université de Caen
Basse-Normandie

Image re-ranking based on statistics of frequent patterns

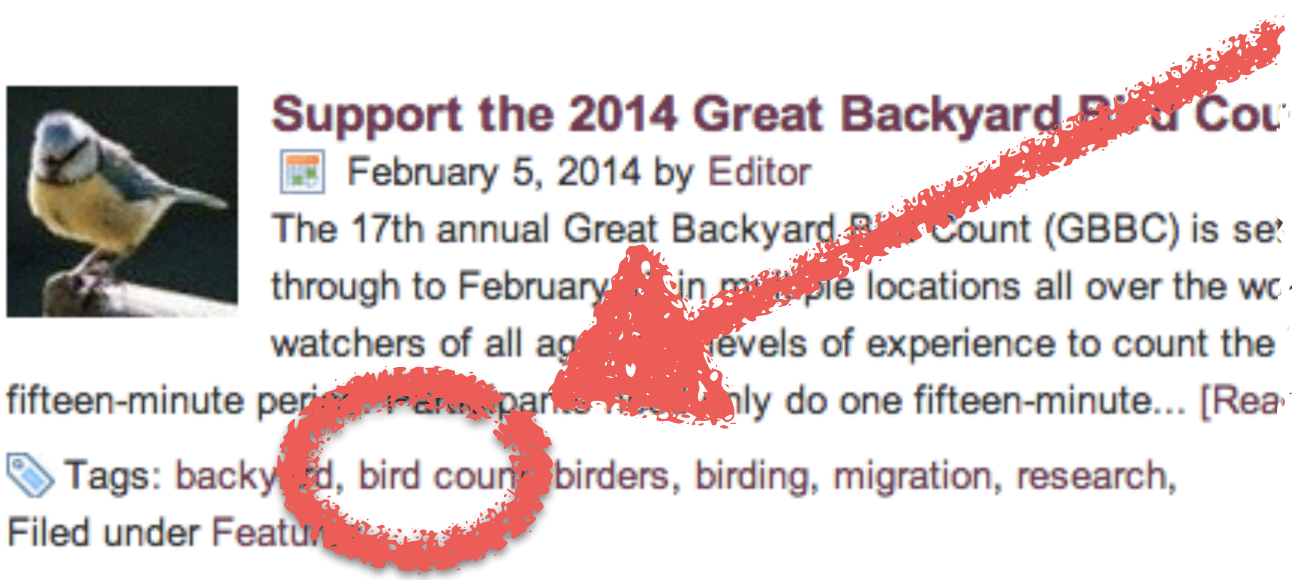
Winn Voravuthikunchai, Bruno Crémilleux and Frédéric Jurie
Université de Caen Basse-Normandie, France
CNRS UMR 6072 – ENSICAEN
firstname.name@unicaen.fr

Glasgow2014 ICMR
University of Glasgow
ACM International Conference on Multimedia Retrieval

Large scale image search

e.g. 10,000,000 images

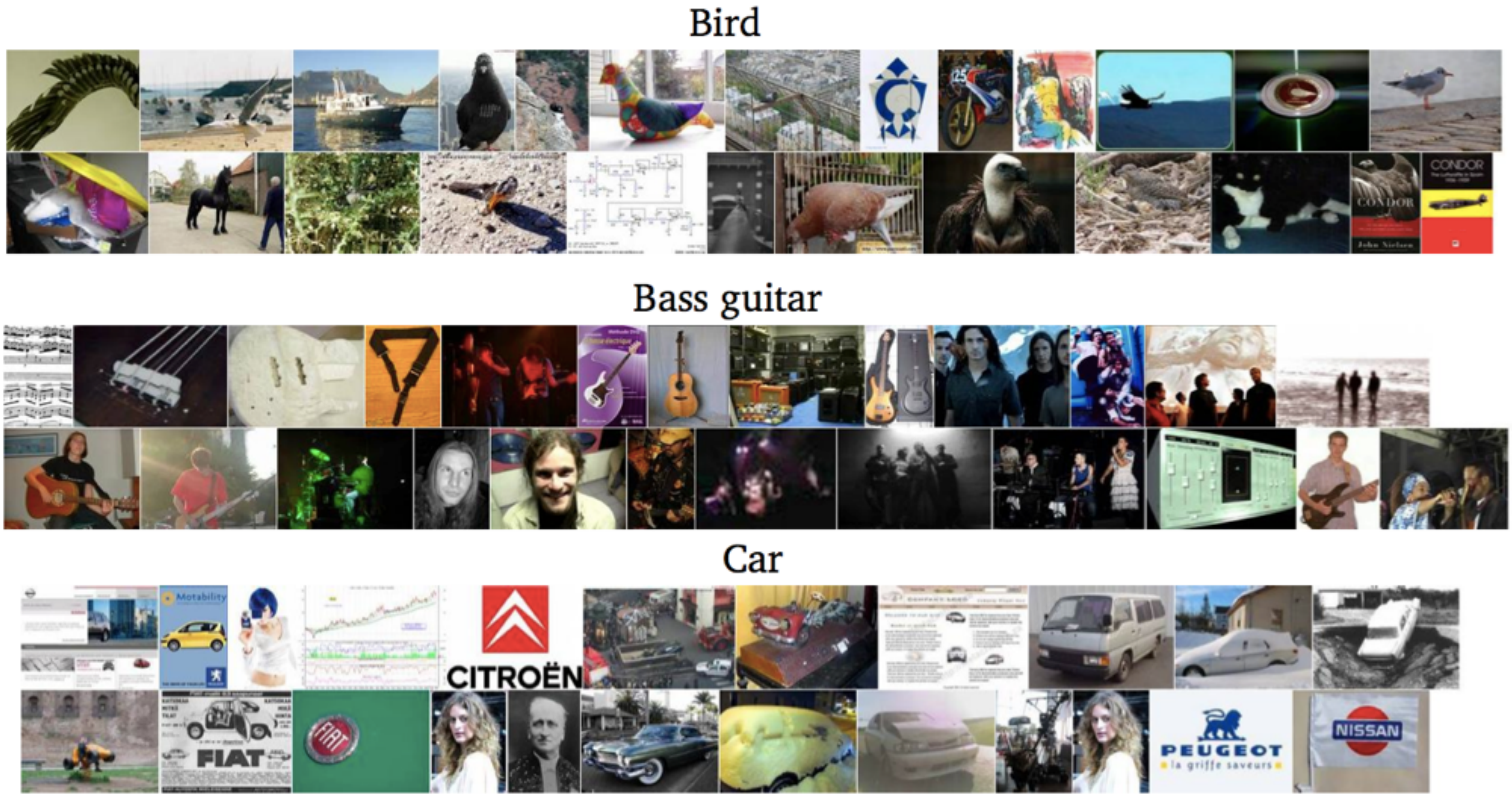
- Free text-based queries (easy to formulate)
- Text: fast indexing structures (inverted files)




```
<title>Birds.com: Online Birds
Guide with Facts, Articles, Videos,
and Photos</title>
<meta name="description"
content="Birds and Birding guide
showcasing information about
species, ornithology, bird watching
activities, care, education, photos
and original articles about
birds." />
```

