Using Saliency-based Visual Attention Methods for Achieving Illumination Invariance in Robot Soccer

F. Serhan Daniş, Tekin Meriçli, and H. Levent Akın

Department of Computer Engineering
Boğaziçi University
Istanbul, Turkey
{serhan.danis,tekin.mericli,akin}@boun.edu.tr

Abstract. In order to be able to beat the world champion human soccer team in the year 2050, soccer playing robots will need to have very robust vision systems that can cope with drastic changes in illumination conditions. However, the current vision systems are still brittle and they require exhaustive and repeated color calibration procedures to perform acceptably well. In this paper, we investigate the suitability of biologically inspired saliency-based visual attention models for developing robust vision systems for soccer playing robots while focusing on the illumination invariance aspect of the solution. The experiment results demonstrate successful and accurate detection of the ball even when the illumination conditions change continuously and dramatically.

1 Introduction

As the deadline for achieving the ultimate goal of RoboCup approaches, where a team of autonomous humanoid robots is expected to play soccer on a standard soccer field against the most recent winner of the World Cup and win the game, robustness to changing visual circumstances remains one of the biggest challenges for developing reliable vision systems for autonomous robots. Using color segmentation as the basis of the developed vision systems still appears to be the most popular approach among the teams of various RoboCup leagues although most of the color segmentation based techniques are not robust against changing illumination conditions. Since the successful operation of the robots primarily depends on the reliability of the vision system, the teams spend a considerable amount of time for color calibration even though the games are still played under carefully controlled illumination conditions.

Biological systems, on the other hand, are very successful in solving such problems; therefore, they have long been the primary source of inspiration for robot vision researchers. The visual attention mechanism is one of the most studied sub-systems of biological vision. Following the visual attention phenomena, researchers aim to obtain more efficient, intelligent, and robust artificial vision systems [1]. In alignment with this goal, in this paper, we present the results of a primary investigation on the suitability of saliency-based visual attention models for the robot soccer domain, focusing on the object detection performance under continuously and drastically changing illumination conditions. Our experiments demonstrate successful detection of the ball even under extreme changes in the illumination conditions where color segmentation based approaches fail to do so.

The rest of this paper is organized as follows. Section 2 gives an overview of the related work in the literature. The methodology followed for this contribution is explained in Section 3, and the details of the experiments and the obtained results are given in Section 4. Section 5 summarizes and concludes the paper while pointing out to potential future work.

2 Related Work

One of the biggest challenges for autonomous mobile robots that perceive their environments through standard cameras is the changing illumination conditions. As a workaround, either restricted configurations in structured domains are considered, or specific models of segmentation and recognition that do not provide generalized solutions [2] are used. Various approaches to address this challenge include describing the problem in terms of illumination [3], surface reflectance [4], and sensor sensitivity [5]. Bayesian decision theory and hierarchical model based approaches also exist in the literature [6, 7]. Methods that do not require domain specific tuning are shown to be computationally less complex and more adaptive, whereas usually the opposite is shown to be true for the classical and model based methods [2]. Illumination invariance has been studied in the robot soccer domain as well since the overall performance of the teams heavily depend on the successful perception of the environment [7, 8].

Visual attention-based approaches are usually used for preprocessing the visual sensory data to determine the parts of the image to further process. For instance, in the work of Rasolzadeh *et al* [9], the visual attention module gets executed prior to the object detection and recognition modules to direct the head saccades and help the robot figure out where to search for important objects. Frintrop *et al* proposed a similar approach [10, 11], where the regions of interest are detected using both bottom up and top down saliency extraction followed by a fast-classifier that classifies regions for object recognition purposes. They applied this method to the problem of detecting balls in a robot soccer environment, and showed that this approach yields to a faster execution compared to a standard classifier and reduces the false detection rates significantly; however, they did not investigate the problem of changing lighting conditions thoroughly.

