position). In literature, the total cost for coding edge points is generally considered to be much smaller (e.g., 1.3 bits per contour position [14, 20]). The high number of bits needed to code the edge points can partly be explained by the low average length of the detected contours. For the "Lena" and "Sails" examples, the average contour length is approximately 10 edge points.

4.5.1   Suggestions for an Alternative Coding Strategy

The compression results in Section 4.5 indicate that more time is needed to obtain acceptable compression of both the edge point locations and the parameter signals θ, ΔV, d and b along the image contours. Although the value of the edge parameters is expected to vary only slightly along each edge curve, the amount of redundancy appears limited when looking at the results presented in Tables 4.II and 4.III. In the parameter signals sharp transitions occur due to the many sign changes (see Figs. 4.6 and 4.7). Also, the parameter signals contain too many additional sharp transitions due to the large number of small contours of limited length. Therefore, the DPCM scheme used for coding the parameter signals should account for the starting points of every contour. Further, the compression scheme presented in Section 4.4 does not make optimal use of redundancies between the different parameter bands. In Figs. 4.6 and 4.7 it can be observed that the parameter bands show significant cross-correlation, which is not yet exploited in the proposed system. Also, separate coding of the edge point positions and the orientation parameter θ is not optimal, since this information can (partly) be deduced from each other.

An improved coding strategy would be to remove the sign information from the parameter signals and code this information separately. Subsequently, lossy coding of the parameter signals within individual chains of edge points could be applied by low-pass filtering and subsampling the parameter signals. The coding of the edge point positions could be done more efficient by designing a higher order Markov model [111, 14]. Due to a lack of time, however, these improvements have not been implemented.

4.6   Discussion

In Chapter 3, the Hermite transform was selected as a tool for image analysis, and it was shown that this transform could be used to detect local pattern orientations. The steered form of the Hermite transform even integrated explicit representation of detected (local) orientations in the image description. Chapter 4 demonstrates that explicit coding of image contours can also be integrated in a Hermite image representation.

Using a Gaussian edge model allows edge parameters other than the edge orientation θ, such as the edge distance d, the edge height ΔV and the edge blur b to be calculated from the frequency coefficients of the steered Hermite transform. Since only a Gaussian edge model is used it is, of course, impossible to obtain a perfect fit for all edge structures present in an image. By extending the modeling of edge elements with non-Gaussian edge elements and, for example, line segments or corner elements, it should be possible to built image

representations that can represent the original image more faithfully. For image compression purposes such an approach would require the coding of additional classification information. In this thesis, only explicit representation of edges is addressed.

Edge parameter coding of the test image "Mondrian" showed that edge descriptions in terms of edge primitives may decrease the entropy of the image representation. For natural images such as "Lena" and "Sails", however, replacing the frequency coefficients by the appropriate edge parameters does not necessarily improve the efficiency of the image representation. The derivation of edge primitives given in Section 4.2 was still limited to the representation of (local) individual edge elements. Therefore, in Section 4.3, these local descriptions were linked into chains in order to construct a contour-based image representation. In the rest of Chapter 4 image reconstruction is focused on contour-coding and images are represented by edge information and low-frequency information only. In Section 4.4 the proposed contour-based image representation is included in a compression scheme in which redundancies along contours are exploited, and the variation of edge parameters along image contours is visualized.

When using the steered Hermite transform as a basis for edge-parameter based coding, information other than edge information does not necessarily have to be discarded. This means that even with integrated edge-parameter coding, residue coding (e.g. for texture reconstruction) remains possible. Image reconstruction based on the DC component and the edge description only is, of course, achieved by setting all residue coefficients to zero. Usually, for high quality image reconstruction, the performance of second-generation coders is much less than that of conventional transform coders such as JPEG or subband coders. Typical contour coding concepts as used by Carlsson [20] and van Beek [14], for instance, always result in low quality images, because no texture information can be coded. For low quality low-bit-rate coding, however, the performance of such algorithms is better than the performance of conventional coding schemes. Carlsson solved this problem by separately coding the residual between the original image and the sketch reconstructed image using the Laplacian pyramid coding algorithm [19]. Unfortunately, the performance of the contour-based scheme of Chapter 4 is less than the performance of the compression scheme based on the steered Hermite transform presented in Chapter 3. In the contour-based scheme, the correlation between adjacent edge parameters was found to be rather limited, and the edge position coding was found to be more expensive than expected.

The typical artifacts of contour-based compression schemes also occur in the Hermite-based contour coder. When, for instance, the compression ratio of an image is increased by course quantization of the edge parameters, the recon-

structed contours show distortions comparable to the contour distortions in the work of Carlsson [20]. The perceived quality of uniform regions, however, is relatively high in the case of the Hermite coder, because uniform regions are reconstructed by decoding and filtering heavily quantized DC coefficients at non-edge window positions (this DC filtering algorithm will be discussed in Chapter 5). In both the edge primitive coder of van Beek [14], and in Carlsson's coder, the uniform regions appear unnatural, due to smooth interpolation between reconstructed contours. Their work does show, however, that on the basis of edge information only, the entire image can be reconstructed.

One may conclude that contour coding is only effective for low-bit-rate coding of specific classes of images containing pronounced contours and pronounced uniform regions (such as the "Mondrian" image used in this chapter and the "scarf" image applied in [14]). In most papers on edge-based coding, only images with pronounced edge structures are used for demonstration purposes. Apparently, for a general coding scheme that can handle any type of images, contour coding is not yet worked out sufficiently. In this chapter it was found that, amongst others, the Hermite contour coder suffers from the large number of (small) contours detected in the images. Therefore, the performance of this scheme could improve considerably if a more robust contour tracing algorithm would be applied in order to increase the average length of the detected contours. Additional smoothing of contours would also make the proposed compression scheme more efficient.

Chapter 5

Reconstruction of Uniform Regions

Abstract

In transform coding schemes, images are usually reconstructed in the decoder by calculating the inverse transform of the coded transform coefficients. Taking the inverse transform usually results in transform specific artifacts in the reconstructed image. Image quality, however, can be improved by spending more effort in the design of the decoder. Since blocking in smooth areas of the image is one of the most important causes of decreasing image quality, many postprocessing algorithms concentrate on removing these blocking artifacts. This chapter shows that the standard JPEG postprocessing method, which is based on prediction of low-frequency AC coefficients in smooth areas of the image, is only effective in a small compression range. In particular, due to its local character, this method fails in removing segmentation artifacts caused by course quantization of the DC coefficients. In this chapter an algorithm is proposed which restores the DC coefficients in uniform areas such that, in combination with the JPEG AC prediction method both segmentation artifacts and blocking artifacts can be reduced significantly. Since the proposed algorithm can be applied to any block coder, the effect of this technique is also demonstrated for a Hermite coding scheme. The algorithm is particularly effective at low bit rates.

5.1 Introduction

The major degradations occurring in block encoded images are *blocking artifacts* and *edge distortion artifacts*. Blocking manifest itself as an artificial boundary discontinuity between adjacent image blocks and is mostly visible in areas of the original image with monotoneously and weakly changing luminance. Edge distortion emerges in areas with sharp luminance transitions and is

characterized by jaggedness and Gibbs-like phenomena along edges. In JPEG, for example, blocking typically occurs at compression ratios higher than 15. Block coders using overlap such as the Hermite coder discussed in Chapter 3 and the Lapped Orthogonal Transform [84] also show blocking at borders. In overlap coders, though, blocking occurs at much higher compression ratios (C>30).

Most compression schemes reconstruct coded images by taking the inverse transform of the decoded transform coefficients. Since quantization is a many-to-one mapping, the same (quantized) coefficient data can be generated by many different source images. This implies that it is impossible to retrieve the original image. In order to be able to select a candidate image which is (perceptually) closer to the original, more sophisticated decoding algorithms are required. For the past few years, several methods have been proposed that deal with reducing the artifacts in image reconstruction.

The blocking effect problem can, for instance, be dealt with by introducing *overlap* between adjacent transform windows. In this way, abrupt boundary discontinuities caused by coding are reduced because the reconstruction process weaves image blocks together. The major disadvantage of this method is an increase in bit rate due to redundancies in the image representation (pixels in overlapping areas are coded more than once). Therefore, the performance of redundant overlap methods is rather poor compared to, for example, post-processing algorithms using simple spatial filtering (Reeve & Lim [114], 1984).

By applying image enhancement after decoding instead of adapting the coding process, reconstructed image quality can be improved without increasing the bit rate. The basis of *spatial filtering* methods is smoothing unwanted discontinuities by low-pass filtering the image at or near block boundaries. In order to avoid degradation of edge content in the image (e.g., blur), most methods applying low-pass filtering are usually combined with edge detection methods eventually extended with filtering along the orientation of edges to remove jaggedness. A sophisticated nonlinear space-variant filter which adapts to the varying shape of the local signal spectrum, and reduces only the locally out-of-band noise, while the signal and the in-band noise is left mostly unchanged, is proposed by Ramamurthi and Gersho [112]. This is achieved by employing a two-dimensional filter in areas away from edges, and when near edges, a one-dimensional filter aligned parallel to the edge. In Macq et al. [78], low-pass filtering is used for block noise estimation and postprocessing is performed by subtraction of the estimated coding noise in the uniform regions of the reconstructed image.

The disadvantage of filtering methods is that there is no a priori guarantee that after being filtered, the original image still matches the compressed data. In some approaches [42], an extra inverse image transform is applied followed by

clipping of the transform coefficients to guarantee image integrity. In this case, however, decoder complexity is enormously increased, especially if the filtering is applied iteratively [42]. Also, filtering is usually applied with small filters in order to retain detail information. This implies that most filtering methods are in particular effective in compression ranges in which artifacts are mainly local (i.e. filtering is usually not effective at very low bit rates).

More advanced methods for postprocessing reduce typical coding artifacts by assuming an explicit model for the original image data. An example is the JPEG recovery method described in Annex K.8 of the JPEG standard [108]. This decoder option assumes the image to be a quadratic surface and missing AC coefficients are predicted to fit this surface (see section 5.2 for more details). The advantage of K.8 is that it is a simple and effective operation, performed in the transform domain. No extra inverse transform is necessary to ensure that decoded coefficients remain within the quantization boundaries.

Other algorithms using a model-based approach are described by Yang et al. [152, 153], who use the theory of projections onto convex sets (POCS) for image recovery. The key idea behind POCS based image reconstruction is that known properties about the original are expressed in the form of convex constraint
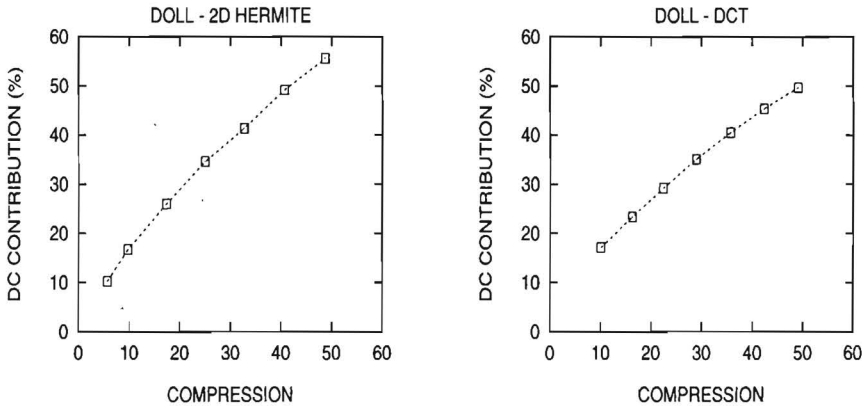

Figure 5.1:   Contribution of the DC coefficient in the compressed bytestream for a typical Hermite-coded image (left) and a DCT coded image using a standard quantization matrix (right). In case of the Hermite coder, for instance, at a compression ratio of 30, almost 40 percent of the compressed bytestream is used for representing the DC image. For both coders, the quantized DC coefficients were compressed using differential pulse code modulation (DPCM) followed by arithmetic coding. Curves are shown for the "doll" test scene.

sets that capture the smoothness properties of the desired image. The image that satisfies all the imposed constraints lies in the intersection of these sets and is found by alternating projections onto these sets. Yang et al. used both spatially invariant prior knowledge to reconstruct the compressed images along with the transmitted data and spatially adaptive smoothness constraint sets. Although the algorithms using POCS perform very well, the simplified algorithm proposed in [153], already requires 10 to 17 times the computations of a conventional decoder without any postprocessing. Another technique is described by Stevenson [131], who restores images by taking the maximum *a posteriori* (MAP) estimate based on probabilistic models for both the noise introduced by the coding and for a "good" image. This technique, however, also requires an additional inverse transform for each iteration.

The goal of this chapter is to take a closer look at the reconstruction of smooth or uniform areas in low bit-rate coding. In particular, the impact of DC quantization on perceived image quality is demonstrated. Since in block transform coding schemes the DC coefficients absorb a relative large part of the compressed bytestream (see Fig. 5.1), more attention should be paid to efficient (de)coding of the DC image. In Section 5.2 the JPEG decoder option K.8 is discussed and an experiment is carried out to measure the effectiveness of this postprocessing as a function of the level of quantization. Section 5.3 proposes an enhanced algorithm in which a low-pass filtering of the DC coefficient is carried out prior to applying the standard K.8 decoder option. An evaluation of this algorithm is presented in 5.4. Section 5.6 reviews the main conclusions of this chapter.

5.2   JPEG Decoding Using Prediction of AC Coefficients

An interesting way in which performance of transform coding schemes can be improved is by applying more sophisticated decoding schemes. An example of such a scheme is proposed in Annex K.8.1 and Annex K.8.2 of the JPEG coding standard [108]. The decoder option described in this standard is based on the recovering of low-frequency AC coefficients at block positions in which, due to course quantization, only the DC coefficient is available. By fitting a quadratic surface through the DC coefficients within a 3x3 array of blocks centered on the block in question, a prediction can be made for the low-frequency AC coefficients of the centered block.

