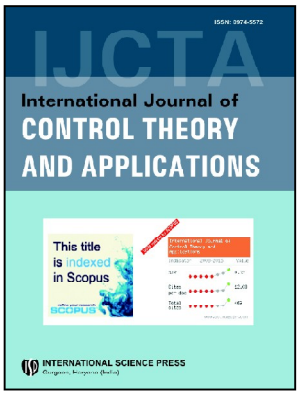

International Journal of Control Theory and Applications

ISSN : 0974-5572

Volume 10 • Number 22 • 2017

Visual Saliency Detection of Stereoscopic 3D images based on Scale-Space Analysis

[1]Rakesh Y. and [2]K. Sri Rama Krishna

[1] *Asst. Prof, Dept. of ECE, SRKIT, Vijayawada, India, E-mail: rakesh.yemineni@gmail.com*
[2] *Prof, Dept. of ECE, VRSEC, Vijayawada, India, E-mail: srk_kalva@yahoo.com*

Abstract: Visual saliency alludes to the particular fixation on prominent or significant regions in a scene that have likewise been appeared to relate with vital objects and their relationships. Visual saliency is crucial for human visual experience furthermore important to numerous applications. Visual consideration - especially stimulus driven, saliency-based detection has been a dynamic research field over the previous decades. In this paper, we present a novel saliency recognition model for 3D pictures in view of highlight difference from luminance, color, and depth. Difference of the stereo pair is extricated utilizing sliding window strategy. Here we consider visual saliency as a different issue in three perspectives. Primarily we consider saliency area as a repeat space examination issue. Secondly, we finish this by using the possibility of non-saliency. At last, we as a whole the while consider the recognizable proof of remarkable areas of different size. The paper proposes another bottom up perspective for distinguishing visual saliency, depicted by a scale-space examination of the ampleness scope of regular images. To detect salient images the image amplitude spectrum is convolved with a low-pass Gaussian part of a reasonable scale. A HFT plays out the examination in the frequency domain.

Keywords: Visual attention, Saliency, luminance, Depth, Stereoscopic images, HFT.

I. INTRODUCTION

VISUAL consideration encourages our capacity to quickly find the most essential data in a scene [1], [2]. Such picture regions are said to be salient since it is expected that they pull in more noteworthy consideration by the visual framework than different parts of the image. These salient regions are relied upon to have particular components at the point when contrasted and others in the picture. The investigation of saliency attention may uncover the detectional instruments of biological visual frameworks, and in addition display their obsession determination conduct.

Then again, as a part of low-level counterfeit vision handling, it encourages consequent preparing, for example, question discovery on the other hand acknowledgment by diminishing computational cost, which is a key thought continuously applications. For object detection, this would dependably be more proficient than dense sampling, if one could guarantee the exactness of the attentional mechanism.

Saliency starts from visual distinctiveness, unconventionality, irregularity, or astonish, and is regularly ascribed to varieties in picture qualities like color, gradient and edges. Visual saliency, being firmly identified with how one sees and infers visual jolts, is researched by numerous controls together with subjective brain science, neurobiology, and Computer vision. Speculations of human consideration conjecture that the human vision framework just procedures parts of a picture in point of interest. Early work by Treisman and Gelade [10], Koch and Ullman [9], and consequent consideration hypotheses proposed by Itti and Wolfe, recommend two phases of visual consideration: quick, pre-mindful, base up, information focussed saliency identification; and gentler, undertaking reliant, top-down, objective focussed saliency extraction.

