Overview of the CLEF 2009 Large-Scale Visual Concept Detection and Annotation Task

Stefanie Nowak, Peter Dunker

Semantic Audio-Visual Systems, Fraunhofer IDMT, Ilmenau, Germany

`stefanie.nowak@idmt.fraunhofer.de, peter.dunker@ieee.org`

Abstract

The large-scale visual concept detection and annotation task (LS-VCDT) in Image-CLEF 2009 aims at the detection of 53 concepts in consumer photos. These concepts are structured in an ontology which implies a hierarchical ordering and which can be utilized during training and classification of the photos. The dataset consists of 18.000 Flickr photos which were manually annotated with 53 concepts. 5.000 photos were used for training and 13.000 for testing. Altogether 19 research groups participated and submitted 73 runs. Two evaluation paradigms have been applied, the evaluation per concept and the evaluation per photo. The evaluation per concept was performed by calculating the Equal Error Rate (EER) and the Area Under Curve (AUC). For the evaluation per photo a recently proposed hierarchical measure was utilized that takes the hierarchy and the relations of the ontology into account and calculates a score per photo. For the concepts, an average AUC of 84% could be achieved, including concepts with an AUC of 95%. The classification performance for each photo ranged between 69% and 100% with an average score of 90%.

Categories and Subject Descriptors

H.3 [**Information Storage and Retrieval**]: H.3.1 Content Analysis and Indexing; H.3.3 Information Search and Retrieval; H.3.4 Systems and Software; H.3.7 Digital Libraries; H.2 [**Database Managment**]: H.2.4 Systems—*Multimedia Databases*

General Terms

Measurement, Performance, Experimentation, Benchmark

Keywords

Image Classification and Annotation, Knowledge Structures, Evaluation

1 Introduction

Automatic methods for archiving, indexing and retrieving multimedia content become more and more important through the steadily increasing amount of digital data on the web and at home. These methods are often difficult to compare as they are evaluated on different kinds of datasets and concerning different concepts to be annotated. CLEF is an evaluation initative that aims at comparing approaches and results in cross-language retrieval for 10 years now. One track of CLEF is ImageCLEF which deals with the evaluation of image-based approaches in the medical and consumer photo domain. This year ImageCLEF posed six task. In the LS-VCDT the participants were asked to annotate a number of photos with a defined set of concepts in a multilabel scenario.

2 Task Description, Database and Ontology

The focus of LS-VCDT lies on the automatic detection and annotation of concepts in a large consumer photo collection. It mainly poses two challenges:

1. Can image classifiers scale to the large amount of concepts and data?

2. Can an ontology (hierarchy and relations) help in large scale annotations?

In this task, the MIR Flickr 25.000 image dataset [10] is utilized. This collection consists of 25.000 photos from Flickr with creative commons license. Most of them contain EXIF data, stored in a separate text file. We used altogether 18.000 of theses photos, annotated them manually with the defined visual concepts and provided them to the participants.

The training set consists of 5.000 and the testset of 13.000 images of the photoset. All images have multiple annotations. Most annotations refer to holistic visual concepts and are annotated at an image-based level. Altogether we provided the annotations for 53 concepts in rdf format and as plain text files. The visual concepts are organized in a small ontology. Participants may use the hierarchical order of the concepts and the relations between concepts for solving the annotation task. It was not allowed to use additional data for the training of the systems to ensure comparability among the groups.

