Chapter 11
Object and Concept Recognition for Image Retrieval

Stefanie Nowak, Allan Hanbury, and Thomas Deselaers

Abstract ImageCLEF introduced its first automatic annotation task for photos in 2006. The visual object and concept detection task evolved over the years to become an inherent part of the yearly ImageCLEF evaluation cycle with growing interest and participation from the research community. Although the task can be solved purely visually, the incorporation of multi–modal information such as EXIF (Exchangeable Image File Format) data, concept hierarchies or concept relations is supported. In this chapter, the development, goals and achievements of four cycles of object and concept recognition for image retrieval are presented. This includes the task definitions and the participation of the research community. In addition, the approaches applied to solve the tasks and the lessons learnt are outlined. The results of all years are illustrated, compared and the most promising approaches are highlighted. Finally, the interactions with the photo retrieval task are presented.

11.1 Introduction

In 2006, ImageCLEF added an 'Automatic annotation task for general photographs', which over the years evolved from an image classification task into an object retrieval task (2007), and then into a hierarchical concept annotation task (2008–2009). It has developed into an inherent part of the annual ImageCLEF evaluation cycle with interactions with other tasks. As the task names indicate, the focus of the task changed over the years but the objective has always been to analyze the

Stefanie Nowak
Fraunhofer IDMT, Ilmenau, Germany, e-mail: `stefanie.nowak@idmt.fraunhofer.de`

Allan Hanbury
Information Retrieval Facility, Operngasse 20b, 1040 Vienna, Austria e-mail: `a.hanbury@ir-facility.org`

Thomas Deselaers
Computer Vision Laboratory, ETH Zurich, e-mail: `deselaers@vision.ee.ethz.ch`

content of images based on their visual appearance only. Object class recognition, automatic image annotation, and object retrieval are strongly related tasks. In object class recognition, the aim is to identify whether a certain object is contained in an image; in automatic image annotation, a textual description of a given image is created; and in object retrieval, images containing certain objects or object classes have to be retrieved out of a large set of images. Each of these techniques is important to allow for semantic retrieval from image collections.

Evaluation campaigns for object detection (Everingham et al, 2006, 2010), content–based image retrieval (Clough et al, 2005) and image classification (Moellic and Fluhr, 2006) have been established since 2005. Although these evaluation initiatives have a certain overlap with the tasks described in this chapter, ImageCLEF has always focused on multi–modal analysis and the integration of detection technologies into actual retrieval systems. For example, in 2006 and 2007 the generalization of object recognition algorithms across different databases was tested. This scenario denies the often made assumption that for training and testing the same database with similar annotation characteristics is present. In 2008, the participants were provided with a taxonomy and in 2009 with an ontology as additional knowledge sources. These knowledge sources structured the visual concepts into sub and super classes. The ontology also specifies additional relations and restrictions. This textual information was available to enhance the visual analysis algorithms and, for example, to validate the output of the classifiers.

In this chapter, we summarize and analyze the development and the insights gained from four years of object and concept recognition in ImageCLEF. This also allows us to analyze the progress of visual image analysis techniques over these years. This chapter is structured as follows: Section 11.2 introduces the ImageCLEF object and concept retrieval tasks of 2006–2009 in detail and illustrates their aims and the data sets used. Section 11.3 summarizes the approaches of the participants to solve the tasks. Section 11.4 presents the results of the individual years and summarizes the most promising methods. Finally, the combinations of the object retrieval task with the photo retrieval task (Chapter 8) are discussed in Section 11.5, and we conclude in Section 11.6.

11.2 History of the ImageCLEF Object and Concept Recognition Tasks

The first automatic image annotation task was organized in ImageCLEF 2006. A summary of the four cycles of the object and concept recognition tasks from 2006 to 2009 is shown in Table 11.1. The task changed significantly from year to year, which is rather unusual in evaluation campaigns. These changes are manifested in the data sets used (see Chapter 2 for a detailed analysis of the data sets) as well as in the task to be solved by the participants. They reflect the aim to move from a classification task to a full image annotation system that can be combined with other modalities. Every year the task was adapted considering the insights of the

Table 11.1: Summary of the ImageCLEF object and concept recognition tasks characteristics. The table illustrates the type of task, the training and test sets used, the number of images each set contains, the number of visual classes and the number of participants and runs for the years 2006–2009 (OC=Object Categorization, CD=Concept Detection).

Year	Task	Training Set	Num. Images	Test Set	Num. Images	Num. Class	Num. Partic.	Runs
2006	OC	LTU	13,963	Photos	1,100	21	4	10
2007	OC	PASCAL	2,618	IAPR TC–12	20,000	10	7	26
2008	CD	IAPR TC–12 suppl.	1,827	IAPR TC–12 suppl.	1,000	17	11	53
2009	CD	MIR Flickr	5,000	MIR Flickr	13,000	53	19	73

visual tasks of the previous years as well as of the other ImageCLEF tasks. One drawback resulting from these changes is that it is difficult to assess the progress of participating methods over the years.

11.2.1 2006: Object Annotation Task

The *Object Annotation Task* in 2006 (Clough et al, 2007) aimed at the analysis of how purely visual information can be made accessible to text–based searches. The task was designed as a plain classification task to keep the entry barrier low for potential participants. Although the 21 classes were labelled by an object name in English, in fact the task was completely language independent: any other language, or just class numbers, could have been used. A further aim was to investigate how well object categorization algorithms can generalize to images of the same objects that do not necessarily have the same acquisition characteristics. This is a commonly occurring situation in practice, as it is usually not viable to collect a training set large enough to cover all variabilities; however, in other object recognition evaluations this is typically not considered. The training images used were generally clean, containing very little clutter and few obscuring features, while the test images showed objects in a more realistic setting without constraints on acquisition parameters. The training images were taken from a manually collected data set of images in 268 classes kindly provided by LTU technologies[1], from which we selected 21 classes, leading to 13,963 training images. The classes chosen were ashtrays, backpacks, balls, banknotes, benches, books, bottles, calculators, cans, chairs, clocks, coins, computer equipment, cups mugs, hifi equipment, knives forks spoons, mobile phones, plates, sofas, tables, and wallets.

For the test set, 1,100 images of these objects were taken by the organizers. In each test image, at least one object of one of the 21 classes appears, although objects not belonging to any of the classes frequently appear as background clutter.

[1] http://www.ltutech.com/

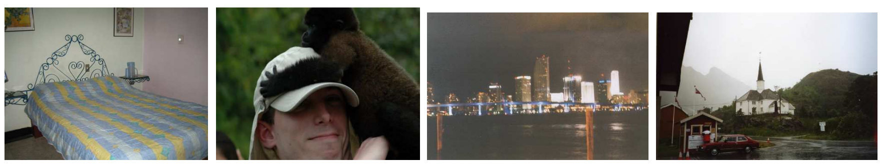
Fig. 11.1: Example images for four of the 21 classes used in the image annotation task in 2006: (left) training set, (right) test set.

The distribution of classes in both training and test sets was non–uniform. Examples of images from the training and test set are shown in Figure 11.1. Along with the training images, 100 of the test images were provided to participants as a development set. The test set was released at a later stage to make training on the testing data difficult. As this was the first time the task was run, and due to its difficulty, only four groups participated, submitting a total of ten runs. For evaluation, the error rate (percentage of incorrectly classified images) was used.

