# Robots Looking for Interesting Things: Extremum Seeking Control on Saliency Maps

Yinghua Zhang, Jinglin Shen, Mario Rotea and Nicholas Gans

***Abstract*— This paper presents a novel approach to increase the amount of visual stimuli in sensor measurements using saliency maps. A saliency map is a combination of normalized feature maps in different channels (i.e. color, intensity) to represent the relative strength of visual stimuli in an image. The total saliency is higher when the camera is looking at a scene with more interesting things in the field of view and vise versa. We employ methods of extremum seeking control to find a camera position that corresponds to local maximum saliency value. We combine the global properties of simplex optimization methods with the local search properties and dynamic response of extremum seeking control to create a novel algorithm that is more likely to find a global maximum than conventional extremum seeking control. Simulations and experiments are presented to show the strength of this approach.**

## I. Introduction

Given limited sensors covering a wide area, a sensor needs to isolate targets of interest to maximize the value of its measurements. Alternately, given abundant sensors, the amount of data may overwhelm communication channels, processor bandwidth, or human observers, necessitating the ability to transmit only the most useful data. This paper presents an initial investigation to control the position of a sensor to collect the most valuable measurements via extremum seeking control (ESC) of sensor configuration. In particular, we seek to maximize the visual stimuli in images or video data to provide the most relevant images.

ESC seeks optimize the value of a measurable cost function [1]. The strength of these methods is that no prior knowledge of the cost function is necessary. A stability proof of ESC was first provided by Krstic and Wang for a general nonlinear SISO system [2]. Multivariable ESC was later studied by Rotea, and a set of detailed design guidelines for ESC were provided [3]. Recently, Global ESC methods were studied by Tan and Nesic [4].

The above methods share a common framework. The control input is the current estimate of the optimal input. A periodic disturbance or dither signal (commonly a sinusoid) is added to the control input. Via a series of filters and modulating signals, an estimate of the gradient is generated. This gradient is integrated to produce the control input. Under certain conditions of the system, output function, dithering functions and filters, the ESC methods can be proven to converge to the extremum. Variation of the ESC methods have been developed to remain stable despite nonlinear dynamics of the system.

J. Shen, Y. Zhang and N. Gans are with the Department of Electrical Engineering, M. Rotea is with the Department of Mechanical Engineering, University of Texas at Dallas, Richardson, TX 75080, USA {yxz102220, shen.jinglin, rotea, ngans}@utdallas.edu

Saliency describes the level of attractiveness of visual stimuli, often modeled on human visual response. Many saliency-based attention models and computational visual attention systems has been developed. Koch and Ullman first [5] proposed to integrate a number of different visual feature maps such as color, orientation and direction of movement into a global measurement of conspicuity, known as saliency map. A saliency map based visual attention model was proposed by Itti et al. [6]. Other top-down modulation methods were proposed [7], [8] in cases when some knowledge of the appearance is known in advance or according to specific task demands.

In this work, we use saliency map in a extremum seeking control problem. The saliency map is computed according to intensity, color and orientation channels. Instead of using a winner-take-all scheme to find the most salient region, the sum of saliency values of every pixel in the image is calculated and used as the cost function. Using ESC to guide a camera to the location of maximum saliency is an ideal approach, as knowledge of the saliency as a function of the sensor workspace is not needed. Saliency as an objective function map often has multiple local maxima, and ESC algorithms can easily attract to a local maximum rather than a global solution. Motivated by this issue, we present a novel approach that combines the global properties of simplex optimization methods and dynamic properties of extremum seeking control. We call this combined method Simplex Guided Extremum Seeking.

Saliency has been extensively utilized in the field of computer vision and robotics in recent years, such as detecting regions of interest [6], video compression [9], robot localization and SLAM [10], [11], as well as robot motion planning and human-robot interaction. Vijayakumar et al. [12] implemented a visual attention system using a humanoid robot, whose peripheral camera followed a moving object recognized in the saliency map. Other systems use visual attention to guild robot in object manipulation problems [13], [14]. In [15], a attention model is built for humanoid robots using both visual and acoustic saliency maps.

Visual attention system has also been used in visual servoing problems, like the visual attention guided robot navigation in [16]. Also, Scheier et al. [17] built a mobile robot that approaches large object using saliency map. Recently, visual servoing methods that are based on image intensities have been developed, such as [18], [19]. These methods do not require any tracking or matching process, but suffer from the sensitivity to illumination variations. Dame et al. [20] proposed a Mutual information-based method, which shows

good robustness to environment changes.

Alternative approaches to sensor placement, and camera placement in particular, have been investigated. Several groups (e.g. Howard et al. [21], Murray et al [22], and Zou and Chakrabarty [23]) focused on coverage, i.e. maximizing the amount of area that is covered by at least one sensor. Mittal and Davis place sensors to avoid occlusions [24]. Research by Zhao et al. focused on arranging multiple sensors to simultaneously measure areas or track targets [25]. Abidi suggested using maximum expected entropy to choose what new camera view of an object would add the most new information [26]. Papanikolopoulos has investigated sensor placement to reduce the amount of processing that must be performed [27] or reducing the expected error in the final estimation [28]. Similar work was done by Ercan et al. [29]. Previous methods differ from the approach proposed in this paper in that they utilized off line optimization methods and knowledge of the scene and environment. The method proposed in this paper runs in real time, can adapt to dynamic environments, and does not require knowledge of the scene or environment.

