科技资料

Visual Communications and Image Processing'91: Visual Communication

Part 1

PROCEEDINGS
SPIE—The International Society for Optical Engineering

Visual Communication and Image Processing '91: Visual Communication

Kou-Hu Tzou
Toshio Koga
Chairs/Editors

11–13 November 1991
Boston, Massachusetts

Sponsored and Published by
SPIE—The International Society for Optical Engineering

Cooperating Organizations
IEEE Circuits and Systems Society
IEEE Acoustics, Speech, and Signal Processing Society
European Association for Signal Processing (EURASIP)

SPIE P PROCEEDINGS SERIES

Volume 1605
Part One of Two Parts

SPIE (Society of Photo-Optical Instrumentation Engineers) is a nonprofit society dedicated to the advancement of optical and optoelectronic applied science and technology.

The papers appearing in this book comprise the proceedings of the meeting mentioned on the cover and title page. They reflect the authors' opinions and are published as presented and without change, in the interests of timely dissemination. Their inclusion in this publication does not necessarily constitute endorsement by the editors or by SPIE.

Please use the following format to cite material from this book:
Author(s), "Title of paper," *Visual Communications and Image Processing '91: Visual Communication*, Kou-Hu Tzou, Toshio Koga, Editors, Proc. SPIE 1605, page numbers (1991).

Library of Congress Catalog Card No. 91-641811
ISBN 0-8194-0742-9

Published by
SPIE—The International Society for Optical Engineering
P.O. Box 10, Bellingham, Washington 98227-0010 USA
Telephone 206/676-3290 (Pacific Time) • Fax 206/647-1445

Printed in the United States of America.

Volume 1605

CONFERENCE COMMITTEE

Conference Chairs

Kou-Hu Tzou, Bell Communications Research
Toshio Koga, NEC Corporation (Japan)

Technical Committee

Ali N. Akansu, New Jersey Institute of Technology
Dimitris Anastassiou, Columbia University
Rashid Ansari, Bell Communications Research
Jan Biemond, Delft University of Technology (Netherlands)
Defu Cai, Institute of Electronics (China)
Allen Gersho, University of California/Santa Barbara
Madan M. Gupta, University of Saskatchewan (Canada)
Barry G. Haskell, AT&T Bell Laboratories
T. R. Hsing, Bell Communications Research
F. Jutand, ENST (France)
Aggelos K. Katsaggelos, Northwestern University
Jae-Kyoon Kim, Korea Advanced Institute of Science and Technology (South Korea)
Hiroshi Kotera, NTT Human Interface Laboratories (Japan)
Murat Kunt, Swiss Federal Institute of Technology (Switzerland)
Petros Maragos, Harvard University
Yoshitaka Morikawa, Okayama University (Japan)
Nasser M. Nasrabadi, Worcester Polytechnic Institute
King N. Ngan, National University of Singapore (Singapore)
Takashi Okagaki, University of Minnesota Medical School
William A. Pearlman, Rensselaer Polytechnic Institute
Soo-Chang Pei, National Taiwan University (Taiwan)
Eliezer Peli, Eye Research Institute
Peter Pirsch, Universität Hannover (FRG)
K. R. Rao, University of Texas/Arlington
Michael M. Skolnick, Rensselaer Polytechnic Institute
Fang-Kuo Sun, The Analytic Sciences Corporation
Ming-Ting Sun, Bell Communications Research
Andrew G. Tescher, Lockheed Palo Alto Research Laboratory
Lance T. Wu, Computer and Communications Laboratory, ITRI (Taiwan)
Yasuhiko Yasuda, University of Tokyo (Japan)
Yehoshua Y. Zeevi, Technion—Israel Institute of Technology (Israel) and Rutgers University
Tian G. Zhuang, Shanghai Jiao Tong University (China)

(continued)

VISUAL COMMUNICATIONS AND IMAGE PROCESSING '91:
VISUAL COMMUNICATION

Volume 1605

Session Chairs

Session 1A—Video Sequence Coding I
Barry G. Haskell, AT&T Bell Laboratories

Session 2A—Hierarchical Image Decomposition
Rashid Ansari, Bell Communications Research

Session 3A—Vector Quantization
Allen Gersho, University of California/Santa Barbara

