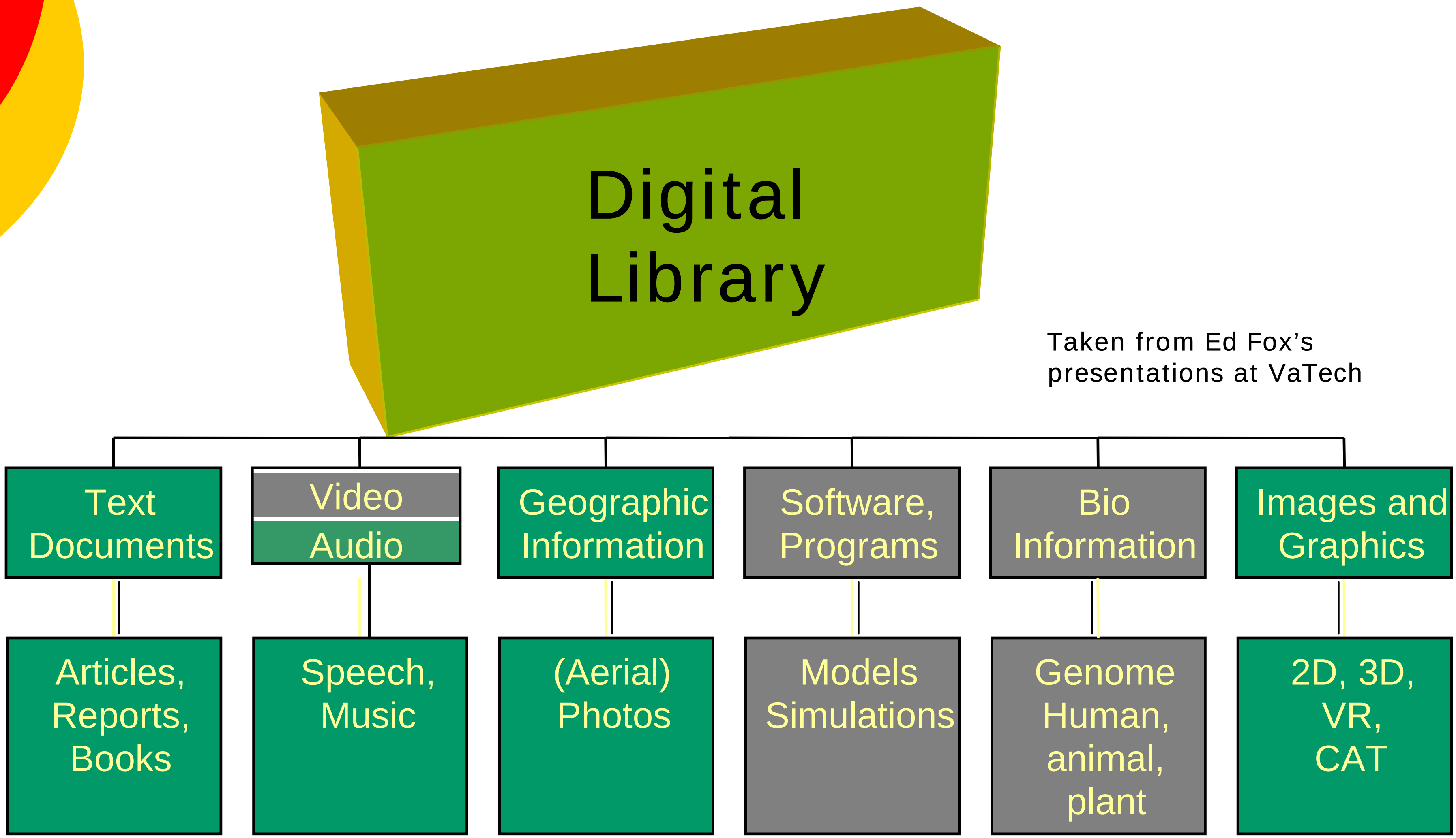
Representation and digitization of multimedia*

Week 2 Min-Yen KAN

*Heavily scaled down from original lecture outline :-(

Media types in the DL



Taken from Ed Fox's presentations at VaTech

Distribution of media types in the library

Library Type	LoC Gov't	NUS Acad	U Toronto Acad
Books and manuscripts	19 M	2.2M	9.1 M
Maps	4 M	22.1 K	278 K
Photographs	12 M		622 K
Music	2.7M		186 K
Motion pictures	.9 M		21 K
CD-ROM Databases		1.4K	2.1 K

Question: is the distribution of what we'd like in the digital library the same as in the automated library?

- - NUS and LOC figures 2003; U Toronto, 2002
- NUS Libraries multimedia increased over 13% but only 2% for books

Outline

Representation / Digitization

- Textual images
- Images
- Audio
- Coordinated multimedia

Textual images

Cost basis for archives

	Year 1	Year 4	Year 7	Year 10
Depository Library				
Storage cost (per volume)	.24	.27	.30	.34
Access cost (per volume)	3.97	4.46	5.02	5.64
Digital Archive				
Storage cost (per volume)	2.77	1.83	1.21	.80
Access cost (per volume)	6.65	4.76	3.51	2.70

From Lesk (99), pg. 75

Digitization

- Scanning
  - Binding
  - Planetary scanner
- Resolution of scan
  - 300 dpi* for access
  - 600 or higher for archival copy

A high speed scanner, 1.2 pps at 200 dpi

Planetary or "bookeye" scanner

*** - Depends the smallest point size you need to resolve**

Digitization

- Purpose:
  - Archival
    - Quality
    - Stability in the long term
  - Accessibility
    - Delivery
    - Editing
    - Annotation

1. Initiate the digitalization project
2. Establish start-up costs and secure funding
3. Prepare a detailed project plan include milestones and deliverables
4. Assess and select materials for digitization
5. Digitize materials (prepare source materials, digitize, check quality)
6. Post-process digital materials: edit, OCR, store, catalog and index
7. Deliver and make materials accessible
8. Support and maintenance of materials

-- From Chowdhury and Chowdhury (03)

Document capture costs in USD (ca. 1999)

	Preparation	Scanning	Post-scan Processing	Total (3 years)
Capital	Tables, jogger, $1,500	Mid-volume scanner plus PC, $25,000	Two PCs, printer, software, $12,000	**$47,500 (11%)**
Maintenance	None	8% per year $2,000 per year	8% per year $1,000 per year	
Labor	Two people $40,000 per year	One person $20,000 per year	Two people $40,000 per year	**$300,000 (71%)**
Space	120 square feet $12,000 per year	40 square feet $4,000 per year	100 square feet $10,000 per year	**$78,000 (18%)**
Total (3 years)	**$157,500 (37%)**	**$103,000 (24%)**	**$165,000 (39%)**	**$425,500 (100%)**

