REGIMRobvid: Objects and scenes detection for Robot vision 2013

Amel Ksibi, Boudour Ammar, Anis Ben Ammar, Chokri Ben Amar and Adel M. Alimi

REGIM (REsearch Group on Intelligent Machines), University of Sfax, National Engineering School of Sfax (ENIS), BP 1173, Sfax, 3038, Tunisia

Abstract. This paper describes the participation of the REGIM team in the ImageCLEF 2013 Robot Vision Challenge. The competition was focused on the problem of objects and scenes classification in indoor environments. Objects and scenes are considered as concepts. During the competition, we aim to classify images according to the room in which they were acquired, using the information provided by the visual images only. Our system is based on PHOW features extraction and PEGASOS SVM algorithm to learn a multi-class classifier that is enable to detect the objects and the adequate scene. For this end, we focus on how to interpret the scores provided by the SVM classifier. Our system was ranked 4th among 6 teams.

Keywords: Scene selection, object detection, PEGASOS SVM, PHOW

1 Introduction

In recent years, robotics has witnessed a large growth and profound change in scope. Visual detection has becoming one of the most popular research topics and it is playing an important role in robotics ([1], [2], [4] and [8]). The ImageCLEF 2013 Robot Vision challenge has been the fifth edition of a competition that started in 2009 within the ImageCLEF as part of the [7]. The challenge addresses the problem of semantic place classification using visual and depth information. This time, the task also addresses the challenge of object and scene recognition. The rooms/categories that appear in the

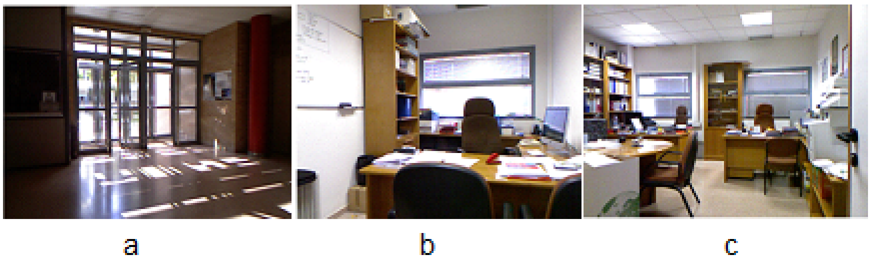

Fig. 1. Some existing rooms in the database (a) Hall (b) Proffessor office (c) Secretary

database are: Corridor, Hall, ProfessorOffice, StudentOffice, TechnicalRoom, Toilet, Secretary, VisioConferene, Warehouse and ElevatorArea (see fig. 1). The eight objects that can appear in any image of the database are: Extinguisher, Computer, Chair, Printer, Urinal, Screen, Trash, and Fridge (see fig. 2) [5].

In order to determine the presence or absence of an object, two thresholds are used (score average and score average plus standard deviation). The object does not exist if its score is below the average. Else, if its score exceeds the mean standard deviation, then the object exists. Else if the score is between the two values, this object is unknown.

The scene is selected if its score is the maximum. If the score is over than the average, then the scene is relevant. Else, the scene is unknown.

The remainder of this paper is organized as follows: the next section explains the proposed scene and object detection process. Section 3 summarizes the experiments of the work and discusses the obtained results. Finally, section 4 draws conclusions and provides suggestions for future work.

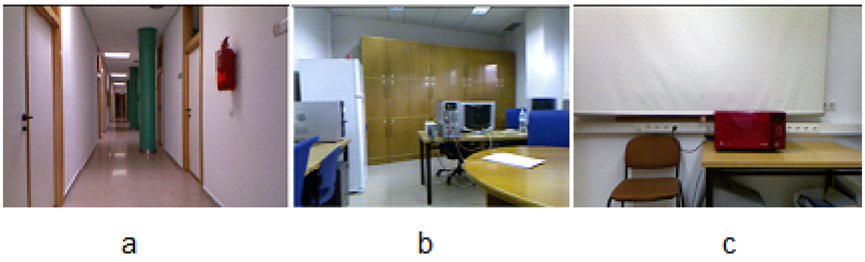

Fig. 2. Some existing objects in the database (a) Extinguisher (b) Computer (c) Chair

2 The Proposed system

2.1 Architecture of the proposed system

In this section, we describe the developed system for RobotVision2013 task participation. Two sets of concepts are used: Object concepts and Scene concepts. In our system, we aim to learn two appropriate classifiers multi-classes using visual features and machines learning for objects detection and scene detection.

Given an image, we hope to describe it using N objects and M locations or scenes. More specifically, we must detect one appropriate location and some objects.