The effect of applying the AC coefficient prediction as just described is demonstrated in Fig. 5.2. Image (a) shows a detail of the "Lena" image reconstructed with a standard JPEG baseline coder using a quantization scaling factor of 6. In this image, due to course quantization of both DC and AC coefficients, the uniform regions suffer from *segmentation* artifacts. Image (b) was reconstruc-
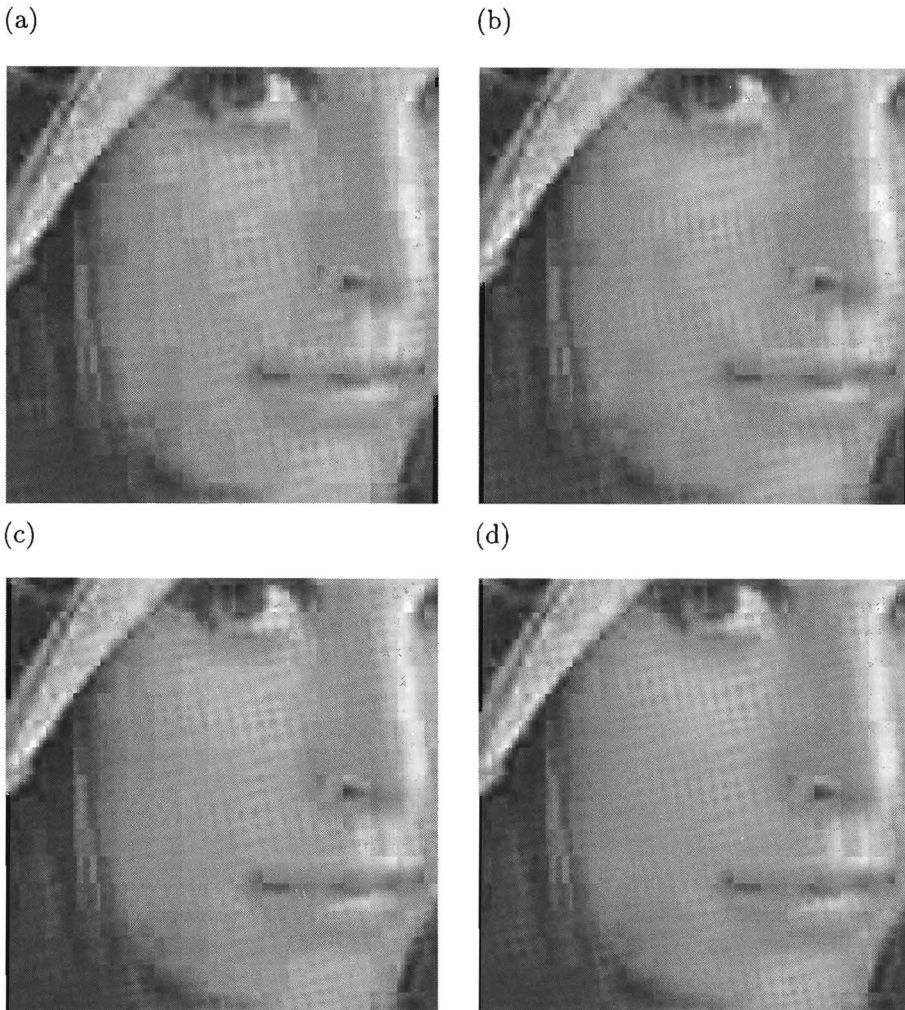
(a) (b)

(c) (d)

Figure 5.2: Effect of low-frequency AC prediction in the decoder. (a) Reconstruction using JPEG with a quantization scaling level of $Q = 6$. (b) Reconstruction using JPEG with a quantization scaling factor of 6 and coefficient prediction according to Annex K.8 of the JPEG standard. In image (b) the segmentation artifacts remain visible. (c) Reconstruction using JPEG with a quantization scaling factor of 6 for the AC coefficients only. (d) Reconstruction using JPEG with a quantization scaling factor of 6 for the AC coefficients and low-frequency AC prediction. In the latter case, the uniform regions are more smoothly reconstructed.

ted from the same compressed bytestream. In this case, however, low-frequency AC prediction was applied in the decoder, resulting in a smoothing of the segment boundaries. Unfortunately, the segmentation artifacts remain visible and the reconstructed smooth regions in the image show a "spotty" and "blurred" appearance.

In Fig 5.2 (c), the image is reconstructed using a quantization scaling factor of 6 for the AC coefficients only (i.e., the DC quantization level for the DC component is not scaled). Since most uniform regions are only represented by the DC coefficients, the boundaries between all individual transform blocks are clearly visible. In image (d), however, coefficient prediction has been applied in the decoder, resulting in a smooth reconstruction of the uniform regions. In this image, the uniform regions clearly suffer less from any blocking and/or segmentation artifacts.

In order to measure the effect of applying AC coefficient prediction at various levels of quantization, a subjective quality assessment was carried out for the 480x240 8-bit grayscale test scenes "lena", and "sails" (see Fig. 2.1). For both scenes, eleven pairs of JPEG-coded images were generated. Six of these pairs were constructed by multiplying the default JPEG quantization matrix with factors ranging from 1 up to 6, and combining each of these 6 JPEG comparison images with images reconstructed from the same JPEG-decoded coefficients using AC prediction in the decoder. The other 5 stimulus pairs were generated by scaling only the AC quantization steps with factors between 2 and 6 and, again, combining them with images reconstructed from the same decoded coefficients using AC prediction.

Three male and three female subjects between 22 and 30 years of age participated in the experiment. They had normal or corrected-to-normal vision and a visual acuity, measured on a Landolt chart, between 1.5 and 2.0. Each test scene was handled in a separate session. Because a repetition rate of 8 was used, the subjects had to judge, in each session, 88 randomly displayed pairs of images of which they were asked to rate the difference in image quality on a numerical scale ranging from -5 to 5. The plus and the minus signs were used to indicate whether the left or the right image was preferred.

The stimuli were displayed on a 50 Hz non-interlaced BARCO CCID7351B monitor which was placed in a dark room against a white, dimly lit 2.5 cd/m2 background. The monitor had a peak luminance of 70 cd/m2 and a gamma of 2.5. The subjects observed the image pairs, simultaneously displayed on the left and on the right hand sides of the screen, from a distance of 5 times the stimulus height. Since each image occupied an area of 22 x 11 cm, this resulted in a horizontal size of 5.7 degrees of visual angle and a resolution of 42 pixels per degree of visual angle. The stimulus presentation time was set to 9 seconds. Before each session a representative training sequence of 12 image pairs was
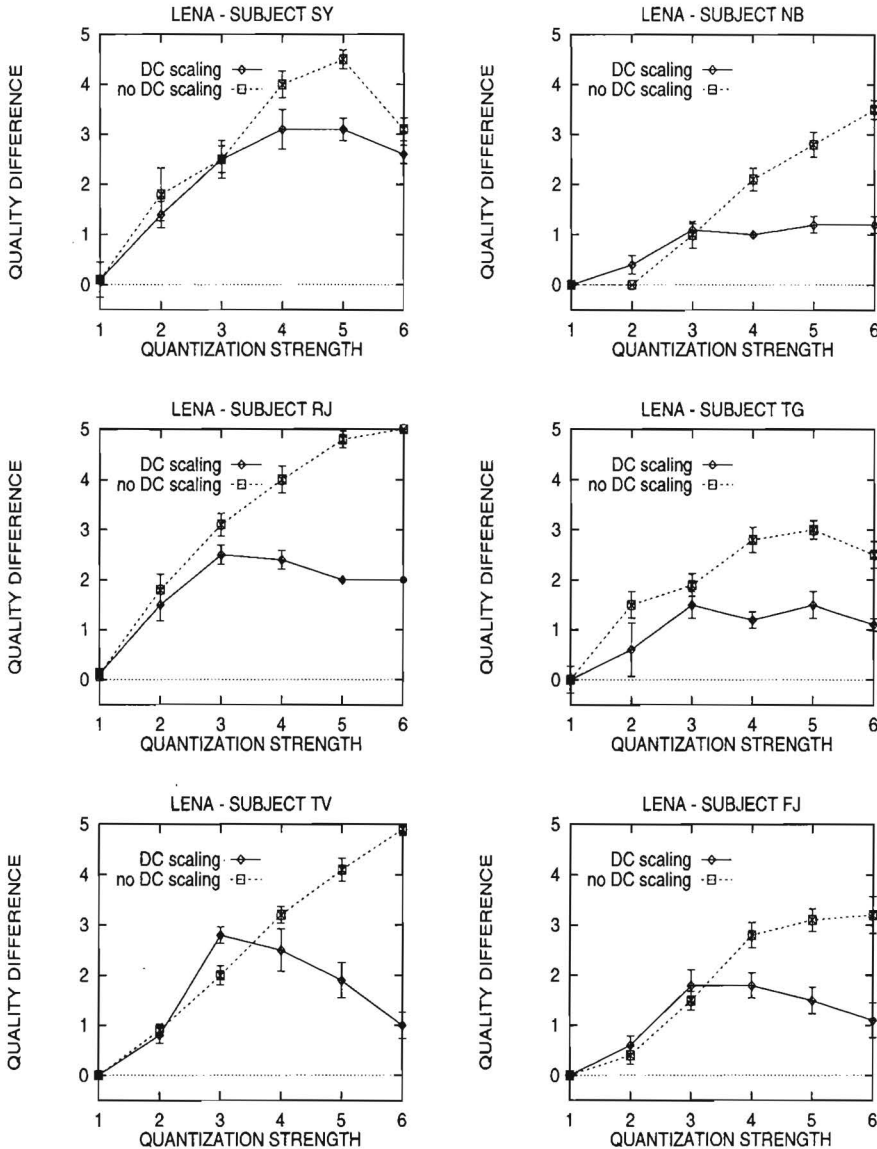
Figure 5.3:   Individual mean raw scores of all subjects for the test scene "lena", indicating the difference in quality for JPEG-decoded images compared to images decoded with JPEG using a prediction of AC coefficients in the decoder. The lengths of the error bars are equal to twice the standard error of the mean.
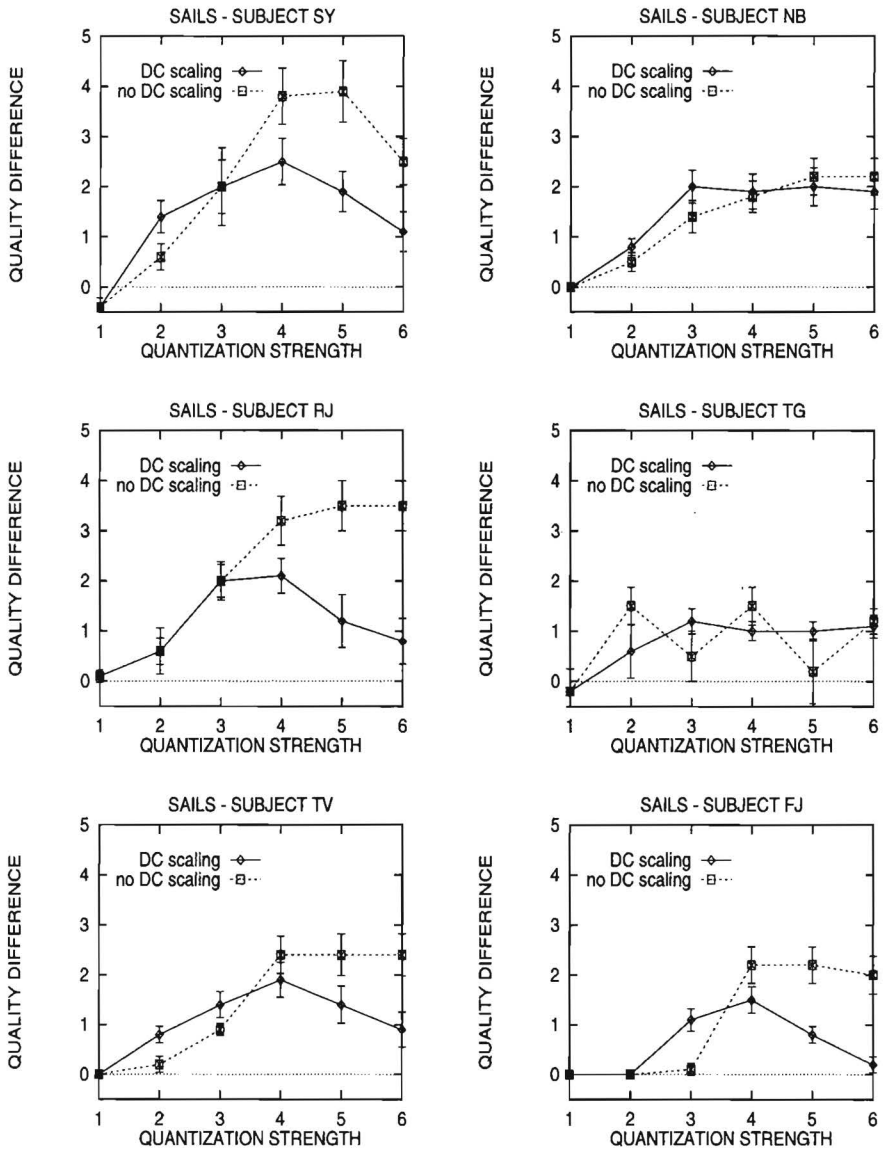
Figure 5.4: Individual scores of all subjects for the test scene "sails",
indicating the difference in quality for JPEG-decoded images com-
pared to images decoded with JPEG using a prediction of AC coeffi-
cients in the decoder. The lengths of the error bars are equal to twice
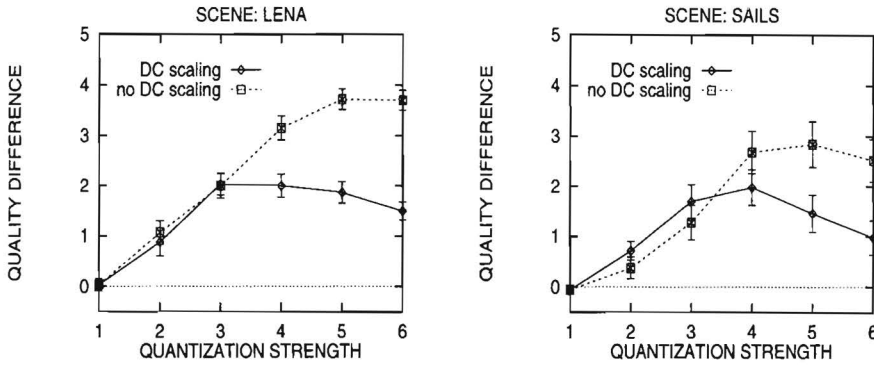the standard error of the mean.

Figure 5.5: Averaged difference-in-quality scores indicating the effect of low-frequency AC prediction on reconstructed image quality at various levels of quantization. Results are shown for the test scenes "lena" and "sails". The lengths of the error bars are equal to twice the average standard error of the mean.

presented to the subject.

Individual difference-in-quality scores for each image pair were obtained by taking the mean of the eight scores corresponding to that particular stimulus. Since the experiment was balanced (i.e., each image was displayed an equal number of times on the right and on the left hand side of the monitor), the raw data were sign-corrected. Figs. 5.3 and 5.4 show the individual results of all subjects collected for the test scenes "lena" and "sails", respectively. Positive scores correspond to a preference of the subject for the image decoded using AC prediction. Averaged results for both test scenes are shown in Fig. 5.5.

