

Digital
Multimedia
Standards
Series

DIGITAL VIDEO: AN INTRODUCTION TO MPEG-2

Barry G. Haskell,
Atul Puri, and
Arun N. Netravali

SNAP-1024
IPR2025-01375

Distributors for North, Central and South America:

Kluwer Academic Publishers

101 Philip Drive

Assinippi Park

Norwell, Massachusetts 02061 USA

Telephone (781) 871-6600

Fax (781) 871-6528

E-Mail <kluwer@wkap.com>

Distributors for all other countries:

Kluwer Academic Publishers Group

Distribution Centre

Post Office Box 322

3300 AH Dordrecht, THE NETHERLANDS

Telephone 31 78 6392 392

Fax 31 78 6546 474

E-Mail <orderdept@wkap.nl>



Electronic Services <<http://www.wkap.nl>>

Library of Congress Cataloging-in-Publication

Haskell, Barry G.

Digital video ; an introduction to MPEG-2 / Harry G. Haskell, Atul Puri, and
Arun N. Netravali.

p. cm.

Includes bibliographical references and index.

ISBN 0-412-08411-2

1. Digital video. 2. Video compression--Standards. 3. Coding theory

I. Puri, Atul. II. Netravali, Arun N. III. Title.

TK6680.5.H37 1996

96-14018

621.388'33--dc 20

CIP

British Library Cataloguing in Publication Data available

Cover design: Curtis Tow Graphics

Copyright © 1997 by Chapman & Hall

Seventh printing by Kluwer Academic Publishers 2000

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, mechanical, photo-copying, recording, or otherwise, without the prior written permission of the publisher, Kluwer Academic Publishers, 101 Philip Drive, Assinippi Park, Norwell, Massachusetts 02061

Printed on acid-free paper.

Printed in the United States of America

Motion Compensation Modes in MPEG

We saw in the last chapter that exploitation of temporal redundancy commonly uses motion compensation. After the removal of temporal redundancy, spatial redundancy is reduced by transformation to the frequency domain, for example, by the Discrete Cosine Transform (DCT), followed by quantization to force many of the transform coefficients to zero.

MPEG-1² was developed specifically for coding progressive scanned video in the range of 1 to 2 Mbits/s. MPEG-2⁴ aims at higher bitrates and contains all of the progressive coding features of MPEG-1. In addition, MPEG-2 has a number of techniques for coding interlaced video.

7.1 CODING OF PROGRESSIVE SCANNED VIDEO

The primary requirement of the MPEG video standards is that for a given bitrate they should achieve the highest possible quality of decoded video. As well as producing good picture quality under normal play conditions, some applications have additional requirements. For instance, multimedia applications may require the ability to randomly access and decode any single video picture*

in the bitstream. Also, the ability to perform a fast search of video in the bitstream, both forward and backward, is extremely desirable. It is also useful to be able to edit compressed bitstreams directly while maintaining decodability. Finally, a variety of video formats and image sizes should be supported.

Both spatial and temporal redundancy reduction are needed for the high compression requirements of MPEG. Some techniques used by MPEG were described in the previous chapter.

7.1.1 Exploiting Spatial Redundancy of Progressive Video

Because video is a sequence of still images, it is possible to achieve some compression using techniques similar to JPEG.¹ Such methods of compression are called *Intraframe* coding techniques, wherein each picture of the video is individually and independently compressed or encoded. Intraframe coding exploits the spatial redundancy that exists between adjacent pels of a picture. Pictures coded using only intraframe coding are called *I-pictures*.

As in JPEG, the MPEG intraframe video coding algorithms employ a Block-based two-dimensional DCT. An I-picture is first divided into 8×8

*Frames and pictures are synonymous in MPEG-1, whereas MPEG-2 has both frame-pictures and field-pictures.

Blocks of pels, and the two-dimensional DCT is then applied independently on each Block. This operation results in an 8×8 Block of DCT coefficients in which most of the energy in the original (pel) Block is typically concentrated in a few low-frequency coefficients. The coefficients are then quantized and transmitted as in JPEG.

7.1.2 Exploiting Temporal Redundancy of Progressive Video

Many of the interactive requirements can be satisfied by intraframe coding. However, for typical video signals the image quality that can be achieved by intraframe coding alone at the low bitrates desired is usually not sufficient.

Temporal redundancy results from a high degree of correlation between adjacent pictures. The H.261 and MPEG algorithms^{3,5-7} exploit this redundancy by computing an interframe difference signal called the *Prediction Error*. In computing the prediction error, the technique of motion compensation is employed to correct the prediction for

motion. As in H.261, the *Macroblock* (MB) approach is adopted for motion compensation in MPEG. In unidirectional motion estimation, called *Forward Prediction*, a *Target MB* in the picture to be encoded is matched with a set of displaced macroblocks of the same size in a past picture called the *Reference* picture. As in H.261, the Macroblock in the Reference picture that *best matches* the Target Macroblock is used as the *Prediction MB*.^{*} The prediction error is then computed as the difference between the Target Macroblock and the Prediction Macroblock. This operation is illustrated in Fig. 7.1.

