Overview of the ImageCLEF@ICPR 2010 Robot Vision Track*

Andrzej Pronobis[1], Henrik I. Christensen[2], and Barbara Caputo[3]

[1] Centre for Autonomous Systems, The Royal Institute of Technology, Stockholm, Sweden
pronobis@kth.se
[2] Georgia Institute of Technology, Atlanta, GA, USA
hic@cc.gatech.edu
[3] Idiap Research Institute, Martigny, Switzerland
bcaputo@idiap.ch
http://www.imageclef.org/2010/ICPR/RobotVision

Abstract. This paper describes the robot vision track that has been proposed to the ImageCLEF@ICPR2010 participants. The track addressed the problem of visual place classification. Participants were asked to classify rooms and areas of an office environment on the basis of image sequences captured by a stereo camera mounted on a mobile robot, under varying illumination conditions. 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 (from different viewpoints and under different conditions), 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). A total of eight groups participated to the challenge, with 25 runs submitted to the obligatory task, and 5 submitted to the optional task. The best result in the obligatory task was obtained by the Computer Vision and Geometry Laboratory, ETHZ, Switzerland, with an overall score of 3824.0. The best result in the optional task was obtained by the Intelligent Systems and Data Mining Group, University of Castilla-La Mancha, Albacete, Spain, with an overall score of 3881.0.

Keywords: Place recognition, robot vision, robot localization.

* 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.

D. Ünay, Z. Çataltepe, and S. Aksoy (Eds.): ICPR 2010, LNCS 6388, pp. 171–179, 2010.

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.

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. Encouraged by this first positive response, the track has been proposed for the second time in 2010, within the context of the ImageCLEF@ICPR2010 initiative. In this second edition of the track, participants were asked to classify rooms and areas on the basis of image sequences, captured by a stereo camera mounted on a mobile robot within an office environment, under varying illumination conditions. The system built by the participants had to be able to answer the question "where are you?" when presented with a test sequence imaging rooms seen during training (from different viewpoints and under different conditions) or additional rooms, not imaged in the training sequence.

The image sequences used for the contest were taken from the previously unreleased COLD-Stockholm database. The acquisition was performed in a subsection of a larger office environment, consisting of 13 areas (usually corresponding to separate rooms) representing several different types of functionality. The appearance of the areas was captured under two different illumination conditions: in cloudy weather and at night. Each data sample was then labeled as belonging to one of the areas according to the position of the robot during acquisition (rather than contents of the images).

The challenge was to build a system able to localize semantically (I'm in the kitchen, in the corridor, etc.) when presented with test sequences containing images acquired in the previously observed part of the environment, or in additional rooms that were not imaged in the training sequences. The test images were acquired under different illumination settings than the training data. The system had to assign each test image to one of the rooms that were present in the training sequences, or to indicate that the image comes 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).

We received a total of 30 submission, of which 25 were submitted to the obligatory task and 5 to the optional task. The best result in the obligatory task was obtained by the Computer Vision and Geometry Laboratory, ETHZ, Switzerland. The best result in the optional task was obtained by the Intelligent

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

Systems and Data Mining Group (SIMD) of the University of Castilla-La Mancha, Albacete, Spain.

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. Conclusions are drawn in Section 5. 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.

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, validation and test sets consisted of a subset of the previously unreleased COLD-Stockholm database. The sequences were acquired using the MobileRobots PowerBot robot platform equipped with a stereo camera system consisting of two Prosilica GC1380C cameras (Figure 1). In order to facilitate the participation of those groups not familiar with stereo images, we allowed the participants to sue monocular as well as stereo image data. The acquisition was performed in a subsection of a larger office environment, consisting of 13 areas (usually corresponding to separate rooms) representing several different types of functionality (Figure 2).

The appearance of the areas was captured under two different illumination conditions: in cloudy weather and at night. The robot was manually driven through the environment while continuously acquiring images at a rate of 5fps. Each data sample was then labeled as belonging to one of the areas according to the position of the robot during acquisition, rather than according to the content of the images.

Four sequences were selected for the contest: two training sequences having different properties, one sequence to be used for validation and one sequence for testing. Each of these four sequences had the following properties:

- *training-easy.* This sequence was acquired in 9 areas, during the day, under cloudy weather. The robot was driven through the environment following a similar path as for the test and validation sequences and the environment was observed from many different viewpoints (the robot was positioned at multiple points and performed 360 degree turns).

