Simulating Human Saccadic Scanpaths on Natural Images

[1,3], C C [1], [1,2]
[1,2], F F [2,4], [2],
[1]N E . V , [2] . M P M E , P U
[3] . U , C A S , 1 4 , C
[4]D P , P U , 1 1, C
S M S , P U , 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 maximi ation, we propose a computational model to simulate human saccadic scanpaths on natural images. he model integrates three related factors as driven forces to guide eye movements se uentially 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. he three filter response maps are combined into multi-band residual filter response maps, on which we compute residual perceptual information (RPI) at each location. he RPI map is a dynamic saliency map varying along with eye movements. he 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.

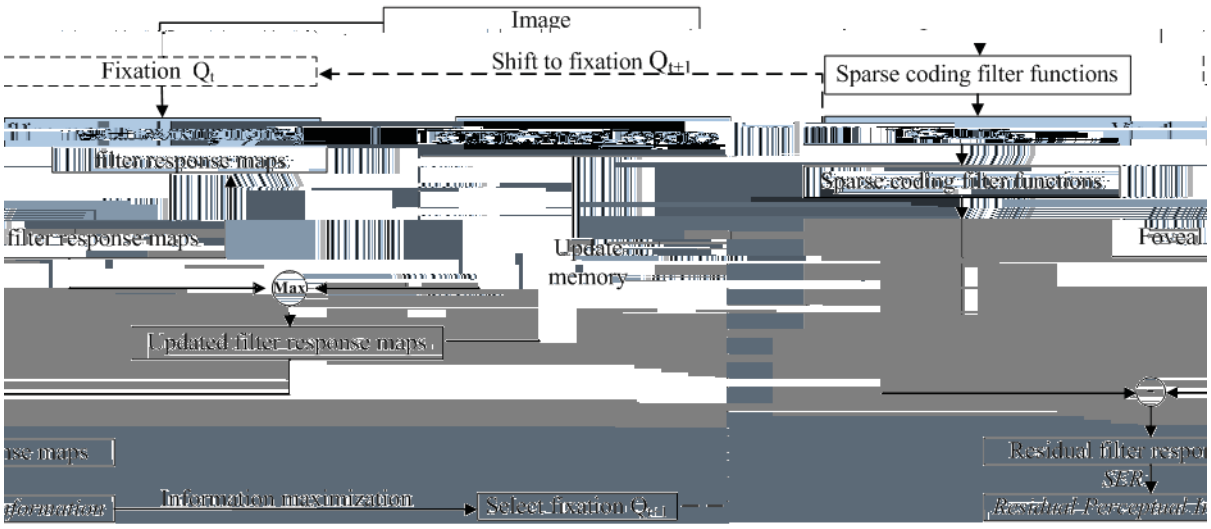
1. Introduction

I . . , .
F . fi
. . O , .
fi
. S fi

. H.
. I ,
fi ,
fi . I ,
I ,
. I
fi
, ,

Proposed method

4
, .
F . 1 .
fi
. N .
fi .
V1
21 .
fi , .
. fi . I F . 1,
fi Q_t fi



F . 1. .

V . .

fi .

, fi

ϵ $0 \leq \epsilon \leq 1$.

fi

fi

fi *so far*

fi

fi .

. C ,

Site Entropy Rate 2 . fi

residual perceptual information (RPI) .

RPI

. I

SER fi Q_{t+1},

A , *multi-band filter response maps* ,

.

1.1. Related work

I , .

fi

. I *et al.* 1

. 1

, ,

, *et al.* 1

. I

fi ,

. C

, *et al.* *self-information*

. H *et al.* 1

M

et al. 13 1

, H . *et al.* 1

. A *et al.* 1

. I

et al. 2 fi

fi ,

fi .
1
. H ,
1 , fi
fi
S 3,
.
dynamic
. C ,
. 1
, ,
. I
, R *et al.* 22 .
. N ,
. H *et al.* 14
,
1 . F .
et al. fi fi
,
fi
1
. *et al.* 2
1 .
. I S
2, .
fi
fi 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 .
, **V**1
. 3 . .
,
fi 21 .
I , fi
multi-band filter response maps ,
, . S fi , I
C A ICA 1
1 2 fi 12 ,
$8 \times 8 \times 3$,
1 . , fi
fi
. 4
F . 2.

