PERCEPTUALLY MEANINGFUL QUADTREE DECOMPOSITION USING DUAL-HOMOGENEOUS CRITERIA

Zhaohui Che†*, Guangtao Zhai*†*, Ke Gu*†*, Yucheng Zhu*† *and Jiantao Zhou*‡

†Insti. of Image Commu. & Infor. Proce., Shanghai Jiao Tong University, Shanghai, China
‡Department of Computer and Information Science, University of Macau, Macau
Email: {†chezhaohui, †zhaiguangtao, †gukesjtuee, †zyc420}@sjtu.edu.cn,‡jtzhou@umac.mo

ABSTRACT

Saliency detection is a powerful tool for many applications because it can provide valuable information reflecting human visual attention. Quadtree decomposition is a classical and efficient image operator which can divide image into square blocks in different scales. The blocks in smaller size contain valuable information such as edges and texture. In image lossy compression or similar applications, we want to reserve valuable regions (salient objects) as much as we can, meanwhile, we also want to suppress non-salient regions human are not interested in. For these reasons, we propose an ingenious saliency guided quadtree decomposition model. The proposed model outperforms traditional methods in two aspects: It reserves the salient regions well, and the edges can be protected simultaneously.

***Index Terms*—** Saliency Detection, Quadtree Decomposition (QTD), Homogeneous Criterion (HC).

1. INTRODUCTION

Saliency detection is a popular topic which aims at extracting regions drawing human visual attention dramatically from images or videos [1]. Plenty of saliency models are proposed in past several years. Itti and Koch proposed the classical Itti&Koch saliency model [2] using center-surrounding operator to extract conspicuous regions from images in 1998. Bruce and Tsotsos proposed the AIM saliency model [3] using maximizing information principle to predict human fixation in 2005. Harel *et al.* came up with GBVS saliency model [4] based on graph theory and markovian approach in 2006. Xiaodi Hou *et al.* proposed an efficient saliency model [5] based on image signature in 2012. Ming Ming Cheng *et al.* proposed the regional contrast based saliency model [6] in 2014. Ke Gu *et al.* proposed a succinct and effective saliency model [7] based on free energy theory in 2015.

Both of saliency modelling and saliency application are potential research directions. Saliency detection have been commendably applied in segmentation [8], compression [9-10], recognition, tracking and retargeting so far. In most cases, saliency map containing valuable human visual attention information can provide guidance for specific tasks.

Quadtree decomposition is a classical and simple image operator which is widely used in information retrieval [11], image segmentation [12] and image compression [13]. As reported by Eli Shusterman and Meir Feder [14], the advantages of QTD are simplicity, adaptivity and useful output for further applications. But its biggest drawback is blocking artifact in reconstructed image.

In QTD approach, we start from the original image and divide the regions that dissatisfy homogeneous criterion into four square sub-blocks. The QTD will be executed iteratively until all sub-blocks satisfy homogeneous criterion. Ultimately, the image is divided into many non-overlapping blocks in different scales. The key of the QTD is how to choose homogeneous criterion (HC). HC is the metric which describes region consistency, and it reflects specific properties of image blocks. Traditional homogeneous criteria include mean square error (MSE), range value (the difference between maximum and minimum of the block), the maximal difference between each pixel and corresponding mean value of the block, color difference in RGB or Lab color space, *etc.*. Different HC are suitable for different image properties, and it's necessary to choose appropriate HC for specific tasks. Spann and Wilson [12] combined statistical and spatial information together to improve HC for image segmentation. Shusterman and Feder [14] adopted the improved MSE metric near to optimal choice of the threshold to image compression. We all know that, for human visual perception, the most valuable information of the image is embedded in the conspicuous regions (salient regions) which draw human visual attention dramatically. Therefore it's important to reserve salient regions in reconstructed image when we use QTD for lossy compression task. On the contrary, non-salient regions should be divided into bigger blocks so as to reduce transmission and storage cost. In this paper, we proposed an improved QTD method combined with saliency detection model to reserve the conspicuous regions well while suppressing non-salient regions severely.

The rest of this paper is organized as follows: In section

2, we introduce the saliency model based on free energy and the proposed QTD approach in detail. In section 3, we execute the proposed method in some complex natural images and give the experimental results compared with traditional method. We draw some conclusions and propose future development directions in section 4.

