The Cross Language Image Retrieval Track:

ImageCLEF 2008

Thomas Deselaers, Henning Müller, Thomas Arni, Theodora Tsikrika, Jayashree Kalpathy-Cramer, Mark Sanderson, Paul Clough, Jana Kludas, Allan Hanbury, Michael Grubinger, Thomas M. Deserno, Bill Hersh

1 RWTH Aachen University, Computer Science Dep. Germany
2 University and Hospitals of Geneva, Switzerland
3 Sheffield University, UK
4 CWI, The Netherlands
5 Oregon Health Science University
6 University of Geneva, Switzerland
7 Vienna University, Austria
8 Victoria University, Australia
9 RWTH Aachen University, Medical Informatics, Germany
General overview
  - Participation
  - Problems
Photo retrieval task
Medical image retrieval task
WikipediaMM Task
Visual Concept Detection Task
Medical image annotation
Conclusions
General participation and news

- **Total: 63 groups registered**
- Photo Retrieval: 24 groups, 1042 runs
- Medical Retrieval: 15 groups, 111 runs
- WikipediaMM Retrieval: 12 groups, 77 runs
- Visual Concept Detection: 11 groups, 53 runs
- Medical Image Annotation: 6 groups, 24 runs

**News:**
- WikipediaMM task
- visual concept detection task
- diversity-based ranking for photo retrieval
Goals:
- Address the **growing need for diversity**
- Allows to **measure diversity**
- Make participation **straightforward**
- Attract both **conceptual and visual teams**

Task:
- **Promote Diversity**
- Top 20 results should contain:
  - Maximum number of relevant images
  - Relevant images from as many different clusters as possible
## Participation

### 2008
- **24 groups**
- **1042 runs**

### 2007
- **20 groups**
- **616 runs**

### 2006
- **12 groups**
- **157 runs**

### 2005
- **11 groups**
- **349 runs**

- **AVEIR** - Joint project of the four French labs: LIG, LIP6, LSIS, PTECH - 4 runs
- **Budapest-ACAD** - Computer and Automation Research Institute, Hungarian Academy of Sciences, Budapest, Hungary - 8 runs
- **CLAC** - Computational Linguistics at Concordia (ClaC) Lab, Concordia University, Montreal, Canada – 6 runs
- **CUT** - Chemnitz University of Technology, Chemnitz, Germany – 4 runs
- **DCU** - School of Computing, Dublin City University, Dublin Ireland – 733 runs
- **GITS** - KAMEYAMA Lab, GITS, Waseda University, Japan – 4 runs
- **INAEOE** - National Institute of Astrophysics, Optics and Electronics, Computer Science Department, Puebla, Mexico – 16 runs
- **IPAL** - Image Perception, Access & Language (IPAL), Singapore & National Center for Scientific Research, France & Institute for Infocomm Research, Singapore & University of Joseph Fourier, Grenoble, France – 10 runs
- **LIG** - Laboratory of Informatics of Grenoble (LIG), Grenoble, France – 4 runs
- **LSIS** - System and Information Sciences Lab, France – 15 runs
- **Meiji** - Department of Computer Science, Meiji University, Japan – 8 runs
- **MirFI** - Computer Science Faculty, Daedalus, Madrid, Spain – 41 runs
- **MirGSI** - Intelligent System Group, Daedalus, Madrid, Spain - 14 runs
- **MMIS** - Imperial College London & Open University, UK – 9 runs
- **NII** - National Institute of Informatics, Tokyo, Japan – 10 runs
- **NTU** - National Taiwan University, Taipei, Taiwan – 7 runs
- **Ottawa** - School of Information Technology and Engineering, University of Ottawa, Canada - 13 runs
- **PTECH** - Institut TELECOM, TELECOM ParisTech, Paris, France – 15 runs
- **Shef** - Department of Information Studies, University of Sheffield, Sheffield, UK – 37 runs
- **SINAI** - Sinai group of the University of Jaén, Jaén, Spain – 6 runs
- **TEXMESS** - Department of Software and Computing Systems, University of Alicante, Spain & University of Jaén, Jaén, Spain – 17 runs
- **UA-GPLSI** - Department of Software and Computing Systems, University of Alicante, Spain – 18 runs
- **UPMC** - Pierre & Marie Curie University, Paris, France -
- **XRCE** - Xerox Research Centre Europe - 28 runs
Collection: IAPR TC-12 Benchmark