2.2 Concept learning

a)PHOW features extraction

Visual images are represented by a Pyramid Histogram of Visual Words (PHOW) [3],

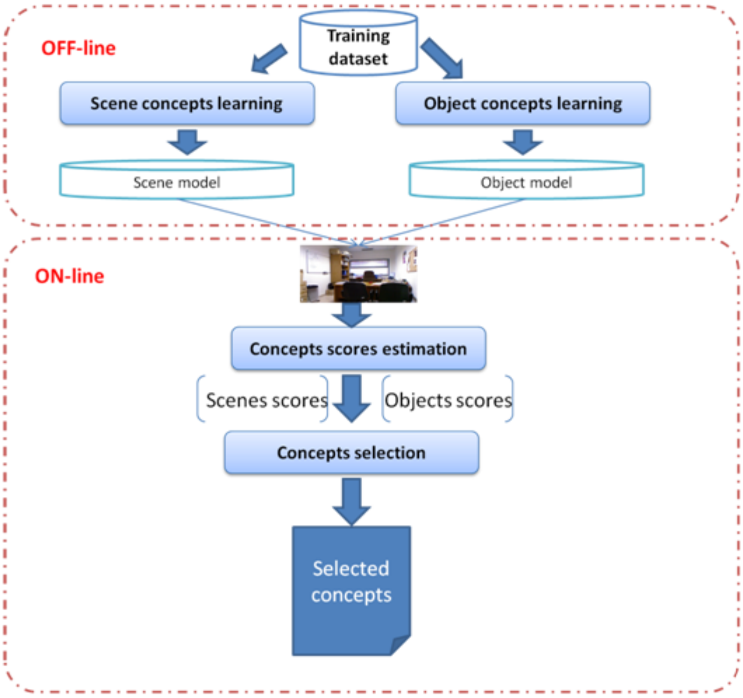

Fig. 3. Overview of the proposed system

which are a variant of dense SIFT descriptors, extracted at multiple scales. In fact, the PHOW descriptor involves computing visual words on a dense grid [10].
To compute these features, a dictionary of visual words was first generated by quantizing the SIFT descriptors that capture the local spatial distribution of gradients. ELKAN kmeans clustering is selected to perform quantization. We fixed, here, the dictionary size to 300 visual words. Then, in order to characterize the joint distribution of appearance and location of the visual words in an image, each image is divided into regions at multiple scales (44 subdivisions). So, a spatial histogram is computed for each image sub-region at each scale.

b) Object and scene learning

Object and location detection, in Robot Vision task, requires extremely fast classification. Or, automatic concept learning relying on computationally heavy kernel-based classifiers, such as non-linear SVMs, are disabled to accomplish this need resulting in a computational bottleneck. In fact, we argue that the critical efficiency criterion is the classifier evaluation cost. There have been numerous approaches to reduce the computational complexity from the level of standard non-linear SVMs such as the homogeneous kernel map which is used in our process over the Chi2 kernel SVM.

Firstly, given the obtained spatial histograms, we train the PEGASOS stochastic gradient descent as a linear SVM classifier in order to build efficiently, concepts models. While the PEGASOS SVM is very fast to train [9], it cannot typically match the performance of non-linear, as it is limited to use an inner product to compare descriptors. Therefore, we perform a step of data pre-transforming through computing the homogeneous kernel map that provides a linear representation of a Chi2 kernel. This linear approximation is used, then, to train a Chi2-kernel SVM, by applying the linear SVM solver PEGASOS.

2.3 Concept scores estimation

Given a test image, we classify it using two obtained models, respectively, for objects and location detection. The outputs of each classifier are the concept having the best score and two detection scores vectors. Since an image can contain more than one object, the decision of the object classifier is enough sufficient. So, we need to perform a process of object selection to deduce other relevant concepts underlying this image.
In addition, the selected concept by the scene classifier can have a low score despite having the highest one among other concepts. Therefore, the process of scene selection needs to be improved in a way that for the corresponding case, the system will give the result ”unknown”.

2.4 Concept Selection

Concept selection can be either objects selection or scene selection. Objects selection aims to choose an optimal subset of a predefined concepts list that is able to capture the semantic content of the corresponding image. In contrast, scene selection aims to find the most adequate scene.
As the obtained concepts scores are sparse, we need, firstly, a step of normalization and thresholding to discriminate the most representative objects and the most probably detected scene [6].

a. Scores normalization
The conventional normalization formula is as follows:

$$Vector[i] = \frac{vector[i] - min}{max - min} \tag{1}$$

where i=1..N.

i)Scores normalization by each image
For each image, we perform the normalization process using the above formula. We obtain in each one, obligatory one object having score equal to ”1” and one object having a score equal to ”0”.
In case where all objects are present, this normalization will discard the object having the low scores. In contrast, in case where any object is present, this formula will detect, always, an object with score equal to ”1”.
To overcome this problem, we propose to use normalization by each concept.
ii) Scores normalization by each concept
Given a matrix of all obtained scores for all images in the validation collection, we perform for each concept the above normalization formula.

b. Threshold for concepts selection
After the normalization of the two probability scores, we perform a process of concept selection which aims to define an adequate threshold that separate relevant concepts from others. Defining an optimal threshold for all concepts is suboptimal. So, we need to estimate for each one the corresponding threshold.

i) Object selection

Given an object, the threshold is calculated with respect to the distribution of scores of this object in all images in the validation dataset.

