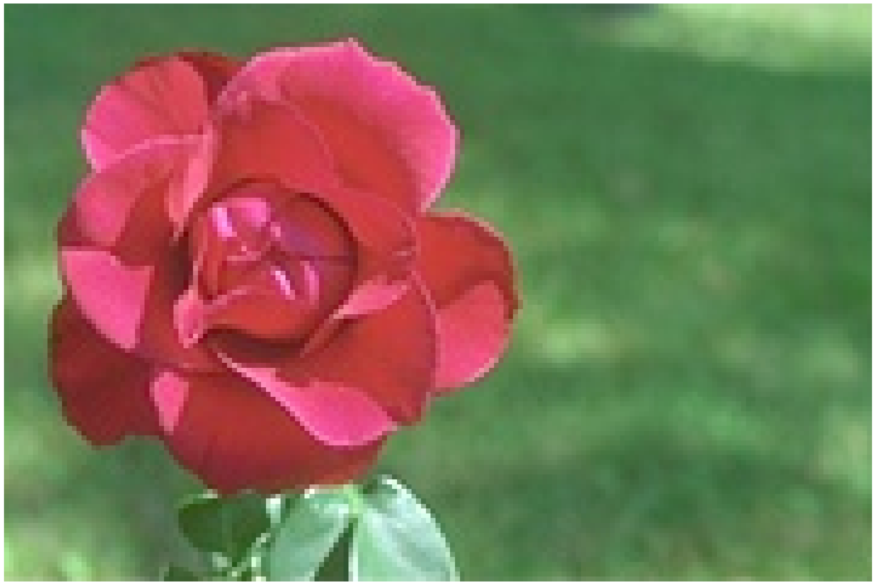
Linking Visual Features with Text for Multimedia Data Mining

Pinar Duygulu

Informedia Project, Carnegie Mellon University
Bilkent University, Turkey

Visual data with annotated text

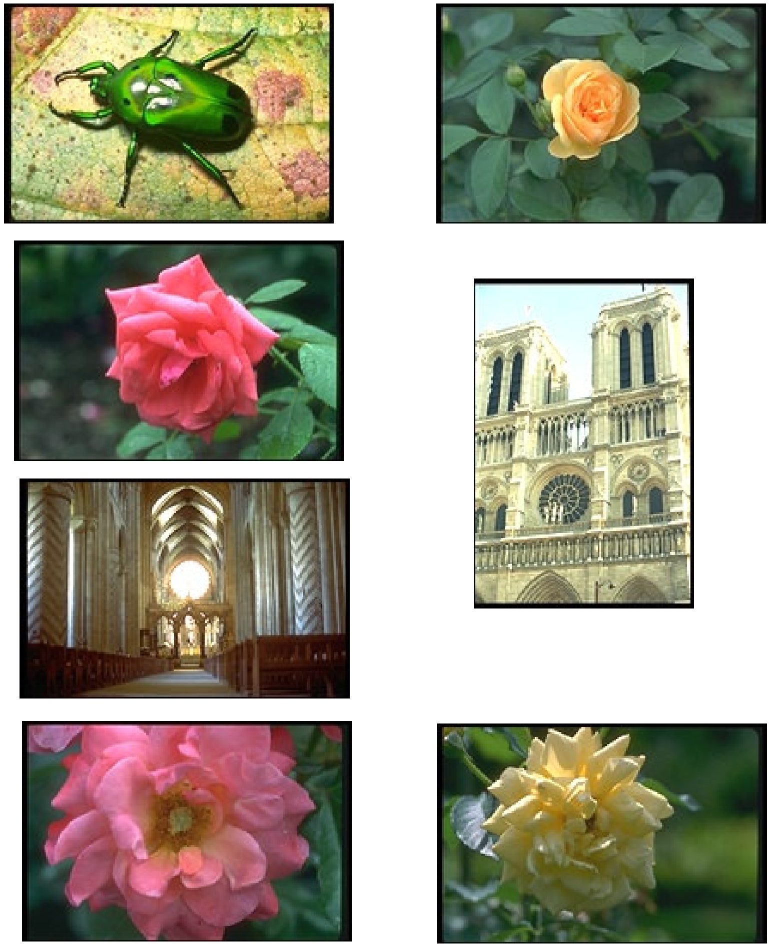


Keywords : rose flower plant leaves

Textual Query

Query on

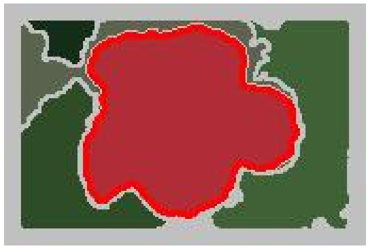
"Rose"



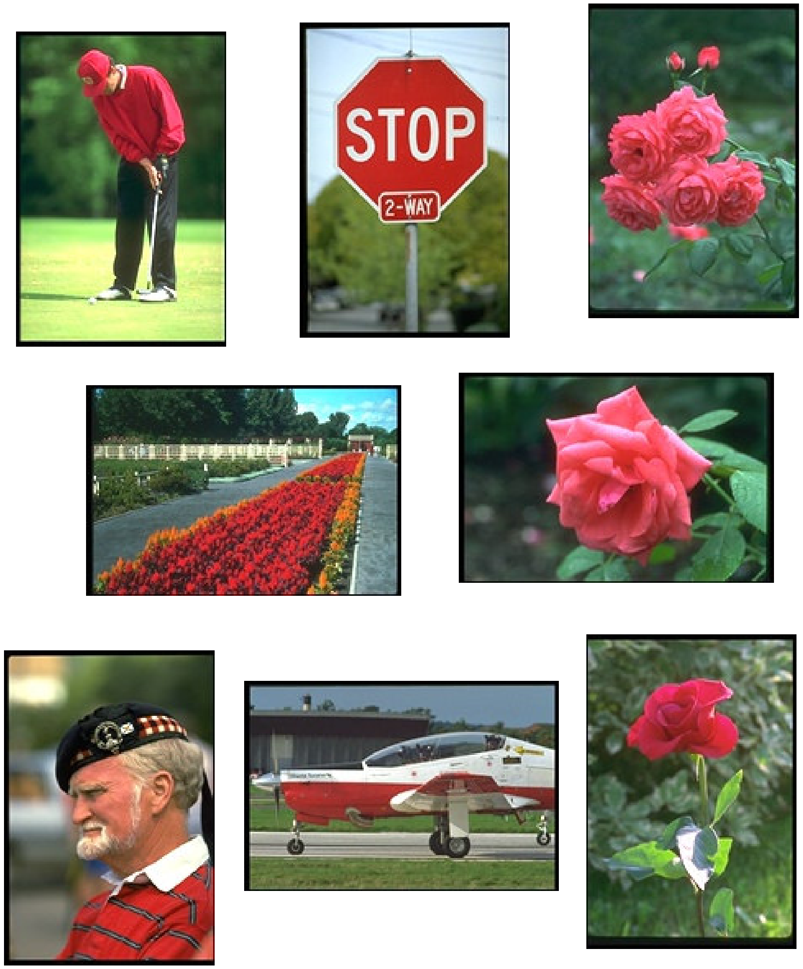
Example from Berkeley Blobworld system

Visual Query

Query on



Example from Berkeley Blobworld system

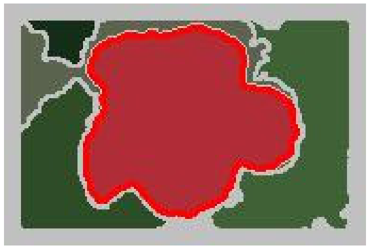


Query using both text and visual features

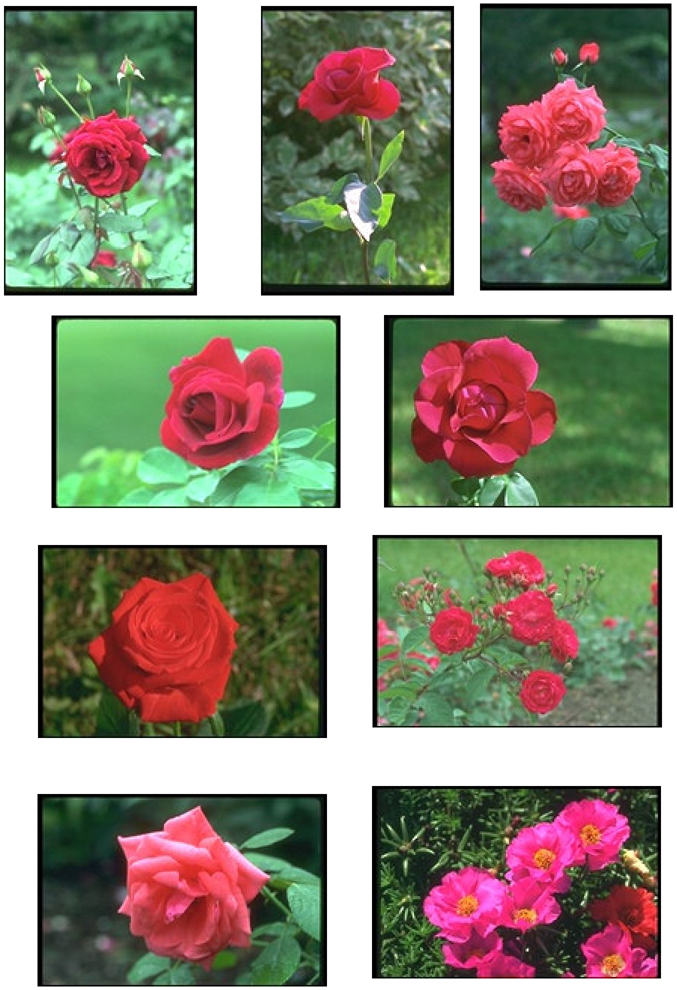
Query on

"Rose"

and

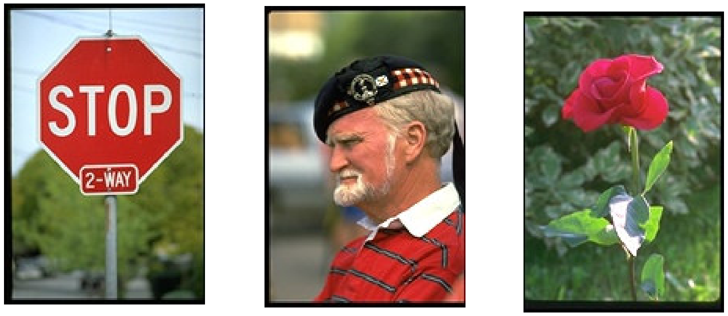


Example from Berkeley Blobworld system



Combination of text with visual features

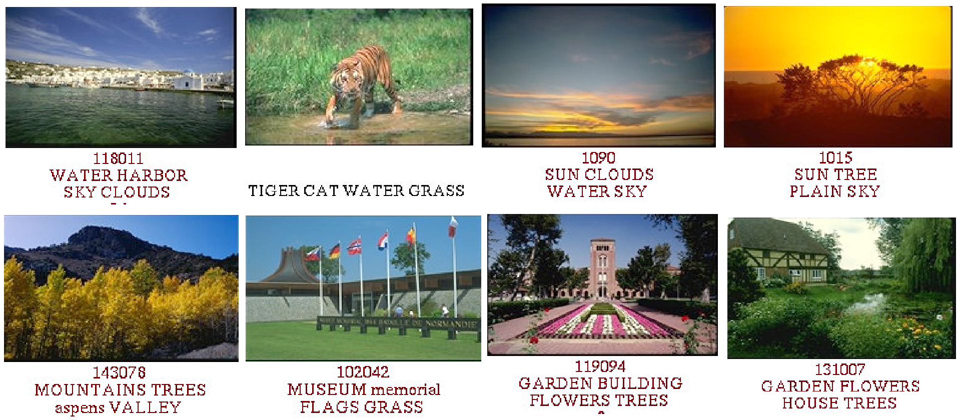
Appearance counts!



Semantics counts!

