THE UNIVERSITY OF TOKYO

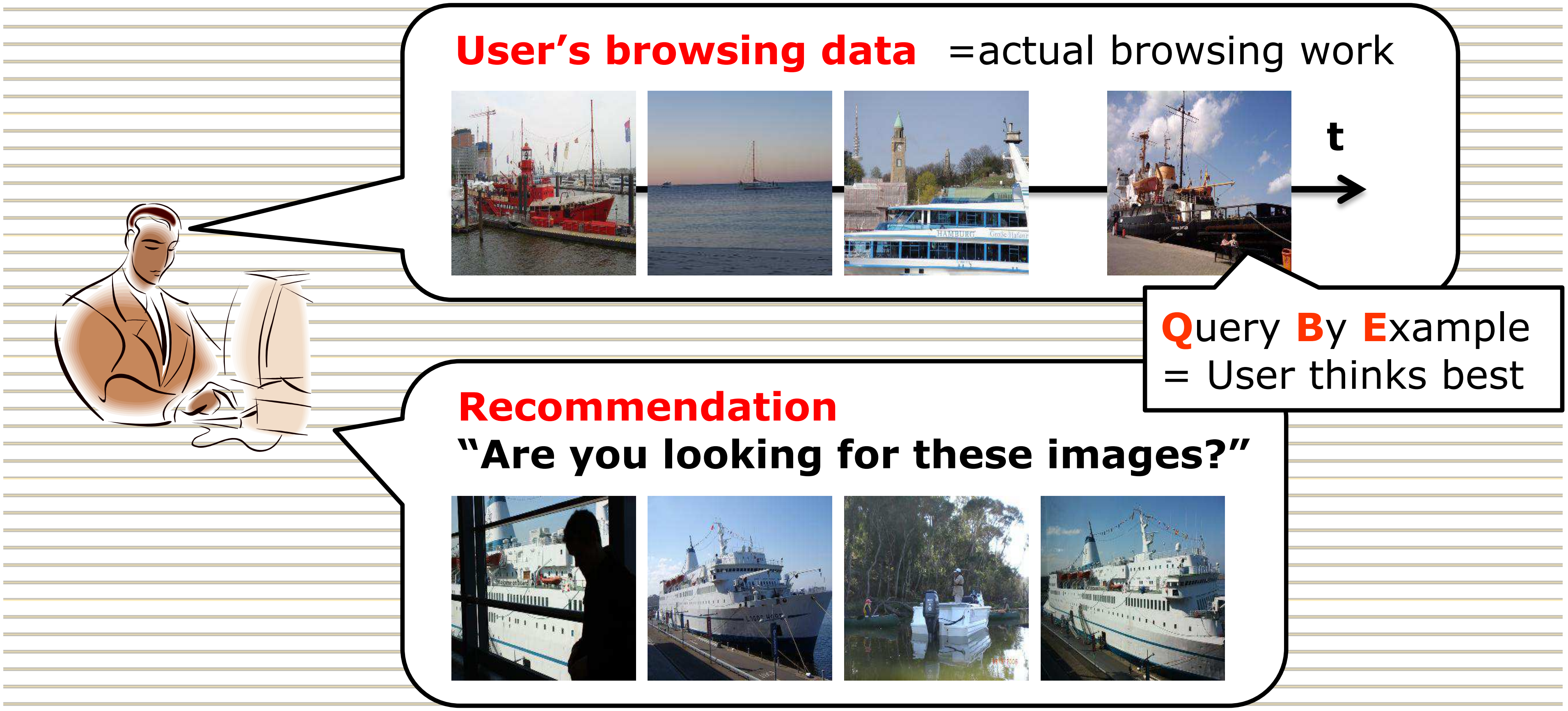
MIL at ImageCLEF 2013 Personal Photo Retrieval

Masaru Mizuochi, Takayuki Higuchi,
Chie Kamada, and Tatsuya Harada

Machine Intelligence Laboratory, The University of Tokyo

Subtask2: Personal Photo Retrieval

The system which can help users to retrieve images from a lot of personal photo collections using browsing data.



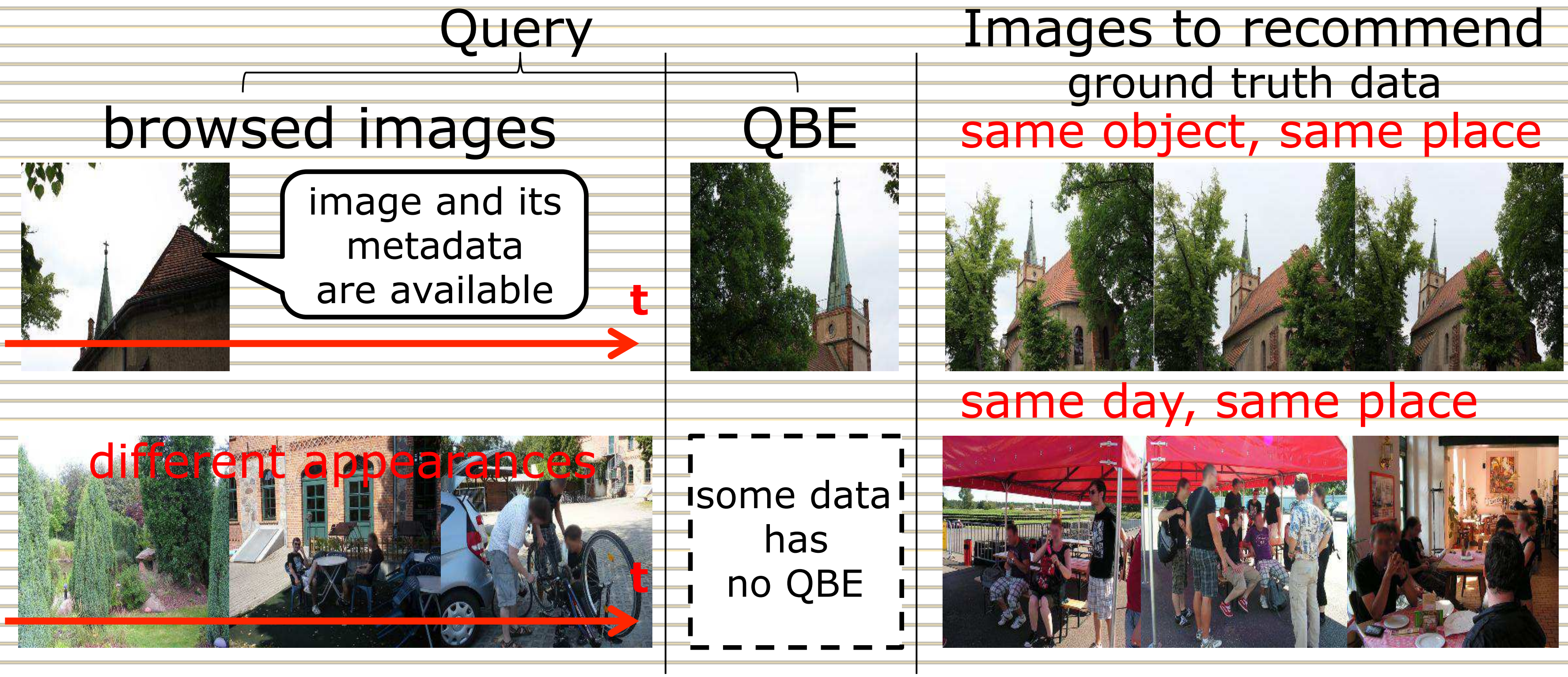
User's browsing data =actual browsing work

t

Query By Example
= User thinks best

Recommendation
"Are you looking for these images?"

