VIDEO FORMATS AND MPEG COMPRESSION

N. C. State University

CSC557 ♦ Multimedia Computing and Networking

Fall 2001

Lecture # 15

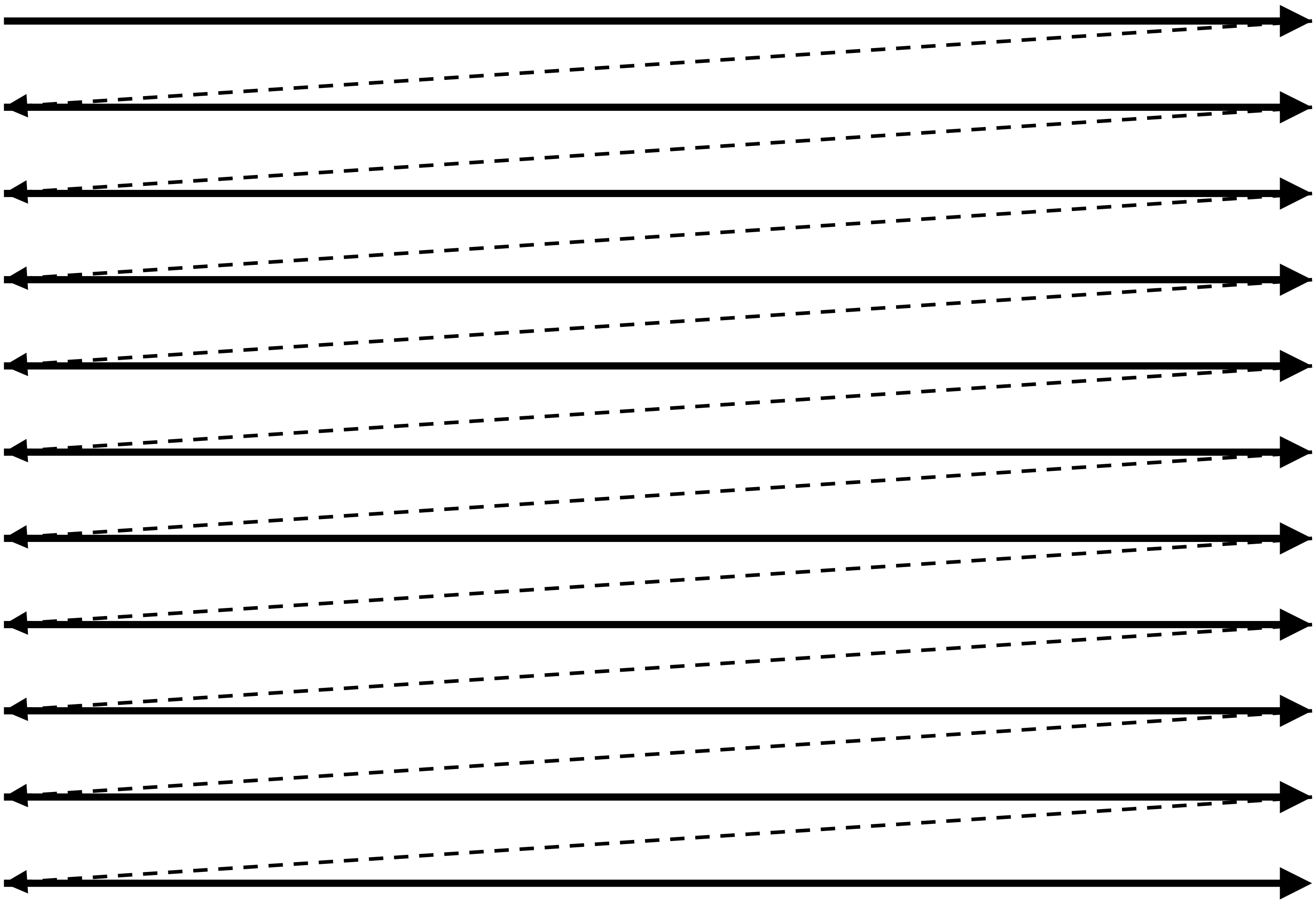
Video Standards

- BroadcastTV (analog)
- VCR (analog)
- Film (analog)
- Computer video
- Digital TV, HDTV
- Digital (compressed) video

