

Simulating Human Saccadic Scanpaths on Natural Images

¹N. E. ... ³ U ... ⁴D. M. P. S. ... M. E. P. U. ... A. S. ... P. U. ... S. P. U. ...

wwang@jdl.ac.cn, {chencheng880829, yizhou.wang, ttjiang, ffang, yuany}@pku.edu.cn

Abstract

Human saccade is a dynamic process of information pursuit. Based on the principle of information maximization, we propose a computational model to simulate human saccadic scanpaths on natural images. The model integrates three related factors as driven forces to guide eye movements sequentially: reference sensory responses, fovea-periphery resolution discrepancy, and visual working memory. For each eye movement, we compute three multi-band filter response maps as a coherent representation for the three factors. The three filter response maps are combined into multi-band residual filter response maps, on which we compute residual perceptual information (RPI) at each location. The RPI map is a dynamic saliency map varying along with eye movements. The next fixation is selected as the location with the maximal RPI value. On a natural image dataset, we compare the saccadic scanpaths generated by the proposed model and several other visual saliency-based models against human eye movement data. Experimental results demonstrate that the proposed model achieves the best prediction accuracy on both static fixation locations and dynamic scanpaths.

1. Introduction

In this paper, we propose a computational model to simulate human saccadic scanpaths on natural images. The model integrates three related factors as driven forces to guide eye movements sequentially: reference sensory responses, fovea-periphery resolution discrepancy, and visual working memory. For each eye movement, we compute three multi-band filter response maps as a coherent representation for the three factors. The three filter response maps are combined into multi-band residual filter response maps, on which we compute residual perceptual information (RPI) at each location. The RPI map is a dynamic saliency map varying along with eye movements. The next fixation is selected as the location with the maximal RPI value. On a natural image dataset, we compare the saccadic scanpaths generated by the proposed model and several other visual saliency-based models against human eye movement data. Experimental results demonstrate that the proposed model achieves the best prediction accuracy on both static fixation locations and dynamic scanpaths.

Human saccade is a dynamic process of information pursuit. Based on the principle of information maximization, we propose a computational model to simulate human saccadic scanpaths on natural images. The model integrates three related factors as driven forces to guide eye movements sequentially: reference sensory responses, fovea-periphery resolution discrepancy, and visual working memory. For each eye movement, we compute three multi-band filter response maps as a coherent representation for the three factors. The three filter response maps are combined into multi-band residual filter response maps, on which we compute residual perceptual information (RPI) at each location. The RPI map is a dynamic saliency map varying along with eye movements. The next fixation is selected as the location with the maximal RPI value. On a natural image dataset, we compare the saccadic scanpaths generated by the proposed model and several other visual saliency-based models against human eye movement data. Experimental results demonstrate that the proposed model achieves the best prediction accuracy on both static fixation locations and dynamic scanpaths.

Proposed method

Figure 1 shows the proposed method. It consists of three main steps: (1) Computing three multi-band filter response maps for reference sensory responses, fovea-periphery resolution discrepancy, and visual working memory. (2) Combining these three maps into multi-band residual filter response maps. (3) Computing residual perceptual information (RPI) at each location and selecting the next fixation as the location with the maximal RPI value.

