Vision Intelligence II: Visual Information Retrieval

Wei Qi Yan

Auckland University of Technology

Table of Contents

Image Query by Keywords

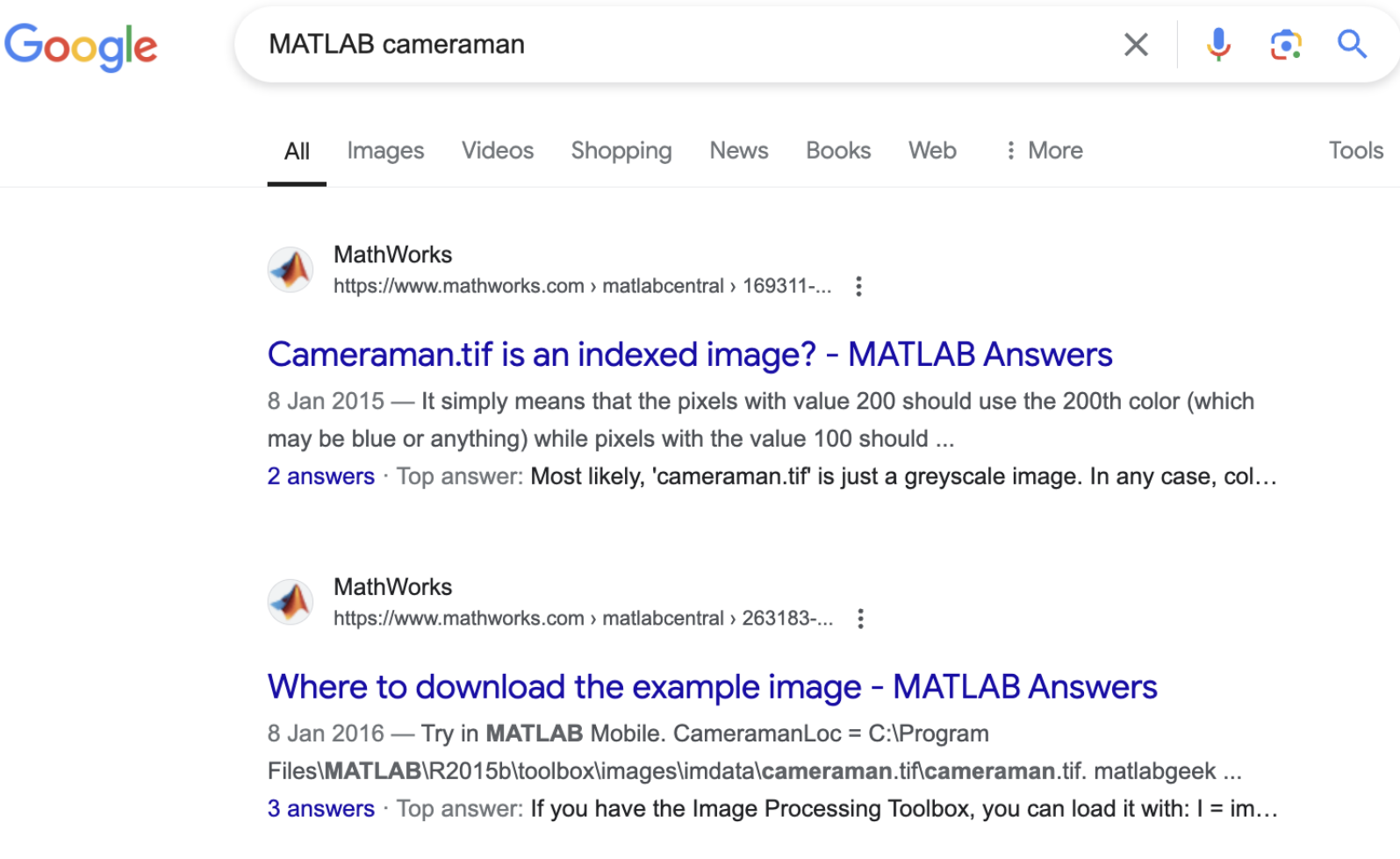