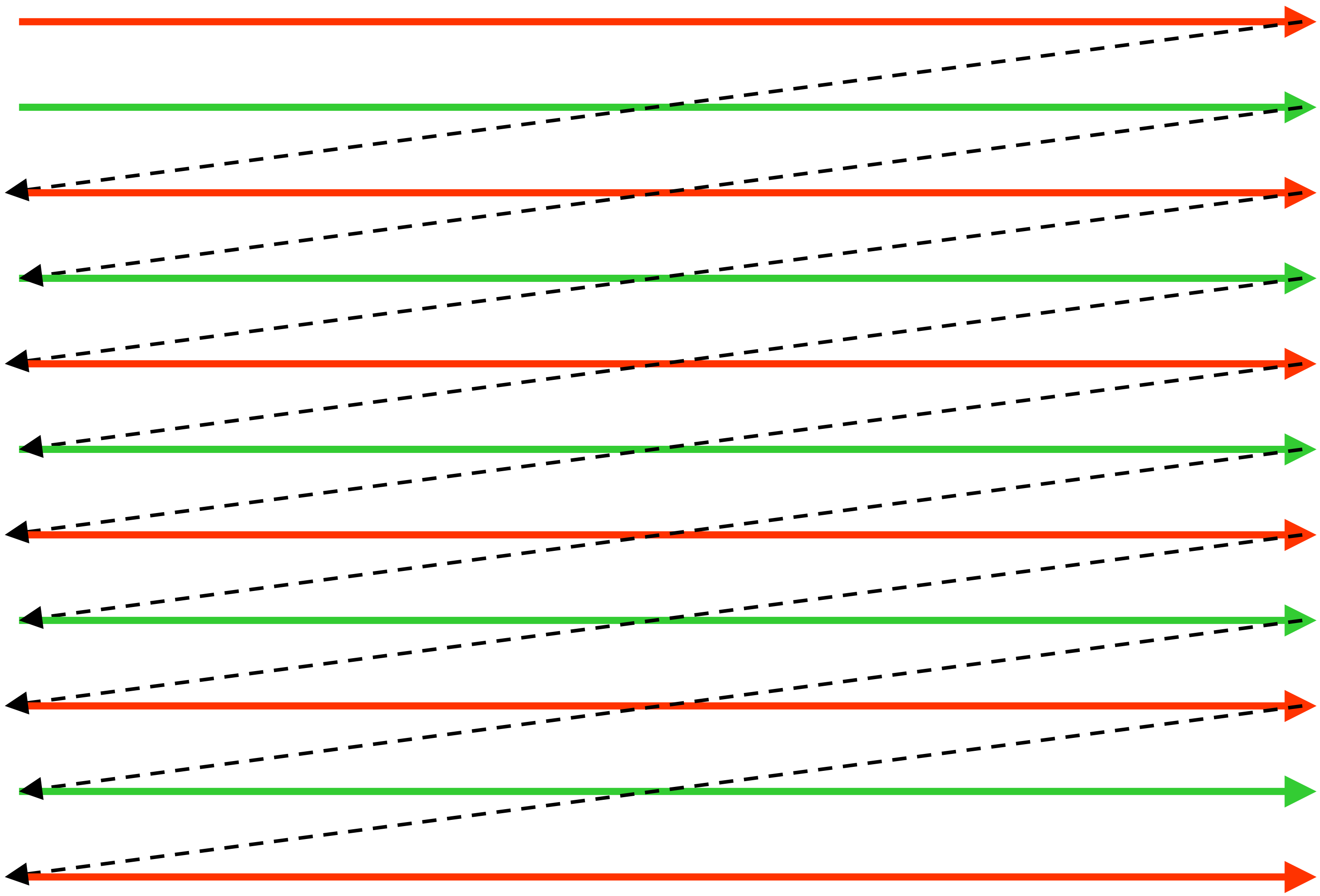
Frame Rates

- Sampling rates must be high enough to avoid motion "aliasing"
  - At least 50 frames/sec needed in the ideal case
  - Few video standards support this high a rate

Raster Scanning Order

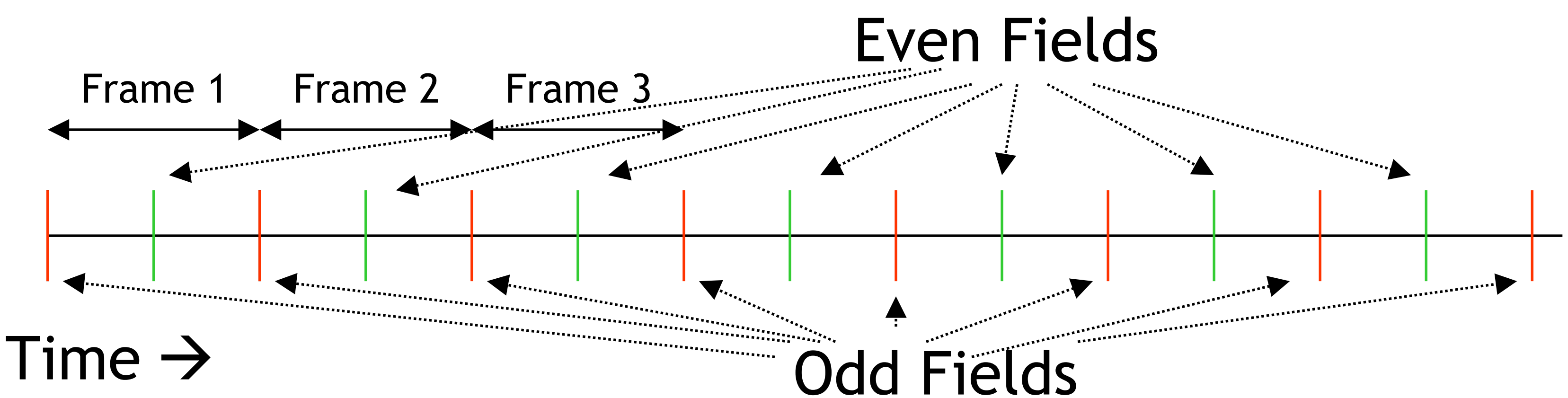


Interlaced Video



Interlaced Video

- Display of alternating, interleaved "fields" at 2X the frame rate
  - The net bandwidth, or number of bits being transmitted, remains the same
  - The fields are taken at half-frame intervals; vertical lines don't line up exactly!



- "Fools" the eye into thinking the sampling rate is greater than the frame rate

Video Standards

Standard	Digital or Analog	Color	VxH Resolution	Aspect Ratio	Frame Rate (fps)	Inter-laced?	Comments
TV (NTSC standard)	analog	YIQ	525 lines (40 not displayed)	4x3 (1.33)	29.97	yes	Used in North America and Japan
TV (PAL standard)	analog	YUV	625 lines		25	Yes	Used in Europe
VHS	Analog		240 lines				
Film (motion pictures)	Analog	Subtractive		1.85	24		
VGA (computer video)	Digital	RGB	640x480—1600x1200	4:3 (1.33)	(70-80 fps)	No	
CCIR-601 (Digital Component Video): SIF	Digital	YCbCr	352x240		29.97	No	ITU-R standard, many resolutions supported. SIF is typical size used in MPEG-I
HDTV	Digital		1920x1080	16x9 (1.78)	30	Yes	Many competing proposals exist.

Issues in Converting Between Standards

- Differences in pixel aspect ratios (square vs. non-square)
- Difference in screen resolution and aspect ratio
- Interlaced vs. Non-interlaced
- Differences in frame rates
  - E.g., speedup up film by 4% for 25 fps TV
  - E.g., use 3:2 pulldown to convert 2 frames at 24 fps into 5 fields (= 2.5 frames) at 30 fps
- Differences in color systems
- Differences in chroma sub-sampling

Video Data Requirements (example)

- 30 frames / second *
  640x480 pixels / frame (VGA) *
  3 bytes / pixel
  = 26.4 MB/second! (211 Mb/s)

- Or, 92 GB for a one hour video!

- Many computer components are too slow/small for this

MPEG Compression Standard

- An international (open) standard
  - MPEG = Motion Picture Expert Group
- Best quality / compression combination
- Assymetrical
  - not too hard to decode: software
  - encoding requires a lot of computation: hardware
- MPEG standard specifies syntax and semantics of the bit stream
  - *not* the method of compression!

Mpeg Video Standards

- MPEG-I
  - Targeted for 1.2-1.5 Mb/s
  - 352x240, 30 frames / second
  - "VHS quality"
- MPEG-II
  - Targetted for 4-15 Mb/s
  - Up to 1920x1152, 30 frames/sec
  - Interlaced video
  - "TV and HDTV quality"
- MPEG-IV
  - Very low bandwidth: 10-64 kb/sec
  - "Videoconference quality"