11.2.2 2007: Object Retrieval Task

In the *Object Retrieval Task* in 2007 (Deselaers et al, 2008), the aim was to identify all images showing objects of a certain class. For training, the ‘training and validation set’ of the PASCAL VOC 2006 data set was used (2618 images). Objects in these images are annotated with a class label and bounding boxes, having a total of 4,754 objects in ten classes. For testing, the 20,000 images in the IAPR TC–12 database (Grubinger et al, 2006) were used. Examples of images from the training and test sets are shown in Figure 11.2. The task was formulated as a retrieval task with ten queries corresponding to the ten object classes. The relevance assessments on the IAPR TC–12 were obtained in three ways: 1. Pooling: a Web–interface allowed the relevance of the obtained image categorizations to be manually assessed. These categorizations were obtained by pooling all runs (Braschler and Peters, 2003); 2. Additional relevance judgments: the Web interface also offered the assessors the ability to provide additional information on the objects present in the image. The web interface allowed relevance to be judged rapidly by members of the research groups of the organizers; 3. Manual categorization: Ville Viitaniemi of the Helsinki University of Technology judged all 20,000 images for relevance to the ten queries with stricter definitions of the relevances. Seven groups participated and submitted 26 runs. Performance was measured using Mean Average Precision (MAP).

Fig. 11.2: Example images for four of the ten classes used in the object retrieval task in 2007: bicycle, car, motorbike, person, with (top) PASCAL training set, (bottom) IAPR TC–12 test set. Note the bounding boxes in the training set, and that more than one object can appear in an image.

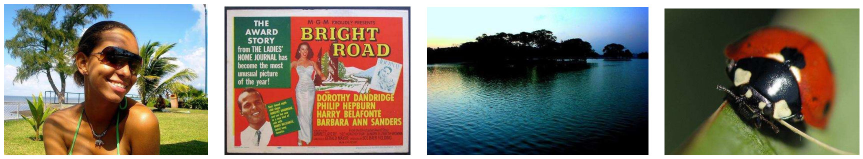


indoor	outdoor, person, day, vegetation, animal	outdoor, night, water, buildings	outdoor, day, road, vegetation, mountains, buildings, sky, overcast

Fig. 11.3: Example images and their concepts from the visual concept detection task in 2008.

11.2.3 2008: Visual Concept Detection Task

In the 2008 *Visual Concept Detection Task* (VCDT) (Deselaers and Hanbury, 2008), the focus was moved from recognizing objects to recognizing concepts, such as indoor/outdoor, day/night, buildings, beach, etc. This is a task that has direct application in annotating images with concepts that are often considered as too obvious to be added to images manually, but have a large potential as useful search terms. 17 hierarchically arranged concepts were chosen. The use of training and test sets with differing characteristics was not continued for the concept detection task. The data set consisted of 2,827 images that were taken from the same pool as those used to create the IAPR TC–12 data set, but were not included in the IAPR TC–12 data set. Example images are shown in Figure 11.3. The data set was divided into 1,827 training images and 1,000 test images. As in 2006, a Web interface was used to annotate the images. Eleven groups participated and submitted 53 runs. The Equal Error Rate (EER) and Area Under Curve (AUC) evaluation measures were used.



Family-Friends, Sky, Summer, Outdoor, Trees, Clouds, Day, Sunny, Portrait, Single Person, Neutral Illumination, No Blur

Canvas, No Blur, Neutral Illumination, Small Group, No Visual Season, No Visual Place, No Visual Time

Landscape, Outdoor, Water, Trees, Sky, Day, Overexposed, No Blur, No Persons, No Visual Season

Plants, Outdoor, Partly Blurred, Macro, Animals, No Visual Time, Neutral Illumination, No Persons, Summer, Aesthetic Impression

Fig. 11.4: Example images from the visual concept detection task in 2009.

11.2.4 2009: Visual Concept Detection Task

In 2009, the Visual Concept Detection Task was carried out at a larger scale (Nowak and Dunker, 2009), with 53 hierarchically organized concepts and a database of 18,000 images from the MIR Flickr 25,000 image data set (Huiskes and Lew, 2008). Examples of the images and concepts are shown in Figure 11.4. The annotation was carried out more carefully and included a validation step as well as a test of inter–annotator agreement. 5,000 images were used for training, and the remaining 13,000 for testing. Participation continued to increase, with 19 groups submitting 73 runs. The EER and AUC evaluation measures were again used, but a new ontology–based measure (OS) (Nowak and Lukashevich, 2009) was also introduced.

11.3 Approaches to Object Recognition

Over the four years, 29 research groups participated in total. Of these, nine research groups participated in the task several times. The participation of the groups is summarized in Table 11.2. For readability, all participating groups are listed together with the group acronyms and the citations of their approaches as follows:

- **apexlab** (Nowak and Dunker, 2009): Shanghai Jiaotong University, Shanghai, China;
- **AVEIR** (Glotin et al, 2009): joint consortium of the four groups: Telecom ParisTech, LSIS, MRIM–LIG and UPMC;
- **budapest / sztaki** (Deselaers et al, 2008; Daróczy et al, 2008, 2009): Data Mining and Web search Research Group, Informatics Laboratory, Computer and Automation Research Institute, Hungarian Academy of Sciences, Hungary;
- **CEA LIST** (Deselaers and Hanbury, 2008; Nowak and Dunker, 2009): Lab of Applied Research on Software–Intensive Technologies of the CEA, France;
- **CINDI** (Clough et al, 2007): Concordia University in Montreal, Canada;

- **DEU** (Clough et al, 2007): Department of Computer Engineering of the Dokuz Eylul University in Tinaztepe, Turkey;
- **FIRST** (Binder and Kawanabe, 2009): Fraunhofer FIRST, Berlin, Germany;
- **HJFA** (Jiang et al, 2008): Microsoft Key Laboratory of Multimedia Computing and Communication of the University of Science and Technology, China;
- **HUTCIS** (Deselaers et al, 2008): Adaptive Informatics Research Centre / Laboratory of Computer and Information Science, Helsinki University of Technology, Finland;
- **I2R** (Deselaers and Hanbury, 2008; Ngiam and Goh, 2009): IPAL French–Singaporean Joint Lab of the Institute for Infocomm Research, Singapore;
- **IAM** (Hare and Lewis, 2009): Intelligence Agents Multimedia Group of the University Southampton, UK;
- **INAOE TIA** (Deselaers et al, 2008; Deselaers and Hanbury, 2008; Escalante et al, 2009): TIA Research Group, Computer Science Department, National Institute of Astrophysics, Optics and Electronics, Tonantzintla, Mexico;
- **ISIS** (van de Sande et al, 2009)]: Intelligent Systems Lab of the University of Amsterdam, The Netherlands;
- **Kameyama** (Sarin and Kameyama, 2009): Graduate School of Global Information and Telecommunication Studies, Waseda University, Japan;
- **LEAR** (Douze et al, 2009): LEAR team of INRIA, Montbonnot, France;
- **LSIS** (Zhao and Glotin, 2008; Dumont et al, 2009): Laboratory of Information Science and Systems, France;
- **Makere** (Deselaers and Hanbury, 2008): Faculty of Computing and Information Technology, Makerere University, Uganda;
- **MedGIFT** (Clough et al, 2007): University and Hospitals of Geneva, Switzerland;
- **MMIS** (Llorente et al, 2008, 2009): Knowledge Media Institute, Open University, Milton Keynes, UK;
- **MRIM-LIG** (Pham et al, 2009): Multimedia Information Modelling and Retrieval group at the Laboratoire Informatique de Grenoble, Grenoble University, France;
- **MSRA** (Deselaers et al, 2008): Microsoft Research Asia;
- **NTU** (Deselaers et al, 2008): School of Computer Engineering, Nanyang Technological University, Singapore;
- **PRIP** (Deselaers et al, 2008): Institute of Computer–Aided Automation, Vienna University of Technology, Vienna, Austria; Intelligent Systems Lab Amsterdam, University of Amsterdam, The Netherlands;
- **RWTH** (Clough et al, 2007; Deselaers et al, 2008; Deselaers and Hanbury, 2008): Human Language Technology and Pattern Recognition Group from the RWTH Aachen University, Germany;
- **Telecom ParisTech** (Ferecatu and Sahbi, 2009): Institut TELECOM, TELECOM ParisTech, Paris, France;
- **UAIC** (Iftene et al, 2009): Faculty of Computer Science of Alexandru Ioan Cuza University, Romania;

