A Study on Detecting Three-Dimensional Balls using Boosted Classifiers

Mitchell Metcalfe
The University of Newcastle
Callaghan NSW 2308,
Australia
mitchell.metcalfe@
newcastle.edu.au

Brendan Annable
The University of Newcastle
Callaghan NSW 2308,
Australia
brendan.annable@
newcastle.edu.au

Monica Olejniczak
The University of Newcastle
Callaghan NSW 2308,
Australia
monica.olejniczak@
newcastle.edu.au

Stephan K. Chalup
The University of Newcastle
Callaghan NSW 2308,
Australia
stephan.chalup@
newcastle.edu.au

ABSTRACT

Many recent approaches to ball detection in robot soccer reduce the task to edge-based circle detection, or training a classifier to detect specific balls with known colour or surface texture. In the present work, a more general approach to ball detection is investigated, where spherical 3D objects must be detected under unknown lighting, colouring and texturing. Pilot experiments applied techniques stemming from the face detection literature, namely boosted-classifiers using extended Haar features, and Local Binary Patterns (LBPs) as features. Disk-like objects were included as negative samples in the training set in order to produce a detector that does not misclassify circular, disk-like objects as 3-dimensional balls. The resulting classifiers were able to detect homogeneously or moderately textured balls while robust detection of balls with unknown strong patterns still remains a challenge.

Keywords

computer vision, sphere detection, adaboost, classification

1. INTRODUCTION

Central to the game of football or soccer is the task of detecting and tracking the ball. While two competing teams interact in gaining control over the ball on the field there is also a large crowd of emotional spectators that all try to solve the question “Where is the ball?” and to follow it visually. While it appears to come naturally to humans, fast and reliable ball detection has presented a significant challenge that has attracted much research, for example, in the robot soccer community [7]. Robust 3D object detection will also be important for developers of interactive entertainment applications, such as augmented reality games and networked exergames, where remote players may interact with real objects such as balls [13].

Early attempts at ball detection algorithms used in the international robot soccer competition, RoboCup, used simple histogramming techniques targeted at a specific range of colour intensities to find a coloured ball on a green field. As the RoboCup playing field has become less structured over the years, competing teams have needed to account for unpredictable colours and have increasingly implemented methods that detect the shape of the ball as well [14]. Schulz et al. [17] used a neural network on subsampled luminance images of the ball to detect the shape of the ball. Recent approaches have focused on detecting the approximately circular shape of the ball in typical images. These include clustering, Hough filters [8], and RANSAC [1, 4].

Many current ball detection methods make assumptions about the ball or the environment that limit their applicability in a more general case. Methods based on colour classification or similarity can suffer from false positive detections due to other objects having similar colours. These methods may also detect both false positives and false negatives due to unexpected changes in lighting [6]. Methods based on circle or ellipse detection can be prone to false positives due to the presence of disk-like objects in the environment, or due to objects that appear circular when viewed from specific angles.

To avoid the limitations of methods based on assumptions such as these, we developed a sphere detection method which uses the shading patterns characteristic of spherical objects to classify objects as spheres.

Nillius et al. [15] perform shading based sphere detection using Principal Component Analysis (PCA) with a basis derived analytically from a given Bidirectional Reflectance Distribution Function (BRDF) and assumptions on scene illumination. While this method works well for plain untextured spheres, it is not designed to work on spheres with patterns printed on them, like many soccer balls.

ACSW ’16 Multiconference, February 02-05, 2016, Canberra, Australia

DOI: http://dx.doi.org/10.1145/2843043.2843473

To build a detector that is more robust to differently textured spheres, we investigated the application of techniques popularised in the realm of face detection to the task of sphere detection. We consider this a promising approach, because the 3D shading features of spheres tend to have similar spatial constraints to facial features in many cases. We assumed that the balls to be detected are resting on the ground and are illuminated from above to constrain the likely positions of shadows and specular highlights.

