Principal Components Analysis-Based Visual Saliency Detection

Bing Yang, Xiaoyun Zhang, Jing Liu, Li Chen and Zhiyong Gao
Institute of Image Communication and Network Engineering
Shanghai Key Laboratory of Digital Media Processing and Transmissions
Shanghai Jiao Tong University, Shanghai 200240, China
{yangbing,xiaoyun.zhang,hilary,hilichen,zhiyong.gao}@sjtu.edu.cn

***Abstract*—In this paper, a novel patch-wise saliency detection algorithm is proposed based on Principal Component Analysis (PCA). As a powerful statistical procedure in data analysis, PCA are fully exploited to convert color space and produce compact patch representation. Specifically, images are first converted to linearly uncorrelated channels and divided into non-overlapped patches. Then the patches are represented by the coefficients of principal components using PCA analysis. Based on the compact representation of patches, two types of distinctiveness are introduced: center-surround contrast and global rarity. Experimental results demonstrate that the PCA-based color space conversion and patch representation can improve the accuracy of human fixations prediction, and the proposed algorithm outperforms the mainstream algorithms on predicting human fixations.**

***Index Terms*—Saliency detection, Principal Component Analysis (PCA), Center-surround, Rarity.**

I. Introduction

Visual attention has been extensively studied recently due to its wide applications in computer vision, contrast enhancement and video compression [1]–[3]. Numerous saliency models have been proposed by psychologists, neurophysiologists and computer scientists to imitate human visual attention. According to the motivation of inferring visual attention, the saliency detection models can be categorized into bottom-up models and top-down models. Top-down models are task-driven while bottom-up models are data-driven and more related to the nature of human visual system. So we focus on bottom-up model in this paper.

Although there is no consensus on the mechanism of human visual system, it is widely accepted that human tends to focus on distinct regions that stand out from the entire image. Based on this motivation, many researchers have proposed various saliency models to estimate visual saliency [4]–[19]. Despite the variety in existing saliency detection models, they all deal with the following three key issues: color space, visual units (pixel-wise or patch-wise) and features, and distinctiveness definition.

Color space that comes along with color images has been widely researched in the past years. Many image processing algorithms operate on separate color channel and fuse the output of each channel either uniformly or non-uniformly. When these methods perform on redundant color space (e.g., RGB), over-emphasis on redundant information may be caused. Even though LAB color space reduces the correlation from a statistical point of view, it does not decorrelate specific images.

Patch-based methods, which have been widely used in literatures, have intrinsic advantages over pixel-based methods for saliency detection problem. First, saliency is a region-based concept since single pixel with high distinctiveness can not be captured. Second, patch-based methods process the patch as a whole. They always have lower computational complexity, which is desired for the pre-processing procedure of high-level tasks such as image understanding. Therefore, patch-based method is used in the proposed algorithm and patch representation comes as an intrinsic problem. Direct representation using pixel values introduces noises and neglects the intrinsic spatial correlations among patches. In [15], dimensionality reduction is adopted for co-similarity matrix, which is equivalent to a PCA representation of image patches. In our paper, we formulate a compact and informative representation of patches directly from PCA point of view. Principal components are extracted to represent each patch and the dimensions corresponding to noises are thrown away to make the representation more effective.

The last key issue associated with a saliency detection algorithm is how to define distinctiveness. Different definition of distinctiveness results in different algorithm. A commonly used definition is center-surround contrast [7]. Large contrast between center and surrounding regions indicates highly informative regions, hence, attracting more attentions. Another type of distinctiveness is defined as global rarity [4], [12]. A patch with features that rarely appear over the entire image is believed distinct and draws more attentions. As pointed out by [5], [9], center-surround contrast and global rarity work complementarily for saliency detection of images.

