Document No. 52

CCITT Study Group XV
Working Party on Visual Telephony
Specialists Group on Coding for Visual Telephony

Date:24 September 1985

SOURCE: CSELT

TITLE : INTEGRATION OF STANDARDS CONVERSION WITH SOURCE CODING

1. Introduction

This contribution reflects the result of some considerations on the possible architecture of a "Universal Video Codec", the adjective meaning "able to operate with any video standard and on a wide bitrate range" stimulated by the problem of introduction of video communication into a digital video distribution network.

The architecture corresponding to a "hybrid" codec will be taken as an example, but it should be borne in mind that other coding schemes can be treated according to the same principles.

2. Video distribution and video communication

A possible video distribution system is based on a downstream subscriber link at 140 Mbit/s (composed of four video programs at 34 Mbit/s) and an upstream subscriber link at 34 Mbit/s . Video communication is thus possible at 34 Mbit/s in the local area, but a lower bitrate (say the primary rate) will probably be required for long distance and interregional communication.

It would be highly desirable that transparent end-to-end video communication be possible with the same equipment without intermediate down and up bitrate conversion be carried out in the network.

The viability of an approach based on a video codec able to operate according to the same basic coding scheme was assessed by computer simulation. Then an investigation was made to find out a possible architecture allowing a wide operation range and avoiding equipment duplication.

3. A video codec architecture for operation on a wide bitrate range.

Fig. 1 gives a representation of a "hybrid" coder and decoder.

Simulation results have shown that at the coder side the processing required as a function of the bitrate can be subdivided as in Fig. 2. Here "high-speed processing" (broken line) means DPCM, transform coding on PCM data ("changed areas") etc. and "adaptivity" (solid line) means block activity, quantiser and prediction switch, bit assignment rule etc.

In principle, a coder having the capability to operate over a wide bitrate range should thus be equipped with two subsystems: each designed for the maximum work load, corresponding to 34 Mbit/s for the for the "high-speed processing" and to 2 Mbit/s for the "adaptivity", unless the processing involved cannot be carried out with a processing unit whose resources can be flexibly shared.

One way to obtain this flexibility is to insert a "changed area buffer" before the source coder so that the "high-speed processing" part can be made to operate at a speed dependent on the amount of changed areas and not at video speed, so that the "adaptivity" part is allowed more processing resources.

Fig. 3 is a concept block diagram of a conditional replenishment coder suited to operation on a wide bitrate range.

4. Standards conversion revisited

Following the CCIR Recommendation 601, the European video standard (henceforth called 625/50 standard) can be considered as a sequence of frames with a temporal spacing of 40 ms, each with 625 lines and 864 samples/line. Each frame is composed of 2 fields displaced half a line in the vertical direction with 312.5 lines/field. The visible part is 576 lines/frame (288 lines/field) and 720 samples/line. The exact location of the visible window is indicated in Fig. 4.

The American/Japanese standard (henceforth called 525/60 standard) has the same structure, but the spacing between frames is 143x525/2250 = 33.37 ms, the number of lines/frame is 525 and the number of samples/line is 858. The two fields have 262.5 lines each. The visible window is 484 lines/frame (242 lines/field) and 720 samples/line (as for the European standard). The exact location is indicated in Fig. 5.

In both standards the samples have a frequency of 13.5 MHz, this creates a continuous slipping of lines and frames with a periodicity of 1001 European frames equivalent to 1200 American frames.

The basic problem of standards conversion is thus formulated in the following way:

Given a sequence of 625/50 (525/60) samples, reconstruct the corresponding sequence of 525/60 (625/50) samples.

The apparently 3-D filtering problem can be reduced to a 2-D filtering problem (in the v,t domain - v being the vertical coordinate) thanks to equality of sampling frequency and number of active samples/line. A typical situation is depicted in Fig. 6, where it is required to construct a horizontal line with coordinates v0,t0 from the available lines. A reasonable choice is to use three adjacent lines and a linear interpolation formula (holding for every sample of the line):

$$Y(v0,t0)=K1 \times Y(v1,t1) + K2 \times Y2(v2,t2) + K3 \times Y3(v3,t3)$$

where

$$Ki=K/di \text{ and } K=1/(1/d1+1/d2+1/d3)$$

A different number of interpolating lines e.g. 2 or 4 could be used and the formula would change accordingly.