In a similar setup that we present in this paper, Garcia *et al* [12] used an attention mechanism to detect balls in the “any ball challenge” scenario of the RoboCup Standard Platform League (SPL), where balls of various sizes, textures, and colors are scattered over the field and the robot is expected to detect them and score by kicking them into the opponent goal. They used the saliency map to extract balls on the field as the regions of the images containing the balls popped out as the “salient regions” compared to the plain green field carpet. Their work differs from the original method of saliency map generation by Itti *et al* [13] in two aspects. First, they use only the color and intensity information and discard the other channels such as motion, orientation, and flicker. Second, the sizes of the images are reduced using the fovea mask for computational efficiency purposes. They present a model performance improvement and a neat integration of the method to the SPL domain; however, they do not consider the changing illumination conditions as we investigate in detail in this paper.

3 Methodology

In this section, we present the working principles of both the saliency-based visual attention and color segmentation and scanline based object detection methods as the comparison of these two approaches constitute the main motivation behind this work.

3.1 Saliency-based visual attention

Although primates have neuronal hardware with limited speed, they are capable of interpreting complex scenes in real time. Such capability is believed to be achieved by the selection of a subset of available visual information by higher visual areas before further processing [14]. Inspired by the remarkable scene interpretation ability of primates and building on a biologically plausible architecture presented by Koch and Ullman [15], which explains human visual search strategies [16], Itti *et al* proposed a model of attention that is based on the same working principles [13]. In this computational model of saliency-based visual attention, the visual input (i.e. the image) is decomposed into feature maps; primarily color, intensity, and orientation, which compete for the final saliency map. During the saliency map generation process, the input image is first used to generate color, intensity, and orientation layers. These different layers are used to generate multi-scale Gaussian pyramids, which correspond to progressive low-pass filtering and sub-sampling into lower resolutions. The feature maps are then obtained by a set of linear center-surround operations and across-scale combinations of these multi-scale pyramids. Finally, these feature maps are linearly combined to generate the final saliency map [17–19]. This model is depicted in Fig. 1.

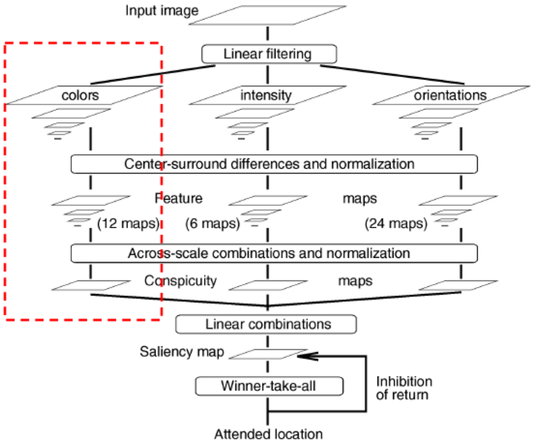


Fig. 1. The attention model proposed by Itti *et al* [13].

Based on the attention model proposed by Itti *et al* [13], in this work, we use the color conspicuity map (marked with the red frame in Fig. 1) to investigate whether it is possible to detect important objects in the image even when the illumination conditions of the environment fluctuate drastically. We also employ and test the performance of a slightly modified version of this model, which uses wavelet low-pass pyramids instead of Gaussian ones [20]. The steps of the saliency map generation process are explained in detail in the following sections.

Generation of color conspicuity maps The first step is to obtain the intensity image I by averaging the R, G, and B color channels, which we denote as matrices of the same size as the original image. Pixel values that are smaller than $1/10$ of the maximum intensity are set to zero before further processing as it gets difficult to perceive the color information of a pixel when its intensity value is very small. New color components, RN, GN, BN, and YN are then computed in terms of R, G, and B as follows.

$$RN \quad = R - (G+B)/2 \tag{1}$$

$$GB \quad = G - (R+B)/2 \tag{2}$$

$$BN \quad = B - (R+G)/2 \tag{3}$$

$$YN = G + R - |R-G| - B \tag{4}$$

Negative values of these color components are set to zero and the component pyramids are constructed as RN_k, GN_k, BN_k and YN_k by either using the Gaussian low pass filter and progressively down-scaling the image into its half size [13] or using the wavelet low pass filter [20]. The subscript k denotes the level of the pyramid, where level 0 is the top level color component of the size of the original image.

