Robot Detection with a Cascade of Boosted Classifiers Based on Haar-like Features

Anonymous

No Institute Given

Abstract. Accurate world modeling has great importance for efficient multi-robot planning in robot soccer. Visual detection of the robots on the field in addition to all other objects of interest is crucial to achieve this goal. The problem of robot detection gets even harder when robots with only on board sensing capabilities, limited field of view, and restricted processing power are used. This work extends the real-time object detection framework proposed by Viola and Jones, and utilizes the unique chest and head patterns of Nao humanoid robots to detect them in the image. Experiments demonstrate rapid detection with an acceptably low false positive rate, which makes the method applicable for real-time use.

1 Introduction

An autonomous robot is expected to operate in its environment by perceiving its surroundings, generating a plan as a sequence of actions, and executing that plan to achieve its goals. Most of the time, instantaneous perception of the world is not sufficient to make reliable decisions; therefore, the robot needs to maintain a *world model* that contains information about how the the environment is changing over time. When we consider multiple robots operating in the same environment, the robots should be aware of each other in order to plan their actions accordingly and maximize the overall utility. Accurate world modeling through detection of other robots gets even more important for a robot when operating in adversarial and collaborative environments, a perfect example of which is the game of soccer.

In soccer, players shape their plans based on the positions and actions of the teammates and the opponent players. The ability to detect the players and keep track of their locations on the field paves the way for more efficient utilization of multiple agents through higher level plans, team play, and tactics. Therefore, in order to improve the quality of the games and increase the chance of scoring by means of proper positioning and passing, detection of the robots on the field as a part of the world modeling process becomes essential. In this paper, we present a robot detection method based on Haar-like features specifically for the RoboCup Standard Platform League (SPL) [1], where Aldebaran Nao humanoid robots [2] with on board sensing capabilities, a limited field of view, and a 500 MHz processor are used.

Nao robots come with red and blue colored patches in various shapes on some parts of their bodies that are used as the "uniforms" to distinguish between the robots of different teams during SPL games. Their human-like faces and chest patches with a unique shape are the two main features that we utilize in the detection method we use. Fig. 1 gives an idea about the physical appearance of the Nao.

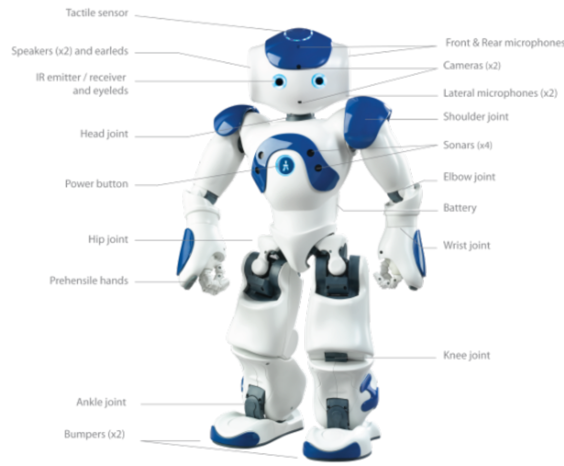


Fig. 1. Aldebaran Nao humanoid robot.

Fasola and Veloso [3] used a similar approach to detect the uniforms of the Sony Aibo robots [4], which were the previously used robot platforms of the SPL, with their main motivation being the prediction of whether the ball is occluded by a robot. They also compared this approach with a method that uses colored blobs in a color segmented image to detect the uniforms. However, as opposed to the Aibo case, there are more colored patches on the Nao and they are so far apart from each other, which makes it difficult to use colored blob based algorithms for robot detection purposes.

The rest of this paper is organized as follows. Section 2 elaborates on the methodology we followed. The experiments are explained and the obtained results are discussed in Section 3. Section 4 summarizes and concludes the paper while pointing to some possible future extensions.

