Overview of the CLEF 2009 Robot Vision Track*

Andrzej Pronobis[1], Li Xing[1], and Barbara Caputo[2]

[1] Centre for Autonomous Systems, The Royal Institute of Technology, Stockholm, Sweden
{pronobis,lixing}@kth.se
[2] Idiap Research Institute, Martigny, Switzerland
bcaputo@idiap.ch

Abstract. The robot vision track has been proposed to the ImageCLEF participants for the first time in 2009 and attracted considerable attention. The track addressed the problem of visual place recognition applied to robot topological localization. Participants were asked to classify rooms of an office environment on the basis of image sequences captured by a perspective camera mounted on a mobile robot. The algorithms proposed by the participants had 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 participants were asked 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). Robustness of the algorithms was evaluated in presence of variations introduced by changing illumination conditions and dynamic variations observed across a time span of almost two years. The participants submitted 18 runs to the obligatory task, and 9 to the optional task. The best results were obtained by the Idiap Research Institute, Martigny, Switzerland for the obligatory task and the University of Castilla-La Mancha, Albacete, Spain for the optional task.

1 Introduction

ImageCLEF[1] [1, 2, 3] started in 2003 as part of the Cross Language Evaluation Forum (CLEF[2], [4]). 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.

* 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 was supported by the EU FP7 project ICT-215181-CogX. The support is gratefully acknowledged.

[1] http://www.imageclef.org/

[2] http://www.clef-campaign.org/

C. Peters et al. (Eds.): CLEF 2009 Workshop, Part II, LNCS 6242, pp. 110–119, 2010.

The robot vision track has been proposed to the ImageCLEF participants for the first time in 2009. The track attracted a considerable attention, with 19 inscribed research groups, 7 groups eventually participating and a total of 27 submitted runs. The track 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[3]. 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, 18 were submitted to the obligatory task, and 9 to the optional task. The best result in the obligatory task was obtained by the Idiap Research Institute, Martigny, Switzerland with an approach based on integration of global and local visual features and a Support Vector Machine. 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.

This paper provides an overview of the Robot Vision track and reports on the runs submitted by the participants. First, details concerning the setup of the robot vision track are given in Section 2. Then, Section 3 presents the participants and Section 4 provides the ranking of the obtained results. Finally, an overview of the approaches used by the participants is given in Section 5. Conclusions are drawn in Section 6. Additional information about the task and on how to participate in the future robot vision challenges can be found on the ImageCLEF web pages.

2 The RobotVision Track

This section describes the details concerning the setup of the robot vision track. Section 2.1 describes the dataset used. Section 2.2 gives details on the tasks proposed to the participants. Finally, section 2.3 describes briefly the algorithm used for obtaining a ground truth and the evaluation procedure.

[3] `http://www.cas.kth.se/IDOL/`

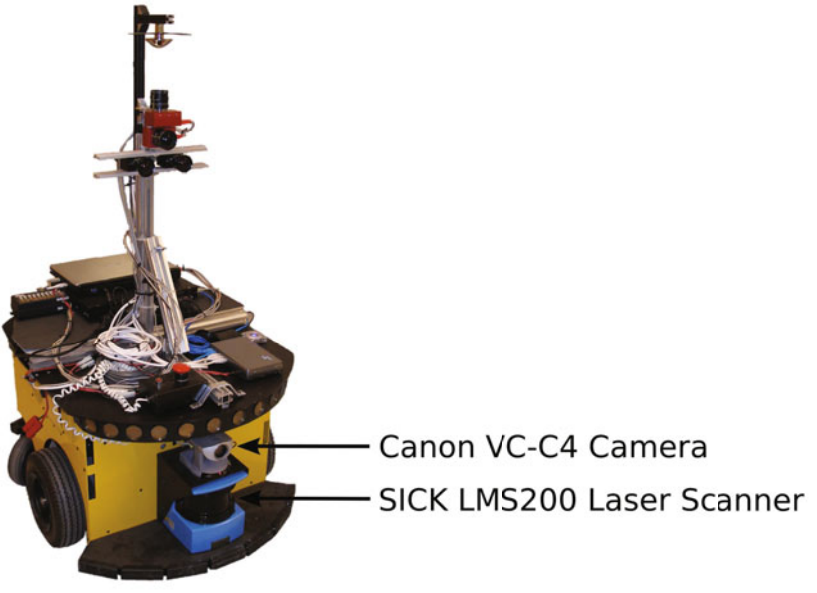


Fig. 1. The MobileRobots PowerBot mobile robot platform used for data acquisition

