Plan for Today's Lecture(s)

- Introduction to Multimedia Resource Description
- "Professional" Metadata
  for Multimedia and Non-Text Resources
- "Computational" Metadata
  for Multimedia and Non-Text Resources
- "Contextual" Metadata
- Resource Description for Music

UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

INFO 202
"Information Organization & Retrieval"
Fall 2013

Robert J. Glushko
glushko@berkeley.edu
@rjglushko

1 October 2013
Lecture 10.1 – Introduction to Describing Multimedia Resources

What is Multimedia?

- Media:
  - Text
  - Audio (speech, music, sound)
  - Graphics and Images
  - Moving Images
- Multimedia -- Composed of more than one form of media
- Time-based media -- Audio and video

Concepts and Relationships for Multimedia Resources

- playlist, album, queue
- composite, collage, mix, remix
- clip, sample, bite
- remake, cover version
- layout, presentation, performance
- broad-/narrow-/simul-/tele-/web-/pod- cast
- installation, environment

What's Different About Describing Multimedia?

- Sensory Gap
- Semantic Gap
- Proliferation Problem

The Sensory Gap (1)

The Sensory Gap (2)

- There is a gap between an object and a computer's ability to sense and describe the object
- An infinite number of different "signals“ can be produced by the same object
- …And different objects can produce similar signals
- Human perceptual machinery excels at recognizing when different "signal patterns" are the same object or when similar patterns are different ones
- But the problem is difficult for computers

Don't Blame the Computer

- It is asking a lot of the computer to reconstruct and recognize a 3D or 4D world from 2D (photos) or 3D (video) projections of it that lose massive amounts of information in the process
- It is fairer to see the sensory gap as evidence of the incredible power of the human visual and cognitive systems to apply neural computation and knowledge to enable visual perception

The Semantic Gap

- Instruments, devices, and sensors encode data in formats that are optimized for efficient capture, storage, decoding, or other processing
- This data is often limited to characteristics that involve coarse categorization and syntactic processing
- The representation of the object can't be (easily) processed to understand what the object "means"
- So there is a semantic gap between the descriptions that people assign and those that can be assigned by automated mechanisms

What's this Picture About?

Name That Tune

Text has a Smaller Semantic Gap

```
<title>Concepts and Relationships for Multimedia
"Documents"</title>
 <list>
  <item><p>playlist, album, queue</p></item>
  <item><p>composite, collage, mix, remix</p></item>
  <item><p>clip, sample, bite</p></item>
  <item><p>remake, cover version</p></item>
  <item><p>layout, presentation, performance</p></item>
  <item><p>broad-/narrow-/simul-/tele-/web-/
     pod- cast</p></item>
  <item><p>installation, environment</p></item>
</list>
```

The Proliferation Problem (1)

- Digital cameras, video recorders, and smart phones have increased resolution and greater storage capacity every year
- Non-professional consumers create very large collections of multimedia objects (20 B / year)
- The average object has less value and doesn't justify much effort to organize and describe it
- These “objects” include audio or video “snippets” that can stand alone

The Proliferation Problem (2)

- Without ways to manage collections of digital multimedia resources we lose the advantages of digital media
- Many current approaches are insufficient, perhaps misguided, and certainly ignore a lot of what we already know
- Great opportunity for innovation and invention, especially interdisciplinary approaches that combine technology and social considerations

Are these “Gaps” New Problems?

- Museums face some of the same or similar problems in describing art works and artifacts:
- There may be many artifacts that represent the same "work" - this is like the "sensory" gap
- The materials or medium in which the artifact is embodied don't convey semantics "on their surface" - this is the semantic gap
- There may be so many artifacts of a particular type that some get only limited descriptions - this is like the proliferation problem

Just a Matter of Degree?

Horse and Rider

Enlarge

Unknown
Greek, Boeotia, about 550 B.C.
Terracotta
82.AD.84

"Black-Throated Wind"
Grateful Dead
1973-02-09
Maples Pavillion, Stanford U.

