

LOW-ENERGY GRAPH FOURIER BASIS FUNCTIONS SPAN SALIENT OBJECTS

Junaid Malik, Caglar Aytekin, Moncef Gabbouj

Tampere University of Technology

ABSTRACT

There is an emerging interest aiming at defining principles for signals on general graphs, which are analogous to the basic principles in traditional signal processing. One example is the Graph Fourier Transform which aims at decomposing a graph signal into its components based on a set of basis functions with corresponding graph frequencies. It has been observed that most of the important information of a graph signal is contained inside the low frequency band, which leads to several applications such as denoising, compression, etc. In this paper, we show that the low frequency basis functions span the salient regions in an image, which can also be considered as important regions. Motivated by this, we present a novel simple and unsupervised method to utilize a number of low-energy basis functions and show that it improves the performance of seven state-of-the-art salient object detection methods in five datasets under four different evaluation criteria, with only minor exceptions.

Index Terms— graph fourier transform, salient object detection, graph signal processing

1. INTRODUCTION

There is an increasing interest towards the extension of classical discrete signal processing principles to signals defined over graphs, commonly referred to as graph signals. The need for this arises from the fact that some data is naturally encoded in an irregular mesh and can only be represented as network-like structures such as graphs. Examples include data generated by interactions of cellular systems in biology, social networks, power grids and transportation networks [1]. Moreover, graphs provide a more sophisticated representation, e.g. by offering the flexibility to incorporate deeper connectivity, which makes them suitable also for data that can be encoded in regular grids. Therefore, a variety of tools proposed in the recent literature introduce the classical signal processing notions of filtering, transforms and sampling for graph signal [1, 2].

In classical signal processing, Fourier transform provides the expansion of a signal in terms of complex weighted exponentials (spectral components), collectively termed as the Fourier basis functions. An accurate approximation of a time domain signal can be achieved by utilizing only a portion of

the low frequency Fourier basis functions, provided that the signal is smooth i.e. has small local variations. In order to exploit this phenomenon for signals defined on graphs, the notion of "Graph Fourier Transform" (GFT) has been introduced [1, 3]. When dealing with graph signals, smoothness refers to the fact that signal values associated with neighbouring nodes are close to one another [4]. Similar to the case of the time-domain signals, smooth graph signals can be approximated using only a few low frequency graph Fourier basis functions. A popular approach based on spectral graph theory is to employ the eigenvectors of graph matrices, such as Laplacian, as the graph Fourier basis [4]. GFTs have found particularly successful application in coding, denoising and compression of digital images [3, 5, 6, 7].

In this paper we utilize the properties of GFT for the task of salient object detection in images. Salient object detection aims at finding a saliency distribution map identifying the most visually appealing and distinct regions of an image [8]. In this paper, we claim that the saliency map, defined as a signal on a graph representation of the image, can accurately be approximated using only the lower frequency basis functions. We test this claim by projecting ground truth saliency maps to a few low energy graph Fourier basis functions and observe that the resulting representation is accurate across a variety of evaluation metrics. Motivated by this, we claim that a saliency map, obtained using any salient object detection method, can be improved by mapping it to a space spanned only by the low frequency basis functions. Our contribution in this paper is thus twofold: 1) graph signals that can be accurately spanned by low frequency graph basis functions have earlier been shown to correspond to only smooth graph signals. Similarly, we show that saliency maps are also mainly encoded in a low-frequency band, although they are not necessarily smooth. 2) Exploiting this finding, we present an efficient unsupervised denoising method for any saliency map.

The rest of the paper is organized as follows. In Section 1.1, we present a brief overview of earlier work in image processing utilizing graph-based transforms. Section 2 elucidates our proposed method to improve the results of salient object detection methods. Section 3 presents the results of our method evaluated using a variety of evaluation metrics. Finally, Section 4 concludes the paper and suggests ideas for further research.

