Saliency Detection Using Quaternion Sparse Reconstruction

Yi Zeng
Shanghai Jiao Tong Univ.
Shanghai, China
bdcae@sjtu.edu.cn

Yi Xu
Shanghai Jiao Tong Univ.
Shanghai, China
xuyi@sjtu.edu.cn
http://icne.sjtu.edu.cn/info/1061/1075.htm

Abstract

We proposed a visual saliency detection model for color images based on the reconstruction residual of quaternion sparse model in this paper. This algorithm measures saliency of color image region by the reconstruction residual and performs more consistent with visual perception than current sparse models. In current sparse models, they treat the color images as multiple independent channel images and take color image pixel as a scalar entity. Consequently, the important information about interrelationship between color channels is lost during sparse representation. In contrast, the quaternion sparse model treats the color image pixels as a quaternion matrix, completely preserving the inherent color structures during the sparse coding. Therefore, the salient regions can be reliably extracted according to quaternion sparse reconstruction residual since these regions cannot be well approximated using its neighbouring blocks as dictionaries. The proposed saliency detection method achieves better performance on Bruce-Tsotsos dataset and OSIE dataset as compared with traditional sparse reconstruction based models and other state-of-art saliency models. Specifically, our model can achieve higher consistency with human perception without training step and gains higher AUC scores than traditional sparse reconstruction based models.

1. Introduction

In last two decades, visual saliency has been studied by researchers in domains of psychology, neurophysiology and computer vision. It is a very important mechanism for human beings to catch critical information effectively, especially in dynamic visual scenes of complex environments. Meanwhile, with an explosive growth of image information, it becomes more significant to automatically extract the salient regions from images.

In recent years, some visual saliency detection models have been proposed to be extensively used in object detection, target recognition and image comprehension. Most of these models take efforts to explain the cognitive process of humans [1], [2], [3]. Physiological experiments show that the neuron response is suppressive when the surrounding items are close to the center while the response is excitatory when they show a lot of difference from the center. Itti et al. [4] are motivated to define a visual attention model as center-surround contrast based on multi-scale image analysis, where a salient region pops up from a scene due to big difference from its neighbouring regions in the appearance of color, intensity and orientation.

Physiological data have suggested that primary visual cortex (area V1) uses a sparse code to efficiently represent natural scenes and the mechanisms in the area V1 contribute to the high saliency of pop-up objects [5]. In recent years, the researchers are motivated to use sparse representation model for saliency computation, where the salient regions are extracted according to sparse reconstruction residual since these regions cannot be well approximated using its neighbouing blocks as dictionaries. Han et al. [6] proposed a weighted sparse coding residual model for bottom-up saliency detection, where the reconstruction residual are weighted with the L_0 norm of sparse coefficients to produce the saliency map. In [7], the saliency value of each region is measured by the Incremental Coding Length (ICL), where the ICL is the description length of the sparse coding and increases when the center block is more informative than its surrounding blocks. All these methods used traditional sparse models to compute the reconstruction residual. However, these traditional sparse models cannot provide a good approximation of the entire spatial color structures of the image since them treat a color image as multiple independent channel images and vectorise the image patches, ignoring the interrelationship between color channels.

In order to avoid color distortions during sparse representation, in our previous work, we established quaternion-based sparse models to represent color images, and achieved better results than traditional sparse models in color image reconstruction, denoising, inpainting and super-resolution [8]. In this paper, we are motivated to propose a saliency detection model based on quaternion sparse reconstruction method and center-surround mechanism of biological vision. It is expected that the quaternion sparse model will provide a good solution of saliency detection problem due to well-preservation of color structure.

The remainder of this paper is organized as follows. Some basic concepts of quaternion algebra and the theory of sparse coding is presented in Section 2. The introduction of quaternion sparse model is presented in Section 3. A saliency detection scheme is designed to extract salient regions in Section 4. Experimental results and comparative analysis are shown in Section 5. Finally, we give some conclusion remarks in Section 6.

2. Theory of sparse coding

In color images, the difference between each block and its surrounding blocks usually is not so obvious, which indicates the possibility to use image block to predict its surrounding blocks. In recent years, great improvement has been made in Human Vision System (HVS) research field. It is pointed out that the main function of the retina is to remove the redundancy of visual information. Physiological data have suggested that primary visual cortex (area V1) uses a sparse code to efficiently represent natural scenes and the mechanisms in the area V1 contribute to the high saliency of pop-up objects [5].