The position of this best matching Prediction Macroblock is indicated by a *Motion Vector* (MV) that describes the horizontal and vertical displacement[†] from the Target MB to the Prediction MB. Unlike H.261, MPEG MVs are computed to the nearest half pel of displacement, using bilinear interpolation to obtain brightness values between pels. Also, using values of past transmitted MVs, a *Prediction Motion Vector* (PMV) is computed. The difference $MV - PMV$ is then encoded and transmitted along with the pel prediction error signal. Pictures coded using Forward Prediction are called *P-pictures*.

MOTION-COMPENSATED PREDICTION

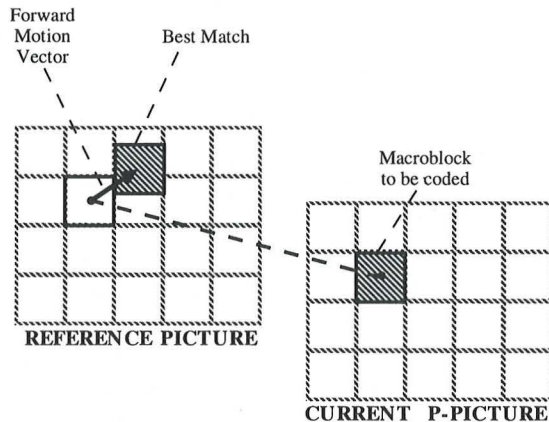


Fig. 7.1 Motion-Compensation using Unidirectional prediction. A displaced macroblock in a previous reference picture is used as the prediction of the moving-area macroblock in the current P-picture. The prediction error signal is then coded and transmitted.

^{*}Prediction MBs do not, in general, align with coded MB boundaries in the Reference picture.

[†]Positive horizontal and vertical displacements imply a rightward and downward shift, respectively, from the Target MB to the Prediction MB.

MOTION-COMPENSATED INTERPOLATION

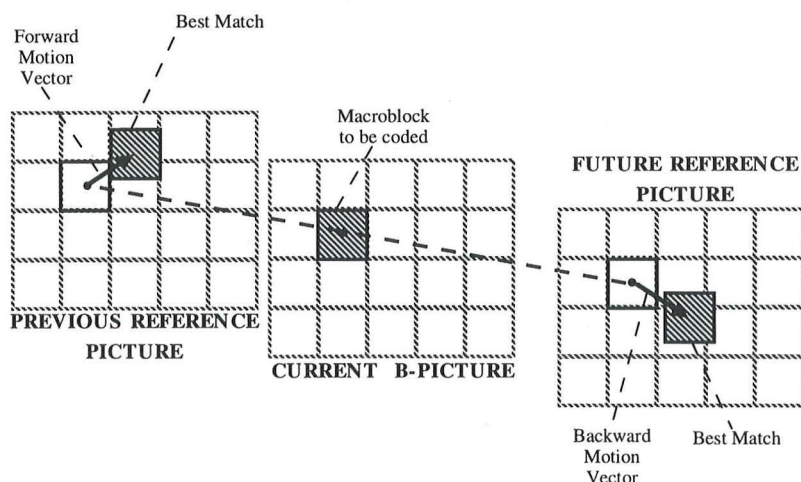


Fig. 7.2 Motion-Compensated Interpolation using Bidirectional prediction. A displaced macroblock in the previous picture is used as one prediction of the moving-area macroblock in the current B-picture. A displaced macroblock in the future picture is used as another prediction of the moving-area macroblock in the current B-picture. One, or the other, or an average of both can be used as the final prediction. The prediction error signal is then coded and transmitted.

The pel prediction error signal itself is transmitted using the DCT-based intraframe encoding technique as described in Chapter 6 and summarized above. In MPEG-1 video (as in H.621), the Macroblock size for motion compensation is chosen to be 16×16 luminance pels, representing a reasonable tradeoff between the compression provided by motion compensation and the cost associated with transmitting the motion vectors.

7.1.3 Bidirectional Temporal Prediction for Progressive Video

Bidirectional temporal prediction,^{2,3,7} also called Motion-Compensated Interpolation, is a key feature of MPEG video. Pictures coded with Bidirectional prediction use two Reference pictures, one in the past and one in the future. A Target Macroblock in bidirectionally coded pictures can be predicted by a Prediction Macroblock from the past Reference picture (*Forward Prediction*), or one from the future Reference picture (*Backward Prediction*), or by an average of two Prediction Macroblocks, one from each Reference picture (*Interpolation*). In every case, a Prediction Macroblock from

a Reference picture is associated with a motion vector, so that up to two motion vectors per MB may be used with Bidirectional prediction. Motion-Compensated Interpolation for a Macroblock in a Bidirectionally predicted picture is illustrated in Fig. 7.2.