Fig. 7 gives the block diagram of the standards converter described above: the delay line on the input video sequence has three taps sliding rightward or leftward in time according to the standard of the input sequence and AU is an arithmentic unit performing multiplications, additions and address calculation. If the input sequence is 525/60, when the nearest (to AU) tap arrives at the end of the delay line, it jumps to the beginning. If the input sequence is 625/50, when the farthest (from AU) tap arrives at the beginning of the delay line, it jumps to the end.

It appears that no architectural difference exists between the two kinds of standards conversion.

5. Integration of standards converter

An equivalent form of the decoder of Fig. 4, where "Memory" now contains all the information required by the "Interpolation" block i.e. "Frame memory" and "Updated area buffer" of Fig. 3, is given by Fig. 8. Here the source decoder writes into "Memory" the (possibly subsampled) decoded data. "Interpolation" performs all necessary filtering (including standards conversion) on decoded data. This avoids performing twice a spatial interpolation when the "video out" standard is different from the received video standard. Therefore it would be desirable that the source decoder too works on non-interpolated interframe data, but if the algorithm so requires, spatial interpolation (only) needs to be carried out in the interframe prediction loop.

A different case may arise when "Memory" does not contain PCM data but e.g. DCT coefficients. In this case interpolation should comprise an inverse DCT block prior to interpolation. Using known properties of DFT it would be possible to carry out spatial conversion between the two standards at the same time of calculation of the inverse DCT if a 18 x 18 block size is adopted for 625/50 and 15 x 18 for 525/60.