2. ALGORITHM DESIGN

In this section, we will introduce the proposed method in detail, and the specific pseudo-code is shown in Algorithm 1.

Algorithm 1 Saliency Guided QTD Algorithm

Input: original image I, threshold A

1. Pad I and make sure the size of I is $2^n \times 2^n$
2. Perform $T_1 = QT_{CD}(I, A)$
3. Perform $T_2 = QT_{SD}(I, A)$
4. Perform $Out_1 = Recover(T_2)$
5. Perform $Out_2 = EdgeRec(Out_1, T_1)$

Output: the reconstructed image Out_2 (Lossy Compression)

FUNCTION $QT_{CD}(I, A)$
while block k exists **do**
  Calculate CD_k as (1)
  if $CD_k > A$ **then**
    divide k into four blocks as k_1, k_2, k_3, k_4.
    in stack: $k+1$=k_1, $k+2$=k_2, $k+3$=k_3, $k+4$=k_4.
  else
    out stack: block k is labelled as $stop$
  end if
  $QT_{CD}(k+1, A)$
end while
ENDFUNCTION
FUNCTION $QT_{SD}(I, A)$
function QT_{SD} is just different from QT_{CD} in HC
We just change (Calculate CD_k as (1)) into (Calculate SD_k as $(2-5)$)
ENDFUNCTION

FUNCTION $Out_1 = Recover(T_2)$
for each block k in T_2 **do**
  Set all pixels of k as mean value of k in RGB space
end for
ENDFUNCTION

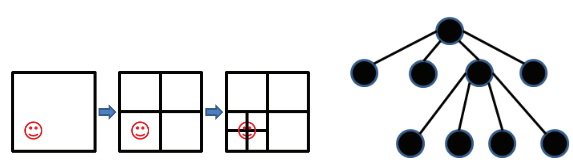
FUNCTION $Out_2 = EdgeRec(Out_1, T_1)$
set $Out_2 = Out_1$
for each block k in scale 1 in T_1 **do**
  Set corresponding block k in Out_2 as value of k in T_1.
end for
ENDFUNCTION

2.1. FES Saliency Model

The basic intention of this paper is taking advantages of saliency model's ability of predicting human visual attention to improve QTD approach. We design an ingenious homogeneous criterion to highlight salient regions in QTD while suppressing non-salient blocks. How to choose the appropriate saliency information is the key point of our method. The ideal saliency model should be accurate and has low computation complexity. Firstly, it's necessary to fix salient objects' positions accurately, therefore the saliency model should get good performance in AUC, CC and NSS metrics [1]. Secondly, natural images always contain multi-objects. As a consequence, the requisite saliency model should find out all of the salient objects. Thirdly, the requisite saliency model should be efficient in order to reduce the calculation complexity. Above all, we choose the FES saliency model proposed by Ke Gu and Guangtao Zhai *et.al* [7] in 2015 as our prior information. FES saliency model is based on free energy theory which reflects human brain feelings about "surprise" of stimuli. The "surprise" is related to saliency. The FES model calculates the local entropy of the gap between an input image signal and its predicted counterpart that is reconstructed from the input one with a semi-parametric model. The FES saliency model satisfies the three principles mentioned above.

2.2. Quadtree Decomposition

QTD is a classical image operator which divides image into non-overlapping square sub-blocks in different scales. The output of this operator is useful for image compression, image segmentation, edge detection, and image coding. Quadtree is also a kind of efficient data structure. This tree structure contains one root node and many leaf nodes. The root node represents the original image, and the leaf nodes represent the regions that are extremely dissimilar from surrounding regions based on the homogeneous criterion. Each intermediate node has one parent node and four child nodes. The quadtree structure is shown in Fig.1. After decomposition, the larger blocks are flat and contain less information, on the contrary, the smaller blocks contain important objects such as smiling face in Fig.1 .



(a) Quadtree Decomposition (b) Data Structure

Fig. 1. A simple example of QTD and its data structure. The region contains smiling face will be divided into sub-blocks because it's dissimilar from surrounding area.