In this paper, we propose a patch-wise image saliency detection algorithm using PCA analysis. As a powerful statistical procedure, PCA plays an important role throughout the proposed scheme. More specifically, the RGB color space is first transformed into uncorrelated color space where correlations among different channels are discarded through PCA analysis. Then, the image of each channel is divided into patches and the principal components of each patch are extracted as patch features based on PCA. With such a compact patch representation, visual saliency is measured by patch distinctiveness both locally and globally. Local distinctiveness focuses on center-

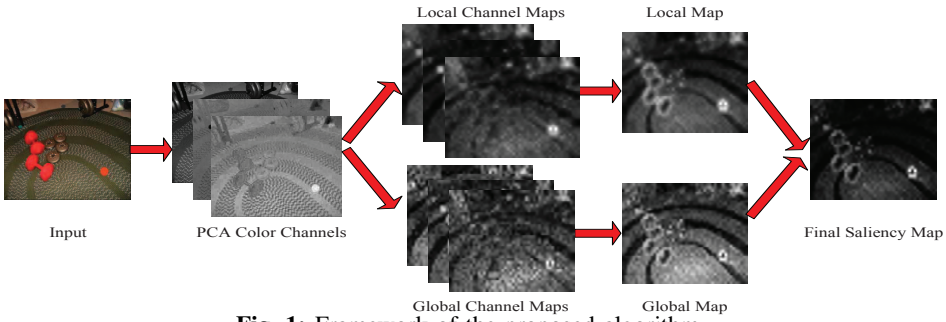


Fig. 1: Framework of the proposed algorithm.

surround contrast while global distinctiveness deals with rarity of features, i.e., the coefficients of principal components. Finally, the saliency maps are normalized and fused together. The framework of the proposed algorithm is summarized in Fig. 1. As shown in the figure, the power of PCA is fully exploited to make a good prediction of patch-wise saliency.

The remainder of this paper is organized as follows. The linearly uncorrelated color space conversion using PCA is presented in Section II. Section III proposes a compact patch representation using PCA analysis. Section IV presents two complementary operations to measure distinctiveness. The experimental results are illustrated in Section V. Finally, Section VI concludes this paper.

II. Color space decorrelation using PCA

Let I be a RGB image represented by a $M \times N \times 3$ matrix. If R, G, and B components are regarded as three variables associated with each pixel, I can be seen as $M \times N$ observations. Since PCA is mathematically defined as a procedure to transform correlated variables to linearly uncorrelated variables, it offers a good solution to find the desired color space.

Reshape each channel of I into a row vector and subtract the mean for each channel, we have $X = [\mathbf{x}_r, \mathbf{x}_g, \mathbf{x}_b]^T$, where $\mathbf{x}_c = \text{vec}(C) - \overline{C}, (c, C) \in \{(r, R), (g, G), (b, B)\}$. $\text{vec}(\cdot)$ and $\bar{\cdot}$ are the vectorization and average function. The covariance matrix of X is:

$$\text{COV}(X) = \frac{1}{MN-1} XX^T. \tag{1}$$

According to linear algebra, the 3×3 symmetric matrix $\text{COV}(X)$ can be orthogonally diagonalized as $\text{COV}(X) = V_X \Lambda_X V_X^T$, where $\Lambda_X = [\lambda_X^1, \lambda_X^2, \lambda_X^3]$ is a diagonal matrix. Let the $T_X = V_X^T$ be the transformation matrix, $Y = V_X^T X$, we have

$$\begin{aligned} \text{COV}(Y) &= \frac{1}{MN-1} YY^T \\ &= \frac{1}{MN-1} (V_X^T X)(V_X^T X)^T \\ &= \frac{1}{MN-1} V_X^T XX^T V_X \\ &= V_X^T V_X \Lambda V_X^T V_X \\ &= \Lambda_X. \end{aligned} \tag{2}$$

Obviously, the transformed color channels of Y are linearly uncorrelated. Finally, Y is reshaped into a $M \times N \times 3$ matrix I_{decorr} corresponding to original image I.

