

The Cross Language Image Retrieval Track ImageCLEF 2009

Henning Müller₁, Barbara Caputo₂, Tatiana Tommasi₂, Theodora Tsikrika₄, Jayashree Kalpathy-Cramer₅, Mark Sanderson₃, Paul Clough₃, Jana Kludas₆, Thomas M. Deserno₇, Stefanie Nowak₈, Peter Dunker₈, Mark Huiskes₉, Monica Lestari Paramita₃, Andrzej Pronobis₁₀, Patric Jensfelt₁₀

¹University and Hospitals of Geneva, Switzerland
 ² Idiap Research Institute, Martigny, Switzerland
 ³Sheffield University, UK, ⁴CWI, The Netherlands
 ⁵Oregon Health Science University, ⁶University of Geneva, Switzerland
 ⁷RWTH Aachen University, Medical Informatics, Germany
 ⁸ Fraunhofer Institute for Digital Media Technology, Ilmenau, Germany
 ⁹ Leiden Institute of Advanced Computer Science, Leiden University, The Netherlands
 ¹⁰ Centre for Autonomous Systems, KTH, Stockholm, Sweden



General overview
 news, participation, problems

CLEF

- Medical Annotation Task
 x-rays & nodules
- Medical Image Retrieval Task
- WikipediaMM Task
- Photo annotation Task
- Photo Retrieval Task
- Robot Vision Task
- Conclusions



General participation

- Total: 84 groups registered, 62 submitted results
 - medical annotation: 7 groups
 - medical retrieval: 17 groups
 - photo annotation: 19 groups
 - photo retrieval: 19 groups
 - robot vision: 7 groups
 - wikipediaMM: 8 groups
- 3 retrieval tasks, 3 purely visual tasks concentrate on language independence
- Collections in English with queries in several languages
 combinations of text and images





- New robot vision task
- New nodule detection task
- Medical retrieval
 new database
- Photo retrieval
 new database
- Photo annotation
 - new database and changes in the task



ImageCLEF Management

New online management system for participants

All users

All Users (1 - 10 of 87)				_	_	-	
6+ 0								
Туре 🔩	Groupname ቱ	E-mail 4	†4	Country 📬	signature OK ↑↓			
Admin	Test	monica_lestari@yahoo.com		United Kingdom	I	Update	Detail	Delete
Admin	Ivan Eggel	ivan.eggel@hevs.ch		Switzerland		Update	Detail	Delete
Admin	MedGIFT	henning.mueller@sim.hcuge	.ch	Switzerland	I	Update	Detail	Delete
Admin	CVM	Theodora.Tsikrika@cwi.nl		Netherlands	I	Update	Detail	Delete
Admin	viperUNIGE	jana.kludas@unige.ch		Switzerland		Update	Detail	Delete
Admin	ldiap	emma@idiap.ch		Switzerland	I	Update	Detail	Delete
Admin	OHSU	kalpathy@ohsu.edu		United States	I	Update	Detail	Delete
Admin	IDMT	nwk@idmt.fraunhofer.de		Germany	I	Update	Detail	Delete
Participant	HUG Geneva	julien.gobeill@sim.hcuge.ch		Switzerland		Update	Detail	Delete
Participant	LIP6	clef@poleia.lip6.fr		France	I	Update	Detail	Delete
	Page: 1 of	9 Go 🕨 🕅 🕅)+C]				

User detail of User No. 39

	Group name	TELECOM ParisTech
	E-mail	sahbi@telecom-paristech.fr
	Address	46 rue Barrault
	Country	France
	Firstname	Hichem
	Lastname	Sahbi
	Phone Number	
	Registration date	Apr 14, 2009
	Account type	Participant
	Signature OK	Ø
1		

Subscribed to the following ImageCLEF subtrack(s)

ImageCLEFphoto-annotation

User has submitted the following runs

Subtrack	† 4	Retrieval Type	†4	Submission Date	†4	Primary Run	†4	Validated ቱ	Detail
ImageCLEFphoto-annota	tion	Mixed		Jun 19, 2009		0		8	Detail
ImageCLEFphoto-annota	tion	Mixed		Jun 19, 2009		8		8	Detail



- Unique access point to all info on the now 7 sub-tasks and information on past events
- Use of a content-management system, so all 15 organizers can edit it directly
- Very appreciated!!
 - 2000 unique accesses per months, >5000 page views, ...
- Access also to collections created in the context of ImageCLEF



ImageCLEF web page

Very international access!