The ability to focus on areas of high visual stimuli may help to reduce transmission bandwidth, and improve accuracy of estimation or recognition algorithms. This paper investigates what environmental conditions allow for stable ESC of image saliency. We also investigate what ESC design parameters, such as frequency of the differ signal, are necessary for stability and performance given the slow sampling rate of most cameras (approximately 30Hz). Experiments are performed to show the strength of proposed method.

## II. Background

### A. Extremum Seeking Control

ESC is designed to optimize a cost function in real time, without any prior knowledge of the input-to-cost mapping. References [2]–[4] concentrated on developing ESC methods. Fig. 1 shows a common scheme of ESC. The current estimate of the optimal state of the system is $\overline{\theta}(t) \in \mathbb{R}^n$. A dither signal $d_1(t) \in \mathbb{R}^n$ is added to $\overline{\theta}(t)$ to give the current state $\theta(t)$. The signal $d_1(t)$ is typically given by a vector of sinusoids $a_i \sin(\omega_i t)$, $i = 1 \ldots n$.

The output $y(t) \in \mathbb{R}$ can be expressed by the Taylor Series

$$y(t) = f(t,\bar{\theta}) + d_1(t)^T \frac{\partial f(t,\bar{\theta})}{\partial \theta} + H.O.T. \qquad (1)$$

Neglecting higher order terms, passing $y(t)$ through a high pass filter block gives a signal correlated to the gradient vector $\partial f(t,\bar{\theta})/\partial\theta$. The gradient is extracted via a demodulation scheme that multiplies the output of the high-pass filter by the dither signal $d_2(t) \in \mathbb{R}^n$, followed by application of a low-pass filter. The resulting signal $\zeta(t) \in \mathbb{R}^n$ is an estimate of the gradient. A signed scalar gain term $k$ determines the direction and speed of motion (i.e. whether we seek maximum or minimum and the rate of convergence). Integrating $k\xi$ gives the current estimate of the optimal state $\overline{\theta}(t) \in \mathbb{R}^n$.

ESC systems are generally nonlinear. However, when the dither signal frequencies $\omega_i$ are large enough, averaging theory [30], [31] can provide a linear system that approximates

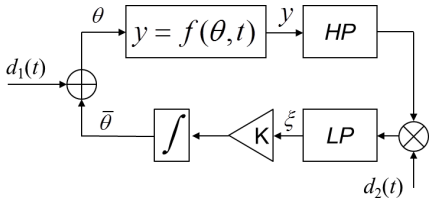



Fig. 1. Block Diagram of the Extremum Seeking Loop

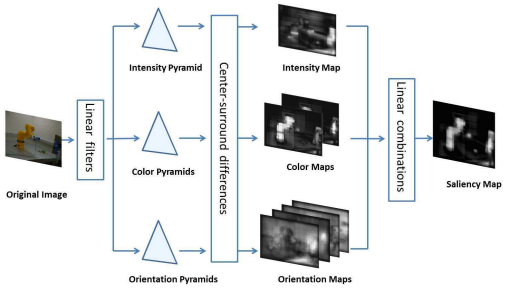



Fig. 2. Saliency model

the dynamics of the ESC loop. Using this linear approximation, reference [3] provides guidelines for selecting dither signals and filters to ensure closed loop stability of the ESC loop. In Section III-B we use these guidelines to design an ESC system for saliency maximization.

### B. Saliency

A saliency map is a presentation of visual stimulation in an image. It is typically the combination of different feature maps. The use of different channels and the weight of each feature are decided according to applications and desired tasks. In this work, the computation of saliency map follows the bottom-up procedure described in [6].

Fig. 2 shows the saliency model in [6]. Three features are used for generating the saliency map: intensity, color and orientation. The color image is first converted to monochrome images in each of the channels. An intensity image $I$ is created as $I = (r+g+b)/3$. Four color channels are used

$$R = r - (g+b)/2, \qquad G = g - (r+b)/2$$
$$B = b - (r+g)/2, \quad Y = (r+g)/2 - |r-g|/2 - b.$$

Gaussian pyramids are then built for all channels as $I(\sigma)$, $R(\sigma)$, $G(\sigma)$, $B(\sigma)$ and $Y(\sigma)$, where $\sigma \in [1..9]$. For the orientation channels, four Gaussian pyramids $O(\sigma,\theta)$ are built by convolving the intensity pyramid with an oriented Gabor filter [32], where $\sigma \in [1..9]$ and $\theta \in \{0°, 45°, 90°, 135°\}$.