2 Methodology

Haar-like features are image features used for generic object recognition purposes. They are named after Haar wavelets for their similarity. The use of a cascade of boosted haar-like features for object detection is proposed by Viola and Jones, and a robust face detector based on this method has nearly become the industry standard [5, 6].

OpenCV is a very popular computer vision software library that provides a wide range of operators, classifiers, and detectors [7]. The OpenCV library features an implementation of Viola-Jones classifier, with a training system called the Haartraining [8] which allows the users to train classifiers for arbitrary objects.

The training of such classifiers requires both positive image samples containing only the patterns of the object of interest, and negative image samples which are known for certain to not contain the object of interest. We treated different parts of the Nao's body as different objects of interest, and images containing those parts of the robot were fed to the training system. We tried training the system using the following parts of the robot, as indicated in Fig. 2:

- The face
- The chest pattern
- The whole body as a single object

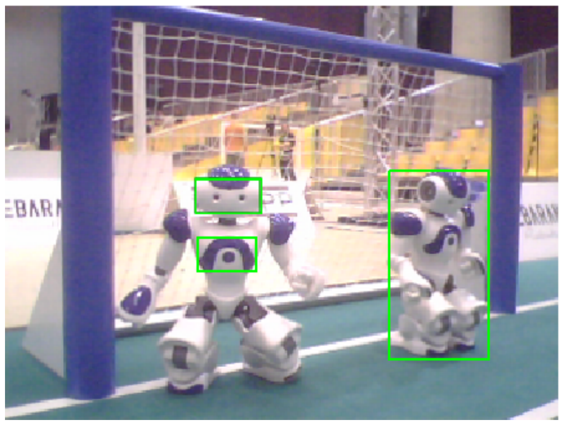


Fig. 2. Two Aldebaran Nao humanoid robots on the field, in front of the blue goal. The face and the chest parts are highlighted with rectangles. Notice the similarity between the color of the goal and the color of the robot uniforms. It is especially challenging to recognize robots accurately in such cases.

The performance of the resulting classifier heavily depends on the variety and the quality of the training data set. To increase the variety of the training set, we used artificially modied images populated from the original images in the training set. The following distortions were applied:

- Adding random skew to the image to the left or to the right
- Rotating the image clockwise or counterclockwise by a random amount
- Shearing the image randomly

These distortions allow us to bootstrap a relatively small training set to a much bigger and a more diverse training set.

2.1 Training Data Collection and Generation

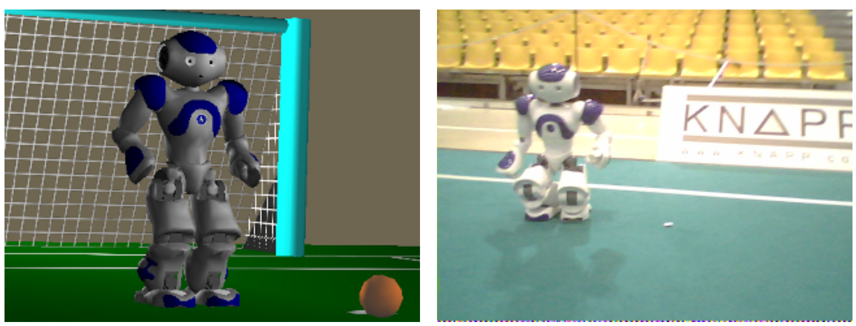
We generated two different training data sets per robot part to be recognized:

- Artificial images generated from a simulated environment

– Actual images from RoboCup 2009 competitions containing real robots on the field with realistic lighting and unpredictable noise sources on the competition site (spectators, advertisement boards, etc.)

We used Webots mobile robot simulator [9] (Fig. 3(a)) for capturing images from a simulated environment. Webots simulate the competition environment realistically; however, the obtained images from the simulator are better compared to the real images. Due to the advanced rendering capabilities of the simulator, the quality of the obtained images is between a color quantized image and a normal image captured from the robot's camera.

