MPEG-1 AND MPEG-2 Video Standards

Supavadee Aramvith and Ming-Ting Sun

Information Processing Laboratory, Department of Electrical Engineering, Box 352500
University of Washington, Seattle, Washington 98195-2500
{supava,sun}@ee.washington.edu

1. MPEG-1 Video Coding Standard

1.1 Introduction

1.1.1 Background and structure of MPEG-1 standards activities

The development of digital video technology in the 1980s has made it possible to use digital video compression in various kinds of applications. The effort to develop standards for coded representation of moving pictures, audio, and their combination is carried out in the Moving Picture Experts Group (MPEG). MPEG is a group formed under the auspices of the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC). It operates in the framework of the Joint ISO/IEC Technical Committee 1 (JTC 1) on Information Technology, which was formally Working Group 11 (WG11) of Sub-Committee 29 (SC29). The premise is to set the standard for coding moving pictures and the associated audio for digital storage media at about 1.5 Mbit/s so that a movie can be compressed and stored in a CD-ROM (Compact Disc – Read Only Memory). The resultant standard is the international standard for moving picture compression, ISO/IEC 11172 or MPEG-1 (Moving Picture Experts Group - Phase 1). MPEG-1 standards consist of 5 parts, including: Systems (11172-1), Video (11172-2), Audio (11172-3), Conformance Testing (11172-4), and Software Simulation (11172-5). In this chapter, we will focus only on the video part.

The activity of the MPEG committee started in 1988 based on the work of ISO JPEG (Joint Photographic Experts Group) [1] and CCITT Recommendation H.261: "Video Codec for Audiovisual Services at px64 kbits/s" [2]. Thus, the MPEG-1 standard has much in common with the JPEG and H.261 standards. The MPEG development methodology was similar to that of H.261 and was divided into three phases: Requirements, Competition, and Convergence [3]. The purpose of the Requirements phase is to precisely set the focus of the effort and determine the rule for the competition phase. The document of this phase is a "Proposal Package Description" [4] and a test methodology [5]. The next step is the competition phase in which the goal is to obtain state of the art technology from the best of academic and industrial research. The criteria are based on the technical merits and the trade-off between video quality and the cost of implementation of the ideas and the subjective test [5]. After the competition phase, various ideas and techniques are integrated into one solution in the convergence phase. The solution results in a document called the simulation model. The simulation model implements, in some sort of programming language, the operation of a reference encoder and a decoder. The simulation model is used to carry out simulations to optimize the performance of the coding scheme [6]. A series of fully documented experiments called core experiments are then carried out. The MPEG committee reached the Committee Draft (CD) status in September 1990 and the Committee Draft (CD 11172) was approved in December 1991. International Standard (IS) 11172 for the first three parts was established in November 1992. The IS for the last two parts was finalized in November 1994.

1.1.2 MPEG-1 target applications and requirements

The MPEG standard is a generic standard, which means that it is not limited to a particular application. A variety of digital storage media applications of MPEG-1 have been proposed based on the assumptions that the acceptable video and audio quality can be obtained for a total bandwidth of about 1.5 Mbits/s. Typical storage media for these applications include CD-ROM, DAT (Digital Audio Tape), Winchester-type computer disks, and writable optical disks. The target applications are asymmetric applications where the compression process is performed once and the decompression process is required often. Examples of the asymmetric applications include video CD, video on demand, and video games. In these asymmetric applications, the encoding delay is not a concern. The encoders are needed only in small quantities while the decoders are needed in large volumes. Thus, the encoder complexity is not a concern while the decoder complexity needs to be low in order to result in low-cost decoders.

The requirements for compressed video in digital storage media mandate several important features of the MPEG-1 compression algorithm. The important features include normal playback, frame-based random access and editing of video, reverse playback, fast forward / reverse play, encoding high-resolution still frames, robustness to uncorrectable errors, etc. The applications also require MPEG-1 to support flexible picture-sizes and frame-rates. Another requirement is that the encoding process can be performed in reasonable speed using existing hardware technologies and the decoder can be implemented using small number of chips in low cost.

Since MPEG-1 video coding algorithm is based heavily on H.261, in the following sections, we will focus only on those which are different from H.261.

1.2 MPEG-1 Video Coding vs. H.261

1.2.1 Bi-directional motion compensated prediction

In H.261, only the previous video frame is used as the reference frame for the motion compensated prediction (forward prediction). MPEG-1 allows the future frame to be used as the reference frame for the motion compensated prediction (backward prediction), which can provide better prediction. For example, as shown in figure 1, if there are moving objects, and if only the forward prediction is used, there will be uncovered areas (such as the block behind the car in Frame N) for which we may not be able to find a good matching block from the previous reference picture (Frame N-1). On the other hand, the backward prediction can properly predict these uncovered areas since they are available in the future reference picture, i.e. frame N+1 in this example. Also shown in the figure, if there are objects moving into the picture (the airplane in the figure), these new objects cannot be predicted from the previous picture, but can be predicted from the future picture.

