The Seventeenth International Symposium on Artificial Life and Robotics 2012 (AROB 17th '12),
B-Con Plaza, Beppu, Oita, Japan, January 19-21, 2012

Density map of attentional capacity allocation

Satoshi Kambara[1,3], Shigeyuki Oba[1,2]

[1] Graduate School of Informatics, Kyoto University, Gokasyo, Uji-city, Kyoto 611-0011, Japan
[2] PRESTO, Japan Science and Technology Agency
(Tel: 81-774-38-3938)

[3] kanbara-s@sys.i.kyoto-u.ac.jp

Abstract: Allocation of attentional capacity is an important consequence of visual attention, but its psychophysical mechanism has not been understood very well. We, in this study, investigate a procedure to estimate a high-resolution density map of the attentional capacity allocation on a visual field by analyzing a set of cognition performances on randomly located tasks. We propose a logistic regression model with multi-scale basis functions in order to achieve high-resolution density map, and an experimental scheme with different sizes of square shaped regions of attractors. Our preliminary results on two subjects showed that the corresponding shapes of attentional capacity allocation were different from those of the attractors which may reflect a hidden allocation mechanism of computational resource in brain.

Keywords: visual attention, attentional capacity, logistic regression

1 INTRODUCTION

Attention is a function of human or mammals brain, to select objective information from sensory organ. We must quickly and efficiently detect and distinguish important information in the environment such as food and predators, otherwise probability to survive can decrease. However, limited computational resource of brain does not allow us to process all of sensory input information simultaneously with high accuracy. Therefore, we concentrate computational resources to a selected set of important objects. This selection process is called *allocation of attentional capacity*. A large resource to a specific area enhances quickness and accuracy of the corresponding information processing.

Visual attention, especially, concentrates information processing resources to specific visual items; namely positions, objects, and features[11]. A large resource to a specific visual item enhances corresponding cognition performances, such as sensitivity to subtle differences in colors, directions, and shapes[8]. Several experiments had proven that the enhancement in human-subject's cognition performances in visual tasks, shortened reaction times and improvement in cognition accuracies[9].

Although we have not fully understood the mechanism of attentional capacity allocation yet, we can indirectly observe the amount of attentional resource allocation via a fluctuation of cognition performance. Thus, attentional capacity allocation is regarded as one of important topics in recent studies on visual attention. Deep relationship has been found between attentional capacity and various problems, processing of bottom-up information (ex. saliency) and top-down information (ex. intention, goal, and preferences)[1], serial and parallel search of objects with visual attention[3, 10], and problems in a relationship between visual attention and eye movement (ex. premotor theory of attention)[2, 4].

In this study, we investigate spatial resolution of attentional capacity allocation. When a human subject put an attention to a region of interest with a certain size and shape, how does the corresponding resource allocation reflect the size and shape? How can we estimate the detail shape of allocation from an observed change in cognition performance? We apply logistic regression model to represent a density map of cognition performance as a set of multi-resolution spatial bases. According to our model, spatial distribution of visual attention itself is estimated by removing background factors coming from direction of gaze effect which directly affects cognitive performances. We have conducted experiments in which subjects have to direct attention to a specific region and assign cognition tasks whose accuracy depends on attentional capacity allocation at specific location in visual field. We construct density maps of attentional capacity allocation based on the logistic regression model and investigated differences between conditions.

There have been many recent works that constructed density maps on visual field. Famous examples are saliency map[6] and Bayesian surprise[5], which predicted direction of gaze. In general, however, directions of gaze and visual attention are not always consistent, and density map of spatial attention itself should be constructed by subtracting gaze effect. There have a few attempt (ex. [7]) to estimate the attention map, however, the estimated shape of a density map is limited to several candidate patterns and did not represent general spatial patterns. We, in this study, extended these studies to investigate general shape of density map of visual attention capacity allocation with subtracting gaze effect.

