

Simulating Human Saccadic Scanpaths on Natural Images

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

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

The proposed method is based on the principle of information maximization. It 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.

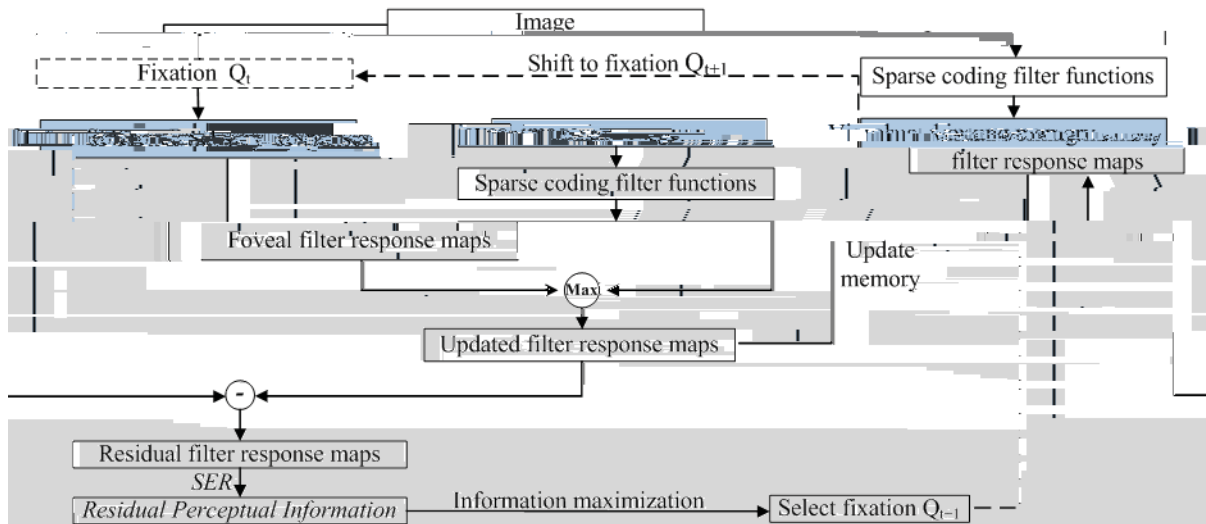


Fig. 1.

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 information gained from the image.

In the context of visual search, the process of selecting a fixation point can be seen as a search for the most informative location in the image. This is often done by calculating the entropy of the image at different locations and selecting the location with the highest entropy.

The Site Entropy Rate (SER) is a measure of the information gained from a fixation point. It is defined as the residual perceptual information (RPI) after accounting for the information gained from the previous fixation point.

The RPI is calculated as the difference between the total information in the image and the information gained from the previous fixation point. This is often done using a multi-band filter response map.

The process of selecting a fixation point is often iterative, where the next fixation point is selected based on the RPI of the current fixation point. This process is often used in applications such as visual search and image compression.

The process of selecting a fixation point is often done using a multi-band filter response map. This map is used to calculate the RPI for each location in the image.

The process of selecting a fixation point is often done using a multi-band filter response map. This map is used to calculate the RPI for each location in the image.

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. Our Approach

I. F. 1.

2.1. Coherent representation of three factors

fi. fi.

2.1.1 Sparse coding filters

S. 3. I. 21. multi-band filter response maps C. A. ICA. 1. S. fi. I. 12. 8 × 8 × 3. fi. 4. F. 2.



F. 2. 4

2.1.2 Foveal imaging

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



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

2.1.3 Visual working memory

Visual working memory is a system that temporarily holds and manipulates visual information. It is often described as a 'scratchpad' for visual data. The capacity of visual working memory is limited, typically to about 3-4 items. This is often measured using the 'change blindness' paradigm, where subjects are asked to detect changes in a scene. The number of items that can be held in working memory is often referred to as the 'magical number 7 ± 2'.

Simulating the forgetting properties.

The forgetting properties of visual working memory can be simulated using a decay constant ϵ . The decay constant is a value between 0 and 1, where 0 represents no forgetting and 1 represents complete forgetting. The decay constant is often estimated from experimental data using a method like the one described in [3, 4].

