Multimedia Content

Streaming Media

section 7.4

27 Feb 2012

Topics

- Physics
  - 7.4.1
  - 7.4.2
  - compression
- Protocols
  - 7.4.3-7.4.5
  - stored media (multiple encodings)
  - live media
  - real time conf

Download
vs
Stream

Definitions?

Challenges?

Advantages?

Network Requirements

- What does an application need from the network?
- Data Loss
- Bandwidth
- Time-sensitivity
- email/web/IM/p2p/DNS/streaming media

- Aside from the Internet, where do we stream multimedia content?
  - how does that work?
  - how is the Internet different?

Your media player

- How does your media player use RTSP and RTP to play music?
  - UI
  - handle transmission errors (must fight loss of data)
  - decompress content
  - must fight jitter

protocol/server

➔fight congestion

➔sample responses

➔change bit rate

➔degrade quality

Streaming Audio

- delays smaller than 150 milliseconds are not noticed
- delays of 400+ milliseconds are annoying
  - ever see a TV news reporter in Iraq interviewed by a news anchor in NYC?
  - human ear is sensitive to short variations (milliseconds)
- We don’t need to get transmission perfect, the human ear/brain is good at interpreting a noisy signal
- Human ear can distinguish sounds in the 20 Hz to 20,000 Hz range
- Loudness is logarithmic

Streaming audio

- TCP vs UDP
- Gaps in data are ok, if they are very small
- How do you expect data to be lost in the network?

Advantages to TCP
  Get complete data*
  Run via unblock port 80
  buffer space is cheap

*streaming stored media

Filling Gaps

- Skip a frame
- Stretch out sounds / interpolate

- Forward Error Correction
  - erasure
  - Parity Packet

- Interleaving
  - even/odd samples in separate packets
  - more tricky with compression

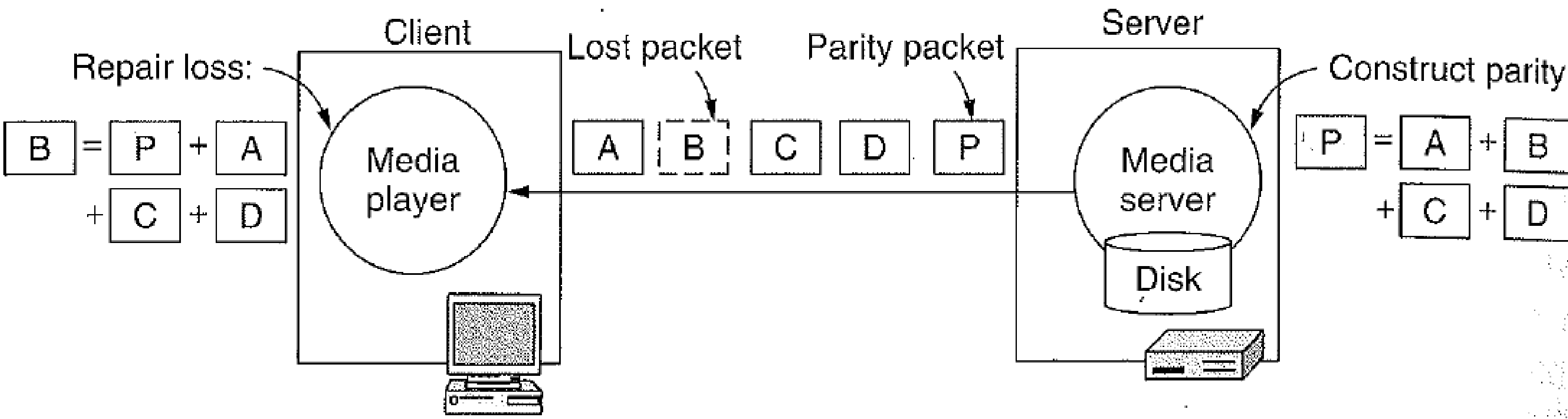


Figure 7-52. Using a parity packet to repair loss.

