PREMIER COLLOQUE IMAGE

Traitement, Synthèse, Technologie et Applications

BIARRITZ – Mai 1984 –

CRITÈRES VISUELS POUR CODAGE D'IMAGE EFFICACE

ON VISUAL CRITERIA IN EFFICIENT PICTURE CODING

Prof.Dr.ing. ZORAN S. BOJKOVIĆ

University of Belgrade, Faculty of Transport and Traffic Engineering

RESUME

Cet article expose le problème des critères visueles pour codage d'image efficace. La fonction de visibilité est analysée sous les limites suivantes:(a) les statistiques des signaux d'images sont nonstationnaires et (b) la sensibilité spatiale pour le bruit de quantification est appliquée. Dans la deuxième partie de l'article, modulation par impulsions codées differentielles d'images MICD et la technique de codage d'images par transformation, sous les critères visuels, sont presentées séparément.

SUMMARY

This paper treats the problem of visual criteria in efficient picture coding. The visibility function is analysed under the following constraints:(a) statistics of picture signals are nonstationary and (b) the spatial sensitivity to quantization noise is used. In the second part of the paper, differential pulse code modulation DPCM and transform coding technique for picture transmission under visual criteria are treated separately.

ON VISUAL CRITERIA IN EFFICIENT PICTURE CODING

Critères Visuels pour Codage d'Image Efficace

Prof.Dr.ing. ZORAN S. BOJKOVIĆ

1. INTRODUCTION

Differential pulse code modulation DPCM as well as transform coding have received attention in the contemporary literature and have been used in digital picture transmission. In order to adapt the quantization strategies for coding the picture, masking properties of the human observer for adapting the quantization strategies are used [1]. The quantization noise weighting function is called the visibility function. This function gives the relationship between the amplitude accuracy and measure of luminance spatial activity.

Differential pulse code modulation is the most basic of the efficient image codings for TV signals. This coding is simple in its scheme and has advantages in picture quality since the coding noises occur mostly in the edges of pictures. Visually, the effect of quantization errors in DPCM is to cause local degradations in areas of large slopes or edges in the picture.

Improvement in visual appearance could be made by designing the quantizer which attaches weight to quantization errors according to visibility rather than probability of a given prediction error [2].

Compared to the DPCM, a major attribute of transform picture coding is that this transform compacts the image energy to a few of the transform domain samples. The transform picture coding systems distribute the coding degradation in a manner less objectionable to a human viewer and show less sensitivity to picture variation and to channel noise [3].

If the fidelity of the signal reproduction is evaluated in terms of mean square errors, then the characteristics of these coding methods may be obtained and their optimum design determined. A distortion measure for the quantization process is some statistic of the quantization error [4].

This paper treats the problem of visual criteria in efficient picture coding. Taking into account that the statistics of picture signals are nonstationary and using the spatial sensitivity to quantization noise, the visibility function is analyzed firstly. After that, differential pulse code modulation and transform coding for picture transmission under visual criteria are treated separately.

2. VISIBILITY FUNCTION

The required fidelity of picture reproduction demanded by the human eye varies from picture element to picture element. Consequently, for efficient digital representation of pictures it is desirable to adapt coding strategies to those local properties of the picture signal which determine the visual sensitivity to quantizazion noise.

Visibility functions measure the relative visibility of noise added to a picture at those points where some measure of local activity exceeds a given treshold. The functions vary with the content of the picture. Visibility functions are used to design quantizing characteristics for coding of monochrome and colour picture signals. The visibility function consists of two parts: a picture dependent component and viewer dependent component which is called masking function. Spatial masking is defined as the reduction in the ability of a person to visually discriminate amplitude errors which occur at or in the neighbourhood of significant spatial changes in the luminance.

In the picture element domain, the masking function at a pel is defined as the weighted sum of the luminance slopes at the pel under consideration and at the neighbouring pels. For example, at point (i,j) the two-dimensional masking function $M_{i,j}$, using a 3 x 3 neighbourhood of slopes is given by

$$M_{i,j} = \sum_{n=i-1}^{i+1} \sum_{t=j-1}^{j+1} \alpha^{[(n,t)-(i,j)]} \cdot \frac{1}{2} \left[|m^H_{n,t}| + |m^V_{n,t}| \right] \quad (1)$$

where $[(n,t) - (i,j)]$ is the Euclidean distance between points (n,t) and (i,j) normalized by the distance between horizontaly adjacent picture elements. On the other hand, $m^H_{n,t}$, $m^V_{n,t}$ are the horizontal and vertical slopes of the image intensity at point (n,t) while α is a constant, i.e $\alpha = 0{,}35$ [1].