Figure 5.5 indicates that for both test scenes, the prediction of low-frequency AC coefficients indeed improves image quality. The effect of AC prediction, however, strongly depends on the level of quantization. If the quantization level of the DC component is not changed (dashed lines), the effect of AC prediction increases for coarser levels of quantization of the AC coefficients. This is explained by the fact that a coarse quantization results in a larger number of blocks in which all AC coefficients are set to zero and thus an increased number of blocks at which AC prediction can be applied. As expected, for a quantization scaling factor of 1, the AC prediction does not show any effect.

If both DC and AC quantization levels are scaled, the effect of AC prediction already saturates at a quantization scaling factor of about 3. In this case, the coarse quantization of the DC component results in a segmentation effect rather than visibility of many individual blocks. Due to its local character, the postprocessing only results in smoothing of segments boundaries. Recon-

structed images appear to be less sharp and perceived image quality may even drop. This reduces the effect of AC coefficient prediction on reconstructed image quality.

The plots in Figs. 5.3 and 5.4 show that most of the individual scores show the same trend. For most subjects, in the case of a standard JPEG coder in which the quantization levels of both DC and AC coefficients are scaled, the effect of AC coefficient prediction shows a maximum approximately between quantization levels Q=3 and Q=4. For higher levels of quantization, the effect of AC prediction decreases.

5.3   Image Restoration by DC Filtering and AC Prediction

In the previous section it was shown that AC prediction in the decoder is significantly more effective if the DC coefficients have not been heavily quantized. Due to the local character of the AC prediction, annoying segmentation artifacts remain visible at coarse DC quantization levels. Accurate coding of DC values, however, is very expensive and rather impossible at high compression ratios (see Fig. 5.1). A possible solution to this problem would be to precede the local AC prediction by a restoration of the DC component. In this way it is possible to suppress the typical segmentation artifacts caused by the DC quantization. In this section, an algorithm is proposed which restores the DC coefficients at the decoder before any AC prediction is applied.

5.3.1   Algorithm

The algorithm described in this section prevents segmentation artifacts in reconstructed images. Since segmentation artifacts are caused by inaccurate coding of DC coefficients, the algorithm concentrates on restoring the DC image in the decoder. More specific, restoration is done by low-pass filtering the decoded DC image. Fig. 5.6 shows the block diagram of the proposed DCT (discrete cosine transform) decoding scheme. In this scheme, image smoothing is achieved by including two extra stages in the decoder: (1) DC restoration by low-pass filtering, and (2) AC prediction by quadratic surface fitting. Since low-pass filtering of the DC image results in artifacts in high frequency areas of the image, smoothing of the DC image is only applied in uniform regions of the image.

After JPEG dequantization, the DC image is low-pass filtered in the uniform regions of the image. First, uniform regions are detected by selecting only those positions for which only the DC coefficient has remained. The selection, however, can be made less strict by setting a threshold $T_{dc} > 0$ on the

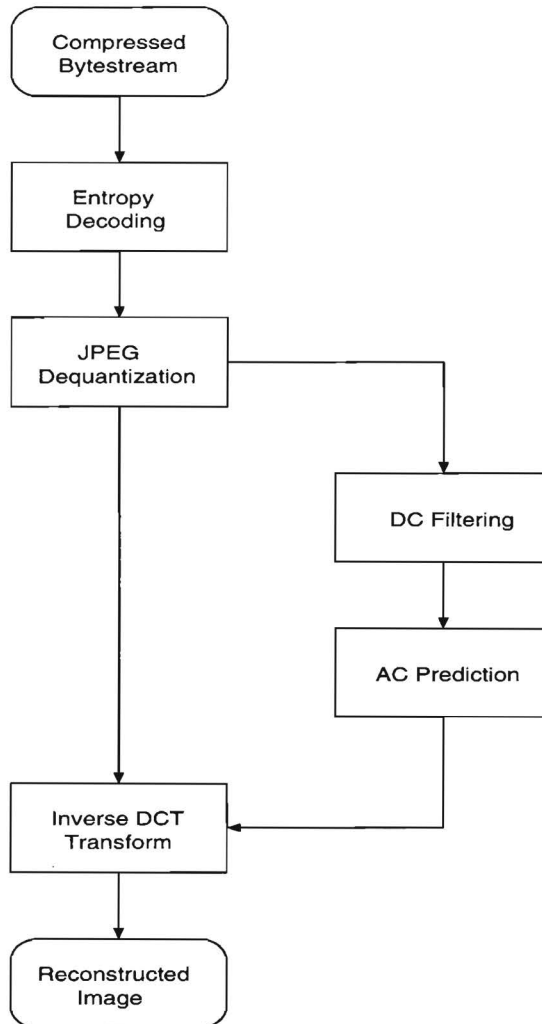Figure 5.6:   Proposed DCT decompression scheme using postprocessing. In a standard JPEG decoder, dequantization of the transform coefficients is directly followed by an inverse DCT transform. The postprocessing proposed in section 5.3, however, first low-pass filters the decoded DC image and performs AC prediction before the inverse transform is applied. In this way, blocking artifacts are reduced.

maximum amount of energy remaining in the decoded AC coefficients of each transform window. In this way, also locations with little remaining AC energy are classified as being uniform. The next step is to filter the DC image at the selected uniform locations. Fig. 5.7 shows two examples of low-pass filters that can be used.

In order to be able to filter nearby edge locations, the filtering in controlled by a binary mask, indicating for each filter position which filter positions should be used and which positions should be discarded. This mask is generated at each selected position in the DC image. Only mask positions corresponding to uniform locations are set to one. Furthermore, a selected mask position is set to zero if the absolute difference between the corresponding DC value and the center DC value exeeds a threshold T_{mask} (Usually, T_{mask} is set equal to the DC quantization stepsize). At each position, the result of the filter operation is divided by the sum of all filter tabs corresponding to the filter mask. If the result of the filter operation lays outside the quantization interval of the original decoded DC-value, the result is set equal to either the lower bound or the upper-bound of this interval. Clipping of the DC values is necessary to ensure image integrity.

1	2	1
2	4	2
1	2	1

1	4	6	4	1
4	16	24	16	4
6	24	36	24	6
4	16	24	16	4
1	4	6	4	1

0	0	0	0	1
0	0	0	1	1
0	0	1	1	1
1	1	1	1	1
1	1	1	1	1

Figure 5.7:   Binomial low-pass filters of size 3x3 and 5x5 used for DC filtering, and an example 5x5 filter mask indicating valid filter positions.

Although only one filter operation is already sufficient for smoothing the image, in principle, the filtering operation can be applied iteratively. Finally, after the DC restoration, AC coefficient prediction according to the JPEG K.8 decoder option is applied in order to remove all visible boundaries between the individual transform blocks.

The advantages of the method described in this section are clear. First, the filtering operation is only applied on the DC component of the DCT representation. Second, no extra inverse DCT transform is needed to ensure image

integrity, since the filtering operation is performed in the transform domain and the filtered result can directly be forced to remain within the quantization interval. In other methods, complexity is increased by performing multiple inverse DCT transforms [42, 131]

5.3.2   Results

This section briefly describes the effect of the algorithm proposed in Section 5.3.1. Fig. 5.8 shows the effect of DC restoration on the test scene "sails". In this case, the two-dimensional 5x5 binomial filter shown in Fig. 5.7 was used for filtering the DC image. It can be seen that all segmentation artifacts are completely removed by the filtering. From the image indicating the difference between the standard JPEG reconstructed image and the image reconstructed using the DC filtering, it can be seen that only uniform regions are affected by the filter operation.

In the next section a subjective evaluation of the proposed DC filtering method is discussed. In general, the perceptual image quality of reconstructed smooth areas improves significantly. However, the peak-signal-to-noise ratio of the reconstructed images (see Section 3.2 for a definition of the PSNR) is hardly changed if the DC filtering algorithm is applied (see also Fig. 5.13 in Section 5.4). The same goes for the standard K.8 AC prediction, which even results in a slight decrease of the PSNR for most images. By visual inspection, it was found that the DC filtering shows a significant result for DCT quantization multiplication factors higher than 3. Although the best results were found for a filter size 5x5, especially for high compression ranges, DC filtering using a filter size 3x3 also showed nice results. Further, the threshold T_{dc} should not be set too high in order to prevent edge positions to be classified as being uniform.

The DC filtering method described is only effective, of course, in the smooth regions of the decoded image. This implies that image quality in high-energy regions in the reconstructed image is not improved and, in particular along edge-like structures, there is still need for another form of postprocessing to remove typical DCT coding artifacts.

5.4   Evaluation of DC Filtering and AC Prediction

In this section, an experiment comparing three different coders is discussed: a standard JPEG coder, a JPEG coder using AC prediction according to the JPEG decoder option K.8, and a JPEG coder using both DC filtering and AC prediction. The experiment was carried out for three test scenes: "lena", "sailboats" and "doll". For each coder, four reconstructed images were generated
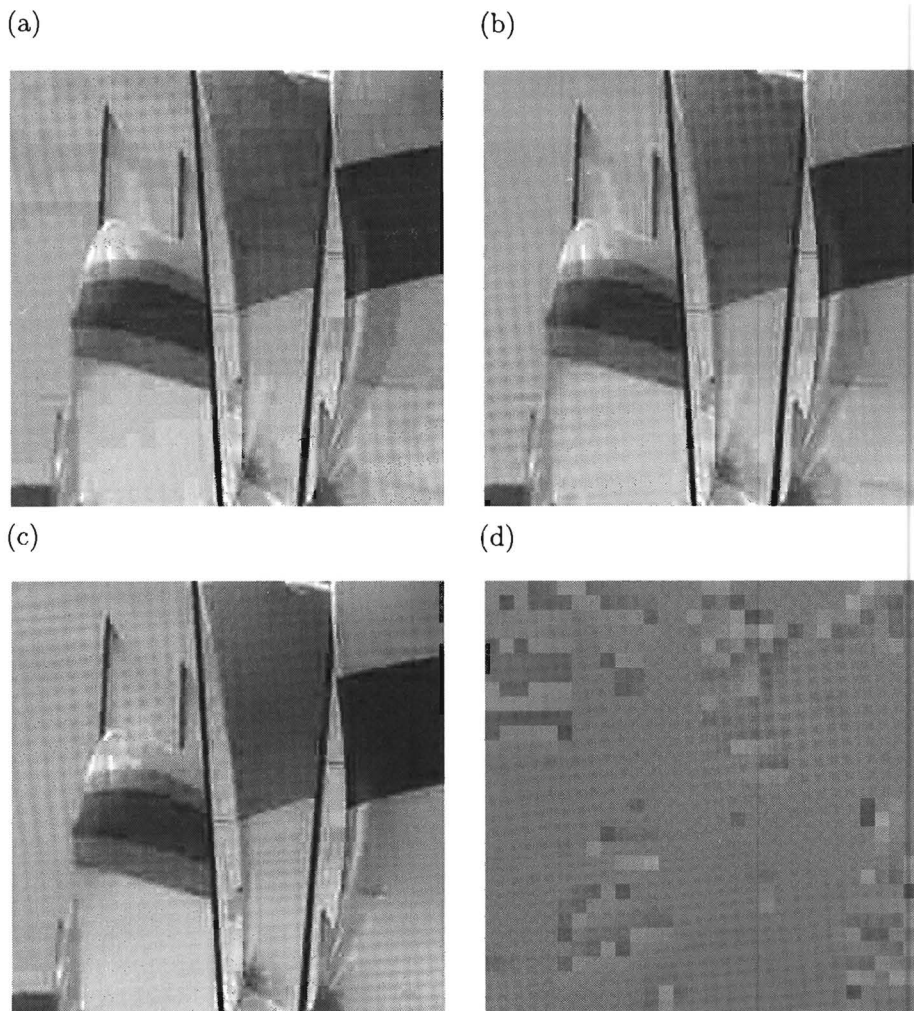
(a)   (b)

(c)   (d)

Figure 5.8: Effect of DC restoration. (a) Reconstruction using JPEG
with a quantization multiplication factor of $Q = 5$. (b) Reconstruc-
tion from the same compressed bytestream with DC restoration in the
decoder. (c) Reconstruction from the same compressed bytestream
with both DC restoration and AC prediction in the decoder. The
threshold T_{dc} applied in the DC filtering was set to 5000. (d) Dif-
ference between (a) and (b). For display purposes the gray values of
the difference image were scaled by a factor of 4, and a value of 128
was added.

Figure 5.9:   Individual quality scores calculated with the MDS program "Multiscale" for the test scene "lena". Results are shown for the subjects SY, RJ, TV, TR, MB and SH.