After the feature pyramids are created for all channels, feature maps are obtained by calculating the center surrounded difference between different levels in the pyramids, which is denoted as $\ominus$. Specifically, the center surrounded difference of a finer scale $c$ and a coarser scale $s = c + \delta$ is given by interpolation of $s$ to the finer scale, followed by point-by-point subtraction. If multiple scales in the pyramid are used, multiple feature maps are generated. Feature maps for Intensity and Orientation are given by:

$$\mathscr{I}(c,s) = |I(c) \ominus I(s)|$$
$$\mathscr{O}(c,s,\theta) = |O(c,\theta) \ominus O(s,\theta)|.$$

For color channels, two color maps (i.e. red/green and blue/yellow) are generated out of the four color pyramids.

$$\mathscr{RG}(c,s) = |(R(c) - G(c)) \ominus (G(s) - R(s))|$$
$$\mathscr{BY}(c,s) = |(B(c) - Y(c)) \ominus (B(s) - Y(s))|.$$

Finally, linear combinations of the feature maps give the saliency map $\mathscr{I}$ as shown in Fig. 2.

$$\mathscr{S} = \sum \mathscr{I}(c,s) + \sum \mathscr{RG}(c,s) + \sum \mathscr{BY}(c,s) + \sum \mathscr{O}(c,s,\theta).$$

In most previous applications,a normalization step is applied on feature maps when linear combination is performed in order to convert all maps to the same scale, and eliminate modality-dependency.

## III. ESC Design for Saliency Maximization

In this Section, we provide the design parameters of the ESC loop to steer a camera to maximize the saliency of captured images. In this initial investigation, robot/camera motion is limited to a 2D vertical plane. Saliency is maximized by adjusting the horizontal and vertical position of the camera. That is, the optimization variables are $\theta = [x, y]^T$ where $x$ and $y$ are the coordinates of the camera in a plane perpendicular to the ground.

### A. Saliency mappings

As mentioned in section II-B, feature maps are normalized when they are combined for building a saliency map. However, normalization is not done in this work due to a different use of the saliency map. A saliency value $S$ of an image is defined as the sum of the entire saliency map.

$$S = \sum_{i=1}^{width} \sum_{j=1}^{height} \mathscr{S}(i,j). \qquad (2)$$

$S$ in (2) is used as the cost function for the extremum seeking algorithm, and the goal is to find the camera position that maximize the amount of visual stimulation in the environment. Therefore, information from all channels directly contributes to the final saliency map without being normalized. How much effect a feature has on the saliency map is controlled by the assigned weight to each feature map. For example, there are more orientation maps than intensity maps, so a smaller weight for orientation is used to balance the effect of orientation channels.

Fig. 3 further illustrates this idea. Comparing the two scenes in Fig. 3(a) and (c), a higher saliency value for Fig. 3(c) is expected according to our definition of $S$. However, the saliency value in (c) is smaller as can be seen at the bottom of Fig. 3(a) and (c). This is because the normalized saliency map reflects only relative strength of visual stimuli in a single image, and thus provides little information when comparing with saliency maps of other images. Therefore, absolute magnitude of each feature map is used for generating the saliency map in order to search for the maximum saliency value $S$ among all possible camera poses.

ESC is an approach for unconstrained optimization, thus a condition for stability or convergence is that the saliency value has local maximum in the interior of the camera workspace. Maps of saliency value as a function of camera

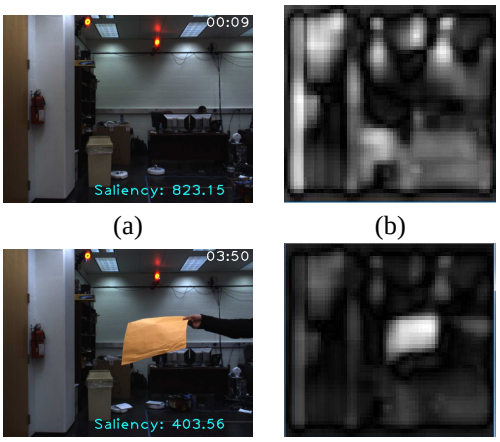



Fig. 3. (a) shows a lab environment where no single object visually stands out. (b) is the saliency map of (a) with all normalization steps proposed in [6]. (c) and (d) are generated in the same way, but with a manually added salient object in the field of view. It can be seen that several regions are bright in (b) since they have similar amount of visual attraction. However, the existence of the yellow folder in (c) makes the all the bright regions in (b) dimmer because the folder visually stands out from the environment.

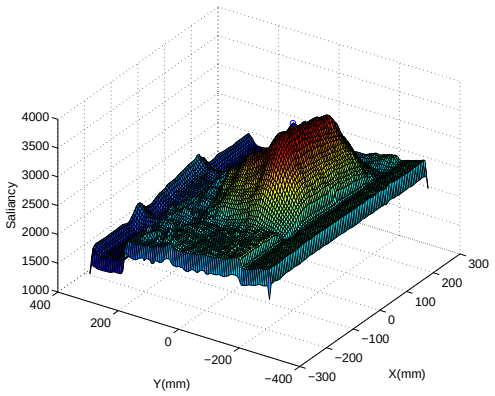



Fig. 4. Saliency Mapping with a few maxima

location are generated to verify this condition in two scenarios. A camera is mounted on the end effector of a Staubli TX90 robot arm. The camera is moved to uniformly sample images in the environment.

In the first scenario, a monochrome poster board was placed in front of the robot with a picture taped in the center. This represents a single area of interest in a largely uninteresting field. The saliency value, as a function of the camera position $[x,y]^T$ is shown in Fig. 4. A clear global maximum can be seen in the interior of the workspace.

In the second scenario, the camera looks at normal lab environment. The saliency mapping is given in Fig. 5. Multiple local maxima can be observed. The global maximum is close to the edge but still in interior of the workspace. Simulation and experiment results for these two scenarios are given in section V and VI.

