



INTERNATIONAL TELECOMMUNICATION UNION

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

H.261

(03/93)

LINE TRANSMISSION OF NON-TELEPHONE SIGNALS

VIDEO CODEC FOR AUDIOVISUAL SERVICES AT $p \times 64$ kbits

ITU-T Recommendation H.261

(Previously "CCITT Recommendation")

FOREWORD

The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of the International Telecommunication Union. The ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Conference (WTSC), which meets every four years, established the topics for study by the ITU-T Study Groups which, in their turn, produce Recommendations on these topics.

ITU-T Recommendation H.261 was revised by the ITU-T Study Group XV (1988-1993) and was approved by the WTSC (Helsinki, March 1-12, 1993).

NOTES

1 As a consequence of a reform process within the International Telecommunication Union (ITU), the CCITT ceased to exist as of 28 February 1993. In its place, the ITU Telecommunication Standardization Sector (ITU-T) was created as of 1 March 1993. Similarly, in this reform process, the CCIR and the IFRB have been replaced by the Radiocommunication Sector.

In order not to delay publication of this Recommendation, no change has been made in the text to references containing the acronyms "CCITT, CCIR or IFRB" or their associated entities such as Plenary Assembly, Secretariat, etc. Future editions of this Recommendation will contain the proper terminology related to the new ITU structure.

2 In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

© ITU 1994

All rights reserved. No part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from the ITU.

CONTENTS

	<i>Page</i>
1 Scope.....	1
2 Brief specification	1
2.1 Video input and output	2
2.2 Digital output and input	2
2.3 Sampling frequency	2
2.4 Source coding algorithm.....	2
2.5 Bit rate	2
2.6 Symmetry of transmission	3
2.7 Error handling	3
2.8 Multipoint operation	3
3 Source coder.....	3
3.1 Source format.....	3
3.2 Video source coding algorithm.....	3
3.3 Coding control	6
3.4 Forced updating	6
4 Video multiplex coder.....	7
4.1 Data structure.....	7
4.2 Video multiplex arrangement	7
4.3 Multipoint considerations	18
5 Transmission coder	19
5.1 Bit rate	19
5.2 Video data buffering	19
5.3 Video coding delay	20
5.4 Forward error correction for coded video signal	20
Annex A – Inverse transform accuracy specification.....	21
Annex B – Hypothetical reference decoder.....	22
Annex C – Codec delay measurement method.....	23
Annex D – Still image transmission	24

Recommendation H.261**VIDEO CODEC FOR AUDIOVISUAL SERVICES AT $p \times 64$ kbit/s***(Geneva, 1990; revised at Helsinki, 1993)*

The CCITT,

considering

- (a) that there is significant customer demand for videophone, videoconference and other audiovisual services;
- (b) that circuits to meet this demand can be provided by digital transmission using the B, H_0 rates or their multiples up to the primary rate or H_{11}/H_{12} rates;
- (c) that ISDNs are likely to be available in some countries that provide a switched transmission service at the B, H_0 or H_{11}/H_{12} rate;
- (d) that the existence of different digital hierarchies and different television standards in different parts of the world complicates the problems of specifying coding and transmission standards for international connections;
- (e) that a number of audiovisual services are likely to appear using basic and primary rate ISDN accesses and that some means of intercommunication among these terminals should be possible;
- (f) that the video codec provides an essential element of the infrastructure for audiovisual services which allows such intercommunication in the framework of Recommendation H.200;
- (g) that Recommendation H.120 for videoconferencing using primary digital group transmission was the first in an evolving series of Recommendations,

appreciating

that advances have been made in research and development of video coding and bit rate reduction techniques which lead to the use of lower bit rates down to 64 kbit/s so that this may be considered as the second in the evolving series of Recommendations,

and noting

that it is the basic objective of the CCITT to recommend unique solutions for international connections,

recommends

that in addition to those codecs complying to Recommendation H.120, codecs having signal processing and transmission coding characteristics described below should be used for international audiovisual services.

NOTES

- 1 Codecs of this type are also suitable for some television services where full broadcast quality is not required.
- 2 Equipment for transcoding from and to codecs according to Recommendation H.120 is under study.

1 Scope

This Recommendation describes the video coding and decoding methods for the moving picture component of audiovisual services at the rates of $p \times 64$ kbit/s, where p is in the range 1 to 30.

2 Brief specification

An outline block diagram of the codec is given in Figure 1.