Figure 5.10: Individual quality scores calculated with the MDS program "Multiscale" for the test scene "sails". Results are shown for the subjects SY, RJ, TV, TR, MB and SH.
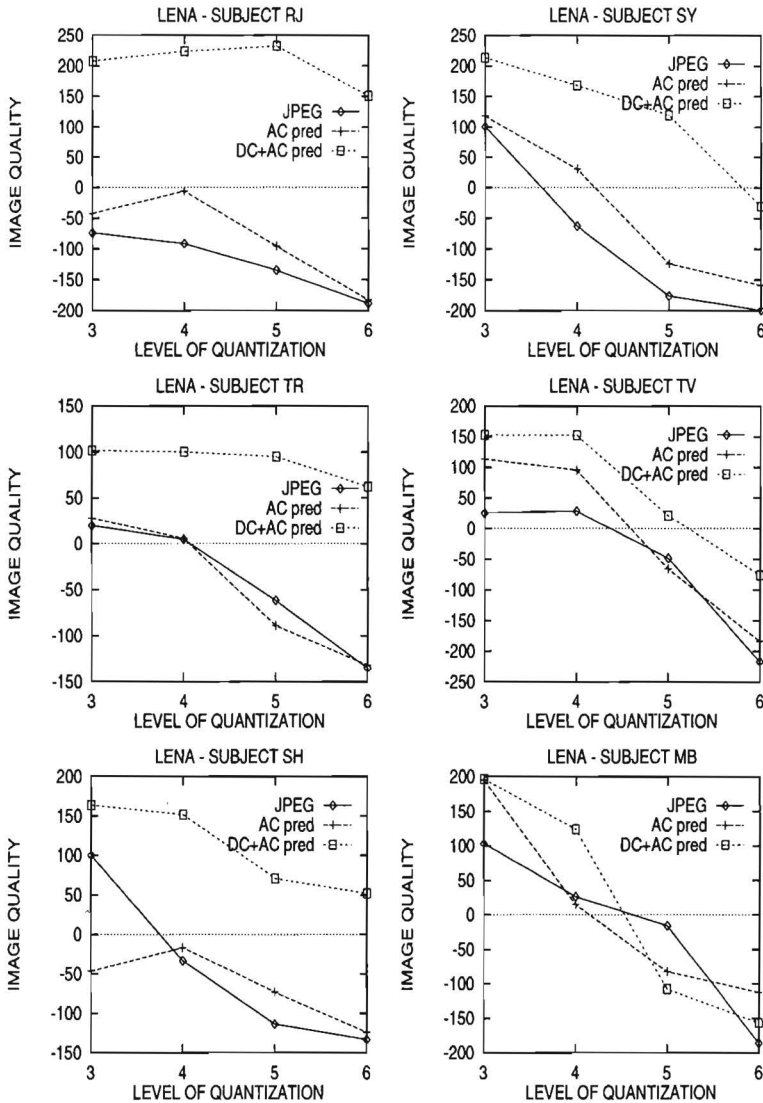
Figure 5.11: Individual quality scores calculated with the MDS program "Multiscale" for the test scene "doll". Results are shown for the subjects SY, RJ, TV, TR, MB and SH.
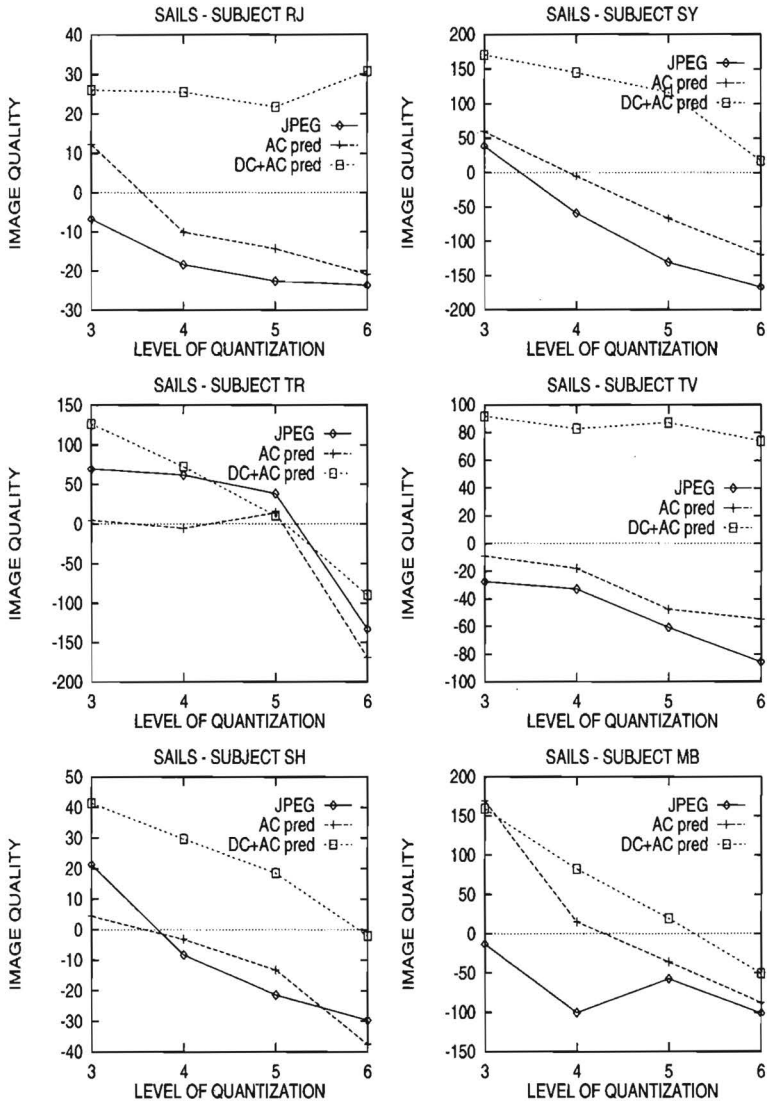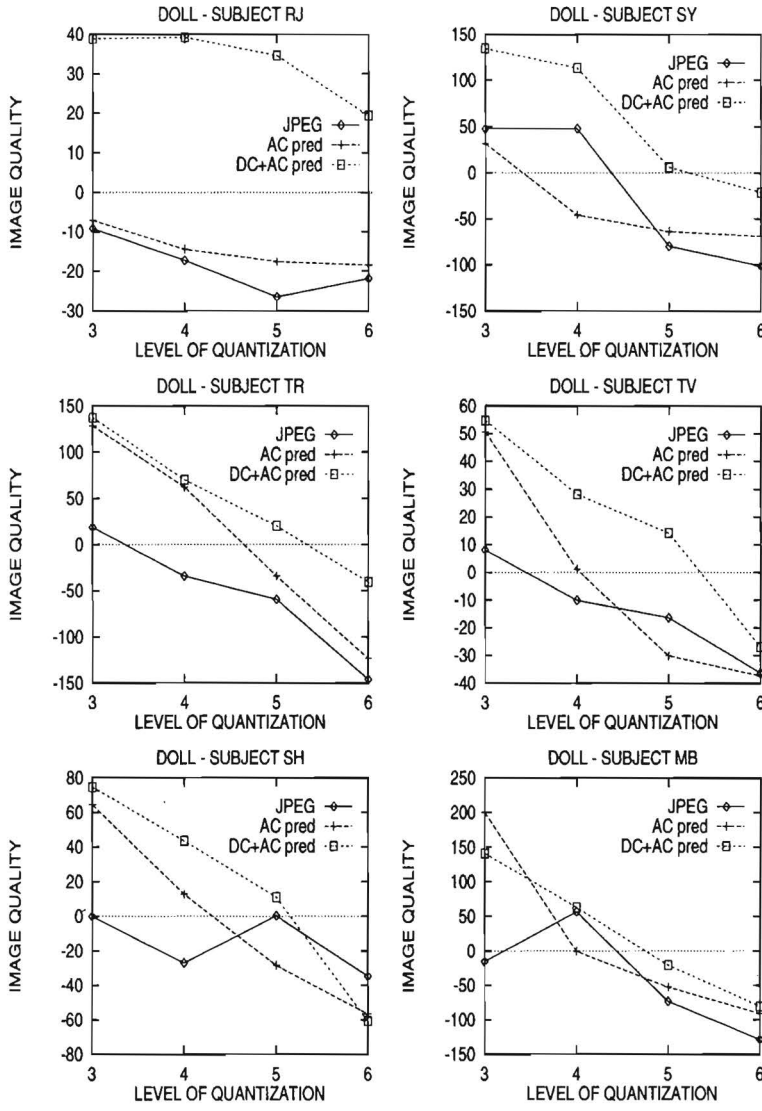
by taking quantization scaling levels 3, 4, 5 and 6. Thus, for each test scene, 12 different images were generated. For the DC filtering, the 5x5 filter of Fig. 5.7 was used (the threshold T_{dc} was set to 5000). The subjects were asked to rate the difference-in-quality of all combinations of the 12 images with the 4 images of the coder in which both DC filtering and AC prediction was used. Quality differences were rated on a numerical scale between -10 and 10, the plus and minus sign indicating whether the left or the right image was preferred.

The experiment was carried out under the same conditions as the experiment described in section 5.2. Since all test scenes were handled in the same session, subjects had to judge a total of 144 stimuli, preceded by a training sequence of 12 image pairs containing a representative subset of the images.
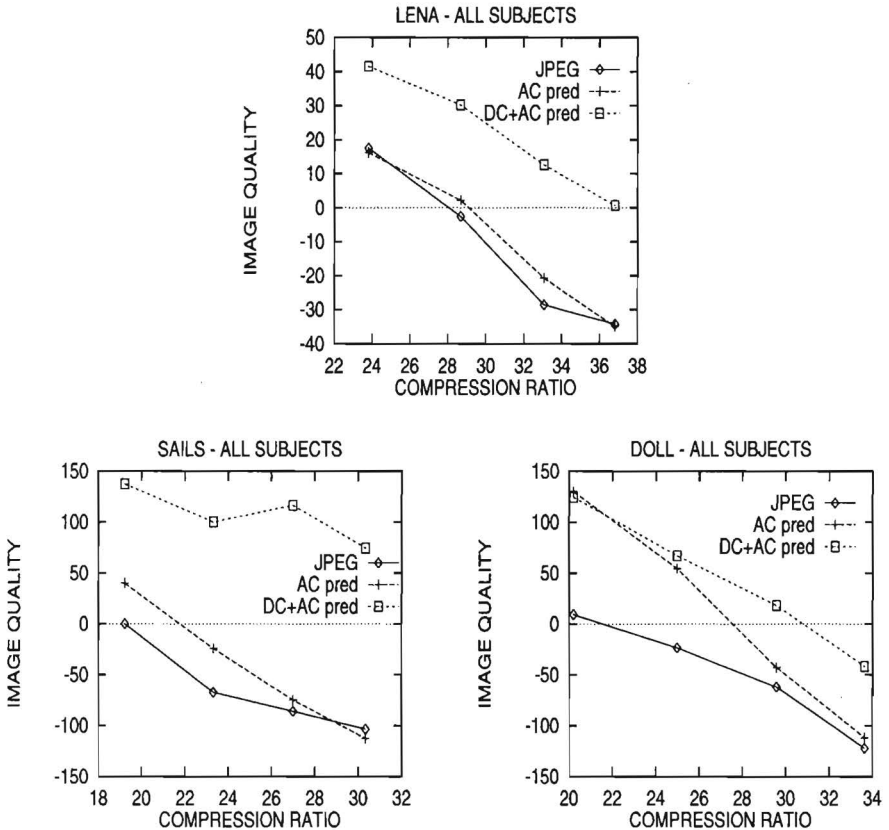

Figure 5.12: Overall quality scores calculated with Multiscale. Results are shown for all three test scenes.

From each session, a 4x12 matrix with difference-in-quality scores was obtained for each test scene. Each matrix was separately used as input for the multi-dimensional scaling (MDS) program "Multiscale" (see also Section 2.6), resulting in individual quality scores for all test scenes. All quality scores calculated for the individual subjects are given in Figs. 5.9, 5.10 and 5.11. Fig. 5.12 shows the overall "Multiscale" results calculated from the data of all subjects together.

For the "lena" image, the DC filtering results in a significant improvement in image quality at all quantization levels tested. For the "sails" image, the improvement is even more pronounced , because this image contains many important uniform regions. The effect of DC filtering in the "doll" image is much less because the improvements only manifest in the uniform regions in the background, while the objects in the foreground are hardly processed. Again, the results also show that AC prediction only has hardly any effect on image quality at high levels of quantization. Fig. 5.13 shows that the differences in image quality after postprocessing are not caused by an improvement in PSNR.
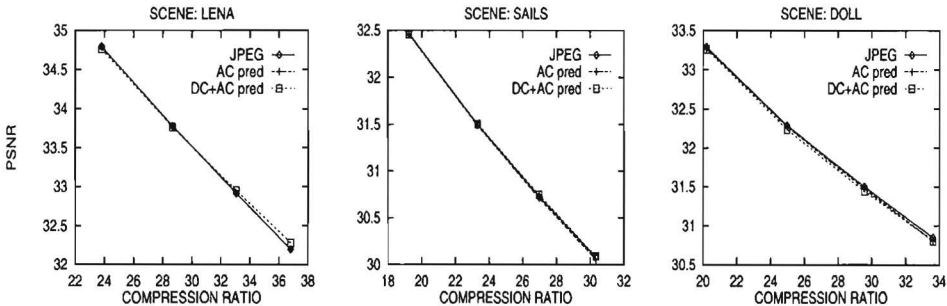
Figure 5.13: PSNR curves of all reconstructed images used in the experiment described in Section 5.4. The plots shows that for all scenes, the postprocessing applied does not improve the PSNR of the decoded images.

5.5 Application of Adaptive DC Filtering in a Hermite Coder

In the previous sections adaptive DC filtering was applied in a JPEG coding scheme. The same algorithm can, of course, be used for filtering DC images in any other block transform compression scheme. This section demonstrates the effect of DC filtering on reconstructed image quality of Hermite-coded images.

(a)

(b)

(c)

(d)
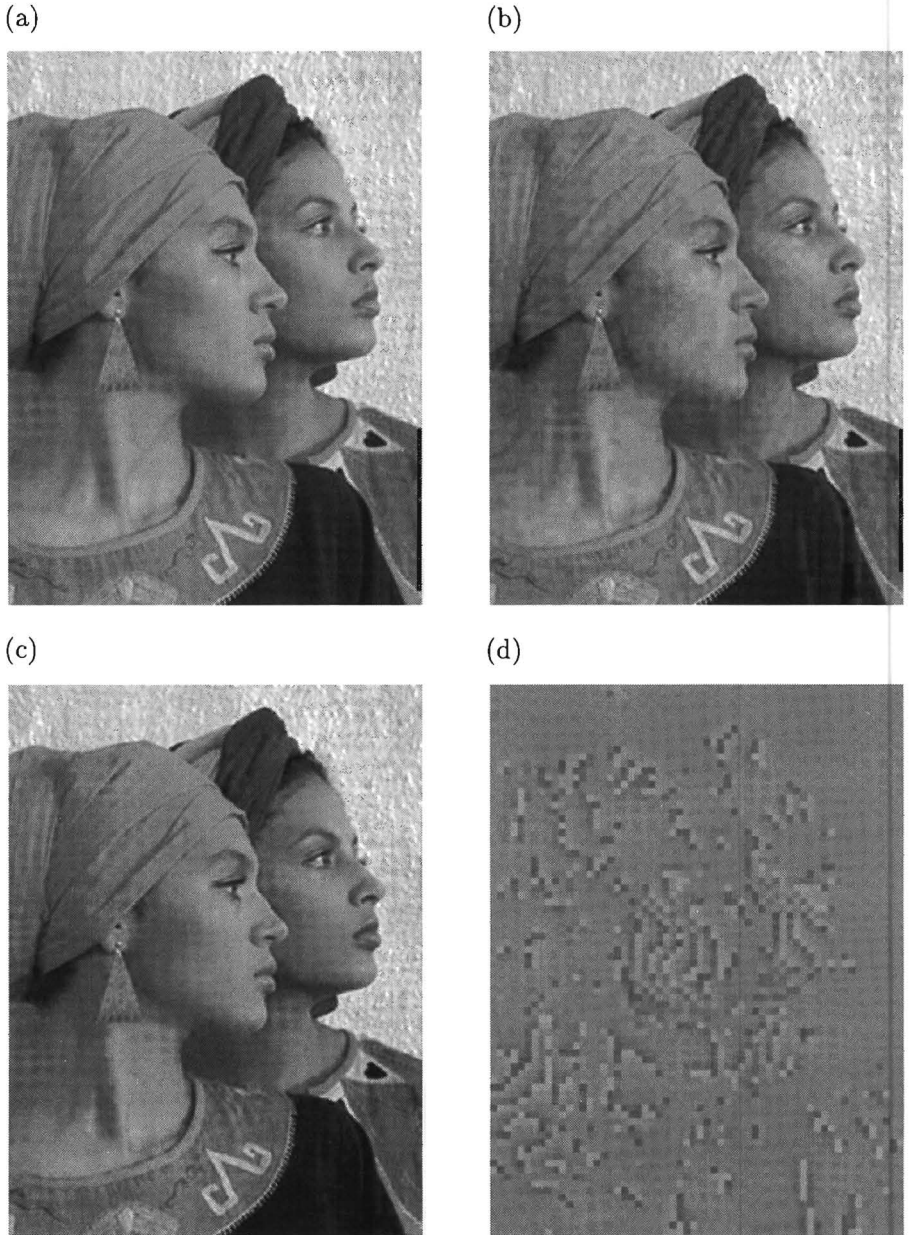
Figure 5.14: Effect of DC filtering in a Hermite coder with transform parameters $(d, \sigma, N) = (4, 1.3, 5)$. (a) Original image (size 320x240 pixels). (b) Hermite reconstructed image with $Q_{DC} = 16$ and $Q_{AC} = 1$. (c) Image reconstructed with DC filtering. (d) Difference between (b) and (c), scaled with a factor of 4 and 128 added for display. For the DC filtering, a binomial filter of length 5 was used and T_{dc} was set to 20.