Session 4A—Model-Based Image Coding
Nasser M. Nasrabadi, Worcester Polytechnic Institute

Session 5A—Video Sequence Coding II
Mitsuharu Yano, NEC Corporation (Japan)

Session 6A—Superhigh-Definition Image Systems
Sadayasu Ono, NTT Transmission Systems Laboratories (Japan)

Session 7A—Still Image Coding
Jan Biemond, Delft University of Technology (Netherlands)

Session 8A—Motion Estimation and Motion Analysis
Aggelos K. Katsaggelos, Northwestern University

Session 9A—Image Transmission and Communication Systems
K. R. Rao, University of Texas/Arlington

Session 10A—3-D Motion Analysis
Dimitris Anastassiou, Columbia University

Session 11A—Hierarchical Image Coding
John W. Woods, Rensselaer Polytechnic Institute

Session 12A—Entropy Coding
Cheng-Tie Chen, Bell Communications Research

Session 13A—Visual Communication Hardware
Cheng-Tie Chen, Bell Communications Research

VISUAL COMMUNICATIONS AND IMAGE PROCESSING '91:
VISUAL COMMUNICATION

Volume 1605

INTRODUCTION

Since its inception in 1986, the Visual Communications and Image Processing conference has grown steadily: this year the total number of submissions reached nearly 260. Members of the review committee spent several weeks coordinating a review of the abstracts submitted and accepted 190, separating them into three parallel sessions. Due to the large number of accepted papers and the limited time slots, some papers that were scheduled as oral presentations were assigned to poster sessions.

The papers are divided into two volumes, according to subject, it is hoped that such division can make future reference more convenient. Volume 1605, Visual Communications, addresses video sequence coding, hierarchical image decomposition, vector quantization, model-based image coding, superhigh-definition image systems, still image coding, image transmission and communication systems, motion estimation and motion analysis, 3-D motion analysis, hierarchical image coding, entropy coding, and visual communication hardware. Volume 1606, Image Processing, contains papers on image analysis, morphology and fractals, pattern recognition, motion perception and moving target detection, image restoration and filtering, human visual system models, image segmentation and classification, digital image processing in medicine, image sequence restoration and filtering, digital image processing algorithms, VLSI implementation and hardware architectures, applications of digital image processing, and neural networks in image processing. Since there is no clear line between visual communication and image processing, we were confronted by the question of where to assign certain topics. In such situations, we tried to do the assignment by majority in that session.

Special thanks are due to the program committee and session organizers for their efforts in soliciting papers of topical interest and high quality. We would also like to extend our appreciation to the SPIE staff for taking care of all the details associated with the conference.

Kou-Hu Tzou
Bell Communications Research

Toshio Koga
NEC Corporation (Japan)

VISUAL COMMUNICATIONS AND IMAGE PROCESSING '91:
VISUAL COMMUNICATION

Volume 1605

CONTENTS

VISUAL COMMUNICATIONS AND IMAGE PROCESSING '91:
VISUAL COMMUNICATION

Volume 1605

(continued)

VISUAL COMMUNICATIONS AND IMAGE PROCESSING '91:
VISUAL COMMUNICATION

Volume 1605

ADDITIONAL PAPERS

SESSION 1A

Video Sequence Coding I

Chair
Barry G. Haskell
AT&T Bell Laboratories

A Digital Video Codec for Medium Bitrate Transmission

Touradj Ebrahimi, Frédéric Dufaux,
Iole Moccagatta, T. George Campbell,
Murat Kunt
Laboratoire de Traitement des Signaux
EPFL-Ecublens (DE)
CH-1015 Lausanne, Switzerland

Abstract

A digital video codec is presented, using a fast Gabor-like wavelet transform to produce a multiresolution data set. Three different strategies are introduced to code different levels of the resolution in the pyramidal data, according to their visual importance. A hierarchical vector quantization is performed to exploit the inter/intra subbands correlations. A multiresolution block matching is proposed to generate the motion field in the scene. These motion vectors are then used to reduce the temporal redundancies. Simulations show good results in quality for medium bitrate transmission.