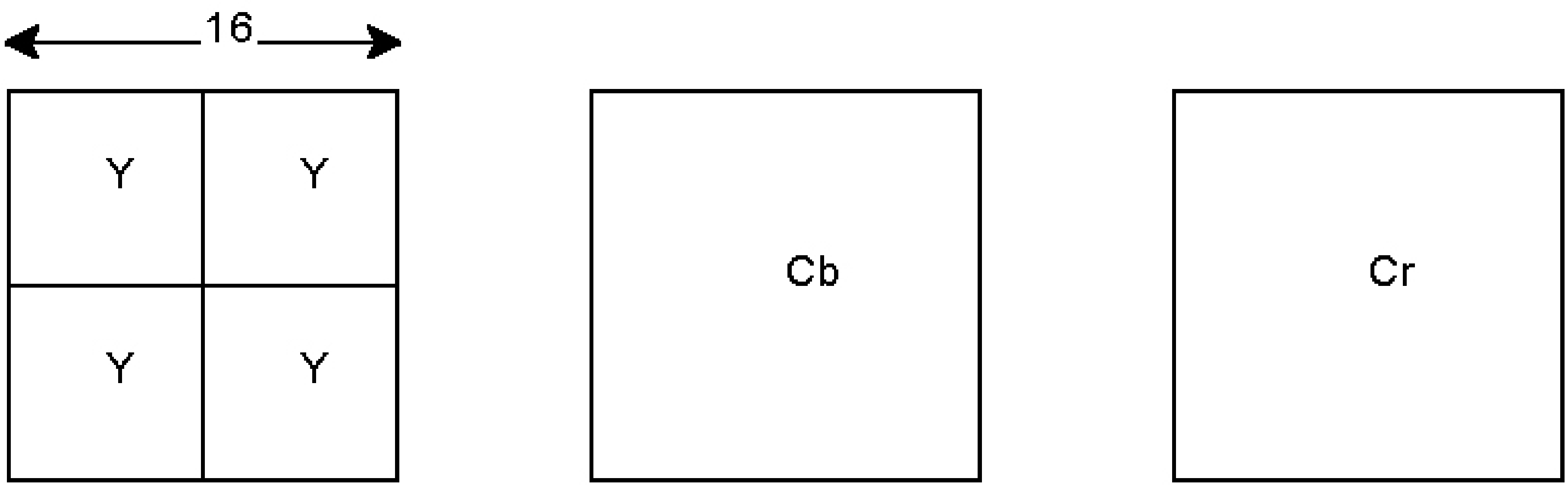
Basic Idea

- Use spatial compression on some frames (I.e., JPEG)
- Add temporal compression ("motion compensation") as well
  - Improves compression ratio by 2-4X

Macro Blocks

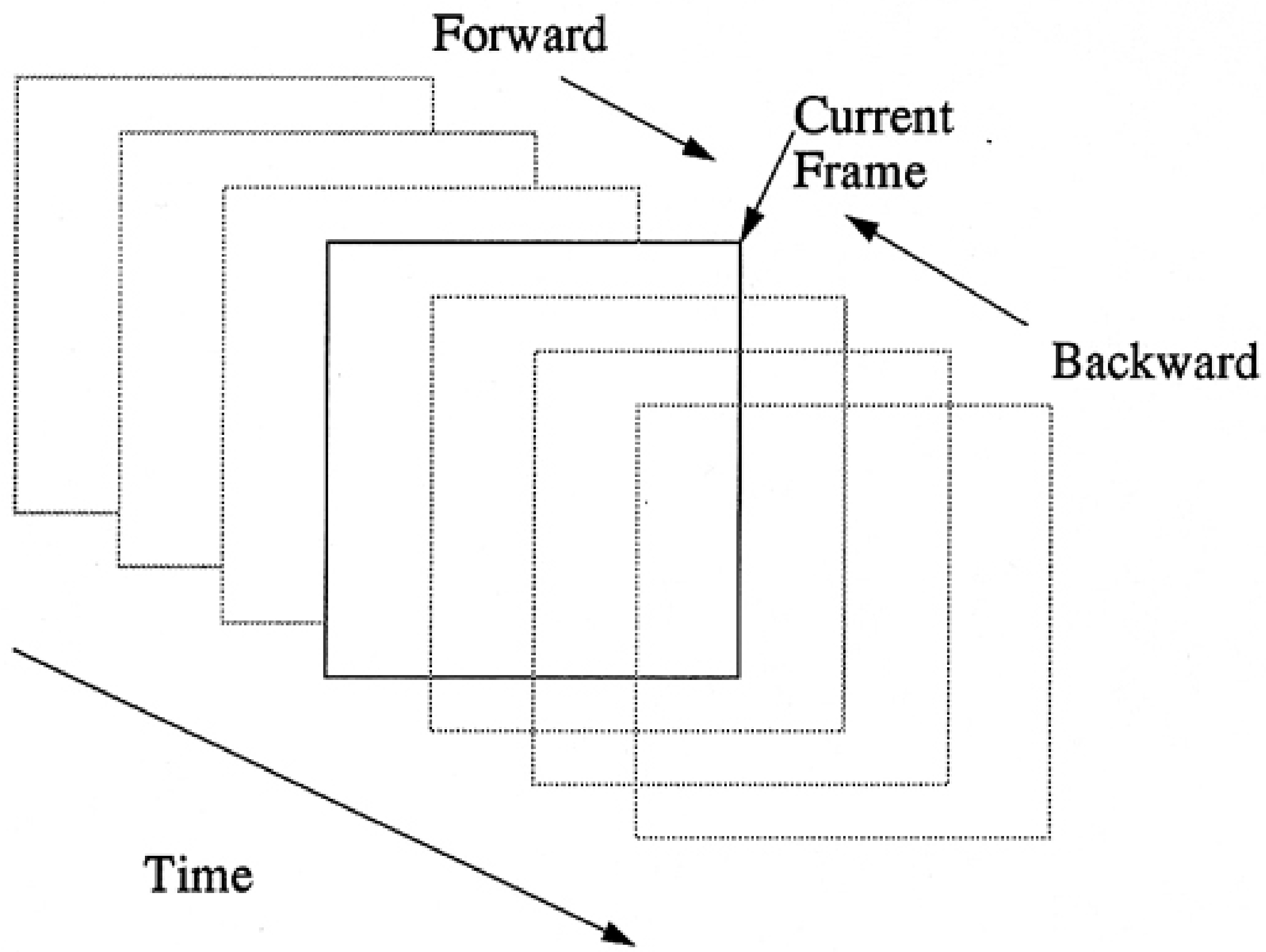
- Macro block = 16x16 pixel region
  - 4 8x8 blocks of Y
  - 1 8x8 block (subsampled) of Cb
  - 1 8x8 block (subsampled) of Cr
- Macro block is basic unit of temporal coding