000	Dashboard – Goog	le Analytics
▲ ► C 😥 + ☑http	s://www.google.com/analytics/reporting/?reset=1&id=5625419&	pdr=20090818-20090917 O ^ Q+ Google O
C Konoba Taurvel Channel	Apple Google Maps YouTube Wikipedia News (287) * Popul	ar¥
Google Analytic	S	hen.mueller@gmail.com Settings My Account Help Sign Out
Analytics Settings View Reports	www.imageclef.org	My Analytics Accounts: www.imagecief.org
Dashboard	Export * 🖸 Email	Bela Advanced Segments: Al Visits 💌
8 Visitors	Dashboard	Jan 1, 2009 - Sep 18, 2009 -
S Traffic Sources		
Content	Visits -	Graph by: 🔛 🥅
- ooas	200	200
Custom Reporting Bets	mmmmm	wwwwwwwwwww
Advanced Segments Bets Email	January 5, 2001 January 28, 200 February 28, 20 March 15, 2009 Apr Site Usage	i 7, 2009 April 30, 2009 May 23, 2009 June 15, 2009 July 8, 2009 July 31, 2009 August 23, 200
Help Resources	Junitability 15,889 Visits	42.02% Bounce Rate
Conversion University	Address 44,896 Pageviews	00:03:59 Avg. Time on Site
(*) Common Questions	2.83 Pages/Visit	44.14% % New Visits
	E Visitors Overview	Map Overlay
	200	
		200



ImageCLEF web page

• Very international access!





Medical Image Annotation Task



Medical Image Annotation Task

- Purely Visual Task
- **2005**:
 - 9000 training images / 1000 test images
 - Assign one out of 57 possible labels to each image
- 2006:
 - 10000 training images / 1000 test images
 - Assign one out of 116 possible labels to each image
- **2007**:
 - 11000 training images / 1000 test images
 - Assign a textual label to each image (one out of 116)
- 2008:
 - 12076 training images / 1000 test images
 - more classes (196), unbalancing, use of hierarchy required

2009: A survey of the past experience

12677 training images / 1733 test images



Label Settings

IRMA CODE: DDDD-AAA-BBB-TTT 1121 -127 -720 -500

- D direction: coronal, anterior-posterior, supine
- A anatomy: abdomen, middle, unspec.
- B biosystem: uropoietic system, unspec. unspec.
- T technique: radiography, plain, analog, overview





2005: 22 (11-4-91-7) 2006: 54 2007: 1121-4a0-914-700 2008: 1121-4a0-914-700



2005: 1 (11-1-50-0) 2006: 1 2007: 1123-127-500-000 2008: 1123-127-500-000



2005: 50 (11-2-45-7) 2006: C 2007: CCCC-CCC-CCC-CCC 2008: 1121-230-451-700



2005: C 2006: C 2007: CCCC-CCC-CCC-CCC 2008: 1127-310-600-625

Clutter Class: images belonging to new classes or described with a higher level of detail in the final 2008 setting



Evaluation Criterion

2005/2006:	classified	error score
 capability of the algorithm to make the correct decision 	18	0.0
2007/2008:	$^{21}_{*}$	$1.0 \\ 0.5$
 incomplete codes 		
 not predicting a position is better than a wrong prediction. 	classified	error count
 incorrect prediction in one position invalidates all the later 	463	0.000000
prediction in this axis	46^{*}	0.025531
\circ axes are independent	461	0.051061
 early errors are worse than late ones 	4^{*1}	0.069297
Clutter Class	4^{**}	0.069297
	47^{*}	0.138594
	473	0.138594
does not influence the	477	0.138594
error score	**	0.125000
	731	0.250000

classified 2005-06	error count	classified 2007-08	error count
18	0.0	111	0.000000
21	0.0	11*	0.000000
*	0.0	1**	0.000000
C	0.0	***	0.000000
		C	0.000000



- TAU biomed:Medical Image Processing Lab, Tel Aviv University, Israel
- Idiap: The Idiap Research Institute, Martigny, Switzerland
- FEITIJS: Faculty of Elecrical Engineering and Information Technologies, University of Skopje, Macedonia
- VPA: Computer Vision and Pattern Analysis Laboratory, Sabanci University, Turkey
- medGIFT: University Hospitals of Geneva, Switzerland
- **DEU**: Dokuz Eylul University, Turkey
- IRMA: Medical Informatics, RWTH Aachen University, Aachen, Germany





				best		
	# runs	score 05	score 06	score 07	score 08	score TOT
TAU	1	365	263	64.30	169.50	852.80
Idiap	4	393	260	67.23	178.93	899.16
FEITIJS	1	549	433	128.10	242.46	1352.56
VPA	5	578	462	155.05	261.16	1456.21
MedGIFT	3	618	507	190.73	317.53	1633.26
IRMA	1	790	638	207.55	359.29	1994.84
DEU	4	1368	1183	487.50	642.50	3681.00

Conclusions

- top performing runs do not consider the hierarchical structure of the task;
- local features outperform global ones;
- discriminative SVM classification methods outperform other approaches;
- 2005 -- 06 decrease in error score: 57 wide classes difficult to model;
- 2007 -- 08 increase in error score: increasing number of classes and unbalancing.



