KNOWLEDGE MEDIA KMi INSTITUTE

Visual mining - interpreting image data

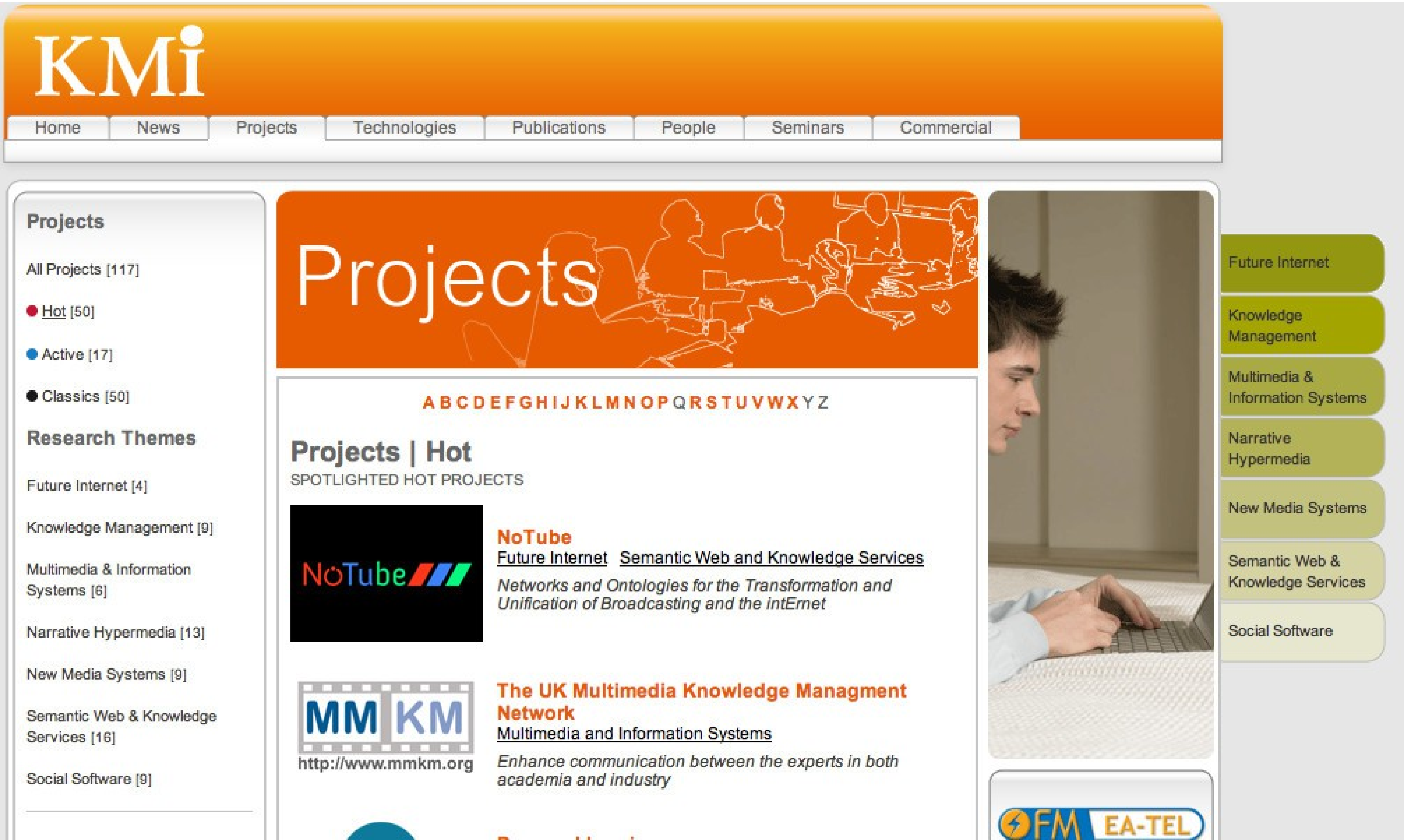
Professor Stefan Rüger

Multimedia and Information Systems
Knowledge Media Institute
The Open University

http://kmi.open.ac.uk/mmis

KNOWLEDGE MEDIA KMi INSTITUTE

kmi.open.ac.uk



Overview

Near Duplicate Detection

- **Examples**
- Principal techniques
- Shape invariant feature vector

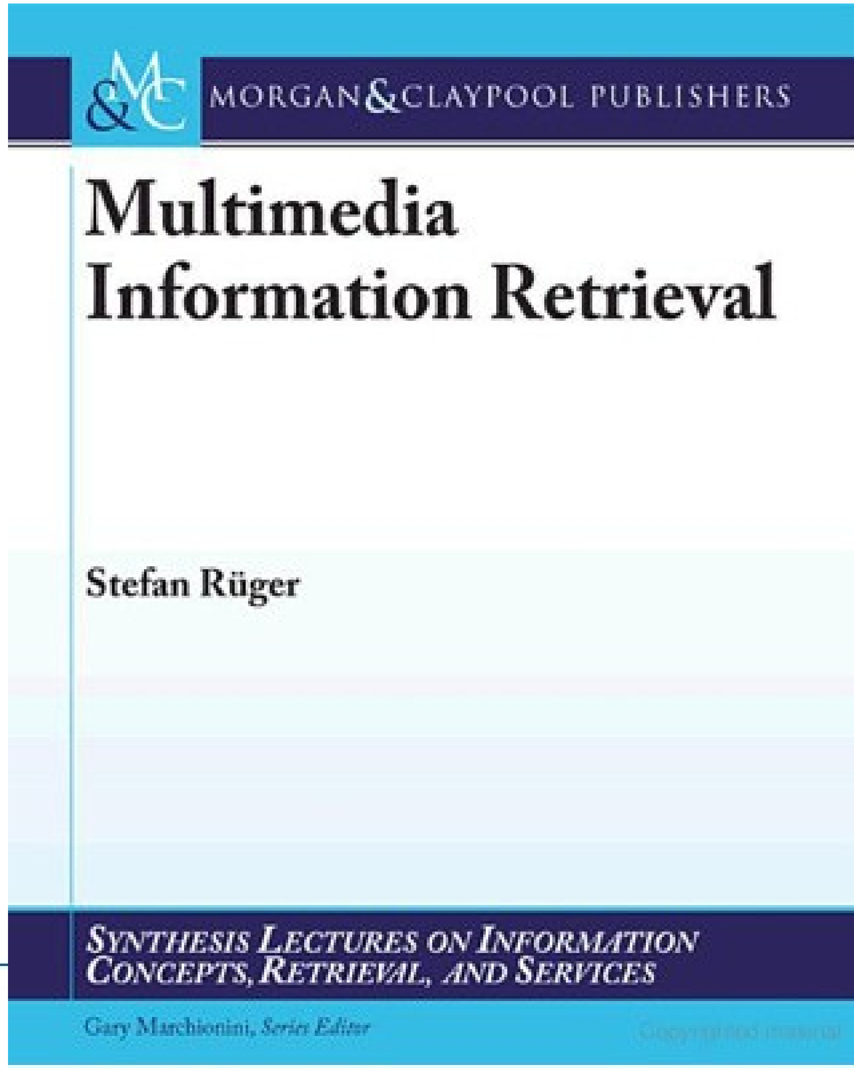
Machine Learning + Computer Vision

- Judging the aesthetics of pictures (example)
- Food analysis (example)

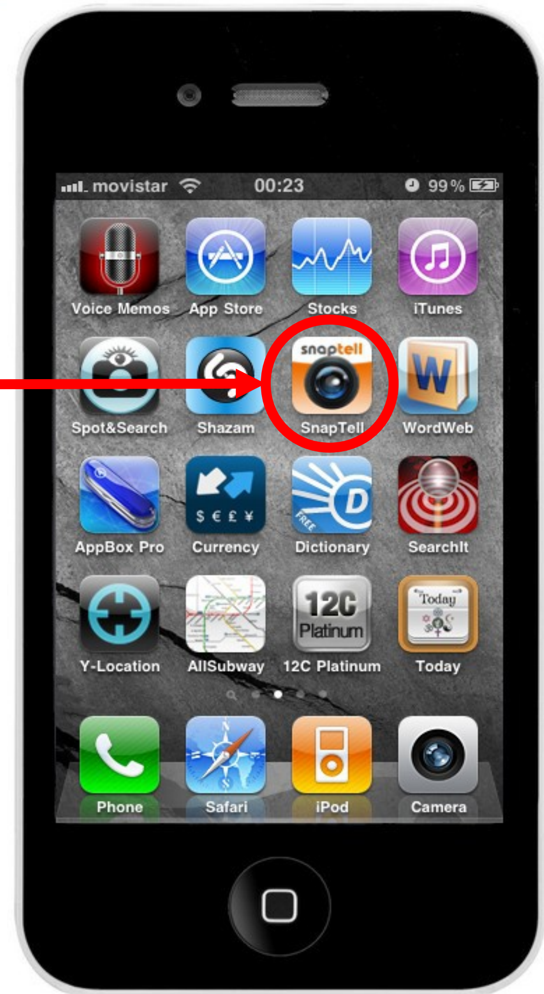
Media understanding

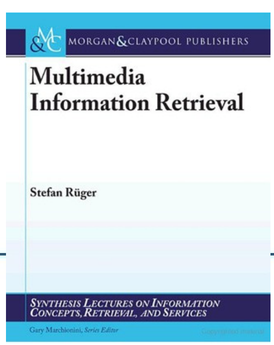
- “learning from watching TV”

Near-duplicate detection: Cool access mode!