Macroblocks



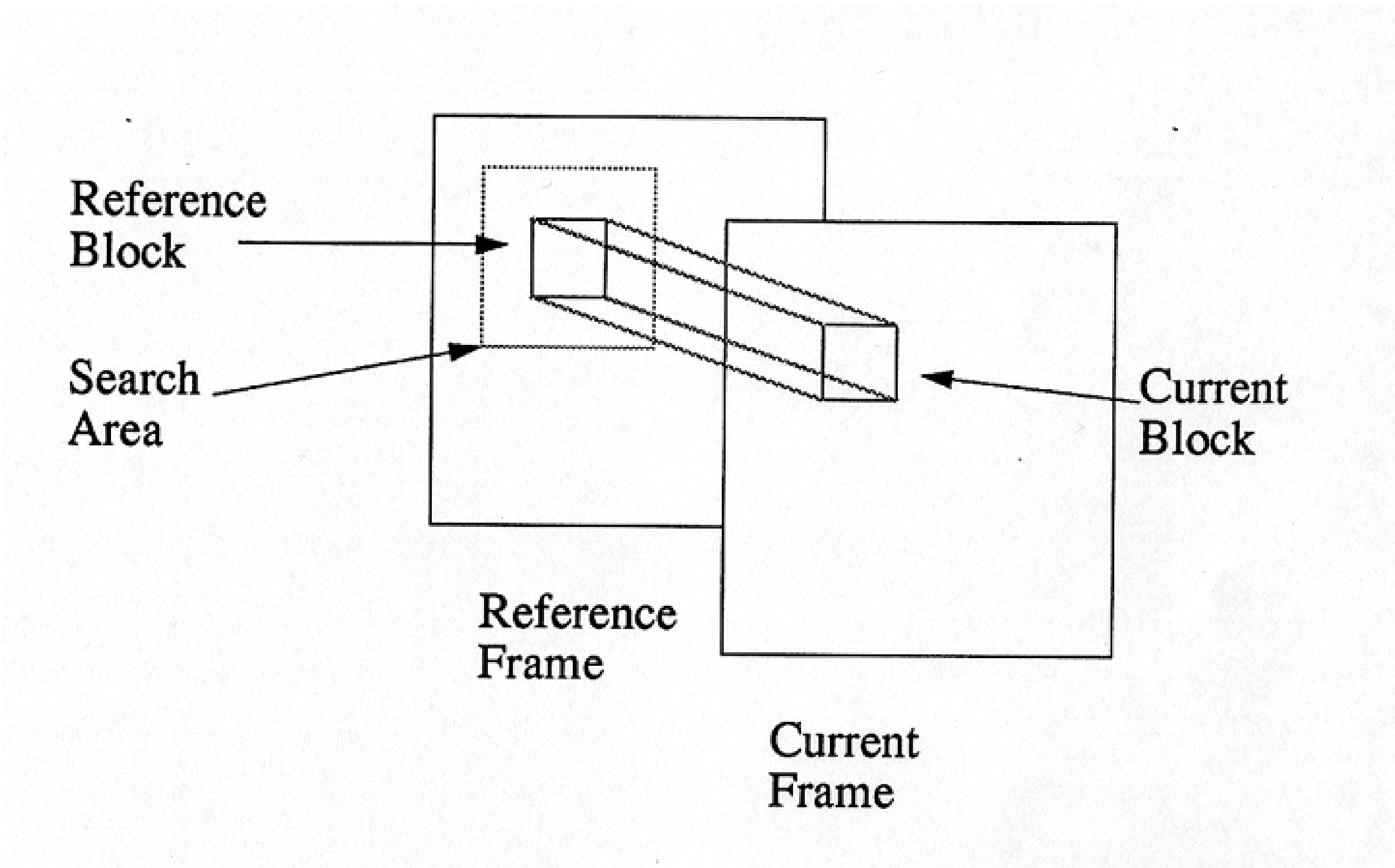
Temporal Encoding

- Search forward or backward (or both) in time for a "similar" macro block
- Interpolated match: average of forward and backwards block


Forward
Current
Frame
Backward
Time

Temporal Encoding

- Search a spatial region for a similar macro block



Searching For Similar Macroblocks

- Search method not specified by standard
- Can search on full pixel or half-pixel boundaries
  - Interpolated images are effectively low-pass filtered
- How define "best" match?
  - Minimize absolute difference =

$$\sum_{i=0}^{15}\sum_{j=0}^{15}|V_n(x+i,y+j)-V_m(x+\Delta x+i,y+\Delta y+j)|$$

  - Where V_n is macroblock, V_m is matching macroblock
  - (V_m is the compressed/decompressed representation of the matching macroblock, I.e., the prediction algorithm is feedforward)

Size Of Search Region

- How big is big enough?
- More separation in time means greater amount of motion possible

Motion Vector Encoding

- Motion vector = ΔX, Δ Y (I.e., offset or displacement) to best match
  - Sign is always relative to the current picture
- Number of vectors needed
  - 1 for forward or backwards predictions
  - 2 for bi-directional (interpolated) prediction
- Then, differentially encoded relative to motion vectors of the previous (adjacent) macroblock
- Entropy code the resulting differentially-coded vectors

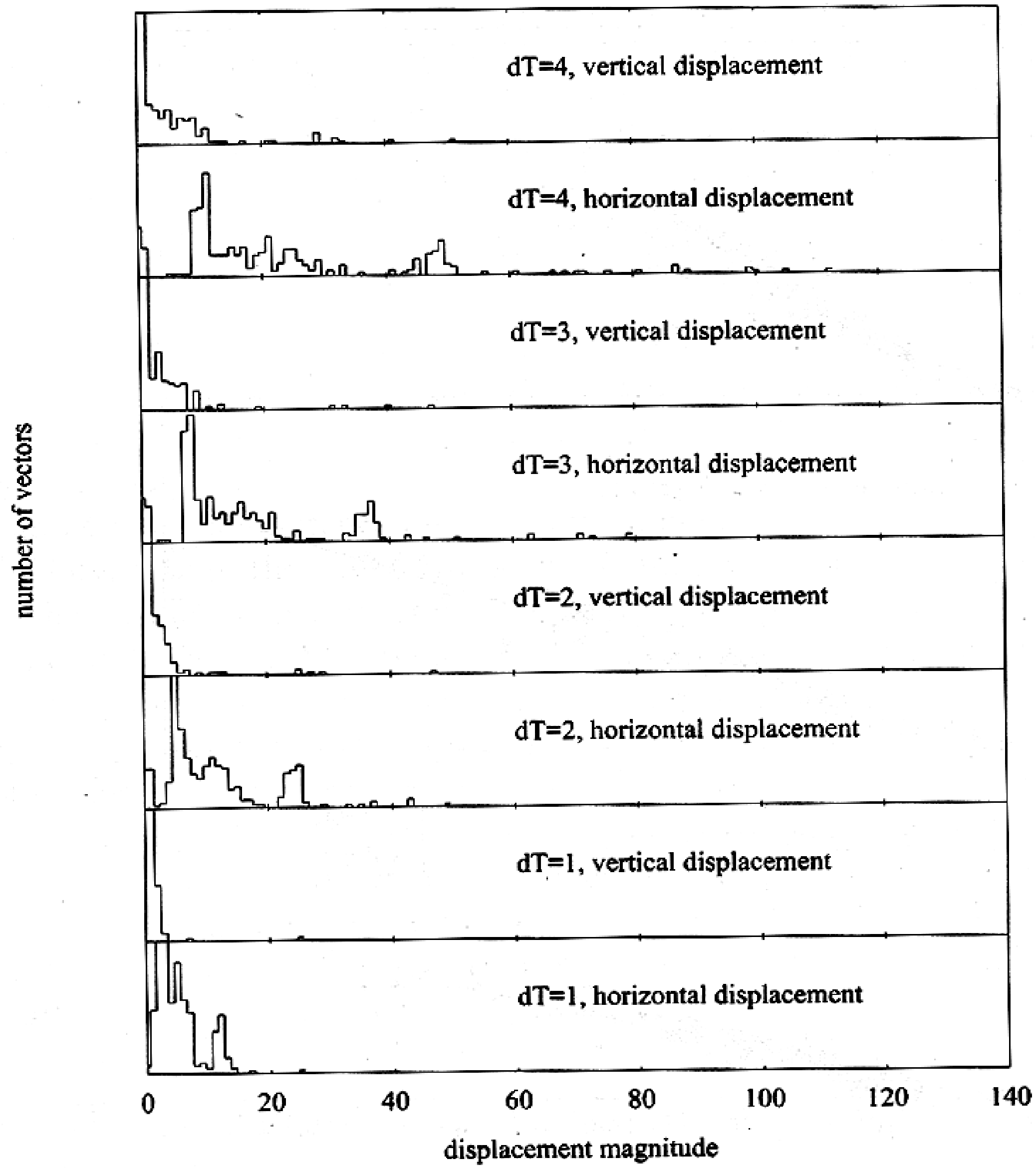


Figure 13.3: Histograms of motion displacement magnitudes for the flower garden sequence for four temporal distances, dT. The displacement unit is half-pel. Some counts near zero displacement are off-scale.