Motivation



Query

browsed images

image and its metadata are available

t

different appearances

t

QBE

some data has no QBE

Images to recommend

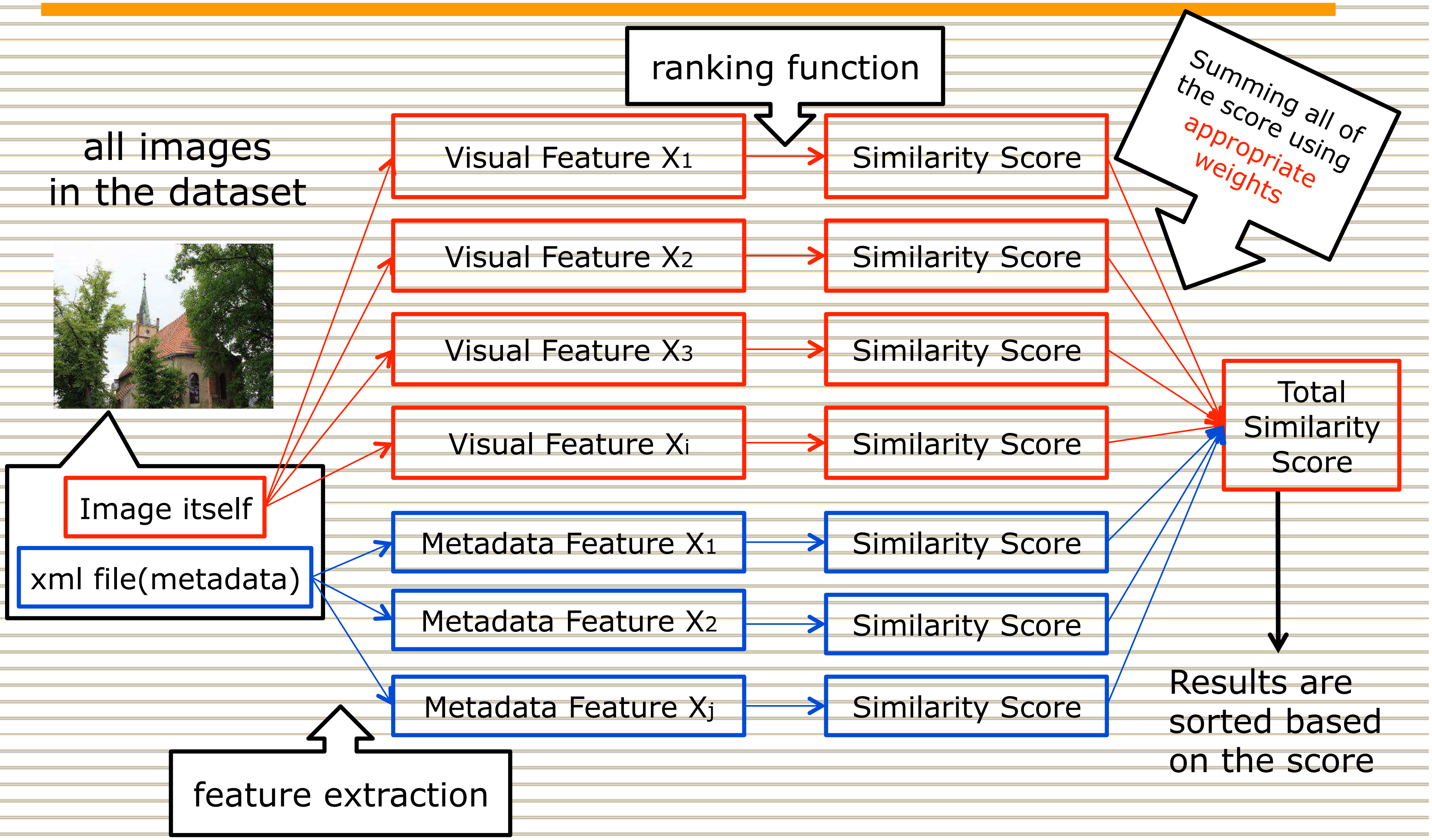
ground truth data

same object, same place

same day, same place

□ Task: estimating a topic from few query data and retrieve images which have the topic

General Photo Retrieval



all images in the dataset

Image itself

xml file(metadata)

feature extraction

Visual Feature X_1

Visual Feature X_2

Visual Feature X_3

Visual Feature X_i

Metadata Feature X_1

Metadata Feature X_2

Metadata Feature X_j

ranking function

Similarity Score

Similarity Score

Similarity Score

Similarity Score

Similarity Score

Similarity Score

Similarity Score

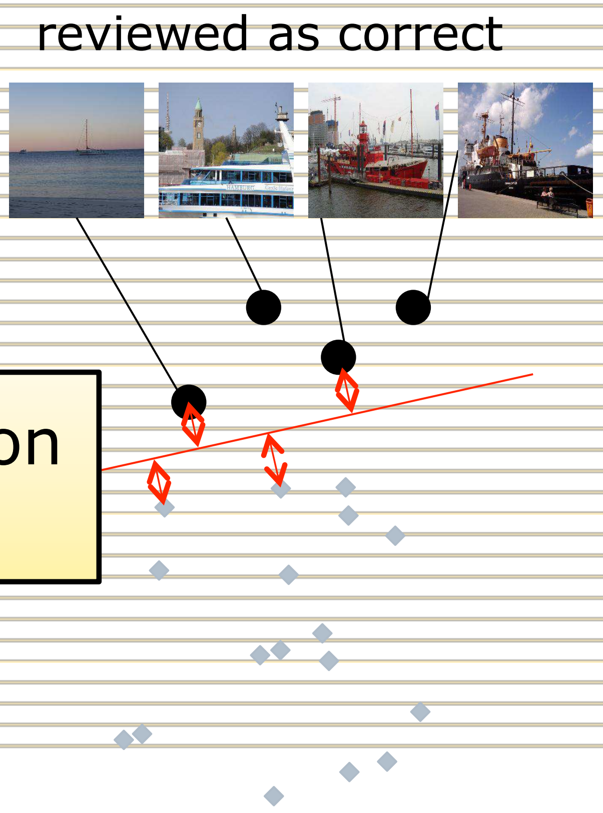
Summing all of the score using appropriate weights

Total Similarity Score

Results are sorted based on the score

Summing scores using appropriate weights

- □ Relevance Feedback: A Power Tool for Interactive Content-Based Image Retrieval [Y. Rui et al., 1998]
  - Learning with SVM classifier
  - Several visual descriptor
  - Similarity score is obtained by combining the scores of each feature with relevance feedback



- subjectivity of human's perception
- dynamically update weights

Calculate Similarity Score

- Learning to Rank for Content-Based Image Retrieval [F. Faria et al., 2010]
  - Learning a ranking by using the multi-stage evaluation by user

- K-Nearest Neighbors directed synthetic images injection[L. Piras et al., 2010]
  - No learning and simple

Methods depend on the query which is available.

Subtask2: Personal Photo Retrieval

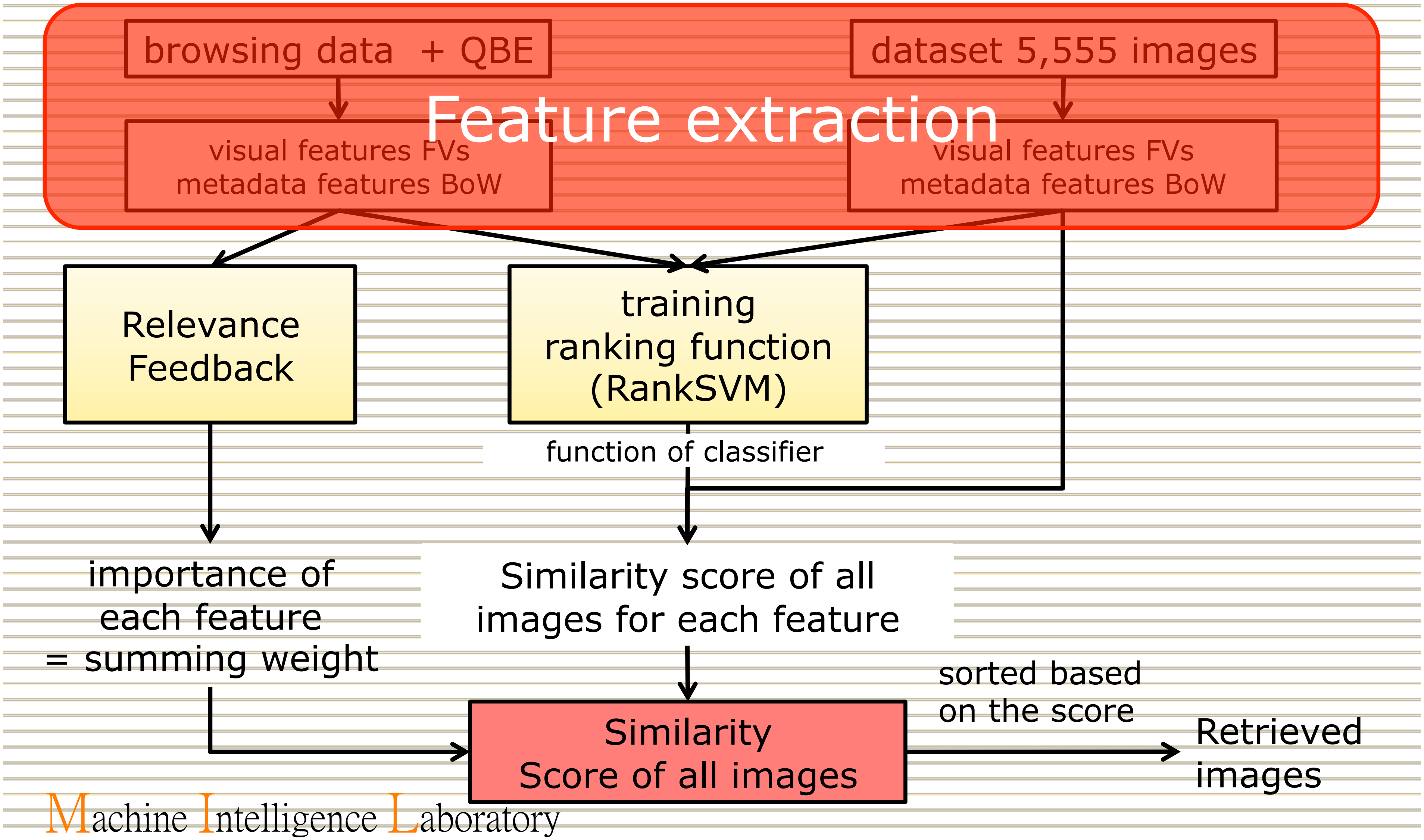
- □ Time-series data are available.
  - ■ browsing data is obtained sequentially

  ⇒Images that user browses later represent the topic better

  ⇒Ranking SVM [T. Joachims, 2003]
- □ The task requires a higher level object recognition to topic detection
  - ■ The latest feature coding for object recognition

  ⇒Fisher Vectors [F. Perronnin et al., ECCV 2010]