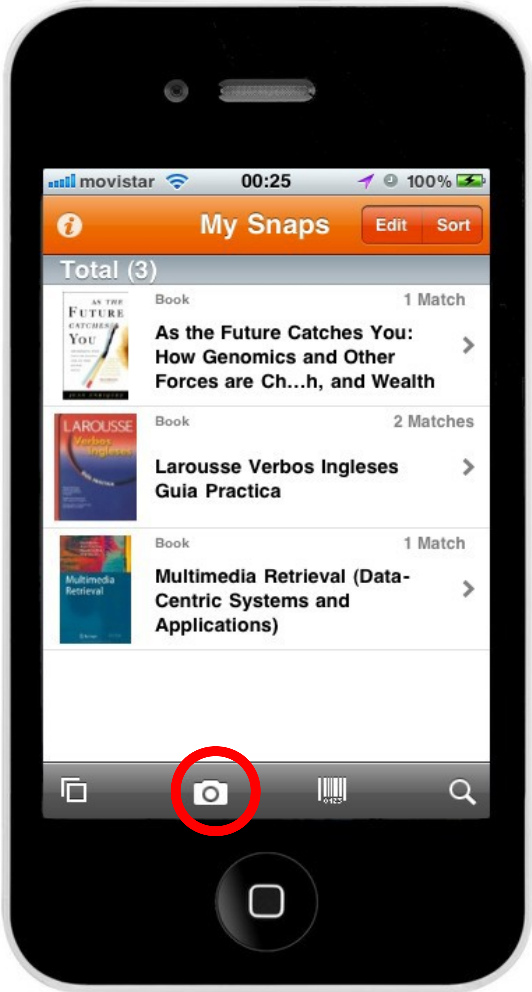


Snaptell: Book, CD and DVD covers

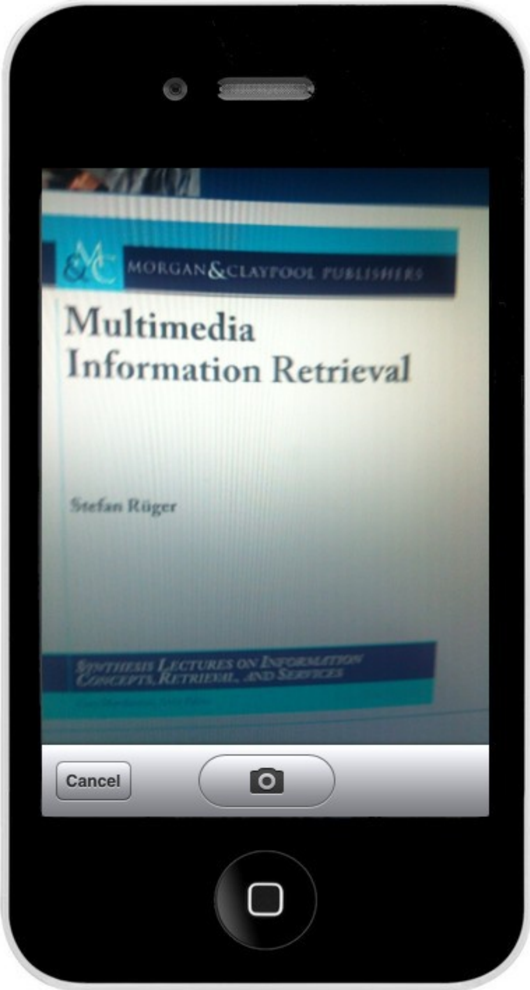




Snaptell: Book, CD and DVD covers

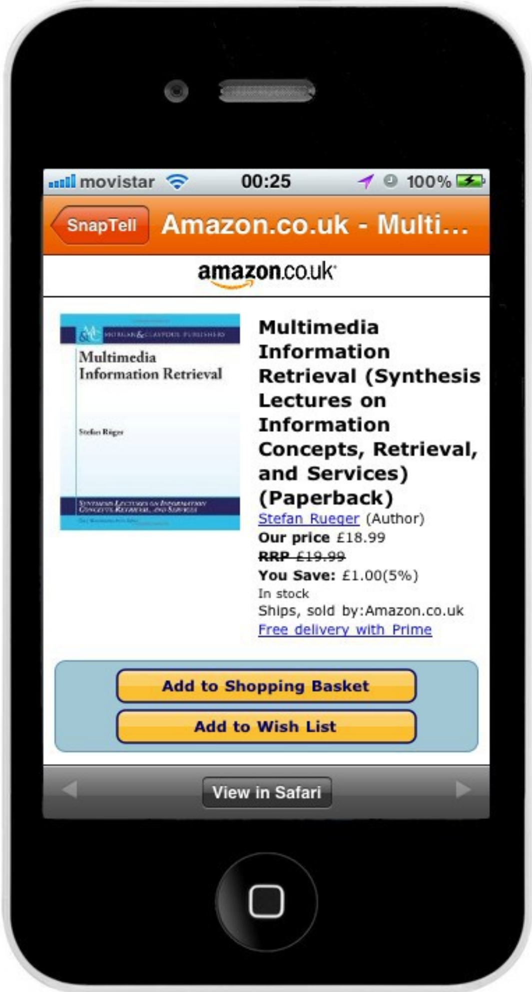


Snaptell: Book, CD and DVD covers

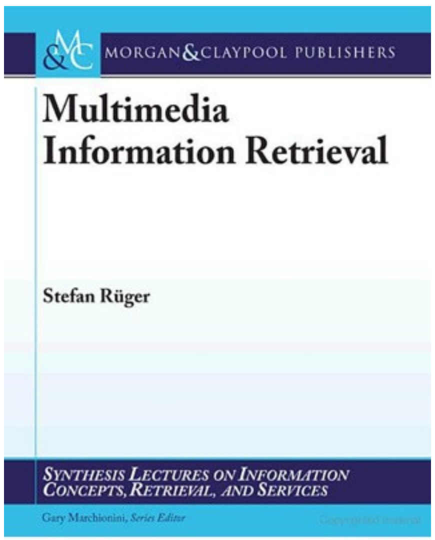


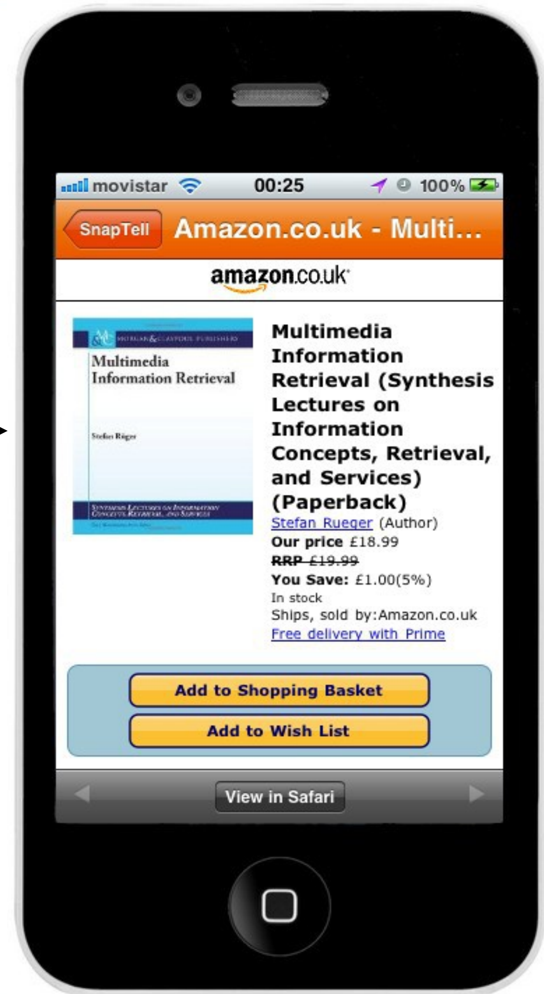
Snaptell: Book, CD and DVD covers

Snaptell: Book, CD and DVD covers



Link from real world to databases





doi: 10.2200/S00244ED1V01Y200912ICR010

The Open Univerity's Spot & Search

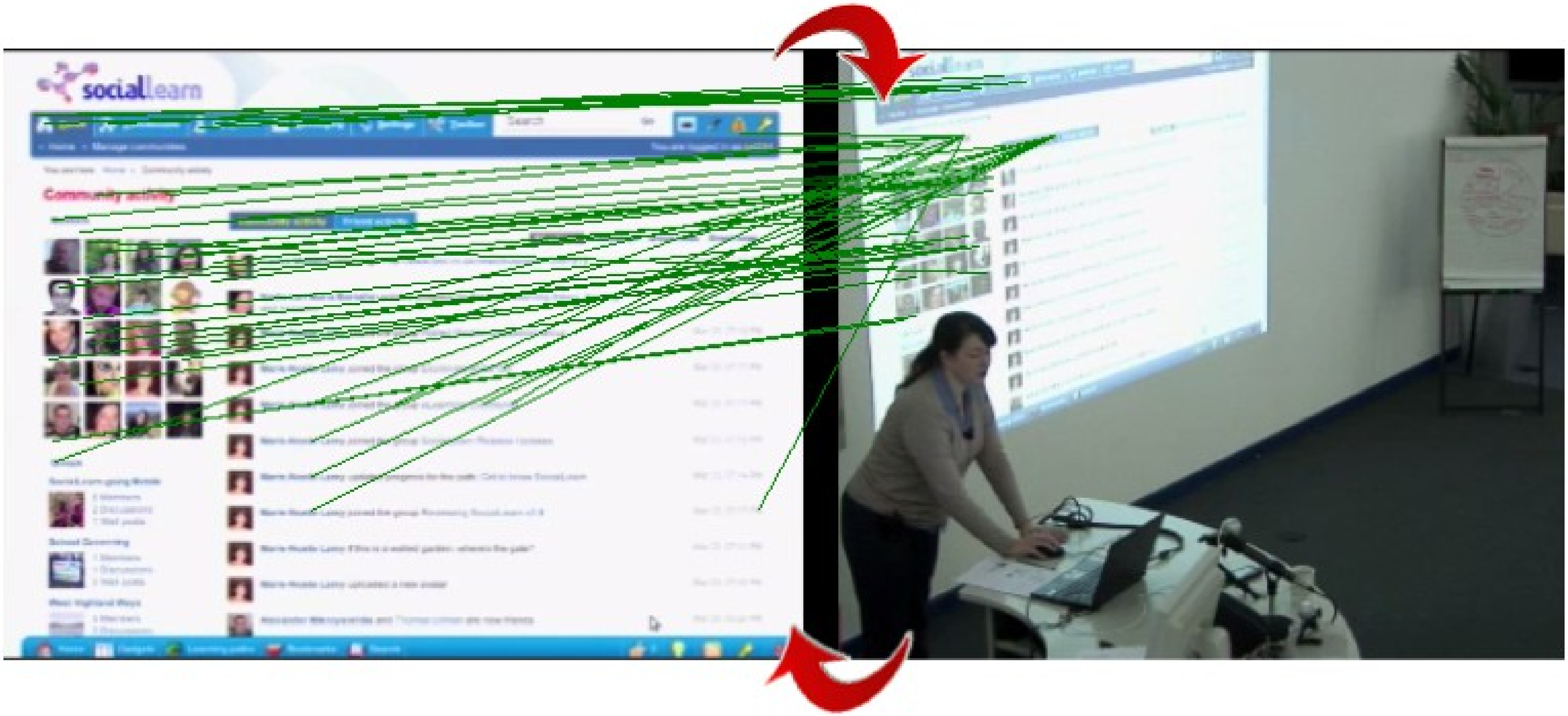
Scott Forrest: E=MC squared

"Between finished surface texture and raw quarried stone. Between hard materials and soft concepts. Between text and context."

More information

[with Suzanne Little]

Linking media



[with Suzanne Little]

Near duplicate detection

Works well in 2d: CD covers, wine labels, signs, ...

Less so in near 2d: buildings, vases, …

Not so well in 3d: faces, complex objects, ...

Overview

Near Duplicate Detection

- Examples
- **Principal techniques**
- Shape invariant feature vector

Machine Learning + Computer Vision

- Judging the aesthetics of pictures (example)
- Food analysis (example)

Media understanding

- “learning from watching TV”

Near-duplicate detection: How does it work?

Fingerprinting technique

1 Compute salient points

2 Extract "characteristics" from vincinity (feature)

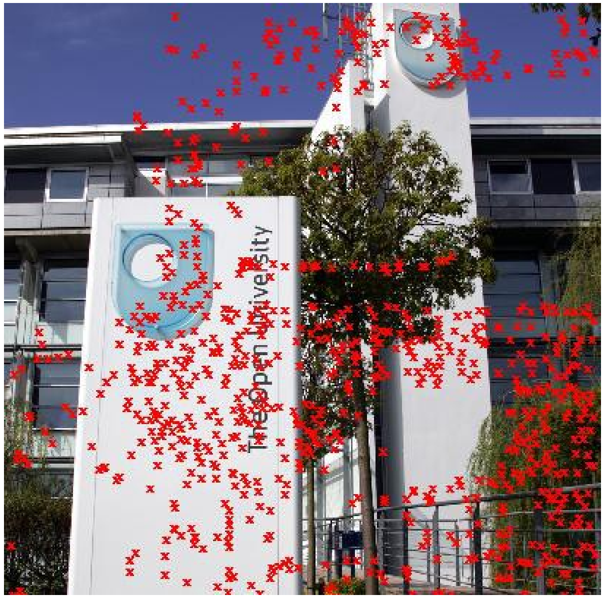
3 Make invariant under rotation & scaling