1 Introduction

This paper describes a digital video compression system suitable for medium range bitrate transmission. The system currently works on CIF sequences (common intermediate format), but an extension to other formats is straightforward. The block diagram of the coder is reported in Figure 1.
A Gabor-like wavelet transform with an efficient implementation is performed at the first stage in both intraframe and interframe modes. The use of the Gabor-like decomposition is motivated because of the optimal joint localization of these functions. This property is of importance in the design of any subband decomposition scheme. On the other hand, according to recent experiments, the majority of the receptive field profiles of the mammalian visual system can be fit quite well by this type of function. The wavelet used in this system is then perceptually derived and more suitable for visual data compression. Moreover a partitioning of the spatial-frequency domain into octave bands is achieved which is motivated by typical image statistics and also by the sensitivity of the human visual system. Finally the multiresolution (pyramidal) structure of the data after the transformation lets us to process the different levels of the information with different techniques, according to their importance. Thus, the DC component of the data (the lowest resolution) is encoded using PCM. The middle levels of resolution in the pyramidal data are coded using a hierarchical tree structured vector quantization. The highest spatial frequencies (highest resolutions) are scalar quantized and coded using an adaptive arithmetic coder. The system is designed currently to work in an ATM (asynchronous transfer mode) environment. But as it can be seen from the block diagram, a classical buffer control strategy can be used to produce constant bitrate at the output of the coder. The buffer control strategy is not discussed in this paper.
To eliminate the interframe redundancies between frames, a motion compensated differential interframe technique is used. The motion field is obtained using a hierarchical block matching which takes into account the multiresolution structure of the data. The motion vectors in the field can be sent or not, depending on the amount of the motion present in the scene. Thus, a motion/no-motion mode is defined. When the codec is working at low bitrates or when the amount of the motion in the scene is not large, no motion information is sent and the decoder computes the motion field from the two previous frames. This procedure is specially efficient in the case of bitrates close to the videoconferencing bitrate. The interframe coding alternates between one predictive and one or several interpolative modes. In order to avoid accumulation of different errors, a fully intraframe technique is applied within a fixed interval to update all the coefficients completely. This mechanism is also reset after a scene change.
Section 2 discusses the fast wavelet transform with optimal localization performed in the codec. Three different strategies for quantization and bit assignment of the multiresolution data are detailed in sections 3 and 4. The