2 MODEL

We applied following logistic regression model that represents spatial allocation of attentional capacity. In this model, an accuracy of a visual task is represented as a function of target position $\boldsymbol{x}$ at which the visual task appeared. The odds ratio of the accuracy is regarded as a sum of attentional capacity term and background term. The background term reflects a difference in cognition accuracy between fovea and peripheral vision that is believed to be task independent, which should be subtracted in order for the attentional capacity term to reflect the effect of task dependent allocation of attentional capacity. The parameters of each terms are determined to fit the experimental data.

2.1 Logistic Regression Model

The odds ratio of accuracy in trial j, is written as a function of the corresponding target position $\boldsymbol{x}_j$ as follows:

$$\ln\left(\frac{p(\boldsymbol{x}_j)}{1-p(\boldsymbol{x}_j)}\right) = \beta^{(0)} + \beta^{(g)} f^{(g)}(\boldsymbol{x}_j) + \sum_{k=1}^{K} \beta_k^{(a)} f_k^{(a)}(\boldsymbol{x}_j) = \boldsymbol{\beta}^T \boldsymbol{f}(\boldsymbol{x}_j) \quad (1)$$

$$f^{(g)}(\boldsymbol{x}_j) = \exp\left(-\frac{(\boldsymbol{x}_j - \boldsymbol{x}^{(\text{fix})})\Sigma^{-1}(\boldsymbol{x}_j - \boldsymbol{x}^{(\text{fix})})}{2}\right)$$

$$f_k^{(a)}(\boldsymbol{x}_j) = \begin{cases} 1 & (\text{if } \boldsymbol{x}_j \text{ in } k \text{ th grid}) \\ 0 & (\text{otherwise}) \end{cases}$$

where $\boldsymbol{x}_j$ and $\boldsymbol{x}^{(\text{fix})}$ denotes 2-d coordinates of the target position at the jth trial and the fixation point of eye-gaze in the experiment, respectively. $p(\boldsymbol{x})$ is a true probability of correct answer that is a function of the target position. $\beta^{(0)}$, $\beta^{(g)}$, and $\beta_k^{(a)}$ are parameters corresponding to bias term, fixation term, and attentional effect at the kth grid, respectively. The kth grid is defined by the basis function $f_k^{(a)}(\boldsymbol{x}_j)$ which takes value of either one or zero if a position $\boldsymbol{x}_j$ is included in the kth region of interest or not, respectively. $f^{(g)}$ denotes background effect determined as a Gauss function peaked at the fixation point $\boldsymbol{x}^{(\text{fix})}$.

In a set of experiment, we observe a corresponding set of results $t_n \in \{0, 1\}$, where $t_n = 1$ and 0 denote correct and incorrect answer at the nth task trial, respectively. The nth task links to the nth position $\boldsymbol{x}_n$. Thus, total log-likelihood of the unknown parameters based on the set of experimental results is given as,

$$L(\boldsymbol{\beta}) = \sum_{n=1}^{N} \{t_n \ln p_n + (1 \quad t_n) \ln(1 \quad p_n)\} \quad (2)$$

$$p_n = \frac{1}{1 + \exp(\quad \boldsymbol{\beta}^T \boldsymbol{f}(\boldsymbol{x}_n))}. \quad (3)$$

We estimate the parameters $\beta^{(0)}, \beta^{(g)}$, and $\beta_k^{(a)}$ that fit the observed data by maximum likelihood estimation with regularization term. Using the estimated parameters, we achieve corresponding density map of attentional capacity by $\text{MAP}(\boldsymbol{x}) = \sum_{k=1}^{K} \beta_k^{(a)} f_k^{(a)}(\boldsymbol{x})$ that reflects a level of correspondence of attentional capacity to the odds ratio of task performance with subtracting the other factors, bias and gaze effect.