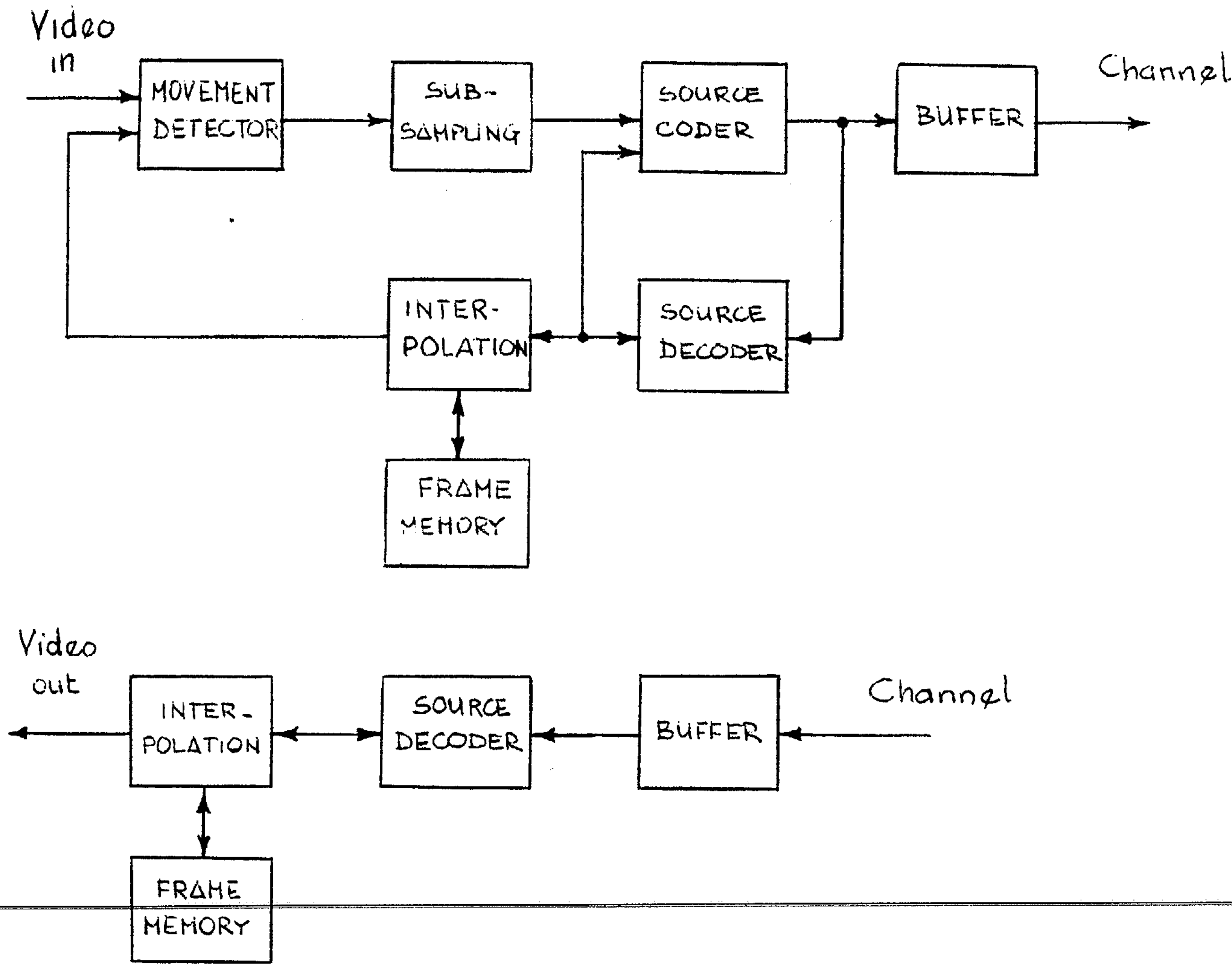

Fig. 1. Block diagram of a "hybrid" video codec

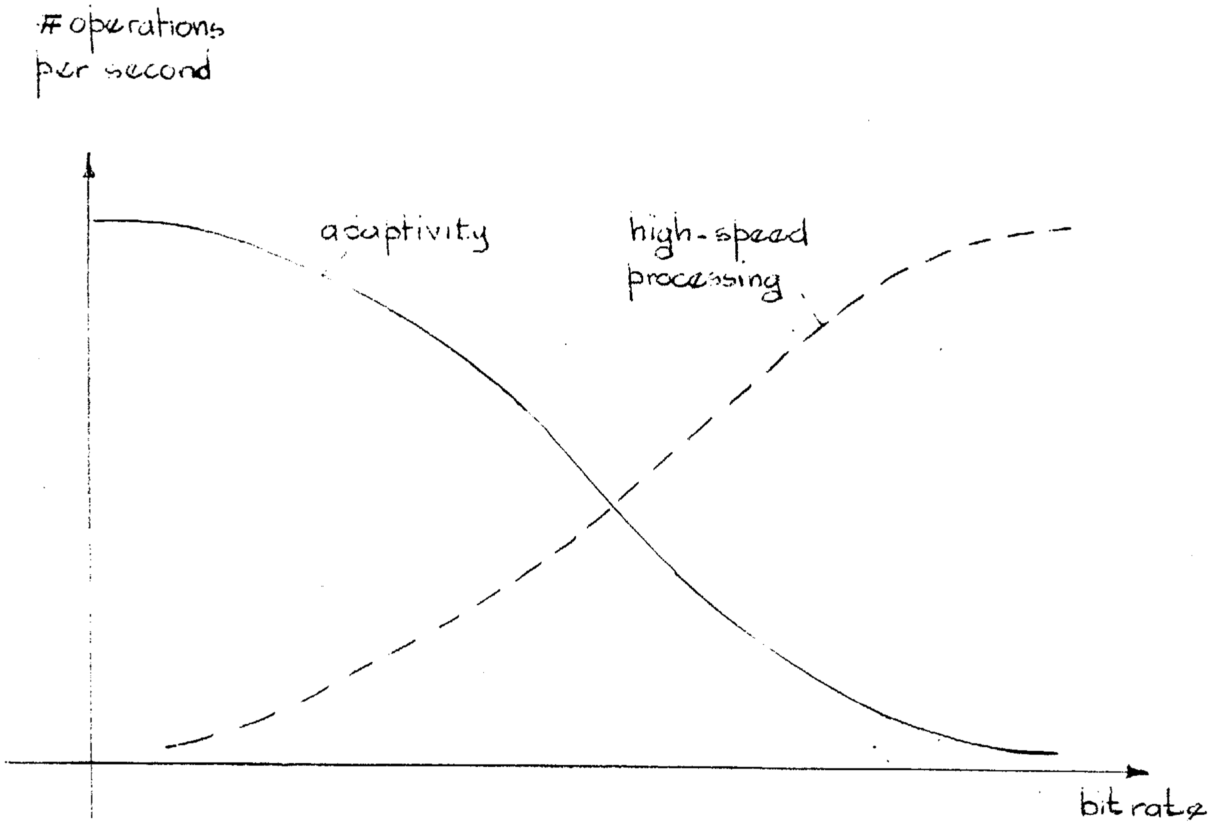

Fig. 2. "High speed processing" and "adaptivity" processes at coder side.