Combat Jitter

- Buffering....
- Low/high water mark
- RTSP - control
- RTP - data
  - UDP or TCP
  - sometimes HTTP over TCP

Streaming Live Audio/Video

- Buffering challenges?
- Audience challenges?
  - multicast
  - often, just normal TCP connection
- RTP over UDP
- Big operations use a Content Distribution Network (CDN)

Multicast (RFC3170)!

- Like (smart) broadcast TV for the Internet
- This is really in the transport layer
  - it runs on top of IP
  - we will talk about specific multicast routing algorithms later
- Send to a subset of the network
  - use an overlay network
  - uses *address indirection*
  - single IP address is used to represent all the receivers
    - multicast group
  - class D multicast address
    - class D: 224.0.0.0 to 239.255.255.255
    - left most bits are 1110
  - IGMP: Internet Group Management Protocol (RFC 4604)

Not available across networks/ service providers

Streaming Audio

- Often have a data channel and a control channel
  - why is that?
- RTP (RFC1889/3550) (chapter 6 pg. 546 in your book)
  - Real-time *transport* protocol
  - built on top of UDP
  - **data** channel
  - "to carry data that has real-time properties." RFC 3550
- RTCP (RFC 1889/3550)
  - sends control/statistical information for a RTP stream
  - "to monitor the quality of service and to convey information about the participants in an on-going session." RFC 3550

RTP

- Generic, real-time transport protocol
- http://www.cs.columbia.edu/~hgs/rtp/
- Media applications use this like MathPacket uses **`receive`**()/**`send`**()
- How does this fit into the protocol stack?
- Can combine many media streams into one RTP stream
  - audio and video may combine into one stream

- Android
- http://developer.android.com/guide/appendix/media-formats.html
  - you can launch an intent.....

RTP

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|V=2|P|X|  CC   |M|     PT      |       sequence number         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                           timestamp                           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           synchronization source (SSRC) identifier            |
+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+
|            contributing source (CSRC) identifiers             |
|                             ....                              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

– http://www.ietf.org/rfc/rfc1889.txt

RTSP

- RTSP (RFC 2326)
  - Real Time Streaming Protocol
    - "Internet VCR remote control protocol"
      - pause, fast forward, reverse, and absolute positioning
    - modeled on HTTP
  - control channel
    - often called an “out-of-band” protocol
  - RTSP does not specify the data (media) packet
  - may use RTP or RDT as data channel
  - RDT (RealNetworks *proprietary* transport protocol)
    - released under the RealNetworks Community Source License
    - https://protocol.helixcommunity.org (defunct?)
  - http://www.cs.columbia.edu/~hgs/rtsp/

http://www.cs.columbia.edu/~hgs/rtsp/faq.html#rtp_rtcp_rtsp

RTSP (in-depth)

- Has notion of session built into the protocol
- Maintains state
- Control may be sent over multiple *TCP* connections
  - why would this be a TCP connection?
  - it is possible to embed the **data** into an RTSP channel
    - why would we want to do this?
    - why would this be a bad idea?

RTSP Control Messages

SETUP

PLAY

PAUSE

REDIRECT

PING

GET_PARAMETER

SET_PARAMETER

OPTIONS

DESCRIBE

TEARDOWN

RTSP example

```
C: SETUP rtsp://audio.example.com/twister/audio RTSP1.0
   Cseq: 1
   Transport: rtp/udp; compression; port=3056; mode=PLAY
S: RTSP/1.0 200 OK
   Cseq: 1
   Session: 4231
C: PLAY rtsp://.....
   Cseq: 2
   ...
S: RTSP/1.0 200 OK
   Cseq: 2
   Session: 4231
C: PAUSE rtsp://.....
   Cseq: 3
   ...
S: RTSP/1.0 200 OK
   Cseq: 3
   Session: 4231
```

Kurose, Ross; Computer Networking, A Top-Down Approach, 5th edition, p 615

Teleconference

- Skype
- Requirements?
  - 150 msec delay becomes annoying
  - PDX to Washington, DC: 3779 km ~ 25 msec RTT (minimum!)