Obtaining the feature maps Feature maps are computed using a method that is inspired by the working principles of the "color double-opponent cells" that were proven to exist in the human primary visual cortex. These neurons are excited by one color in the center of their receptive fields, and inhibited by another, while the opposite is true in the surround. Human primary visual cortex is shown to have such spatial and chromatic opponency for the red/green, green/red, blue/yellow, and yellow/blue color pairs [21]. Analogously, we compute the center surround differences to obtain the feature maps, where $c \in \{2,3,4\}$ are centers and $s = c + p$ are their surrounds with $p \in \{3,4\}$, and * denotes the pyramid levels that are resized to a finer resolution, which in this work is determined by the finest available resolution in the data.

$$RG_{c,s} = |(RN_c - GN_c) - (RN_s^* - GN_s^*)| \tag{5}$$

$$BY_{c,s} = |(BN_c - YN_c) - (BN_s^* - YN_s^*)| \tag{6}$$

While the across-scale combination step of the conspicuity map generation process in the model of Itti *et al* [13] is performed by integrating all color feature maps at different scales, the color feature maps in the work of Li *et al* [20] are first resized to the size of the original image and then squared, resulting in few redundant salient areas. At each step of the algorithm, normalizations should also be applied on the feature maps in order to eliminate modality-dependent amplitude differences. These two proposed methods are identical aside from the functions used for the generation of the color component pyramids, namely wavelet transform and Gaussian filter, and the additional square operation used in the method proposed by Li *et al*.

Merging into a saliency map In our experiments, we used five different saliency maps; three of them are generated using the method proposed by Itti *et al* [13] (*M1)*, and the other two are obtained using the method proposed by Li *et al* [20] (*M2)*.

Table 1. Listing of saliency maps.

Map	Description
M1-a	Color conspicuity map from (*M1)*
M1-b	Combination of the color and intensity conspicuity maps
M1-c	Only the red-green channel of the color conspicuity map
M2-a	Color conspicuity map from (*M2)*
M2-b	Squared *M2-a*

The first saliency map corresponds to the color conspicuity map from *M1.* The second one is the saliency map generated by equally combining the color conspicuity map with the intensity map. Although Garcia *et al* [12] used the two conspicuity maps to generate the final saliency map for detecting the balls on the field, we anticipated that color feature channels would give better results for objects of specific colors. On the basis of this anticipation, we used a third map that is obtained by only utilizing the red-green (*RG*) channel, which made sense considering that our task is finding a red-orange ball on a green field. The fourth map is obtained directly through *M2,* and the fifth one is generated by taking the squares of the pixel values of the fourth map, with the expectation of reducing the number of redundant salient areas as reported by Li *et al* [20]. Table 1 summarizes the compounds of the generated maps.

3.2 Color segmentation and scanline based object detection

Considering computational efficiency and real-time constraints of the robot soccer domain, it is not feasible to process each pixel of the image to find the objects of interests. Therefore, scanlines are used to process the image in a sparse manner, hence speeding up the entire process. This method is especially popular within the RoboCup SPL community as a commercially available standard robot platform with very limited computational resources need to be used for the competitions. A previously trained and stored color table (CT) is utilized for checking the colors of the pixels that a scanline runs through. We utilize a Generalized Regression Neural Network (GRNN) [22] for mapping the real color space to the pseudo-color space composed of a smaller set

