A proactive interface for image retrieval

Master's thesis presentation

Student: László Kozma
Supervisor: Professor Samuel Kaski
Instructor: Phd. Arto Klami

28 May 2009

Contents

Introduction

Image retrieval

- Finding relevant images in a large collection
- Increasing number of applications: photo collections, publishing/advertising, industrial quality control, surveillance, etc.
- Common approach: represent images as vectors: $X_i = [x_1 x_2 \ldots x_n]$
- Assumption is that the relevant images are located in a cluster, along a manifold, etc.
- Most systems in wide use today rely on textual meta-data
- Examples: Google Images, Flickr Search

Introduction

Image representations

"St. Francis preaching to the birds", Giotto (~1297)

"man", "birds", "tree", "sky", "hill"

???

MPEG-7, SIFTcolor

simple features of color, texture

RGB pixel values of digital image

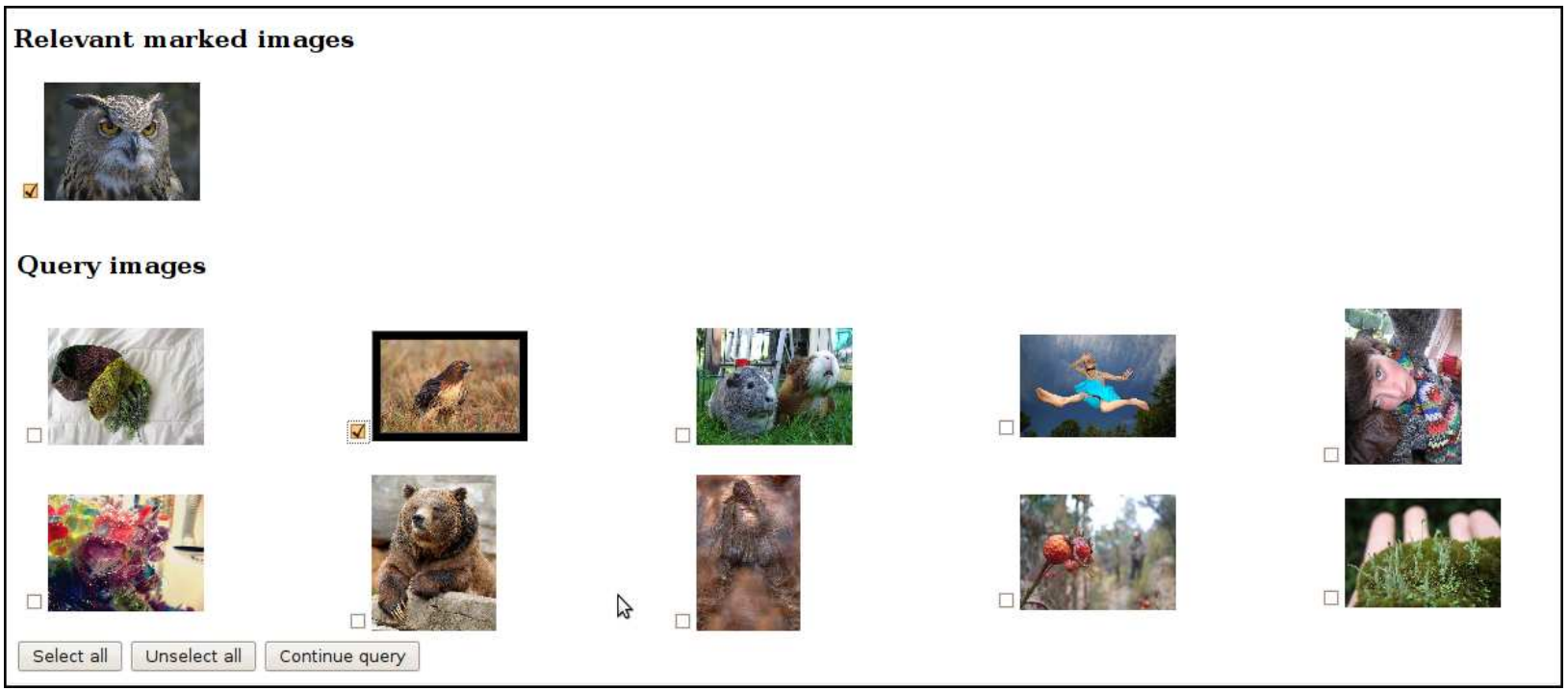
Introduction

Content-based image retrieval (CBIR)

