Color Based Saccades for Attention Control

Özgür Erkent, Isil Bozma Aydin

Color Based Saccades for Attention Control

Özgür Erkent and Işıl Bozma Aydın

Bogazici University, Electric & Electronic Engineering Dept.,
Intelligent Systems Lab. Istanbul 34342, Turkey,
ozgur.erkent@boun.edu.tr,
WWW home page: http://www.isl.ee.bou.edu.tr

Abstract

This paper addresses a simplified version of attention control - where a robot is asked to attend to scene points with *a priori* specified color. Differing from the classical approaches, in which generating fixations is based on explicit search, we introduce the requirement that the search strategy must be accompanied by a series of saccades whose nature control the fixation process. The robot must start from an arbitrary initial fixation point, start looking through the scene and find the target color points as it is doing so. In the explicit search approach, first, the scene point whose color is most similar to the ”looked-for” color is determined and then the camera is made to move to that point. In the artificial potential functions approach, the two stages are merged together where the camera simply starts moving towards a point whose color is similar to the target color – although not necessarily the most similar. We present working implementations of the two approaches – reporting actual experiments with an attentive robot and comparing the resulting search behaviors using quantitative measures on saccadic selectivity and optimization.

1 Introduction

Attention control through the generation of a series of saccades is a fundamental issue in vision applications [1], [2], [3]. Various computational models of the underlying mechanisms have been proposed [4], [5]. The classical approach to generating fixations is based on explicit search where first a planner determines the next fixation point which is then input to a mover module that directs the cameras to this point. As an alternative approach, in artificial potential functions (APF) based attention control, at each fixation point, a corresponding artificial potential surface is constructed and a saccade occurs via sliding from the current fixation point to a minimal point on this surface. In contrast to the classical approach where the planning and the movement stages occur one after the other, here the planning and movement stages are executed together [5]. The contribution of this paper is to study whether APF based attention control can yield saccading behaviour similar to that of explicit search or not. If the resulting behaviours are shown to have comparable properties, APF based attention control may be preferred in robotic applications since with the integration of the planning and movement stages, time required for each saccade goes down considerably [5].

1.1 Related Literature

There has been considerable research on attention control. One of the earlier models is the Feature Integration Theory [6] which proposes that search is conducted along certain feature dimensions which are then possibly combined via an attention mechanism for each candidate item. A two stage of attention model is proposed by the Guided Search Model [7] where first target locations are created in parallel and then visual attention is used to inspect items sequentially starting from the one with the highest activation. Given that eye movements are usually accompanied by shifts of attention [8], it has been suggested that it is possible to selectively attend to a critical subset of items rather than a one-by-one search. In part motivated by these findings, various computational models have also been proposed. Most define a attention mechanism that always seeks the maximally salient point by utilizing a winner-take-all network through parallel computation [9]. The Area Activation Model proposes several alternative mathematical models that maximize color based saccadic selectivity in human subjects [10]. An alternative approach formulates it as a series of motion planning problems and proposes using a family of artificial potential functions for generating saccades [5]. This approach is generalized to arbitrary saliency measures including a detailed performance analysis in [5]. The fundamental serial-parallel dichotomy present in pre-attentive processing for fast target detection is addressed by having saliency mechanisms work on compressed global information [11]. Of course, much work still remains as existing models and theories provide only partial explanations to the experimental data [11].

1.2 Problem Statement

Consider a robot with the following properties:

- The robot has a field of view comprised of high resolution fovea in the center and low resolution periphery[1].
- The robot is capable of changing its field of view through pan and tilt movements of its head.
- Attention control is achieved via employing one of two alternative saccading models – explicit search or APF.

Suppose it is assigned the task of finding locations with a target color in an unknown environment. Some of these locations may fall within its field of view while most will be outside. As there is no model of the scene given *a priori*, it does not know anything about what's outside the original field. It uses only the information available from its current periphery to predict where to saccade next and hence change its field of view.