Masselli et al. [11] successfully apply a boosted Haar classifier [20] to the problem of ball recognition. They show that the Haar classifier outperformed a more classical approach, based on a Hough transform, in the task of detecting uniformly yellow, green, and white balls.

A similar approach [21] attempted to detect a wide variety of generic FIFA-style balls by using extended Haar features [10] as weak classifiers. Zhang et al. [21] reported improved performance when modified Haar features that used a division operation between their area sums, instead of the usual subtraction, were included. This suggests that exploring alternate weak classifiers could lead to valuable performance improvements.

Mitri et al. [12] applied a Sobel filter and a threshold function to each image as preprocessing steps, passing only the detected edge images to the classifiers. The method learnt Classification and Regression Trees (CARTs) of Haar features instead of directly using Haar features as weak classifiers. Their system performed sufficiently well for ball tracking, but detected other round objects as false positives. It also performed significantly better when a more complex training dataset was applied, which included images under different lighting conditions and environments. We consider it likely that their poor false positive rate was a symptom of ignoring the shading information of the spheres by using only an edge image.

Treptow and Zell [18] achieved a much lower false positive rate using Haar features directly, but only trained and tested their detector on a single ball.

As a result of targeting our approach toward the 3D features that distinguish 3-dimensional balls from objects such as disks, our method is expected to be particularly robust against detecting false positives.

2. TRAINING FOR SPHERE DETECTION

The aim of this study is to create a robust sphere detector that does not detect disk-like objects as false positives. To achieve this, we trained Viola-Jones detector cascades [19] using each of the following two different feature types: Extended Haar features [10] and Local Binary Patterns (LBPs) [9]. We hypothesised that including a large proportion of disk-like negative images in the training set would increase the precision of the resulting sphere detectors.

3. EXPERIMENT

In order to test the hypothesis using the two chosen feature types, two sets of six classifiers were trained. Each set of classifiers used a single feature type, and a range of six different proportions of disk-like negative images in the training set. A summary of the training parameters used in the trials is presented in Table 1.

The parameters listed in Table 1 are defined as follows:

Table 1: A summary of the 12 experimental trials

Parameter	Value
Set size	9300
Pos. %	40
Hard Neg. %	{0, 10, 20, 30, 40, 50}
NumPos	200
NumNeg	350
Feature type	{Haar, LBP}

1. **Set size:** Number of samples in the training set.
2. **Positive %:** The percentage of samples in the training set that are positive (i.e. that are images of spheres).
3. **Hard Negative %:** The percentage of negative samples in the training set that are 'hard negatives' (i.e. they are images of disk-like objects, as opposed to miscellaneous background images).
4. **NumPos:** The number of positive samples shown to each stage of the classifier cascade during training.
5. **NumNeg:** The number of negative samples shown to each stage of the classifier cascade during training.

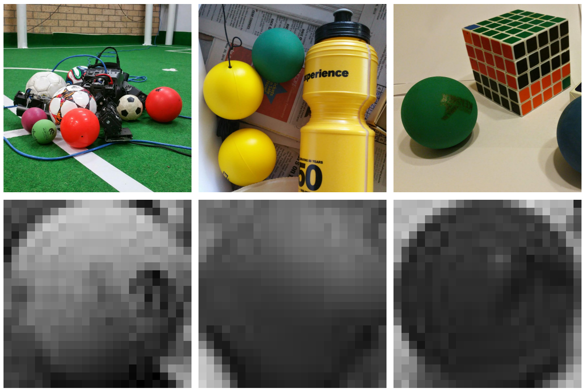
3.1 Image Dataset Compilation

ImageNet [3] is an image database organised according to the WordNet hierarchy [2] that was used as a source of training and testing data. This resource offers a wide range of images along with associated class and bounding box information. The use of ImageNet has facilitated the collection of many more training samples than would be possible to collect manually within the time-frame of this project.