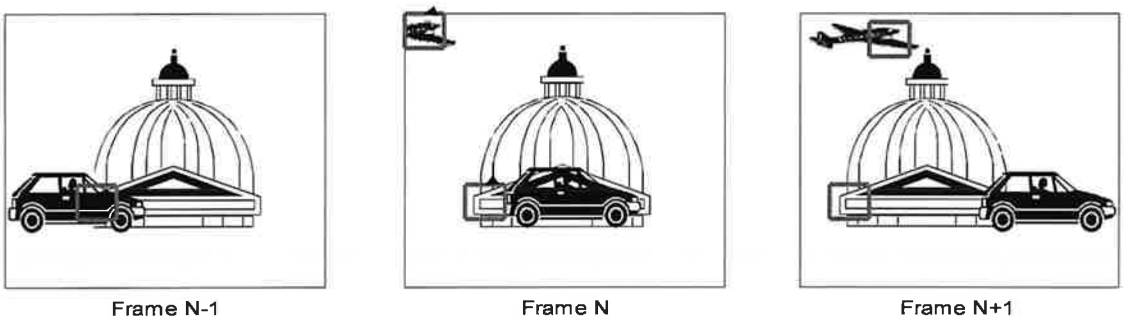

Figure 1: A video sequence showing the benefits of bi-directional prediction.

1.2.2 Motion compensated prediction with half-pixel accuracy

The motion estimation in H.261 is restricted to only integer-pixel accuracy. However, a moving object often moves to a position which is not on the pixel-grid but between the pixels. MPEG-1 allows half-pixel-accuracy motion vectors. By estimating the displacement at a finer resolution, we can expect improved prediction and, thus, better performance than motion estimation with integer-pixel accuracy. As shown in Figure 2, since there is no pixel-value at the half-pixel locations, interpolation is required to produce the pixel-values at the half-pixel positions. Bi-linear interpolation is used in MPEG-1 for its simplicity. As in H.261, the motion estimation is performed only on luminance blocks. The resulting motion vector is scaled by 2 and applied to the chrominance blocks. This reduces the computation but may not necessarily be optimal. Motion vectors are differentially encoded with respect to the motion vector in the preceding adjacent macroblock. The reason is that the motion vectors of adjacent regions are highly correlated, as it is quite common to have relatively uniform motion over areas of picture.

1.3 MPEG-1 video structure

1.3.1 Source Input Format (SIF)

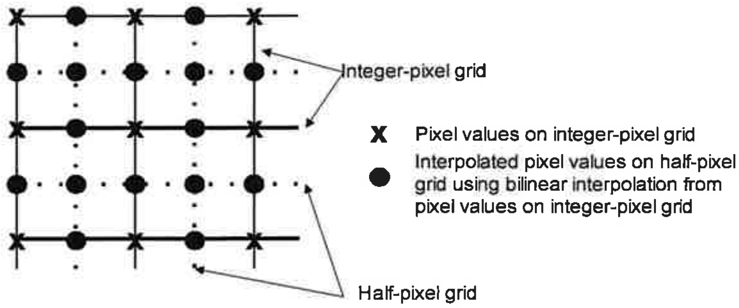

Figure 2: Half-pixel motion estimation.

The typical MPEG-1 input format is the Source Input Format (SIF). SIF was derived from CCIR601, a worldwide standard for digital TV studio. CCIR601 specifies the Y Cb Cr color coordinate where Y is the luminance component (black and white information), and Cb and Cr are two color difference signals (chrominance components). A luminance sampling frequency of 13.5 MHz was adopted. There are several Y Cb Cr sampling formats, such as 4:4:4, 4:2:2, 4:1:1, and 4:2:0. In 4:4:4, the sampling rates for Y, Cb, and Cr are the same. In 4:2:2, the sampling rates of Cb and Cr are half of that of Y. In 4:1:1 and 4:2:0, the sampling rates of Cb and Cr are one quarter of that of Y. The positions of Y Cb Cr samples for 4:4:4, 4:2:2, 4:1:1, and 4:2:0 are shown in Figure 3.

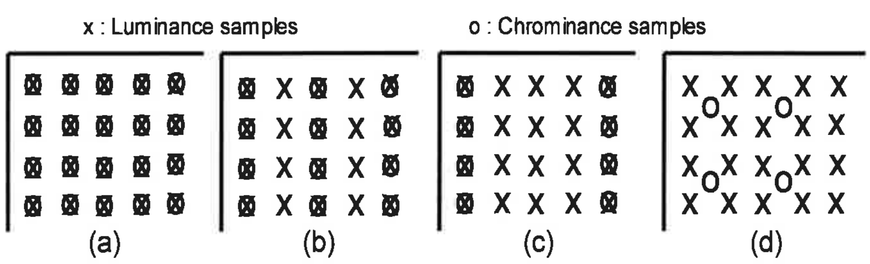

Figure 3: Luminance and chrominance samples in (a) 4:4:4 format (b) 4:2:2 format (c) 4:1:1 format (d) 4:2:0 format.