Nodule Detection Task



Nodule Detection

- Introduced the lung nodule detection task in 2009.
- CT images LIDC
 - 100–200 slices per study
 - manually annotated by 4 clinicians.
- More than 25 groups had registered for the task
- More than a dozen had downloaded the data sets
- Only two groups submitted three runs



Medical Image Retrieval Task



Medical Retrieval Task

- Updated data set with 74,902 images
- Twenty five ad-hoc topics were made available, ten each that were classified as visual and mixed and five that were textual
- Topics provided in English, French, German
- Five case-based topics were made available for the first time
 - Ionger text with clinical description
 - potentially closer to clinical practice
- 17 groups submitted 124 official runs
- Six groups were first timers!
- Relevance judgments paid using TrebleCLEF and Google grants
 - Many topics had duplicate judgments



- Subset of Goldminer collection
 - Radiology and Radiographics
 - images
 - figure captions
 - access to the full text articles in HTML
 - Medline PMID (PubMed Identifier).
- Well annotated collection, entirely in English
- Topics were supplied in German, French, and English



Ad-hoc topics

- Realistic search topics were identified by surveying actual user needs.
- Google grant funded user study conducted at OHSU during early 2009
- Qualitative study conducted with 37 medical practitioners
- Participants performed a total of 95 searches using textual queries in English.
- Randomly selected 25 candidate queries from the 95 searches to create the topics for ImageCLEFmed 2009



Ad-hoc topics

1 Photos of erythema



jaykc logout

Show Only Non-Judged Pool Entries

Frequency	Topic	Image	Title	Caption	Relevant?
70	1	Link 227775	Common and unusual diseases of the nipple-areolar complex	Subareaolar abscess in a 30-year-old woman with pain and erythema of the areola. Photographand US imageshow a well-circumscribed oval intradermal mass (arrow). The mass appears anechoic in , a finding suggestive of benignity.	Relevant O Partially Relevant O Not Relevant
66	1	link 227776	Common and unusual diseases of the nipple-areolar complex	Subareaolar abscess in a 30-year-old woman with pain and erythema of the areola. Photographand US imageshow a well-circumscribed oval intradermal mass (arrow). The mass appears anechoic in , a finding suggestive of benignity.	Relevant Partially Relevant Not Relevant



Case-based topics

- Scenario: provide clinician with articles from the literature are similar to the case (s)he is working on
- Five topics were created based on cases from the teaching file Casimage.
- The diagnosis and all information about the treatment was removed
- In order to make the judging more consistent, the relevance judges were provided with the original diagnosis for each case.

Case-based topics

A 63 year old female remarked an unpainful mass on the lateral side of her right tight. Five months later she visited her physician because of the persistence of the mass. Clinically, the mass is hard and seems to be adherent to deep planes.

CLEF

RX : there is slight thinning, difficult to perceive, of the outer cortex of the right femur of approximately 3-4 cm in length, situated at the junction of the upper and middle third, without periosteal reaction or soft tissue calcifications. US : demonstrates a 6x4x3cm intramuscular mass of the vastus lateralis. This mass is well delineated, hypoechoic, contains some internal echoes and shows posterior enhanced transmission. MRI : The intramuscular mass of the vastus lateralis is in contact with the femoral cortex. There is thinning of the cortex but no intramedullary invasion.