- User can not provide the features directly
- Methods using query-by-example have been developed
- A typical CBIR-interface:
  - System shows a small set of example images to the user
  - User gives feedback on images
  - System updates its state using feedback
  - Repeat ...
- Most CBIR research focuses on better features and better retrieval algorithms
- In this work we take the CBIR engine for granted

Introduction

PicSOM: a content-based image retrieval engine (Laaksonen, Koskela, et al., 2002)



Relevance feedback

- Most systems rely on explicit feedback from user
- Recently, growing interest in using other sources:
  - History of the user-interaction
  - Activity of other users
  - Context
  - Implicit relevance cues:
    - Mouse movement, scrolling pattern
    - **Eye movements**
    - Brain activity
    - Heart rate
    - etc.
- Is there sufficient information in such sources to replace or complement explicit feedback ?

Contributions of this thesis

- Study of the feasibility of using eye movements as a source of implicit relevance feedback in image retrieval
- Design and implementation of an interface for retrieving images using eye movements, based on an existing CBIR engine (PicSOM)

Image navigator

Interface

- Design goals:
  - Integrate with an existing CBIR engine
  - Connect with an eye tracking device to collect gaze data
  - Real-time estimation of relevance and retrieval of new images
  - Allow user to navigate and backtrack if necessary
  - Display images such as to facilitate natural viewing patterns
- Several alternative approaches tried out
- The final interface is inspired by ideas from various zooming user interfaces

