
**Information technology — Coding of
moving pictures and associated audio for
digital storage media at up to about
1,5 Mbit/s —**

Part 2:
Video

*Technologies de l'information — Codage de l'image animée et du son
associé pour les supports de stockage numérique jusqu'à environ
1,5 Mbit/s —*

Partie 2: Vidéo



Contents

Page

Foreword.....	iii
Introduction.....	iv
Section 1: General.....	1
1.1 Scope.....	1
1.2 Normative references.....	1
Section 2: Technical elements.....	3
2.1 Definitions.....	3
2.2 Symbols and abbreviations.....	11
2.3 Method of describing bitstream syntax.....	13
2.4 Requirements.....	15
Annexes	
A 8 by 8 Inverse discrete cosine transform.....	39
B Variable length code tables.....	40
C Video buffering verifier.....	49
D Guide to encoding video.....	51
E Bibliography.....	108
F List of patent holders.....	109

© ISO/IEC 1993

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 publisher.

ISO/IEC Copyright Office • Case Postale 56 • CH1211 Genève 20 • Switzerland

Printed in Switzerland.

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

International Standard ISO/IEC 11172-2 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Sub-Committee SC 29, *Coded representation of audio, picture, multimedia and hypermedia information*.

ISO/IEC 11172 consists of the following parts, under the general title *Information technology — Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s*:

- *Part 1: Systems*
- *Part 2: Video*
- *Part 3: Audio*
- *Part 4: Compliance testing*

Annexes A, B and C form an integral part of this part of ISO/IEC 11172. Annexes D, E and F are for information only.

Introduction

Note -- Readers interested in an overview of the MPEG Video layer should read this Introduction and then proceed to annex D, before returning to clauses 1 and 2.

0.1 Purpose

This part of ISO/IEC 11172 was developed in response to the growing need for a common format for representing compressed video on various digital storage media such as CDs, DATs, Winchester disks and optical drives. This part of ISO/IEC 11172 specifies a coded representation that can be used for compressing video sequences to bitrates around 1,5 Mbit/s. The use of this part of ISO/IEC 11172 means that motion video can be manipulated as a form of computer data and can be transmitted and received over existing and future networks. The coded representation can be used with both 625-line and 525-line television and provides flexibility for use with workstation and personal computer displays.

This part of ISO/IEC 11172 was developed to operate principally from storage media offering a continuous transfer rate of about 1,5 Mbit/s. Nevertheless it can be used more widely than this because the approach taken is generic.

0.1.1 Coding parameters

The intention in developing this part of ISO/IEC 11172 has been to define a source coding algorithm with a large degree of flexibility that can be used in many different applications. To achieve this goal, a number of the parameters defining the characteristics of coded bitstreams and decoders are contained in the bitstream itself. This allows for example, the algorithm to be used for pictures with a variety of sizes and aspect ratios and on channels or devices operating at a wide range of bitrates.

Because of the large range of the characteristics of bitstreams that can be represented by this part of ISO/IEC 11172, a sub-set of these coding parameters known as the "Constrained Parameters" has been defined. The aim in defining the constrained parameters is to offer guidance about a widely useful range of parameters. Conforming to this set of constraints is not a requirement of this part of ISO/IEC 11172. A flag in the bitstream indicates whether or not it is a Constrained Parameters bitstream.

Summary of the Constrained Parameters:

Horizontal picture size	Less than or equal to 768 pels
Vertical picture size	Less than or equal to 576 lines
Picture area	Less than or equal to 396 macroblocks
Pel rate	Less than or equal to 396x25 macroblocks/s
Picture rate	Less than or equal to 30 Hz
Motion vector range	Less than -64 to +63,5 pels (using half-pel vectors) [backward_f_code and forward_f_code <= 4 (see table D.7)]
Input buffer size (in VBV model)	Less than or equal to 327 680 bits
Bitrate	Less than or equal to 1 856 000 bits/s (constant bitrate)

0.2 Overview of the algorithm

The coded representation defined in this part of ISO/IEC 11172 achieves a high compression ratio while preserving good picture quality. The algorithm is not lossless as the exact pel values are not preserved during coding. The choice of the techniques is based on the need to balance a high picture quality and compression ratio with the requirement to make random access to the coded bitstream. Obtaining good picture quality at the bitrates of interest demands a very high compression ratio, which is not achievable with intraframe coding alone. The need for random access, however, is best satisfied with pure intraframe coding. This requires a careful balance between intra- and interframe coding and between recursive and non-recursive temporal redundancy reduction.