The real world images are extracted from the extensive log database of our team, and particularly the images that we captured during the RoboCup 2009 competitions were used (Fig. 3(b)). Most images in the database were captured from the head camera of a Nao. We specifically used these logs instead of taking a set of pictures in the laboratory to keep the various sources of noise and uncertainty (lighting changes, uncontrollable objects such as people around the fields, etc.) that do exist at the competition site but usually not in isolated laboratory environments.



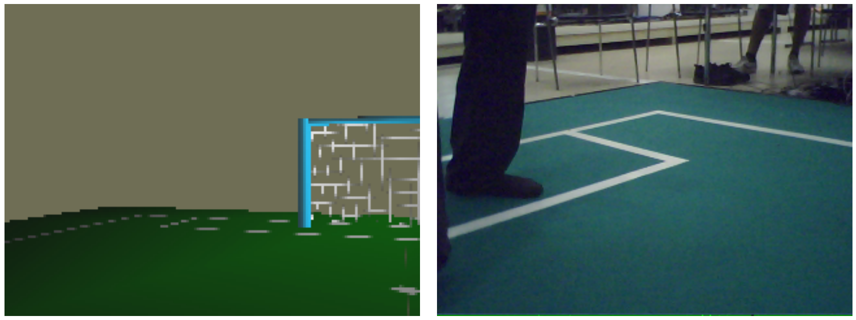
(a) The simulation environment. (b) The real competition environment.

Fig. 3. Training data set images. Notice that the image from the simulated environment lacks the noise that exists in the real world image.

Different training sets were generated for the face, the chest, and the whole body of the robot using the aforementioned images. In each training set, a subset of negative images that do not contain the object of interest are also included, an example of which is given in Fig. 4(b), 4(a). These negative images are used to check false positives once the training is completed. Several variations of positive training data with different number of images were populated per object of interest.

Artificial Images

Body: We started with training a classifier for the whole body of the robot. We prepared a set of positive training images captured from the simulator consisting of frontal and



(a) The simulation environment. (b) The real competition environment.

Fig. 4. Negative images that do not contain objects of interest.

lateral views of the Nao robot, as illustrated in Fig. 5. The result of this training was quite bad; therefore, we decided to exclude the whole body features from the process of training using real world images.

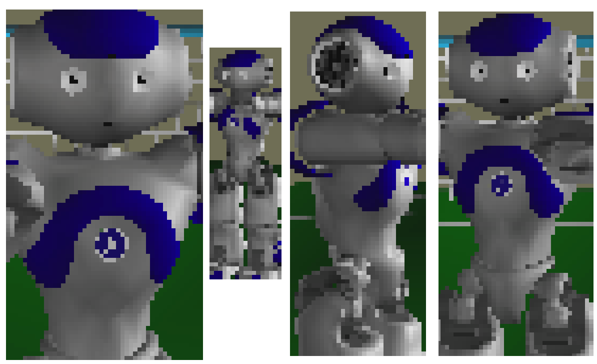


Fig. 5. Positive images of the body from the simulation world.

Face: To train a classifier for detecting face patterns of the Nao, we first prepared a set consisting of images captured from the simulator that include different views of the head of the robot from different angles, examples of which are depicted in Fig. 6.

Real World Images

Face: To obtain training images for the face pattern from the real world logs, we cropped positive images in such a way that the resulting sub-images would contain only the frontal view of the robot's head. The two variations of frontal view used were the frontal view of the whole head (Fig. 7(a)), and the frontal view of the eye region only (Fig. 7(b)).

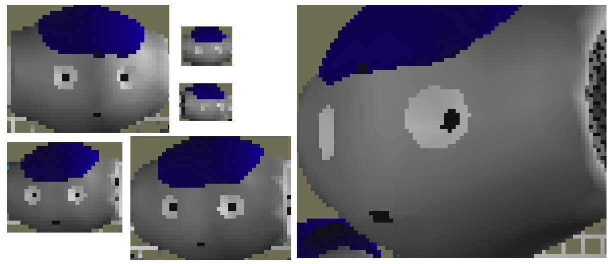


Fig. 6. Positive images of the face from the simulation world.