What can be done by combining text with features

- Information retrieval
- Browsing
- Auto illustration
- Auto annotation
- Multimedia translation

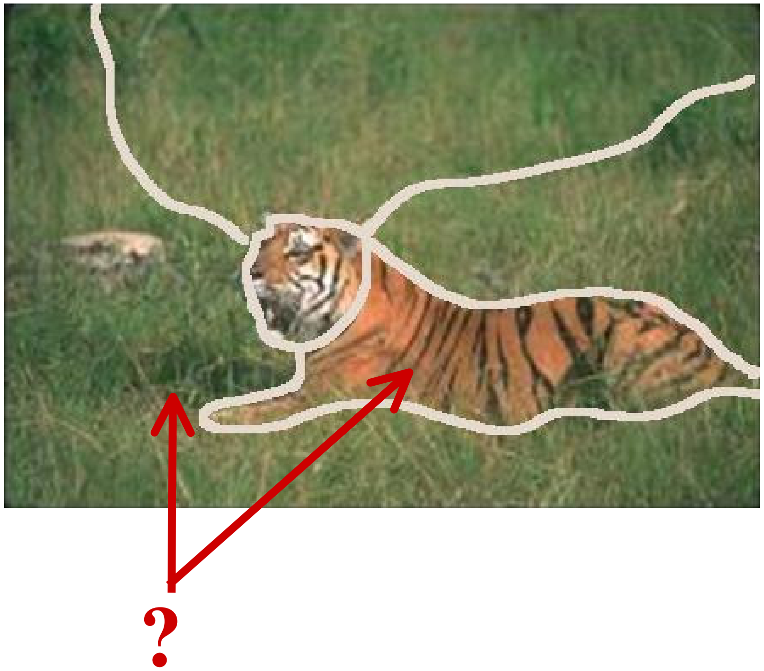


Annotated data sets

Annotation vs Recognition

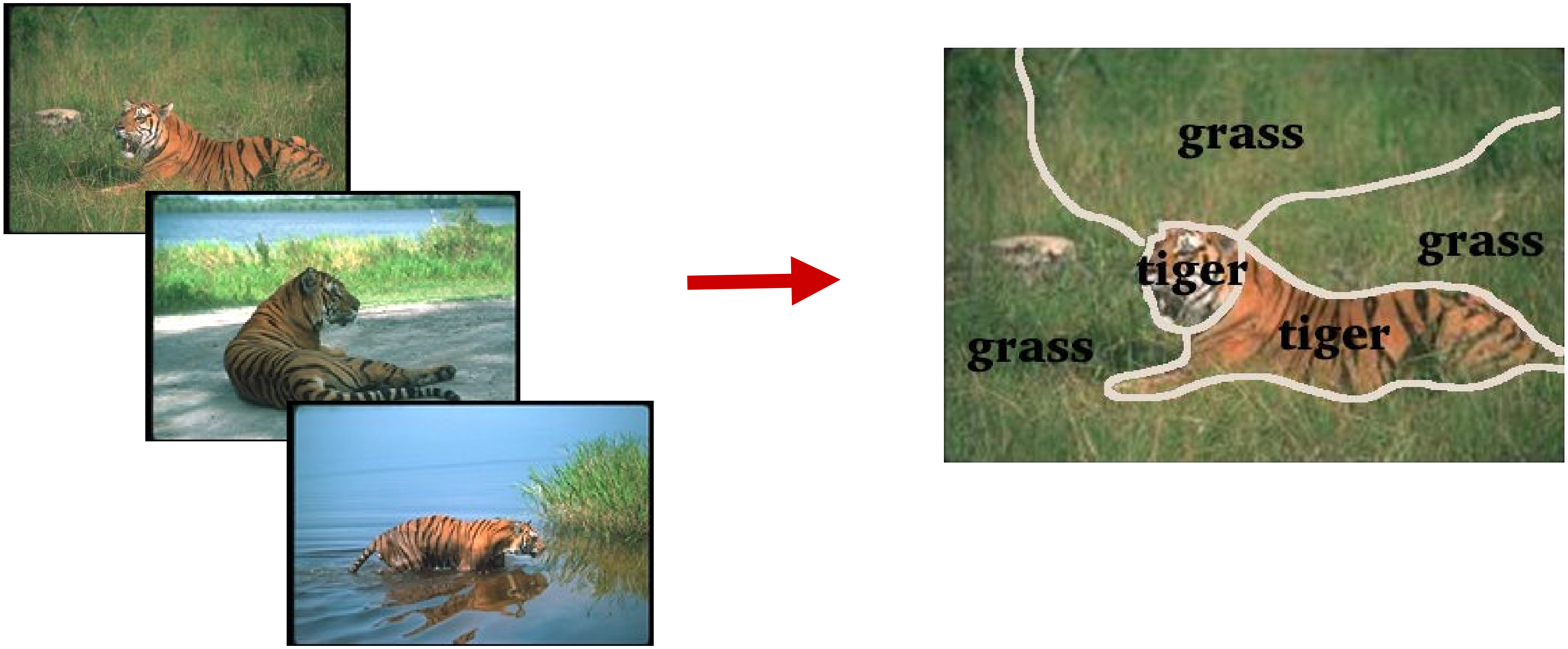
tiger grass cat

Cannot be learned from a single image



tiger grass cat

Learning recognition from large data

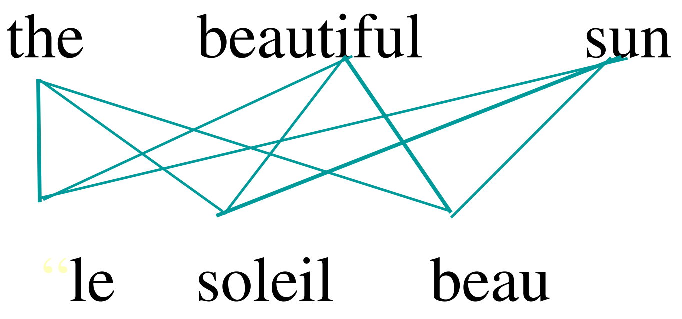


Object recognition on large scale is linking image regions with words

Use joint probability of words and images in large data sets.

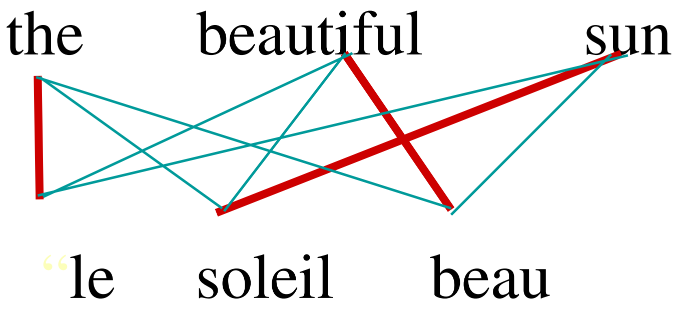
Statistical Machine Translation

Data : aligned sentences
But word correspondences
are unknown



•Given the correspondences, we can estimate the translation p(sun | soleil)

•Given the probabilities, we can estimate the correspondences



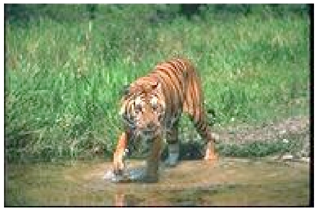
Solution: enough data + EM

Brown et. al 1993

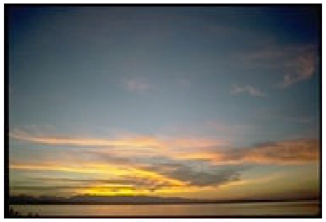
Multimedia Translation

Data :

118011
WATER HARBOR
SKY CLOUDS



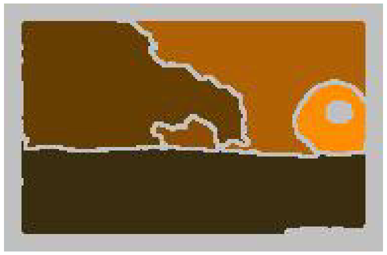
TIGER CAT WATER GRASS



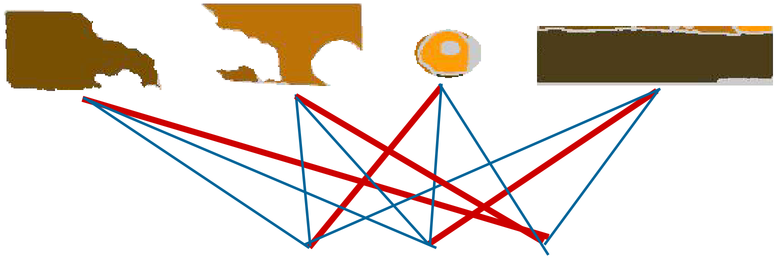
1090
SUN CLOUDS
WATER SKY

Words are associated with the images

But correspondences between image regions and words are unknown



“sun sea sky”



“sun sea sky”

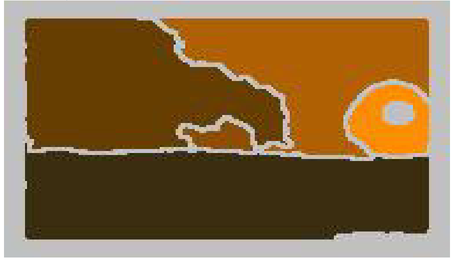
Duygulu et.al, ECCV 2002

Input Representation

sun sky waves sea

word tokens

↓ segmentation



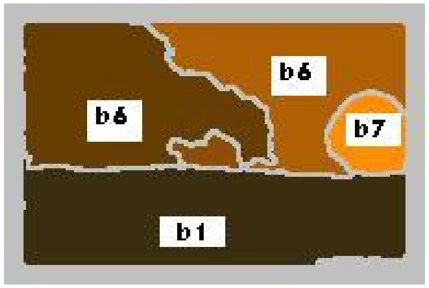
Each blob is a large vector of features

- Region size
- Position
- Colour
- Oriented energy (12 filters)
- Simple shape features