III. Patch Representation using PCA

Given an image I_{decorr} ($M \times N \times 3$) in the proposed uncorrelated color space, we extract non-overlapped patches of size $k \times k$ for each channel. So there are $L = \lfloor M/k \times N/k \rfloor$ patches in each channel. Reshape the L patches in a $k^2 \times L$ matrix P_0 where each column is a sample of observation, i.e., a vectorized patch. Subtract the average patch value from each column of P_0, we have $P = [\mathbf{p}_1, \mathbf{p}_2, \cdots, \mathbf{p}_L]$ where $\mathbf{p}_i$ is the normalized patch. Similar to color space conversion, the principal components of patches are extracted using PCA analysis.

Let $\text{COV}(P)$ be the covariance matrix of P, it can be orthogonally diagonalized as $\text{COV}(P) = V_P \Lambda_P V_P^T$. $\Lambda_P = [\lambda_P^1, \lambda_P^2, \cdots, \lambda_P^{k^2}]$ is a diagonal matrix and we assume that $\lambda_P^1 > \lambda_P^2 > \cdots > \lambda_P^{k^2}$. If the transformation matrix T_P equals V_P^T, the transformed patches are $Q = T_P P = V_P^T P = [\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_L]$. Similar to equation (2), the covariance matrix of Q is the diagonal matrix Λ_P. Therefore, the transformed patches $\mathbf{q}_i$ are linearly uncorrelated. The variance of each component is determined by diagonal values in Λ_P. To make balanced operation on each component, Q is further normalized as

$$NQ = \Lambda_P^{-0.5} * Q. \tag{3}$$

Suppose $\mathbf{nq}_i = [nq_i^1, nq_i^2, \cdots, nq_i^{k^2}]$ is the i-th column of NQ. nq_i^j is the coefficient of the j-th principal component for the i-th patch. For robust and compact representation, we extract the first K principal components, i.e., K most important components. Thus,

$$\widetilde{NQ} = [\widetilde{\mathbf{nq}}_1, \widetilde{\mathbf{nq}}_2, \cdots, \widetilde{\mathbf{nq}}_L], \tag{4}$$

where $\widetilde{\mathbf{nq}}_i = [nq_i^1, nq_i^2, \cdots, nq_i^K]^T$ is the normalized compact representation for the i-th patch $\mathbf{p}_i$. Since PCA analysis is applied on the entire image, $\widetilde{\mathbf{nq}}_i$ is globally unified representation of $\mathbf{p}_i$. Therefore, it is well suited for local and global distinctiveness detection discussed in the following section.

TABLE I: Maximum sAUC values of different models. The sigma values in the second rows are where models achieve their best performance. Accuracies of three best models over each dataset are highlighted in bold font. L, G, and LG stand for the proposed Local, Global and Local+Global models, respectively.

Dataset	**AIM** [4]	**Borji** [5]	**DCT** [6]	**Itti** [7]	**ASW** [8]	**CA** [9]	**LARK** [10]	**FT** [11]	**RARE** [12]	**SR** [13]	**HFT** [14]	**SWD** [15]	**L**	**G**	**LG**
TORONTO	0.695	0.696	0.712	0.646	**0.713**	0.696	0.707	0.582	0.707	0.693	0.672	0.621	**0.714**	0.707	**0.717**
Optimal σ	0.03	0.03	0.04	0.03	0.01	0.06	0.04	0.04	0.02	0.03	0.01	0.02	0.04	0.04	0.04
MIT	**0.685**	0.674	0.671	0.621	**0.685**	0.673	0.652	0.514	0.675	0.658	0.636	0.613	**0.685**	**0.687**	**0.688**
Optimal σ	0.03	0.03	0.04	0.03	0.01	0.03	0.06	0.03	0.02	0.03	0.01	0.02	0.04	0.03	0.03
KOOTSTRA	0.592	0.600	0.604	0.580	**0.617**	0.602	0.598	0.561	0.611	0.591	0.569	0.543	**0.616**	0.613	**0.619**
Optimal σ	0.02	0.02	0.04	0.02	0.01	0.02	0.03	0.04	0.02	0.01	0.01	0.01	0.03	0.03	0.03
NUSEF	**0.650**	0.636	0.631	0.575	**0.644**	0.630	0.619	0.550	0.637	0.634	0.602	0.577	**0.644**	**0.646**	**0.647**
Optimal σ	0.04	0.04	0.04	0.03	0.02	0.04	0.05	0.04	0.02	0.03	0.01	0.02	0.04	0.04	0.04

TABLE II: Average Runtime (in second) obtained with non-optimized MATLAB implementations of different models.