Updating visual working memory.

Visual working memory is updated as new information is presented. The update process involves comparing the new information with the current state of working memory. The update is often modeled using a maximum function, where the new information is only added if it is different from the current state. This is often referred to as the 'maximum function' or 'max' function.

$$f_k^w(x, y, t) \leftarrow \max(f_k^v(x, y, t), \epsilon \cdot f_k^w(x, y, t - 1)). \quad (1)$$

Computing residual filter response maps.

The residual filter response maps are computed by comparing the original image with the current state of working memory. The residual is the difference between the original image and the current state of working memory. This is often referred to as the 'residual filter response map' or 'RFRM'. The residual is computed for each pixel in the image, resulting in a residual map of the same size as the original image.

2.2. Measuring residual perceptual information

Residual perceptual information (RPI) is a measure of the information that is not captured by the current state of working memory. It is often measured using the Site Entropy Rate (SER) method. The SER is a measure of the entropy of the residual filter response map. The SER is computed by summing the entropy of the residual filter response map over all sites in the image. The SER is often used to measure the amount of information that is lost when working memory is updated.

The Site Entropy Rate (SER) is a measure of the entropy of the residual filter response map. It is computed by summing the entropy of the residual filter response map over all sites in the image. The SER is often used to measure the amount of information that is lost when working memory is updated.

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

The Site Entropy Rate (SER) is a measure of the entropy of the residual filter response map. It is computed by summing the entropy of the residual filter response map over all sites in the image. The SER is often used to measure the amount of information that is lost when working memory is updated.

2.3. Saccadic amplitude

Saccadic amplitude is a measure of the size of saccades. It is often measured using the 'saccadic amplitude' method. The saccadic amplitude is the angle between the starting and ending points of a saccade. The saccadic amplitude is often used to measure the amount of information that is lost when working memory is updated.

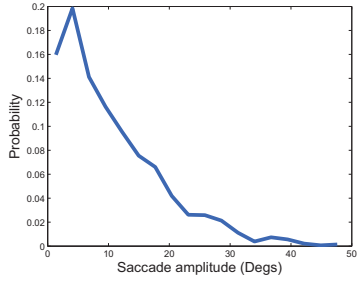


Figure 4.

$$Q_{t+1} \sim N(\mu, \sigma^2), \quad p(z \leq Z/2),$$

Figure 4.

A fixation Q_{t+1} ,

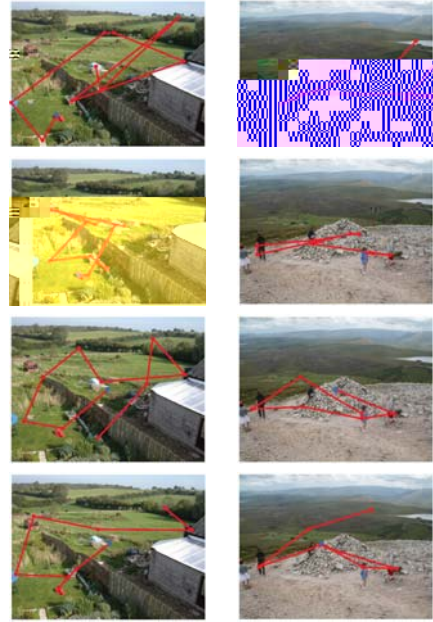


Figure 4.C

3. Experimental Results

..., 2, ..., ,
 ..., fi
 . O
 fi

3.1. Dataset and eye movement data collection

I 2
 24 . SMI
 H . S.
 3 21 CR
 3 1 . A
 fi fi

3.2. Evaluation of fixation order

. A
 1 .
 R et al. 22.
 . N , I et al. 1 .

..., 2, ..., ,
 ..., fi
 A I R
 H , I et
 . M ,
 A I R 1 ,
 1 .
 fi
 , fi
 fi
 8 ~ 10 fi
 , I et
 al. 1 , 2 ,
 . R . S .
 . A F . .

