Simulating Human Saccadic Scanpaths on Natural Images

Wei Wang[1,3], Cheng Chen[1], Yizhou Wang[1,2]
Tingting Jiang[1,2], Fang Fang[2,4], Yuan Yao[2,5]
[1]Natl Eng. Lab for Video Technology, [2]Key Lab. of Machine Perception (MoE), Peking University
[3]Graduate University, Chinese Academy of Sciences, Beijing 100049, China
[4]Department of Psychology, Peking University, Beijing 100871, China
[5]School of Mathematical Sciences, Peking University, Beijing 100871, China
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 human visual system, neurons representing different retinal eccentricities have different spatial frequency tuning. Foveal neurons have a smaller average receptive field size and are better tuned to high spatial frequencies. They are capable of processing visual information at very high spatial resolution. On the other hand, cortical neurons representing peripheral vision have larger receptive fields and are more sensitive to the lower range of spatial frequencies. They are capable of processing information at low spatial resolution [8]. So the information from a foveal image at one fixation is very limited. Human saccadic eye movement is an important mechanism to compensate for the loss of visual acuity in the periphery and to actively pursue information around the scene. In a highly dynamic and cluttered world, to acquire visual information efficiently and rapidly, it is important for our brain to decide not only where we should look at, but also the sequence of fixations. Indeed, both of them are essential for us to understand human saccadic behavior.

In the paper, we propose a computational model to simulate human saccadic scanpaths on natural images without a particular task. Investigating this topic is not only helpful in understanding the computational aspects of visual perception, but also beneficial to many important applications such as image and video compression, object detection, and web-page design.

Proposed method The proposed saccade model integrates three factors that drives human attention reflected by eye movements: reference sensory responses, fovea-periphery resolution discrepancy, and visual working memory.

Fig. 1 shows the framework of the proposed model. The three modules highlighted with a grey-blue background are the key factors. (i) The reference sensory responses of an image are multi-band filter responses of sparse coding functions extracted from the stimuli. They are considered as a representation of the raw input signal and serve as the reference information in the proposed system. Neurophysiological evidence shows that the receptive fields of neurons in the primary visual cortex (V1) are similar to sparse codes learned from natural images [21]. The reference sensory responses simulate neuronal responses of the primary visual cortex to an image at a uniform high resolution. (ii) The fovea-periphery resolution discrepancy provides detailed information around a fixation location, but coarse information in periphery. This information discrepancy directly leads to sequential fixation transitions. In Fig. 1, a foveal image at fixation Q_t is generated and multi-band fil-

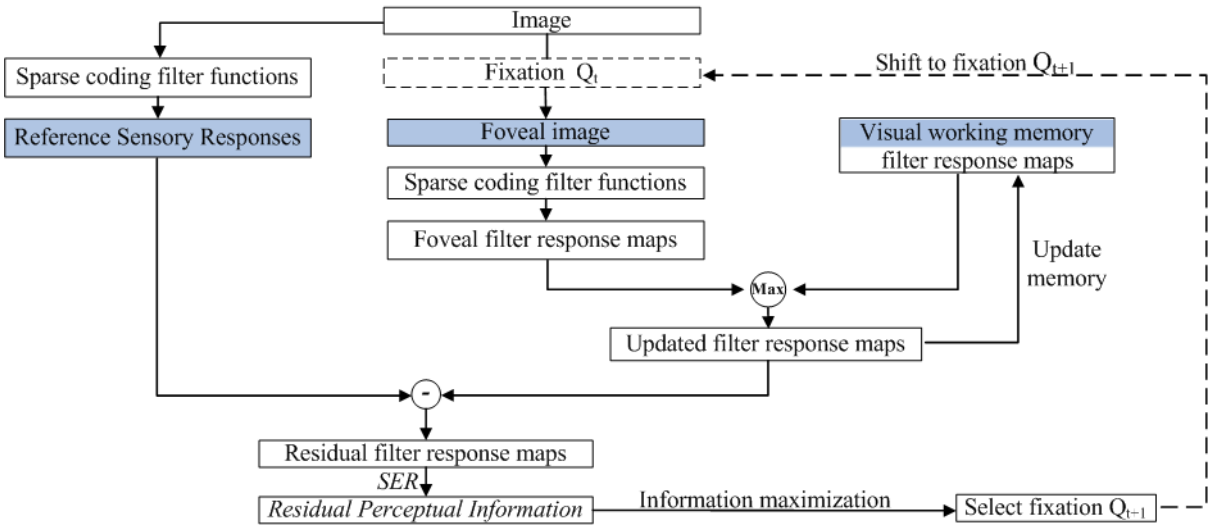

Figure 1. The proposed framework.