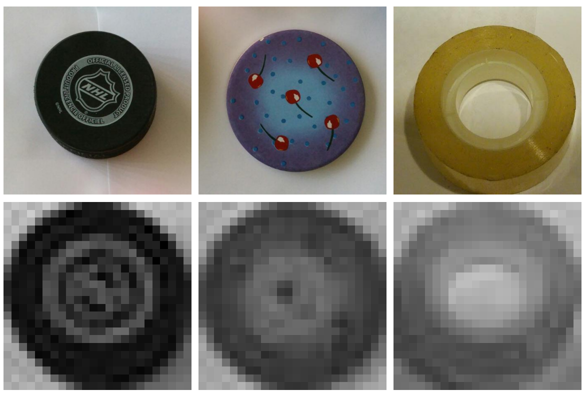
Only images that contained bounding box information were extracted from ImageNet for use as positive training samples. Prior to training, each positive sample was cropped from its original image, using the provided bounding box information, before being resized to 24×24 pixels and converted to grayscale. An example of this process has been demonstrated in Figure 1.

Several WordNet categories (known as *synsets*) were used to compile the training and test set for this project. Table 2 represents the positive samples used in the training set. The images for the negative training set were drawn from both the list of categories of disk-like objects in Table 3, and the list of categories of background images given in Table 4. All *hyponyms* (synsets representing subcategories) of the listed synsets were also included in the dataset, except for those listed in Table 5, as they contained images of objects that were considered too far from either spheres or circles in shape. A randomly selected 20% of the samples of each of the two sample sets was set aside for use as a test set before beginning training.

ImageNet provided a sufficient quantity of positive samples along with bounding box information, but relatively few negative samples (of non-spherical circular objects) with bounding boxes. Additional bounding box extraction was performed to generate the negative training and test sets using ELSD (Ellipse and Line Segment Detector) [16]. The reference implementation of ELSD was run on each of the background images that did not have associated bounding



(a) Positive samples.



(b) Hard negative samples.

Figure 1: Examples of positive (a) and hard negative (b) samples used for training. The first row represents the original image, while the second row represents the same samples that have been cropped, resized and converted to grayscale prior to training.

box data. A tight bounding box around each detected ellipse was cropped and resized to form a 24×24 sample image. Only ellipses that ELSD output to SVG as closed arcs were used, and ellipses with width or height less than 10% of their containing image were rejected.

The background training set was constructed by randomly sampling one square window from each background image, with side lengths uniformly sampled to be between 20% and 100% of the size of the smaller dimension of the image.

4. RESULTS

The performance of the classifier trained in each trial was evaluated by calculating its *precision* (the number of true positives divided by the sum of the numbers of true positives and false positives) and *recall* (the number of true positives divided by the sum of the numbers of true positives and false negatives) when run on the test set. The results of the classifier training and testing performed for the experiment are presented graphically in Figure 2.

5. DISCUSSION

A clear trend is visible in Figure 2, where the precision of the classifiers that used Haar features was much greater

Table 2: ImageNet synsets used for the positive samples in the training set and their associated WordNet identifiers.

Id	Name
n02778669	Ball
n02779435	Ball
n02799071	Baseball
n03131967	Cricket ball
n03445777	Golf ball
n13899404	Ball, globe, orb

Table 3: ImageNet synsets used for the negative samples in the training set and their associated WordNet identifiers.

Id	Name
n03032811	Circle, round
n13873502	Circle
n13873917	Circle
n13875185	Disk, disc, saucer
n13875392	Ring, halo, annulus, doughnut, anchor ring
n13875970	Coil, whorl, roll, curl, curlicue, ringlet, gyre, scroll
n13902336	Rim

than those that used LBPs across the entire range of hard negative percentages used for the training set. No significant difference in the recall of the two classifier types was observed. Based on these observations, the results suggest that Haar features are more appropriate than LBP features for the sphere detection task, as the Haar classifiers had the same or better precision and recall in every training scheme. We suggest that this may be because the differences of intensity of image regions used in Haar features allow them to better capture the low frequency features that distinguish spheres from disks, i.e. shading due to curvature, whereas LBP patterns mostly describe edges, corners, and points, and so more often classify other rounded objects as spheres.