Image navigator

Interface

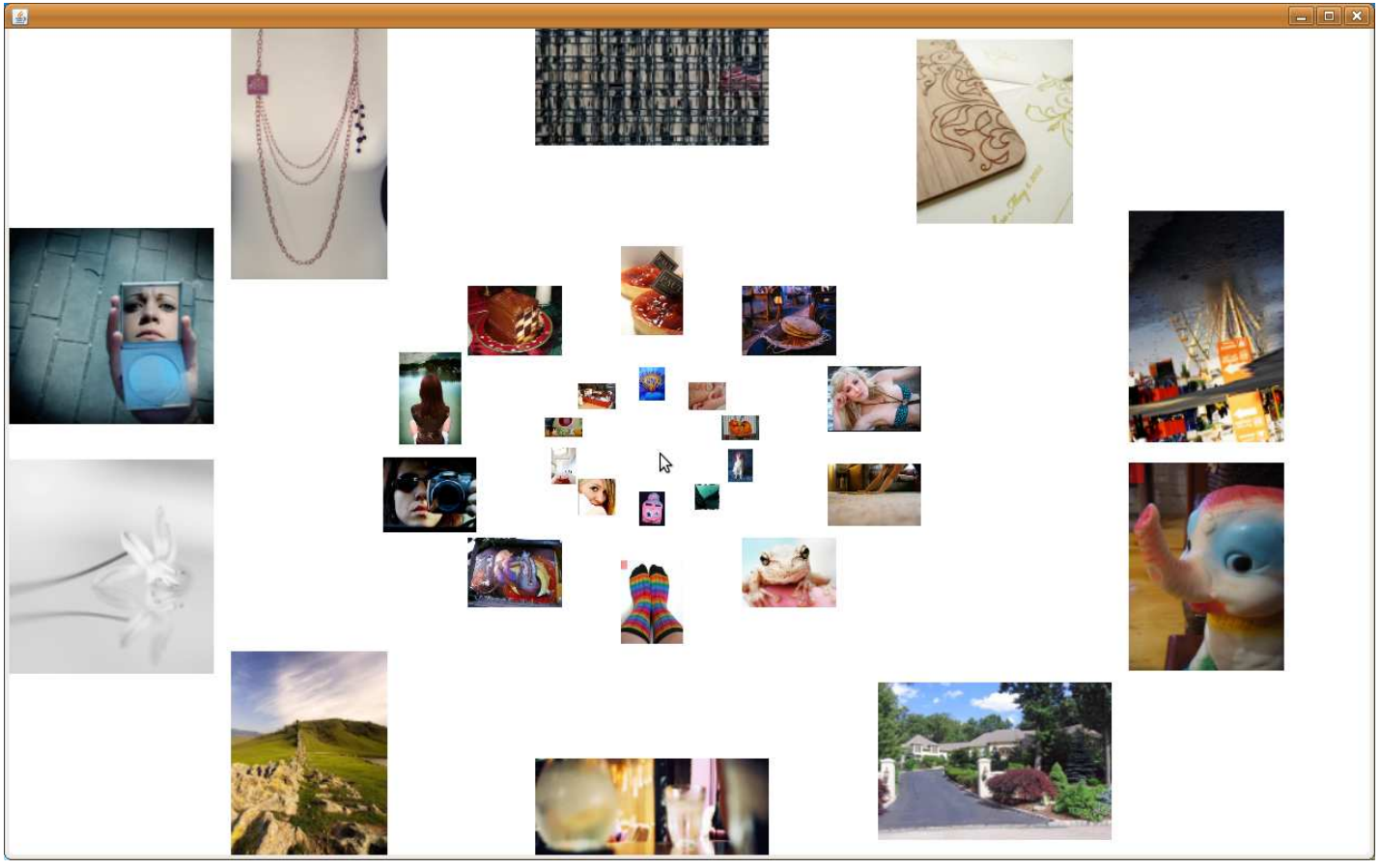


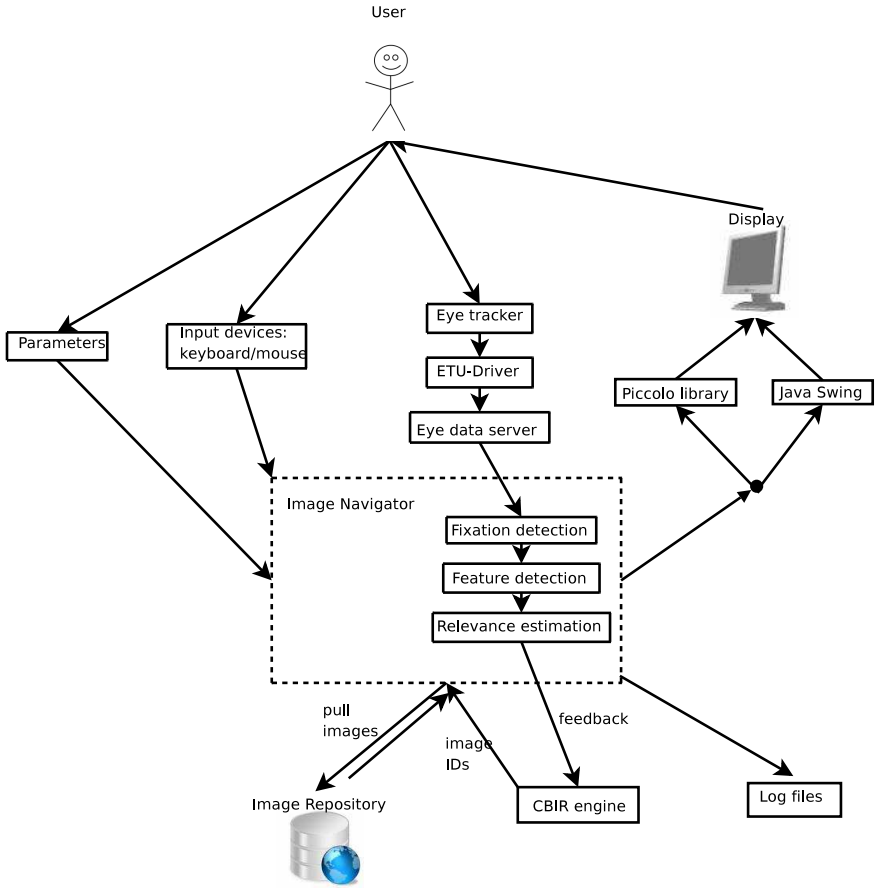
Image navigator

Interface

- Features
  - Circular layout
  - Selection methods
  - Zooming methods
  - Indicating relevance
- (demo)
- Implementation:
  1600 lines of Java code using Piccolo library for zooming interfaces