1.1. Related Work

We first introduce some basic concepts related to graph Fourier transforms. Given a weighted graph, $G = (V, E)$ consisting of nodes V connected by edges E , the weight matrix W is a square matrix of size $N \times N$ ($N = |V|$) and w_{ij} is the weight of the edge connecting nodes i and j . In case of an undirected graph, W is a symmetric matrix as $w_{ij} = w_{ji}$. The degree matrix D , is a diagonal matrix of size $N \times N$ where each diagonal entry, d_{ii} , is the summation of edge weights incident on node i . A well-known matrix describing the graph is the Laplacian matrix, defined as

$$L = D - W \quad (1)$$

Eigen-decomposition of this matrix is commonly used for frequency analysis of graphs [4]. The eigenvectors are considered as the graph Fourier basis functions and the associated eigenvalues provide a notion of graph frequencies [1]. This property of the graph Laplacian has been utilized in a variety of image processing applications. In [9], a few low energy eigenvectors of a modified Laplacian matrix are linearly superposed to generate category independent object proposals. In [5], the eigenvectors of the Laplacian matrix are used for calculating transform coefficients for coding edge-maps of image-blocks to achieve compression with a lower bitrate as compared to traditional methods. Similarly, a multiresolution Graph-Based Transform (GBT) introduced in [7] employs the eigenbasis of the Laplacian for compressing depth-maps. In [3], the generalized eigenvectors obtained using Jordan decomposition of the adjacency matrix are employed as Fourier basis, which is then utilized for image compression. In [6], images are denoised by projecting them onto the space spanned by the first few eigenvectors (lower frequency components) of the Laplacian. This is based on the assumption that the structural information about the regions in the uncorrupted image is encoded in the lower frequency basis functions.

Similar to the findings highlighted in the works mentioned above, we claim that in images, the salient regions can be regarded as the useful information. We then experimentally validate our claim by showing that salient regions of an image can accurately be spanned by low-energy graph basis functions. Similar to the image denoising process in [6], we denoise a saliency map, by mapping it to a low energy basis space. This results in a notable improvement in performance across a variety of evaluation metrics.

2. PROPOSED METHOD

2.1. Graph Structure

Given an input image, we first oversegment it into multiple homogeneous regions (superpixels) using SLIC algorithm [10]. In the graph-based representation of the image, the

nodes represent the superpixels. The edge weight connecting two nodes is inversely related to the distance in the Lab colour space, as in [11]. Moreover, the mean value of Lab colour for all pixels inside a superpixel is chosen as its representative colour. The edge weights are also normalized based on the number of neighbours the connecting nodes have. Finally, an increased neighbourhood approach is adopted where each node is connected to up to its fifth set of neighbours. The n_{th} set of neighbourhood of node i refers to the nodes, which are between, $\text{floor}(2^{n-2} + 1)$ and 2^{n-2} spatial connections away from node i . This expanded neighbourhood helps incorporate longer connectivity information and helps in embedding global contrast information in addition to local contrast. The final edge weight assignment between two nodes i and j , has the following mathematical form:

$$w_{ij} = \left(\frac{1}{\epsilon + d_{LAB}(i, j)^2} \right) * M(i, j) \quad (2)$$

$$M(i, j) = \frac{1}{(|N_{i, C(i, j)}| * |N_j|)^2 + (|N_{j, C(j, i)}| * |N_i|)^2} \quad (3)$$

In (2), $d_{LAB}(., .)$ is the Euclidean distance in the Lab colour space, ϵ is a small number to avoid division by zero and $M(i, j)$ is the normalization term defined as in (3). In (3), $|N_i|$ is the number of nodes in the neighbourhood of node i and $|N_{i, C(i, j)}|$ refers to the number of nodes in the $C(i, j)$ th set of neighbourhood of i . $C(i, j)$ denotes the set containing the neighbourhood levels between nodes i and j . We make the asymmetric weight assignment of [11] symmetric by adding $(|N_{j, C(j, i)}| * |N_i|)^2$ to the denominator of the normalization term $M(i, j)$.

