

# Overview of the CLEF 2009 Robot Vision Track

Barbara Caputo<sup>1</sup>, Andrzej Pronobis<sup>2</sup>, Patric Jensfelt<sup>2</sup>

<sup>1</sup>Idiap Research Institute, Martigny, Switzerland

<sup>2</sup>Centre for Autonomous Systems, Royal Institute of Technology, Stockholm, Sweden

bcaputo@idiap.ch

## Abstract

The robot vision task has been proposed to the ImageCLEF participants for the first time in 2009. The task attracted a considerable attention, with 19 inscribed research groups, 7 groups eventually participating and a total of 27 submitted runs. The task addressed the problem of visual place recognition applied to robot topological localization. Specifically, participants were asked to classify rooms on the basis of image sequences, captured by a perspective camera mounted on a mobile robot. The sequences were acquired in an office environment, under varying illumination conditions and across a time span of almost two years. The training and validation set consisted of a subset of the IDOL2 database<sup>1</sup>. The test set consisted of sequences similar to those in the training and validation set, but acquired 20 months later and imaging also additional rooms. Participants were asked to build a system able to answer the question “where are you?” (I am in the kitchen, in the corridor, etc) when presented with a test sequence imaging rooms seen during training, or additional rooms that were not imaged in the training sequence. The system had to assign each test image to one of the rooms present in the training sequence, or indicate that the image came from a new room. We asked all participants to solve the problem separately for each test image (obligatory task). Additionally, results could also be reported for algorithms exploiting the temporal continuity of the image sequences (optional task).

Of the 27 runs, 21 were submitted to the obligatory task, and 6 to the optional task. The best result in the obligatory task was obtained by the Multimedia Information Retrieval Group of the University of Glasgow, UK with an approach based on local feature matching. The best result in the optional task was obtained by the Intelligent Systems and Data Mining Group (SIMD) of the University of Castilla-La Mancha, Albacete, Spain, with an approach based on local features and a particle filter.

## 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

## General Terms

Measurement, Performance, Experimentation

## Keywords

Place recognition, robot vision, robot localization

---

<sup>1</sup><http://www.cas.kth.se/IDOL/>

# 1 Introduction

ImageCLEF<sup>2</sup> [1, 2, 5] has started in 2003 as part of the Cross Language Evaluation Forum (CLEF<sup>3</sup>, [6]). Its main goal has been to promote research on multi-modal data annotation and information retrieval, in various application fields. As such it has always contained visual, textual and other modalities, mixed tasks and several sub tracks.

This year, for the first time, ImageCLEF has hosted a Robot Vision task. This paper reports on it, while other papers describe the other five tasks of ImageCLEF 2009. More information on the tasks and on how to participate to CLEF can also be found on the ImageCLEF web pages.

## 2 Participation

In 2009, a new record of 85 research groups registered for the seven sub tasks of ImageCLEF. Of these 85, 19 registered to the Robot Vision task. 7 of the registered groups submitted at least one run:

- Multimedia Information Retrieval Group, University of Glasgow, United Kingdom
- Idiap Research Institute, Martigny, Switzerland
- Faculty of Computer Science, The Alexandru Ioan Cuza University (UAIC), Iași, 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

A total of 27 runs were submitted, with 21 runs submitted to the obligatory task and 6 runs submitted to the optional task. In order to encourage participation, there was no limit to the number of runs that each group could submit.

## 3 Data Sets, Tasks, Ground Truthing

This section describes the details concerning the setup of the robot vision task. Section 3.1 describes the dataset used. Section 3.2 gives details on the tasks proposed to the participants. Finally, section 3.3 describes briefly the algorithm used for obtaining a ground truth and the obtained results.

### 3.1 Dataset

Training and validation set consisted of a subset of the publicly available IDOL2 database [3, 4]. An additional, previously unreleased image sequence was used for testing. The part of the IDOL2 database used for training and validation comprises 12 image sequences acquired using a MobileRobots PowerBot robot platform. The image sequences are accompanied by laser range data and odometry data; however use of that data was not permitted in the competition.