2.1 Dataset

Three datasets were made available to the participants. Annotated training and validation data were released when the competition started. Unlabeled testing set was released two weeks before the results submission deadline. The training and validation sets consisted of a subset of the publicly available IDOL2 database [5,6]. 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 presented in Figure 1. The image sequences in the database 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 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 map of the environment is presented in Figure 2. 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. Approximate path followed by the robot during acquisition of the data sequences is plotted with a solid line in Figure 2. Each data sample was then labelled as belonging to one of the rooms according to the

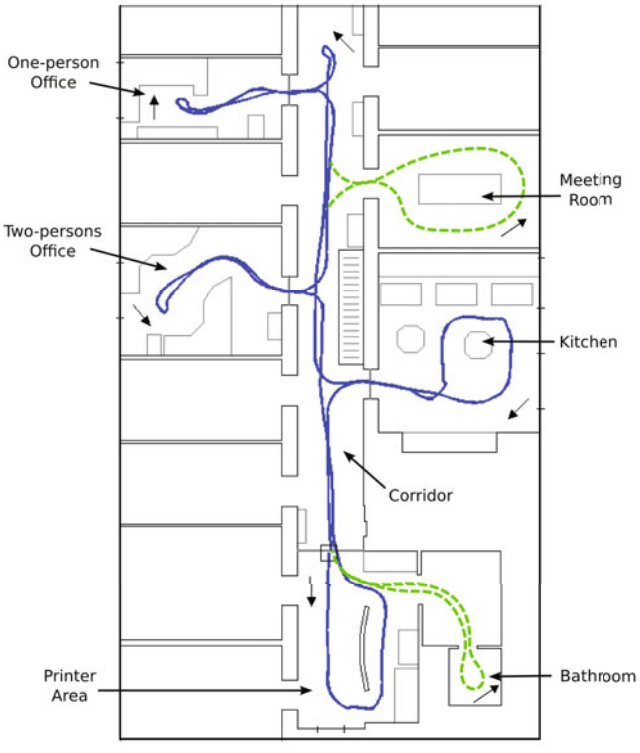
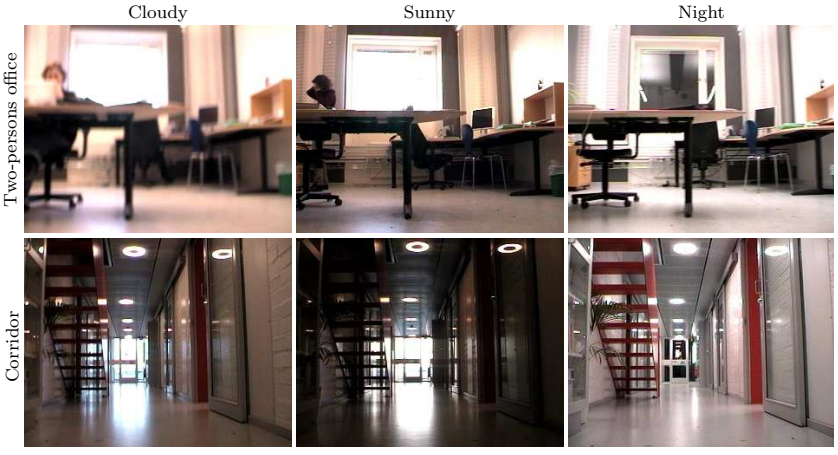


Fig. 2. Map of the environment with approximate path followed by the robot during acquisition of the training, validation and testing data. The dashed segments of the path correspond to the rooms available only in the test set.

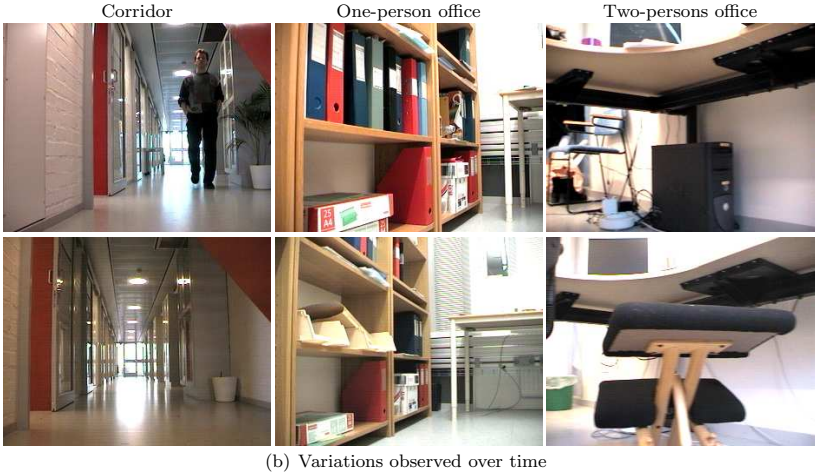
position of the robot during acquisition (rather than contents of the images). Examples of images showing the interiors of the rooms, variations observed over time and caused by activity in the environment as well as introduced by changing illumination are presented in Figure 3.