- 20,000 colour photographs
- Accompanied by semi-structured captions
  - English and Random
- Many images have similar visual content but varying
  - illumination
  - viewing angle
  - background
- Used in ImageCLEF in 2006, 2007
Group photo with Machu Picchu and Huayna Picchu in the background
tourists are sitting on a
grey gravel road in the foreground;
a ruin with grey walls and many green terraces and a distinctive, rocky, steep mountain behind it;
a wooden mountain range and white clouds in the background;
Machu Picchu, Peru
26 October 2004
● 39 topics with full information
  ○ Based on realistic topics (log-file analysis and interviews)
● Available in English only
● Augmented by a cluster tag
  ○ defines how the rel. images should be clustered

Sample topic images:
## Participation

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Type</th>
<th>2008</th>
<th>2007</th>
<th>2006</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Runs</td>
<td>Groups</td>
<td>Runs</td>
</tr>
<tr>
<td>Annotation language</td>
<td>EN</td>
<td>514</td>
<td>24</td>
<td>271</td>
</tr>
<tr>
<td></td>
<td>RND</td>
<td>495</td>
<td>2</td>
<td>32</td>
</tr>
<tr>
<td>Modality</td>
<td>Text Only</td>
<td>404</td>
<td>22</td>
<td>167</td>
</tr>
<tr>
<td></td>
<td>Mixed (text and image)</td>
<td>605</td>
<td>19</td>
<td>255</td>
</tr>
<tr>
<td></td>
<td>Image Only</td>
<td>33</td>
<td>11</td>
<td>52</td>
</tr>
<tr>
<td>Run type</td>
<td>Manual</td>
<td>3</td>
<td>1</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>Automatic</td>
<td>1039</td>
<td>25</td>
<td>455</td>
</tr>
</tbody>
</table>
S-Recall = Subtopic-Recall = Cluster Recall

\[ K \bigcup_{i=i}^{K} \text{subtopics}(d_i) \]

\[ n_A \]

F-Measure = Harmonic mean

\[ \frac{2 \times (P20 \times CR20)}{(P20 + CR20)} \]
Results: Modality Overview

- 33 runs
- 399 runs
- 607 runs
Results: Modality Overview

The graph shows the mean F-Measure for different topics across various modalities. The x-axis represents the topic number, ranging from 2 to 60. The y-axis represents the mean F-Measure, ranging from 0.000 to 0.700. The modalities include TXT, IMG, and TXT IMG, each represented by a different line color.

- **TXT**: Blue line
- **IMG**: Green line
- **TXT IMG**: Red line

The graph indicates fluctuations in F-Measure across topics for each modality, with some topics showing higher peak values than others.
Results: Annotation Language

- EN: 511 runs
- RND: 495 runs
- IMG: 33 runs
Conclusions and Findings

- Choice of annotation language is almost negligible
- Combining concept and content-based retrieval methods can improve retrieval performance
- Purely visual runs performed poorly
- More participants than ever used visual retrieval techniques
- Record number of participants
Medical retrieval 2008

- New data set with almost 66,000 images
- Thirty topics were made available, ten in each of three categories: visual, mixed, and semantic
- 15 groups submitted 111 official runs
- Relevance judgments paid by NSF grant
Database used

- 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
The topics used in 2008 were a subset of the 85 topics used in 2005-2007.

Pulmonary embolism all modalities.
Lungenembolie alle Modalitäten.
Embolie pulmonaire, toutes les formes.

Show me Doppler ultrasound images (colored).
Participants in 2008

- Hungarian Academy of Sciences, Budapest, Hungary
- National Library of Medicine (NLM), National Institutes of Health NIH, Bethesda, MD, USA
- Bania Luka University, Bosnia-Hercegovina;
- MedGIFT group, University of Geneva, Switzerland
- Natural Language Processing group, University Hospitals of Geneva, CH
- GPLSI group, University of Alicante, Spain
- Multimedia Modelling Group, LIG, Grenoble, France
- Natural Language Processing at UNED. Madrid, Spain
- Miracle group, Spain
- Oregon Health and Science University (OHSU), Portland, OR, USA
- IRIT Toulouse, France
- University of Jaen, Spain
- Tel Aviv University, Israel
- National University of Bogota, Colombia
- TextMess group, University of Alicante, Spain
<table>
<thead>
<tr>
<th>Category</th>
<th>Visual</th>
<th>Textual</th>
<th>Mixed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automatic</td>
<td>8</td>
<td>65</td>
<td>31</td>
</tr>
<tr>
<td>Interactive</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Manual</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>
Histogram of MAP for mixed and textual runs

- X-axis: MAP
- Y-axis: Frequency

- Black bars represent mixed runs.
- Grey bars represent textual runs.