ter response maps of this foveal image are extracted. (iii) Visual working memory retains perceptual information for a certain period of time and inhibits immediate return of attention. We also represent the content in working memory with multi-band filter response maps. To simulate the decay of working memory, we multiply the filter response maps in the memory with a forgetting factor ϵ ($0 \leq \epsilon \leq 1$). We update visual working memory by taking the maximal filter response values of the foveal image filter responses and the current decayed filter responses in the memory at each location. The resulted filter responses encode the *so far* obtained perceptual information of the scene in brain.

Then, by subtracting the updated filter response maps in the working memory from the reference sensory responses, we obtain multi-band residual filter response maps. This simulates the dynamic interaction among the three factors along with eye movements. Consequently, we compute the *Site Entropy Rate* [26] on residual filter response maps to obtain a *residual perceptual information (RPI)* map. The resulted RPI map represents the spatial distribution of the discrepancy between the amount of information stored in brain and that contained in input stimulus. In order to pursue the maximal amount of "new" information of the scene and reduce this perceptual discrepancy, according to the information maximization principle, we choose the spot of the maximal SER value as the next fixation Q_{t+1}, and then start the next iteration.

At each eye movement, the three *multi-band filter response maps* are computed as a coherent representation for the three factors. Based on this representation and the strategical implementation of the framework, we provide experimental results to demonstrate the model's capability in simulating human saccadic scanpaths.

1.1. Related work

In the literature, there are two categories of research related to modeling saccadic behavior in terms of generating scanpaths. The first category focuses on computing static visual saliency maps which describe the importance of each image location by image feature contrast or image feature rarity. Itti *et al*. [18] propose a biologically-plausible visual saliency model based on the center-surround contrast mechanism. By arguing that [18]'s linear model of the similarity measure on color, intensity, and orientation is inconsistent with the properties of higher level human judgement, Gao *et al*. [10] propose a discriminant center-surround hypothesis based on mutual information. It treats saliency detection as a classification problem, and obtains an optimal solution from the perspective of decision theory. Considering that information pursuit is the driving force behind attentive sampling, Bruce *et al*. [5] adopt the *self-information* of sparse features as a saliency measure. Harel *et al*. [15] propose a graph-based visual saliency model which computes the equilibrium distribution of a Markov chain on a fully-connected graph as a saliency measure. Gopalakrishnan *et al*. [13] extend [15]'s work and apply it in salient region detection. By analyzing the log-spectrum of natural images, Hou *et al*. [17] compute the spectral residual of an input image and transform the spectral residual to spatial domain to obtain the saliency map. Achanta *et al*. [1] propose a frequency-tuned method to detect salient regions. Itti *et al*. [20] propose a concept of Bayesian surprise mechanism to interpret visual attention. The "surprise" is defined as the difference between the posterior and prior distributions of beliefs of an observer over the hypotheses about the world. These models use static information to predict eye fixations, but ignore many important dynamic proper-

ties in human saccadic behaviors such as fixation order. The work of [18] adopts the principles of winner-take-all and inhibition of return to generate a scanpath out of a static saliency map. However, By comparing scanpaths generated by [18]'s model with human eye movement data, we find that this kind of static saliency model cannot predict fixation order well as shown in Section 3, which suggests that the dynamic properties should be taken into account to simulate human saccade.

The second category focuses on *dynamic* visual saliency maps during human saccade from the view of information theory. Compared with the static visual salience research, the dynamic aspect of human saccade is much less studied. Lee and Yu [19] propose an information maximization framework to explain saccadic eye movements, however, it is lack of experiments to justify their model. Inspired by the framework, Renninger *et al*. [22] simulate human saccade on novel shape silhouettes with both global and local information. Nevertheless, it is not trivial to extend this model to natural images.

There are some other studies about the effective factors of eye movement. Harding *et al*. [14] manipulate low-level image features such as luminance and chromaticity to measure the effects of these changes on human scanpaths, and compare these effects to [18]'s predicted results. Foulsham *et al*. [9] find that accumulation of scanpaths facilitates fixation prediction, and scanpaths can be explained by bottom-up guidance. Both of the work find that scanpaths generated by [18]'s strategy based on static visual saliency do not predict real human saccadic behavior well. Bahill *et al*. [2] discover that most naturally occurring human saccades are within 15 degrees of visual angle. This is a fact exploited in our model.