### B. ESC Design

As shown in [3], there are design variables that affect the stability and performance of ESC systems. These variables include the two dither signals, the high-pass filter, the low-pass filters, and the gain $k$. The sensor modality

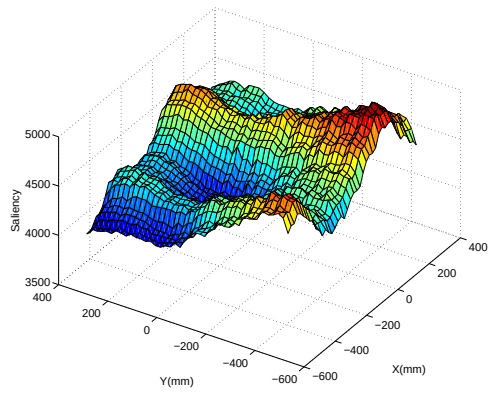



Fig. 5. Saliency Mapping with multiple local maxima

and system plant, i.e. images from a video camera and a robot manipulator, place further restrictions on the choice of design variables. In this section we briefly recap the stability requirements given in [3] and present our determination of best practices for the ESC task at hand.

In this initial investigation only two degrees of freedom are used, translation in a plane perpendicular to the ground. Hence we need a dither signal of the form $d_1(t) = [a_1 \sin(\omega_1 t), a_2 \sin(\omega_2 t)]^T$ The choice of frequencies for the dither vector must satisfy the following constraints [3]:

1) $\omega_1 \neq \omega_2$
2) $\omega_1$ and $\omega_2$ are in the pass band of the high-pass filter and stop band of the low-pass filter.
3) $\omega_1$ and $\omega_2$ are smaller than one half the sampling rate.

Video systems generally have frame rate around 30Hz. After the time required for image processing and control calculation, the sampling rate is approximately 20Hz. We employ dither frequencies in the range 5Hz - 9Hz, which can be tuned for performance using trial and error.

The dither signal amplitudes $a_1$ and $a_2$ affect the seeking accuracy and convergence speed of the ESC loop. A larger dither amplitude lowers the time for convergence but decreases the accuracy of the final estimate for $\bar{\theta}$. Designers must also consider the forces necessary to generate dither signals with large amplitudes. For the plant employed in this paper, we must considering the speed limit of robot actuators. We chose $a_1 = a_2 = 5cm$ to provide good performance in speed and accuracy while not stressing the robot.

The cutoff frequency of the high-pass filter should be lower than $\omega_1$ and $\omega_2$. We employ a second order Chebyshev type I filter with a cutoff frequency of 3Hz, given by

$$G_1(z) = \frac{0.5768z^2 - 1.1536z + 0.5768}{z^2 - 1.0900z + 0.4735}.$$

The cutoff frequency of low-pass filter should also be lower than $\omega_1$ and $\omega_2$. We use an FIR filter with a cutoff frequency of 1Hz, given by

$$\begin{aligned} G_2(z) \quad = \quad & 0.0100 + 0.0249z^{-1} + 0.0668z^{-2} + 0.1249z^{-3} \\ & + 0.1756z^{-6} + 0.1249z^{-7} + 0.1756z^{-4} \\ & + 0.1957z^{-5} + 0.0668z^{-8} + 0.0249z^{-9}. \end{aligned}$$

## IV. Simplex Guided Extremum Seeking

As shown in Fig. 5, the saliency function can have multiple local extrema. ESC is very likely to be trapped at a local maximum and can never reach the global maximum. Therefore, the final result can be heavily affected by the initial position. It has been shown that increasing the amplitude of dither signal can improve the chance to reach the global extremum [33]. However, high amplitude signals can saturate the actuators and make it difficult to demodulate the signal to gather the gradient information.

Alternately, a multi-directional algorithm that searches for extrema through the whole workspace, is more likely to find a global maximum [34]–[36]. Multi-directional search algorithms are approaches to linear programming that construct a point simplex and iteratively optimize the points to converge to the extremum, therefore they are referred to as simplex methods. The maximum point in the simplex is always kept, and a group of linear combinations (reflection, extension, and contraction) are used to predict points with a better value. This continues until the best point is located or a termination condition is met. The downside to simplex methods is poor dynamic response, in that they are not well suited to changing maps, such as the saliency of a changing scene.

Therefore, we propose a combined ES algorithm that can employ simplex methods to make extremum seeking more global while preserving the dynamic tracking abilities of ESC. We call this method Simplex Guided Extremum Seeking (SGES), which uses a simplex method for large scale searching, and ESC for small-scale local searching. SGES shows strong promise for optimizing the cost functions that have many local extrema along with the global one.

For a $n$ dimensional search space, SGES executes ESC at $n+1$ initial trial points to obtain $n+1$ local maxima. The maxima are taken as simplex vertices and denoted as $\mathbf{x}_0^0$, $\mathbf{x}_1^0$, $\mathbf{x}_2^0 ... \mathbf{x}_n^0$. The superscript represents iteration time, and they are ordered after every vertex update, such that $f(\mathbf{x}_0^k) > f(\mathbf{x}_i^k)$ for $i = 1,2...n$ in any $k$th iteration.