Image Query by Example

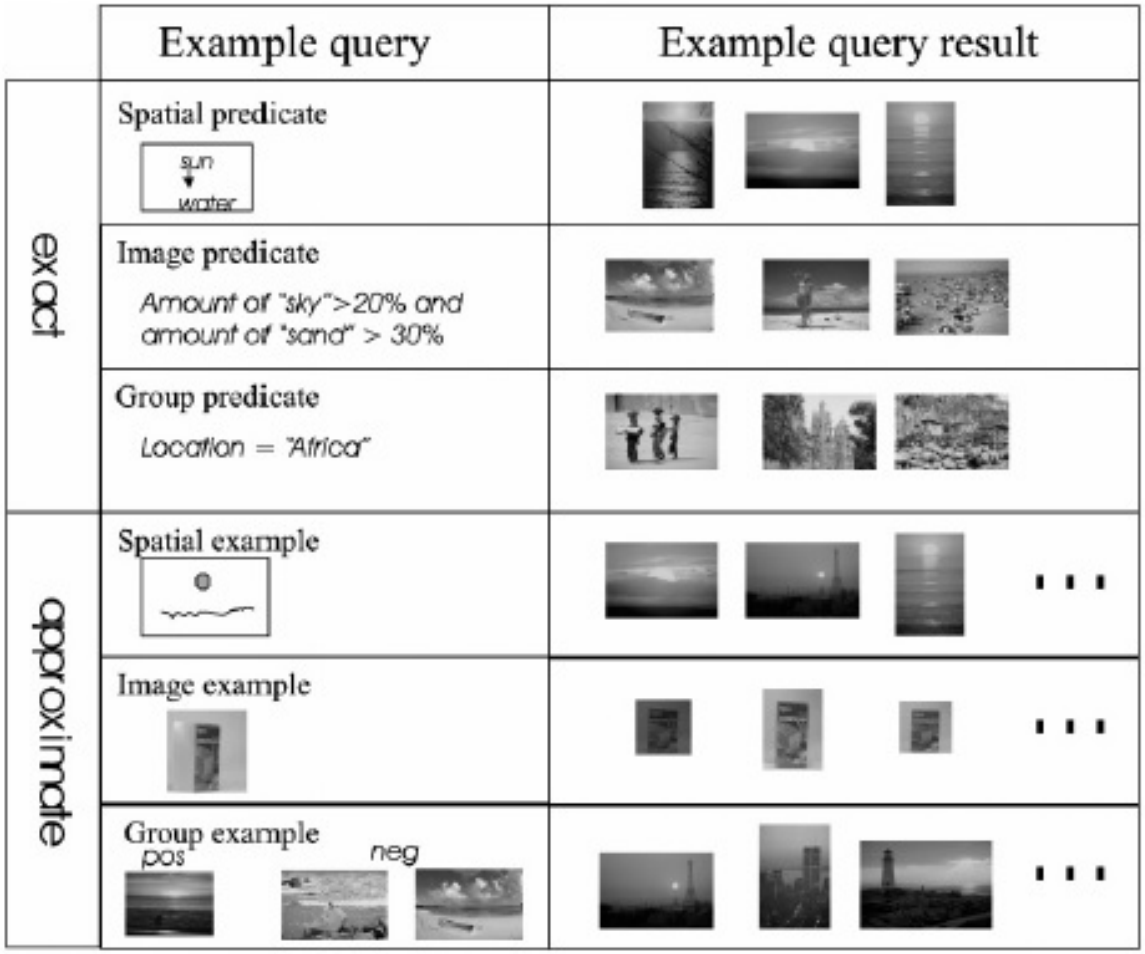


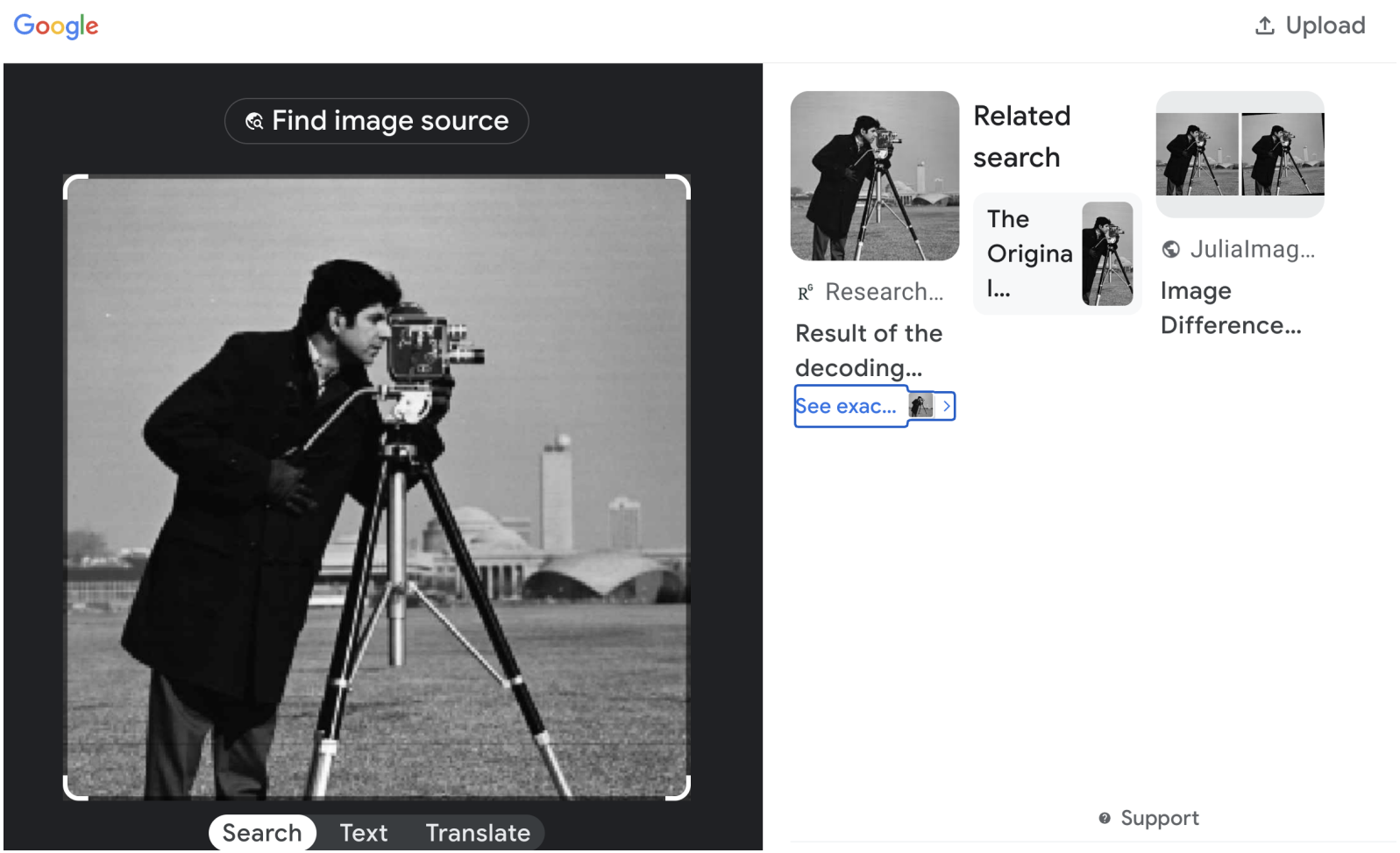
Image Query By Example

- Manual tagging can never completely describe image contents.
- Image content is automatically extracted as features such as color, texture, shape, etc.
- Image content is described by using its feature vector.