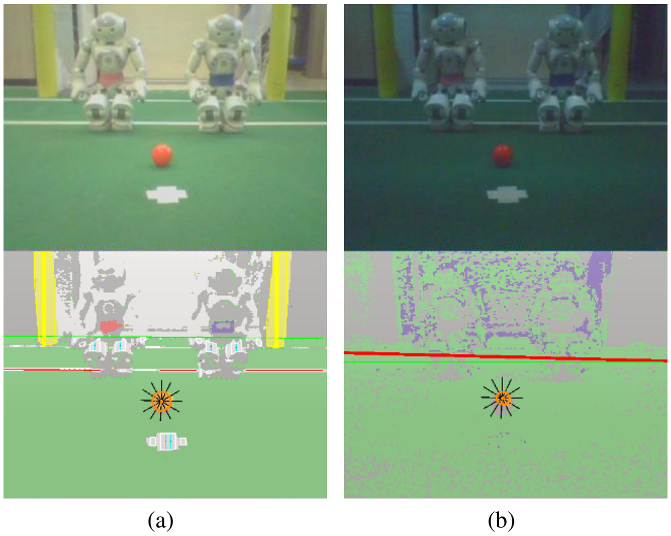


Fig. 2. Color segmentation and scanline based object detection. The classified image looks very clear and the important objects in the image are successfully detected when a suitable color table is used (a); however, this approach becomes unreliable when the lighting conditions change (b).

of pseudo-colors, namely, white, green, yellow, blue, robot-blue, orange, red, and "ignore". In order to obtain the outputs of the trained GRNN in a time-efficient manner, a look up table is constructed for all possible inputs. Y, U, and V values are used to calculate the unique index and the value at that index gives the color group ID to determine the color group of a pixel.

Regions of interest are formed by grouping the same-colored scanline segments that are spatially adjacent and "touching" each other; that is, these segments are on two consecutive scanlines and either of them has a start or end point within the borders of the other one. These regions are then passed to the so called the *region analyzer* module to be further filtered and processed for the detection of the ball, the field lines and intersections of them, the goal posts, and the robots. The ball detector uses additional star-shaped scanlines that originate from the centers of the candidate regions to find the borders of the region and use these border points to check whether the region has circular properties by using a voting-based circle fitting algorithm. Fig. 2(a) shows the result of this process when the used color table matches the lighting conditions; however, this approach may fail when the lighting conditions change as shown in Fig. 2(b), which constituted the main motivation behind this research. The objects are either not detected at all (e.g. lines and goal posts), or the detection result is misleading (e.g. smaller-than-actual size of the ball, which results in a farther-than-actual projected ball location on the field).

4 Experiments

For our experiments, we utilize the iLab Neuromorphic Vision C++ Toolkit (iNVT) software developed and released by Itti *et al* [23]. The experiments are run on the grayscale saliency maps extracted via the methods mentioned in Section 3 as well as the raw images processed by our color classification and scanline based vision module that utilizes previously trained color tables, which is explained in Section 3.2. We particularly focus on the detection of the ball in the images.

4.1 The robot platform

We performed our initial experiments offline on the images captured from one of the cameras of the *Nao V3* humanoid robot manufactured by Aldebaran Robotics [24], which has been used as the common robot platform of the Standard Platform League (SPL) of RoboCup [25] since 2008. The *Nao*'s camera is capable of providing images with 640×480 resolution at 30Hz; however, most teams prefer using 320×240 images due to the processing power limitations of the robot. Our team also utilizes the images provided in 320×240 resolution for the competitions; therefore, in order to be able to compare the performances of the saliency-based methods with the performance of the color segmentation based method, we kept the image resolutions identical for the two methods in our experiments.

4.2 Illumination configurations

In our experiments, we use 6 different illumination configurations controlled by 3 factors. We denote these factors as *fluo* for the fluorescent lamps, *spot* for the spot lights, and *day* for the daylight coming in through the windows. The combinations of these configurations are labeled as $\{C_1, ..., C_6\}$. Table 2 summarizes these configurations. Even though there are 8 possible combinations of these 3 factors, in our experiments, we exclude the $\langle \neg fluo, spot, \neg dayl \rangle$ and $\langle \neg fluo, spot, dayl \rangle$ configurations as the presence of the spot lights provides the majority of the lux value, which is covered by the cases C_5 and C_6.

Table 2. Listing of the illumination configuration used in our experiments.