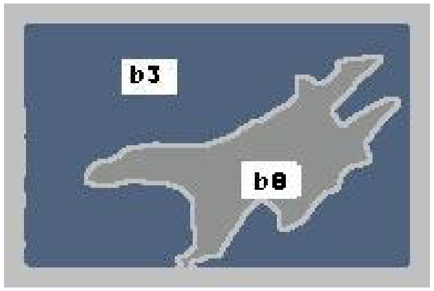
k-means to cluster features

For each blob label of the

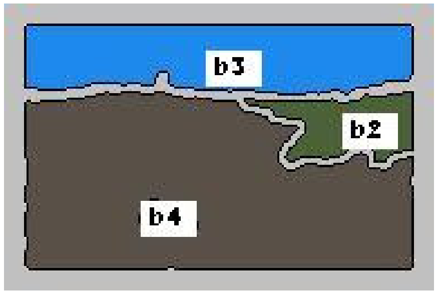
Closest cluster→ blob tokens



w6 w7 w8 w1

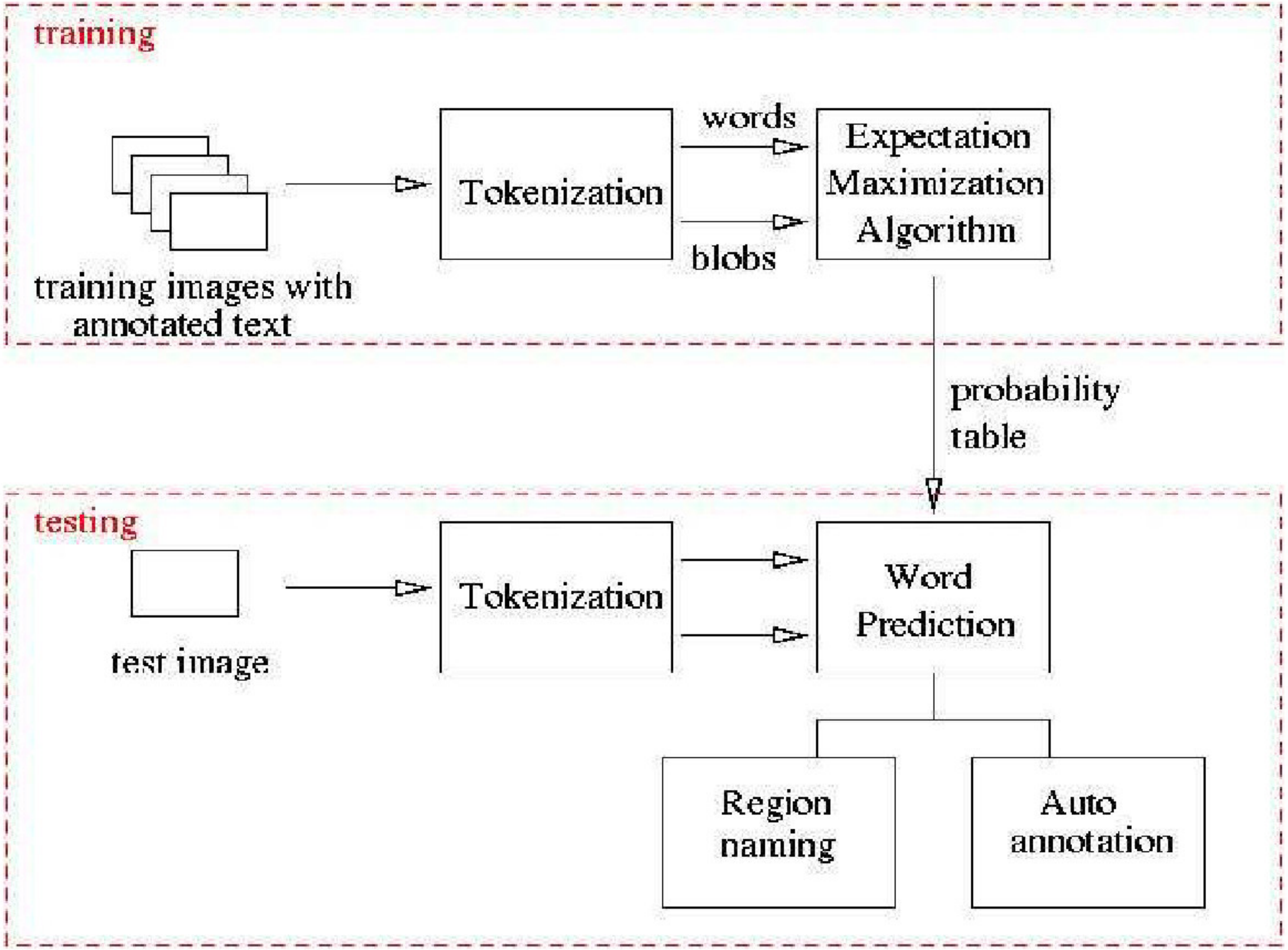


w3 w4 w5 w1

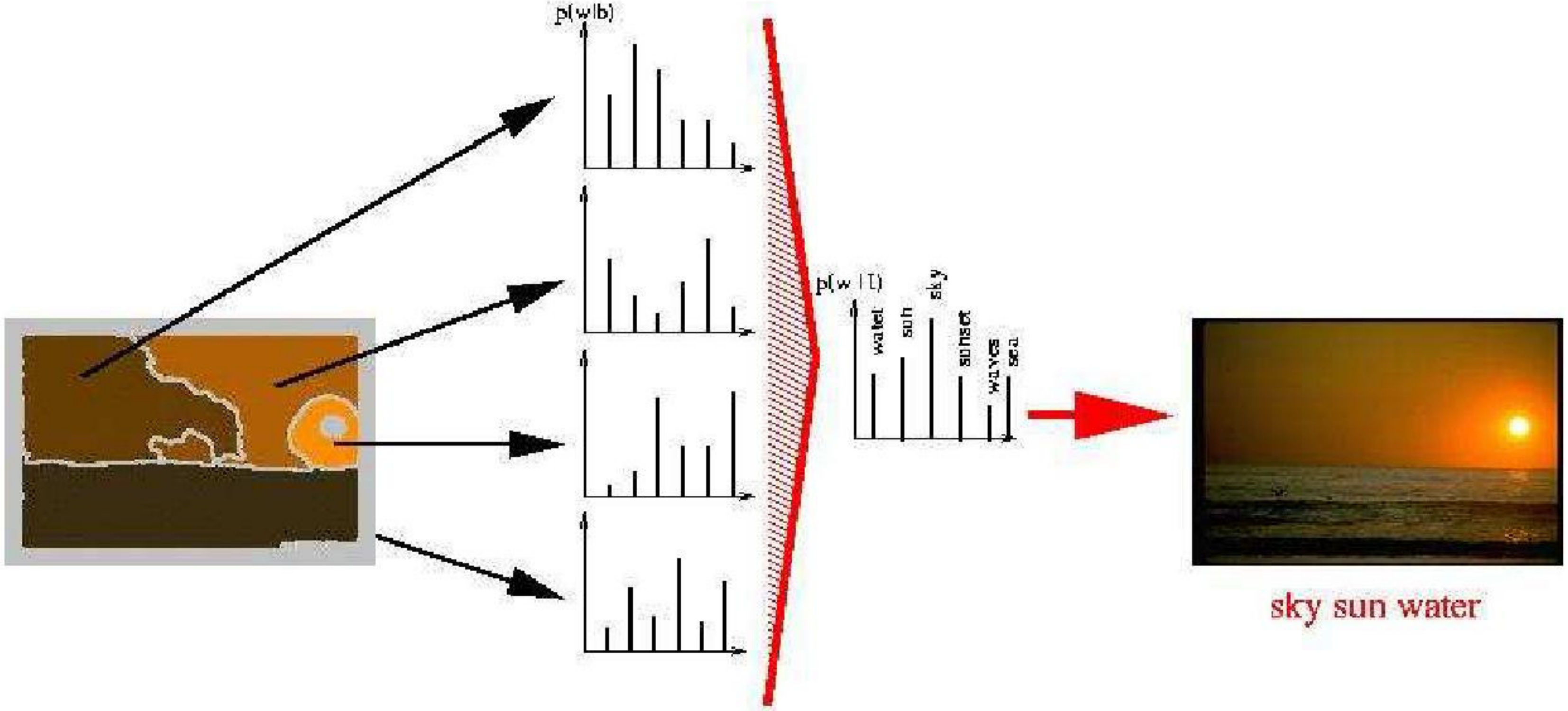


w12 w2 w1

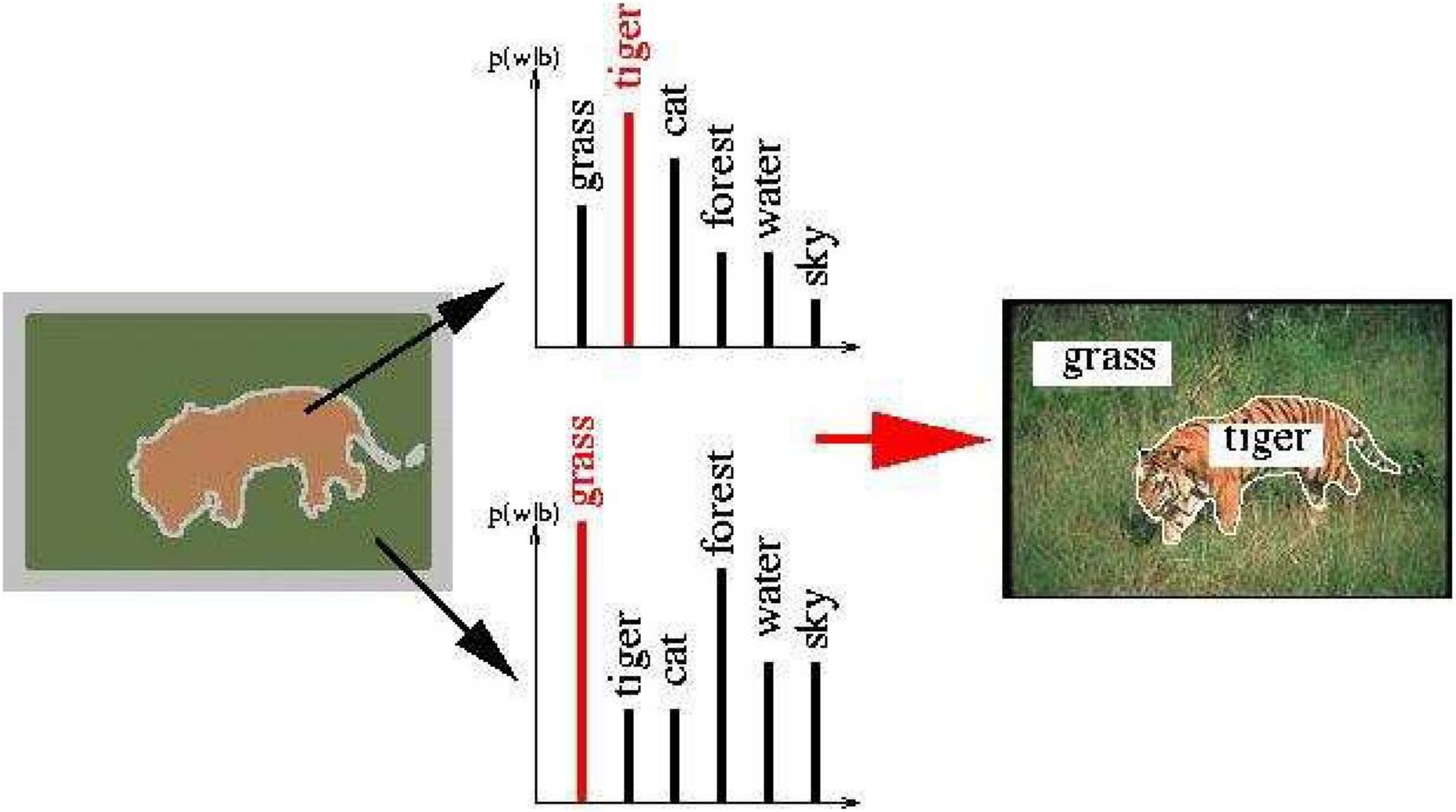
Overview of the system



Auto-Annotation

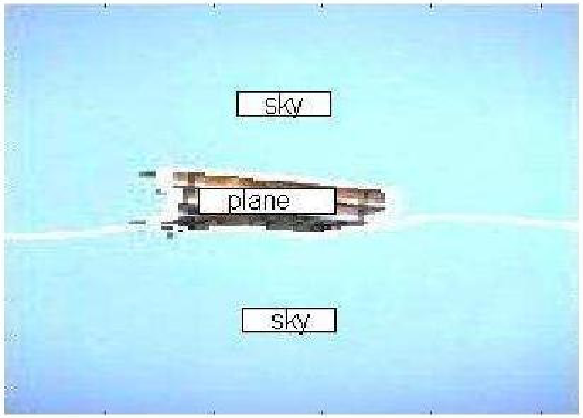


Region Naming

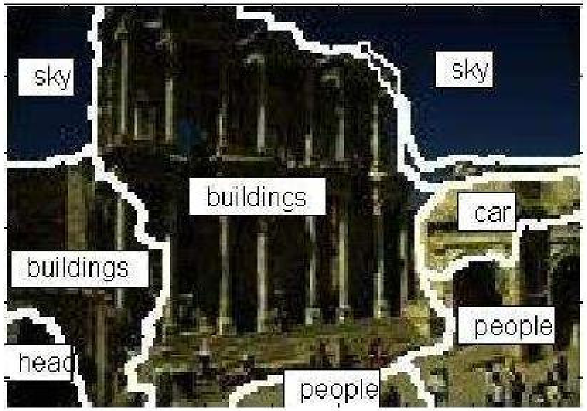


Results

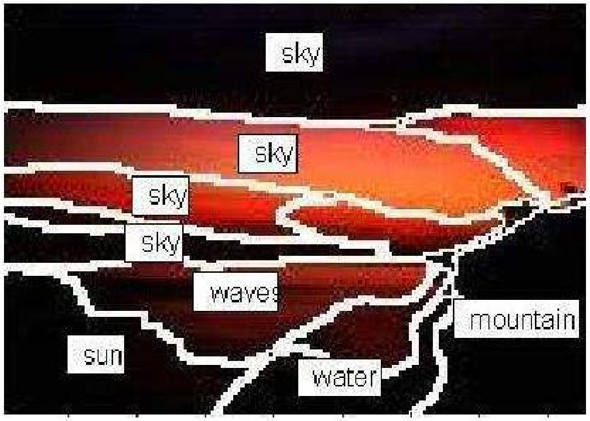
plane sky

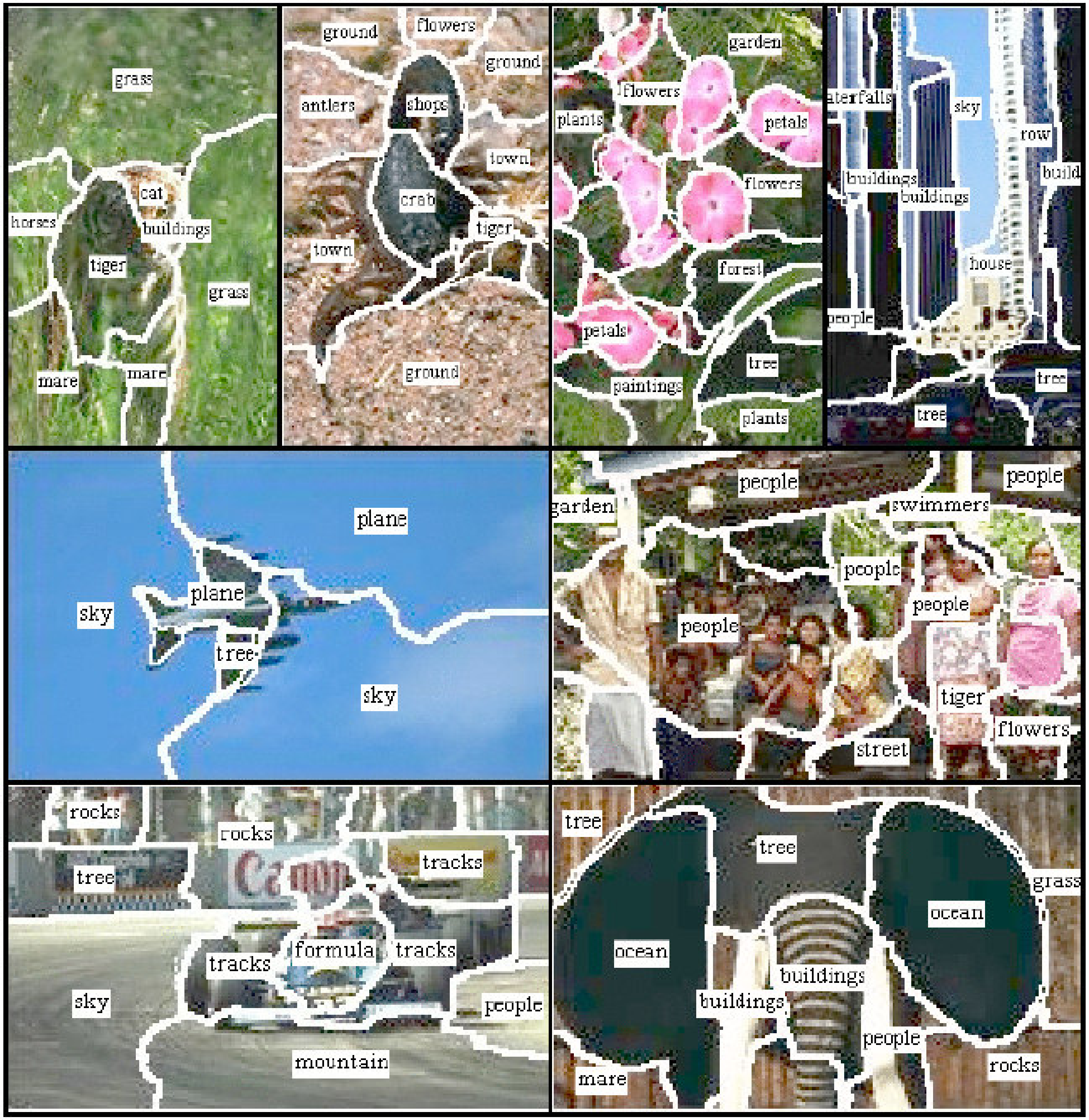


people ruins stone



sunset tree water




grass
cat
horses
buildings
tiger
grass
mare
mare
ground
flowers
ground
antlers
shops
town
crab
tiger
town
ground
garden
flowers
plants
petals
flowers
forest
petals
tree
paintings
plants
aterfalls
sky
row
build
buildings
buildings
house
people
tree
tree
plane
plane
sky
tree
sky
people
people
garden
swimmers
people
people
people
tiger
flowers
street
rocks
rocks
tracks
tree
formula
tracks
tracks
sky
people
mountain
tree
tree
grass
ocean
ocean
buildings
buildings
people
rocks
mare

