Chapter 10
The Robot Vision Task

Andrzej Pronobis and Barbara Caputo

Abstract In 2009, ImageCLEF expanded its tasks with the introduction of the first robot vision challenge. The overall focus of the challenge is semantic localization of a robot platform using visual place recognition. This is a key topic of research in the robotics community today. This chapter presents the goals and achievements of the first edition of the robot vision task. We describe the task, the method of data collection used and the evaluation procedure. We give an overview of the obtained results and briefly highlight the most promising approaches. We then outline how the task will evolve in the near and distant future.

10.1 Introduction

A fundamental competence for a mobile robot is to know its position in the world. Providing robots with the ability to build an internal representation of the surrounding space, so as to be able to derive robust information about their location therein, can be considered as one of the most relevant research challenges for the robotics community today. The topic has been vastly researched, resulting in a broad range of approaches spanning from the purely metric (Jogan and Leonardis, 2003; Dissanayake et al, 2001; Wolf et al, 2005), to topological (Ulrich and Nourbakhsh, 2000; Ullah et al, 2008; Cummins and Newman, 2008), and hybrid (Thrun, 1998; Brunskill et al, 2007). As robots break down the barriers and start to interact with people (Zender et al, 2008) and operate in large–scale environments (Cummins and Newman, 2008; Ullah et al, 2008), topological models are becoming more popular

Andrzej Pronobis
Department of Computer Science, Royal Institute of Technology, Stockholm, Sweden e-mail: pronobis@nada.kth.se

Barbara Caputo
Idiap Research Institute, Centre du Parc, Rue Marconi 19, PO Box 592, CH-1920 Martigny, Switzerland e-mail: bcaputo@idiap.ch

H. Müller et al. (eds.), *ImageCLEF*, The Information Retrieval Series 32,
DOI 10.1007/978-3-642-15181-1_10,

as a way to augment, or even replace, purely metric space representations. In particular, research on building topological maps has been pushing for methods suitable for place recognition.

Traditionally, sonar and/or laser have been the sensory modalities of choice for place recognition and topological localization (Nourbakhsh et al, 1995; Martínez Mozo0s et al, 2007). The assumption that the world can be represented in terms of two dimensional geometrical information proved convenient for many applications. However, the inability to capture important aspects of complex realistic environments leads to the problem of perceptual aliasing (Kuipers and Beeson, 2002), and vastly limits the usability of purely geometrical methods. Recent advances in vision have made this modality emerge as a natural and viable solution. Vision provides richer sensory input allowing for better discrimination. It opens up new possibilities for building cognitive systems, actively relying on the semantic context. Not unimportant is the cost effectiveness, portability and popularity of visual sensors. As a result, this line of research is attracting more and more attention, and several methods have been proposed using vision alone (Torralba et al, 2003; Pronobis and Caputo, 2006; Siagian and Itti, 2007; Cummins and Newman, 2008), or combined with more traditional range sensors (Kortenkamp and Weymouth, 1994; Tapus and Siegwart, 2005; Pronobis et al, 2008).

In spite of significant progress, vision–based localization still represents a major challenge. Firstly, visual information tends to be noisy and difficult to interpret. The visual appearance of places varies with time because of illumination changes (day and night, artificial light either on and off) and because of human activities (furniture moved around, objects being taken out of drawers, and so on). Thus, the solutions must be highly robust, provide good generalization abilities and in general be adaptive. Additionally, the application puts strong constraints on the computational complexity and the increased resolution, while dimensionality of the visual data still constitutes a problem. The fact that so many different parameters influence the accuracy of a vision–based localization system is another challenge in itself, proving especially burdensome at the design stage. As the results depend greatly on the choice of training and test input data, which are unstable over time, it is hard to measure the influence of the different parameters on the overall performance of the system. For the same reason, it becomes nearly impossible to compare solutions in a fair way, as they are usually evaluated in different environments, in different conditions, and under varying assumptions. This is a major obstacle slowing down progress in the field. There is a need for standardized benchmarks and databases, which would allow for fair comparisons, simplify the experimental process and boost development of new solutions. Databases are heavily exploited in the computer vision community, especially for object recognition and categorization (Griffin et al, 2007; MIT-CSAIL, 2006). As the community acknowledges the need for benchmarking, a lot of attention is directed towards designing new data sets, reflecting the increasing capabilities of visual algorithms (Ponce et al, 2006). Also in robotics, research on Simultaneous Localization and Mapping (SLAM) makes use of several publicly available data sets (Howard and Roy, 2003; Nebot, 2006).