 0-8194-0742-9/91/$4.00

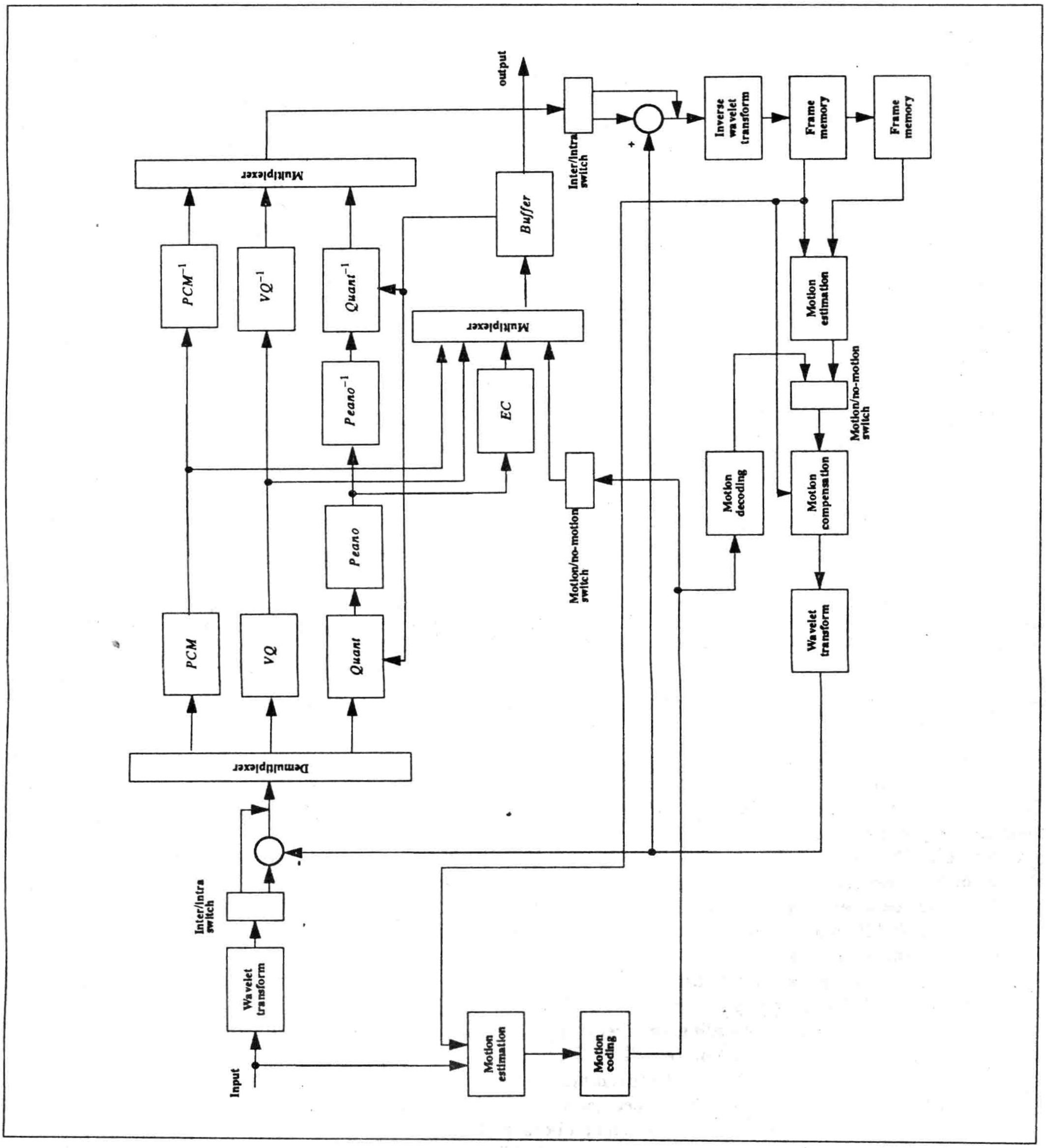

Figure 1: The block diagram of the codec.