The rest of the paper is organized as follows. In Section 2, we introduce the details of our model. The experimental results and a new evaluation method of fixation order and fixation location are presented in Section 3. Finally, we discuss some issues of the model and conclude the paper in Section 4.

2. Our Approach

In this section, we introduce each component of the proposed model (shown in Fig. 1) in details and elaborate the information pursuit strategy of eye movements based on the residual perceptual information measure.

2.1. Coherent representation of three factors

We propose multi-band filter responses as a coherent representation for the three factors. The integration of them at each eye movement produces a dynamic saliency map, which decides the next fixation.

2.1.1 Sparse coding filters

Single-unit recording evidence shows that when a natural image is presented, it only activates a small number of V1 neurons [3]. To simulate this sparse property of simple cells in the primary visual cortex, the sparse coding theory is proposed to extract the intrinsic local structure of natural images for efficient coding [21].

In this paper, we use sparse coding filter functions to compute three *multi-band filter response maps* for the reference sensory responses, fovea-periphery resolution discrepancy, and visual working memory. Specifically, Independent Component Analysis (ICA) [16] is adopted to learn a set of 192 color sparse filter functions from 120,000 image patches of size $8 \times 8 \times 3$ pixels, which are randomly extracted from 1500 natural color images. Then, each filter convolves an image and generates a sub-band filter response map. 64 samples of learned color sparse bases are shown in Fig. 2.

Figure 2. 64 learned color sparse coding bases.

2.1.2 Foveal imaging

To simulate the foveal imaging of human eyes, we adopt Geisler and Perry's multi-resolution pyramid method to create foveal images [12]. First, the spatial resolution (acuity) at each pixel relative to the current fixation is computed by a function of eccentricity [11]. Second, we build a multi-resolution pyramid of the original image, typically of 6 or 7 levels. Third, the desired spatial resolution at each pixel in the foveal image is computed as a weighted sum of corresponding pixel values in different layers of the pyramid. An original image and one of its foveal images are shown in Fig. 3 (a) and (b).

(a) (b)

Figure 3. An example of foveal imaging. (a) Original image (b) Foveal image at the fixation labeled by a red cross.

2.1.3 Visual working memory

Once an image location is visited by the fovea, information at that fixation is acquired. Visual working memory integrates the information across previous eye movements, meanwhile, it loses the stored information at a certain rate. This forgetting property will steer eyes moving back to previously visited salient spots when the "residual information" becomes trivial, in other words, the information at the previous fixations has been forgotten. In the following we explain the mechanism of updating the filter responses in the visual working memory.

Simulating the forgetting properties. In our model, we multiply the current filter responses in working memory with a constant forgetting factor ϵ ($0 \leq \epsilon \leq 1$) to simulate its forgetting property. If $\epsilon = 1$, no forgetting effect; if $\epsilon = 0$, it is memoryless. In Section 3.4, we compare simulated scanpaths under different values of ϵ. The result shows that the forgetting property is an important factor to be modeled in simulating human saccades.

Updating visual working memory. Visual working memory is the place where perceptual information from current fixation and previous ones is dynamically updated.

In our model, the updating process of working memory is implemented as taking the maximal filter response values between the foveal image filter responses and the current decayed filter responses in working memory at each location. A *Max* operation is to simulate the transient activation in the caudal superior frontal sulcus and posterior parietal cortex when updating the attentional focus [4]. This is another key process that happens in working memory. To be specific, let $f_k^v(x, y, t)$ and $f_k^w(x, y, t)$ represent the k-th sub-band filter responses of a foveal image and visual working memory at time t and position (x, y) respectively. Then the updating process is as follows:

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

Computing residual filter response maps. The updated filter response maps in visual working memory will interact with the reference sensory responses to predict the next fixation (see Fig. 1). Psychological evidence shows that people will shift attention when they move fixation to another point [7]. Also it is known that attention shift-away is functionally equivalent to reducing stimulus strength [6]. Therefore in our model, we subtract the updated filter responses in visual working memory from the reference sensory responses to simulate the reduction of the stimulus strength. The residual filter responses are computed as $r_k = |f_k^o - f_k^w|$, where f_k^o represents the k-th sub-band of the reference sensory response map.