The visibility of quantization noise added to a picture element as a function of masking function at that pel is shown in Fig. 1. It can be seen that the visibility

ON VISUAL CRITERIA IN EFFICIENT PICTURE CODING
Critères Visuels pour Codage d'Image Efficace

Prof. Dr. ing. ZORAN S. BOJKOVIĆ

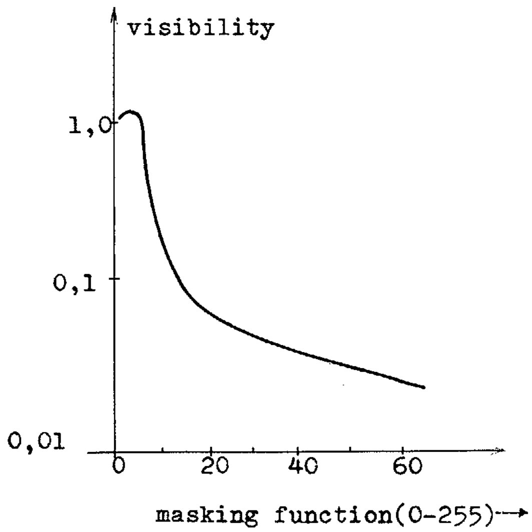

Figure 1.Visibility of quantization noise and masking function

dicrease, implying that at higher values of spatial detail the visibility of quantization noise is lower. The visibility at a point is proportional to the power of noise and the proportionality constant is given by the value of the visibility function evaluated using the spatial detail at that picture element. Thus, visibility functions, generally speaking, allow to judge the visibility of the quantization noise . Moreover, the visibility functions are used to change dinamically the input-output mapping of a single quantizer to reduce the bit rates. This is done by reassigning the input of the quantizer to a different representative level than normal in such a way to reduce the entropy of the quantized output, while keeping the visibility quantization noise below a certain treshold.

A visibility function of a particular picture is measured by determing the visual sensitivity to distortions added at points where the horizontal slope of the picture exceeds a treshold. The measurement is made by comparing the visibility of the added noise to the visibility of reference noise added to the picture. Let the visibility distribution function be V(x), while v(x) denotes visibility density function.v(x) is a subjective measure of the visibility of distortion added to a picture at those points where the slope is equal to x. Visibility of distortion dicreases with x. Some other factors like viewing conditions, semantic content of the scene as well as the disposition of the observer can also affect the shape of the visibility function. Visibility density function can be approximated by some combination of probability density p(x) and masking density function m(x), i.e

$$v(x)=F\left[p(x),\ m(x)\right] \qquad (2)$$

where masking function can be treated as the psychophysical weighting function. Block diagram of the system used in the subjective tests to obtain the visibility function is shown in Fig.2. The video signal is stored on the drum with an accuracy. The pictures are processed on the computer such that the least significant bit is set to a "one" if the element difference is above treshold and the pictures are restored on the drum with this information. On display, the pictures are read from the drum in real time and the least significant bit is used to switch the noise source in or out of the appropriate pels.

3. MASKING FUNCTION IN DIFFERENTIAL PULSE CODE MODULATION USED FOR PICTURE TRANSMISSION

One of the ways for digital picture transmission is to use DPCM technique. The continuous video signal is sampled and the difference between the present sample and its estimate called prediction error is quantized and coded for transmission. Usually the difference signal is quantized to 8 levels and coded with a 3 or 4 bit code. The band-

