# Image and Video Retrieval

LBSC 796/INFM 718R

Session 13, May 4, 2011

Douglas W. Oard

# Agenda

- Questions
- Image retrieval
- Video retrieval
- Finishing up Speech and Music
- Project presentations

# Image Retrieval

- We have already discussed three approaches
  - Controlled vocabulary indexing
  - Ranked retrieval based on associated captions
  - Social filtering based on other users' ratings

- Today's focus is on content-based retrieval
  - Analogue of content-based text retrieval

# Applications

- Retrospective retrieval applications
  - Manage stock photo archives
  - Index images on the web
  - Organize materials for art history courses
- Information filtering applications
  - Earth resources satellite image triage
  - News bureau wirephoto routing
  - Law enforcement at immigration facilities

# Yahoo! Image Surfer


**Computers and Internet:Internet:World Wide We b:Searching the Web**

The following is a list of **Image Surfer** categories at **Yahoo**. Stay tuned because more interesting categories are added every week for your viewing pleasure.

Search

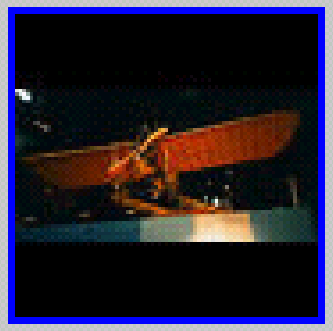
---

## Image Surfer Categories

- **Arts**
  Dance, Landscapes, Photography, ...
- **Entertainment**
  Comics and Animation, Rock and Pop, ...
- **People**
  Actors and Actresses, Models, ...
- **Recreation**
  Computer Games, Sports, ...
- **Science**
  Animals, Space and Astronomy, Museums, ...
- **Vehicles**
  Automobiles, Planes, Motorcycles, ...

[Visual Search]
[Image Info]

*Flattop, Portuguese AF A7D*



[Visual Search]
[Image Info]

*Phil's Aircraft Page- Early Flight, Bleriot Monopl...*

[Visual Search]
[Image Info]

*Diamond Aviation Home Page,*

[Visual Search]
[Image Info]

,

[Visual Search]
[Image Info]

*Evan's Airliner Photo Index Edition #8, Federal Ex...*

[Visual Search]
[Image Info]

*Lockheed P2 Neptune Aircraft, , VAH-21, P2 Picture...*

[Visual Search]
[Image Info]

*Lockheed P2 Neptune Aircraft, , VAH-21, P2 Picture...*

[Visual Search]
[Image Info]

*Aircraft Base: Military Aircraft, RAF Tornado*

[Visual Search]
[Image Info]

*Jets, Douglas A4 Skyhawk , Cars, A4*

[Visual Search]
[Image Info]

*Aircraft Base: Military Aircraft, RAF Tornado*

[Visual Search]
[Image Info]

*Russian Aviation, Redesignated Su 27ib the Su 34 o...*

[Visual Search]
[Image Info]