2.2 Designing basis functions $f_k^{(a)}$

$f^{(a)}(\boldsymbol{x})$ in eq. (1) denotes a basis function that determines spatial extension in the visual field, where we considered multiple layers of base sets corresponding to multiple scales of unit bases. The first scale consists 20 32 grids of bases whose size are 1 1 as like in the left panel in the fig. 1. And, the second scale is made by picking all 2 2 grid in the first scale as in the right panel in the fig. 1. Similarly, we repeated the same procedure until the sixth scale which has all 6 6 size of grid in the first scale.

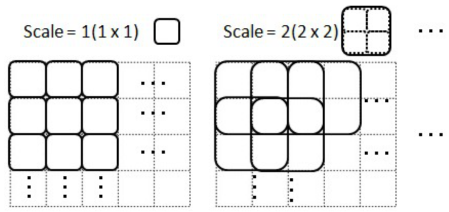

Fig. 1. Determination of multi-scale square grid. The k-th scale grid is made from all possible $k \quad k$ patches that are overlapped each other.

2.3 Ridge Regularlized Maximum Likelihood Estimation

This model is highly redundant because of overlap between the six scales, which makes it impossible to solve without regularization. We calculate parameters that maximizes likelihood function with a Ridge regularization term.

$$\hat{\boldsymbol{\beta}}_{ML \quad R(\lambda)} = \arg\max L(\boldsymbol{\beta}) \quad \lambda||\boldsymbol{\beta}||^2 \quad (4)$$

where $L()$ is the log-likelihood function and λ is a regularization parameter. We calculated $\boldsymbol{\beta}$ in (4) by Newton-Raphson method. The regularization parameter is determined in a cross-validation procedure.

3 EXPERIMENT

In order to investigate spatial allocation of attentional capacity of human subjects, we conducted the following experiment.

The experiment consists of hundreds of unit tasks for each condition. A single unit task includes the following steps.

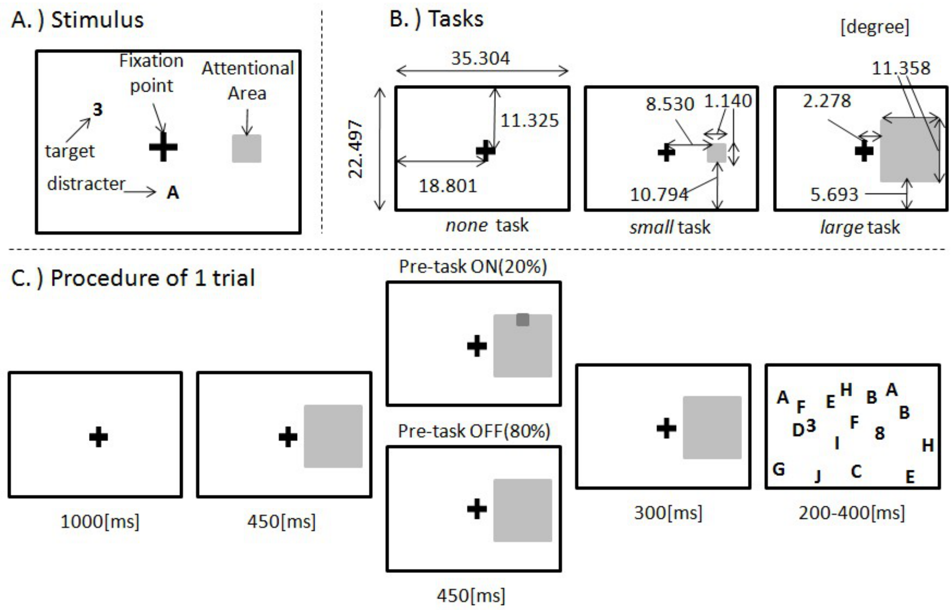

Fig. 2. A.) Stimulus in our experiment. Size of targets and distractors is 1.424 degree. B.) Position and size of pre-cue. C.) Procedure of 1 trial. *None* task do not have pre-task.