2.2. Graph Fourier Transform

In addition to the graph Laplacian formed by the weights in (2), [11] incorporates the boundary potential by adding a matrix V to the Laplacian L . V is a diagonal matrix whose diagonal entries have a positive value for nodes representing the boundary of the image and are zero elsewhere. This introduces a bias in favour of treating the boundary of the image as background and is based on the assumption that salient objects are less likely to occupy the borders of an image [12, 13, 14, 15]. In this work, we keep this value quite small in order to handle objects touching the image boundaries as well. The modified Laplacian takes the form

$$L_M = D - W + V \quad (4)$$

Being a real and symmetric graph, L_M has real eigenvalues and a complete set of corresponding orthonormal eigenvectors, which form an orthogonal graph Fourier basis. This orthogonality is a desired trait, as such transforms are known to redistribute energy of a signal in such a way that most of the useful information is encapsulated within a small fraction of components [16].

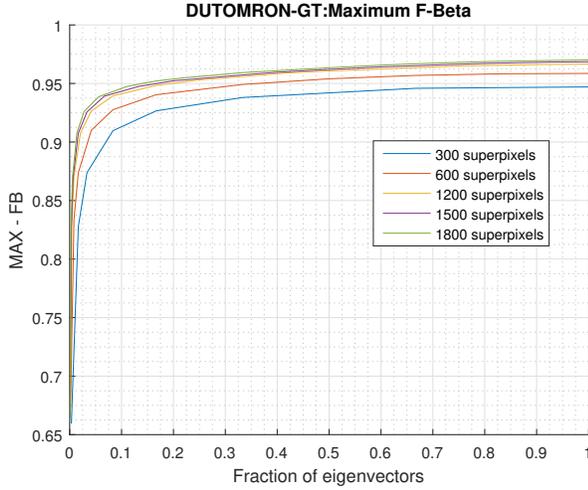


Fig. 1. Performance of projected ground truth saliency maps vs fraction of eigenvectors utilized

2.3. Methodology

We first construct the graph representation of the original image and obtain the modified Laplacian matrix as described in Sections 2.1 and 2.2. The eigenvalue decomposition of L_M yields the eigenvectors $\{\mathbf{u}_{i=0}^{n-1}\}$ (basis functions) and the corresponding eigenvalues (frequencies). The K eigenvectors, corresponding to the K lowest eigenvalues, are chosen as the basis functions, assumed to be spanning the salient object. The choice of K is critical as it defines the fraction of spectral components employed. In our experiments, we restrict ourselves to the lower 1/3rd of the spectrum and keep the exact value of K as a design parameter.

The pixel-wise saliency map \mathbf{S} of the image, obtained using any salient object detection method, is first converted to a superpixel representation $\tilde{\mathbf{S}}$ using the mapping obtained by running SLIC on the original image and averaging the per-pixel saliency value inside each superpixel. We now have an $N \times 1$ vector, $\tilde{\mathbf{S}}$, providing a superpixel representation of the original saliency map, where N is the total number of superpixels. This is a signal defined on the graph and its projection onto the space spanned by the selected basis functions, K lowest energy eigenvectors, is given as follows:

$$\mathbf{S}_{\text{proj}}^K = \mathbf{u}_1 c_1 + \mathbf{u}_2 c_2 + \mathbf{u}_3 c_3 + \dots + \mathbf{u}_K c_K \quad (5)$$

where the coefficients $c_1, c_2, c_3, \dots, c_K$ are calculated as

$$c_n = \tilde{\mathbf{S}} \cdot \mathbf{u}_n \quad (6)$$

In (6), \cdot represents the scalar product between two vectors. Finally, $\mathbf{S}_{\text{proj}}^K$ is converted back to the image domain by replicating the normalized saliency value of each superpixel across all the pixels constituting that superpixel.

3. EVALUATION

3.1. Metrics

In order to evaluate the performance of the projected saliency maps, we make use of multiple evaluation metrics as in [25]. The mean square error (MSE) measures the average pixel wise square of the difference between the ground truth and the saliency map. The F_β measure is a weighted harmonic mean of the precision and recall of the saliency map and provides an evaluation based on both of these metrics as defined in (7)

$$F_\beta = \frac{(1 + \beta^2) \text{Precision} * \text{Recall}}{(\beta^2 * \text{Precision}) + \text{Recall}} \quad (7)$$

As in [17], we set β^2 to 0.3 in order to give more weight to precision. We analyze both the maximum and mean values for the measure in our experiments. We also calculate the AUC measure, which is the area under the ROC (Receiver Operating Characteristics) curve. The ROC curve is obtained by plotting the true positive rate against the false positive rate of detection.