(a)

(b)

(c)

(d)

Figure 5.15: Effect of DC filtering in a Hermite coder with transform parameters $(d, \sigma, N) = (8, 2.5, 9)$. (a) Original image (size 320x240 pixels). (b) Hermite reconstructed image with $Q_{DC} = 32$ and $Q_{AC} = 1$. (c) Image reconstructed with DC filtering. (d) Difference between (b) and (c), scaled with a factor of 4 and 128 added for display. For the DC filtering, a binomial filter of length 5 was used and T_{dc} was set to 5000.
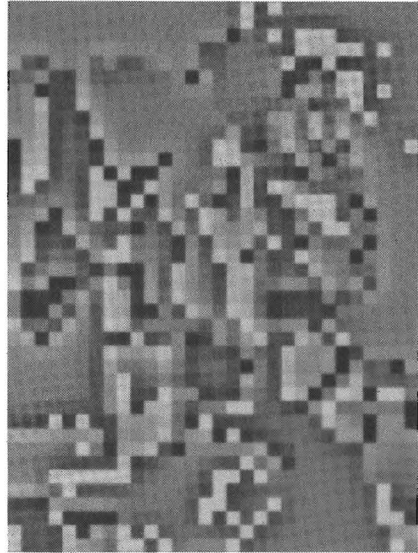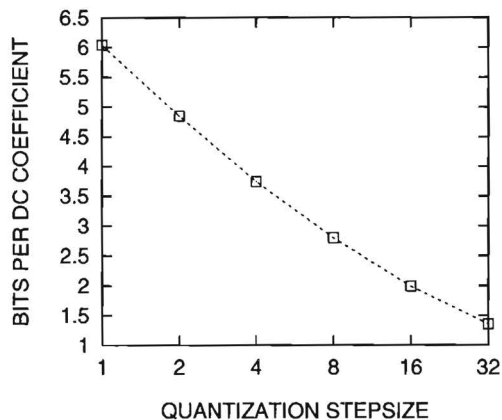
Figure 5.16: Measured number of bits in the compressed bytestream per DC coefficient as a function of the level of quantization. The data were obtained for "lena" from a Hermite coder with transform parameters $(d, \sigma, N) = (4, 1.3, 5)$. The DC coefficients were coded using a DPCM scheme followed by arithmetic coding.

In a Hermite coder, high compression of the DC image can also only be achieved at the cost of severe segmentation artifacts. Figure 5.16 gives an impression of the average number of bits needed to code a DC coefficient in a Hermite coder as a function of the level of quantization. For a Hermite coder, a DC quantization stepsize of 1 corresponds to 256 different gray levels. Since segmentation already becomes visible at a DC quantization stepsize equal to 4, it is clear that accurate coding of the DC coefficients is rather expensive. The Hermite transform has the advantage, however, that by using overlapping Gaussian windows, reconstructed uniform regions are less susceptible to local blocking artifacts. Therefore, in a Hermite coder, DC filtering alone, without AC prediction already suppresses most artifacts in uniform regions caused by DC quantization. For the Hermite coder used in this section, no orientation adaptivity was applied.

Figs. 5.14 and 5.15 show the effect of DC filtering on a detail of the image "girls" for two different parameter settings of the Hermite transform. In both figures, image (a) is given as a reference image, reconstructed from the Hermite coefficients quantized with quantization stepsize 1. Image (b) is reconstructed with strongly quantized DC coefficients (stepsizes $Q_{DC} = 16$ and $Q_{DC} = 32$, in Figs. 5.14 and 5.15, respectively. The quantization of the AC coefficients is not changed ($Q_{AC} = 1$ for all AC coefficients). It can be seen that in both cases,

the DC quantization results in unacceptable degradation of image quality.

Image (c) shows the Hermite-coded images reconstructed from the same Hermite coefficients. Most of the segmentation artifacts have been removed. The artifacts remaining manifest themselves mostly at high energy positions such as edges and lines. Fig. 5.15 does show, however, that even extremely high values of threshold T_{dc} (in this example the threshold was set to 5000) do not necesarily lead to additional distortions in high-energy areas. Note that the Figs. 5.14 and 5.15 also demonstrate that details in the image are not affected by low-pass filtering the DC image. Image (d) was obtained by scaling the difference between image (b) and image (c) with a factor 4 and adding a gray value of 128. No subjective evaluations were carried out to measure the effects of DC filtering in a Hermite coder.

5.6 Discussion

In block transform coding, smooth areas in the image are susceptible to annoying blocking and segmentation artifacts, which highly effect overall perceived image quality. In Chapter 2 this was already demonstrated by Watson's DCTune technique: at high compression ratios, more accurate coding of low frequency regions and more severe coding of contours and texture improves overall image quality significantly. More accurate reconstruction of smooth areas, however, can also be achieved by applying sophisticated decoding algorithms.

In this chapter, it was shown that the standard JPEG postprocessing method, which is based on prediction of low-frequency AC coefficients in smooth areas of the image, is only effective in a small compression range. By combining this method with adaptive low-pass filtering of the DC image, however, both segmentation and blocking artifacts can be reduced without introducing a high complexity in the decoder. By applying postprocessing in the transform domain, the postprocessing becomes more flexible and less complex. Furthermore, checking whether the postprocessing remains within the possible quantization errors does not require any inverse transformations.

The proposed method is only effective in large uniform regions. For images with a lot of detail, only the background is nicely smoothed by the DC filtering, and the overall image quality is only slightly improved. For images with pronounced smooth regions (e.g. the test scenes "lena" and "sails"), the method shows significant improvement in perceived image quality. Differences between subjects, of course, remain. Curves indicating the peak-signal-to-noise ratio (PSNR) of the JPEG-coded images showed that neither the AC prediction nor the DC filtering improved the PSNR of the reconstructed images. For

improving the perceived quality of image elements other than almost homogeneous regions, such as edges and texture, other postprocessing on the decoded transform coefficients is needed.

Although the DC filtering shown in this chapter was initially demonstrated for JPEG-coded images only, this technique can be applied in any transform block coder. In Chapters 3 and 4 it was shown that the Hermite transform is particularly effective at representing contours and edges. Since blocking and segmentation artifacts also occur in Hermite coded images, although only at very high compression ratios, the decoder option described can be applied to improve the coding efficiency of the Hermite transform in uniform regions. The effect of the adaptive DC filtering algorithm on Hermite-coded images was demonstrated in Section 5.5. One may conclude that by simple modeling, significant gain in image quality can be achieved for target regions and that block coding can also be very effective for low-bit-rate image compression.

Chapter 6

Epilogue

Using knowledge of visual perception is inevitable in the design of high perform-ance image processing algorithms. First of all, in conventional image compres-sion techniques based on (local) frequency representations of the image data, the role of perception is usually directed at the removal of information that can not be perceived by the human eye. Characteristics of the human visual system such as luminance masking, contrast sensitivity and contrast masking can be used to obtain optimal image quality for a given bit rate [102, 146, 123]. Second, the same characteristics of the visual system can be exploited in order to develop objective measures for the perceived quality of reconstructed images [85, 26]. This thesis uses a different philosophy by assuming that for high com-pression ratios, it is essential to focus on the representation of those structures in the image that contribute most to perceived image quality. With respect to the measurement of image quality, the use of subjective evaluation methods is proposed.

It is well known that the notion of image quality can not easily be tackled with objective quality measures [120]. In particular for low-bit-rate coding, measurement of perceived image quality still requires evaluation experiments in which subjects are asked for their opinion. Also in this thesis, a number of examples can be found that indicate that the objective peak-signal-to-noise ratio (PSNR) may deviate substantially from perceived image quality, even in ranges of moderate compression. Selecting a valid quality assessment method, however, is also not trivial. Experiments in **Chapter 2** indicated that prob-lems with a single stimulus or direct category scaling experiment may occur when reconstructed images of different coders are evaluated. Due to recon-struction artifacts of a very different nature, subjects who are asked to judge image quality may unconsciously judge the images at the level of individual artifacts, rather than at the level of overall image quality. This single stimulus method is one of the evaluation techniques recommended by the International Telecommunication Union (ITU). Other research [117] has indicated that this method is also susceptible to contextual effects due to stimulus spacing and frequency of occurrence of stimuli. The experiments in [117] demonstrate that

these contextual effects also occur in the double stimulus method using separate scaling rates for both images.

By using methods based on difference ratings, such as the method based on Anderson's functional measurement theory (FMT), both the contextual effects and the biasing effects due to different coding artifacts as described above, can be reduced significantly. In particular, quality-difference ratings ensure that subjects actually judge image quality, instead of the strength of individual artifacts. Although comparison experiments are relatively time-consuming, they should be clearly preferred. In order to limit the amount of comparisons, variants on existing experimental techniques, such as the FMT method used in this thesis, will have to be constructed (e.g., comparison methods that require only a small subset of all possible image pairs to be judged, or adaptive methods in which the sequence of displayed stimuli depends on previous judgements). Next to the set-up of experiments, processing and interpretation of the data that are obtained is equally important. For instance, by normalizing the data of the individual subjects, more reliable results are obtained. Also, more sophisticated processing can be applied (e.g. by using Thurstone's law of categorical judgement). In particular it was shown that metric multi-dimensional scaling methods provide a powerful tool for data analysis. For the development of image compression techniques, however, next to methods for image quality measurement in order to evaluate their performance, it would be of great interest to obtain more knowledge on how people actually experience image quality and how the quality of different kinds of image structures contributes to overall quality perception. Additional research, however, is needed to indicate the relative quality contributions of image structures such as texture, contours and almost homogeneous regions.

In this thesis, the representation of images is directed at the reconstruction of edge structures and uniform regions. For image analysis, a frequency-based decomposition technique is applied that uses overlapping Gaussian windows and projects images locally onto a basis of orthogonal polynomials. This technique, the Hermite transform, was originally developed in order to model the information analysis of the retinal and cortical receptive fields in the early visual system. With respect to image compression, the Hermite transform has been applied in hierarchical and scale-space algorithms [89, 91]. In particular, the Hermite transform has proved to be efficient in detection and extraction of image features [65].

Chapter 3 introduces a new form of the Hermite transform and shows that the $N + 1$ Hermite filters of order N provide all information necessary to calculate the responses of those filters in an arbitrary orientation. This so called steering property results in an orientation adaptive form of the Hermite transform that enables rotation of the Hermite filters in any direction. By steering the

Hermite filters locally along the direction of maximum energy, signal energy is compacted into a small number of frequency coefficients. In order to demonstrate the efficiency of the Hermite transform for image compression purposes, an image compression scheme that adapts to local image orientations is presented. By discarding the 2D energy components in this compression scheme, an efficient representation of images is obtained, based on (local) one-dimensional structures alone.

Comparisons with other compression techniques such as JPEG show that the proposed scheme performs very well at high compression ratios, not only in terms of peak-signal-to-noise ratio but also in terms of perceptual image quality. It was found that on the basis of a PSNR criterion, for most test images, JPEG should be preferred over a Hermite coder at compression ratios lower than 25. However, results of quality assessment experiments using subjects indicate that the perceptual quality of Hermite-coded images is much better than would be expected on the basis of the PSNR. An explanation for the high perceptual quality of the Hermite-coded images, is the difference in the nature of the artifacts. Due to overlapping orientation adaptive frequency patterns, highly annoying blocking artifacts are much less pronounced and edges are smoothly reconstructed. The evaluation of new image compression concepts does, of course, remain a delicate matter, partly due to the various components of which a compression algorithm consists. Especially the entropy coding of the remaining coefficients of an image representation my have a strong effect on the performance of the overall compression algorithm.

Despite the redundancies in the signal representation due to both overlapping Gaussian windows and highly overlapping frequency selectivities of the Hermite filters, it is rather surprising that the Hermite transform enables the construction of high performance image compression schemes. From a perceptual point of view, however, the steered Hermite transform results in less visible distortions near edges, because most of the quantization noise is aligned with the orientation of the edge and therefore masked by the edge [112]. Also, the steered Hermite transform is based on local orientation adaptivity, while other techniques map the entire image on subimages that are bandpass in orientation [74, 142, 130]. A disadvantage of the steered Hermite transform is a higher computational complexity due to local orientation adaptivity, thus more processing time.

In **Chapter 4** the orientation adaptive form of the Hermite transform is used as a basis for deriving an edge-parameter based representation of local edge elements. By using a Gaussian edge model, the edge parameters such as the edge orientation θ, the edge distance d, the edge height ΔV and the edge blur b can be calculated from the frequency coefficients of the steered Hermite transform. Edge descriptions in terms of edge parameters decrease the entropy of

the image representation. In natural images, however, replacing the frequency coefficients by the appropriate edge parameters does not necessarily have to improve the efficiency of the image representation. Since only a Gaussian edge model is used, it is impossible to capture all edge structures present in the image. By extending the modelling of image elements with, for example, line segments, corners and non-Gaussian edge elements, it should be possible to built image representations that can represent the original image more faithfully. For image compression purposes, however, such an approach requires the coding of additional classification information.

The derivation of edge primitives in Chapter 4 is limited to the representation of edge elements within a local window. Therefore, these local descriptions are integrated in a contour-based image representation technique by detecting chains of local edge elements. The impressions of the variation of edge parameters along image contours show that a more efficient coding can be obtained when exploiting the redundancies along these contours. In the rest of Chapter 4, image reconstruction is based on contour information supplemented with low-frequency DC information, and the proposed contour-based image representation is included in an image compression scheme. The performance of this compression scheme, however, is much less than the performance of the scheme discussed in Chapter 3. More time is needed to exploit cross-correlation between the different parameter bands and to construct a more efficient entropy coding scheme.