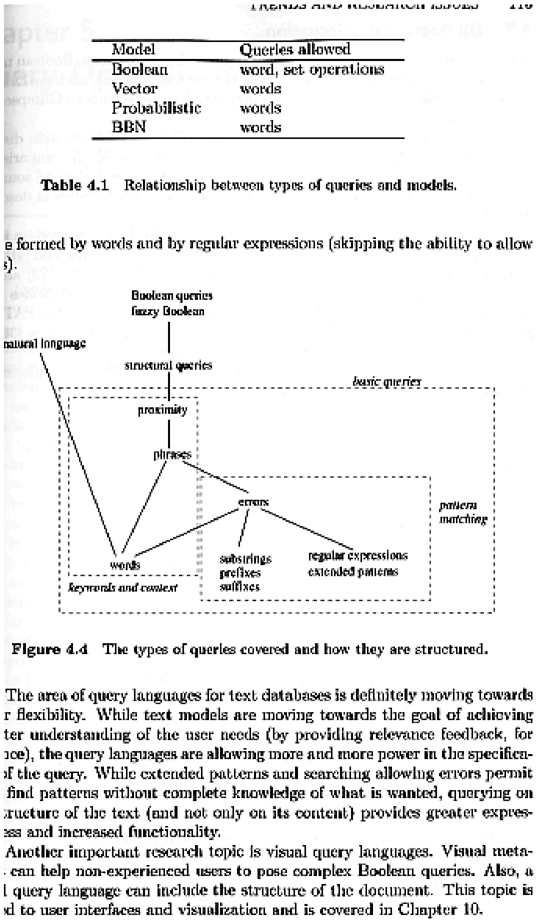
Capacity = ~1,000 page per hour x 6.5 hours x 250 days x 3 years = 4.8 M.
Cost per page is $425,500 / 4,875,000 = $0.09 (8.7 cents)

Images of text

You’ve scanned in an image like this…

What to do with it?

How would we like to store and access this information?



TRENDS AND RESEARCH ISSUES 115

Model	Queries allowed
Boolean	word, set operations
Vector	words
Probabilistic	words
BBN	words

Table 4.1 Relationship between types of queries and models.

e formed by words and by regular expressions (skipping the ability to allow
s).

Figure 4.4 The types of queries covered and how they are structured.

The area of query languages for text databases is definitely moving towards r flexibility. While text models are moving towards the goal of achieving ter understanding of the user needs (by providing relevance feedback, for ıce), the query languages are allowing more and more power in the specifica- of the query. While extended patterns and searching allowing errors permit find patterns without complete knowledge of what is wanted, querying on ructure of the text (and not only on its content) provides greater expres- ss and increased functionality.

Another important research topic is visual query languages. Visual meta- can help non-experienced users to pose complex Boolean queries. Also, a l query language can include the structure of the document. This topic is d to user interfaces and visualization and is covered in Chapter 10.

Storing a textual image

- Mostly bi-level (two-tone) until recently

1. CCITT Fax III and IV
   - Bi-level transmission and storage standard
   - Optimized for Roman alphabet
2. Textual image compression
   - Codebook of marks
   - A level for access and one for preservation

Horizontal run length encoding

- Always starts with white run (possibly of length zero)
- Huffman code stores a *terminating code* (TC) of all lengths shorter than 64 pixels
- Longer length encoded by 64, 128, 256 = 2^k make up code + TC

code table

run length	color of run: white	black
0	00110101	0000110111
1	000111	010
2	0111	11
3	1000	10
4	1011	011
5	1100	0011
6	1110	0010
7	1111	00011
8	10011	000101
9	10100	000100
...	...	...

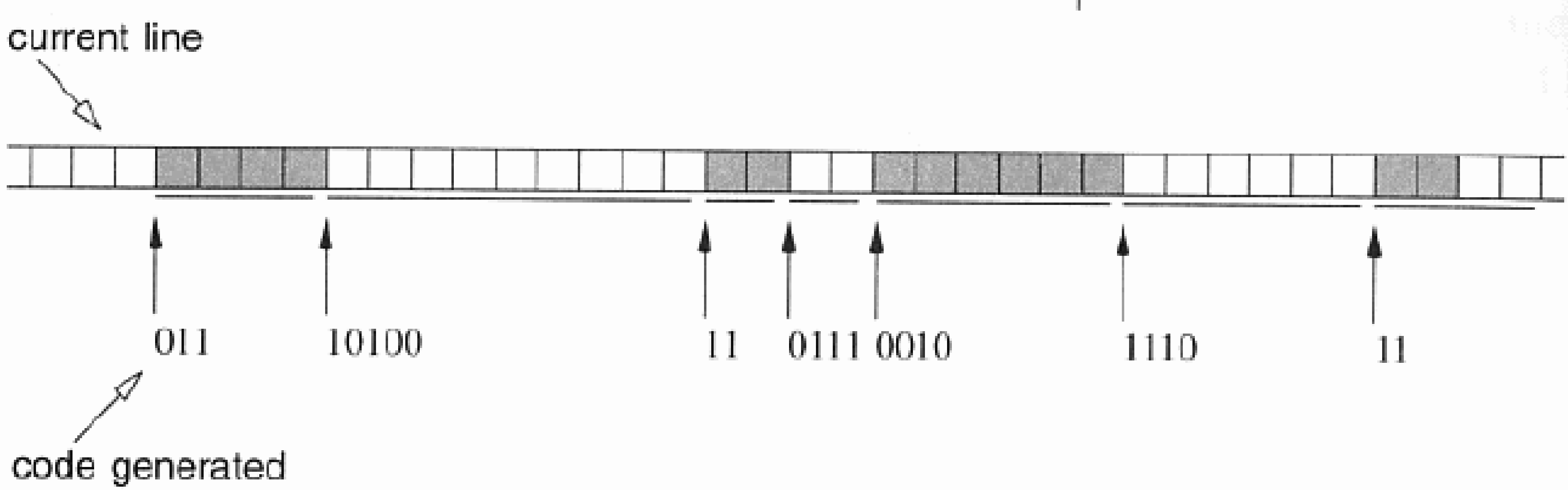
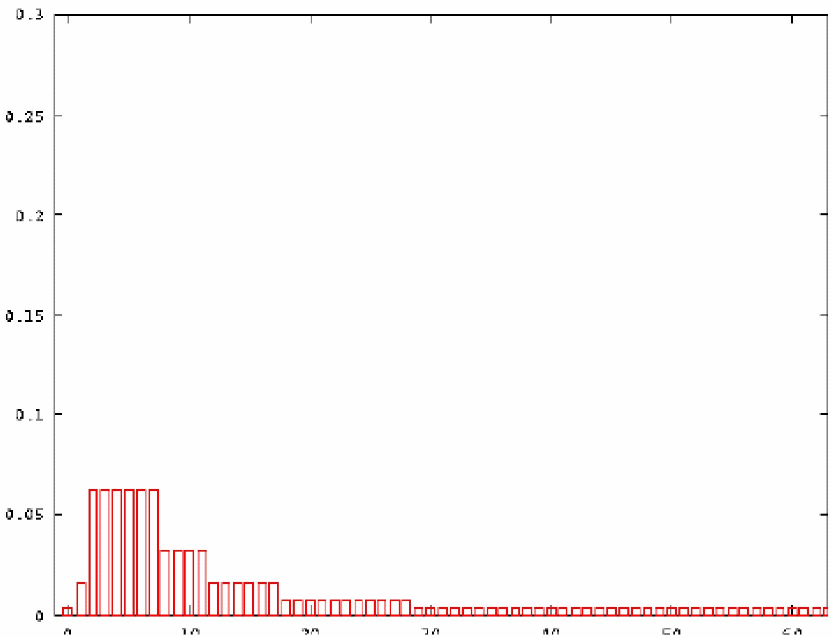