The LS-VCDT is an extension of the former VCDT 2008 concerning the amount of data available and the amount of concepts to be annotated. In 2008, the database was quite small with about 1.800 images for training, 1.000 images for testing and 17 concepts to be detected.
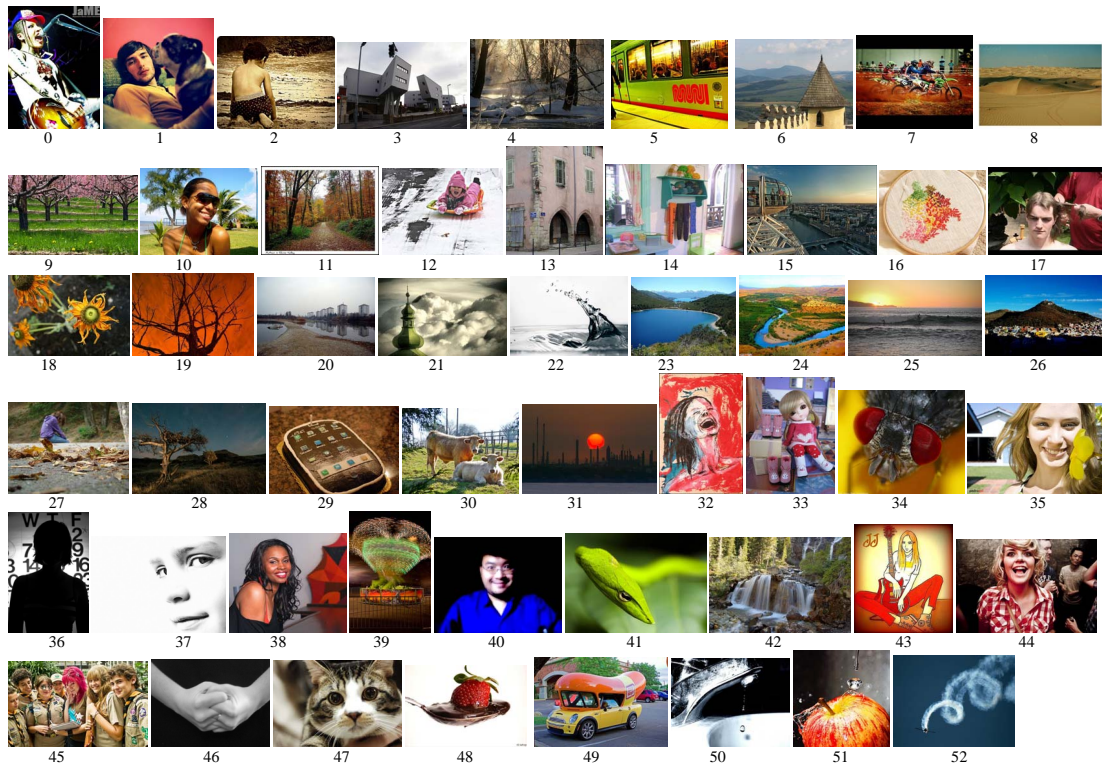

Figure 1: Example photos for each concept. The numbers below the photos denote the concept number (see also Table 1).

2.1 Annotation Process

The annotation process was realized in three steps. First the annotation of all photos was performed by several annotators, second a validation step of these annotations was conducted and third an agreement between different annotators for the same concepts and photos was calculated.

The annotation of 18.000 photos was performed by 43 persons from the Fraunhofer IDMT. The number of photos that were annotated by one person varied between 30 and 2.500 images. All annotators were provided with a definition of the concepts and example images with the goal to allow a consistent annotation amongst the large number of persons. It was important that the concepts are represented over the whole image. Some of the concepts exclude each other, others can be depicted simultaneously. One example photo per concept is illustrated in Fig. 1 and a complete list of all concepts can be found in Table 1. The frequency of each concept in the training and test sets is also depicted.

After this first annotation step, a validation of the annotations was performed. Due to the number of people, the number of photos and the ambiguity of some image contents, the annotations were not consistent throughout the database. Three persons performed a validation by screening only those photos that a) were annotated with concept X and b) that were not annotated with concept X. In the first case they had to delete all annotations for concepts that were not depicted at the photo and so were wrongly assigned. In the second case the goal was to find the photos where an annotation for concept X was missing but where the concept was visible.

Additionally, a subset of 100 photos was annotated by 11 different persons. These annotations are used to calculate an agreement between annotators for different concepts and photos. The agreement on concepts is illustrated in Table 1. For each photo and each concept, the annotation of the majority of annotators was regarded as correct and the percentage of annotators that annotated correct is utilized as agreement factor. This agreement is used in the Hierarchical Score (HS) as scaling factor (see Sec. 2.3). In case of a low agreement the algorithm assumes that the concept is ambiguous and therefore reduces the costs if the system wrongly assigns this concept. Regrettably, there was no possibility to annotate each photo by two or three persons to get a validation and an agreement on concepts over the whole set.