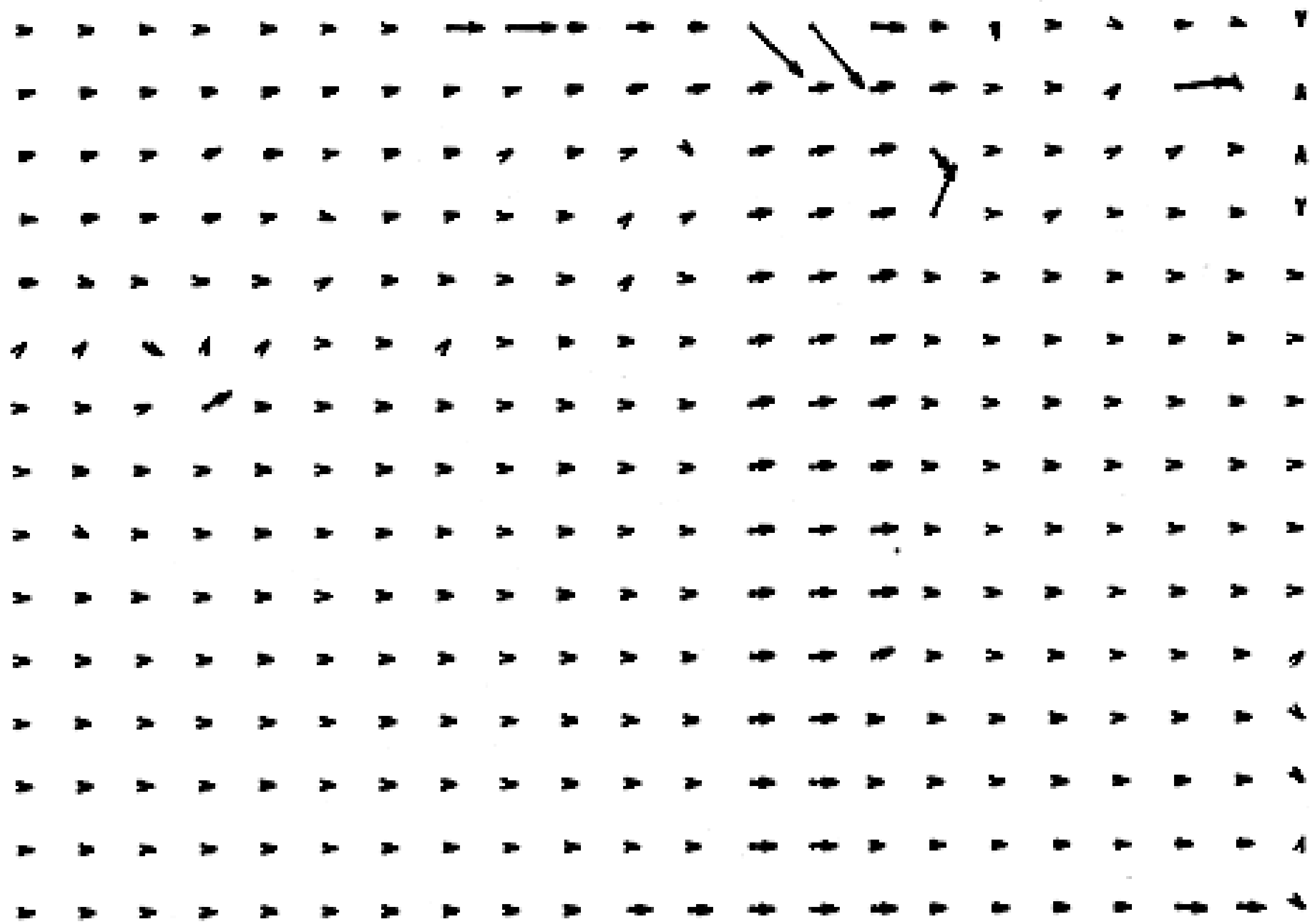


Figure 13.8: Motion vectors for flower garden picture 3 for prediction from 2.