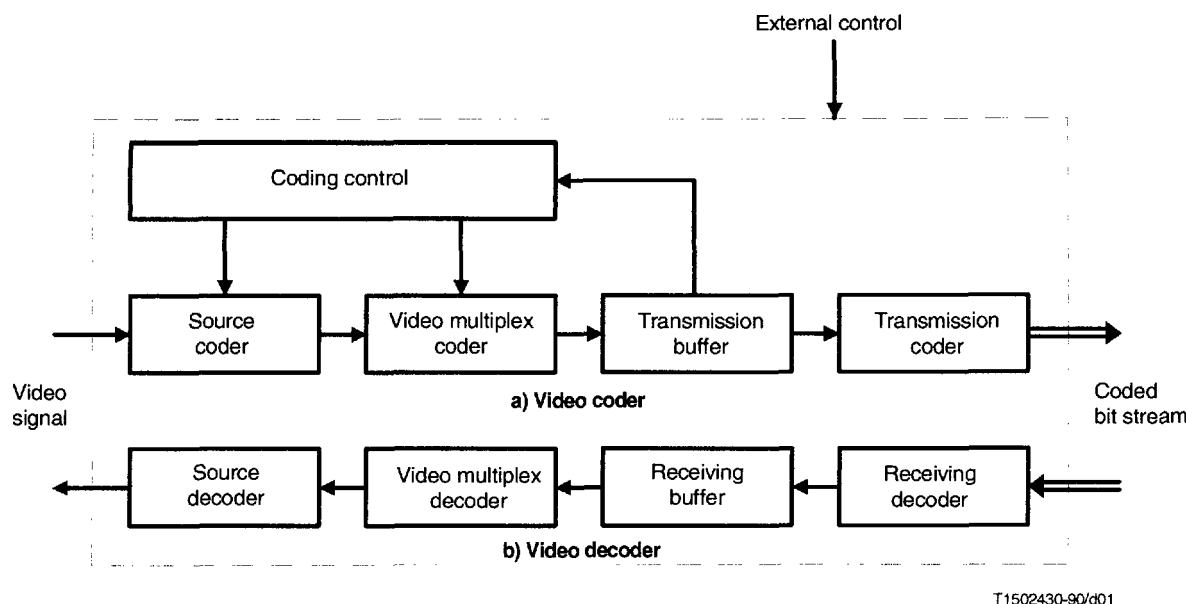


FIGURE 1/H.261

Outline block diagram of the video codec

2.1 Video input and output

To permit a single Recommendation to cover use in and between regions using 625- and 525-line television standards, the source coder operates on pictures based on a common intermediate format (CIF). The standards of the input and output television signals, which may, for example, be composite or component, analogue or digital and the methods of performing any necessary conversion to and from the source coding format are not subject to Recommendation.

2.2 Digital output and input

The video coder provides a self-contained digital bit stream which may be combined with other multi-facility signals (for example as defined in Recommendation H.221). The video decoder performs the reverse process.

2.3 Sampling frequency

Pictures are sampled at an integer multiple of the video line rate. This sampling clock and the digital network clock are asynchronous.

2.4 Source coding algorithm

A hybrid of inter-picture prediction to utilize temporal redundancy and transform coding of the remaining signal to reduce spatial redundancy is adopted. The decoder has motion compensation capability, allowing optional incorporation of this technique in the coder.

2.5 Bit rate

This Recommendation is primarily intended for use at video bit rates between approximately 40 kbit/s and 2 Mbit/s.

2.6 Symmetry of transmission

The codec may be used for bidirectional or unidirectional visual communication.

2.7 Error handling

The transmitted bit-stream contains a BCH code (Bose, Chaudhuri and Hocquengham) (511,493) forward error correction code. Use of this by the decoder is optional.

2.8 Multipoint operation

Features necessary to support switched multipoint operation are included.

3 Source coder

3.1 Source format

The source coder operates on non-interlaced pictures occurring 30 000/1001 (approximately 29.97) times per second. The tolerance on picture frequency is ± 50 ppm.

Pictures are coded as luminance and two colour difference components (Y , C_B and C_R). These components and the codes representing their sampled values are as defined in CCIR Recommendation 601.

Black = 16

White = 235

Zero colour difference = 128

Peak colour difference = 16 and 240.

These values are nominal ones and the coding algorithm functions with input values of 1 through to 254.

Two picture scanning formats are specified.

In the first format (CIF), the luminance sampling structure is 352 pels per line, 288 lines per picture in an orthogonal arrangement. Sampling of each of the two colour difference components is at 176 pels per line, 144 lines per picture, orthogonal. Colour difference samples are sited such that their block boundaries coincide with luminance block boundaries as shown in Figure 2. The picture area covered by these numbers of pels and lines has an aspect ratio of 4:3 and corresponds to the active portion of the local standard video input.

NOTE – The number of pels per line is compatible with sampling the active portions of the luminance and colour difference signals from 525- or 625-line sources at 6.75 and 3.375 MHz, respectively. These frequencies have a simple relationship to those in CCIR Recommendation 601.