3.2. Experiments

For our experiments, we evaluate on five different datasets; SED2 [18], JUDD [8], ECSSD [14], DUTOMRON [15] and MSRA10K [19]. Moreover, we explore five superpixel resolutions corresponding to 300, 600, 1200, 1500 and 1800 superpixels per image. For each resolution, we use the lower 1/3rd of the spectrum by only considering the top 1/3rd of the total number of eigenvectors (sorted in ascending order of eigenvalues). We start from the lowest energy basis function and then gradually increase the coverage by utilizing comparatively higher energy basis functions. As a preliminary experiment, we project the ground truth saliency map onto the space spanned by the selected basis functions. The performance of the projected ground truth saliency map is then evaluated using the metrics defined in Section 3.1. In Figure 1, we plot the evaluation results of this experiment on DUTOMRON dataset for the maximum F_β metric. As shown in the figure, the rate of increase in performance is maximum in the lower end of the spectrum, thus reaffirming our hypothesis that most of the information concerning the salient object is contained inside the lower energy basis functions. This trend is generally consistent across all datasets and evaluation metrics [9].

We then proceeded to apply the same operation to saliency maps obtained from the methods of DRFI[20], DSR[21], RBD[22], MC[12], ST[23], MDF[24] and DCL[25]. The experiments were conducted on each of the five datasets and the denoised saliency maps were evaluated using the metrics described in Section 3.1. The saliency maps for the methods of [20],[21],[22],[12] and [23] were obtained from a publicly available source [17]. For [24] and [25], the original maps were acquired from the authors' online page [26].

Table 1. Comparison of original vs denoised saliency maps on ECSSD dataset

| | DRFI | MC | ST | RBD | DSR | MDF | DCL+ | GT |
|-------------------|----------------------------------|--------------------------------|---------------------------------|---------------------------------|--------------------------------|----------------------------------|----------------------------------|---------------------------|
| AUC | 0.947 | 0.924 | 0.918 | 0.918 | 0.912 | 0.912 | 0.912 | 0.999 _{600,1800} |
| | 0.950 _{300,1800} | 0.937 _{1,600} | 0.937 _{1,600} | 0.937 _{1,600} | 0.937 _{1,600} | 0.948 _{600,1800} | 0.969 _{50,1200} | |
| MSE | 0.086 | 0.107 | 0.098 | 0.116 | 0.119 | 0.087 | 0.051 | 0.019 _{600,1800} |
| | 0.083 _{100,300} | 0.103 _{10,600} | 0.096 _{50,600} | 0.109 _{50,1800} | 0.109 _{10,300} | 0.080 _{100,300} | 0.051 _{200,600} | |
| F_{β} -MAX | 0.786 | 0.742 | 0.752 | 0.718 | 0.737 | 0.831 | 0.898 | 0.980 _{600,1800} |
| | 0.804 _{25,600} | 0.781 _{1,1500} | 0.781 _{1,1500} | 0.781 _{1,1500} | 0.781 _{1,1500} | 0.845 _{50,1200} | 0.900 _{400,1500} | |
| F_{β} -MEAN | 0.649 | 0.596 | 0.621 | 0.610 | 0.629 | 0.800 | 0.824 | 0.866 _{600,1800} |
| | 0.653 _{50,300} | 0.618 _{1,1800} | 0.622 _{100,300} | 0.618 _{1,1800} | 0.619 _{100,300} | 0.741 _{200,600} | 0.800 _{600,1800} | |