We hypothesised that including a greater proportion of images of disk-like objects as negative samples in the training set would increase the precision of the resulting trained classifiers. Figure 2 shows that the precision of Haar-based classifiers in the experiment increased by over 10% (from 61.7% to 72.4%) as the proportion of hard negative images in the training set was increased from 0% up to 20%. Further increases in the proportion of hard negatives caused the precision to fall to 65.5% The precision of the LBP-based classifiers grew from 48% when trained with no hard negatives to a maximum of 50% when trained with 20% hard negatives in the training set. The precision fell to 37% as the proportion of hard negative samples in the training set increased to 50%. These results do not support the hypothesis for LBP classifiers, as the initial increase in precision was small, and further increases in the hard negative percentage caused the precision to quickly fall below its initial value (when trained without hard negatives). The results for the classifiers using Haar features do support the hypothesis, as the precision improved significantly when a small proportion of hard negatives were used, and remained above its initial value for the full range of hard negative proportions tested.

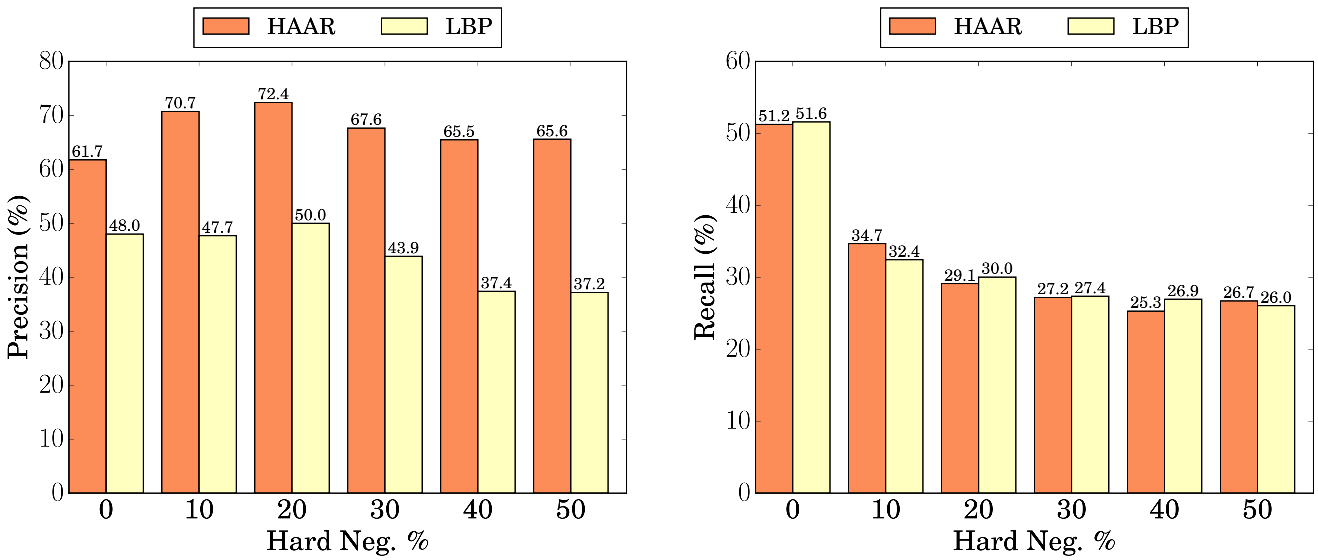


Figure 2: Column graphs comparing the proportion of the negative samples in the training set of each trial that were 'hard negatives' (images of round, non-spherical objects) with the resulting rates of precision and recall.

Table 4: ImageNet synsets used for the background samples in the training set and their associated WordNet identifiers.