Methodology Overview



Feature extraction

browsing data + QBE

dataset 5,555 images

visual features FVs
metadata features BoW

visual features FVs
metadata features BoW

Relevance Feedback

training
ranking function
(RankSVM)

function of classifier

importance of
each feature
= summing weight

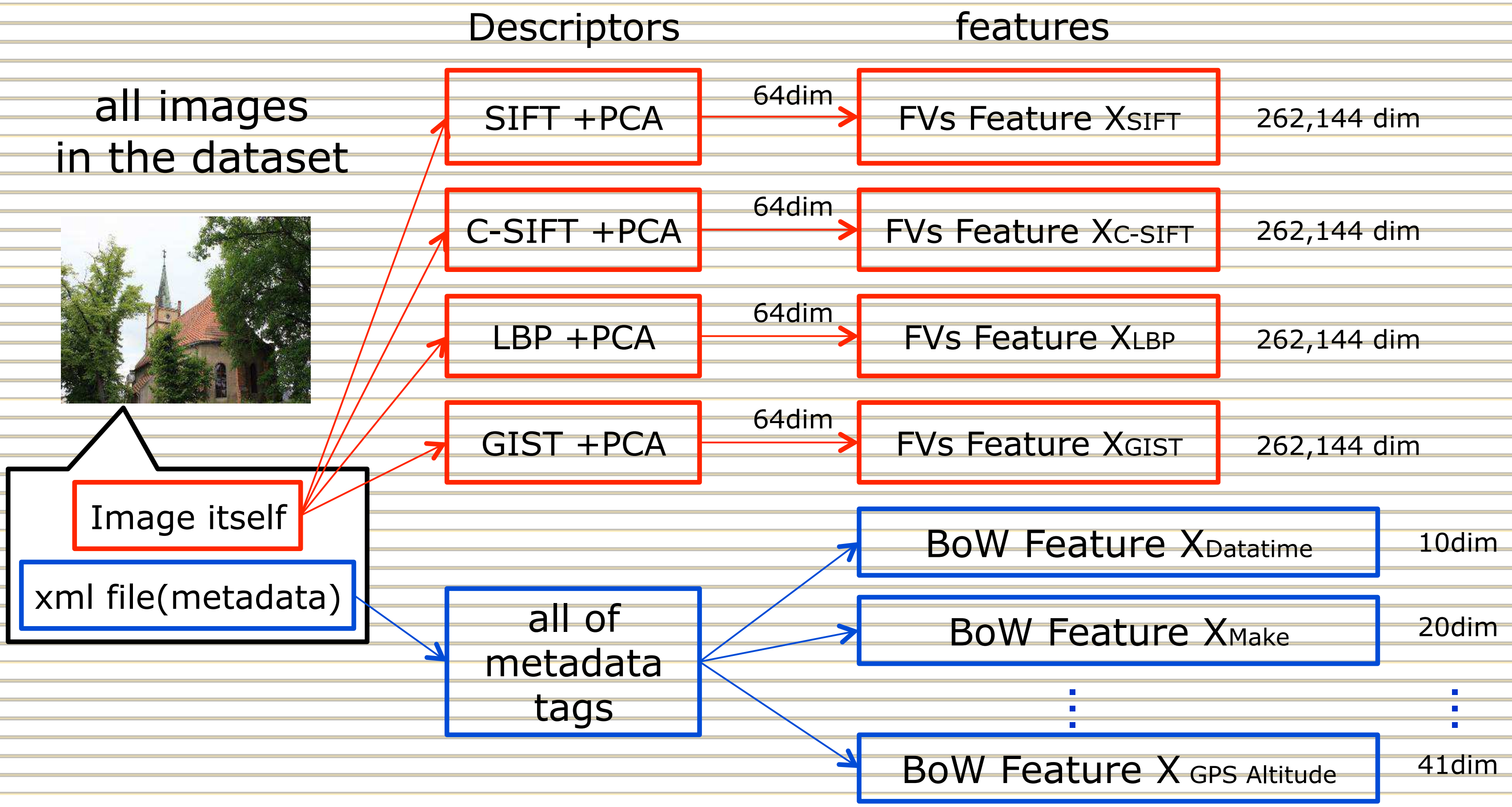
Similarity score of all
images for each feature

Similarity
Score of all images

sorted based
on the score

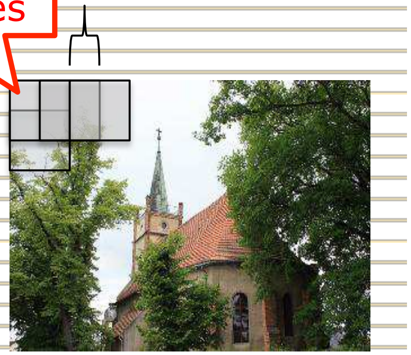
Retrieved
images

Feature Extraction Overview



Visual Feature Extraction

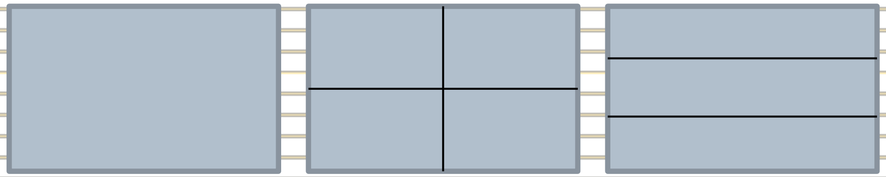
- We used the Improved Fisher Vectors (IFV)[F. Perronnin et al., ECCV 2010]
  - Dimension of IFV = 262,144
    - Local descriptors
      - 4 descriptors: SIFT, C-SIFT, GIST, LBP
      - use Global descriptors as Local one
        - 5 scales of local patches
        - Sampling: each 6 grid step



extract global feature from local patches

each 6 grid step

- Dimension reduction of local feature with PCA :64
- components in GMM :256
- spatial pyramid divided into 1x1, 2x2, and 3x1 cells



Metadata Feature Extraction

- □ Bag of Words representation (⇒ [0,0,0,1,0,...])
- □ Extract 10 Exif data from xml file given

EXIF data name	dimension
Make (Canon, NIKON, SONY, ...)	20
Model (Canon PowerShot, CYBERSHOT, ...)	38
Flash (auto, fired, ...)	13
SceneCaptureType (Portrait, Night scene, ...)	4
DateTime (2011, 2009, ...)	10
GPS Altitude (0 metres , 102 metres, ...)	41
GPS Latitude Ref (S, N)	2
GPS Latitude (8˚32‘42", 8˚17‘16", ...)	143
GPS Longitude Ref (E, W, ...)	2
GPS Longitude (150˚19’53.4", 6˚15’33.6", ...)	151

didn’t use about 30 metadata
“orientation” , “shutter speed”, ...

Retrieval Methods

- Similarity Score
  - train the classifier so that

QBE gets higher score than browsed images and Later browsed images are regarded as higher ranked than earlier ones.