The histogram shows the distribution of MAP values for both mixed and textual runs, with a peak for mixed runs around MAP 0.1.
Topic Analysis

Average MAP by topic

<table>
<thead>
<tr>
<th>Topic #</th>
<th>visual</th>
<th>mixed</th>
<th>semantic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.05</td>
<td>0.10</td>
<td>0.08</td>
</tr>
<tr>
<td>2</td>
<td>0.12</td>
<td>0.15</td>
<td>0.20</td>
</tr>
<tr>
<td>3</td>
<td>0.18</td>
<td>0.22</td>
<td>0.35</td>
</tr>
<tr>
<td>4</td>
<td>0.07</td>
<td>0.11</td>
<td>0.10</td>
</tr>
<tr>
<td>5</td>
<td>0.13</td>
<td>0.17</td>
<td>0.18</td>
</tr>
<tr>
<td>6</td>
<td>0.09</td>
<td>0.14</td>
<td>0.12</td>
</tr>
<tr>
<td>7</td>
<td>0.10</td>
<td>0.16</td>
<td>0.11</td>
</tr>
<tr>
<td>8</td>
<td>0.17</td>
<td>0.23</td>
<td>0.21</td>
</tr>
<tr>
<td>9</td>
<td>0.25</td>
<td>0.30</td>
<td>0.29</td>
</tr>
<tr>
<td>10</td>
<td>0.19</td>
<td>0.24</td>
<td>0.20</td>
</tr>
<tr>
<td>11</td>
<td>0.16</td>
<td>0.22</td>
<td>0.15</td>
</tr>
<tr>
<td>12</td>
<td>0.12</td>
<td>0.17</td>
<td>0.10</td>
</tr>
<tr>
<td>13</td>
<td>0.08</td>
<td>0.13</td>
<td>0.08</td>
</tr>
<tr>
<td>14</td>
<td>0.11</td>
<td>0.16</td>
<td>0.11</td>
</tr>
<tr>
<td>15</td>
<td>0.09</td>
<td>0.14</td>
<td>0.09</td>
</tr>
<tr>
<td>16</td>
<td>0.10</td>
<td>0.15</td>
<td>0.10</td>
</tr>
<tr>
<td>17</td>
<td>0.15</td>
<td>0.20</td>
<td>0.15</td>
</tr>
<tr>
<td>18</td>
<td>0.18</td>
<td>0.23</td>
<td>0.18</td>
</tr>
<tr>
<td>19</td>
<td>0.22</td>
<td>0.27</td>
<td>0.22</td>
</tr>
<tr>
<td>20</td>
<td>0.25</td>
<td>0.30</td>
<td>0.25</td>
</tr>
<tr>
<td>21</td>
<td>0.20</td>
<td>0.25</td>
<td>0.15</td>
</tr>
<tr>
<td>22</td>
<td>0.15</td>
<td>0.20</td>
<td>0.10</td>
</tr>
<tr>
<td>23</td>
<td>0.18</td>
<td>0.23</td>
<td>0.18</td>
</tr>
<tr>
<td>24</td>
<td>0.21</td>
<td>0.26</td>
<td>0.21</td>
</tr>
<tr>
<td>25</td>
<td>0.24</td>
<td>0.29</td>
<td>0.24</td>
</tr>
<tr>
<td>26</td>
<td>0.27</td>
<td>0.32</td>
<td>0.27</td>
</tr>
<tr>
<td>27</td>
<td>0.30</td>
<td>0.35</td>
<td>0.30</td>
</tr>
<tr>
<td>28</td>
<td>0.29</td>
<td>0.34</td>
<td>0.29</td>
</tr>
<tr>
<td>29</td>
<td>0.32</td>
<td>0.37</td>
<td>0.32</td>
</tr>
<tr>
<td>30</td>
<td>0.35</td>
<td>0.40</td>
<td>0.35</td>
</tr>
</tbody>
</table>
Four topics were each judged by two judges