Model	**AIM** [4]	**Borji** [5]	**DCT** [6]	**Itti** [7]	**ASW** [8]	**CA** [9]	**LARK** [10]	**FT** [11]	**RARE** [12]	**SR** [13]	**HFT** [14]	**SWD** [15]	**L**	**G**	**LG**
Times(s)	15.68	12.11	0.12	0.58	32.54	31.3	2.26	0.12	1.55	0.03	0.62	0.74	0.73	0.58	0.76

IV. Saliency Map Estimation

With the compact representation $\widetilde{NQ}$, two complementary (local and global) operations are applied to each channel. The first one detects center-surround contrast while the latter one focuses on the rarity of a single patch over the entire image. Then, an inter-channel fusion is followed by a local-global combination to produce final saliency map.

A. Center-Surround Contrast

For local saliency, we evaluate the dissimilarity between center patch with its surroundings. Suppose $\mathbf{p}_j \in \mathcal{S}_i, j \in [1, \cdots, 8]$ are eight-connected surrounding patches of the center patch $\mathbf{p}_i$. Center-surround contrast is defined as:

$$S_l(\mathbf{p}_i) = \frac{1}{8}\sum_{j=1}^{8} \frac{D_{\text{pca}}(\mathbf{p}_i, \mathbf{p}_j)}{D_{\text{loc}}(\mathbf{p}_i, \mathbf{p}_j) + \delta}, \tag{5}$$

where δ is a constant. $D_{\text{loc}}(\mathbf{p}_i, \mathbf{p}_j)$ is the Euclidean distance between spatial locations of $\mathbf{p}_i$ and $\mathbf{p}_j$. $D_{\text{pca}}(\mathbf{p}_i, \mathbf{p}_j)$ is the Euclidean distance between PCA-based compact representation discussed in Section III, i.e.,

$$D_{\text{pca}}(\mathbf{p}_i, \mathbf{p}_j) = \|\widetilde{\mathbf{nq}}_i - \widetilde{\mathbf{nq}}_j\|_2^2. \tag{6}$$

D_{loc} emphasizes the contrast between nearby patches and eliminates the influence of patches far away. This is consistent with our motivation for local distinctiveness.

B. Global Rarity

Although it is common for saliency regions to have high center-surround contrast, it often happens that saliency regions have similar surroundings but they are globally distinct over the entire image. A simple example is a patch within the saliency object. It has low local contrast but high rarity in the image. Therefore, a complementary global operation is necessary for saliency detection algorithm.

Motivated by the information-theoretic saliency method of Bruce and Tsotsos [4], we measure the rarity of patch using the probability of each coefficient in the PCA-based patch representation. More specifically, we calculate the binned (100 bins here) histograms for each principal component coefficient. The produced histograms are denoted as $\mathbf{h}_1, \cdots, \mathbf{h}_K$. For a certain patch $\mathbf{p}_i$ represented by $\widetilde{\mathbf{nq}}_i = [nq_i^1, nq_i^2, \cdots, nq_i^K]$, the probability of nq_i^j is approximated by $P(nq_i^j) = \mathbf{h}_j(\text{Bin}(nq_i^j))/L$ where $\text{Bin}(nq_i^j)$ is the bin that nq_i^j falls into. Assuming that coefficients of different components are conditionally independent, the probability of $\mathbf{p}_i$ is the product of probability of individual components. That is

$$P(\mathbf{p}_i) = \prod_{j=1}^{K} P(nq_i^j). \tag{7}$$

And the rarity of this patch is measured by:

$$S_g(\mathbf{p}_i) = -\log P(\mathbf{p}_i). \tag{8}$$

C. Fusion of Saliency maps

After the local or global mechanisms performed on each channel, 6 saliency maps are fused into one single map. The fusion is achieved in two steps: an inter-channel fusion followed by a local-global combination. For a better fusion, we exploit a smart normalization operator $\mathcal{N}(\cdot)$ [7]. This operator assigns higher weight to maps with smaller number of strong peaks. Guaranteed by the decorrelated color space, uniform fusion of the normalized maps is applied for inter-channel fusion,