Pictures coded using Bidirectional Prediction are called *B-pictures*. Pictures that are Bidirectionally predicted are never themselves used as Reference pictures, that is, Reference pictures for B-pictures must be either P-pictures or I-pictures. Similarly, Reference pictures for P-pictures must also be either P-pictures or I-pictures.

Bidirectional prediction provides a number of advantages. The primary one is that the compression obtained is often higher than can be obtained using Forward (unidirectional) prediction alone. To obtain the same subjective picture quality, Bidirectionally predicted pictures can often be encoded with fewer bits than pictures using only Forward prediction.

However, Bidirectional prediction does introduce extra delay in the encoding process, because pictures must be encoded and transmitted out of sequence as shown in Fig. 7.3. Further, it entails extra encoding complexity because Macroblock matching (the most

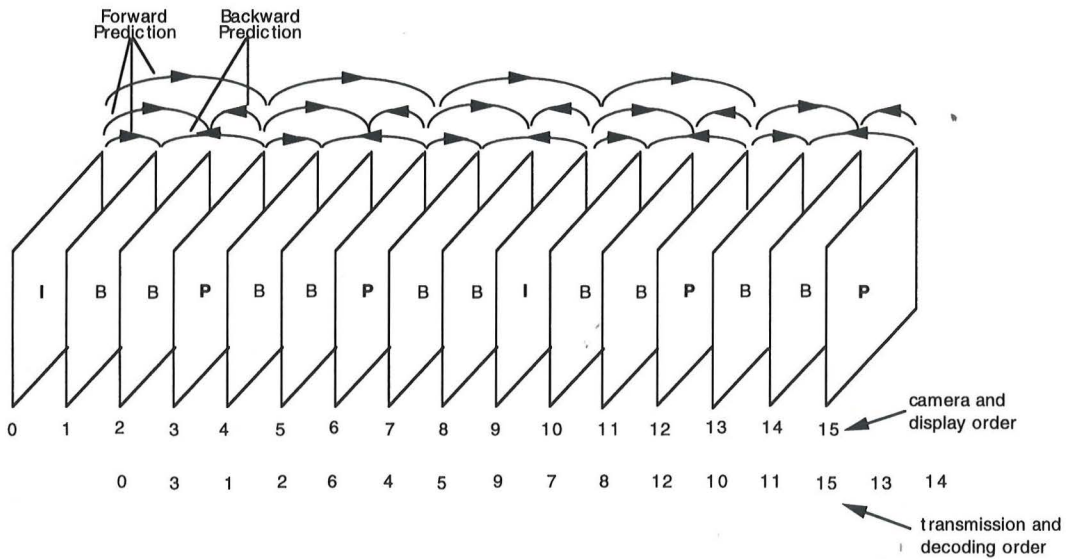


Fig. 7.3 With Bidirectional Prediction, the transmission order of the video frames is not the same as the Camera/Display Order. Note that the coded sequence starts with an I-picture followed by a P-picture.

computationally intensive encoding procedure) has to be performed twice for each Target Macroblock, once with the past Reference picture and once with the future Reference picture.

Bidirectional prediction is most useful in normal scenes containing slow to moderate motion or during camera panning over a relatively stationary background. It is less useful if there are numerous and frequent scene changes or rapid, erratic motion of large areas of the picture. For this type of material, where coding artifacts are much more difficult for viewers to detect, P-pictures by themselves offer a reasonable tradeoff between image quality and updating the picture as quickly as possible.

7.2 CODING OF INTERLACED VIDEO

The original objective of MPEG-2 was to code interlaced CCIR* 601 video at a bitrate that would serve a large number of consumer applications⁶. In

fact, one of the main differences between MPEG-1 and MPEG-2 is that MPEG-2 handles interlace efficiently.^{4,5,8-10} Since the picture resolution of CCIR 601 is approximately four times that of the SIF of MPEG-1, the bitrate chosen for MPEG-2 optimization was 4 Mbits/s. However, MPEG-2 allows rates as high as 429 Gbits/s.

For interlace, MPEG-2 provides a choice of two *Picture Structures*, as shown in Fig. 7.4. *Field-pictures* consist of individual fields that are each divided into macroblocks and coded separately.[†] With *Frame-pictures*, on the other hand, each interlaced field pair is interleaved together into a frame that is then divided into macroblocks and coded. MPEG-2 requires interlaced video to be displayed as alternate top and bottom fields.[‡] However, within a frame either the top or bottom field can be temporally first and sent as the first Field-picture of the frame.[§]

MPEG-2 also allows for progressive coded pictures, but interlaced display. In fact, coding of a 24 frame/s progressive film source with 3:2 pulldown interlaced display (see Section 5.6.3) is also supported.

*CCIR has been renamed to International Telecommunications Union—Radio Sector (ITU-R).

[†]Each pair of fields still belongs to a unique frame, however.

[‡]The top field contains the top line of the frame. The bottom field contains the second (and bottom) line of the frame.

[§]If the temporal ordering of the fields in a frame differs from the previous frame, due perhaps to editing, then typically one field must be displayed twice to maintain the top/bottom alternation. See **repeat_first_field** in Chapter 10.