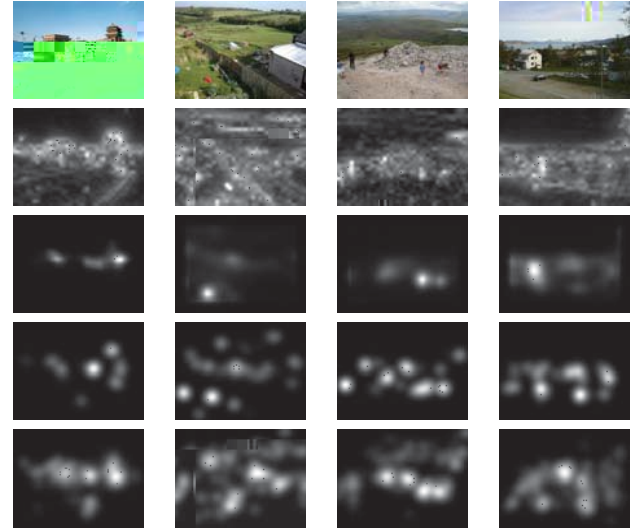
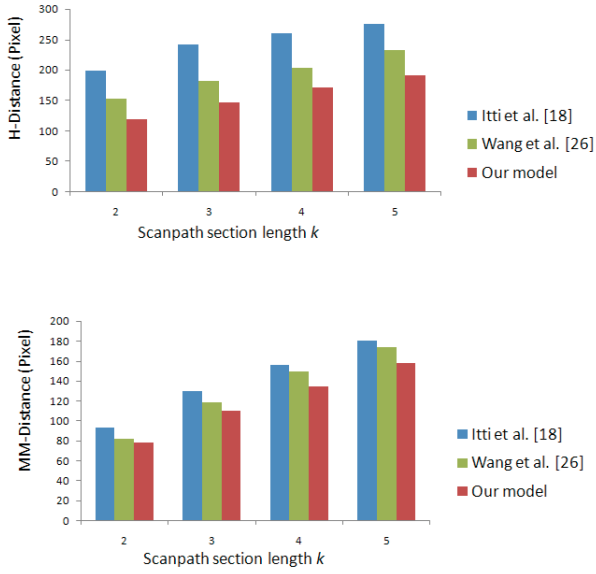


Figure 1: Comparison of scanpath visualizations. The top row shows original images with scanpaths overlaid. The second row shows grayscale versions of the scanpaths. The third and fourth rows show binary representations of the scanpaths, with the bottom row showing a more refined binary representation.

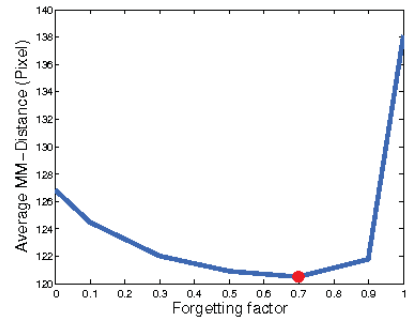
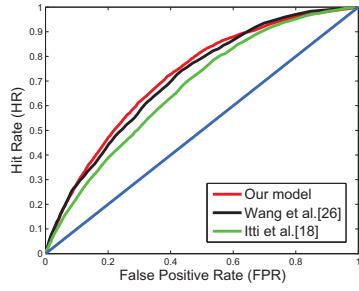
Figure 1: Comparison of scanpath visualizations. The top row shows original images with scanpaths overlaid. The second row shows grayscale versions of the scanpaths. The third and fourth rows show binary representations of the scanpaths, with the bottom row showing a more refined binary representation.

3.2.1 Distance of scanpaths

In this section, we define the distance of scanpaths. We use the *time-delay embedding*, $C_m^k(t) = (c_m(t), \dots, c_m(t+k-1))$ to represent the scanpath m at time t . The set of all scanpaths is $X = \{C_m^k(t)\}_t \subseteq \mathbb{R}^k$. Similarly, the set of all scanpaths is $Y = \{C_h^k(\tau)\}_\tau$. We define the distance between two scanpaths x and y as $d_k(x, Y) = \min_{\tau} \{\|x - C_h^k(\tau)\|_2\} / k$. The distance between two scanpaths x and y is $d_k(x, Y)$.