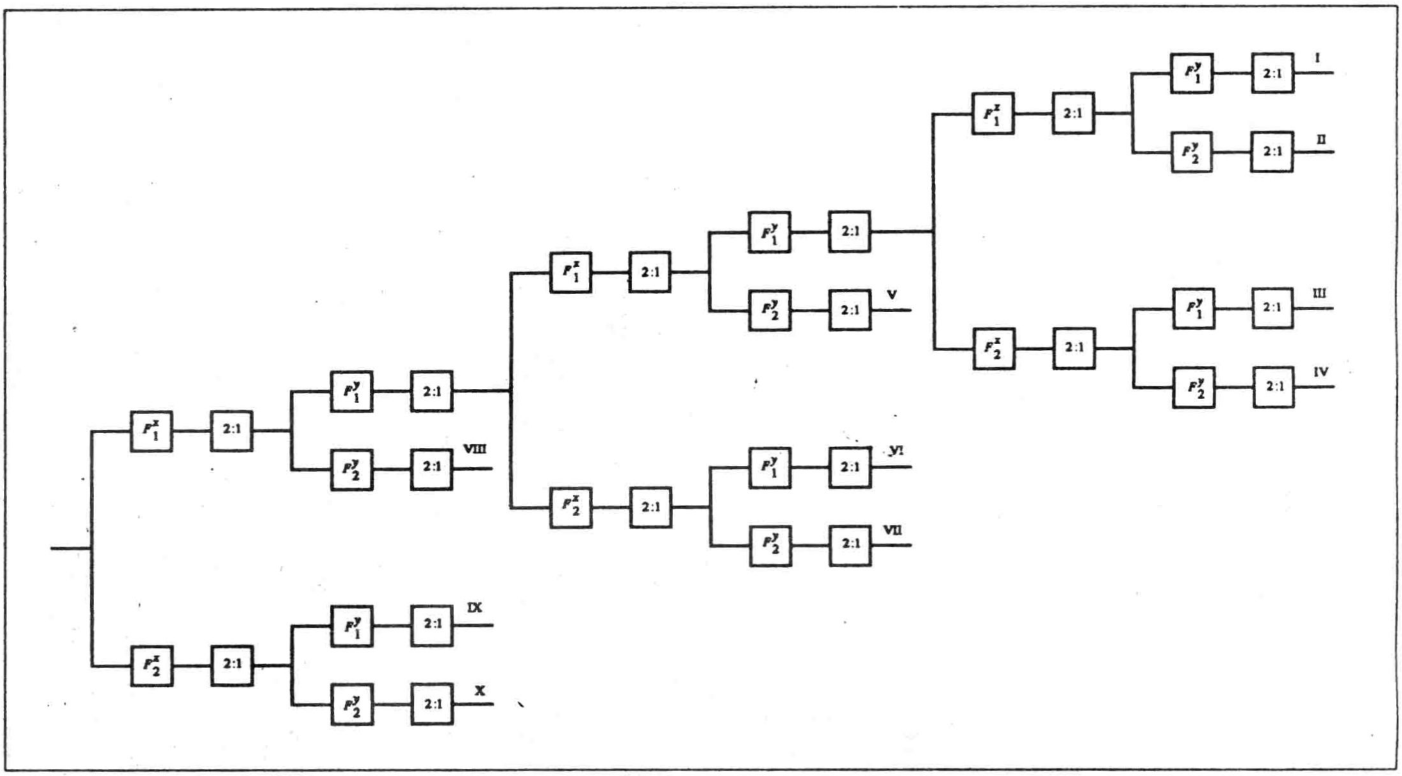

Figure 2: Tree structure recursive filtering to perform the fast Gabor-like wavelet transform.

technique to reduce the temporal redundancies is developed in section 5. In section 6, the motion field estimation is explained. Finally, experimental results are reported in section 7.

2 Gabor-like wavelet with fast implementation

The video codec system presented in this paper performs a Gabor-like wavelet transform by recursive filtering [1]. The basic structure is a 1-D two channel frequency decomposition. The high frequency channel is obtained from the low frequency channel filter by a π shift. The low frequency filter in its turn is an approximation of a gaussian where the frequency response at points π and $-\pi$ have been set to zero. This operation makes the high frequency filter's response have a zero at DC. By applying the basic two channel decomposition in a recursive way, all the frequency channels, except the low pass one, will have a zero at DC (see Figure 2). Moreover, the synthesis filters are approximated by a filter having coefficients in powers-of-two. The analysis filters are obtained from the relation between the analysis and the synthesis filters for an exact reconstruction, and approximated by filters having coefficients in power-of-two. Figure 3 gives the analysis and the synthesis filters frequency response, for real and power-of-two coefficients.
The use of the Gabor-like wavelet transform is motivated on the one hand by the fact that Gabor functions, which are gaussians modulated by complex exponentials, have optimal joint localization in the spatial/spatial-frequency domain. On the other hand, according to recent experiments, the majority of the receptive field profiles of the mammalian visual system can be fit quite well by this type of functions [2].
A number of remarks seem necessary to be given here. The complexity of the synthesis filters is less than those of the analysis filters. This makes these filters become very suitable for broadcasting applications. Both synthesis and analysis filters are very short in the time domain and both of them contain only coefficients in powers-of-two. This is a further simplification in order to make these filters suitable for video systems with medium bit rates. The relation between the low pass and high pass filters allows a polyphase implementation of these filters.
It is also important to notice that, because of the approximations, the reconstruction obtained by these filters is not perfect. However in all our experiments the signal to noise ratio of the reconstructed images has been more than $46dB$. This is in a large number of applications a good enough approximation. For other applications in which a

Filter	Coefficients Value									
$g_0(\cdot)$	2^{-7}	2^{-3}	2^{0}	2^{0}	2^{-3}	2^{-7}				
$g_1(\cdot)$	-2^{-7}	2^{-3}	-2^{0}	2^{0}	-2^{-3}	2^{-7}				
$f_0(\cdot)$	2^{-6}	0	-2^{-3}	-2^{-7}	2^{0}	2^{0}	-2^{-7}	-2^{-3}	0	2^{-6}
$f_1(\cdot)$	-2^{-6}	0	2^{-3}	-2^{-7}	-2^{0}	2^{0}	2^{-7}	-2^{-3}	0	2^{-6}

Table 1: The value of coefficients in the synthesis and the analysis filters.