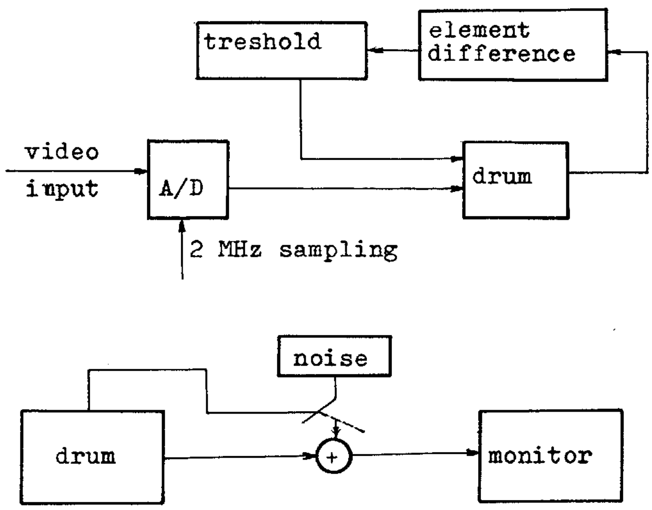

Figure 2. Block diagram of the system used to obtain the visibility function

width reduction is from 6 to 8 bits per pel of conventional PCM to 3 bits per pel for DPCM. At the receiver, in order to provide a reconstruction of the original video signal, the decoded difference signal is reconstructed and combined with an estimate from a predictor identical to the one at the transmitter. The difference between input and output signals assuming errorless transmission of the digits is the quantizing error. Three types of degradation can be seen due to improper design of the quantizer of a DPCM coder. These are reffered to as granular noise, edge busyness and slope overload. When the flat areas are coarsely quantized, they have the appearance of random noise added to the picture. If the largest level of the quantizer is small, then for every high contrast edge it takes several samples for the output to follow the input, resulting in slope overload. For edges whose contrast changes gradually, the quantizer output oscilates around the signal value and may change from line to line or frame to frame giving the appearance of a "busy edge" Usually, the effect of quantization errors in DPCM is to cause local degradations in areas of large slopes or edges in the picture.

The visibility of prediction error depends on a combination of factors such as its probability, perceptibility, etc. In this case, a visibility function is defined as one which relates the subjective visibility of noise added to an image pel to the magnitude of prediction error at the pel. For the previous picture element prediction rule, the prediction error is called the slope of illuminance function.

The design of quantizer involve specifying the positions of the input or decision levels denoted by x_0, x_1, ..., x_M, and the output or representative levels y_1, y_2, ..., y_N. For a small number of quantizing levels there are a large number of variables to be adjusted. Fortunately, picture quality changes slowly with the position of the quantizer levels. The best setting of the levels will change with the type of picture being encoded. It means that pictures with many large changes in amplitude require coarser setting of the levels than pictures having relatively little detail.

The mean square quantization error can be represented by

$$\varepsilon_q^2 = \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} (y_i - x)^2 p(x)\, dx \tag{3}$$

where x is the input signal to the quantizer, p(x) is the pdf and x_i and y_i are the decision and representative levels, respectively.

The amplitude of the input signal to the quantizer with previous element prediction is an approximation to the local horizontal slope of the signal. This is an approximate measure of the amount of local masking. Thus, the mean square error is weighted by a function of the local slope. Let the weighting function be w(x). This weighting function can be determined from subjective experiments in which the treshold of a narrow line stimulus is measured adjacent to an edge as a function of the amplitude of the edge. Thus, the weighted mean square eroor is

$$\varepsilon_{qw}^2 = \sum_{i=1}^{N} \int_{x_{i-1}}^{x_i} (y_i - x)^2 w(x) p(x)\, dx \tag{4}$$

and is functionally equivalent to ε_q^2 if the product w(x)p(x) is replaced by a single function of x.

For prediction errors in a DPCM coder, their visibility function for a given image can be measured as follows. For some fixed interval $[x, x+\Delta x]$ and for suitable small Δx add white noise to all those pels in the original image where the prediction error magnitude or the masking function lies in this interval. Let P_m be the power of the noise. Then obtain another image by adding white noise of power P_w to all the pels such that two images are subjectively equivalent. The visibility density function in this condition is

$$v(x) = \frac{-dV(x)}{dx} \tag{5}$$

where

$$V(x) = \frac{P_w}{P_m} \tag{6}$$

This function varies with the scene.