Two formulas are tested in experiments:

$$\tau = \begin{cases} \frac{1}{N}\sum_{i=1}^{N} c_i^q \\ \frac{1}{N}\sum_{i=1}^{N} c_i^q + \sigma \end{cases} \quad (2)$$

Where: N is the number of images, c_i^q is the score of concept q in image i.

σ is the standard deviation of all scores for N images.

ii) Scene selection

For a given image, the system selects the scene having the highest score. Meanwhile, we need to verify this decision. In fact, we define a threshold to decide if the system is sure or has an ambiguity. This threshold is equal to the average of all scores of images according to this scene concept.

3 Experiments and results

This section presents the results of the Robot Vision task of ImageCLEF 2013 for the subtask: task1. Six groups registered to the Robot Vision 2013. A total of 16 runs were submitted. The limit of the number of runs that could be submitted was 3.

For the competition we submitted two systems: the first is a system based on PHOW features extraction and PEGASOS SVM classifier with normalization and one threshold, the second uses the same techniques PHOW + PEGASOS SVM but with normalization and two thresholds. Our system ranked forth, achieving 4638.250 points on this task.

4 CONCLUSIONS and future work

We have described in this article the fifth edition of the Robot Vision task at ImageCLEF 2013, which attracted an attention of 6 groups submitting runs.

We propose an approach, which detects the location and some objects using PHOW features extraction and PEGASOS SVM classifier.

First, an off line module was performed before starting the test. It consists of PHOW extraction descriptors for object and scene concepts and the training step using PEGASOS SVM learning method.

The online process is the concepts scores estimation and the selection of concepts using the PEGASOS SVM model. Future work aims at exploiting the described methods to help elderly and disabled persons seated in wheeled chairs. Furthermore, it is also planned to use different techniques of tracking like Extended Kalman filter or incremental PCA (Principal Component Analysis) to track the detected objects.

ACKNOWLEDGMENT

The authors would like to acknowledge the financial support of this work by grants from General Direction of Scientific Research (DGRST), Tunisia, under the ARUB program.

References

1. Ammar B., Rokbani N., Alimi A. M., ”Learning System for Standing Human Detection”, IEEE International Conference on Computer Science and Automation Engineering (CSAE 2011), Page(s): 300 - 304, Shanghai 10-12 June 2011
2. Ammar B., Wali A., Alimi A. M., ”Incremental Learning Approach for Human Detection and Tracking”, 7th International Conference on Innovations in Information Technology (Innovations’11), Page(s): 128-133, Abu Dhabi 25-27 April 2011
3. A. Bosch, A. Zisserman, and X. Munoz. Image classifcation using random forests and ferns. In Proc. ICCV, 2007
4. Bousnina S., Ammar B., Baklouti N., Alimi M. A., ”Learning system for mobile robot detection and tracking” , International Conference on Communications and Information Technology (ICCIT), Page(s): 384-389, Hammamet Tunisia, June 2012.
5. B. Caputo, H. Mller, B. Thomee, M. Villegas, R. Paredes, D. Zellhofer, H. Goeau, A. Joly, P. Bonnet, J. Martinez Gomez, I. Garcia Varea, M. Cazorla, ImageCLEF 2013: the vision, the data and the open challenges. Proceedings of CLEF 2013, Springer LNCS, 2013.
6. Feki G., Ksibi A., Ben Ammar A. and Ben Amar C., REGIMvid at ImageCLEF2012: Improving Diversity in Personal Photo Ranking Using Fuzzy Logic, Proceedings of CLEF 2011, Springer LNCS, 2011.
7. Jesus Martinez-Gomez, Ismael Garcia-varea, Miguel Cazorla, Barbara Caputo, Overview of the ImageCLEF 2013 Robot Vision Task, Working notes of CLEF 2013, Valencia, Spain, 2013.
8. J. Ruiz-del-Solar and P. A. Vallejos , ”Motion Detection and Object Tracking for an AIBO Robot Soccer Player”, Robotic Soccer, Pedro Lima (Ed.), 2007.
9. Shalev-Shwartz, S., Singer, Y., Srebro, N., Cotter, A. (2011). Pegasos: primal estimated sub-gradient solver for SVM. Mathematical Programming,127(1), 3-30.
10. D. G. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision, Volume 60 issue 2, pages: 91-110, 2004.