NEUVIEME COLLOQUE SUR LE TRAITEMENT DU
SIGNAL ET SES APPLICATIONS

NICE du 16 au 20 MAI 1983

SOME EXPERIMENTS IN PARAMETER ENCODING OF VIDEO SIGNALS

Luciano Capo, Angelo Giglio, Giovanni Zarone

Istituto Elettrotecnico, Facoltà d'Ingegneria, Università di Napoli, 21 via Claudio, 80125 Napoli, Italia

RESUME

SUMMARY

Unlike "waveform encoding" (whose aim is simply to replicate original waveforms), "parameter encoding" consists in extracting and transmitting only the basic features necessary for a specific application.

An essential feature of an image, which has to be sent to a human observer, is its contour set. Every contour may be represented by a few parameters. Therefore a dramatic bit rate reduction is expected to be achieved by parameter encoding, i.e. by characterizing an image with only a few parameters and some additional information.

The present work deals with some new experiments in contour tracking, in intraframe and interframe contour coding. A tracking algorithm, which takes into account the contour regularity properties and the high correlation between grey levels of contour elements has been, in fact, investigated. Moreover the high temporal redundancy of video signals has been exploited by an interframe predictive contour coding technique.

The quality of a typical videotelephonic sequence, processed by these techniques, was evaluated by means of MSSTE, a parameter closely related to the subjective opinions of the observers

The quite satisfactory results are capable of further improvement.

1. INTRODUCTION

The aim pursued by image coding is "to minimize the number of bits required to transmit pictures for a given level of distortion". The hitherto deeply investigated methods fall into the "waveform encoding", i.e. they try to replicate the original waveforms by mainly removing the statistical redundancy, present to a high degree in a video signal.

The rate distortion theory gives a solution to this problem. However there are some difficulties in using the results of this theory. Above all it does not tell us how to synthesize a practical coder. Moreover it is difficult to compute the necessary "rate distortion function" for many realistic models of the picture source and distortion criteria, due to the lack of exaustive statistical models for video signals and to the lack of a widely-accepted distortion criterion |1|.

A completely different approach is introduced by parameter encoding, whose aim is not the faithful reconstruction of the original waveforms, but the extraction and transmission of the only basic features necessary for a specific application. Since the ultimate recipient of the information is a normal observer, great importance must be given to perceptual and psycho-physical aspects of human vision in determining the basic features to be taken into account.

It is well known that a lot of information is given to the viewer by the contours of the image objects |2|,|3|. The only contour set, in fact, allows one to recognize a picture immediately and unambigously, although it does not provide the quality required by tv standards. The contour may be segmented in prefixed geometric parts (i.e. circle arcs); these parts may be represented by fairly efficient parameter codes. Rendition of details, shading etc. within object boundaries requires the transmission of additional data. Therefore a dramatic bit rate reduction is expected to be achieved by characterizing an image with only a few parameters and some additional information.

The present paper will deal with some experiments in contour extracting and coding. Because of extremely high temporal redundancy of typical videotelephonic signals, not only intraframe but also interframe coding techniques will be investigated. Then the coding of regions within boundaries (implemented by interpolative methods) will be examined. Finally the results, obtained by applying these techniques to a real videotelephonic sequence, will be given.

The simulated coding process is represented by the block scheme of fig. 1.

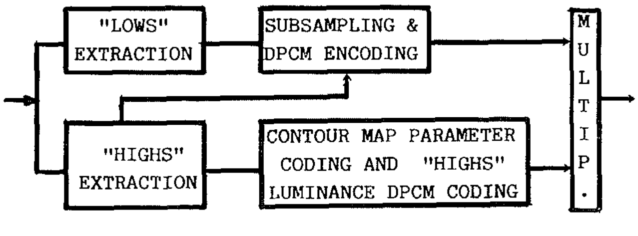


Fig. 1 - Block scheme of the investigated coder.

2. CONTOUR EXTRACTION

2. Generalities

The contour extraction is made up of three steps: detection, thinning and tracking.

