DIXIEME COLLOQUE SUR LE TRAITEMENT DU
SIGNAL ET SES APPLICATIONS

NICE du 20 au 24 MAI 1985

Vision Adapted Transform Coding of the Colour Information of Natural Images

Uwe Franke

Institut für Elektrische Nachrichtentechnik, RWTH AACHEN, D-5100 AACHEN

RESUME

Cet exposé présente un codeur de chrominance avec un grand taux de compression adapté à la perception visuelle humaine. Le système de codage, appliqué sur les signaux couleurs de l'espace YIQ est basé sur la transformation discrète de Cosine de dimensions 8 x 8 points d'images.
Des seuils de visibilité des fonctions de base 8 x 8 DCT ont été évalués pour les 2 composants couleurs. L'adaption se manifeste dans l'utilisation de ces seuils dans la sélection et quantification des coefficients à transmettre.
Une nouvelle méthode de classification, utilisée pour informer le décodeur sur les coefficients transmis est déscrite. Cette méthode est basée sur la probabilité d'occurence des combinations des coefficients de DCT situés au-delà de leurs seuils correspondants.
Ensemble avec un schéma de prédiction bidimensional pour les coefficients dc exploitant les grandes corrélations statistiques dans des blocs adjacents, le codeur de chrominance se suffit de moins de 0.1 bit/point d'images pour coder toute l'information couleur. Malgré ce grand taux de compression, aucune dégradation visible de la qualité d'image a pu être enregistrée.

SUMMARY

A chrominance coder with very high compression ratio is presented, which is adapted to the visual perception of the human eye. The coding system, working on the chrominance components of the YIQ-colour space, is based on a discrete cosine transform (DCT) with block dimension 8 x 8 pels.
Visibility thresholds of the 8 x 8 DCT basis functions have been determined for both components. Adaptivity is achieved by application of these thresholds to the selection of the coefficients to be transmitted as well as to their quantization.
A new classification method used to inform the decoder which coefficients are transmitted is described. This method is based on the probability of occuring combinations of DCT coefficients above their corresponding thresholds.
In conjunction with a two-dimensional prediction scheme for the dc-coefficients exploiting the high statistical dependencies between adjacent subblocks, the vision adapted chrominance coder requires less than 0.1 bit per pel to code the relevant colour information. In spite of this very low bit rate no apparent degradation of the image quality is visible.

Vision Adapted Transform Coding of the Colour Information of Natural Images

I INTRODUCTION

The transmission of colour images makes high data compression ratios desirable to ensure short transmission times even on narrowband channels. Many coding schemes are known from literature, most of which are designed for luminance signals. This paper describes a coder which has specially been designed for chrominance signals.

Transform coding has proved to be a highly efficient method to code grey level pictures. An adaptation of the coder to the local image detail enables image transmission with very low bit rates.
Beside zonal coding schemes, where the parameters of the coder vary with respect to some statistical properties of the subblocks /1/, threshold coding schemes are being studied increasingly. Such coder transmit only those coefficients with amplitudes above a given threshold. The selection of appropriate thresholds is important for the coder design. CHEN /2/ applies an identical threshold to all coefficients, only depending on the globally desired bit rate. Much better adapted to the human vision are visibility thresholds (VTs), which have been introduced by LOHSCHELLER /3/ for a luminance coder. In this paper, this concept is extended do the encoding of chrominance signals.

Compatibility to existing luminance coders is achieved by a linear colour space conversion. Fig. 1 shows the corresponding system configuration.

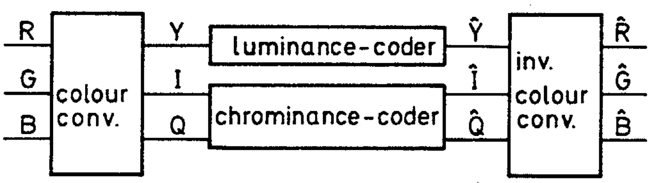

Fig. 1: luminance compatible system configuration

The coder uses the chrominance signals I and Q corresponding to the NTSC standard, which can be obtained from the R,G,B signals by the linear conversion

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & -0.322 \\ 0.211 & -0.523 & 0.312 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

(please notice, that the matrix element a(3,2) has the value -0.523 and not -0.253, as erroneously written in many articles).
The selected colour space is superior to the YUV colour representation of the PAL or SECAM television systems for two reasons:
- first, the conversion to YIQ yields a higher energy compaction between the two chrominance components,
- second, the orientations of the axes I and Q are better adapted to the human spectral perception /4/. As a result, the signal Q can be transmitted with a much smaller spectral resolution than the signals U and V, yielding better results when irrelevance reducing coding schemes are used.