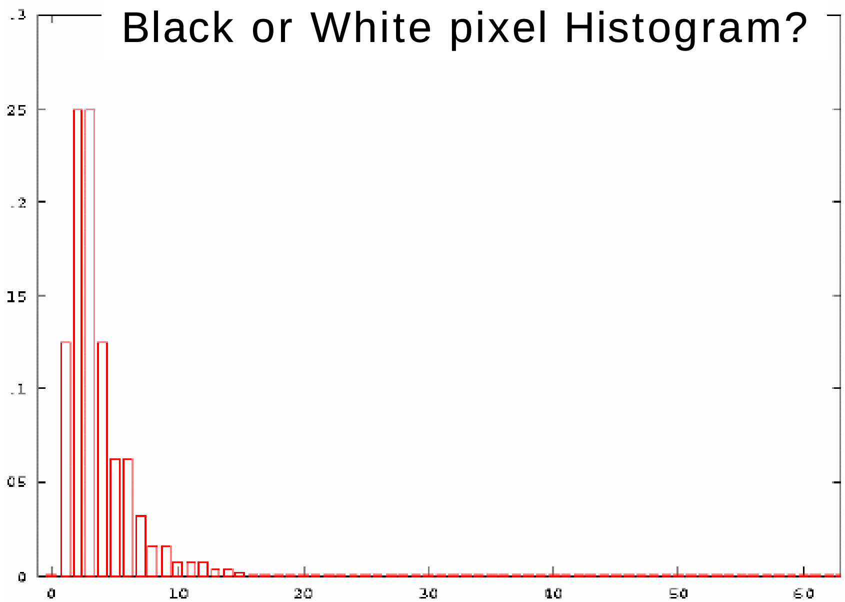
Figure 6.2 Example of one-dimensional coding.

From Witten et al, MG

Which histogram is which?



Black or White pixel Histogram?



- CCITT Fax group III uses Huffman encoding to decide close to optimal encoding
- We show a black pixel histogram and white pixel histogram here. Which is which?