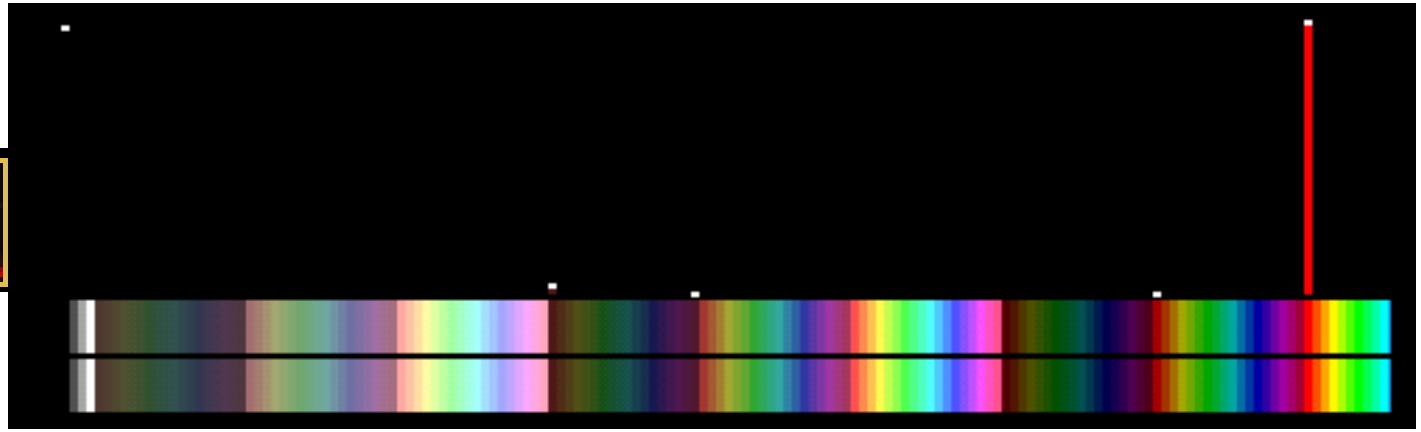
*Manitoba Aviation Page, [Photo 3]*

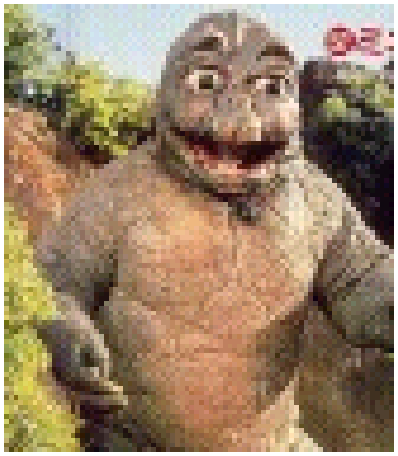
# Color Histogram Matching

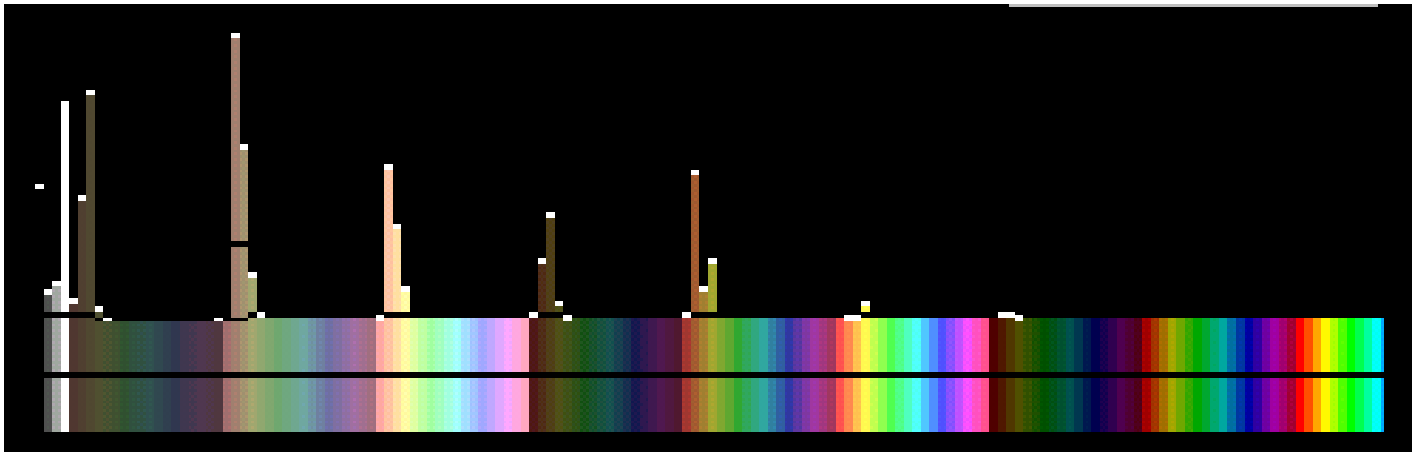
- Represent image as a rectangular pixel raster
  - e.g., 1024 columns and 768 rows
- Represent each pixel as a quantized color
  - e.g., 256 colors ranging from red through violet
- Count the number of pixels in each color bin
  - Produces vector representations
- Compute vector similarity
  - e.g., normalized inner product


http://www.ctr.columbia.edu/webseek/

# Color Histogram Example

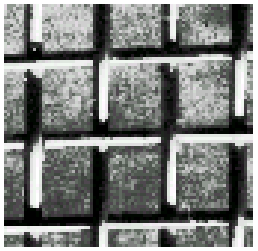


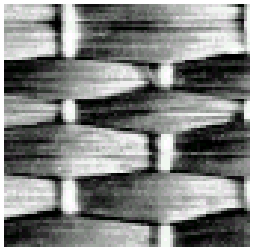


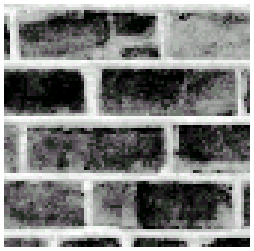
# Texture Matching

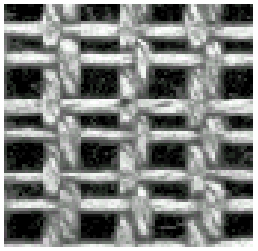
- Texture characterizes small-scale regularity
  - Color describes pixels, texture describes regions
- Described by several types of features
  - e.g., smoothness, periodicity, directionality
- Match region size with image characteristics
  - Computed using filter banks, Gabor wavelets, …
- Perform weighted vector space matching
  - Usually in combination with a color histogram