[1] This is achieved through using a two-camera system as explained in [12]. However, here as the focus is on saccades which are part of the pre-attention stage, the attentive processing of information coming from the high resolution fovea is beyond the scope of the paper. The reader is referred to [12, 13] for a comprehensive treatment including the complete modelling of attentive vision.

2 Attentive Robots & Saccades

Consider a robot and let its camera's configuration space $\mathcal{F}$ be defined by a subset of the unit sphere S^2. $\mathcal{F}$ corresponds to the pan and tilt movements of the robot head. Suppose that at instant $k\delta T$, the robot is fixated at $f_k \in \mathcal{F}$ such that the image plane I_k lies in the plane tangent to $\mathcal{F}$ at f_k. Fig. 1(left) shows one dimensional view of $f_k \in \mathcal{F}$ and the associated image plane I_k. An intrinsic two-dimensional Cartesian coordinate frame is defined such that its origin coincides with f_k as shown in Fig. 1(right). The visual field P_k^v is a subset of I_k of size $d_v \times d_v$ whose origin coincides with f_k while the fovea $P_k^f \subset P_k^v$ is a smaller region of size $d_f \times d_f$ where $d_f << d_v$. Note that both P_k^v and P_k^f are compact and connected components. Let $\pi_k : \mathcal{F} \rightarrow P_k^v$ denote the transformation map from the configuration space to the visual field such that $\pi_k(f_k) = [0\ 0]^T$ and π_k^{-1} denote the corresponding inverse map. The current visual field is represented by the HSV color map[2] $c_k : P_k^v \rightarrow Z_H \times Z_C \times Z_C$ where $Z_H = \{0, \ldots 359\}$ and $Z_C = \{0, \ldots 255\}$ – adopting a formulation as presented in [14]. Once the visual processing associated with the current fixation is completed, the robot starts saccading until it reaches the next fixation point $f_{k+1} \in \mathcal{F}$ at which it stops as shown in Fig. 1(left). After a saccade, the image plane changes to I_{k+1} as shown in Fig.1(left). The origin of the two image planes are translated by δf. A scanpath S is an ordered sequence of N saccades where each saccade starts at one fixation point f_k and stops at the consecutive fixation point f_{k+1}.

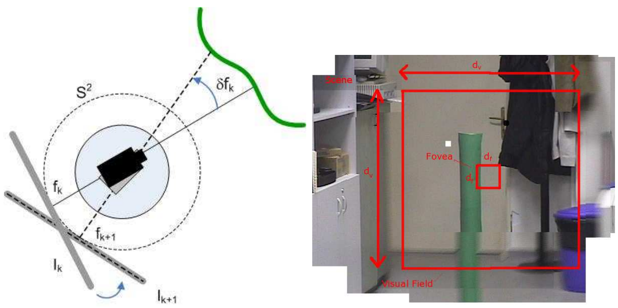

Fig. 1. Left: Top down view (1D projection) of the fixation points and image planes before and after a saccade; Right: A fovea P_k^f centered around fixation point f_k of size $d_f \times d_f$ with a visual field P_v of size $d_v \times d_v$ in a given image plane.

2.1 Color as Saliency

Define a set of "color saliency" functions $\varphi_{f_k} : P_k^v \rightarrow [0, 1]$ parametrized by the fixation points f_k. Each φ_{f_k} is based on $\hat{\varphi}_{f_k} : P_k^v \rightarrow (0, \infty)$ where $\hat{\varphi}_{f_k}(x) = \frac{1}{\beta(x, f_k)}$;

---

[2] $c_k(x) = \left[h_k(x)\ s_k(p)\ v_k(x) \right]^T$. The hue component h_k is the primary color sensed and is associated with the mean wavelength, the colorfulness s_k is related to the width of its spectral density and the brightness v_k is related to intensity.

the saliency of any point x on the image. The saliency measure β is constructed as the weighted product of two factors – $\beta(x, f_k) = \beta_1(x)^{w_1} \beta_2(x, f_k)^{w_2}$.