Some Problems May Be New

- The temporal structure of multimedia, especially video, mandates new descriptive vocabulary and new ways to identify meaningful components
- Video and music meet emotional/psychological needs that are more complex than those for "documents" - so the descriptions of the latter have to be able to address these needs
- People don't usually access or retrieve music or video "to satisfy information requirements“

UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

INFO 202
"Information Organization & Retrieval"
Fall 2013

Robert J. Glushko
glushko@berkeley.edu
@rjglushko

1 October 2013
Lecture 10.2 – "Professional" Metadata
for Multimedia and Non-Text Resources

"Professional" Metadata for Multimedia and Non-textual Works

- Museum works and other non-text artifacts have long been described by professionals using structured metadata according to classification rules
- The descriptions are often "layered" beginning with accessible properties with more analytic descriptions added for those objects that warrant them

"Panofsky's Three Levels of Meaning

- Description ("Preiconographic“ or “Primary”)
- Identification ("Iconographic“ or “Secondary”)
- Interpretation ("Iconologic“)

Description / "Preicongraphic" / Primary Level

- The "primary" or "natural" subject matter of a work
- The generic elements or things depicted in, on, or by an object / image / art work
- These would be recognizable by anyone, regardless of their expertise or training

Identification / "Icongraphic" / Secondary Level

- The name of the subject or thing depicted in the work
- Can require research or fact assembly from sources of social and cultural knowledge

Interpretation / "Iconologic" Level

- The meaning or theme represented by the subject matter or iconography of the work
- This meaning is often symbolic, and deeply grounded in the culture in which the work was created
- Done by domain and methodology experts

Panovsky's 3-Level Scheme

- PRIMARY: Marble statue of nude woman standing on a seashell
- SECONDARY: Statue made in 2005 by Lucio Carusi of Carrara, Italy, titled "Venus", made of local marble
- INTERPRETIVE: 3D transformation of 1486 painting by Italian painter Sondro Botticelli, titled "Birth of Venus," now in the Uffizi Gallery in Florence. Carusi's Venus is substantially slimmer in proportions because of changing notions of female beauty

(TDO Figure 4.5)

Botticelli 1486

Carusi 2005

Getty Categories for the Description of Works of Art (CDWA)

- The CDWA is a massive metadata schema with 532 elements and sub-elements
- Each element typically has a controlled vocabulary or recommends one (e.g, Art and Architecture Thesaurus)
- There is a core set of 36 elements
- CDWA-Lite is the XML specification for another simple subset of CDWA

CDWA's First Question

- You should have been able to guess the first question raised in the CDWA documentation
- Section 1.1 of the CDWA (the top level element) is "Object/Work"

> *Works of art or architecture may be considered a single item, or they may be made up of many physical parts or arranged in separate physical groupings. It is necessary to define the particular work of art, architecture, or group of objects in question, whether it be a single painted canvas or an altarpiece made up of many panels...*

CDWA in XML - Examples

```
<cdwalite:objectWorkTypeWrap>
    <cdwalite:objectWorkType>rhyton</cdwalite:objectWorkType>
</cdwalite:objectWorkTypeWrap>

<cdwalite:objectWorkTypeWrap>
    <cdwalite:objectWorkType termsource="AAT">painting </cdwalite:objectWorkType>
    <cdwalite:objectWorkType termsource="AAT">altarpiece</cdwalite:objectWorkType>
 </cdwalite:objectWorkTypeWrap>

<cdwalite:objectWorkTypeWrap>
    <cdwalite:objectWorkType termsource="AAT" termsourceID="aat300127141">cartes-de-
    visite </cdwalite:objectWorkType>
    <cdwalite:objectWorkType termsource="AAT" termsourceID="aat300265164">boudoir
    photographs </cdwalite:objectWorkType>
```

Art and Architecture Thesaurus

ID: 300011576 **Record Type:** concept

Carrara marble (white marble, <marble by color or pattern>, ... Materials)

Note: Refers to a type of marble quarried in the area around Carrara, in Tuscany, Italy. It is characterized by a fine, compact grain and varies in color from pure white to creamy white, sometimes with a bluish tinge; it is a saccharoidal rock that can appear translucent in the finer grades. It has been a favorite stone of sculptors from antiquity to the present, including Michelangelo Buonarotti. Luna marble was the name used in ancient Rome.