However, no database has yet emerged as a standard benchmark for visual place recognition applied to robot localization.

The robot vision task aims at filling this gap, and provides a benchmark to the research community working on the issues described above. The task has been introduced for the first time in 2009 and it has immediately attracted a considerable attention, with seven participating groups and a total of 24 valid runs submitted. These very encouraging first results support us in our vision and make us foresee several future editions of the challenge.

In the rest of the chapter we describe in detail the first edition of the task in Section 10.2, then give a brief overview on how the task is currently evolving in its 2010 implementations in Section 10.3. We conclude with an overall discussion and discussion on future goals.

10.2 The Robot Vision Task at ImageCLEF 2009: Objectives and Overview

The two main objectives of the robot vision task at ImageCLEF are to push forward research on semantic spatial modeling for robot localization, while at the same time making this research field easier to approach also by groups from other research fields, with no previous experience on robotics and robot vision.

To achieve this last objective, we are committed to provide to participants data sequences acquired from mobile robot platforms. This is in contrast with existing benchmark evaluation challenges in robotics, where participants are requested to operate their algorithms on robot platforms (Nebot, 2006; Howard and Roy, 2003). By making this choice, we aim at attracting the attention of researchers from the pattern recognition, computer vision and machine learning fields, who usually test their algorithms on benchmark databases but who would find it daunting to approach a full robotic system for the same task.

The achievement of the first objective requires the definition of a set of subsequent tasks, of increasing complexity over the years, so as to progressively raise the bar and focus on the open challenges that are timely to attack. In the rest of this section we describe in detail the first edition of the robot vision task, which was held in 2009, and where the focus was on topological localization from data acquired by a perspective camera. Here, the challenge was to achieve robustness under varying imaging conditions. We first give a general description of the task (Section 10.2.1). Then, we describe the data set used in more detail (Section 10.2.2). Section 10.2.3 describes how we evaluated the performance of the submitted runs. A thorough description of the outcome of the task is given in Section 10.2.4. The two coming editions of the robot vision task, organized for 2010, shifted the focus onto the place categorization problem (Section 10.3).

10.2.1 The Robot Vision Task 2009

The robot vision task at ImageCLEF 2009 addressed the problem of topological localization of a mobile robot using visual information. We asked participants to determine the topological location of a robot based on images acquired with a perspective camera, mounted on a robot platform. The image sequences were recorded in a five room subsection of an indoor environment, under fixed illumination conditions and at a fixed time. 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 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 it had to indicate that the image came from a room that was not seen during training.

The overall task was further divided in two separate sub–tasks, one mandatory and one optional. In the mandatory task, the algorithm had to provide information about the location of the robot separately for each test image. In the optional task, the algorithm was allowed to exploit the continuity of the sequences and to rely on the test images already seen.

10.2.2 Robot Vision 2009: The Database

The image sequences consisted of a subset of the publicly available IDOL2 database (Luo et al, 2007) for the training and validation set, and of a previously unreleased sequence for test. All sequences were acquired with a Canon VC–C4 perspective camera, using the resolution of 320 x 240 pixels, mounted on a MobileRobots PowerBot robot platform (see Figure 10.1). The acquisition was performed in a five room subsection of a larger office environment, selected so that each of the five rooms represented a different functional area: a one–person office, a two–person office, a kitchen, a corridor, and a printer area. Figure 10.2 shows the map of the environment.

For the training and validation sequences, the visual appearance of the rooms was captured under three different illumination conditions: in cloudy weather, in sunny weather, and at night. The robot was manually driven through each of the five rooms while continuously acquiring images and laser range scans at a rate of 5 fps. Each data sample was then labeled as belonging to one of the rooms according to the position of the robot during acquisition, rather than according to the content of the images. Examples of images showing the interior of the rooms, variations observed over time and caused by activities in the environments, as well as induced by changes in illumination, are shown in Figure 10.3.