First, a pre-cue that specifies a certain region of a screen panel is shown for a subject to draw attention. Then, 10-15 alphabet and two digit letters are displayed simultaneously in a screen panel. The subject is asked to detect the two digit letters and answer them with fixing eye-gaze on the fixation point at the center of the screen. Since there are alphabet letters as distractors and all the letters disappears in a short while, the subject cannot answer correctly if the digit letters are displayed in the area where attentional capacity is not payed enough. (See Fig. 2(a), (c) for detail.)

We compared three types of pre-cue settings; they are (1) large square (2) small square, and (3) none. (See Fig 2(b) for detail.) The duration time to show the letters is set at a value from 200msec to 400msec depending on individual subject so that the task is not too easy nor difficult and the accuracy at a location takes a value from 10 to 90 percent.

In order to confirm that the subject really payed attention to the pre-cued area, a sub-task is asked; color of a small square in the pre-cued area is slightly changed with a certain probability and the subject is asked if the change occured after the main task. In order to confirm that the subjects fixed their eye gaze to the fixation point, the eye movement is observed by an eye-tracker device, EYE-LINK.

Two healthy adult male subjects conducted 360 trials of the unit task for each of three pre-cue settings with 5 minutes rest in less than every 120 trials. The resultant performance data, set of record of correct/wrong answers with the locations of digit letters, is integrated to form a density map of attentional capacity allocation in each condition.

In order to obtain the density maps of attentional capacity allocation in the *small* and *large* tasks, we treated the *none* task as an origin of common background factor. And, since there were statistical variance in the estimated background factor, we estimated a null standard deviation σ of odds ratio of the *none* task by calculating 4-fold cross-validation. We normalized the estimated odds ratio of the *small* and *large* tasks by using the estimated mean and standard deviation of the common background factor. Regularization factor was set at $\lambda = 400$ at which cross-validation likelihood was maximized.

$$\ln\left(\frac{p(\boldsymbol{x}_j)}{1 \quad p(\boldsymbol{x}_j)}\right) = {}^{(0)} + \sum_{k=1}^{K} {}^{(a)}_{k} f_k^{(a)}(\boldsymbol{x}_j) + \mathrm{MAP}^{none}(\boldsymbol{x}_j) \quad (5)$$

$\mathrm{MAP}^{none}(\boldsymbol{x}_j)$: $\mathrm{MAP}(\boldsymbol{x})$ learned by data of *none* task.

The result is shown in Fig. 3. Strong allocation of attentional capacity was observed around the center of pre-cued attractor region for both of small and large square tasks. But, detail size and shape of the allocated regions were different between the two conditions. Attentional capacity allocation in small task was concentrated at smaller area than those in large square task, which suggested that total resource of the subject was limited and had to distribute it when a large area was attracted. In detail, however, the size of the region of allocated attentional capacity did not seem proportional to those of the corresponding region of attractors. The shape were also different between the attention allocation and the region of attractors; in small task, significant enhancement of attentional capacity was observed even in the regions around the small attractors; in large task, the enhanced region did not fill all the corners of the large square shaped region of attractors. These differences suggested a kind of limitation in spatial resolution of attentional capacity allocation.