Here we address the issue from three points of view. In [17], essentially we do consider saliency location as an issue in frequency domain. Not in any way like later methodologies which indicate saliency as an area marvel, we propose another recurrence space worldview, which concedes the full use of worldwide data. Second, as opposed to exhibiting saliency in a picture, we describe the possibility of non-saliency using worldwide data. Despite the fact that as a piece of this paper we are solely worried with choosing saliency computationally, it is similarly entrancing to consider the natural viewpoint. Examine has suggested that things saw by the human visual system are thought to battle with each other to explicitly focus our thought on a subset of images. Objects that appear in the visual field will affect how they are seen by covering each other. Therefore, various are obstructed, while those that are not, will finally win in the visual cortex to give a convergence of thought. In this paper, we exhibit these stifled locales as non-saliency. Differentiated and remarkable districts, which are especially specific in the picture, non-saliency can generally speaking be shown by normal or uniform zones. These are then covered, in like manner permitting striking articles to genuinely pop out. In this paper, non-saliency is shown in the recurrence area. Third, we in like manner address another issue, that of recognizing remarkable zones of different sizes. To date there is no dependable importance of saliency in the composition. The models of saliency are different.

We propose another structure for saliency identification, which evidently, at first sight, is from every angle like the Convolution Theorem yet in assurance we will show that it is unquestionably not. We will display that the convolution of the sufficiency range with a Gaussian bit of an appropriate scale is proportionate to a saliency finder. The proposed framework can both highlight pretty much nothing and incomprehensible remarkable districts and to stifle reiterated distractors in messed pictures.

Nowadays, with the advancement of stereoscopic presentation, there are different developing products for 3D interactive media. In the paper [12], the creators presented the contention on how these contentions may be restricted and how visual solace may be enhanced by the visual consideration model. Contrasted and different saliency location models proposed for 2D pictures, just a couple examining the 3D saliency identification exist at present. Not the same as saliency recognition for 2D pictures, the profundity variable needs to be considered in saliency location for 3D pictures. To accomplish the profundity recognition, binocular depths, (for example, binocular difference) are presented and consolidated together with others, (for example, monocular divergence) in a versatile route taking into account the review space conditions. In any case, this change of profundity discernment additionally to a great extent impacts the human survey conduct. Hence, how to assess the saliency from profundity prompts and how to join the saliency from profundity with those from other 2D low-level components are two imperative variables in planning 3D saliency recognition models. Two sorts of the visual consideration component: bottom up and top down methodologies. The bottom up methodology is boost driven, generally acquired from early components, and errand autonomous. In any case, the top-down methodology, which is objective driven, comprises of abnormal state information preparing and earlier learning to bolster the undertakings, for example, object acknowledgment, scene order, target location, recognizable proof of the logical data, and so on.

II. DISPARITY EXTRACTION

Disparity alludes to the distinction in picture area of an article seen by the left and right eyes, coming about because of the eyes' even detachment (parallax). The mind utilizes binocular uniqueness to concentrate profundity

data from the two-dimensional retinal pictures in stereopsis. In computer vision, binocular divergence alludes to the distinction in directions of comparable elements inside two stereo pictures. A comparable disparity can be utilized as a part of range finding by a fortuitous event rangefinder to decide separation and/or height to an objective. In cosmology, the dissimilarity between various areas on the Earth can be utilized to decide different cosmic parallax, and Earth's circle can be utilized for stellar parallax. Human eyes are on a level plane isolated by around 50–75 mm (inter-pupillary separation) contingent upon every person. In this manner, every eye has a somewhat diverse perspective of the world around. This can be effectively seen when on the other hand shutting one eye while taking a gander at a vertical edge. The binocular difference can be seen from evident even move of the vertical edge between both perspectives. At any given minute, the observable pathway of the two eyes meet at a point in space. This point in space tasks to the same area (i.e. the inside) on the retinae of the two eyes. In view of the diverse perspectives saw by the left and right eye in any case, numerous different focuses in space don't fall on relating retinal areas. Visual binocular dissimilarity is characterized as the distinction between the purpose of projection in the two eyes and is normally communicated in degrees as the visual angle.

The expression "binocular difference" alludes to geometric estimations made outside to the eye. The difference of the pictures on the genuine retina relies on upon elements inside to the eye, particularly the area of the nodal focuses, regardless of the fact that the cross segment of the retina is an impeccable circle. Divergence on retina adjusts to binocular dissimilarity when measured as degrees, while entirely different if measured as separation because of the confounded structure inside eye.