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 presented in Figure 1. The acquisition was performed 20 months after the acquisition of the IDOL2 database. The robot followed a very similar path to the one used for acquisition of the IDOL2 database. However, this time the path was extended with two additional rooms: a meeting room and a



(a) Variations introduced by illumination



(b) Variations observed over time

(c) Remaining rooms (at night)

Fig. 3. 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

bathroom. The dashed segments of the path shown in Figure 2 correspond to those rooms available only in the test set.

2.2 The Task

The Robot Vision track 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 mobile 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 were available or the sequences were scrambled). This corresponds to the problem of global topological localization. We called this the obligatory task. However, results could also be reported for the case when the algorithm was allowed to exploit continuity of the sequences and relied on the test images acquired before the classified image. We called this the optional task.

2.3 Ground Truth and Evaluation

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 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 granted 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 granted 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.

3 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:

- Faculty of Computer Science, The Alexandru Ioan Cuza University (UAIC), Iaşi, Romania
- Idiap Research Institute, Martigny, Switzerland
- Computer Vision & Image Understanding Department (CVIU), Institute for Infocomm Research, Singapore
- Laboratoire des Sciences de l'Information et des Systèmes (LSIS), La Garde, 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
- Multimedia Information Retrieval Group (MIRG), University of Glasgow, United Kingdom

A total of 27 runs were submitted, with 18 runs submitted to the obligatory task and 9 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.

4 Results

This section presents the results of the robot vision track of ImageCLEF 2009. Table 1(a) shows the results for the obligatory task, while Table 1(b) shows the result for the optional task. Scores are presented for each of the submitted runs that complied with the rules of the contest.

We see that the majority of runs were submitted to the obligatory task. A possible explanation is that the optional task requires a higher expertise in robotics that the obligatory task, which therefore represents a very good entry point. Additional three runs were submitted to the optional track by MRIM. However, since no runs were submitted to the obligatory track, the results could not be accepted in the official ranking. The next section provides an overview of the approaches used by the participants.

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

(a) Obligatory task.

#	Group	Score
1	Idiap	793.0
2	UAIC	787.0
3	UAIC	787.0
4	CVIU	784.0
5	UAIC	599.5
6	UAIC	599.5
7	LSIS	544.0
8	SIMD	511.0
9	LSIS	509.5
10	MRIM	456.5
11	MRIM	415.0
12	MRIM	328.0
13	UAIC	296.5
14	MRIM	25.0
15	LSIS	-32.0
16	LSIS	-32.0
17	LSIS	-32.0
18	LSIS	-32.0

(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

5 Approaches

The submissions used a wide range of techniques for representing visual information, building models of the appearance of the environment and spatio-temporal integration. It is interesting to note though that most of the groups, including the two groups that ranked first in the two tasks, employed approaches based on local features, either used as the only image representation or in combination with other visual cues. 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. At the same time, the algorithms used for place recognition spanned from statistical methods to approaches transplanted from the language modeling community.

The Scale Invariant Feature Transform (SIFT) [7] was employed most frequently as a local descriptor and the groups winning in both tasks used SIFT in order to represent visual information. The approach used by Idiap [8] which ranked first in the obligatory task, used SIFT combined with several other descriptors including two global image representations: Composed Receptive Field Histograms (CRFH) and PCA Census Transform Histograms (PACT). The algorithm employed by SIMD [9] relied mainly on the SIFT descriptor complemented with lines and squares detected using the Hough transform. Other participants also used SIFT (UAIC [10]); color SIFT (SIFT features extracted from the red, green and blue channels) combined with HSV color histograms and

multi-scale canny edge histograms (MRIM [11]); local features extracted from patches formed around interest points found using the Harris corner detector in images pre-processed using an illumination filter based on the Retinex algorithm (MIRG [12]); or Profile Entropy Features (PEF) encoding RGB color and texture information (LSIS [13]). Techniques using color descriptors ranked lower in general in the obligatory task, which might suggest that color information was not sufficiently robust to the large variations in illumination captured in the dataset.