The second format, quarter-CIF (QCIF), has half the number of pels and half the number of lines stated above. All codecs must be able to operate using QCIF. Some codecs can also operate with CIF.

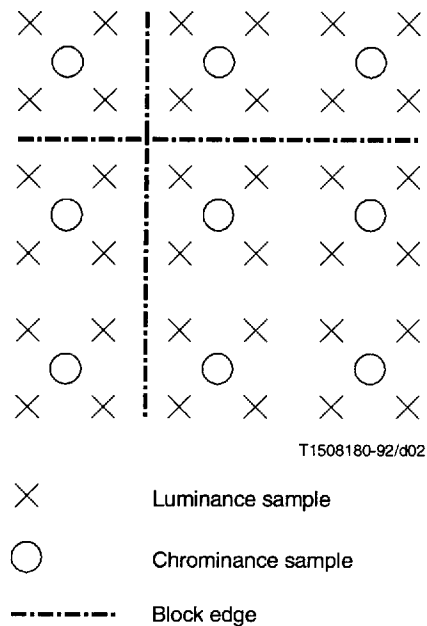
Means shall be provided to restrict the maximum picture rate of encoders by having at least 0, 1, 2 or 3 non-transmitted pictures between transmitted ones. Selection of this minimum number and CIF or QCIF shall be by external means (for example via Recommendation H.221).

3.2 Video source coding algorithm

The source coder is shown in generalized form in Figure 3. The main elements are prediction, block transformation and quantization.

The prediction error (INTER mode) or the input picture (INTRA mode) is subdivided into 8 pel by 8 line blocks which are segmented as transmitted or non-transmitted. Further, four luminance blocks and the two spatially corresponding colour difference blocks are combined to form a macroblock as shown in Figure 10.





T1508180-92/d02

FIGURE 2/H.261
Positioning of luminance and chrominance samples

The criteria for choice of mode and transmitting a block are not subject to recommendation and may be varied dynamically as part of the coding control strategy. Transmitted blocks are transformed and resulting coefficients are quantized and variable length coded.

3.2.1 Prediction

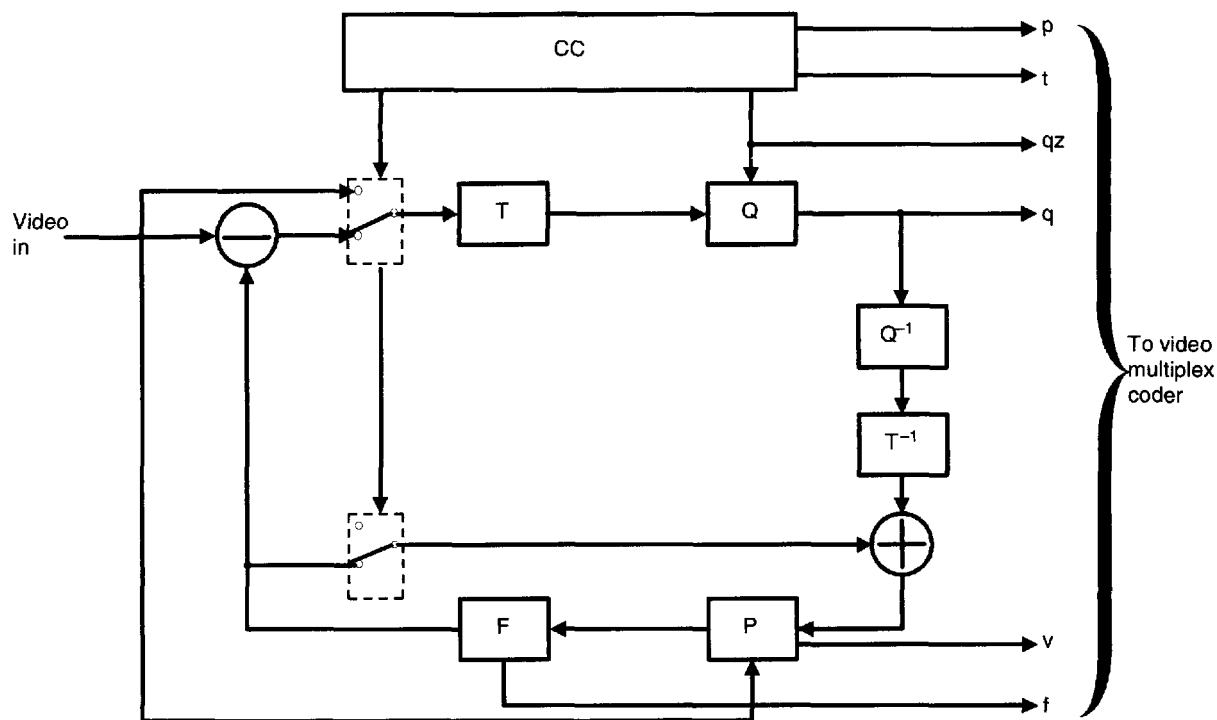
The prediction is inter-picture and may be augmented by motion compensation (see 3.2.2) and a spatial filter (see 3.2.3).