Terms:

Carrara marble (preferred, C,U,D,American English-P)
Luna marble (C,U,UF,American English)
marble, Carrara (C,U,UF,American English)
marble, Luna (C,U,UF,American English)
marmor Lunense (C,U,UF,American English)
marmor Luniense (C,U,UF,American English)

Facet/Hierarchy Code: M.MT

Hierarchical Position:

Materials Facet
.... Materials
..... materials
..... <materials by composition>
..... inorganic material
..... rock
..... metamorphic rock
..... marble
..... <marble by color or pattern>
..... white marble
..... Carrara marble

Union List of Artist Names

Buonarroti, Michelangelo (preferred, index, V)
Michelangelo Buonarroti (display, V)
Michelangelo Buonarotti (V)
Michelangelo (V)
Michelagnolo di Lodovico Buonarroti Simoni (V)
Michelagniolo di Lodovico de Lionardo di Buonarroto Simoni (V)
Michelagniolo di Lodovico di Lionardo di Buonarroto Simoni (V)
Michelangelo di Lodovico di Lionardo di Buonarroto Simoni (V)
Michelangelo di Ludovico Buonarroti (V)
Buonarroti, Michel Angelo (V)
Buonarroti, Michelagniolo (V)
Bonarroti, Michelangelo (V)
Bonorotti, Michelangelo (V)
Buonarota, Michelangelo (V)

Michael Angelo Buonaroti (V)
Michael Angelo Buonarotti (V)
Michelagniolo Buonarroti (V)
Michelagnolo Buonarotti (V)
Michelagnolo Buonarruoti (V)
Michelange Bonaroti (V)
Michelang.o Bonarota (V)
Michel Angel de Bonarrotta (V)
Michelangelo Bonarota (V)
Michelangelo Bonaroti (V)
Michelangelo Buonarota (V)
Michelangelo Buonaroti (V)
Michelangelo Buonarrota (V)
Michelangelo Buonnaroti (V)
Michelangiolo Buonaroti (V)
Micheleangelo Buonarota (V)
Michel Angelo (V)
Michel'Angelo (V)
Michael Angelo (V)

Thesaurus of Geographic Names

ID: 7006077 **Record Type: administrative**

Carrara (inhabited place)

Coordinates:
Lat: 44 05 00 N *degrees minutes* Lat: 44.0833 *decimal degrees*
Long: 010 06 00 E *degrees minutes* Long: 10.1000 *decimal degrees*

Note: Located on the Carrione river; gained importance as marble-quarrying center, noted for fine white marble suitable for sculpture; was important medieval center; was center of Massa-Carrara principality & duchy (after 1633).

Names:
Carrara (**preferred**, C,V,N)
Apuania (H,V,N) ..... ancient name

Hierarchical Position:
World (facet)
.... Europe (continent)
..... Italy (nation)
..... Tuscany (region) (C) part of new kingdom of Italy from 1861
..... Massa-Carrara (province)
..... Carrara (inhabited place)

Place Types:
inhabited place (**preferred**, C) ..... was an ancient center
city (C)
quarrying center (C)
episcopal see (C)

Metadata for Video

- Across types of video, the basics still apply
  - Name, network/channel or person uploading, episode, time, date, etc.
  - MPEG-7 standard defines the key elements
- Digital broadcast TV is governed by an alphabet soup of standards (PSIP, PMCP, ATSC…)
- Standards define what information about a program or a station shows up on your DVR's channel guide

Professional Metadata-Making for Video

- When an ad pops up during a YouTube video, that's not random
- Advertisers want to be sure they (and we) can find people/characters drinking Pepsi, wearing Levi's, etc. buy clothes worn in TV shows
- NFL Films: Video loggers tag game footage with details from drop-down menus (team, date, yardage) and with keyword phrases like 'funky fans,' 'torn uniform,' ...'"

Metadata in/on Video

Comparing “Professional” Metadata with “Amateur" Metadata

- Flickr, YouTube, and other collections of multimedia works are characterized by non-standard classification schemes
- There is no "professional" metadata to start with and build upon
- Result 1: Unstructured tagging with little use of controlled vocabularies within each collection
- Result 2: No standardization or interoperability across collections

UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

INFO 202
"Information Organization & Retrieval"
Fall 2013

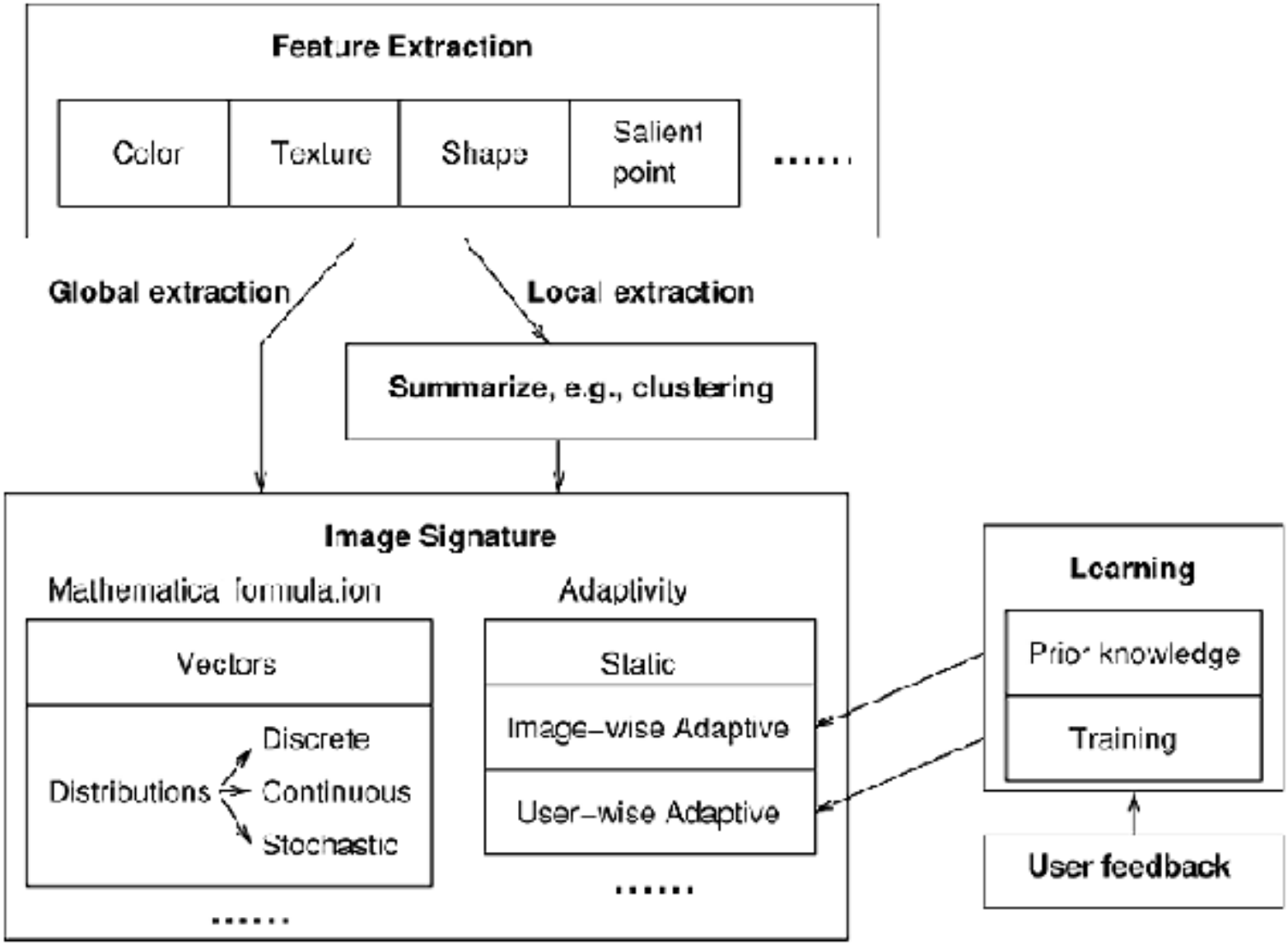
Robert J. Glushko
glushko@berkeley.edu
@rjglushko

1 October 2013
Lecture 10.3 – "Computational" Metadata
for Multimedia and Non-Text Resources

Bridging the Semantic Gap in Describing Images

- Bridging the gap means extracting features from multimedia content and finding ways to infer semantic-level descriptions from them
- Some key questions: What features best support semantic inference? What inference techniques work the best? Can we exploit multi-modal or cross-modal information?