Model Selection

Model for joint probability of text and blobs

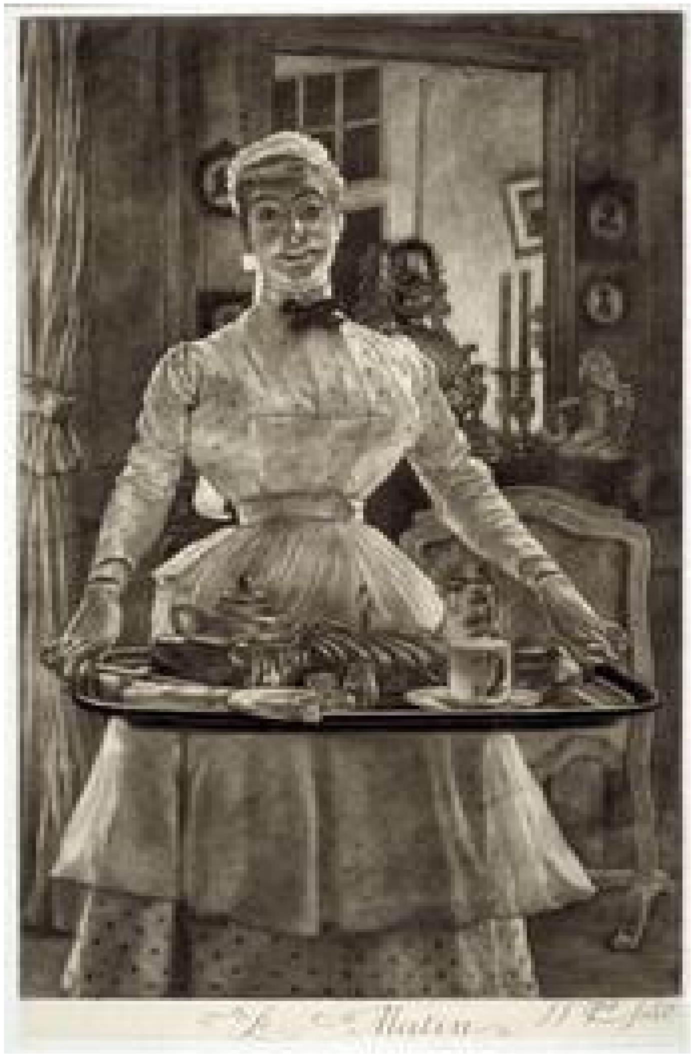
- Clustering models
- Aspect models
- Hierarchical models
- Bayesian models
- Co-occurrence models

Many of these based on models proposed for text [Brown, Della Pietra, Della Pietra & Mercer 93; Hofmann 98; Hofmann & Puzicha 98]
A comparison paper is published in JMLR
‘Matching words and Pictures’, Barnard, Duygulu, Forsyth, Freitas, Blei, Jordan

Other data sets

Corel Image Data	40,000 images
Fine Arts Museum of San Francisco	83,000 images online
Cal-flora	20,000 images, species information
News photos with captions (yahoo.com)	1,500 images per day available from yahoo.com
Hulton Archive	40,000,000 images (only 230,000 online)
internet.archive.org	1,000 movies with no copyright
TV news archives (televisionarchive.org, informedia.cs.cmu.edu)	Several terabytes already available
Google Image Crawl	>330,000,000 images (with nearby text)
Satellite images (terrarserver.com, nasa.gov, usgs.gov)	(And associated demographic information)
Medical images	(And associated with clinical information)

FAMSF Data (83,000 images online)



Web number: 4359202410830012

rec number: 2

Title: Le Matin

Primary class: Print

Artist: Tissot

Description:
serving woman stands in a dressing room, in front of vanity with chair, mirror and mantle, holding a tray with tea and toast

Display date: 1886

Country: France

Pictures from words (Auto-Illustration)

Text Passage (Moby Dick)

"The large importance attached to the harpooneer's vocation is evinced by the fact, that originally in the old Dutch Fishery, two centuries and more ago, the command of a whale-ship …"

Extracted Query

large importance attached fact old dutch century more command whale ship was person was divided officer word means fat cutter time made days was general vessel whale hunting concern british title old dutch ...

Retrieved Images

PRINT NAVAL BATTLE JAPANESE SHIP CHINESE BEING SHIP WATER

PRINT SHIP SURROUNDED ICE SEVERAL SHIP SEEN WHALE OTHER CURRIER

PRINT ATTACK WAGON ROAD FOREST CALLOT

PRINT WAR FRIGATE UNITED STATE ENGLISH SHIP AMERICAN SHIP CURRIER

VIEW GREEK CHURCH DRAWING D'OYLEY

PRINT SMALL BOAT APPROACHING BLOWING WHALE SHIP MOUNTAIN BACKGROUND CURRIER

PLAY BOAT PRINT KUNISADA

PRINT MEN SMALL MOUNTAIN HAS COME SEVERAL SMALL FOREGROUND POLITICAL

PRINT WHITE HOUSE GROUNDS BACKGROUND POLITICAL TYPE INDIAN ARMS TREE

PRINT FIGURE STANDING DOORWAY HORSE TRAP FOREGROUND WHISTLER

PRINT FISHING BOAT BEACH

PRINT WINTER LOW LAND COUNTRY HORSE DRAWN TENT HORIZON WINDMILL

PRINT RESIDENCE HAS LONG BEEN SIDE WOODLAND LAKE GROVE FOREGROUND

PRINT WARSHIP GUN PORT OPEN SHIP CASTLE GARDEN CURRIER

PRINT LOADED CART BEING PULLED LOUIS PHILIPPE PUSHED JEAN CHARLES

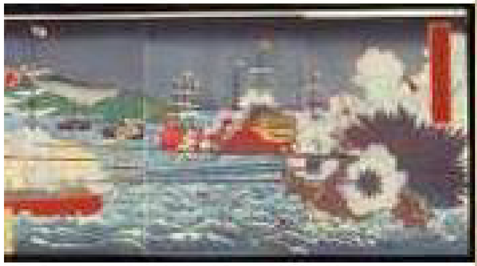
PRINT NIGHT FULL MOON PADDLE WHEEL BOAT JAMES RACE CURRIER_AND_IVES

DRAWING VIEW NEW YORK CITY BRIDGE WATER VARIOUS SAILING SHIP

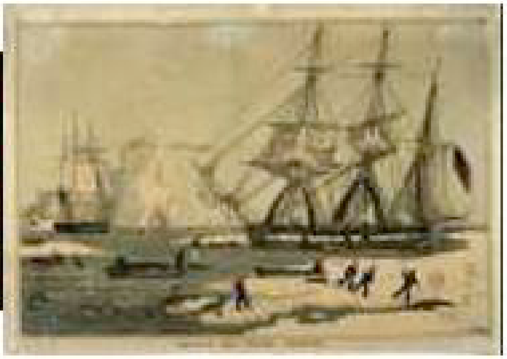
PRINT PEOPLE LARGE SQUARE RIVER BRIDGE TURBANED MEN FIGURE SLAVE

PRINT PEOPLE LARGE SQUARE RIVER BRIDGE TURBANED MEN FIGURE SLAVE

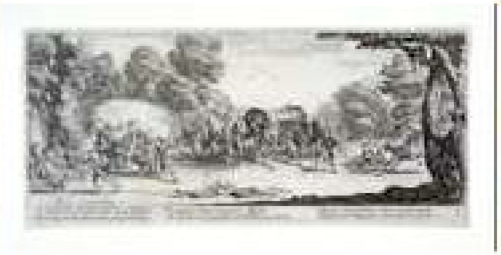
PRINT FISHERMAN ROWBOAT LOBSTER TRAP LOBSTER ROUGH WATERS DISTANCE MAST



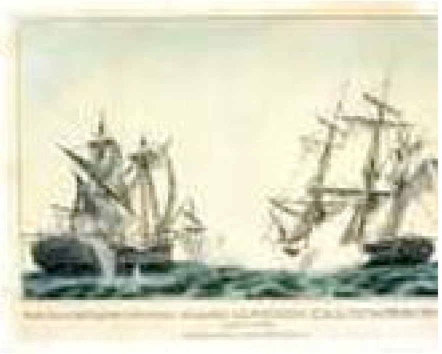
PRINT NAVAL BATTLE JAPANESE SHIP CHINESE BEING SHIP WATER