4 Quantise: create visterms

5 Index as in text search engines

6 Check/enforce spatial constraints after retrieval

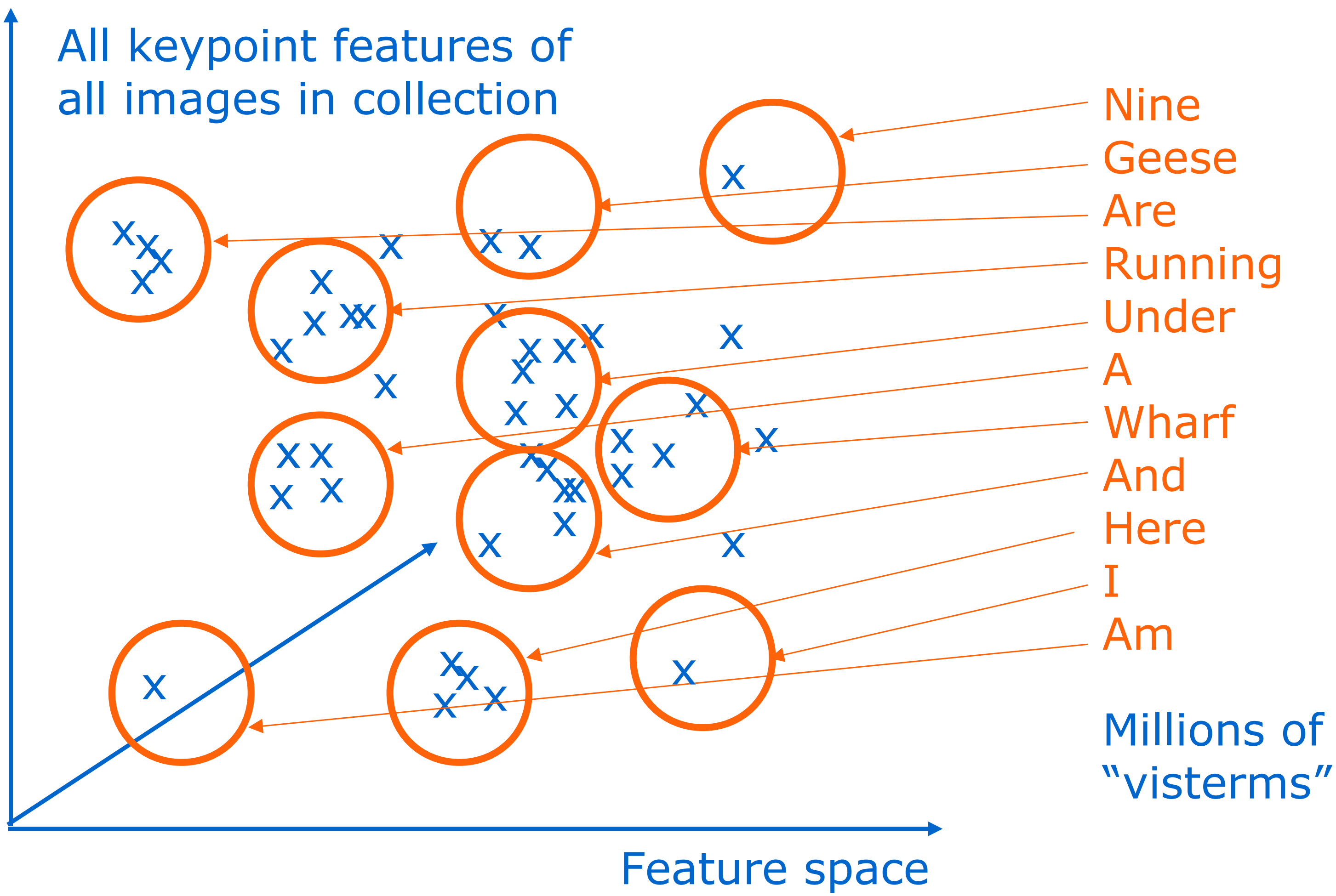
NDD: Compute salient points and features



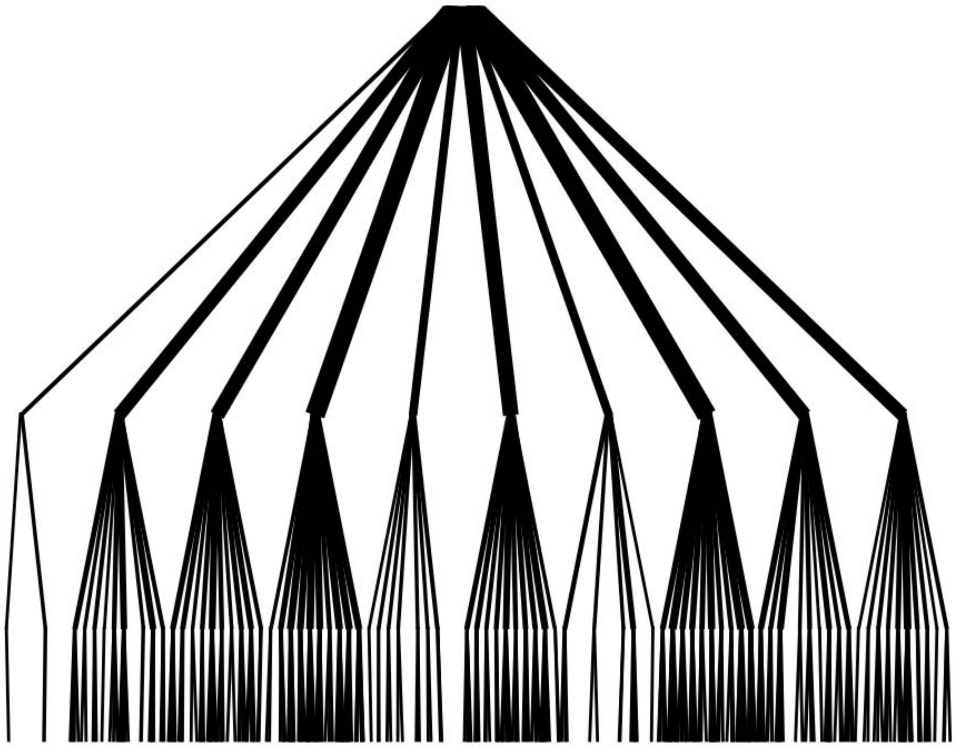
Eg, SIFT features: each salient point described by a feature vector of 128 numbers; the vector is invariant to scaling and rotation

[Lowe2004 – http://www.cs.ubc.ca/~lowe/keypoints/]

NDD: Keypoint feature space clustering

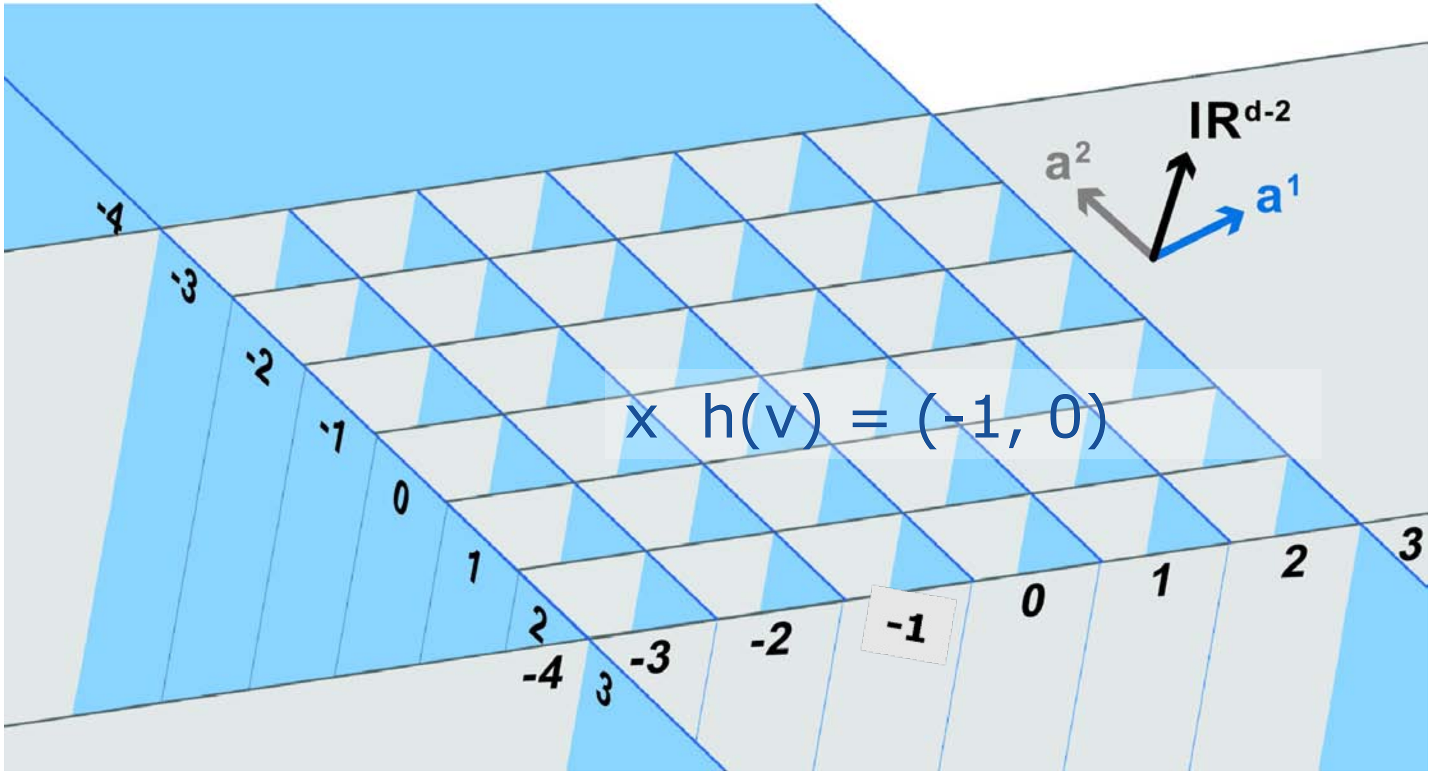


Clustering
Hierarchical k-means

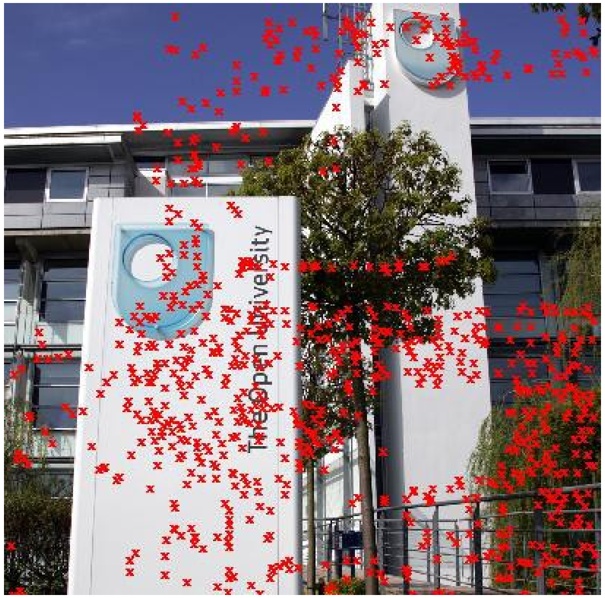


[Nister and Stewenius, CVPR 2006]

Quantisation
LSH hashes



NDD: Encode all images with visterms



Jkjh Geese Bjlkj Wharf Ojkkjhhj Kssn Klkekjl Here Lkjkll Wjjkll Kkjlk Bnm Kllkgjg Lwoe Boerm ...

NDD: query like text

At query time compute salient points, keypoint features and visterms

Query against database of images represented as bag of vistems

Query

Joiu Gddwd Bipoi Wueft
Oiooiuui Kwwn Kpodoip Hdfd
Loiopp Wiiopp Koipo Bnm
Kppoyiy Lsld Bldfm ...

NDD: Check spatial constraints

[with Suzanne Little, SocialLearn project]

Near-duplicate detection: Summary

Fingerprinting technique

1 Compute salient points

2 Extract “characteristics” from vincinity (feature)

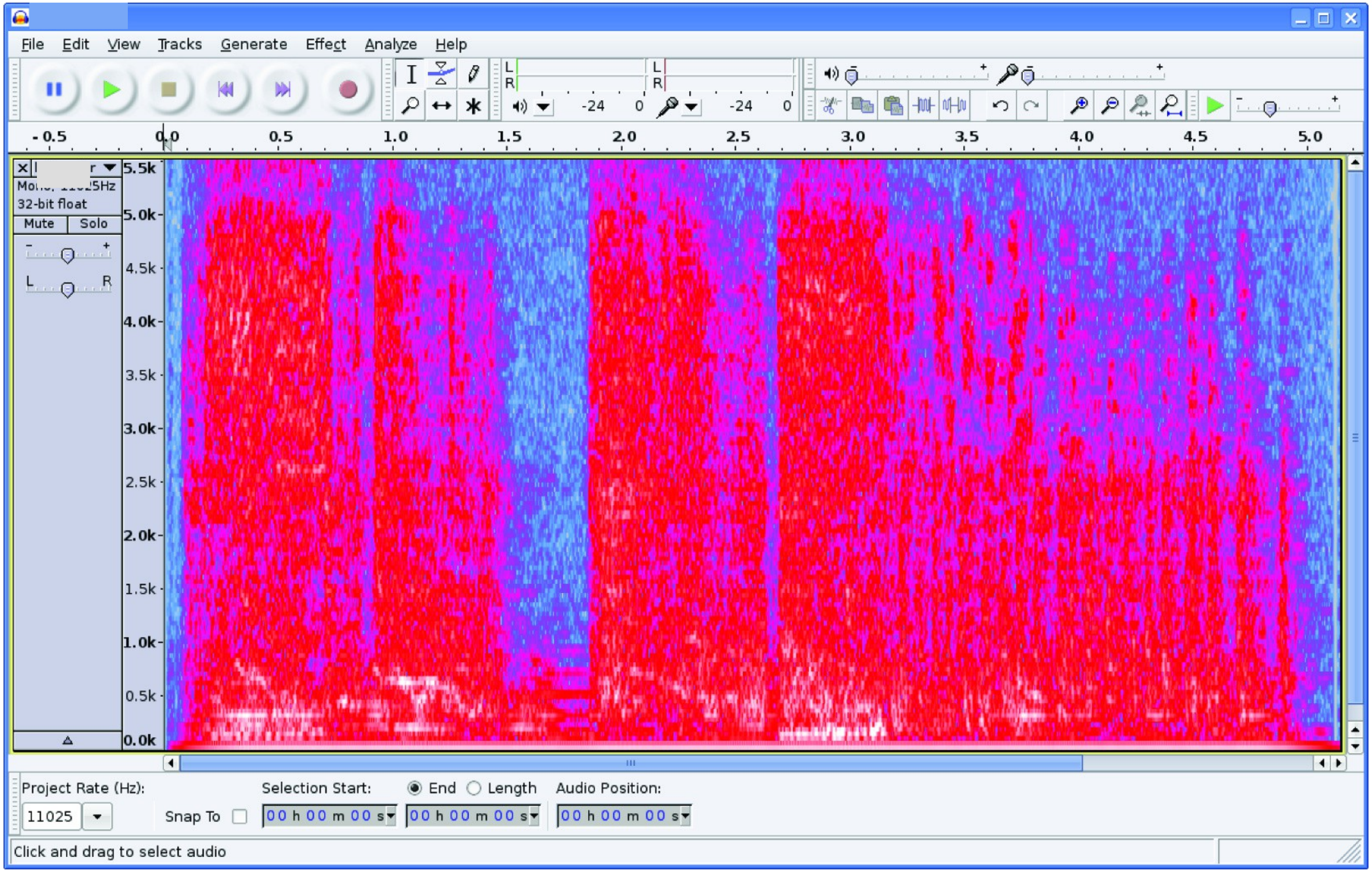
3 Make invariant under rotation & scaling