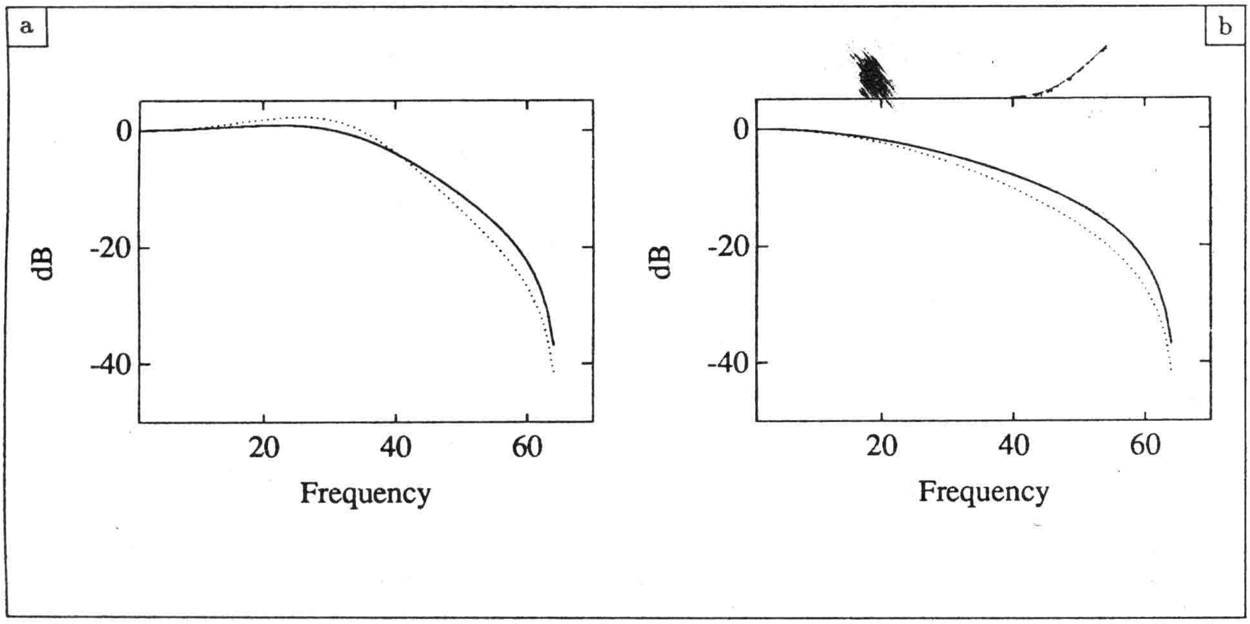

Figure 3: Filters characteristics in the frequency domain. The continuous lines are the filters with coefficients in power-of-two and the dotted lines are the ones with real values a) analysis filters b) synthesis filters.

higher quality is required, the error of reconstruction can be coded and transmitted through a separate channel.
The tree structure based on a two channel decomposition is a further simplification in implementation, since the same filters with the same coefficients are used on different levels.
Figure 4 shows an original image, its transformed coefficients and the reconstructed image, obtained from these coefficients. The scaled error of reconstruction has also been reported in the same figure. As it can be seen from this figure, the error is larger in bright areas than in dark areas. This property is in accordance with the human visual system.
The Table 1 gives the exact values of the coefficients in the analysis and the synthesis filters

3 Quantization

In this section we will present the coding strategy which performs an efficient quantization of the data produced by the transformation stage of the coder. Three different coding philosophies take into account the different properties of the multiresolution data. The lowest resolution level in the pyramidal data (DC component) is the most important and the most sensitive to distortions. In addition the amount of information relative to this part is small as compared to the other resolution levels. Because of these reasons a PCM technique is applied to transmit this part of the data. The intermediate levels in the multiresolution data are encoded using a vector quantization (VQ) with a vector tailoring capable to exploit the cross-level correlation inside the pyramidal data. Because of the importance and the originality of this approach, an entire section is devoted to it. Simulation results applying this pyramidal coding scheme to the entire multiresolution data are reported. The highest resolution levels (high frequency subbands) are uniformly scalar quantized and scanned along a Peano curve, before being processed by an adaptive arithmetic coder.