Id	Name
n02782778	Ballpark, park
n02913152	Building, edifice
n03841666	Office, business office
n04335209	Street
n08524735	City, metropolis, urban center
n08659446	Field

Table 5: Blacklisted ImageNet synsets and their associated WordNet identifiers.

Id	Name
n04023962	Punching bag, punch bag, punching ball, punchball
n04118538	Rugby ball
n04186051	Shaving cream, shaving soap
n09229709	Bubble

The precision of both classifiers quickly fell from close to 50% to below 35% as the hard negative proportion increased from 0% to 10%. It gradually decreased to a minimum of 25.3% as the hard negative proportion increased up to 50%.

5.1 Live Classification Performance Evaluation

The live classification performance of one of the best performing classifiers from the experiment was tested to qualitatively evaluate the quality of the training process. This test involved the use of a HD webcam to perform live object detection. The webcam was presented with a number of balls with the intention of observing the robustness of the trained sphere detector. This experiment appeared to perform quite well and detected most of the balls in the scene. The sphere detector, however, did not perform well when presented with a highly textured ball as seen in Figure 4.

Figure 4: Ball that often could not be detected.

It was also noted that the detector did not detect the balls when the webcam was significantly rotated, such that the ball (and its shading pattern) was upside down with respect to the camera. This effect is illustrated in Figure 3. This supports the suggestion that the detectors using Haar features perform shading based detection and not simply circle detection.

6. CONCLUSION

This paper approached the problem of sphere detection using boosted classifiers. We hypothesised that including a larger proportion of disk-like images as negative samples in the training set would increase the precision of the resulting sphere detectors. The hypothesis was tested in a controlled experiment that compared the appropriateness of two dif-

Figure 3: Sequence of rotated webcam images. In the training data it was assumed that natural light usually comes from above. The displayed results indicate that the ball detector uses the geometric positioning of shadows and highlights as a characteristic feature that is common to 3D balls.

ferent feature types to the problem: extended Haar features and LBPs.

The results of the pilot experiments supported the hypothesis for classifiers using Haar features. Further experiments would be required to determine whether the small increase in precision seen in the LBP results is significant, and whether the recall of both classifier types can be increased to a more useful level through the use of a larger training set. Additionally, our results suggested that Haar features are better suited to the problem of sphere detection than LBPs.

Future work could repeat the experiments described in this paper with a greater number and variety of training schemes. Additional feature descriptors such as Histograms of Oriented Gradients (HoG) could be investigated as well as the use of image preprocessing techniques such as histogram equalisation, gamma intensity correction and high and low pass filtering [5]. The preprocessing techniques would be applied to all sample windows prior to training and upon classification, in an effort to emphasise the image features that best distinguish spheres from disks and background samples.

The pilot results of this study indicate that a 3D ball detector can be trained using similar techniques as are common for face detection. It remains a challenge to devise a training scheme that leads to a classifier that can detect the relatively weak 3D shading characteristics of a sphere in the presence of other strong patterns (such as shown in Figure 4). This study aims to contribute to the development of a universal ball detector that is not trained on a specific ball with a specific pattern and is robust to changes of lighting conditions so that it can be employed without recalibration. Future research plans where the results of the present study can become important include interactive game scenarios where autonomous robot soccer play is merged with robotic telepresence through virtual reality helmets.

7. REFERENCES

[1] B. Annable, D. Budden, S. Calland, S. K. Chalup, S. K. Fenn, M. Flannery, J. Fountain, R. A. King, A. Mendes, and M. Metcalfe. The NUbots team description paper 2013, RoboCup Humanoid League, 2013.

[2] F. Christiane. *WordNet: An Electronic Lexical Database*. Bradford Books, 1998.

[3] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *IEEE Conference on Computer Vision and Pattern Recognition, 2009 (CVPR 2009)*, pages 248–255, 2009.