- Color saliency $\beta_1 : P_k^v \rightarrow R^+$ defined as $\beta_1(x) = \left\| c_k(x) - c^G \right\|^2$ where $c^G = [h^G s^G v^G]^T$ is the target color.
- Inhibition of return and short term memory $\beta_2(x, f_k) = \log(\beta_{21}(x, f_k)\beta_{22}(x))$ where the log operator is used to limit its magnitude,
  - The next fixation point should be maximally away from the current fixation point f_k which may simply be measured by $\beta_{21}(x, f_k) = \|x\|^2$
  - The next fixation point should be maximally away from the previous N_m fixation points as quantized by $\beta_{22}(x) = \prod_{l=1}^{N_m} \|x - \pi_k(f_l)\|^2$.

As $\hat{\varphi}_{f_k}$ can possibly blow up in P_k^v, in order to make $\hat{\varphi}$ admissible, it is squashed by the function $\sigma : (0, \infty) \rightarrow [0, 1]$ defined as: $\sigma(x) = \frac{1}{1+x}$. Finally, φ_{f_k} is constructed as the composition $\varphi_{f_k}(x) = \sigma \circ \hat{\varphi}_{f_k}(x)$.

2.2 Explicit Search

Explicit search simply finds the most salient point and then moves to it. In this case, all points in $P_k^v - P_k^f$ are considered to be candidate foveas. In order to ensure inhibition of return, the fovea P_k^f containing the current fovea is not included in the search.

1. Let k be the saccade number and consider the current fixation point f_k.
2. Compute the color saliency function φ_{f_k} on $P_k^v - P_k^f$
3. Find $x^* \in \arg\min_{x \in P_k^v - P_k^f} \varphi_{f_k}(x)$
4. Move the camera to $f_{k+1} = \pi_k^{-1}(x^*)$.
5. Increment k and go to step 1.

2.3 Artificial Potential Functions

Suppose that the robot is currently fixated on f_k. Assume that the camera dynamics of each saccade can be satisfactorily described by a simple first order model:

$$\dot{x} = -D_x \varphi_{f_k}(x) \tag{1}$$

with the initial condition $x(0) = \pi_k(f_k)$ and where $-D_x \varphi_{f_k}$ is the control input. Let it be noted that the next fixation point f_{k+1} lies in the inverse map π_k^{-1} of the limit set. As this set can contain local minima, the new fixation point is ensured of being locally most salient. The algorithm is defined as follows:
1. Inhibition region: A small region inside the visual field containing the fovea P_k^f is inhibited from containing the next fovea as follows:

- Convolve the fovea P_k^f with a Gaussian filter having standard deviation $\sigma = \frac{1}{3}(d_v - d_f)$ followed by a high-pass gradient filter.
- Let $v_f = -\sum_{x \in P_k^f} D_x \varphi_{f_k}$

– Move to a point $\widetilde{x}_k$ which is located at the intersection of the line starting at the origin and in the direction v_f with the boundary of the inhibition region δP_k^f. Let it be noted that if $v_f = 0$ (which may occur in case of no saliency and prior fixations), the robot is made to move to a randomly selected point on δP_k^f.

If this step is not applied, then the gradient at the current fixation f_k will be so small that the control law $\dot{x} = -D_x \varphi_{f_k}(x)$ would yield no movement.
2. Visual field inhibition: The new fixation point must be inside the visual field P_k^v. This is achieved by the following steps:

– Switch to the control law $\dot{x} = -D_x \varphi_{f_k}(x)$ with the initial point at $\widetilde{x}_k$ and apply it to the camera using the inverse map π_k^{-1} until one of the two following conditions become valid:
  1. $D_x \varphi_{f_k}(x) = 0$.
  2. $x \in \delta P_k^v$ (the boundary of visual field)

As a result of this step, the camera simply slides to a local optimal minima on the saliency surface without having to reconstruct the complete surface.