The participants applied a wide range of techniques to the place recognition problem in the obligatory task. Several variations of a simple image matching strategy were used by SIMD [9], UAIC [10] and MIRG [12]. Idiap [8] built models of places using Support Vector Machines (SVM), separately for several visual cues, and combined the outputs using a Discriminative Accumulation Scheme (DAS). The group of CVIU, also used Support Vector Machines, while LSIS [13] used Least Squares Support Vector Machines (LS-SVM). Finally, MRIM [11] applied a framework based on visual vocabulary and a language model (Conceptual Unigram Model).

Four groups submitted runs to the optional task. The approach used by SIMD [9], which ranked first in this track, employed a particle filter to perform Monte Carlo localization. MIRG [12] used decision rules to process the results obtained for separate frames. CVIU and Idiap [8] applied simple temporal smoothing techniques which obtained lower scores than the other approaches.

6 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 cues provided by stereo vision and proposing new challenges to the participants. In particular, we plan to focus on the problem of place categorization and use objects as an important source of information about the environment.

References

1. Clough, P., Müller, H., Deselaers, T., Grubinger, M., Lehmann, T.M., Jensen, J., Hersh, W.: The CLEF 2005 cross–language image retrieval track. In: Peters, C., Gey, F.C., Gonzalo, J., Müller, H., Jones, G.J.F., Kluck, M., Magnini, B., de Rijke, M., Giampiccolo, D. (eds.) CLEF 2005. LNCS, vol. 4022, pp. 535–557. Springer, Heidelberg (2006)
2. Clough, P., Müller, H., Sanderson, M.: The CLEF cross–language image retrieval track (ImageCLEF) 2004. In: Peters, C., Clough, P., Gonzalo, J., Jones, G.J.F., Kluck, M., Magnini, B. (eds.) CLEF 2004. LNCS, vol. 3491, pp. 597–613. Springer, Heidelberg (2005)

3. Müller, H., Deselaers, T., Kim, E., Kalpathy-Cramer, J., Deserno, T.M., Clough, P., Hersh, W.: Overview of the ImageCLEFmed 2007 medical retrieval and annotation tasks. In: Peters, C., Jijkoun, V., Mandl, T., Müller, H., Oard, D.W., Peñas, A., Petras, V., Santos, D. (eds.) CLEF 2007. LNCS, vol. 5152, pp. 472–491. Springer, Heidelberg (2008)
4. Savoy, J.: Report on CLEF–2001 experiments. In: Peters, C., Braschler, M., Gonzalo, J., Kluck, M. (eds.) CLEF 2001. LNCS, vol. 2406, pp. 27–43. Springer, Heidelberg (2002)
5. Luo, J., Pronobis, A., Caputo, B., Jensfelt, P.: The KTH-IDOL2 database. Technical Report CVAP304, Kungliga Tekniska Hoegskolan, CVAP/CAS (October 2006), http://www.cas.kth.se/IDOL/
6. Luo, J., Pronobis, A., Caputo, B., Jensfelt, P.: Incremental learning for place recognition in dynamic environments. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2007), San Diego, CA, USA (October 2007)
7. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2) (2004)
8. Xing, L., Pronobis, A.: Multi-cue discriminative place recognition. In: Peters, C., Tsikrika, T., Müller, H., Kalpathy-Cramer, J., Jones, G.J.F., Gonzalo, J., Caputo, B. (eds.) CLEF 2009 Workshop, Part II. LNCS, vol. 6242, Springer, Heidelberg (2010)
9. Martínez-Gómez, J., Jiménez-Picazo, A., García-Varea, I.: A particle- Iter-based self-localization method using invariant features as visual information. In: Working Notes for the CLEF 2009 Workshop, Corfu, Greece (2009)
10. Boroş, E., Roşca, G., Iftene, A.: Uaic: Participation in imageclef 2009 robot vision task. In: Working Notes for the CLEF 2009 Workshop, Corfu, Greece (2009)
11. Pham, T.T., Maisonnasse, L., Mulhem, P.: Visual language modeling for mobile localization. In: Working Notes for the CLEF 2009 Workshop, Corfu, Greece (2009)
12. Feng, Y., Halvey, M., Jose, J.M.: University of glasgow at imageclef 2009 robot vision task. In: Working Notes for the CLEF 2009 Workshop, Corfu, Greece (2009)
13. Glotin, H., Zhao, Z.Q., Dumont, E.: Fast lsis prole entropy features for robot visual self-localization. In: Working Notes for the CLEF 2009 Workshop, Corfu, Greece (2009)