Table 11.2: Participation in the object retrieval task over the years. The rows denote in which year the single groups participated and the number illustrates the number of run configurations that were submitted. Please note that in 2009 the maximum number of runs was restricted to five.

Group	CINDI	DEU	RWTH	MedGIFT	budapest / sztaki	HUTCIS	INAOE TIA	MSRA	NTU	PRIP	CEA LIST	HJFA	I2R	LSIS	MMIS	Makere	UPMC	XRCE	apexlab	AVEIR	FIRST	IAM	LEAR	ISIS	Kameyama	MRIM-LIG	Telecom ParisTech	UAIC	Wroclaw Uni
2006	4	2	2	3																									
2007			1		2	13	4	3	1	2																			
2008			1		13		7				3	1	8	7	4	1	6	2											
2009					5		5				4		2	5	5		5	1	3	4	4	3	5	5	5	4	2	1	5

- **UPMC** (Tollari et al, 2008; Fakeri-Tabrizi et al, 2009): University Pierre et Marie Curie in Paris, France;
- **Wroclaw Uni** (Nowak and Dunker, 2009): Wroclaw University of Technology, Poland;
- **XRCE** (Ah-Pine et al, 2008, 2009): Textual and Visual Pattern Analysis group from the Xerox Research Center Europe, France;

In the following, we outline commonly used techniques to solve the object retrieval and detection tasks. To this end, all 162 submissions from 29 research groups and 17 countries are analyzed. The submitted approaches are categorized regarding descriptors, codebook generation, classification methods, and post–processing.

11.3.1 Descriptors

A large variety of visual descriptors was used throughout the four cycles of this ImageCLEF task. Most groups apply combinations of descriptors. One broad distinction is whether a descriptor describes an image as a whole (global features) or only a region of the image (local features). Among the local features, different sizes of the described regions are considered: some descriptors only consider small square image regions while others consider large portions of the image. Furthermore, the positions from which local features are extracted vary widely. Local descriptors can, for example, be extracted from sparse interest points or from a dense grid. Frequently, a set of local features is further represented as a histogram over a visual codebook (e.g. ISIS, RWTH, IAM, LSIS, see Section 11.3.2).

Some groups also extract local features from image regions that were obtained by an unsupervised image segmentation engine. Here, often the entire image is covered with regions in a jigsaw–like way [budapest, INAOE TIA].

Among the global features many groups used color histograms [DEU, CINDI, HUTCIS, Makere, CEA–LIST, etc.] amongst other features or texture features such as edge histograms, Tamura histograms [MMIS] or Gabor features [LSIS, NTU]. Also the Gabor–based GIST features [kameyamalab, INRIA-LEAR, apexlab] and profile entropy features [LSIS] were applied.

Bag–of–words representations of SIFT, color–SIFT or image patches were common among the local features [ISIS, FIRST, MSRA, PRIP, IAM, HUTCIS, HJ-FA]. These features were often extracted from Harris-Laplace interest points or using a dense grid. Region-based local features also allow for using shape [Makere, budapest, INAOE TIA] and spatial layout [CEA-LIST].

Several approaches extract global and local features and analyze the combination of both feature types [LEAR, INAOE TIA, I2R, budapest, CEA LIST, kameyama, AVEIR, MRIM, HUTCIS].

Some groups tried to make use of additional information such as EXIF tags [UAIC] and concept names [Telecom Paristech, AVEIR]. Others obtained higher level features, e.g. with the application of a face detector [UAIC].

11.3.2 Feature Post–processing and Codebook Generation

While global image descriptors directly describe an entire image, local features are often summarized in a bag–of–visual–words descriptor. Many variations of bag–of–visual–words approaches were proposed. The most common approach is to cluster a set of representative local descriptors using k-means into 500–2,000 cluster prototypes. Then each image is represented by a histogram counting how many of its local descriptors belong into which of the clusters.

Such approaches were adopted by many groups over the four years. MSRA and RWTH followed this approach, while ISIS additionally investigated different settings for codebook generation. IAM uses a hierarchical k–means for clustering. LSIS's approach applies a Euclidean distance on multigrid features for visual word assignment after the k-means clustering and budapest replaces the k-means clustering step with a Gaussian Mixture Model.

11.3.3 Classifier

Given the image descriptors, a classifier is applied to predict the class(es) of the test images. The parameters of the classifier are trained on the training data and tuned using the validation data.

Classifiers are often grouped into generative, discriminative, or model–free approaches. Generative probabilistic models estimate the distribution of observations for each class and use this to predict which class is most likely for a certain observation. Discriminative approaches directly model the posterior probability for the classes. Another option is to combine or blend both approaches.

In the ImageCLEF object retrieval tasks, a large variety of discriminative and generative classifiers have been used. By far the most prominent approach was the classification with Support Vector Machines (SVMs) with different kernels, multiple-kernels, and multi-class extensions [CINDI, HUTCIS, MSRA, CEA-LIST, LSIS, I2R, INRIA-LEAR, ISIS, MRIM-LIG, UPMC and FIRST]. Other discriminative approaches include log–linear models [RWTH] and logistic regression [budapest, XRCE], fuzzy decision forest [UPMC] or random forests [INAOE TIA].

The most popular model–free approach was the nearest neighbor classifier. It has often been used as baseline for more sophisticated approaches using different distance functions [CINDI, DEU, INAOE TIA, CEA-LIST, HJ-FA, Makere, PRIP and Kameyama Lab] or weighted neighbors [INRIA-LEAR].

Furthermore, a variety of language models have been applied. MSRA uses a visual topic model and a trigram language model and IAM investigated a cross–language latent indexing method with a cosine distance decision function. Non–parametric density estimation functions [MMIS], Markov Random Fields [INAOE TIA] and Self Organizing Maps [HUTCIS] are further adopted methods.

Some groups used a number of classifiers and applied a fusion step of the results after classification, e.g. [HUTCIS, AVEIR].

11.3.4 Post–Processing

After the classification step, some groups further refined the results. This step was mainly applied in 2008 and 2009, as in these years a taxonomy and an ontology were offered as additional knowledge bases. Some participants incorporated this knowledge to improve their classifiers, partly also directly in the classification step. A popular approach was the co–occurrence and correlation analysis of concept context in the training data [MMIS, INAOE TIA, UPMC, budapest]. One group applied semantic similarities that were determined by word correlations in Google, WordNet and Wikipedia [MMIS]. Furthermore, thresholds were adapted in case of mutually exclusive concepts [I2R, XRCE].

