Improving image processing, enhancement and retrieval using automatic classification

Claudio Cusano

cusano@disco.unimib.it

www.ivl.disco.unimib.it

Why?

- Information about the content of digital images allows processing strategies that can meet the increasing demand for quality,speed and easy of use in imaging applications
  - Optimal color processing in image reproduction
  - Adaptive enhancement (the right algorithm for the right image)
  - Better organization, storage and retrieval of images databases

Text-based retrieval

Google

sunset | Search Images | Search the Web | Advanced Im… | Preferences

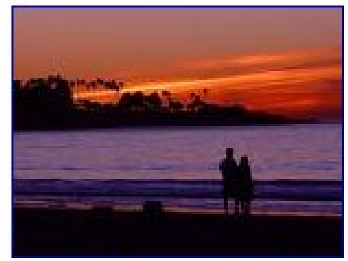
Moderate SafeSearch is on

Images Showing: All image sizes | Any content

Results **1** - **20** of about **39,800,000** for **sunset** [definition]. (**0.07** seconds)

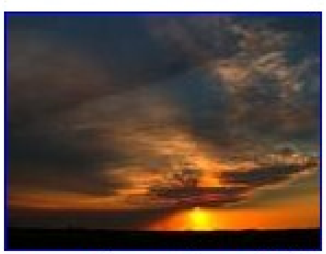
An amazing **sunset** scene. **Sunset**
500 x 375 - 126k - jpg
www.freedigitalphotos.net

A Photo of an amazing **sunset** over ...
1152 x 864 - 114k - jpg
www.travelblog.org
[More from www.travelblog.org]



Sunset in La Jolla.
1024 x 768 - 96k - jpg
www.pdphoto.org
[More from www.pdphoto.org]

Sunset at Casey Station Antarctica
500 x 323 - 27k - jpg
optics.kulgun.net



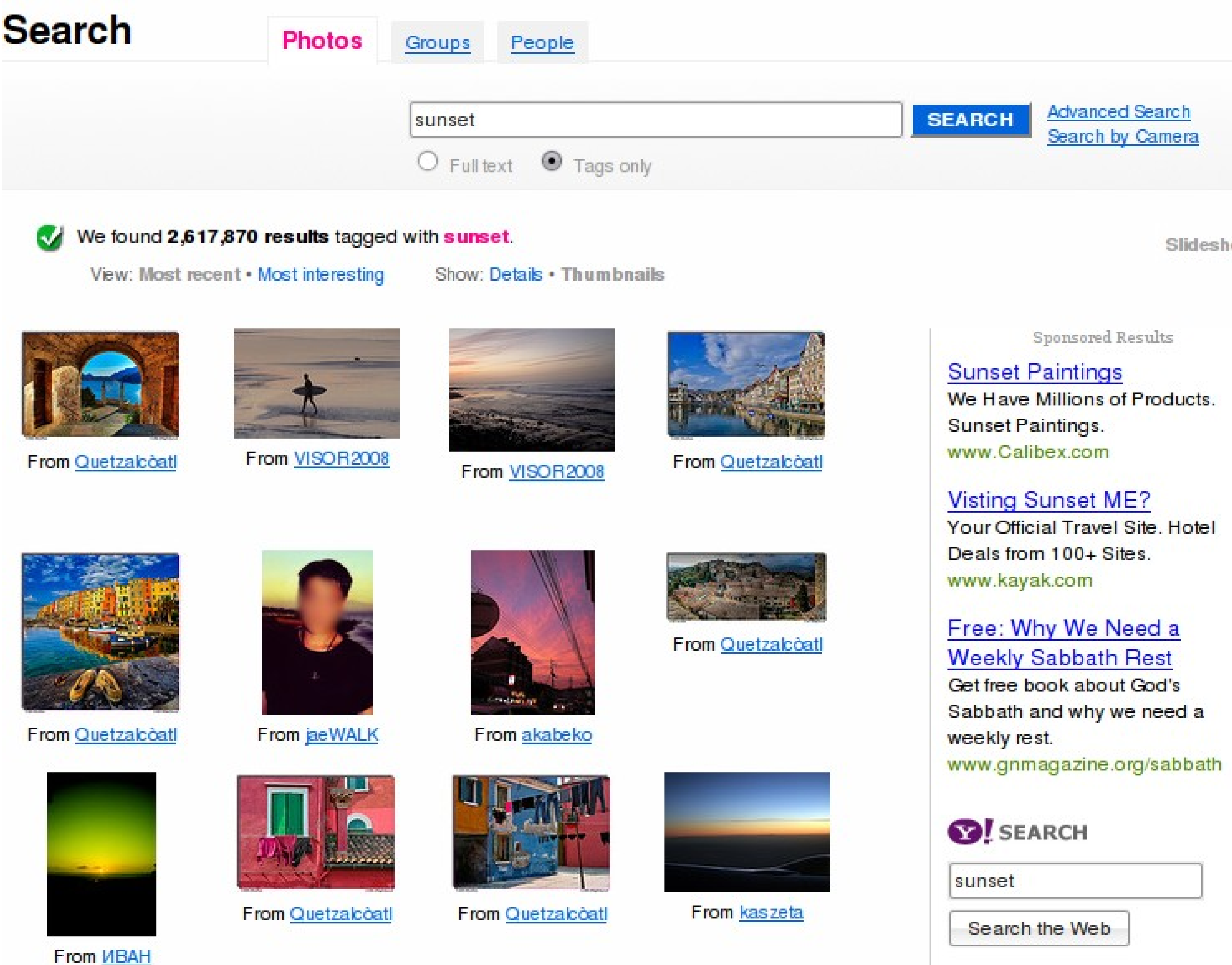
The **sunset** was beautiful as if I can ...
800 x 600 - 71k - jpg
sweetnostalgia.wordpress.com

Sunset sailing.
1024 x 768 - 108k - jpg
www.travelooce.com

Picture of Cross at **Sunset** - Free ...
600 x 400 - 59k - jpg
www.freefoto.com

Seattle **Sunset** background image
1024 x 768 - 110k - jpg
www.zenhaiku.com

Keywords-based retrieval

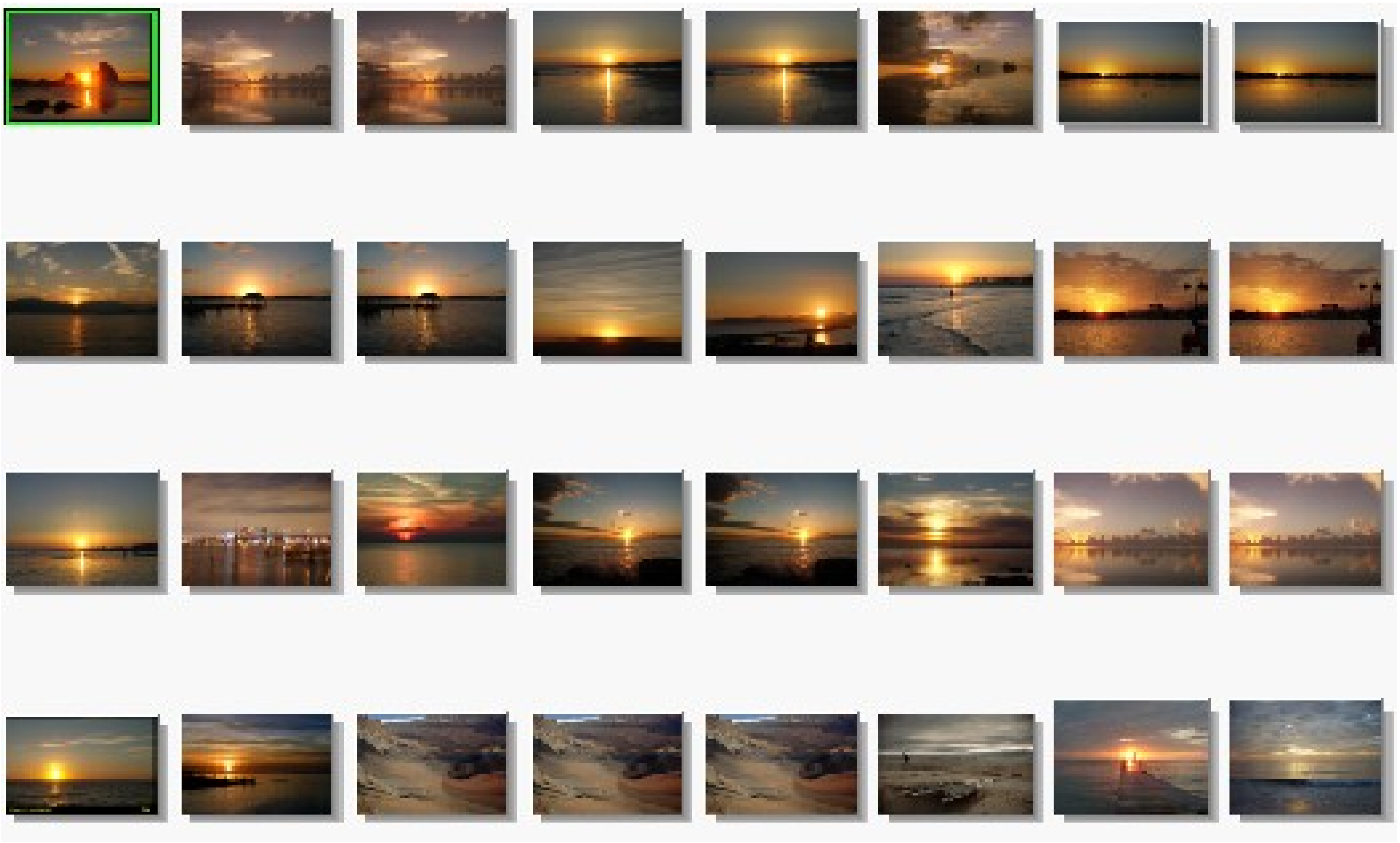


Description by low-level features

- Images are indexed by their by their visual content
  - Color distribution
  - Texture
  - Shape
  - Composition
  - ...
- Low-level features are descriptors that can be reliably computed without any *a priori* knowledge about the content of the images