2.3. Dual-Homogeneous Criteria

Homogeneous criterion (HC) is the key of the QTD. The same image will be divided into different structures by different HC. How to design HC is an important topic for specific tasks. In this paper, we want to achieve an perceptually meaningful QTD which can not only highlight salient objects but also protect edges. We design a dual-homogeneous criteria to protect salient regions and edges respectively.

Color difference (CD) is a popular and efficient HC in QTD especially for color images. We introduce the color difference in RGB color space in equation 1.

$$CD_k = \sqrt{\frac{\sum_{i,j=1}^{size_k}((R_{ij}-\bar{R}_k)^2+(G_{ij}-\bar{G}_k)^2+(B_{ij}-\bar{B}_k)^2)}{size_k^2}} \tag{1}$$

Where R_{ij}, G_{ij}, B_{ij} are red, green and blue components of the pixel (i, j) in the block k. Accordingly, $\bar{R}_k, \bar{G}_k, \bar{B}_k$ represent mean values of red, green and blue components in block k. $size_k$ is the scale (height of the square block) of the block k. The CD is a kind of mean square error (MSE) in RGB color space. We find that this CD HC has excellent edge extraction ability. We will show its edge extraction performance in section 3.

Before we execute the QTD, we calculate the saliency map SM of the original image by FES saliency model. After getting SM, we will introduce three terms $AvSM_k$, $RaSM_k$ and $JuSM_k$ in equation 2, 3 and 4. $AvSM_k$ is the mean value of the kth block's saliency value. $RaSM_k$ is the range value (the difference between maximum and minimum) of the kth block's saliency value. $JuSM_k$ is the maximum of $AvSM_k$ and $RaSM_k$.

$$AvSM_k = \frac{\sum_{i,j=1}^{size_k} SM_{ij}}{size_k \times size_k} \tag{2}$$

$$RaSM_k = Max(SM_{ij}|(i,j) \in k) - Min(SM_{ij}|(i,j) \in k) \tag{3}$$

$$JuSM_k = Max(AvSM_k, RaSM_k) \tag{4}$$

We design the saliency guided SD HC in equation 5.

$$SD_k = \alpha \times \frac{1}{1+e^{-(JuSM_k - \bar{SM})}} \times CD_k + (1-\alpha) \times AvSM_k \tag{5}$$

Where α is the free parameter (default value is 0.5), and the $\bar{SM}$ is the global mean value of the original image's saliency map. The $\bar{SM}$ is regarded as a threshold to judge if the block k is salient region. The equation 5 is made up of two parts. The former contains three factors: coefficient α, a sigmoid function, and the color difference. It's worth noting that the saliency value plays the dominated role in QTD. Therefore, we adopt the sigmoid function as the weight of color difference. As we can see in equation 5, if the $JuSM_k - \bar{SM}$ is positive, the $\frac{1}{1+e^{-(JuSM_k - \bar{SM})}}$ tends to 1, otherwise it tends to 0. Above all, this sigmoid function highlights the blocks whose $JuSM_k$ exceeds threshold $\bar{SM}$ while suppressing blocks whose $JuSM_k$ is less than threshold. The reason why we adopt $JuSM_k$ is that we want to guarantee that the QTD will not stop until the saliency objects are divided into subtle blocks. To be specific, when we start from initial image, the observation tells us that the $AvSM_k$ is vary small since there are plenty of non-salient pixels, but the $RaSM_k$ is large. In addition, in deep-level decomposition, blocks contain salient objects are made up of pixels with similar saliency value, therefore the $RaSM_k$ is very small. Fortunately, we find that the $AvSM_k$ of these blocks are large. The second part of the equation 5 is $AvSM_k$. Above all, the SD HC can protect salient objects well, and the CD HC can protect edges well.

3. EXPERIMENTAL RESULTS

We select some natural images from Bruce and Tsotsos database [3] as test images. Some experimental results for demonstrating the proposed model are shown in Fig. 2. The second column in Fig.2 are saliency maps from FES model. We find that these saliency maps are accurate and integrated. Both of foreground and background can be extracted so that the complex images with multi salient objects can be reserved well such as the first three rows in Fig.2. Our method also produce an accurate edge map shown in the third column. The edges are extracted by CD HC because either sides near the edges have strong color contrast. The edge map is applied to compensate for the edge detail of non-salient regions to improve the final result.