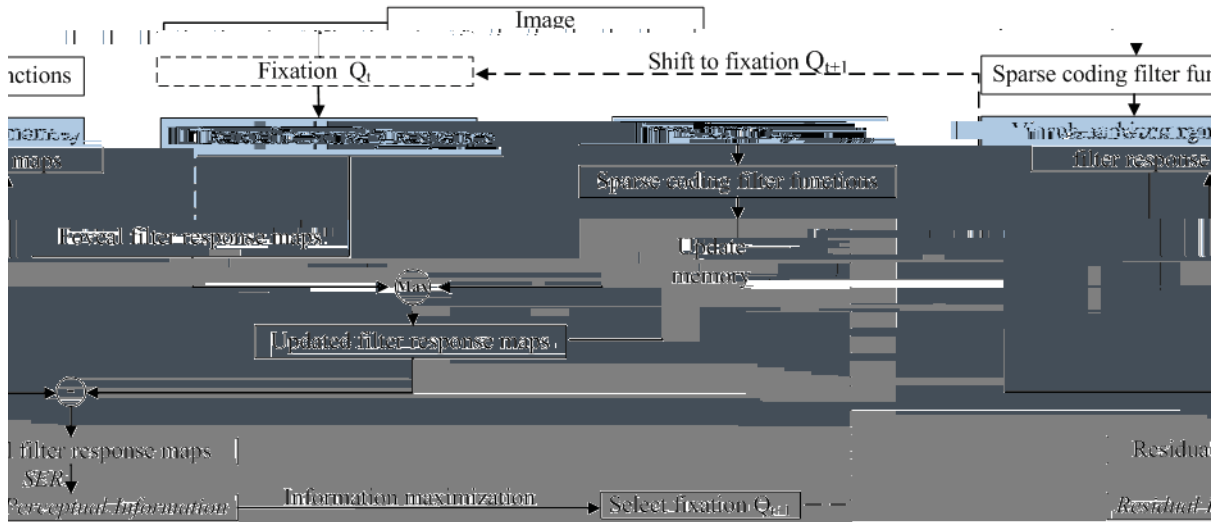


FIG. 1.

1.1. Related work

The process of selecting a fixation point is a complex task that involves understanding the visual information in an image. This process is often modeled using information theory, where the goal is to maximize the amount of information gained from each fixation.

In the context of visual search, the amount of information gained from a fixation is often measured using the Site Entropy Rate (SER) and the Residual Perceptual Information (RPI). The SER is a measure of the entropy of the filter response maps, and the RPI is a measure of the information that remains after the filter response maps have been processed.

The process of selecting a fixation point is often modeled as an iterative process. In each iteration, the filter response maps are calculated for a given fixation point, and the SER and RPI are calculated. The fixation point that maximizes the SER and RPI is selected for the next iteration.

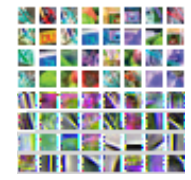
This process is often modeled using information maximization, where the goal is to maximize the amount of information gained from each fixation. This process is often modeled using the multi-band filter response maps, which are used to calculate the SER and RPI.

The process of selecting a fixation point is often modeled using the multi-band filter response maps, which are used to calculate the SER and RPI. The process is often modeled using information maximization, where the goal is to maximize the amount of information gained from each fixation.

1. H. S. 3, dynamic C. 1. I. R. et al. 22. N. H. et al. 14. et al. 1. F. 1. et al. 2. I. S. 3. F. S. 4.

2.1.1 Sparse coding filters

S. 3. I. 21. multi-band filter response maps C. A. ICA. 1. S. fi, I. 12. 1. 4. F. 2.



F. 2. 4

2.1.2 Foveal imaging

P. 12. F. 11. S. A. F. 3.

2. Our Approach

I. F. 1.

2.1. Coherent representation of three factors

fi. fi.



F. 3. A. F. fi. O.

2.1.3 Visual working memory

On each trial, the visual working memory is updated with a new set of visual stimuli. The visual stimuli are represented by a set of feature maps f_k^v and f_k^w . The visual stimuli are processed by a set of residual filters r_k to produce a set of residual filter response maps f_k^o . The residual filter response maps are then used to compute the residual filter response maps f_k^w .

Simulating the forgetting properties.

The forgetting properties are simulated by a set of parameters ϵ and δ . The parameters ϵ and δ are defined as follows:

$$\epsilon = \frac{1}{S} \sum_k f_k^o \quad (3.4)$$

Updating visual working memory.

The visual working memory is updated by a set of parameters α and β . The parameters α and β are defined as follows:

$$f_k^w(x, y, t) \leftarrow \alpha f_k^v(x, y, t) + \beta f_k^w(x, y, t-1) \quad (4)$$

The visual working memory is updated by a set of parameters α and β . The parameters α and β are defined as follows:

$$f_k^w(x, y, t) \leftarrow \alpha f_k^v(x, y, t) + \beta f_k^w(x, y, t-1) \quad (1)$$

Computing residual filter response maps.