Sparse representation has been widely used in image signal processing since the image can be regarded as a multiple-dimensional signal [9].

In sparse coding model we can use a linear combination of a set of dictionary elements to represent the input signal $y \in R^n$:

$$y = D_{n\times k}\alpha,\ \ k > n \tag{1}$$

The over complete base $D = \{d_1, d_2, \dots, d_K\}$ is called a dictionary. Each of these columns is usually referred as an atom. Vector $\alpha \in R^k$ is the sparse representation coefficient, which is expected to contain the minimum number of nonzero coefficients.

In order to select this kind of solution, we need to solve the following equation:

$$\widehat{\alpha} = \arg\min_{\alpha} J(\alpha)\ \ s.t.\, y = D\alpha \tag{2}$$

where $J(\alpha) = \|\alpha\|_0$ is l_0 norm. It represents the sparsity of α, meaning the number of nonzero elements in the sparse vector α. Thus, (2) can be rewritten as:

$$\widehat{\alpha} = \arg\min_{\alpha} \|\alpha\|_0\ \ s.t.\, y = D\alpha \tag{3}$$

Actually there usually has residual in the solution:

$$\widehat{\alpha} = \arg\min_{\alpha} \|\alpha\|_0\ s.t.\, y = D\alpha + r \tag{4}$$

where r is the residual or the error term that is generated in the reconstruction. Sparse representation model is shown in Figure 1.

3. Quaternion sparse model

In this section, we give some basic concepts of quaternion algebra. A more complete introduction of quaternion algebra can be referred to [8] and [10].

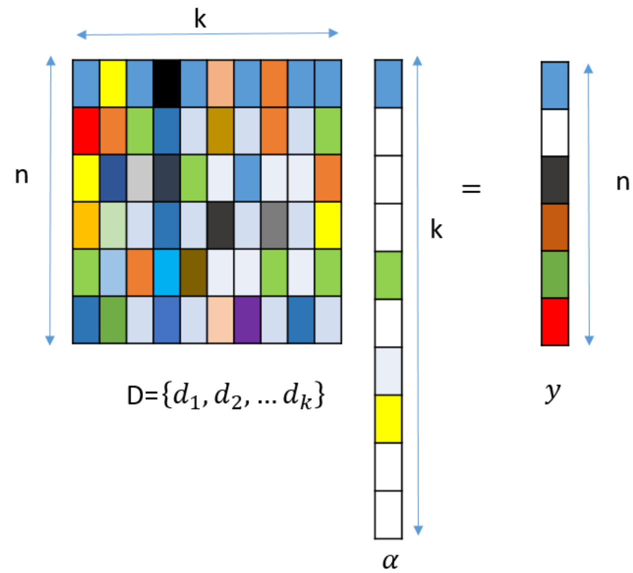


Figure 1: An illustration of sparse coding

3.1. Definition of Quaternion Algebra

Quaternion was first introduced by W. Hamilton [11] in 1832. It is of great interest because of its significance in vector analysis. Let $\dot{x}$ be a quaternion, then

$$\dot{x} = x_0 + x_1 \cdot i + x_2 \cdot j + x_3 \cdot k \tag{5}$$

where $x_i \in R$, i = 0, 1, 2 , 3, and the imaginary units *i, j, k* obey the quaternion rules that $i^2 = j^2 = k^2 = -1$ and $i \cdot j = -j \cdot i = k, j \cdot k = -k \cdot j = i, k \cdot i = -i \cdot k = j$. As a vector entity, the quaternion is associative but not commutative and it can simultaneously manipulate all its four components.

Just as complex number can be represented as matrices, so as quaternion [12]. Using 4×4 real matrices, the quaternion $\dot{x}$ can be written as:

$$\begin{bmatrix} x_0 & x_1 & x_2 & x_3 \\ -x_1 & x_0 & -x_3 & x_2 \\ -x_2 & x_3 & x_0 & -x_1 \\ -x_3 & -x_2 & x_1 & x_0 \end{bmatrix} \tag{6}$$

In this kind of representation, quaternion addition and multiplication correspond to matrix addition and matrix multiplication.