CLEF P

Participants

- NIH (USA)
- Liris (France)
- ISSR (Egypt)
- UIIP Minsk (Belarus)
- MedGIFT (Switzerland)
- Sierre (Switzerland)
- SINAI (Spain)
- Miracle (Spain)
- BiTeM (Switzerland)

- York University (Canada)
- AUEB (Greece)
- University of Milwaukee (USA)
- University of Alicante (Spain)
- University of North Texas (USA)
- OHSU (USA)
- University of Fresno (USA)
- DEU (Turkey)



Runs submitted

Ad-hoc	Visual	Textual	Mixed
Automatic	15	52	25
Interactive	1	7	3
Manual	0	0	2

Case-based	Visual	Textual	Mixed
Automatic	15	52	25



Topic Analysis

Easy Topics CT Images of an inguinal hernia Lobar pneumonia x-ray Glioblastoma multiforme MR Pneumoconiosis

Difficult Topics Mesothelioma image lung disease, gross or micro pathology Gallbladder histology





Inter-rater agreement

Topic	Judge 1	Judge 2	Kappa
1	3	4	0.341
3	3	11	0.715
7	4	6	0.302
8	6	15	0.639
10	4	12	0.15
13	7	12	0.021
14	11	12	0.0298
15	6	7	0.885
17	3	4	0.821
10			0.004
18	4	15	0.884
18 20	4	15 12	0.884 0.0388
20 Topic	4 7 judge1	15 12 judge2	0.884 0.0388 Kappa
18 20 Topic 26	4 7 judge1 4	15 12 judge2 7	0.884 0.0388 Kappa 0.06
18 20 Topic 26 27	4 7 judge1 4 4	15 12 judge2 7 7	0.884 0.0388 Kappa 0.06 -0.10
18 20 Topic 26 27 28	4 7 judge1 4 4 4	15 12 judge2 7 7 7 11	0.884 0.0388 Kappa 0.06 -0.10 0.37
18 20 Topic 26 27 28 29	4 7 judge1 4 4 4 4 4 4	15 12 judge2 7 7 11 7	0.884 0.0388 Kappa 0.06 -0.10 0.37 -0.25
18 20 Topic 26 27 28 29 29 29	4 7 judge1 4 4 4 4 4 4 4 4	15 12 judge2 7 7 11 7 11 7 11	0.884 0.0388 Kappa 0.06 -0.10 0.37 -0.25 0.13
18 20 Topic 26 27 28 29 29 29 29	4 7 judge1 4 4 4 4 4 4 4 7	15 12 judge2 7 7 11 7 11 7 11 11	0.884 0.0388 Kappa 0.06 -0.10 0.37 -0.25 0.13 0.28

- 16 of 30 topics had multiple judges
- Some judges overly lenient
 - not used for final qrels
- Familiarity with topic seems to impact leniency
- Correlation of measures with different judges depends on level or leniency





Inter-rater agreement

Topic	Judge 1	Judge 2	Kappa
1	3	4	0.341
3	3	11	0.715
7	4	6	0.302
8	6	15	0.639
10	4	12	0.15
13	7	12	0.021
14	11	12	0.0298
15	6	7	0.885
17	3	4	0.821
10	4	1 8	0.004
10	4	15	0.004
20	47	15	0.0388
20 Topic	7 judge1	12 judge2	0.884 0.0388 Kappa
20 Topic 26	4 7 judge1 4	15 12 judge2 7	0.884 0.0388 Kappa 0.06
20 Topic 26 27	4 7 judge1 4 4	15 12 judge2 7 7	0.884 0.0388 Kappa 0.06 -0.10
20 Topic 26 27 28	4 7 judge1 4 4 4	15 12 judge2 7 7 7 11	0.884 0.0388 Kappa 0.06 -0.10 0.37
20 Topic 26 27 28 29	4 7 judge1 4 4 4 4 4 4	15 12 judge2 7 7 11 7	0.884 0.0388 Kappa 0.06 -0.10 0.37 -0.25
20 Topic 26 27 28 29 29 29	4 7 judge1 4 4 4 4 4 4 4 4	15 12 judge2 7 7 11 7 11 7 11	0.884 0.0388 Kappa 0.06 -0.10 0.37 -0.25 0.13
20 Topic 26 27 28 29 29 29 29	4 7 judge1 4 4 4 4 4 4 4 7	15 12 judge2 7 7 11 7 11 7 11 11	0.884 0.0388 Kappa 0.06 -0.10 0.37 -0.25 0.13 0.28

- 16 of 30 topics had multiple judges
- Some judges overly lenient
 - not used for final qrels
- Familiarity with topic seems to impact leniency
- Correlation of measures with different judges depends on level or leniency





Conclusions

- Focus for this year was text-based retrieval (again!)
 - Almost twice as many text-based runs compared to multimedia runs
- As in 2007 and 2008, purely textual retrieval had the best overall run
 - Purely textual runs performed well (MAP >0.42)
 - Purely visual runs performed poorly
- Combining text with visual retrieval can improve early precision
 - Combinations not (always) robust
- Semantic topics combined with a database containing high quality annotations in 2008 and 2009
 - less impact of using visual techniques as compared to previous years.