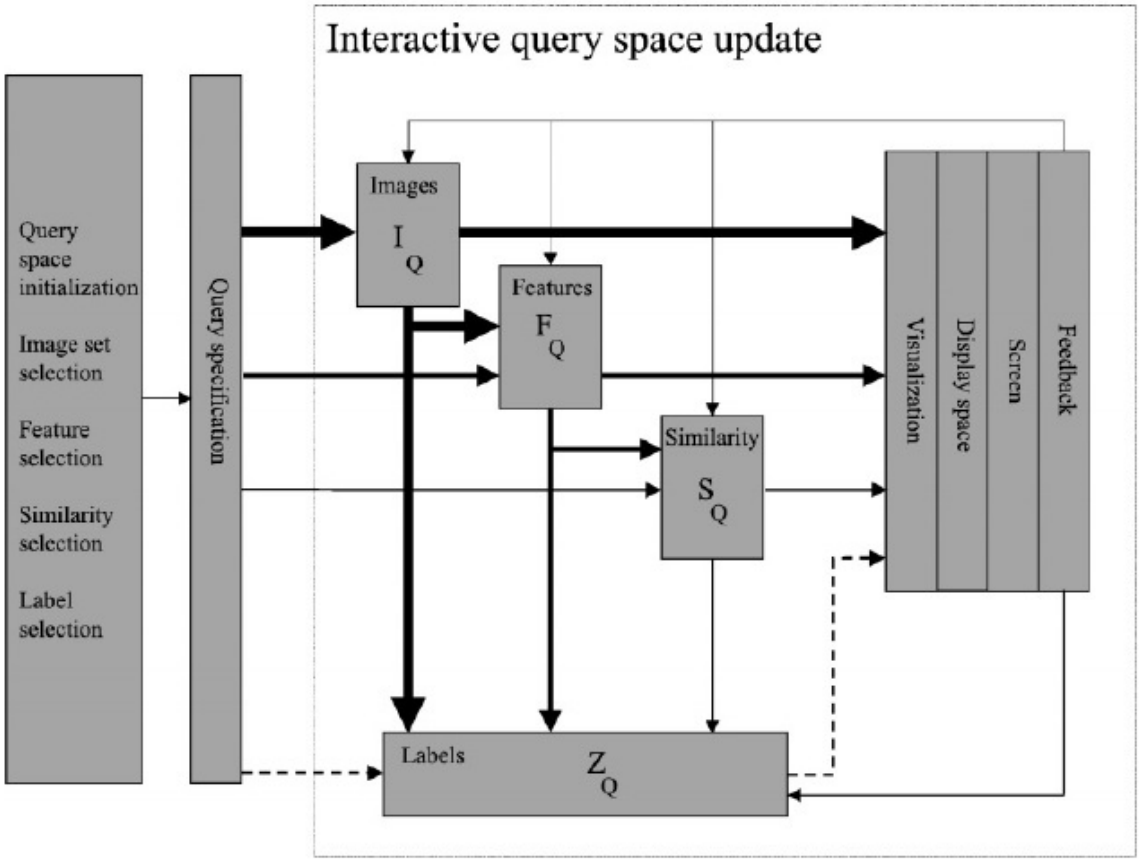
Key Steps of Image Query By Example

- Similarity matching
- Ranked list of items
- Evaluation of results

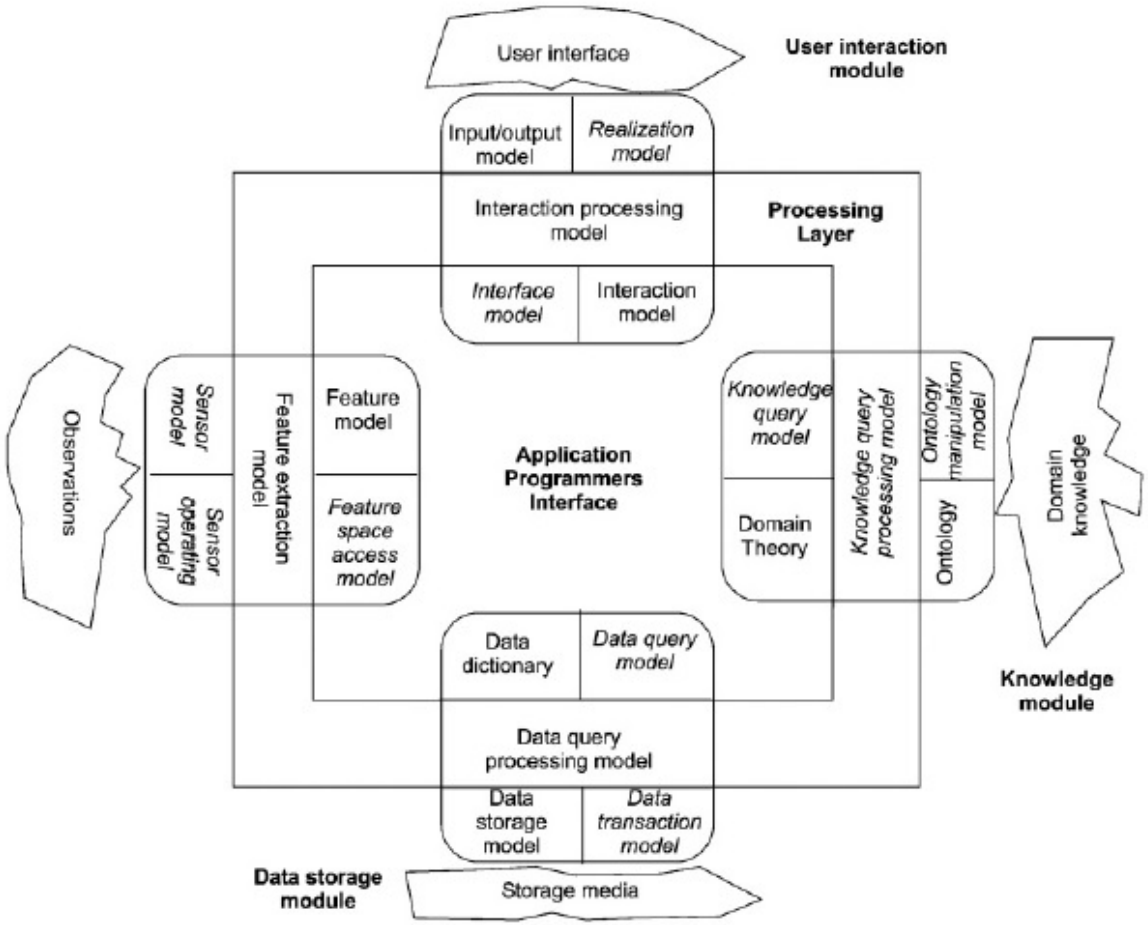
Image Query by Example



A Sketch


Interactive query space update
Query space initialization
Image set selection
Feature selection
Similarity selection
Label selection
Query specification
Images
I Q
Features
F Q
Similarity
S Q
Labels
Z Q
Visualization
Display space
Screen
Feedback

The Architecture



Questions?