The results of the compression schemes presented in this thesis show that the Hermite transform is particularly effective at representing one-dimensional structures such as contours and edges. The representation of uniform regions, however, remains a problem. In block coding algorithms, smooth areas in the image are susceptible to annoying blocking and segmentation artifacts. In Chapter 2 this was already demonstrated by Watson's DCTune technique: at high compression ratios, more accurate coding of low frequency regions and less accurate coding of contours and texture improves overall image quality significantly. Accurate reconstruction of smooth areas, however, can also be achieved by applying sophisticated decoding algorithms [114, 42, 153].

Chapter 5 shows that the standard JPEG postprocessing method, which is based on prediction of low-frequency AC coefficients in smooth areas of the image, is only effective in a small compression range. By combining this method with a new adaptive low-pass filtering algorithm for restoring the DC image, however, both segmentation and blocking artifacts can be reduced without introducing unacceptable complexity in the decoder. By applying postprocessing in the transform domain, the postprocessing becomes more flexible and less complex. Furthermore, checking whether the postprocessing results remain within the possible quantization errors does not require any inverse transform-

ations.

Although the DC filtering proposed in Chapter 5 was originally developed for JPEG-coded images, this technique can, of course, be applied in any transform block coder. Since blocking and segmentation artifacts also occur in Hermite coded images, though only at high compression ratios, the decoder option proposed in this thesis can be applied to improve coding efficiency of the Hermite transform in uniform regions. The effect of adaptive DC filtering in Hermite-coded images is visualized in Chapter 5 for the "girls" image for two different parameter settings of the Hermite transform. The same DC filtering algorithm is used in Chapter 4 in order to reconstruct the low frequency parts of the image in combination with explicit coding of edges.

It is clear that with sophisticated decoding and/or postprocessing, perceived image quality can be significantly improved. The gain of postprocessing would probably be much higher, though, if one would integrate coder and decoder strategies. In principle, it is possible to include the decoding strategy in the coder [121]. In the specific case of the Hermite transform, improvement of the quality of other image structures than smooth regions, such as edges and texture, requires additional postprocessing on the decoded transform coefficients. Anyhow, the coding algorithms described in this thesis show that block coding can be very effective in low bit rate coding applications.

In this thesis, new concepts are considered in order to construct a second-generation image representation based on the Hermite transform. The steered Hermite transform was shown to result in an efficient compression scheme, especially at low bit rates. Nevertheless, other results in this thesis show that the contour-based compression algorithm based on the steered Hermite transform does not meet the high expectations. Therefore, more research effort will have to be directed at efficient entropy coding of the coefficients and parameters of this contour-based representation. Although this thesis has attempted to bridge the gap between transform coders and second-generation coder, the concepts of second-generation coding are not yet worked out sufficiently in order to be able to construct compression algorithms that outperform transform coders for high-quality image reconstruction. For the moment, one has to conclude that at moderate compression ratios conventional waveform compression should still be preferred. It is, however, important to realize that the coefficients and parameters of an image representation do not entirely determine the performance of an image compression system. This performance is determined by the combination of the image representation, the strategy for quantization and entropy coding and, of course, the decoding algorithm. For optimum image compression, these components cannot be designed independent of each other. With respect to image quality evaluation, this thesis stresses the importance of selecting an appropriate quality evaluation method. In particular for low-

bit-rate compression, the use of objective quality measures should be avoided. Hopefully the reader will benefit from the range of examples presented in this thesis.

Bibliography

[1] M. Abramowitz and I. Stegun, *Handbook of mathematical functions*, New York: Dover, 1965.

[2] A.J. Ahumada, Jr. and H.A. Peterson, "Luminance-model-based DCT quantization for color image compression," *SPIE Human Vision, Visual Processing, and Digital Display III*, vol. 1666, pp. 365-374, 1992.

[3] A.J. Ahumada and A.B. Watson, "A visual detection model for DCT quantization," *AIAA Version of 12/9/93*, pp. 1-5.

[4] A.N. Akansu and R.A. Haddad, *Multiresolution signal decomposition*, Academic Press, 1992.

[5] N.H. Anderson, "Algebraic models in perception," in: E.C Carterette and M.P. Friedman, ed., *Handbook of perception, Vol. II: Psychophysical Judgement and Measurement*, Academic Press, New York, 1974, pp. 216-298.

[6] M.J.J.C. Annegarn et al., "HD-MAC: een stap vooruit in de evolutie van de televisietechniek," *Philips Technisch Tijdschrift*, vol. 43, pp. 213-229, Dec. 1986.

[7] R. Aravind et al., "Image and video coding standards," *AT&T Technical Journal*, pp. 67-77, Jan./Feb. 1993.

[8] R.B Arps and T.K. Truong, "Comparison of international standards for lossless still image compression," *Proceedings of the IEEE*, vol. 82, no. 6, pp. 889-899, June 1994.

[9] R.J. Baddeley and P.J.B. Hancock, "A statistical analysis of natural images matches psychophysically derived orientation tuning curves," *Proc. R. Soc. Lond. B*, 246, pp. 219-223, 1991.

[10] D.H. Ballard and C.M. Brown, *Computer vision*, Prentice-Hall, 1982.

[11] C.S. Barreto and G.V. Mendonca "Enhanced zerotree wavelet transform image coding exploiting similarities inside subbands," *Proc. ICIP-96*, Vol. II, September 1994, pp. 549-551.

[12] P. Barten, "Evaluation of subjective image quality with the square-root integral method," *Journal of the Optical Society of America A*, vol 7, no. 10, pp. 2024-2031, 1990.

[13] S. Bech, R Hamberg et al., "The RaPID perceptual image description method (RaPID)," *Proc. of the SPIE Conference on Human Vision and Electronic Imaging* vol. SPIE-2657, pp. 317-328, 1996.

[14] P.J.L. van Beek, "Edge-based image representation and coding," Ph.D. thesis, Delft University of Technology, Dec. 1995.

[15] M. Benard and M. Kunt, "Linear prediction in directional images," *Proc. of the EUSIPCO-86*, The Hague, pp. 805-808, 2-5 Sept. 1986.

[16] J. Brandt and A.K. Jain, "A Medial Axis transform algorithm for compression and vectorization of document images," *Proc. ICASSP 89*, vol. 3, pp. 1850-1853, 1989.

[17] V. O'Brien, "Contour perception, Illusion and reality," *Journal of the Optical society of America*, vol. 48, no. 2, pp. 112-119, Feb. 1985.

[18] A.C. den Brinker and J.A. Roufs, "Evidence for a generalized Laguerre transform of temporal events by the visual system," *Biol. Cybern.*, vol. 67, pp. 395-402, 1992.

[19] P.J. Burt and E.H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. COM-31, pp. 532-540, 1983.

[20] S. Carlsson, "Sketch based coding of grey level images," *Signal Processing* 15, pp. 57-83, 1988.

[21] CCIR, "Recommendation 500-3: Method for the subjective assessment of the quality of television pictures," *Recommendations and Reports of the CCIR*, International Telecommunication Union, Geneva, 1986.

[22] C. Chatfield, *Statistics for technology*, third edt., Chapman and Hall, London, 1983.

[23] R.J. Clarke, *Transform coding of images*, Academic Press, London, 1985.

[24] S. Comes et al., "Post-processing of decoded pictures by filtering of the unmasked noise," *IEEE Trans. Image Processing*, submitted for publication, 1996.

[25] P.C. Cosman, R.M. Gray ans R.A. Olshen, "Evaluating quality of compressed medical images: SNR, subjective rating, and diagnostic accuracy," *Proceedings of the IEEE*, vol. 82, no. 6, pp. 919-932, June 1994.

[26] S. Daly, "The visible differences predictor: an algorithm for the assessment of image fidelity," in: A.B. Watson, ed., *Digital images and human vision*, Massachusetts Institute of Technology, Cambridge, Massachusetts, pp. 179-206, 1993.

[27] J.G Daugman, "Two-dimensional spectral analysis of cortical receptive field profiles," *Vision Research*, vol. 20, pp. 847-856, 1980.

[28] J.G. Daugman, "Six formal properties of two-dimensional anisotropic visual filters: structural principles and frequency/orientation selectivity," *IEEE Trans. Syst. Man Cybern.*, vol. SMC-13, no. 5, pp. 882-887, Sept./Oct. 1983.

[29] J.G. Daugman, "Six formal properties of two-dimensional anisotropic visual filters: structural principles and frequency/orientation selectivity," *IEEE Trans. Syst. Man Cybern.*, vol. SMC-13, no. 5, pp. 882-887, Sept./Oct. 1983.

[30] J.G. Daugman, "Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-36, pp. 1169-1179, July 1988.

[31] A.M. van Dijk, J.B. Martens and A.B. Watson, "Quality assessment of coded images using numerical category scaling," *Proc. of the European Symposium on Advanced Networks and Services*, Amsterdam, March 20-24, pp. 90-101, 1995.

[32] A.M. van Dijk, "Transform coding of images using local orientation adaptivity," *IPO Annual Progress Report*, vol. 30, pp. 57-65, 1995.

[33] A.M. van Dijk and J.B. Martens, "Feature-based image compression with steered Hermite transforms," *International Conference on Image Processing, ICIP-96*, Lausanne, September 16-19, vol. I, pp 205-208, 1996.

[34] A.M. van Dijk and J.B. Martens, Quality assessment of compressed images: A comparison between two methods," *International Conference on Image Processing, ICIP-96*, Lausanne, September 16-19, vol. II, pp 25-28, 1996.

[35] A.M. van Dijk and J.B. Martens, "Image representation and compression using steered Hermite transforms," *Signal Processing*, vol. 56, no. 1, pp. 1-16, 1997.

[36] A.M. van Dijk and J.B. Martens, "Subjective quality assessment of compressed Images," *Signal Processing*, vol. 58, no. 3, pp. 235-252 ,1997.

[37] A.M. van Dijk, "Reconstruction of transform coded images using DC filtering," *Signal Processing*, submitted for publication, 1997.

[38] M.P. Eckert and G. Buchsbaum, "Efficient coding of natural time varying images in the early visual system," *Phil. Trans. R. Soc. Lond. B*, pp. 385-395, 1993.

[39] A.W.M. van den Enden and N.A.M. Verhoeckx, *Digitale signaalbewerking*, Delta Press BV, 1987.

[40] B. Escalante Ramirez, *Perceptually-assessed digital processing of medical images*, Ph.D. thesis, Eindhoven University of Technology, Sept. 1992.

[41] B. Escalante Ramirez and J.B. Martens, "Noise reduction in computerized tomography images by means of polynomial transform," *Journal of Visual Communication and Image Representations*, vol. 3, no. 3, pp. 272-285, September 1992.

[42] Z. Fan and R. Eschbach, "JPEG decompression with reduced artifacts," *Proc. of the SPIE Conference on Image and Video Compression*, vol. SPIE-2186, pp. 50-55, 1994.

[43] L.M. Florack, B.M. ter Haar, J.J Koenderink et al., "Scale and the differential structure of images," *Image and vision computing*, vol. 10, no. 6, pp. 376-388, July/Aug.1992.

[44] R. Forchheimer and T. Kronander, "Image coding - from waveform to animation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-37, pp. 2008-2023, Dec. 1989.

[45] W.T. Freeman and E.H. Adelson, "Steerable filters," *OSA Topical Meeting on Image Understanding and Machine Vision*, vol. 14, pp. 144, Technical Digest series, June 1989.

[46] W.T. Freeman and E.H. Adelson, "The design and use of steerable filters," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. PAMI-13, no. 9, pp. 891-906, Sept. 1991.

[47] W.T. Freeman, *Steerable filters and local analysis of image structure*, Ph.D. Dissertation, June 1992.

[48] P.P. Gandhi, "JPEG-based image compression for low bit-rate coding," *Proc. of the SPIE Conference on Still Image Compression II*, vol. SPIE-2669, pp. 82-94, 1996.

[49] W.E. Glenn, "Digital image compression based on visual perception and scene properties," *SMPTE Journal*, pp. 392-397, May 1993.

[50] R.C Gonzalez and R.E. Woods, *Digital image processing*, Addison Wesley Publishing Company, 1992.

[51] D.N. Graham, "Image transmission by two-dimensional contour coding," *Proc. IEEE*, vol. 55, no. 3, pp. 336-346, March 1967.

[52] Y.H. Gu, "Adaptive multiscaled polynomial transform for image compression based on local image propery," *IPO report 931*, Sept. 1993.

[53] R. Hamberg and H. de Ridder, "Continuous assessment of perceptual image quality," *Journal of the Optical Society of America A*, vol. 12, No 12, pp. 2573-2577, December 1995.

[54] R.W. Hamming, *Digital filters*, Prentice-Hall, 1992.

[55] W.L. Hays, *Statistics*, fourth edt., Holt, Rinehart and Winston Inc., New York, 1988.

[56] R. Heusdens, *Overlapped transform coding of images: theory, realization and application*, Ph.D. thesis, Delft University of Technology, 1997.

[57] M.L. Hilton, B.J. Jawerth and A. Sengupta, "Compressing still and moving images with wavelets," *Multimedia systems*, vol. 2, no. 3, 1994.

[58] D.H. Hubel and T.N. Wiesel, "Brain mechanisms of vision," in: *The brain (Scientific American)*, pp. 84-96, 1979.

[59] D.H. Hubel, "Exploration of the primary visual cortex, 1955-78," *Nature*, vol. 299, pp. 515-522, Oct. 1982.

[60] A.K. Jain, "Image data compression: A Review," *Proc. IEEE*, vol. 69, pp. 349-391, March 1981.

[61] A.K. Jain, *Fundamentals of digital image processing*, Prentice Hall, Englewood Cliffs, N.J., 1989.

[62] S.A. Karunasekera and N.G. Kingsbury, "A distortion measure for blocking artifacts in images based on human visual sensitivity," *IEEE Trans. Image Processing*, vol. 4, no. 6, pp. 713-724, June 1995.

[63] J. Katto, K. Onda and Y. Yasuda, "Variable bit-rate coding based on human visual system," *Signal Processing: Image Communications*, vol. 3, pp. 313-320 Elsevier, 1991.

[64] V. Kayargadde and J.B. Martens, "Estimation of edge parameters and image blur using polynomial transforms," *CVGIP: Graphical models and image processing*, vol. 56, no. 6, pp. 442-461, Nov. 1994.

[65] V. Kayargadde, *Feature extraction for image quality prediction*, Ph.D. thesis, Eindhoven University of Technology, April 1995.