Kappa measurements

<table>
<thead>
<tr>
<th>Topic</th>
<th>Judge 1</th>
<th>Judge 2</th>
<th>Strict kappa</th>
<th>Lenient Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>User 3</td>
<td>User 4</td>
<td>0.91</td>
<td>0.95</td>
</tr>
<tr>
<td>5</td>
<td>User 5</td>
<td>User 7</td>
<td>0.7</td>
<td>0.79</td>
</tr>
<tr>
<td>6</td>
<td>User 3</td>
<td>User 5</td>
<td>0.48</td>
<td>0.48</td>
</tr>
<tr>
<td>25</td>
<td>User 7</td>
<td>User 10</td>
<td>0.69</td>
<td>0.7</td>
</tr>
</tbody>
</table>
Conclusions

- Focus for this year was text-based retrieval
  - Almost twice as many text-based runs compared to multi-media runs
  - Most groups performed better on the semantic topics than visual or mixed topics
  - As in 2007, purely textual retrieval had the best overall run
    - Mixed runs performed worse than corresponding textual run
  - Purely visual runs performed poorly
  - Combining text with visual retrieval can improve early precision
    - Combinations can be fragile
  - Semantic topics combined with a database containing high quality annotations in 2008
    - less impact of using visual techniques as compared to previous years.
Our goal in the upcoming ImageCLEF medical retrieval task is to increase the number of visual runs or mixed submitted.

- Modify the task to favor more integrated approaches.

Interactive retrieval has always had poor participation

- Relevance feedback and query modification have a potential to significantly improve results

Transition to more “find similar case”

- Same database?
- Database with annotations for regions of interest?
wikipediaMM: Task

**History:**
- 2008 wikipediaMM task @ ImageCLEF
- 2007 MM track @ INEX
- 2006 MM track @ INEX

**Description:**
- ad-hoc image retrieval
- collection of Wikipedia images
  - large-scale
  - heterogeneous
  - user-generated annotations
  - availability of multi-lingual data
- diverse multimedia information needs

**Aim:**
- investigate mono-media and cross-media 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: Yearly cycle

- CFP
- Topic development
- Data release
- Task definition
- Experiments Submissions
- Relevance assessments
- Evaluation Results
- Results analysis
- Workshop

Organisers
Participants
Both
wikipediaMM: Collection

- 151,590 images
  - wide variety
  - global scope
  - JPEG, PNG formats

- Annotations
  - user-generated
    - highly heterogeneous
    - varying length
    - noisy
  - semi-structured
  - monolingual (English)

- Used in INEX MM 2006 - 2007
wikipediaMM: Additional resources

- provided by University of Amsterdam
- 101 MediaMill concepts
- classifiers trained on TRECVID 2005 data

Features

Concepts

Natural Image statistics:
- 0.486683, 0.453243, 0.421476, 0.371388, 0.279351, 0.410819, 0.321526, 0.461151, 0.391605, 0.360453, 0.086648, 0.271732, 0.231170, 0.334546, 0.270454, 0.577713, 0.529623, 0.501511, 0.482971, 0.355653, 0.544112, 0.392626, 0.562681, 0.463030, 0.438386, 0.101318, 0.335869, 0.275694, 0.456026, 0.335002, 0.354639, 0.308430, 0.279414, 0.278727, 0.268225, 0.325531, 0.237821, 0.308108, 0.317694, 0.294898, 0.106288, 0.266600, 0.231708, 0.296142, 0.263102, 0.642913, 0.622363, 0.624570, 0.536831, 0.629639, 0.624990, 0.456589, 0.646821, ...

- Boat: 0.12
- Building: 0.84
- Bus: 0.08
- Bush: 0.01
- ....
<topic>

<number> 62 </number>

<title> cities by night </title>
<concept> building </concept>
<image> http://www.bushland.de/hksky2.jpg </image>

<narrative> I am decorating my flat and as I like photos of cities at night, I would like to find some that I could possibly print into posters. Photos of cities (or the earth) from space are not relevant. I would like to find photos of skylines or photos that contain parts of a city at night (including streets and buildings). </narrative>

</topic>

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of topics</td>
<td>75</td>
</tr>
<tr>
<td>Average number of terms in title</td>
<td>2.64</td>
</tr>
<tr>
<td>Number of topics with image(s)</td>
<td>43</td>
</tr>
<tr>
<td>Number of topics with concept(s)</td>
<td>45</td>
</tr>
<tr>
<td>Number of topics with both image and concept</td>
<td>28</td>
</tr>
<tr>
<td>Number of topics with text only</td>
<td>15</td>
</tr>
</tbody>
</table>
wikipediaMM: Participation
(topic development, submissions, assessments)