3 Experiments

We now report on experimental results from a vision-endowed robot to suggest the nature of the color based search behaviour with two alternative models – explicit search and APF. The robot is given a task of finding a target color c^G . In these experiments, the visual field $d_v = 350$, the fovea size $d_f = 40$ and the short term memory $N_m = 6$. The comparison is based on two measures:

– **Saccadic Selectivity:** Saccadic selectivity captures the perceived "similarity of colors" precisely and is a quantitative measure of the attention received by a color given a target color [10]. We assess the goodness of the mathematical models as proposed therein in predicting saccadic selectivity.
– **Optimization of saliency, saccade length and area coverage:** Planning saccades in search tasks need to balance the saliency attained at the destination with the effort required to reach it [15]. Consequently, saccades do not necessarily end up at the most salient points. Hence, for each approach, we evaluate the optimization performance via measuring the degree of finding the target color (as measured by minimal color error) and the required effort (as measured by minimized saccade length while having maximized area coverage).

3.1 Synthetic Scene

The first set of experiments are set up in a manner similar to that conducted on human subjects as presented in [16]. Here, the robot is made to look at an image as shown in Fig.2 and is asked to find for example blue colored squares. Each scanpath consists of 15 saccades. A sample scanpath created using explicit search method is as shown in Fig.2 (left) while that of APF approach is as shown in Fig.2 (right). The robot is made

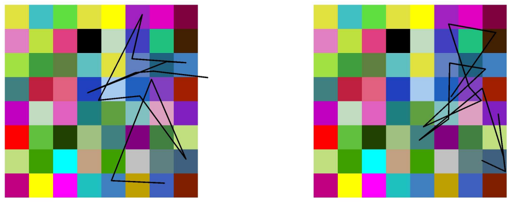

Fig. 2. Sample scanpaths in a synthetic scene where target color is on the third row from top-right. Left: Explicit search; Right: APF.

to repeat the task with random initial fixation points and the resulting behaviours are analyzed.

The optimality of the search scanpath is investigated with respect to three different optimization measures – 1.) Saliency as measured by the color error β_1 function; 2.) Mean saccade distance as measured by the sum of lengths of all saccades divided by the number of saccades; and 3.) Area coverage of the scanpath [3]. There is a trade off between the maximization of the saliency (minimal color error) and minimization of effort (minimal mean saccade length) while having maximal area coverage. The results as presented in Table 1 reveal that the color error for the explicit search method is smaller as expected. However, the mean saccade distance in APF approach is much less while area coverage is higher. These results indicate that the APF approach has a higher tendency to achieving a balance between minimal color error, minimal scanpath distance and maximal area coverage with respect to explicit search. Hence, if the task is to find the best match, explicit search works better while requiring more effort. However, if optimization is important, APF approach seems to have a better performance.

Table 1. Average scanpath measures in synthetic scene: Explicit Search (ES) vs APF.

	ES	APF
Color Error	0.0453	0.1202
Mean saccade length (pixels)	245.8	236.1
Mean Saccade Length (radians)	0.204	0.191
Area Coverage %	3.5	5.7

3.2 Real Environment

In the second set of experiments, the robot is given the task of finding a green colored object in a cluttered environment. The robot is made to repeat the task six times from

[3] The definition is as given in Appendix.