4 Quantise: create visterms

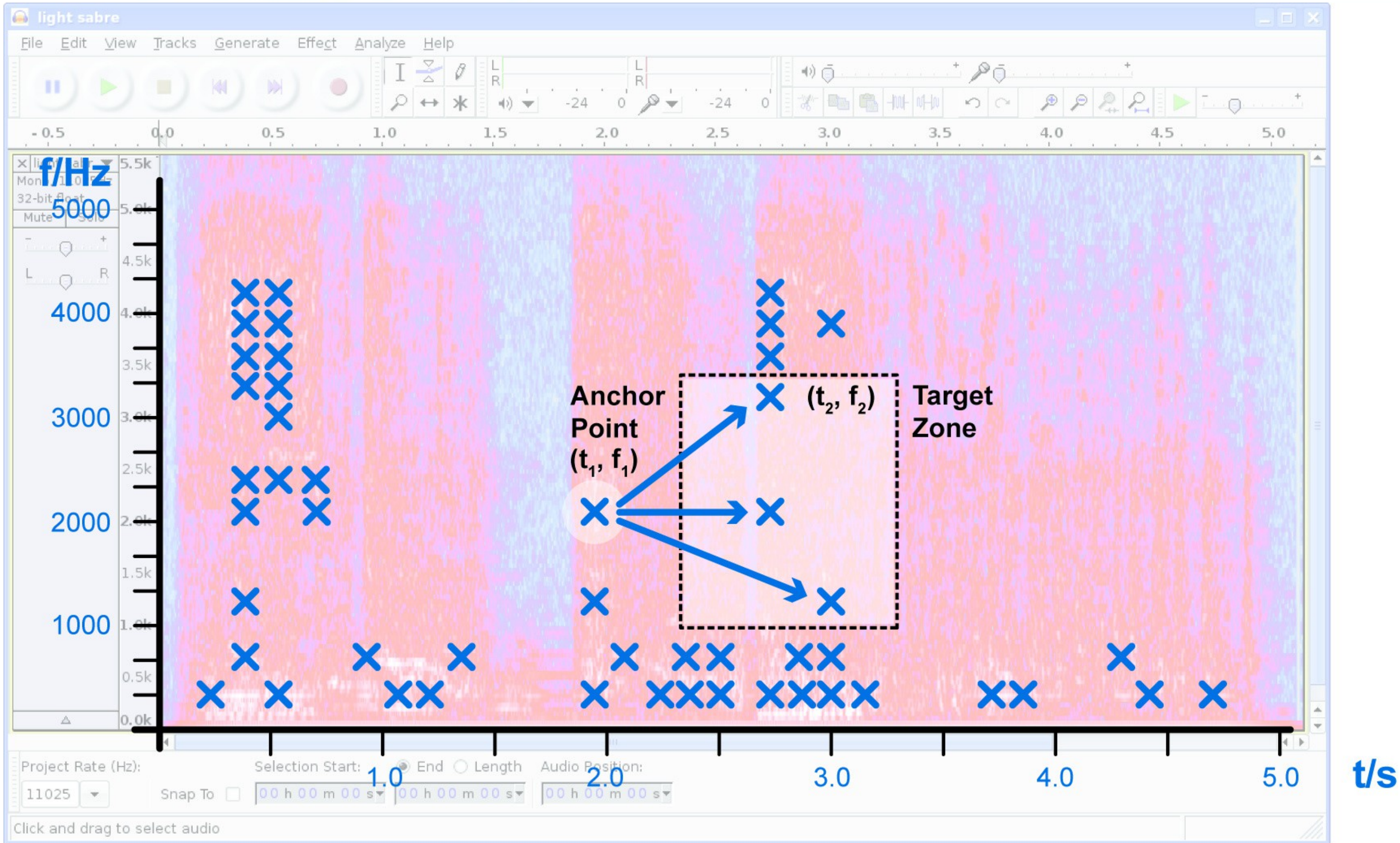
5 Index as in text search engines

6 Check/enforce spatial constraints after retrieval

Audio fingerprinting



Salient points



Encoding: (f_1, f_2, t_2-t_1) hashes to (t_1, id)

Overview

Near Duplicate Detection

- Examples
- Principal techniques
- **Shape invariant feature vector**

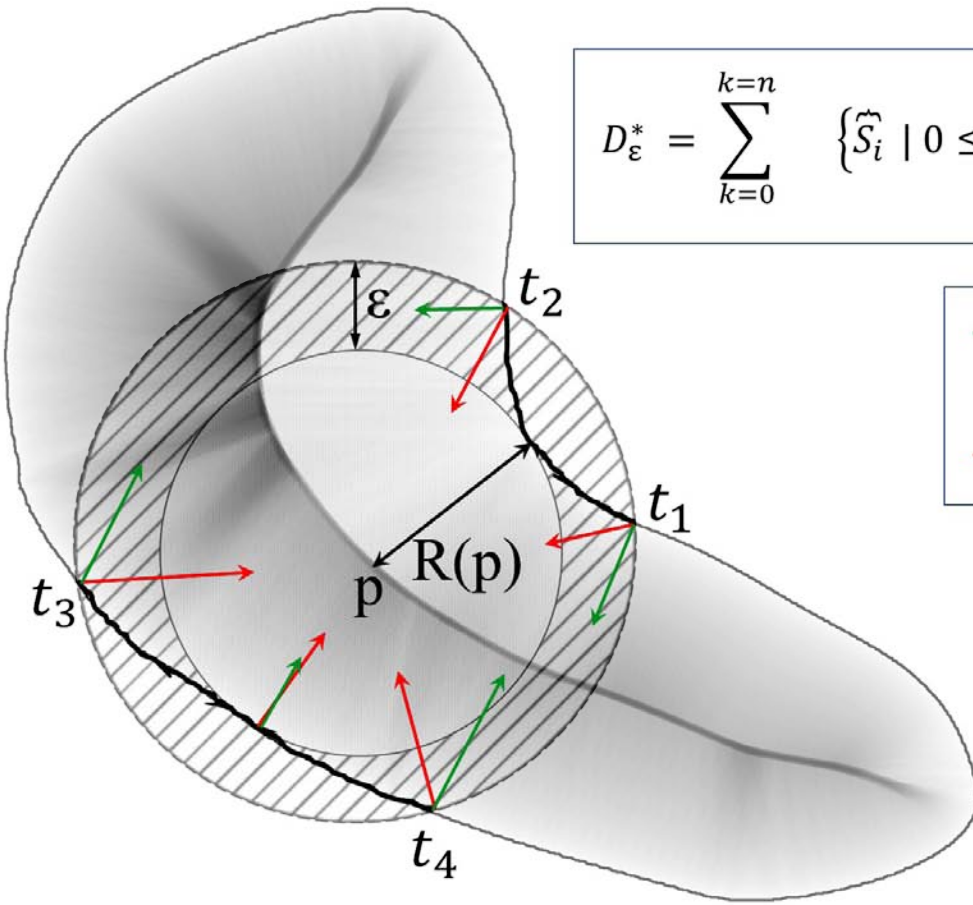
Machine Learning + Computer Vision

- Judging the aesthetics of pictures (example)
- Food analysis (example)

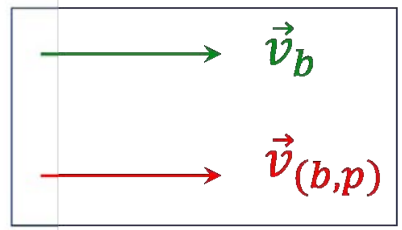
Media understanding

- "learning from watching TV"

Features are Key: Medialness

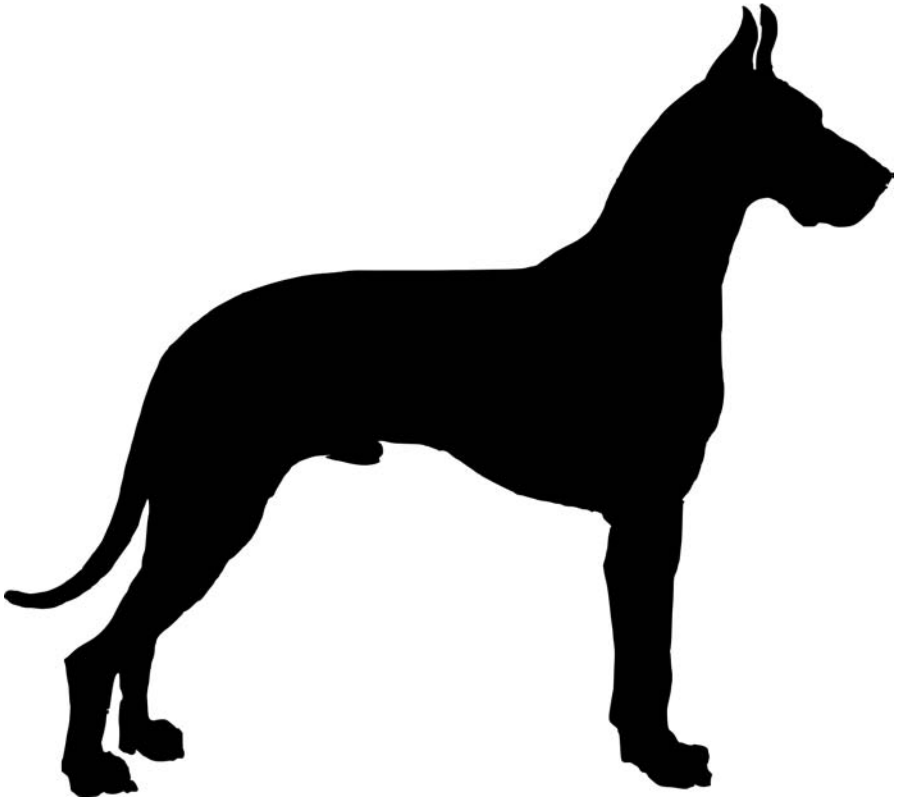


$$D^*_{\varepsilon} = \sum_{k=0}^{k=n} \left\{ \overset{\frown}{S}_i \mid 0 \le \vec{v}_b \cdot \vec{v}_{(b,p)} \right\}, i = 2k+1, n \ge 1$$