11.4 Results

In this section, the results of the individual years are summarized. The task and the databases changed over the years, as outlined in Section 11.2. Therefore, the

Table 11.3: Results from the object annotation task in 2006 sorted by error rate.

Rank	Group ID	Descriptor	Classifier	Error Rate [%]
1	RWTH	dense BoW	log-linear	77.3
2	RWTH	sparse BoW	log-linear	80.2
3	cindi	global edge, color	SVM	83.2
4	cindi	global edge, color	SVM	85.0
5	cindi	global edge, color	SVM	85.2
6	cindi	global edge, color	KNN	87.1
7	DEU	global edge	generative Gauss	88.2
-	medGIFT	collection frequencies	GIFT-NN	90.5
-	medGIFT	collection frequencies	GIFT-NN	91.7
8	DEU	global colorlayout	generative Gauss	93.2

results of the different years cannot be compared directly to each other. However, we compare results across different years where possible.

11.4.1 2006: Object Annotation Task

Table 11.3 shows the results for three participating groups of the object annotation task in 2006. The results of MedGIFT are not ranked, because they submitted their runs after the deadline. The runs were evaluated using the error rate. Error rates are very high and range from 77.3% to 93.2%. Further analysis revealed that many of the test images could not be classified correctly by any method. Summarizing, the discriminative classification methods outperformed the others by a small amount.

11.4.2 2007: Object Retrieval Task

The submissions of the Object Retrieval Task in 2007 were evaluated according to average precision (AP) per class and ranked by the MAP over all classes. Table 11.4 presents the results. HUTCIS obtained the best result with a MAP of 2.9%. Considering the class–wise results, the best overall results were obtained for the `car` class with an AP of 11.1%. Also, the classes `person` and `bicycle` could be detected well with an AP of 8.6% and 4.1%, respectively. The worst results were achieved for the classes `dog` and `sheep`, which could be detected with an AP of just 0.1%. Except the classes `sheep` and `cat`, all best results per class were obtained by one of the SVM configurations of HUTCIS.

The low performance of all methods shows that the task is very difficult and that the varying number of relevant images per topic further complicates it.

Table 11.4: Results from the ImageCLEF 2007 object retrieval task with complete relevance information obtained by manual categorization for the whole database. All values have been multiplied by 100 to make the table more readable. The results for each class are presented in the corresponding columns. The MAP over all classes is in the last column. The highest AP per class is shown in bold. Please note that the results of the budapest group are not fully comparable as they assigned just a single class per photo instead of multiple classes and used a different, more strongly labelled training set.

Group ID	Descriptor	Classifier	Bicycle	Bus	Car	Mbike	Cat	Cow	Dog	Horse	Sheep	Person	MAP
HUTCIS	BoW (global and local)	SVM	**4.1**	**1.2**	10.6	0.4	0.0	0.6	0.1	**3.8**	0.0	8.3	2.9
HUTCIS	SIFT, color	SVM	2.6	1.0	**11.1**	1.0	0.0	1.0	0.1	3.2	0.0	8.1	2.8
HUTCIS	SIFT	SVM	2.4	1.1	10.3	**1.8**	0.0	**1.1**	0.1	3.0	0.0	8.1	2.8
HUTCIS	BoW (global and local)	SVM	3.0	1.1	4.2	0.6	0.0	0.7	**0.1**	2.5	0.0	**8.6**	2.1
HUTCIS	BoW (global and local)	SVM	1.6	0.9	0.5	0.3	0.0	0.6	0.1	1.5	0.0	8.3	1.4
HUTCIS	SIFT, color	SVM	1.4	1.0	0.7	0.3	0.0	0.5	0.1	1.1	0.0	8.4	1.4
HUTCIS	SIFT, color	SVM	2.0	0.8	0.4	0.3	0.0	0.8	0.1	1.1	0.0	8.2	1.3
HUTCIS	BoW (global and local)	SOM	0.9	0.7	4.5	0.6	0.0	0.3	0.1	0.7	0.0	5.6	1.3
MSRA	SIFT	pLSA + SVM	0.9	0.5	3.6	0.6	0.7	0.1	0.1	0.4	0.0	6.0	1.3
HUTCIS	SIFT, color	SVM	1.3	0.8	0.5	0.2	0.0	0.5	0.1	0.8	0.0	8.4	1.3
HUTCIS	BoW (global and local)	SOM	0.8	0.6	4.2	0.5	0.0	0.3	0.1	0.4	0.0	5.4	1.2
HUTCIS	SIFT	SVM	1.1	0.7	0.4	1.4	0.0	0.3	0.0	1.0	0.0	7.2	1.2
HUTCIS	SIFT	SVM	1.1	0.8	0.3	0.3	0.0	0.4	0.1	0.9	0.0	6.9	1.1
HUTCIS	SIFT	SVM	0.3	0.9	0.3	0.3	0.0	0.3	0.0	1.1	0.0	6.6	1.0
RWTH	dense BoW	log-linear	0.4	0.2	1.4	0.2	0.0	0.1	0.0	0.2	0.0	5.5	0.8
budapest	BoW	segment NN	0.1	0.1	0.8	0.0	0.0	0.2	0.0	0.2	0.0	4.1	0.5
NTU	global color, texture, shape	SVM	1.2	0.7	2.4	0.0	0.0	0.0	0.1	0.1	0.0	0.8	0.5
budapest	BoW	segment NN	0.4	0.0	0.4	0.1	0.0	0.0	0.1	0.1	0.0	3.9	0.5
MSRA	patch-based texture	tri-gram language model	0.4	0.3	0.7	0.1	0.1	0.0	0.0	0.3	0.0	2.5	0.4
MSRA	patch-based texture	tri-gram language model	0.3	0.2	0.5	0.0	0.0	0.1	0.0	0.2	**0.1**	2.5	0.4
INAOE-TIA	BoW	naïve Bayes + AdaBoost	0.1	0.0	0.1	0.0	0.0	0.2	0.0	0.2	0.0	3.2	0.4
INAOE-TIA	BoW	KNN + MRF	0.5	0.1	0.6	0.0	0.0	0.2	0.0	0.0	0.0	2.2	0.4
INAOE-TIA	BoW	KNN + MRF	0.5	0.1	0.6	0.0	0.0	0.2	0.0	0.0	0.0	2.2	0.4
PRIP	SIFT	EMD + NN	0.1	0.0	0.3	0.1	**1.4**	0.1	0.0	0.0	0.0	1.5	0.4
INAOE-TIA	BoW	KNN	0.3	0.0	0.5	0.1	0.0	0.0	0.0	0.0	0.0	2.2	0.3
PRIP	SIFT	EMD + NN	0.1	0.0	0.1	0.5	0.1	0.0	0.1	0.1	0.0	0.8	0.2

11.4.3 2008: Visual Concept Detection Task

The results of the 2008 Visual Concept Detection Task are presented in Table 11.5 and Table 11.6. Runs were ranked according to their EER and AUC scores. Table 11.5 shows the performance for the best run of each group in terms of EER and AUC and the descriptors and classifiers applied. The best scores of each group range from 16.7% EER to 49.3% EER. In terms of AUC, the best run achieved 90.7% AUC, while the values fall to 20% AUC.