Creating an Image "Signature"



Computational Description of Images

- Shatford (as described by Christel) expanded Panovsky's 3 levels to 10 to make finer distinctions about the amount of semantic information required (or extracted by computational means)

1. Type/technique
2. Global distribution
3. Local structure
4. Global composition
5. Generic objects
6. Generic scene
7. Specific objects
8. Specific scene
9. Abstract objects
10. Abstract scene

Query with Level 2 Description (low-level perceptual features)

Christel Figure 2.2: Example of syntactically correct matches of blue-green images to blue-green key, but with results spanning a variety of objects unlikely to be the true information targets for a user.

Video Analytics (1)

- Millions of video surveillance cameras are deployed worldwide
- Most of their video feeds are not monitored by people in real time, so the cameras can't prevent crime, accidents, terrorist acts, etc. - the recordings have to be analyzed after the fact

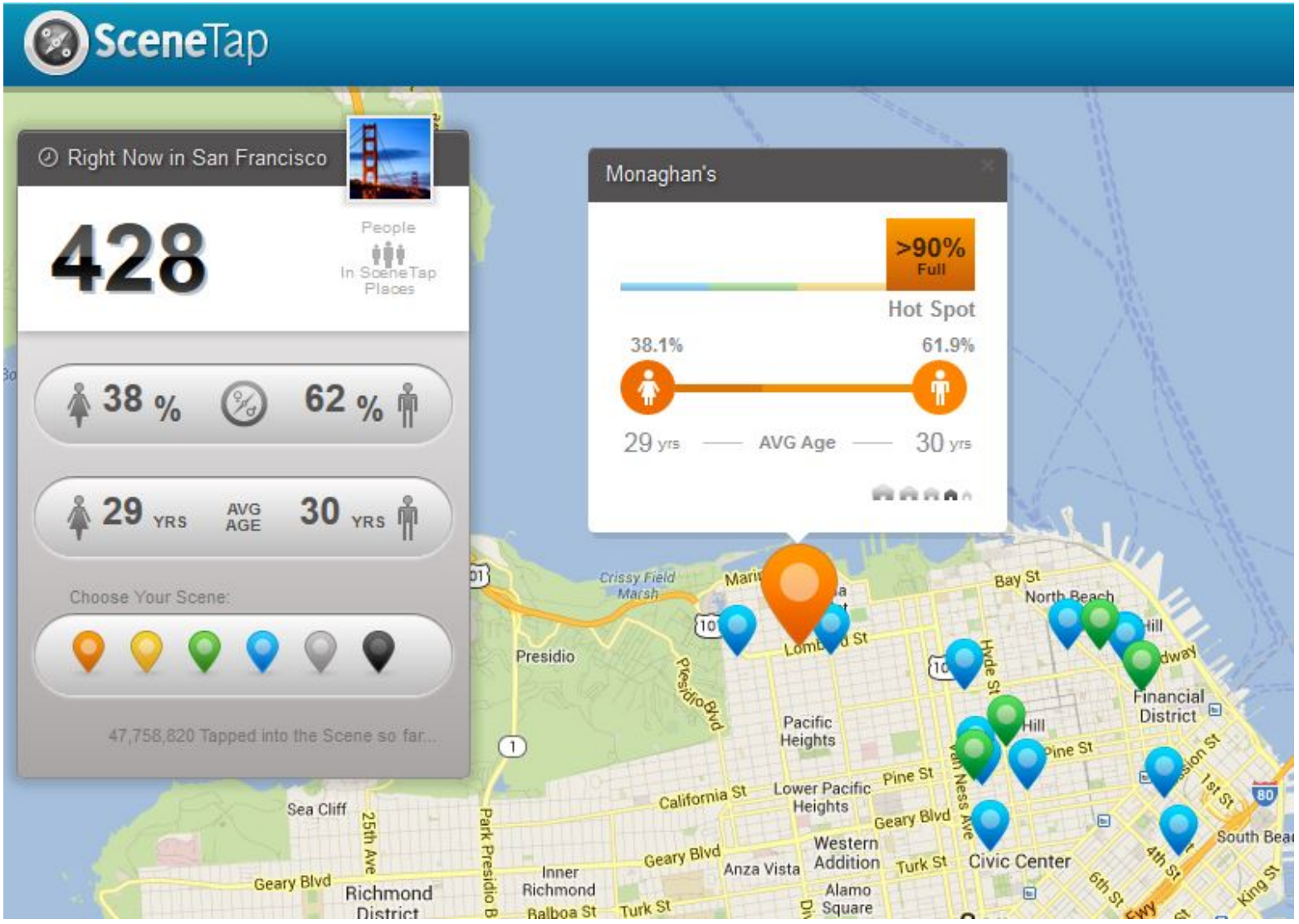
Video Analytics (2)

