More than a 1000 words: multimedia information retrieval

Professor Stefan Rüger
Multimedia and Information Systems
Knowledge Media Media Institute
The Open University
http://kmi.open.ac.uk/mmis
Projects

Spotlighted Hot Projects

NoTube
Future Internet  Semantic Web and Knowledge Services
Networks and Ontologies for the Transformation and Unification of Broadcasting and the Internet

The UK Multimedia Knowledge Management Network
Multimedia and Information Systems
Enhance communication between the experts in both academia and industry
Since 1995: 117 projects & 67 technologies

Current year

17 live projects
typically £2.5m ext, £1m internal

• 10 EU
• 3 UK
• 1 US
• 3 internal (iTunes U, SocialLearn, radar)
Find near-duplicate images
or find visually similar images
or find semantically similar images
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
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
This brilliant Big Art Project in the form of the Open University's letters is the result of OpenUnlimited activities in the Church on Saturday 27th June 2009.

The theme for the project was "What does the OU mean to you?". It was very popular as children and adults expressed themselves freely, by sticking things, adding glitter or drawing with pens and paints to decorate the letters, and as you can see, they did a fantastic job.

Estates Fabrics Team constructed these massive letters from 8ft sheets of plywood, added a matt coat of colour and then once our
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, ...
Near-duplicate detection: Applications

Tourism
Museum guide (“What is this?”)
Linking learning objects (slide and video, figures, ...)
Plagiarism detection
Word spotting for handwritten documents
“What is this?” (parts of an engine, ...)
Shopping (where does this [take photo] sell for less?)
Shazam: name that song
Near-duplicate detection: How does it work?

**Fingerprinting technique**

1. Compute salient points
2. Extract “characteristics” from vicinity (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

All keypoint features of all images in collection

Feature space

Nine Geese Are Running Under A Wharf And Here I Am

Millions of "visterms"
Clustering
Hierarchical k-means

[Nister and Stewenius, CVPR 2006]
Quantisation through hashing

\[ h^i: \mathbb{R}^d \rightarrow \mathbb{Z} \]
\[ v \mapsto h^i(v) = \left\lfloor \frac{a^i v + b^i}{w} \right\rfloor \]

- \( a^i \in \mathbb{R}^d \) is a random Gaussian-distributed vector.
- \( w \in \mathbb{R}^+ \) is a constant.
- \( b^i \in [0, w) \) is a random number.

\[ h(v) = (h^1(v), h^2(v), \ldots, h^k(v)) \] is the LSH hash vector.
Quantisation
LSH hashes

\[ x \cdot h(v) = (-1, 0) \]
NDD: Encode all images with visterms

Jkjh Geese Bjlkj Wharf Ojkkjhhj Kssn Klkekjl Here Lkjkll Wjjkl Kkjlk Bnm Klkgjg Lwoe Boerm ...
At query time compute salient points, keypoint features and visterms

Query against database of images represented as bag of visterms
NDD: Check spatial constraints

[with Suzanne Little, SocialLearn project]
Fingerprinting technique

1. Compute salient points
2. Extract “characteristics” from vicinity (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 \((t_1, id)\)
Entropy considerations

Specificity: Encoding \((f_1, f_2, t_2 - t_1)\) to use 30 bit
Redundancy
Probability of matching

![Graph showing the probability of at least one of n=30 surviving vs. the probability of one correctly classified.](image-url)
Shazam

Rueger, Multimedia IR, 2010 explains it all! Buy it now 😊
Near-duplicate detection
Where are the challenges?

3d at scale
Glare
Diversity
Segmentation

[Victoria and Albert museum, London, ceramics collection, 2010]
Leaf detection

10,000s of species
1m and more specimens
High inter-class similarity
High intra-class variability

No longer NDD!
Machine Learning at its finest
Feature representation vital
Eminently difficult

[with Natural History Museum, London, and Goldsmiths]
Shock graph attractors: Deviation from symmetry

[with Frederic Fol Leymarie, Goldsmiths, 2011]
Find near-duplicate images 
or **find visually similar images** 
or find semantically similar images

More than a 1000 words...

Image retrieval

[illustration from Suzanne Little, SocialLearn project]
Similarity:
Features and distances
Which features?
The semantic gap

1m pixels with a spatial colour distribution

faces & vase-like object

victory, triumph, ...