Table 11.5: Summary of the results of the visual concept detection task in ImageCLEF 2008. The table shows the results for the best run per group.

GroupID	Descriptor	Classifier	rank	EER	AUC
XRCE	local color and texture	Fisher-Kernel SVM + logistic regression	1	16.7	90.7
RWTH	BoW	log linear model	3	20.5	86.2
UPMC	local color	fuzzy decision forests	4	24.6	82.7
LSIS	profile entropy features + others	SVM	5	25.9	80.5
MMIS	color, Tamura texture	non-parametric density estimation	13	28.4	77.9
CEA-LIST	color, spatial layout	NN, SVM	17	29.0	73.4
IPAL-I2R	variety of descriptors	—	19	29.7	76.4
budapest	global and local	logistic regression	20	31.1	74.9
TIA	global and local	SVM, random forest	24	32.1	55.6
HJ-FA	color, SIFT	KNN	47	45.1	20.0
Makere	luminance, color, texture, shape	NN	51	49.3	30.8

Table 11.6 presents the results per concept. For each concept, the best and the worst EER and AUC are shown, along with the average EER and AUC over all runs submitted. The best results were obtained for all concepts by the XRCE group, with budapest doing equally well on the `night` concept. The best AUC per concept is at least 80% for the concept `road/pathway` and rises up to 97.4% for the concepts `indoor` and `night`. The rather poor results for the concept `road/pathway` can be explained by the high variability in the appearance of this concept. The concept with the highest average score, in other words, the concept that was detected best in most runs is `sky`. Again, the concept with the worst average score is `road/pathway`.

Summarizing, discriminative approaches with local features achieved the best results. Further, the results demonstrate that the concept detection task could be solved reasonably well.

11.4.4 2009: Visual Concept Detection Task

The evaluation of the concept detection task in 2009 focused on two evaluation paradigms, the evaluation per concept and the evaluation per photo. The evaluation per concept was conducted with the EER and AUC as in the previous year. For the evaluation per photo, a new evaluation measure, the Ontology Score (OS), was introduced (Nowak et al, 2010).

The results are given in Table 11.7. The group with the best concept–based results, ISIS, achieves an EER of 23% and an AUC of 84% on average for their best run. The next three groups in the ranking closely follow these results with an EER of about 25% and an AUC of 82% and 81%. The performance of the groups at the end of the list goes up to 53% EER and falls to 7% AUC.

The evaluation per photo reveals scores in the range of 39% to 81% for the best run per group. The best results in terms of OS were achieved by the XRCE group with 81% annotation score over all photos. It can be seen from the table that the ranking of the groups is different than for the EER/AUC measures.

In Table 11.8, the results for each concept are illustrated in terms of EER and AUC over all runs submitted. All concepts could be detected at least with 44% EER and 58% AUC, but on average with an EER of 23% and an AUC of 84%. The majority of the concepts were classified best by the ISIS group. It is obvious that the aesthetic concepts (`Aesthetic_Impression`, `Overall_Quality` and `Fancy`) are classified worst (EER greater than 38% and AUC less than 66%.). This is not surprising due to the subjective nature of these concepts which also made the ground truthing difficult. The best classified concepts are `Clouds` (AUC: 96%), `Sunset-Sunrise` (AUC: 95%), `Sky` (AUC: 95%) and `Landscape-Nature` (AUC: 94%).

Summarizing, the groups that used local features such as SIFT achieved better results than the groups relying solely on global features. Most groups that investigated the concept hierarchy and analyzed, for example, the correlations between the concepts, could achieve better results in the OS compared to the EER. Again, the discriminative methods outperformed the generative and model–free ones.

Table 11.6: Overview of the results per concept of the visual concept detection task 2008.

		best			average		worst	
#	concept	EER	AUC	group	EER	AUC	EER	AUC
00	indoor	8.9	97.4	XRCE	28.0	67.6	46.8	2.0
01	outdoor	9.2	96.6	XRCE	30.6	70.5	54.6	13.3
02	person	17.8	89.7	XRCE	35.9	62.2	53.0	0.4
03	day	21.0	85.7	XRCE	35.4	64.9	52.5	9.7
04	night	8.7	97.4	XRCE/budapest	27.6	72.5	73.3	0.0
05	water	23.8	84.6	XRCE	38.1	57.8	53.0	3.2
06	road/pathway	28.8	80.0	XRCE	42.6	50.7	56.8	0.0
07	vegetation	17.6	89.9	XRCE	33.9	67.4	49.7	30.7
08	tree	18.9	88.3	XRCE	36.1	62.8	59.5	1.0
09	mountains	15.3	93.8	XRCE	33.1	61.2	55.8	0.0
10	beach	21.7	86.8	XRCE	35.8	57.6	51.4	0.0
11	buildings	17.0	89.7	XRCE	37.4	60.8	64.0	0.5
12	sky	10.4	95.7	XRCE	24.0	78.6	50.8	37.3
13	sunny	9.2	96.4	XRCE	30.2	66.5	55.4	0.0
14	partly cloudy	15.4	92.1	XRCE	37.5	58.9	55.5	0.0
15	overcast	14.1	93.7	XRCE	32.1	67.6	61.5	0.0
16	animal	20.7	85.7	XRCE	38.2	54.2	58.4	0.0

Table 11.7: Summary of the results for the concept detection task in 2009. The table shows the EER and AUC performance for the best run per group ranked by EER for the concept–based evaluation and the performance with the OS measure for the best run per group for the photo–based evaluation. Note, that the best run for the EER measure is not necessarily the same run as for the OS measure.

Group ID	Descriptor	Classifier	Rank	EER	AUC	Rank	OS
ISIS	color SIFT	SVM	1	0.23	0.84	14	0.77
LEAR	BoW (global and local)	SVM / NN	5	0.25	0.82	12	0.77
I2R	global and local	SVM	7	0.25	0.81	2	0.81
FIRST	SIFT, color	multiple kernel SVM	8	0.25	0.82	4	0.80
XRCE	BoW	sparse logistic regression	14	0.27	0.80	1	0.81
budapest	various global and local features	logistic regression	17	0.29	0.77	35	0.68
MMIS	color, Tamura, Gabor	non-parametric density estimation	21	0.31	0.74	42	0.58
IAM	SIFT	cosine distance of visual terms	23	0.33	0.72	61	0.41
LSIS	various features	SVM(LDA) / Visual Dictionary	24	0.33	0.72	49	0.51
UPMC	HSV histogram	SVM	33	0.37	0.67	58	0.44
MRIM	RGB histogram, SIFT, Gabor	SVM	34	0.38	0.64	28	0.72
AVEIR	various global and local features, text	SVM / Visual Dictionary / canonical correlation	41	0.44	0.55	50	0.50
Wroclaw Uni	various features	Multivariate Gaussian Model + NN	43	0.45	0.22	11	0.78
Kameyama	global and local	KNN	47	0.45	0.16	7	0.80
UAIC	face detection, exif	NN + default values	54	0.48	0.11	32	0.69
apexlab	various features	KNN	56	0.48	0.07	17	0.76
INAOE TIA	various global features	KNN	57	0.49	0.10	20	0.74
Random	-	-	-	0.50	0.50	-	0.38
CEA LIST	global and local	Multiclass boosting	68	0.50	0.47	29	0.71
TELECOM	global, text features	Canonical Correlation Analysis + thresholds	72	0.53	0.46	65	0.39

11.4.5 Evolution of Concept Detection Performance

Comparisons of performance across years can best be made from 2008 to 2009. Although the database changed between these evaluation cycles, the methodology of the tasks was similar. Comparing the results from 2008 to 2009, the average AUC over all concepts for the best run drops from 90% to 84%, while increasing the number of concepts with a factor of about three. The most comparable concepts `indoor` and `outdoor` dropped by 13% and 7%, respectively, which can be explained with the third concept `NoVisualPlace` in the group in 2009. Other concepts could be annotated with a similar quality, e.g. `mountains` and `sky` -1%, `day` and `trees` +/-0%. The concept `person` was substituted by four concepts `single person`,

Table 11.8: Overview of the best results per concept over all submitted runs in 2009 in terms of the EER and AUC and the name of the group which achieved these results.