A number of techniques are used to achieve a high compression ratio. The first, which is almost independent from this part of ISO/IEC 11172, is to select an appropriate spatial resolution for the signal. The algorithm then uses block-based motion compensation to reduce the temporal redundancy. Motion compensation is used for causal prediction of the current picture from a previous picture, for non-causal prediction of the current picture from a future picture, or for interpolative prediction from past and future pictures. Motion vectors are defined for each 16-pel by 16-line region of the picture. The difference signal, the prediction error, is further compressed using the discrete cosine transform (DCT) to remove spatial correlation before it is quantized in an irreversible process that discards the less important information. Finally, the motion vectors are combined with the DCT information, and coded using variable length codes.

0.2.1 Temporal processing

Because of the conflicting requirements of random access and highly efficient compression, three main picture types are defined. Intra-coded pictures (I-Pictures) are coded without reference to other pictures. They provide access points to the coded sequence where decoding can begin, but are coded with only a moderate compression ratio. Predictive coded pictures (P-Pictures) are coded more efficiently using motion compensated prediction from a past intra or predictive coded picture and are generally used as a reference for further prediction. Bidirectionally-predictive coded pictures (B-Pictures) provide the highest degree of compression but require both past and future reference pictures for motion compensation. Bidirectionally-predictive coded pictures are never used as references for prediction. The organisation of the three picture types in a sequence is very flexible. The choice is left to the encoder and will depend on the requirements of the application. Figure 1 illustrates the relationship between the three different picture types.

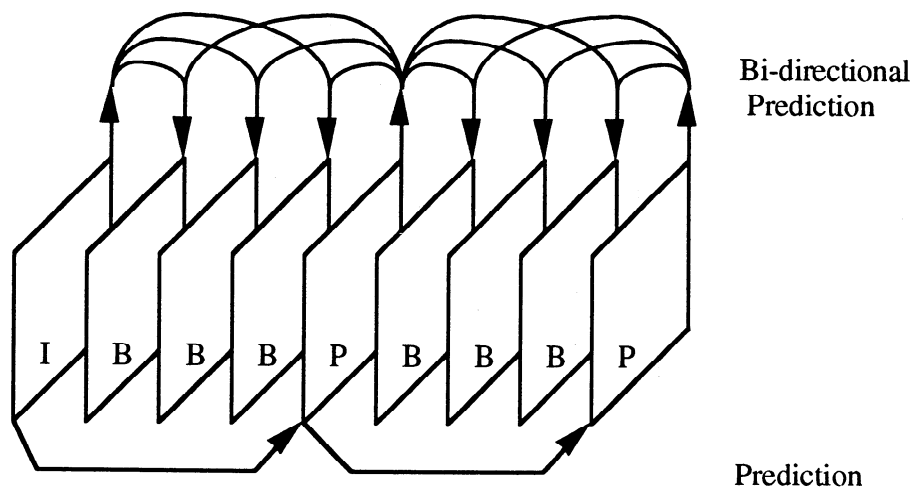


Figure 1 -- Example of temporal picture structure

The fourth picture type defined in this part of ISO/IEC 11172, the D-picture, is provided to allow a simple, but limited quality, fast-forward playback mode.

0.2.2 Motion representation - macroblocks

The choice of 16 by 16 macroblocks for the motion-compensation unit is a result of the trade-off between increasing the coding efficiency provided by using motion information and the overhead needed to store it. Each macroblock can be one of a number of different types. For example, intra-coded, forward-predictive-coded, backward-predictive coded, and bidirectionally-predictive-coded macroblocks are permitted in bidirectionally-predictive coded pictures. Depending on the type of the macroblock, motion vector information and other side information are stored with the compressed prediction error signal in each macroblock. The motion vectors are encoded differentially with respect to the last coded motion vector, using variable-length codes. The maximum length of the vectors that may be represented can be programmed, on a picture-by-picture basis, so that the most demanding applications can be met without compromising the performance of the system in more normal situations.