[66] V. Kayargadde and J.B. Martens, "Perceptual characterization of images degraded by blur and noise: experiments," *Journal of the Optical Society of America A*, vol. 13, no. 6, pp. 1167-1177, 1996.

[67] D. Kersten, "Predictability and redundancy of natural images," *J. Opt. Soc. Am. A*, vol.4, no. 12, pp. 2395-2400, Dec. 1987.

[68] J.J Koenderink, "The structure of images," *Biol. Cybern.*, vol. 50, pp. 363-370, 1984.

[69] J.J Koenderink and A.J van Doorn, "Representation of local geometrie in the visual system," *Biol. Cybern.*, vol. 55, pp. 367-375, 1987.

[70] J.J Koenderink, "Operational significance of receptive field assemblies," *Biol. Cybern.*, vol. 58, pp. 163-171, 1988.

[71] J.J. Koenderink and A.J. van Doorn, "Receptive field families," *Biological Cybernetics*, vol. 63, pp. 291-297, 1990.

[72] J.J. Koenderink and A.J. van Doorn, "Generic neighborhood operators," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. PAMI-14, pp. 597-605, June 1992.

[73] X. Kong and J. Goutsias, "A study of pyramidal techniques for image representation and compression," *Journal of visual communications and image representation*, vol. 5, no. 2, June, pp. 190-203, 1994.

[74] M. Kunt, A. Ikonomopoulos and M. Kocher, "Second-generation image-coding techniques," *Proc. IEEE*, vol.73, no. 4, pp. 549-574, Apr. 1985.

[75] M. Kunt, M. Benard abd R. Leonardi, "Recent results in high-compression image coding," *IEEE Trans. on circuits and systems*, vol. CAS-34, pp. 1306-1336, Nov. 1987.

[76] G.E Legge and J.M. Foley, "Contrast masking in human vision," *J. Opt. Soc. Am*, vol. 70, no. 12, pp. 1458-1470, Dec. 1980.

[77] J.O. Limb, "Distortion criteria of the human viewer," *IEEE Trans. Syst. Man Cybern.*, vol. SMC-9, pp. 778-793, Dec. 1979.

[78] B. Macq et al., "Image visual quality restoration by cancellation of the unmasked noise," *Proc. ICASSP 94*, vol. 5, pp. 53-56, 1994.

[79] S.G. Mallat, "Multifrequency channel decompositions of images and wavelet models," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-37, pp. 2091-2110, Dec. 1989.

[80] S.G. Mallat and S. Zhong, "Compact image coding from edges with wavelets," *Proc. ICASSP 91*, vol. 3, pp. 2745-2748, 1991.

[81] S.G. Mallat and W. Liang Hwang, "Singularity detection and processing with wavelets," *IEEE Trans. Inform. Theory*, vol. IT-38, no. 2, pp. 617-643, March 1992.

[82] S.G. Mallat and S. Zhong, "Characterization of signals from multiscale edges," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. PAMI-14, pp. 710-732, July 1992.

[83] J.L. Mannos and D.J. Sakrison, "The effects of a visual fidelity criterion on the encoding of images," *IEEE Trans. Inform. Theory*, vol. IT-20, no. 4, pp. 525-536, July 1974.

[84] H.S. Malvar and D.H. Staelin, "The LOT: transform coding without blocking effects," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-37, pp. 553-559, April 1989.

[85] H. Marmolin, "Subjective MSE measures," *IEEE Trans. Syst. Man Cybern.*, vol. SMC-16, no. 3, pp. 486-489, June 1986.

[86] D. Marr, *Vision, a computational investigation into the human representation and processing of visual information*, Freeman, San Francisco, 1982.

[87] J.B. Martens and G. M. Majoor, "Scale-space coding and its perceptual relevance," *IPO Annual Progress Report*, vol. 21, pp. 63-71, 1986.

[88] J.B. Martens, "Applications of polynomial transforms in image deblurring," *SPIE*, Sept. 1989.

[89] J.B. Martens, "Applications of polynomial transforms in image coding and computer vision," *Proc. of the SPIE Conference on Visual Communications and Image Processing IV*, vol. SPIE-1199, pp. 1279-1290, Nov. 1989.

[90] J.B. Martens, "The Hermite transform - Theory," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-38, no. 9, pp. 1595-1606, Sept. 1990.

[91] J.B. Martens, "The Hermite transform - Applications," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-38, no. 9, pp. 1607-1618, Sept. 1990.

[92] J.B. Martens, "Application of scale space to image coding," *IEEE Trans. Commun.*, COM-38, no. 9, pp. 1585-1591, Sept. 1990.

[93] J.B. Martens, "Deblurring digital images by means of polynomial transforms," *Computer Vision, Graphics, and Image Processing*, vol. 50, no. 2, pp. 157-176, May 1990.

[94] J.B. Martens, "Adaptive image processing by means of polynomial transforms," *Proc. of the SPIE*, Feb. 1992.

[95] J.B. Martens, "Local orientation analysis in images by means of the Hermite transform," *IEEE Trans. Image Processing*, accepted for publication, 1997.

[96] H.B. Mitchell, N Zilverberg and M Avraham, "A comparison of different block truncation coding algorithms for image compression," *Signal Processing: Image Communications*, vol. 6, pp. 77-82, 1994.

[97] J.L. Mitchell, W.B. Pennebaker et al., *MPEG video compression standard*, Chapman & Hall, New York, 1997.

[98] A. Montanvert, P. Meer and A. Rosenfeld, "Hierarchical image analysis using irregular tessellations," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. PAMI-13, pp. 307-316, April 1991.

[99] N.M. Nasrabadi and R. A. King, "Image coding using vector quantization: a review," *IEEE Trans. Commun.*, vol. COM-36, Aug. 1988.

[100] T.R. Natarajan and N. Ahmed, "On interframe transform coding," *IEEE Trans. Commun.*, vol. COM-25, no. 11, pp. 1323-1329, 1977.

[101] F.L. van Nes and M.A. Bouman, "Spatial modulation transfer in the human eye," *Journal of the Optical Society of America*, vol. 57, no. 3, pp. 401-406, March 1967.

[102] A.N. Netravali and B. Prasada, "Adaptive quantization of picture signals using spatial masking," *Proc. IEEE*, vol. 65, pp. 536-548, 1977.

[103] A.N. Netravali and B.G. Haskell, *Digital pictures*, Plenum Press, New York, 1988.

[104] J.P.Oakley, M.J. Cunningham and G. Little, "A fourier-domain formula for the least-squares projection of a function onto a repetitive basis in n-dimensional space," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-38, Jan. 1990.

[105] A.V. Oppenheim and R.W. Schafer, *Digital signal processing*, Prentice-Hall International, 1975.

[106] A. Parducci and L.F. Perrett, "Category rating scales: Effects of relative spacing and frequency of stimulus values," *Journal of Experimental Psychology Monograph*, vol. 89, pp. 427-452, 1971.

[107] D.E. Pearson, "Developments in model-based video coding," *Proceedings of the IEEE*, vol. 83, no 6, pp. 892-906, June 1995.

[108] W.B Pennebaker and J.L Mitchell, *JPEG still image data compression*, Van Nostrand Reinhold, New York, 1993.

[109] P.P. Gandhi, "JPEG-based image compression for low bit-rate coding," *Proc. of the SPIE Conference on Still-image Compression II*, vol. SPIE-2669, pp. San Jose, January 29-30, 1996.

[110] J.G. Proakis, *Digital communications*, Mc Graw-Hill Book Company, 1983.

[111] M. Rabbani and P.W. Jones, *Digital Image Compression Techniques, Tutorial Texts in Optical Engineering*, SPIE Optical Engineering Press, 1991.

[112] B. Ramamurthi and A. Gersho, "Nonlinear space-variant postprocessing of block coded images," *IEEE Trans. Acoust, Speech, Signal Processing*, vol. ASSP-34, no. 5, pp. 1258-1268, Oct. 1986.

[113] T.R. Reed et al, "Image sequence coding using concepts in visual perception," *Proc. of the SPIE Conference on Human Vision and Electronic Imaging: Models, Methods, and Applications*, vol. SPIE-1249, pp. 272-283, 1990.

[114] H.C. Reeve and J.S.Lim, "Reduction of blocking effects in image coding," *Optical Engineering*, vol. 23, no. 1, pp. 34-37, Jan./Feb. 1984.

[115] H. de Ridder and G.M. Majoor, "Subjective assessment of impairment in scale-space-coded images," *IPO Annual Progress Report*, vol. 23, pp. 55-64, 1988.

[116] H. de Ridder and G.M. Majoor, "Numerical category scaling: an efficient method for assessing digital image coding impairments," *Proc. of the SPIE Conference on Human Vision and Electronic Imaging: Models, Methods, and Applications*, vol. SPIE-1249, pp. 65-77, Santa Clara, 12-14 Feb. 1990.

[117] H. de Ridder, "Contextual effects in quality judgments," *Proc. of the Workshop on Quality Assessment in Speech, Audio and Image Communication*, pp. 56-61, Darmstadt, 11-13 March, 1996.

[118] J.A.J. Roufs and H. Bouma, "Towards linking perception research and image quality," *Proceedings of the SID*, vol. 21, no. 3, pp. 247-270, 1980.

[119] J.A.J. Roufs, "Brightness contrast and sharpness, interactive factors in perceptual image quality," *Proc. of the SPIE Conference on Human Vision, Visual Processing, and Digital Display*, vol. SPIE-1077, pp. 66-72, 1989.

[120] J.A.J. Roufs, "Perceptual image quality: concept and measurement," *Philips Journal of Research*, vol. 47, no. 1, pp. 35-62, January 1992.

[121] T.P. O'Rourke and R.L. Stevenson, "Vector quantization with distance contraints for enhanced post-processing," *Proc. of the SPIE Conference on Still-image Compression II*, vol. SPIE-2669, pp. 9-20, San Jose, January 29-30, 1996.

[122] A. Saadane, H. Senane and D. Barba, "Masking and quantization laws in a visual subband coding scheme," *IEEE*, 1994.

[123] R. Safranek and J. Johnston, "A perceptually tuned sub-band image coder with image dependent quantization and post-quantization data compression," Proc. ICASSP 89, vol. 3, pp. 1945-1948, 1989.

[124] R. Safranek, J. Johnston and R. Rosenholtz, "A perceptually tuned sub-band coder," *Proc. of the SPIE Conference on Human Vision and Electronic Imaging: Models, Methods and Applications*, vol. SPIE-1249, pp. 284-293, 1990.

[125] J.A. Saghri, P.S. Cheatham and A. Habibi, "Image quality measure based on a human visual system model," *Optical Engineering*, vol. 28, no. 7, pp. 813-818, July 1989.

[126] S.S. Schiffman, M.L. Reynolds and F.W. Young, *Introduction to multidimensional scaling: theory, methods, and applications*, Academic Press, New York, 1981.

[127] W.F. Schreiber, C.F. Knapp and N.D. Kay, "Synthetic highs - an experimental TV bandwidth reduction system," *J. SMPTE*, vol. 68, pp. 525-537, Aug. 1959.

[128] J.M. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Processing*, vol. 41, no. 12, pp. 3445-3462, Dec. 1993.

[129] G. Shulman and P. Mulvanny, "Discrimination of spatio-temporal patterns: the role of sustained and transient mechanisms," *Perception*, vol. 12, pp. 531-543, 1983.

[130] E.P. Simoncelli, W.T. Freeman, E.H. Adelson et al., "Shiftable multi-scale transforms," *IEEE Trans. Inform. Theory*, special issue on wavelets, 1992.

[131] R.L. Stevenson, "Reduction of coding artifacts in transform image coding," *Proc. ICASSP 1993*, vol. V, pp. 401-404, April 1993.

[132] D. Taubman and A. Zakhor, "Orientation adaptive subband coding of images," *IEEE transactions on Image Processing,* vol. 3, no. 4, July 1994.

[133] C. Teunissen, "The validity of CCIR quality indicators along a graphical scale," *SMPTE Journal*, vol. 105, no. 3, pp. 144-149, 1996.

[134] G.J. Tonge, "Time-sampled motion portrayal," *Second int. conf. on image processing and its applications,* vol. 24-26 June 1986.

[135] W.S. Torgerson, *Theory and methods of scaling*, John Wiley & Sons, New York, 1958.

[136] K. Tzou and T.R. Hsing, "Applications of physiological human visual system model to image data compression," *Application of Digital Image Processing VII,Proc.*, vol. 504, pp. 419-424, 1984.

[137] P.P. Vaidyanathan, "Quadrature mirror filter banks, m-band extensions and perfect-reconstruction techniques," *IEEE ASSP Magazine*, July 1987.

[138] M. Vetterli and C. Herley, "Wavelets and filter banks: theory and design," *IEEE Transactions on Signal Processing*, vol. 40, no. 9, Sept. 1992.

[139] G.J. Wallace, R. Vivian and H. Poulsen, "Subjective testing results for still picture compression algorithms for international standardization," *IEEE Global telecommunications conference*, pp. 1022-1027, Nov. 1988.

[140] G.G. Walter, *Wavelets and other orthogonal systems with applications*, CRC Press, 1994.

[141] G.J. Wallace, The JPEG still image compression standard, Comm. of the ACM, vol. 34, no. 4, pp. 31-44, April 1991.

[142] A.B. Watson, "The cortex transform: rapid compution of simulated neural images," *Computer Vision, Graphics, and Image Processing*, vol. 39, pp. 311-327, 1987.

[143] A.B. Watson, "Efficiency of a model human image code," *Journal of the Optical Society of America A*, vol. 4, no. 12, pp. 2401-2417, December 1987.

[144] A.B. Watson, "Perceptual-components architecture for digital video," *J. Opt. Soc. Am. A*, vol. 7, no. 10, pp. 1943-1949, 1990.

[145] A.B. Watson, "Visually optimal DCT quantization matrices for individual images," NASA Ames Research Center, 1992.

[146] A.B. Watson, "DCT quantization matrices visually optimized for individual images," *Human Vision, Visual Processing, and Digital Display IV*, Rogowitz ed. 1993 SPIE. Bellingham, WA.

[147] A.B. Watson, "Perceptual optimization of DCT color quantization matrices," *Proc. ICIP*, Vol. I, November 1994, pp. 100-104.

[148] W.M. Wells, "Efficient synthesis of gaussian filters by cascaded uniform filters," *IEEE Transa. Patt. Anal. Machine Intell.*, vol. PAMI-8, no. 2, pp. 234-239, March 1986.

[149] P.H. Westerink, *Subband Coding of Images*, Ph.D. thesis, Delft University of Technology, Oct. 1989.