2.2 Ontology

In addition to the photos and their annotation an ontology was provided. All concepts are structured in this ontology. Fig. 2 shows a simple hierarchical organization of a part of the concepts. The hierarchy allows to make assumptions about the assignment of concepts to documents. E.g., if a photo is classified to contain `trees`, it also contains `plants`. Then, next to the *is-a* relationship of the hierarchical organization of concepts, additionally other relationships between concepts determine possible label assignments. The ontology restricts e.g., that for a certain sub-node only one concept can be assigned at a time (disjoint items) or that a special concept (like `portrait`) postulates other concepts like `persons` or `animals`.

The ontology allows the participants to incorporate knowledge in their classification algorithms, and to make assumptions about which concepts are probable in combination with certain labels.

2.3 Evaluation Measures

The evaluation of submissions to LS-VCDT considers two evaluation paradigms. We are interested in the evaluation per concept and in the evaluation per photo. For the evaluation per concept, the EER and the AUC of the ROC curves summarize the performance of the individual runs. The EER is defined as the point where the false acceptance rate of a system is equal to the false rejection rate. These scores were also used in the VCDT task 2008 and allow to compare the results of the different groups to some overlapping concepts. The evaluation per photo is assessed with a recently proposed hierarchical measure [15]. It considers partial matches between system output and ground truth and calculates misclassification costs for each missing or wrongly annotated concept per image. The score is based on structure information (distance between concepts in the

No.	Concept	Train (%)	Test (%)	Annotator Agreement
0	Partylife	3,26	4,18	0.97
1	Familiy_Friends	13,18	15,26	0.91
2	Beach_Holidays	1,56	2,82	0.96
3	Building_Sights	10,74	11,17	0.93
4	Snow	1,66	1,35	1.0
5	Citylife	13,6	15,88	0.90
6	Landscape_Nature	15,92	17,13	0.94
7	Sports	1,56	2,41	0.99
8	Desert	0,36	0,35	0.99
9	Spring	2,02	0,62	0.98
10	Summer	12,82	7,80	0.87
11	Autumn	1,7	1,56	0.98
12	Winter	3,0	1,92	0.99
13	No_Visual_Season	80,46	88,10	0.84
14	Indoor	28,58	24,50	0.90
15	Outdoor	53,42	51,48	0.96
16	No_Visual_Place	18	24,02	0.88
17	Plants	11,4	26,84	0.91
18	Flowers	5,08	4,38	0.95
19	Trees	9,86	12,71	0.94
20	Sky	20,92	27,47	0.92
21	Clouds	12,78	14,57	0.95
22	Water	6,7	11,35	0.97
23	Lake	1,04	1,35	0.98
24	River	1,42	1,83	0.99
25	Sea	3,02	3,30	0.97
26	Mountains	2,12	3,93	0.98
27	Day	54,02	52,28	0.88
28	Night	7,24	6,82	0.96
29	No_Visual_Time	38,76	40,90	0.88
30	Sunny	13,66	15,60	0.88
31	Sunset_Sunrise	5,12	4,45	0.99
32	Canvas	2,28	2,78	0.98
33	Still_Life	8,4	8,40	0.91
34	Macro	5,74	14,72	0.94
35	Portrait	10,3	15,79	0.92
36	Overexposed	0,96	1,61	0.99
37	Underexposed	5,92	4,45	0.92
38	Neutral_Illumination	93,12	93,95	0.89
39	Motion_Blur	2,84	3,30	0.97
40	Out_of_focus	1,98	1,55	0.98
41	Partly_Blurred	30,56	26,51	0.86
42	No_Blur	64,62	68,65	0.83
43	Single_Person	21,74	20,82	0.96
44	Small_Group	8,44	9,18	0.96
45	Big_Group	2,8	3,25	0.98
46	No_Persons	67,02	66,78	0.89
47	Animals	9,22	8,46	0.99
48	Food	2,98	3,77	0.98
49	Vehicle	4,26	7,59	0.97
50	Aesthetic_Impression	16,98	18,04	0.75
51	Overall_Quality	25,82	14,29	0.81
52	Fancy	14,88	13,33	0.84