2008
- 12 groups
- 77 runs

2007 (INEX MM)
- 4 groups
- 12 runs

2006 (INEX MM)
- 4 groups
- 16 runs

CEA - CEA LIST, France
CHEMNITZ - Computer Science and Media, Chemnitz University of Technology
CURIEN - Laboratoire Hubert Curien, Universite Jean Monnet, Saint-Etienne, France
CWI - Database Architectures and Information Access, CWI, Netherlands
IMPERIAL - Multimedia and Information Systems, Imperial College, UK
IRIT - SIG-IRIT, Toulouse, France
STZAKI - Data Mining and Web Search, Hungarian Academy of Sciences
UALICANTE - NLP and Information Systems, University of Alicante, Spain
UNIGE - Computer Vision and Multimedia, Universite de Geneve, Switzerland
UPEKING - Digital Media Institute, Peking University, China
UPMC-LUP6 - UPMC/LIP6 - Computer Science Lab, Paris, France
UTOULON - LSIS, UMR CNRS & Universite Sud Toulon-Var, France

CLAC - Computational Linguistics, Concordia University, Montreal, Canada
DEU - Dept. Comp. Engineering, Dokuz Eylul University, Turkey
XRCE - Xerox Research Centre Europe
wikipediaMM: Assessment system

Query 62
- Not relevant
- Relevant

Topic: cities by night
Judged: 1751/1751

This is document 55926 from the ImageCLEF wikipediaMM Corpus. Note: It is a snapshot of a wikipedia page as we crawled the web; the current page may have changed since that time. Assessments are to be made on the crawled page as shown below.

55926: Hamburg1.jpg

Hamburg1.jpg
<table>
<thead>
<tr>
<th>Group</th>
<th># runs</th>
<th>rank</th>
<th>Modality</th>
<th>FB/QE</th>
<th>MAP</th>
<th>P@20</th>
<th>R-prec.</th>
<th>rank</th>
<th>MAP</th>
<th>P@20</th>
<th>R-prec.</th>
</tr>
</thead>
<tbody>
<tr>
<td>upeking</td>
<td>7</td>
<td>1</td>
<td>TXT</td>
<td>QE</td>
<td>0.3444</td>
<td>0.3993</td>
<td>0.3794</td>
<td>37.5</td>
<td>0.1749</td>
<td>0.2020</td>
<td>0.1922</td>
</tr>
<tr>
<td>cea</td>
<td>2</td>
<td>2</td>
<td>TXTCON</td>
<td>QE</td>
<td>0.2735</td>
<td>0.3840</td>
<td>0.3225</td>
<td>3.0</td>
<td>0.2684</td>
<td>0.3757</td>
<td>0.3153</td>
</tr>
<tr>
<td>valencia</td>
<td>24</td>
<td>3</td>
<td>TXT</td>
<td>NOFB</td>
<td>0.2700</td>
<td>0.3040</td>
<td>0.3075</td>
<td>22.5</td>
<td>0.2349</td>
<td>0.2797</td>
<td>0.2271</td>
</tr>
<tr>
<td>sztaki</td>
<td>8</td>
<td>10</td>
<td>TXT</td>
<td>NOFB</td>
<td>0.2551</td>
<td>0.2773</td>
<td>0.3020</td>
<td>15.5</td>
<td>0.2508</td>
<td>0.2777</td>
<td>0.2954</td>
</tr>
<tr>
<td>wpi</td>
<td>2</td>
<td>13</td>
<td>TXT</td>
<td>NOFB</td>
<td>0.2528</td>
<td>0.2833</td>
<td>0.3080</td>
<td>16.0</td>
<td>0.2511</td>
<td>0.2810</td>
<td>0.3023</td>
</tr>
<tr>
<td>curien</td>
<td>6</td>
<td>22</td>
<td>TXT</td>
<td>NOFB</td>
<td>0.2453</td>
<td>0.2860</td>
<td>0.2905</td>
<td>45.5</td>
<td>0.1515</td>
<td>0.1940</td>
<td>0.1822</td>
</tr>
<tr>
<td>chemnitz</td>
<td>4</td>
<td>27</td>
<td>TXTIMCON</td>
<td>QE</td>
<td>0.2195</td>
<td>0.2747</td>
<td>0.2734</td>
<td>32.5</td>
<td>0.2122</td>
<td>0.2770</td>
<td>0.2643</td>
</tr>
<tr>
<td>imperial</td>
<td>6</td>
<td>44</td>
<td>TXT</td>
<td>NOFB</td>
<td>0.1918</td>
<td>0.2647</td>
<td>0.2362</td>
<td>59.5</td>
<td>0.0978</td>
<td>0.1384</td>
<td>0.1235</td>
</tr>
<tr>
<td>irit</td>
<td>4</td>
<td>48</td>
<td>TXT</td>
<td>NOFB</td>
<td>0.1652</td>
<td>0.2353</td>
<td>0.2148</td>
<td>57.5</td>
<td>0.1198</td>
<td>0.1783</td>
<td>0.1647</td>
</tr>
<tr>
<td>geneva</td>
<td>2</td>
<td>52</td>
<td>TXT</td>
<td>NOFB</td>
<td>0.1440</td>
<td>0.1793</td>
<td>0.1806</td>
<td>58.0</td>
<td>0.1179</td>
<td>0.1660</td>
<td>0.1574</td>
</tr>
<tr>
<td>upmc-lip6</td>
<td>7</td>
<td>56</td>
<td>TXT</td>
<td>NOFB</td>
<td>0.1193</td>
<td>0.1820</td>
<td>0.1581</td>
<td>66.5</td>
<td>0.0602</td>
<td>0.0947</td>
<td>0.0817</td>
</tr>
<tr>
<td>utoulon</td>
<td>5</td>
<td>70</td>
<td>TXT</td>
<td>NOFB</td>
<td>0.0399</td>
<td>0.0673</td>
<td>0.0583</td>
<td>70.0</td>
<td>0.0399</td>
<td>0.0353</td>
<td>0.0583</td>
</tr>
</tbody>
</table>
wikipediaMM: Results (cross-media)
Findings:
- text-only approaches perform quite well
  - query expansion (knowledge bases)
  - relevance feedback
  - weighting schemes (BM25, DFR, LM)
- BUT fusion with concepts can be beneficial
- different types of topics benefit from different approaches