The image sequences in the IDOL2 database were captured with a Canon VC-C4 perspective camera using the resolution of 320x240 pixels. The acquisition was performed in a five room

---

<sup>2</sup><http://www.imageclef.org/>

<sup>3</sup><http://www.clef-campaign.org/>

subsection of a larger office environment, selected in such way that each of the five rooms represented a different functional area: a one-person office, a two-persons office, a kitchen, a corridor, and a printer area. The appearance of the rooms was captured under three different illumination conditions: in cloudy weather, in sunny weather, and at night. The robots were manually driven through each of the five rooms while continuously acquiring images and laser range scans at a rate of 5fps. Each data sample was then labelled as belonging to one of the rooms according to the position of the robot during acquisition (rather than contents of the images). Examples of images showing the interiors of the rooms, variations observed over time and caused by activity in the environment as well as introduced by changing illumination are presented in Figure 1.

The IDOL2 database was designed to test the robustness of place recognition algorithms to variations that occur over a long period of time. Therefore, the acquisition process was conducted in two phases. Two sequences were acquired for each type of illumination conditions over the time span of more than two weeks, and another two sequences for each setting were recorded 6 months later (12 sequences in total). Thus, the sequences captured variability introduced not only by illumination but also natural activities in the environment (presence/absence of people, furniture/objects relocated etc.).

The test sequences were acquired in the same environment, using the same camera setup. The acquisition was performed 20 months after the acquisition of the IDOL2 database. The sequences contain additional rooms that were not imaged in the IDOL2 database.

## 3.2 The Task

The robot vision task addressed the problem of visual place recognition applied to topological localization of a mobile robot. Specifically, participants were asked to determine the topological location of a robot based on images acquired with a perspective camera mounted on a robot platform.

Participants were given training data consisting of an image sequence. The training sequence was recorded using a mobile robot that was manually driven through several rooms of a typical indoor office environment. The acquisition was performed under fixed illumination conditions and at a given time. Each image in the training sequence was labeled and assigned to the room in which it was acquired.

The challenge was to build a system able to answer the question 'where are you?' (I'm in the kitchen, in the corridor, etc.) when presented with a test sequence containing images acquired in the previously observed part of the environment or in additional rooms that were not imaged in the training sequence. The test images were acquired 6-20 months later after the training sequence, possibly under different illumination settings. The system had to assign each test image to one of the rooms that were present in the training sequence or indicate that the image came from a room that was not included during training. Moreover, the system could refrain from making a decision (e.g. in the case of lack of confidence).

The algorithm had to be able to provide information about the location of the robot separately for each test image (e.g. when only some of the images from the test sequences are available or the sequences are scrambled). This corresponds to the problem of global topological localization. We called this the obligatory task. However, results can also be reported for the case when the algorithm is allowed to exploit continuity of the sequences and rely on the test images acquired before the classified image. We called this the optional task.

## 3.3 Ground Truth

The image sequences used in the competition were annotated with ground truth. The annotations of the training and validation sequences were available to the participants, while the ground truth for the test sequence was released after the results were announced. Each image in the sequences was labelled according to the position of the robot during acquisition as belonging to one of the rooms used for training or as an unknown room. The ground truth was then used to calculate a



(a) Variations introduced by illumination



(b) Variations observed over time



(c) Remaining rooms (at night)

Figure 1: Examples of pictures taken from the IDOL2 database showing the interiors of the rooms, variations observed over time and caused by activity in the environment as well as introduced by changing illumination.

(a) Obligatory task.			(b) Optional task.		
#	Group	Score	#	Group	Score
1	Glasgow	890.5	1	SIMD	916.5
2	Idiap	793.0	2	CVIU	884.5
3	UAIC	787.0	3	Idiap	853.0
4	UAIC	787.0	4	SIMD	711.0
5	CVIU	784.0	5	SIMD	711.0
6	Glasgow	650.5	6	SIMD	609.0
7	UAIC	599.5			
8	UAIC	599.5			
9	LSIS	544.0			
10	SIMD	511.0			
11	LSIS	509.5			
12	MRIM	456.5			
13	MRIM	415.0			
14	MRIM	328.0			
15	UAIC	296.5			
16	MRIM	25.0			
17	LSIS	-32.0			
18	LSIS	-32.0			
19	LSIS	-32.0			
20	LSIS	-32.0			
21	Glasgow	-188.0			

Table 1: Results for each run submitted to the obligatory (a) and optional (b) tasks.

score indicating the performance of an algorithm on the test sequence. The following rules were used when calculating the overall score for the whole test sequence:

- 1 point was given for each correctly classified image.
- Correct detection of an unknown room was regarded as correct classification.
- 0.5 points was subtracted for each misclassified image.
- No points were given or subtracted if an image was not classified (the algorithm refrained from the decision).