2.2 Detection

The detection consists in recognizing the pels which are near sharp luminance transitions and that are possible contour elements. It takes place with the application of a local detection operator. The application to any pixel of a detection operator furnishes a "gradient" matrix; the highest values of this matrix are at contour lines. The sizes of pel neighbourhood, whose elements concur to determine a gradient value, depend on the required accuracy, on the number of quantization levels and on the S/N ratio |4|. In our experiments we used some known operators, such as Laplacian, Sobel and Robert Cross operator |5|. Moreover we tested the following operator:

$$G = \max(|GX|, |GY|)$$

where $GX = L(0) - L(N)$
$GY = L(6) - L(N)$
$L(N)$ is the luminance value of the pel at which G is to be computed;
$L(0)$ and $L(6)$ are the luminance values of adjacent pels on the same row and on the same column (see fig. 2).

3	2	1
4	N	0
5	6	7

Fig. 2 - Elementary cell.

Note that the neighbourhood size of the last operator is less than the neighbourhood size of the other above-mentioned operators. It allows us to accomplish a higher accuracy in contour detection. In general, as the neighbourhood sizes decrease, the vulnerability of the contour extraction process to noise increases. However, this drawback can be neutralized by an appropriate choice of tracking algorithm, as we will see later.

We enphasize that:

a) high gradient values cluster near contour lines;
b) because of special illumination conditions or because of noise, either contour pels may be characterized by low gradient values or flat area pels may be characterized by high gradient values;
c) moving object blurring, due to camera integration, does not make sharp contours.

These problems need to be taken into account in implementing the tracking algorithm.

2.3. Thinning

The clustering of high gradient values in the contour neighbourhood would force the tracking algorithm to duplicate the actual contours and would introduce false contours. In order to avoid this drawback, a thinning operation is necessary.

It is likely that a pel of the actual contour is located at the highest value in a cluster. The thinning operation was therefore implemented as follows. A gradient value is reset to zero if its value is not the maximum once it has been compared with the gradient values pertinent to the adjacent elements on the same row and on the same column, or (see fig. 2):

```
IF(G(N).NE.M1).AND.(G(N).NE.M2)) THEN G(N)=0
```

where $M1 = \max(G(N), G(0), G(4))$
$M2 = \max(G(N), G(2), G(6))$.

2.4. Tracking

The contour tracking consists in recognizing the pels which belong to the same contour. In accomplishing this aim we exploited the following properties:

a) the contour must be continuous;
b) a contour element must have a high detection operator value;
c) the correlation between following pel luminances of the same contour is high (congruence property);
d) the contour lines have regularity requirements (structural property).

The tracking procedure operates according to two different modes: a "start point" mode and a "contour tracing" mode.

According to the "start point" mode, the gradient matrix is scanned. In the scanning every gradient value is compared with a threshold. The threshold getting through locates a start point.

As soon as a start point is located, "contour tracing" mode is performed. This consists in searching the next contour element. The region where this element is searched is the elementary cell, that is the eight pels neighbouring the recently traced element. For each of these eight elements we calculated the function:

$$f = G + F + W1 + W2$$

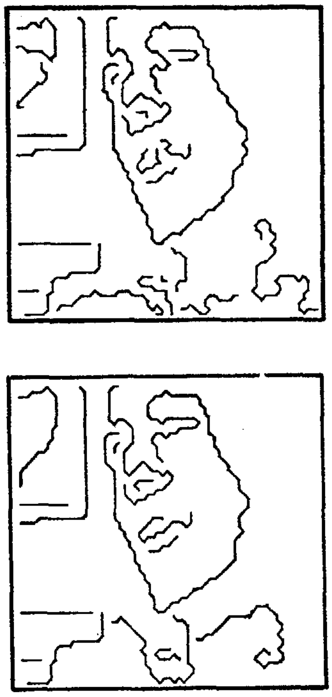
where G is the above detection operator;
F is a non increasing function of luminance difference between present and previous contour element; therefore it takes into account the high correlation existing between luminance values of adjacent pels of the same contour;
$W1 \neq 0$ for the only element in line with the previous ones of the same contour;
$W2 \neq 0$ for the only element which makes the contour closing easier.