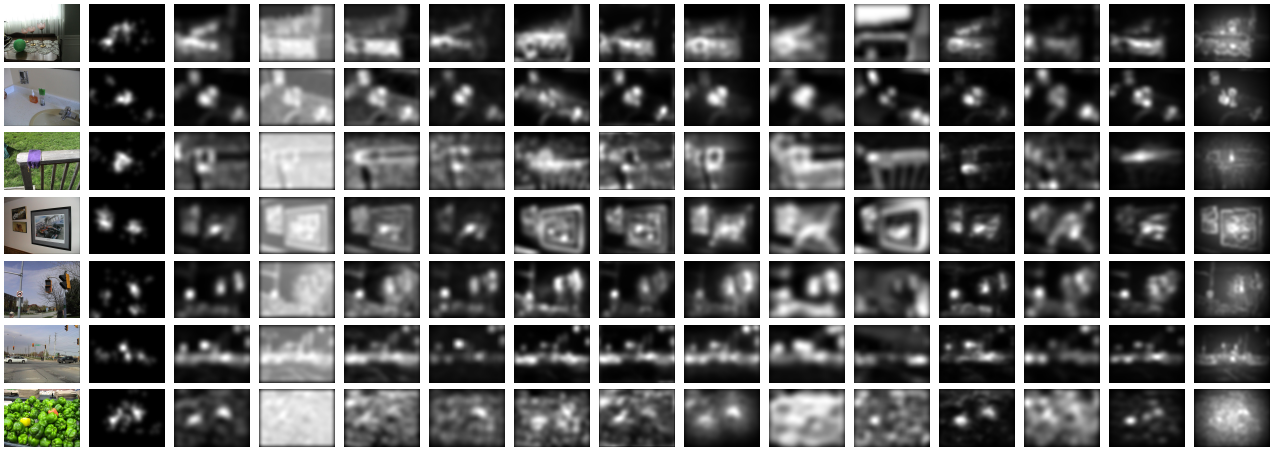
$$\begin{aligned} S_L &= \mathcal{N}(S_l^{\text{channel1}}) + \mathcal{N}(S_l^{\text{channel2}}) + \mathcal{N}(S_l^{\text{channel3}}), \\ S_G &= \mathcal{N}(S_g^{\text{channel1}}) + \mathcal{N}(S_g^{\text{channel2}}) + \mathcal{N}(S_g^{\text{channel3}}). \end{aligned} \tag{9}$$

Finally, local and global saliency maps are combined as

$$S = \mathcal{N}(S_L) \cdot \mathcal{N}(S_G). \tag{10}$$

V. Experimental Results

In this section, we evaluate the proposed saliency model qualitatively and quantitatively by using TORONTO saliency benchmark described in [4], MIT saliency benchmark described in [20], KOOTSTRA saliency benchmark described in [21] and NUSEF saliency benchmark described in [22]. The TORONTO saliency benchmark includes 120 images with



(a) (b) (c) (d) (e) (f) (g) (h) (i) (j) (k) (l) (m) (n) (o)

Fig. 2: Qualitative results for the TORONTO dataset [4]. (a) Original image. (b) Human Fixations Map. (c) Our model. (d) AIM [4]. (e) Borji [5]. (f) DCT [6]. (g) Itti [7]. (h) ASW [8]. (i) CA [9]. (j) LARK [10]. (k) FT [11]. (l) RARE [12]. (m) SR [13]. (n) HFT [14]. (o) SWD [15]. Saliency maps are blurred with optimal blurring factor which are shown in Table I.

resolution of 511×681 from indoor and outdoor environments. The MIT saliency benchmark includes 1003 images collected form Flicker and LabelMe datasets. The KOOTSTRA saliency benchmark includes 101 images with resolution of 768×1024 from 5 different categories: animals, automan, buildings, flowers and natural scenes. The NUSEF saliency benchmark includes 758 images containing emotionally affective scenes or objects. For all the experiments described in this paper, the patch size is 14×14, and the extracted principal components number is 11.