F . 2. 4 .

2.1.2 Foveal imaging

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

F . 3. A . O
F fi .

2.1.3 Visual working memory

O , fi . . V . , , , . fi . I fi . R . .

Simulating the forgetting properties. I . , . . fi ϵ $0 \leq \epsilon \leq 1$. I $\epsilon = 1$, . I $\epsilon = 0$, . I S 3.4, . . ϵ. .

Updating visual working memory. V . . fi . I . , . fi . A *Max* . . 4 . fi , $f_k^v(x,y,t)$ $f_k^w(x,y,t)$ k fi t (x,y) .

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

Computing residual filter response maps. fi F . 1 . P . A . , . fi . fi $r_k = |f_k^o - f_k^w|$, f_k^o k .

2.2. Measuring residual perceptual information

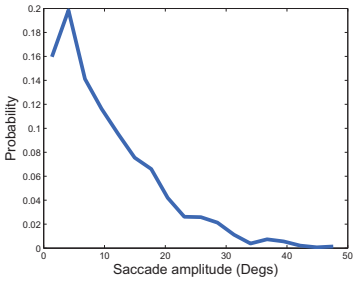
F , . fi . I . , . *Site Entropy Rate* SER 2 . fi . *Site Entropy Rate* fi . R . I , . fi . SER . . SER , SER i

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

π_{ki} i k fi , P_{kij} i j k fi . A 2 , SER . , . P 2 . SER . SER . . F , SER . . fi . M , fi , *e.g.*, , . . S , . . SER . . , . . .

2.3. Saccadic amplitude

U . , . F . 4. I . 90% . 20° . . fi . , , fi Q_{t+1} . fi Q_t. F , $Z \times Z$. fi Q_t , . $Z/2$. 4 20° . . SER .



F . 4. . . .

Q_{t+1}. N , Q_{t+1} $p(z \leq Z/2)$,

F . 4.
A fi Q_{t+1}, .

3. Experimental Results

, .
, .

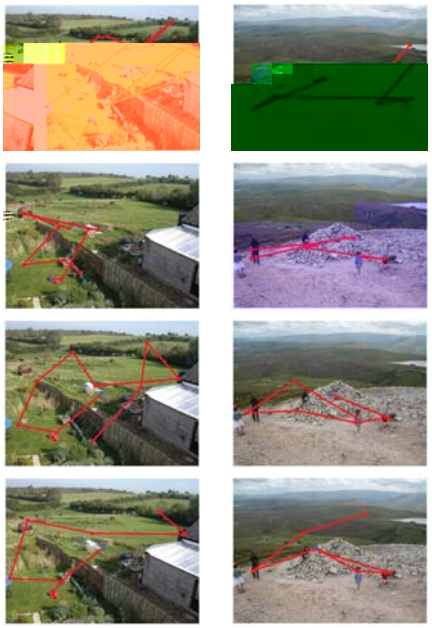
. O . fi
. .
fi .