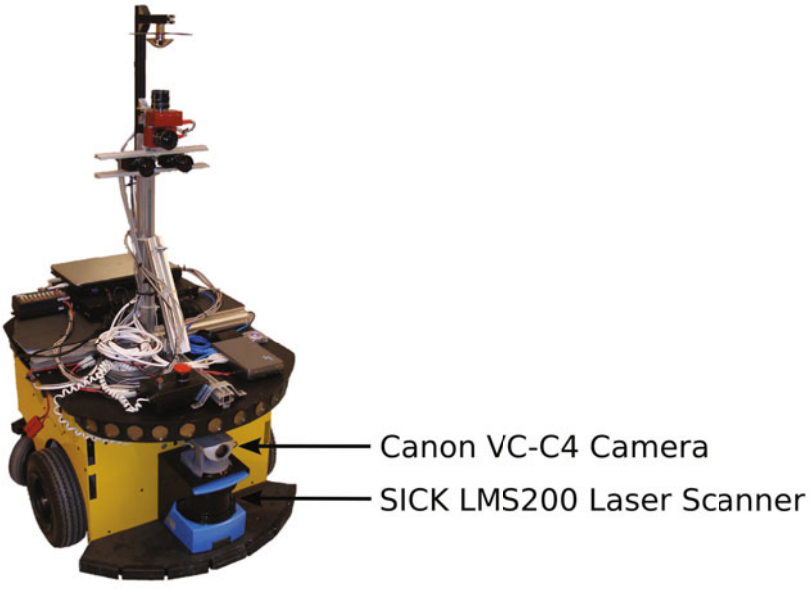

Fig. 1. The MobileRobots PowerBot mobile robot platform used for data acquisition

- *training-hard.* This sequence was acquired in 9 areas, during the day, under cloudy weather. The robot was driven through the environment in a direction opposite to the one used for the training-easy sequence, without making additional turns.
- *validation.* This sequence was acquired in 9 areas, at night. A similar path was followed as for the training-easy sequence, without making additional turns.
- *testing.* This sequence was acquired in similar conditions and following similar path as in case of the validation sequence. It contains four additional areas,for a total of 13, that were not imaged in the training or validation sequences: elevator, workshop, living room and laboratory. Exemplar images for these rooms are shown in Figure 3.

As an additional resource, we made available to participants the camera calibration data for the stereo image sequences.

2.2 The Task

The overall goal of the robot vision track is to stimulate research on semantic place recognition for mobile robot localization. The problem can be mapped to an image annotation task, where participants have to recognize the room type (kitchen, a corridor') on the basis of images acquired with a stereo camera, mounted on a mobile robot platform.

Fig. 2. Example pictures of nine rooms used for the robot vision task at ICPR 2010. From left to right, top: corridor, kitchen, large office 1. From left to right, middle: large office 2, student office, printer area. From left to right, bottom: elevator 1, small office 2, large office 2.

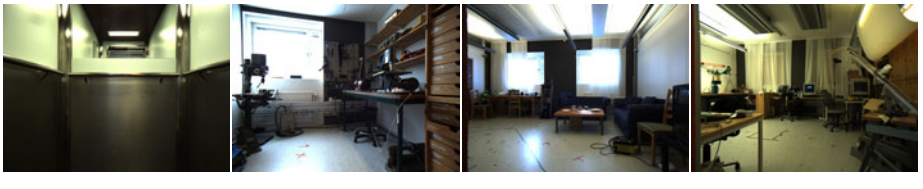

Fig. 3. Example pictures of the four additional rooms in the test sequence, used for the robot vision task at ICPR 2010. From left to right: elevator2, workshop, living room and laboratory.

All data, consisting of training, validation and test sequences, were recorded using a mobile robot, manually driven through several rooms, under fixed illumination conditions. The environment of choice was a standard academic office environment. Images in the sequences were annotated according to the position of the robot, as opposed to their informative content. For instance, an image acquired in the room 'corridor', taken when the robot was facing the entrance of the room 'kitchen', is labeled as 'corridor' even if it shows mostly part of the 'kitchen'.

The test sequences were acquired under different illumination conditions. They imaged the same rooms contained into the training and validation sequences, plus some additional rooms not seen before. Therefore participants had to address at the same time two challenges: (a) recognizing correctly rooms seen before, and (b) recognizing as 'unknown' the new rooms in the test sequence.

We consider two separate tasks, *task 1* (obligatory) and *task 2* (optional). In task 1, the algorithm had to be able to provide information about the location of the robot separately for each test image, without relying on information contained in any other 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. In task 2, the algorithm was allowed to exploit continuity of the sequences and rely on the test images acquired before the classified image, with the constraint that images acquired after the classified image could be used. The same training, validation and testing sequences were used for both tasks. The reported results were compared separately.