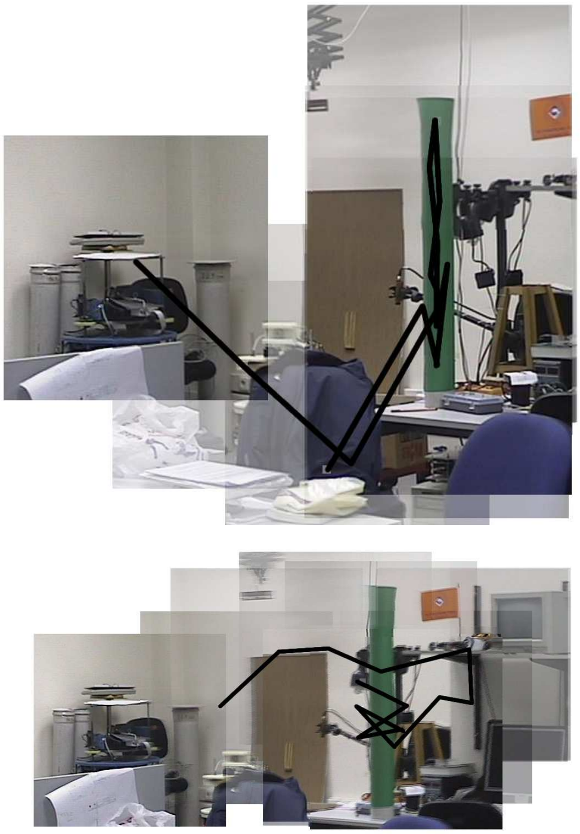

Fig. 3. Finding green colored locations. Top: A sample scanpath created by explicit search method; Bottom: A sample scanpath created by APF.

random initial locations with the object's position being randomly moved in the scene in a manner such that at each location the robot is ensured of seeing the object. Each scanpath consists of 15 saccades. The robot uses the two approaches alternately in generating its scanpaths. A sample scanpath with explicit search approach is as shown in Fig.3(top) while that of APF approach is as shown in Fig.3(bottom). The average color error, mean saccade length and area coverage are computed. Table 2 presents comparative results. It is observed that these results are consistent with those of the synthetic scene.

In summary, experiments have revealed the following:

– The APF approach may not always saccade to the most salient places, however, it seems to have lower energy requirements with shorter saccade lengths and higher area coverage as compared to explicit search.

Table 2. Average scanpath measures in real scene: Explicit search (ES) vs APF.

	ES	APF
Color Error	0.20	0.27
Mean saccade length (pixels)	213.5	135.4
Mean Saccade Length (radians)	0.186	0.112
Area Coverage %	1.6	1.9

– If the target is outside the field of view and there is no *a priori* model of the scene, both approaches can only use the best information available from its periphery to decide where to saccade next and hence change its field of view.

Let it be noted that although the presented results are limited here due to space limitations, our extensive experiments reveal that APF based approach can be used integrated with other tasks such as robot navigation based on color tracking – where due to real-time requirements, optimization becomes more crucial.

4 Conclusion

This paper considers the problem of color based attention control. Two alternative approaches – explicit search vs artificial potential function based saccades – are compared with respect to saccadic selectivity and optimization. In explicit search, first the visual field point whose properties best match the target color is determined and following the camera is made to move to that point. In the alternative artificial potential functions based approach, the camera starts moving towards a point whose color is similar to the target color – although not necessarily the most similar. In assessing the optimization performance, as expected explicit search finds the most salient places. In the APF approach, although saccades are not always to the most salient places, the effort required for doing so is much lower compared to explicit search as the mean saccade length is much lower with an higher area coverage. As part of ongoing work, saccades with saliency functions based on more complex features are being investigated for APF. Artificial potential function based attention control may be preferred in robotic applications since the time required for each saccade goes down considerably as a consequence of the integration of the planning and movement stages.

Acknowledgments

This work has been supported by Bogazici University BAP Project 07A205 and Tübitak MAG Project 107M240.

APPENDIX

A Area Coverage

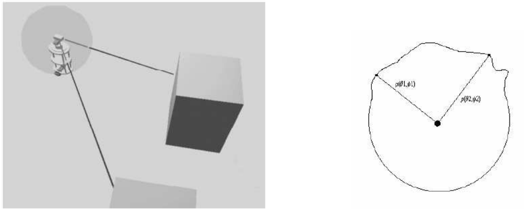

Fig. 4. (Left) A schematic representation of APES, the bubble surface and two consecutive fixation points. (Right) The bubble projected onto a 2D plane in order show the inflation at the two corresponding points [17].