Wikipedia Retrieval Task



WikipediaMM Task

• History:

- 2008 wikipediaMM task @ ImageCLEF
- 2006/2007 MM track @ INEX

• Description:

- ad-hoc image retrieval
- collection of Wikipedia images
 - Iarge-scale
 - heterogeneous
 - user-generated annotations
- diverse multimedia information needs

• **Aim**:

- investigate mono-media and multi-modal retrieval approaches
 - focus on fusion/combination of evidence from different modalities
- attract researchers from both text and visual retrieval communities
- support participation through provision of appropriate resources



wikipediaMM Collection

- 151,590 images
 - wide variety
 - global scope
 - JPEG, PNG formats
- Annotations
 - user-generated
 - highly heterogeneous
 - varying length
 - noisy
 - semi-structured
 - monolingual (English)



wikipediaMM Topics

<topic>

CLEF

<number>108 </number>

<title> bird nest </title>

<image> http://chiotsrun.com/.../birds-nest-at-moms.jpg </image>

<image> http://farm1.static.flickr.com/...21a4.jpg </image>

<narrative> We search for photos of birds and their nests, empty nests are also relevant. Food that has 'nest' in their name is not relevant, neither is the Beijing Olympic stadium.

</topic>

Number of topics	45	
average # of images/topic	1.85	
average # of terms/topic	2.94	
average # of relevant docs	36.03	
Average # words/reldocs	37.96	





- range from easy (eg. 'bikes') to difficult highly semantic topics (e.g. 'aerial photos of non-artificial landscapes')
- challenging for current state-of-the-art retrieval algorithms



wikipediaMM Participation

- 32 groups registered
- 8 groups submitted a total of 57 runs

	TD	А	S
CEA, LIC2M-CEA, Centre CEA de Saclay, France	х	х	х
DCU, Dublin City University, School of Computing, Ireland			х
DEU, Dokuz Eylul University, Dep. of Computer Engineering, Turkey	1	х	х
IIT-Hyderabad, Search and Info Extraction Lab, India			х
AHC, Laboratoire Hubert Curien, UMR CNRS, France		х	х
SZTAKI, Hungarian Academy of Science, Hungary		х	х
SINAI, Intelligent Systems, University of Jaen, Spain			х
UALICANTE, Software & Computer Systems, Uni. of Alicante, Spain	n	х	х
UNIGE, University of Geneva, CVML, Switzerland	х	х	
CWI, Interactive Information Access Group, Netherlands	х	Х	

Participation in topic development (TD), assessment (A) and submission (S)



	Participant	Run	Modality	Feedback/Expansion	MAP	P@10	P@20	R-prec.	Bpref
1	deuceng	deuwiki2009_205	TXT	QE	0.2397	0.4000	0.3133	0.2683	0.2191
2	deuceng	deuwiki2009_204	TXT	QE	0.2375	0.4000	0.3111	0.2692	0.2170
3	deuceng	deuwiki2009_202	TXT	QE	0.2358	0.3933	0.3189	0.2708	0.2217
4	lach	run_TXTIMG_100_3_1_5_meanstd	TXTIMG	NOFB	0.2178	0.3378	0.2811	0.2538	0.2006
5	lach	run_TXTIMG_50_3_1_5_meanstd	TXTIMG	NOFB	0.2148	0.3356	0.2867	0.2536	0.2023
6	cea	cealateblock	TXTIMG	QE	0.2051	0.3622	0.2744	0.2388	0.1938
7	cea	ceaearlyblock	TXTIMG	QE	0.2046	0.3556	0.2833	0.2439	0.2014
8	cea	ceabofblock	TXTIMG	QE	0.1975	0.3689	0.2789	0.2342	0.1886
9	cea	ceatlepblock	TXTIMG	QE	0.1959	0.3467	0.2733	0.2236	0.1847
10	cea	ceabofblockres	TXTIMG	QE	0.1949	0.3689	0.2789	0.2357	0.1890
11	cea	ceatlepblockres	TXTIMG	QE	0.1934	0.3467	0.2733	0.2236	0.1847
12	lach	run_TXTIMG_Siftdense_0.084	TXTIMG	NOFB	0.1903	0.3111	0.2700	0.2324	0.1828
13	lach	run_TXT_100_3_1_5	TXT	NOFB	0.1890	0.2956	0.2544	0.2179	0.1687
14	lach	run_TXT_50_3_1_5	TXT	NOFB	0.1880	0.3000	0.2489	0.2145	0.1715
15	ualicante	Alicante-MMLCA	TXTIMG	FB	0.1878	0.2733	0.2478	0.2138	0.1734