- Three main R & D currents:
  - Detect, track, and classify events, especially "anomalous" ones
  - Exploit "camera networks" to combine multiple observations and perspectives to improve analysis
  - Improve real-time analysis, especially by exploiting application specific knowledge
  - “Lights, Camera, Data” story from WSJ

The Dark Side of Image Analysis

- What is the age and gender of this person?
- Scene Tap uses cameras with facial detection software to scout bar scenes (http://scenetap.com/")
- Facebook whacked for using face recognition sw

Scene Tap



[See “San Francisco Hates Your Startup”]

UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

INFO 202
"Information Organization & Retrieval"
Fall 2013

Robert J. Glushko
glushko@berkeley.edu
@rjglushko

1 October 2013
Lecture 10.4 – "Contextual" Metadata

Contextual Metadata

- Metadata about the context in which some content was "captured"
- Location, time, other people or things present are basic elements, but there are many more
- This kind of information has often been collected, but not usually analyzed and applied to description until afterwards

Non-Automated Contextual Metadata in Museum Catalog

Bone: AN1955.565

Summary page

- *Object Type:* Plastered human skull
- *Period:* Pre-Pottery Neolithic
- *Material:* Bone
- *Dimensions:* H: 15.2cm. W: 16.7cm. L: 22cm.
- *Region:* Palestine
- *Find's spot:* Jericho: 1953 season of excavation: skull D III (Reg. 534); from square D. I, stage XVI-XVII: phase xlii-xliii)
- *Museum Catalogue No.:* Moorey (2004) I

More images:

from the front

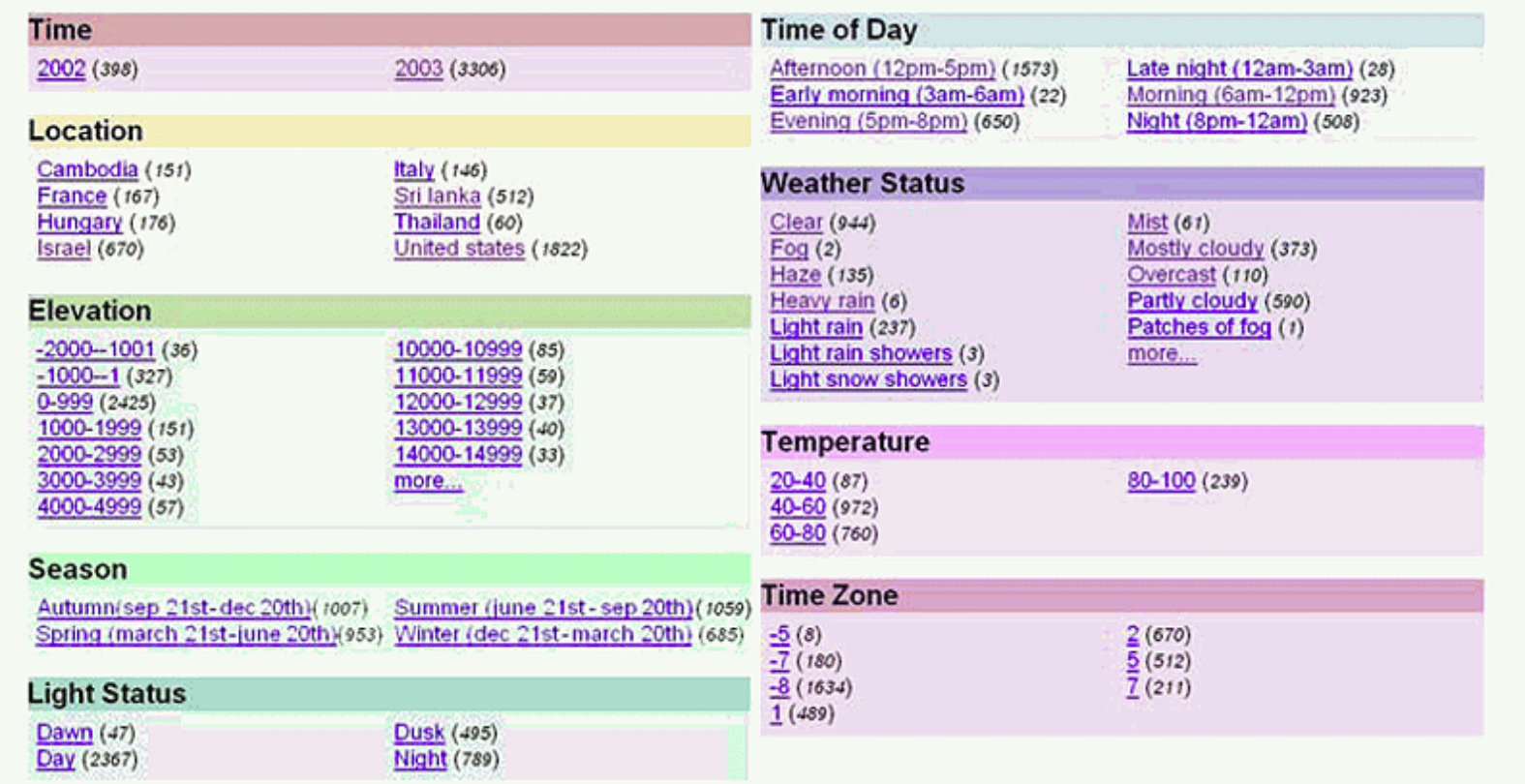
from the side

Automated Contextual Metadata in EXIF (Digital Camera)

Manufacturer	CASIO
Model	QV-4000
Orientation	top - left
Software	Ver1.01
Date and Time	2003:08:11 16:45:32
YCbCr Positioning	centered
Compression	JPEG compression
x-Resolution	72.00
y-Resolution	72.00
Resolution Unit	Inch
Exposure Time	1/659 sec.
FNumber	f/4.0
ExposureProgram	Normal program
Exif Version	Exif Version 2.1
Date and Time (original)	2003:08:11 16:45:32
Date and Time (digitized)	2003:08:11 16:45:32

ComponentsConfiguration	Y Cb Cr -
Compressed Bits per Pixel	4.01
Exposure Bias	0.0
MaxApertureValue	2.00
Metering Mode	Pattern
Flash	Flash did not fire.
Focal Length	20.1 mm
MakerNote	432 bytes unknown data
FlashPixVersion	FlashPix Version 1.0
Color Space	sRGB
PixelXDimension	2240
PixelYDimension	1680
File Source	DSC
InteroperabilityIndex	R98
InteroperabilityVersion	(null)

Augmenting EXIF (time) with Geo-Referenced Additional Resources



Time

2002 (398) 2003 (3306)

Location

Cambodia (151) Italy (146)
France (167) Sri lanka (512)
Hungary (176) Thailand (60)
Israel (670) United states (1822)

Elevation

-2000--1001 (36) 10000-10999 (85)
-1000--1 (327) 11000-11999 (59)
0-999 (2425) 12000-12999 (37)
1000-1999 (151) 13000-13999 (40)
2000-2999 (53) 14000-14999 (33)
3000-3999 (43) more...
4000-4999 (57)

Season

Autumn(sep 21st-dec 20th) (1007) Summer (june 21st-sep 20th) (1059)
Spring (march 21st-june 20th) (953) Winter (dec 21st-march 20th) (685)

Light Status

Dawn (47) Dusk (495)
Day (2367) Night (789)

Time of Day

Afternoon (12pm-5pm) (1573) Late night (12am-3am) (28)
Early morning (3am-6am) (22) Morning (6am-12pm) (923)
Evening (5pm-8pm) (650) Night (8pm-12am) (508)

Weather Status

Clear (944) Mist (61)
Fog (2) Mostly cloudy (373)
Haze (135) Overcast (110)
Heavy rain (6) Partly cloudy (590)
Light rain (237) Patches of fog (1)
Light rain showers (3) more...
Light snow showers (3)