Our model is compared with 12 mainstream models AIM [4], Borji [5], DCT [6], Itti [7], ASW [8], CA [9], LARK [10], FT [11], RARE [12], SR [13], HFT [14], SWD [15][1]. To tackle the well-known center bias and border effects, shuffled Receiver Operating Characteristic Area Under the Curve (sAUC) is adopted to evaluate the performance. The sAUC score is computed to evaluate the consistency between the saliency map and its fixations map [6]. The saliency map of each model is convolved with a variable size Gaussian kernel. The sAUC score of models are obtained over the range of standard deviations σ of the Gaussian kernel in image width (from 0.01 to 0.13 in steps of 0.01). Table I shows the maximum sAUC values of different models under optimal σ. It can be seen that our model outperforms most of the other models. Besides our model, the second best model is AWS focusing on the sAUC score, which is consistent with the conclusion in [23]. The qualitative comparison with the mainstream models on TORONTO saliency benchmark is also shown in Fig. 2. As demonstrated by the experimental results, the fully exploited usage of PCA for saliency detection can improve the precision of human fixations prediction.

[1]The implementation of mainstream models except SR and HFT are downloaded from http://saliency.mit.edu/results_mit300.html. The implementation of SR is downloaded from http://www.its.caltech.edu/~xhou/. The implementation of HFT is downloaded from https://sites.google.com/site/jianlinudt/hft.

Another advantage of the proposed model is that it yields an acceptable computational complexity. Table II shows the average runtime with different models. Computation times reported in Table II are obtained with non-optimized MATLAB implementation on a PC running Windows 7 with Intel CORE^{TM} i7-3770S @ 3.1 GHz CPU and 8GB RAM. The input image size is 511×681. Our model is faster than AIM, Borji, ASW, CA, LARK, RARE, and slower than DCT, Itti, FT, SR, HFT, and SWD models. Our Global model is a little faster than our Local model. And our Local+Global model consumes almost the same computation time as our Local model. This is because that the Global model uses the same patch representation with Local model. Compared to the PCA-based color space conversion and patch representation process, the time consumed for histogram calculation is ignorable.

VI. Conclusion

A patch-wise saliency detection algorithm is proposed based on PCA analysis in this paper. By fully exploiting the power of PCA, the color channels are decorrelated and a compact representation of patch is obtained. And then visual distinctiveness are detected both locally and globally based on this representation. Local distinctiveness measures center-surround contrast while global distinctiveness evaluates the rarity compared to the entire image. Experimental results demonstrate that the PCA-based color space conversion and patch representation can improve the accuracy of human fixations prediction. And the proposed algorithm achieves superior accuracy against the mainstream algorithms on predicting human fixations. Incorporating top-down factors and extending our model into spatio-temporal domain are two future works.

Acknowledgment

This work was supported in part by National Natural Science Foundation of China (61221001, 61133009, 61301116), the 111 Project (B07022), Shanghai Key Laboratory of Digital Media Processing and Transmissions (STCSM 12DZ2272600), National Key Technology R&D Program of China (2013BAH53F04).

REFERENCES

[1] Laurent Itti, "Automatic foveation for video compression using a neurobiological model of visual attention," *Image Processing, IEEE Transactions on*, vol. 13, no. 10, pp. 1304–1318, 2004.

[2] Ali Borji and Laurent Itti, "State-of-the-art in visual attention modeling," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 1, pp. 185–207, 2013.

[3] Ke Gu, Guangtao Zhai, Xiaokang Yang, Wenjun Zhang, and Chang Wen Chen, "Automatic contrast enhancement technology with saliency preservation," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 25, no. 9, pp. 1480–1494, Sept 2015.

[4] Neil Bruce and John Tsotsos, "Saliency based on information maximization," in *Advances in neural information processing systems*, 2005, pp. 155–162.