3.2. Quaternion sparse model

For an input color image, we define an image block for each channel as a vector $y_c \in R^n$, where n represent the dimension of the vector. Most of the traditional sparse models separate color channels, so each channel use a dictionary of their own:

$$y_c = D_C x_c \quad (7)$$

where $c = r, g, b$, D_C is their dictionary and x_c is sparse coefficient. But these traditional sparse models cannot take interrelationship of color channel into account. In the contrast, the quaternion model represent sparse coding in the following way:

$$\dot{y} = \dot{D}\dot{x} \quad (8)$$

where $\dot{y}$ is a representation of the color image block in quaternion model, $\dot{y} = y_r \cdot i + y_g \cdot j + y_b \cdot k$. $\dot{D} = D_r \cdot i + D_g \cdot j + D_b \cdot k$ is the dictionary. $\dot{x} = x_0 + x_1 \cdot i + x_2 \cdot j + x_3 \cdot k$ is sparse coefficient vector. By this method, we can preserve the interrelationship between the color channels.

4. Saliency detection based on quaternion reconstruction residual

Those research works of saliency detection based on traditional sparse models treat RGB channels separately or stack RGB channels as a long vector. However, it is not consistent with the mechanism of human visual system, which in essence processes the color channels parallelly.

In order to tackle this problem, we propose to use quaternion sparse reconstruction residual to measure the saliency of each image region.

In this section, we propose a saliency detection method based on quaternion sparse reconstruction residual model. Compared with other saliency detection based on traditional sparse reconstruction models, our method uses quaternion sparse reconstruction residual to measure the saliency of each image region, it can calculate reconstruction residual accurately without losing inter-channel information.

4.1. Sparse reconstruction residual computation

In our saliency detection method, the input image I is divided into several image blocks. Suppose $\dot{y} \in I$ is one of the image blocks. $S(\dot{y})$ represents the surrounding blocks of $\dot{y}$. In order to capture the structural information of the image, the image blocks are overlapped. By using quaternion model to represent the color image block, $\dot{y}$ can be represented in following way:

$$\dot{y} = \dot{D}\dot{x} + \dot{r} \quad (9)$$

where $\dot{y}$ represents the central image block, $\dot{D}$ is the dictionary extracted from $S(\dot{y})$, $\dot{x}$ is the sparse encoding coefficient, and $\dot{r} = r_0 + r_1 \cdot i + r_2 \cdot j + r_3 \cdot k$ is the residual. The goal is to obtain the best balance between sparsity and information loss, it can be written as the following cost function:

$$\mathrm{E} = \|\dot{r}\|_2^2 + \lambda \cdot Sp(\dot{x}) \quad (10)$$

where λ is the regularization parameter to achieve trade-off between the two cost terms, $Sp(\dot{x})$ represent the sparseness of $\dot{x}$.

The term of $\dot{r}$ in (9) indicates the prediction uncertainty of $\dot{y}$ when surrounding blocks and sparse coefficient x can be obtained. The unpredictability of $\dot{y}$ will increase with the higher value of $\dot{r}$. Accordingly, we define the saliency value Sa of image block $\dot{y}$ as:

$$Sa(\dot{y}) = \left\|\dot{y} - \dot{D}\dot{x}\right\|_2^2 \quad (11)$$

In order to solve the problem of sparse encoding, we use the common used l_0 norm minimization. The formula (9) can be rewritten as an optimization problem:

$$\min \lambda\|\dot{x}\|_0 + \left\|\dot{y} - \dot{D}\dot{x}\right\|_2^2 \quad (12)$$

However, this optimization is difficult to solve. According to the research of Donoho [13], the l_0 norm minimization can be replaced with the l_1 norm:

$$\min \lambda\|\dot{x}\|_1 + \left\|\dot{y} - \dot{D}\dot{x}\right\|_2^2 \quad (13)$$

This problem is a Lasso [14] linear regression problem. Transform quaternion to real matrix and then the solution process is the same as the real-valued Lasso. After that we can get sparse encoding's residual.