- Use image content (more accurate)
- Free queries: precomputing image classifiers can not be done offline
- Reranking: get images based on text, rerank them based on image content … but online computations!

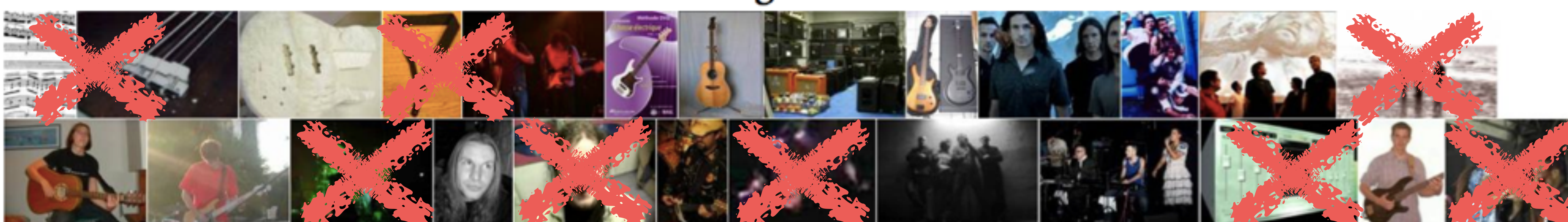
Initial set of images



Large scale image search

Bird

Bass guitar



Car

Initial set of images

Bird

Bass guitar

Car

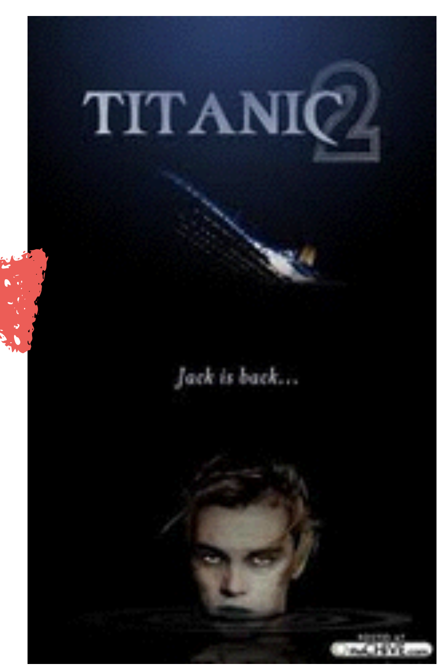
Key Assumption

Relevant images

Visual constancy (at least within groups)