No.	Concept	AUC	EER	Group ID	No.	Concept	AUC	EER	Group ID
0	Partylife	0.83	0.24	ISIS	27	Day	0.85	0.24	ISIS
1	Family_Friends	0.83	0.24	ISIS	28	Night	0.91	0.17	LEAR
2	Beach_Holidays	0.91	0.16	ISIS	29	No_Visual_Time	0.84	0.25	ISIS
3	Building_Sights	0.88	0.20	ISIS	30	Sunny	0.77	0.30	LEAR - ISIS
4	Snow	0.87	0.21	LEAR	31	Sunset_Sunrise	0.95	0.11	ISIS
5	Citylife	0.83	0.25	ISIS	32	Canvas	0.82	0.25	XRCE
6	Landscape_Nature	0.94	0.13	ISIS	33	Still_Life	0.82	0.25	ISIS
7	Sports	0.72	0.34	FIRST	34	Macro	0.81	0.26	ISIS
8	Desert	0.89	0.18	ISIS	35	Portrait	0.87	0.21	XRCE - ISIS
9	Spring	0.83	0.25	FIRST	36	Overexposed	0.80	0.25	UPMC
10	Summer	0.81	0.26	ISIS	37	Underexposed	0.88	0.18	CVIUI2R
11	Autumn	0.87	0.21	ISIS	38	Neutral_Illumination	0.80	0.26	LEAR
12	Winter	0.85	0.23	ISIS	39	Motion_Blur	0.75	0.32	ISIS
13	No_Visual_Season	0.81	0.26	ISIS	40	Out_of_focus	0.81	0.25	LEAR
14	Indoor	0.84	0.25	ISIS	41	Partly_Blurred	0.86	0.22	LEAR
15	Outdoor	0.90	0.19	ISIS	42	No_Blur	0.85	0.23	LEAR
16	No_Visual_Place	0.79	0.29	ISIS	43	Single_Person	0.79	0.28	ISIS - LEAR
17	Plants	0.88	0.21	ISIS	44	Small_Group	0.80	0.28	ISIS
18	Flowers	0.87	0.20	ISIS - FIRST	45	Big_Group	0.88	0.21	ISIS
19	Trees	0.90	0.18	ISIS	46	No_Persons	0.86	0.22	ISIS
20	Sky	0.95	0.12	ISIS	47	Animals	0.83	0.25	ISIS
21	Clouds	0.96	0.10	ISIS	48	Food	0.90	0.19	ISIS
22	Water	0.90	0.18	ISIS	49	Vehicle	0.83	0.24	ISIS
23	Lake	0.91	0.16	ISIS	50	Aesthetic_Impression	0.66	0.38	ISIS
24	River	0.90	0.17	ISIS	51	Overall_Quality	0.66	0.38	ISIS
25	Sea	0.94	0.13	ISIS	52	Fancy	0.58	0.44	ISIS
26	Mountains	0.93	0.14	ISIS					

`small group`, `big group` and `no person` and dropped on average by 7%. Concepts that achieved better scores in 2009 are `beach` +4%, `clouds` +4% and `water` +5%. In case of `clouds`, the 2009 task was easier, because the concepts `overcast` and `partly cloudy` were combined in one concept.

11.4.6 Discussion

In 2006, the bag–of–visual–words approach by RWTH with a log–linear classifier performed best. In 2007, HUTCIS obtained the best result by combining various descriptors (color, edge, SIFT, combinations) and SVM classifiers. In 2008, XRCE achieved the best result using local color and texture features and a combination of Fisher–kernel SVMs and logistic models. In 2009, ISIS obtained the best result using a large variety of local descriptors extracted from different interest points and grids represented in a bag–of–words–descriptor and χ^2-SVM classifiers. For the photo-based evaluation, the XRCE group achieved the best results in 2009 with a system similar to their 2008 approach.

Over all years, the best results were obtained using discriminative classifiers. The classifier itself varied throughout the years and also the features differed. The knowledge provided in form of a taxonomy and an ontology in 2008 and 2009, respectively, was not further considered by most groups. Only in the post–processing step of the XRCE run in 2009, was the probability of the presence of a particular concept adapted by analyzing its likely relationships.

11.5 Combinations with the Photo Retrieval Task

In 2008, two automatic runs provided by participants of the visual concept detection task were made available to the participants of the photo retrieval task. These contained annotations for the database of 20,000 photos used in the photo retrieval task with the VCDT concepts. Two groups that participated in the photo retrieval task of ImageCLEF made use of these annotations. UPMC applied VCDT annotations provided by their own algorithm. They used the detected visual concepts to re-rank the first 50 results returned by text retrieval approaches. The concepts to use for the re–ranking were chosen by two approaches: (i) the concept word appears in the query text and (ii) the concept word appears in the list of synonyms (obtained by WordNet) of the words in the query text. The first approach improved the results of all the queries for which it was applicable, while the second resulted in worse results for some topics. Both approaches achieved a better overall performance than using text alone: the F–measure for the best text only run (using TF–IDF) is 0.273, while the F–measure for the run re–ranked using the first approach is 0.289.

The NII group (Inoue and Grover, 2008) made use of both provided VCDT concept annotation sets. They also used the concepts to re–rank results returned by a text retrieval approach. The best results were obtained by a re–ranking based on a hierarchical clustering which uses distances between vectors to encode the VCDT concepts. This re–ranking decreased the P20 metric while increasing the CR20 metric, resulting in an increase of the F–measure from 0.224 for text only to 0.230 after the re–ranking.

INAOE TIA used one of the provided VCDT concept annotation sets as one part of a group of visual retrieval algorithms whose results were integrated in a late fusion process. It is therefore not possible to determine the effect of only the VCDT concepts on the results.

11.6 Conclusion

This chapter presents an overview of the object and concept recognition tasks of ImageCLEF in the four years from 2006 to 2009. The tasks varied strongly over the years reflecting the objective to start with a flat classification task and going towards a full image annotation that can be used for content–based access to photo repositories. Over the years, 29 groups participated in total and submitted over 163 runs, processing a total of 35,100 test images.

For the future, we will continue to pose challenging tasks for object and concept annotation. In 2010, the task will consider Flickr User Tags, so that the participants can decide whether they solve the concept detection task purely visually, purely based on social data or if they prefer to follow multi–modal approaches. The aim is to analyze if the multi–modal annotation approaches can outperform text only or visual only approaches and which approach is best suited to which type of concepts. Furthermore, the systems are trained and evaluated on 92 concepts, containing also more subjective annotations such as `boring` or `cute` and event concepts such as `birthday` or `work`.