Open issues:
- effective combination of text and visual evidence?
- cross-media relevance feedback?

Next year??
- multilinguiality (topics? annotations?)
- which concepts? ground truth?
- what resources to provide to participants?
- user participation in topic development/relevance assessments?
2006: object annotation
2007: object retrieval
2008: visual concept detection
Visual Concept Detection Task
VCDT: Participants

- CEA-LIST, France
- MSR China, Multimedia Computing and Communications, China
- IPAL-I2R, Infocomm Research Lab, Singapore
- LSIS, Information Sciences and Systems, France
- MMIS, Open University, UK
- Makere, Faculty of Computing and Information Technology, Makere University, Uganda
- RWTH, Human Language Technology and Pattern Recognition Group, Aachen, Germany
- TIA, Group for Machine Learning for Image Processing and Information Retrieval, Mexico
- UPMC, University Pierre et Marie Curie, Paris
- XRCE, Xerox Research Centre Europe, Grenoble, France
- SZTAKI, Hungarian Academy of Science, Budapest, Hungary
<table>
<thead>
<tr>
<th></th>
<th>best run</th>
<th></th>
<th></th>
<th>AVG</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td># runs</td>
<td>rank</td>
<td>EER</td>
<td>AUC</td>
<td>rank</td>
<td>EER</td>
</tr>
<tr>
<td>XRCE</td>
<td>2</td>
<td>1</td>
<td>16.65</td>
<td>90.66</td>
<td>1.5</td>
<td>17.97</td>
</tr>
<tr>
<td>RWTH</td>
<td>1</td>
<td>3</td>
<td>20.45</td>
<td>86.19</td>
<td>3</td>
<td>20.45</td>
</tr>
<tr>
<td>UPMC</td>
<td>6</td>
<td>4</td>
<td>24.55</td>
<td>82.74</td>
<td>11</td>
<td>27.2</td>
</tr>
<tr>
<td>LSIS</td>
<td>7</td>
<td>5</td>
<td>25.88</td>
<td>80.51</td>
<td>20.29</td>
<td>32.8</td>
</tr>
<tr>
<td>MMIS</td>
<td>4</td>
<td>13</td>
<td>28.44</td>
<td>77.94</td>
<td>23.25</td>
<td>32.55</td>
</tr>
<tr>
<td>CEA_LIST</td>
<td>3</td>
<td>17</td>
<td>29.04</td>
<td>73.4</td>
<td>26.33</td>
<td>33.39</td>
</tr>
<tr>
<td>IPAL_I2R</td>
<td>8</td>
<td>19</td>
<td>29.71</td>
<td>76.44</td>
<td>32.13</td>
<td>35.96</td>
</tr>
<tr>
<td>budapest</td>
<td>13</td>
<td>20</td>
<td>31.14</td>
<td>74.9</td>
<td>31.77</td>
<td>35.17</td>
</tr>
<tr>
<td>TIA</td>
<td>7</td>
<td>24</td>
<td>32.09</td>
<td>55.64</td>
<td>39.57</td>
<td>39.87</td>
</tr>
<tr>
<td>HJ_FA</td>
<td>1</td>
<td>47</td>
<td>45.07</td>
<td>19.96</td>
<td>47</td>
<td>45.07</td>
</tr>
<tr>