The residual filter response maps are computed by a set of parameters r_k . The parameters r_k are defined as follows:

$$r_k = |f_k^o - f_k^w|$$

2.2. Measuring residual perceptual information

The residual perceptual information is measured by a set of parameters S_i and S_j . The parameters S_i and S_j are defined as follows:

$$S_i = \sum_k SER_{ki} - \sum_k \pi_{ki} \sum_j P_{kij} \quad (2)$$

The residual perceptual information is measured by a set of parameters S_i and S_j . The parameters S_i and S_j are defined as follows:

$$S_i = \sum_k SER_{ki} - \sum_k \pi_{ki} \sum_j P_{kij} \quad (2)$$

The residual perceptual information is measured by a set of parameters S_i and S_j . The parameters S_i and S_j are defined as follows:

The residual perceptual information is measured by a set of parameters S_i and S_j . The parameters S_i and S_j are defined as follows:

2.3. Saccadic amplitude

The saccadic amplitude is measured by a set of parameters Q_t and Q_{t+1} . The parameters Q_t and Q_{t+1} are defined as follows:

$$Q_t = \sum_k f_k^o \quad (4)$$

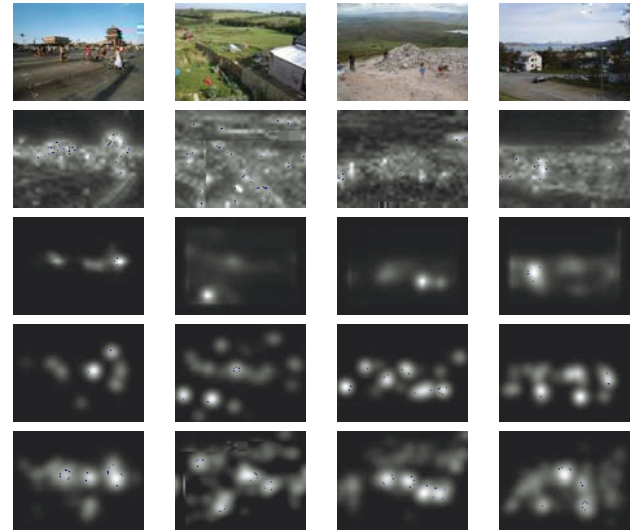
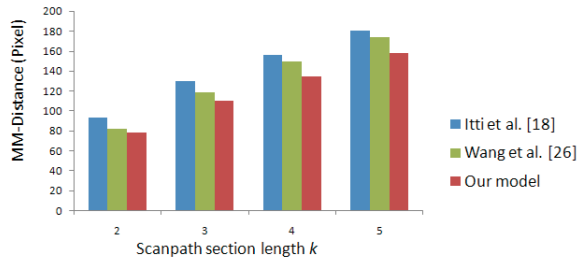
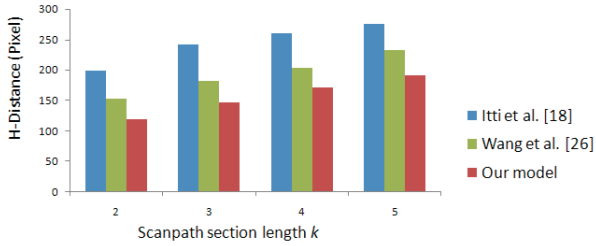


Figure 1. Comparison of H-Distance and MM-Distance for different scanpath sections.

Figure 1. Comparison of H-Distance and MM-Distance for different scanpath sections. The top row shows the original scanpath sections, the middle row shows the H-Distance heatmaps, and the bottom row shows the MM-Distance heatmaps. The columns represent different scanpath sections (1, 2, 3, 4).

3.2.1 Distance of scanpaths

In this section, we define the distance of scanpaths. We use the time-delay embedding, $C_m^k(t)$, to represent the scanpath section of length k . The time-delay embedding is defined as $C_m^k(t) = (c_m(t), \dots, c_m(t+k-1))$. The time-delay embedding is a vector in \mathbb{R}^k . The time-delay embedding is defined as $C_m^k(t) = (c_m(t), \dots, c_m(t+k-1))$. The time-delay embedding is a vector in \mathbb{R}^k . The time-delay embedding is defined as $C_m^k(t) = (c_m(t), \dots, c_m(t+k-1))$. The time-delay embedding is a vector in \mathbb{R}^k .