# Texture Test Patterns


d001

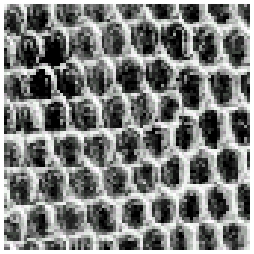

d056

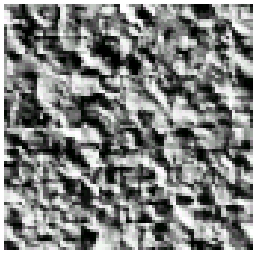

d095

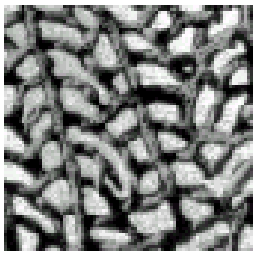

d020


d014

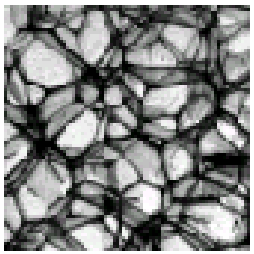

d006

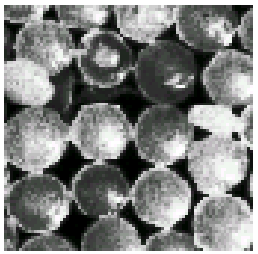

d003

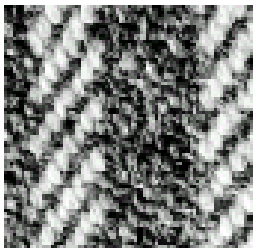

d004

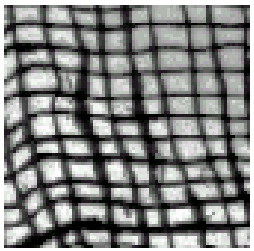

d087

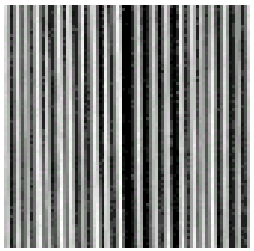

d005

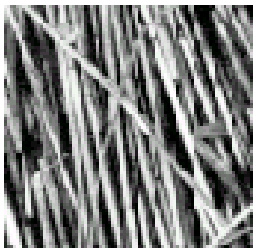

d111


d066


d011


d103


d049


d015

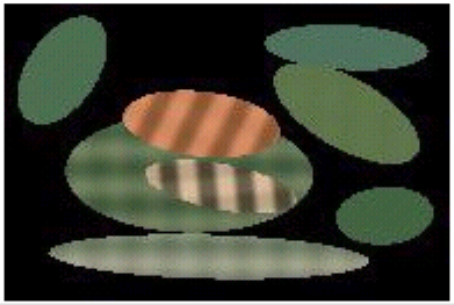
# Image Segmentation

- Global techniques alone yield low precision
  - Color & texture characterize objects, not images
- Segment at color and texture discontinuities
  - Like “flood fill” in Photoshop
- Represent size shape & orientation of objects
  - e.g., Berkeley’s “Blobworld” uses ellipses
- Represent relative positions of objects
  - e.g., angles between lines joining the centers
- Perform rotation- and scale-invariant matching

# Flood Fill in Photoshop

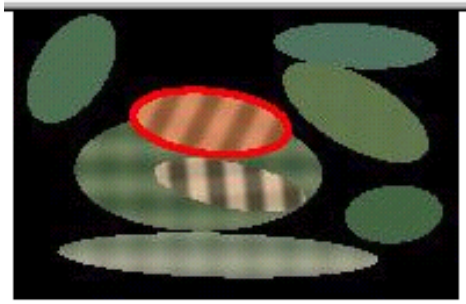
- More sophisticated techniques are needed ☺

# Berkeley Blobworld



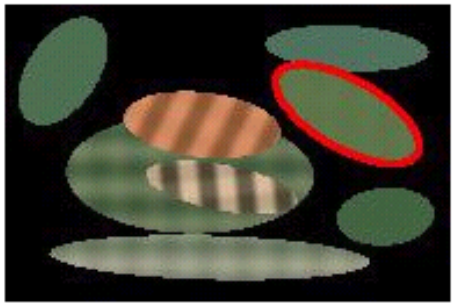
Click on one or two of the blobs in the blobwo
image. Then change the radio buttons to adju
parameter weights. Press one of the Query bu
when you are done.