Retrieval by similarity

Query image



The semantic gap

- Humans tend to use high-level concepts to interpret images
- There is not a direct link between high-level concepts and low-level features

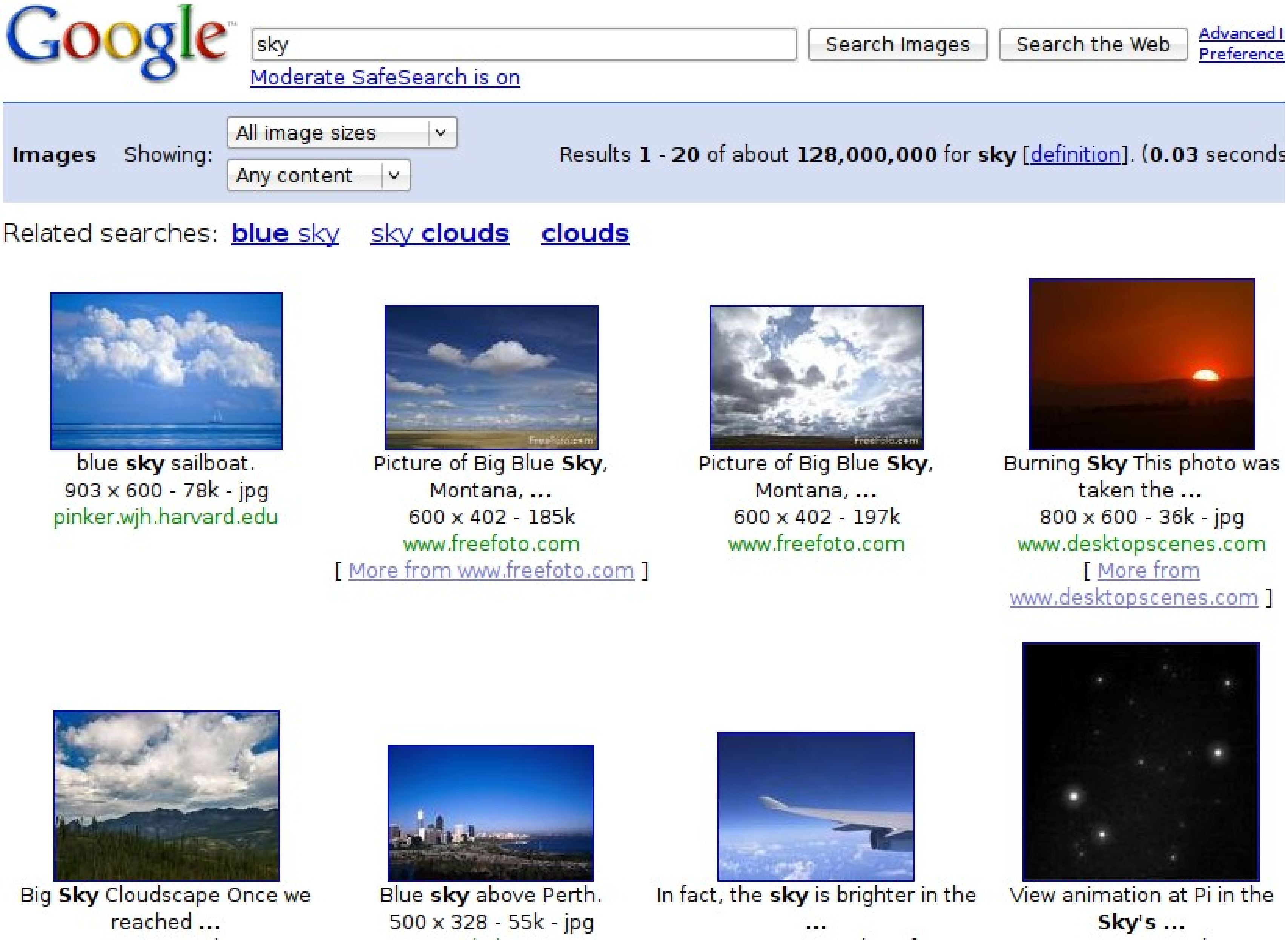
- The performance of state of the art CBIR is far for users' expectations

- Broad domains (web, consumer photographs) are still addressed using text

Bridging the semantic gap

- Interaction with the user (e.g. relevance feedback)
- Data driven: statistical modeling / machine learning
- Knowledge based approaches
  - Sky: “uniform (texture), light blue (color) region in the upper part of the image (composition)”

Bridging the semantic gap

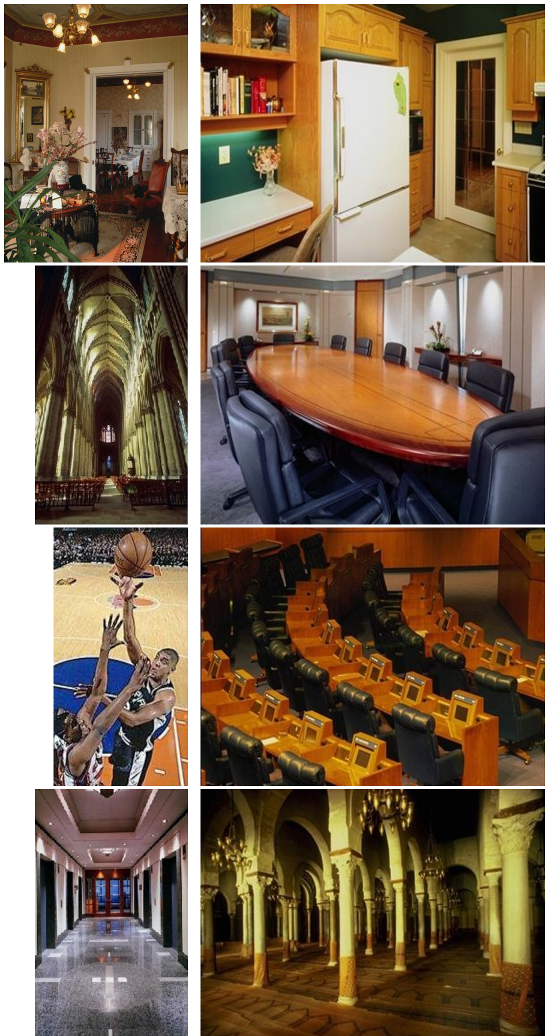


Supervised learning

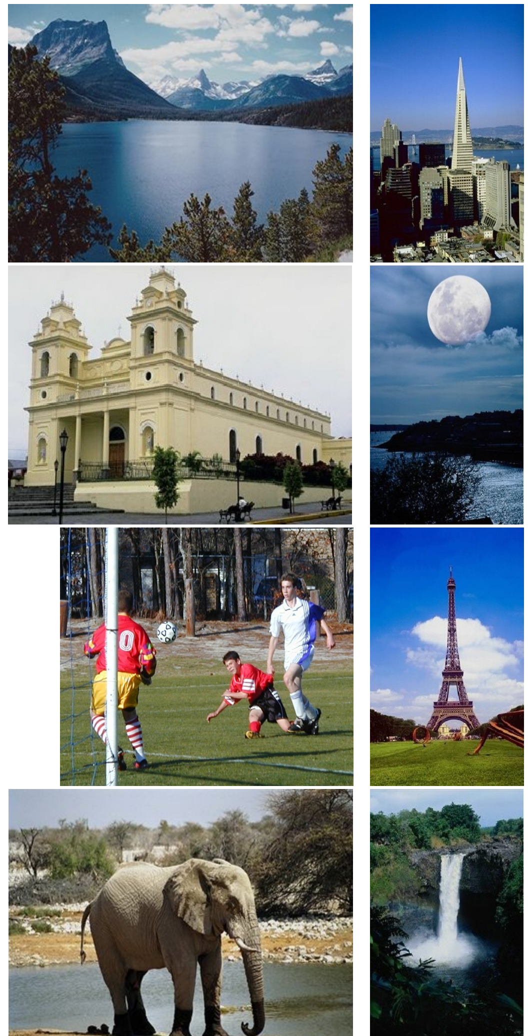
- The link between low-level features and high-level concepts is determined by supervised learning
  - SVMs
  - Neural Networks
  - Decision trees
  - ...
- Concepts (classes) are usually predefined
- High dimensionality → Large training sets
- Non-convex decision surfaces