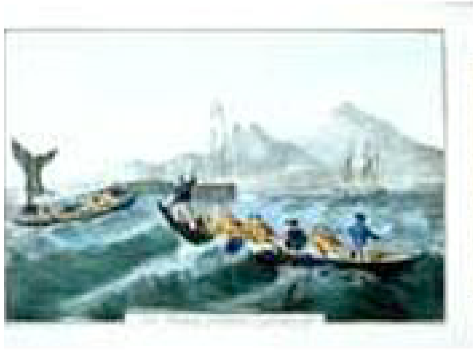
PRINT SHIP SURROUNDED ICE SEVERAL SHIP SEEN WHALE OTHER CURRIER



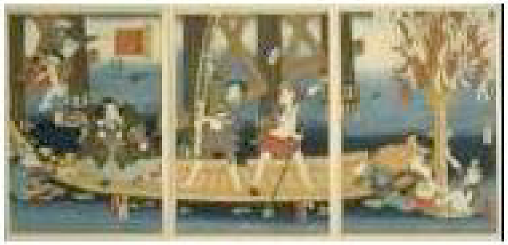
PRINT ATTACK WAGON ROAD FOREST CALLOT



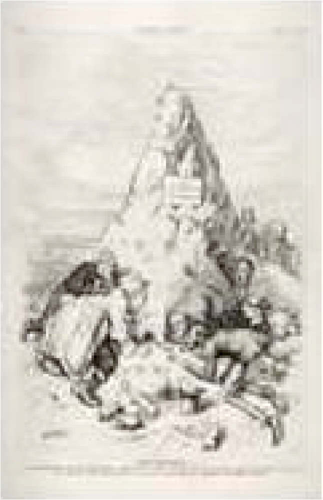
PRINT WAR FRIGATE UNITED STATE ENGLISH SHIP AMERICAN SHIP CURRIER



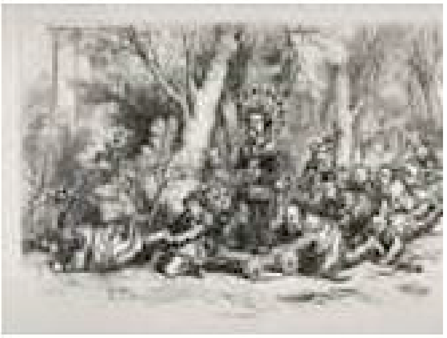
PRINT SMALL BOAT APPROACHING BLOWING WHALE SHIP MOUNTAIN BACKGROUND CURRIER



PLAY BOAT PRINT KUNISADA

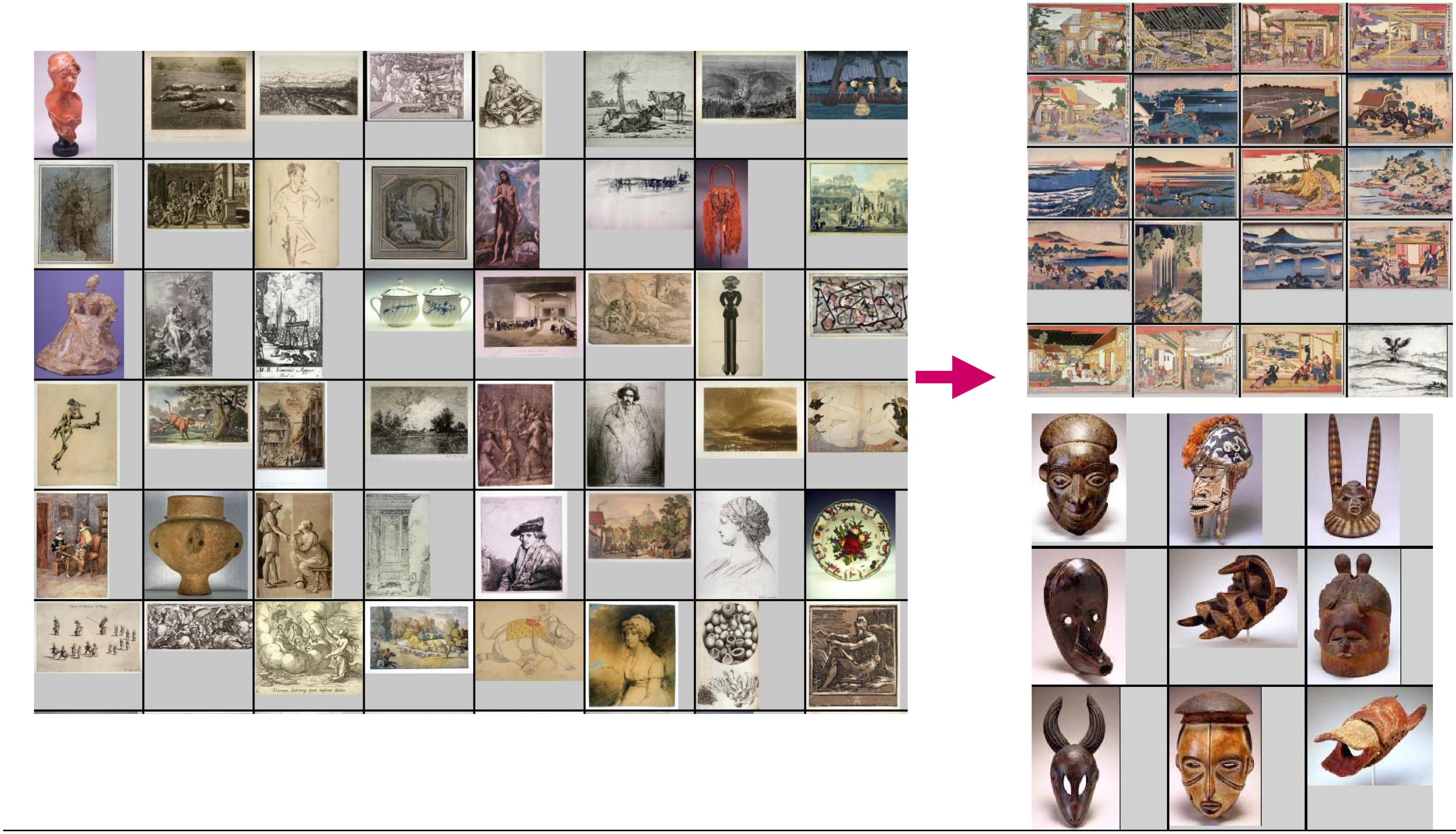


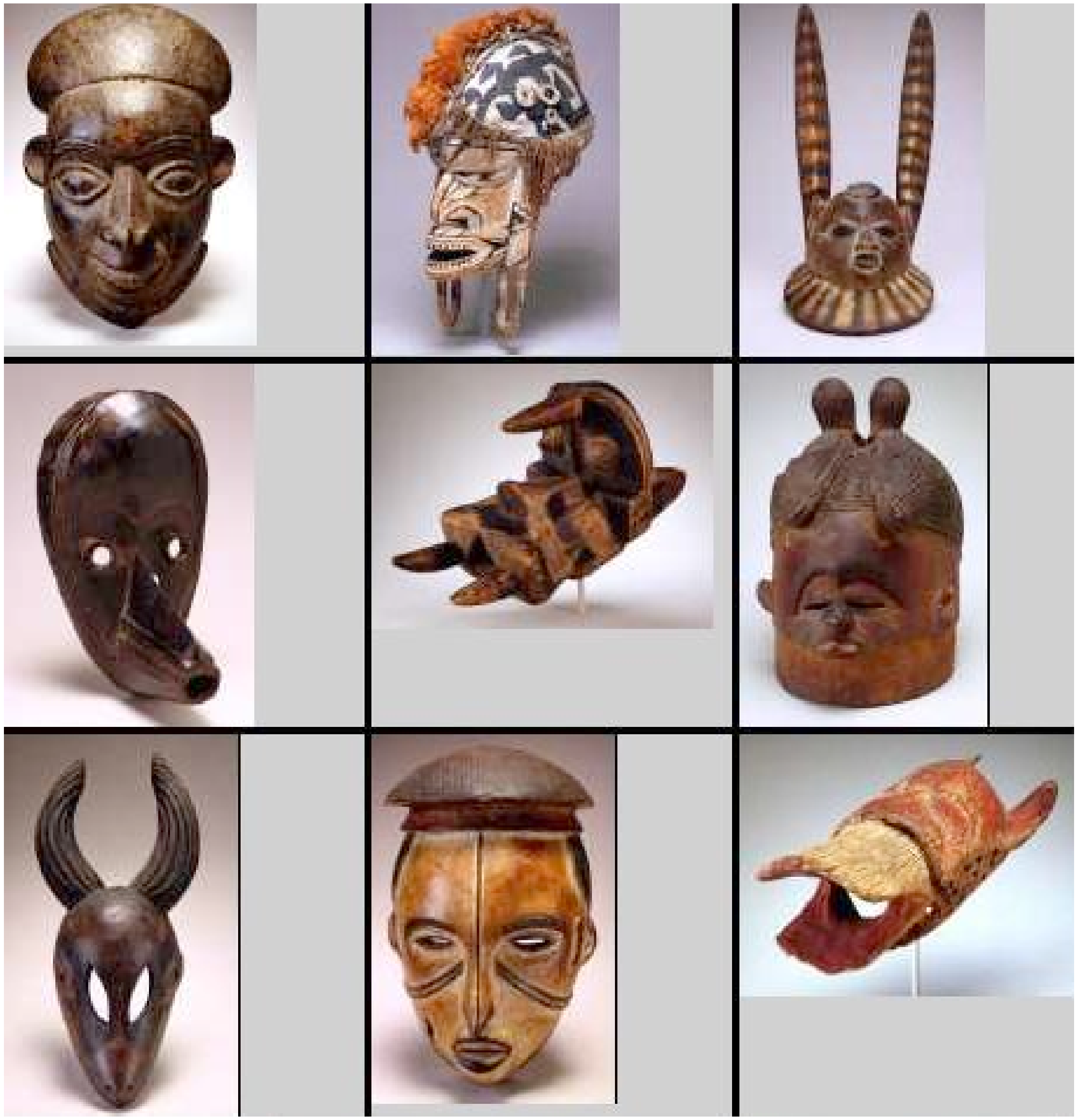
PRINT MEN SMALL MOUNTAIN HAS COME SEVERAL SMALL FOREGROUND POLITICAL

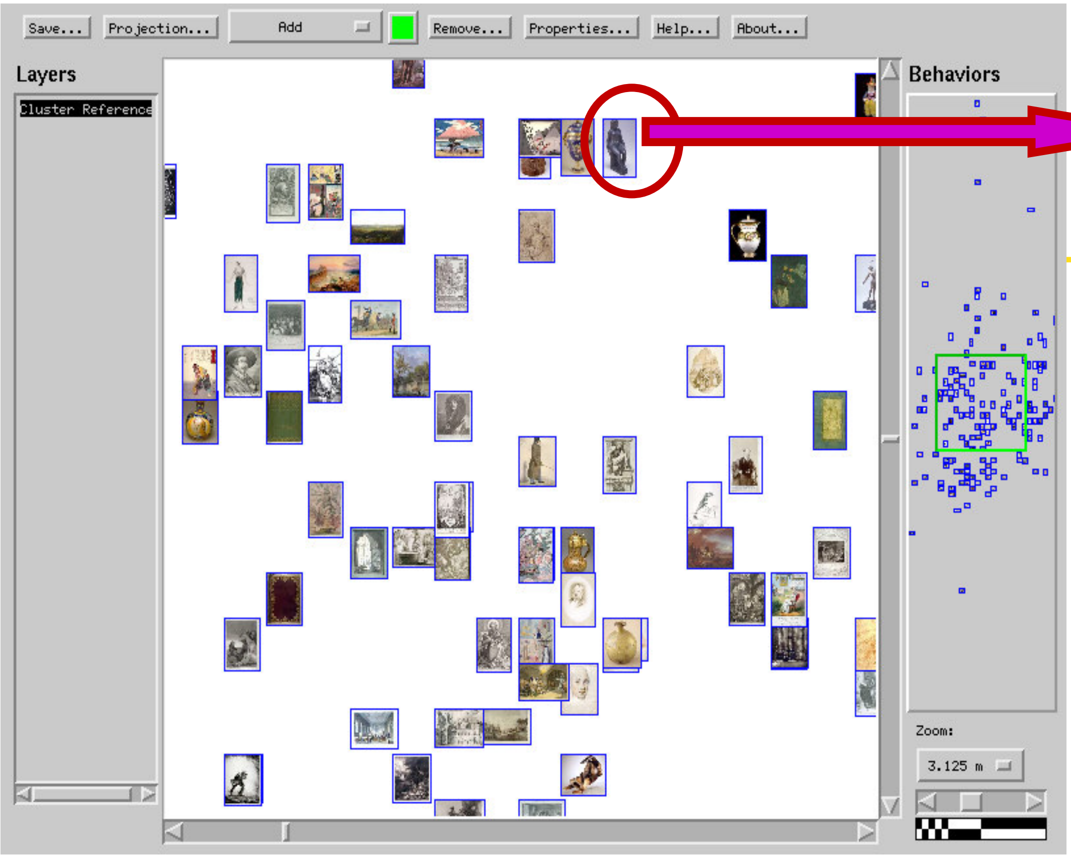


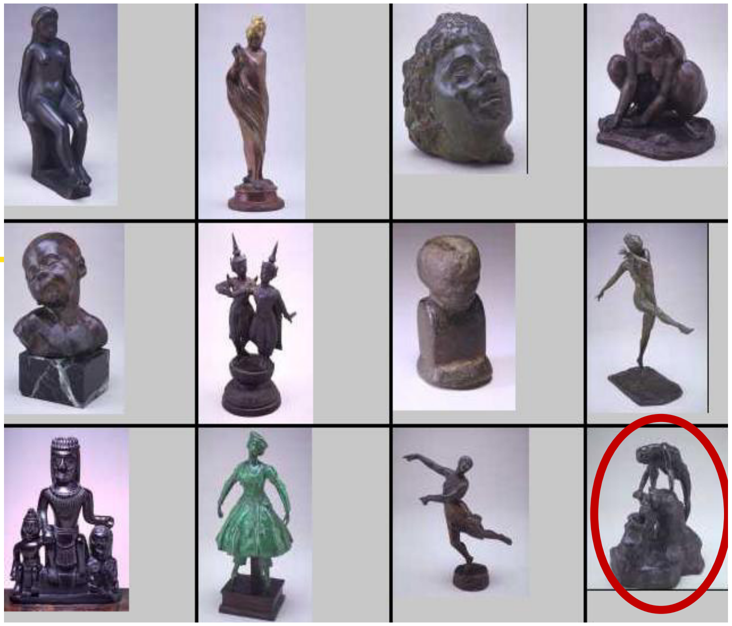
PRINT WHITE HOUSE GROUNDS BACKGROUND POLITICAL TYPE INDIAN ARMS TREE

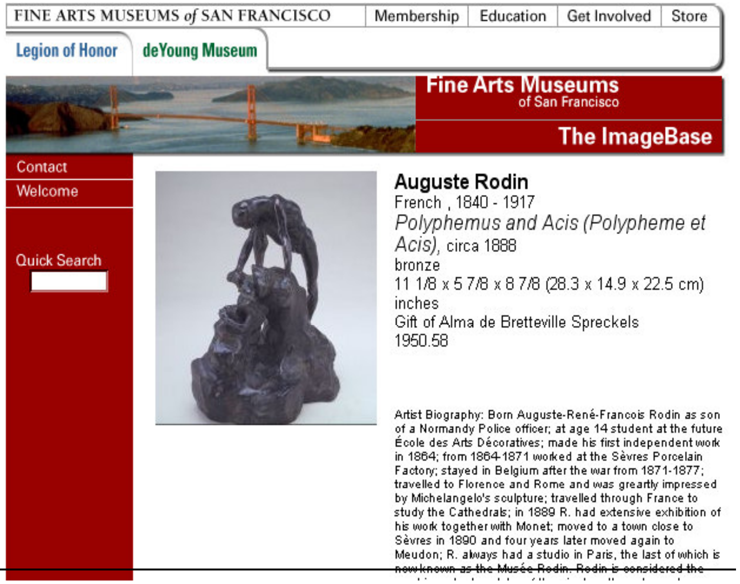
Organizing Image Collections







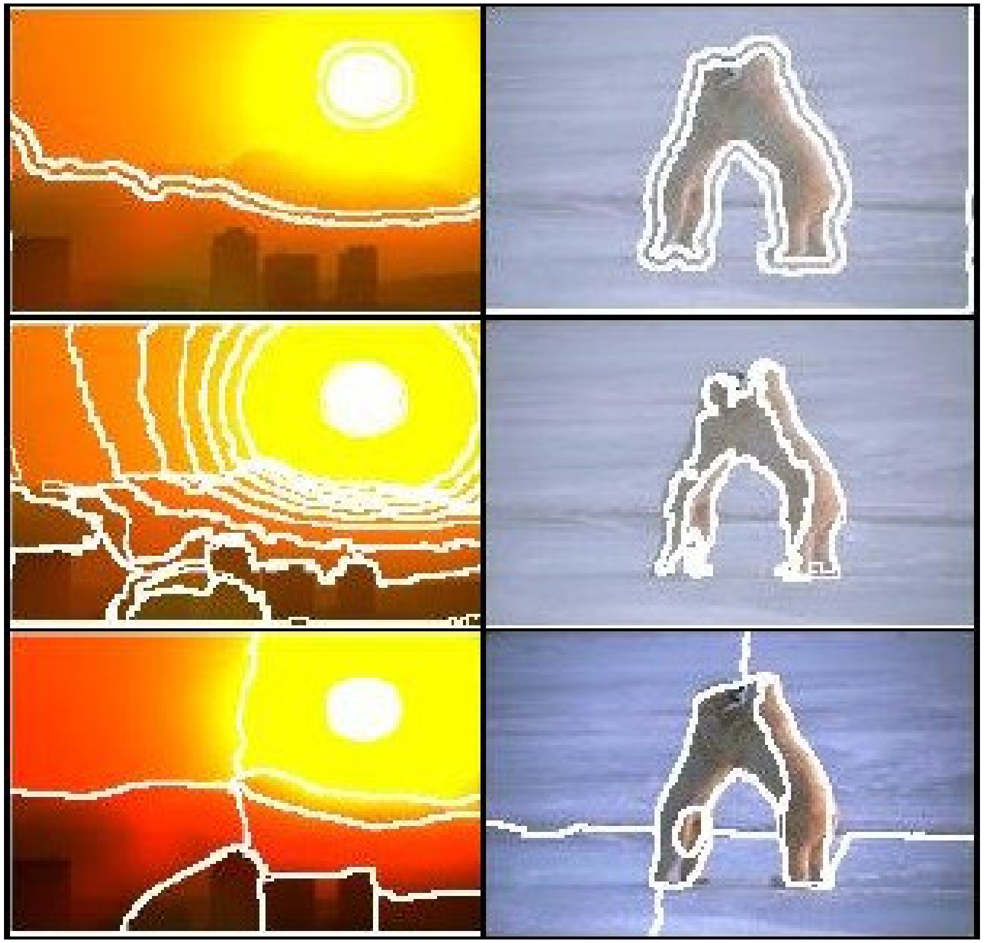




Auguste Rodin
French , 1840 - 1917
Polyphemus and Acis (Polypheme et Acis), circa 1888
bronze
11 1/8 x 5 7/8 x 8 7/8 (28.3 x 14.9 x 22.5 cm) inches
Gift of Alma de Bretteville Spreckels
1950.58