If the point of fixation of a viewer is determined primarily by the semantics of the picture rather than by the syntax, i.e if the viewer is watching the action, probability would have a large effect on visibility. If a picture is viewed with no restriction on viewing time, probability of occurance will only weakly affect visibility. With these facts it can be taken that

$$v(x) = \frac{p(x)^{\alpha}}{m(x)} \tag{7}$$

When the exponent $\alpha = 0$, the pdf p(x) has no

ON VISUAL CRITERIA IN EFFICIENT PICTURE CODING
Critères Visuels pour Codage d'Image Efficace
Prof. Dr. ing. ZORAN S. BOJKOVIĆ

effect, while for $a=1$, plays an important role. From (7), masking density function becomes

$$m(x)=\frac{p(x)^{a}}{V(x)} \quad (8)$$

An alternate definition of masking density function is

$$m(x)=-\frac{dM(x)}{dx} \quad (9)$$

where masking function is

$$M(x)=\frac{P(x)^{a}}{V(x)} \quad (10)$$

Assume that M(x) is a constant function independent of picture statistics. The aim is to find the value for "a" that minimizes the change in M(x) with changes in picture material. We will minimize the variance of log M(x) since visual tresholds are more directly related to log V(x) than to V(x). From (10),

$$\log M(x)=a\log P(x)-\log V(x) \quad (11)$$

For eight values of V and P equation (11) becomes

$$\log M_{ji}=a\log P_{ji}-\log V_{ji} \quad (12)$$

where the index "i" ranges from 1 to 8, while the index "j" associated with the tresholds ranges from 1 to 10. On the other hand, the covariance of log M_{ji} is

$$\mathrm{cov}(\log M_{ji},\log M_{ji})=\frac{1}{LN-L}\sum_{j=1}^{L}\sum_{i=1}^{N}\left(\log M_{ji}-\overline{\log M_{ji}}\right)^{2} \quad (13)$$

From equation (12) it will be

$$\mathrm{cov}(\log M_{ji},\log M_{ji})=\frac{1}{LN-L}\sum_{j=1}^{L}\sum_{i=1}^{N}\left[a\left(\log P_{ji}-\overline{\log P_{ji}}\right)-\left(\log V_{ji}-\overline{\log V_{ji}}\right)\right]^{2} \quad (14)$$

The solution of "a" that minimizes (12) is analogous to a linear regression of V on P and has the form

$$a=\frac{\mathrm{cov}(V,P)}{\mathrm{cov}(V,V)} \quad (15)$$

With the reasonable assumptions that V(x) and P(x) are exponentially distributed, i.e

$$V(x)=v_0\cdot e^{-v_1 x},\ x\geq 0 \qquad P(x)=e^{-p_1 x},\ x\geq 0 \quad (16)$$

we can approximate the masking function M(x) Equations (9),(10) and (16) imply that M(x) is exponential. Thus, the masking function has the form

$$M(x)=\frac{1}{v_0}\,e^{-m_1 x} \quad (17)$$

where $m_1=ap_1-v_1$. The restriction is that the masking function should not change with picture statistics, while the horizontal element difference is used as a measure of activity in the picture.

4. VISUAL CRITERIA IN TRANSFORM PICTURE CODING

In transform coding a long sequence of data samples is devided into blocks of N samples and each block treated as a vector is quantized independently of other blocks. A widely used measure of reconstructed image fidelity for an NxM size image is the average mean square error defined as

$$e_{ms}^{2}=\frac{1}{N\cdot M}\sum_{i=1}^{N}\sum_{j=1}^{M}\overline{\left(u_{i,j}-u_{i,j}^{*}\right)^{2}} \quad (18)$$

where $\{u_{ij}\}$ and $\{u_{ij}^{*}\}$ represent the NxM original and reproduce images respectively. The criterion of mean square error is mathematically simple to treat but does not fully reflect actual characteristics of information receptors. Thus, the application of criterion for lowest weighted mean square error can be expected to give more efficient codings than mean square error.

For high-quality coding, any differences between the coded and uncoded pictures will be subjectively small, which implies that coding distortions will be close to the visual treshold. Here visual treshold is the point at which a stimulus just becomes visible. It is a statistical measure and is usually defined as the amplitude of the stimulus such that it will be detected on 50% of occasions. When we are concerned with coding for high quality images, knowledge of visual treshold will be a valuable guide in determining the relative amount of tolerable distortion at a pel.