The database was designed to test the robustness of place recognition algorithms to variations that occur over a long period of time. Therefore, the acquisition pro-

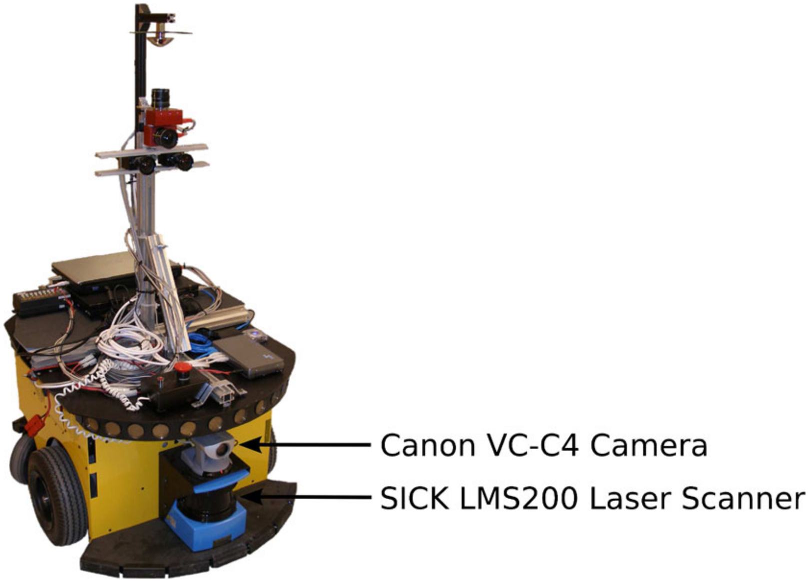

Fig. 10.1: The MobileRobots PowerBot mobile robot platform used for data acquisition in the robot vision task.

cess 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 six months later (12 sequences in total). Thus the sequences captured the variability introduced not only by illumination but also by 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 set–up. The acquisition was performed 20 months after the training data. The sequences contain additional rooms that were not imaged in the IDOL2 database.

10.2.3 Robot Vision 2009: Performance 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

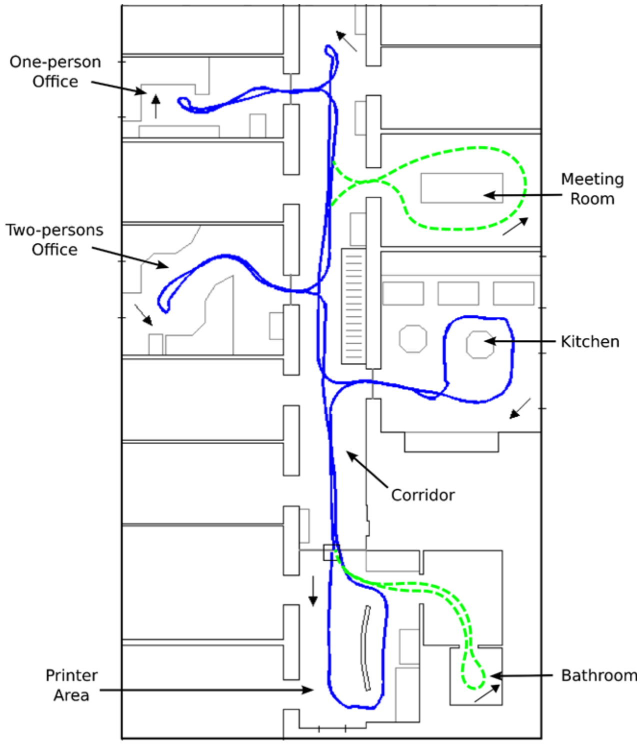

Fig. 10.2: Map of the environment with the approximate path followed by the robot during acquisition of the training, validation and testing data for the 2009 edition of the robot vision task. The dashed segments of the path correspond to the rooms available only in the test set.

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

- +1.0 point was given for each image classified correctly;
- +1.0 point was given for each image identified correctly as an unknown room;
- -0.5 points were given for each image misclassified;
- 0.0 points were given for each image where the algorithm did not provide any indication, i.e. for each not classified image.

The sum of all scores obtained for all images in the test sequences gave the overall score for each submitted run.