Table 1: Summary of the frequencies of each concept in the training and test sets. On the right the agreements between annotators are depicted for each concept.
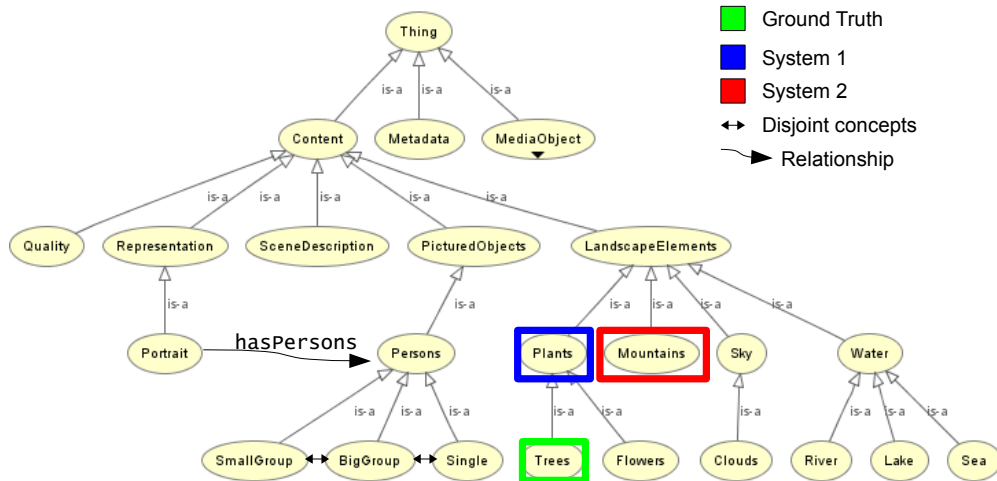
Figure 2: Visualization of an ontology fragment for image annotation. The concepts are hierarchically structured and different types of relationships are exemplarily highlighted.

hierarchy), relationships from the ontology and the agreement between annotators for a concept. The calculation of misclassification costs favours systems that annotate a photo with concepts close to the correct ones more than systems that annotate concepts that are far away in the hierarchy from the correct concepts. (E.g. for the single-label classification case depicted in Fig. 2, system 1 gets lower misclassification costs than system 2.)

3 Results

19 participants submitted results to the LS-VCDT task in altogether 73 runs. The number of runs was restricted to a maximum of 5 runs per group.

In Table 2 the results for the evaluation per concept are illustrated. The team with the best results (ISIS University of Amsterdam) achieves an EER of 23% and an AUC of 84% in average for their best run. One run with pseudo-random numbers was added by the organizers. In this case for each concept a random number between 0 and 1 was generated that denotes the confidence of the annotation for the EER/AUC computation and that was rounded to 0 or 1 for the hierarchical measure per photo. The random numbers achieve an EER and AUC of 50%.

In Table 3 the results for each concept are summarized. In average the concepts could be detected with an EER of 23% and an AUC of 84%. A great amount of these concepts was 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 smaller than 66%.). This is not suprising due to the subjective nature of these concepts which also made the groundtruthing very difficult. The best classified concepts are `Clouds` (AUC: 96%), `Sunset-Sunrise`(AUC: 95%), `Sky`(AUC: 95%) and `Landscape-Nature` (AUC: 94%).

In Table 4 the results for the evaluation per photo are summarized. The classification performance per photo ranges between 69% and 100% with an average of 90%. The best results in terms of HS were achieved by the XRCE group with 83% 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. It seems that some of the groups took the ontology information into account (at least in a post-processing step) and others ignored it. The include of the annotator agreements does not change the results substantially. The scores are a bit worse as the measure is stricter, but the ranking of the groups remains.

		Best Run			Average Runs		
Team	Runs	Rank	EER	AUC	Rank	EER	AUC
ISIS	5	1	0.234	0.839	3.2	0.240	0.833
LEAR	5	5	0.249	0.823	13.2	0.268	0.798
CVIUI2R	2	7	0.253	0.814	9.0	0.255	0.813
FIRST	4	8	0.254	0.817	10.5	0.258	0.803
XRCE	1	14	0.267	0.803	14.0	0.267	0.803
bpacad	5	17	0.292	0.773	20.6	0.312	0.746
MMIS	5	21	0.312	0.744	27.8	0.345	0.699
IAM Southampton	3	23	0.330	0.715	24.7	0.335	0.709
LSIS	5	24	0.331	0.721	42.2	0.418	0.602
LIP6	5	33	0.372	0.673	42.0	0.414	0.554
MRIM	4	34	0.384	0.643	38.0	0.415	0.584
AVEIR	4	41	0.441	0.551	49.8	0.461	0.548
Wroclaw University	5	43	0.446	0.221	45.4	0.449	0.200
KameyamaLab	5	47	0.452	0.164	53.4	0.466	0.133
UAIC	1	54	0.479	0.106	54.0	0.479	0.106
apexlab	3	56	0.483	0.070	60.3	0.487	0.078
INAOE TIA	5	57	0.485	0.099	61.0	0.489	0.080
Random	1	-	0.500	0.499	-	0.500	0.499
CEA LIST	4	68	0.500	0.469	69.5	0.502	0.463
TELECOM ParisTech	2	72	0.526	0.459	72.5	0.527	0.459