4.2. Framework and method

The framework is shown in Figure 2. We use sliding windows to get image blocks. For each central block, the dictionary is gotten from the surrounding blocks. Then reconstruct central block by its own dictionary via quaternion sparse model. In this step the reconstruction residual is calculated. We use the residual as our saliency value. The higher the residual is, the more salient the region is.

In this subsection, we use the quaternion sparse model to treat the color image pixel as a unit of four elements, so that the relationship between the RGB channels can be preserved.

The saliency detection algorithm is given in Algorithm 1.

Algorithm 1 Saliency computation based on quaternion sparse reconstruction residual
Require: Given color image I **Output**: The saliency map Sa 1. For each image block $\dot{y}$ of the image I, establish dictionary D from its surrounding blocks 2. Use Lasso algorithm to obtain the sparse representation cofficient $\dot{x}$ of image block $\dot{y}$ 3. The saliency value of image block $\dot{y}$ is calculated by: $Sa(\dot{y}) = \left\|\dot{y} - \dot{D}\dot{x}\right\|_2^2$ 4. Compute the saliency value of all the image blocks, return Sa

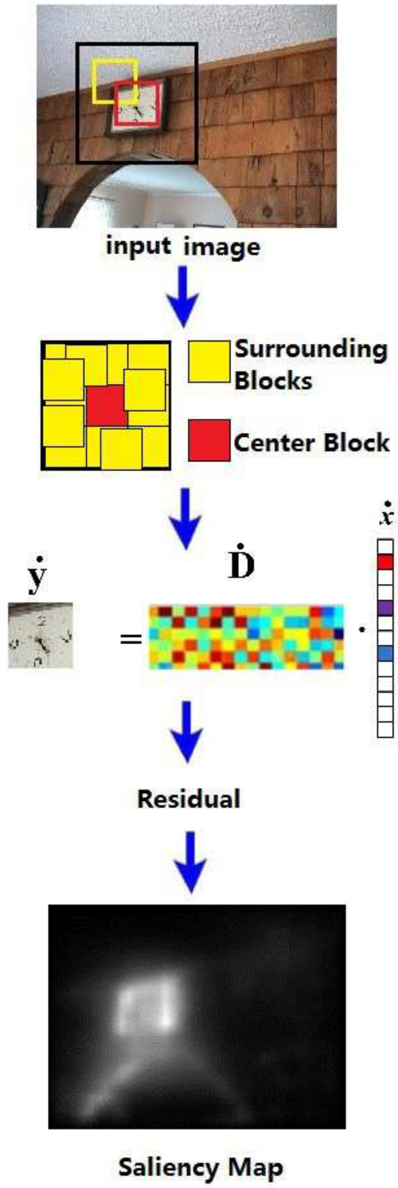


Figure 2: Framework of saliency detection based on quaternion sparse reconstruction

5. Experiment

In this section, we evaluate the performance of our saliency detection method over Bruce-Tsotsos dataset [15, 16] and OSIE (Object and semantic images and eye-tracking) dataset [17]. Bruce-Tsotsos dataset consists in eye movement data collected from 11 participants who observed 120 color images. These color images include both outdoor and indoor scenes, and most of them have cluttered background. OSIE dataset provides object and semantic saliency, including 700 images and 5551 objects with contour outlined and semantic attribute annotated.

5.1. Parameter setting

We set the size of the central image block as 8×8 pixels. We set λ = 0.25 through statistical performance analysis during Lasso linear regression. Considering that the salient regions appear at different scales, we resize the input image and compute saliency map at three scales, e.g. 40*30, 80*60, 160*120 pixels. The constructed saliency map is a linear superposition from these three scales. Our method applies Gaussian blur filter on the constructed saliency maps to preserve piece-wise saliency smoothness. The standard deviation of the two-dimensional Gauss filter is set as $\sigma_f = 4$.