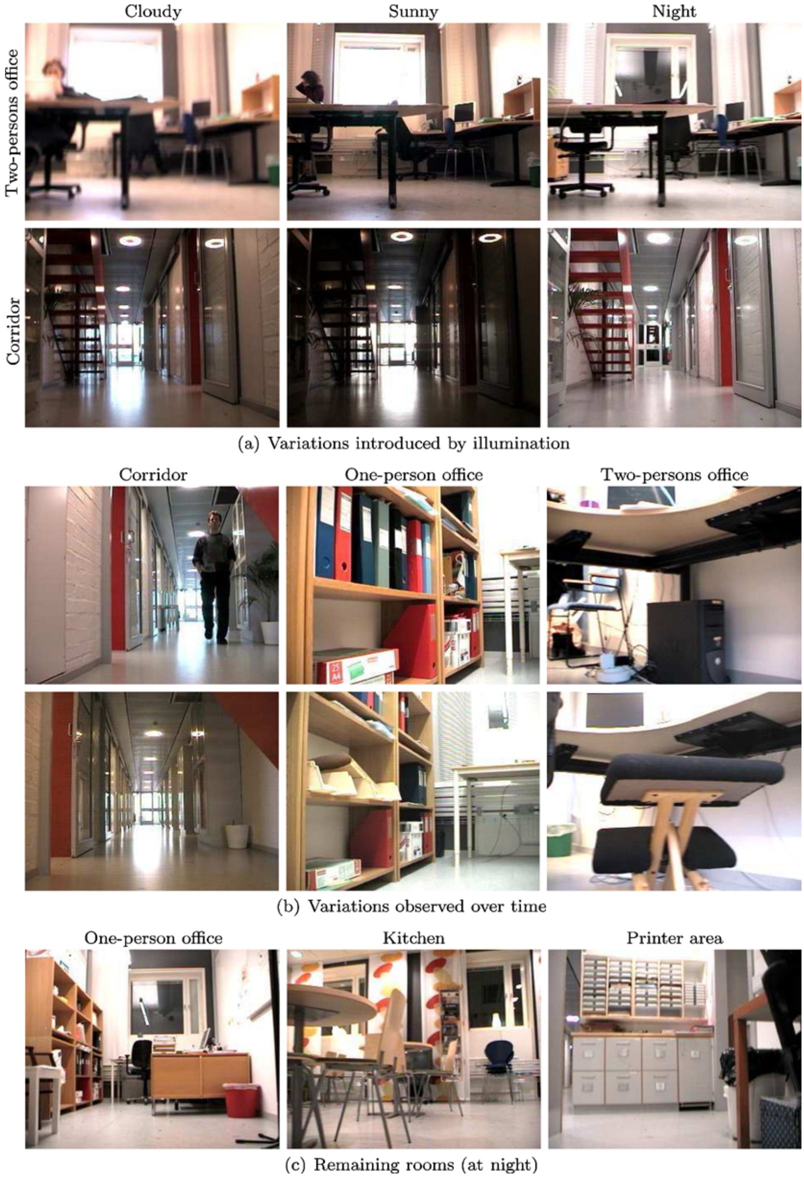

Fig. 10.3: Examples of pictures taken from the IDOL2 database showing the interior of the rooms, variations observed over time and caused by activity in the environment as well as introduced by changing illumination.

Table 10.1: Results for each run submitted to the mandatory task at the robot vision task 2009.

#	Group	Score
1	Idiap Research Institute, Switzerland	793.0
2	Faculty of Computer Science, The Alexandru Ioan Cuza University, Romania	787.0
3	Faculty of Computer Science, The Alexandru Ioan Cuza University, Romania	787.0
4	Computer Vision and Image Understanding Department, Singapore	784.0
5	Faculty of Computer Science, The Alexandru Ioan Cuza University, Romania	599.5
6	Faculty of Computer Science, The Alexandru Ioan Cuza University, Romania	599.5
7	Laboratoire des Sciences de lInformation et des Systemes	544.0
8	Intelligent Systems and Data Mining Group, Spain	511.0
9	Laboratoire des Sciences de lInformation et des Systemes	509.5
10	Multimedia Information Modeling and Retrieval Group, France	456.5
11	Multimedia Information Modeling and Retrieval Group, France	415.0
12	Multimedia Information Modeling and Retrieval Group, France	328.0
13	Faculty of Computer Science, The Alexandru Ioan Cuza University, Romania	296.5
14	Multimedia Information Modeling and Retrieval Group, France	25.0
15	Laboratoire des Sciences de lInformation et des Systemes	-32.0
16	Laboratoire des Sciences de lInformation et des Systemes	-32.0
17	Laboratoire des Sciences de lInformation et des Systemes	-32.0
18	Laboratoire des Sciences de lInformation et des Systemes	-32.0