Non-relevant images

Scatterred



Online mining of visually similar structures

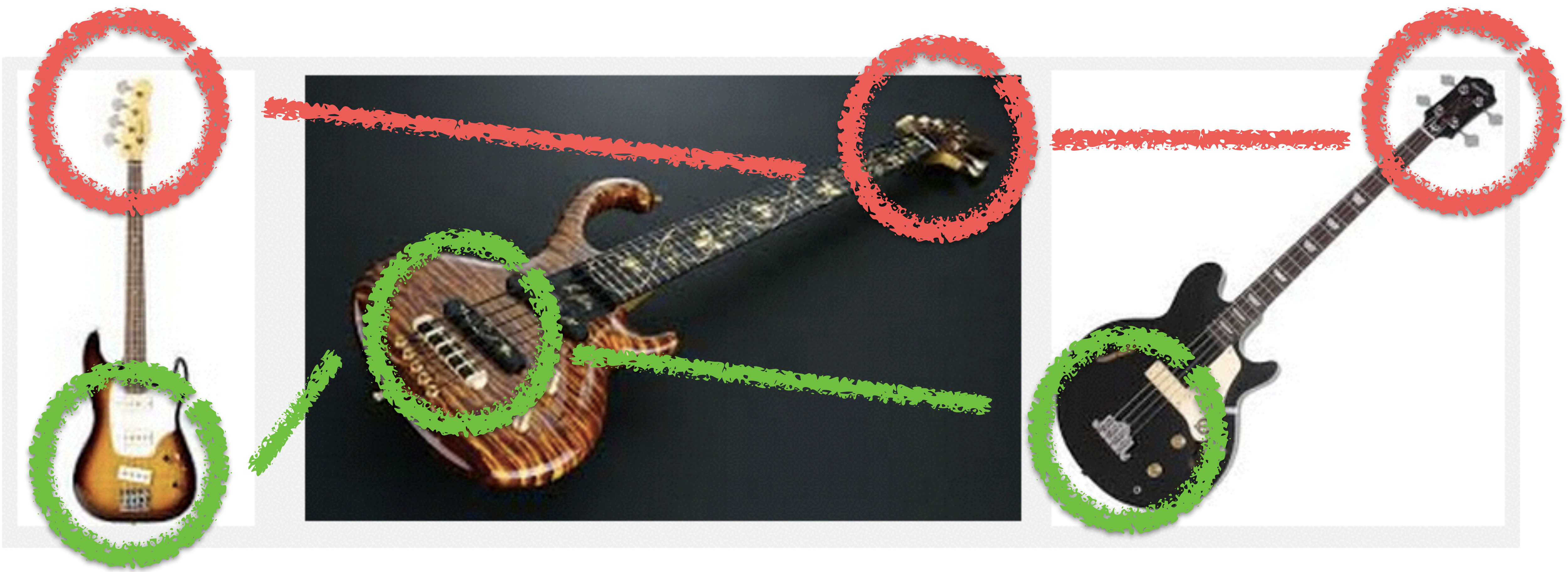


Image	Relevancy	Rank
I	yes	1
I	yes	2
I	no	3
I	yes	4
I	no	5

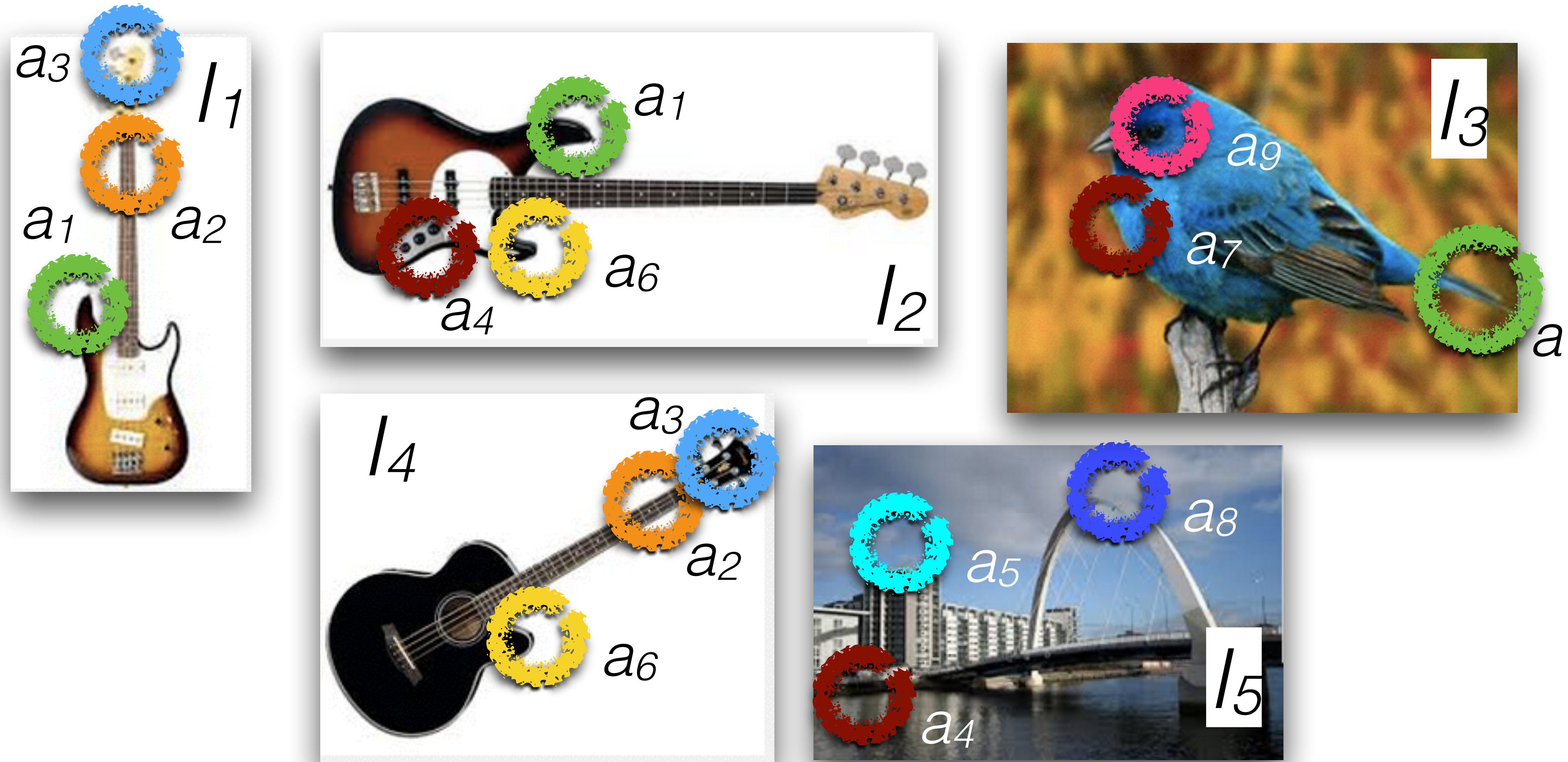


Image I_i	Trans. t_i	rel.
I_1	$\{a_1, a_2, a_3\}$	yes
I_2	$\{a_1, a_4, a_6\}$	yes
I_3	$\{a_1, a_7, a_9\}$	no
I_4	$\{a_2, a_3, a_6\}$	yes
I_5	$\{a_4, a_5, a_8\}$	no

Initial ranking

Image I_i	Trans. t_i	rel.
I_1	$\{a_1, a_2, a_3\}$	yes
I_2	$\{a_1, a_4, a_6\}$	yes
I_3	$\{a_1, a_7, a_9\}$	no
I_4	$\{a_2, a_3, a_6\}$	yes
I_5	$\{a_4, a_5, a_8\}$	no

- *Pattern* (or *Itemset)* X = $\{a_1, \ldots, a_k\}$: a set of items, subsequences, substructures, etc.
- Frequency / support: # of occurrences $\mathcal{F}(a_2, a_3) = 2$
- *Frequent* pattern: a pattern that occurs frequently in a data set (min freq)
- An itemset X is closed if there exists no super-pattern Y, Y⊃X with the same support as X
- Closed pattern is a lossless compression of freq. patterns