3.2.2 Motion compensation

Motion compensation (MC) is optional in the encoder. The decoder will accept one vector per macroblock. Both horizontal and vertical components of these motion vectors have integer values not exceeding ± 15 . The vector is used for all four luminance blocks in the macroblock. The motion vector for both colour difference blocks is derived by halving the component values of the macroblock vector and truncating the magnitude parts towards zero to yield integer components.

A positive value of the horizontal or vertical component of the motion vector signifies that the prediction is formed from pels in the previous picture which are spatially to the right or below the pels being predicted.

Motion vectors are restricted such that all pels referenced by them are within the coded picture area.



T1502441-90/d03

- T Transform
- Q Quantizer
- P Picture memory with motion compensated variable delay
- F Loop filter
- CC Coding control
- p Flag for INTRA/INTER
- t Flag for transmitted or not
- qz Quantizer indication
- q Quantizing index for transform coefficients
- v Motion vector
- f Switching on/off of the loop filter

FIGURE 3/H.261

Source coder

3.2.3 Loop filter

The prediction process may be modified by a two-dimensional spatial filter (FIL) which operates on pels within a predicted 8 by 8 block.

The filter is separable into one-dimensional horizontal and vertical functions. Both are non-recursive with coefficients of 1/4, 1/2, 1/4 except at block edges where one of the taps would fall outside the block. In such cases the 1-D filter is changed to have coefficients of 0, 1, 0. Full arithmetic precision is retained with rounding to 8 bit integer values at the 2-D filter output. Values whose fractional part is one half are rounded up.

The filter is switched on/off for all six blocks in a macroblock according to the macroblock type (see 4.2.3, MTYPE).

3.2.4 Transformer

Transmitted blocks are first processed by a separable two-dimensional discrete cosine transform of size 8 by 8. The output from the inverse transform ranges from -256 to +255 after clipping to be represented with 9 bits. The transfer function of the inverse transform is given by:

$$f(x, y) = \frac{1}{4} \sum_{u=0}^7 \sum_{v=0}^7 C(u) C(v) F(u, v) \cos[\pi(2x + 1)u/16] \cos[\pi(2y + 1)v/16]$$

with $u, v, x, y = 0, 1, 2, \dots, 7$

where x, y = spatial coordinates in the pel domain,

u, v = coordinates in the transform domain,

$C(u) = 1/\sqrt{2}$ for $u = 0$; otherwise 1,

$C(v) = 1/\sqrt{2}$ for $v = 0$; otherwise 1.

NOTE – Within the block being transformed, $x = 0$ and $y = 0$ refer to the pel nearest the left and top edges of the picture, respectively.

The arithmetic procedures for computing the transforms are not defined, but the inverse one should meet the error tolerance specified in Annex A.

3.2.5 Quantization

The number of quantizers is 1 for the INTRA dc coefficient and 31 for all other coefficients. Within a macroblock the same quantizer is used for all coefficients except the INTRA dc one. The decision levels are not defined. The INTRA dc coefficient is nominally the transform value linearly quantized with a stepsize of 8 and no dead-zone. Each of the other 31 quantizers is also nominally linear but with a central dead-zone around zero and with a step size of an even value in the range 2 to 62.

The reconstruction levels are as defined in 4.2.4.

NOTE – For the smaller quantization step sizes, the full dynamic range of the transform coefficients cannot be represented.

3.2.6 Clipping of reconstructed picture

To prevent quantization distortion of transform coefficient amplitudes causing arithmetic overflow in the encoder and decoder loops, clipping functions are inserted. The clipping function is applied to the reconstructed picture which is formed by summing the prediction and the prediction error as modified by the coding process. This clipper operates on resulting pel values less than 0 or greater than 255, changing them to 0 and 255, respectively.

3.3 Coding control

Several parameters may be varied to control the rate of generation of coded video data. These include processing prior to the source coder, the quantizer, block significance criterion and temporal sub-sampling. The proportions of such measures in the overall control strategy are not subject to recommendation.

When invoked, temporal sub-sampling is performed by discarding complete pictures.

3.4 Forced updating

This function is achieved by forcing the use of the INTRA mode of the coding algorithm. The update pattern is not defined. For control of accumulation of inverse transform mismatch error a macroblock should be forcibly updated at least once per every 132 times it is transmitted.