Indoor/outdoor/close-up classification

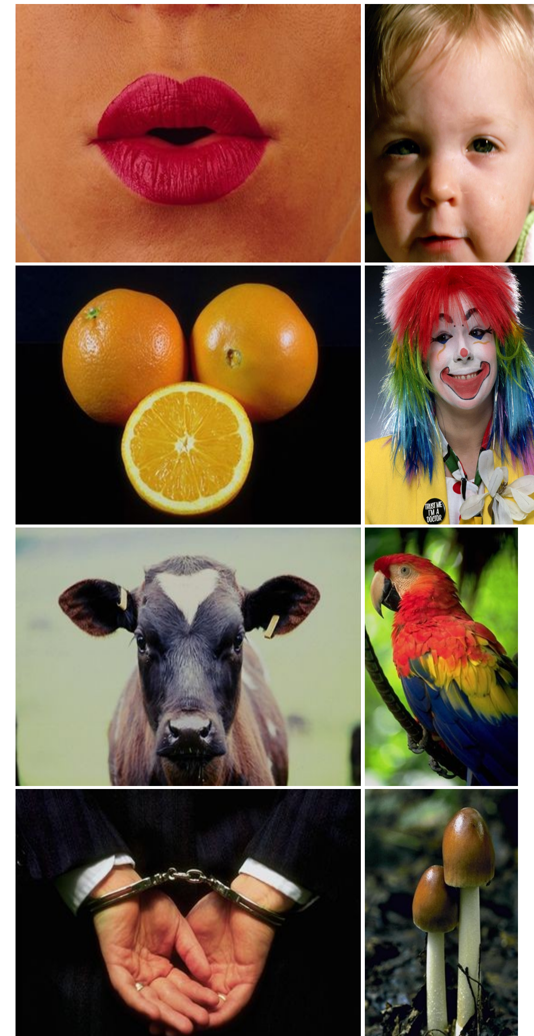
Indoor



Outdoor



Close-up



Indoor/outdoor/close-up classification

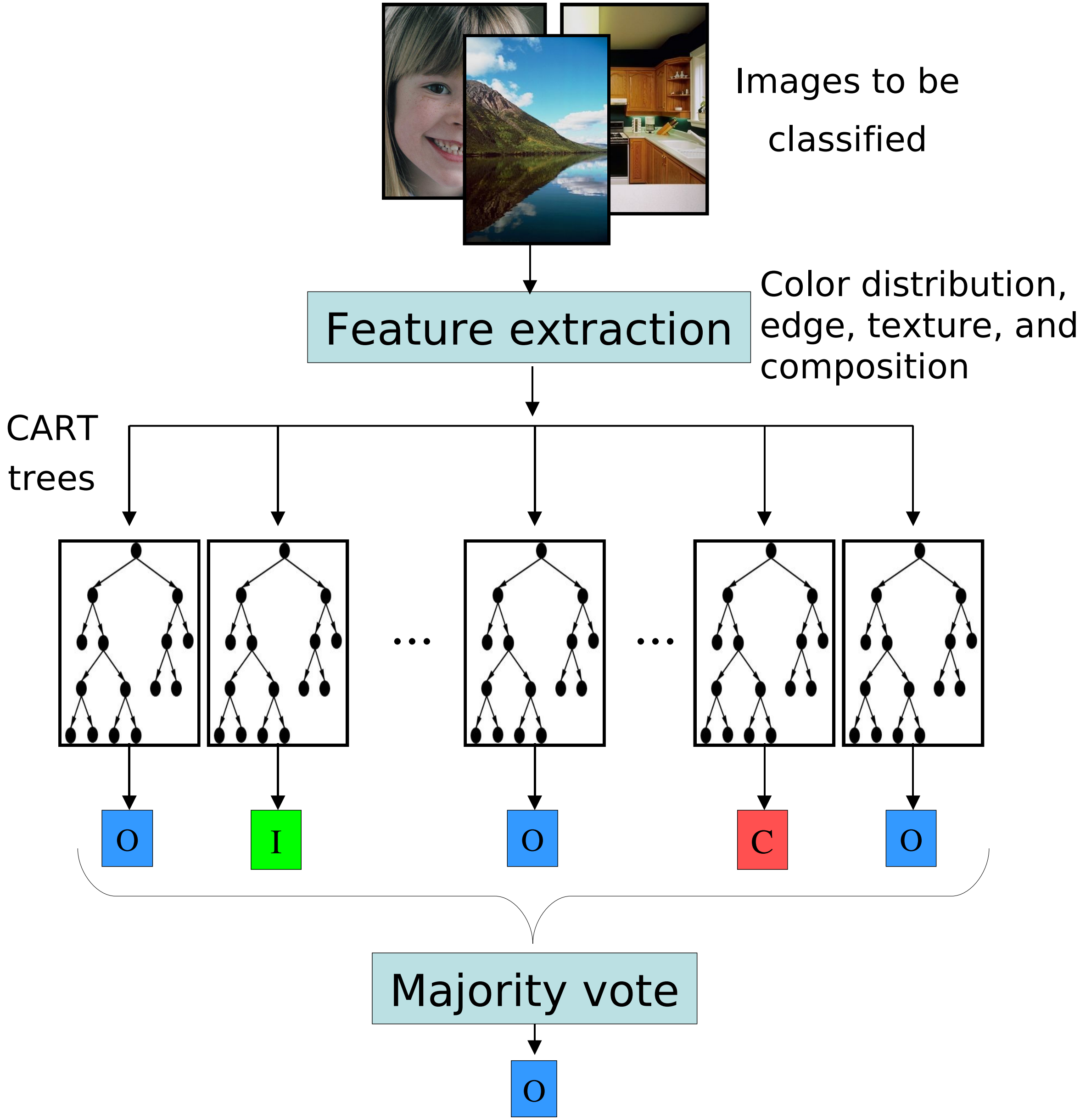


Image description

- Images are described by a set of low-level features related to
  - Color distribution (moments in the YUV color space)
  - Shape (edge direction histogram)
  - Texture (statistics derived from the coefficients of wavelet decomposition)
  - Composition (number of homogeneous regions, dispersion, symmetry ...)
  - ...

Image classification using decision forests

- The decision forest is composed of several decision trees built according to the CART methodology
  - Each tree is trained on a different bootstrap replicate of the training set
- CART is a non-parametric, non-metric classification method
  - no *a priori* assumption about the distribution of the classes in the features space is required
  - feature scaling, and decorrelation are not required

Dataset

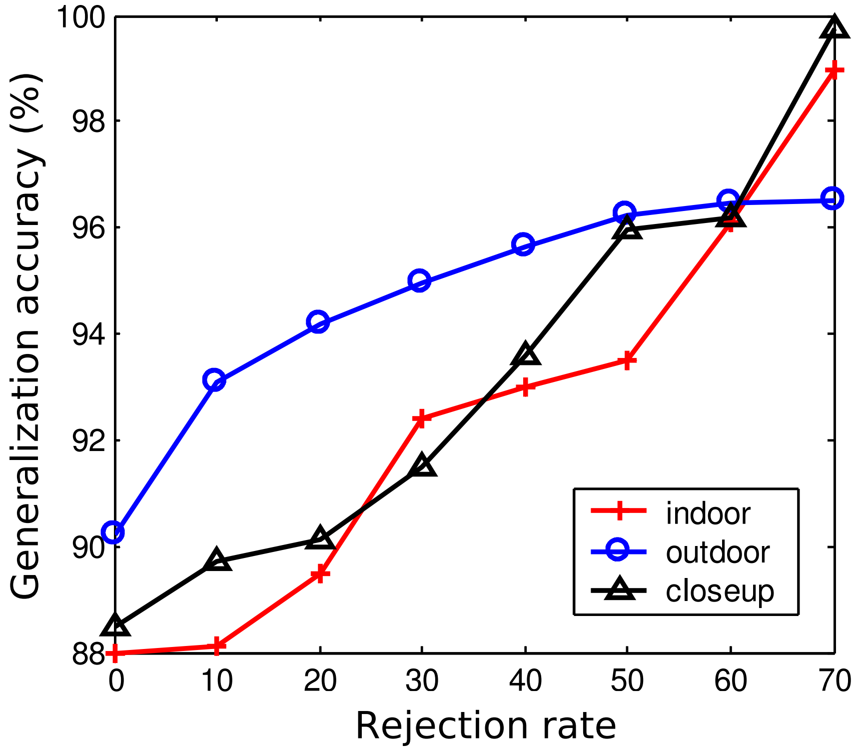
- 9000 images downloaded from the web
- Manually annotated by five users into 2100 indoor 4650 outdoor, and 2250 close-ups (78% of complete agreement)
- 4500 images (1500 per class) randomly selected to form the training set
- The remaining 4500 images have been used for performance evaluation