score of QBE

∨

score of browsed last

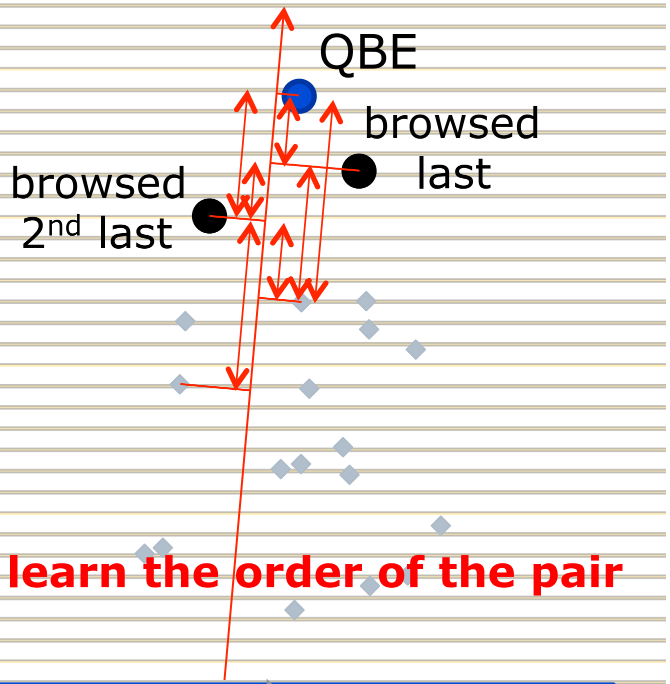
∨

score of browsed 2nd last

∨

score of the others

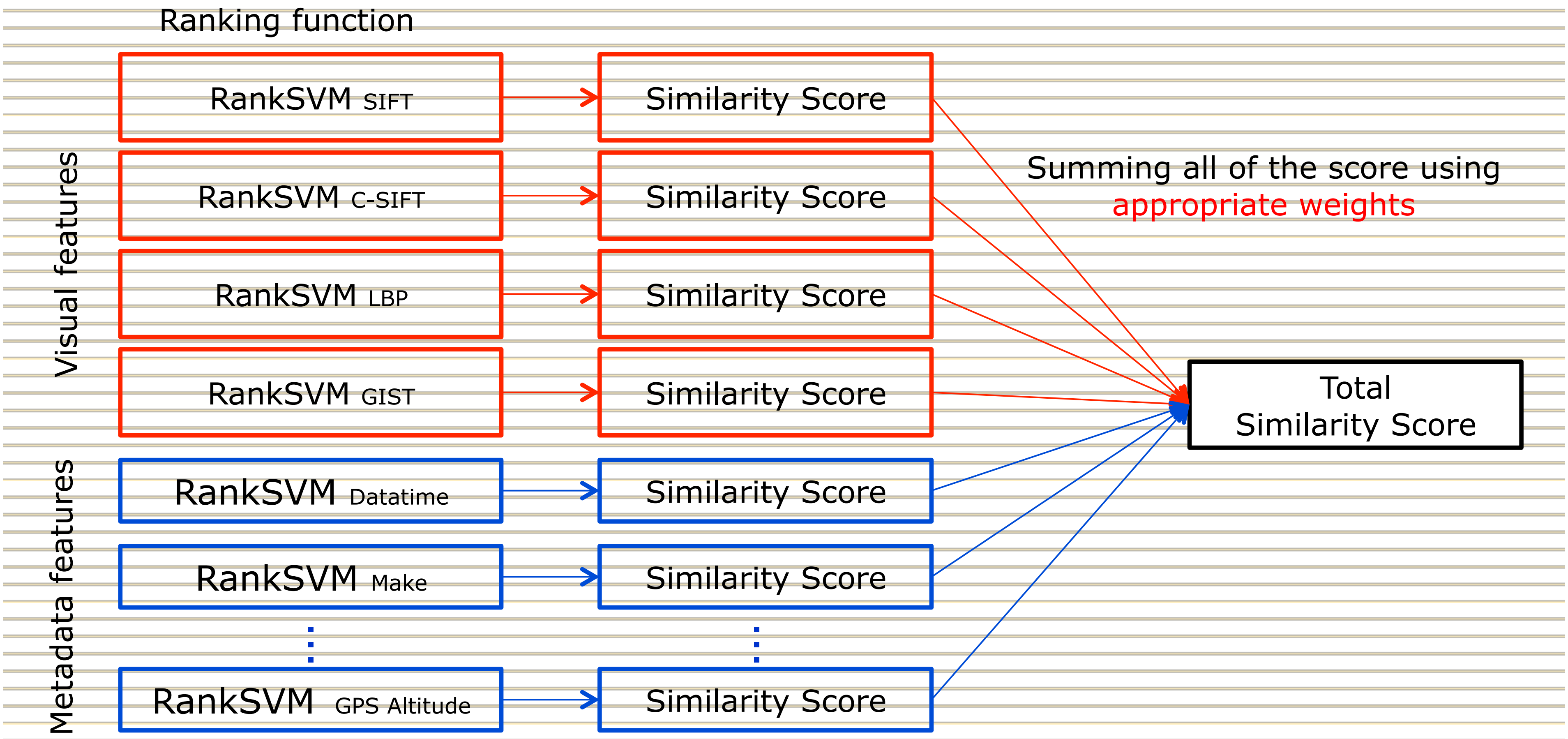
RankSVM



learn the order of the pair

Output from RankSVM ⇒ Score

Relevance Feedback



Ranking function

Visual features

RankSVM SIFT → Similarity Score

RankSVM C-SIFT → Similarity Score

RankSVM LBP → Similarity Score

RankSVM GIST → Similarity Score

Metadata features

RankSVM Datatime → Similarity Score

RankSVM Make → Similarity Score

⋮ ⋮

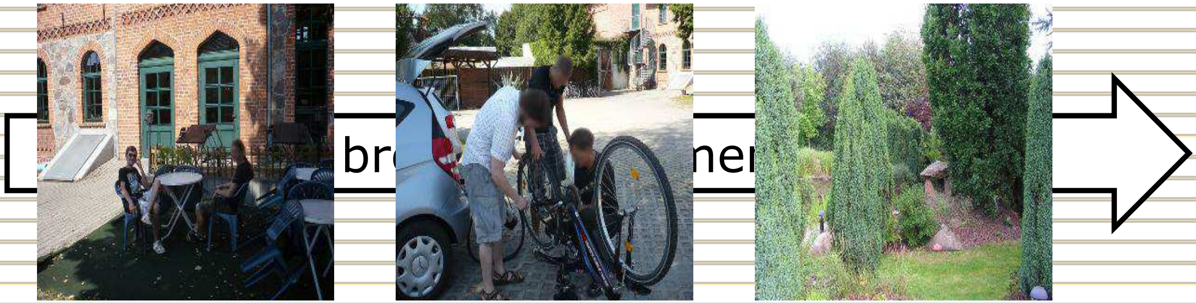
RankSVM GPS Altitude → Similarity Score

Summing all of the score using appropriate weights

Total Similarity Score

Relevance Feedback

- The weights are calculated by utilizing the browsing process.



firstly browsed image

[Visual feature] [DateTime feature] ... [GPS feature]

[Visual 2nd] [DateTime 2nd] ... [GPS 2nd]

Calculate variance

[Visual 3rd] → Visual large

[DateTime 3rd] → DateTime

...

[GPS 3rd] → GPS small

Initial weights / recalc weights

Visual: 1.0 0.8 decrease

DateTime: 1.0

...

GPS: 1.0 1.2 increase

$$\omega_{l,t}^{new} = \frac{\sigma_l^I}{\sigma_l^{B_t}}$$

←variance in all images

←variance in query images

$$\omega_{l,t} = \alpha \times \omega_{l,t}^{new} + (1-\alpha) \times \omega_{l,t}^{old}$$

Experiment

- 1.Ranking function and Feature representations comparison (Visual features)
  - RankSVM vs NN vs SVM
  - FVs coding vs LLCs coding
    - LLCs (Locality-constrained Linear Coding) [Lin et al., CVPR 2011]

      dimension = 1024 * 7 = 7168

      Local descriptors SIFT, C-SIFT, LBP and GIST
- 2.Ranking function comparison (Metadata features)
  - RankSVM vs NN vs SVM
- 3.Combinations of visual and metadata features

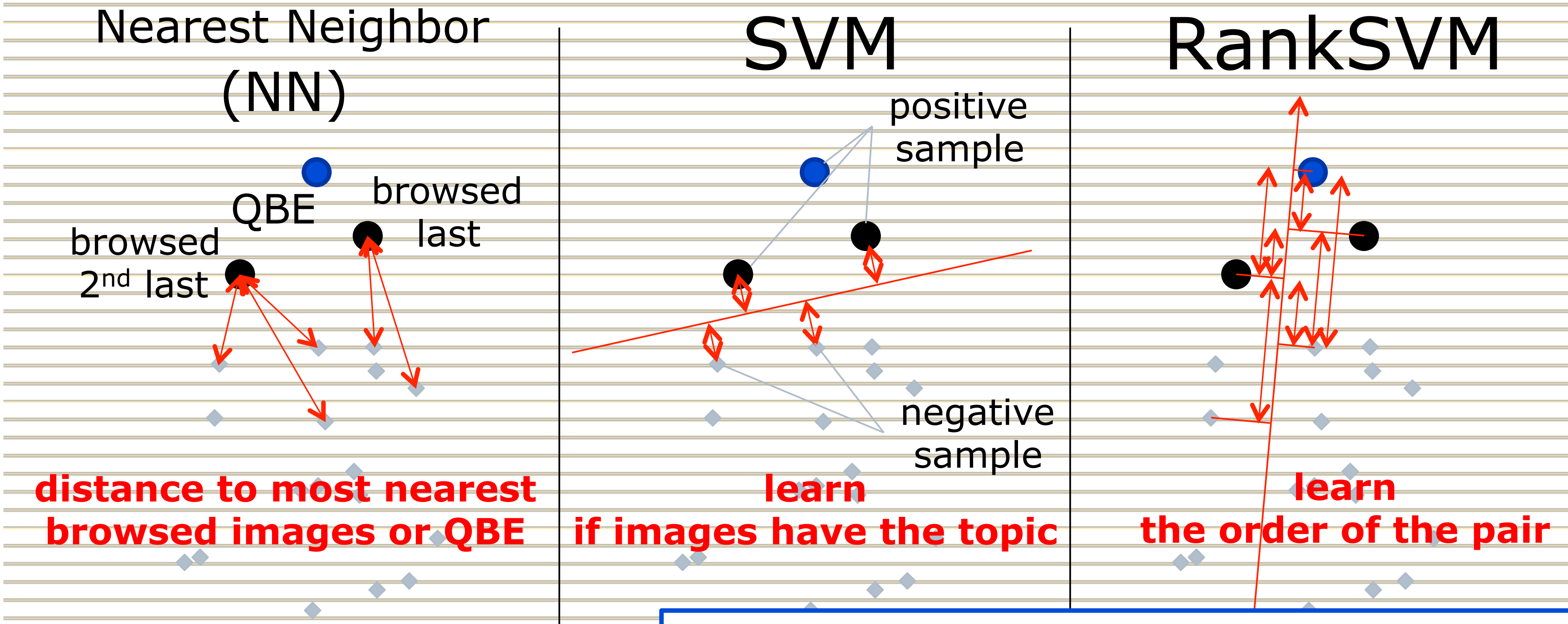
･number of topics(browsing data) : 74 ･Dataset : 5,555 images

･browsing data and QBE : 1~4 images

･Evaluation : NDCG(ndcg_cut_100) on ground truth data

1.Ranking function comparison(Visual features)

□ calculation methods of similarity score (Visual features)



Nearest Neighbor (NN)

SVM

RankSVM

distance to most nearest browsed images or QBE

learn if images have the topic

learn the order of the pair

Distance in NN
Output of SVM
Output of RankSVM
⇒ Score

1.Ranking function and Feature representations comparison (Visual feature only)

ndcg_cut_100	NN	SVM	rankSVM
LLCs+SIFT	0.2946	0.3066	0.3308
LLCs+C-SIFT	0.2856	0.2967	0.3257
LLCs+LBP	0.3043	0.3199	0.3385
LLCs+GIST	0.2796	0.2943	0.3175
FVs+ SIFT	0.3135	0.3278	0.3357
FVs+ C-SIFT	0.3492	0.3486	0.3696
FVs+ LBP	0.3636	0.3363	0.3861
FVs+ GIST	0.3376	0.3145	0.3572

- SVM < NN < RankSVM
- LLCs < FVs

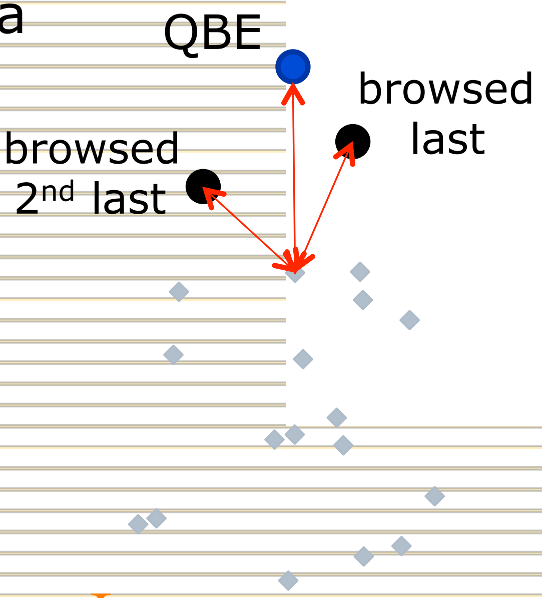
2.Ranking function comparison (Metadata features)

calculation methods of similarity score (Metadata features)