Table 2: Summary of the results for the evaluation per concept. The table shows the EER and AUC for the best run per group and the averaged EER and AUC for all runs of one group.

3.1   Submitted Technologies

This subsection gives a brief overview about the submitted technologies of the participants. Further information about each approach can be found in the corresponding papers.

The **IAM group** [9] focuses on visual-terms, which are created by low-level features mainly based on SIFT and a following codebook quantization. For machine learning, the Cross Language Latent Indexing method was applied which maps the concept names and the visual-terms into a semantic space. The decision classification is handled by estimating the smallest cosine distance between concepts and visual-terms. With the different runs, a successive expansion of the concept hierarchy was tried, which results in no improvement.

The algorithm of the **TELECOM ParisTech group** [7] was designed especially for the large-scale scenario, which means a low complexity for the processing and an easy extension to a variety of concepts, by accepting a decrease of precision. The algorithm utilizes global visual features and text features generated out of the 53 visual concepts via PCA. A Canonical Correlation Analysis is used to capture linear relationships between these different features spaces.

The **UPMC/LIP6 group** [6] utilizes a simple HSV histogram feature calculated in 3 horizontal segments and a linear kernel SVM for learning.

The **MRIM-LIG group** [13] combines RGB histograms, SIFT and Gabor features. For the learning phase, different SVM combinations are trained and as a priori, the best feature/SVM setup for each concept is used.

The **LSIS group** [18] combines different features, e.g. HSV, edge, gabor or profile entropy and applies a Visual Dictionary with a visual-word approach.

The **AVEIR submissions** [8] are from a joint group of the individual participants: Telecom ParisTech, LSIS, MRIM-LIG and UPMC/LIP6. The AVEIR submissions seem to be equal to the individual submissions. An efficient and reliable combination or fusion method based on a carefulness index is discussed only theoretically in the working notes.

The **KameyamaLab group** [16] proposed a system with joint global color and texture features as well as local features based on saliency regions. Additionally, a gist of scene feature is used. For the assignment of concept labels, a KNN classifier is applied.

The **ISIS group** [17] applies a system that is based on four main steps. First, a sampling strat-