  - at 64 kb(it)ps, how long would it take to fill a 1 KByte packet?
    - 125 msec
  - how long to transmit a 1KByte packet at 1Mbps
    - 8 msec

Other time sinks?

- Buffering?

- UDP!

Small Packets!

Setup/tear down calls

- H.323
  - International Telecommunication Union
- SIP
  - IETF
  - application layer
  - UDP or TCP
  - RFC 3261
  - setup/management/termination

Either can use RTP/RTCP

Sound to bits

- Sound is a smooth wave
- To digitize it, the wave is sampled at fixed intervals
  - Nyquist theorem: sufficient to make samples at $2f$, if f is the highest frequency (2.1.3)
- Two axes:
  - sampling rate
  - sampling precision
  - Phone:
    - 8000 samples per second
    - 8 bit sample: 256 different gradations (7 bits data, 1 bit control)
  - CD
    - 44,100 samples per second
    - 16 bit sample: 65,536 different gradations

Total Bandwidth?
- stereo?

Which sounds better?
- any complaints about CD quality sound?

Max Frequency on
- phone
- cd?

Quantization

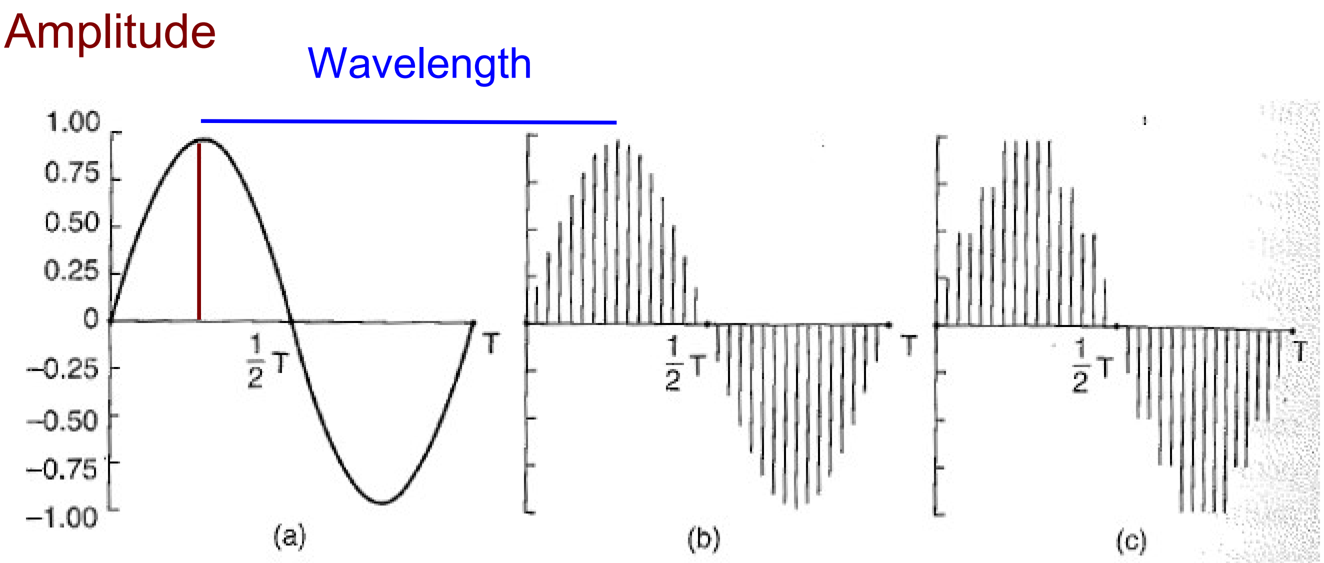


Figure 7-42. (a) A sine wave. (b) Sampling the sine wave. (c) Quantizing the samples to 4 bits.

4 bits represents how many quantities?