(a) Large crops. (b) Smaller crops.

Fig. 7. Facial positive images.

Chest: The specific pattern on the chest area of the Nao, as indicated in Fig. 8, is unique; that is, only a Nao robot has this pattern on the field, under normal conditions. Therefore, that makes the chest region an important part for detection of the robots.

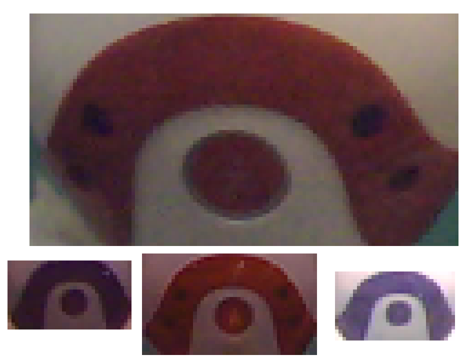


Fig. 8. Positive images of the chest pattern.

2.2 Training

The object detection procedure employs simple features, reminiscent of the Haar basis functions [10]. In the paper of Viola and Jones [5], three types of these Haar-like features are utilized; namely, two, three, and four rectangle features, which help in computing the difference between the sum of pixels within rectangular areas in the image. These features are pictured in Fig. 9.

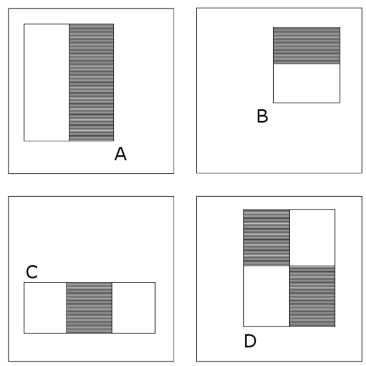


Fig. 9. Rectangular Haar-like features employed in the work of Viola and Jones [5]. A and B are the two-rectangle features. C is the three-rectangle feature and D is the diagonal four-rectangle feature.

The feature extraction method proposed by Viola and Jones, scans the images at 11 different scales and uses an intermediate transformation of the image, called the *integral image*. In contrast to the conventional approach that requires computation of pyramids of 11 images, they use a cascading property that reduces the computation time and increases the detection performance. A variant of AdaBoost, an adaptive meta-algorithm for improving the learning performance by adjusting the strong classifier threshold to minimize the false negatives, is used for deciding which of the extracted features the weak classifiers contribute the most to form the final strong classifier.

The whole system is packed under the name "The Haartraining" and used as a generic object detection training algorithm in the literature [8]. The data collection and training procedures we followed for robot detection task can be summarized as follows:

1. Crop images to generate positive images that contain only the pattern to be detected
2. Pick some negative images that do not contain the patterns of interest
3. Generate M random training samples from the set of positive images
4. Generate N random test samples from the set of negative and positive images
5. Apply Haartraining to M training samples to generate an XML file of the haar cascade descriptor
6. Evaluate the performance of the descriptor on the test samples

3 Experiments and Results

To standardize the testing procedure, we kept the number of distorted images from the original positive images at 1000, and generated 200 test images for each trial. Table 1 summarizes the experiments according to the pattern, the original number of positive images containing the pattern, and the pattern properties in each trial.

Table 1. Comparison of different detection setups.

