# A NEURAL APPROACH TO REAL TIME COLOUR RECOGNITION IN THE ROBOT SOCCER DOMAIN

José Angelo Gurzoni Jr.*, Murilo Fernandes Martins†, Flavio Tonidandel*, Reinaldo A. C. Bianchi*

*Dept. of Electrical Engineering
Centro Universitário da FEI, Brazil

†Dept. of Electrical and Electronic Engineering
Imperial College London, UK

Emails: jgurzoni@fei.edu.br, mfmartin@imperial.ac.uk, flaviot@fei.edu.br, rbianchi@fei.edu.br

**Abstract**— This paper addresses the problem of colour recognition in real time computer vision systems applied to the robot soccer domain. The developed work presents a novel combination between our current model-based computer vision system and neural networks to classify colours and identity the robots on an image. This combination presents an alternative to the considerably high computational cost of the neural network, which reduces the demanding and time consuming task of calibrating the system before a match. The results prove that this new approach copes with luminance variations across the playing field and enhances the accuracy of the overall system.

**Keywords**— Neural Networks, computer vision, mobile robotics

## 1 INTRODUCTION

Since its introduction to the Artificial Intelligence research field, the Robot Soccer domain has become a widely studied area of research. The Robot Soccer domain introduces a real, dynamic and uncertain environment, requiring real time responses from a team of robots that must, autonomously, work together to defeat an opponent team of robots. The reason for such popularity comes from the diversity of challenges addressed by this domain in many areas of research, including machine learning, multi-robot systems, computer vision, control theory and so forth.

A robot soccer team must be endowed with sensory and perceptual capabilities in order to actuate in the environment. The richest perceptual data of a robot soccer system is, for sure, provided by the computer vision system, responsible for the recognition of the robots and ball pose. The computer vision system must retrieve the pose of the objects recognised in the scene accurately, in real time and should cope with luminance variation, occlusion, lens distortion and so forth. Notice that, as different teams and even robots are distinguished by different colours, the computer vision system must recognise these colours in order to provide the right information about the scene observed.

Although largely studied by several researchers (e.g., (Comaniciu and Meer, 1997), (Bruce et al., 2000)), colour segmentation is still an open issue and worthy of consideration, as recent studies like (Tao et al., 2007), (Sridharan and Stone, 2007) and (Zickler et al., 2009) have shown. This is due to difficulties in dealing effectively with luminance variation and colour calibration.

In our previous work (see (Martins et al., 2006), (Martins et al., 2007)) we developed a model-based computer vision system that does not use colour information for object recognition, only for object distinction. As a result, this system proved to be fast to recognise objects and also robust to luminance variation. It also made the task of calibrating colours easier, since this information is only required when the objects are already detected in the image. However, the calibration of colours was still required.

Moving towards an automatic computer vision system, we present in this paper a novel combination between our model-based computer vision system and a Multi-layer perceptron (MLP) neural network to classify colours as an alternative to the high computational cost of neural networks, enhancing the accuracy of the overall system.

We first examine colour and object recognition issues and define the problem in Section 2. Subsequently, in Section 3 we present a brief review of the relevant characteristics of Artificial Neural Networks for the problem of colour classification and, in Section 4, describe our approach to solve the problem stated. Section 5 has descriptions and analysis of the experiments performed, followed by the paper conclusion, Section 6, where we remark the contributions and present future work.

## 2 COLOUR AND OBJECT RECOGNITION IN ROBOT SOCCER

The robot soccer domain is composed of a group of distinct categories, which are regulated by either FIRA or RoboCup. These categories range from small-sized wheeled mobile and dog-like 4-legged robots to humanoid robots, offering to researchers

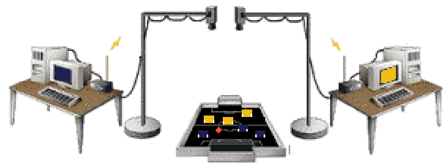

Figure 1: The MiroSot arquitecture proposed by FIRA

two main challenges in computer vision: local and global vision systems.