The element that makes this function maximum is taken as next element. The search stops when the f function falls below a given threshold. This circumstance locates the contour "end point". Once a contour is completely traced, the "start point" mode is performed again.

The presence of F, W1 and W2 weights allows us to favour the direction for which the contour prosecution is more likely and to neutralize noise effects and contour singularities (gaps and ramification points).

The structure of f function is additive to keep the computation time low. However no significant improvement was obtained by giving f a multiplicative structure.

The optimum W1 and W2 values were found by heuristic search. They are: W1=2; W2=1.



F, W1, W2 = 0

F, W1, W2 ≠ 0

Fig. 3 - The effect of terms F, W1 and W2.

Fig. 3 shows the improvement obtained in contour extracting by using the terms F, W1 and W2.

3. CONTOUR CODING

3.1. Intraframe Contour Coding

The well-known Freeman coding requires 3 bits per each contour element. This figure can be lowered to 2 bits if one adopts a differential technique. A further improvement can be accomplished by an entropy coding. However Freeman coding may represent an intermediate step even in a more sophisticated coding, since it presents some very interestin properties |6|.

3.2. Interframe Contour Coding

A more efficient contour coding is obtained if a prediction of present contour, based on the previous frame contour, is available.

In this the circumstance that displacements are very small (in general they are less than 4 pels/frame) can be exploited. Besides, the translations are almost rigid. Prediction is achieved by computing the displacement of a contour as a whole. The present contour will be different from the predicted one owing to deformations due to movement (since contours are two-dimensional projections of three-dimensional objects). Only significant deformations have to be transmitted.

Displacement as a whole is computed by a correlation technique |7|. Let C1(x,y) and C2(x,y) be the functions that describe previous and present contour.

$$C1(x,y),\ C2(x,y) = \begin{cases} 1 & \text{if the } (x,y) \text{ pel belongs to contour;} \\ 0 & \text{else.} \end{cases}$$

It is decided that a displacement $D^* = (\Delta x^*, \Delta y^*)$ occurs if the $(\Delta x^*, \Delta y^*)$ pair makes correlation function between previous and present contour maximum.

$$CORR(\Delta x, \Delta y) = \sum_{x,y} C1(x+\Delta x, y+\Delta y)\cdot C2(x,y)$$

$$x=1,\ldots x_{max}$$
$$y=1,\ldots y_{max}$$
$$\Delta x,\ \Delta y = -4,\ldots 4$$

$$\Delta x^*, \Delta y^*: \quad CORR(\Delta x^*, \Delta y^*) = \max_{\Gamma} CORR(\Delta x, \Delta y)$$

$$\text{where} \quad \Gamma = |-4\ldots 4| \times |-4\ldots 4|$$

Once the displacement is computed, overlap between predicted and present contour generally appears as in figure 4.

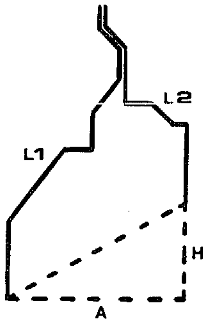


Fig. 4 - An example of overlap between predicted (L1) and present (L2) contour.

The segment that joins end points of two contours is defined "closure". In order to keep computation time low, we assumed:

$$CH = \max(A, H)$$

as an estimate of the closure length. A deformation is regarded as significant, or not, by implementing the test:

```
    IF(CH.GT.T1) THEN the deformation is
                      significant
ELSE IF(AR.GT.T2) THEN the deformation is
                      significant
ELSE the deformation is not significant
```

where T1, T2 are two appropriate threshold and AR is the area limited by closure, present and predicted contour.

An actual overlap of present and predicted contour is depicted in fig. 5.

By using this technique an average 20% bit rate saving has been obtained (see tab. I).

A refinement of displacement estimate is expected to improve the overall performance.

4. "HIGHS" SYNTHESIS AND "LOWS" RECONSTRUCTION

4.1. "Highs" synthesis

Not only the contour position, but also the luminance of elements X and Y (which are near the contour, see fig. 6) needs to be transmitted.

Because of spatial masking phenomenon, it is much more important to replicate the exact contour position than the exact luminance value at the contour. Therefore an alternative coding scheme where the X values were not transmitted but reconstructed at the receiver by interpolation on the ground of Y values, was investigated.