Audio Compression

- Lower bandwidth than video
  - still compressed
- Encode/decode
  - compress/uncompress
- Streaming stored audio
  - what does this allow?
- Live audio
  - still compressed. consequences?
- Lossy

Much work on compressing speech for telephone lines

vocoders

Audio Compression

- Read section 7.4.1

- AAC (Advanced Audio Coding)
  - audio portion of MPEG-4 files
- MP3 - MPEG-1 audio layer (part 3)
  - No RFC, this is not an Internet standard
  - MPEG-1: standard for encoding audio and video
  - MPEG: Moving Pictures Expert Group, part of ISO
  - perceptual encoding
    - exploit “flaws” in the human sense of hearing
    - loud sounds mask quiet sounds

psychoacoustics

    - rock’n’roll compressed to 96 kbps (without loss to listener)
    - piano concert compressed to 128 kbps (without loss to listener)
    - total bandwidth? Stereo?

Psychoacustics

power ≈ loudness

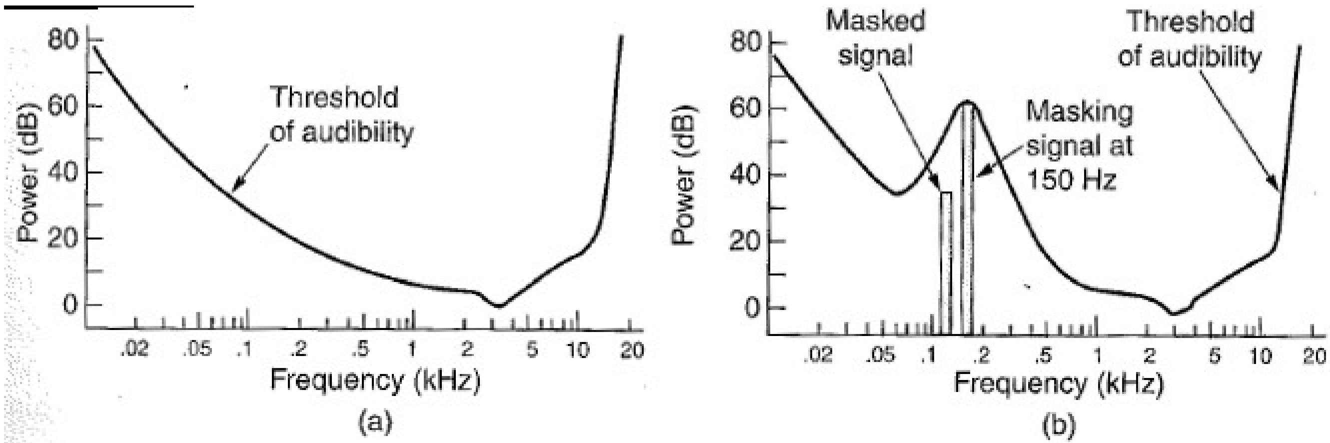


Figure 7-43. (a) The threshold of audibility as a function of frequency. (b) The masking effect.

- get power at each frequency (Fourier transform)
- transmit unmasked frequencies
  - plus other digital compression techniques
    - Huffman encoding

Huffman

- Key idea:
  - encode common symbol with fewer bits
  - encode uncommon symbol with more bits
  - Count the frequency of each symbol
  - build a tree (not unique)
    Pacific, min number of bits to encode Pacific?

Must send string and Tree; with large enough strings the overhead of the tree is minimal.

Frequency	Value
1	p
1	a
1	f
2	i
2	c

http://www.siggraph.org/education/materials/HyperGraph/video/mpeg/mpegfaq/huffman_tutorial.html