Experimental results

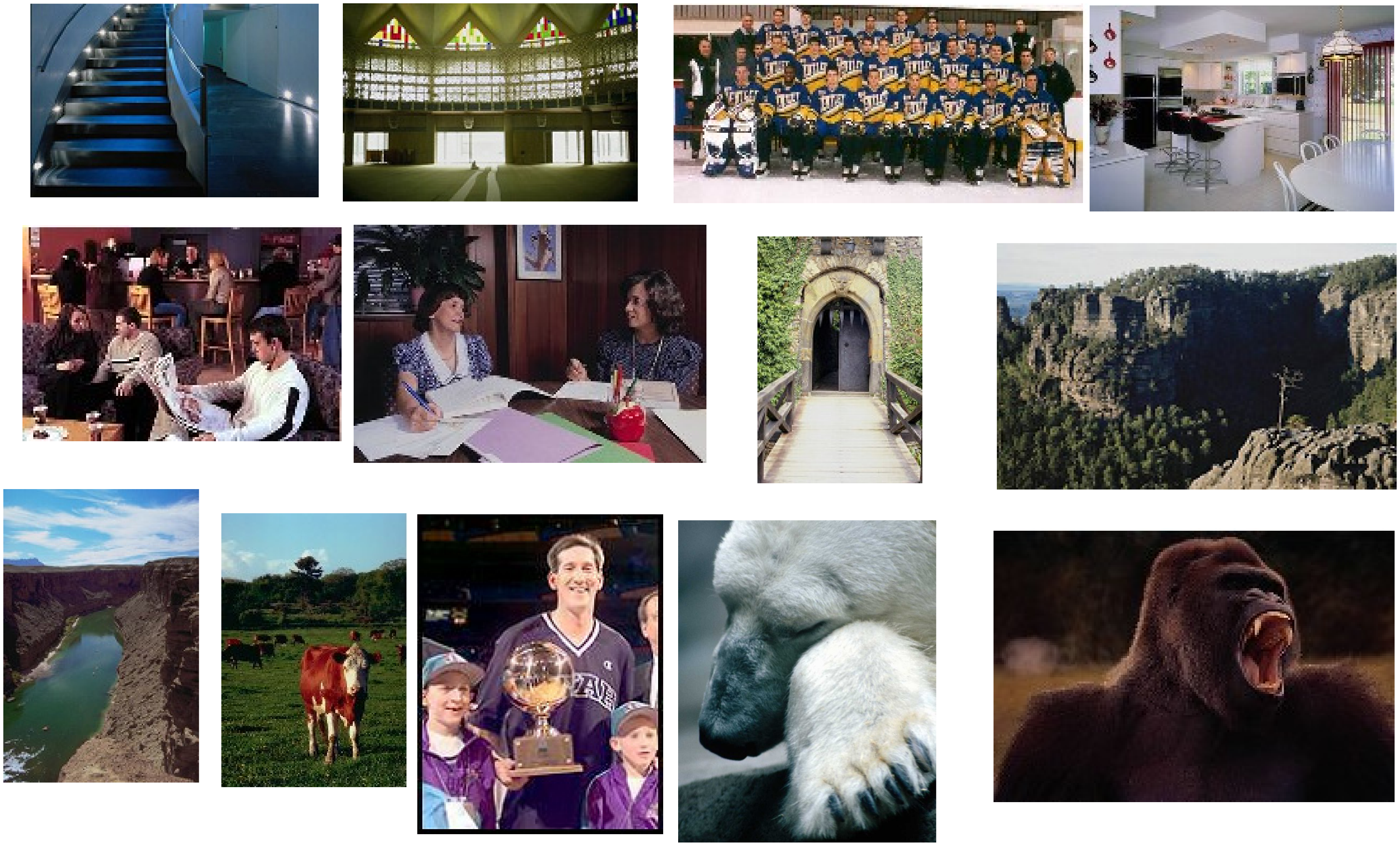
Predicted class

True class	Indoor	Outdoor	Close-up
Indoor	**0.88**	0.05	0.07
Outdoor	0.03	**0.90**	0.07
Close-up	0.05	0.06	**0.89**

Results obtained on the test set



Experimental results



Some misclassified images

Other classification examples

Document classification

Predicted class

True class	Photo	Graphic	Text
Photo	**0.96**	0.03	0.01
Graphic	0.08	**0.89**	0.03
Text	0.01	0.04	**0.95**

Outdoor photographs classification

Predicted class

True class	Day	Sunset	Night
Day	**0.87**	0.05	0.08
Sunset	0.09	**0.82**	0.09
Night	0.10	0.07	**0.83**

Predicted class

True class	Mountain	Sea
Mountain	**0.87**	0.13
Sea	0.18	**0.82**

Predicted class

True class	Urban	Rural
Urban	**0.85**	0.15
Rural	0.16	**0.84**

Examples of applications

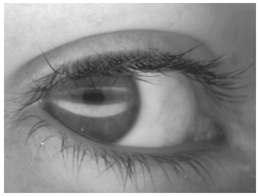
- Image enhancement: apply the right algorithm to the right image (computation color constancy)
- Rendering/visualization (adaptive image cropping)
- Content-based image retrieval

Color constancy on digital devices

The Human Visual System is (almost) able to compensate for illuminants (**color constancy**)

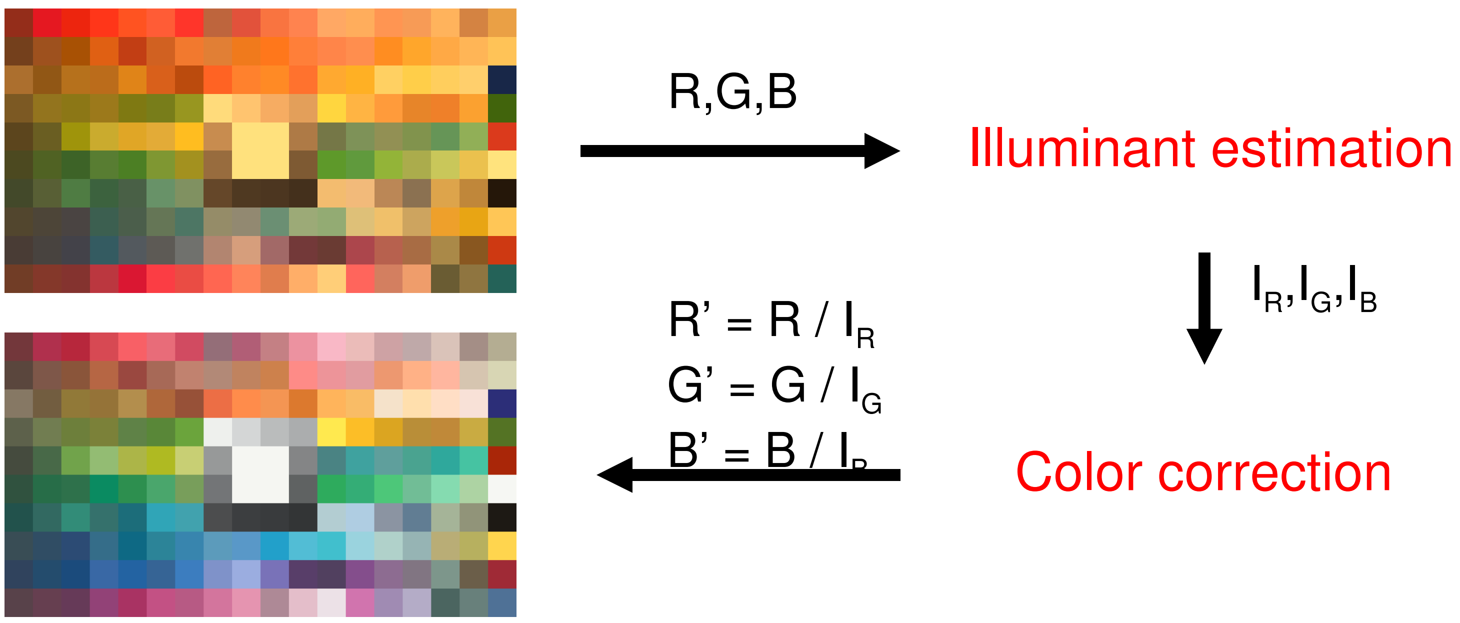
People expect Digital Imaging Acquisition systems to do the same

Computational color constancy tries to emulate this HVS feature on digital devices



Computational color constancy

- A popular approach adopts a two stage procedure:
  - the scene illuminant is estimated from the image data
  - image colors are then corrected on the basis of this estimate



Illuminant estimation

- An ill-posed problem
  - Algorithms usually exploit some assumptions about statistical properties of expected illuminants or of the objects reflectances

Original image

Gray world

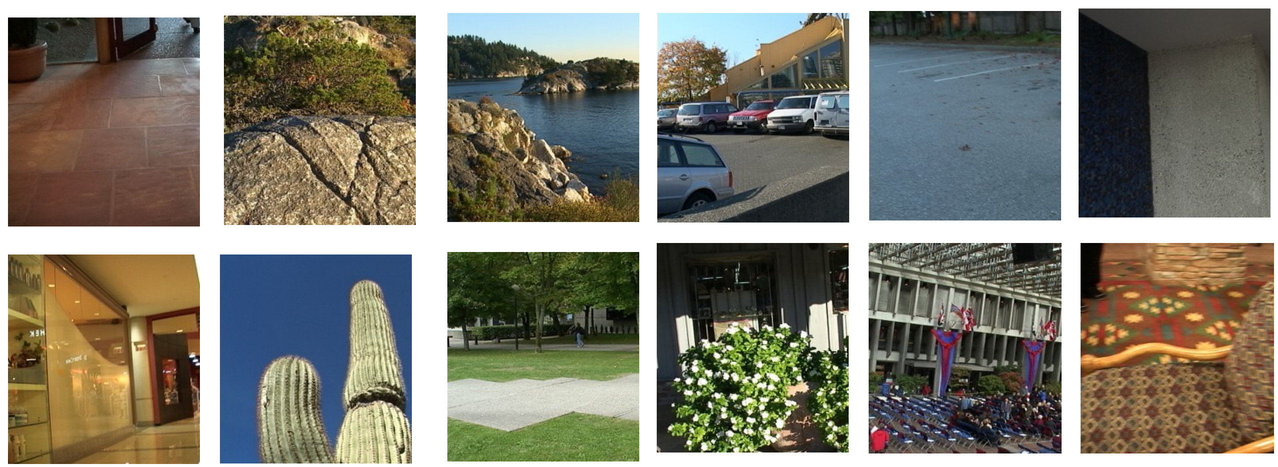
White Point

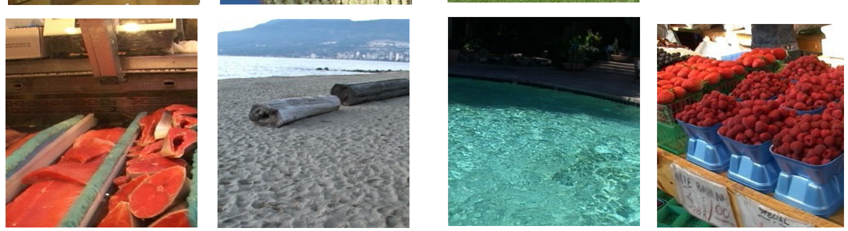
Exploiting classification

- Illuminant estimation can be improved by taking into account information about the content of the images
  - Indoor/outdoor classification has been considered because
    - Indoor/outdoor images present different content
    - Are usually taken under different illumination conditions