Image navigator

Implementation



Relevance prediction

Methodology

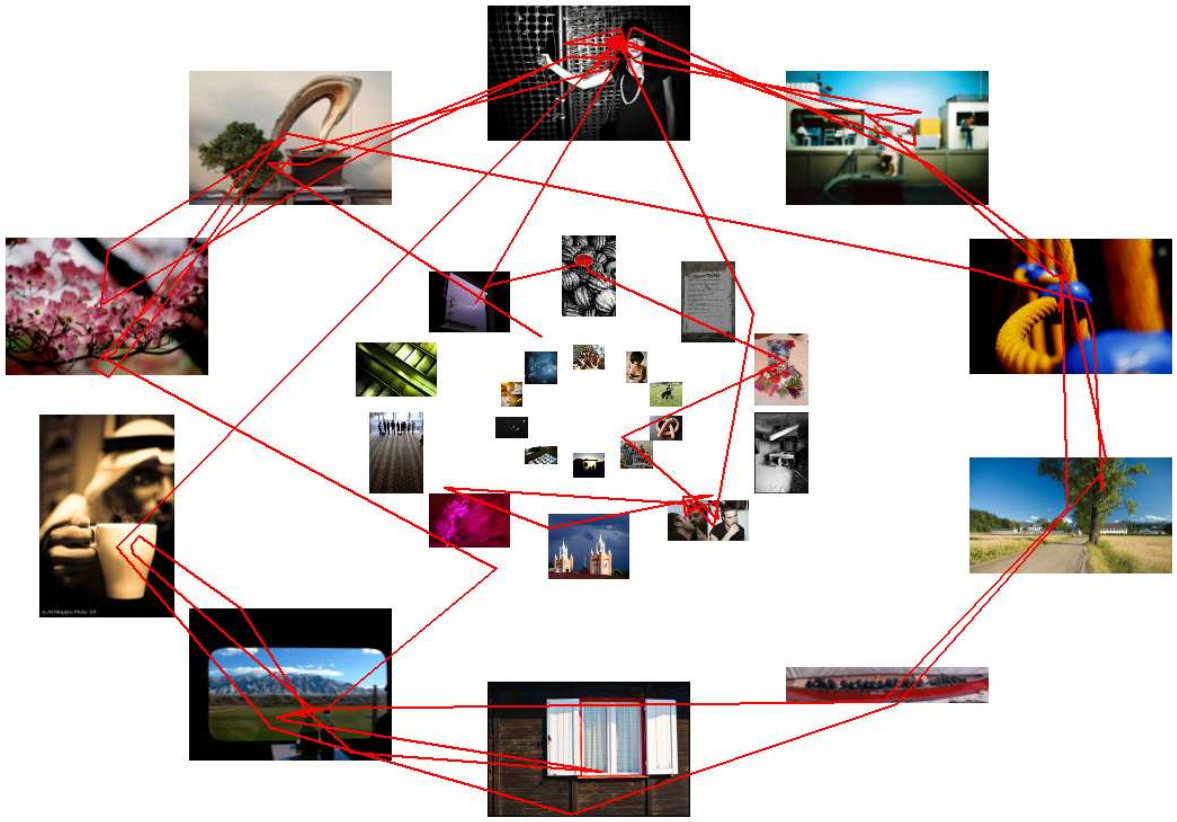
- We collect gaze data from test subjects using the system
- A set of complex features is extracted from data
- A predictor of relevance is learned from the collected data
- The predictor is implemented and used in the online system
- Further data is collected for evaluation