Conclusions:

CLEF

- best performing run: a text-based approach
- half of the submissions combine text and visual evidence (29/57)
- groups with mono-media and multi-modal runs: mm runs always outperform their text-based runs
- multi-modal runs outperform mono-media runs on average
- many (successful) query/document expansion submissions
- participants willing to help voluntarily in assessment



Photo Annotation Task



Photo Annotation Task

- Large-Scale Visual Concept Detection Task (LS-VCDT)
 - annotate the photos with depicted visual concepts
 - provided real-world knowledge
- Main Challenges:
 - Can image classifiers scale to the large amount of concepts and data?
 - Can an ontology (hierarchy and relations) help in large scale annotations?
- Participation:
 - 40 groups registered
 - 19 groups submitted



LS-VCDT: Dataset

- MIR Flickr 25.000 Image Dataset
- 53 visual concepts
 - Most: holistic visual concepts
- Organization in a Photo Tagging Ontology
- Annotation Format:
 - Plain text format
 - Rdf-xml
- Trainingset: 5.000 photos + EXIF data + ground truth annotations
- Testset: 13.000 photos + EXIF data



Citylife Outdoor Night Underexposed Vehicle No_Blur No_Persons No_Visual_Season

_S-VCDT: Annotation Process

1. Annotation Step

- 18.000 photos annotated
- 43 persons (min 30 photos, max 2500 photos)
- Guideline for annotation

2. Validation Step

- 3 persons
- Screening of photos

 a) annotated with X
 b) not annotated with X
- 3. Annotator Agreements

LS-VCDT: Evaluation Measures

1) Evaluation per concept

- Equal Error Rate and Area Under Curve
 - AUC: average 84% per concept
 - EER: average 23% per concept
- 2) Evaluation per photo
 - Correlation between ground truth and annotated label set for each photo
 - Hierarchy of concepts
 - Domain knowledge
 - Annotator agreements
 - between 69% -100% per photo, average 90%

LS-VCDT: Results

TEAM	Best RANK	Best EER	Best AUC	Team	Rank	HS	HS*
ISIS	1	0.234	0.839	XRCE	1	0.829	0.810
LEAR	5	0.249	0.823	CVIUI2R	2	0.828	0.808
CVIUI2R	7	0.253	0.814	FIRST	4	0.815	0.794
FIRST	8	0.254	0.817	KameyamaLab	7	0.809	0.787
XRCE	14	0.267	0.803	LEAR	11	0.792	0.769
bpacad	17	0.292	0.773	Wroclaw University	12	0.790	0.765
MMIS	21	0.312	0.744	ISIS	13	0.783	0.760
IAM Southampton	23	0.330	0.715	apexlab	15 / 14	0.779	0.759
LSIS	24	0.331	0.721	INAOE TIA	20	0.759	0.732
LIP6	33	0.372	0.673	CEA LIST	23 / 24	0.752	0.725
MRIM	34	0.384	0.643	MRIM	27 / 28	0.741	0.711
AVEIR	41	0.441	0.551	UAIC	33	0.724	0.691
Wroclaw University	43	0.446	0.221	bpacad	35	0.707	0.678
KameyamaLab	47	0.452	0.164	MMIS	42	0.618	0.576
UAIC	54	0.479	0.106	LSIS	47 / 49	0.549	0.498
apexlab	56	0.483	0.070	AVEIR	51	0.516	0.479
INAOE TIA	57	0.485	0.099	LIP6	60 / 59	0.445	0.414
Random	-	0.500	0.499	IAM Southampton	63 / 61	0.419	0.396
CEA LIST	68	0.500	0.469	TELECOM ParisTech	66 / 64	0.390	0.361
TELECOM ParisTech	72	0.526	0.459	random	-	0.384	0.351

LS-VCDT: Results

Results per Concept (AUC):

- Detection of landscape elements very well
- Detection of aesthetic concepts bad

- LS-VCDT 2009:
 - 84% AUC average over 53 concepts on 13.000 photos
- VCDT 2008:
 - 90,8% AUC average over 17 concepts on 1.000 photos
- Ontology knowledge (links) rarely used
 - only as post-processing step, not for learning

Photo Retrieval Task

Photo Retrieval Task

- Task:
 - study diversity for image retrieval
 - present as many diverse results in the top 10 results
- Collection:
 - Belga data set
 - 498,039 images with unstructured caption (English)
 - 25 times larger than previous year's collection
 - 50 topics
 - 25 topics containing cluster titles and 25 topics without cluster titles
 - Average of 3.96 clusters for each topic
 - Average of 208.49 relevant documents per clusters
- 44 institutions registered, which was the highest number for this task