Full Query | Cheshire Query



| | Somewhat Important | Very Important |
|---|---|---|
| This blob is: | ○ | ◉ |

| | Not Important | Somewhat Important | Very Important |
|---|---|---|---|
| Color: | ○ | ○ | ◉ |
| Texture: | ○ | ○ | ◉ |
| Location: | ◉ | ○ | ○ |
| Shape/Size: | ○ | ◉ | ○ |



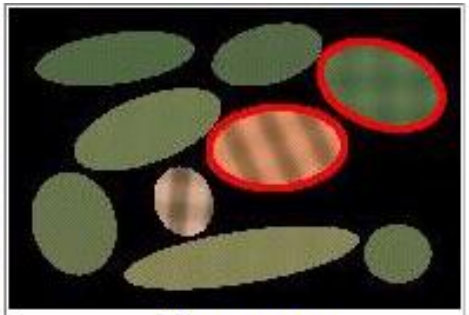
| | Somewhat Important | Very Important |
|---|---|---|
| This blob is: | ◉ | ○ |

| | Not Important | Somewhat Important | Very Important |
|---|---|---|---|
| Color: | ○ | ○ | ◉ |
| Texture: | ○ | ◉ | ○ |
| Location: | ◉ | ○ | ○ |
| Shape/Size: | ◉ | ○ | ○ |

**Blob 1 is** *to the right of* ○ *to the left of* ○ *above* ○ *below* ○ **Blob 2.**

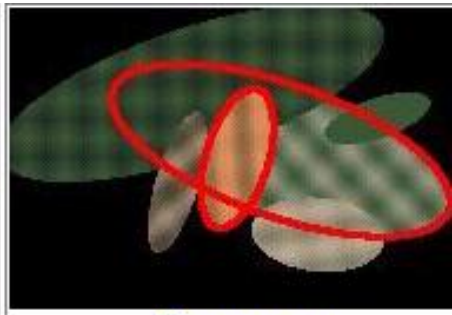
# Berkeley Blobworld

108004 (score = 0.86)



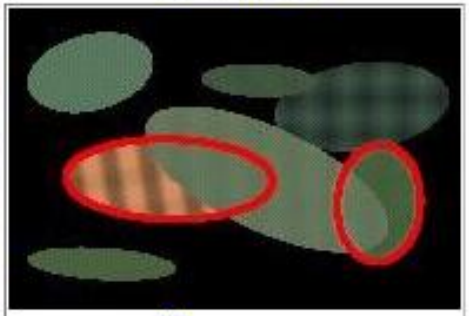
New query

108084 (score = 0.85)



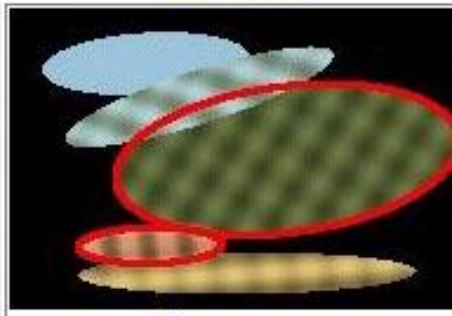
New query

108044 (score = 0.84)



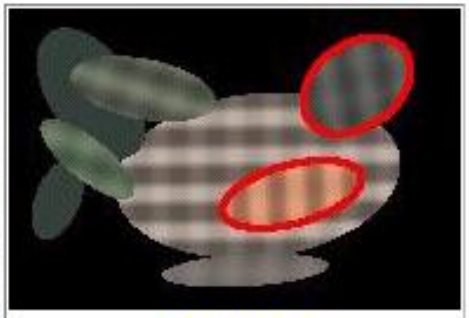
New query

97098 (score = 0.79)



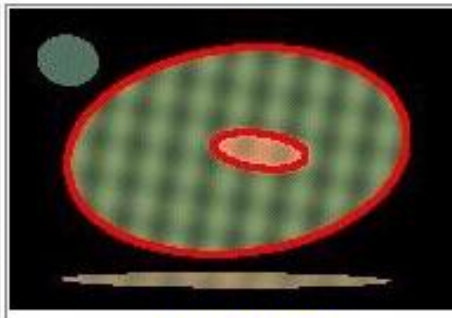
New query

108040 (score = 0.75)