2.2. Measuring residual perceptual information

From the view of information theory, a residual filter response map contains the perceptual discrepancy between the information contained in an image and that stored in brain after a series of fixations. This discrepancy is what we aim to pursue in the following eye movements.

In our model, the residual perceptual information is measured by the *Site Entropy Rate* (SER) [26] computed from the residual filter response maps. The *Site Entropy Rate* model adopts a fully-connected graph representation for the filter response maps to simulate the cortical neuron connectivity. Random walks are deployed on the graph to model the information transmission between neurons of the network. The site entropy rate of the random walk is proposed to measure the average information transmitted from a node to all the others. In this way, each sub-band filter response map produces a SER map. By summing up all the SER maps, the total exchanged information (measured by SER value) at each location i can be obtained as:

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

where π_{ki} is the stationary probability at location i for the k-th filter response map, P_{kij} is the transition probability of a random walk from location i to location j on the k-th filter response map. According to [26], the larger the SER value at a location is, the more salient the location is. Please refer to [26] for more details about the SER model.

There are two major reasons we adopt the SER to measure residual perceptual information. First, the SER model can explain the perceptual mechanism of the center-surround saliency whereas the residual filter response map cannot. Moreover, the residual filter response map is sensitive to noise, *e.g.*, salt-and-pepper type noise, and high frequency texture regions. Second, the residual perceptual information measure derived from SER is from the information theoretic viewpoint. This coincides with our model assumption, i.e. human saccade is a dynamic process of information pursuit.

2.3. Saccadic amplitude

Using our eye movement data, we plot the distribution of saccadic amplitudes in Fig. 4. It is found that more than 90% saccade amplitudes are within 20° of visual angle around the current fixation point. Therefore, in the proposed model, we adopt the following randomized strategy to select the next fixation Q_{t+1} contingent on current fixation Q_t. First, a window of size $Z \times Z$ pixles centered at current fixation point Q_t is selected, where the value of $Z/2$ (about 400 pixels) corresponds to 20° of visual angle under the eye tracking system. Then the locations where the maximal SER values inside and outside the window are selected

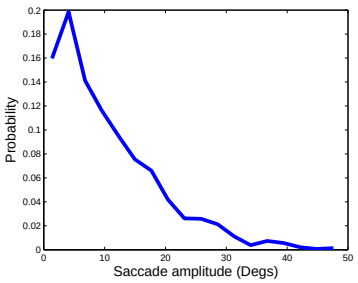

Figure 4. The distribution of saccade amplitudes.

as candidates for Q_{t+1}. Next, we sample Q_{t+1} from these candidates according to the probability $p(z \leq Z/2)$, which is obtained from the statistics of saccade amplitudes shown in Fig. 4.

At the new fixation Q_{t+1}, we generate a new foveal image and start the next iteration.

3. Experimental Results

To test the performance of the proposed model, we collect human eye movement data from a natural image dataset, and compare scanpaths generated by our model and two other approaches against the eye movement data in two aspects. One is the dynamic aspect of saccades such as fixation orders. The other is the static property of scanpaths such as fixation spatial densities.

3.1. Dataset and eye movement data collection

We randomly collected a dataset of 20 color images from the Internet including natural scenes, street and buildings, and indoor images, etc. We collected eye movement data from 24 subjects with this dataset using a high-speed SMI eye-tracker with a 500 Hz sampling rate. Subjects were positioned 0.53m away from a 21-inch CRT monitor. The images were presented in a random order, each was displayed for 3 seconds followed by a blank screen for 1 second. A cross was placed at the center of the blank screen so as to engage the first fixation at the center of the images. The subjects were given no particular instruction except asking them to observe the images.

3.2. Evaluation of fixation order

There is a lack of literature on computational models of the dynamic aspect of visual attention. As mentioned above, Lee and Yu's work [19] is rather a conceptual framework without adequate implementation solution and experimental results; Renninger *et al.* [22] simulate scanpaths on novel shapes, whereas, it is not trivial to adapt their method to natural images. Nonetheless, Itti *et al.* in [18] propose a scanpath generation method from static saliency maps based on winner-takes-all (WTA) and inhibition-of-return (IoR) regulations. To our knowledge, this is the most referred method in literature. Hence, we compare our model with Itti *et al.*'s approach. Moreover, to demonstrate that the advantage of the proposed model does come from incorporating the dynamic process of information pursuit, we substitute the static saliency computation module in [18] with an up-to-date state-of-the-art static saliency model [26], but still using the WTA and IoR regulations in [18], then we compare the generated scanpaths by our method and by the updated model of [18].