Photo Retrieval Topics

Query Part 1	Query Part 2
<title> clinton </title>	<title> obama </title>
<clustertitle> hillary clinton </clustertitle>	
<clusterdesc> Relevant images show photographs of Hillary Clinton. Images of Hillary with other people are relevant if she is shown in the foreground. Images of her in the background are irrelevant. </clusterdesc>	
<image/> belga26/05859430.jpg	<image/> belga30/06098170.jpg
<clustertitle> obama clinton </clustertitle>	
<clusterdesc> Relevant images show photographs of Obama and Clinton. Images of those two with other people are relevant if they are shown in the foreground. Images of them in the background are irrelevant. <!--<br-->clusterDesc></clusterdesc>	
<image/> belga28/06019914.jpg	<image/> belga28/06019914.jpg
<clustertitle> bill clinton </clustertitle>	
<clusterdesc> Relevant images show photographs of Bill Clinton. Images of Bill with other people are relevant if he is shown in the foreground. Images of him in the background are irrelevant. </clusterdesc>	
<image/> belga44/00085275.jpg	<image/> belga30/06107499.jpg

Participation

- 19 groups submitting 84 runs
- Choice of Modality:
 - TXT-IMG: 36 runs
 - TXT: 41 runs
 - IMG: 7 runs
- Choice of Tags
 - Title
 - Cluster Title
 - Cluster Description
 - Image

- Evaluation Measure: P@10, CR@10, F1
- Top 10 Runs for All Queries

No	Group	Run Name	Query	Modality	P@10	CR@10	F1
1	XEROX-SAS	XRCEXKNND	T-CT-I	TXT-IMG	0.794	0.824	0.809
2	XEROX-SAS	XRCECLUST	T-CT-I	TXT-IMG	0.772	0.818	0.794
3	XEROX-SAS	KNND	T-CT-I	TXT-IMG	0.8	0.727	0.762
4	INRIA	LEAR5_TI_TXTIMG	T-I	TXT-IMG	0.798	0.729	0.762
5	INRIA	LEAR1_TI_TXTIMG	T-I	TXT-IMG	0.776	0.741	0.758
6	InfoComm	LRI2R_TI_TXT	T-I	TXT	0.848	0.671	0.749
7	XEROX-SAS	XRCE1	T-CT-I	TXT-IMG	0.78	0.711	0.744
8	INRIA	LEAR2_TI_TXTIMG	T-I	TXT-IMG	0.772	0.706	0.737
9	Southampton	SOTON2_T_CT_TXT	T-CT	TXT	0.8240	0.654	0.729
10	Southampton	SOTON2_T_CT_TXT_IMG	T-CT	TXT-IMG	0.746	0.71	0.727

• Query Category

Queries	P@10 Mean	P@10 SD	CR@10 Mean	CR@10 SD	F1 Mean	F1 SD
All Queries	0.655	0.209	0.547	0.137	0.585	0.166
Query Part 1	0.677	0.221	0.558	0.164	0.6	0.182
- Query Part 1 with CT	0.685	0.2	0.594	0.159	0.625	0.17
- Query Part 1 without CT	0.664	0.254	0.5	0.157	0.558	0.196
Query Part 2	0.632	0.219	0.542	0.133	0.569	0.173

• Modality

Modality	Number of Runs	P@10 Mean	P@10 SD	CR@10 Mean	CR@10 SD	F1 Mean	F1 SD
TXT-IMG	36	0.713	0.116	0.612	0.107	0.656	0.102
TXT	41	0.698	0.142	0.539	0.094	0.598	0.096
IMG	7	0.103	0.027	0.254	0.079	0.146	0.04

Query Tags	Runs	Mean F1	1
T-CT-I	9	0.7288	2 0.9 2 0.8
T-I	7	0.7171	0.7
CT-I	2	0.6925	0.6
СТ	2	0.6687	
T-CT	15	0.6233	
T-CT-CD	9	0.5688	
T-CT-CD-I	15	0.4689	0
Т	17	0.5462	0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.
l	8	0.1786	F1-measure for Query Part 1

Conclusion

- The development of new collection has provided a more realistic framework to evaluate diversity further
- Cluster information is essential for providing diverse results
- When cluster information is not available, image examples are valuable to detect the diversity need
- A combination of T-CT-I maximizes diversity
- Using mixed modality achieved the highest F1

Robot Vision Task

Robot Vision - Intro

- New task in the ImageCLEF 2009 campaign
- Addressed the problem of visual place recognition for robot topological localization