New query

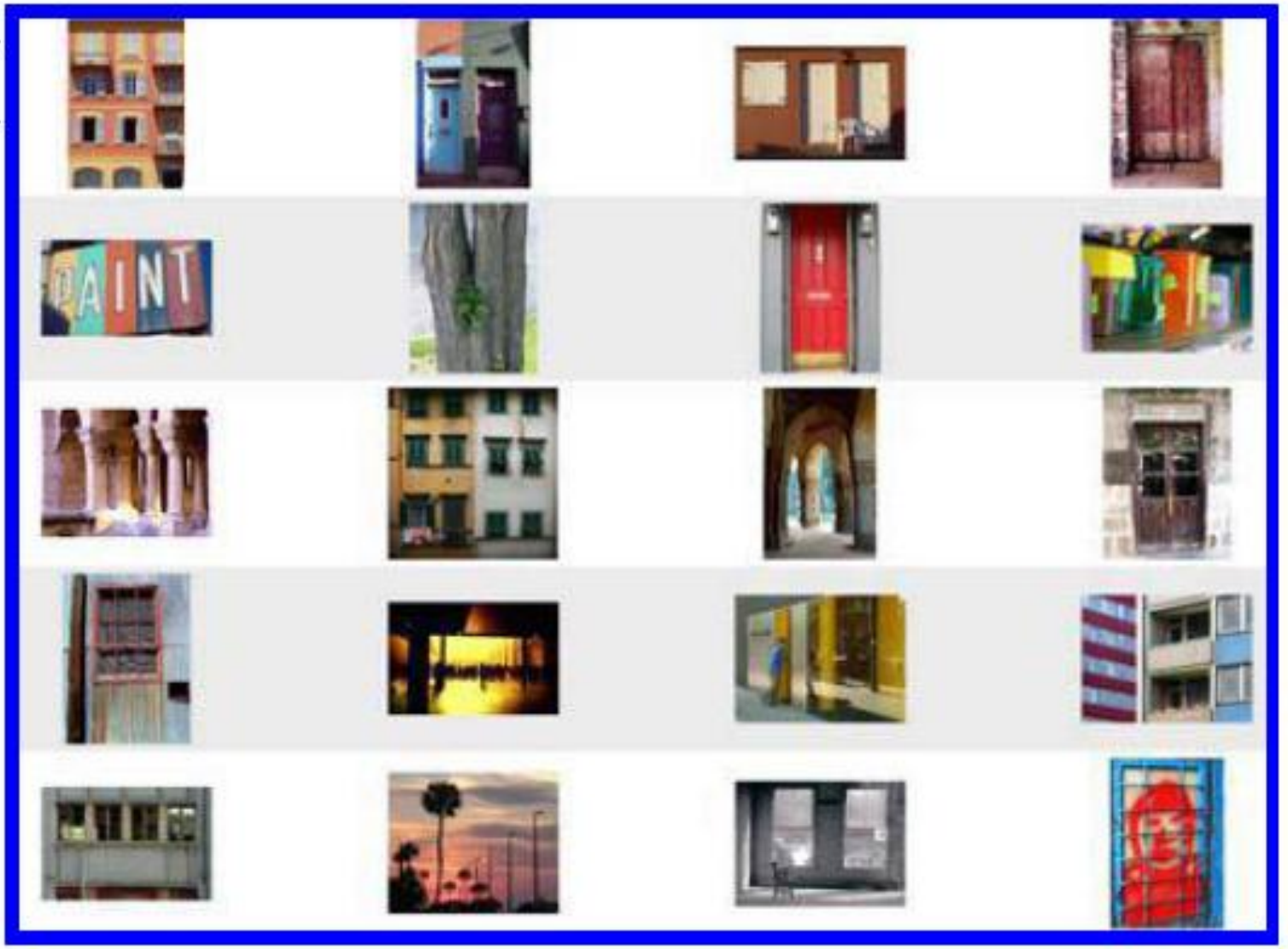
108058 (score = 0.75)



New query

# Automated Annotation

door



# Image Retrieval Summary

- Query
  - Keywords, example, sketch
- Matching
  - Caption text
  - Segmentation
  - Similarity (color, texture, shape)
  - Spatial arrangement (orientation, position)
  - Specialized techniques (e.g., face recognition)
- Selection
  - Thumbnails

# Try Some Systems

- Google Image Search (text)
  - http://images.google.com
- IBM QBIC (color, location)
  - http://wwwqbic.almaden.ibm.com/, select Hermitage

# Multimedia

- A set of time-synchronized modalities
  - Video
    - Images, object motion, camera motion, scenes
  - Audio
    - Speech, music, other sounds
  - Text
    - Closed captioning, on-screen captions, signs, …

# Multimedia Genres

- Television programs
  - News, sports, documentary, talk show, …
- Movies
  - Drama, comedy, mystery, …
- Meeting records
  - Conference, video teleconference, working group
- Others
  - Surveillance cameras, personal camcorders, …