CCITT fax group IV

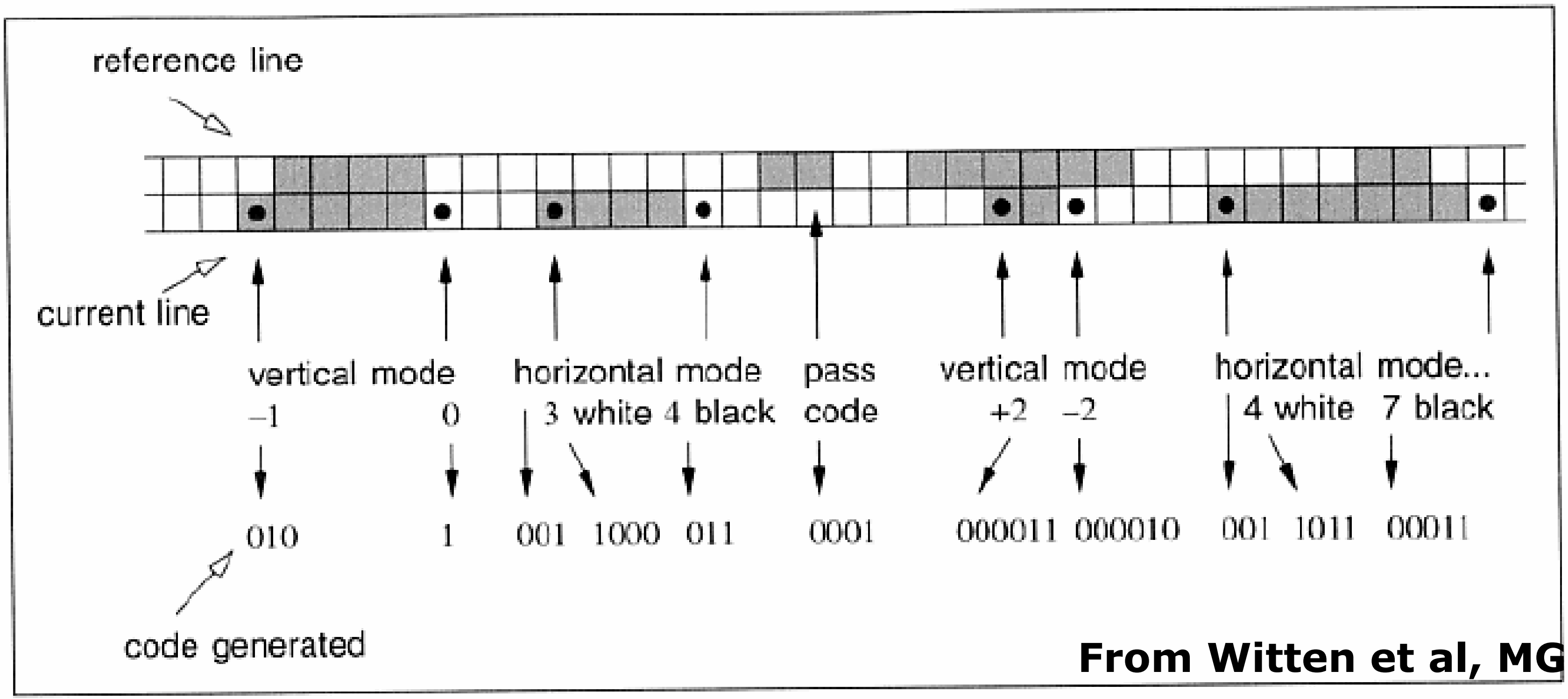


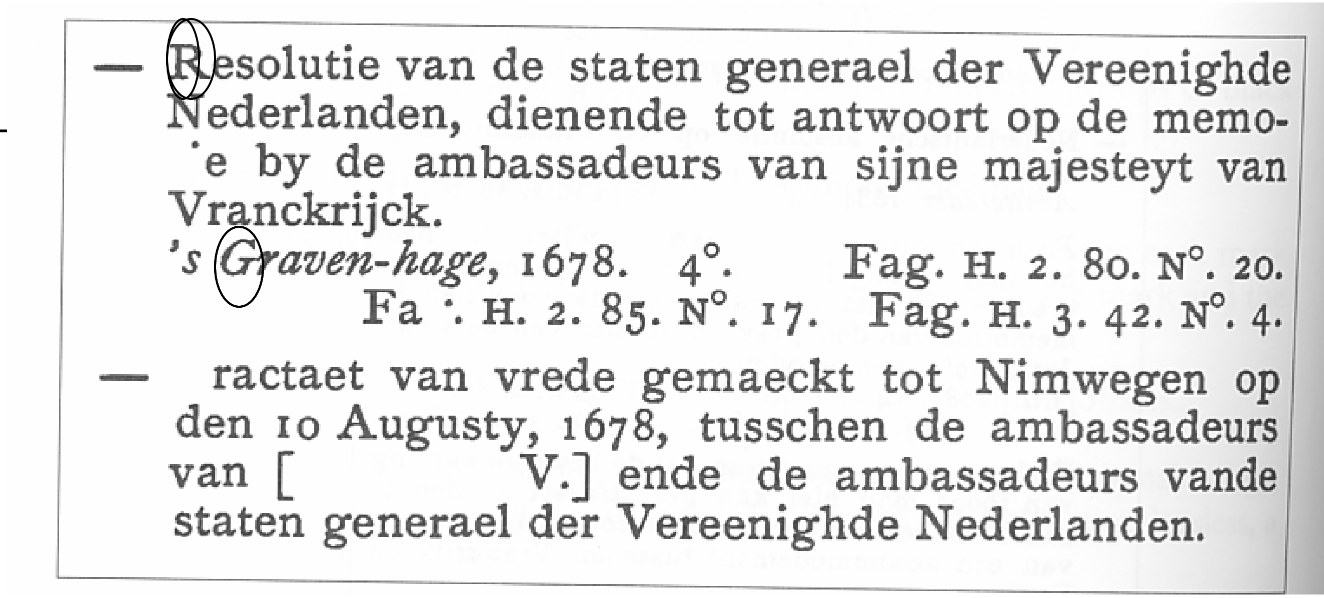
Figure 6.3 Example of two-dimensional coding.

- Takes vertical redundancy into account
- Three methods of encoding: vertical, horizontal and pass

Textual image compression

1. Find and isolate ***marks*** (connected group of black pixels)
2. Construct library of symbols
3. Identify the symbol closes to each mark and get coordinates
4. Store information
5. *Store additional information to reconstruct original image

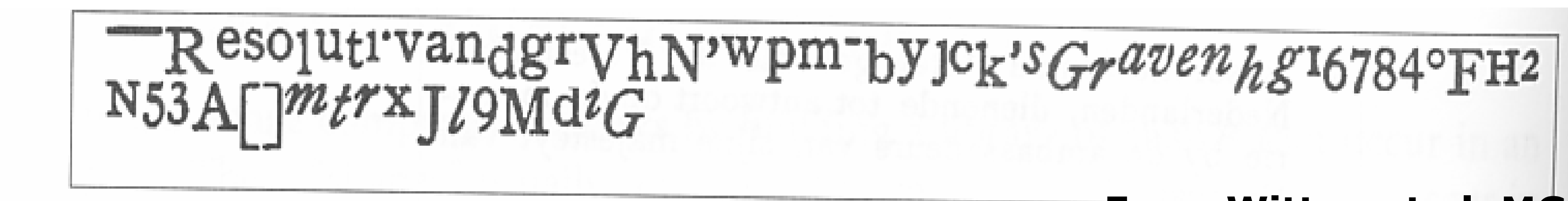
Library


— Resolutie van de staten generael der Vereenighde Nederlanden, dienende tot antwoort op de memo-'e by de ambassadeurs van sijne majesteyt van Vranckrijck.
's Graven-hage, 1678. 4°. Fag. H. 2. 80. N°. 20. Fa ·. H. 2. 85. N°. 17. Fag. H. 3. 42. N°. 4.

— ractaet van vrede gemaeckt tot Nimwegen op den 10 Augusty, 1678, tusschen de ambassadeurs van [V.] ende de ambassadeurs vande staten generael der Vereenighde Nederlanden.

(symbol, x-offset, y-offset)

(1,50,13) (28,73,121)



From Witten et al, MG

Figure 7.2 Library of symbols created from the example image.