The former requires robots endowed with such an embedded system (camera and computer vision algorithms) that allows the robot to localise the ball, the other robots (teammates and opponents) and itself. The latter entails a system where the camera is fixed over the game field, providing a global perception of the environment and centralising the processing of the images captured by the camera, as illustrated in Fig. 1.

### *2.1 The RoboFEI global vision system*

We have been developing a global vision system for the RoboFEI robot soccer team, under the rules of the MiroSot, regulated by FIRA. Basically, each team is represented by a colour, either yellow or blue, and the ball is an orange coloured golf ball. A label must be placed on top of each robot with a blob coloured according to the team the robot represents and a minimum area of 15 $cm^2$.

To distinguish between robots in our team, another blob with a different colour (green, cyan or pink) is placed on the label and we deliberatively use a circle as the shape of the blobs. The circles are posed in such a way that the line segment between their centres forms an angle of 45 degrees with the front of the robot (Fig. 2).

In order to recognise the objects (robots and ball), the computer vision system was designed in seven stages, as shown in Fig. 3 (for a detailed description, see (Martins et al., 2006), (Martins et al., 2007)).

Initially, the images are captured from the camera with resolution of either 320x240 or 640x480 pixels and 24-bit RGB colour depth, at 30 frames per second. Then, the background is subtracted so only the objects remain in the image, and the edges are extracted.

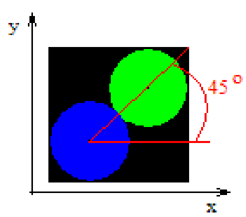


Figure 2: Example of a label used to identify the robots

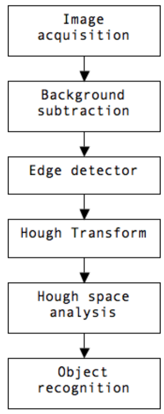


Figure 3: Block diagram of the RoboFEI global vision system

Subsequently, the *Hough Transform*, fitted to detect circles, is computed using only edge information and hence a *Hough space* is generated, which maps cartesian coordinates into probabilities. The higher the probability the higher the certainty that a given point is centre of a real circle in the image. The *Hough space* is thresholded and the points with the highest probabilities are stored to be used afterwards.

Finally, the robots are recognised using colour information, since all the circle centres in the image are already known (the high probability points in the *Hough space*). By using this approach, the influence of luminance variation is considerably minimised and the accuracy required to calibrate the colours is reduced.

However, the colour recognition in the *Object Recognition* stage is based on a set of constant thresholds (minimum and maximum values for *Hue*, *Saturation* and *Value*) for each colour class. The use of constant thresholds implies in two main limitations: (i) the minimum and maximum HSV values can only be set as continuous ranges, while the number of discontinuities is not known, and (ii) the ranges of HSV must be adjusted for each colour in such a way that these ranges are neither so narrow that cannot represent all possible samples of a class, nor too wide that two colour ranges get superposed. Due to these restrictions, the task of calibrating colours is demanding and time consuming, requiring knowledge of theoretical concepts, such as the HSV colour space representation and how each attribute of this colour space influences colour information.

Many researchers have been addressing the colour calibration procedure (e.g., (Gunnarsson et al., 2005), (Sridharan and Stone, 2007)) in different categories, proposing new techniques to solve the problems of manual colour calibration systems. Our approach in this paper (described further in Section 4) makes use of Artificial Neural Networks.

## 3 NEURAL NETWORKS TO RECOGNISE COLOURS

Artificial Neural Networks, particularly Multi Layer Perceptron (MLP) networks, have been applied to a wide range of problems such as pattern recognition and robot control strategies, noticeably for their capability of approximating non-linear functions, by learning from input-output example pairs, and generalisation (see (Mitchell, 1997)). MLP neural networks can learn from complex and noisy training sets without the need to explicitly define the function to be approximated, which is usually multi-variable, non-linear or unknown. In addition, they also have the capability to generalise, retrieving correct outputs when input data not used in the training set is presented. This generalisation allows the use of a small number of samples during the training stage.

