The Robot Vision Track at ImageCLEF 2010*

Andrzej Pronobis[1], Marco Fornoni[3], 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
{mfornoni,bcaputo}@idiap.ch
http://www.imageclef.org/2010/robot

Abstract. This paper describes the robot vision track that has been proposed to the ImageCLEF 2010 participants. The track addressed the problem of visual place classification, with a special focus on generalization. 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, acquired within the same building but at a different floor than the training sequence. The test data contained images of 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). A total of seven groups participated to the challenge, with 42 runs submitted to the obligatory task, and 13 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 677. The best result in the optional task was obtained by the Idiap Research Institute, Martigny, Switzerland, with an overall score of 2052.

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. M. Fornoni was supported by the MULTI project, funded by the Swiss National Science Foundation. A. Pronobis was supported by the EU FP7 project ICT-215181-CogX. The support is gratefully acknowledged.

1 Introduction

ImageCLEF[4] [1–3] started in 2003 as part of the Cross Language Evaluation Forum (CLEF[5], [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. The second edition of the challenge was held in conjunction with ICPR 2010 and saw an increase in participation, with 9 participating groups and 34 submitted runs. As for previous events, in 2010 the challenge addressed the problem of visual place classification, this time with a special focus on generalization.

Participants were asked to classify rooms and functional areas on the basis of image sequences, captured by a stereo camera mounted on a mobile robot within an office environment. The test sequence was acquired within the same building but at a different floor than the training sequence. It contained rooms of the same categorical type (corridor, office, bathroom), and it also contained room categories not seen in the training sequence (meeting room, library). The system built by participants had to be able to answer the question 'where are you?' when presented with a test sequence imaging a room category seen during training, and it had to be able to answer 'I do not know this category' when presented with a new room category.

We received a total of 55 submissions, of which 42 were submitted to the obligatory task and 13 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 Idiap Research Institute, Martigny, Switzerland.

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

[4] `http://www.imageclef.org/`
[5] `http://www.clef-campaign.org/`

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

The image sequences used for the contest are taken from 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. The acquisition was performed on three different floors of an office environment, consisting of 36 areas (usually corresponding to separate rooms) belonging to 12 different semantic and functional categories.

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 contents of the images).

Three sequences were selected for the contest: a training sequence, a sequence that should be used for validation and a sequence for testing:

- training sequence: Sequence acquired in 11 areas, on the 6th floor of the office building, 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).
- validation sequence: Sequence acquired in 11 areas, on the 5th floor of the office building, during the day, under cloudy weather. Similar path was followed as for the training sequence; however without making the 360 degree turns.
- testing sequence - Sequence acquired in 14 areas, on the 7th floor of the office building, during the day, under cloudy weather. The robot followed similar path as in case of the validation sequence.

2.2 The Task

Participants were given training data consisting of a sequence of stereo images. 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 an ID and a semantic category of the area (usually a 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 test sequence containing images acquired in a different environment (different floor of the same building) containing areas belonging to the semantic categories observed previously (present in the training sequence) or to new semantic categories (not

imaged in the training sequence). The test images were acquired under similar illumination settings as the training data, but in a different office environment. The system should assign each test image to one of the semantic categories of the areas that were present in the training sequence or indicate that the image belongs to an unknown semantic category not included during training. Moreover, the system could refrain from making a decision (e.g. in the case of lack of confidence).

We considered 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). In task 2, the algorithm was allowed to exploit the continuity of the sequences and rely on the test images acquired before the classified image (images acquired after the classified image could not be used). The same training, validation and testing sequences were used for both tasks. The reported results were compared separately.

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. The test sequences were acquired in a different environment than the training and validation sequences (one more floor of the same building), under similar conditions, and contained additional rooms belonging to semantic categories that were not imaged previously. The algorithms trained on the training sequence will be used to annotate each of the test images. The same tools and procedure as for the validation were used to evaluate and compare the performance of each method during testing.

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 belonging to one of the known categories.
- 1 points was subtracted for each misclassified image belonging to one of the known or unknown categories.
- No points were granted or subtracted if an image was not classified (the algorithm refrained from the decision).
- 2 points were granted for a correct detection of a sample belonging to an unknown category (true positive).

#	Group	Score
1	CVG	677
2	Idiap MULTI	662
3	NUDT	638
4	Centro Gustavo Stefanini	253
5	CAOR	62
6	DYNILSIS	-20
7	UAIC2010	-77

Table 1. Results obtained by each group in the obligatory task.

#	Group	Score
1	Idiap MULTI	2052
2	CAOR	62
3	DYNILSIS	-67

Table 2. Results obtained by each group in the optional task.

– 2.0 points were subtracted for an incorrect detection of a sample belonging to an unknown category (false positive).

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 2010, 7 groups participated to the Robot Vision task, namely:

– CVG: Computer Vision and Geometry laboratory, ETH Zurich, Switzerland;
– Idiap-MULTI: The Idiap Research Institute, Martigny, Switzerland;
– NUDT: Department of Automatic Control, College of Mechatronics and Automation, National University of Defense Technology, Changsha,China.
– Centro Gustavo Stefanini, La Spezia, Italy.
– CAOR, France.
– DYNILSIS: Univ. Sud Toulon Var R229-BP20132-83957 La Garde CEDEX, France.
– UAIC2010: Facultatea de Informatica, Universitatea Al. I. Cuza, Romania.

A total of 55 runs were submitted, with 42 runs submitted to the obligatory task and 13 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 2010. 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.

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 other editions of the robot vision task.

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: Cross Language Evaluation Forum (CLEF 2005). Springer Lecture Notes in Computer Science (September 2006) 535–557
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.: Multilingual Information Access for Text, Speech and Images: Result of the fifth CLEF evaluation campaign. Volume 3491 of Lecture Notes in Computer Science (LNCS)., Bath, UK, Springer (2005) 597–613
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: CLEF 2007 Proceedings. Volume 5152 of Lecture Notes in Computer Science (LNCS)., Budapest, Hungary, Springer (2008) 473–491
4. Savoy, J.: Report on CLEF–2001 experiments. In: Report on the CLEF Conference 2001 (Cross Language Evaluation Forum), Darmstadt, Germany, Springer LNCS 2406 (2002) 27–43