Dataset

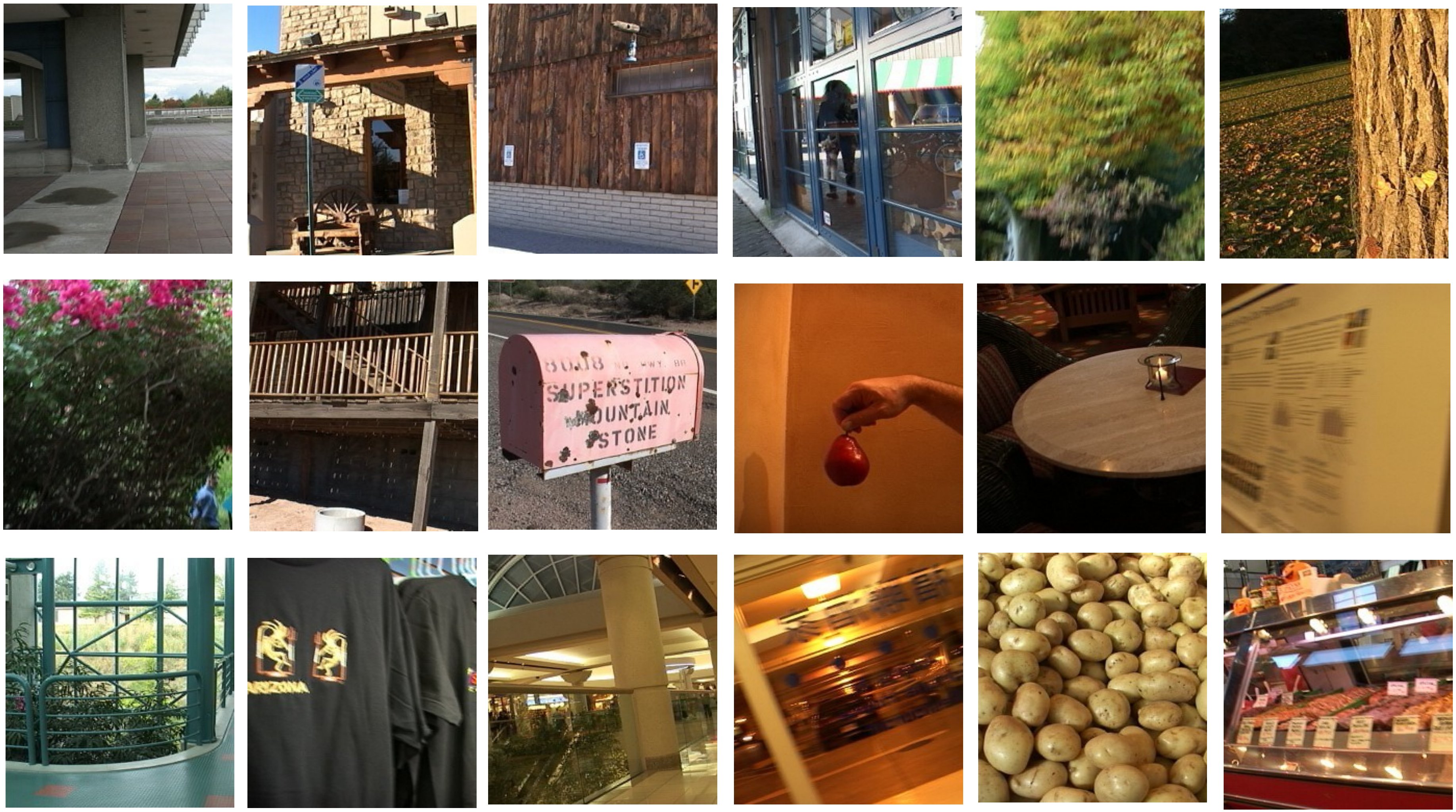
- The Ciurea and Funt image database
  - 15 video clips at 15 fps
  - More than 11000 frames
  - A gray sphere is used to estimate the illuminant color
  - Video summarization techniques are used to select 1135 uncorrelated images





F. Ciurea, B. Funt, “A Large Image Database for Color Constancy Research,” Proc. IS&T/SID 11th Color Imaging Conf., pp. 160-164, 2003.

Some misclassified images



Overall, about 85% of generalization accuracy on the Ciurea & Funt dataset

Color constancy algorithms

- From the framework proposed by Van de Weijer et al.

$$\left(\iint \left| \nabla^n \boldsymbol{\rho}_\sigma(x,y) \right|^p \mathrm{d}x\mathrm{d}y \right)^{\frac{1}{p}} = k\mathbf{I}$$

J. van de Weijer, T. Gevers, A. Gijsenij, "Edge-based Color Constancy," *IEEE Trans. on Image Processing*, 16(9), pp. 2207–2214, 2007

Six, widely used, algorithms have been selected

- Gray World (GW): $n = 0$, $p = 1$, $\sigma = 0$
- White Point (WP): $n = 0$, $p = \infty$, $\sigma = 0$
- Shades of Gray (SG): $n = 0$, $\sigma = 0$
- General Gray World (gGW): $n = 0$
- Gray Edge (GE1): $n = 1$
- Second Order Gray Edge (GE2): $n = 2$

- Some algorithms require a tuning for the parameters p and σ

Color constancy algorithms

- Two combining algorithms have also been included:
  - Least Mean Squares (LMS): the output of the six algorithms are linearly combined (weights need to be estimated)
  - No2Max: the two estimations with the highest distance from the others are discared. The remaining four are averaged (S. Bianco, F. Gasparini, R. Schettini, “A Consensus Based Framework For Illuminant Chromaticity Estimation,” J. of Electronic Imaging, 17, pp. 023013-1–9, 2008)

Experimental results

- A training set of 300 images has been used to determine the optimal parameters of the algorithms (via pattern search)
  - On the two classes
  - On the whole training set

	Indoor Images		Outdoor Images		Whole Training	
	Median	WST	Median	WST	Median	WST
GW	4.91	3	7.86	0	5.62	1
WP	11.83	0	2.81	2	7.76	0
SG	**4.31**	**6**	2.81	2	5.56	1
gGW	4.32	**6**	2.81	2	5.57	1
GE1	5.40	1	3.72	1	5.45	1
GE2	5.57	1	**2.48**	**7**	5.47	1
N2M	5.13	3	2.83	2	5.02	6
LMS	4.58	5	2.71	2	**4.50**	**7**

Experimental results

- On the 835 images of the test set (331 indoor 504 outdoor)
  - Content Independent Strategy (CI): algorithms tuned on the whole training set
  - Content Dependent Parameterization (CDP): class specific parameters, chosen on the basis of the output of the classifier

	CI strategy		CDP Strategy	
	Median	WST	Median	WST
GW	5.95	0	5.95	0
WP	5.48	3	5.48	3
SG	5.80	0	4.08	**4**
gGW	5.80	0	5.39	1
GE1	4.47	5	4.32	3
GE2	4.65	5	**3.94**	**4**
N2M	4.79	4	4.01	**4**
LMS	**4.18**	**7**	4.05	**4**

Experimental results

- Algorithm selection
  - Class-Dependent Algorithms (CDA): for each class the best algorithm (and its corresponding parameters) is selected
  - Class-Dependent Algorithms with Uncertainty Class (CDAUC): introduction of the uncertainty class. Images falling in that class are processed by the algorithm that has proved to be the best class-independent algorithm

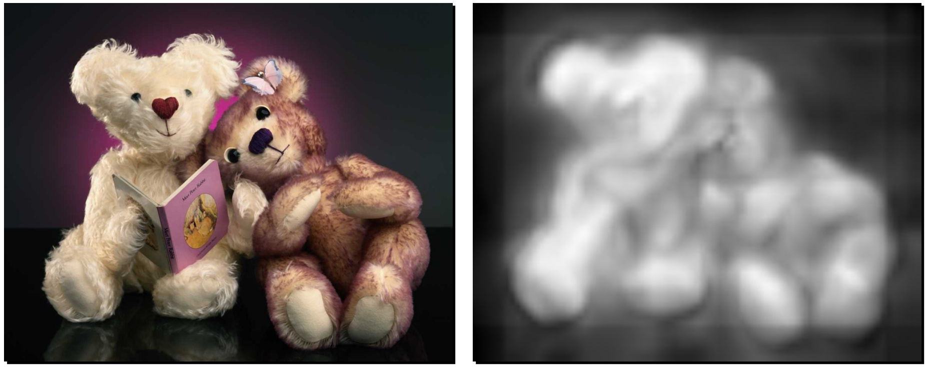
Strategy	Underlying algorithms	Median	WST
CI	LMS, general purpose parameters	4.18	0
CDP	GE2, indoor and outdoor parameters	3.94	1
CDA	SG for indoor and GE2 for outdoor	3.78	1
CDAUC	SG ind., GE2 out., LMS uncertain	**3.54**	**3**