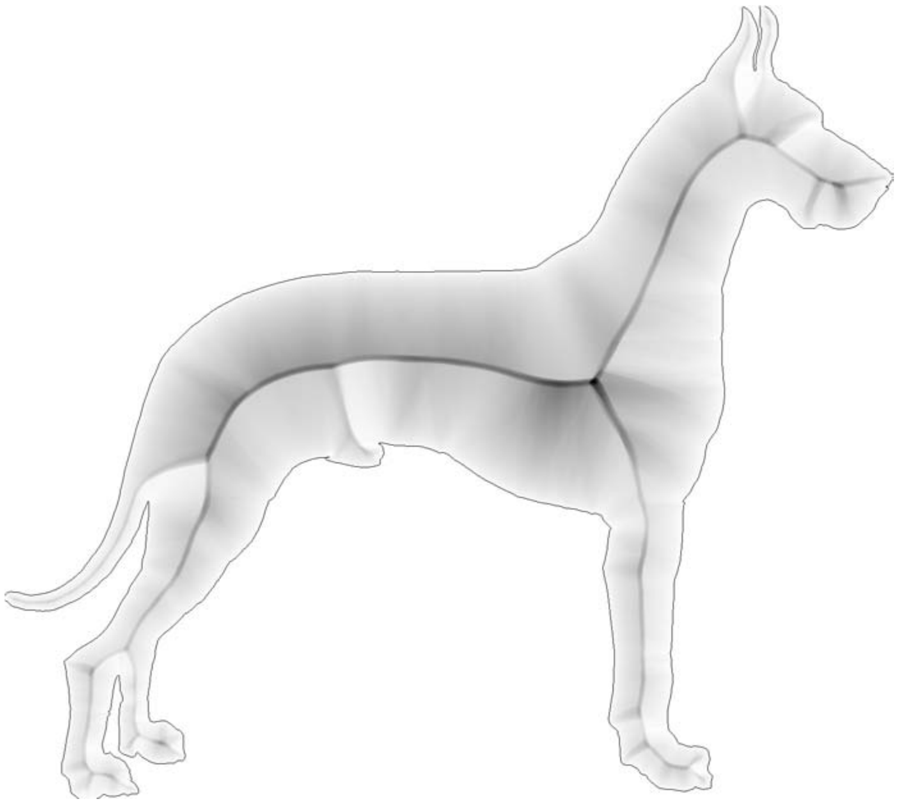
[F Fol Leymarie – after Kowacs]

Medialness: Example



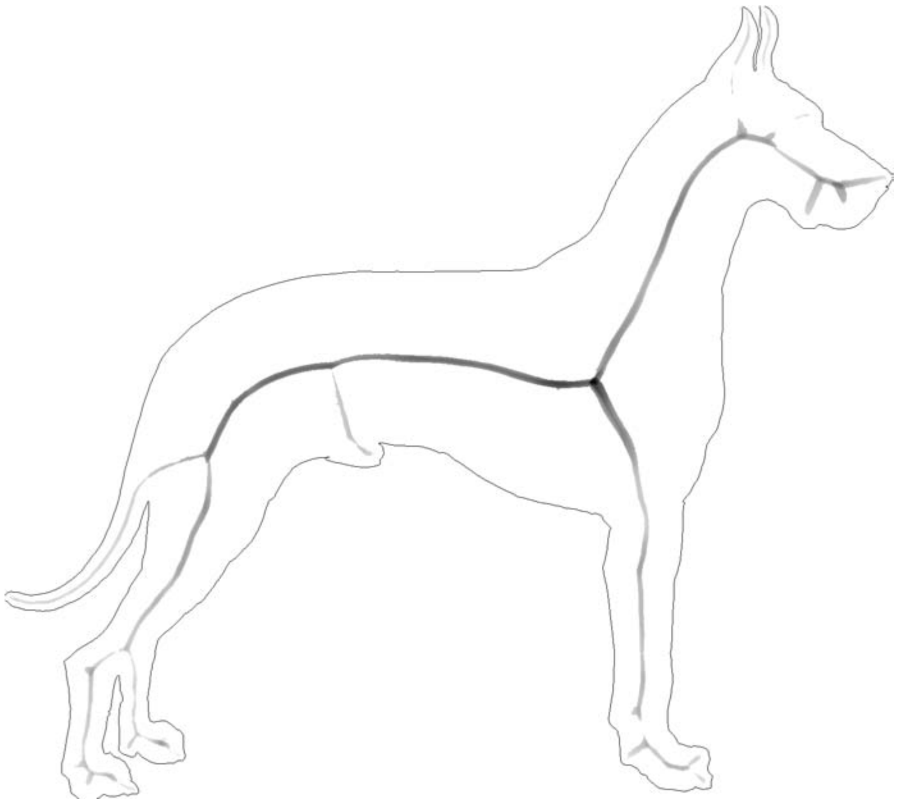
[F Fol Leymarie and P Aparajeya]

Medialness



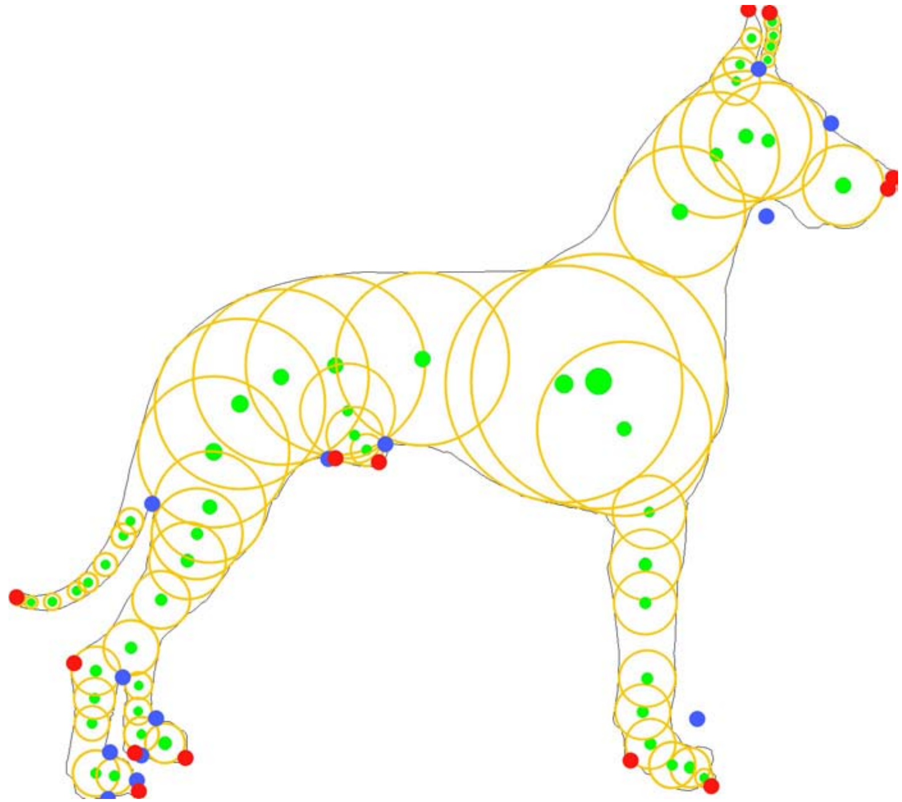
[F Fol Leymarie and P Aparajeya]

Medialness: top hat



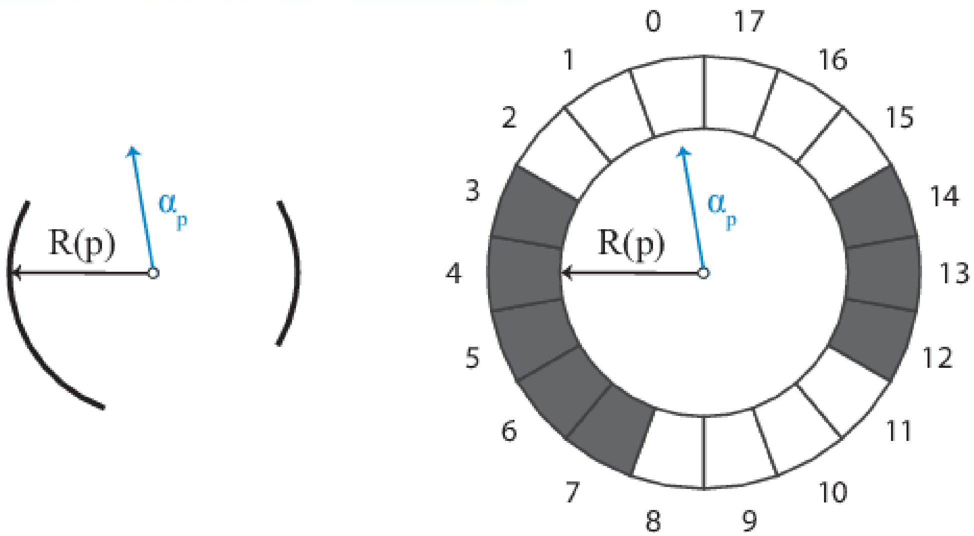
[F Fol Leymarie and P Aparajeya]

Medialness: dominant points



[F Fol Leymarie and P Aparajeya]

Shape Invariant Feature Vector: ShIFT



α_p = Orientation of Keypoint

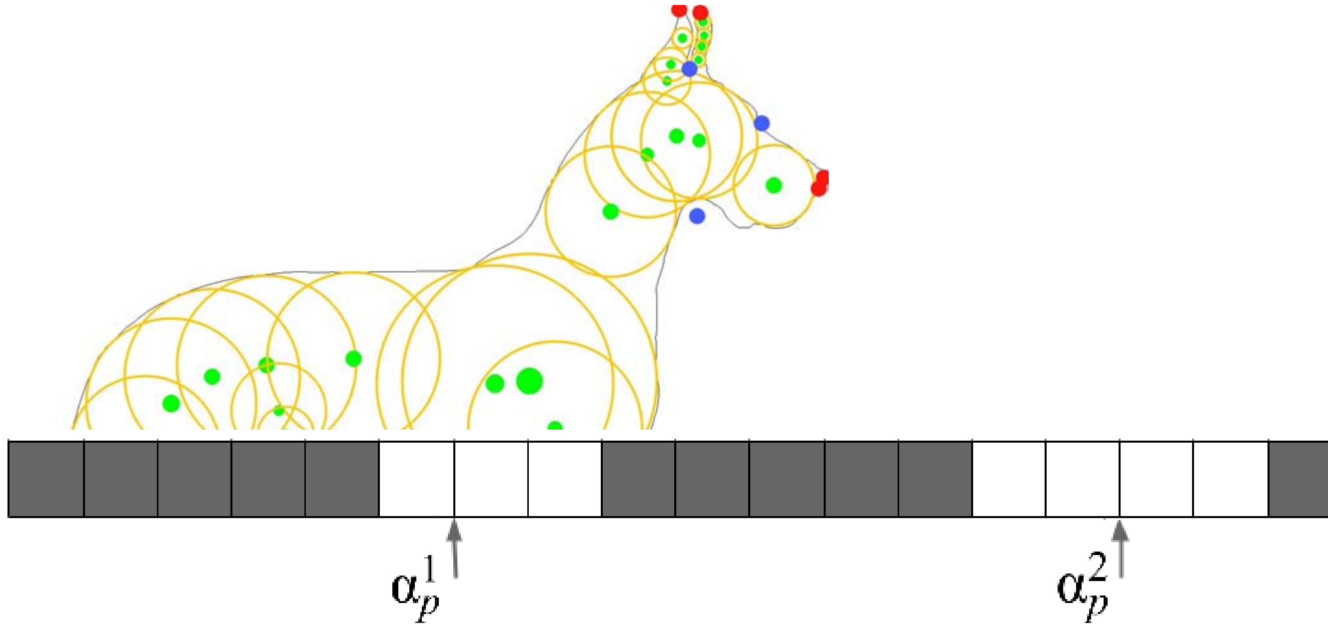
R(p) = Minimum Radial Distance

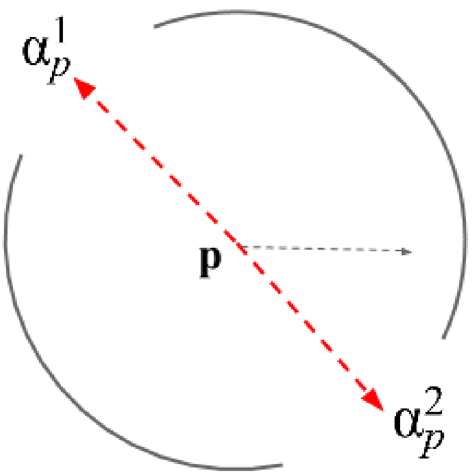
Bin No.	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Binary Code	0	0	0	1	1	1	1	1	0	0	0	0	1	1	1	0	0	0

Feature Vector <0,0,0,1,1,1,1,1,0,0,0,0,1,1,1,0,0,0>

[with F Fol Leymarie and P Aparajeya]