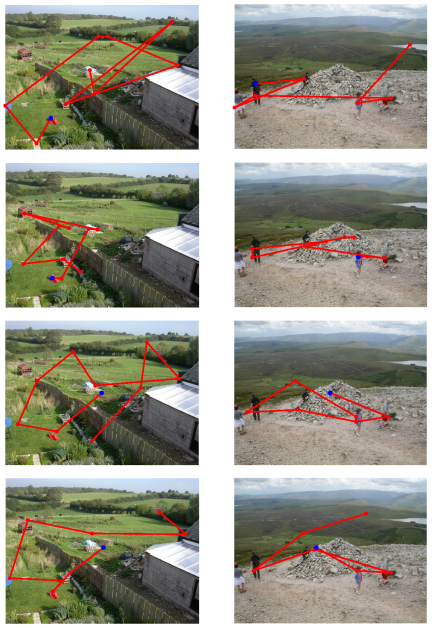

Figure 5. Comparison of scanpaths generated by our model and the other two methods. The rows from top to bottom show the scanpaths simulated by [18]'s method, [26]'s method, our method, and example scanpaths from eye tracking data, respectively. The blue dots mark the starting fixation points.

To be consistent with the setting of the eye movement experiments, our model places the initial fixation at the image center and then generates a series of fixations. When simulating eye movements, first we decide a length of the scanpath for an image. The length is sampled from the statistics of eye movement data on that image (usually $8 \sim 10$ fixations). Then, we generate three fixation sequences of that length using the following three models, our model, Itti *et al.*'s model [18] and its updated model using [26], respectively. Repeating this on the dataset, we obtain three groups of scanpaths. Subsequently, we compare the three groups against human scanpaths using the evaluation method introduced below. As shown in Fig. 5, the simulated scanpaths by our model are more similar to human saccades than the other two methods.

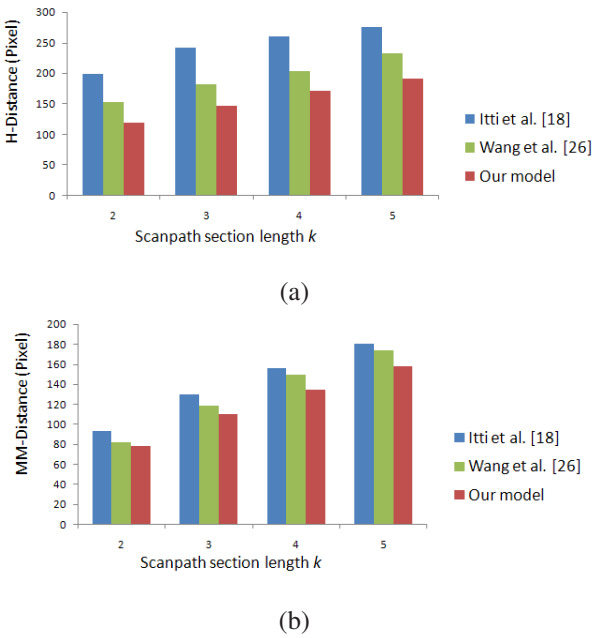

Figure 6. Comparison results between our model, [18] and [26] using Hausdorff distance (a) and mean minimal distance (b) at different scanpath length k.

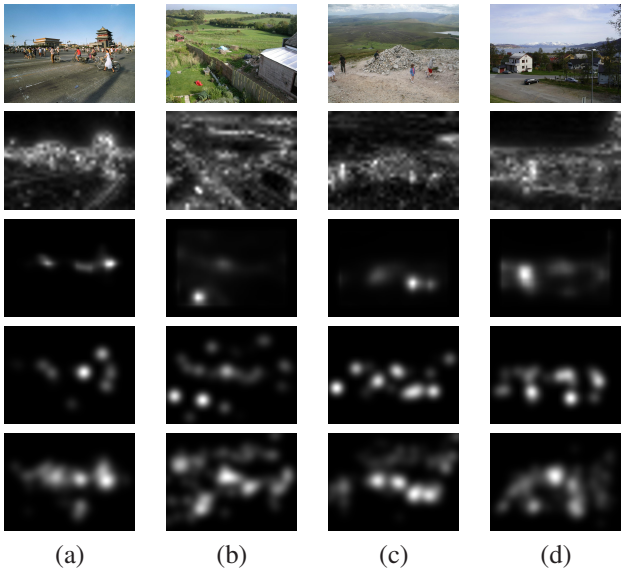

Figure 7. Comparison results about fixation density maps between our model and two other methods. The rows from top to bottom are: input images, the saliency maps by [18]'s model, [26]'s model, our model, and human fixation density maps, respectively.