Questions?

The key steps of image query by example include:

1. Find the corresponding keywords
2. Similarity matching
3. Raked list of items
4. Evaluations of results

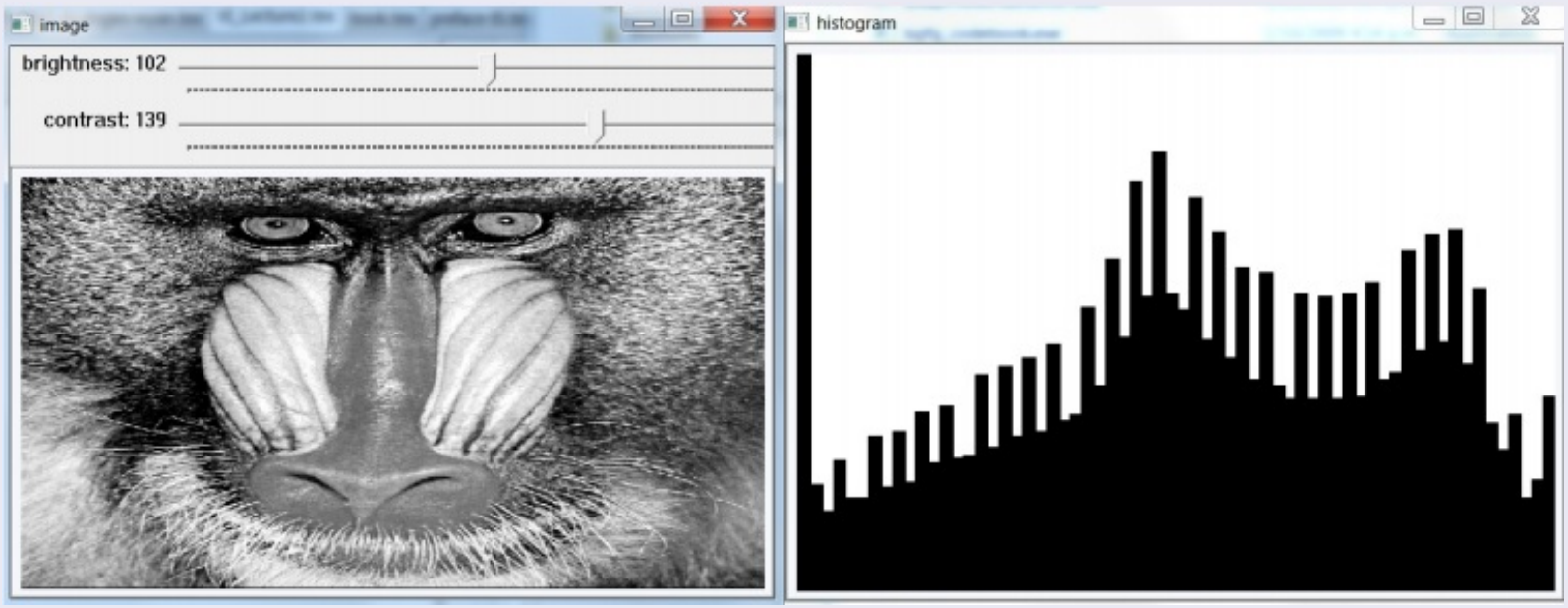
The wrong answer is:___

Questions?

Image Features

- Color
- Histogram
- Texture
- Shape/edge/silhouette
- Corners
- .....

Image Histogram



Histogram Distance: Inner Product

For histograms $Q = (H_1^q, H_2^q, \cdots, H_b^q)$ and $D = (H_1^d, H_2^d, \cdots, H_b^d)$,

$$< Q, D >= Q \cdot D = |Q| \cdot |D| \cos(\alpha) = \sum_{i=1}^{b} (H_i^q \cdot H_i^d)$$

$$\cos(\alpha) = \frac{< Q, D >}{|Q| \cdot |D|} = \frac{Q \cdot D}{|Q| \cdot |D|} = \frac{\sum_{i=1}^{b} (H_i^q \cdot H_i^d)}{\sqrt{\sum_{i=1}^{b} (H_i^q)^2} \sqrt{\sum_{i=1}^{b} (H_i^d)^2}}$$

Question: If $Q = (0.1, 0.1, 0.5)$ and $D = (0.2, 0.5, 0.0)$, what are $< Q, D >$ and $\cos(\alpha)$?