10.2.4 Robot Vision 2009: Approaches and Results

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.

Table 10.1 shows the results for the mandatory task, while Table 10.2 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 mandatory task. A possible explanation is that the optional task requires a higher expertise in robotics that the mandatory task, which therefore represents a very good entry point.

In the following we provide an overview of the approaches used by the participants. The Scale Invariant Feature Transform (SIFT) (Lowe, 2004) 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 (Xing and Pronobis, 2010), which ranked first in the mandatory task, used SIFT combined with several other descriptors including two global image representations: Composed Receptive Field Histograms (CRFH) and PCA Census Transform Histograms

Table 10.2: Results for each run submitted to the optional task of the robot vision task 2009.

#	Group	Score
1	Intelligent Systems and Data Mining Group, Spain	916.5
2	Computer Vision and Image Understanding Department, Singapore	884.5
3	Idiap Research Institute, Switzerland	853.0
4	Intelligent Systems and Data Mining Group, Spain	711.0
5	Intelligent Systems and Data Mining Group, Spain	711.0
6	Intelligent Systems and Data Mining Group, Spain	609.0

(PACT). The algorithm employed by SIMD (Martínez-Gómez et al, 2009) relied mainly on the SIFT descriptor complemented with lines and squares detected using the Hough transform. Other participants also used SIFT (UAIC: (Boroş et al, 2009)); color SIFT (SIFT features extracted from the red, green and blue channels) combined with Hue/Saturation/Value (HSV) color histograms and multi–scale canny edge histograms (MRIM: (Pham et al, 2009)); 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: (Feng et al, 2009)); or Profile Entropy Features (PEF) encoding RGB color and texture information (LSIS: (Glotin et al, 2009)). Techniques using color descriptors ranked lower in general in the mandatory task, which might suggest that color information was not sufficiently robust to the large variations in illumination captured in the data set.

The participants applied a wide range of techniques to the place recognition problem in the mandatory task. Several variations of a simple image matching strategy were used by SIMD (Martínez-Gómez et al, 2009), UAIC (Boroş et al, 2009) and MIRG (Feng et al, 2009). Idiap (Xing and Pronobis, 2010) 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 CVIU group also used Support Vector Machines, while LSIS (Glotin et al, 2009) used Least Squares Support Vector Machines (LS-SVM). Finally, MRIM (Pham et al, 2009) 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 (Martínez-Gómez et al, 2009), which ranked first in this track, employed a particle filter to perform Monte Carlo localization. MIRG (Feng et al, 2009) used decision rules to process the results obtained for separate frames. CVIU and Idiap (Xing and Pronobis, 2010) applied simple temporal smoothing techniques, which obtained lower scores than the other approaches.

10.3 Moving Forward: Robot Vision in 2010

In this section we describe how the robot vision task has evolved during 2010. The editions of the challenge maintained the focus on visual place classification for topological localization. Specifically, we organized two editions of the task, one in conjunction with the International Conference on Pattern Recognition (ICPR 2010) and one, ongoing at the time of writing, under the ImageCLEF 2010 umbrella. The level of difficulty of the tasks proposed grew mainly in two directions:

- Image sequences were acquired by a stereo camera, as opposed to a perspective camera as in 2009.
- The number of areas to be recognized grew from five in 2009 (kitchen, corridor, one person office, two person office, printer area) to nine for the robot vision task organized jointly with ICPR2010 (elevator, corridor, kitchen, large office 1, large office 2, student office, laboratory, printer area) up to ten for the robot vision task at ImageCLEF2010 (corridor, elevator, kitchen, large office, meeting room, printer area, recycle area, small office, toilet, large meeting room).