Acknowledgements We would like to thank the organizers of CLEF and ImageCLEF for the support of the object and concept recognition tasks. This work has been partly supported by grant No. 01MQ07017 of the German research program THESEUS funded by the Ministry of Economics and the European Commission Network of Excellence MUSCLE (FP6-507752).

References

Ah-Pine J, Cifarelli C, Clinchant S, Csurka G, Renders J (2008) XRCE's Participation to ImageCLEF 2008. In: Working Notes of CLEF 2008, Aarhus, Denmark

Ah-Pine J, Clinchant S, Csurka G, Liu Y (2009) XRCE's Participation in ImageCLEF 2009. In: Working Notes of CLEF 2009, Corfu, Greece

Binder A, Kawanabe M (2009) Enhancing Recognition of Visual Concepts with Primitive Color Histograms via Non–sparse Multiple Kernel Learning. In: Peters C, Tsikrika T, Müller H, Kalpathy-Cramer J, Jones J, Gonzalo J, Caputo B (eds) Multilingual Information Access Evaluation Vol. II Multimedia Experiments: Proceedings of the 10th Workshop of the Cross–Language Evaluation Forum (CLEF 2009), Revised Selected Papers, Springer, Corfu, Greece, Lecture Notes in Computer Science (LNCS)

Braschler M, Peters C (2003) CLEF methodology and metrics. In: Peters C, Braschler M, Gonzalo J, Kluck M (eds) Evaluation of Cross–Language Information Retrieval Systems, Evaluation of Cross–Language Information Retrieval Systems, Springer, Darmstadt, Germany, Lecture Notes in Computer Science (LNCS), vol 2406, pp 394–404

Clough PD, Müller H, Sanderson M (2005) The CLEF 2004 cross–language image retrieval track. In: Peters C, Clough P, Gonzalo J, Jones G, Kluck M, Magnini B (eds) Multilingual Information Access for Text, Speech and Images Fifth Workshop of the Cross–Language Evaluation Forum, CLEF 2004, Springer, Bath, UK, Lecture Notes in Computer Science (LNCS), vol 3491, pp 597–613

Clough PD, Grubinger M, Deselaers T, Hanbury A, Müller H (2007) Overview of the ImageCLEF 2006 photographic retrieval and object annotation tasks. In: Peters C, Clough P, Gey F, Karlgren J, Magnini B, Oard D, de Rijke M, Stempfhuber M (eds) Evaluation of Multilingual and Multimodal Information Retrieval 7th Workshop of the Cross–Language Evaluation Forum, CLEF 2006, Springer, Alicante, Spain, Lecture Notes in Computer Science (LNCS), vol 4730, pp 579–594

Daróczy B, Fekete Z, Brendel M, Rácz S, Benczúr A, Siklósi D, Pereszlényi A (2008) SZTAKI@ ImageCLEF 2008: visual feature analysis in segmented images. In: Peters C, Deselaers T, Ferro N, Gonzalo J, Jones G, Kurimo M, Mandl T, Peñas A, Petras V (eds) Evaluating Systems for Multilingual and MultiModal Information Access 9th Workshop of the Cross–Language

Evaluation Forum, Springer, Aarhus, Denmark, Lecture Notes in Computer Science (LNCS), vol 5706, pp 644–651

Daróczy B, Petrás I, Benczúr A, Fekete Z, Nemeskey D, Siklósi D, Weiner Z (2009) Interest Point and Segmentation-Based Photo Annotation. In: Peters C, Tsikrika T, Müller H, Kalpathy-Cramer J, Jones J, Gonzalo J, Caputo B (eds) Multilingual Information Access Evaluation Vol. II Multimedia Experiments: Proceedings of the 10th Workshop of the Cross–Language Evaluation Forum (CLEF 2009), Revised Selected Papers, Springer, Corfu, Greece, Lecture Notes in Computer Science (LNCS)

Deselaers T, Hanbury A (2008) The visual concept detection task in ImageCLEF 2008. In: Peters C, Deselaers T, Ferro N, Gonzalo J, Jones G, Kurimo M, Mandl T, Peñas A, Petras V (eds) Evaluating Systems for Multilingual and MultiModal Information Access 9th Workshop of the Cross–Language Evaluation Forum, Springer, Aarhus, Denmark, Lecture Notes in Computer Science (LNCS), vol 5706, pp 531–538

Deselaers T, Hanbury A, Viitaniemi V, Benczúr A, Brendel M, Daróczy B, Escalante Balderas H, Gevers T, Hernández Gracidas C, Hoi S, Laaksonen J, Li M, Marín Castro H, Ney H, Rui X, Sebe N, Stöttinger J, Wu L (2008) Overview of the ImageCLEF 2007 Object Retrieval Task. In: Peters C, Jijkoun V, Mandl T, Müller H, Oard D, Peñas A, Petras V, Santos D (eds) Advances in Multilingual and MultiModal Information Retrieval 8th Workshop of the Cross–Language Evaluation Forum, CLEF 2007, Springer, Budapest, Hungary, Lecture Notes in Computer Science (LNCS), vol 5152, pp 445–471

Douze M, Guillaumin M, Mensink T, Schmid C, Verbeek J (2009) INRIA–LEARs participation to ImageCLEF 2009. In: Working Notes of CLEF 2009, Corfu, Greece

Dumont E, Zhao ZQ, Glotin H, Paris S (2009) A new TFIDF Bag of Visual Words for Concept Detection. In: Peters C, Tsikrika T, Müller H, Kalpathy-Cramer J, Jones J, Gonzalo J, Caputo B (eds) Multilingual Information Access Evaluation Vol. II Multimedia Experiments: Proceedings of the 10th Workshop of the Cross–Language Evaluation Forum (CLEF 2009), Revised Selected Papers, Springer, Corfu, Greece, Lecture Notes in Computer Science (LNCS)

Escalante H, Gonzalez J, Hernandez C, Lopez A, Montex M, Morales E, Ruiz E, Sucar L, Villasenor L (2009) TIA–INAOE's Participation at ImageCLEF 2009. In: Working Notes of CLEF 2009, Corfu, Greece

Everingham M, Zisserman A, Williams C, Van Gool L, Allan M, et al (2006) The 2005 PASCAL Visual Object Classes Challenge. In: Machine Learning Challenges. Evaluating Predictive Uncertainty, Visual Object Classification, and Recognising Textual Entailment (PASCAL Workshop 2005), Springer, Southampton, UK, no. 3944 in Lecture Notes in Artificial Intelligence (LNAI), pp 117–176

Everingham M, Gool LV, Williams CKI, Winn J, Zisserman A (2010) The pascal visual object classes (voc) challenge. International Journal of Computer Vision 88:303–538

Fakeri-Tabrizi A, Tollari S, Usunier N, Gallinari P (2009) Improving Image Annotation in Imbalanced Classification Problems with Ranking SVM. In: Peters C, Tsikrika T, Müller H, Kalpathy-Cramer J, Jones J, Gonzalo J, Caputo B (eds) Multilingual Information Access Evaluation Vol. II Multimedia Experiments: Proceedings of the 10th Workshop of the Cross–Language Evaluation Forum (CLEF 2009), Revised Selected Papers, Springer, Corfu, Greece, Lecture Notes in Computer Science (LNCS)

Ferecatu M, Sahbi H (2009) TELECOM ParisTech at ImageCLEF 2009: Large Scale Visual Concept Detection and Annotation Task. In: Working Notes of CLEF 2009, Corfu, Greece