Residue

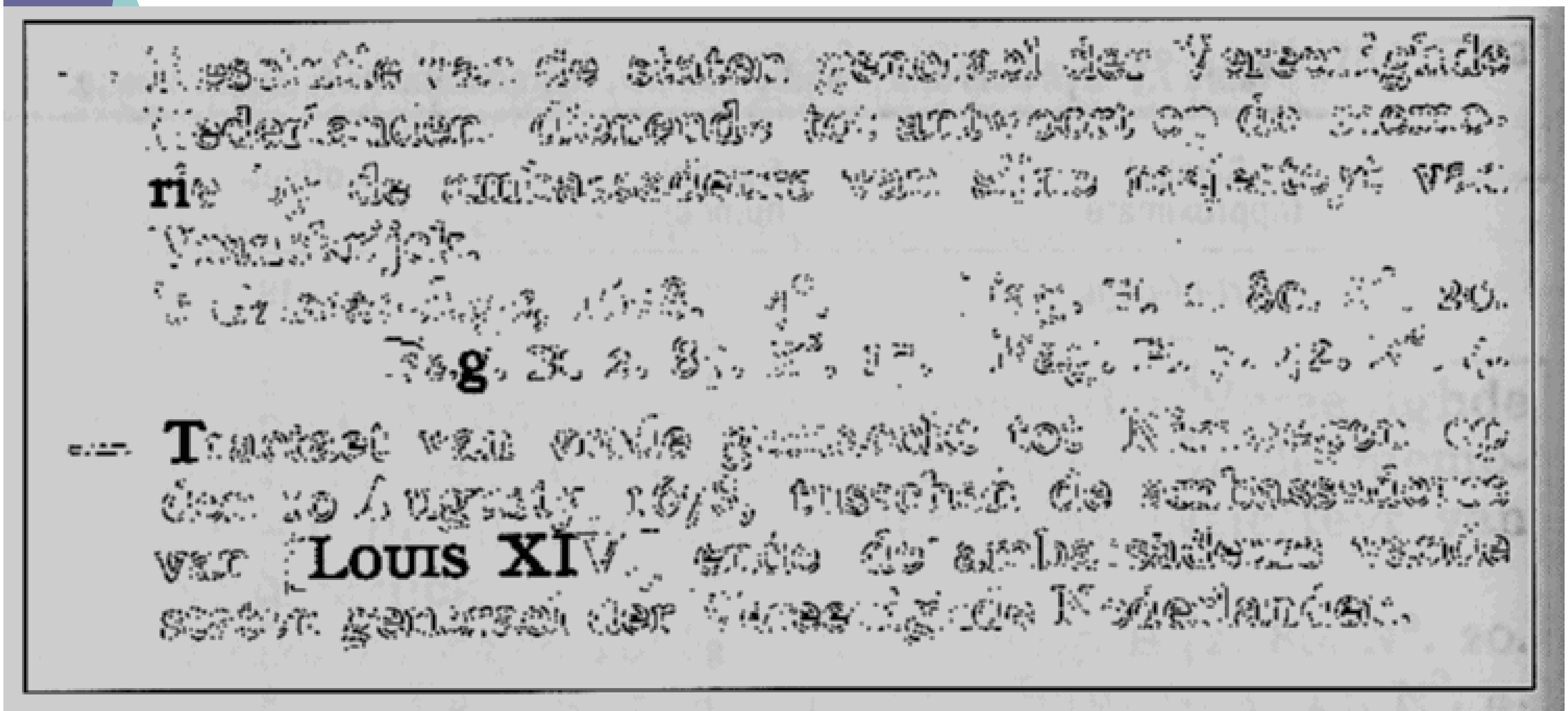


Figure 7.4 The "residue" image.

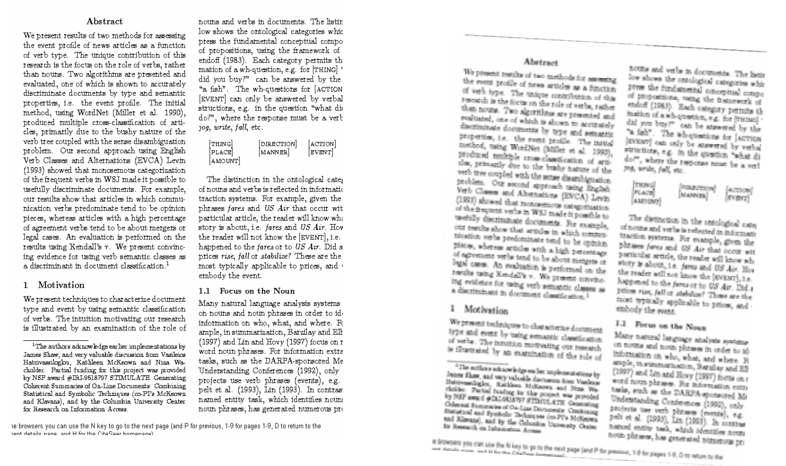
From Witten et al, MG

Text image outline

- Storage √
  - CCITT Fax Group III and IV √
  - Textual image compression √
- Access
  - De-skew
  - Segmentation
  - Media detection

De-Skew

- Projection profile
  1. Accumulate Y-axis pixel histogram
  2. Rotate to find most crisp histogram



- One of three common algorithms

From Witten et al, MG

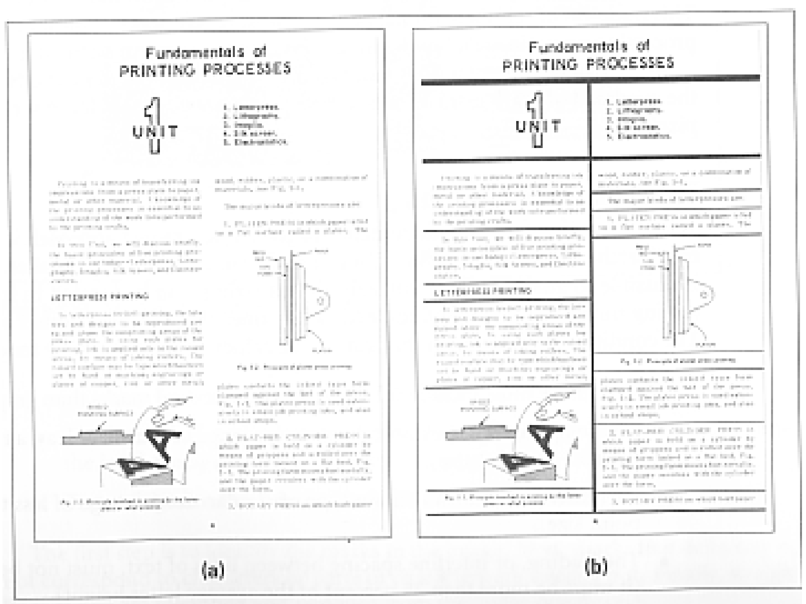
Segmentation

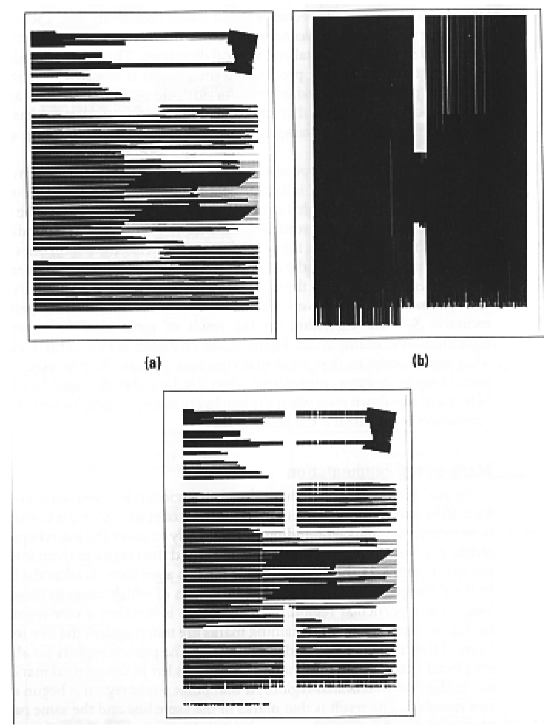
- Top-down

(e.g., X-Y cut)

- Bottom-up

(e.g. smearing)