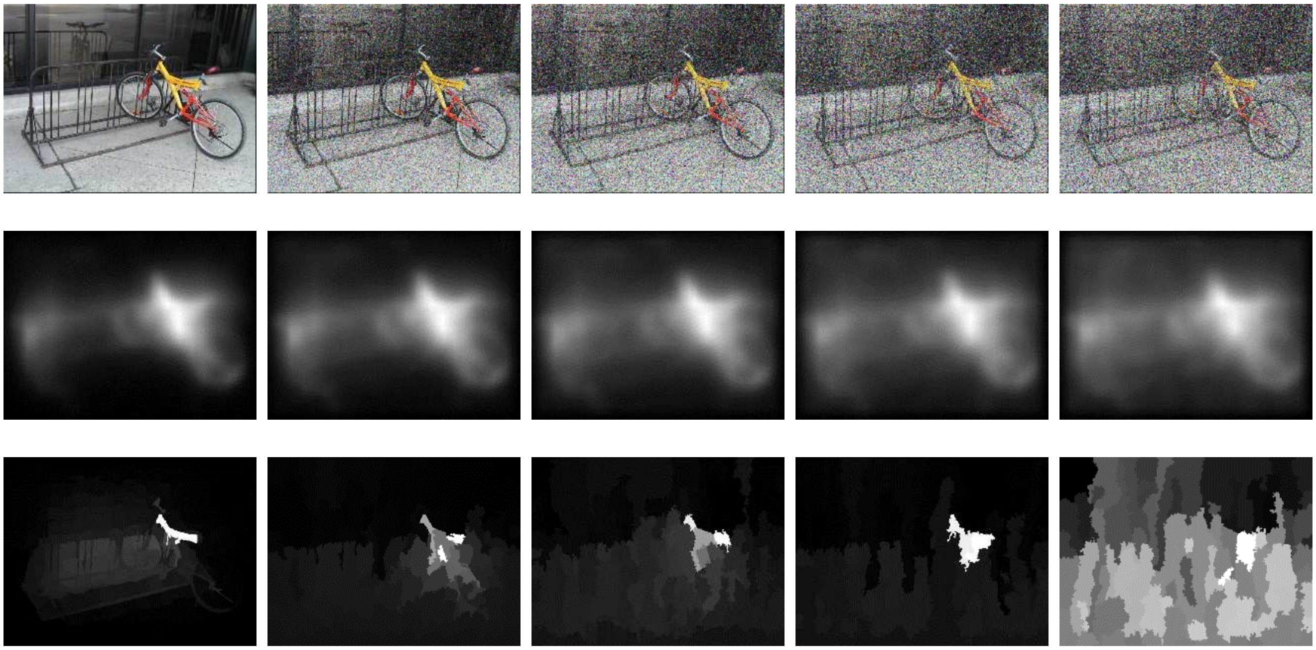
5.2. Comparisons with state-of-art algorithms

To verify the benefits of quaternion sparse model, we compare the proposed saliency detection framework with several state-of-art saliency detection methods including Incremental Coding Length (ICL) [7], Quaternion Discrete Cosine Transform (QDCT) [18], Phase spectrum of Quaternion Fourier Transform (PQFT) [19], Saliency filter(SF) [20], geodesic saliency(GS) [21], manifold ranking(MR) [22] and background connectivity(BC) [23]. MR is one of the best saliency detection algorithms so far.

We use the area under the ROC curve (AUC) to quantitatively evaluate the performance of these saliency detection methods. The AUC is a widely-used metric for performance evaluation of saliency detection. We list mean AUC scores in Table 1 for a statistical analysis from dataset.

TABLE 1.
COMPARISON OF THE MEAN AUC SCORES

Methods	Bruce-Toronto dataset	OSIE dataset
OUR	0.772872	**0.805443**
QDCT[18]	0.752091	0.785043
PQFT[19]	0.722421	0.764131
ICL[7]	0.731416	0.783006
SF[20]	0.545176	0.631455
GS[21]	0.760593	0.798070
MR[22]	**0.788485**	0.757710
BC[23]	0.761727	0.778706



a)The original image b) $\sigma_g = 0.1$ c) $\sigma_g = 0.2$ d) $\sigma_g = 0.3$ e) $\sigma_g = 0.4$

Figure 3: The saliency detection results of the image added Gaussian noise (from top to bottom: original image, saliency map of our method, saliency maps of MR method)

a) $\sigma_{s\&p} = 0.1$ b) $\sigma_{s\&p} = 0.2$ c) $\sigma_{s\&p} = 0.3$ d) $\sigma_{s\&p} = 0.4$ e) $\sigma_{s\&p} = 0.5$

Figure 4: The saliency detection results of the image added salt and pepper noise (from top to bottom: original image, saliency map of our method, saliency maps of MR method)

As we know, the AUC reflects the prediction accuracy of the saliency map for the fixation point of human eyes. Higher mean AUC score we get, more accurate prediction the algorithm can achieve. From Table 1, we observe that our method achieves the highest mean AUC scores in OSIE dataset and has good performance in Bruce-Toronto dataset.