Shape Invariant Feature Vector: ShIFT



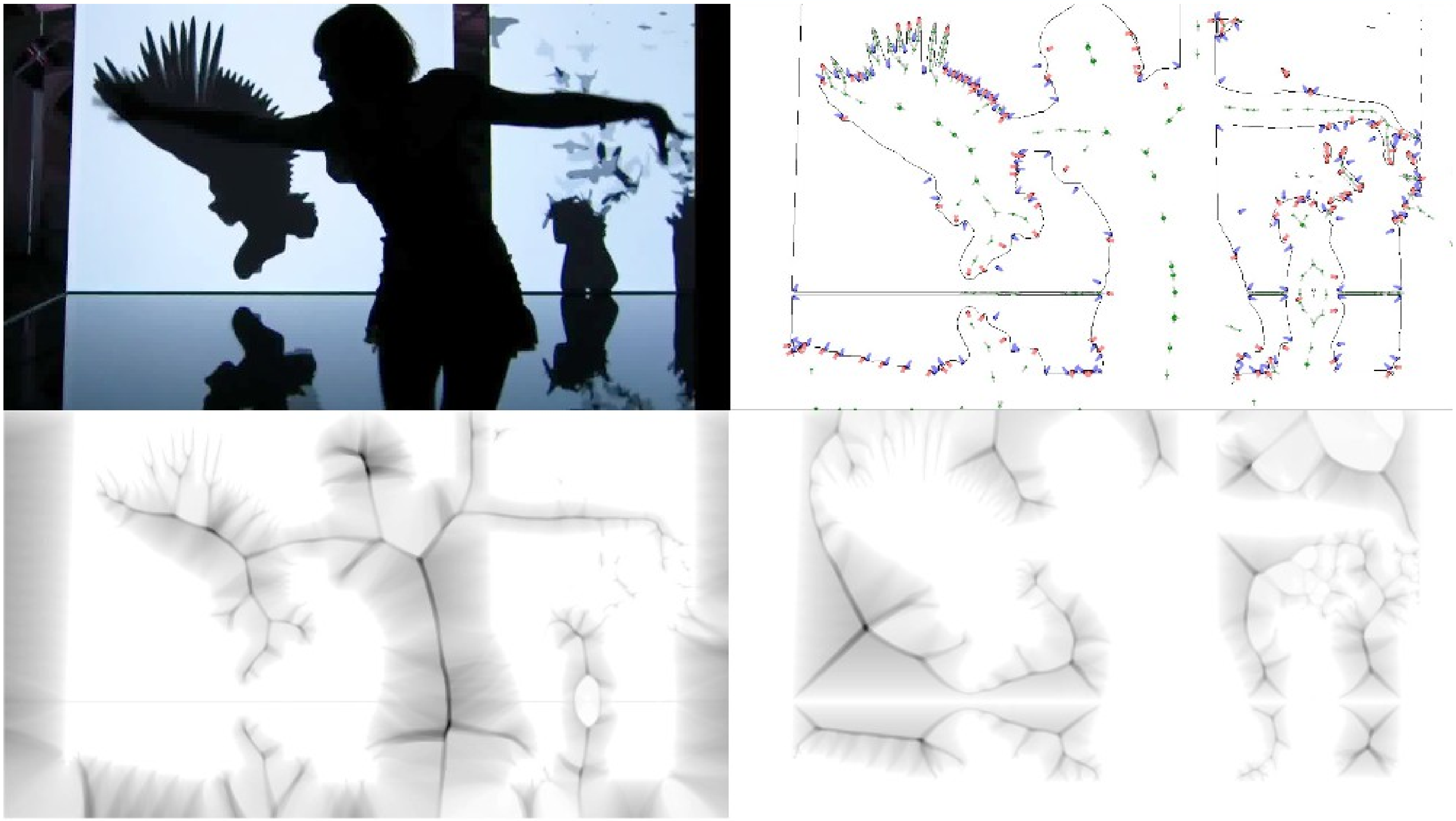


ShIFT Vectors

α_p^1	0	0	1	1	1	1	1	0	0	0	0	1	1	1	1	1	1	0
α_p^2	0	0	1	1	1	1	1	1	0	0	0	1	1	1	1	1	0	0

[with F Fol Leymarie and P Aparajeya]

No more marker MoCap?



[with P Aparajeya, F Fol Leymarie and V Petresin, CMVP 2015]

Overview

Near Duplicate Detection

- Examples
- Principal techniques
- Shape invariant feature vector

Machine Learning + Computer Vision

- **Judging the aesthetics of pictures**
- Food analysis

Media understanding

- “learning from watching TV”

Good photography

Mean: 6.6/10, 180 scores

[with Faria, S Bagley, T Breckon, WIAMIS 2013]

Snapshots

Mean: 2.3/10, 245 scores

Mean: 2.3/10, 279 scores

Research questions

- Is it possible to extract image aesthetics automatically?

- How well can ML methods perform on this task?

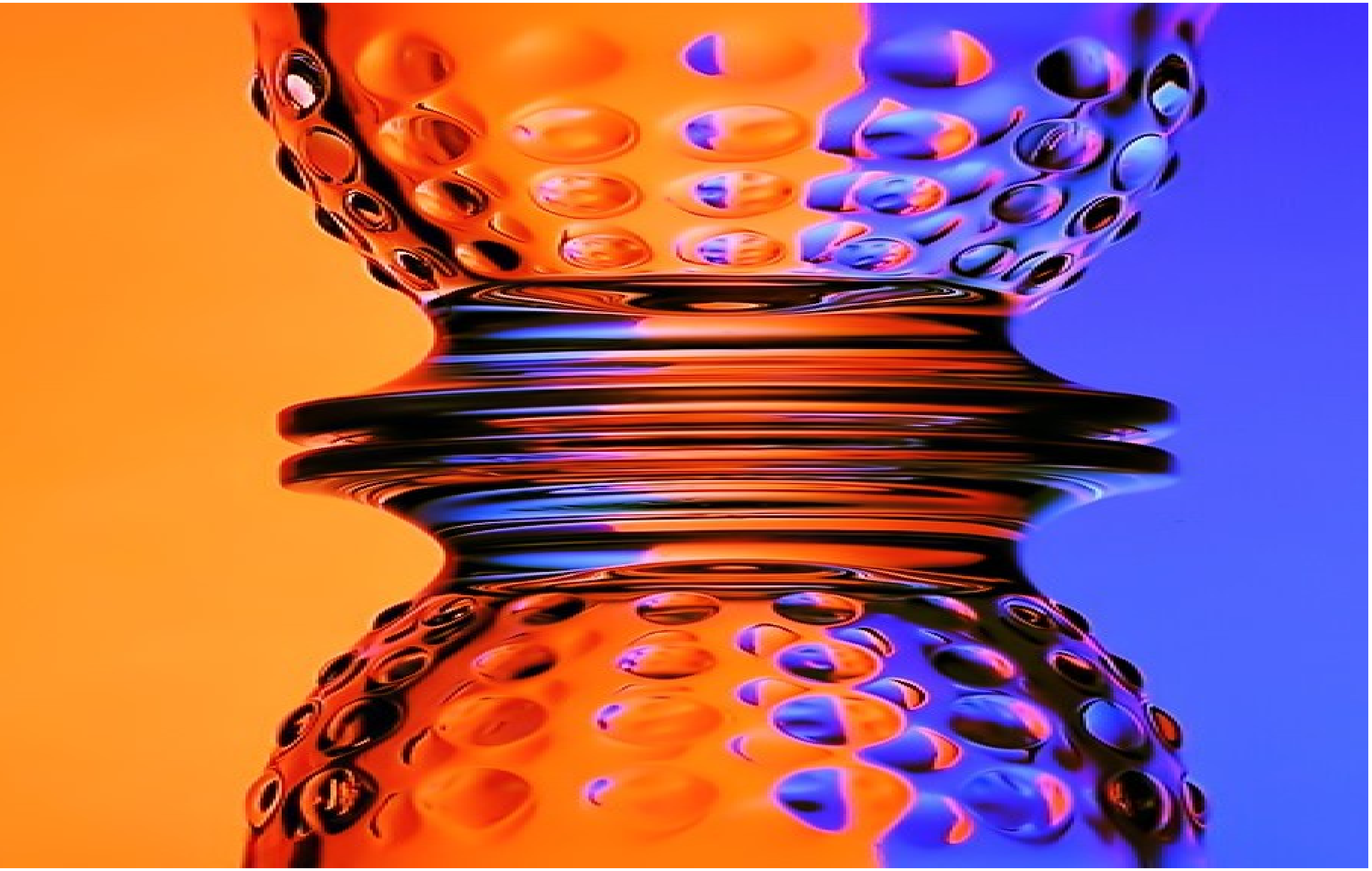
- How can we evaluate the results?

[with Faria, S Bagley, T Breckon, WIAMIS 2013]

Simplicity

Mean: 5.8/10, 265 scores

Complementary colours



Mean: 7.0/10, 304 scores

[with Faria, S Bagley, T Breckon, WIAMIS 2013]

Low contrast

Mean: 5.6/10, 227 scores

[with Faria, S Bagley, T Breckon, WIAMIS 2013]

High contrast

Mean: 6.5/10, 145 scores

[with Faria, S Bagley, T Breckon, WIAMIS 2013]

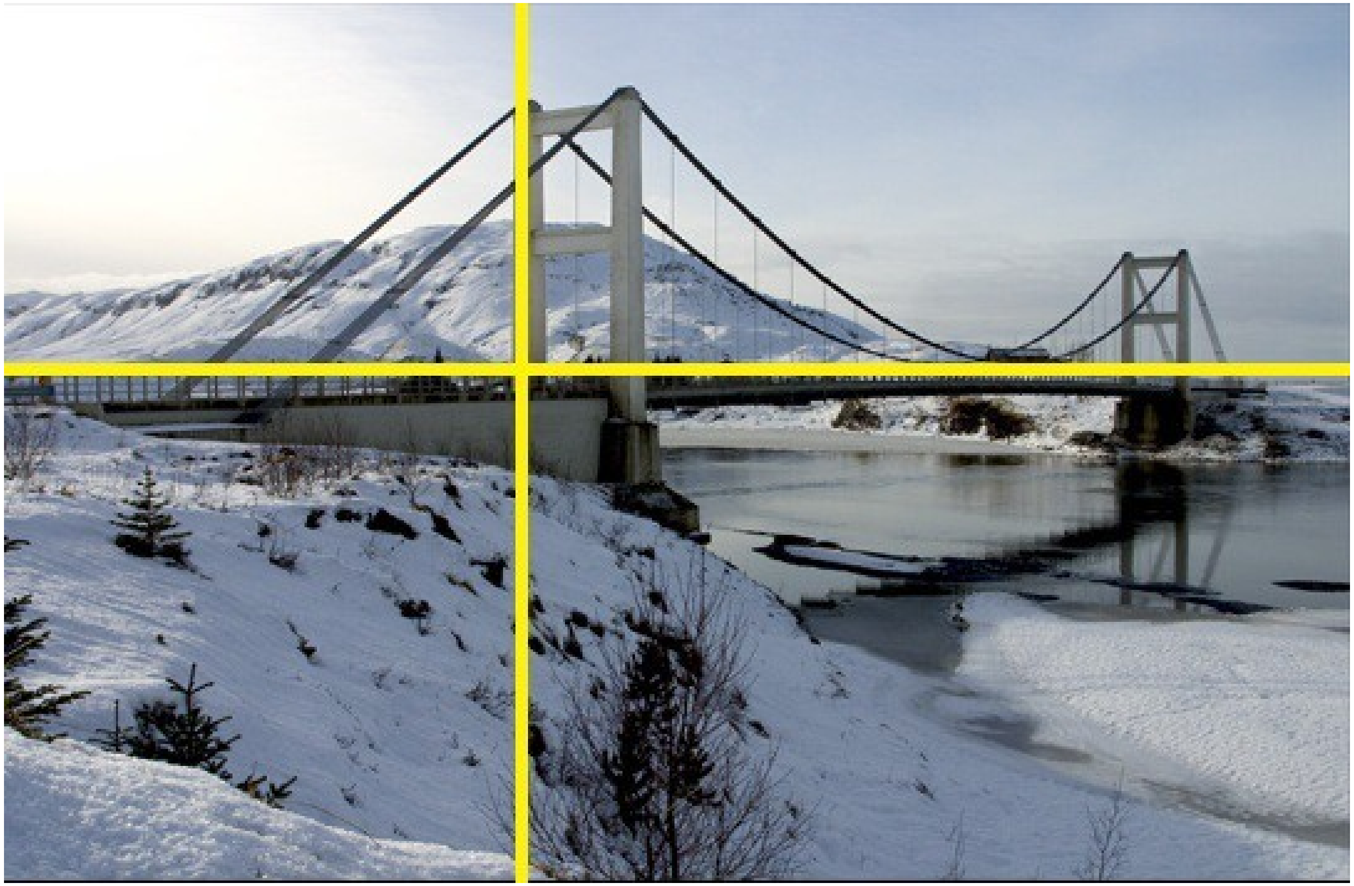
The Open University

The rule of thirds

Mean: 6.6/10, 185 scores

[with Faria, S Bagley, T Breckon, WIAMIS 2013]

Golden ratio



Mean: 6.5/10, 209 scores

[with Faria, S Bagley, T Breckon, WIAMIS 2013]

Diagonals

Mean: 5.5/10, 255 scores

Blur

Mean: 5.6/10, 230 scores

[with Faria, S Bagley, T Breckon, WIAMIS 2013]

How good is this?