# Video Structures

- Image structure
  - Absolute positioning, relative positioning
- Object motion
  - Translation, rotation
- Camera motion
  - Pan, zoom, perspective change
- Shot transitions
  - Cut, fade, dissolve, …

# Object Motion Detection

- Hypothesize objects as in image retrieval
  - Segment based on color and texture
- Examine frame-to-frame pixel changes
- Classify motion
  - Translation
    - Linear transforms model unaccelerated motion
  - Rotation
    - Creation & destruction, elongation & compression
  - Merge or split

# Camera Motion Detection

- Do global frame-to-frame pixel analysis

- Classify the resulting patterns
  - Central tendency -> zoom out
  - Balanced exterior destruction -> zoom in
  - Selective exterior destruction -> pan
  - Coupled rotation and translation -> perspective
    - Coupled within objects, not necessarily across them

# Shot-to-Shot Structure Detection

- Create a color histogram for each image
- Segment at discontinuities (cuts)
  - Cuts are easy, other transitions are also detectable
- Cluster representative histograms for each shot
  - Identifies cuts back to a prior shot
- Build a time-labeled transition graph

# Shot Classification

- Shot-to-shot structure correlates with genre
  - Reflects accepted editorial conventions
- Some substructures are informative
  - Frequent cuts to and from announcers
  - Periodic cuts between talk show participants
  - Wide-narrow cuts in sports programming
- Simple image features can reinforce this
  - Head-and-shoulders, object size, …

# Exploiting Multiple Modalities

- Video rarely appears in isolation
  - Sound track, closed captions, on-screen captions
- This provides synergy, not just redundancy
  - Some information appears in only one modality
- Image analysis complements video analysis
  - Face detection, video OCR

# Story Segmentation

- Video often lacks easily detected boundaries
  - Between programs, news stories, etc.
- Accurate segmentation improves utility
  - Too large hurts effectiveness, to small is unnatural
- Multiple segmentation cues are available
  - Genre shift in shot-to-shot structure
  - Vocabulary shift in closed captions
  - Intrusive on-screen text
  - Musical segues

# Closed Captions

- Designed for hearing-impaired viewers
  - Speech content, speaker id, non-speech audio
- Weakly synchronized with the video
  - Simultaneously on screen for advance production
  - Significant lag for live productions
- Missing text and significant errors are common
  - Automatic spelling correction can produce nonsense

# Aligning Closed Captions

- Speech and closed caption are redundant, but:
  - Each contains different types of errors
  - Each provides unique information
- Merging the two can improve retrieval
  - Start with a rough time alignment
  - Synchronize at points of commonality
    - Speech recognition provides exact timing
  - Use the words from both as a basis for retrieval
    - Learn which to weight more from training data

# On-Screen Captions

- On-screen captions can be very useful
  - Speaker names, event names, program titles, …
- They can be very challenging to extract
  - Low resolution, variable background
- But some factors work in your favor
  - Absolutely stable over multiple frames
  - Standard locations and orientations

# Video OCR

- Text area detection
  - Look for long thin horizontal regions
    - Bias towards classic text locations by genre
  - Integrate detected regions across multiple frames
- Enhance the extracted text
  - Contrast improvement, interpolation, thinning
- Optical character recognition
  - Matched to the font, if known