3.2.1 Distance of scanpaths

In order to quantitatively compare the stochastic and dynamic scanpaths of varied lengths, we propose to employ *time-delay embedding*, which has been used widely in the study of dynamical systems [23]. Specifically, we divide scanpaths into pieces of length k, e.g. $C_m^k(t) = (c_m(t), \cdots c_m(t+k-1))$ denotes a k-dimensional time-delay embedding vector, starting at the t'th fixation generated by a model m. By varying the initial point t, the collection of all such k-dimensional vectors gives rise to the model space $X = \{C_m^k(t)\}_t \subseteq \mathbb{R}^k$. Similarly $Y = \{C_h^k(\tau)\}_\tau$ denotes all the k-dimensional vectors in eye movement data of the same image. In particular when $k = 2$, these vectors are discrete approximation of vector fields. Comparison between such point cloud data X and Y in $\mathbb{R}^k$ will reflect the dynamical similarities between models and human.

For each model-generated k-dimensional vector $x = C_m^k(t) \in X$, define its normalized distance to the human scanpaths as $d_k(x, Y) = \min_\tau\{\|x - C_h^k(\tau)\|_2\}/k$. In other words, we search among all the length k scanpath sections of human eye tracking data on the image, $d_k(x, Y)$ is the one with the minimum distance to the given model vector x. The smaller such a distance, the closer/more similar to human scanpath, thus a better prediction.

We use two distance measures to evaluate the scanpaths generated by a model w.r.t. human data: (i) Hausdorff distance (H-Distance) computes the maximal value of all the minimal distances between two sets of scanpaths, which is defined as

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

$$= \max_t\{d_k(C_m^k(t), Y)\}. \quad (4)$$

(ii) The *mean minimal distance* (MM-Distance) , as its name tells, is defined as $d_M^k = \mathrm{E_t}[d_k(C_m^k(t), Y)]$.

In the evaluation experiments, our model parameters are set as follows: the forgetting factor $\epsilon = 0.7$, the window size $Z = 800$ pixels, the half-resolution of foveal imaging is $2.3°$. Fig. 6 (a) and (b) show comparison results between our model and the other two methods in terms of the Hausdorff distances and the mean minimal distances, respectively. We compare scanpaths at different lengths $k = 2$ to 5. From the comparison results, we can see that our model performs better in simulating the dynamics of saccadic scanpaths. When $k > 5$, the distance measures increase with k due to the stochastic property of human saccadic behavior, whereas, the ranking of the compared methods remains the same. To study the individual difference in scanpaths is one of our future tasks.

3.3. Evaluation of fixation distribution

We compare our model with those introduced in [26] and [18] using two types of measures described in [5]: (1) the static saliency map of each image is computed as a fixation density map using a kernel-based density estimation.

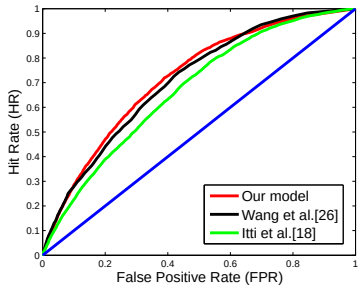

Figure 8. ROC curves of our model and two other models from [26] and [18].

Table 1. ROC area comparison

	Itti *et al*. [18]	Wang *et al*. [26]	Our model
ROC area	0.6706	0.7081	0.7183

Fig. 7 (a∼d) shows the comparison results. (2) We also use ROC curves and ROC areas to compare these three models in Fig. 8 and Tab. 1. Both the ROC curves and ROC areas are generated by classifying the locations in a saliency map into fixations and non-fixations with varying quantization thresholds. The larger is the ROC area, the better prediction does the model make. From both figures and the table, it can be seen that our model predicts human fixations more accurately than [18], and comparable to [26].