Configuration	Tuple	Description	Illuminance
C_1	$\langle \neg fluo, \neg spot, \neg dayl \rangle$	no lights	39 lux
C_2	$\langle \neg fluo, \neg spot, dayl \rangle$	only daylight	134 lux
C_3	$\langle fluo, \neg spot, \neg dayl \rangle$	only fluorescent lights	200 lux
C_4	$\langle fluo, \neg spot, dayl \rangle$	fluorescent lights and daylight	350 lux
C_5	$\langle fluo, spot, \neg dayl \rangle$	fluorescent and spot lights	908 lux
C_6	$\langle fluo, spot, dayl \rangle$	all lights	1067 lux

The scene we set up for our experiments includes the essential visual elements of the SPL; which are one red and one blue player placed in front of the yellow goal, the field lines and the cross-shaped penalty mark, and the ball placed between the robots and the penalty mark. During our experiments, we keep the robot stationary while the contributions of the different light sources to the environment's illumination characteristics are changed to generate different lighting conditions. The sample color images from these configurations are given in Fig. 3 with the corresponding average lux values of the environment.

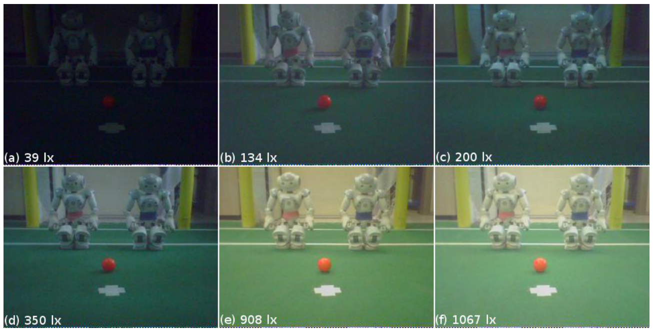
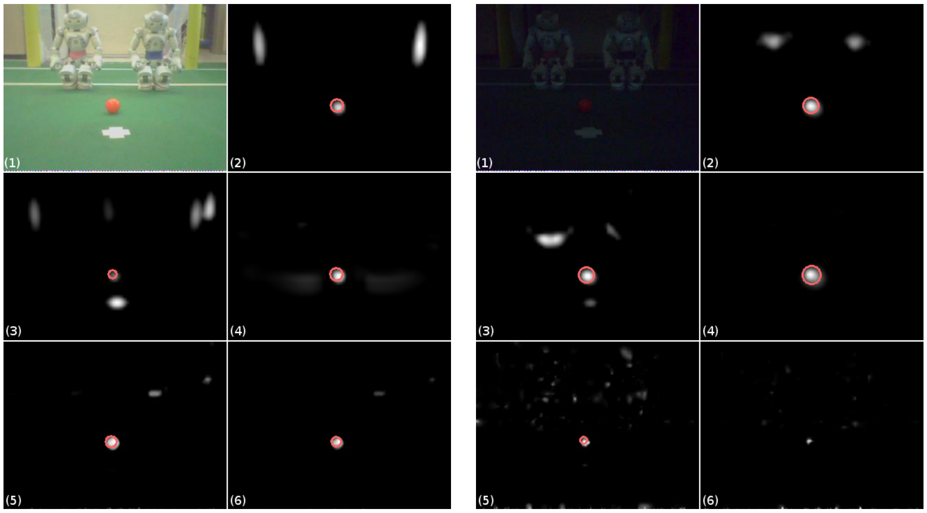


Fig. 3. Scenes captured by the robot under different illumination configurations: (a) C_1 - no lights, (b) C_2 - only daylight, (c) C_3 - only fluorescent lights, (d) C_4 - fluorescent lights and daylight, (e) C_5 - fluorescent and spot lights, and (f) C_6 - all lights.

4.3 Results

We used the five different saliency map generation methods listed in Table 1 in our saliency-based experiments. Sample saliency maps for configurations C_1 and C_6 are shown in Fig. 4. Ball detection is performed by trying to fit a circle to the salient regions after performing a simple thresholding on them. The circle fit operation is considered successful if the error value is lower than 10 pixels. Detected balls on the saliency maps are also shown as orange circles in Fig. 4. Additionally, an error analysis is performed by utilizing human perception as the source of the ground truth information and reporting the difference between the ground truth and the output of the detection algorithm. For each image, we report a *hit* when the actual ball is found accurately (center and radius errors are below 8 pixels), a *false location* when some other regions is confused for the ball, and a *miss* when the ball is not detected at all. The left column and the right column of Fig. 5 show the *hit*, *false location*, and *miss* rates obtained after running the color segmentation based and saliency based algorithms, respectively, on 80 frames captured under each of the six different illumination configurations.