Artist Biography: Born Auguste-René-Francois Rodin as son of a Normandy Police officer; at age 14 student at the future École des Arts Décoratives; made his first independent work in 1864; from 1864-1871 worked at the Sèvres Porcelain Factory; stayed in Belgium after the war from 1871-1877; travelled to Florence and Rome and was greartly impressed by Michelangelo's sculpture; travelled through France to study the Cathedrals; in 1889 R. had extensive exhibition of his work together with Monet; moved to a town close to Sèvres in 1890 and four years later moved again to Meudon; R. always had a studio in Paris, the last of which is nowknown as the Musée Rodin. Rodin is considered the

Evaluating Segmentation Algorithms



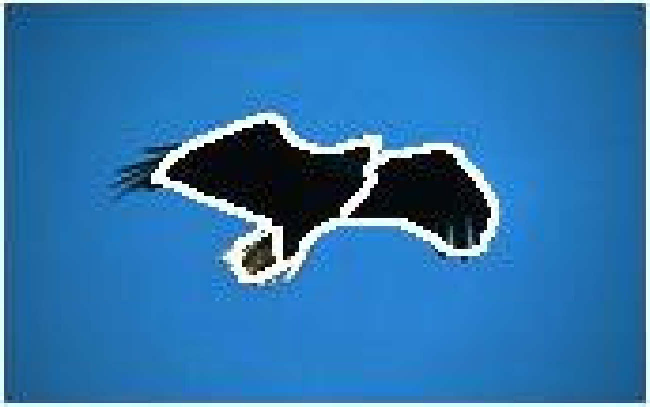
Blobworld

Mean-Shift

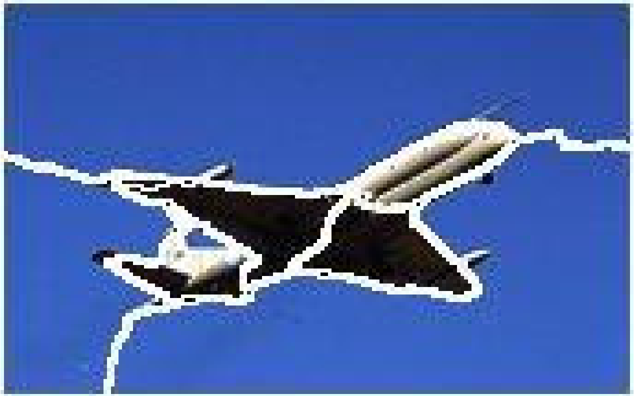
Normalized cuts

Feature Selection

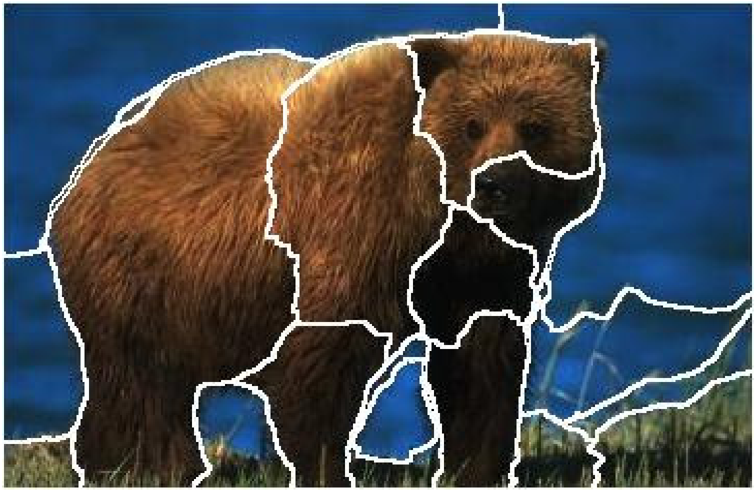
Propose good features to differentiate words that are not distinguishable (e.g., eagle and jet)

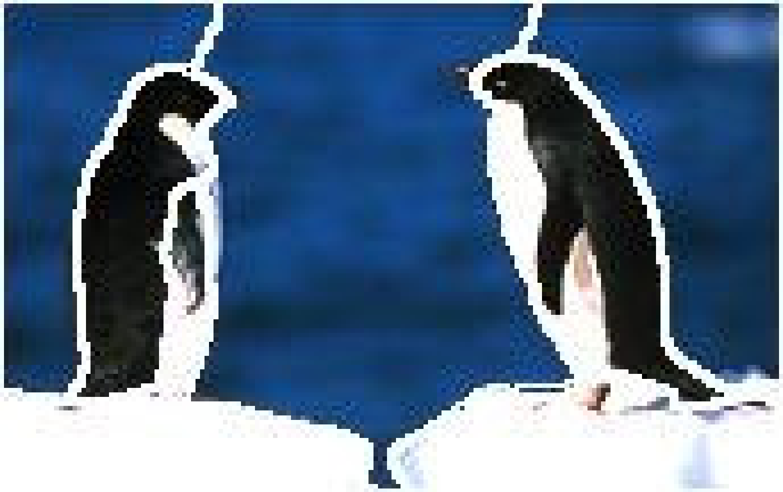


On Corel data set color is the dominant feature



Merging Regions with Word Prediction



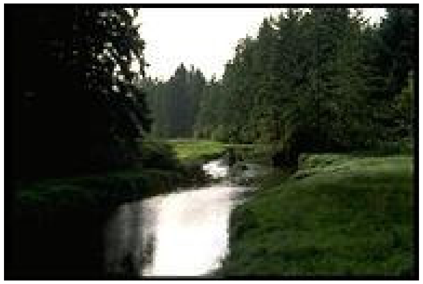


Low level segmenters split up objects and cannot group disparate regions belonging to semantic entities

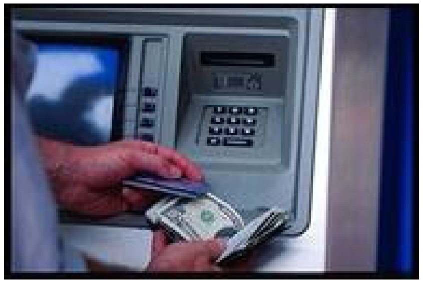
Using word prediction gives a way to incorporate high-level semantic information in the merging process

Propose a merge between regions that have similar posterior distributions over words

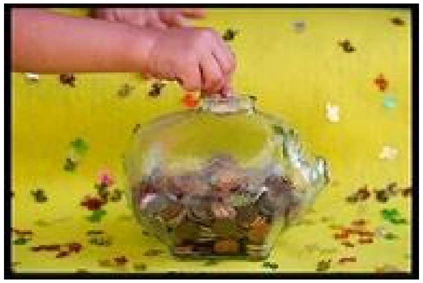
Sense Disambiguation



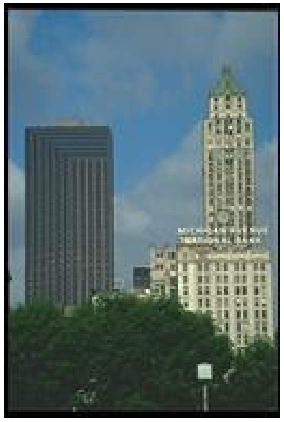
26078 water grass trees **bank**s



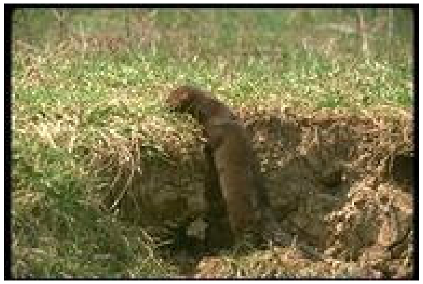
125090 **bank** machine money currency bills



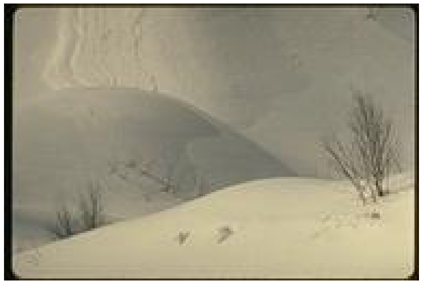
125084 piggy **bank** coins currency money



212001 **bank** buildings trees city

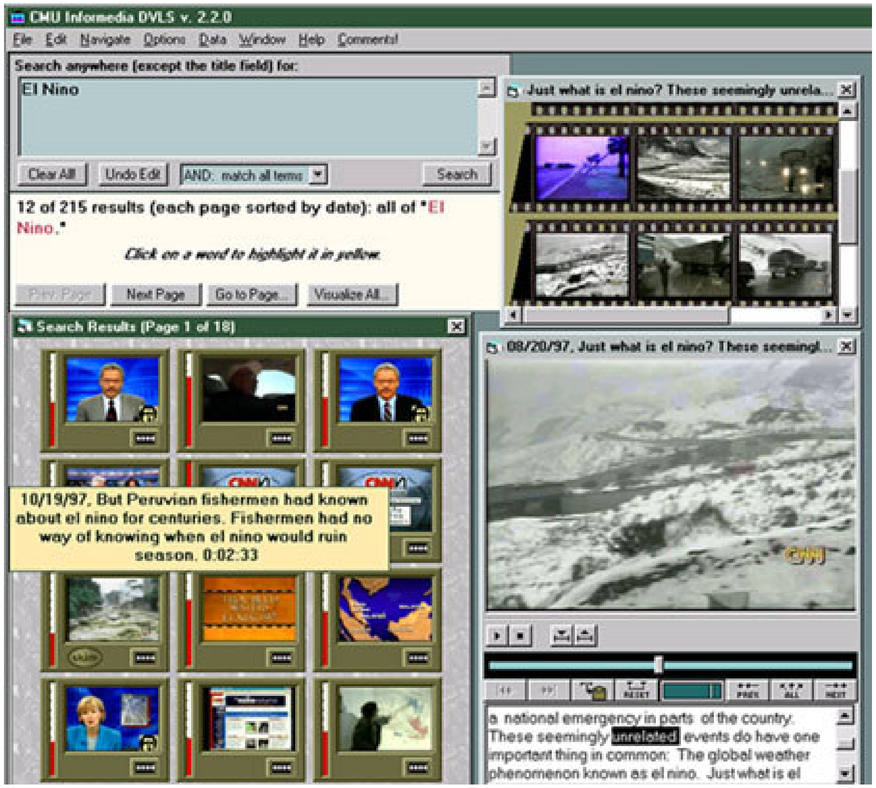


173044 mink rodent **bank** grass



151096 snow **bank**s hills winter

Informedia Digital Video Library Project



CMU Informedia DVLS v. 2.2.0
File Edit Navigate Options Data Window Help Comments!
Clear All
Undo Edit
AND: match all terms
Search
Prev Page
Next Page
Go to Page...
Visualize All...
Search Results (Page 1 of 18)
CNN

Search anywhere (except the title field) for:

El Nino

12 of 215 results (each page sorted by date): all of "El Nino."

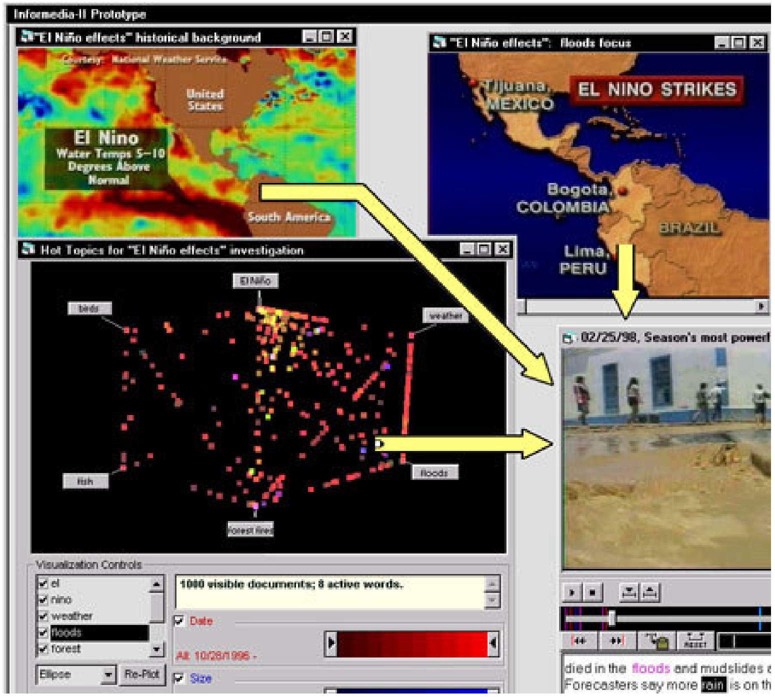
Click on a word to highlight it in yellow.

Just what is el nino? These seemingly unrela...

10/19/97, But Peruvian fishermen had known about el nino for centuries. Fishermen had no way of knowing when el nino would ruin season. 0:02:33

08/20/97, Just what is el nino? These seemingl...

a national emergency in parts of the country. These seemingly unrelated events do have one important thing in common: The global weather phenomenon known as el nino. Just what is el



☑ el
☑ nino
☑ weather
☑ floods
☑ forest

1000 visible documents; 8 active words.