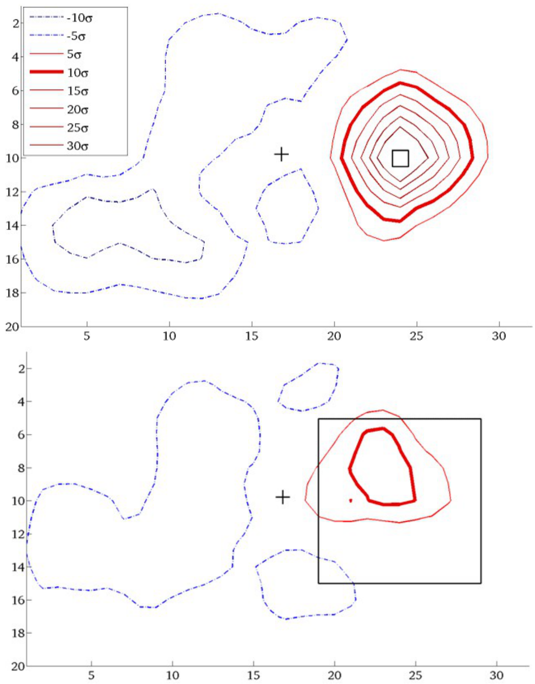

Fig. 3. Estimated density maps of *small* and *large* tasks are shown in top and bottom panels, respectively. Maps of a single subject in two subjects are shown. Contour lines denotes normalized odds ratios whose values are normalized by an estimated standard deviation σ of *none* task. Black square indicates pre-cued region of attention attractor. Black cross indicates fixation point.

4 DISCUSSION

We have constructed spatial density map of attentional capacity allocation by using a logistic regression model with multi-scale basis functions. Based on this model, we have investigated spatial resolution of the attentional capacity allocation. For different settings of pre-cued region of attractors, small and large squares, we found different results in size and strength of attentional capacity allocation, however, the shape of the attentional capacity did not obviously reflect the square shape of pre-cued region of attractor, which suggested a limited spatial resolution of attentional capacity allocation.

Our final goal is to understand mechanism of allocation of attentional capacity. We will increase the number of subjects and unit tasks of experiment, and establish the model of allocation of attentional capacity. In this paper, we have confirmed that allocation of attentional capacity is done as the shape of area where subjects attend and added some modification. In the future, we aim to discover the factors and mechanism related to the model of allocation of attentional capacity.

In addition, we will improve logistic regression model. For example, we change the designing basis functions $f_k^{(a)}$ to that has more biological validity. Furthermore, we add the factors which affect cognition performance(ex. saliency, Bayesian surprise) to our model in addition to attentional term and fixation term. By these improvement, we aim to represent density of attentional capacity more accurately.

REFERENCES

[1] Belopolsky AV, Zwaan L, Theeuwes J, et al (2007), The size of an attentional window modulates attentional capture by color singletons, Psychonomic Bulletin and Review 16(5):934-938.

[2] Hunt AR, Kingstone A (2003), Covert and overt voluntary attention: Linked or independent?, Acognitive Brain Research 18(1):102-105.

[3] Awh E, Pashler H (2000), Evidence for split attentional foci, Journal of Experimental Psychology Human Perception and Performance 26(2):834-846.

[4] Klein RM (1980), Does Oculomotor Readiness Mediate Cognitive Control Of Visual Attention?, Attention and performance 8:259-276.

[5] Itti L, Baldi P (2009), Bayesian Surprise Attracts Human Attention, Vision Research 49(10):1295-1306.

[6] Itti L, Koch C, Niebur E (1998), A Model of Saliency-Based Visual Attention for Rapid Scene Analysis, Pattern Analysis and Machine Intelligence, IEEE Transactions on 20(11):1254-1259 .

[7] Gobell JL, Tseng C, Sperling G (2004), The Spatial Distribution of Visual Attention, Vision Research 44(12):1273-1296.

[8] Huang L, Pashler H (2005), Attention Capacity and Task Difficulty in Visual Search, Cognition 94(3):B101-B109.

[9] Carrasco M, Ling S, Read S (2004), Attention Alters Appearance , Nature Neuroscience 7(3), 308-313.

[10] VanRullen R, Carlson T, Cavanagh P (2007), The Blinking Spotlight of Attention, Proceedings of the National Academy of Sciences 104(49):19204-19209.

[11] Luck SJ, Hillyard SA, Mouloua M, et al (1996), Mechanisms of Visual-spatial Attention: Resource Allocation or Uncertainty Reduciton?, Journal of Experimental Psychology: Human Perception and Performance 22(3):725-737.