Table 2. Comparison of original vs denoised saliency maps on DUTOMRON dataset

| | DRFI | MC | ST | RBD | DSR | MDF | DCL+ | GT |
|-------------------|---------------------------------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|----------------------------------|----------------------------------|---------------------------|
| AUC | 0.934 | 0.886 | 0.895 | 0.888 | 0.899 | 0.761 | 0.906 | 0.999 _{600,1800} |
| | 0.927 _{600,1800} | 0.903 _{1,600} | 0.903 _{1,600} | 0.903 _{1,600} | 0.903 _{1,600} | 0.923 _{500,1500} | 0.936 _{300,1500} | |
| MSE | 0.072 | 0.091 | 0.091 | 0.083 | 0.085 | 0.073 | 0.059 | 0.014 _{200,600} |
| | 0.0712 _{5,1200} | 0.079 _{1,600} | 0.079 _{5,600} | 0.077 _{5,600} | 0.078 _{5,600} | 0.067 _{50,300} | 0.057 _{100,300} | |
| F_{β} -MAX | 0.665 | 0.627 | 0.631 | 0.630 | 0.626 | 0.694 | 0.751 | 0.959 _{600,1800} |
| | 0.681 _{10,300} | 0.673 _{1,600} | 0.673 _{1,600} | 0.673 _{1,600} | 0.673 _{1,600} | 0.720 _{10,600} | 0.756 _{25,600} | |
| F_{β} -MEAN | 0.541 | 0.517 | 0.523 | 0.540 | 0.546 | 0.668 | 0.680 | 0.861 _{600,1800} |
| | 0.577 _{1,600} | 0.577 _{1,600} | 0.577 _{1,600} | 0.577 _{1,600} | 0.577 _{1,600} | 0.628 _{100,300} | 0.669 _{200,600} | |

Table 1 and Table 2 document the evaluation results on ECSSD and DUTOMRON datasets. These datasets are selected because the saliency maps for all methods are available for them. In each table, a pair of rows corresponds to a single evaluation metric and the columns represent the methods. The top row for each metric shows the values for the original saliency maps while the bottom row shows the values for the best performing projection. The subscripts refer to the fraction of available spectrum utilized. For example, 600,1800 implies that 600 out of the available spectrum of 1800 basis functions are used. The evaluation metric value is emphasized in cases where the denoised map outperforms the original one. We observe that in majority of the cases, our method enhances the original saliency map performance. For detailed results on all datasets and methods, the reader is referred to publicly available supplement ¹.

Figure 2 shows some visual examples of the improvements made on an image from DUTOMRON dataset. Furthermore, the proposed method takes less than 3 seconds, including SLIC, for an image of size 1024x768 at the highest superpixel resolution of 1800 superpixels per image. The running time is significantly reduced for lower superpixel resolutions.

4. CONCLUSION AND FUTURE WORK

In this paper, we explore a new application for graph Fourier transform by utilizing it to improve the performance of salient object detection methods. In our preliminary experiments, we found that most of the information pertaining to the salient

object in an image is contained inside the lower frequencies of the graph representing the image. Our experiments proved that we can utilize the Fourier basis functions corresponding to these low frequencies to improve the performance of saliency maps. Furthermore, the percentage improvement achieved was found to be dependent on the two design parameters; superpixel resolution and the fraction of spectrum defined as low frequency. Our future work involves investigating ways to learn or infer these parameters directly from the properties of the image and the saliency map.



Fig. 2. Saliency Maps: (from left to right) Ground Truth, Original obtained using method of [20] and its denoised version using the proposed method.

5. ACKNOWLEDGEMENTS

This work is funded by Qatar National Research Funding (QNRF) within the NPRP project, NPRP9-390-1-088, SIMUPOR.

6. REFERENCES

- [1] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The Emerging Field of Signal Processing on Graphs: Extending High-Dimensional Data