□ Nearest Neighbor

- Distance metric like RBF kernel between images

  Euclidean distance is not appropriate for BoW
- Summing similarity scores of image and all browsed data



$$d(\boldsymbol{x}_i^m, \boldsymbol{x}_j^m) = 1 - e^{-\tau||\boldsymbol{x}_i^m - \boldsymbol{x}_j^m||^2}$$

$$c_{i,j} = \frac{1}{1 + d(\boldsymbol{x}_i^m, \boldsymbol{x}_j^m)}$$

$$s_{t,i}^m = \sum_{k \in B_t \cup q_t} c_{k,i}$$

calculation methods of score comparison

ndcg_cut_100

□ RankSVM ⇒ 0.6508

□ SVM ⇒ 0.6367

□ Nearest Neighbor ⇒ 0.6228
with RBF kernel

□ Nearest Neighbor ⇒ 0.6203
with RBF kernel
without Relevance Feedback

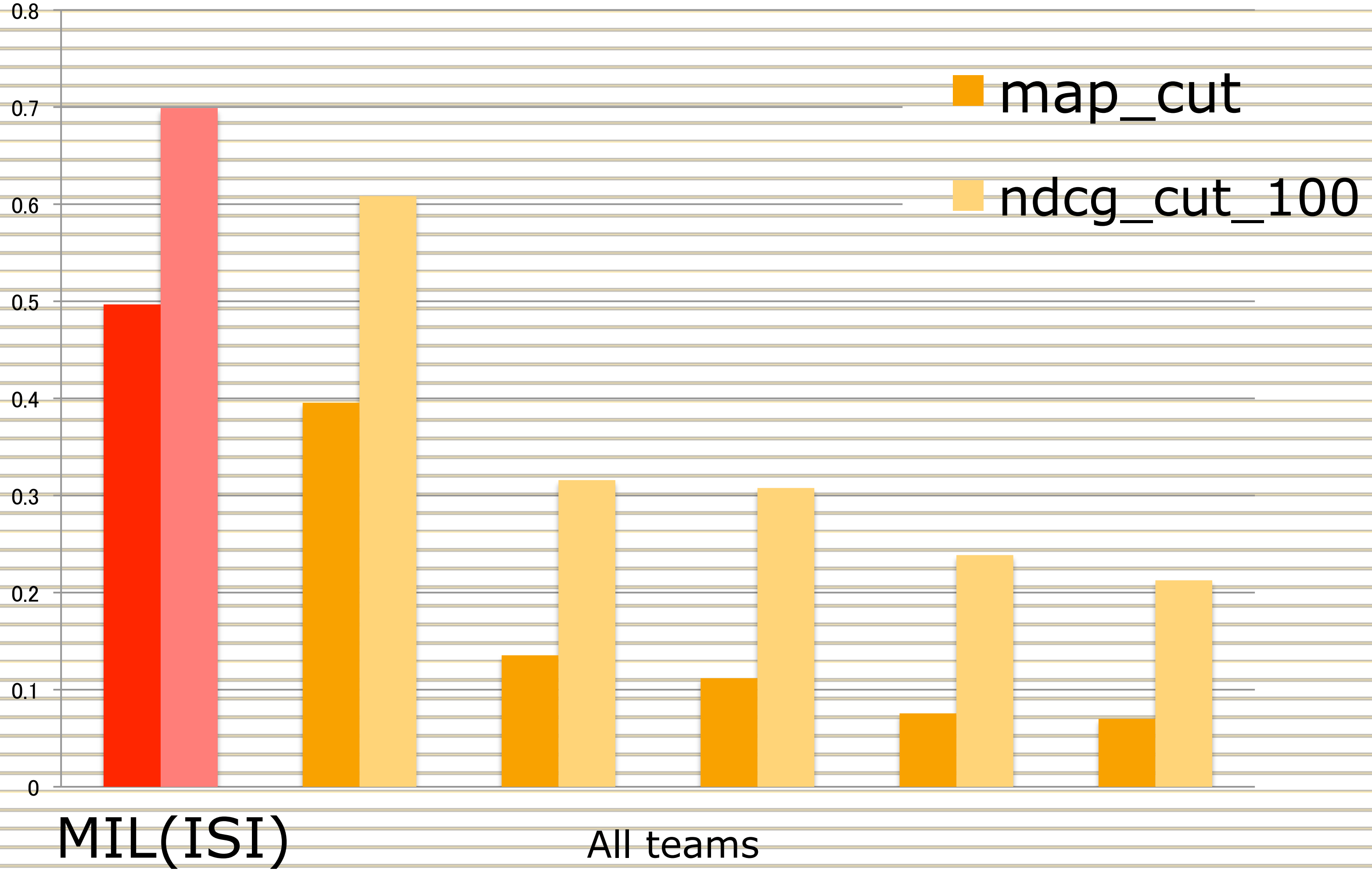
NN < SVM < RankSVM

3.Combinations of visual and metadata features

- □ We used RankSVM as ranking function and FVs for visual features.

SIFT	✓	–	–	–	✓	–
C-SIFT	✓	✓	–	✓	✓	✓
LBP	✓	✓	–	✓	✓	✓
GIST	✓	✓	–	✓	✓	–
10 Metadata	✓	✓	✓	–	–	–
ndcg_cut_100	0.7039	0.7040	0.6508	0.4236	0.4186	0.4166
ndcg_cut_20	**0.7477**	0.7463	0.6689	0.5193	0.5134	0.5171

Result



Conclusions

□ Motivation

Estimating a topic from few query data and retrieve images which have the topic

□ Methodology

Train RankSVM for

visual features(**FVs** of SIFT, C-SIFT, LBP, GIST) and

metadata features(**BoW** of 10 Exif data).

- Make
- Model
- Flash
- SceneCaptureType
- DateTime
- GPS Altitude
- GPS Latitude Ref
- GPS Latitude
- GPS Longitude Ref
- GPS Longitude

Combine similarity score with relevance feedback

□ Result

LLCs < FVs (Visual)

SVM < NN < RankSVM (Visual)

NN < SVM < RankSVM (Metadata)

- Thank you for listening.

UNIVERSITAT POLITÈCNICA DE VALÈNCIA
CLEF 2013
Conference and Labs of the Evaluation Forum
Information Access Evaluation meets Multilinguality, Multimodality, and Visualization
23 - 26 September
Valencia - Spain
Yandex

Topic:
CLEF2013@Valencia

Index

- Outline of subtask
- Methodology
  - Outline
  - Feature Extraction
  - Retrieving Methods
  - Relevance Feedback
- Results
- Conclusions

What is needed?

- another example

QBE

browsed images

same day, place, camera

What is needed?