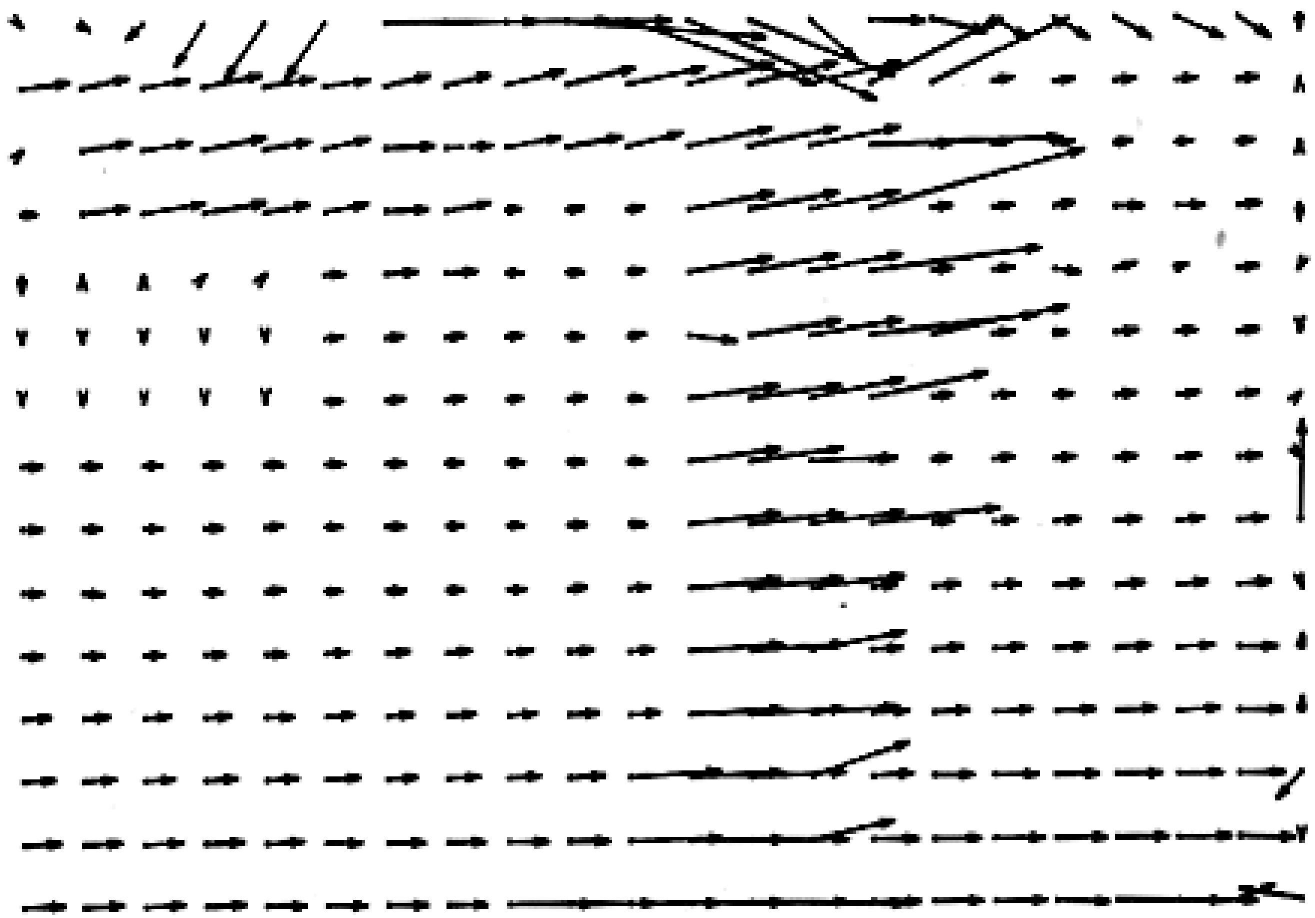


Figure 13.11: Motion vectors for flower garden picture 6 for prediction from 2, for a cost slope of 0.137. Figure 13.9 shows the same search without a cost function.

Effects Of Scene Changes

- Temporal compression not nearly as effective after scene changes

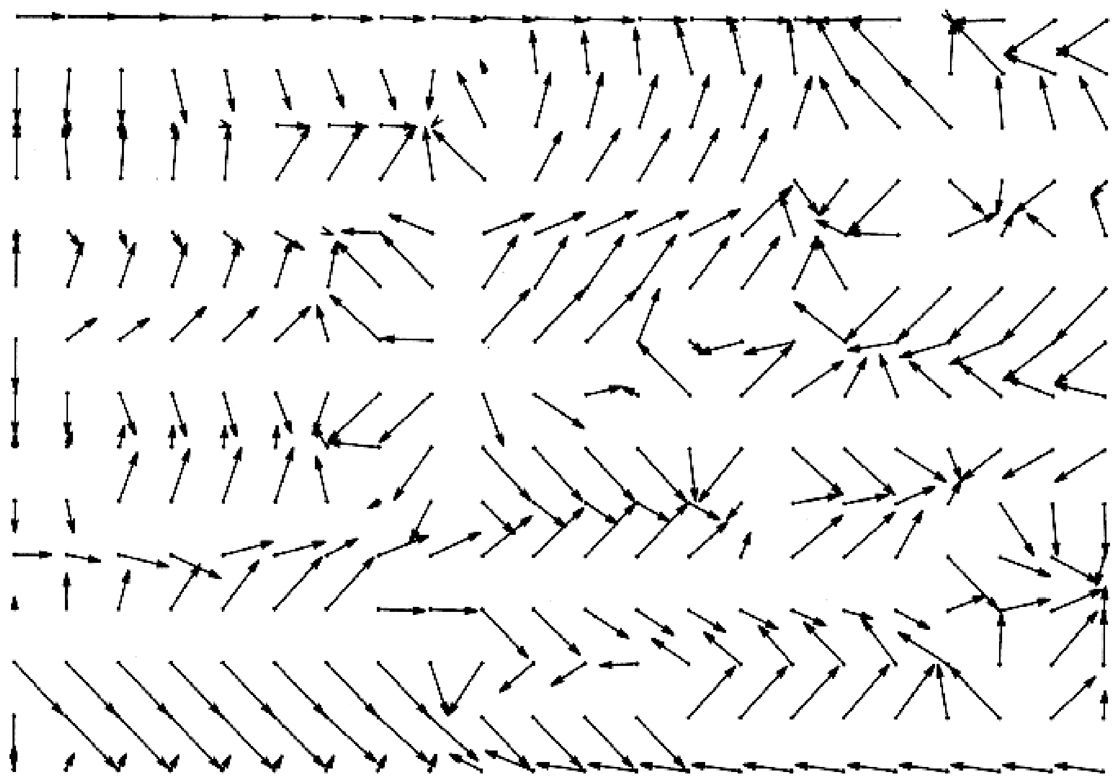


Figure 13.13: Motion vectors for picture 66 of the table tennis sequence for prediction from 67. A scene change occurs between pictures 66 and 67.