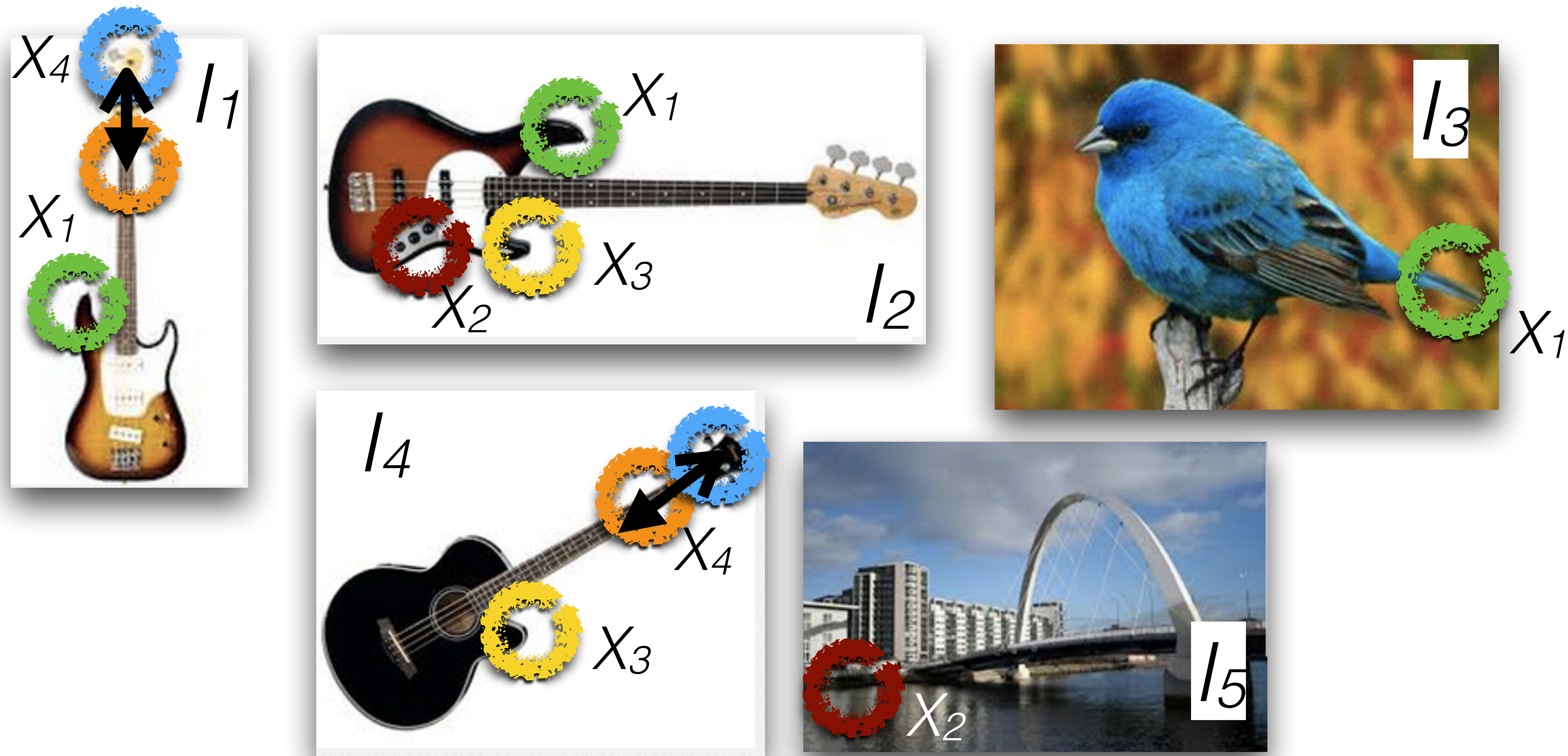
Initial ranking

Image I_i	*Trans.* t_i	*rel.*
I_1	$\{a_1, a_2, a_3\}$	*yes*
I_2	$\{a_1, a_4, a_6\}$	*yes*
I_3	$\{a_1, a_7, a_9\}$	*no*
I_4	$\{a_2, a_3, a_6\}$	*yes*
I_5	$\{a_4, a_5, a_8\}$	*no*

Frequent closed patterns

Patterns $\mathcal{X}_j$
$\mathcal{X}_1 = \{a_1\}$
$\mathcal{X}_2 = \{a_4\}$
$\mathcal{X}_3 = \{a_6\}$
$\mathcal{X}_4 = \{a_2, a_3\}$

$\mathcal{F} \geq 2$



X_4
I_1
X_1
X_1
X_2
X_3
I_2
I_3
X_1
I_4
X_4
X_3
I_5
X_2

Image I_i	*Trans.* t_i	*rel.*
I_1	$\{a_1, a_2, a_3\}$	*yes*
I_2	$\{a_1, a_4, a_6\}$	*yes*
I_3	$\{a_1, a_7, a_9\}$	*no*
I_4	$\{a_2, a_3, a_6\}$	*yes*
I_5	$\{a_4, a_5, a_8\}$	*no*

Initial ranking

Patterns $\mathcal{X}_j$
$\mathcal{X}_1 = \{a_1\}$
$\mathcal{X}_2 = \{a_4\}$
$\mathcal{X}_3 = \{a_6\}$
$\mathcal{X}_4 = \{a_2, a_3\}$

Frequent closed patterns

$\mathcal{F} \geq 2$

Scoring Function

$$S(I_i) = |\mathcal{F}(\mathcal{T}, minfr) \subseteq t_i|$$

Set of patterns >= min freq

Minimum frequency

Transaction of the image

$Image\ I_i$	$Trans.\ t_i$	$rel.$
I_1	$\{a_1, a_2, a_3\}$	yes
I_2	$\{a_1, a_4, a_6\}$	yes
I_3	$\{a_1, a_7, a_9\}$	no
I_4	$\{a_2, a_3, a_6\}$	yes
I_5	$\{a_4, a_5, a_8\}$	no

Initial ranking

$Patterns\ \mathcal{X}_j$
$\mathcal{X}_1 = \{a_1\}$
$\mathcal{X}_2 = \{a_4\}$
$\mathcal{X}_3 = \{a_6\}$
$\mathcal{X}_4 = \{a_2, a_3\}$

Frequent closed patterns
$\mathcal{F} \geq 2$

$Image\ I_i$	$\mathcal{X}_j$ in t_i	$\#\mathcal{X}_j$ in t_i	rel.
I_2	$\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3$	3	yes
I_1	$\mathcal{X}_1, \mathcal{X}_4$	2	yes
I_4	$\mathcal{X}_3, \mathcal{X}_4$	2	yes
I_3	$\mathcal{X}_1$	1	no
I_5	$\mathcal{X}_2$	1	no

Re-ranking

An even better scoring function!