5.3. Noise robustness

In this subsection, we evaluate the robustness of our algorithm. An instance is shown in Figure 3 and Figure 4. Given an image in Toronto database, we add Gaussian noises and salt and pepper noises to evaluate robustness of our algorithm.

We add a set of Gaussian white noises with the mean of 0 and the variance of [0.1, 0.2, 0.3, 0.4], then use our algorithm to detect saliency region, as shown in Figure 3. Also, we add salt and pepper noises with the noise density of [0.1, 0.2, 0.3, 0.4, 0.5], then use our algorithm to detect saliency region, as shown in Figure 4.

From Figure 3 and Figure 4, we can find that our method has good robustness under noises. The robustness of MR is relatively weak. Moreover, its false detection becomes more and more serious with the increase of noise level. In contrast, our method can still detect saliency region when the noise is serious. This is mainly because that we use the quaternion sparse reconstruction residual to measure the center-surrounding contrast. When the surrounding area and the central area have the same kind of noise, the influence of the noise on the reconstruction residual can be weaken during sparse coding process.

5.4. Adaptability under different scenarios

In order to verify the applicability of our algorithm in different scenarios, we select 50 indoor images, 50 outdoor images, 50 humans and 50 animal images from Toronto and OSIE databases. We calculate the accuracy of the algorithm using the mean AUC scores gained in different scenarios. Classified scenario images are shown in Figure 5 and mean AUC scores are listed in Table 2.

TABLE 2
MEAN AUC SCORES OF OUR METHOD IN DIFFERENT SCENARIOS

Indoor	Outdoor	Animal	People
0.772940	0.765085	0.846291	0.797359

From Table 2 we observe that our method has a stable performance in different scenarios. It is noted that our method achieves higher mean AUC scores in animal and people scenarios than in the outdoor scenes. This is mainly due to the background of outdoor scenes are more cluttered, which introduces reconstruction errors during sparse representation due to serious noises.

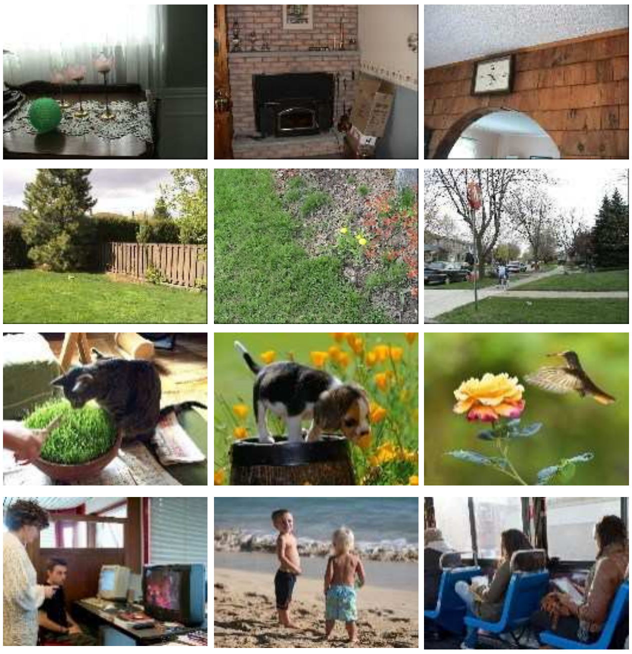


Figure 5: Classified images (from top to bottom: indoor, outdoor, animal, people)

5.5. Comparison with traditional sparse model via subjective visual evaluation results

In order to show that in real scenes our method is more accurate in general saliency detection than traditional sparse reconstruction models, we list a set of results for subjective visual evaluation in Figure 6. We can observe in the first column of Figure 6 that our saliency maps get the outline of salient object shown by human eye-tracking data accurately while ICL cannot. From the second column to the fifth column, we can observe that our saliency maps are more consistent with the human eye-tracking data than ICL.