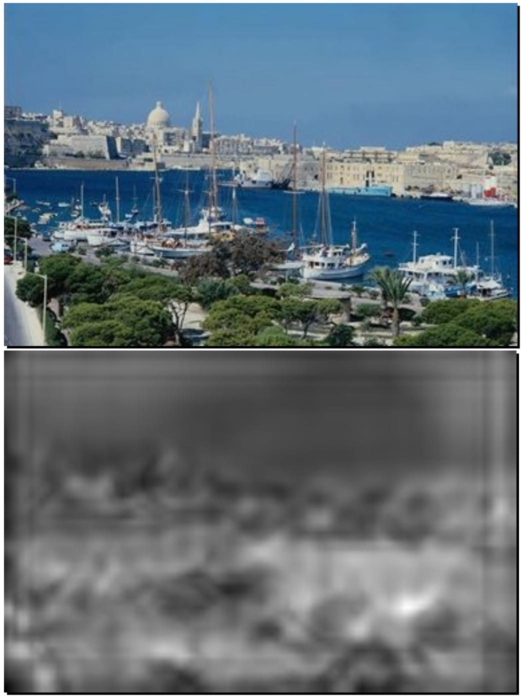
Experimental results

- Image misclassification approximately doubles angular errors
  - 4.85 vs. 9.79 on indoor images
  - 2.31 vs. 5.07 on outdoor images
- How much improvement is expected using a better classifier?
  - We obtained a median angular error of 3.48 degrees using an “optimal” classifier
  - An error of 5.63 has been obtained using a “random” classifier

Adaptive Image Cropping

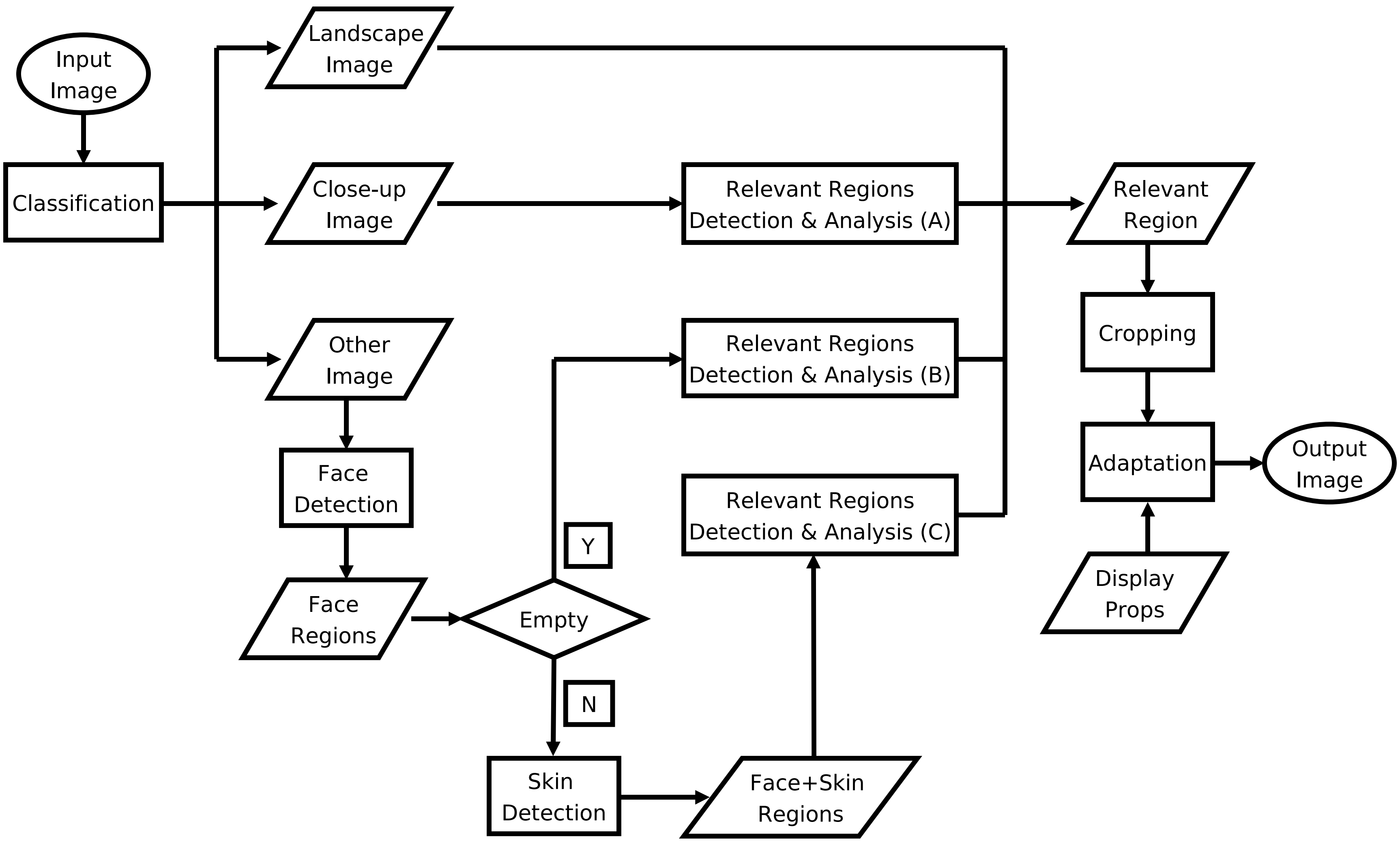
Saliency Regions Detection





L. Itti, C. Koch: “A model of saliency based visual attention of rapid scene analysis.” *IEEE Trans. on Pattern Analysis and Machine Intelligence,* 20, 1254–1259 (1998)

Adaptive Image Cropping

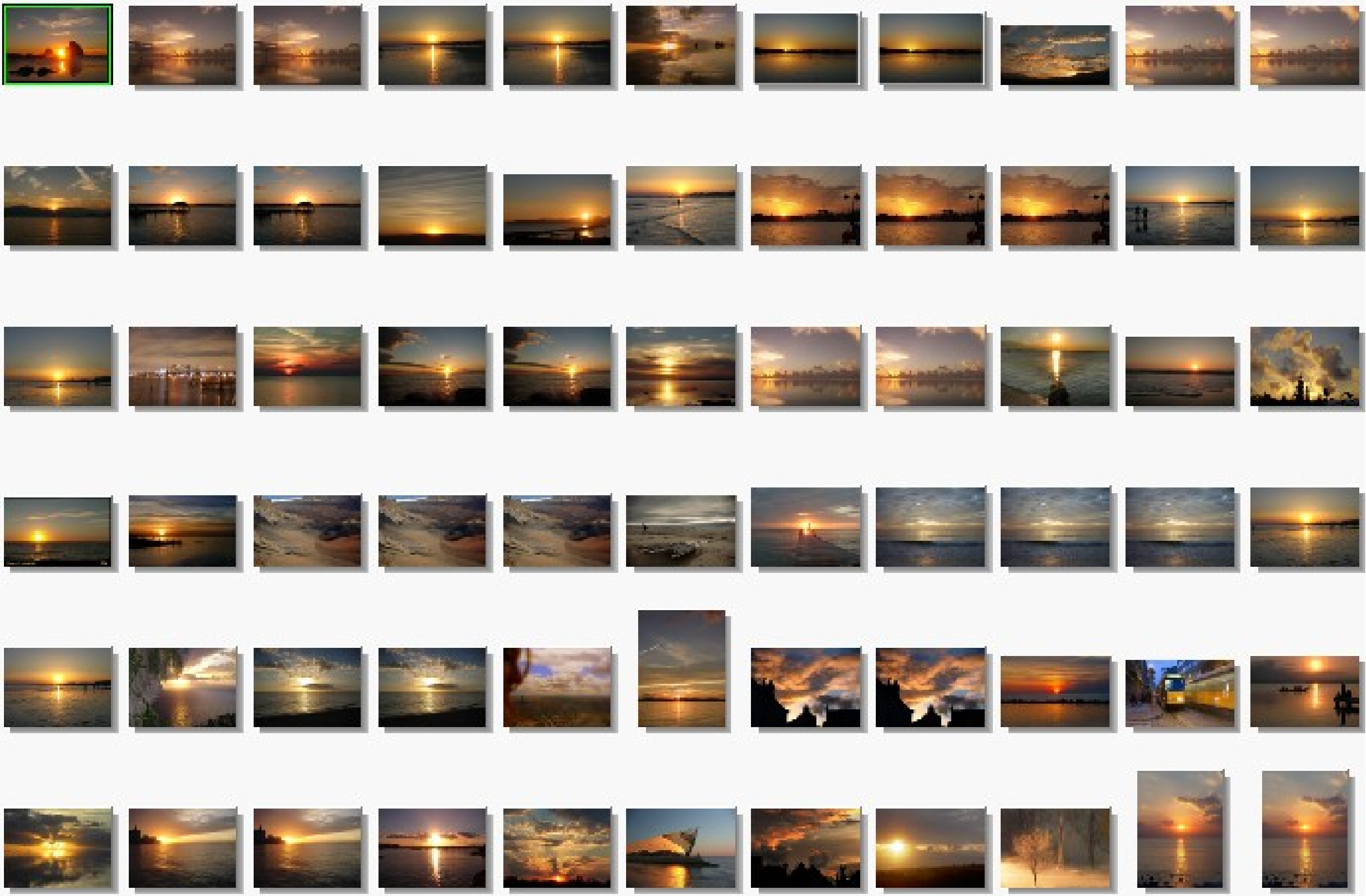


Adaptive Image Cropping

Judgment of five non professional photographers on 300 images:
7% worse than original, **40**% equivalent, **53**% better