It can be interpreted from Fig. 5(a), 5(c), and 5(e) that the color segmentation based methods work quite well when the color table used matches the illumination condition; however, usually even small changes in the illumination characteristics results in failure, as also shown in Fig. 2. Although it is possible to prepare several color tables for various illumination conditions and switch between them based on some image statistics, this method becomes ineffective for continuously changing illumination conditions.



(a) Configuration $C_1 = \langle fluo, spot, dayl \rangle$: (1) original frame, saliency methods (2) I-a, (3) I-b, (4) I-c, (5) II-a, and (6) II-b.

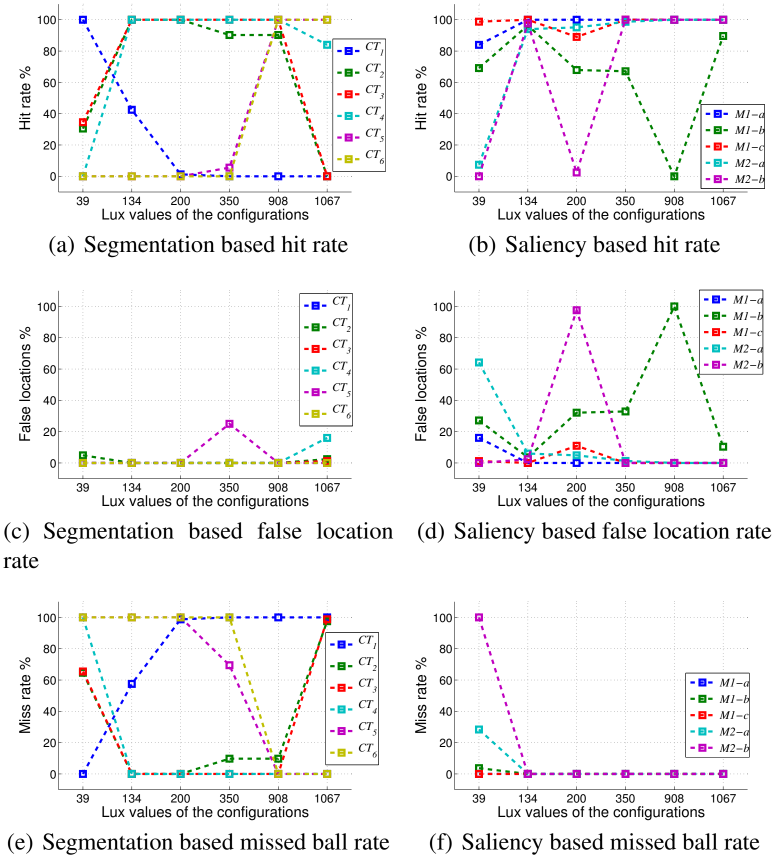
(b) Configuration $C_6 = \langle \neg flue, \neg spot, \neg dayl \rangle$: (1) original frame, saliency methods (2) I-a, (3) I-b, (4) I-c, (5) II-a, and (6) II-b.

Fig. 4. Sample saliency maps.

In Fig. 5(b), 5(d), and 5(f), we see that using only the color channels yields better results in general. Using the intensity map combined with the color map, which corresponds to *M1-b*, results in low hit rates and high false location rates. The results obtained with *M2-a* and *M2-b* show that *M2* works well for object recognition in bright environments; however, it performs poorly when there is not enough light. The highest accuracy in finding the ball is achieved with *M1-c*.

Fig. 6 shows the mean errors between the radius reported by the color segmentation based and the saliency based methods and the radius marked by a human as the ground truth. Considering that Fig. 6(a) shows the consistency of only the rare occasions that the ball is detected, saliency based methods tend to be more consistent whereas the color segmentation based method reports inconsistent results especially for the lighting configurations that the used color table is not trained for. The most consistent results seem to be achieved with *M1-c*. In addition to the reported radius consistency analysis, we also performed a consistency analysis for the reported center of the ball, the results of which can be seen in Fig. 7. The color segmentation based method reports a consistent center for the detected ball when a ball is found in the image; however, the biggest problem with this method is that it usually cannot find the ball at all when there is a mismatch between the color table and the illumination configuration (Fig. 5(e)).