Histogram Distance: Inner Product

For histograms $Q = (H_1^q, H_2^q, \cdots, H_b^q)$ and $D = (H_1^d, H_2^d, \cdots, H_b^d)$,

$$< Q, D >= Q \cdot D = |Q| \cdot |D| \cos(\alpha) = \sum_{i=1}^{b} (H_i^q \cdot H_i^d)$$

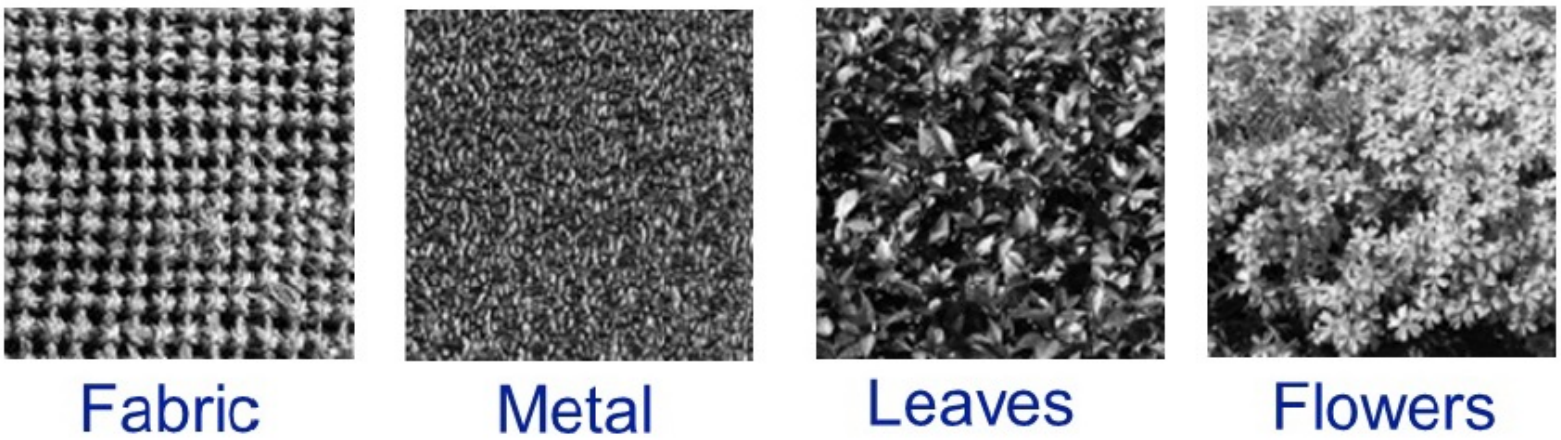
$$\cos(\alpha) = \frac{< Q, D >}{|Q| \cdot |D|} = \frac{Q \cdot D}{|Q| \cdot |D|} = \frac{\sum_{i=1}^{b} (H_i^q \cdot H_i^d)}{\sqrt{\sum_{i=1}^{b} (H_i^q)^2} \sqrt{\sum_{i=1}^{b} (H_i^d)^2}}$$

Question: If $Q = (0.1, 0.1, 0.5)$ and $D = (0.2, 0.5, 0.0)$, what are $< Q, D >$ and $\cos(\alpha)$?

Hints: $< Q, D > = 0.1 \cdot 0.2 + 0.1 \cdot 0.5 + 0.5 \cdot 0.0$;
$|Q| = \sqrt{0.1^2 + 0.1^2 + 0.5^2}$; $|D| = \sqrt{0.2^2 + 0.5^2 + 0.0^2}$;