As $\mathbf{x}_0^k$ is the current best maximum, it is reasonable to assume this vertex lies in a more optimal region of the workspace. So we perform reflection to generate $n$ initial trial points $\mathbf{r}_i^k = \mathbf{x}_0^k - \alpha(\mathbf{x}_i^k - \mathbf{x}_0^k)$ for $i = 1,2...n$, where $\alpha > 0$ is a constant. ESC is performed from each trial point, leading to a new group of local maxima, denoted as $\hat{\mathbf{r}}_i^k$.

If there is a local maximum $\hat{\mathbf{r}}_{j_r}^k$, such that $f(\hat{\mathbf{r}}_{j_r}^k) > f(\mathbf{x}_0^k)$ and $0 < j_r \leq n$, it is possible that better points could be found further along this direction. So we perform on extension step, generating $n$ initial trial points $\mathbf{e}_i^k = \mathbf{x}_0^k - \lambda(\mathbf{x}_i^k - \mathbf{x}_0^k)$ for $i = 1,2...n$, where $\lambda > \alpha$ is a constant. ESC is ran from each trial point, producing one more group of maxima $\hat{\mathbf{e}}_i^k$.

If there is a local maximum $\hat{\mathbf{e}}_{j_e}^k$, such that $f(\hat{\mathbf{e}}_{j_e}^k) > f(\hat{\mathbf{r}}_{j_r}^k) > f(\mathbf{x}_0^k)$, and $0 < j_e \leq n$, we accept $\hat{\mathbf{e}}_i^k$ to update the vertices, i.e. $\mathbf{x}_i^{k+1} = \mathbf{e}_i^k$ for $i = 1,2...n$, else we accept $\hat{\mathbf{r}}_i^k$, i.e. $\mathbf{x}_i^{k+1} = \hat{\mathbf{r}}_i^k$ for $i = 1,2...n$.

If there is no $\hat{\mathbf{r}}_{j_r}^k$, such that $f(\hat{\mathbf{r}}_{j_r}^k) > f(\mathbf{x}_0^k)$, we accept the $\mathbf{x}_0^k$ as the current global optimal and contract all $\mathbf{x}_i^k$ for $i = 1,2...n$ towards $\mathbf{x}_0^k$. In this case, generate $n$ initial trial points

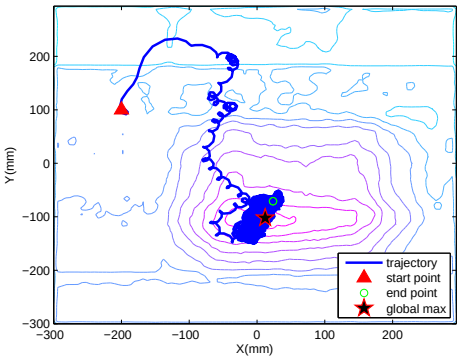



Fig. 6. Simulation result using ESC for the first scenario

$\mathbf{c}_i^k = \mathbf{x}_0^k + \theta(\mathbf{x}_i^k - \mathbf{x}_0^k)$, for $i = 1,2...n$, where $0 < \theta < 1$. Again performing ESC from each trial point leads to one more group of local maxima $\hat{\mathbf{c}}_i^k$. If there is some $\hat{\mathbf{c}}_{j_c}^k$, such that $f(\hat{\mathbf{c}}_{j_c}^k) > f(\mathbf{x}_0^k)$, we accept contraction step and update vertices as $\mathbf{x}_i^{k+1} = \hat{\mathbf{c}}_i^k$, for $i = 1,2...n$. Otherwise, we update vertices as $\mathbf{x}_i^{k+1} = \mathbf{c}_i^k$, for $i = 1,2...n$ to guarantee convergence.

We will show in our future works that, under specific conditions, all the simplex vertices of SGES will converge to some optimal local maximum. For better search performance, the $n+1$ initial trial points should construct a simplex with $n$ linear independent edges.

## V. Simulation Results

In this section, simulations are conducted to demonstrate the performance of both basic ESC and SGES, using the saliency mappings shown in Figs. 4 and 5. ESC is tested for the first map, and simulations of both ESC and SGES are presented for the second map.

The simulation result using ESC on the saliency value map with few local maxima very near the global maximum (shown in Fig. 4), is given by Fig. 6. The background of the figure is a contour plot of the saliency value mapping. The initial point is denoted as a red triangle, and the end point is denoted as a green circle. It can be seen that the camera trajectory converges to the global maximum, which is shown as a star on the map. The "curly" nature of the trajectory is due to the dither signals necessary for ESC.

In cases where multiple local maxima are widely distributed on the map, such as shown in Fig. 5, ESC does not guarantee to converge to the global maximum. Three trials are done for ESC on the multiple local maxima map with different starting points: $[x,y]^T = [-100,-100]$, $[100,100]$ and $[0,100]$. The simulation result is given in Fig. 7. For all three trials, the camera settles at a local maximum instead of going to the global maximum.