Let $X = \{C_m^k(t)\}_t \subseteq \mathbb{R}^k$ and $Y = \{C_h^k(\tau)\}_\tau \subseteq \mathbb{R}^k$ be two sets of scanpath sections. The distance between X and Y is defined as $d_k(X, Y) = \frac{1}{|X|} \sum_{x \in X} d_k(x, Y)$.

For a scanpath section $x = (x_1, \dots, x_k) \in \mathbb{R}^k$, the distance between x and Y is defined as $d_k(x, Y) = \frac{1}{|Y|} \sum_{y \in Y} \frac{\|x - y\|_2}{k}$. The distance between x and Y is defined as $d_k(x, Y) = \frac{1}{|Y|} \sum_{y \in Y} \frac{\|x - y\|_2}{k}$.

The distance between X and Y is defined as $d_k(X, Y) = \frac{1}{|X|} \sum_{x \in X} d_k(x, Y)$. The distance between X and Y is defined as $d_k(X, Y) = \frac{1}{|X|} \sum_{x \in X} d_k(x, Y)$.

Figure 1. Comparison of H-Distance and MM-Distance for different scanpath sections. The top row shows the original scanpath sections, the middle row shows the H-Distance heatmaps, and the bottom row shows the MM-Distance heatmaps. The columns represent different scanpath sections (1, 2, 3, 4).

The H-Distance is defined as $d_H^k = \frac{1}{|X|} \sum_{x \in X} \frac{\|x - C_h^k(\tau)\|_2}{k}$. The H-Distance is defined as $d_H^k = \frac{1}{|X|} \sum_{x \in X} \frac{\|x - C_h^k(\tau)\|_2}{k}$.

$$d_H^k = \frac{1}{|X|} \sum_{x \in X} \frac{\|x - C_h^k(\tau)\|_2}{k} \quad (3)$$

$$d_H^k = \frac{1}{|X|} \sum_{x \in X} \frac{\|x - C_h^k(\tau)\|_2}{k} \quad (4)$$

The mean minimal distance (MMD) is defined as $d_M^k = \frac{1}{|X|} \sum_{x \in X} d_k(x, Y)$. The mean minimal distance (MMD) is defined as $d_M^k = \frac{1}{|X|} \sum_{x \in X} d_k(x, Y)$.

The MMD is defined as $d_M^k = \frac{1}{|X|} \sum_{x \in X} d_k(x, Y)$. The MMD is defined as $d_M^k = \frac{1}{|X|} \sum_{x \in X} d_k(x, Y)$.

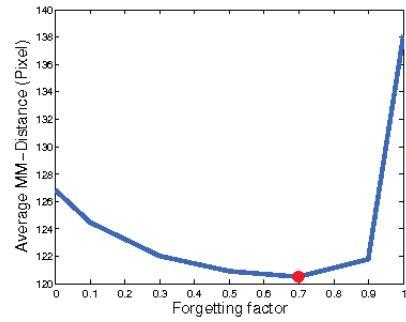
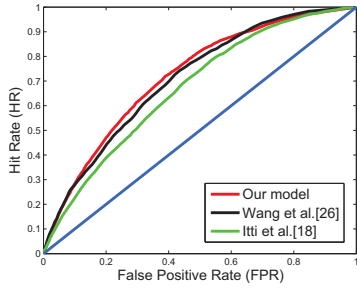
The MMD is defined as $d_M^k = \frac{1}{|X|} \sum_{x \in X} d_k(x, Y)$. The MMD is defined as $d_M^k = \frac{1}{|X|} \sum_{x \in X} d_k(x, Y)$.

The MMD is defined as $d_M^k = \frac{1}{|X|} \sum_{x \in X} d_k(x, Y)$. The MMD is defined as $d_M^k = \frac{1}{|X|} \sum_{x \in X} d_k(x, Y)$.

The MMD is defined as $d_M^k = \frac{1}{|X|} \sum_{x \in X} d_k(x, Y)$. The MMD is defined as $d_M^k = \frac{1}{|X|} \sum_{x \in X} d_k(x, Y)$.