For both editions of the task, the image sequences were acquired using a MobileRobots PowerBot robot platform equipped with a stereo camera system consisting of two Prosilica GC1380C cameras (Figure10.1). We now give a general overview of the robot vision task@ICPR2010 (Section 10.3.1) and look at the ongoing edition of the task at ImageCLEF 2010 (Section 10.3.2).

10.3.1 The Robot Vision Task at ICPR2010

In the second edition of the robot vision task the challenge was again 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 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 indicate that the image came from an unknown room. We also allowed the system to abstain from decision in the case of low confidence in the decision.

We considered two separate tasks: task 1 (mandatory) and task 2 (optional) as we did in 2009. 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 a 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 image sequences used for the contest were taken from

Fig. 10.4: Example pictures of the nine rooms used for the robot vision task at ICPR 2010.

the previously unreleased COLD–Stockholm database (Figure 10.4). The following rules were used for calculating the final score for a run:

- +1.0 point for each correctly classified image;
- correct detection of an unknown room was treated the same way as correct classification;
- -0.5 points for each misclassified image;
- 0.0 points for each image that was not classified (the algorithm refrained from the decision);
- the final score was a sum of points obtained for both sets (easy and hard).

Nine groups participated in the competition, submitting a total of 34 runs. At the time of writing, evaluation and reporting of the results are still ongoing.

Fig. 10.5: Example pictures showing the room categories used for the robot vision task at ImageCLEF 2010.

10.3.2 The Robot Vision Task at ImageCLEF2010

The third edition of the challenge, running at the time of writing, has a special focus on generalization. Participants are being 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 (Figure 10.5). The challenge is to build a system able to answer the question 'where are you?' when presented with test sequences 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 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 can refrain from making a decision (e.g. in the case of lack of confidence).

We consider two separate tasks: task 1 (mandatory) and task 2 (optional). The following rules are used when calculating the final score for a run:

- +1.0 point for each correctly classified image belonging to one of the known categories;
- -1.0 point for each misclassified image belonging to one of the known or unknown categories;
- 0.0 points for each image that was not classified (the algorithm refrained from the decision);
- +2.0 points for a correct detection of a sample belonging to an unknown category (true positive);
- -2.0 points for an incorrect detection of a sample belonging to an unknown category (false positive).

10.4 Conclusions

This chapter presents an overview over the newly established robot vision task of ImageCLEF. The overall aim of the task is to push forward research in the field of semantic robot localization using visual information. Therefore, the three editions of the task that have been held so far have addressed the issues of place recognition under varying imaging conditions (robot vision task at ImageCLEF 2009, robot vision task at ICPR 2010) and the visual place categorization problem (robot vision task at ImageCLEF 2010). Participation has been encouraging since its first edition, and it has been growing steadily over the editions.

For the future, we plan to continue posing challenging tasks on the visual place recognition problem for mobile robots, with the aim of attracting contributions to this problem from as many groups outside of the robotics community as possible. A strong focus that we foresee for the near future is the place categorization problem, that is currently one of the most baffling research issues in computer vision as well as robot vision. By introducing stereo data, we will also gently push participants to use more and more 3–D information and temporal continuity in the image sequences.

Acknowledgements We would like to thank the organizers of CLEF and ImageCLEF for their support in organizing the robot vision task.

References

Boroş E, Roşca G, Iftene A (2009) Uaic: Participation in ImageCLEF 2009 robot vision task. In: Working Notes of CLEF 2009, Corfu, Greece. 978-88-88506-84-5

Brunskill E, Kollar T, Roy N (2007) Topological mapping using spectral clustering and classification. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)

Cummins M, Newman P (2008) FAB–MAP: Probabilistic localization and mapping in the space of appearance. The International Journal of Robotics Research 27(6):647–665

Dissanayake M, Newman P, Clark S, Durrant-Whyte H, Csorba M (2001) A solution to the simultaneous localization and map building (slam) problem. IEEE Transactions on Robotics and Automation 17(3):229–241

Feng Y, Halvey M, Jose JM (2009) University of glasgow at ImageCLEF 2009 robot vision task. In: Working Notes of CLEF 2009, Corfu, Greece.