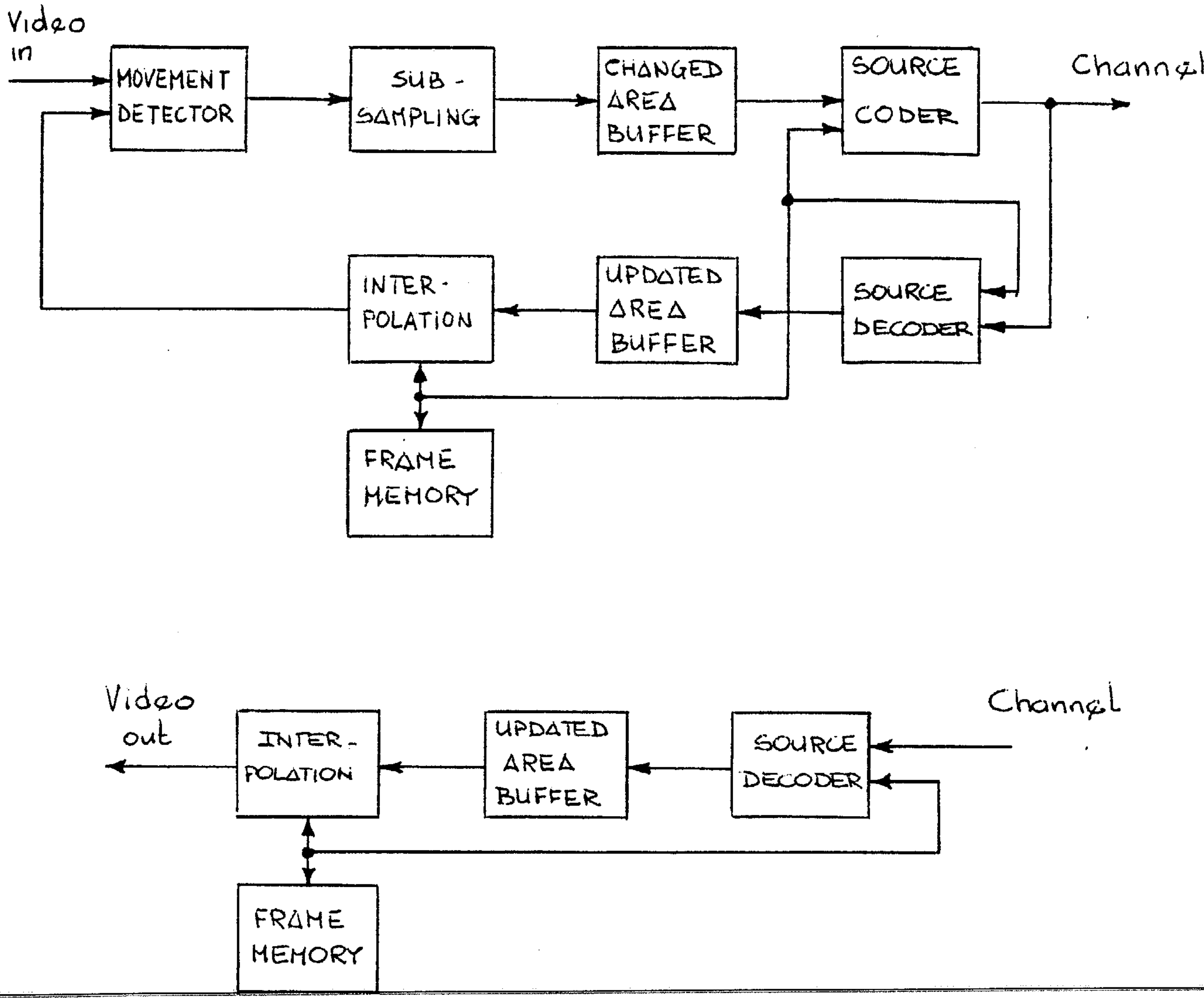

Fig. 3. New block diagram of "hybrid" video codec.