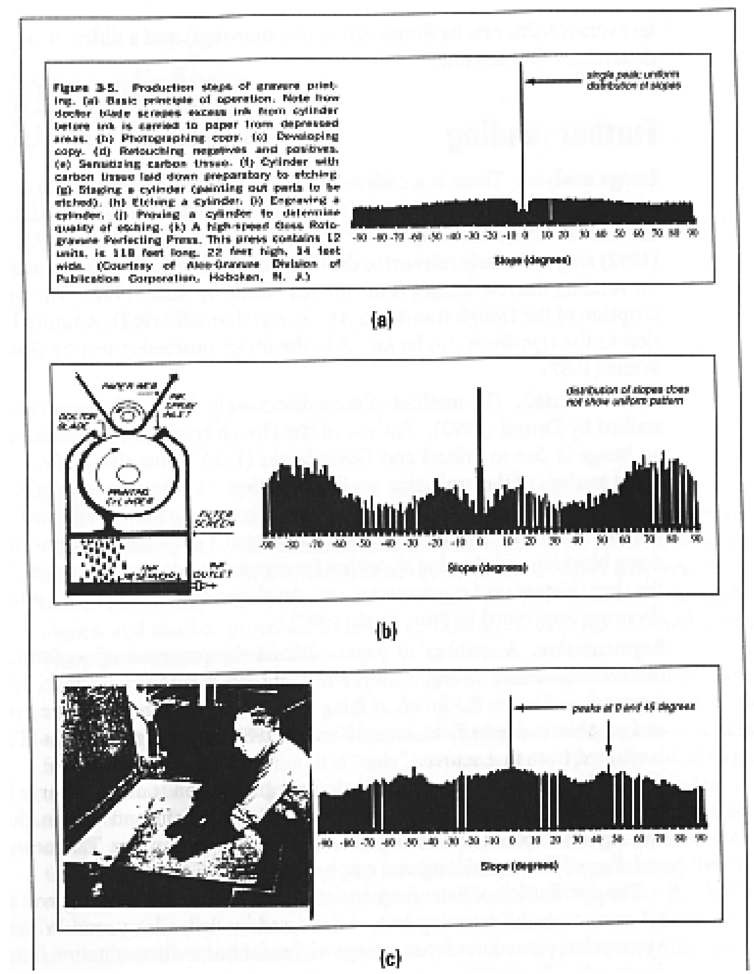




Classification

- Separate:
  - Images
  - Text
  - Line art
  - Equations
  - Tables

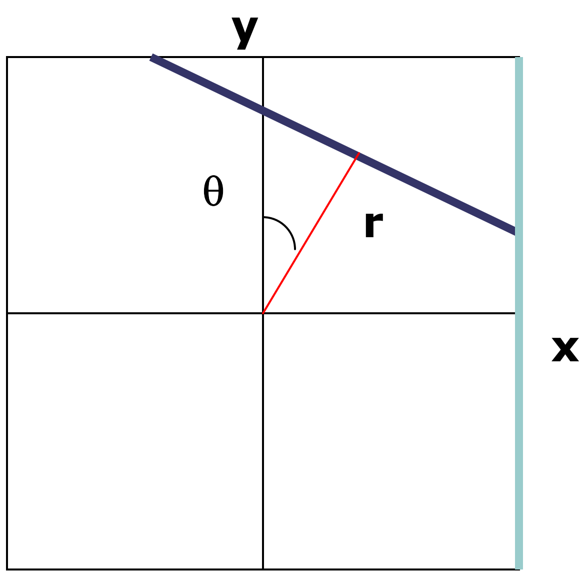
- One technique:
  - Slope Histogram (Hough transform)

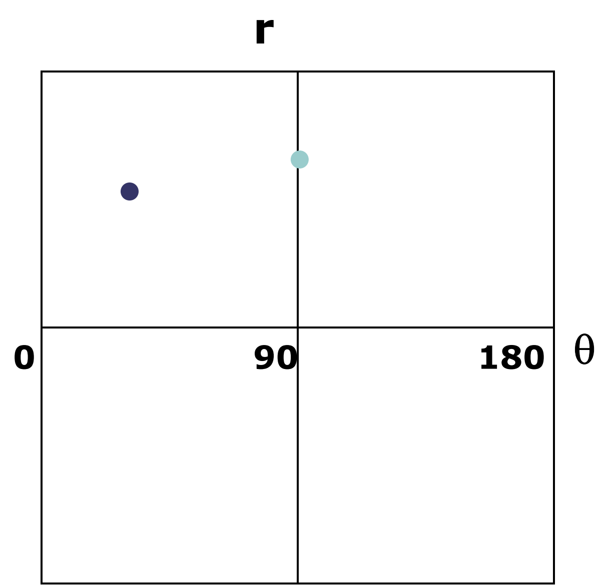


From Witten et al, MG

Hough Transform

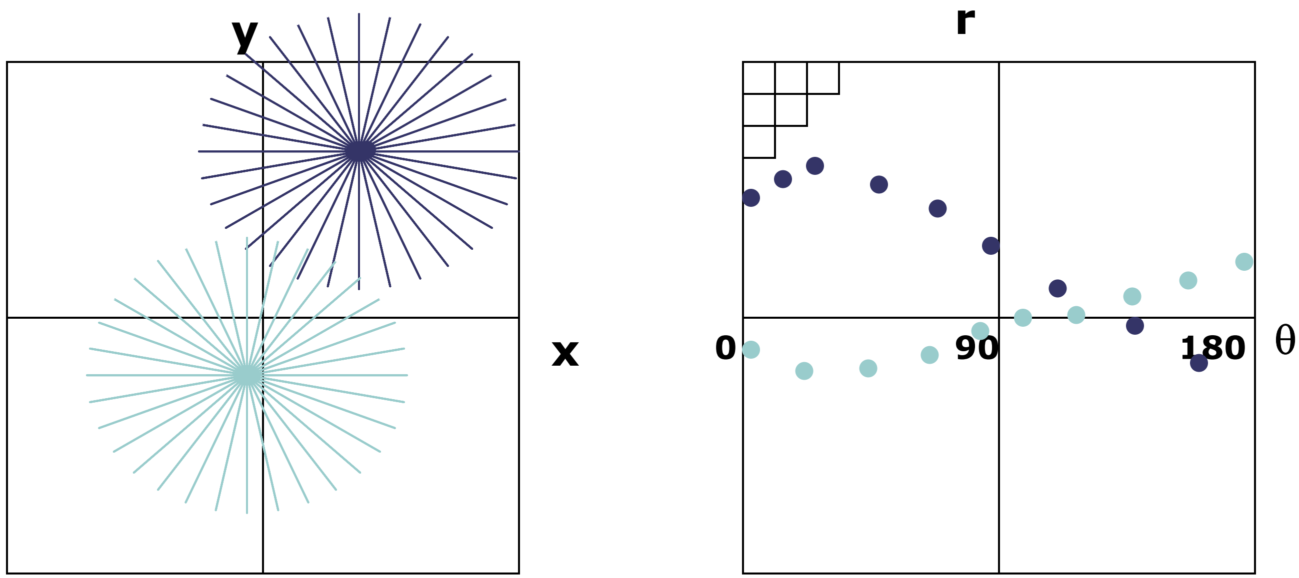
- A line-to-point transform
- In practice, used to find lines in an image (e.g., set of pixels on a line)





Hough Transform

- Create virtual lines for each point
- Accumulate counts for bin in Hough space



Effective as not doing pairwise comparison

Robust Document Understanding

- OCR and document understanding are (currently) fragile technologies
  - Full scan ⇒ OCR ⇒ store pipeline makes many assumptions
  - What are some?
    - __________________
    - __________________
    - __________________
    - __________________
    - __________________

Scholarly and historical DL are much harder!

A solution (one of many)

- Courtesy Henry Baird’s ICDAR 03 slides.
  - http://www.cse.lehigh.edu/~baird/Talks/icdar03.ppt#21

Image data

- Raster graphics
  - As an array of pixels
- Vector graphics
  - As a collection of vectors
- Which format appropriate for which images?
  - Maps
  - Photographs
  - Line art
- For which use?
  - Fidelity?
  - Re-scaling?
  - Compression?