II VISION ADAPTED CODER

The presented coder uses the discrete cosine transform (DCT), which has a superior energy compaction property /5/ in the class of transforms with known fast computational algorithms /6,7/.

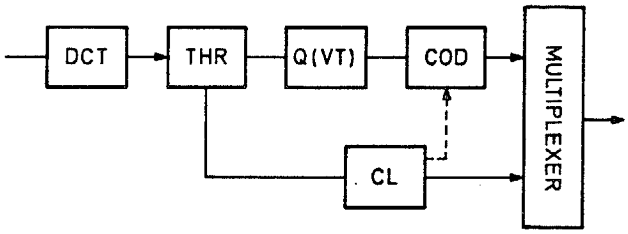

Fig. 2: block diagram of the Vision Adapted Coder

Fig. 2 shows the block diagram of the coder. Subsequent to the DCT with block dimension 8 x 8 pels the coefficients undergo a threshold operation. The supra-threshold coefficients are then quantized. Based on the results of the thresholding a classifier determines which coefficients are to be transmitted; the coefficients are coded depending on this classification. The necessary overhead information has to be sent in advance of the coefficients of the particular subblock.

In the following sections, the different coder components will be decribed in detail.

A. Visibility Thresholds

Within the scope of coder design the selection of the thresholds is of special importance because the resulting bit rate as well as the achievable image quality depends on the chosen thresholds.
The consequence of using an identical threshold for all coefficients is the transmission of coefficients which do not improve picture quality. The aim therefore is to transmit only those coefficients which represent an optical stimulus lying above the visibility threshold of the human eye.

Assuming a perception model of the human vision with independent spectral sensitivity as proposed by SAKRISON /8/ also for the chrominance, the VTs can be determined independently. This was done under viewing conditions according to the CCIR Rec. 500-2 using a uniform grey background of mean luminance.
The VTs for 16 relevant coefficients of the component I and 8 coefficients of the component Q have been measured in the following way: Inside a bounded area on the monitor one of the basis pictures was presented with slowly temporally increasing intensity until it became visible for the test persons.
The visibility thresholds averaged on 20 persons are displayed in Fig. 3 (k,l are indices for the matrix representation of the DCT coefficients).
The increase of the VTs with growing local frequency of the basis functions corresponds to the lowpass characteristic of the human eye for the chrominance signals.