- Previous score do not use textual information
- Textual information is embedded into the original ranking
- Weight frequent itemsets (sum of the inverse of the original rank of images containing them)

$$w(\mathcal{X}) = \sum_{k \in K_{\mathcal{T}}(\mathcal{X})} \frac{1}{k}$$

$Image\ I_i$	$Trans.\ t_i$	$rel.$
I_1	$\{a_1, a_2, a_3\}$	yes
I_2	$\{a_1, a_4, a_6\}$	yes
I_3	$\{a_1, a_7, a_9\}$	no
I_4	$\{a_2, a_3, a_6\}$	yes
I_5	$\{a_4, a_5, a_8\}$	no

Example

$$\mathcal{X}_4 = (a_2, a_3)$$

$$w(\mathcal{X}_4) = \frac{1}{1} + \frac{1}{4}$$

An even better scoring function

$$S(I_i) = \sum_{\mathcal{X} \in \{\mathcal{F}(\mathcal{T}, F_{min}) \subseteq t_i\}} w(\mathcal{X})$$

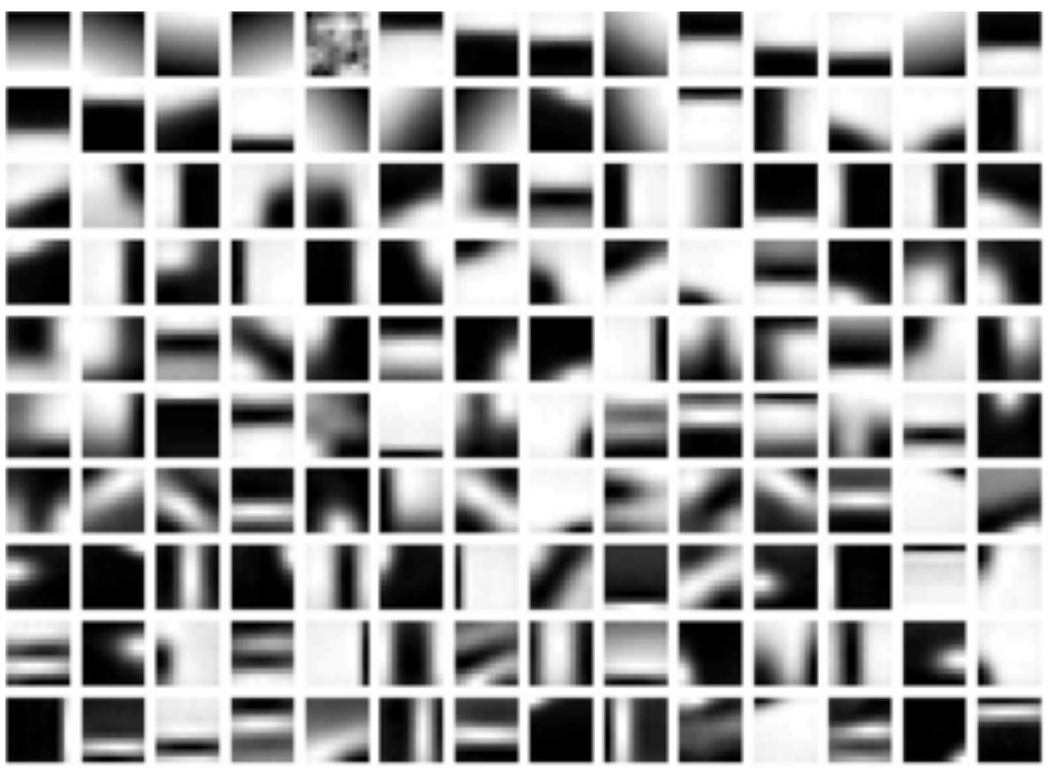
With

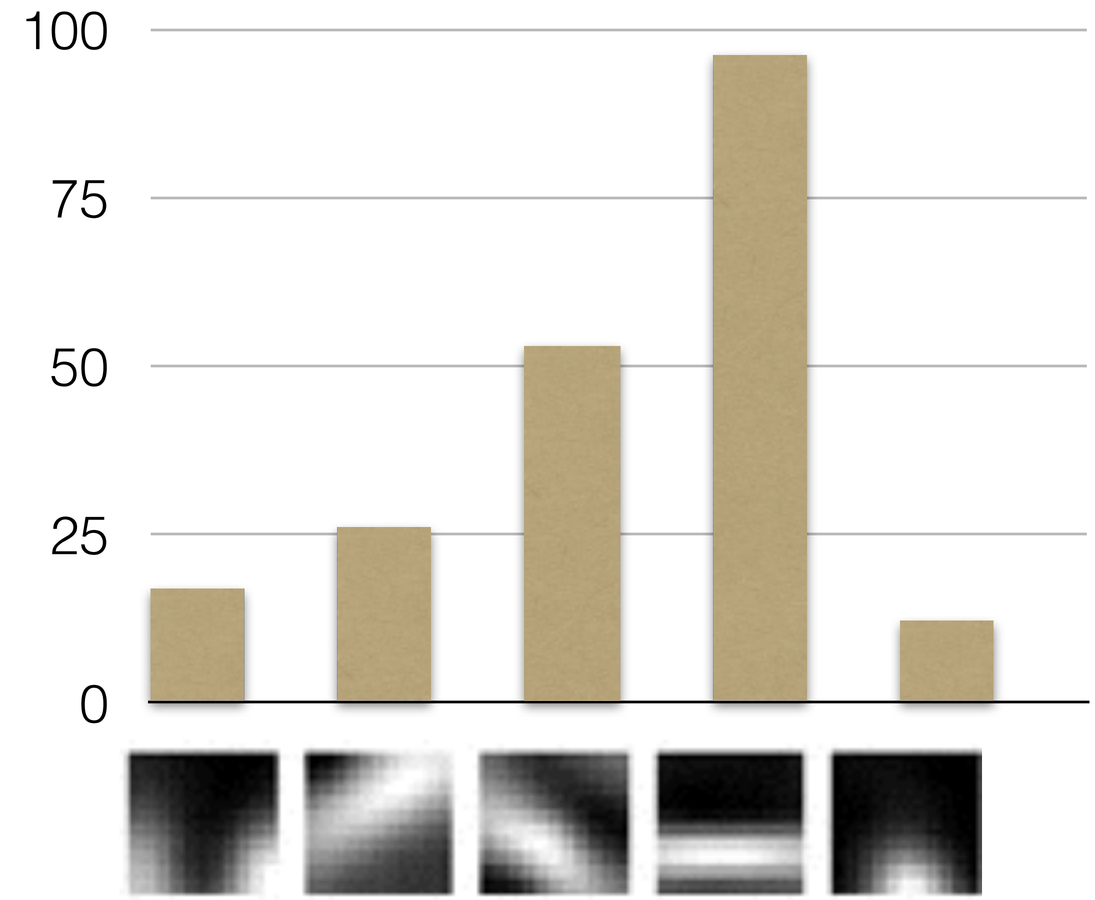
$$w(\mathcal{X}) = \sum_{k \in K_{\mathcal{T}}(\mathcal{X})} \frac{1}{k}$$