No.	Concept	Best AUC	Best EER	Group
0	Partylife	0.83	0.24	ISIS
1	Familiy_Friends	0.83	0.24	ISIS
2	Beach_Holidays	0.91	0.16	ISIS
3	Building_Sights	0.88	0.20	ISIS
4	Snow	0.87	0.21	LEAR
5	Citylife	0.83	0.25	ISIS
6	Landscape_Nature	0.94	0.13	ISIS
7	Sports	0.72	0.34	FIRST
8	Desert	0.89	0.18	ISIS
9	Spring	0.83	0.25	FIRST
10	Summer	0.81	0.26	ISIS
11	Autumn	0.87	0.21	ISIS
12	Winter	0.85	0.23	ISIS
13	No_Visual_Season	0.81	0.26	ISIS
14	Indoor	0.84	0.25	ISIS
15	Outdoor	0.90	0.19	ISIS
16	No_Visual_Place	0.79	0.29	ISIS
17	Plants	0.88	0.21	ISIS
18	Flowers	0.87	0.20	ISIS - FIRST
19	Trees	0.90	0.18	ISIS
20	Sky	0.95	0.12	ISIS
21	Clouds	0.96	0.10	ISIS
22	Water	0.90	0.18	ISIS
23	Lake	0.91	0.16	ISIS
24	River	0.90	0.17	ISIS
25	Sea	0.94	0.13	ISIS
26	Mountains	0.93	0.14	ISIS
27	Day	0.85	0.24	ISIS
28	Night	0.91	0.17	LEAR
29	No_Visual_Time	0.84	0.25	ISIS
30	Sunny	0.77	0.30	LEAR - ISIS
31	Sunset_Sunrise	0.95	0.11	ISIS
32	Canvas	0.82	0.25	XRCE
33	Still_Life	0.82	0.25	ISIS
34	Macro	0.81	0.26	ISIS
35	Portrait	0.87	0.21	XRCE - ISIS
36	Overexposed	0.80	0.25	LIP6
37	Underexposed	0.88	0.18	CVIUI2R
38	Neutral_Illumination	0.80	0.26	LEAR
39	Motion_Blur	0.75	0.32	ISIS
40	Out_of_focus	0.81	0.25	LEAR
41	Partly_Blurred	0.86	0.22	LEAR
42	No_Blur	0.85	0.23	LEAR
43	Single_Person	0.79	0.28	ISIS - LEAR
44	Small_Group	0.80	0.28	ISIS
45	Big_Group	0.88	0.21	ISIS
46	No_Persons	0.86	0.22	ISIS
47	Animals	0.83	0.25	ISIS
48	Food	0.90	0.19	ISIS
49	Vehicle	0.83	0.24	ISIS
50	Aesthetic_Impression	0.66	0.38	ISIS
51	Overall_Quality	0.66	0.38	ISIS
52	Fancy	0.58	0.44	ISIS

Table 3: Overview of concepts and results per concept in terms of the best EER and best AUC per concept and the name of the group which achieved these results.

		Best Run			Average Runs		
Team	Runs	Rank	HS	HS*	Rank	HS	HS*
XRCE	1	1	0.829	0.810	1.0	0.829	0.810
CVIUI2R	2	2	0.828	0.808	2.5	0.828	0.808
FIRST	4	4	0.815	0.794	6.0 / 5.75	0.812	0.791
KameyamaLab	5	7	0.809	0.787	25.6 / 25.2	0.690	0.668
LEAR	5	11	0.792	0.769	19.4 / 19.6	0.765	0.740
Wroclaw University	5	12	0.790	0.765	40.2 / 39.0	0.592	0.571
ISIS	5	13	0.783	0.760	17.2 / 17.0	0.773	0.750
apexlab	3	15 / 14	0.779	0.759	30.0 / 29.7	0.699	0.675
INAOE TIA	5	20	0.759	0.732	31.6 / 31.4	0.699	0.671
CEA LIST	4	23 / 24	0.752	0.725	27.3 / 27.5	0.739	0.712
MRIM	4	27 / 28	0.741	0.711	35.5 / 35.8	0.681	0.646
UAIC	1	33	0.724	0.691	33.0	0.724	0.691
bpacad	5	35	0.707	0.678	40.2	0.634	0.602
MMIS	5	42	0.618	0.576	47.8 / 49.0	0.564	0.515
LSIS	5	47 / 49	0.549	0.498	57.0 / 58.0	0.459	0.423
AVEIR	4	51	0.516	0.479	56.3 / 58.0	0.471	0.419
LIP6	5	60 / 59	0.445	0.414	69.2 / 69.0	0.298	0.274
IAM Southampton	3	63 / 61	0.419	0.396	64.7 / 63.3	0.398	0.374
TELECOM ParisTech	2	66 / 64	0.390	0.361	67.5 / 66.5	0.367	0.341
random	1	-	0.384	0.351	-	0.384	0.351

Table 4: Summary of the results for the evaluation per photo. The table illustrates the average hierarchical score (HS) over all photos for the best run per group and the average HS per group. HS* denotes the scores if the annotator agreements are ignored during computation.

egy is applied that combines a spatial pyramid approach and saliency points detection. Second, SIFT features are extracted in different color spaces. To reduce the amount of visual features, a codebook transformation is utilized in the third step and the frequency information of predefined codewords is used as final feature. The final learning step is based on SVM with χ^2 kernel. The runs differ mainly in the number of SIFT features used and the codebook generation.

The **FIRST group** [2] used SIFT features on different color channels and pyramid histograms over color intensities. The SIFT features are combined by the bag of words approach. For classification, a SVM was applied with average kernel, sparse $L1$ MKL and non-sparse L^p MKL kernel.