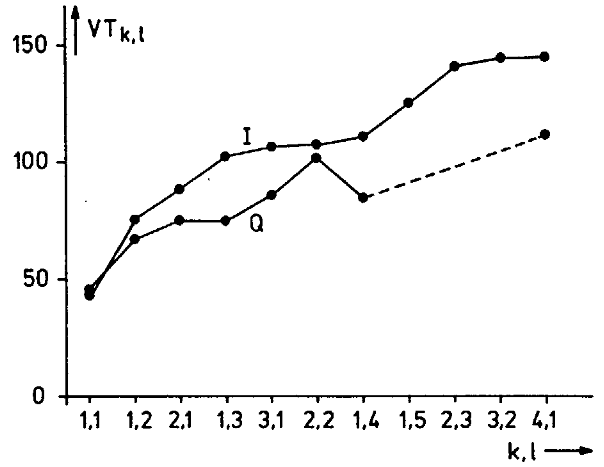

Fig. 3: averaged visibility thresholds

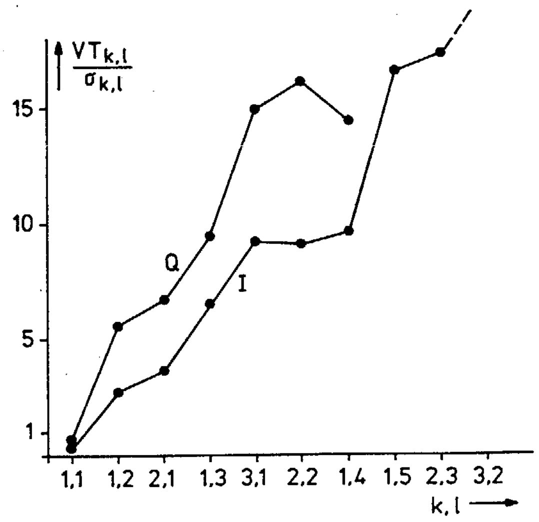

Fig. 4: normalized visibility thresholds

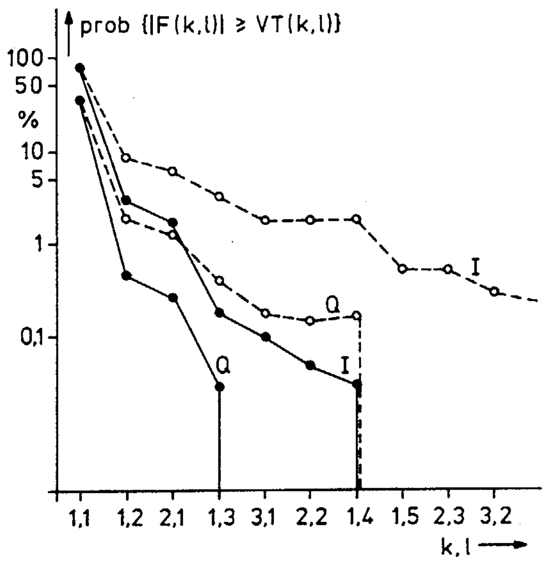

Fig 5: probability of particular coefficients lying above their VTs

While the values of the VTs increase, the spectral variances of the corresponding coefficients decrease strongly, leading to normalized VTs as shown in Fig. 4.
Because of the highly peak shaped density functions of the coefficients (see section B.) only a small fraction of the coefficients have to be transmitted, as can be drawn from Fig. 5 (based on approx. 50000 transform blocks). The appliance of an identical threshold - here the VT of the first coefficient - to all coefficients, as for example proposed by CHEN, would yield results represented by the broken lines.

The advantage of using vision adapted thresholds is evident. The determined TVs are the foundation for all further investigations.

B. Quantization

The data reduction of transform coding systems is based on the energy compaction into only a few coefficients, which is exploited by more or less rough quantization. Usually the necessary quantizers are optimized for minimal mean square error (MSE).
In preliminary investigations, the distribution densities of the coefficients have been measured. They can be modelled by the generalized exponential function given by

$$p(\xi,\sigma,\nu) = \alpha(\sigma,\nu)\cdot\exp\{-[\beta(\sigma,\nu)\cdot|\xi|]^{\nu}\} \quad (2)$$

The shape of this function depends on the standard deviation σ and the shape-factor ν. The investigations revealed, that the dc-coefficients are approximately Laplacian distributed (ν=1.0), while the distribution density of the interesting ac-coefficients can be well modelled with a shape-factor of 0.5. Fig. 6 shows the corresponding density function for normalized variance and the boundaries of MSE-quantizers according to LLOYD-MAX /9,10/ for 4 and 3 bit, respectively.

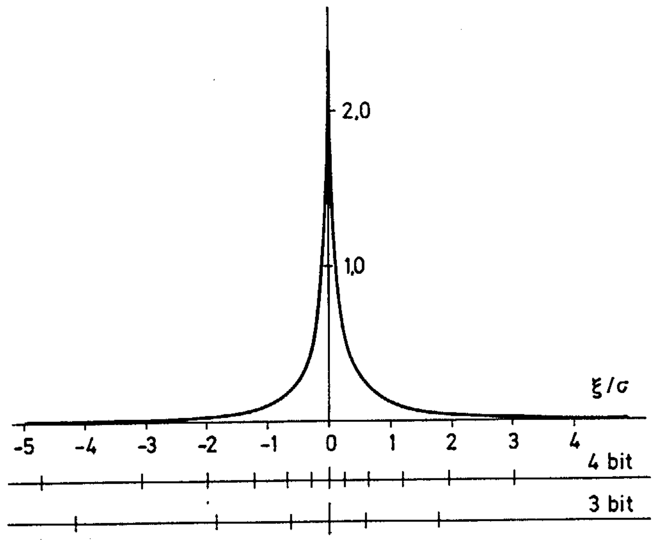

Fig. 6: modelled distribution density function of the ac-coefficients and the boundaries of MAX-quantizer /10/

The MAX-quantizers show two disadvantages: First, high saturated colours are reproduced unsatisfactoryly because the MAX-quantizers either quantize too rough or are in their saturation due to the small variance of the dc-coefficient of the component Q (highly saturated green colours cannot be reproduced). Second, the values of the VTs increase with growing local frequency of the basis function, while the variance of the corresponding coefficients decrease. The result is (see Fig. 4 and 6) that many of the representation levels of the quantizer lie in the range [-VT,VT] and would not be used by the threshold coder described here (similar circumstances arise in the presence of luminance signals, too).
It is evident, that statistically optimized quantizer characteristics are not suited for the quantization of chrominance spectral coefficients if high subjective image quality and low bit rates are desired. At the same time, this means that the MSE is not a proper error criterion at least for the design of a chrominance coding system. Lacking an appropriate vision adapted error criterion, the coder design has to be done according to the results of subjective tests.