3.1. Dataset and eye movement data collection

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

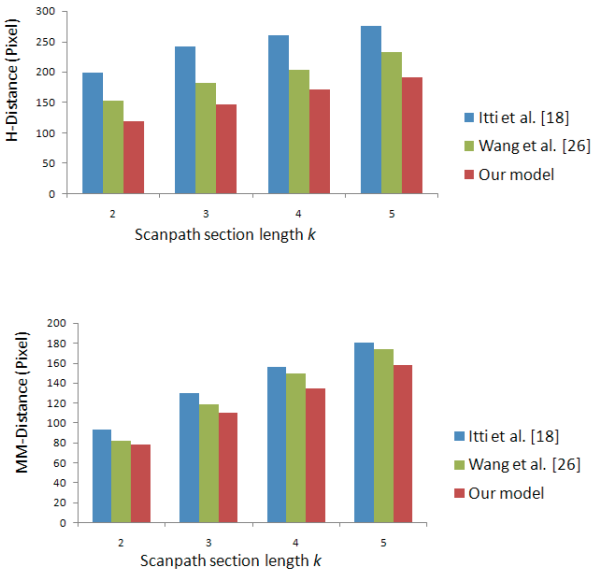
3.2. Evaluation of fixation order

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



F . . C .
. .
. 1 , 2 , . ,
, .
. fi .

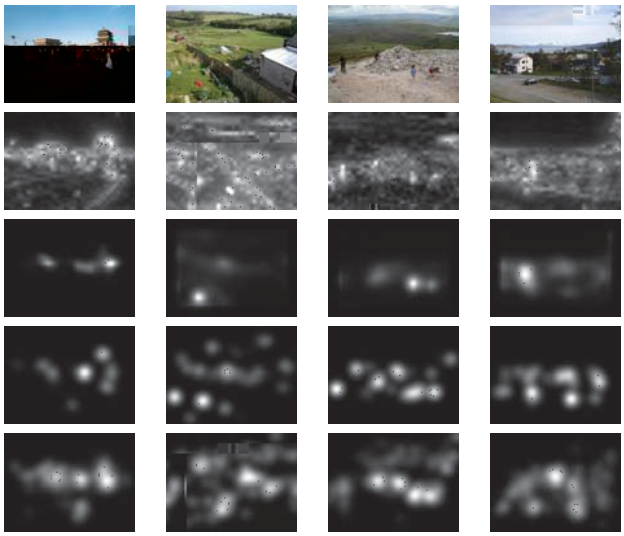
A . I R
. . . , .
. H , . I *et al.* . M ,
. . .
. . 1 .
2 , .
. A I R . 1 ,
. .
1 .
.
fi
, . fi .
fi .
, fi .
. . 8 ∼ 10 fi
. , fi .
. , , I *et al.* 1 . 2 ,
. R , .
. S. , .
. . .
. A F . , .
.
.



F . . C . . , 1 2
. H . ‹‹ ‹
‹ k.

3.2.1 Distance of scanpaths

I .
‹ ,
time-delay embedding, .
‹ 23 . S fi ,
‹ k, . . $C_m^k(t)$ =
$(c_m(t), \cdots c_m(t+k-1))$ k
, t fi
m. t,
‹ . k
$X = \{C_m^k(t)\}_t \subseteq \mathbb{R}^k$. S
$Y = \{C_h^k(\tau)\}_\tau$ k
‹ . I .
$k = 2$,
fi . C . . X Y
$\mathbb{R}^k$ 4
. .
F k $x =$
$C_m^k(t) \in X$, fi .
$d_k(x, Y) = \min_\tau\{\|x - C_h^k(\tau)\|_2\}/k$. I
, k
‹ . k , $d_k(x, Y)$
x. ,
. , . .
. . .
. . . H . ‹‹

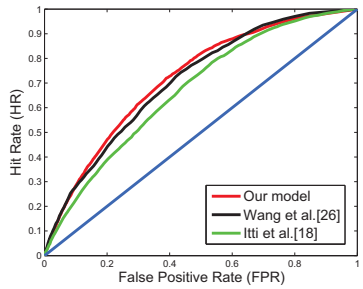


F . . C . . fi
. . ‹
. , 1 , 2
, . , . fi , .

H D . . ‹
‹ ,
fi

$$d_H^k = \max_t\{\min_\tau\{\|C_m^k(t) - C_h^k(\tau)\|_2\}\}/k \quad 3$$
$$= \max_t\{d_k(C_m^k(t), Y)\}. \quad 4$$

mean minimal distance MM D ,
, fi $d_M^k = \mathsf{E}_t[d_k(C_m^k(t), Y)]$.
I . , .
‹ ‹ ‹ $\epsilon = 0.7$,
$Z = 800$, ‹ . ‹ ‹
2.3°. F .