<td>Makere</td>
<td>1</td>
<td>51</td>
<td>49.25</td>
<td>30.83</td>
<td>51</td>
<td>49.25</td>
</tr>
<tr>
<td>#</td>
<td>concept</td>
<td>best</td>
<td>average</td>
<td>worst</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>------------------</td>
<td>-------------------------------</td>
<td>---------</td>
<td>----------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>EER</td>
<td>AUC</td>
<td>group</td>
<td>EER</td>
<td>AUC</td>
</tr>
<tr>
<td>0</td>
<td>indoor</td>
<td>8.9</td>
<td>97.4</td>
<td>XRCE</td>
<td>28</td>
<td>67.6</td>
</tr>
<tr>
<td>1</td>
<td>outdoor</td>
<td>9.2</td>
<td>96.6</td>
<td>XRCE</td>
<td>30.6</td>
<td>70.5</td>
</tr>
<tr>
<td>2</td>
<td>person</td>
<td>17.8</td>
<td>89.7</td>
<td>XRCE</td>
<td>35.9</td>
<td>62.2</td>
</tr>
<tr>
<td>3</td>
<td>day</td>
<td>21</td>
<td>85.7</td>
<td>XRCE</td>
<td>35.4</td>
<td>64.9</td>
</tr>
<tr>
<td>4</td>
<td>night</td>
<td>8.7</td>
<td>97.4</td>
<td>XRCE/budapest</td>
<td>27.6</td>
<td>72.5</td>
</tr>
<tr>
<td>5</td>
<td>water</td>
<td>23.8</td>
<td>84.6</td>
<td>XRCE</td>
<td>38.1</td>
<td>57.8</td>
</tr>
<tr>
<td>6</td>
<td>road/pathway</td>
<td>28.8</td>
<td>80</td>
<td>XRCE</td>
<td>42.6</td>
<td>50.7</td>
</tr>
<tr>
<td>7</td>
<td>vegetation</td>
<td>17.6</td>
<td>89.9</td>
<td>XRCE</td>
<td>33.9</td>
<td>67.4</td>
</tr>
<tr>
<td>8</td>
<td>tree</td>
<td>18.9</td>
<td>88.3</td>
<td>XRCE</td>
<td>36.1</td>
<td>62.8</td>
</tr>
<tr>
<td>9</td>
<td>mountains</td>
<td>15.3</td>
<td>93.8</td>
<td>XRCE</td>
<td>33.1</td>
<td>61.2</td>
</tr>
<tr>
<td>10</td>
<td>beach</td>
<td>21.7</td>
<td>86.8</td>
<td>XRCE</td>
<td>35.8</td>
<td>57.6</td>
</tr>
<tr>
<td>11</td>
<td>buildings</td>
<td>17</td>
<td>89.7</td>
<td>XRCE</td>
<td>37.4</td>
<td>60.8</td>
</tr>
<tr>
<td>12</td>
<td>sky</td>
<td>10.4</td>
<td>95.7</td>
<td>XRCE</td>
<td>24</td>
<td>78.6</td>
</tr>
<tr>
<td>13</td>
<td>sunny</td>
<td>9.2</td>
<td>96.4</td>
<td>XRCE</td>
<td>30.3</td>
<td>66.5</td>
</tr>
<tr>
<td>14</td>
<td>partly cloudy</td>
<td>15.4</td>
<td>92.1</td>
<td>XRCE/budapest</td>
<td>37.5</td>
<td>58.9</td>
</tr>
<tr>
<td>15</td>
<td>overcast</td>
<td>14.1</td>
<td>93.7</td>
<td>XRCE</td>
<td>32.1</td>
<td>67.6</td>
</tr>
<tr>
<td>16</td>
<td>animal</td>
<td>20.7</td>
<td>85.7</td>
<td>XRCE</td>
<td>38.2</td>
<td>54.2</td>
</tr>
</tbody>
</table>
VCDT: Conclusions