In practice, mining complexity is linear with the number of images (but grows exponentially with the number of items)

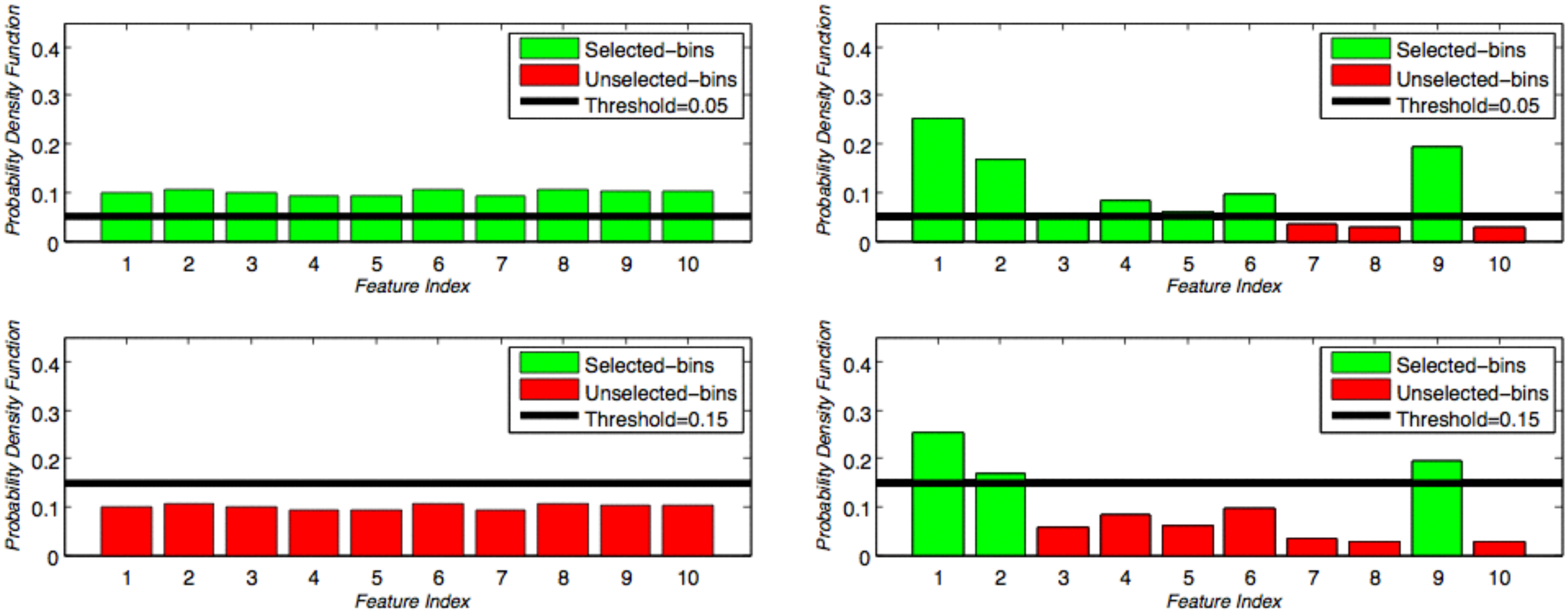
Items are binary elements.
How to produce them from images?

Binarization of bag-of-words representations

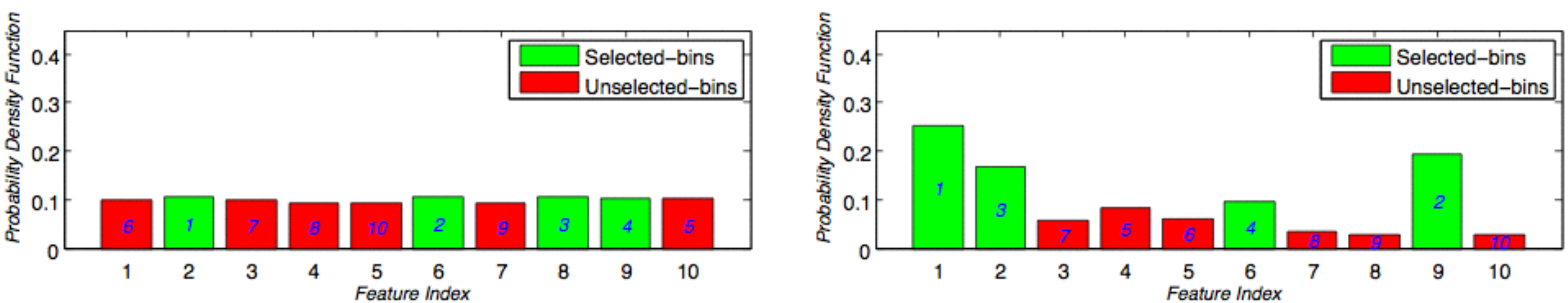




Images are represented as histograms of visual words



(a) Fixed threshold binarization.



(b) Top-K bins binarization.