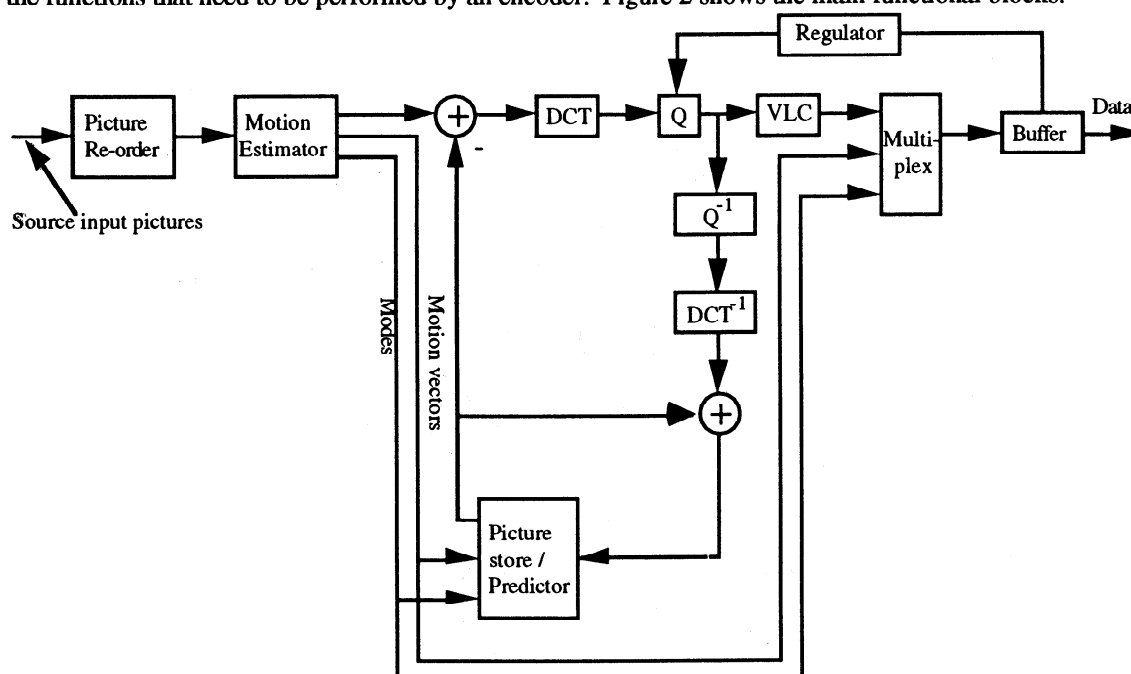
It is the responsibility of the encoder to calculate appropriate motion vectors. This part of ISO/IEC 11172 does not specify how this should be done.

0.2.3 Spatial redundancy reduction

Both original pictures and prediction error signals have high spatial redundancy. This part of ISO/IEC 11172 uses a block-based DCT method with visually weighted quantization and run-length coding. Each 8 by 8 block of the original picture for intra-coded macroblocks or of the prediction error for predictive-coded macroblocks is transformed into the DCT domain where it is scaled before being quantized. After quantization many of the coefficients are zero in value and so two-dimensional run-length and variable length coding is used to encode the remaining coefficients efficiently.

0.3 Encoding

This part of ISO/IEC 11172 does not specify an encoding process. It specifies the syntax and semantics of the bitstream and the signal processing in the decoder. As a result, many options are left open to encoders to trade-off cost and speed against picture quality and coding efficiency. This clause is a brief description of the functions that need to be performed by an encoder. Figure 2 shows the main functional blocks.



where

DCT is discrete cosine transform
 DCT⁻¹ is inverse discrete cosine transform
 Q is quantization
 Q⁻¹ is dequantization
 VLC is variable length coding

Figure 2 -- Simplified video encoder block diagram

The input video signal must be digitized and represented as a luminance and two colour difference signals (Y , C_b , C_r). This may be followed by preprocessing and format conversion to select an appropriate window, resolution and input format. This part of ISO/IEC 11172 requires that the colour difference signals (C_b and C_r) are subsampled with respect to the luminance by 2:1 in both vertical and horizontal directions and are reformatted, if necessary, as a non-interlaced signal.

The encoder must choose which picture type to use for each picture. Having defined the picture types, the encoder estimates motion vectors for each 16 by 16 macroblock in the picture. In P-Pictures one vector is needed for each non-intra macroblock and in B-Pictures one or two vectors are needed.

If B-Pictures are used, some reordering of the picture sequence is necessary before encoding. Because B-Pictures are coded using bidirectional motion compensated prediction, they can only be decoded after the subsequent reference picture (an I or P-Picture) has been decoded. Therefore the pictures are reordered by the

encoder so that the pictures arrive at the decoder in the order for decoding. The correct display order is recovered by the decoder.

The basic unit of coding within a picture is the macroblock. Within each picture, macroblocks are encoded in sequence, left to right, top to bottom. Each macroblock consists of six 8 by 8 blocks: four blocks of luminance, one block of Cb chrominance, and one block of Cr chrominance. See figure 3. Note that the picture area covered by the four blocks of luminance is the same as the area covered by each of the chrominance blocks. This is due to subsampling of the chrominance information to match the sensitivity of the human visual system.