Temperature

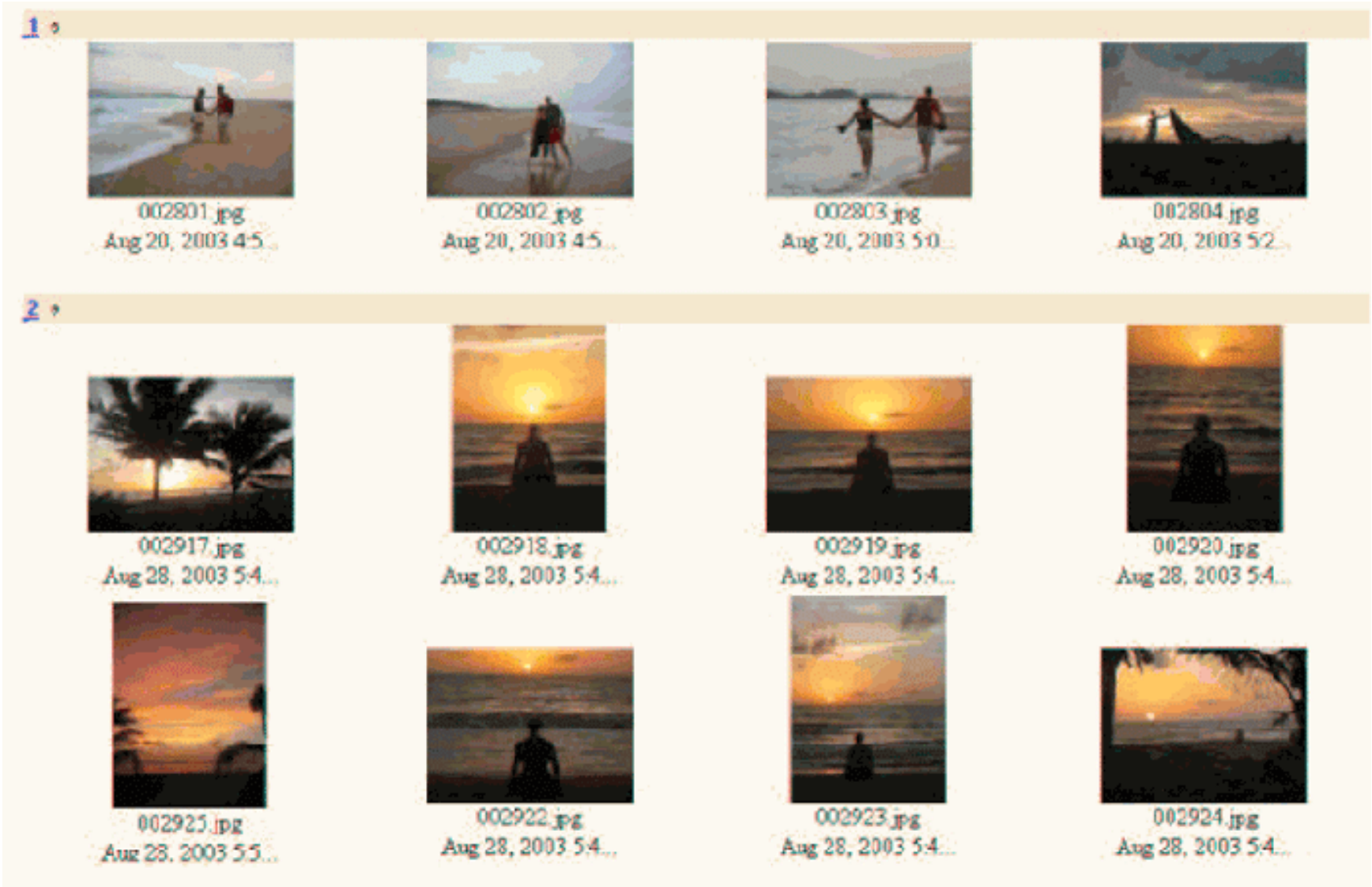
20-40 (87) 80-100 (239)
40-60 (972)
60-80 (760)

Time Zone

-5 (8) 2 (670)
-7 (180) 5 (512)
-8 (1634) 7 (211)
1 (489)

Naaman, Mor, et al. "Context data in geo-referenced digital photo collections." 2004.

Using Contextual Metadata in Image Retrieval (Naaman et al)



UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

INFO 202
"Information Organization & Retrieval"
Fall 2013

Ryan Baker

1 October 2013
Lecture 10.5 – Metadata for Music

Readings for Next Lecture

- TDO 5 through 5.4
- Kent 4
- Fellbaum, Chrisiane. Wordnet link.springer.com/chapter/10.1007%2F978-90-481-8847-5_10 (skip sections 10.12-10.15)