Converting analog TV signal to digital video with the 13.5 MHz sampling rate of CCIR601 results in 720 active pixels per line (576 active lines for PAL and 480 active lines for NTSC). This results in a 720x480 resolution for NTSC and a 720x576 resolution for PAL. With 4:2:2, the uncompressed bit-rate for transmitting CCIR601 at 30 frames/s is then about 166 Mbits/s. Since it is difficult to compress a CCIR601 video to 1.5 Mb/s with good video quality, in MPEG-1, typically the source video resolution is decimated to a quarter of the CCIR601 resolution by filtering and sub-sampling. The resultant format is called Source Input Format (SIF) which has a 360x240 resolution for NTSC and a 360x288 resolution for PAL. Since in the video coding algorithm, the block-size of 16x16 is used for motion compensated prediction, the number of pixels in both the horizontal and the vertical dimensions should be multiples of 16. Thus, the four left-most and right-most pixels are discarded to give a 352x240 resolution for NTSC systems (30 frames/s) and a 352x288 resolution for PAL systems (25 frames/s). The chrominance signals have half of the above resolutions in both the horizontal and vertical dimensions (4:2:0, 176x120 for NTSC and 176x144 for PAL). The uncompressed bit-rate for SIF (NTSC) at 30 frames/s is about 30.4 Mbits/s.

1.3.2 Group Of Pictures (GOP) and I-B-P Pictures

In MPEG, each video sequence is divided into one or more groups of pictures (GOPs). There are four types of pictures defined in MPEG-1: I-, P-, B-, and D-pictures of which the first three are shown in figure 4. Each GOP is composed of one or more pictures; one of these pictures must be an I-picture. Usually, the spacing between two anchor frames (I- or P-pictures) is referred to as M, and the spacing between two successive I-pictures is referred to as N. In Figure 4, M=3 and N=9.

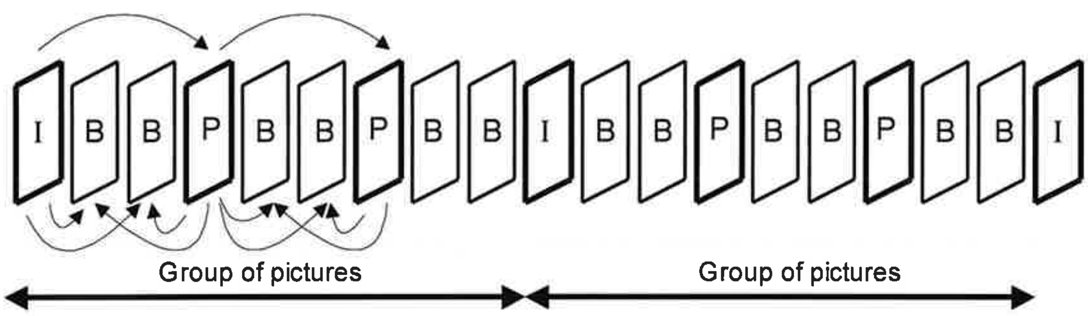

Figure 4: MPEG Group Of Pictures.

I-pictures (Intra-coded pictures) are coded independently with no reference to other pictures. I-pictures provide random access points in the compressed video data, since the I-pictures can be decoded independently without referencing to other pictures. With I-pictures, an MPEG bit-stream is more editable. Also, error propagation due to transmission errors in previous pictures will be terminated by an I-picture since the I-picture does not have a reference to the previous pictures. Since I-pictures use only transform coding without motion compensated predictive coding, it provides only moderate compression.

P-pictures (Predictive-coded pictures) are coded using the forward motion-compensated prediction similar to that in H.261 from the preceding I- or P-picture. P-pictures provide more compression than the I-pictures by virtue of motion-compensated prediction. They also serve as references for B-pictures and future P-pictures. Transmission errors in the I-pictures and P-pictures can propagate to the succeeding pictures since the I-pictures and P-pictures are used to predict the succeeding pictures.

B-pictures (Bi-directional-coded pictures) allow macroblocks to be coded using bi-directional motion-compensated prediction from both the past and future reference I- or P-pictures. In the B-pictures, each bi-directional motion-compensated macroblock can have two motion vectors: a forward motion vector which references to a best matching

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.

WHAT WILL YOU BUILD? | **sales@docketalarm.com** | **1-866-77-FASTCASE**

fastcase®
Smarter legal research.