6. Conclusions

In this paper, we propose a method for saliency detection based on quaternion sparse reconstruction residual and center-surround contrast model. Experimental results demonstrated that the proposed saliency detection framework can provide more consistent results with HVS than those methods based on traditional sparse models in most cases. The main reason is that the current sparse

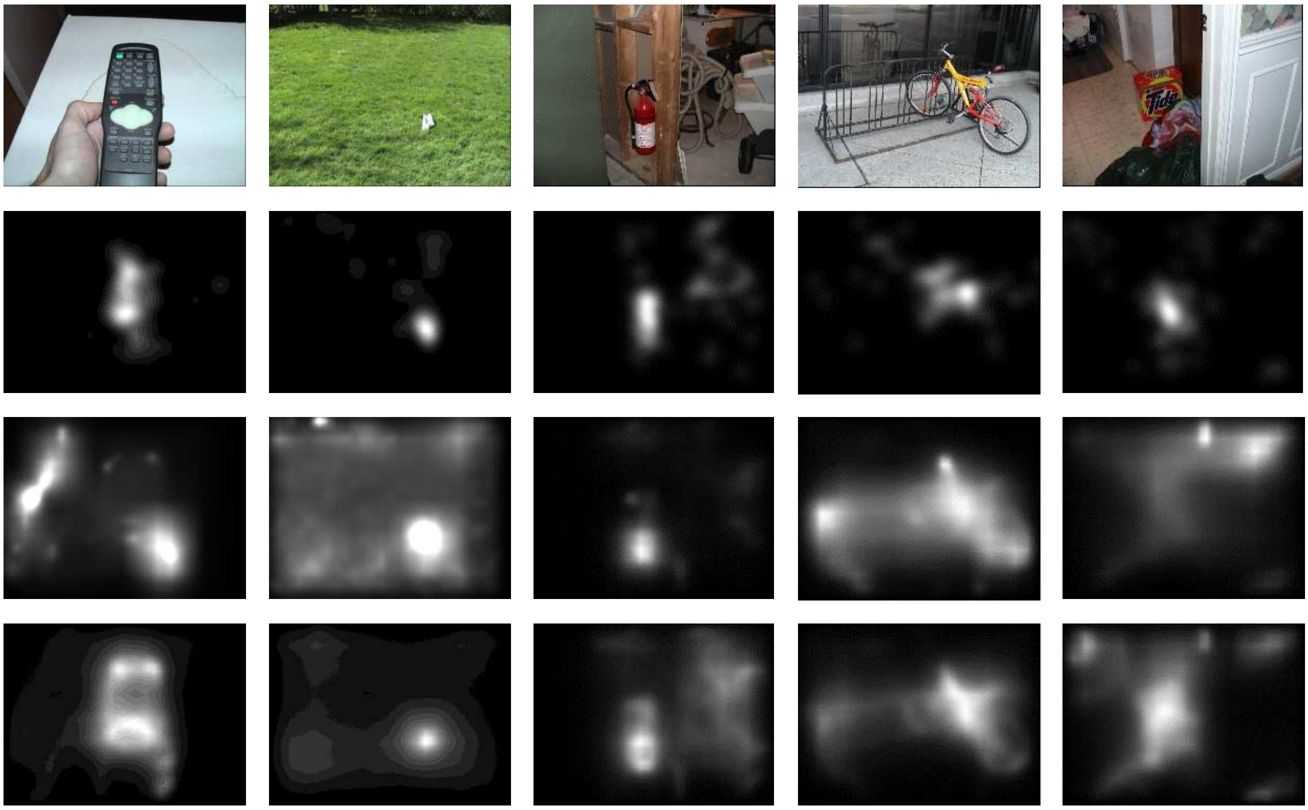


Figure 6: Visual comparison with typical saliency detection algorithms (from top to bottom: original images, human eye-tracking data, saliency map of ICL, saliency map of our method)

models lose color structure information during the reduced order approximation of the color image. In contrast, we use quaternion sparse model to represent high order signal without losing information between channels during sparse coding. With a view of center-surrounding contrast model, the reconstruction residual from quaternion sparse representation is more accurate to measure the saliency value of an image region.