Specifically in the problem of colour classification, the representation of the colour classes in the HSV domain (as stated in Section 2.1) can be spread along a variable, and unknown a priori, number of discontinuous portions of the space, which a manually-defined thresholding system cannot represent unless it has one set of thresholds for each of these discontinuous portions. The creation of multiple thresholds also considerably increase the expertise and time required to calibrate the colours, if the task is to be carried manually by the operator. On the other hand, this representation can be quickly learnt by neural networks from the training set, without the need of operator's expertise, making them a much more suitable solution for the task.

However, MLP neural networks have a relatively high computational cost, due to the fact that each new input sample requires a new calculation of the neuron matrices, which cannot be computed a priori, thus making MLPs to be commonly deemed unsuitable for real time applications. Next section details a way to overcome this inconvenience and have neural networks applied to a real time application.

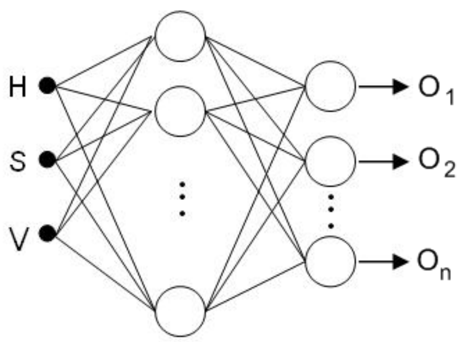


Figure 4: The MLP neural network implemented to recognise colours

## 4 THE IMPLEMENTED SYSTEM

In order to be fully compatible with the model-based vision system used in this paper, the MLP neural network algorithm was implemented in C/C++ language, and the machine learning portion of the well-known Intel OpenCV library due to its very good performance.

The MLP neural network is composed of 3 layers of neurons, activated by sigmoid functions. Because the input is a 3 dimensional vector, the HSV colour space, 3 neurons form the input layer, while the output layer has 7 neurons, six outputs representing a given colour of interest in which a pixel must be classified (orange, yellow, blue, green, pink and cyan) and an additional output representing the background. Regarding the hidden layer, after some preliminary investigations we found the best accuracy using 20 neurons. Therefore, the topology of the MLP neural network was set as 3-20-7. To train the network, the backpropagation with momentum algorithm was used, and the parameters adjusted to learning rate of 0.1, momentum of 0.1 and stop criteria (also called performance) $\epsilon$ of $10^{-6}$. This tight value for $\epsilon$ serves to really ensure the training set was learnt correctly and the scalar value was defined during the test phase, looking at different values of $\epsilon$ that could still produce high fidelity results.

Moving towards an automatic global vision system, we developed a solution to the colour calibration problem, inspired by the work developed by (Simoes and Costa, 2000). In order to bypass the high computational cost constraint, we make use of our model-based vision system to detect, through a Hough transform, the points with high probability of being the actual centres of circles that form the real objects in the image. Once the circles are found, we apply the MLP neural network (shown in Fig. 4) to classify the pixels of a 5x5 mask inside this area.

In contrast to our previous method to calibrate colours described in Section 2.1, the task of training the MLP neural network is straightforward to anyone who is able to distinguish colours, since no specific knowledge is needed and the only requirement is to select samples of a given colour of interest and define to which output that samples should be assigned.

This yields a more efficient use of the MLP neural network evaluation function, since only a small number of pixels need to be evaluated, rather than evaluating the whole image. For instance, in an image with resolution of 320x240 pixels (the lowest resolution we use), the neural network would need to classify 76800 points, while with the combination of our model-based system the number of points to be classified is reduced to 1250 (about 50 points with probability higher than a fixed threshold - for more details, see (Martins et al., 2006), (Martins et al., 2007)), because only a 5x5 pixel mask of each detected circle needs to be evaluated. This reduction, as shown by the results further presented, is enough to allow the execution of the network in real time.

## 5 EXPERIMENTS

In order to validate our proposal and analyse the performance of the MLP neural network implemented, we performed experiments to measure accuracy and execution time of the algorithm, as well as a complete real-world scenario test. Below we describe the MLP neural network training stage and the three experiments performed.