Fig.3 shows a transform picture coding scheme that takes into account the visual criterion. The image luminance field is first converted to a contrast field via a memoryless nonlinear transformation. This image field is then Fourier transformed. The transform domain elements are multiplied by a frequency weighting function $H(\omega_1,\omega_2)$ and the resulting samples are quantized using the usual mean square criterion. Inverse weighting followed by inverse Fourier transformation gives the reconstructed contrast field. For large image block sizes, the frequency weighting function $H(\omega_1,\omega_2)$ can be applied. To apply this method for coding images block by block via an arbitrary transform, the image contrast field should first be convolved with h(x,y) which is the Fourier inverse of $H(\omega_1,\omega_2)$. For practical implementation it would then be desirable to seek discrete finite approximations of h and h^{-1}. The resulting field $\{z_{i,j}\}$ could then be coded by any desired method. At the receiver, the encoded field $\{z_{i,j}\}$ must now be convolved with the inverse function $h^{-1}(i,j)$. The transform domain quan-

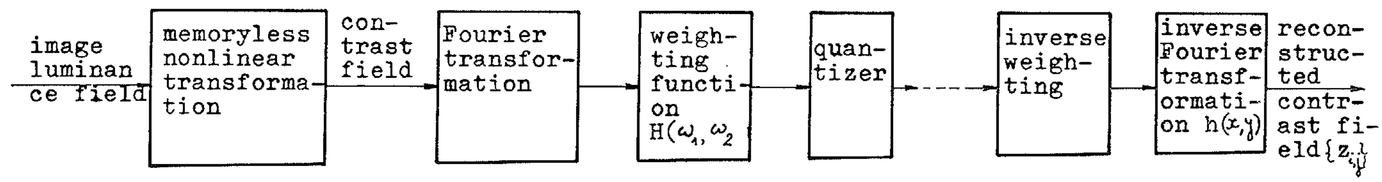

Figure 3. Transform picture coding scheme under the visual criterion

tizer design and bit allocation depend on the statistics of the field $\{z_{i,j}\}$.

In the Hadamard transform domain, using a 2x2 transform of a block of pels in the same field defined as in Fig.4, the measure of spatial activity in a block is taken as

$$H = \max(|H_2|,\ |H_4|) \tag{19}$$

Figure 4 shows that H is the maximum magnitude of the average line or element difference of pels in the block. A,B,C, and D, are the pel positions, while H_1, H_2, H_3 and H_4 are the Hadamard coefficients, given by

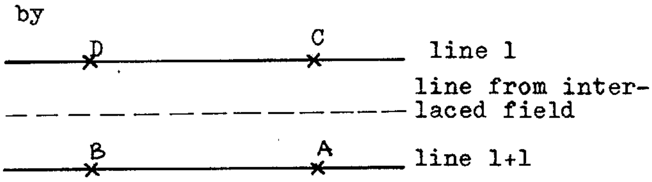

Figure 4. Definition of Hadamard coefficients

$H_1 = A+B+C+D$, $H_2 = A+B-C-D$, $H_3 = A-B-C-D$, $H_4 = A-B+C-D$.

The masking of the quantization noise is related to the spatial detail as measured by masking functions M or H. The test picture is obtained by adding varying amounts of noise to simulate the quantization noise to all pels or blocks. The measure of spatial detail has a given value.

5. CONCLUSION

In efficient picture coding, adaptation of coding strategies to local properties of the picture which determine the visual sensitivity to quantization noise is very important. Weighting functions called visibility functions for the visibility of quantization noise can be derived for relatively large number of pictures under different viewing conditions. One can design quantizers for pictures and viewing conditions by measuring the probability density function of the picture signal.

Visibility function depends on the particular picture being viewed. There are two important effects in the visibility function: statistical and psychophysical. Thus, the visibility function can be approximated by some function of the probability function and a masking function.

Using visibility function as well as the statistics of prediction error, measures of picture quality can be constructed. On the other hand, in transform picture coding, knowing the amount of error tolerable in each transform coefficient is of great importance.

REFERENCES

[1] A.N.Netravali,B.Prasada, "Adaptive Quantization of Picture Signals Using Spatial Masking", Proceedings of IEEE, April 1977, pp 536-548.

[2] J.C.Candy, R.H. Bosforth, "Method for Designing Differential Quantizers Based on Subjective Evaluations of Edge Busyness" Bell Syst.Techn.Journ.,Sept.1972, No.7, Vol.51, pp 1495-1516.

[3] Z.S. Bojković, "A Bound onSignal-to-Quantizing noise ratio in linear transform picture coding systems", Archiv Für Elektronik und Übertragungstechnik AEÜ, band 32 (1978) pp 430-433.

[4] J.B. Millard,H.I.Manusell, "Digital Encoding of the Video Signal", Bell Syst. Tech. Journ. Vol.50, No.2., Feb,1971, pp 459-480.