Fig. 8 represents the simulation result of SGES for the multiple maxima case. The same three initial conditions are used to construct the initial simplex. The final position of the camera converges very close to the global maxima. Each dashed line represents one simplex update step motion. Fig. 9 shows how the simplex is updated. The red and the green dashed lines are minimum point and medium point in the

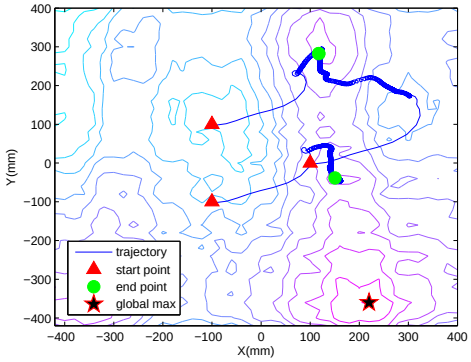



Fig. 7. Simulation result using ESC for the second scenario where multiple local maxima exists

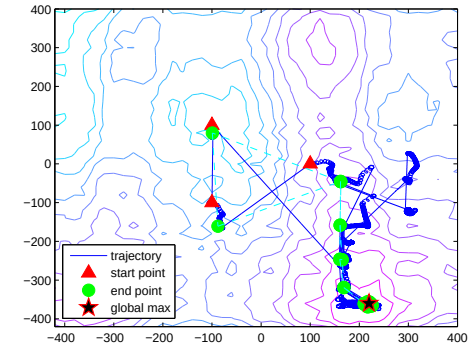



Fig. 8. Simulation result using SGES for the second scenario where multiple local maxima exists

simplex set respectively. The solid line is the maximum value point in the simplex set, which increases monotonically. At the same time, the area of the triangle constructed by the three simplex points monotonically decreases, meaning that the SGES algorithm converges in both space and the value of $S$.

The saliency extrema found for each initial condition in the three simulations is given in Table I. The results show that the proposed SGES method can improve the ability of extremum seeking to find the global maximum.

## VI. Experimental Results

Experiments are performed to demonstrate the performance of the proposed ESC and SGES methods. A Staubli

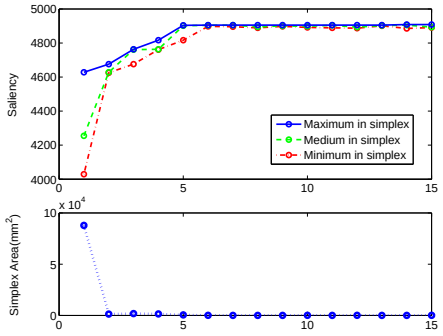
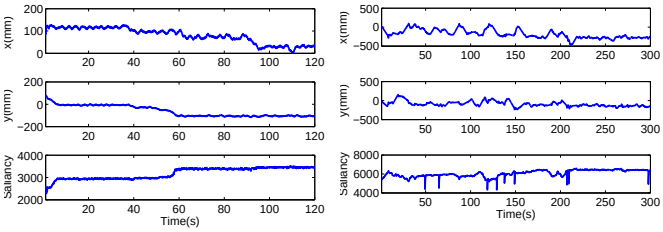



Fig. 9. Simplex update for SGES. The simplex points converge in the measure of saliency. At the same time, the area of the polygon the simplex points define gets smaller, indicating they converge in space.

| Sim. # | Extremum S achieved | Global Max. S |
|---|---|---|
| 1 | 3643.6 | 3730.6 |
| 2 | 4631, 4632.8, 4631 | 4910 |
| 3 | 4891, 4893.4, 4908.3 | 4910 |

TABLE I

EXTREMA FOUND IN SIMULATION OF THE DIFFERENT ALGORITHMS




(a) (b)

Fig. 10. ESC Experiment results for single and multiple maxima scenario

TX90 robot arm with 6 degrees of freedom is used. A camera is mounted on the end effector of the robot. In this initial investigation, only 2 degrees of freedom are used, that is, translation in the $x-y$ plane.

In the first experiment, ESC is used for the first scenario described in Section III-A, featuring a monochrome poster board with a picture fixed at the center. The camera was placed at an initial position away from the global maximum. The experiment result is given in Fig. 10(a). It can be seen that the camera converges to the position that the maximizes the saliency value. By maximizing the saliency value, the camera bring the picture into full view, which is visually the most interesting thing in the environment. The $x$ position is not completely steady after the saliency value converges, since the saliency value does not change much as long as the picture stays entirely in the field of view.

Next, SGES is tested using the second scene described in Section III-A, with three salient objects in front of the camera. Unlike in the first experiment, the saliency mapping has multiple local maxima distributed widely in the camera work space. We start SGES with three random simplex points. At each point, ESC is employed to find the local maxima. When the system detects a gradient less than 1 for 2 seconds, the system moves to the next predicted simplex point. This switch condition can be tuned for desired performance. The position of the camera and saliency measure over time and the simplex update plot are shown in Fig. 10(b) and Fig. 11, respectively. The simplex points both converge in terms of saliency value, and the area of the polygon defined by the points becomes smaller, meaning the location of the points converges. Fig. 12 gives the scenes taken at the starting and finishing camera locations. At the beginning, the camera only sees two of the three salient objects on the table. The third one is brought into full view at the end.