Relevance prediction

Experimental set-up

Relevance prediction

Typical gaze patterns



Relevance prediction

Features

Nr.	Name	Description
1	FirstVisit	Time passed between image displayed and first visit
2	MeanLength	Mean length of fixations
3	FixTimeSpread	Standard deviation of fixation occurance times
4	SumLength	Total length of fixations
5	MaxContView1	Max. continuous viewing time without viewing other images
6	MeanContView	Mean length of continuous viewing sessions of image
7	MaxContView2	Max. cont. viewing time without fixating at empty space or other images
8	RatioTotal	The ratio of viewing times over the image and over all other images
9	RatioRing	The ratio of viewing time over the image and over all images in same ring
10	MeanSaccLength	Mean Length of saccade before fixation
11	MeanPrevImage	Proportion of times when previous fixation also over this image
12	MeanPrevEmpty	Proportion of times when previous fixation over empty space
13	MeanPrevRing	Proportion of times when previous fixation over the same ring
14	FirstVisitIndex	How many images viewed on this ring before the first fixation on image
15	RevisitCount	How many times image re-visited in total
16	PrevDist	Average distance from previously viewed image on the same ring

Relevance prediction

- We use classical logistic regression
- The feature vector of ith image is denoted by $\mathbf{x}_i$
- The model estimates the probability of that image being relevant as:

$$p(r_i|\mathbf{x}_i) = \frac{1}{1 + \exp(-\mathbf{w}^T\mathbf{x}_i - \alpha)}$$

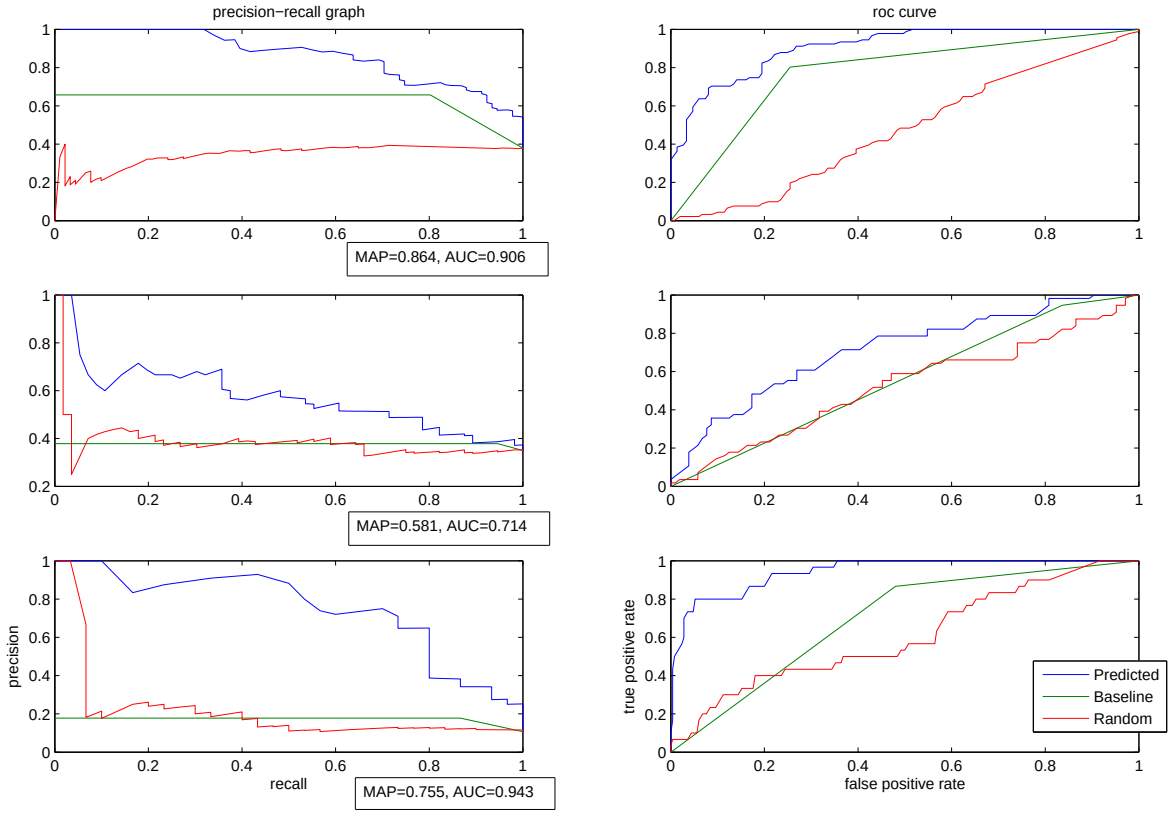
- $\mathbf{w}$ is a projection vector that weights the different features
- α is a bias term
- $\mathbf{w}$ and α are learned to maximize the likelihood of the true relevance in labeled training data
- The relevance predictions are obtained by thresholding the probabilities: all images with probability over t are deemed relevant.