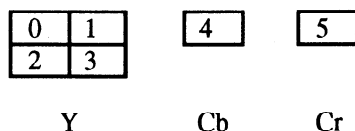


Figure 3 -- Macroblock structure

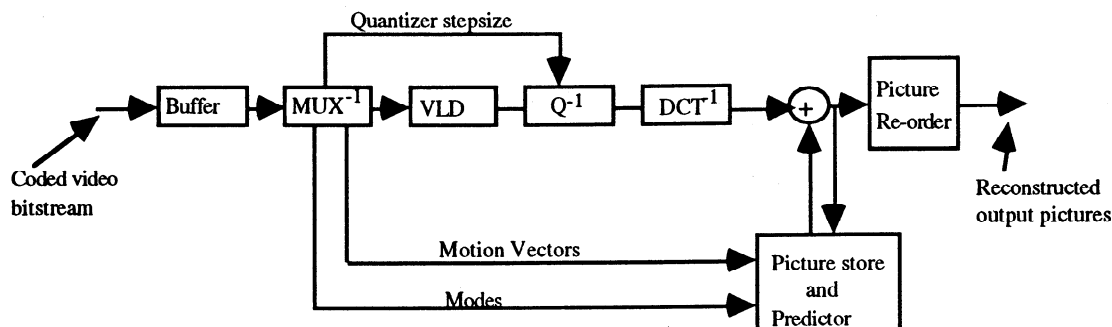
Firstly, for a given macroblock, the coding mode is chosen. It depends on the picture type, the effectiveness of motion compensated prediction in that local region, and the nature of the signal within the block. Secondly, depending on the coding mode, a motion compensated prediction of the contents of the block based on past and/or future reference pictures is formed. This prediction is subtracted from the actual data in the current macroblock to form an error signal. Thirdly, this error signal is separated into 8 by 8 blocks (4 luminance and 2 chrominance blocks in each macroblock) and a discrete cosine transform is performed on each block. Each resulting 8 by 8 block of DCT coefficients is quantized and the two-dimensional block is scanned in a zig-zag order to convert it into a one-dimensional string of quantized DCT coefficients. Fourthly, the side-information for the macroblock (mode, motion vectors etc) and the quantized coefficient data are encoded. For maximum efficiency, a number of variable length code tables are defined for the different data elements. Run-length coding is used for the quantized coefficient data.

A consequence of using different picture types and variable length coding is that the overall data rate is variable. In applications that involve a fixed-rate channel, a FIFO buffer may be used to match the encoder output to the channel. The status of this buffer may be monitored to control the number of bits generated by the encoder. Controlling the quantization process is the most direct way of controlling the bitrate. This part of ISO/IEC 11172 specifies an abstract model of the buffering system (the Video Buffering Verifier) in order to constrain the maximum variability in the number of bits that are used for a given picture. This ensures that a bitstream can be decoded with a buffer of known size.

At this stage, the coded representation of the picture has been generated. The final step in the encoder is to regenerate I-Pictures and P-Pictures by decoding the data so that they can be used as reference pictures for subsequent encoding. The quantized coefficients are dequantized and an inverse 8 by 8 DCT is performed on each block. The prediction error signal produced is then added back to the prediction signal and limited to the required range to give a decoded reference picture.

0.4 Decoding

Decoding is the inverse of the encoding operation. It is considerably simpler than encoding as there is no need to perform motion estimation and there are many fewer options. The decoding process is defined by this part of ISO/IEC 11172. The description that follows is a very brief overview of one possible way of decoding a bitstream. Other decoders with different architectures are possible. Figure 4 shows the main functional blocks.



Where

- DCT⁻¹ is inverse discrete cosine transform
- Q⁻¹ is dequantization
- MUX⁻¹ is demultiplexing
- VLD is variable length decoding

Figure 4 -- Basic video decoder block diagram

For fixed-rate applications, the channel fills a FIFO buffer at a constant rate with the coded bitstream. The decoder reads this buffer and decodes the data elements in the bitstream according to the defined syntax.

As the decoder reads the bitstream, it identifies the start of a coded picture and then the type of the picture. It decodes each macroblock in the picture in turn. The macroblock type and the motion vectors, if present, are used to construct a prediction of the current macroblock based on past and future reference pictures that have been stored in the decoder. The coefficient data are decoded and dequantized. Each 8 by 8 block of coefficient data is transformed by an inverse DCT (specified in annex A), and the result is added to the prediction signal and limited to the defined range.