The third experiment explores the dynamic property of SGES. The camera looks at the lab environment with several books and a cup placed on a table, as shown in Fig. 13(a). Fig. 13(b) shows that SGES first converges to a place that includes the door in the field of view. In Fig. 13(c) another book is placed on the table, and SGES finally converges to include the added book into full view in Fig. 13(d). The

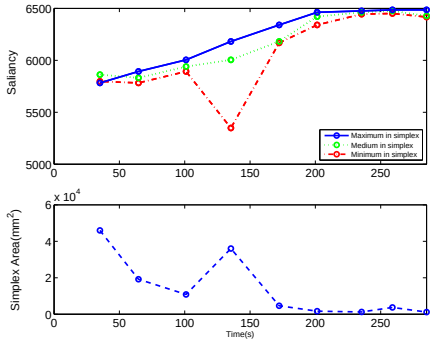



Fig. 11. Simplex update for SGES Experiment

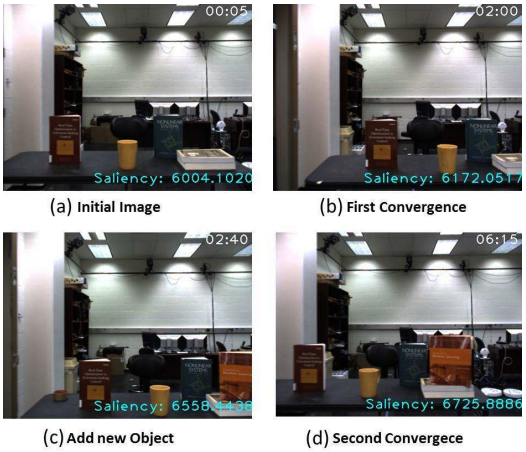

(a) (b)

Fig. 12. Starting and end position for SGES experiment

simplex update is shown in Fig. 14. The simplex area is very small at the end, showing that the position of simplex points converges. However, the final minimum and medium in simplex are not as close to the maximum as in previous experiments. This could be due to the large gradient around the global maximum point. The result shows that SGES works for image scenes that change dynamically.

## VII. CONCLUSION AND FUTURE WORK

We propose a extremum seeking control method to maximize the amount of visual stimuli in an image scene. This method is used to guide the camera to look at interesting things in the environment. Since the saliency distribution often has multiple local maxima, a novel Simplex Guided Extremum Seeking approach is employed that combines the global properties of simplex optimization methods and dynamic properties of extremum seeking control. Simulation and experiment results shows strong potential of the proposed method. SGES efficiently converges to the global




(a) Initial Image (b) First Convergence

(c) Add new Object (d) Second Convergece

Fig. 13. Critical Scenes for dynamic SGES Experiment

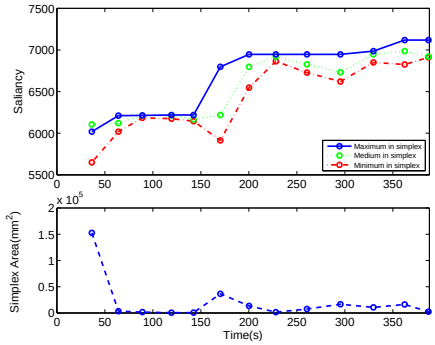



Fig. 14. Simplex Update for dynamic SGES Experiment

maxima. SGES may have application in many other real time optimization problems that experience local maxima.

This work is an early investigation, and there are several avenues open for future work. A formal analysis must be done to prove the stability and convergence of SGES approach. Only a subset of the 6 degrees of freedom of the robot is used in this work. Extensive experiments are needed to explore the performance of proposed method adding more degrees of freedom. Finally, more channels can be added when building saliency map, such as variance, entropy and visual surprise. Other sensors and information measures like uniformity can be explored, and may offer better performance in some circumstances.

## REFERENCES


[1] J. Sternby, "Extremum control systems: An area for adaptive control," in *Preprints of the Joint American Control Conference*, 1980.

[2] M. Krstic and H.-H. Wang, "Design and stability analysis of extremum seeking feedback for general nonlinear systems," *Proc. Conf. Desision and Control*, pp. 1743–1748, Dec. 1997.

[3] M. Rotea, "Analysis of multivariable extremum seeking algorithms," in *Proc. American Control Conference*, vol. 1, no. 6, Sep. 2000, pp. 433 –437.

[4] Y. Tan, D. Nesic, I. Mareels, and A. Astolfi, "On global extremum seeking in the presence of local extrema," *Automatica*, vol. 45, no. 1, Jan. 2009.

[5] C. Koch and S. Ullman, "Shifts in selective visual attention: towards the underlying neural circuitry," *Human Neurobiology*, vol. 4, no. 4, pp. 219–227, 1985.

[6] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254 –1259, Nov. 1998.

[7] J. M. Wolfe, "Guided search 2.0: A revised model of visual search," *Psychonomic Bulletin & Review*, vol. 1, no. 2, pp. 202–238, 1994.

[8] E. Niebur and C. Koch, "Control of selective visual attention: modeling the 'where' pathway." in *Advances in Neural Information Processing Systems*, D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo, Eds. Cambridge, MA: MIT Press, 1996, pp. 802–808.

[9] Z. Li, S. Qin, and L. Itti, "Visual attention guided bit allocation in video compression," *Image and Vision Computing*, vol. 29, no. 1, pp. 1–14, 2011.

[10] S. Frintrop and P. Jensfelt, "Attentional landmarks and active gaze control for visual slam," *IEEE Trans. on Robotics*, vol. 24, no. 5, pp. 1054 –1065, 2008.