In the first experiment, the MLP neural network was tested detached from the model-based vision system, aiming at measuring its accuracy. Subsequently, the second experiment tested our combined solution to object recognition, focusing on the accuracy and execution time of the complete vision system when compared with the manual calibration of the colour ranges. The third test is a real-world test, with measurements of accuracy taken directly from a game scenario.

In all cases, the tests were performed on a computer equipped with an Intel Pentium D 3.4GHz processor running Microsoft Windows XP operating system.

### *5.1 The MLP neural network training stage*

In our approach, the training stage happens before a game, when a human supervisor selects the training points in the image and defines to which output (either a colour or background) these points should be classified. In this scenario, we take advantage of capturing the input samples in a live video streaming, what takes into account the noise. As points are captured from different frames, they intrinsically carry the noise to the training set, improving the robustness of the neural network. Once enough training points are captured, the human triggers the network's training function.

### *5.2 Neural Network accuracy test*

In this test, we trained the neural network using 140 training points, 70 from an image with luminance of 750 lux and other 70 from another image with luminance of 400 lux, with the training points equally distributed among the 7 output classes. The network was then executed against 10x76800 pixels (10 images with resolution of 320x240 pixels and luminance varying from 400 to 1000 lux). The result was an average of 83% correctly classified pixels. The source image and resultant representation of the classified pixels are shown in Fig. 5. From the image is noticeable that most of the incorrectly classified pixels belong to the

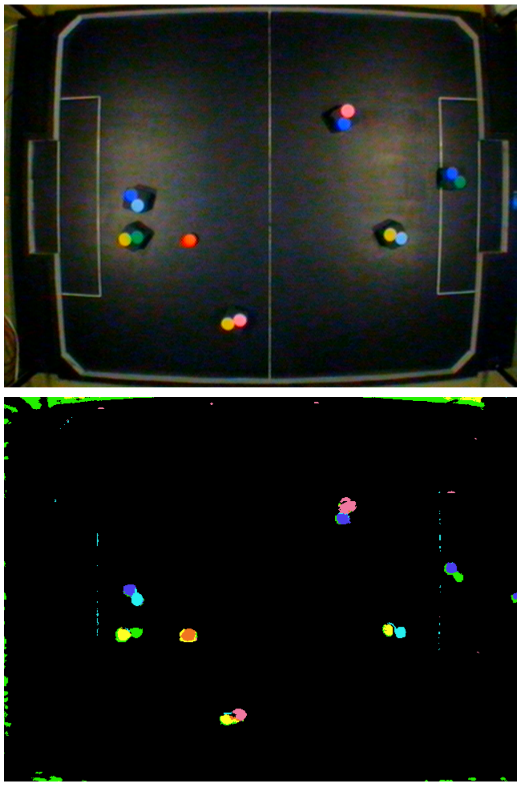

Figure 5: (top) Source image from noisy camera. (bottom) Resultant image, generated by the Neural network classification

background class, what does not pose a problem for the algorithm, since the colour information is only retrieved from points belonging to the foreground image, after background subtraction. This experiment proves that our MLP neural network is robust to luminance variation and the accuracy is high enough to be used on our approach in the robot soccer domain. However, this test is a proof of concept, aimed at validating the algorithm. The practical application of the proposed algorithm is presented in the following experiments.

### *5.3 Manual thresholding x Neural approaches comparison*

In this test, we used 30 images of the game field with the objects (6 robots and a ball) randomly positioned on the field, captured at the resolution of 640x480 pixels.

Three people (first two with basic knowledge and one with advanced skills in regards to our computer vision system) were asked to calibrate the manual colour recognition system. In addition, each one was asked to capture 140 training points (20 for each output class) to train the MLP neural network, always using the same image. A fourth person then presented the 30 images to the computer vision system and obtained the results shown in Table 1 for both algorithms. During this experiment, the averaged execution times of the complete vision system were also measured, within a loop of 500 executions for each test image, and the results are shown in Table 2.