Acknowledge

This work was supported in part by the National Natural S cience Foundation of China under Grant 61201384 and Gr ant61527804, in part by the 111 Project B07022.

References

[1] V. Navalpakkam and L. Itti. "An integrated model of top-down and bottom-up attention for optimizing detection speed". In CVPR, pages 2049–2056, 2006.

[2] U. Rutishauser, D.Walther, C. Koch, and P. Perona. "Is bottomup attention useful for object recognition?" In CVPR, pages 37–44, 2004.

[3] L. Itti and C. Koch. "Computational modeling of visual attention." Nature Reviews Neuroscience, 2(3):194–201, 2001.

[4] L. Itti, C. Koch, and E. Niebur. "A model of saliency-based visual attention for rapid scene analysis." In PAMI, 20:1254–1259, 1998.

[5] B A, Olshausen, Field D J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images.[J]. Nature, 1996, 381(6583):607-609.

[6] Han, B., Zhu, H., Ding, Y.: "Bottom-up saliency based on weighted sparse coding residual." In: Proceedings of the 19th ACM International Conference on Multimedia (MM), pp. 1117–1120 (2011)

[7] Li Y, Zhou Y, Xu L, et al. "INCREMENTAL SPARSE SALIENCY DETECTION". IEEE International Conference on Image Processing, 2009:3093 - 3096.

[8] Yi X, Licheng Y, Hongteng X, et al. Vector sparse representation of color image using quaternion matrix analysis.[J]. IEEE Trans Image Process, 2015, 24(4):1315 - 1329.

[9] Wright J, Ma Y, Mairal J, et al. "Sparse Representation for Computer Vision and Pattern Recognition". Proceedings of the IEEE, 2010, 98(6):1031 - 1044.

[10] Lilong Shi, Exploration in quaternion colour, Ph.D. thesis, School of Computing Science-Simon Fraser University, 2005.

[11] William Rowan Hamilton, "On quaternions; or on a new system of imaginaries in algebra," The London, Edinburgh,

and Dublin Philosophical Magazine and Journal of Science, vol. 25, no. 163, pp. 10–13, 1844.
[12] Farebrother, Richard William; Groß, Jürgen; Troschke, Sven-Oliver (2003). "Matrix representation of quaternions". Linear Algebra and its Applications (362): 251–255.
[13] Donoho D L. For most large underdetermined systems of linear equations the minimal $\ell 1$ - norm solution is also the sparsest solution[J]. Communications on pure and applied mathematics, 2006, 59(6): 797-829.
[14] Tibshirani R. Regression shrinkage and selection via the lasso[J]. Journal of the Royal Statistical Society. Series B (Methodological), 1996: 267-288.
[15] Bruce N, Tsotsos J. Attention based on information maximization[J]. Journal of Vision, 2007, 7(9): 950-950.
[16] Bruce N D B, Tsotsos J K. Saliency, attention, and visual search: An information theoretic approach[J]. Journal of vision, 2009, 9(3): 5
[17] Xu J, Jiang M, Wang S, et al. Predicting human gaze beyond pixels[J]. Journal of vision, 2014, 14(1): 28.
[18] Schauerte B, Stiefelhagen R. Predicting human gaze using quaternion dct image signature saliency and face detection[C]//Applications of Computer Vision (WACV), 2012 IEEE Workshop on. IEEE, 2012: 137-144.
[19] Guo C, Zhang L. A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression[J]. Image Processing, IEEE Transactions on, 2010, 19(1): 185-198.
[20] Perazzi F, Krahenbuhl P, Pritch Y, et al. Saliency filters: Contrast based filtering for salient region detection[C]//Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 2012: 733-740.
[21] Wei Y, Wen F, Zhu W, et al. Geodesic saliency using background priors[M]//Computer Vision–ECCV 2012. Springer Berlin Heidelberg, 2012: 29-42.
[22] Yang C, Zhang L, Lu H, et al. Saliency detection via graph-based manifold ranking[C]//Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on. IEEE, 2013: 3166-3173.
[23] Zhu W, Liang S, Wei Y, et al. Saliency optimization from robust background detection[C]//Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on. IEEE, 2014: 2814-2821.