[150] J.W. Woods and S.D. O'Neil, "Subband coding of images," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, no. 5, pp. 1278-1288, Oct. 1986.

[151] A.P. Witkin, "Scale space filtering: a new approach to multi-scale description," *Proc. Int. Joint Conf. Artificial Intell.*, 1983.

[152] Y. Yang and N.P. Galatsanos, "Iterative projection algorithms for removing the blocking artifacts of block-DCT compressed images," *Proc. ICASSP-93*, vol. V, pp. 405-408, April 1993.

[153] Y. Yang, N.P. Galatsanos and A.K. Katsaggelos, "Projection-based spatially adaptive reconstruction of block-transform compressed images," *IEEE Trans. Image Processing*, vol. IP-4, no. 7, pp. 896-908, July 1995.

[154] R. Young, "The Gaussian derivative model for spatial vision: I. Retinal mechanisms," *Spatial vision*, vol. 2, no. 4, pp. 273-293, 1987.

[155] R.A. Young, "Oh say, can you see? The physiology of vision," *Proc. of the SPIE Conference on Human Vision, Visual Processing, and Digital Display II*, vol. SPIE-1453, 1991.

[156] B. Zeng and A.N. Venetsanopoulos, "A JPEG-based interpolative image coding scheme," *Proc. ICASSP*, vol. V, pp. 393-396, April 1993.

[157] C. Zetzsche, E. Barth and B. Wegmann, "The importance of intrinsically two-dimensional image features in biological vision and picture coding," in: A.B. Watson, ed., *Digital images and human vision*, Massachusetts Institute of Technology, Cambridge, Massachusetts, pp. 109-138, 1993.

List of Abbreviations

1D	One-dimensional
2D	Two-dimensional
AC coefficient	Coefficient related to the alternating part of a signal
C	Compression ratio
CCIR	International institute for standardization (currently ITU)
DC coefficient	Coefficient related to the constant part of a signal
DCT	Discrete Cosine Transform
DPCM	Differential Pulse Code Modulation
EOB	End-of-Block
EOC	End-of-Chain
FMT	Functional Measurement Theory
HVS	Human Visual System
IPO	Institute for Perception Research
ISDN	Integrated Services Digital Network
ITU	International Telecommunication Union, formerly CCIR
JPEG	Joint Photographic Experts Group
KLT	Karhunen Loève Transform
LOT	Lapped Orthogonal Transform
MDS	Multi-Dimensional Scaling
MPEG	Motion Picture Experts Group
PCM	Pulse Code Modulation
POCS	Projections onto Convex Sets
PSNR	Peak-Signal-to-Noise Ratio
RMSE	Root-Mean-Squared Error

Summary

Multi-media applications in which multiple types of digital information such as sound, text and images are integrated make high demands upon technical systems. In particular the huge amount of data needed to represent images at relatively high image quality, may cause problems due to the limited storage and transmission capacity of these systems.

This thesis discusses new methods for reducing the number of bits needed to represent digital images. For the design of such compression methods, using our knowledge of visual perception is essential. When removing part of the image information, conventional image compression algorithms exploit, for instance, specific limitations of the visual system such as insensitivity for high spatial frequencies. At high compression ratios this approach results, however, in an unacceptable distortion of perceived image quality. In this thesis an alternative approach is followed, based on the assumption that optimum image quality for a given bit rate can only be obtained if the image representation is focused at the explicit description of those image structures which are of great importance for perceived image quality. Examples of perceptually important image structures are edges and uniform regions.

The image representation described in this thesis has been developed on the basis of the Hermite transform. After a discussion of the major characteristics of the Hermite transform with respect to image reconstruction, a new orientation-adaptive form of this transformation is introduced. This *steered Hermite transform* makes it possible to rotate the local frequency patterns that are used by the Hermite transform for image description, in an arbitrary orientation. This local rotation property makes that image descriptions with a steered Hermite transform are very efficient due to the compaction of signal energy into a small number of frequency components. In particular edge elements and other locally one-dimensional patterns are represented very efficiently. Application of the steered Hermite transform in a compression algorithm demonstrates that with this scheme good results can be obtained compared to well known reference coders such as JPEG and advanced wavelet coders.

With the steered Hermite transform it is also possible to describe edges locally in terms of edge primitives other than only the edge orientation. Also edge primitives such as position, height and sharpness can be calculated from the frequency coefficients of the orientation adaptive Hermite transform. Subsequently, by chaining local edge elements into contours, an image description is obtained that consists of explicit contour information, supplemented by DC information for the reconstruction of uniform regions. The integration of this contour-based representation in an image compression algorithm in which the redundancies between adjacent edge elements are exploited still requires additional research effort in order to improve the coder performance. For the time being it seems that, for high-quality image reconstruction, conventional transformation coders outperform so called *second-generation coders* that make use of explicit description of real-world image structures.

For obtaining acceptable image quality, however, reliable representation of edge information alone is insufficient. The reconstruction of almost homogeneous areas in the image is equally important. For so called block coders, a high compression ratio causes blocking structures and segmentation artifacts in the image. These artifacts are usually perceived as highly annoying. In this thesis a new, general applicable method is described that avoids these artifacts when decoding the image. The proposed method decreases both blocking and segmentation artifacts by adaptive low-pass filtering of the decoded DC coefficients. The effect is demonstrated with both a standard JPEG coder and a Hermite coder.

In this thesis, next to the representation and compression of images, attention is paid to methods for measuring image quality. Since objective measures such as the signal-to-noise ratio do not always correlate well with perceived image quality, it is necessary to measure image quality via experiments with subjects. Selecting the appropriate experimental method, however, is not always trivial. In this thesis, a number of methods based on numerical scaling are discussed that can be used to evaluate image quality of compressed images. Also, some techniques are presented that can be used to process experimentally obtained data. In particular, experiments in this thesis show that evaluation methods based on comparisons between coded images should be preferred over direct quality judgments.

Samenvatting

Multimedia-toepassingen waarin verschillende soorten informatie zoals geluid, tekst en beeld zijn geïntegreerd, stellen hoge eisen aan technische systemen. Door de beperkte opslag- en transmissiecapaciteit van deze systemen, kan vooral de grote hoeveelheid data die nodig is om beeldmateriaal met relatief goede kwaliteit te kunnen reproduceren tot problemen leiden.

In dit proefschrift worden nieuwe methoden beschreven om vergaande reductie mogelijk te maken van het aantal bits dat nodig is voor de representatie van beeldmateriaal. Bij het ontwerpen van dergelijke compressiemethoden speelt visuele perceptie een belangrijke rol. Conventionele algoritmen voor beeldcompressie maken bij het gedeeltelijk weglaten van beeldinformatie bijvoorbeeld gebruik van de beperkingen van het visuele systeem, zoals ongevoeligheid voor hoge spatiële frequenties. Bij hoge compressiefactoren leidt deze aanpak echter tot een onacceptabele aantasting van de beeldkwaliteit. In dit proefschrift wordt een alternatieve aanpak gevolgd, gebaseerd op het uitgangspunt dat optimale beeldkwaliteit bij een gegeven compressiefactor slechts dan bereikt kan worden indien de beeldrepresentatie gericht is op het expliciet weergeven van die structuren in het beeld die voor de perceptieve kwaliteit van beelden het belangrijkst zijn. Voorbeelden van perceptief belangrijke beeldstructuren zijn randen en vlakke gebieden.

De in dit proefschrift beschreven beeldrepresentatie is ontwikkeld op basis van de Hermite transformatie. Na bespreking van de voor beeldreconstructie belangrijkste eigenschappen van de Hermite transformatie, wordt een nieuwe, oriëntatie-adaptieve vorm van deze transformatie geïntroduceerd. Deze *steered Hermite transform* maakt het mogelijk om de lokale frequentiepatronen die door de Hermite transformatie worden gebruikt om een beeld te beschrijven, te roteren in een willekeurige richting. Juist deze lokale rotatie-eigenschap maakt dat beeldbeschrijvingen met een oriëntatie-adaptieve Hermite transformatie zeer efficiënt zijn door de compactie van signaalenergie in een klein aantal frequentiecomponenten. Met name randelementen en andere lokaal eendimensionale patronen worden zeer efficiënt weergegeven. Toepassing van de steered Hermite transform in een compressiealgoritme laat zien dat dit schema

goede resultaten behaalt ten opzichte van bekende referentiecoders zoals JPEG en geavanceerde wavelet coders.

Met de zojuist genoemde oriëntatie-adaptieve Hermite transformatie is het ook mogelijk om randen lokaal te beschrijven in termen van randprimitieven zoals randpositie, hoogte, oriëntatie en scherpte. Deze primitieven worden berekend uit de frequentiecoëfficiënten van de oriëntatie-adaptieve Hermite transformatie. Door het aan elkaar koppelen van lokale randelementen wordt vervolgens een beeldbeschrijving verkregen bestaande uit een expliciete representatie van contouren, aangevuld met DC-informatie om de reconstructie van tussenliggende uniforme gebieden mogelijk te maken. De integratie van deze op contouren gebaseerde beschrijving in een algoritme voor beeldcompressie, waarbij gebruik wordt gemaakt van redundantie tussen naast elkaar gelegen randelementen op een contour, kan echter nog verbeterd worden. Voorlopig laat het zich dus aanzien dat voor compressie met behoud van hoge beeldkwaliteit conventionele transformatiecoders een beter resultaat halen dan zgn. *second-generation coders* waarbij beeldstructuren op expliciete wijze worden beschreven.

Het goed weergeven van randinformatie is echter niet genoeg om een goede beeldkwaliteit te kunnen bereiken. Het is ook belangrijk om vrijwel vlakke gebieden in het beeld met hoge kwaliteit te kunnen reconstrueren. Bij zgn. *block coders* onstaan bij een hoge compressie echter blokkerige structuren in het beeld, die als bijzonder storend worden ervaren. In dit proefschrift wordt een nieuwe, algemeen toepasbare methode beschreven om deze blokstructuren bij het reconstrueren van het gecodeerde beeld te voorkomen. De gepresenteerde methode bereikt een vermindering van blok- en segmentatieartefacten door adaptieve low-pass filtering van de gedecodeerde DC coëfficiënten. Het effect wordt gedemonstreerd aan de hand van zowel een JPEG coder als een Hermite coder.

Naast de representatie en compressie van beelden wordt in dit proefschrift aandacht besteed aan methoden voor het meten van beeldkwaliteit. Omdat objectieve maten voor beeldkwaliteit zoals de signaal-ruis-verhouding niet altijd correleren met perceptieve beeldkwaliteit, is het noodzakelijk om beeldkwaliteit te meten aan de hand van experimenten met proefpersonen. Het selecteren van de juiste experimentele methode is echter niet altijd eenvoudig. In dit proefschrift worden een aantal op numerieke schaling gebaseerde methoden beschreven die gebruikt kunnen worden om beeldkwaliteit van gecodeerde beelden te evalueren. Ook worden diverse technieken aangegeven die gebruikt kunnen worden om experimenteel verkregen data te verwerken. In het bijzonder wordt aangetoond dat evaluatiemethoden gebaseerd op vergelijkingen tussen gecodeerde beelden de voorkeur genieten boven directe kwaliteitsbeoordelingen.

Curriculum Vitae

Toon van Dijk werd op 23 september 1968 geboren in Wijchen, alwaar hij op vierjarige leeftijd zijn onderwijscarrière begon op kleuterschool 'Tienakker'. Na de kleuterschoolperiode bezocht hij de Mariaschool om zes jaar lang lager onderwijs te volgen. In 1980 ging hij naar het Dukenburg College te Nijmegen en haalde het Atheneum-β diploma dat hem toegang verschafte tot de Katholieke Universiteit te Nijmegen. Aan de universiteit studeerde hij Informatica met als specialisatierichting Telecommunicatie en digitale signaal- bewerking, wat uiteindelijk leidde tot een afstudeerstage bij het Dr. Neher Laboratorium van KPN Research in Leidschendam. Na in 1992 cum laude te zijn afgestudeerd, aanvaardde hij een promotieplaats op het Instituut voor Perceptie Onderzoek (IPO) te Eindhoven, waar hij promoveerde op het toe- passen van nieuwe beeldrepresentatietechnieken in algoritmen voor compressie van digitale beelden. Sinds oktober 1996 is hij in dienst bij de afdeling Research & Development van Océ-Technologies B.V. in Venlo. Zijn huidige interesse richt zich op ontwikkeling van informatie- en kennissystemen en user- systeem communicatie. Of hij met zijn promotie in 1997 na 25 jaar zijn carrière in het onderwijs definitief heeft beëindigd, zal de toekomst uitwijzen ...

Venlo, 4 augustus 1997.

Stellingen

1. Leven in een informatie-intensieve wereld vereist niet zozeer de kunde om relevante informatie *op te nemen* danwel de kunde om niet-relevante informatie te *negeren*.

2. Hoewel geavanceerde user-interfaces het voor steeds meer mensen mogelijk maken om computers te gebruiken, vergroten zij tegelijkertijd de afstand tussen een gebruiker en de computer.

3. Het gegeven dat het bij promoties naast wetenschappelijke glamour ook om financiële belangen gaat, maakt promoveren tot een ultiem voorbeeld van *showbusiness*.

4. Het huidige AiO-stelsel en de aanstelling van beurspromovendi aan Nederlandse universiteiten hebben de discussie over het in stand houden van het minimumloon tot een zinloze gemaakt.

5. Door de voortschrijdende techniek op het gebied van transmissie- en opslagmedia en de schijnbaar onbeperkte behoefte aan informatie, lijken methoden voor datacompressie eerder aangewend te worden om produkten sneller op de markt te brengen dan om capaciteit zo zuinig mogelijk te benutten.

6. De ontwikkeling van een perfect model om het verloop van beurskoersen te voorspellen zal tot stabilisering van het koersverloop leiden, aangezien koersveranderingen juist voortkomen uit verschillende opvattingen over toekomstige ontwikkelingen.

7. Het voornemen om bij sollicitaties een verbod op leeftijdsdiscriminatie in te stellen is niet in het voordeel van de sollicitant, zolang dit verbod niet gepaard gaat met een algemene maatschappelijke aanvaarding van de juistheid van een dergelijke maatregel.

8. Door verschuivingen in vraag en aanbod heeft de aanschaf van duurzaam geproduceerd hardhout eenzelfde vernietigend effect op het voortbestaan van tropische regenwouden als het aanschaffen van niet-duurzaam geproduceerd hardhout.

9. De door de *Linux smart quotes* gegeven wetten m.b.t. het uitvoeren van experimenten,

    (a) If reproducibility may be a problem, conduct the test only once

    (b) If a straight line fit is required, obtain only two data points,

    wijzen nog eens op de oplettendheid die de lezer in acht moet nemen bij het bestuderen van experimentele resultaten.