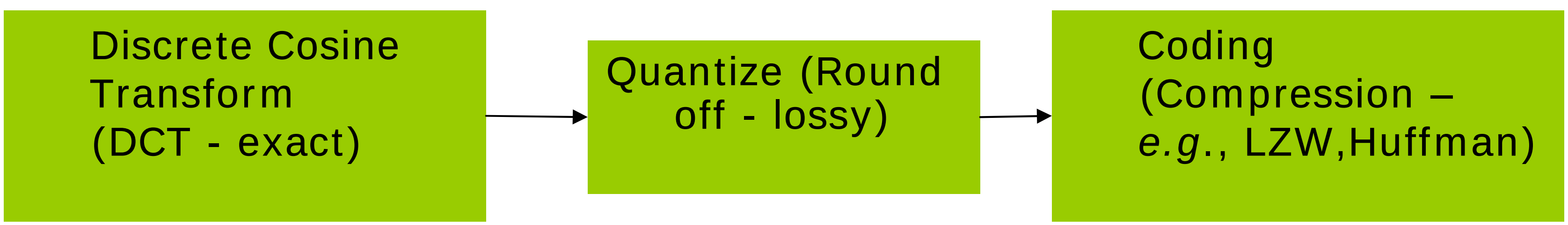
GIF / PNG

- **GIF** ('jiff', Graphics Interchange Format)
  - Stable, lossless color format
  - Compression achieved by:
    - 8-bit format (256 colors)
    - LZW encoding (Unisys patent)
  - __________________________________.
  - Interlacing options for low-bandwidth accessibility

- **PNG** ('ping', Portable Network Graphics)
  - Uses _____________________________
  - Up to 48 bits of color (compared to 8 in GIF)
  - Support for alpha channels (transparency) and gamma correction (white balancing)

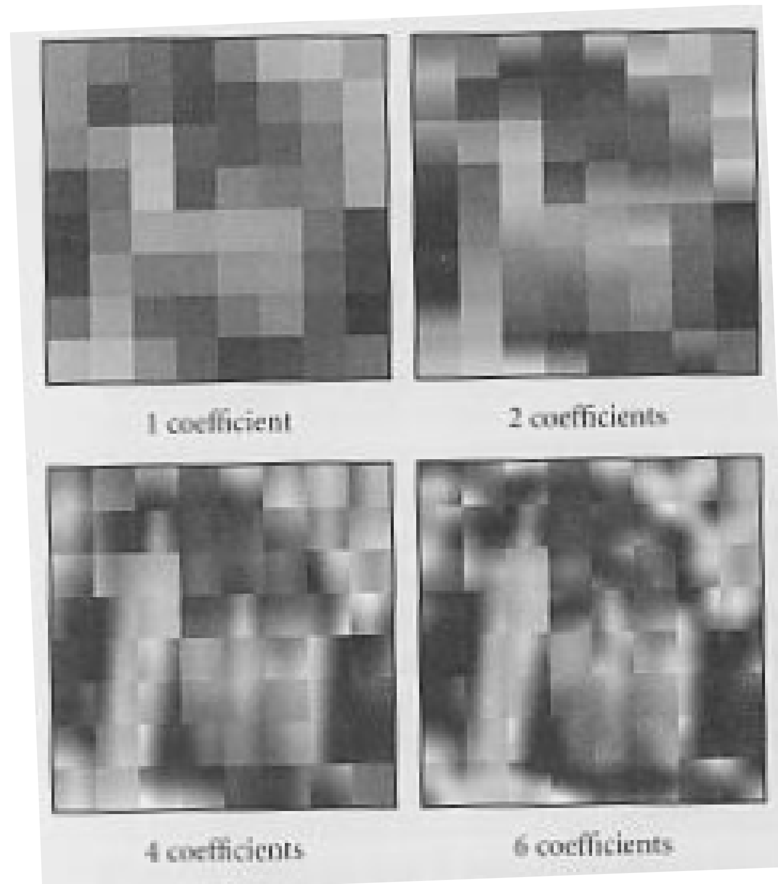
Joint Photography Experts Group

- Breaks image into 8×8 pixel blocks, each pixel 24 bits (YUV channels = 3×8 bits each)
- Compresses each block separately, ____________________



JPEG, continued

- Transform yields coefficients
- Ordered from low frequency (gradual change) to high frequency
- Gradual changes well represented
  - Good for scenery, natural images
- JPEG 2000 incorporates wavelet compression
  - Better for sharp edges



Postscript

- A programming language whose operators draw graphics on the page.
  - Text is a deemed a type of graphic
  - To “draw” a page, you construct a paths used to create the image.
- A stack based, usually interpreted language
- Uses reverse polish notation

A simple Postscript example

A method to place some text down the left margin of the a page.

- You can use this after the marker for the beginning of a page.

```
gsave                          % save graphics state on stack
90 rotate                      % rotate 90 degrees
100 .55 -72 mul moveto         % go to coords 100, (.55*-72)
/Times-Roman findfont          % Get the font (set of operators) Times-Roman
10 scalefont                   % set the font size
setfont                        % Use the specified font
0.3 setgray                    % Change the color to gray
(PUT NOTE HERE) show           % call the individual operators P,U,T ...
                               % to draw letters
grestore                       % restore the graphics state
```

Portable Document Format

- An object database
  - Subset of Postscript, makes it faster to process
  - Can use several different compression techniques (*e.g.*, LZW and Huffman)
  - Proprietary
  - Has capabilities for hyperlinks

Geospatial Datasets

- Which image format is best for maps?

Hmm, let's think about it. What goes into a map?

1. ____________________, which provides the position and shapes of specific geographic features.
2. _____________________, which provides additional non-graphic information about each feature.
3. ___________________, which describes how the features will appear on the screen.

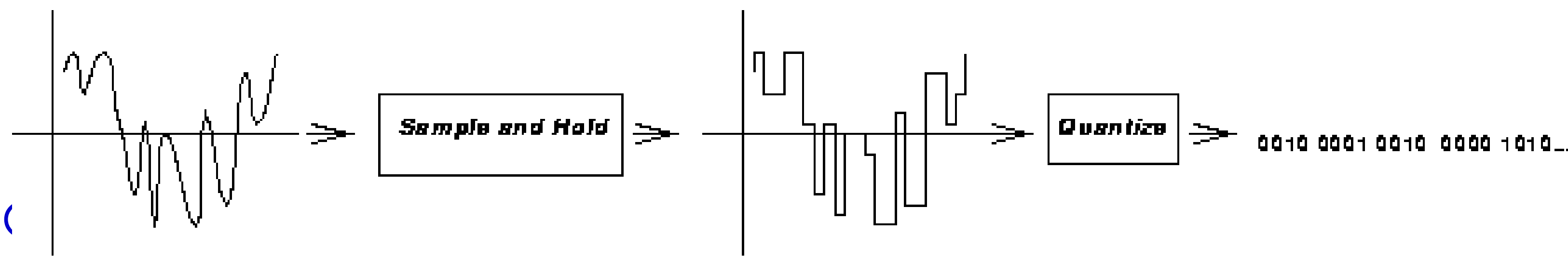
-- Excerpted from Geo Community, 04

Audio

- Limit representation to what people can hear
  - Humans: ~ _____________ KHz
- Highest frequency (pitch) determines storage size.
  - Speech: limited range: up to 3 KHz
  - Music: full dynamic range, 20 KHz
  - Can be referred to as its *bandwidth*