[11] C. Siagian and L. Itti, "Biologically inspired mobile robot vision localization," *IEEE Trans. on Robotics*, vol. 25, no. 4, pp. 861–873, 2009.

[12] S. Vijayakumar, J. Conradt, T. Shibata, and S. Schaal, "Overt visual attention for a humanoid robot," in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2001.

[13] M. Bollmann, R. Hoischen, M. Jesikiewicz, C. Justkowski, and B. Mertsching, "Playing domino: A case study for an active vision system," in *Proc. of the First International Conference on Computer Vision Systems*, 1999.

[14] A. Rotenstein, A. Andreopoulos, E. Fazl, D. Jacob, M. Robinson, K. Shubina, Y. Zhu, and J. Tsotsos, "Towards the dream of intelligent, visually-guided wheelchairs," in *Proc. of the 2nd International Conference on Technology and Aging*, 2007.

[15] J. Ruesch, M. Lopes, A. Bernardino, J. Hornstein, J. Santos-Victor, and R. Pfeifer, "Multimodal saliency-based bottom-up attention a framework for the humanoid robot icub," in *IEEE International Conference on Robotics and Automation*, 2008.

[16] A. Borji, "Interactve learning of task-driven visual attention control," Ph.D. dissertation, Institute for Research in Fundamental Sciences, School of Cognitive Sciences, 2009.

[17] C. Scheier and S. Egner, "Visual attention in a mobile robot," in *Proc. of the IEEE International Symposium on Industrial Electronics*, 1997.

[18] C. Collewet and E. Marchand, "Photometric visual servoing," *IEEE Transactions on Robotics*, vol. PP, no. 99, pp. 1–7, 2011.

[19] K. Deguchi, "A direct interpretation of dynamic images and camera motion for vision guided robotics," in *IEEE/SICE/RSJ International Conference on Multisensor Fusion and Integration for Intelligent Systems*, dec 1996, pp. 313 –320.

[20] A. Dame and E. Marchand, "Mutual information-based visual servoing," *IEEE Transactions on Robotics*, vol. PP, no. 99, pp. 1 –12, 2011.

[21] A. Howard, M. J. Mataric, and G. S. Sukhatme, "Mobile sensor network deployment using potential field: A distributed scalable solution to the area coverage problem," in *Proc. of the International Conference on Distributed Autonomous Robotic Systems*, 2002, pp. 299–308.

[22] A. T. Murray, K. Kim, J. W. Davis, R. Machiraju, and R. Parent, "Coverage optimization to support security monitoring," *Computers, Environment and Urban Systems*, vol. 31, no. 2, pp. 133 – 147, 2007.

[23] Y. Zou and K. Chakrabarty, "Sensor deployment and target localization in distributed sensor networks," *ACM Trans. Embed. Comput. Syst.*, vol. 3, no. 1, pp. 61–91, 2004.

[24] A. Mittal and L. S. Davis, "A general method for sensor planning in multi-sensor systems: Extension to random occlusio," *International Journal of Computer Vision*, vol. 76, no. 1, pp. 31–52, 2008.

[25] J. Zhao, S.-C. Cheung, and T. Nguyen, "Optimal camera network configurations for visual tagging," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 2, no. 4, pp. 464–479, Aug. 2008.

[26] B. R. Abidi, "Automatic sensor placement," in *Intelligent Robots and Computer Vision XIV: Algorithms, Techniques, Active Vision, and Materials Handling*, D. P. Casasent, Ed., vol. 2588, no. 1. SPIE, 1995, pp. 387–398.

[27] D. Fehr, L. Fiore, and N. Papanikolopoulos, "Issues and solutions in surveillance camera placement," in *Proc. IEEE Conf. Intelligent Robots and Systems*, 2009, pp. 3780–3785.

[28] R. Bodor, A. Drenner, P. Schrater, and N. Papanikolopoulos, "Optimal camera placement for automated surveillance tasks," *Journal of Intelligent and Robotic Systems*, vol. 50, no. 3, pp. 257–295, 2007.

[29] A. O. Ercan, D. B. Yang, A. E. Gamal, and L. J. Guibas, *Distributed Computing in Sensor Systems*. Springer Berlin, 2006, ch. Optimal Placement and Selection of Camera Network Nodes for Target Localization, pp. 389 – 404.

[30] H. K. Khalil, *Nonlinear Systems*, 3rd ed. New Jersey: Prentice Hall, 2002.

[31] J. A. Sanders, F. Verhulst, and J. A. Murdock, *Averaging Methods in Nonlinear Dynamical Systems*, ser. Applied Mathematical Sciences. Springer, 2007, no. 59.

[32] T. S. Hans G. Feichtinger, Ed., *Gabor Analysis and Algorithms: Theory and Applications*. Birkhuser, 1998.

[33] Y. Tan, D. Nesic, and I. Mareels, "On non-local stability properties of extremum seeking control," *Automatica*, vol. 42, Mar. 2006.

[34] G. Dantzig, *Linear programming and extensions*, ser. Landmarks in Physics and Mathematics. Princeton University Press, 1998.

[35] J. A. Nelder and R. Mead, "A simplex method for function minimization," *Computer Journal*, vol. 7, pp. 308–313, 1965.

[36] V. Torczon, "On the convergence of the multidirectional search algorithm," *SIAM Journal on Optimization*, no. 1, 1991.