Texture Analysis

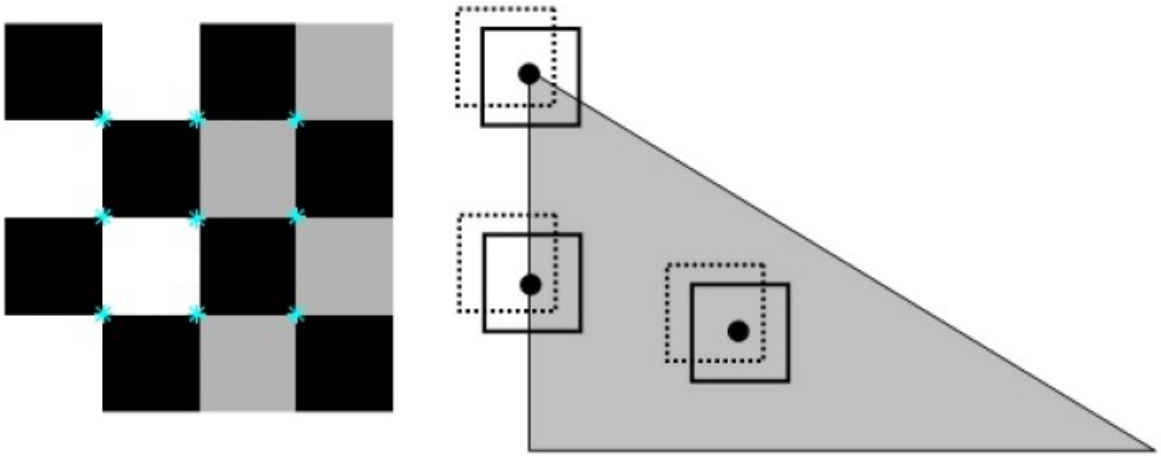
- Texture: Patterns repeat with variation
- Texture feature of GLCM (Gray-level Co-occurrence Matrix): Contrast, correlation, energy, homogeneity.



URL: https://au.mathworks.com/help/images/ref/graycoprops.html

Corners

- A corner in an image is given at a pixel where two edges of different directions intersect.
- Corners usually lie on high-contrast regions of the image.
- Relative positions between corners in the original scene shouldn't change.



Corners

- Corners are invariant to scaling, orientation, and distortions.
- The best match for each pair of corners is found by identifying its nearest neighbor of corners.
- The nearest neighbors are defined as the corners with the minimum distance from the given descriptor vector.

MATLAB: RANSAC

Questions?

Questions?

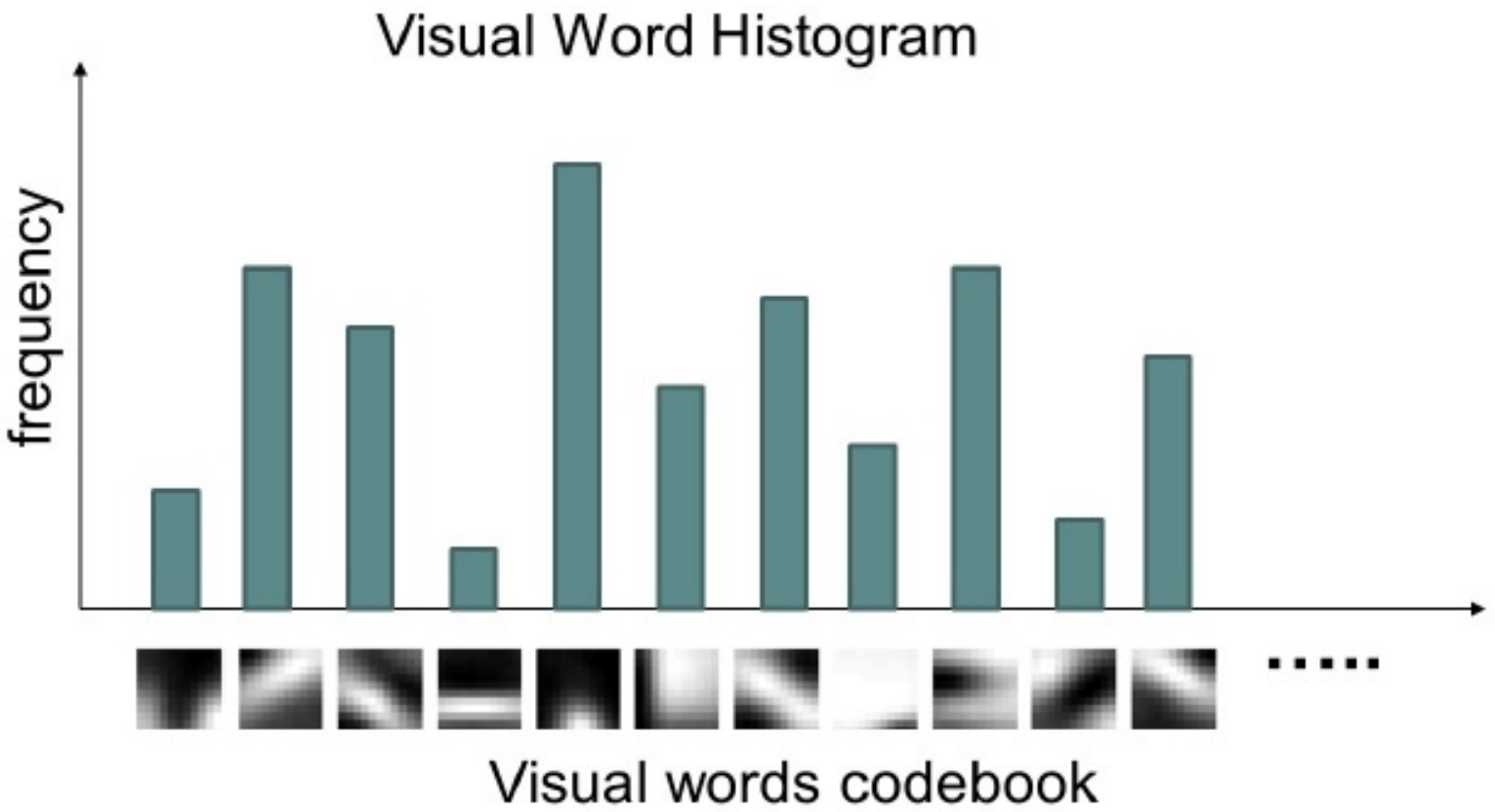
Pertaining to image corners:

1. Relative positions between image corners in the original scene should not change.
2. Relative positions between image corners in the original scene should change.
3. Relative positions between image corners in the original scene should be uncertain.
4. None of the given options.

The right answer is:___

Questions?

Visual Words



Visual Words

- Visual word: A small patch on the image (array of pixels) which can carry any kind of interesting information in any feature space (color changes, texture changes,$\cdots$, etc.).
- Visual term: The clustering result in the feature space (centers of the clusters), more than one patch can give nearest information in feature space.
- Visual language: A set of visual words & visual terms or visual vocabulary.
- Bag of visual words: A collection of visual words which together can give information about the meaning of the image.

Inverted Index

1. Sentence 1: “what is it”
2. Sentence 2: “it is a banana”
3. Sentence 3: “it is”

Inverted Index

1. Sentence 1: “what is it”
2. Sentence 2: “it is a banana”
3. Sentence 3: “it is”

- “a”:{2}
- “banana”:{2}
- “is”:{1,2,3}
- “it”:{1,2,3}
- “what”:{1}

Inverted Index

1. Sentence 1: “what is it”
2. Sentence 2: “it is a banana”
3. Sentence 3: “it is”

- “a”:{2}
- “banana”:{2}
- “is”:{1,2,3}
- “it”:{1,2,3}
- “what”:{1}

Search: “what is it”

Calculate: $\{1\} \cap \{1,2,3\} \cap \{1,2,3\} = \{1\}$

Remark: It is the intersection of these posting lists

Questions?

Questions?

In visual information retreval, the inverted index is:

1. The intersection of these posting lists.
2. The union of these posting lists.
3. The complement of these posting lists.
4. The disjoint of these posting lists.

The right answer is: ___

Questions?

Metadata-Based Image Search

- EXIF (Exchangeable Image File Format): Timestamp, focal length, shutter speed, aperture, etc.
  1. Privacy and security
  2. FlashPix extensions
  3. Exif audio files
- Semantic tags: Visual words, concepts and ontology
- Visual keywords search

Note: Metadata can improve the accuracy of the pure content-based methods.

EXIF Entities

Tag	Value
Manufacturer	CASIO
Model	QV-4000
Orientation (rotation)	top - left [8 possible values[21]]
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)

Extension EXIF Entities

Exif audio files

Tag	Value
Encoding	Microsoft PCM
Num Channels	1
Sample Rate	7872
Avg Bytes Per Sec	7872
Bits Per Sample	8
Date Created	2005:08:08
Exif Version	0220
Related Image File	IMGP1149.JPG
Time Created	16:23:35
Make	PENTAX Corporation
Model	PENTAX Optio WP
MakerNote	(2064 bytes of data)

FlashPix extensions

Tag	Value
Code Page	1200
Used Extension Numbers	1
Extension Name	Screen nail
Extension Class ID	10000230-6FC0-11D0-BD01-00609719A180
Extension Persistence	Invalidated By Modification
Extension Create Date	2003:03:29 17:47:50
Extension Modify Date	2003:03:29 17:47:50
Creating Application	Picoss
Extension Description	Presized image for LCD display
Storage-Stream Pathname	/.Screen Nail_bd0100609719a180
Screen Nail	(124498 bytes of data containing 640x480 JPEG preview image)

Questions?

Questions?

Image EXIF data includes:

1. Timestamp
2. Focal length
3. Shutter speed
4. Visual tokens

The wrong answer is:___

Questions?

Vision Intelligence II: Visual Information Retrieval

Learning Objectives

- Extract the features that can be used to understand visual information.
- Demonstrate knowledge of how to apply algorithms and techniques for vision modelling.