F . . ROC . .
2 1 .

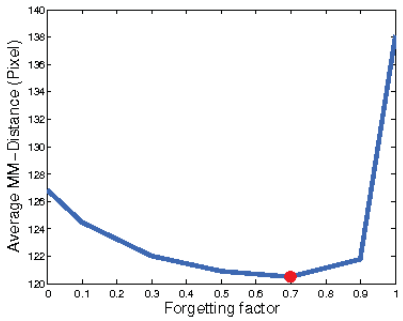
1. ROC

	I *et al.* 1	*et al.* 2	O.
ROC	.	. 1	. 1 3

F . ∼ . . 2 .
ROC . ROC
F . . 1. ROC . ROC
fi fi .
. ROC ,
. F fi . ,
. fi
. 1 , 2 .

3.4. Assessment of the forgetting factor

, .
. ϵ .
1 . . F .
. ϵ
. N .
H . , , .
k ϵ.
. F . 1 $\epsilon = 0.7$,
.
D. ,
. ϵ , . . $\epsilon = 0$.
, . .
. A . ,
fi
. I , , .
. .
ϵ , . . $\epsilon = 1$, .
fi . C . ,
fi .
. . . **V**

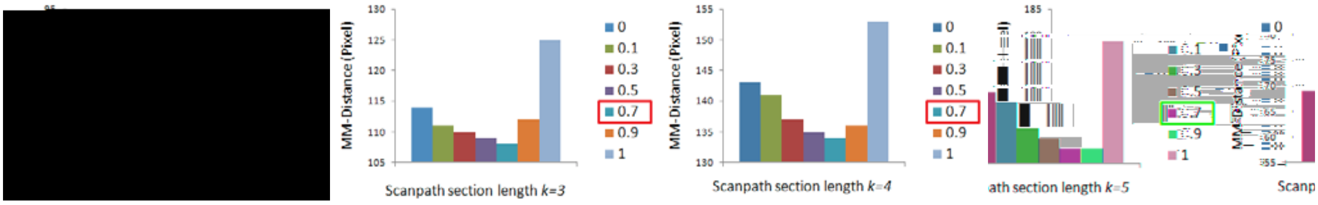


F . 1 . k
ϵ. $\epsilon = 0.7$.

, fi
$\epsilon = 0.7$ . .
fi . F
, .
. .
1 22 .

4. Conclusion, Discussion and Future Work

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



F . . A . ϵ .

I . , . edit distance

. F . . U .

ournal of Vision, 2 .

D. , V. M , N. V . NIPS, 2 .

. M , fi . I .

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. . F . . Computer Vision and Pattern Recognition, 2 .

2 A. , D. A , . S . M . . 1 . Investigative Ophthalmology, 1 .

3 H. . U . . Neural Computation, 1 .

4 C. , . R , . R . . ournal of Neuroscience, 2 .

N. . . S . NIPS, 2 .

. ournal of Neuroscience, 2 .

M. C , C. P , M. E . S . Vision Research, 2 .

A. C . V . Annual Review of Psychology, 1 1.

. F . . U . ournal of Vision, 2 .

1 D. , V. M , N. V . NIPS, 2 .

11 . . P . A . SPIE Proceedings: Human Vision and Electronic Imaging, 1 .

12 . . P . R . fi . ACM Symposium on Eye racking Research & Applications, 2 2.

13 V. , . H. , D. R . R . CVPR, 2 .

14 . H M. . R . ournal of Vision, 2 1 .

1 . H , C. , P. P . NIPS, 2 .

1 . H A. S . I fi . Proc. R. Soc. ond. B, 1 .

1 . H . . S A . 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. V , . C . . ournal of Vision, 2 .

23 . S , . , M. C . E . ournal 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 .