3.4. Assessment of the forgetting factor

To assess modeling the decay property of working memory, we generate a series of scanpaths on each image of the dataset by tuning the forgetting factor ϵ from 0 (memoryless) to 1 (remember every detail) in our model. Fig. 9 shows the average mean minimal distances between the simulated scanpaths and eye tracking data with different ϵ over all the images of the dataset. Note that the results using Hausdorff distance are similar, hence, omitted. To select the optimal forgetting factor, we also compute the average distance of different saccade length k at each ϵ. The plotted curve in Fig. 10 clearly shows that when $\epsilon = 0.7$, the model predicts the scanpaths best.

During the experiment, we observe two interesting phenomena. (i) When ϵ is small, e.g. $\epsilon = 0$ for an extreme case, the model quickly forgets just acquired information in working memory. As a result, the simulated scanpath will be trapped into the oscillation in between two or three fixations. In other words, in this case, there is no inhibition of return. This obviously deviates from human behavior. (ii) When ϵ is large, e.g. $\epsilon = 1$ for an extreme case, the model retains most of the acquired information from the previous fixations in working memory. Consequently, the simulated fixations seldom returns to the visited locations. This is also not true compared to the human data. Via generating scanpaths under different forgetting factor and comparing them with eye tracking data, the proposed model finds the best parameter $\epsilon = 0.7$ to simulate human saccadic behavior (with fixation revisits). From both psychological evidence and the comparison results to other models, we argue that considering the forgetting effect of visual working memory plays a key role in simulating eye movements. This is one of the key differences between the proposed model and those proposed in [19] and [22].

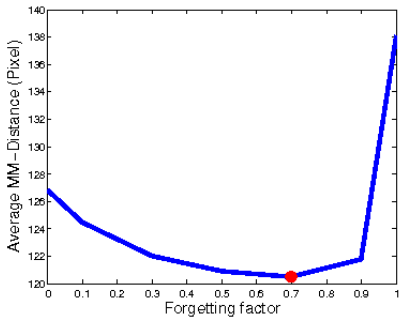

Figure 10. The average distance of different scanpath length k at each ϵ. $\epsilon = 0.7$ is selected as model parameter.

4. Conclusion, Discussion and Future Work

In the paper, we propose a computational model to simulate human saccadic scanpaths on natural images without a specific task based on the principle of information maximization. The proposed model identifies and integrates three important factors as the driving forces to guide eye movements based on a coherent multi-band sparse code filter response representation. The computational model embodies several well known psychological phenomena, such as center-surround saliency, inhibition of return, fixation revisit, algorithmic implementation of information transmission on the neural network and the forgetting effect of short term working memory, etc. We propose a new evaluation method to compare fixation order of scanpaths. Extensive experiments show the advantage of the proposed model in predicting human saccadic scanpaths.

This model is inspired by some biological evidences, but not aims to copy the exact mechanisms of visual processing. For examples, although people have different views on the accessibility of raw signal [25], the proposed model adopts the *reference sensory responses* as a reference information or knowledge to compare with the content in visual working memory and consequently guide eye movements. Also note that sparse coding may not be universally acknowledged due to its limitation in accounting for the high-order statistics of natural images, but many work still use it to represent an image because it is consistent with a lot of experimental evidence (see [24]).

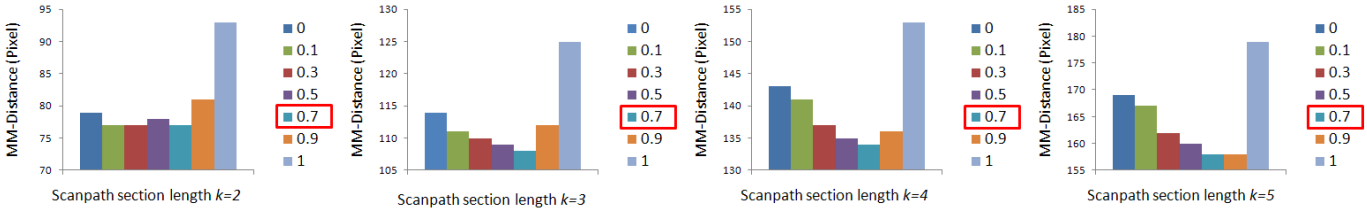

Figure 9. Average mean minimal distances between the simulated scanpaths and eye tracking data with different ϵ over the image dataset.