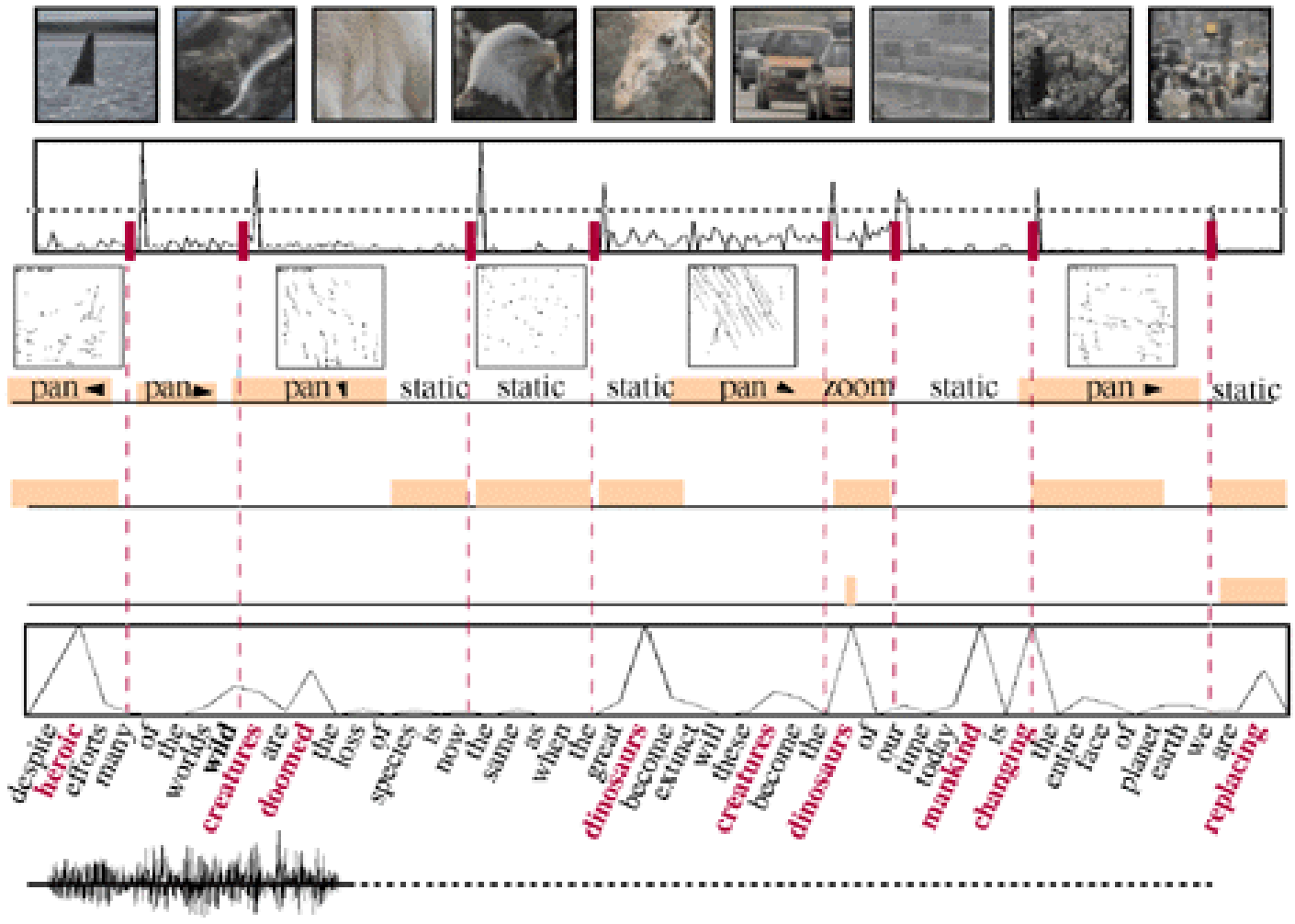
# Face Recognition

- Segment from images based on shape
  - Head, shoulders, and hair provide strong cues
- Track across several images
  - Using optical flow techniques
- Select the most directly frontal view
  - Based on eye and cheek positions, for example
- Construct feature vectors
  - “Eigenface” produces 16-element vectors
- Perform similarity matching

# Identity-Based Retrieval

- Face recognition and speaker identification
  - Both exploit information that is usually present
  - But both require training data
- On-screen captions provide useful cues
  - Confounded by OCR errors and varied spelling
- Closed captions and speech retrieval help too
  - If genre-specific heuristics are used
    - e.g., announcers usually introduce speakers before cuts

# Combined Technologies Integration



# Summary So Far

- Multimedia retrieval builds on all we know
  - Controlled vocabulary retrieval & social filtering
  - Text, image, speech and music retrieval
- New information sources are added
  - Video structure, closed & on-screen captions
- Cross-modal alignment adds new possibilities
  - One modality can make another more informative
  - One modality can make another more precise

# Video Selection Interfaces

- Each minute of video contains 1,800 frames
  - Some form of compaction is clearly needed
- Two compaction techniques have been used
  - Extracts select representative frames or shots
  - Abstracts summarize multiple frames
- Three presentation techniques are available
  - Storyboard, slide show, full motion

# Key Frame Extraction

- First frame of a shot is easy to select
  - But it may not be the best choice
- Genre-specific cues may be helpful
  - Minimum optical flow for director's emphasis
  - Face detection for interviews
  - Presence of on-screen captions
- This may produce too many frames
  - Color histogram clusters can reveal duplicates

CMU Informedia DVLS v. 1.06

File Edit Navigate Options Video Window Help Comments!

tell me about the evolution of species

Clear All! More Options... OR: Search

The results set shows the best 12 of 628 matches on any of "tell me about the evolution of species."

*Click on a word to focus on it. Press the shift key while clicking to have multi-word focus.*

Search Results


Darwin observed Galapagos and developed theor...