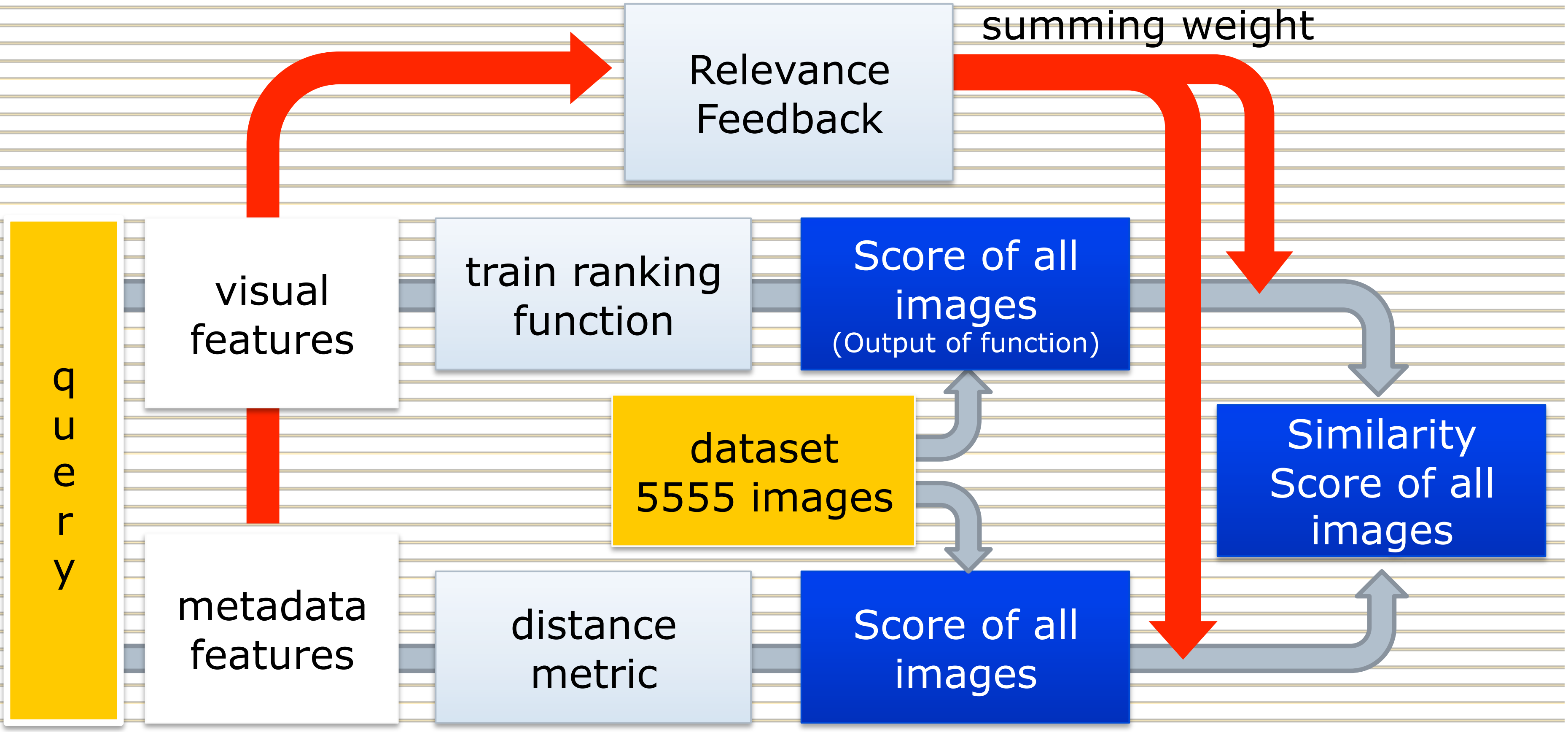
- another example

QBE

browsed images

same object

Methodology Outline



Retrieving Methods

- Metadata Similarity Score
  - Distance between image i & j

$$d(\boldsymbol{x}_i^m, \boldsymbol{x}_j^m) = 1 - e^{-\tau||\boldsymbol{x}_i^m - \boldsymbol{x}_j^m||^2}$$

$$c_{i,j} = \frac{1}{1 + d(\boldsymbol{x}_i^m, \boldsymbol{x}_j^m)}$$

  - Similarity of image i for topic t

$$s_{t,i}^m = \sum_{k \in B_t \cup q_t} c_{k,i}$$

summing the scores with all of the query browsed images

Result 2 Feature combination

□ Combinations of FVs visual features only

SIFT	–	✓	–	✓	–	✓
C–SIFT	✓	✓	✓	✓	–	✓
LBP	✓	✓	✓	✓	✓	–
GIST	✓	✓	–	–	✓	✓
10 Metadata	–	–	–	–	–	–
ndcg_cut_100	0.4236	0.4186	0.4186	0.4118	0.4058	0.4008

Result 3

- □ metadata features only

	SIFT	–	–
	C-SIFT	✓	–
	LBP	✓	–
	GIST	✓	–
10	Metadata	–	✓
ndcg_cut_100		0.4236	0.6228

Result 4 Feature combination

- □ Top combinations of visual and metadata features

	SIFT	✓	–	–	✓	✓	✓
	C-SIFT	✓	✓	✓	✓	–	✓
	LBP	✓	✓	✓	✓	✓	–
	GIST	✓	✓	–	–	✓	✓
10	Metadata	✓	✓	✓	✓	✓	✓
ndcg_cut_100		0.6998	0.6986	0.6985	0.6983	0.6982	0.6967

We submitted this score

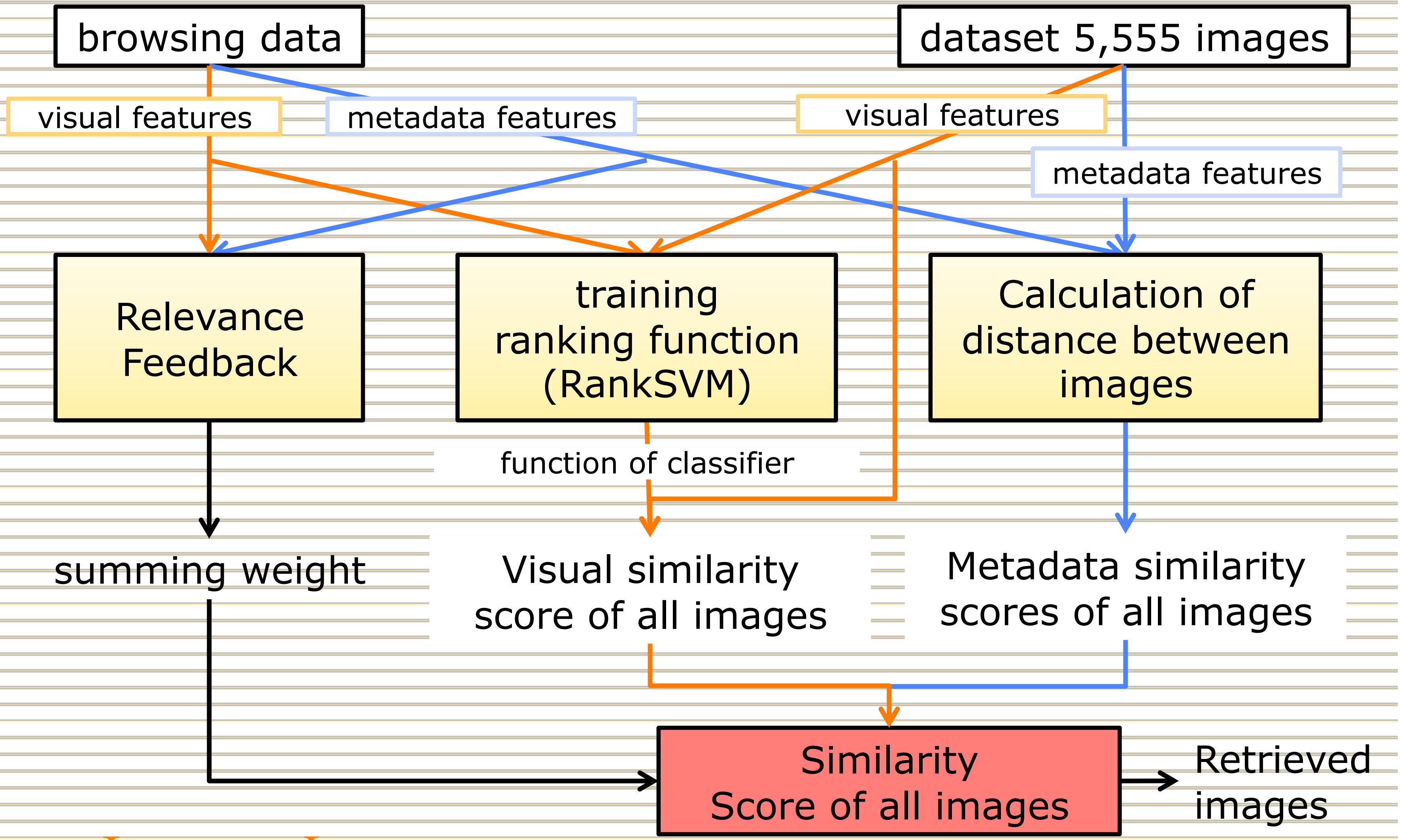
Result 2 Feature combination

- □ Combinations of FVs visual features only

SIFT	✓	–	–	–	✓	–
C-SIFT	✓	✓	–	✓	✓	✓
LBP	✓	✓	–	✓	✓	✓
GIST	✓	✓	–	✓	✓	–
10 Metadata	✓	✓	✓	–	–	–
ndcg_cut_100	0.6998	0.6986	0.6228	0.4236	0.4186	0.4186

We submitted this score

Methodology Outline



browsing data

dataset 5,555 images

visual features

metadata features

visual features

metadata features

Relevance Feedback

training ranking function (RankSVM)

Calculation of distance between images

function of classifier

summing weight

Visual similarity score of all images

Metadata similarity scores of all images

Similarity Score of all images

Retrieved images

- 1.Ranking function and Feature representations comparison (Visual features)
  - RankSVM vs NN vs SVM
  - FVs coding vs LLCs coding
    - LLCs (Locality-constrained Linear Coding) [Lin et al., CVPR 2011]

      dimension = 1024 * 7 = 7168

      Local descriptors SIFT, C-SIFT, LBP and GIST
- 2.Ranking function comparison (Metadata features)
  - RankSVM vs SVM vs Distance metric

・number of topics(browsing data) : 74

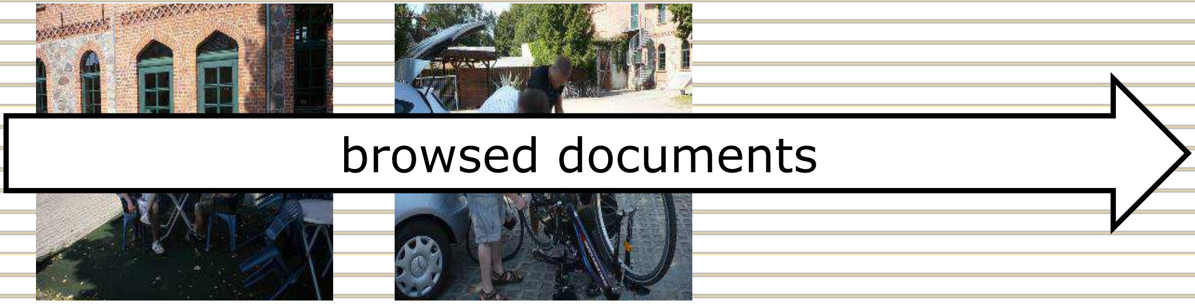
・Dataset : 5,555 images

・browsing data and QBE : 1~4 images

・Evaluation : NDCG(ndcg_cut_100) on ground truth data

Relevance Feedback

- The weights are calculated by utilizing the browsing process.