 Considerable attention: 19 inscribed groups, 7 groups participating, 27 submitted runs

• Sequences of images acquired using mobile robot platform

- Sequences of images acquired using mobile robot platform
- Divided into training / validation / testing sequences
- Training/validation sequences acquired within 5 room subsection of an office environment

•

One-person office

Corridor

Two-persons office

Kitchen

Printer area

- Additional new rooms in the testing sequence
- Images labeled with the room ID based on the robot position

• Appearance captured under three illumination settings: cloudy weather, sunny weather, night

• Across a time span of almost two years

Robot Vision - Task

- Task:
 - Determine the topological location of a robot for each image in a single unlabeled test image sequence
 - Indicate new rooms not present in the training set
- Training: a single labeled image sequence acquired under (possibly) different illumination, 6-20 months earlier
- Two sub-tasks:
 - Obligatory classify each image independently (global topological localization)
 - Optional exploit continuity of the sequence
- Score based on the number of correctly classified images
 - punishment for incorrect classification

Robot Vision - Participants

- Multimedia Information Retrieval Group, University of Glasgow, United Kingdom
- Idiap Research Institute, Martigny, Switzerland

CLEF

- Faculty of Computer Science, The Alexandru Ioan Cuza University (UAIC), Iaşi, Romania
- Computer Vision & Image Understanding Department (CVIU), Institute for Infocomm Research, Singapore
- Laboratoire des Sciences de l'Information et des Systèmes (LSIS), France
- Intelligent Systems and Data Mining Group (SIMD), University of Castilla-La Mancha, Albacete, Spain
- Multimedia Information Modeling and Retrieval Group (MRIM), Laboratoire d'Informatique de Grenoble, France

Robot Vision - Results

(0	() Obligatory	task.
#	Group	Score
1	Glasgow	890.5
2	Idiap	793.0
3	UAIC	787.0
4	UAIC	787.0
5	CVIU	784.0
6	Glasgow	650.5
7	UAIC	599.5
8	UAIC	599.5
9	LSIS	544.0
10	SIMD	511.0
:	:	:

(a) Obligatory tools

(b) Optional task.

#	Group	Score
1	SIMD	916.5
2	CVIU	884.5
3	Idiap	853.0
4	SIMD	711.0
5	SIMD	711.0
6	SIMD	609.0

•Winners:

- Multimedia Information Retrieval Group, University of Glasgow, United Kingdom
- Intelligent Systems and Data Mining Group (SIMD), University of Castilla-La Mancha, Albacete, Spain

Robot Vision - Conclusions

- The first RobotVision tasks attracted a considerable attention
- An interesting complement to the existing tasks
- Diverse and original approaches to the place recognition problem
- Local-feature based approaches dominate
 illumination filtering can improve results
- Unknown class detection is a difficult problem
- We plan to continue the task in the next years
 - Introducing new challenges (categorization)
 - Adding new sources of information (laser, odometry)
 - bridging the gap between robot vision and other tasks

ImageCLEF 2009 Parallel Session

- Thursday October 1, 5PM
- Ballroom
 - Presentations from each task

Breakout session Friday noon Ballroom
 Discussion and feedback

Problems/Issues

 Photo Annotation Task Ontology knowledge only used for post-processing • Wikipedia task Visual baseline similarity scores were provided late and a bit buggy Medical Annotation task Not many participants, no significant improvement over previous years Lung detection task • Too few runs. Not enough interest? Too difficult? Medical retrieval task Did not provide general baselines Robot Vision Part of one task was very difficult (unknown classes) • Photo retrieval task • Evaluation measure for diversity

Highlights of ImageCLEF

Record participation in most sub-tasks

 New task with many participants
 Many ImageCLEF first timers

 Text-based retrieval still superior for many task

 Multimodal runs often improve purely textual runs (Wiki, Photo)

- Higher early precision with multi-modality over textual runs
- Case-based medical retrieval with some very good results
- New "retrieval" approaches in robot vision task

- Several Ideas for next year!
- What do you expect?
- What are our ideas?
- What data is available?
- Breakout Session:
- Fill in the survey

 www.imageclef.org/survey

Future Plans

- ICPR contest accepted
 - ImageCLEF 2009 data
 - Another try with interactive retrieval
- Tasks that will continue
 - Medical retrieval
 - Wikipedia task (maybe sharing other databases)
 - Robot vision
 - Photo annotation
- Task that will stop
 - medical annotation
- Uncertain
 - Lung nodule
 - Photo retrieval