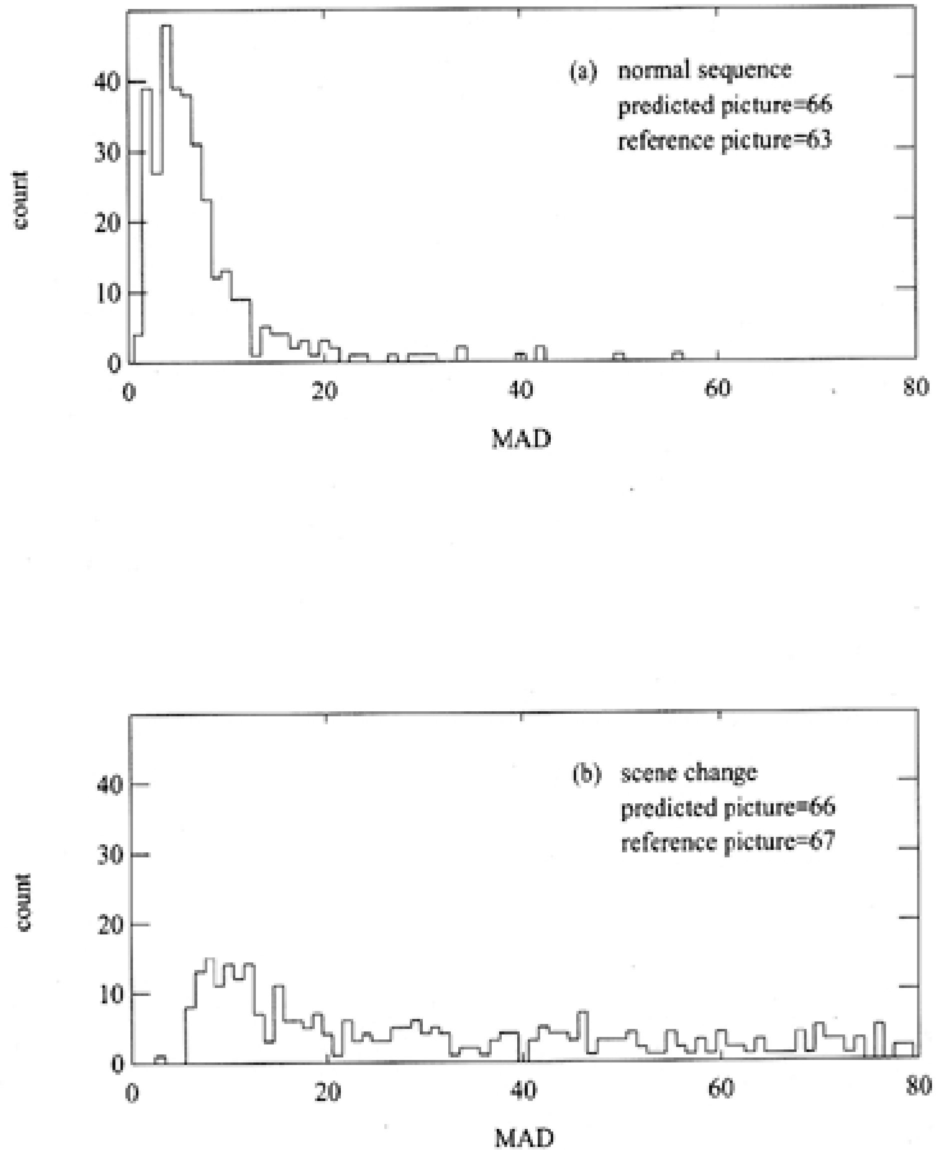


Figure 13.12: Histograms of MAD values for pictures from the table-tennis sequence. A scene change occurs between pictures 66 and 67.

Temporal Results

- If closest match is "close enough", just output the motion vector
- If not close enough, then spatially code (JPEG) all 6 blocks (I.e., no temporal compression possible)
- But if “almost close enough”…

"Almost" Close Enough

- Output motion vector
- Create 6-bit block code
- For each block...
  - Subtract from the reference (matched) block
  - Spatially compress the difference block
  - If the result is completely zero, clear the block bit and skip the block
  - Otherwise, set the block bit and output the (spatially compressed) difference block

Quantization

- Can tolerate higher quantization (I.e., greater loss of quality) for difference blocks than for I blocks
- Possible to adapt the quantizer scale to video content
  - Q factor can be part of the macroblock
- Default quantization matrices for predicted and non-predicted blocks
  - Can be overridden for entire video, or frame-by-frame

Default "non-predicted block" quantization matrix

1	16	19	22	26	27	29	34
16	16	22	24	27	29	34	37
19	22	26	27	29	34	34	38
22	22	26	27	29	34	37	40
22	26	27	29	32	35	40	48
26	27	29	32	35	40	48	58
26	27	29	34	38	46	56	69
27	29	35	38	46	56	69	83

Default "predicted block" quantization matrix

16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16

Run-length and Huffman Coding

- Zigzag ordering: like JPEG
- Run-length coding: like JPEG
- Huffman coding: just encode a subset of the possible inputs, consisting of the most probable symbols
- "Escape" sequence used to represent remaining (infrequent) symbols in unencoded form

Types Of Macroblocks

- I macro blocks
  - Spatial information only
- Temporal macro blocks
  - Motion vector(s)
  - Block code
  - Spatially-coded difference blocks
- "Skip" macro block (no change)

Types Of Frames

- I frames
  - Spatial macro blocks only
  - Serves as "anchor" frame
- P frames
  - Spatial, forward motion, and skip macro blocks
  - Can refer backwards to previous I and P frames
  - Can serve as "anchor" frame

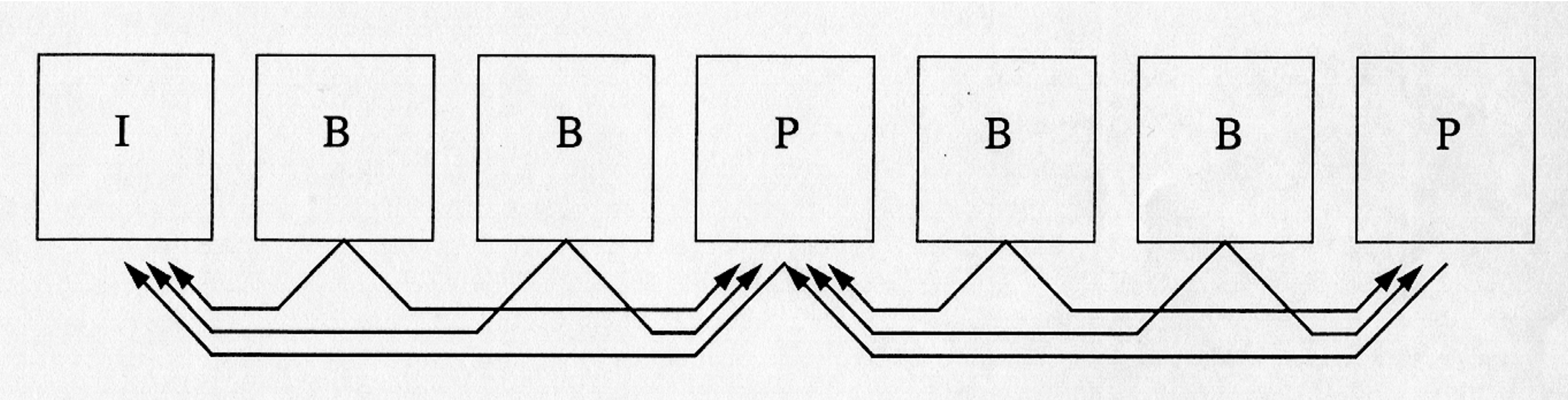
Frames (Cont.)

- B frames
  - Spatial, forward/backward/interpolated motion, and skip macro blocks
  - Can refer backwards or forwards to I and P frames
  - Never used as "anchor" frame
  - Depends on a "future" frame which must be received and decoded first
- "Anchor" frames are transmitted before the frames that reference them
  - Reordering required at decoder
- "Typical" size ratios of I:P:B frames = 5:3:1

Group Of Pictures (GOP)

- Repetitive Sequence of Frame Types
  - (IB^m(PB^m)^n)^o
- Examples
  - I I I ....
  - IPP IPP IPP...
  - common: IBBPBBPBB IBBPBBPBB IBBPBBPBB ...