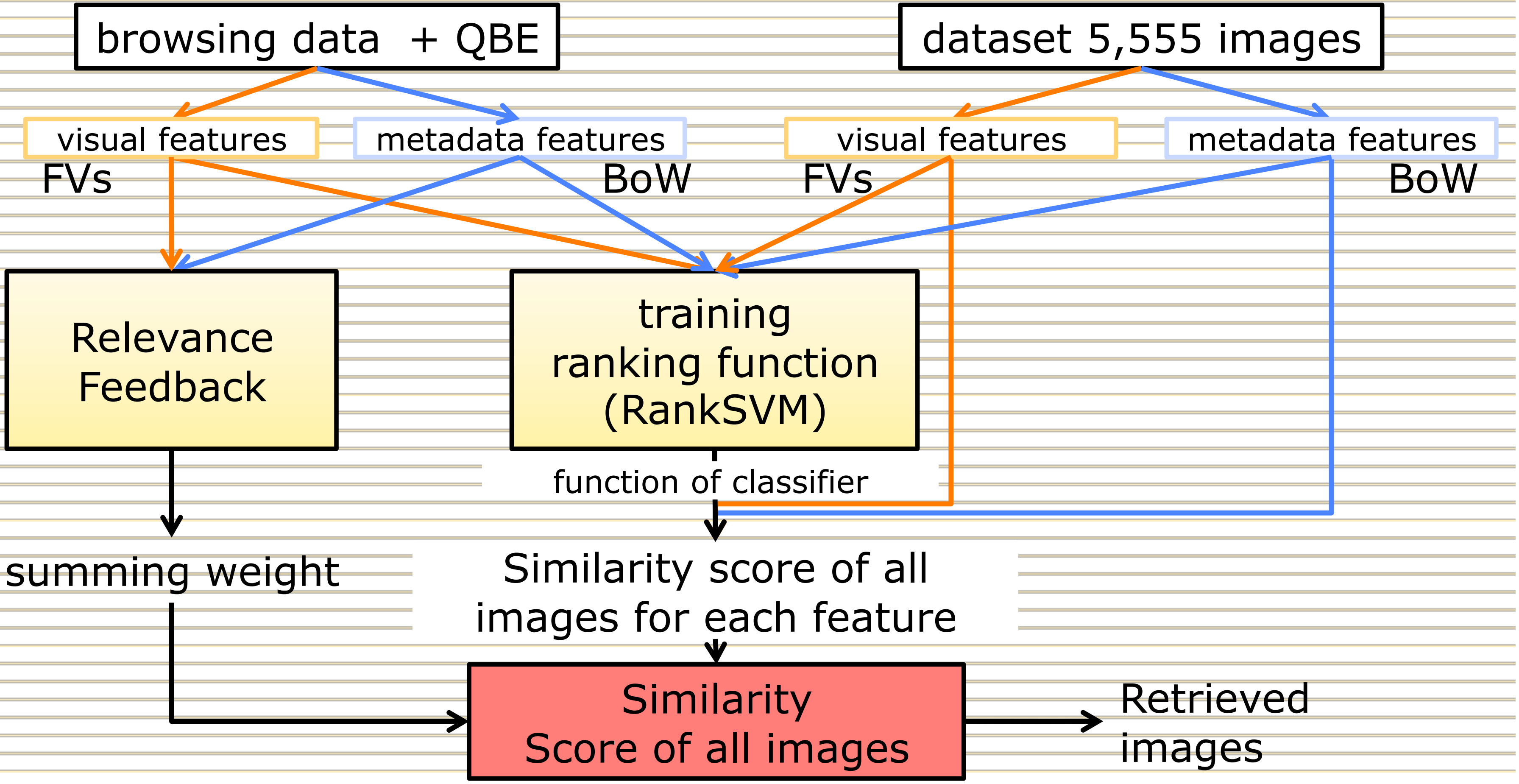


firstly browsed image		Calculate variance	recalc weights
[Visual feature]	[Visual 2nd]	$\Rightarrow \sigma_{\mathrm{Visual}}$ large	Visual: 0.8 decrease
[DateTime feature]	[DateTime 2nd]	$\Rightarrow \sigma_{\mathrm{DateTime}}$	DateTime: 1.0
...	...	. ..	...
[GPS feature]	[GPS 2]	$\Rightarrow \sigma_{\mathrm{GPS}}$ small	GPS: 1.2 increase

$$\omega_{l,t}^{new} = \frac{\sigma_l^I}{\sigma_l^{B_t}}$$

$$\omega_{l,t} = \alpha \times \omega_{l,t}^{new} + (1-\alpha) \times \omega_{l,t}^{old}$$

Methodology Outline



browsing data + QBE

dataset 5,555 images

visual features

metadata features

FVs

BoW

Relevance Feedback

training ranking function (RankSVM)

function of classifier

summing weight

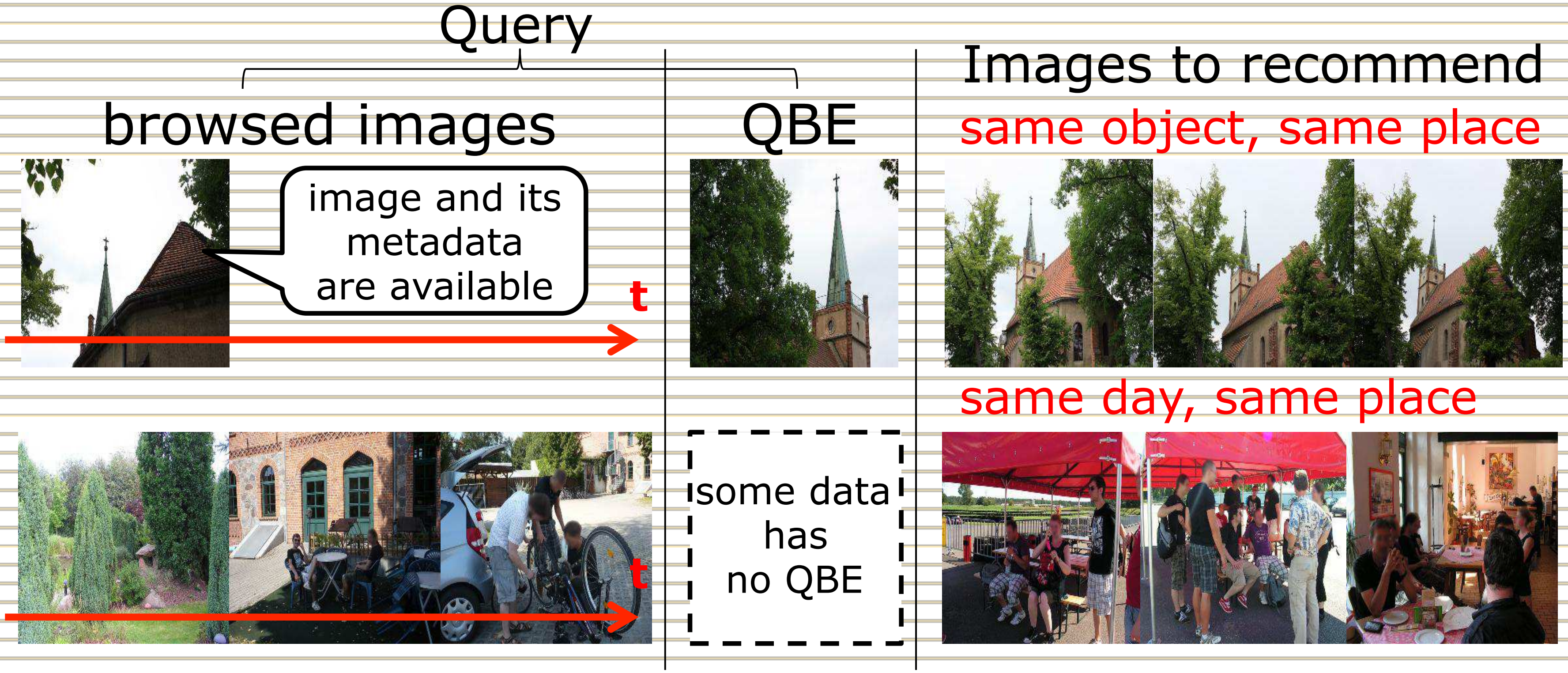
Similarity score of all images for each feature