Figure 2: Comparison of scanpath visualizations. The top row shows original images with scanpaths overlaid. The second row shows grayscale versions of the scanpaths. The third and fourth rows show binary representations of the scanpaths, with the bottom row showing a more refined binary representation.

Figure 2: Comparison of scanpath visualizations. The top row shows original images with scanpaths overlaid. The second row shows grayscale versions of the scanpaths. The third and fourth rows show binary representations of the scanpaths, with the bottom row showing a more refined binary representation.



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

F... 1... k
ε. ε = 0.7

1. ROC

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

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

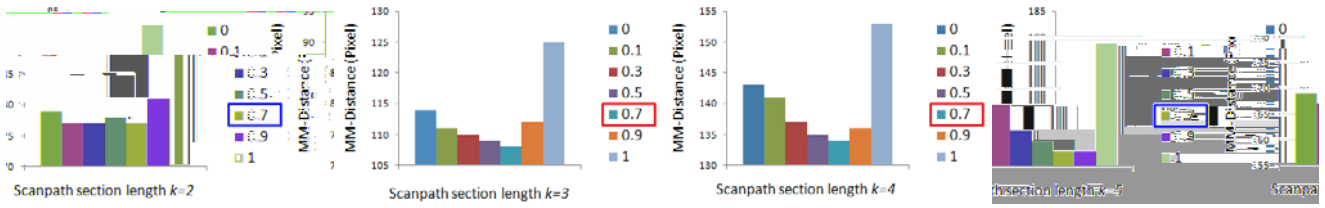
fi
ε = 0.7
F
1... 22...

3.4. Assessment of the forgetting factor

1... ε... F...
N... ε...
H... k... ε.
F... 1... ε = 0.7,
D... ε... ε = 0
A...
I... ε... ε = 1
C
Θ

4. Conclusion, Discussion and Future Work

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



F. . . A . . .

ε . . .

I . . . , . . . edit distance

. . . F . . . U . . .

. I I M I

. Journal of Vision, 2 D. . . , M . . . , N. NIPS, 2 P . . . A SPIE Proceedings: Human Vision and Electronic Imaging, 1 P . . . R ACM Symposium on Eye Tracking Research & Applications, 2 H. . . , . H. . . , D. R . . R CVPR, 2 M. / . . R Journal of Vision, 2 1 H . . . , C. . . , P. P NIPS, 2 H . . . A. . . S . . I Proc. R. Soc. ond. B, 1 N. H S . . . A Computer Vision and Pattern Recognition, 2 I . . . , C. . . , E. N . . . A IEEE PAMI, 1 S. . . A Advanced in Neural Information Processing System, 1 I . . . P. / . . / NIPS, 2 O . . . D. F . . E Nature, 1 R . . . , P. . . , . C Journal of Vision, 2 S . . . , . . . M. C . . . E Journal of Statistical Physics 65: 579 16, 1 1 E. S . . . O . . . N Annual Review of Neuroscience, 2 1 F. . . P Nature Reviews Neuroscience, 2 3 , . . . , H . . . , . M CVPR, 2 1

Acknowledgments

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

References

1. R. A . . . , S. H . . . , F. E . . . , S. S Computer Vision and Pattern Recognition, 2
2. A. / . . . , D. A . . . , . . . S . . . M Investigative 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

1. . . . H . . . , C. . . , P. P NIPS, 2 H . . . A. . . S . . I Proc. R. Soc. ond. B, 1 N. H S . . . A Computer Vision and Pattern Recognition, 2 I . . . , C. . . , E. N . . . A IEEE PAMI, 1 S. . . A Advanced in Neural Information Processing System, 1 I . . . P. / . . / NIPS, 2 O . . . D. F . . E Nature, 1 R . . . , P. . . , . C Journal of Vision, 2 S . . . , . . . M. C . . . E Journal of Statistical Physics 65: 579 16, 1 1 E. S . . . O . . . N Annual Review of Neuroscience, 2 1 F. . . P Nature Reviews Neuroscience, 2 3 , . . . , H . . . , . M CVPR, 2 1