GOP Sequence (example)



GOP Tradeoffs

- How much compression desired?
  - More anchor frames = larger file size
- How much buffer space required?
  - More anchor frames = less reordering buffer space needed
- How much compression / decode delay is tolerable?
  - More anchor frames = less delay
- What degree of random access desired?
  - More anchor frames = easier and faster random access, with less memory required
- How much error recovery is needed?
  - More anchor frames = less sensitive to errors

MPEG Systems

- The size of the decoder buffer is specified
  - An encoder constraint is that the decoder buffers must never overflow or underflow

MPEG Layers

- Sequence: picture size, picture rate, buffer sizes, quantization matrices, GOP pattern
- GOP: SMPTE time code
- Picture (frame): frame type, motion vector range for frame
- Slice: unit of resynchronization, collection of consecutive macroblocks preceded by resynchronization code
- Macroblock: unit of motion compensation (16x16)
- Block: unit of spatial coding (8x8)

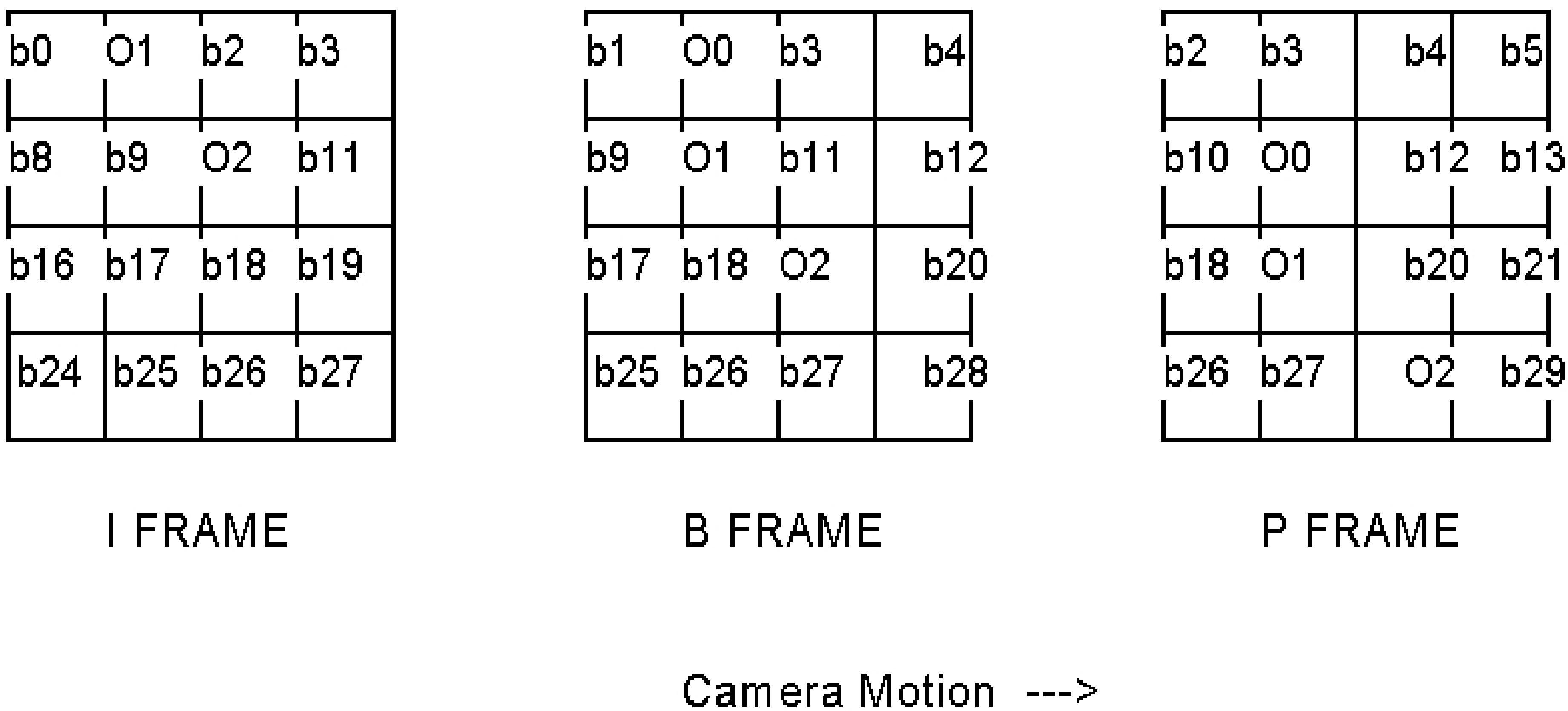
Levels and Profiles

- MPEG I "constrained" parameters (reduce possible combinations):
  - # columns <= 720, #rows <= 576
  - # macroblocks/frame <= 396
  - # macroblocks / sec <= 396*25 = 330 * 30
  - Frames rate <= 30 / sec
  - Bit rate <= 1.86 Mb/s
  - Decoder buffer <= 368 Kb
- MPEG-2
  - Level = resolution
  - Profile = features used

Levels and Profiles (cont.)

- Levels
  - Low: 352H x 288V
  - Main: either 720H x 576V x 25 fps, or 720H x 480v x 29.97 fps
  - High: up to 1920H x 1152V
- Profiles
  - Simple: no B pictures allowed
  - Main: features we have discussed
  - SNR: layered coder (low quality picture and quality enhancement layer both encoded)
  - Spatial: low resolution and resolution enhancement layer both encoded
  - High: all of the above, plus a low frame rate and frame rate enhancement layer both encoded

Compression Example



Exhaustive Strategy

- Check every possible match within some spatial region

Search Strategies

- Not part of the standard, but the most computationally expensive part of compression
- One possibility (expensive): exhaustive search
- Example: offset from -15 to +15 in X and Y directions
  - Approximately 3600 half-pixel possible matches
  - 3600 * 256 subtracts / match * # predicted macroblocks / frame * # predicted frames / sec = around 7 billion operations / second, just for matching
  - More for bidirectionally-predicted!

Pyramid Search

- Form lower and lower resolution versions of the frames to match and be matched, using low-pass filtering and subsampling
- Compare lowest-resolution versions first to find best match
- Then, do search within this region at a higher resolution to find best match
- etc., until regions are 16x16 macroblocks and best match is found

Logarithmic Search

- One method: search corners and center of a diagonal square
  - If a corner is the best match, shift diagonal square to be centered on that, and repeat
  - If the center is the best match, do a local exhaustive search around the center
- Another method: search boundary and center of a large square
  - If a boundary point is the best match, shift the square to center on that point, reduce the square size, and repeat
  - If the center is the best match, do a local exhaustive search

Decompression

- Much faster; No searching or matching required
- If motion vector present
  - Use tables to undo entropy coding of motion vector
  - Undo differential encoding of motion vector
  - Read reference block(s) from buffer
- If quantizer present
  - Scale the quantization matrix

Decompression (Cont.)

- If block code present
  - Use tables to entropy decode coefficients
  - Undo RLE
  - Undo zigzag
  - Undo quantization
  - Undo DCT
- Combine differential and reference blocks

Decompression (Cont.)

- Combine 6 blocks into a macro block (with unsubsampling)
- Combine macroblocks into an image
- Convert YCbCr to RGB for display

H.261

- Standard for videoconferencing
- Much lower bit rates / quality than MPEG
- Also called px64, p = 1..30
- YCbCr color coding
- 288 lines by 352 pixels (CIF)or 144 by 176 (QCIF)

H.261 (cont.)

- Macro blocks like MPEG
- Intraframe coding like JPEG
- Interframe coding like P frames from MPEG

Sources Of Info

- [Gibson98] Digital Compression for Multimedia
- [Mitchell97] MPEG Video Compression Standard
- [Poynton96] A Technical Introduction to Digital Video
- [Fluckiger95] Understanding networked multimedia : applications and technology