In the literature, some work uses *edit distance* to measure the difference between scanpaths [9]. We do not adopt this measure for two reasons: (i) The method introduces more free parameters, e.g. the cost for every operator. It is rather subjective to tune these parameters to make the distance measure reasonable. (ii) We observe that although there exist shared sections among scanpaths, the stochastic nature of saccade introduces much variation among scanpaths. This makes the design of operations and the operator cost tuning particular hard.

In the future, we will extend the current model to explain individual difference in scanpaths and propose some new evaluation criteria for such extension. Moreover, the fixation duration is another important attribute of saccadic behavior. It will also be considered in our future work.

Acknowledgments

This work was supported by the National Science Foundation of China under grant No. 90920012 and 60872077, the National Basic Research "973" Program of China under grant No. 2009CB320904.

References

[1] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk. Frequency-tuned salient region detection. *Computer Vision and Pattern Recognition*, 2009.

[2] A. Bahill, D. Adler, and L. Stark. Most naturally occurring human saccades have magnitudes of 15 degrees or less. *Investigative Ophthalmology*, 1975.

[3] H. Barlow. Unsupervised learning. *Neural Computation*, 1989.

[4] C. Bledowski, B. Rahm, and J. Rowe. What 'works' in working memory? separate systems for the selection and updating of critical information. *Journal of Neuroscience*, 2009.

[5] N. Bruce and J. Tsotsos. Saliency based on information maximization. *NIPS*, 2006.

[6] G. Buracas and G. Boynton. The effect of spatial attention on contrast response functions in human visual cortex. *Journal of Neuroscience*, 2007.

[7] M. Carrasco, C. Penpeci-Talgar, and M. Eckstein. Spatial covert attention increases contrast sensitivity across the csf: support for signal enhancement. *Vision Research*, 2000.

[8] A. Chapanis. Vision. *Annual Review of Psychology*, 1951.

[9] T. Foulsham and G. Underwood. What can saliency models predict about eye movements? spatial and sequential aspects of fixations during encoding and recognition. *Journal of Vision*, 2008.

[10] D. Gao, V. Mahadevan, and N. Vasconcelos. The discriminant center-surround hypothesis for bottom-up saliency. *NIPS*, 2007.

[11] W. Geisler and J. Perry. A real-time foveated multiresolution system for low bandwidth video communication. *SPIE Proceedings: Human Vision and Electronic Imaging*, 1998.

[12] W. Geisler and J. Perry. Real-time simulation of arbitrary visual fields. *ACM Symposium on Eye Tracking Research & Applications*, 2002.

[13] V. Gopalakrishnan, Y. Hu, and D. Rajan. Random walks on graphs to model saliency in images. *CVPR*, 2009.

[14] G. Harding and M. Bloj. Real and predicted influence of image manipulations on eye movements during scene recognition. *Journal of Vision*, 2010.

[15] J. Harel, C. Koch, and P. Perona. Graph-based visual saliency. *NIPS*, 2006.

[16] J. Hateren and A. van der Schaaf. Independent component filters of natural images compared with simple cells in primary viusal cortex. *Proc. R. Soc. Lond. B*, 1998.

[17] X. Hou and L. Zhang. Saliency detection: A spectral residual approach. *Computer Vision and Pattern Recognition*, 2007.

[18] L. Itti, C. Koch, and E. Niebur. A model of saliency based visual attention for rapid scene analysis. *IEEE TPAMI*, 1998.

[19] T. Lee and S. Yu. An information-theoretic framework for understanding saccadic eye movements. *Advanced in Neural Information Processing System*, 1999.

[20] l.Itti and P. Baldi. Bayesian surprise attracts human attention. *NIPS*, 2006.

[21] B. Olshausen and D. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 1996.

[22] L. Renninger, P. Verghese, and J. Coughlan. Where to look next? eye movements reduce local uncertainty. *Journal of Vision*, 2007.

[23] T. Sauer, J. Yorke, and M. Casdagli. Embedology. *Journal of Statistical Physics 65: 579 16*, 1991.

[24] E. Simoncelli and B. Olshausen. Natural image statistics and neural representation. *Annual Review of Neuroscience*, 2001.

[25] F. Tong. Primary visual cortex and visual awareness. *Nature Reviews Neuroscience*, 2003.

[26] W. Wang, Y. Wang, Q. Huang, and W. Gao. Measuing visual saliency by site entropy rate. *CVPR*, 2010.