Resume > |<< Prev Hit 00047340 Next Hit >>| 0:00:42

<< Prev Para. << All >> Next Para. >> 0:57:11

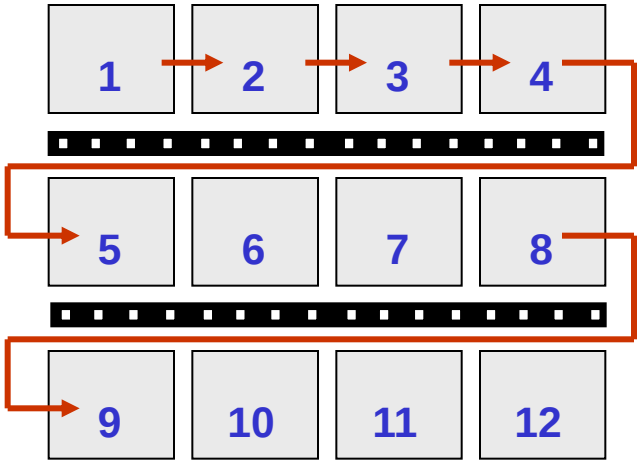
When Charles Darwin came here 150 years ago he was puzzled by the differences between similar animals from

# Salient Stills Abstracts

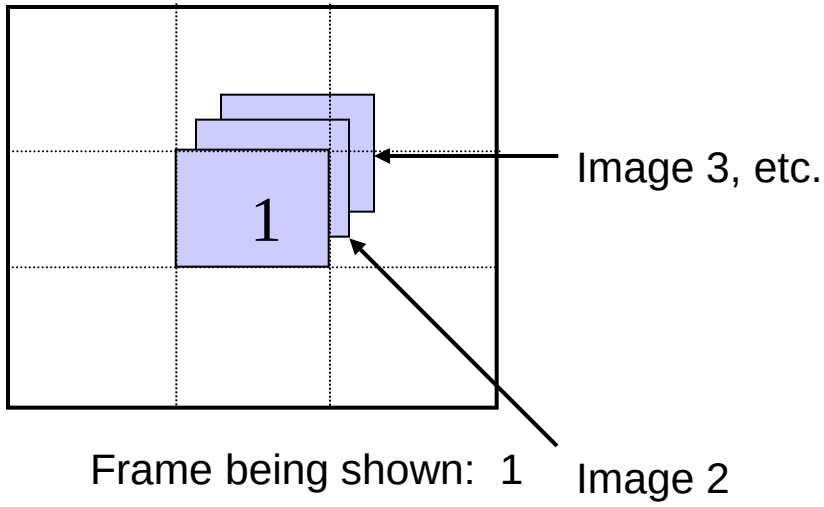
- Composite images that capture several scenes
  - And convey a sense of space, time, and/or motion
- Exploits familiar metaphors
  - Time exposures, multiple exposures, strobe, …
- Two stages
  - Modeling (e.g., video structure analysis)
  - Rendering
    - Global operators do time exposure and variable resolution
    - Segmentation supports production of composite frames

# Storyboards and Slide Shows

Storyboard
(Static)




Slide Show Interface
(Dynamic)



# Storyboards

- Spatial arrangement of still images
  - Linear arrangements depict temporal evolution
    - Overlapped depictions allow denser presentations
  - Graph can be used to depict video structure
    - But temporal relationships are hard to capture
- Naturally balances overview with detail
  - Easily browsed at any level of detail
- Tradeoff between detail and complexity
  - Further limited by image size and resolution

# Static Filmstrip Abstraction

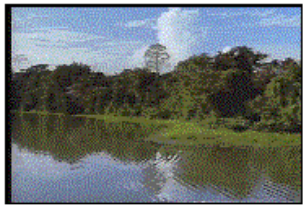
# Slide Shows

- Flip through still images in one spot
  - At a rate selected by the user
- Conserves screen space
  - But it is hard to process several simultaneously
- Several variations possible
  - Content-sensitive dwell times
  - Alternative frame transitions (cut, dissolve, …)

# Static vs. Dynamic Surrogates

**Click on the images to examine further images from the clip:**

1 of 4



2 of 4

3 of 4

4 of 4

Which clip(s) illustrate(s) best wildlife in and around rivers?

☐ #1 ☐ #2 ☐ #3 ☐ #4

NEXT

# Full Motion Extracts

- Extracted shots, joined by cuts
  - The technique used in movie advertisements
- Conveys more information using motion
  - Optionally aligned with extracted sound as well
- Hard to build a coherent extract
  - Movie ads are constructed by hand

# Project Presentations

- Five ~20-minute slots:
  - 15 minute presentation
  - 5 minutes for questions
- Two projectors
  - Laptop or second instructor console for system
  - Primary instructor console for slides

# Things to Discuss

- What you did
- Why you did it
- **Overview** of how you did it
- What you how about how well it works
  - Batch evaluation
  - User study
- Big things you learned