Evaluation

- How well the predictor generalizes on the collected data:
  - Within data from one user
  - Between different users
- How stable are the results (using high number of random split-ups of the data)
- Interpret the model
  - What are the obtained weights
  - How consistent are the weights between different runs
  - What are some false/true positives/negatives
- How does the choice of threshold affect the performance of the predictor
- How is the performance of PicSOM affected by the noise in the predictions
- What is the retrieval accuracy if the predictions are used in a real retrieval scenario instead of explicit feedback

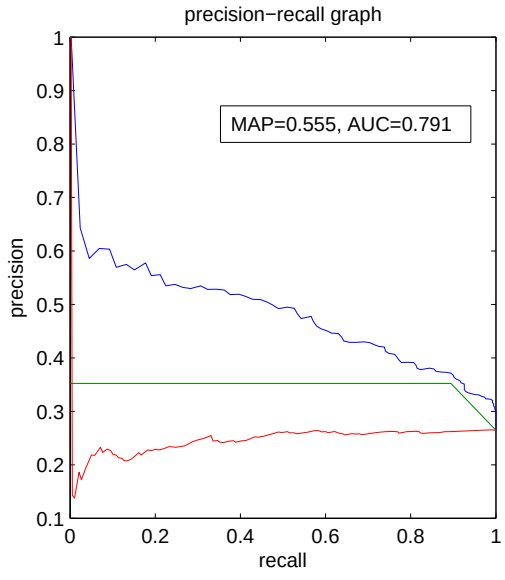
Evaluation

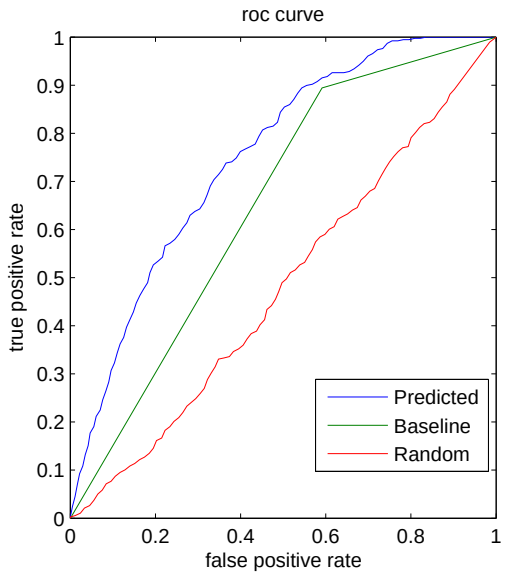
Within-subject data

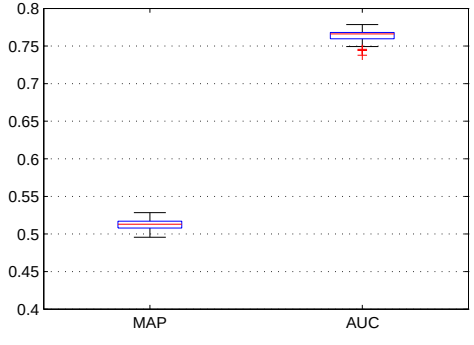


Evaluation

Different divisions of data from all subjects

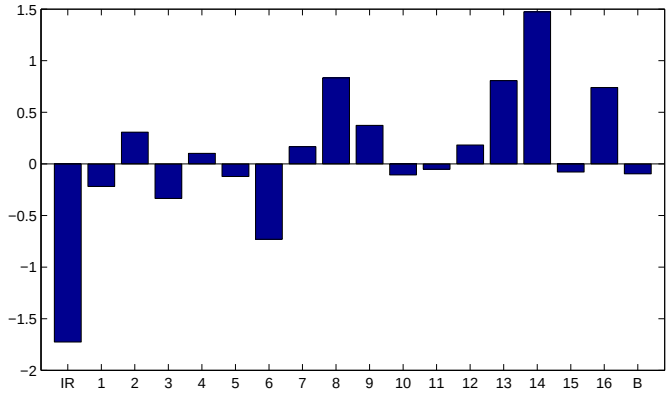


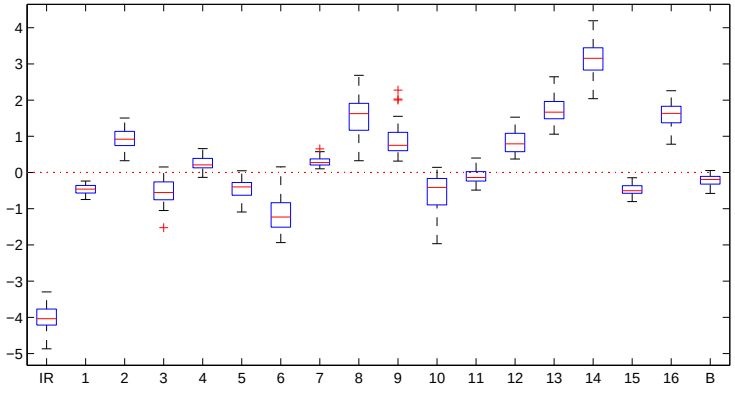




Evaluation

Model weights





Evaluation

Sample images

0.09

0.50

0.68

0.08

0.50

0.627

Evaluation

Choice of threshold

Table: Prediction results with different threshold values.

threshold	TP	TN	FP	FN	pn	np
0.3	232	767	285	146	39%	27%
0.4	175	882	170	203	53%	16%
0.5	115	951	101	263	70%	10%
0.6	71	998	54	307	81%	5%
0.7	41	1020	32	337	89%	3%

Evaluation

Error-sensitivity of PicSOM

Table: MAP of PicSOM while retrieving 400 images with true class “people”.

np \ pn	0	0.1	0.2	0.3	0.4	0.5	0.6