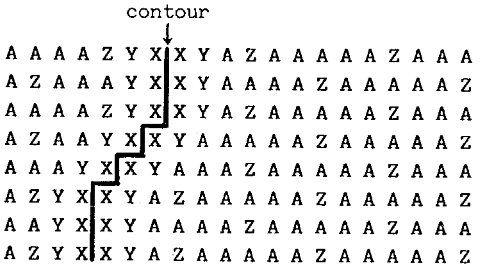
The X and Y luminances were coded by a spatial DPCM.

4.2. "Lows" reconstruction

In order to code the regions inside the detected contours, the Z luminances (obtained by subsampling, see fig. 6) were transmitted. Several

Tab. I - Contour Coding Average Cost

	whole sequence	frame #2
Intraframe (bits/pel)	0.28	0.24
Interframe (bits/pel)	0.22	0.12



```
A A A A Z Y X X Y A Z A A A A A Z A A A
A Z A A A Y X X Y A A A A Z A A A A A Z
A A A A Z Y X X Y A Z A A A A A Z A A A
A Z A A Y X X Y A A A A A Z A A A A A Z
A A A Y X X Y A A A Z A A A A A Z A A A
A Z Y X X Y A Z A A A A A Z A A A A A Z
A A Y X X Y A A A A Z A A A A A Z A A A
A Z Y X X Y A Z A A A A A Z A A A A A Z
```

Fig. 6 - "Highs" (X and Y pels) and "Lows" (A and Z pels) characterization.

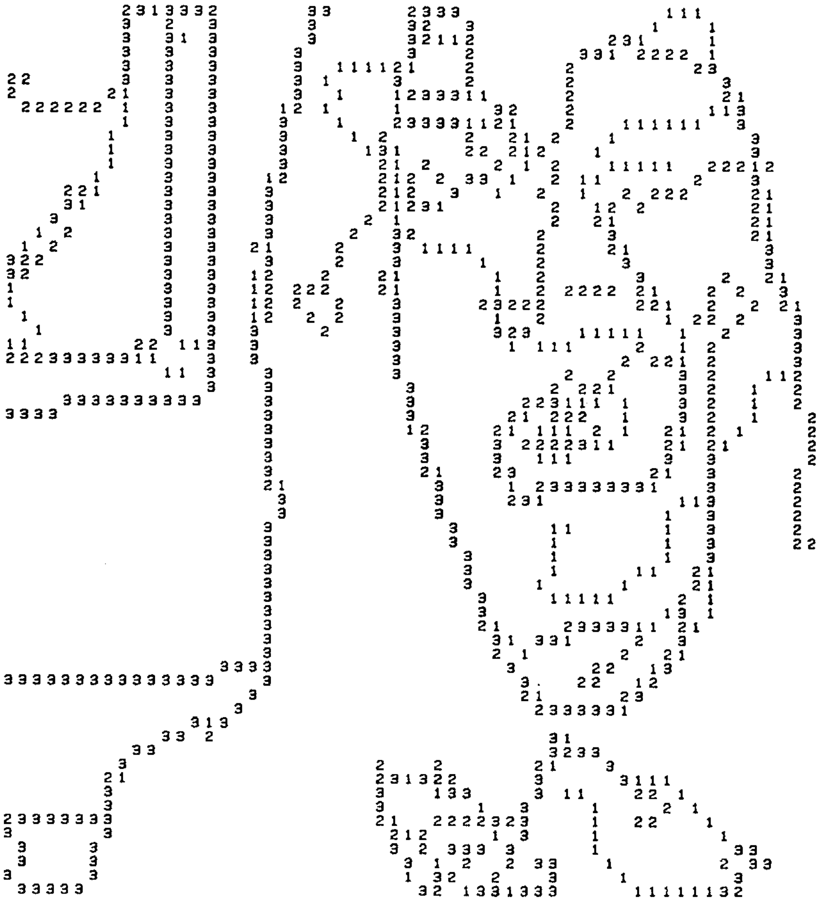


Fig. 5 - Actual overlap between predicted (L1) and present (L2) contour.
label 1: L1 - (L1 ∩ L2)
label 2: L2 - (L1 ∩ L2)
label 3: L1 ∩ L2.

horizontal (HF) and vertical (VF) spatial sampling frequencies were utilized. The A pels were constructed at the receiver by interpolation on the ground of X and Z grey levels. The Z values were coded by a spatial DPCM, following a given path (intraframe coding) or by a temporal DPCM (interframe coding).