Table 3 summarizes the success rates obtained when the most successful saliency based methods *M1-a* and *M1-c*, and the color segmentation based method running with the available color tables are applied on a dataset of 618 frames collected under continuously changing illumination conditions. Perfect and near perfect hit rates are achieved with *M1-a* and *M1-c*, respectively.



(a) Segmentation based hit rate

(b) Saliency based hit rate

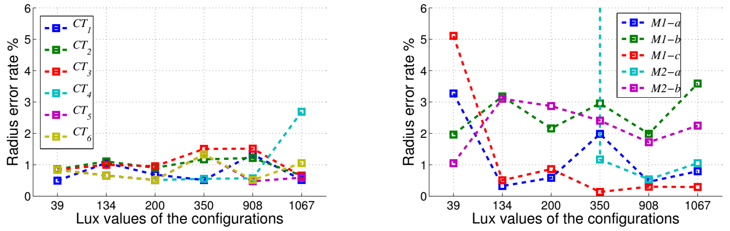
(c) Segmentation based false location rate

(d) Saliency based false location rate

(e) Segmentation based missed ball rate

(f) Saliency based missed ball rate

Fig. 5. Performances of the color segmentation based (left) and saliency based methods (right).



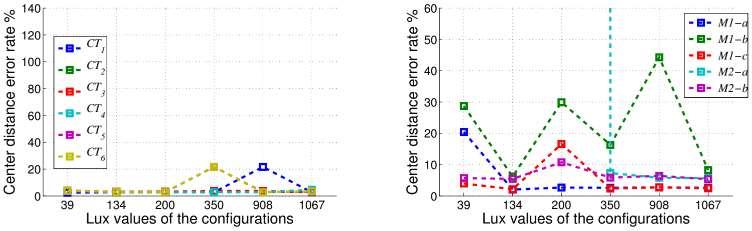
(a) Color segmentation based

(b) Saliency based

Fig. 6. Radial consistency check for all methods.

5 Conclusions and Future Work

Illumination independent robust visual perception of the environment has been one of the biggest challenges for computer and robot vision researchers. Being capable of



(a) Color segmentation based

(b) Saliency based

Fig. 7. Central consistency check for all methods.

Table 3. Comparison of the ball detection performances (%) of the individual color tables (CT_i) and *M1-c* on a dataset collected under continuously changing illumination conditions.

	CT_1	CT_2	CT_3	CT_4	CT_5	CT_6	*M1-a*	*M1-c*
hit rate	13.13	10.37	89.14	90.76	8.91	2.59	**100**	**99.68**
false locations	18.48	0.49	0.16	9.24	1.94	0	0	0.32
miss rate	68.39	89.14	10.69	0	89.14	97.40	0	0

solving this problem almost effortlessly, biological systems have been a great source of inspiration for the proposed solutions thus far. In this paper, we make use of one such biologically inspired saliency based method with some modifications to investigate its suitability for illumination independent object detection in robot soccer domain. Our experiments demonstrate successful and consistent detection of the ball even when the lighting conditions of the environment change drastically, while the standard color classification based methods fail in such cases. Even though the experiments were performed on an off-board computer as the processor of the available *Nao* robot platform cannot meet the real-time requirements when executing the saliency based method, this approach can still be applicable in other leagues of RoboCup, such as the Middle Size League, where robots equipped with more powerful computational resources are used. Potential future work includes the development of a computationally efficient version of this method for achieving real-time on-board computations, a complete object detection framework with additional sanity checks and filters, and testing of the method on a moving robot in a regular robot soccer game.

Acknowledgments: This project was supported by Boğaziçi University Research Fund project 09M105.