Glotin H, Zhao ZQ, Dumont E (2009) Fast LSIS profile entropy features for robot visual self–localization. In: Working Notes of CLEF 2009, Corfu, Greece.

Griffin G, Holub A, Perona P (2007) Caltech–256 Object Category Dataset. Technical Report 7694. Available at http://authors.library.caltech.edu/7694/

Howard A, Roy N (2003) The Robotics Data Set Repository (Radish). Available at http://radish.sourceforge.net/

Jogan M, Leonardis A (2003) Robust localization using an omnidirectional appearance-based subspace model of environment. Robotics and Autonomous Systems 45(1):51–72

Kortenkamp D, Weymouth T (1994) Topological mapping for mobile robots using a combination of sonar and vision sensing. In: Proceedings of the 12th National Conference on Artificial Intelligence

Kuipers B, Beeson P (2002) Bootstrap learning for place recognition. In: Proceedings of the 18th National Conference on Artificial Intelligence (AAAI'02)

Lowe D (2004) Distinctive image features from scale–invariant keypoints. International Journal of Computer Vision 60(2)

Luo J, Pronobis A, Caputo B, Jensfelt P (2007) 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

Martínez-Gómez J, Jiménez-Picazo A, García-Varea I (2009) A particle–filter-based self–localization method using invariant features as visual information. In: Working Notes of CLEF 2009, Corfu, Greece.

Martínez Mozos O, Triebel R, Jensfelt P, Rottmann A, Burgard W (2007) Supervised semantic labeling of places using information extracted from sensor data. Robotics and Autonomous Systems 55(5)

MIT-CSAIL (2006) The MIT–CSAIL database of objects and scenes. Available at http://web.mit.edu/torralba/www/database.html.

Nebot E (2006) The Sydney Victoria Park Dataset. Available at http://www-personal.acfr.usyd.edu.au/nebot/dataset.htm

Nourbakhsh I, Powers R, Birchfield S (1995) Dervish: An office navigation robot. AI Magazine 16(2):53–60

Pham TT, Maisonnasse L, Mulhem P (2009) Visual language modeling for mobile localization. In: Working Notes of CLEF 2009, Corfu, Greece.

Ponce J, Berg T, Everingham M, Forsyth D, Hebert M, Lazebnik S, Marszalek M, Schmid C, Russell C, Torralba A, Williams C, Zhang J, Zisserman A (2006) Dataset issues in object recognition. In: Towards Category–Level Object Recognition

Pronobis A, Caputo B (2006) A discriminative approach to robust visual place recognition. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'06)

Pronobis A, Martínez Mozos O, Caputo B (2008) SVM–based discriminative accumulation scheme for place recognition. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'08)

Siagian C, Itti L (2007) Biologically–inspired robotics vision monte– carlo localization in the outdoor environment. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'07)

Tapus A, Siegwart R (2005) Incremental robot mapping with fingerprints of places. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'05)

Thrun S (1998) Learning metric–topological maps for indoor mobile robot navigation. Artificial Intelligence 1:30–42

Torralba A, Murphy K, Freeman B, Rubin M (2003) Context–based vision system for place and object recognition. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV'03)

Ullah MM, Pronobis A, Caputo B, Luo J, Jensfelt P, Christensen H (2008) Towards robust place recognition for robot localization. In: Proceedings of the 2008 IEEE International Conference on Robotics and Automation,

Ulrich I, Nourbakhsh I (2000) Appearance–based place recognition for topological localization. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA 2000)

Wolf J, Burgard W, Burkhardt H (2005) Robust vision–based localization by combining an image retrieval system with monte carlo localization. IEEE Transactions Transactions on Robotics 21(2):208–216

Xing L, Pronobis A (2010) Multi–cue discriminative place recognition. In: Peters C, Tsikrika T, Müller H, Kalpathy-Cramer J, Jones GJF, Gonzalo J, Caputo B (eds) Multilingual Information Access Evaluation Vol. II Multimedia Experiments, Springer

Zender H, Martinez Mozos O, Jensfelt P, Kruijff GJ, Burgard W (2008) Conceptual spatial representations for indoor mobile robots. Robotics and Autonomous Systems 56(6):493–502