The problems that have been described can be solved by a modified linear quantization rule given by

$$F_q(k,l) = \begin{cases} \mathrm{NINT}[F(k,l) \,/\, VT(k,l)] & \text{if } |F(k,l)| > VT(k,l) \\ 0 & \text{else} \end{cases} \quad (3)$$

Notice, that the interval widths of the quantizers are controlled by the VTs representing the human spectral sensitivity. Within a realization in hardware one quantizer can be multiplexed for all ac-coefficients as a further advantage.

The investigation by subjective tests revealed, that no quantization errors are visible when the following bit assignment is used:

component I

5	4	3	3	0	0	0	0
4	3	0	0	0	0	0	0
3	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

component Q

4	2	0	0	0	0	0	0
2	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

The chosen bit assignment guarantees that all supra-threshold coefficients occuring in the available 15 test images are quantized without any quantizer overload.

Because of their very low probability of occurance (compare Fig. 5) an additional entropy coding of the quantized ac-coefficients would yield only a negligible further data reduction.

C. Classification

The designed threshold coder based on the measured visibility thresholds transmits only those coefficients with amplitudes above the corresponding VTs. The receiver has to be informed which coefficients are transmitted for every particular subblock; within the chrominance coding scheme investigated here the values of all coefficients often lie below their VTs so that no coefficients at all have to be transmitted, as can be drawn from table I.
The aim is to minimize the mean number of necessary overhead bits.

Preliminary studies revealed, that only 3 coefficients are necessary to code the chrominance component Q without loss of image quality. This means that 8 combinations of coefficients lying above their VTs are possible with an entropy of approximately 1 bit/block. The appliance of a Huffman coding scheme requires nearly 1.4 bit/block on an average to code the class information of the component Q.
The method proposed here enables quantizers without zero representation levels for this component.

As can be drawn from Fig. 5, maximal 7 particular coefficients of the component I lie above the corresponding VTs. That means that 256 different supra-threshold coefficient patterns (STCPs) are possible. To avoid large lookup tables, a new classification scheme is introduced here.

The determination of the probability of occurance of particular STCPs yields interesting results which are shown in table I.

pattern	h(I)	h(Q)
no coefficient	27.0%	64.4%
only first coeff.	68.4%	34.9%
coeff. 1,1 + 1,2	2.4%	0.3%
coeff. 1,1 + 2,1	1.4%	0.2%
	99.2%	99.8%

Table I: probability of occurance of particular STCPs

This distribution of the probability of occurance enables an efficient classification and coding scheme.

The approach is to define N classes given by their STCPs. Each subblock has to be classified to one of these classes according to the following rules:

1. To achieve the required reconstruction quality, for each subblock the selection of the class has to ensure that all supra-threshold coefficients are contained in the chosen class.
2. To achieve a minimal data rate, the selected class should contain as few coefficients as possible which lie under their VTs.

Beside the selection of the best classes the number N of classes to be defined must be determined.

Fig. 7 shows the total number of data bits and overhead bits necessary to code the component I as a function of the number of used classes N assuming the bit assignment from section B. The upper curve results from a fixed length coding of the class information, the lower one from using a variable length coding scheme (Huffman-code). It can be seen that no significant gain can be achieved by using more than 5 classes. The necessary 5.27 bit/block include 1.38 bit for the class information.