Application to CBIR

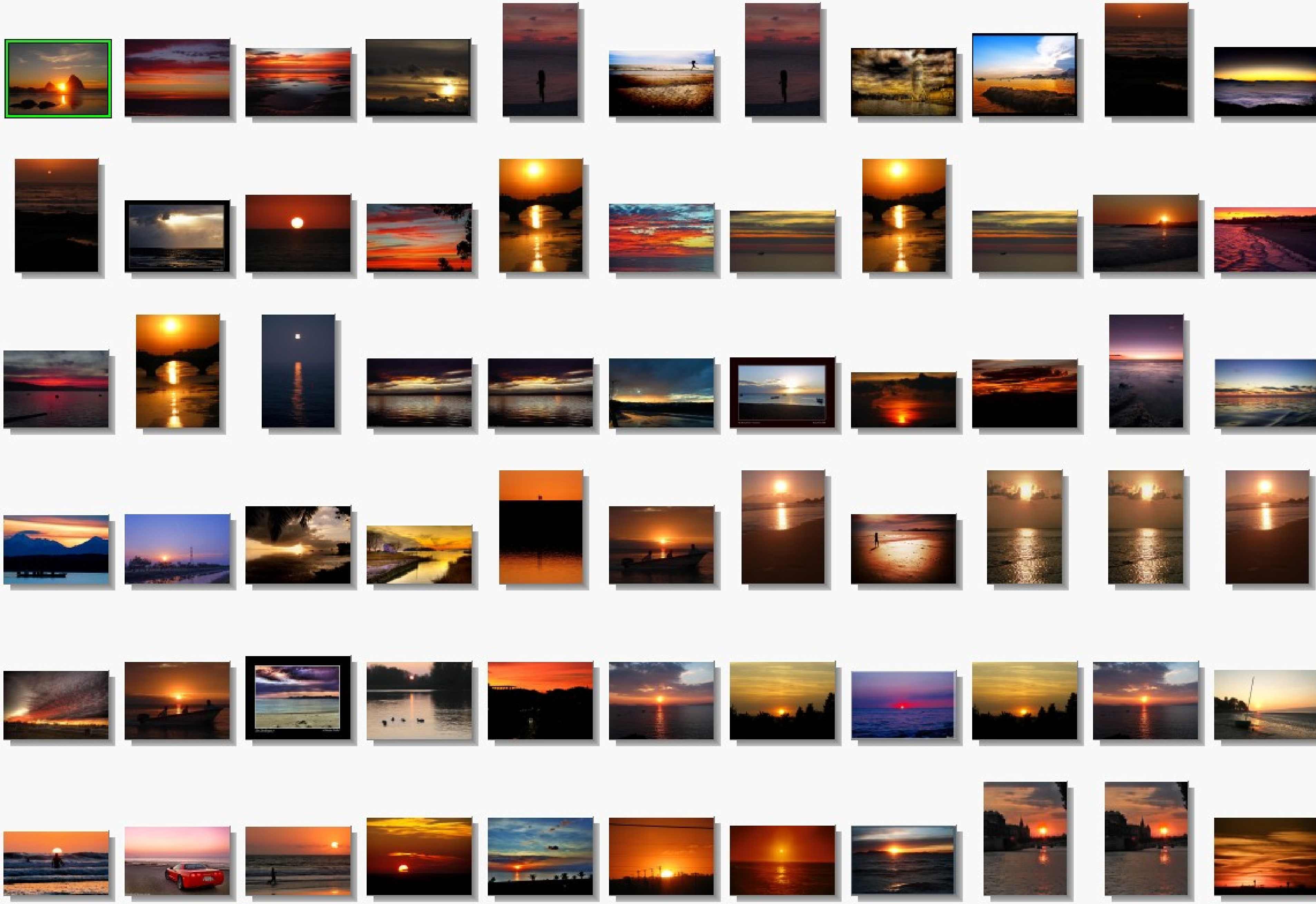
- Relevance feedback is an on-line processing which tries to learn the user's intentions on the fly
- Retrieval by similarity
  - The user selects one or more examples
  - It may also selects negative examples
  - The system updates the weights associated to low-level features and refines the search



G. Ciocca, G., I. Gagliardi, and R. Schettini, “Quicklook2: An Integrated Multimedia System,” Intl. J. of Visual Languages and Computing, **12**, pp. 81-103 (2001)

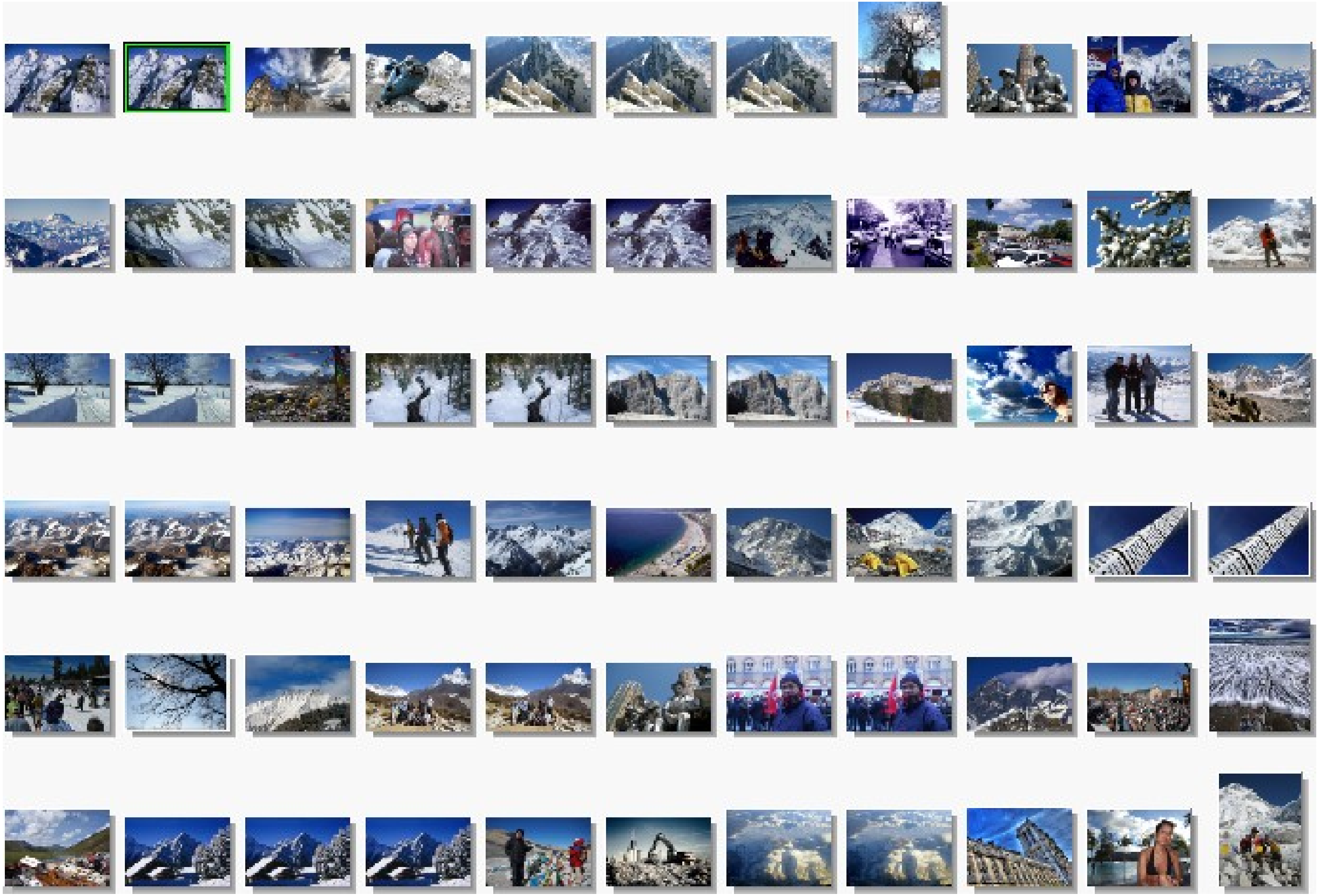
Exploiting classification

- Three classifiers have been considered
  - Day/sunset/night
  - Urban/rural
  - Mountain/sea
- For each image in the dataset (46000 images) a “semantic vector” is formed
- The vector includes a measure of the confidence of the classifiers about each class