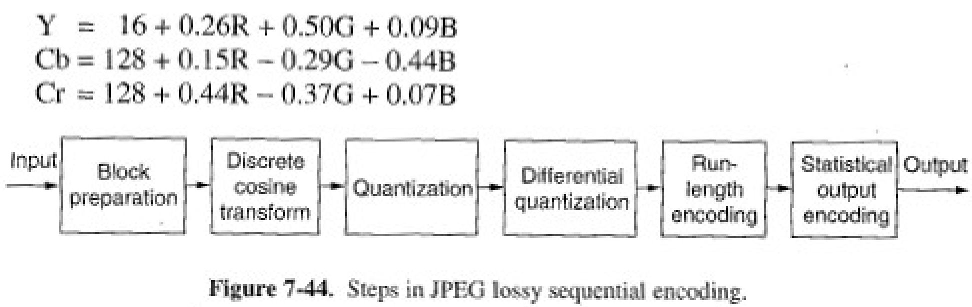
Moving Pictures to bits

- Pixel: picture element
  - rgb, 8 bits per pixel for 24 bit color
  - how many colors at 24 bits?
- 50 images/second = movement!
- Broadcast TeeVee
  - 30 frames/sec (US)
  - only show half (odd/even) rows per frame
  - interlace
  - 60 frames/sec!
  - your eye is a forgiving organ.
    - image decay

Video Compression, Why?

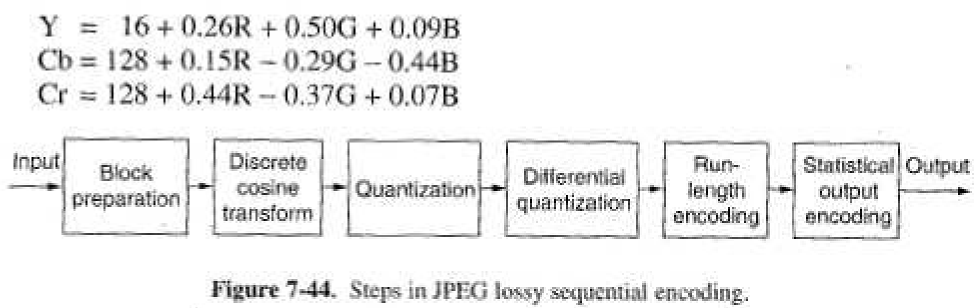
- 640 x 480, 24 bits per pixel, 30 frames a second
  - bandwidth required?
  - How much compression do we need?
    - what is the common download speeds offered by ISPs?
    - don't forget we also need (stereo) audio!

Compression

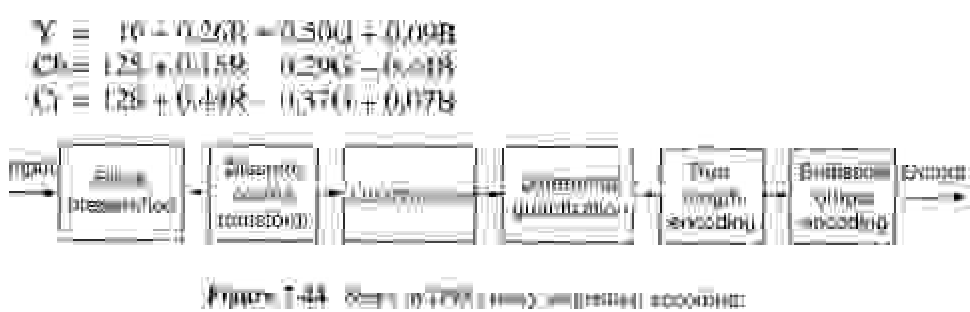


35 KB

Where does the compression happen?



11 KB



2 KB

JPEG

- Joint Photographic Experts Group
  - 4 modes
  - we care about lossy, sequential mode

- Luminance - brightness
  - Y
- Chrominance - color
  - Cb, Cr
- Calculate a matrix for Y, Cb, Cr
  - average 4 pixels in Cb, Cr matrices to reduce data
  - build blocks of 8x8 in each matrix

Assumption:
Most pixels have similar neighbors

Your eyes care more about brightness than color.