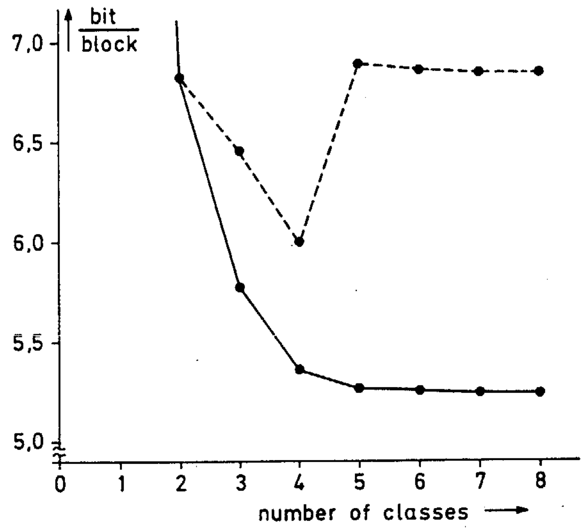

Fig. 7: total number of necessary bits per block as function of the number of classes

A Huffman coding of all 256 possible patterns would require 5.24 bit/block on an average. It appears that by application of a Huffman coding scheme with all its known disadvantages practically no reduction of the data rate is achieved.
The optimal class division using 5 classes is given in the following table II.

class	coefficients	code
1	1,1	O
2	-	IO
3	1,1 1,2	IIO
4	1,1 2,1	IIIO
5	all 7	IIII

Table II optimal class division

D. Predictive coding of the dc-coefficients

The total number of bits required for the image transmission is composed of the data bits and overhead bits. Decisive for the number of necessary data bits is the representation of the dc-coefficients, since in most cases only these coefficients - if any at all - have to be transmitted (see Fig. 5). The quantized dc-coefficients of adjacent subblocks are still highly correlated with horizontal and vertical normalized covariance coefficients of about 0.9.

This statistical dependencies can be exploited for a further significant data reduction by one of the following two possibilities:

1. a second transform coding of a set of 4, 16 or more adjacent dc-coefficients
2. application of a predictive coding scheme.

Besides of a good utilization of the existing redundancy, the algorithm to be used should enable an error free reconstruction of the dc-coefficients. Since the first method only results in high compression ratios if reconstruction errors of the dc-coefficients are tolerable, here a predictive coding scheme is investigated, which in addition can be realized much easier in hardware.

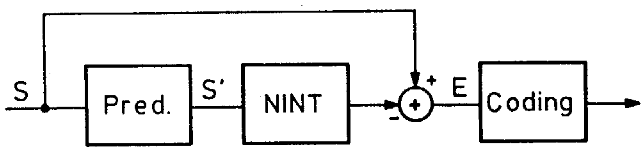

Fig. 8: block diagram of predictive coder

Fig. 8 shows the block diagram of the two-dimensional linear prediction scheme studied here. The prediction value S' of the actual dc-coefficient undergoes a rounding operation (NINT). This operation is necessary to enable an efficient coding of the prediction error E. Using an error free coding a feed-back of the quantized prediction error is avoided.

The prediction value can be obtained by

$$S' = \sum_{i=1}^{3} a_i \cdot x_i \qquad (4)$$

with the (fixed) predictor coefficients arranged in the following way:

$$\begin{matrix} a_1 & a_2 \\ a_3 & x \end{matrix} .$$

Under the conditions

$$\sum_{i=1}^{3} a_i = 1 \text{ and } a_i = \frac{n_i}{8}, \; n_i \in \{-8,8\} \qquad (5)$$

the values of the predictor coefficients have been optimized to achieve a minimal mean bit rate. The used procedure ensures, that the optimal predictor coefficients depending on the chosen code are found for every given code! Using a Huffman code for the (integer) prediction error E the mean bit rates for the dc-coefficients can be reduced to 2.0 bit/block and 1.4 bit/block for the components I and Q, respectively. Compared with 5 bit/block (I) and 4 bit/block (Q) if no prediction is applied the gain is remarkable.

An additional reduction of the necessary mean number of bits is possible since the decoder knows a priori from the received class information which blocks are 'colourless'; i.e. the dc-coefficients lie below their VTs and do not have to be transmitted. This is valid for 27% and 65% of the blocks of the chrominance components I and Q, respectively.

As can be drawn from table III, the mean number of bits are reduced to 1.55 bit/block and 0.6 bit/block for the chrominance components. Correctely, this values have to be compared with the number of bits necessary using the class information without prediction.

without class inf.	I	Q
without predicton	5 bit	4 bit
with predicton	2 bit	1.4 bit
with class inf.		
without predicton	3.75 bit	1.4 bit
with predicton	1.55 bit	0.6 bit

Table III: mean number of bits necessary for the coding of the dc-coefficients

In conclusion, the exploitation of the decoder's prescience due to the received class information enables a considerable improvement of the coding performance. Using a Huffman coding scheme and the predictor coefficients optimized for this code, which can be taken from table IV, the transmission of the dc-coefficients of both chrominance components requires only about 2.15 bit/block on an average.

	a_1	a_2	a_3
component I	-2/8	6/8	4/8
component Q	-0/8	4/8	4/8

Table IV: optimized predictor coefficients (Huffman code)

Fig. 9 shows the histogram of the prediction error E obtained by using the predictor coefficients listed above.