Similarity Score of all images

Retrieved images

Machine Intelligence Laboratory

Motivation



Query

browsed images

image and its metadata are available

QBE

some data has no QBE

Images to recommend

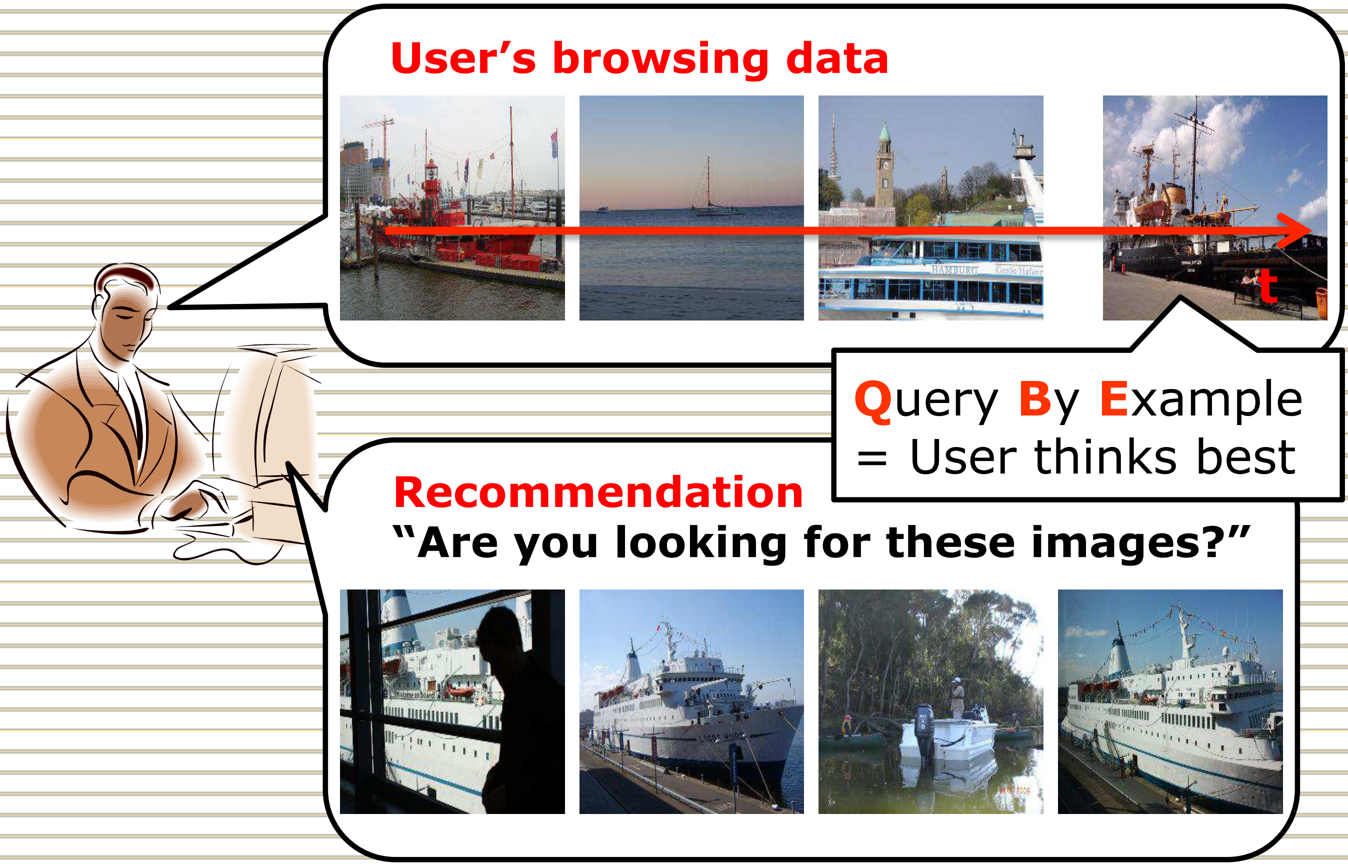
same object, same place

same day, same place

- □ Task: estimating a topic from few query data and retrieve images which have the topic

Subtask2: Personal Photo Retrieval

The system which can help users to retrieve images from a lot of personal photo collections using browsing data.



User's browsing data

Query **B**y **E**xample
= User thinks best

Recommendation
"Are you looking for these images?"

Visual Feature Extraction

- □ FVs (Fisher Vectors)

  [F. Perronnin et al., ECCV 2010]

  dimension = 262,144

  Local descriptors

  SIFT, C-SIFT, LBP and GIST

  using Global descriptors as Local one (densely extracted from five scales of patches on a regular grid every six pixels)

  256 GMM components

  Spatial pyramid divided into 1x1, 2x2, and 3x1 cells

Metadata Feature Extraction

- □ Bag of Words representation (⇒ [0,0,0,1,0,…])

EXIF data name	dimension
Make (Canon, NIKON, SONY, …)	20
Model (Canon PowerShot, CYBERSHOT, …)	38
Flash (auto, fired, …)	13
SceneCaptureType (Portrait, Night scene, …)	4
DateTime (2011, 2009, …)	10
GPS Altitude (0 metres , 102 metres, …)	41
GPS Latitude Ref (S, N)	2
GPS Latitude (8˚32‘42", 8˚17‘16", …)	143
GPS Longitude Ref (E, W, …)	2
GPS Longitude (150˚19’53.4", 6˚15’33.6", …)	151

Retrieval Methods

- Similarity Score
  - train the classifier so that

QBE gets higher score than browsed images and Later browsed images are regarded as higher ranked than earlier ones.

score of QBE

∨

score of browsed last

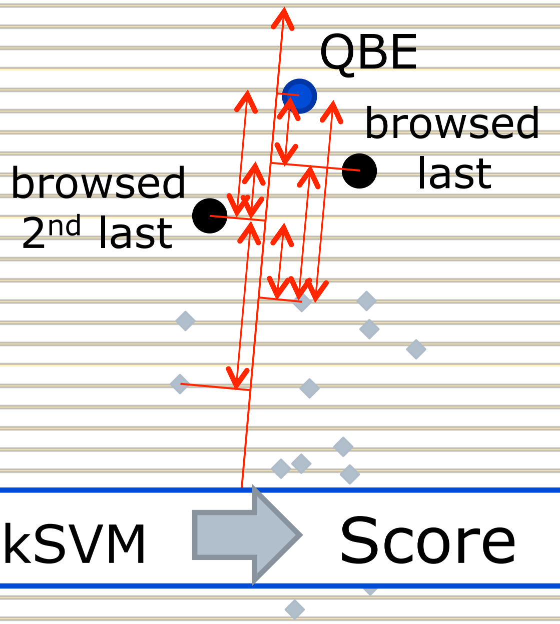
∨

score of browsed 2nd last

∨

score of the others

RankSVM



Output from RankSVM ⇒ Score

Conclusions

- Motivation

  To estimate a topic from few query data and retrieve images which have the topic

- Methodology

  visual (**RankSVM** + **FVs** of C-SIFT, LBP, GIST)

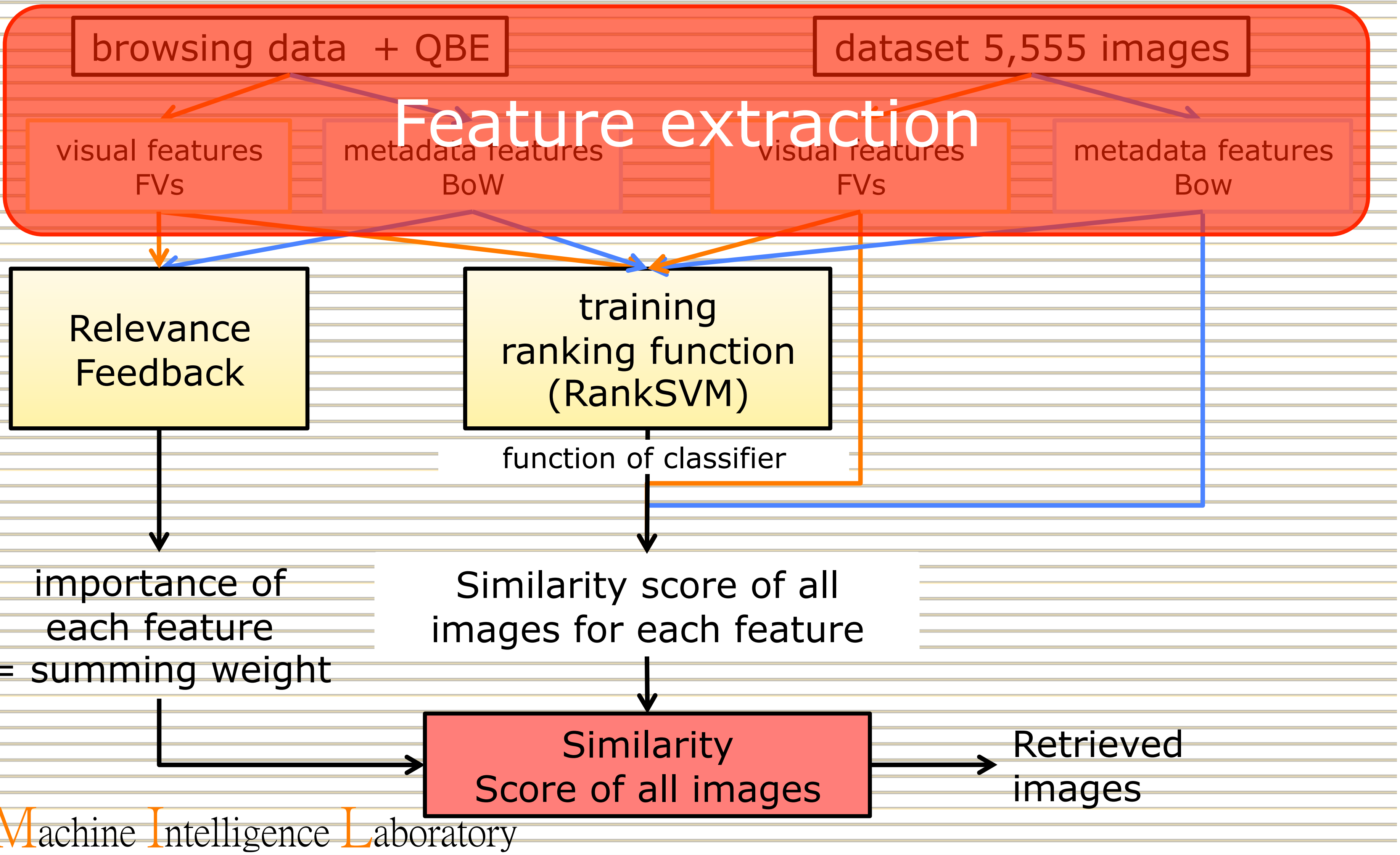
  + relevance feedback

  metadata (**RankSVM** + Bow of 10 Exif data)

- Result

  LLCs < FVs
  SVM < NN < RankSVM (Visual)
  Distance metric < SVM < RankSVM (Metadata)

Methodology Overview



browsing data + QBE

dataset 5,555 images

Feature extraction

visual features
FVs

metadata features
BoW

visual features
FVs

metadata features
Bow

Relevance
Feedback

training
ranking function
(RankSVM)

function of classifier

importance of
each feature
= summing weight

Similarity score of all
images for each feature

Similarity
Score of all images

Retrieved
images