☑ Date

All: 10/28/1996 -

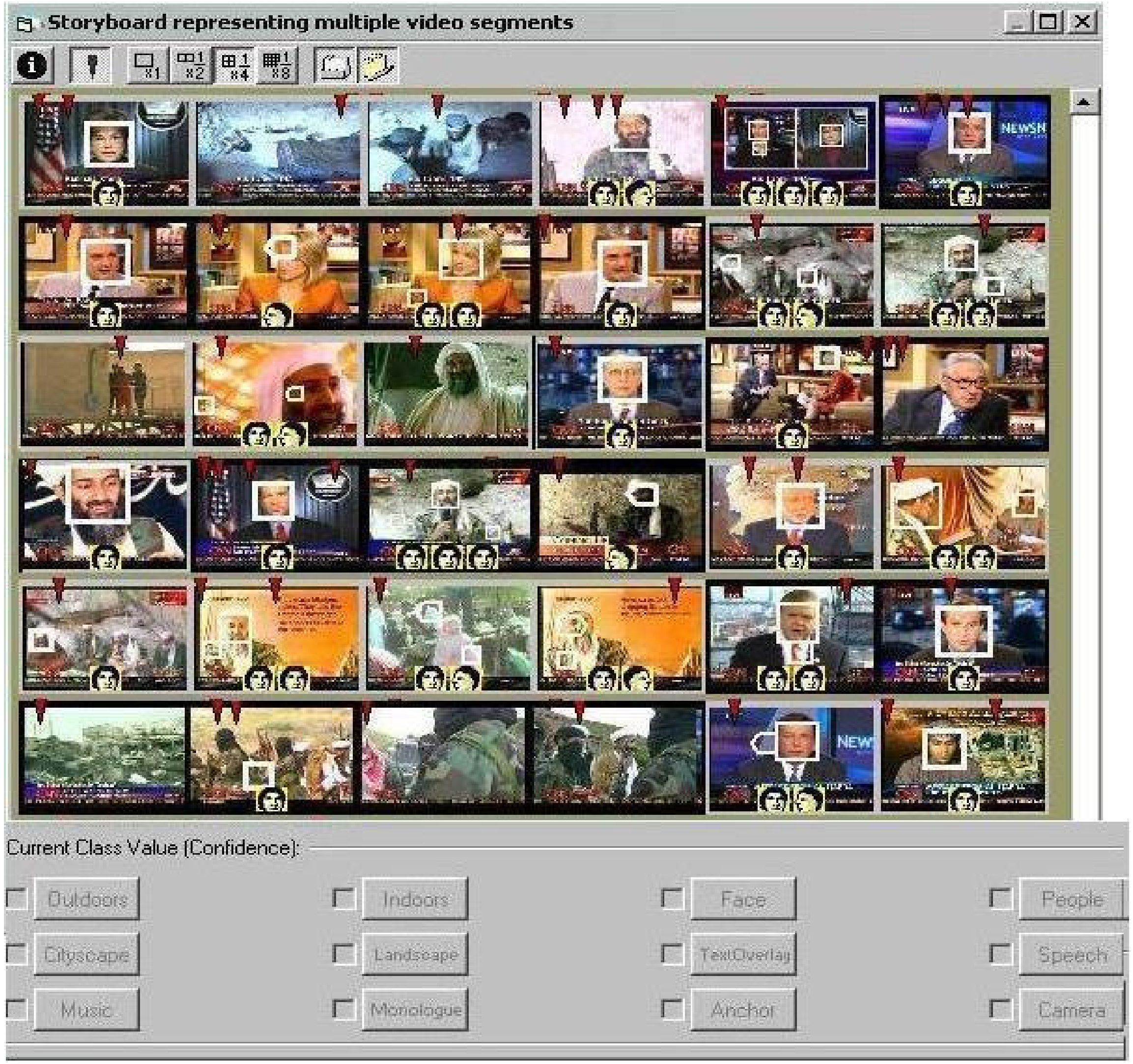
☑ Size

02/25/98, Season's most powerf

died in the floods and mudslides a
Forecasters say more rain is on th

IDVL interface returned for "El Nino" query along with different multimedia abstractions from certain documents.

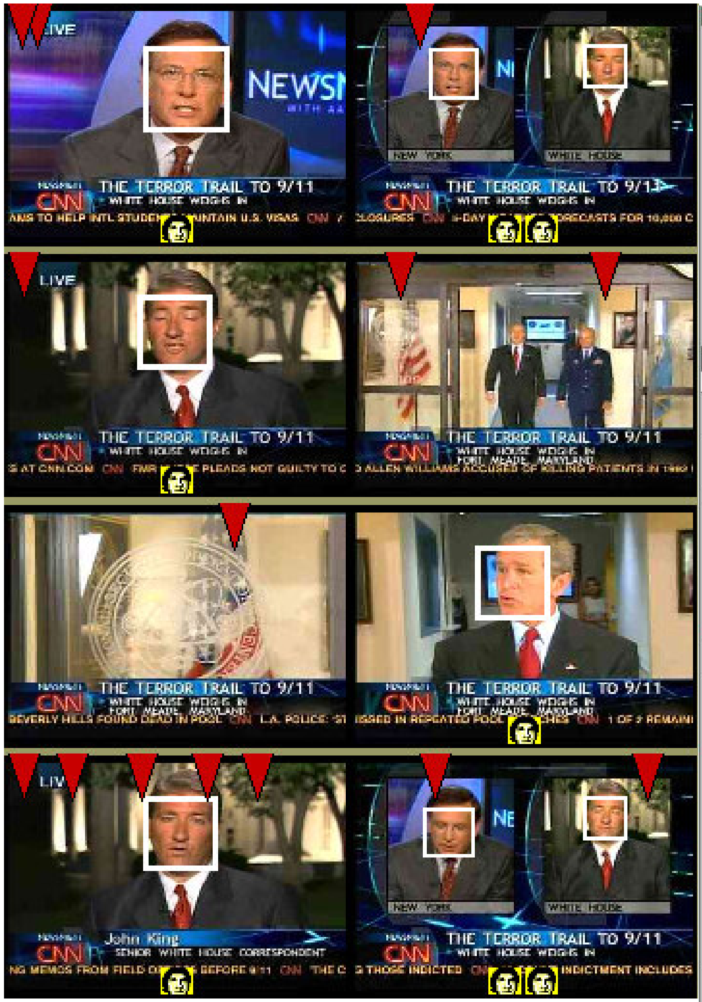
Informedia Digital Video Library Project

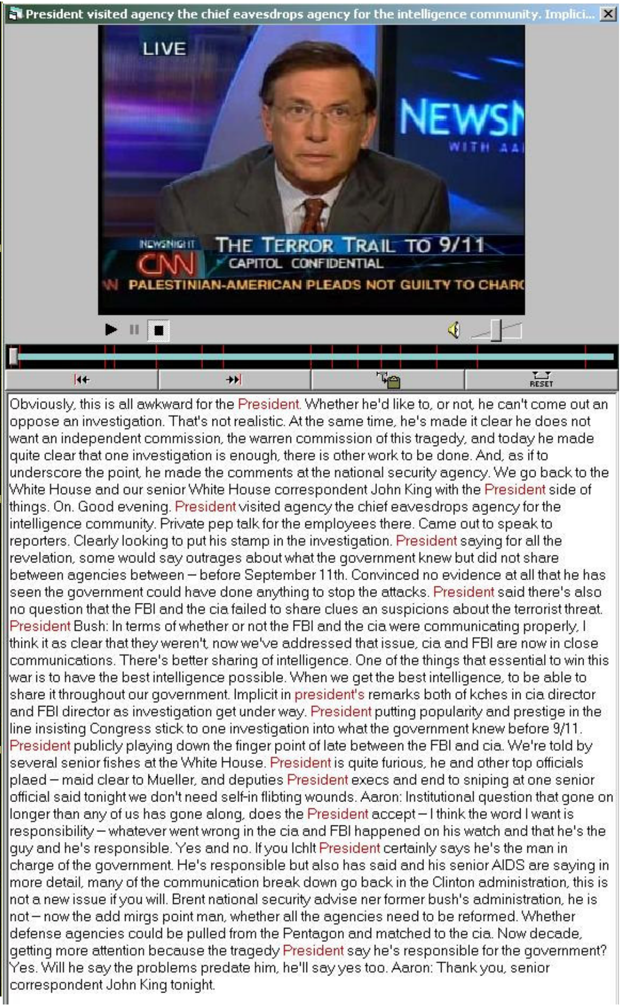


IDVL interface returned for “bin ladin" query

The results can be tuned using many classifiers

Associating video frames with text



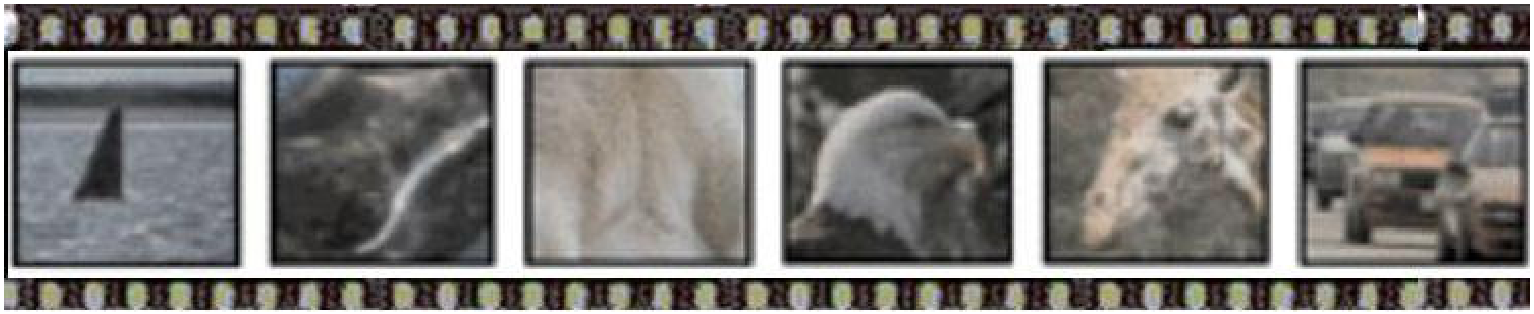

President visited agency the chief eavesdrops agency for the intelligence community. Implici...

Obviously, this is all awkward for the President. Whether he'd like to, or not, he can't come out an oppose an investigation. That's not realistic. At the same time, he's made it clear he does not want an independent commission, the warren commission of this tragedy, and today he made quite clear that one investigation is enough, there is other work to be done. And, as if to underscore the point, he made the comments at the national security agency. We go back to the White House and our senior White House correspondent John King with the President side of things. On. Good evening. President visited agency the chief eavesdrops agency for the intelligence community. Private pep talk for the employees there. Came out to speak to reporters. Clearly looking to put his stamp in the investigation. President saying for all the revelation, some would say outrages about what the government knew but did not share between agencies between – before September 11th. Convinced no evidence at all that he has seen the government could have done anything to stop the attacks. President said there's also no question that the FBI and the cia failed to share clues an suspicions about the terrorist threat. President Bush: In terms of whether or not the FBI and the cia were communicating properly, I think it as clear that they weren't, now we've addressed that issue, cia and FBI are now in close communications. There's better sharing of intelligence. One of the things that essential to win this war is to have the best intelligence possible. When we get the best intelligence, to be able to share it throughout our government. Implicit in president's remarks both of kches in cia director and FBI director as investigation get under way. President putting popularity and prestige in the line insisting Congress stick to one investigation into what the government knew before 9/11. President publicly playing down the finger point of late between the FBI and cia. We're told by several senior fishes at the White House. President is quite furious, he and other top officials plaed – maid clear to Mueller, and deputies President execs and end to sniping at one senior official said tonight we don't need self-in flibting wounds. Aaron: Institutional question that gone on longer than any of us has gone along, does the President accept – I think the word I want is responsibility – whatever went wrong in the cia and FBI happened on his watch and that he's the guy and he's responsible. Yes and no. If you IchIt President certainly says he's the man in charge of the government. He's responsible but also has said and his senior AIDS are saying in more detail, many of the communication break down go back in the Clinton administration, this is not a new issue if you will. Brent national security advise ner former bush's administration, he is not – now the add mirgs point man, whether all the agencies need to be reformed. Whether defense agencies could be pulled from the Pentagon and matched to the cia. Now decade, getting more attention because the tragedy President say he's responsible for the government? Yes. Will he say the problems predate him, he'll say yes too. Aaron: Thank you, senior correspondent John King tonight.