- visual concept detection works quite well
- some concepts are far easier than others
  - water & road are very hard
  - sunny, night, indoor/outdoor is easy
- local features and discriminative classifiers outperform other methods

- only one group used VCDT outcome for photo retrieval
  - improvements are consistent

Open Questions:
- which concepts? which training data?
- how to combine/fuse VCDT and photo retrieval!
Medical Image Annotation Task

- Purely visual task
- Given an image, find a textual description

2005:
  - 9,000 training images/1,000 test images
  - Assign one out of 57 possible labels to each image

2006:
  - 10,000 training images/1,000 test images
  - Assign one out of 116 possible labels to each image

2007:
  - 11,000 training images/1,000 test images
  - Assign a textual label to each image (one out of 116)

2008:
  - 12,076 training images/1,000 test images
  - More classes, use of hierarchy required (~200 classes)
Example: 1121-127-720-500  
DDDD-DDDD-AAA-BBB-TTT

D Direction:  
coronal, anterior-posterior, supine

A Anatomy:  
abdomen, middle, unspec.

B Biosystem:  
uroopoetic system, unspec., unspec.

T Technique:  
radiography, plain, analog, overview

Aim: Predict complete code  
• as far as possible  
• correctly
Evaluation Criterion

- incomplete codes 11___-12_-7__-5__
- not predicting a position: better than a wrong prediction
- incorrect prediction in one position invalidates all later predictions in this axis
- axes are independent
- early errors are worse than late

Examples
(for one axis): correct 318a

<table>
<thead>
<tr>
<th></th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>318a</td>
<td>0</td>
</tr>
<tr>
<td>318*</td>
<td>0.06</td>
</tr>
<tr>
<td>3187</td>
<td>0.12</td>
</tr>
<tr>
<td>31**</td>
<td>0.14</td>
</tr>
<tr>
<td>32**</td>
<td>0.52</td>
</tr>
<tr>
<td>8988</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Example Images

12,076 train images
1,000 test images
196 unique codes
Participants

- **FEIT** - Faculty of Electrical Engineering and Information Technology, Skopje, Macedonia
- **medGIFT** - University Hospitals of Geneva, Switzerland
- **Miracle** - Miracle Lab, Daedalus University, Madrid, Spain
- **TAU** - Medical Image Processing Lab, Tel Aviv University, Israel
- **IDIAP** - IDIAP research institute, Martigny, Switzerland
- **IRMA** - Medical Informatics, RWTH Aachen University, Aachen, Germany
### Results

<table>
<thead>
<tr>
<th></th>
<th>best</th>
<th>average</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td># runs</td>
<td>rank</td>
</tr>
<tr>
<td>IDIAP</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>TAU</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>IRMA</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>MIRACLE</td>
<td>4</td>
<td>13</td>
</tr>
<tr>
<td>medGIFT</td>
<td>4</td>
<td>17</td>
</tr>
<tr>
<td>FEIT</td>
<td>2</td>
<td>22</td>
</tr>
</tbody>
</table>

### Conclusion/Findings:

- machine learning techniques from IDIAP work best
- local features outperform global ones
- use of wildcards necessary for good results
- results of last year published in PRL SI
Highlights of ImageCLEF 2008

- large number of participants in ImageCLEF
- good registration/participation ratio in photo retrieval
- new Wikipedia MM task
- Quaero sponsored pre workshop
Thursday 14:30

Visual Concept Detection Task
Gabriella Csurka, Xerox Research Center Europe, France
Image Fisher Vector based visual concepts detection and image retrieval

Photo Retrieval
Gareth Jones, Dublin City University, Ireland
DCU at ImageCLEFPhoto 2008

WikipediaMM Retrieval
Adrian Popescu, CEA-LIST, France
Conceptual image retrieval over the Wikipedia corpus

Medical Image Annotation
Tatiana Tommasi, IDIAP, Switzerland
CLEF2008 Image Annotation Task: an SVM Confidence-Based Approach

Medical Retrieval
Manuel Carlos Díaz Galiano, SINAI, U Jaen, Spain
SINAI at ImageCLEFmed
Several Ideas for next year!
What do you expect?
What are our ideas?
What data is available?

Breakout Session:
○ Friday 11:00h

Fill in the survey
○ www.imageclef.org/survey