¹<http://104.131.82.198/salobjdetection/gft/>

- Analysis to Networks and Other Irregular Domains,” oct 2012.
- [2] A. Sandryhaila and J. M. F. Moura, “Discrete Signal Processing on Graphs,” *IEEE Transactions on Signal Processing*, vol. 61, no. 7, pp. 1644–1656, apr 2013.
- [3] A. Sandryhaila and J. M. F. Moura, “Discrete signal processing on graphs: Graph fourier transform,” in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, may 2013, pp. 6167–6170, IEEE.
- [4] X. Zhu and M. Rabbat, “Approximating signals supported on graphs,” in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, mar 2012, pp. 3921–3924, IEEE.
- [5] G. Shen, W.-S. Kim, S. K. Narang, A. Ortega, J. Lee, and H. Wey, “Edge-adaptive transforms for efficient depth map coding,” in *28th Picture Coding Symposium*, dec 2010, pp. 566–569, IEEE.
- [6] F. Zhang and E. R. Hancock, “Graph spectral image smoothing using the heat kernel,” *Pattern Recognition*, vol. 41, no. 11, pp. 3328–3342, 2008.
- [7] W. Hu, G. Cheung, X. Li, and O. Au, “Depth map compression using multi-resolution graph-based transform for depth-image-based rendering,” in *2012 19th IEEE International Conference on Image Processing*, sep 2012, pp. 1297–1300, IEEE.
- [8] A. Borji, “What is a Salient Object? A Dataset and a Baseline Model for Salient Object Detection,” *IEEE TIP*, vol. 24, no. 2, pp. 742–756, feb 2015.
- [9] J. Malik, Ç. Aytekin, and M. Gabbouj, “Category Independent Object Proposals Using Quantum Superposition,” in *2017 IEEE International Conference on Image Processing, 2017*, number September.
- [10] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, “SLIC Superpixels,” *EPFL Technical Report 149300*, , no. June, pp. 15, 2010.
- [11] Ç. Aytekin, E. C. Ozan, S. Kiranyaz, and M. Gabbouj, “Extended quantum cuts for unsupervised salient object extraction,” *Multimedia Tools and Applications*, 2016.
- [12] B. Jiang, L. Zhang, H. Lu, C. Yang, and M.-H. Yang, “Saliency Detection via Absorbing Markov Chain,” in *2013 IEEE ICCV*, dec 2013, pp. 1665–1672, IEEE.
- [13] W. Zhu, S. Liang, Y. Wei, and J. Sun, “Saliency Optimization from Robust Background Detection,” in *2014 IEEE CVPR*, jun 2014, pp. 2814–2821, IEEE.
- [14] Q. Yan, L. Xu, J. Shi, and J. Jia, “Hierarchical Saliency Detection,” in *2013 IEEE CVPR*, jun 2013, pp. 1155–1162, IEEE.
- [15] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang, “Saliency Detection via Graph-Based Manifold Ranking,” in *2013 IEEE CVPR*, jun 2013, pp. 3166–3173, IEEE.
- [16] R. Wang, *Introduction to orthogonal transforms : with applications in data processing and analysis*, Cambridge University Press, 2012.
- [17] A. Borji, M.-M. Cheng, H. Jiang, and J. Li, “Salient Object Detection: A Benchmark,” *IEEE TIP*, vol. 24, no. 12, pp. 5706–5722, dec 2015.
- [18] S. Alpert, M. Galun, A. Brandt, and R. Basri, “Image Segmentation by Probabilistic Bottom-Up Aggregation and Cue Integration,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 2, pp. 315–327, feb 2012.
- [19] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S.-M. Hu, “Global Contrast Based Salient Region Detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 569–582, mar 2015.
- [20] J. Wang, H. Jiang, Z. Yuan, M.-M. Cheng, X. Hu, and N. Zheng, “Salient Object Detection: A Discriminative Regional Feature Integration Approach,” *International Journal of Computer Vision*, pp. 1–18, 2016.
- [21] X. Li, H. Lu, L. Zhang, X. Ruan, and M.-H. Yang, “Saliency Detection via Dense and Sparse Reconstruction,” in *2013 IEEE International Conference on Computer Vision*, dec 2013, pp. 2976–2983, IEEE.
- [22] P. Viola and M. Jones, “Robust Real-Time Face Detection,” *IJCV*, vol. 57, no. 2, pp. 137–154, 2004.
- [23] Zhi Liu, Wenbin Zou, and O. Le Meur, “Saliency Tree: A Novel Saliency Detection Framework,” *IEEE TIP*, vol. 23, no. 5, pp. 1937–1952, may 2014.
- [24] Guanbin Li and Y. Yu, “Visual saliency based on multi-scale deep features,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, jun 2015, pp. 5455–5463, IEEE.
- [25] G. Li and Y. Yu, “Deep Contrast Learning for Salient Object Detection,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, jun 2016, pp. 478–487, IEEE.
- [26] L. Guanbin and Y. Yizhou, “hkuis - Guanbin Li’s Homepage,” 2016, Available: <https://sites.google.com/site/lighb86/hkuis> [Accessed April 4, 2017].