The tasks employed two sets of training, validation and testing sequences. The first, easier set contained sequences with constrained viewpoint variability. In this set, training, validation and testing sequences were acquired following similar path through the environment. The second, more challenging set contained sequences acquired following different paths (e.g. the robot was driven in the opposite direction). The final score for each task was calculated based on the results obtained for both sets.

The competition started with the release of annotated training and validation data. Moreover, the participants were given a tool for evaluating performance of their algorithms. The test image sequences were released later and were acquired in the same environment, under different conditions. They also contained additional rooms that were not imaged previously.

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. Each of the two test sequences consisted of a total of 2551 features. Therefore, according to the rules listed above, the maximum possible score is of 2551, both for the easy and the hard test sequences, with a maximum overall score of 5102.

3 Participation

In 2010, 28 groups registered to the Robot Vision task. 8 of them submitted at least one run, namely:

- CVG: Computer Vision and Geometry laboratory, ETH Zurich, Switzerland;
- TRS2008: Beijing Information Science and Technology University, Bejing, China;
- SIMD: Intelligent Systems and Data Mining Group, University of Castilla-La Mancha, Albacete, Spain;
- CAS IDIAP: Center for Autonomous Systems, KTH, Stockholm, Sweden and Idiap Research Institute, Martigny, Switzerland;
- PicSOM TKK: Helsinki University of Technology, TKK Department of Information and Computer Science, Helsinki, Finland;
- Magrit: INRIa Nancy, France;
- RIM at GT: Georgia Institute of Technology, Atlanta, Georgia, USA;

A total of 30 runs were submitted, with 25 runs submitted to the obligatory task and 5 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@ ICPR2010. Table 1 shows the results for the obligatory task, while Table 2 shows the result for the optional task. Scores are presented for each of the submitted runs that complied with the rules of the contest.

Table 1. Results obtained by each group in the obligatory task. The maximum overall score is of 5102, with a maximum score of 2551 for both the easy and the had sequence.

#	Group	Overall Score	Score Easy	Score Hard
1	CVG	3824.0	2047.0	1777.0
2	TRS2008	3674.0	2102.5	1571.5
3	SIMD	3372.5	2000.0	1372.5
4	CAS IDIAP	3344.0	1757.5	1372.5
5	PicSOM TKK	3293.0	2176.0	1117.0
6	Magrit	3272.0	2026.0	1246.0
7	RIM at GT	2922.5	1726.0	1196.5
8	UAIC	2283.5	1609.0	674.5

Table 2. Results obtained by each group in the optional task. The maximum overall score is of 5102, with a maximum score of 2551 for both the easy and the had sequence.

#	Group	Overall Score	Score Easy	Score Hard
1	SIMD	3881.0	2230.5	1650.5
2	TRS2008	3783.5	2135.5	1648.0
3	CAS IDIAP	3453.5	1768.0	1685.5
4	RIM at GT	2822.0	1589.5	1232.5

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. The same behavior was noted at the first edition of the robot vision task in 2009.

Concerning the obtained results, we notice that all groups perform considerably better on the easy sequence, compared to the hard sequence. This is true for both the obligatory and the optional task. For the obligatory task, the best performance on the easy sequence was obtained by the PicSOM TKK group with a score of 2176.0. This is only 375 points lower than the maximum possible score of 2551, also considering that the images in the sequence are annotated according to the robots position, but they are classified according to their informative content. As opposed to this, we see that the best performance on the easy sequence was obtained by the CVG group, with a score of 1777.0. This is 774 points lower than the maximum possible score of 2551, more than twice the difference in score between the nest performance and the maximum one for the easy sequence.

This pattern is replicated in the optional task, indicating that the temporal continuity between image frames does not seem to alleviate the problem. We see that the best performance for the easy sequence is obtained by the SIMD group, with a score of 2230.5. For the hard sequence, the best performance (obtained by the CAS IDIAP group) drops to 1685.5.

These results indicate quite clearly that the capability to recognize visually a place under different viewpoints is still an open challenge for mobile robots. This is a strong motivations towards proposing similar tasks to the community in the future editions of the robot vision task.

5 Conclusions

The robot vision task at ImageCLEF@ICPR2010 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, 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., 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. 597–613. Springer, Heidelberg (2006)
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 medical 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. 473–491. 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)