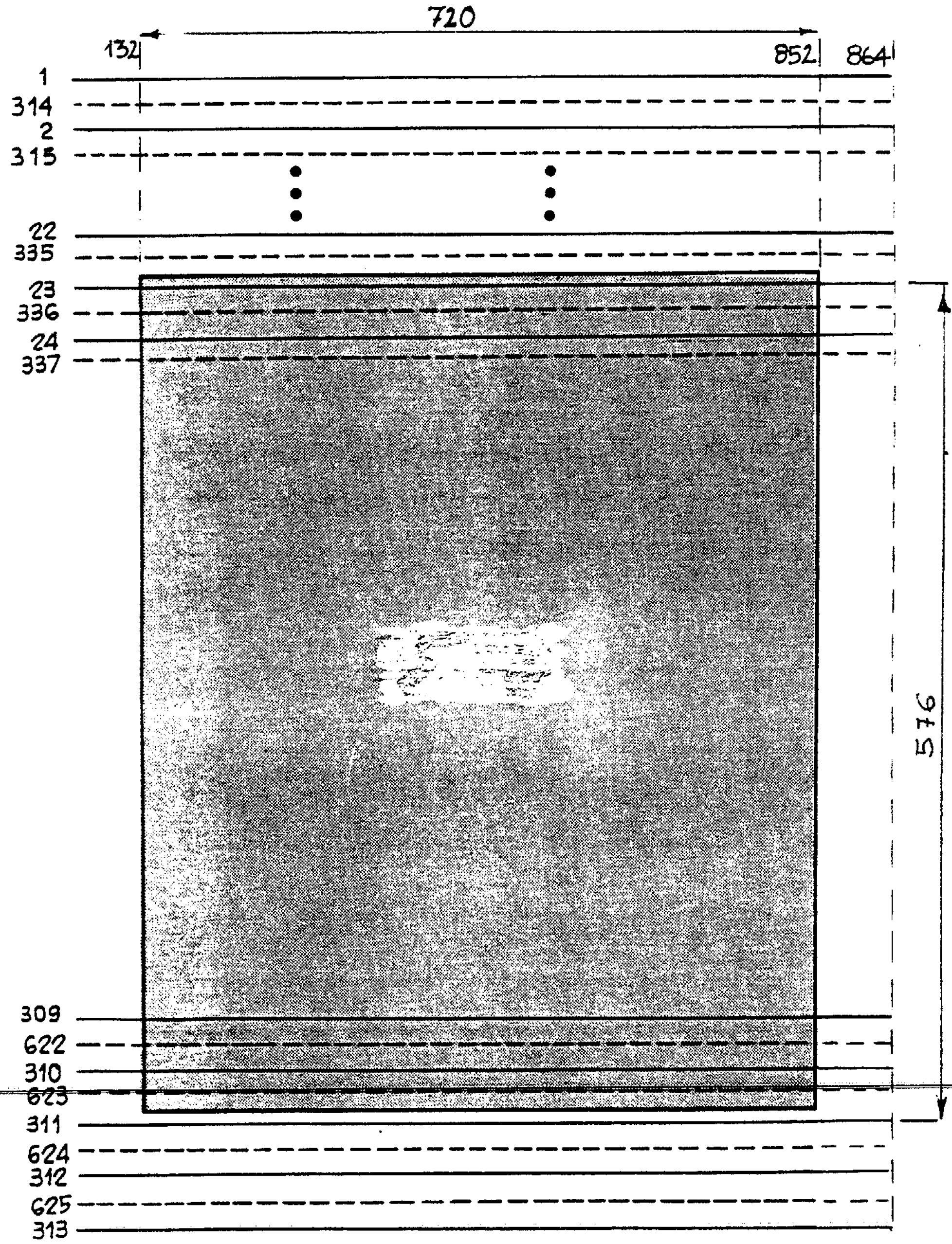

Fig. 4 - The 625/50 video standard

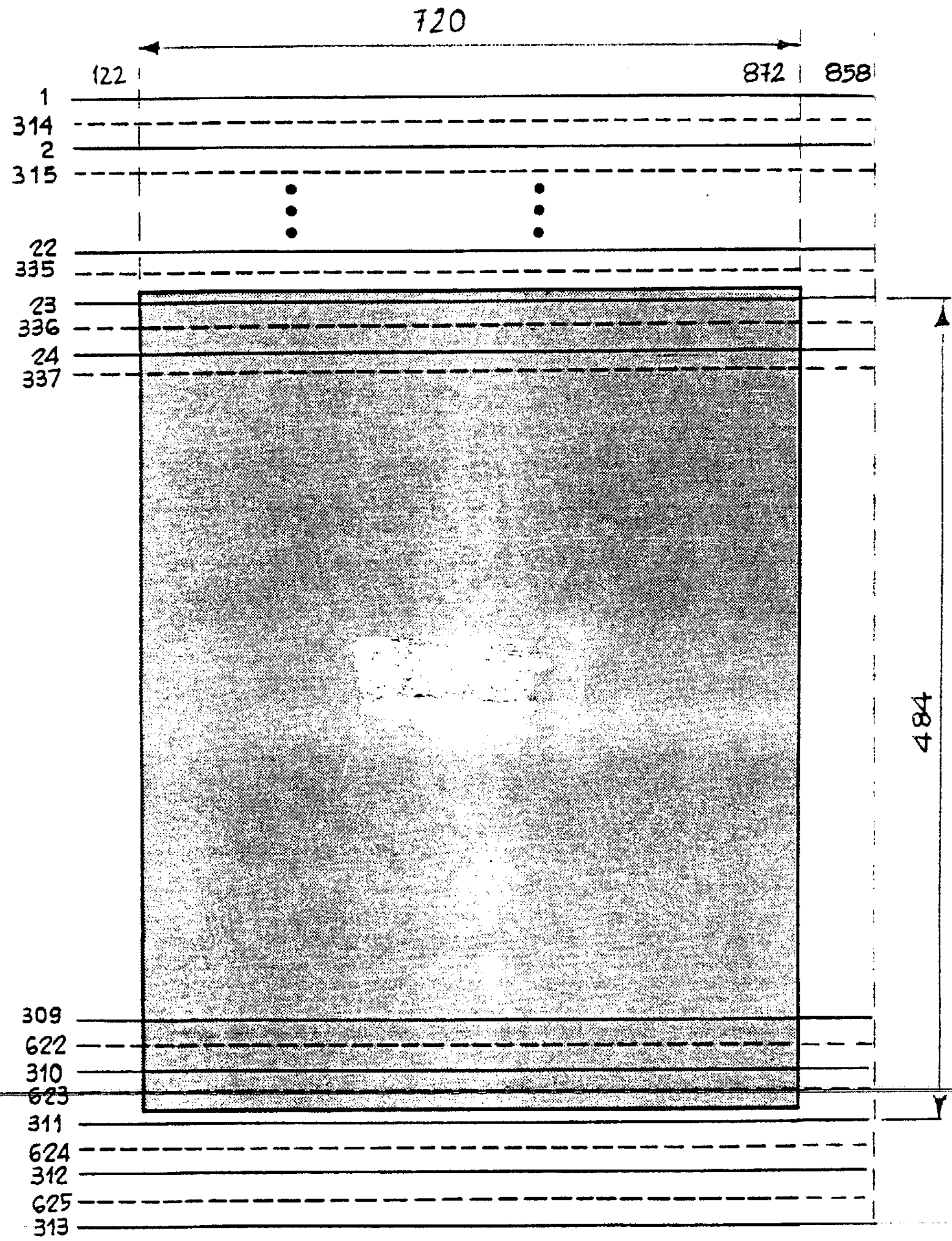

Fig. 5. The 525/60 video standard