0	0.351	0.364	0.343	0.326	0.334	0.319	0.328
0.1	0.343	0.335	0.344	0.307	0.327	0.319	0.294
0.2	0.352	0.333	0.327	0.308	0.304	0.291	0.279
0.3	0.341	0.336	0.323	0.261	0.306	0.256	0.246
0.4	0.296	0.318	0.308	0.301	0.294	0.202	0.248
0.5	0.302	0.344	0.331	0.307	0.269	0.253	0.243
0.6	0.329	0.308	0.298	0.272	0.228	0.245	0.214

Evaluation

Retrieval accuracy

Table: Retrieval accuracy of CBIR with different feedback methods.

	clouds	dog	people	flower	sunset	animals
Explicit	0.532	0.014	0.422	0.126	0.407	0.119
Random	0.107	0.015	0.206	0.011	0.020	0.012
Predicted	0.245	0.087	0.401	0.022	0.041	0.233

Conclusions

- Image navigator
  - A new type of interface for content-based image retrieval
  - Suitable for usage in retrieval with iterative query refinement using relevance feedback
  - Automates the collection of relevance feedback by using eye tracking
  - Eye movement measurements are fed into a machine learning method predicting the relevances, and the predictions are given for the engine
- Study of feasibility of using eye movements for relevance prediction
  - There is information about relevance in eye movements
  - In short term more realistic to complement explicit feedback than to completely replace
  - Better integration with image retrieval process can lead to improvements

Conclusions

Possible improvements

- Implementation
- Interface
- Experimental set-up
- Machine learning

Questions / comments ...