[4] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[5] R. Gross and V. Brajovic. An image preprocessing algorithm for illumination invariant face recognition. In *Audio-and Video-Based Biometric Person Authentication - 4th International Conference, AVBPA 2003 Guildford, UK, June 9âĂŞ11, 2003 Proceedings*, volume 2688 of *Lecture Notes in Computer Science*, pages 10–18, Berlin, 2003. Springer-Verlag.

[6] N. Henderson, R. King, and S. Chalup. An automated colour calibration system using multivariate gaussian mixtures to segment hsi colour space. In J. Kim and R. Mahony, editors, *Proceedings of the 2008 Australasian Conference on Robotics and Automation (ACRA 2008)*. ARAA (on-line), 2008.

[7] H. Kitano, M. Asada, Y. Kuniyoshi, I. Noda, E. Osawa, and H. Matsubara. Robocup. a challenge problem for ai. *AI Magazine*, 18(1):73–85, 1997.

[8] X. Li, H. Lu, D. Xiong, H. Zhang, and Z. Zheng. A survey on visual perception for RoboCup MSL soccer robots. *International Journal of Advanced Robotic Systems*, 10:110–120, 2013.

[9] S. Liao, X. Zhu, Z. Lei, L. Zhang, and S. Z. Li. Learning multi-scale block local binary patterns for face recognition. In *Advances in Biometrics - International Conference, ICB 2007, Seoul, Korea, August 27-29, 2007. Proceedings*, volume 4642 of *Lecture Notes in Computer Science*, pages 828–837. Springer, 2007.

[10] R. Lienhart and J. Maydt. An extended set of haar-like features for rapid object detection. In

International Conference on Image Processing, 2002, volume 1, pages 900–903. IEEE, 2002.

[11] A. Masselli, R. Hanten, and A. Zell. Robust real-time detection of multiple balls on a mobile robot. In *European Conference on Mobile Robots (ECMR 2013)*, pages 355–360, September 2013.

[12] S. Mitri, K. Pervölz, H. Surmann, and A. Nüchter. Fast color-independent ball detection for mobile robots. In *Proceedings of the IEEE Computer Society International Conference Mechatronics and Robotics*, pages 900–905, 2004.

[13] F. F. Mueller, D. Edge, F. Vetere, M. R. Gibbs, S. Agamanolis, B. Bongers, and J. G. Sheridan. Designing sports: A framework for exertion games. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 2651–2660, New York, NY, USA, 2011. ACM.

[14] C. L. Murch and S. K. Chalup. Combining edge detection and colour segmentation in the four-legged league. In *Proceedings of the 2004 Australasian Conference on Robotics & Automation (ACRA'2004)*. ARAA (on-line), 2004.

[15] P. Nillius, J. Sullivan, and A. Argyros. Shading models for illumination and reflectance invariant shape detectors. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2014)*, pages 1–8, Los Alamitos, CA, USA, 2008. IEEE.

[16] V. Patraucean, P. Gurdjos, and R. G. von Gioi. A parameterless line segment and elliptical arc detector with enhanced ellipse fitting. In A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid, editors, *Computer Vision - ECCV 2012, 12th European Conference on Computer Vision Florence, Italy, October 7-13, 2012 Proceedings, Part II*, volume 7573 of *Lecture Notes in Computer Science (LNCS)*, pages 572–585, 2012.

[17] H. Schulz, H. Strasdat, and S. Behnke. A ball is not just orange: Using color and luminance to classify regions of interest. In *Proceedings of the Second Workshop on Humanoid Soccer Robots, Pittsburgh, PA, USA*, 2007.

[18] A. Treptow and A. Zell. Real-time object tracking for soccer-robots without color information. *Robotics and Autonomous Systems*, 48(1):41–48, 2004.

[19] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001)*, volume 1, pages 511–518. IEEE, 2001.

[20] P. Viola and M. Jones. Robust real-time object detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.

[21] H. Zhang, H. Lu, P. Dong, D. Xiong, and Z. Zheng. A novel generic ball recognition algorithm based on omnidirectional vision for soccer robots. *International Journal of Advanced Robotic Systems*, 10(388), 2013.