Glotin H, Fakeri-Tabrizi A, Mulhem P, Ferecatu M, Zhao Z, Tollari S, Quenot G, Sahbi H, Dumont E, Gallinari P (2009) Comparison of Various AVEIR Visual Concept Detectors with an Index of Carefulness. In: Working Notes of CLEF 2009, Corfu, Greece

Grubinger M, Clough P, Müller H, Deselaers T (2006) The IAPR TC–12 benchmark — a new evaluation resource for visual information systems. In: Proceedings of the International Workshop OntoImage'2006, pp 13–23

Hare J, Lewis P (2009) IAM@ImageCLEFPhotoAnnotation 2009: Naive application of a linear–algebraic semantic space. In: Working Notes of CLEF 2009, Corfu, Greece

Huiskes MJ, Lew MS (2008) The MIR Flickr Retrieval Evaluation. In: MIR 2008: Proceedings of the 2008 ACM International Conference on Multimedia Information Retrieval, ACM press

Iftene A, Vamanu L, Croitoru C (2009) UAIC at ImageCLEF 2009 Photo Annotation Task. In: Peters C, Tsikrika T, Müller H, Kalpathy-Cramer J, Jones J, Gonzalo J, Caputo B (eds) Multilingual Information Access Evaluation Vol. II Multimedia Experiments: Proceedings of the 10th Workshop of the Cross–Language Evaluation Forum (CLEF 2009), Revised Selected Papers, Springer, Corfu, Greece, Lecture Notes in Computer Science (LNCS)

Inoue M, Grover P (2008) Query Types and Visual Concept–Based Post–retrieval Clustering. In: Peters C, Deselaers T, Ferro N, Gonzalo J, Jones G, Kurimo M, Mandl T, Peñas A, Petras V (eds) Evaluating Systems for Multilingual and MultiModal Information Access 9th Workshop of the Cross–Language Evaluation Forum, Springer, Aarhus, Denmark, Lecture Notes in Computer Science (LNCS), vol 5706, pp 661–668

Jiang J, Rui X, Yu N (2008) Feature Annotation for Visual Concept Detection in ImageCLEF 2008. In: Working Notes of CLEF 2008, Aarhus, Denmark

Llorente A, Overell S, Liu H, Hu R, Rae A, Zhu J, Song D, Rüger S (2008) Exploiting Term Co–occurrence for Enhancing Automated Image Annotation. In: Peters C, Deselaers T, Ferro N, Gonzalo J, Jones G, Kurimo M, Mandl T, Peñas A, Petras V (eds) Evaluating Systems for Multilingual and MultiModal Information Access 9th Workshop of the Cross–Language Evaluation Forum, Springer, Aarhus, Denmark, Lecture Notes in Computer Science (LNCS), vol 5706, pp 632–639

Llorente A, Motta E, Rüger S (2009) Exploring the Semantics Behind a Collection to Improve Automated Image Annotation. In: Peters C, Tsikrika T, Müller H, Kalpathy-Cramer J, Jones J, Gonzalo J, Caputo B (eds) Multilingual Information Access Evaluation Vol. II Multimedia Experiments: Proceedings of the 10th Workshop of the Cross–Language Evaluation Forum (CLEF 2009), Revised Selected Papers, Springer, Corfu, Greece, Lecture Notes in Computer Science (LNCS)

Moellic PA, Fluhr C (2006) ImageEVAL 2006 official campaign. Tech. rep., ImagEVAL

Ngiam J, Goh H (2009) Learning Global and Regional Features for Photo Annotation. In: Peters C, Tsikrika T, Müller H, Kalpathy-Cramer J, Jones J, Gonzalo J, Caputo B (eds) Multilingual Information Access Evaluation Vol. II Multimedia Experiments: Proceedings of the 10th Workshop of the Cross–Language Evaluation Forum (CLEF 2009), Revised Selected Papers, Springer, Corfu, Greece, Lecture Notes in Computer Science (LNCS)

Nowak S, Dunker P (2009) Overview of the CLEF 2009 Large–Scale Visual Concept Detection and Annotation Task. In: Peters C, Tsikrika T, Müller H, Kalpathy-Cramer J, Jones J, Gonzalo J, Caputo B (eds) Multilingual Information Access Evaluation Vol. II Multimedia Experiments: Proceedings of the 10th Workshop of the Cross–Language Evaluation Forum (CLEF 2009), Revised Selected Papers, Springer, Corfu, Greece, Lecture Notes in Computer Science (LNCS)

Nowak S, Lukashevich H (2009) Multilabel Classification Evaluation using Ontology Information. In: The 1st Workshop on Inductive Reasoning and Machine Learning on the Semantic Web — IRMLeS 2009, co–located with the 6th Annual European Semantic Web Conference (ESWC), Heraklion, Greece

Nowak S, Lukashevich H, Dunker P, Rüger S (2010) Performance measures for multilabel evaluation: a case study in the area of image classification. In: Proceedings of the international conference on Multimedia information retrieval, ACM press, pp 35–44

Pham T, Maisonnasse L, Mulhem P, Chevallet JP, Quénot G, Al Batal R (2009) MRIM–LIG at ImageCLEF 2009: Robot Vision, Image annotation and retrieval tasks. In: Peters C, Tsikrika T, Müller H, Kalpathy-Cramer J, Jones J, Gonzalo J, Caputo B (eds) Multilingual Information Access Evaluation Vol. II Multimedia Experiments: Proceedings of the 10th Workshop of the Cross–Language Evaluation Forum (CLEF 2009), Revised Selected Papers, Springer, Corfu, Greece, Lecture Notes in Computer Science (LNCS)

van de Sande K, Gevers T, Smeulders A (2009) The University of Amsterdams Concept Detection System at ImageCLEF 2009. In: Peters C, Tsikrika T, Müller H, Kalpathy-Cramer J, Jones J, Gonzalo J, Caputo B (eds) Multilingual Information Access Evaluation Vol. II Multimedia Experiments: Proceedings of the 10th Workshop of the Cross–Language Evaluation Forum

(CLEF 2009), Revised Selected Papers, Springer, Corfu, Greece, Lecture Notes in Computer Science (LNCS)

Sarin S, Kameyama W (2009) Joint Contribution of Global and Local Features for Image Annotation. In: Working Notes of CLEF 2009, Corfu, Greece

Tollari S, Detyniecki M, Fakeri-Tabrizi A, Marsala C, Amini M, Gallinari P (2008) Using visual concepts and fast visual diversity to improve image retrieval. In: Peters C, Deselaers T, Ferro N, Gonzalo J, Jones G, Kurimo M, Mandl T, Peñas A, Petras V (eds) Evaluating Systems for Multilingual and MultiModal Information Access 9th Workshop of the Cross–Language Evaluation Forum, Springer, Aarhus, Denmark, Lecture Notes in Computer Science (LNCS), vol 5706, pp 577–584

Zhao Z, Glotin H (2008) Enhancing Visual Concept Detection by a Novel Matrix Modular Scheme on SVM. In: Peters C, Deselaers T, Ferro N, Gonzalo J, Jones G, Kurimo M, Mandl T, Peñas A, Petras V (eds) Evaluating Systems for Multilingual and MultiModal Information Access 9th Workshop of the Cross–Language Evaluation Forum, Springer, Aarhus, Denmark, Lecture Notes in Computer Science (LNCS), vol 5706