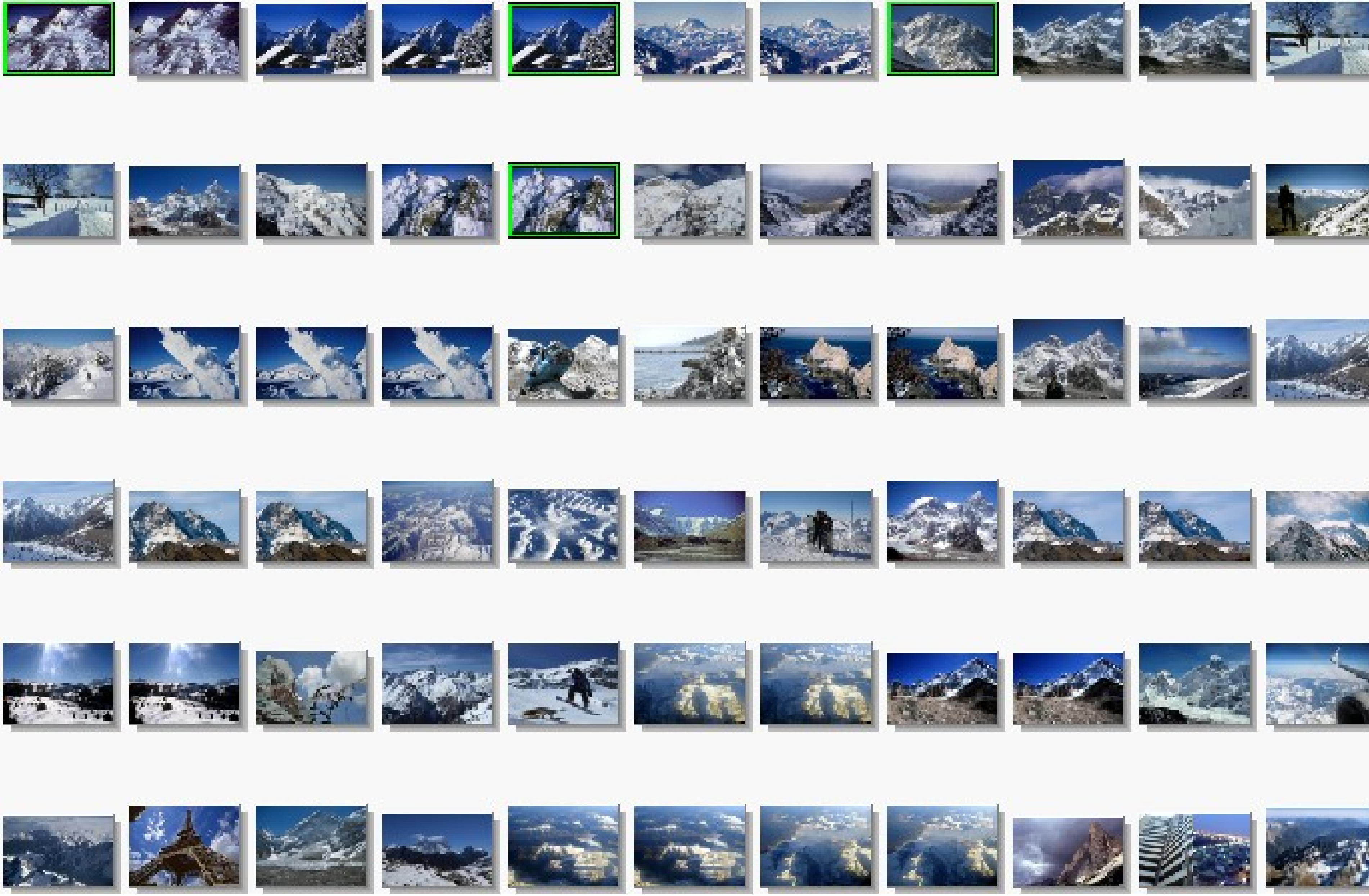


Classification only

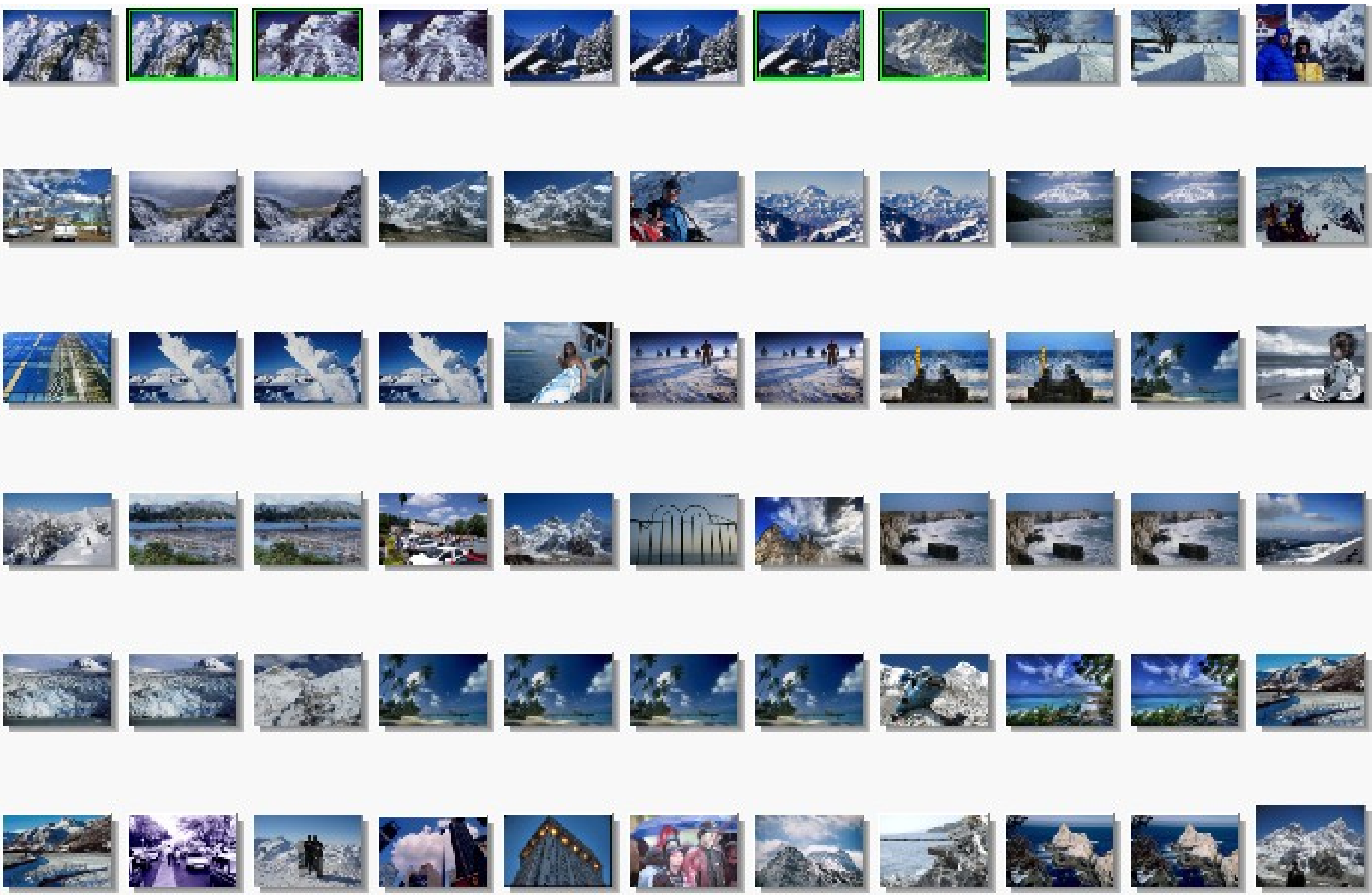
Classification + low-level features



First iteration



Classification + low-level features



Low-level only

Conclusions

- The problem of image classification has yet to be solved
  - But for some applications the accuracy of the methods in the state of the art allows for a significant improvement
- Content-based image retrieval is probably not yet ready for broad domain applications
  - But the performance of hybrid systems seems promising

Related Publications

R. Schettini, C. Brambilla, C. Cusano, and G. Ciocca, Automatic classification of digital photographs based on decision forests. *Int. Journal of Pattern Recognition and Artificial Intelligence* 18(5), 819–845 (2004)

S. Bianco, G. Ciocca, C. Cusano, and R. Schettini. Improving color constancy using indoor-outdoor image classification. *IEEE Trans. on Image Processing* (in print)

G. Ciocca, C. Cusano, F. Gasparini, and R. Schettini. Self-adaptive cropping for small displays. *IEEE Trans. on Consumer Electronics*, 53(4), 1622-1627 (2007)

G. Ciocca, C. Cusano, and R. Schettini. Semantic Classification, Low Level Features and Relevance Feedback for Content-Based Image Retrieval, *Multimedia Processing and Applications, Electronic Imaging, IS&T and SPIE 21st Annual Symposium* (submitted)

G. Ciocca, C. Cusano, F. Gasparini, and R. Schettini. Content-aware image enhancement. In *Proc. Of Artificial Intelligence and Human-Oriented Computing (AI*IA 2007)*, Lecture Notes in Artificial Intelligence, LNAI 4733, 686-697 (2007)

C. Cusano, F. Gasparini, and R. Schettini. Image annotation for adaptive enhancement of uncalibrated color images. In *Proc. of 8th International Conference on VISual information systems*: Lecture Notes in Computer Science, LNCS 3736, 216-255 (2005)

G. Ciocca, C. Cusano, and R. Schettini. Image annotation using SVM. In *Proc Internet imaging V SPIE*, vol. 5304, 330–338 (2004)