5. RESULTS

The test signal is a typical videotelephonic sequence made up of 16 non interlaced pictures. Every frame consists of 64 pels x 64 pels. Every pel luminance was digitized by a uniform quantizer with 256 output levels (8 bits/pel). Some statistics on this signal is given by |8|.

To characterize the performance of the investigated coding techniques tested by the above signal, we used bit rate and MSSTE (Mean Square Supra Threshold Error). The last one is a parameter closely related to the subjective opinions of the observers on picture quality |9|.

Table II shows the results obtained by intraframe and interframe coding. We kept the first 8 and the last 8 frames of the sequence apart. SF is the original spatial sampling frequency.

Tab. II - Main Results

Frame				Intraframe			Inter.
	HF/SF			1/5	1/7	1/7	1/7
	VF/SF			1/2	1/2	1/2	1/2
	Transmitted Highs Pels			X,Y	X,Y	Y	Y
1-8	Lows Luminance		bits/pel	1.06	0.74	0.74	0.27
	Highs	position	bits/pel	0.31	0.31	0.31	0.12
		luminan.	bits/pel	1.38	1.45	0.44	0.40
	Total		bits/pel	2.75	2.50	1.49	0.79
	MSSTE			1.22	1.80	2.84	3.41
9-16	Lows luminance		bits/pel	1.03	0.72	0.72	0.33
	Hihs	position	bits/pel	0.26	0.26	0.26	0.10
		luminan.	bits/pel	1.21	1.18	0.50	0.38
	Total		bits/pel	2.50	2.16	1.48	0.81
	MSSTE			1.03	1.69	2.42	3.80

6. CONCLUSIONS

Some contour extraction algorithms, intraframe and interframe contour coding techniques and lows reconstruction (by interpolative methods) have been investigated in the framework of the parameter picture coding.

The results available are partial. The experimented methods need to be optimized. This study, in fact, is only a preliminary report on a deeper analysis to be carried out.

The parameter encoding deserves a particular consideration because it will undoubtedly allow us to achieve extreme bit rate reduction |10|.

REFERENCES

| 1| A.N. Netravali, J.O. Limb, "Picture coding: a review", PIEEE, vol. 68, March 1980, pp.366-406.

| 2| D. Gabor, P.C.J. Hill, "Television band compression by contour interpolation", PIEE, May 1961, pp. 303-315.

| 3| D.H. Hubel, T.N.Wisel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex", J. Physiol., vol. 160 (1962), pp. 106-154.

| 4| J.K. Aggarwal, R.O. Duda, A. Rosenfeld, "Segmentation: boundaries" in "J.K.Aggarwal, R.O.Duda, A.Rosenfeld (eds), Computer methods in image analysis" IEEE Press, New York (1977), pp. 181-182.

| 5| R.C.Gonzales, P.Wintz, "Digital image processing", (Addison-Wesley Pub. Co., 1977).

| 6| H.Freeman, "On the encoding of arbitrary geometric configurations", IRE Trans. on Electr. Comput., vol. EC-10 (1961), pp. 260-268.

| 7| D.I.Barnea, H.F.Silverman, "A class of algorithms for fast digital image registration", IEEE Trans. on Computer, vol. C-21 (1972), pp. 179-186.

| 8| L.Arena, G.Zarone, "3-D filtering of tv signal", Alta Frequenza, vol. 46 (1977), pp. 108-116.

| 9| D.Sharma, A.N.Netravali, "Design of quantizers for DPCM coding of picture signals", IEEE Trans. on Comm., vol. COM-25 (1977), pp. 1267-1274.

|10| B.G.Haskell, R.Steele, "Audio and video bit rate reduction", PIEEE, vol. 69 (1981), pp. 252-262.