Experimental results of traditional method are shown in the fourth column. Traditional method highlights texture and edges regardless of conspicuousness. For examples, in the first row, the boundaries of the "desk" are reserved well, but the "tellurion" on the desk and "sundries" under the desk contain severe blocking artifacts. In the second row, the traditional method highlights texture of the "tablecloth" but ignores salient "tomatoes" and "teacup". In the third case, we can find that there are some blocking artifacts in the "red dumbbells" and "gray iron plates". On the contrary, the texture of "carpet" are reserved well. In the final case, the "human face" is compressed severely. These results are unsatisfactory because non-salient texture wastes too much storage and transmission cost.

The last two columns are results of the proposed method. The difference between them are edges. The edge maps shown in the third column can provide coordinates and color information for the final results such as "chair" in the second row, texture of "carpet" in the third row and "books" in the right region of the fourth row. Experimental results imply that all of the salient objects and subtle edges can be protected well to lower damage to image quality after lossy compression. Non-salient regions are divided into big blocks so as to reduce storage and transmission cost dramatically.

Fig. 2. From left to right each column represents original images, saliency maps, edge maps, compressed image by traditional method whose HC is color difference, results of our method before edge compensation, and final results of the proposed method.

4. CONCLUSION

We propose a saliency guided QTD model which can reserve salient objects and edges well in this paper. The proposed model is very helpful to many applications related to human visual perception, such as lossy compression. Traditional QTD approach is blind and wastes much cost to store non-salient texture and edges. Experimental results implicit that the proposed model gets good performance in most of complex natural images. We hope to develop more perceptually meaningful applications combined with saliency model in the future.

5. ACKNOWLEDGEMENT

This work was supported in part by the National Science Foundation of China under Grants 61422112, 61371146, 61221001, the Foundation for the Author of National Excellent Doctoral Dissertation of China under Grant 201339.

6. REFERENCES

[1] Borji A, Itti L. "State-of-the-art in visual attention modeling." *Pattern Analysis and Machine Intelligence,* IEEE Transactions on, 2013, 35(1): 185-207.

[2] Itti L, Koch C, Niebur E. "A model of saliency-based visual attention for rapid scene analysis." IEEE Transactions on *Pattern Analysis & Machine Intelligence*, 1998 (11): 1254-1259.

[3] Bruce N, Tsotsos J. "Saliency based on information maximization." *Advances in neural information processing systems.* 2005: 155-162.

[4] Harel J, Koch C, Perona P. "Graph-based visual saliency." *Advances in neural information processing systems.* 2006: 545-552.

[5] Hou X, Harel J, Koch C. "Image signature: Highlighting sparse salient regions." *Pattern Analysis and Machine Intelligence,* IEEE Transactions on, 2012, 34(1): 194-201.

[6] Cheng M, Mitra N J, Huang X, et al. "Global contrast based salient region detection." *Pattern Analysis and Machine Intelligence,* IEEE Transactions on, 2015, 37(3): 569-582.

[7] Gu K, Zhai G, Lin W, et al. "Visual Saliency Detection With Free Energy Theory." *Signal Processing Letters,* IEEE, 2015, 22(10): 1552-1555.

[8] Kim T H, Lee K M, Lee S U. "Generative image segmentation using random walks with restart." *Computer Vision ECCV 2008.* Springer Berlin Heidelberg, 2008: 264-275.

[9] Guo C, Zhang L. "A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression." *Image Processing,* IEEE Transactions on, 2010, 19(1): 185-198.

[10] Dhavale N, Itti L. "Saliency-based multifoveated MPEG compression." *Signal processing and its applications,* 2003. Proceedings. Seventh international symposium on. IEEE, 2003, 1: 229-232.

[11] Finkel R A, Bentley J L. "Quad trees a data structure for retrieval on composite keys." *Acta informatica,* 1974, 4(1): 1-9.

[12] Spann M, Wilson R. "A quad-tree approach to image segmentation which combines statistical and spatial information." *Pattern Recognition,* 1985, 18(3): 257-269.

[13] Markas T, Reif J. "Quad tree structures for image compression applications." *Information Processing & Management,* 1992, 28(6): 707-721.

[14] Shusterman E, Feder M. "Image compression via improved quadtree decomposition algorithms." *Image Processing,* IEEE Transactions on, 1994, 3(2): 207-215.