The **XRCE group** [1] uses a set of different features, a GMM image representation, a Fisher vector and local RGB statistics and SIFT features. The local features are extracted on a multi-level image-grid. For the classification, a Sparse Logistic Regression approach was applied. In the post-processing, the hierarchical structure, disjoint-concepts and relating concepts were considered.

The **SZTAKI[1] group** [3] used SIFT features and a graph-based segmentation algorithm. Based on the segments, color histograms, shape and DFT features are estimated. The SIFT features are post-processed with a GMM, and a Fisher kernel is applied on the features derived from the segmentation. For classification, a binary logistic regression approach is utilized. As one of a few groups the SZTAKI group applied the connections between concepts in the provided ontology among others to estimate correlations of appearing concepts.

The **INRIA-LEAR group** [4] utilizes a bag-of-features setup with global features namely a gist of scene descriptor and different color histograms applied in three horizontal regions of the image and the local SIFT feature quantized with k-means. In two runs a weighted nearest neighbor tag prediction method is applied and in two runs a SVM for each concept is used. The fifth run uses a SVM classifier, trained for multi-class separation. The SVM runs performed better than the tag prediction, whilst the tag prediction was ten times faster. No post-processing on the ontology rules was performed, therefore the results of the HS are worse than the EER.

The **TIA-INAOE group** [5] provided an algorithm based on global features, e.g. color and edge histograms. As baseline run, a KNN classifier is used and the most often appearing concepts in the top nearest neighbor training images were assigned. A further label refinement process

---

[1]SZTAKI equals the bpacad submissions

concentrates on co-occurrence statistics of the disjoint concepts of training set.

The **CVIU I2R group** [14] provides a system that utilizes various global and local features, e.g. color and edge histograms, color coherence vector, census transform and different SIFT features. Furthermore, a local region search algorithm is used to previously select a relevant bounding box for each concept. In combination with a SVM with χ^2 kernel a feature selection process is applied to choose the most relevant features for each concept. For the disjoint concepts, the probabilities were manipulated in order to have a single concept above 0.5.

The **UAIC group** [11] utilizes four modules. First, a face detection software was used to estimate the number of faces in images. Unfortunately, this modules breaks the rules of the LS-VCDT, because the face detector is based on a Viola and Jones detector which was trained with data that was not provided in this task. The second module concentrates on clustering of training images, whereas most concepts were set to a score of 0.5 if no decision could be made. The same process was applied in an EXIF data processing module. The last module sets default values to disjoint concepts depending on their occurrence in the training data.

The **MMIS submissions** [12] utilize global color histogram, Tamura texture and Gabor features. Selected features are estimated in nine subregions and concatenated to the overall feature vector. A non-parametric density estimation is applied and a baseline approach by global feature weighting was submitted. The other four submissions use different parameter combinations for word correlations to semantic similarity. The runs differ by the source of the semantic similarity space, which was estimated by the training data, by Google Web search, WordNet and Wikipedia measure. The submission based on the training data achieved the best results.

Summarizing the approaches, some facts can be driven. The groups that used local features like SIFT achieved better results than the groups relying solely on global features. Most groups that investigated the concept hierarchy and analyzed, e.g. the correlations between the concepts, could achieve better ranks evaluated with the hierarchical measure than with the EER. The information about the computational performance are difficult to compare, because the information range from 72 hours for the complete process, to 1 second for training and testing. In the 2010 task, a more detailed specification for this information is needed.

4 Conclusion

This paper summarises the ImageCLEF 2009 LS-VCDT. Its aim was to automatically annotate photos with 53 concepts in a multilabel scenario. An additionally provided ontology could be used to enrich the classification system. The results show that in average the task could be solved reasonably well with the best system achieving an AUC of 84% for all photos. Four other groups got an AUC score over or equal to 80%. Evaluated on the concept basis, the concepts could be annotated in average with an AUC of 84%. In terms of HS, the best system annotated all photos with an average annotation rate of 83%. Three other systems were very close to these results with 83%, 82% and 81%. Part of the groups used the ontology for post-processing or to learn correlations of concepts. No participant integrated the ontology in a reasoning system and tried to apply this system for the classification task. The large number of concepts and photos posed no problem to the classification systems.