References

1. S. Frintrop, E. Rome, and H. I. Christensen. Computational Visual Attention Systems and Their Cognitive Foundations : A Survey. *ACM Transactions on Applied Perception*, 7(1):1–39, 2010.
2. M. Sridharan and P. Stone. Color learning and illumination invariance on mobile robots: A survey. *Robotics and Autonomous Systems*, 57(6-7):629–644, June 2009.
3. D. A. Forsyth. A novel algorithm for color constancy. *International Journal of Computer Vision*, 5(1):5–35, 1990.

4. G. J. Klinker, S. A. Shafer, and T. Kanade. A physical approach to color image understanding. *International Journal of Computer Vision*, 4:7–38, 1990.
5. G. D. Finlayson, S. D. Hordley, and P. M. Hubel. Color by correlation: A simple, unifying framework for color constancy. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(11):12091221, 2001.
6. D. H. Brainard and W. T. Freeman. Bayesian color constancy. *Journal of the Optical Society of America. A, Optics, image science, and vision*, 14(7):1393411, July 1997.
7. D. Schulz and D. Fox. Bayesian color estimation for adaptive vision-based robot localization. In *IROS*, 2004.
8. X. Luan, W. Qi, D. Song, M. Chen, T. Zhu, and L. Wang. Illumination invariant color model for object recognition in robot soccer. In *ICSI (2)'10*, pages 680–687, 2010.
9. B. Rasolzadeh, M. Björkmann, K. Huebner, and D. Kragic. An Active Vision System for Detecting, Fixating and Manipulating Objects in the Real World. *The International Journal of Robotics Research*, 29(2-3):133–154, August 2009.
10. S. Frintrop. *VOCUS: A Visual Attention System for Object Detection and Goal-Directed Search*, volume 3899 of *Lecture Notes in Computer Science*. Springer, 2006.
11. S. Frintrop, A. Nüchter, K. Pervölz, H. Surmann, S. Mitri, and J. Hertzberg. Attentive Classification. *International Journal of Applied Artificial Intelligence in Engineering Systems*, 1(1), 2009.
12. J. F. Garcia, F. J. Rodríguez, V. Matellán, and C. Fernández. Saliency map based attention control for the RoboCup SPL. In *Workshop of Physical Agents*, 2010.
13. L. Itti, C. Koch, and E. Niebur. A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259, Nov 1998.
14. J.K. Tsotsos, S.M. Culhane, W.Y. Kei Wai, Yuzhong Lai, Neal Davis, and Fernando Nuflo. Modeling visual attention via selective tuning. *Artificial intelligence*, 78(1-2):507–545, 1995.
15. C. Koch and S. Ullman. Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology*, 4(4):219–227, 1985.
16. A.M. Treisman and G. Gelade. A feature-integration theory of attention. *Cognitive psychology*, 136(12):97–136, 1980.
17. L. Itti. *Models of Bottom-Up and Top-Down Visual Attention*. PhD thesis, California Institute of Technology, 2000.
18. L. Itti and C. Koch. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision research*, 40(10-12):1489–506, January 2000.
19. L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(11):1254–1259, March 2002.
20. Z. Li, T. Fang, H. Huo, and J. Zhu. Color conspicuity map based on wavelet low-pass pyramid for popping out contours of salient objects. *Optical Engineering*, 49(5):050502, 2010.
21. S. Engel, X. Zhang, and B. Wandell. Colour tuning in human visual cortex measured with functional magnetic resonance imaging. *Nature*, 388(6637):68–71, 1997.
22. D. F. Specht. A general regression neural network. *IEEE Transactions on Neural Networks*, 2(6):568–576, November 1991.
23. L. Itti, G. Rees, and J.K. Tsotsos. Models of bottom-up attention and saliency. *Neurobiology of attention*, 582(1980):1–11, 2005.
24. D. Gouaillier, V. Hugel, P. Blazevic, C. Kilner, J. Monceaux, P. Lafourcade, B. Marnier, J. Serre, and B. Maisonnier. Mechatronic design of NAO humanoid. In *Proceedings of the 2009 IEEE International conference on Robotics and Automation*, ICRA'09, pages 2124–2129, Piscataway, NJ, USA, 2009. IEEE Press.
25. The RoboCup Standard Platform League. http://www.tzi.de/spl.