[5] Ali Borji and Laurent Itti, "Exploiting local and global patch rarities for saliency detection," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2012, pp. 478–485.

[6] Xiaodi Hou, Jonathan Harel, and Christof Koch, "Image signature: Highlighting sparse salient regions," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 1, pp. 194–201, 2012.

[7] Laurent Itti, Christof Koch, and Ernst Niebur, "A model of saliency-based visual attention for rapid scene analysis," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, , no. 11, pp. 1254–1259, 1998.

[8] Antón Garcia-Diaz, Xosé R Fdez-Vidal, Xosé M Pardo, and Raquel Dosil, "Saliency from hierarchical adaptation through decorrelation and variance normalization," *Image and Vision Computing*, vol. 30, no. 1, pp. 51–64, 2012.

[9] Stas Goferman, Lihi Zelnik-Manor, and Ayellet Tal, "Context-aware saliency detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 10, pp. 1915–1926, 2012.

[10] Hae Jong Seo and Peyman Milanfar, "Static and space-time visual saliency detection by self-resemblance," *Journal of vision*, vol. 9, no. 12, pp. 15, 2009.

[11] Ravi Achanta, Sheila Hemami, Francisco Estrada, and Sabine Susstrunk, "Frequency-tuned salient region detection," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2009, pp. 1597–1604.

[12] Nicolas Riche, Matei Mancas, Matthieu Duvinage, Makiese Mibulumukini, Bernard Gosselin, and Thierry Dutoit, "Rare2012: A multi-scale rarity-based saliency detection with its comparative statistical analysis," *Signal Processing: Image Communication*, vol. 28, no. 6, pp. 642–658, 2013.

[13] Xiaodi Hou and Liqing Zhang, "Saliency detection: A spectral residual approach," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2007, pp. 1–8.

[14] Jian Li, Martin D Levine, Xiangjing An, Xin Xu, and Hangen He, "Visual saliency based on scale-space analysis in the frequency domain," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 4, pp. 996–1010, 2013.

[15] Lijuan Duan, Chunpeng Wu, Jun Miao, Laiyun Qing, and Yu Fu, "Visual saliency detection by spatially weighted dissimilarity," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2011, pp. 473–480.

[16] M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S. Hu, "Global contrast based salient region detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 37, no. 3, pp. 569–582, 2015.

[17] Yin Li, Yue Zhou, Lei Xu, Xiaochao Yang, and Jie Yang, "Incremental sparse saliency detection," in *IEEE International Conference on Image Processing*. IEEE, 2009, pp. 3093–3096.

[18] Olivier Le Meur, Patrick Le Callet, Dominique Barba, and Dominique Thoreau, "A coherent computational approach to model bottom-up visual attention," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 5, pp. 802–817, 2006.

[19] Ke Gu, Guangtao Zhai, Weisi Lin, Xiaokang Yang, and Wenjun Zhang, "Visual saliency detection with free energy theory," *Signal Processing Letters, IEEE*, vol. 22, no. 10, pp. 1552–1555, 2015.

[20] Tilke Judd, Krista Ehinger, Frédo Durand, and Antonio Torralba, "Learning to predict where humans look," in *IEEE international conference on Computer Vision*. IEEE, 2009, pp. 2106–2113.

[21] Gert Kootstra, Arco Nederveen, and Bart De Boer, "Paying attention to symmetry," in *British Machine Vision Conference*. The British Machine Vision Association and Society for Pattern Recognition, 2008, pp. 1115–1125.

[22] Subramanian Ramanathan, Harish Katti, Nicu Sebe, Mohan Kankanhalli, and Tat-Seng Chua, "An eye fixation database for saliency detection in images," in *European Conference on Computer Vision*. 2010, pp. 30–43, Springer.

[23] Ali Borji, Dicky N Sihite, and Laurent Itti, "Quantitative analysis of human-model agreement in visual saliency modeling: a comparative study," *Image Processing, IEEE Transactions on*, vol. 22, no. 1, pp. 55–69, 2013.