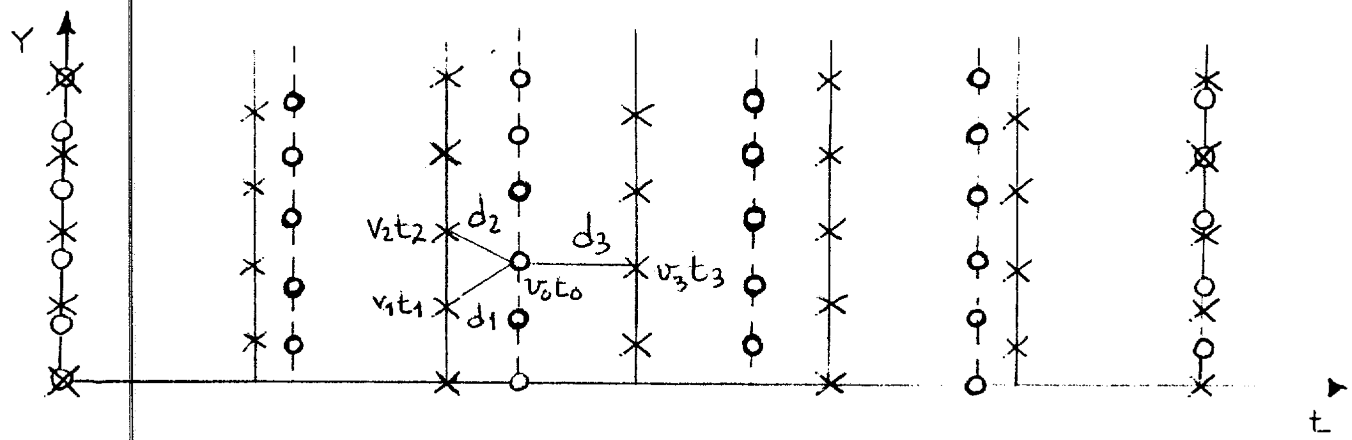

Fig. 6. Geometrical arrangement of samples in standards conversion

× × 525/60

○ ○ 625/50

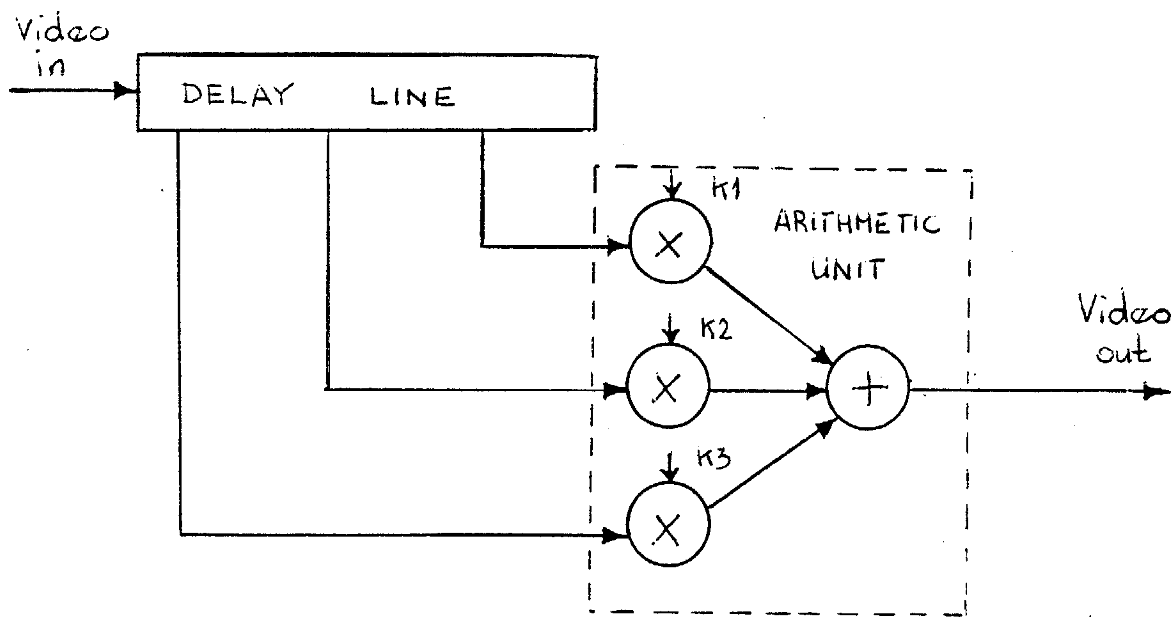

Fig. 7 - Block diagram of the standards converter

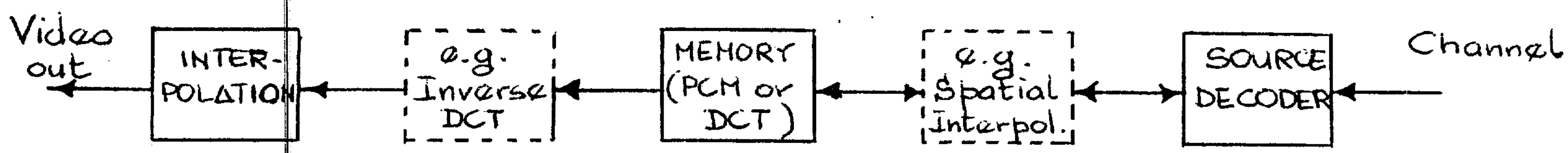

Fig. 8. Hybrid decoder incorporating standards conversion.