The MMD is defined as $d_M^k = \frac{1}{|X|} \sum_{x \in X} d_k(x, Y)$. The MMD is defined as $d_M^k = \frac{1}{|X|} \sum_{x \in X} d_k(x, Y)$.

The MMD is defined as $d_M^k = \frac{1}{|X|} \sum_{x \in X} d_k(x, Y)$. The MMD is defined as $d_M^k = \frac{1}{|X|} \sum_{x \in X} d_k(x, Y)$.



F... ROC
2... 1...

F... 1... k
ε. ε = 0.7

1. ROC

ROC	I et al. [1]	et al. [2]	O.
ROC		1	1 3

F... ~
ROC... ROC
F... 1. / ROC... ROC
fi... fi
ROC... fi
F... fi
1... 2...

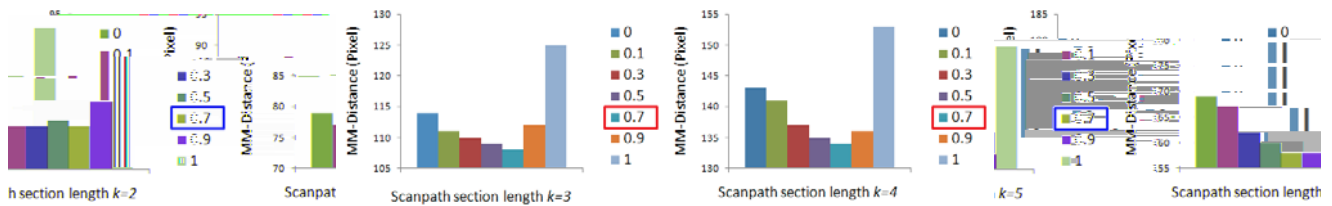
fi... F...
ε...
1... 22...

3.4. Assessment of the forgetting factor

1... ε... F...
N... ε...
H... k... ε...
F... 1... ε...
D... ε... ε...
A...
I... ε... ε...
C...
Θ

4. Conclusion, Discussion and Future Work

I...
fi... fi
fi... fi
E...
F... 2...
reference sensory responses
A...
24...



F. . . A . . .

€ . . .

I . . . , . . . edit distance

. F . . . U . . .

. I
 I
 M
 fi I

. journal of Vi-
 sion, 2
 1. D. . . , M . . . , N. . .
 NIPS, 2
 11. P . . . A SPIE
 Proceedings: Human Vision and Electronic Imaging, 1
 12. P . . . R
 ACM Symposium on Eye Tracking Research &
 Applications, 2
 13. , H. . . , D. R . . . R CVPR, 2
 14. . H M. / . . . R

Acknowledgments

. N S F
 C N 2 12 2
 N R 3 P C
 N 2 C / 32 4

. journal of Vision, 2 1
 1. . H . . . , C . . . , P . P
 NIPS, 2
 1. . H . . . A S I
 fi Proc. R. Soc. Lond. B, 1
 1. N. H S A
 X Computer Vision and Pattern Recognition, 2
 1. I . . . , C . . . , E. N A
 IEEE PAMI, 1
 1. S A
 Advanced in Neural
 Information Processing System, 1
 2. . I . . . P. / . . . /
 NIPS, 2
 21. / . . . O D. F E
 fi
 Nature, 1
 22. R , P. , C
 ?
 journal of
 Vision, 2
 23. . S , M. C E
 journal
 of Statistical Physics 65: 579 16, 1 1
 24. E. S / O N

 Annual Review of Neuroscience, 2 1
 2. F P
 Nature
 Reviews Neuroscience, 2 3
 2. , , H , M

 CVPR, 2 1

References

1. R. A . . . , S. H . . . , F. E . . . , S. S
 F Computer Vision
 and Pattern Recognition, 2
 2. A. / . . . , D. A . . . , . . . S . . . M
 1 In-
 vestigative Ophthalmology, 1
 3. H. / . . . U
 Neural Computation,
 1
 4. C. / . . . , / . . . R . . . , . . . R
 ?
 journal of Neuroscience, 2
 N. / S
 NIPS, 2
 / /
 journal
 of Neuroscience, 2
 M. C . . . , C. P . . . , M. E . . . S

 Vision Research, 2
 A. C Annual Review of Psychology, 1 1