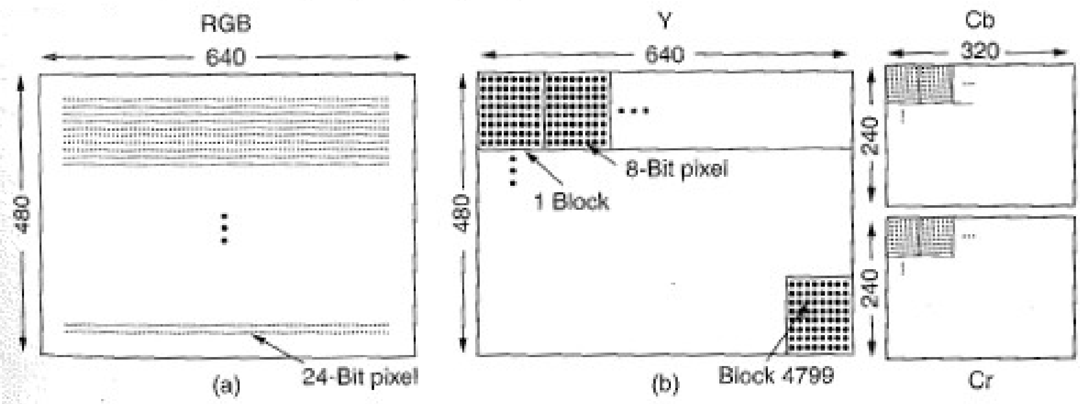


Figure 7-45. (a) RGB input data. (b) After block preparation.

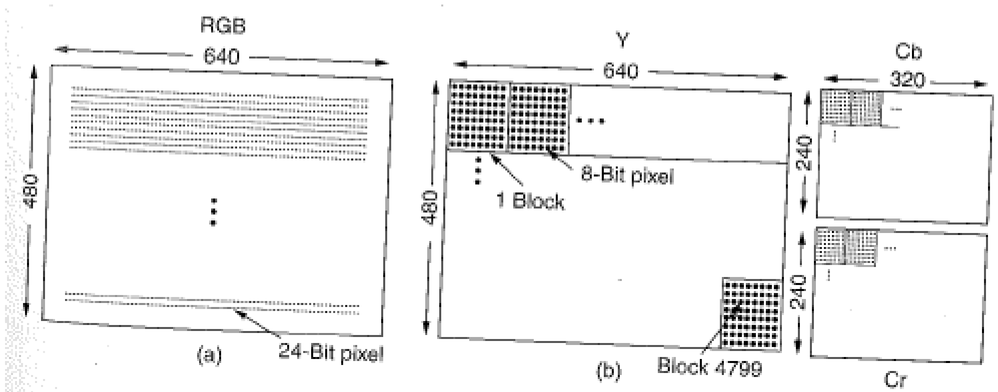


Figure 7-45. (a) RGB input data. (b) After block preparation.