Mupliple random projections + adaptative thresholding

Vis. words distrib.						After proj. R1				Transactions
A	B	C	D	E	F	A	C	E	F	
.2	.3	.2	.0	.1	.2	.2	.2	.1	.2	{A,C,F}
.0	.3	.4	.1	.1	.1	.0	.4	.1	.1	{C,E,F}
.3	.3	.0	.2	.1	.1	.3	.0	.1	.1	{A,E,F}
.2	.2	.0	.3	.1	.2	.2	.0	.1	.2	{A,E,F}
.0	.3	.1	.1	.4	.1	.0	.1	.4	.1	{C,E,F}
.1	.3	.0	.2	.1	.3	.1	.0	.1	.3	{A,E,F}

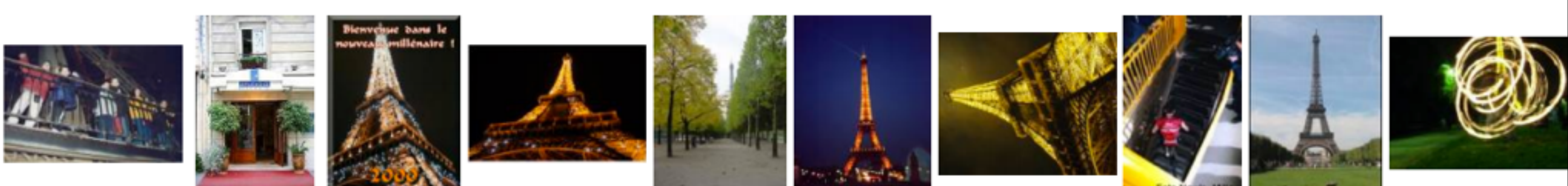
Experimental validation

Experimental validation

- LCM library / minfreq=2
- BOW: VLfeat library
- Visual words are pooled from a 3-level spatial pyramid (1x1,2x2, 4x4)
- 3 different datasets

INRIA Web Queries dataset

- 71,478 images / 353 queries / about 200 images per query
- +/- 40% of the images are relevant to the query
- Diverse queries: general object classes (‘car’, ‘bird’, ‘mountain’, etc.) + specific names of objects, places, events, or persons (‘Nike Logo’, ‘Eiffel tower’, ‘Cannes festival’, ‘Cameron Diaz’,etc.)
- Annotations giving the relevance to each query is provided.
- The evaluation protocol: Average Precision AP is calculated per each query and the mean Average Precision mAP is reported.

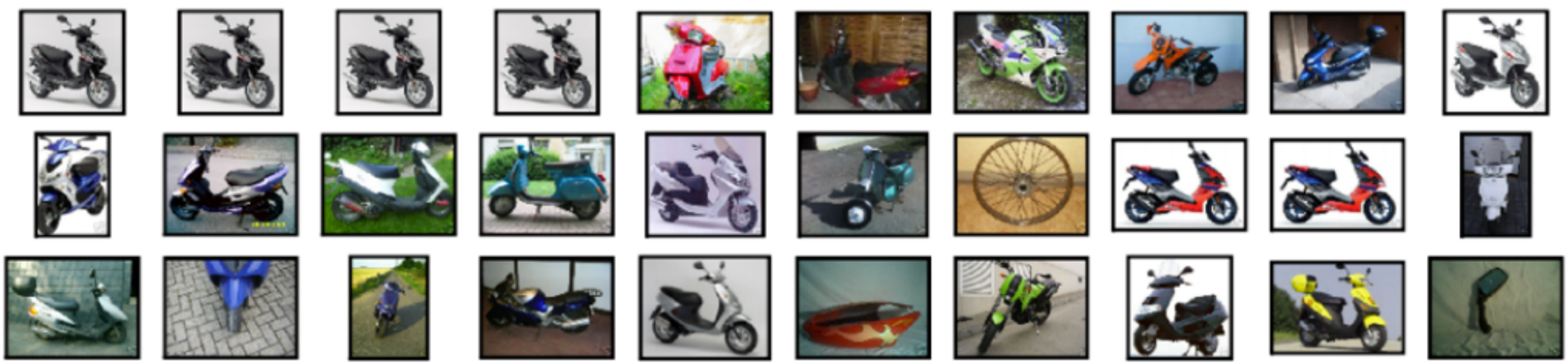


QUAERO’s visual concepts image dataset

- More image (950 images/query)
- More querie (519 queries)