[http://www.sierraexpressmedia.com/archives/11148]
Polysemy
Fig. 17A

103d

light
Where are the challenges?

Image content analysis (diversity, semantic gap, polysemy)
Mapping to higher level (semantic representation)
Time taken and resources needed ("scalability")
Automation, scale and coping with errors
Find near-duplicate images
or find visually similar images
or **find semantically similar images**
From visual to semantic similarity

\{\text{Snow, ice, bear, grass, ...} \} \quad \leftrightarrow \quad \{\text{City, water, building, sky, ...} \}
Example: grass classifier

- very likely
- may be
- probably not
Automated annotation as machine translation

water  grass  trees

the beautiful sun

le soleil beau
Modelling semantic concepts

From training data

- outdoor
- town
- crowd
- sky
- grass
- tarmac
- faces
Automated annotation as machine learning

Probabilistic models:
- maximum entropy models
- models for joint and conditional probabilities
- evidence combination with Support Vector Machines

[with Magalhães, SIGIR 2005]
[with Yavlinsky and Schofield, CIVR 2005]
[with Yavlinsky, Heesch and Pickering: ICASSP May 2004]
[with Yavlinsky et al CIVR 2005]
[with Yavlinsky SPIE 2007]
[with Magalhães CIVR 2007, best paper]
A simple Bayesian classifier

\[
P(w|I) = \frac{P(w, I)}{P(I)} = \frac{\sum_J P(w, I|J)P(J)}{\sum_J P(I|J)P(J)}
\]

\[
= \frac{\sum_J P(I|w, J)P(w|J)P(J)}{\sum_J \sum_w P(I|w, J)P(w|J)P(J)}
\]

Use training data \( J \) and annotations \( w \)

\( P(w|I) \) is probability of word \( w \) given unseen image \( I \)

The model is an empirical distribution \( (w, J) \)
Automated annotation

[with Yavlinsky et al CIVR 2005]
[with Yavlinsky SPIE 2007]
[with Magalhaes CIVR 2007, best paper]

Automated: water buildings city sunset aerial

[Corel Gallery 380,000]
The good door

[beholdsearch.com, 19.07.2007, now behold.cc (Yavlinksy)]
[images: Flickr creative commons]
The bad wave

[beholdsearch.com, 19.07.2007, now behold.cc (Yavlinksy)]
[images: Flickr creative commons]
The ugly
Machine Learning is hard, and data may not help.
Automated annotation: State of the art 2008

- Effectiveness
  - (Yavlinsky, Schofield, Rüger 2005)
  - (Feng, Lavrenko, Manmatha 2004)
  - (Carneiro, Vasconcelos 2005)
  - (Magalhaes, Rüger 2007)

- Flexibility
  - (Blei, Jordan, 2003)
  - (Lavrenko, Manmatha, Jeon, 2003)
  - (Jeon, Lavrenko, Manmatha, 2003)

- Efficiency
  - (Amir, Argillander et al 2005)
  - (Snoek, Worrying, et al. 2006)
  - (Snoek, Gemert, et al. 2006)
Handling annotation errors

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

Visual language?

Salient keypoints / visterms
Horizontal vs vertical

IR, CV, ML, NLP, IE, Clustering, Systems, ...
Eyeballing
Roping the user in
Stanley Milgram experiment 1967

Link randomly selected people from Wichita, Kansas, and Ohama, Nebraska to a stockbroker in Boston
Connect similar objects

Focus

Structure

[with Heesch ECIR 2004, CIVR 2005]
[PhD thesis Heesch, nominated for best UK thesis 2005]

Colour

Text
Browsing example

TRECVID topic 102: Find shots from behind the pitcher in a baseball game as he throws a ball that the batter swings at

In your mind
[dataset TRECVID 2003]
Navigate 100,000's of products in seconds...

Visual Search allows you to find the products you can't describe...

Search

Visit Store

Loriblu
Multicolor Leather Trim
Denim Wedge Sandal Shoes
$342
Ambition (future work)

Horizontal research: automated media understanding
Ambition (future work)

Vertical research: specific applications
More than a 1000 words:
multimedia information retrieval

Professor Stefan Rüger
Multimedia and Information Systems
Knowledge Media Media Institute
The Open University
http://kmi.open.ac.uk/mmis