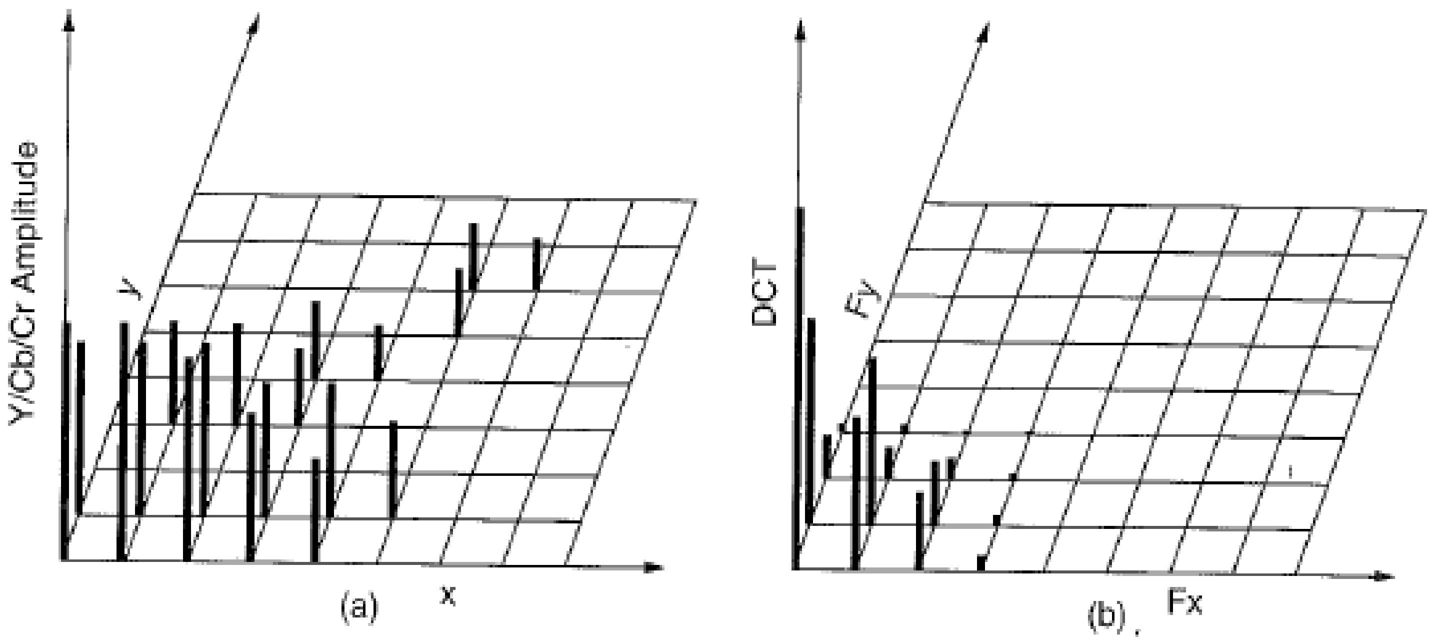


Figure 7-46. (a) One block of the *Y* matrix. (b) The DCT coefficients.

Differential quantization

DCT coefficients

150	80	40	14	4	2	1	0
92	75	36	10	6	1	0	0
52	38	26	8	7	4	0	0
12	8	6	4	2	1	0	0
4	3	2	0	0	0	0	0
2	2	1	1	0	0	0	0
1	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0

Quantization table

1	1	2	4	8	16	32	64
1	1	2	4	8	16	32	64
2	2	2	4	8	16	32	64
4	4	4	4	8	16	32	64
8	8	8	8	8	16	32	64
16	16	16	16	16	16	32	64
32	32	32	32	32	32	32	64
64	64	64	64	64	64	64	64

Quantized coefficients

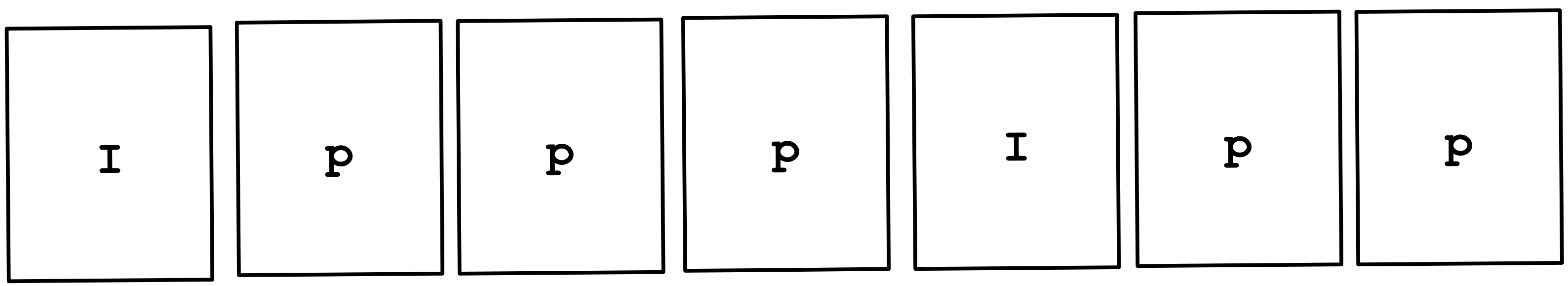
150	80	20	4	1	0	0	0
92	75	18	3	1	0	0	0
26	19	13	2	1	0	0	0
3	2	2	1	0	0	0	0
1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

Figure 7-47. Computation of the quantized DCT coefficients.

Run Length Encoding

MPEG

- Sequence of frames
  - **I** (intra frame)
  - P (predictive) motion vector of objects from previous frame
  - B (bi-directional) motion vector of objects from previous/next frame



  - Macroblock: 16x16 pixels (4 luminance, 2 chrominance)
    - each has a motion vector