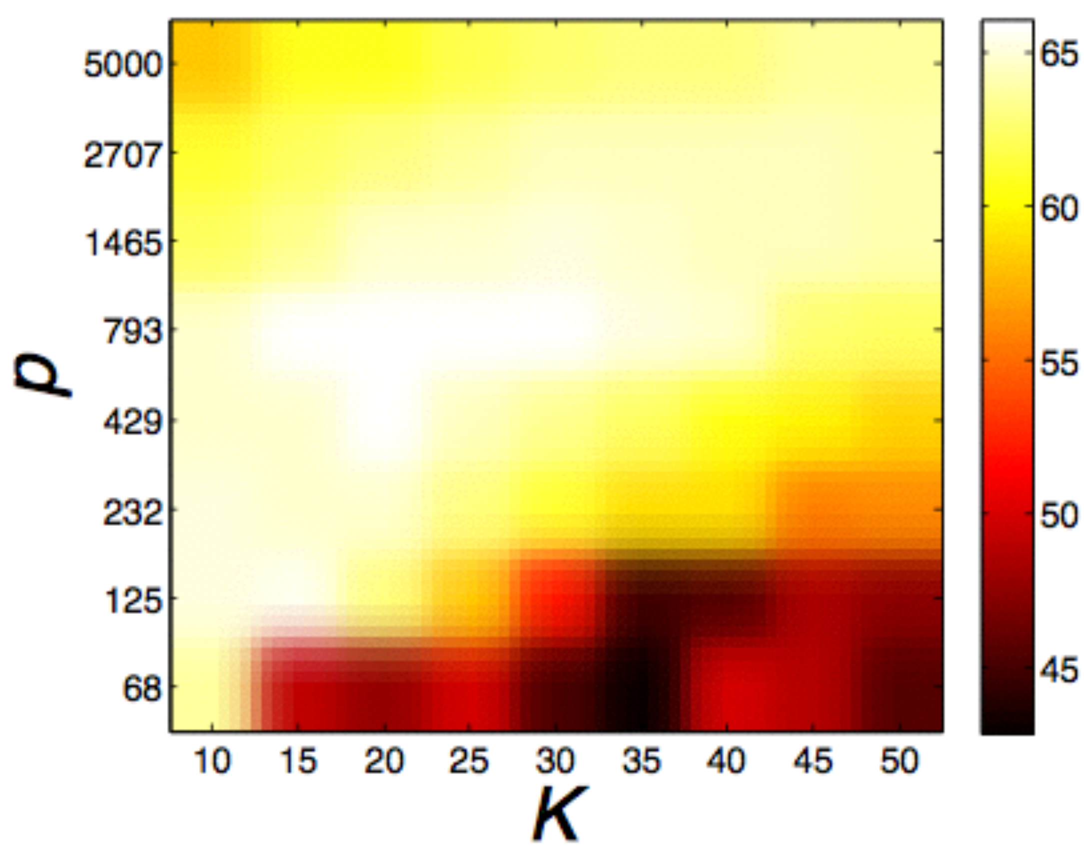
eBay Motorbike dataset

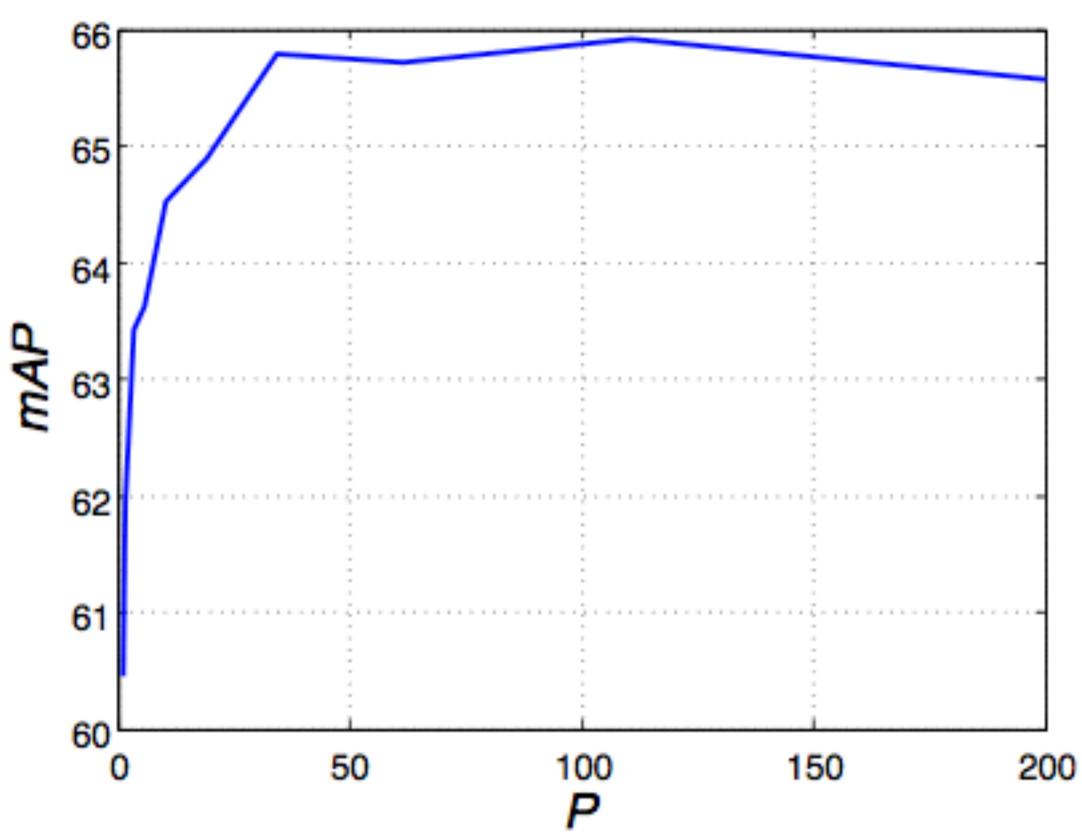
- 5,245 images of different types of motorbikes collected from eBay
- Outliers removal (97% of relevant images)
- EER



Parameters

- **K**: number of items given by the adaptive thresholding (K=20)
- **P**: number of projections (P=20)
- **p**: dimensionality of the projected space (p=800)





Parameters

- K: number of items given by the adaptive thresholding (K=20)
- P: number of projections (P=500)
- p: dimensionality of the projected space (p=800)

Vocab. size	100	1,000	2,000
mAP	62.9	65.7	66.4

Complexity and computational time

Query	Pat. Extract(s)		Scoring(s)		#Pat.	
	Tr.	*Rp.*	*Tr.*	*Rp.*	*Tr.*	*Rp.*
Maradona	0.18	0.10	0.01	0.03	2k	7k
Giraffe	0.30	0.10	0.02	0.04	4k	8k
Times square	0.49	0.14	0.03	0.07	7k	14k
Grand canyon	0.81	0.12	0.05	0.06	10k	11k
Logo Chelsea	3.53	0.13	0.40	0.06	42k	11k
Map World	8.62	0.13	1.03	0.07	100k	12k
Mean 30 queries	*1.58*	*0.11*	0.17	0.04	17k	9k
Std. 30 queries	*6*	*0.02*	0.02	0.02	30k	3k

Alternate mining procedure. Explained in the paper but not in this presentation (less efficient in general)

Method	$mAP(\%)$
Original Search Engine	56.9
Query-ind.+Query-dep. [24]	65.5
LogReg (visual) [15]	64.9
SpecFilter+MRank [17]	**73.8**
Ours	72.2

INRIA Web Queries dataset

Method	$mAP(\%)$
Original ranking	70.4
Query-ind.+Query-dep. [24]	72.7
Ours	**76.1**

QUAERO's visual concepts image dataset

Method	$EER(\%)$
Implicit Shape Model [6]	71.0
FP+SVM [18]	72.6
Ours	**80.0**

eBay Motorbike dataset


Eiffel tower
Original
Ranked
Logo Chelsea
Original
Ranked
Bob Marley
Original
Ranked
Bird
Original
Ranked

Conclusions

- New approach for image re-ranking relying on the effective use of pattern mining techniques.
- Efficient scoring function relying on the hypothesis that non-relevant images are more scattered than relevant ones
- Updates the original scores by measuring the amount of frequent patterns contained in the images.
- How to produce binary items from real valued histograms
- Experimentally validated: in addition of being fast enough for on the fly usage, the approach gives state-of-the-art results on three different challenging datasets.


Eiffel tower
Original
Ranked
Logo Chelsea
Original
RHS Chelsea Flower Show
googolplex
EASTPOINTE
PROJECT
AMERICAN
Life
Ranked
CHELSEA
Bob Marley
Original
Ranked
Bird
Original
Ranked
Pigeon