After all the macroblocks in the picture have been processed, the picture has been reconstructed. If it is an I-picture or a P-picture it is a reference picture for subsequent pictures and is stored, replacing the oldest stored reference picture. Before the pictures are displayed they may need to be re-ordered from the coded order to their natural display order. After reordering, the pictures are available, in digital form, for post-processing and display in any manner that the application chooses.

0.5 Structure of the coded video bitstream

This part of ISO/IEC 11172 specifies a syntax for a coded video bitstream. This syntax contains six layers, each of which either supports a signal processing or a system function:

Layers of the syntax	Function
Sequence layer	Random access unit: context
Group of pictures layer	Random access unit: video
Picture layer	Primary coding unit
Slice layer	Resynchronization unit
Macroblock layer	Motion compensation unit
Block layer	DCT unit

0.6 Features supported by the algorithm

Applications using compressed video on digital storage media need to be able to perform a number of operations in addition to normal forward playback of the sequence. The coded bitstream has been designed to support a number of these operations.

0.6.1 Random access

Random access is an essential feature for video on a storage medium. Random access requires that any picture can be decoded in a limited amount of time. It implies the existence of access points in the bitstream - that is segments of information that are identifiable and can be decoded without reference to other segments of data. A spacing of two random access points (Intra-Pictures) per second can be achieved without significant loss of picture quality.

0.6.2 Fast search

Depending on the storage medium, it is possible to scan the access points in a coded bitstream (with the help of an application-specific directory or other knowledge beyond the scope of this part of ISO/IEC 11172) to obtain a fast-forward and fast-reverse playback effect.

0.6.3 Reverse playback

Some applications may require the video signal to be played in reverse order. This can be achieved in a decoder by using memory to store entire groups of pictures after they have been decoded before being displayed in reverse order. An encoder can make this feature easier by reducing the length of groups of pictures.

0.6.4 Error robustness

Most digital storage media and communication channels are not error-free. Appropriate channel coding schemes should be used and are beyond the scope of this part of ISO/IEC 11172. Nevertheless the compression scheme defined in this part of ISO/IEC 11172 is robust to residual errors. The slice structure allows a decoder to recover after a data error and to resynchronize its decoding. Therefore, bit errors in the compressed data will cause errors in the decoded pictures to be limited in area. Decoders may be able to use concealment strategies to disguise these errors.

0.6.5 Editing

There is a conflict between the requirement for high coding efficiency and easy editing. The coding structure and syntax have not been designed with the primary aim of simplifying editing at any picture. Nevertheless a number of features have been included that enable editing of coded data.

Information technology — Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s —

Part 2: Video

Section 1: General

1.1 Scope

This part of ISO/IEC 11172 specifies the coded representation of video for digital storage media and specifies the decoding process. The representation supports normal speed forward playback, as well as special functions such as random access, fast forward playback, fast reverse playback, normal speed reverse playback, pause and still pictures. This part of ISO/IEC 11172 is compatible with standard 525- and 625-line television formats, and it provides flexibility for use with personal computer and workstation displays.

ISO/IEC 11172 is primarily applicable to digital storage media supporting a continuous transfer rate up to about 1,5 Mbit/s, such as Compact Disc, Digital Audio Tape, and magnetic hard disks. Nevertheless it can be used more widely than this because of the generic approach taken. The storage media may be directly connected to the decoder, or via communications means such as busses, LANs, or telecommunications links. This part of ISO/IEC 11172 is intended for non-interlaced video formats having approximately 288 lines of 352 pels and picture rates around 24 Hz to 30 Hz.

1.2 Normative references

The following International Standards contain provisions which, through reference in this text, constitute provisions of this part of ISO/IEC 11172. At the time of publication, the editions indicated were valid. All standards are subject to revision, and parties to agreements based on this part of ISO/IEC 11172 are encouraged to investigate the possibility of applying the most recent editions of the standards indicated below. Members of IEC and ISO maintain registers of currently valid International Standards.

ISO/IEC 11172-1:1993 *Information technology - Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s - Part 1: Systems.*

ISO/IEC 11172-3:1993 *Information technology - Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s - Part 3 Audio.*

CCIR Recommendation 601-2 *Encoding parameters of digital television for studios.*

CCIR Report 624-4 *Characteristics of systems for monochrome and colour television.*

CCIR Recommendation 648 *Recording of audio signals.*

CCIR Report 955-2 *Sound broadcasting by satellite for portable and mobile receivers, including Annex IV Summary description of Advanced Digital System II.*

CCITT Recommendation J.17 *Pre-emphasis used on Sound-Programme Circuits.*

IEEE Draft Standard P1180/D2 1990 *Specification for the implementation of 8x 8 inverse discrete cosine transform*".

IEC publication 908:1987 *CD Digital Audio System*.