The pattern	training images	image properties	hits	missed	false
Robot body (simulation)	30	noisy	20	180	1200
Robot face (simulation)	10	noisy	173	27	180
Robot face	27	noisy	176	24	844
Robot face	4	clear	170	30	365
Robot chest	33	noisy	169	31	113
Robot chest	30	noisy	144	56	16
Robot chest	4	clear	157	43	5

The artificial dataset was used mostly for exploration purposes. The performance evaluation after the training using the artificial dataset yielded bad results. In the case of the robot body detection, the algorithm detected 1200 "ghost" robots in addition to 20 hits, which corresponds to 600%. Obtaining a hit ratio of 10% (20 out of 200) indicated that using full body of the robot as the positive patterns was not a good idea. The hit ratio was 86% for the artificial robot face detection; however, the false positive rate was 93%, which is also very high. Both results obtained in the simulation environment directed us to use more specific patterns for detection, such as the face or the chest pattern instead of the whole body pattern.

The experiments were repeated several times using the real world dataset. In the first attempts the positive sets included many noisy images. Compared to the robot face detection results in the artificial dataset, the results of robot face detection in the real world data set were worse, with a hit ratio of 88% and a false positive rate of 400%, which can be interpreted as an average of four "ghost" robots were detected in each image. Using less noisy and clearer positive images, the false positive rate could be reduced to 200%, which is still not acceptable since it would result in finding many "ghost" robots on the field.

The results of the chest pattern were quite acceptable compared to the previous results. As the chest pattern is specific to the Nao humanoid robot, the success of the results were expected. The false positive rate in the chest detection was only 2.5%, and the hit rate was 79%, which are highly acceptable results as the robot would not detect many false robots on the field, and would detect approximately four of the five robots within its field of view. The required time to process an image frame for detection is 10 ms on a conventional computer. This corresponds to about 40 ms on the processor of the Nao robot ($\tilde{2}$0fps), which is also reasonable for real-time robot detection implementation on the actual robot.

4 Conclusions and Future Work

We contribute a robust robot detection algorithm utilizing a cascade of boosted classifiers based on Haar-like features for quickly and efficiently detecting Aldebaran Nao humanoid robots during the soccer games of RoboCup Standard Platform League. We treated the robot as a whole and as a collection of parts, training separate classifiers for each part. To improve the classier performance, we investigated the possibility of augmenting the training set using artificial images captured from the simulation environment. Furthermore, we used a variety of distortions like rotation, skew, and shear to create modified versions of the training images, enlarging the training set both in size, and in variety of features. Experimental results show that despite the intuitive expectation of the face detection of the robot to yield better results, the chest pattern was the most distinguishable and easy to recognize part of the robot. The resulting recognition system runs at around 20fps, which makes it suitable for use in actual games.

Investigation of preparing better stratified training sets, developing a method for making the selection process easy for extensive actual competition logs, and trying to use haar features on the image channels Cb and Cr along with the channel Y are among the possible future extensions.

References

1. The RoboCup Standard Platform League. http://www.tzi.de/spl.
2. Aldebaran-Nao. http://www.aldebaran-robotics.com/eng/Nao.php.
3. Juan Fasola and Manuela Veloso. Real-Time Object Detection using Segmented and Grayscale Images. In *Proceedings of ICRA'2006*, pages 4088–4093, May 2006.
4. The Sony Aibo robots. http://support.sony-europe.com/aibo/.
5. Paul Viola and Michael Jones. Rapid Object Detection using a Boosted Cascade of Simple Features. In *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, volume 1, page 511, Los Alamitos, CA, USA, 2001. IEEE Computer Society.
6. Paul A Viola and Michael J Jones. Robust Real-Time Face Detection. *International Journal of Computer Vision*, 57(2):137—-154, 2004.
7. G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.
8. Naotoshi Seo. Tutorial: OpenCV haartraining (Rapid Object Detection With A Cascade of Boosted Classifiers Based on Haar-Like Features), 2007.
9. Webots Mobile Robot Simulation Environment. http://www.cyberbotics.com.
10. Constantine P. Papageorgiou, Michael Oren, and Tomaso Poggio. A general framework for object detection. In *ICCV '98: Proceedings of the Sixth International Conference on Computer Vision*, page 555, Washington, DC, USA, 1998. IEEE Computer Society.