While both algorithms produced good classification levels, the neural network showed not to be easily susceptible to performance drops caused by different trainer experiences, expressed by its much smaller deviation on the experiments, and to match and exceed the thresholding algorithm performance. In terms of execution time, we conclude that no significant time is added with the introduction of the neural network algorithm.

Table 1: Accuracy (%)

| | *Thresholding* | *Neural Network* |
|---|---|---|
| 1st trainer | $90 \pm 14.2$ | $97.5 \pm 2.3$ |
| 2nd trainer | $76 \pm 29.0$ | $97.5 \pm 3.5$ |
| 3rd trainer | $96 \pm 21.0$ | $96 \pm 2.3$ |

Table 2: Execution times ($ms$)

| | *Image Size* | |
|---|---|---|
| *Algorithm* | 320x240 | 640x480 |
| Neural Network | $7.92 \pm 1.15$ | $27.30 \pm 1.76$ |
| Thresholding | $6.10 \pm 0.01$ | $25.55 \pm 0.28$ |

### *5.4 Real Game Scenario*

At last, we tested and compared the algorithm in a real game scenario, with luminance variations and using real time images from the camera. This is a very similar complete system validation test we performed in our previous work (Martins et al., 2007), with the difference that, this time, we used a camera with noisy signal. The noisy colour signal is a very common noise found specially in analog cameras, caused by imperfect or old cabling connections, electric wiring close to video signal cables and aging cameras. Under luminance variation and noise conditions, we sought to stress the capabilities of the neural algorithm to generalise, correctly classifying pixels of different luminance levels, even when trained and executed with noisy samples.

To do so, the neural algorithm was calibrated with 70 samples from the live camera image, 35 of these obtained at a luminance level of 1000 lux and 35 obtained at 600 lux. As the thresholding algorithm would need to be used as comparison, it was calibrated as well, with some static images taken at different luminance levels, as the paragraph of result analysis will explain. Then, the robots were set to randomly run across the field and a counter summed the number of times each robot was correctly identified, in a cycle of 4000 frames, in a process that was repeated for each

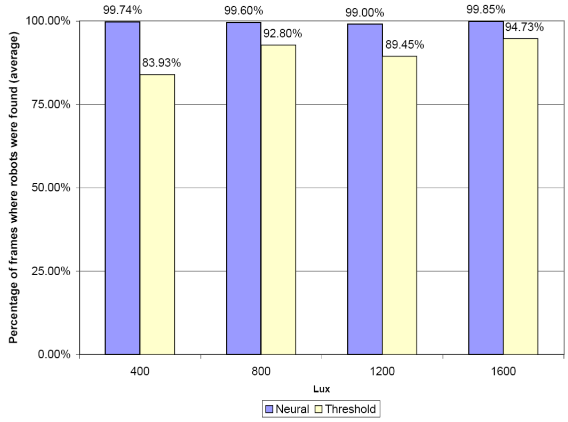


Figure 6: Percentage of frames where the robots were found, averaged among 6 robots of different colour combinations, by each algorithm.

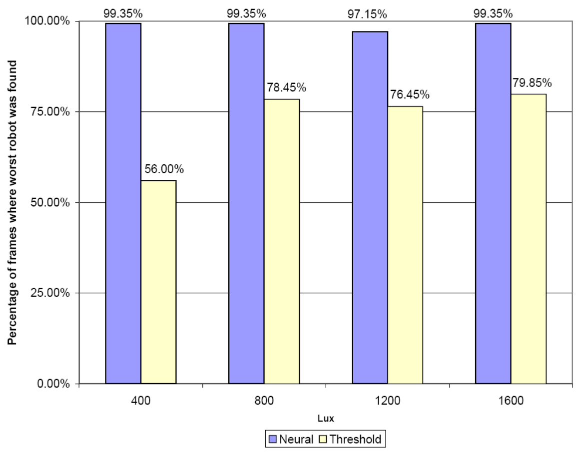


Figure 7: Percentage of frames where the robot with smaller detection count was found by each algorithm.