The area coverage is a measure which computes the percentage of the total scene that is covered during the course of the scanpath. It is based on the bubble [17] which is a deformable surface hypothetically placed around the robot as shown in Fig. 4. The bubble surface is defined by spherical coordinates and is initialized to be a partial sphere with radius r and area A before the scanpath begins. Each fixation f_k is assumed to cover a region $\delta A(f_k)$ of size ξA on the sphere. A good measure of area coverage is then based on total area covered during the course of scanpath with respect to the total bubble surface defined as follows:

$$a_c = \frac{\left\| \bigcup_{k=1}^{N_i} \delta A(f_k) \right\|}{A} \tag{2}$$

where $\|\bullet\|$ is the area operator. Note that this measure considers the union of all the foveas visited and is different from simple addition of the areas. The closer this value is 1, the higher is the total area coverage.

References

1. Yang, G., Dempere-Marco, L., Hu, X., Rowe, A.: Visual search: Psychophysical models and practical applications. Image and Vision Computing **20** 291–305 (2002)
2. Gilchrist, I., M.Harvey: Refixation frequency and memory mechanisms in visual search. Current Biology **10**(19) 1209–1212 (2000)
3. Motter, B.C., Belky, E.J.: The zone of focal attention during active visual search. Vision Research **38**(7) 1007–1022 (1998)
4. Itti, L., Koch, C.: Computational modeling of visual attention. Nature Reviews Neuroscience **2**(3) 194–203 (2001)

5. Ilhan, B.D., Erkent, O., Bozma, H.I.: Saccades and fixating using artificial potential functions,. In: Proceedings of IEEE/IRJ Conference on Intelligent Robots and Systems. 5819–5824 (2006)
6. Treisman, A., Gelade, G.: A feature integration theory of attention. Cognitive Psychology **12** 97–136 (1980)
7. Wolfe, J.M.: Visual search. In Pashler, H., ed.: Attention. Hove, England UK 13–71 (1998)
8. Hoffman, J.E.: Visual attention and eye movements. In Pashler, H., ed.: Attention. England UK: Hove (1998)
9. Itti, L., Koch, C.: A saliency-based search mechanism for overt and covert shifts of visual attention. Vision Research 1489–1506 (2000)
10. Pomplun, M., Reingold, E.M., Shen, J., Williams, D.E.: The area activation model of saccadic selectivity in visual search. In Gleitman, L., Joshi, A., eds.: Proceedings of the Twenty Second Annual Conference of the Cognitive Science Society, Mahwah, NJ: Erlbaum 375–380 (2000)
11. Cohen, E., Levy, N., Ruppin, E.: Global vs. local processing of compressed representations: A computational model of visual search. Neurocomputing **32-33** 667–671 (2000)
12. Soyer, C., Bozma, H., Istefanopulos, Y.: Apes - biologically motivated attentive robot. Autonomous Robots **20** 61–80 (2006)
13. Soyer, C., Bozma, H., Istefanopulos, Y.: Attentional sequence based recognition: Markovian and evidential reasoning. IEEE Transactions on Systems, Man and Cybernetics **33** 937–950 (2003)
14. Gonzalez, R.C., Woods, R.E.: Digital Image Processing. Pearson Prentice Hall (2007)
15. Araujo, C., Kowler, E., Pavel, M.: Eye movements during visual search: the costs of choosing optimal path. Vision Research **41** 3613–3625 (2001)
16. Xu, Y., Higgins, E.C., Xiao, M., Pomplun, M.: Mapping the color space of saccadic selectivity in visual search. Cognitive Science **31** 877:887 (2007)
17. Soyer, C., Bozma, H., Istefanopulos, Y.: A new memory model for selective perception systems. In: Proceedings of IEEE/RSJ International Conference on Robots and Systems. 2304–2309 (October 2000)