A script was available to the participants that automatically calculated the score for a specified test sequence given the classification results produced by an algorithm.

## 4 Results

This section describes the results of the robot vision task at ImageCLEF 2009. Table 1(a) shows the results for the obligatory task, while Table 1(b) shows the result for the optional task.

We see that the majority of runs were submitted to the obligatory task: of the 27 total submissions, 21 were submitted to the obligatory run and only 6 to the optional task. A possible explanation is that the optional task requires a higher expertise on robotics than the obligatory task, which therefore represents a very good entry point.

The submissions used a wide range of techniques, spanning from local descriptors combined with statistical methods to approaches transplanted from the language modeling community. It is interesting to note though that the two groups that ranked first in the two sub tasks both used a local features based approach. This confirms a consolidated trend in the robot vision community that treats local descriptors as the off the shelf feature of choice for visual recognition.

## 5 Conclusions

The first robot vision task at ImageCLEF 2009 attracted a considerable attention and proved an interesting complement to the existing tasks. The approach presented by the participating groups were diverse and original, offering a fresh take on the topological localization problem. We plan to continue the task in the next years, adding laser information and odometry to the visual information, and proposing new challenges to the perspective participants.

## 6 Acknowledgements

We would like to thank the CLEF campaign for supporting the ImageCLEF initiative. B. Caputo was supported by the EMMA project, funded by the Hasler foundation. A. Pronobis and P. Jensfelt were supported by the EU FP7 project CogX ICT-215181. The support is gratefully acknowledged.

## References

- [1] Paul Clough, Henning Müller, Thomas Deselaers, Michael Grubinger, Thomas M. Lehmann, Jeffery Jensen, and William Hersh. The CLEF 2005 cross-language image retrieval track. In *Cross Language Evaluation Forum (CLEF 2005)*, Springer Lecture Notes in Computer Science, pages 535–557, September 2006.
- [2] Paul Clough, Henning Müller, and Mark Sanderson. The CLEF cross-language image retrieval track (ImageCLEF) 2004. In Carol Peters, Paul Clough, Julio Gonzalo, Gareth J. F. Jones, Michael Kluck, and Bernardo Magnini, editors, *Multilingual Information Access for Text, Speech and Images: Result of the fifth CLEF evaluation campaign*, volume 3491 of *Lecture Notes in Computer Science (LNCS)*, pages 597–613, Bath, UK, 2005. Springer.
- [3] J. Luo, A. Pronobis, B. Caputo, and P. Jensfelt. The KTH-IDOL2 database. Technical Report CVAP304, Kungliga Tekniska Högskolan, CVAP/CAS, October 2006. Available at: <http://www.cas.kth.se/IDOL/>.
- [4] J. Luo, A. Pronobis, B. Caputo, and P. Jensfelt. Incremental learning for place recognition in dynamic environments. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS07)*, San Diego, CA, USA, October 2007.
- [5] Henning Müller, Thomas Deselaers, Eugene Kim, Jayashree Kalpathy-Cramer, Thomas M. Deserno, Paul Clough, and William Hersh. Overview of the ImageCLEFmed 2007 medical retrieval and annotation tasks. In *CLEF 2007 Proceedings*, volume 5152 of *Lecture Notes in Computer Science (LNCS)*, pages 473–491, Budapest, Hungary, 2008. Springer.
- [6] Jacques Savoy. Report on CLEF–2001 experiments. In *Report on the CLEF Conference 2001 (Cross Language Evaluation Forum)*, pages 27–43, Darmstadt, Germany, 2002. Springer LNCS 2406.