Sampling

- Take continuous signal and discretize
- Higher sampling rate = better fidelity



sampling rate = 2 × bandwidth

- Music: full dynamic range: ~ 22K × 2 = 44K
- Speech: 4K × 2 = 8 K

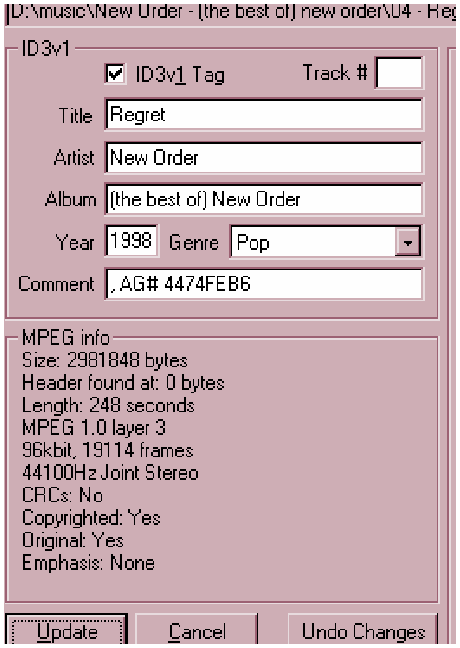
Amplitude and Channels

- Sampling at these time intervals to get *amplitude* of signal
  - a total of ~30-60 dB in loudness
  - Human ear more sensitive to soft sounds
  - *Compand* amplitude (____)
  - 1 or 2 bytes

- For each time interval, may have to sample one or more channels
  - Differential coding (joint stereo)
  - Dolby AC 3 = ____ channels
  - Stereo = 2 channels

Storage Requirements (bitrate)

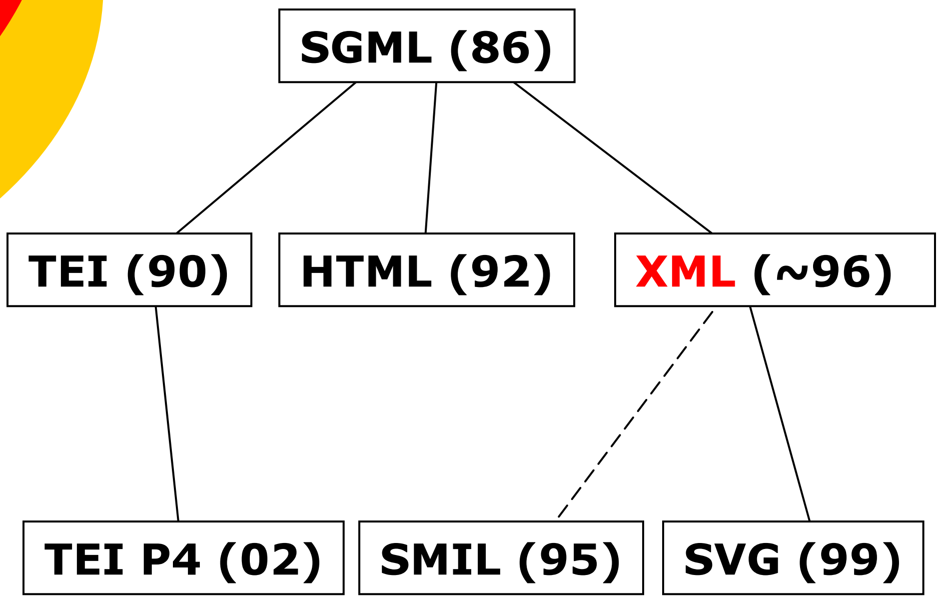
- Digital Music:
  - 44 K samples/sec × 16 bits/sample × 2 channels = ~1.4 M bits/sec
- Digital Voice:
  - 8 K samples/sec × 8 bits/sample × 1 channel = ~64 K bits/sec
- Analog
  - FM stereo: 40 K samples/sec × 8 bits/sample × 3 channels = ~900 K bits/sec
  - Telephony: ~6 K samples/sec × 2 bits/sample × 1 channel = ~12 K bits/sec
- Formats
  - AAC: ________________________
  - MP3: ________________________
  - GSM: ________________________



Putting media together

Have multimedia, will travel...

XML



XML says: “My family tree!”

- A basis for many other technologies
- No semantics (eXtensible, not rigid), just allows for hierarchical containment
- A meta markup language

XML, continued

- Features:
  - Separation of content from presentation
    - Content: Document Type Definition (DTD), optional
    - Presentation: ________________________, ____________________
  - Enhanced hyperlinking capabilities
    - Bidirectional linking
    - Finer grained linking (XPointer)

Text Encoding Initiative

To encode knowledge "of literary and linguistic texts for online research and teaching"

- better interchange and integration of scholarly data
- support for all texts, in all languages, from all periods
- guidance for the perplexed: ____ to encode --- hence, a user-driven codification of existing best practice
- assistance for the specialist: ____ to encode --- hence, a loose framework into which unpredictable extensions can be fitted

- From the TEI Pizza talk

The "**beef**" in XML. All the semantics and none of the filling. It's quite filling, weighing in at 600 K words! (Think 8 kg of books)

Synchronized Multimedia Integration Language :-)

- A script for orchestrating a presentation
  - Think TV news
- Basics:
  - Define a root window
  - Layers
- Timing
  - <par> parallel playback
  - <seq> sequential playback
  - Media clips have begin and end attributes

To think about: what's the alternative format to SMIL? How does it enhance presentation?

Summary

- Representation of knowledge
  - The more you know about the media, the faster, smaller you can transmit and store it
  - Different formats for different purposes, difference isn't superficial

- Multimedia representation
  - Trend toward accessibility, not compressibility
  - Separation of compression from format

References

- More on SMIL: http://www.bu.edu/webcentral/learning/smil1/
- SMIL demos: http://www.ludicrum.org/demos/SMILTimingForTheWeb-Demos.html
- http://www.geocomm.com/ and http://www.usgs.gov are good spots for GIS information.
- Genomic DL indexing and retrieval: http://goanna.cs.rmit.edu.au/~jz/fulltext/ieeekade02.pdf
- JPEG: Pennebaker and Mitchell (93), *The JPEG Still Image Data Compression Standard*
- TEI Pizza talk: http://www.tei-c.org/Talks/