Query on "president"

Association problem

Associating video frames with text



…despite heroic efforts many of the worlds wild creatures are doomed the loss of species is now the same as when the great dinosaurs become extinct will these creatures become the dinosaurs of our time today…

Associating video frames with text

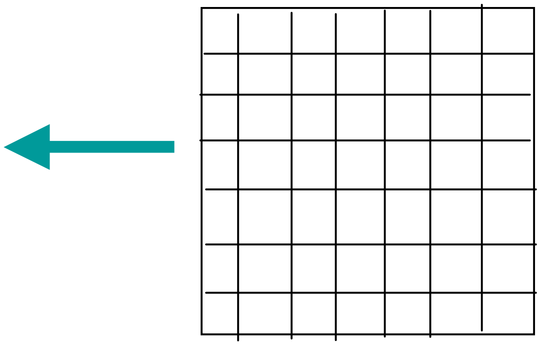
Position,

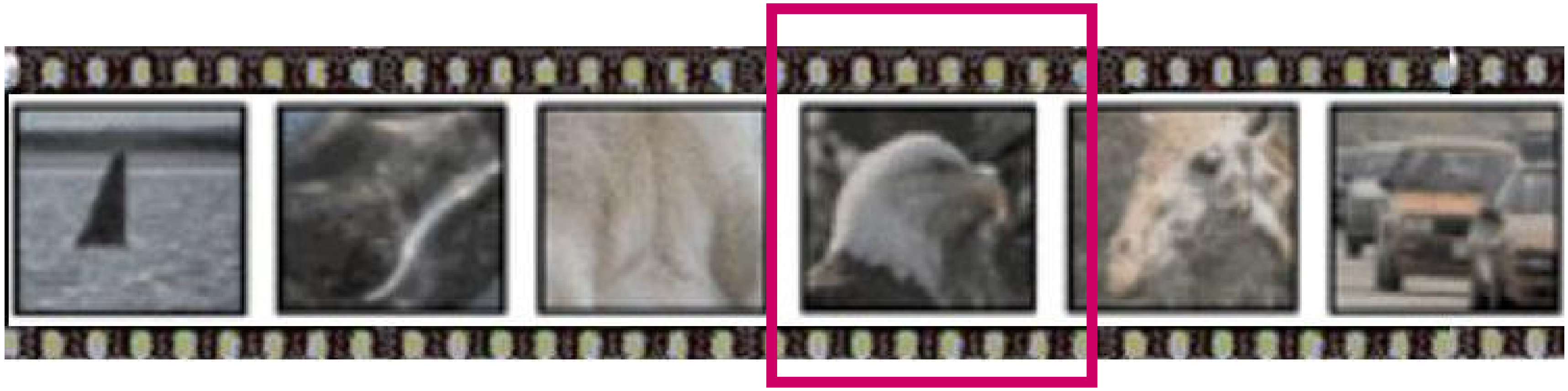
Color

(RGB and Lab, mean and std)

Texture

(Oriented energy filters, DoG)



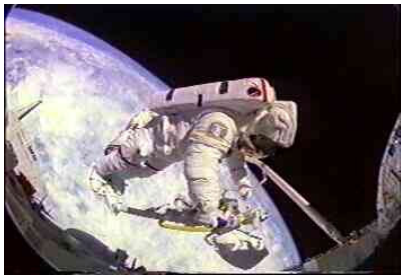


…efforts many | of the worlds | wild creatures | are doomed | the loss of | species …

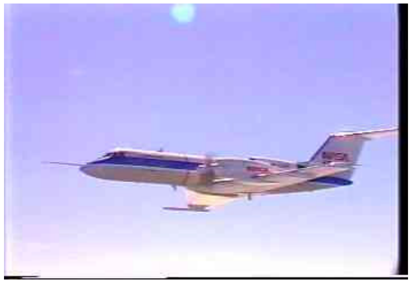
Duygulu & Wactlar, ACMSIGIR-MIR 2003

TREC-2001 data

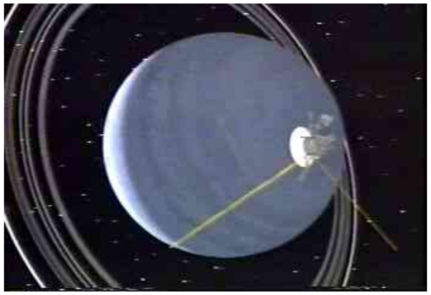
Auto-annotation results



space (6), astronaut(7)

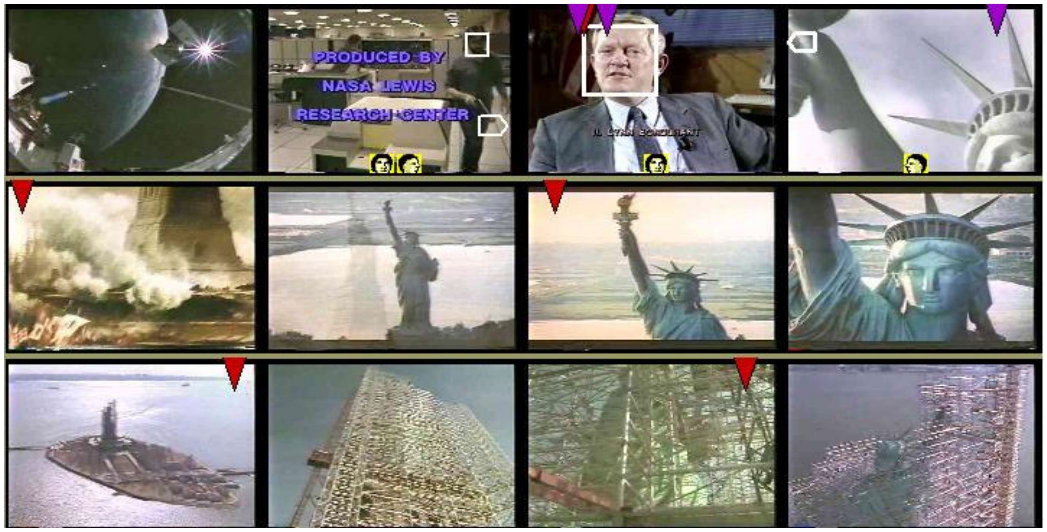


plane(2)



space (1), telescope(10)

Query for "Statue of Liberty" on current Informedia system



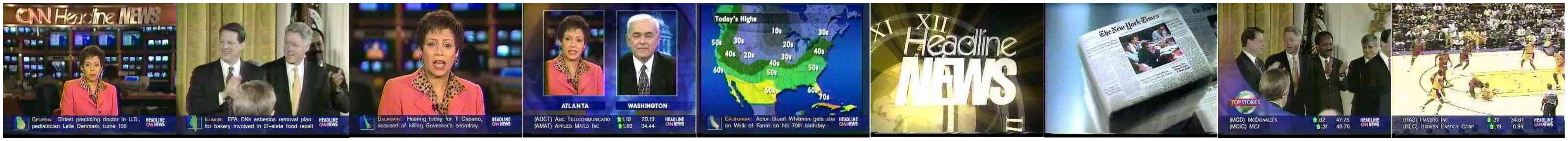
Corrected with auto-annotation

statue(1) liberty(3)

statue(1) liberty(3)

News videos - structured

Taking the surrounding words are problematic
Segments are defined in some close caption text
If it is not available use structure to obtain segments



anchor

News story

anchor – reporter dialogs

weather

logos

commercials

overview

sports

Get only news stories

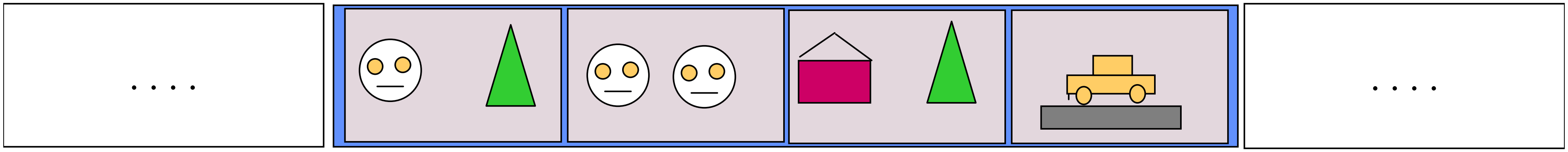
Remove commercials

Remove graphics

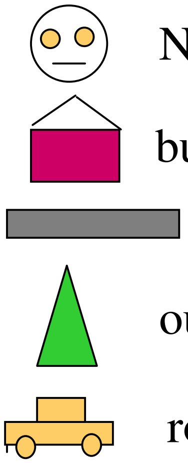
Remove anchor images but use text

Associating text with frames

w1 w2 w10 w1 w5 w6 w2 w1 w4 w10 w5 w3 w11



Color tokens : 1-230



Num faces (1 /2 / >=3)

building

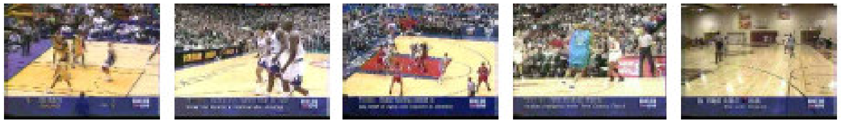
road

outdoor

road

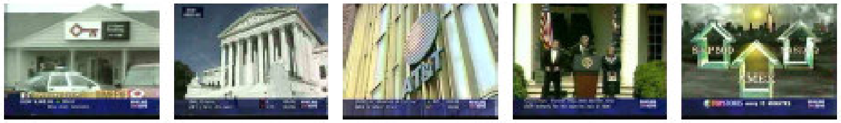
Token words

stock, wall, market, street, investor, report, news, business, jones, industrials, interest, deal, thanks, cnnfn, company, susan, yesterday, morris, number, merger



series, bull, jazz, playoff, game, conference, final, karl, lead, indiana, utah, difference, combination, board, night, ball, point, pair, front, team

pilot, veteran, family, rescue, foot, effort, crew, search, security, troop, fact, affair, member, survivor, tobacco, field, department, health, communication, leader

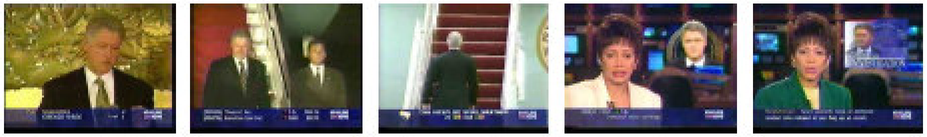


company, market, line, worker, street, union, profit, wall, cost, news, strike, yesterday, rate, quarter, stock, check, report, level, fact, board

Semantic retrieval

!! only single occurrence per segment

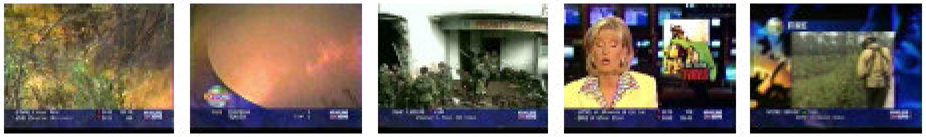
Search on clinton



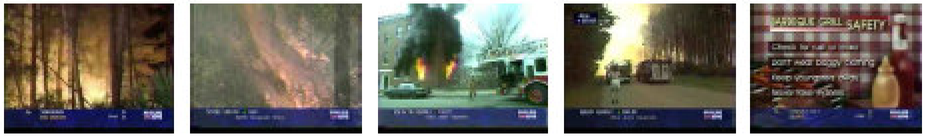
20 / 130 (15%)

27 / 133 (20%)

Search on fire



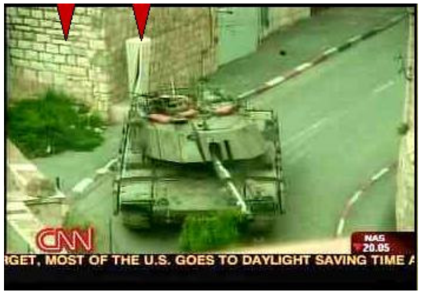
11 / 44 (25%)



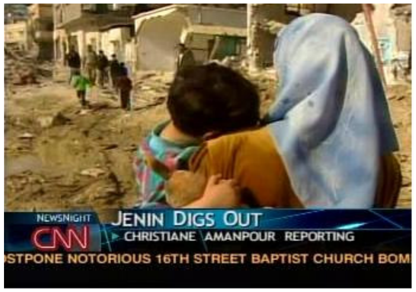
15 / 38 (40%)

Future work

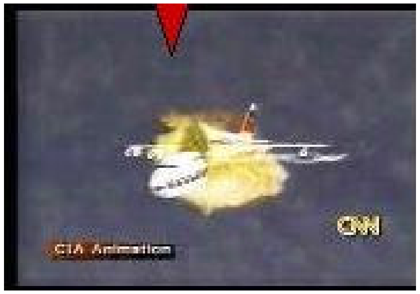
Solving correspondences in broadcast news for better retrieval



..tanks on the street …

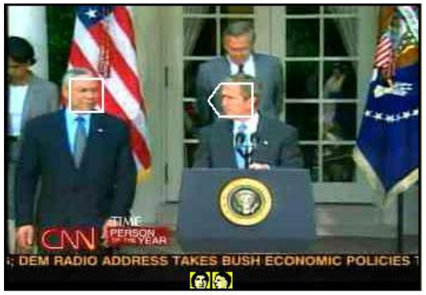

NEWSNIGHT
CNN
JENIN DIGS OUT
CHRISTIANE AMANPOUR REPORTING
STPONE NOTORIOUS 16TH STREET BAPTIST CHURCH BOM

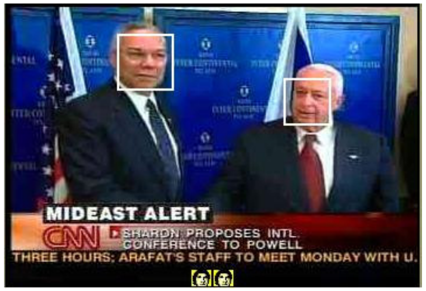
..start attacking on houses by helicopters and tanks…

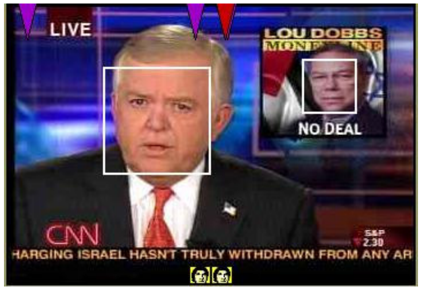


..fuel tank…

Face Recognition by resolving correspondences between named entities and faces




MIDEAST ALERT
CNN
SHARON PROPOSES INTL. CONFERENCE TO POWELL
THREE HOURS; ARAFAT'S STAFF TO MEET MONDAY WITH U.



Summary

When text and visual features are combined it is possible to do many interesting tasks
Including better retrieval, browsing, auto-annotation and auto-illustration

Object recognition on the very large scale can be viewed as translation of regions to words

There are many other available multi-modal data sets

Video is a huge source of information where audio, text, and visual features appear together

Current systems that are based on text should be improved with the help of multi-modal data