different luminance level. Fig. 6 shows the averaged detection count among the 6 robots, comparing the neural and threshold algorithms. Fig. 7 shows a worst case detection count, of the robot detected the lesser number of times during one cycle of 4000 frames. The game test results show that the neural algorithm is extremely robust to luminance variation, as well as to noise. On the contrary, the thresholding system is highly vulnerable to noise. It is important to note also that, to obtain a calibration for the thresholding algorithm that could be used in the test, more than 45 minutes where spent in trials, for even when the results in the test images were satisfactory, after the robots started to move the results were showed to be poor. The neural network calibration did not take more than 3 minutes and was executed only one time.

## 6 CONCLUSIONS AND FUTURE WORK

The results of the different experiments confirm that our approach, combining object detection and neural network colour classification, considerably increases the accuracy of the vision system under real-world conditions such as noise and light intensity variation while, at same time, creates an easier to operate method for colour recognition. We successfully removed the time consuming and demanding task of manually adjusting thresholds for colour calibration and replaced it by a more robust and efficient classifier, the neural network, but avoided its high computational cost with the use of the circle detection algorithm based in Canny and Hough filters. The limitation of the system is its inability to detect objects with unknown geometric form, which we intend to solve in future work. We would like also, as a future work, to explore the use of non supervised classifiers and possibly RBF (Radial Basis Function) neural networks, aiming to produce a fully automatic calibration system.

## 7 ACKNOWLEDGMENTS

Murilo Fernandes Martins would like to thank the support of CAPES through the grant 5031/06-0.

The authors would like to thank the members of the RoboFEI team, for their valuable contributions and time spared.

### References

Bruce, J., Balch, T. and Veloso, M. (2000). Fast and inexpensive color image segmentation for interactive robots, *Proceedings of IROS-2000*, Japan.

Comaniciu, D. and Meer, P. (1997). Robust analysis of feature spaces: color image segmentation, *CVPR '97: Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97)*, IEEE Computer Society, Washington, DC, USA.

Gunnarsson, K., Wiesel, F. and Rojas, R. (2005). The color and the shape: Automatic online color calibration for autonomous robots, *in* A. Bredenfeld, A. Jacoff, I. Noda and Y. Takahashi (eds), *RoboCup*, Vol. 4020 of *Lecture Notes in Computer Science*, Springer, pp. 347–358.

Martins, M. F., Tonidandel, F. and Bianchi, R. A. C. (2006). A fast model-based vision system for a robot soccer team, *in* A. F. Gelbukh and C. A. R. García (eds), *MICAI*, Vol. 4293 of *Lecture Notes in Computer Science*, Springer, pp. 704–714.

Martins, M. F., Tonidandel, F. and Bianchi, R. A. C. (2007). Towards model-based vision systems for robot soccer teams, *in* P. Lima (ed.), *Robotic Soccer*, I-Tech Education and Publishing, chapter 5, pp. 95–108.

Mitchell, T. (1997). *Machine Learning*, McGraw-Hill Education (ISE Editions).

Simoes, A. S. and Costa, A. H. R. (2000). Using neural color classification in robotic soccer domain, *International Joint Conference IBERAMIA'00 Ibero-American Conference on Artificial Intelligence and SBIA'00 Brazilian Symposium on Artificial Intelligence* pp. 208–213.

Sridharan, M. and Stone, P. (2007). Color learning on a mobile robot: Towards full autonomy under changing illumination, *in* M. M. Veloso (ed.), *IJCAI*, pp. 2212–2217.

Tao, W., Jin, H. and Zhang, Y. (2007). Color image segmentation based on mean shift and normalized cuts, *Systems, Man, and Cybernetics, Part B, IEEE Transactions on* **37**(5): 1382–1389.

Zickler, S., Laue, T., Birbach, O., Wongphati, M. and Veloso, M. (2009). Ssl-vision: The shared vision system for the robocup small-size league, *Proceedings of the International RoboCup Symposium 2009 (RoboCup 2009).*