[from http://www.gold.ac.uk/pg/msc-music-mind-brain/]

[with Faria, S Bagley, T Breckon, WIAMIS 2013]

Global features

- Normalized area enclosing 75% energy in the Laplacian image
- Mean of the saliency map
- Standard deviation of the saliency map
- Hue count
- Simplicity measure: number of segments after mean shift algorithm
- Luminance histogram width
- Weber contrast
- Michelson contrast
- Average brightness
- Average saturation
- Colour harmony
- Blur measure
- Size ratio
- Compositional balance

[with Faria, S Bagley, T Breckon, WIAMIS 2013]

Local features

- Standard deviation of hue, saturation and luminance in the subject area
- Average saturation in the subject area
- Simplicity: number of background colours
- Hue count in a background area
- Background contrast: Michelson, luminance RMS, hue RMS, saturation RMS

- Subject/background brightness
- Squared difference of the subject/background average values for brightness, hue, and saturation
- Subject/background Weber and Michelson contrast
- RMS of subject/background relation for luminance, hue, and saturation

- Rule of thirds measure
- Subject area measure
- Blur measure for background
- Blur relation between subject and background

[with Faria, S Bagley, T Breckon, WIAMIS 2013]

Datasets

	Photo.net	DPChallenge	CUHKPQ	MIRFlickr	AVA
Images	3,581	16,509	17,613	1,000,000	255,000
Scores p. image	6	>100	6	6	>100
Num. of scores	☑	☑	6	☑	☑
Mean of scores	☑	☑	☑	☑	☑
Distri-bution	6	☑	6	☑	☑
Scale	1 - 7	1 - 10	1 - 10	1 - 10	1 - 10

[with Faria, S Bagley, T Breckon, WIAMIS 2013]

Dataset in our experiments

	Photo.net	DPChallenge	CUHKPQ	MIRFlickr	AVA
Images	3,581	16,509	17,613	1,000,000	255,000
Scores p. image	6	>100	6	6	>100
Num. of scores	☑	☑	6	☑	☑
Mean of scores	☑	☑	☑	☑	☑
Distri-bution	6	☑	6	☑	☑
Scale	1 - 7	1 - 10	1 - 10	1 - 10	1 - 10

[with Faria, S Bagley, T Breckon, WIAMIS 2013]

ML methods

- AdaBoost

  weak decision tree classifier set of 110, weight trim rate 0.98 and maximum depth 3

- SVM

  sigmoid kernel

- Random Forest

  maximum depth value of 32

  n_trees = 100

[with Faria, S Bagley, T Breckon, WIAMIS 2013]

Features selected

- Normalized area enclosing 75% energy in the Laplacian image
- Hue count
- Average brightness
- Average saturation
- Colour harmony
- Standard deviation of hue in the subject area
- Hue count of the background
- RMS of subject/background relation for saturation
- Subject/background Weber contrast
- Subject/background Michelson contrast

[with Faria, S Bagley, T Breckon, WIAMIS 2013]

Recall and precision on fixed values

	Recall when precision = 0.99	Precision when recall = 0.81
Ke et al.	<0.01	0.65
Luo et al.	0.16	0.86
Yeh et al.	0.16	0.79
Current project	0.81	0.99

[with Faria, S Bagley, T Breckon, WIAMIS 2013]

Overview

Near Duplicate Detection

- Examples
- Principal techniques
- Shape invariant feature vector

Machine Learning + Computer Vision

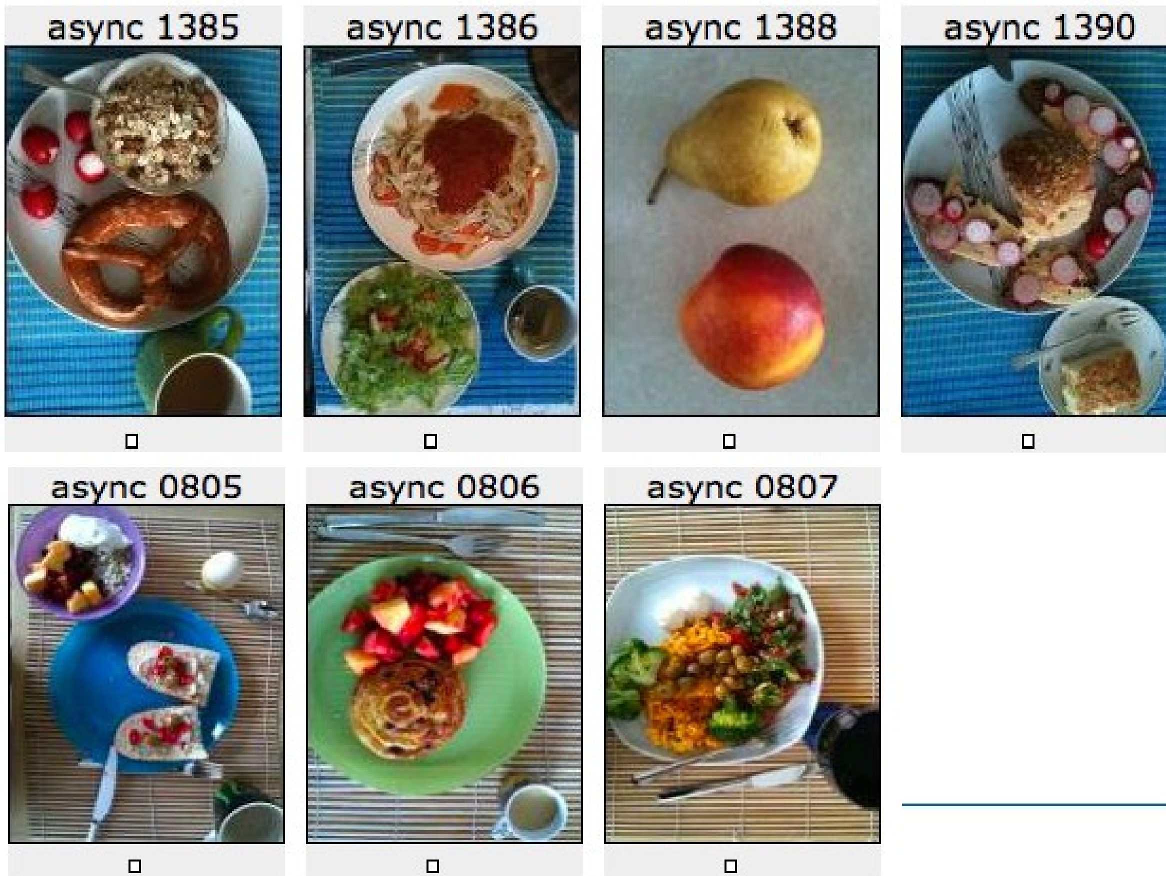
- Judging the aesthetics of pictures
- **Food analysis**

Media understanding

- “learning from watching TV”

ML Example 2: Automated food log

Can only eat what you have photo-logged



Automated food log

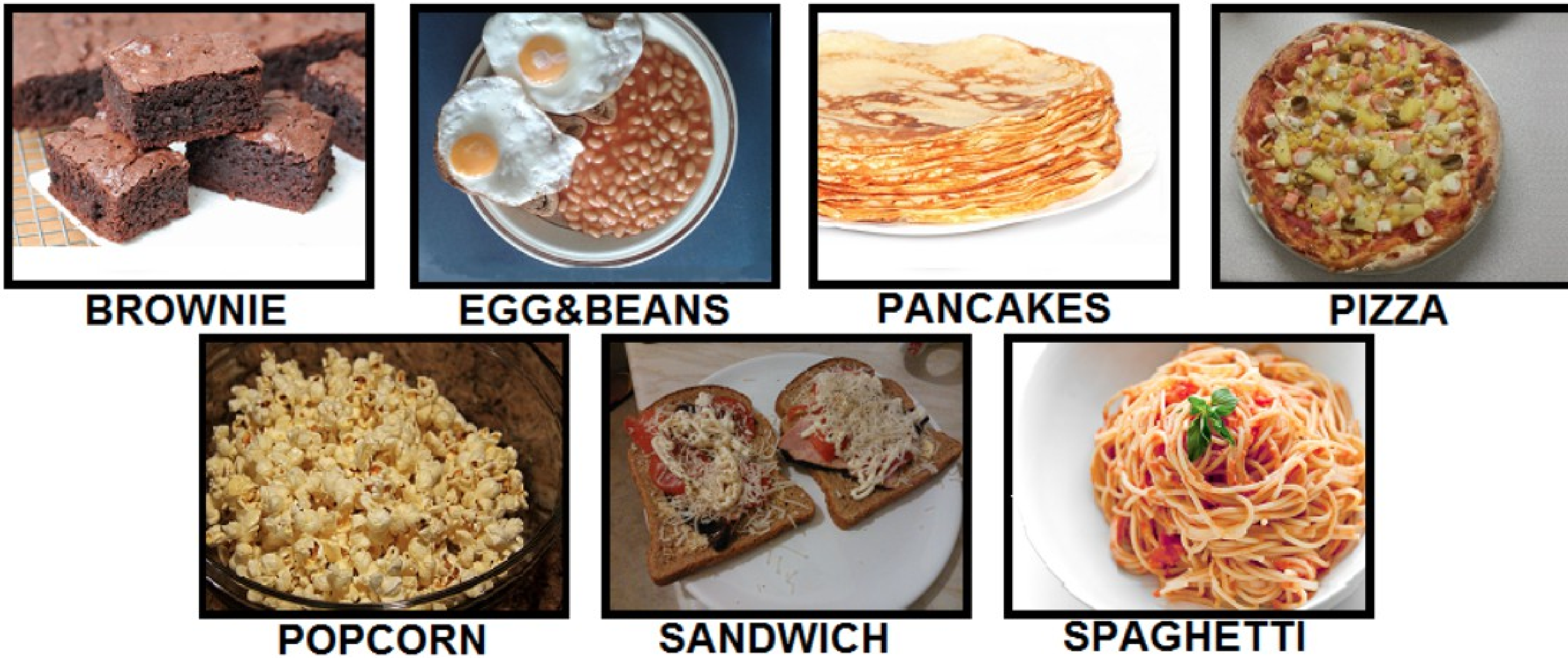


FIGURE 3.3: The 7 classes dish database - in-house.