5 Acknowledgment

References

[1] J. Ah-Pine, S. Clinchant, G. Csurka, and Y. Liu. XRCE's Participation in ImageCLEF 2009. *CLEF working notes 2009, Corfu, Greece*, 2009.

[2] A. Binder and M. Kawanabe. Fraunhofer FIRST's Submission to ImageCLEF2009 Photo Annotation Task: Non-sparse Multiple Kernel Learning. *CLEF working notes 2009, Corfu, Greece*, 2009.

[3] B. Daroczy, I. Petras, A.A. Benczur, Z. Fekete, D. Nemeskey, D. Siklosi, and Z. Weiner. SZTAKI @ ImageCLEF 2009. *CLEF working notes 2009, Corfu, Greece*, 2009.

[4] M. Douze, M. Guillaumin, T. Mensink, C. Schmid, and J. Verbeek. INRIA-LEARs participation to ImageCLEF 2009. *CLEF working notes 2009, Corfu, Greece*, 2009.

[5] H.J. Escalante, J.A. Gonzalez, C.A. Hernandez, A. Lopez, M. Montex, E. Morales, E. Ruiz, L.E. Sucar, and L. Villasenor. TIA-INAOE's Participation at ImageCLEF 2009. *CLEF working notes 2009, Corfu, Greece*, 2009.

[6] A. Fakeri-Tabrizi, S. Tollari, L. Denoyer, and P. Gallinari. UPMC/LIP6 at ImageCLEFannotation 2009: Large Scale Visual Concept Detection and Annotation. *CLEF working notes 2009, Corfu, Greece*, 2009.

[7] M. Ferecatu and H. Sahbi. TELECOM ParisTech at ImageClef 2009: Large Scale Visual Concept Detection and Annotation Task. *CLEF working notes 2009, Corfu, Greece*, 2009.

[8] H. Glotin, A. Fakeri-Tabrizi, P. Mulhem, M. Ferecatu, Z. Zhao, S. Tollari, G. Quenot, H. Sahbi, E. Dumont, and P. Gallinari. Comparison of Various AVEIR Visual Concept Detectors with an Index of Carefulness. *CLEF working notes 2009, Corfu, Greece*, 2009.

[9] J.S. Hare and P.H. Lewis. IAM@ImageCLEFPhotoAnnotation 2009: Naive application of a linear-algebraic semantic space. *CLEF working notes 2009, Corfu, Greece*, 2009.

[10] Mark J. Huiskes and Michael S. Lew. The MIR Flickr Retrieval Evaluation. In *MIR '08: Proceedings of the 2008 ACM International Conference on Multimedia Information Retrieval*, New York, NY, USA, 2008. ACM.

[11] A. Iftene, L. Vamanu, and C. Croitoru. UAIC at ImageCLEF 2009 Photo Annotation Task. *CLEF working notes 2009, Corfu, Greece*, 2009.

[12] A. Llorente, S. Little, and S. Rüger. MMIS at ImageCLEF 2009: Non-parametric Density Estimation Algorithms. *CLEF working notes 2009, Corfu, Greece*, 2009.

[13] P. Mulhem, J-P. Chevallet, G. Quenon, and R. Al Batal. MRIM-LIG at ImageCLEF 2009: Photo Retrieval and Photo Annotation tasks. *CLEF working notes 2009, Corfu, Greece*, 2009.

[14] J. Ngiam and H. Goh. I2R ImageCLEF Photo Annotation 2009 Working Notes. *CLEF working notes 2009, Corfu, Greece*, 2009.

[15] S. Nowak and H. Lukashevich. 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*, 2009.

[16] S. Sarin and W. Kameyama. Joint Contribution of Global and Local Features for Image Annotation. *CLEF working notes 2009, Corfu, Greece*, 2009.

[17] K.E.A. van de Sande, T. Gevers, and A.W.M. Smeulders. The University of Amsterdam's Concept Detection System at ImageCLEF 2009. *CLEF working notes 2009, Corfu, Greece*, 2009.

[18] Z-Q. Zhao, H. Glotin, and E. Dumont. LSIS Scale Photo Annotations: Discriminant Features SVM versus Visual Dictionary based on Image Frequency. *CLEF working notes 2009, Corfu, Greece*, 2009.