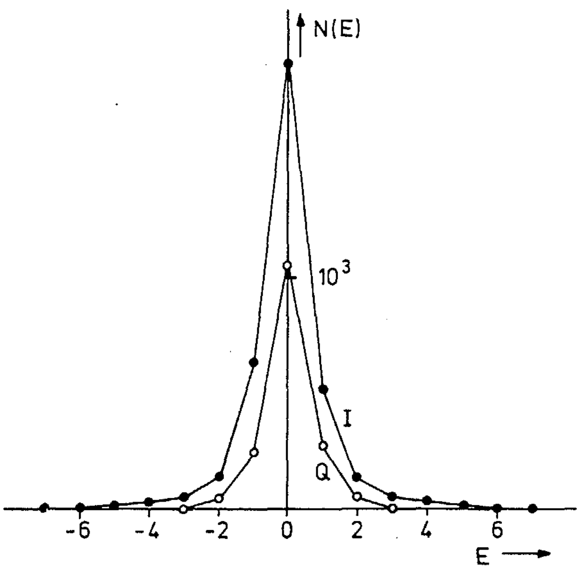

Fig. 9: histogram of prediction error E

III SUMMARY

A high efficient chrominance coder adapted to the human visual perception has been introduced. The coder is based on the DCT with block dimension 8 x 8 pels. Subjectively determined visibility thresholds of the DCT basis functions are used to select the coefficients to be transmitted. These coefficients are quantized by means of a modified linear quantizer controlled by the visibility thresholds. A new classification scheme based on the probability of occuring suprathreshold coefficient patterns is used to inform the decoder which coefficients are sent. To achieve a very low bit rate on the one hand a two-dimensional prediction scheme is applied to the dc-coefficients of the chrominance components I and Q and on the other hand the decoder's prescience about colourless blocks is exploited which is obtained from the class information received in advance.

Computer simulations confirmed the good performance of the vision adapted chrominance coder. It reveals that 3.1 bit/block and 2.0 bit/block are required on an average to code the chrominance components I and Q, respectively. This results in a total mean bit rate of only 0,08 bit/pel for the whole relevant colour information.

Acknowledgement

I wish to thank the ANT Nachrichtentechnik GmbH, Backnang, West-Germany, for their financial support given to this research.

Literatur

/1/ PRATT,W.K.: Digital Image Processing, John Wiley & SONS, New York, 1978

/2/ CHEN,W.-H.; PRATT,W.K.: Scene Adaptive Coder, IEEE Trans. on commun., Vol. COM-32, No. 3, March 1984, pp. 225

/3/ LOHSCHELLER,H.: Vision Adapted Progressive Image Transmission, Second European Signal Processing Conference, EUSIPCO 83

/4/ MIDDLETON,K.; HOLMES,M.C.: The Apparent Colors of Surfaces of Small Subtense - A Preliminary Report, Jou. of the opt. Soc. of America, Vol. 39, July 1949, pp. 582

/5/ CHEN.W.-H.: Adaptive Coding of Monochrome and Color Images, IEEE Trans of commun., Vol. COM-25, November 1977, pp. 1285

/6/ CHEN,W.-H.: A Fast Computation Algorithms for the DCT, IEEE Trans. on commun., Vol. COM-25, September 1977, pp. 1004

/7/ KAMANGAR,F.A.; RAO,K.R.: Fast Algorithms for the 2-D Discrete Cosine Transform, IEEE Trans. on computers, Vol. C-31, No. 9, September 1982, pp. 899

/8/ SAKRISON,D.J.: On the role of the observer and a distortion measure in image transmission, IEEE Trans. on commun., COM-25, pp. 1251

/9/ MAX,J.: Quantizing for minimum distortion, IRE Trans. IT-6 (1960), pp. 7-12

/10/ GOETZE, M.: Kombinierte Quellen- und Kanalcodierung in adaptiven Transformationscodierungssystemen, Ph.D. thesis, Techn. Univ. of Aachen, W.-Germany, 1985