[with P Walachowska, MSc project 2014]

Automated food log

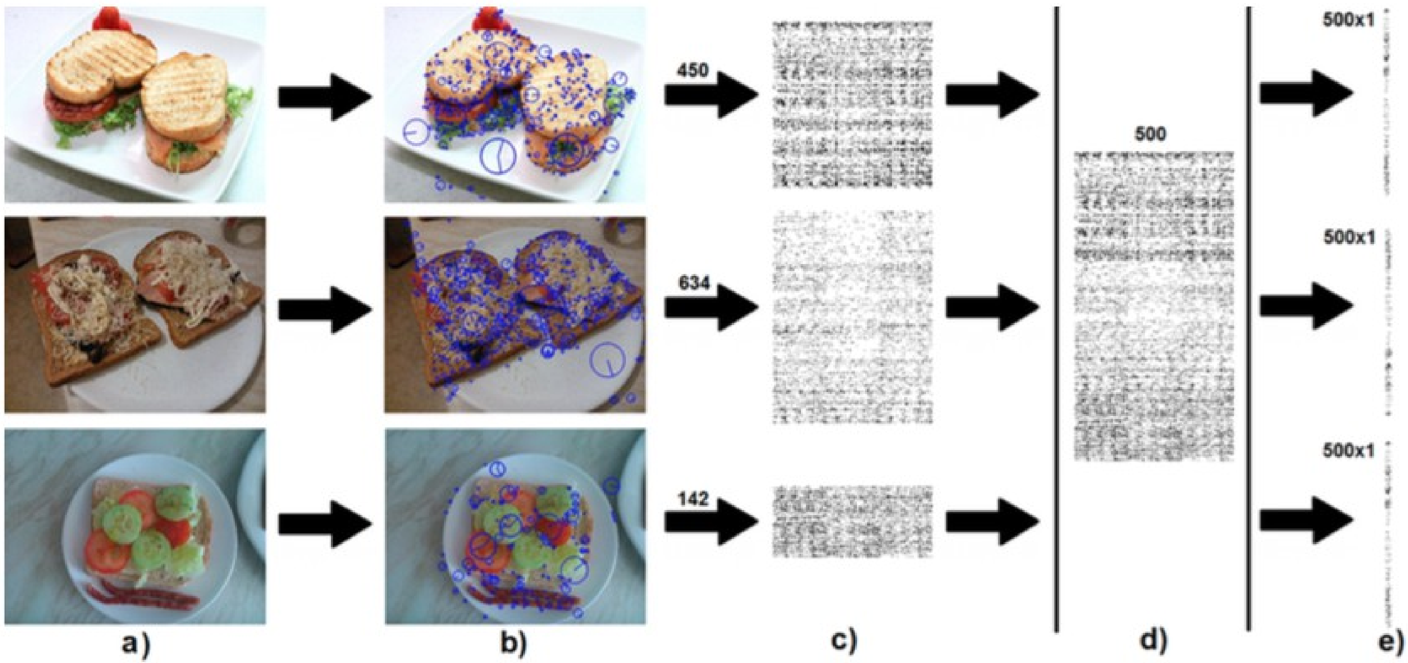


FIGURE 6.7: The example of constructing the bag of features in accordance to the dictionary.

[with P Walachowska, MSc project 2014]

Automated food log

	pizza	popcorn	egg&beans	spaghetti	sandwich	brownie	pancakes
pizza	**9**	0	0	0	0	0	1
popcorn	0	**7**	0	0	0	0	0
egg&beans	**2**	0	**7**	1	0	1	0
spaghetti	1	0	0	**10**	0	0	0
sandwich	**2**	0	1	**2**	**10**	1	0
brownie	0	0	0	0	0	**9**	0
pancakes	0	0	0	0	0	0	**8**

TABLE 8.2: The confusion matrix for in-house dataset. Descriptor: Color SIFT without segmentation.

[with P Walachowska, MSc project 2014]

Automated food log

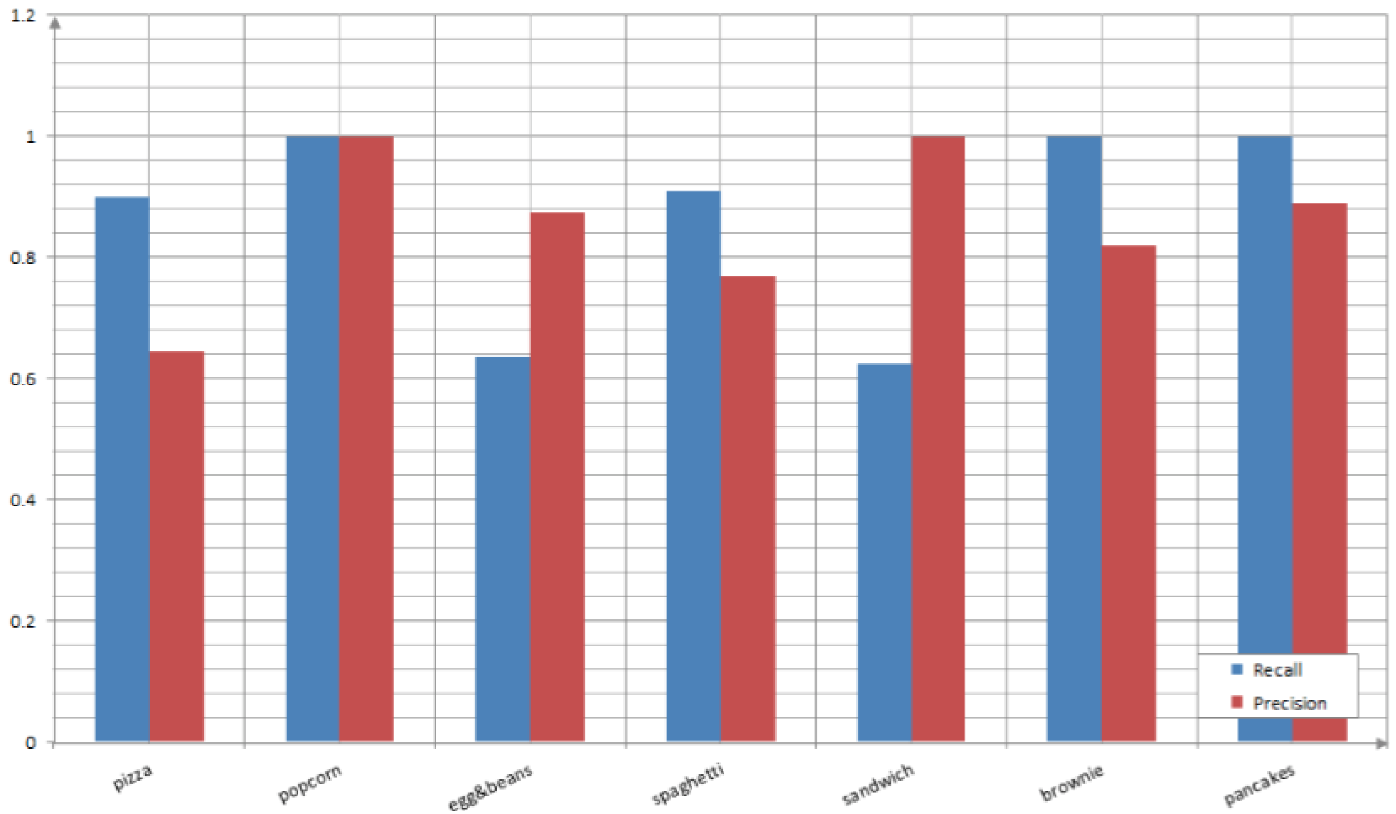


FIGURE 8.8: The graph illustrating the Recall and Precision ratio for results presented in 8.2.

[with P Walachowska, MSc project 2014]

	acerolas	apples	apricots	avocados	bananas	blackberries	blueberries	cantaloupes	cherries	coconuts	figs	grapefruits	grapes	guava	kiwifruit	lemons	limes	mangos	olives	oranges	passionfruit	peaches	pears	pineapples	plums	pomegranates	raspberries	strawberries	tomatoes	watermelons
acerolas	**6**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
apples	0	**1**	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
apricots	0	0	**3**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
avocados	0	0	0	**5**	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0
bananas	0	0	0	0	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**2**	0	0	0	0	0	0	0
blackberries	0	0	0	0	0	**3**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
blueberries	0	0	0	1	0	0	**6**	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
cantaloupes	0	0	0	0	0	0	0	**7**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
cherries	0	0	0	0	0	0	0	0	**7**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
coconuts	0	0	0	0	0	0	0	0	0	**5**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
figs	0	0	0	0	0	0	0	0	0	0	**4**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
grapefruits	0	0	0	0	0	0	0	0	0	0	0	**5**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
grapes	0	0	0	0	0	0	0	0	0	0	0	0	**5**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
guava	0	0	1	0	0	0	0	1	0	0	0	0	1	**6**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
kiwifruit	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**6**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
lemons	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**2**	0	0	0	1	0	0	0	0	0	0	0	0	0	0
limes	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**2**	0	0	1	0	0	0	0	0	0	0	0	0	0
mangos	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	**1**	0	0	0	0	1	0	0	0	0	0	0	0
olives	**2**	**2**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**2**	0	0	0	0	0	0	1	**2**	0	0	0
oranges	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	**4**	0	0	0	0	0	0	1	0	0	0
passionfruit	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**6**	0	0	0	0	1	0	0	0	0
peaches	0	0	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	1	0	0	**2**	0	0	0	0	1	0	1	0
pears	0	0	1	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	**4**	0	0	0	0	0	0	0
pineapples	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**6**	0	0	0	0	0	0
plums	**2**	0	0	0	0	0	**2**	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	**3**	0	0	0	0	0
pomegranates	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**8**	0	0	1	0
raspberries	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**8**	0	0	0
strawberries	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**4**	0	0
tomatoes	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**3**	1
watermelons	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**5**

TABLE 8.4: The confusion matrix for FIDS30 dataset. Descriptor: Color SIFT without segmentation.

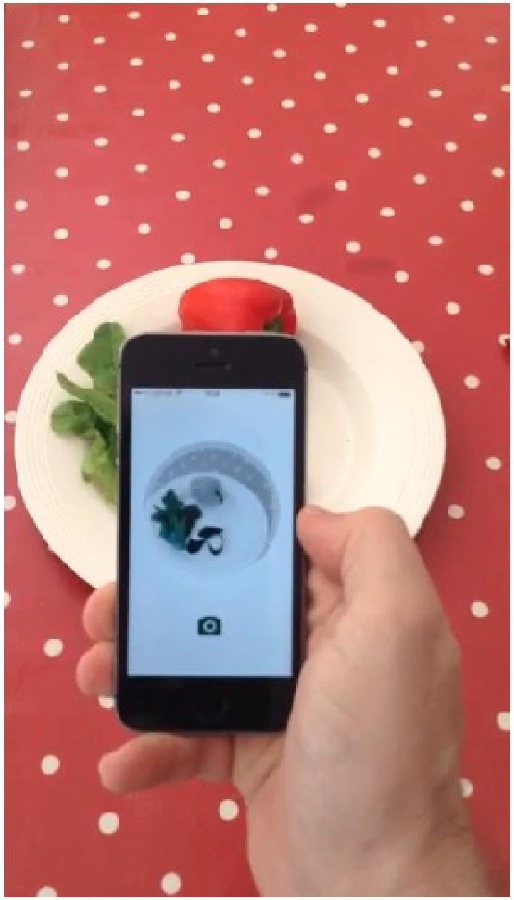
[with P Walachowska, MSc project 2014]

Automated food log

FIGURE 8.14: The samples of the dataset taken from different class which can be confusing for the classifier.

[with P Walachowska, MSc project 2014]

Application



[with mc schraefel and Alex Rogers, Southampton]

Application

[with mc schraefel and Alex Rogers, Southampton]

Overview

Near Duplicate Detection

- Examples
- Principal techniques
- Shape invariant feature vector

Machine Learning + Computer Vision

- Judging the aesthetics of pictures (example)
- Food analysis (example)

Media understanding

- “learning from watching TV”

Visual language?

Abstract concepts (victory, triumph, religion)
Complex concepts (barbecue, car)
Simple material (grass, tarmac, sky)
Part-based models

Visual language?

Salient keypoints / visterms

KNOWLEDGE MEDIA KMi INSTITUTE

Crystal ball question

What can computers learn from watching 20,000 hours TV news?

Play this story

Organisations: Abu Ghraib, Abu Mussab, Baghdad Airport, Daily Telegraph, Eugene Armstrong, Foreign Office, Iraqi National Guard, The embassy

People: Al-Jazeera, Al-Qaeda, Al-Zarqawi, Jack Hensley, Jack Straw, Ken Bigley, Mr Big, Mr Blair, Nicholas Witchell, President Bush, Sir Davidmaning, Tony Blair

Locations: Baghdad, Basra, Britain, Europe, Iraq, Northern Ireland, US, Washington

Dates: 48, today

Date : Sat Sep 18 2004
Length : 399.46 seconds
Full Story : Link
Summary : *uk Kid snappers release pictures of British and American hostages kidnapped in Iraq.* The Arabic television station Al-Jazeera has shown images of ten hostages, it is reported they will be killed if their company doesn't leave Iraq within three days. Opposition parties accused the Prime Minister of misleading the public after leaked documents suggested he was warned two years ago of the potential for pOgs war problems in Iraq. In a memo seen by the Daily Telegraph the Daily Telegraph the Foreign Secretary, Jack Straw, said there was no clear post-conflict plan. They didn't put in place a pronner plan. It is game set and match with no sense of pleasure to Despite the continuing security problems across Iraq, the Prime Minister says the US-led coalition was prepared for life after Saddam. Indeed, we have unfolded tht plan, but there are people in Iraq, outsiders, as well as former regime elements who are determinedto to stop us.

Visual mining - interpreting image data

Professor Stefan Rüger

Multimedia and Information Systems
Knowledge Media Institute
The Open University

http://kmi.open.ac.uk/mmis