(A) Processing disparity utilizing advanced stereo pictures

The disparity of components between two stereo pictures are normally figured as a movement to one side of a picture highlight when seen in the privilege image.[3] For instance, a solitary point that shows up at the x coordinate t (measured in pixels) in the left picture might be available at the x coordinate t - 3 in the right picture. For this situation, the dissimilarity at that area in the right picture would be 3 pixels.

Stereo pictures may not generally be accurately adjusted to consider fast divergence figuring. For instance, the arrangement of cameras might be marginally pivoted off level. Through a procedure known as picture amendment, both pictures are turned to take into account differences in just the even bearing (i.e. there is no difference in the y picture coordinates). [3] This is a property that can likewise be accomplished by exact arrangement of the stereo cameras before picture catch.

(B) Calculation

After amendment, the correspondence issue can be explained utilizing a calculation that sweeps both the left and right pictures for coordinating picture highlights. A typical way to deal with this issue is to frame a littler picture patch around each pixel in the left picture. These picture patches are contrasted with all conceivable inconsistencies in the right picture by looking at their relating picture patches. For instance, for a dissimilarity of 1, the patch in the left picture would be contrasted with a comparable measured patch in the privilege, moved to one side by one pixel. The examination between these two patches can be made by achieving a computational measure from one of the accompanying conditions that looks at each of the pixels in the patches.

- Set the size of window used when smoothing
- Set the tolerance - how close R-L and L-R values need to be
- Set the weight on gradients opposed to color
- Determine pixel correspondence Right-to-Left and Left-to-Right
    - Get gradient for each image

- Shift image and derivatives
- Calculate CSAD - Cost from Sum of Absolute Differences
- Calculate CGRAD - Cost from Gradient of Absolute Differences
- Calculate total 'difference' score

o Put corresponding disparity into correct place in image

o Keep only high-confidence pixels

o Calculate the best disparity when image 1 is slided on to image 2 and vice versa.

o Eliminate bad pixels

o Form the final disparity model

III. SALIENCY USING THE HYPERCOMPLEX FOURIER TRANSFORM (HFT)

The saliency calculation utilizing one and only feature map (of intensity). In any case, keeping in mind the end goal to get better execution, more components are required, for instance, color and movement information. We can utilize the alleged hypercomplex matrix to consolidate different feature maps. Thus, the HFT is utilized to supplant the Fourier Transform utilized as a part of saliency computing.

(A) Hypercomplex Fourier Transform

The input to the conventional DFT is a genuine matrix. Every picture element is a component of the information network and is a real number. In any case, in the event that we consolidate more than one element into a hypercomplex matrix, every component is a vector and this hypercomplex framework is a vector field. Accordingly, the customary Fourier Transform gets to be inadmissible for the purpose computations. The HFT was proposed in [18], in which the hypercomplex information was determined to be a quaternion5. For a hypercomplex network f(x,y) is given by:

$$f(x,y) = a + ib + jc + kd, \tag{1}$$

the discrete form of the HFT of (1) is given by:

$$\mathrm{FH}[m,n] = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} e^{-\mu 2\pi((\frac{mx}{M})+(\frac{ny}{N}))} f(x,y) \tag{2}$$

where μ is a pure unit of quaternion and μ2 =-1. Note that $FH[m,n]$ is also a hypercomplex matrix. The IHFT is given as:

$$f(x,y) = \frac{1}{\sqrt{MN}} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} e^{\mu 2\pi((\frac{mx}{M})+(\frac{ny}{N}))} \mathrm{FH}[m,n] \tag{3}$$

(B) Hypercomplex Representation of Multiple Feature Maps

The Hypercomplex representation can be utilized to consolidate numerous components. We characterize the input hypercomplex grid as follows:

$$f(x,y) = w1f1 + iw2f2 + jw3f3 + kw4f4 \tag{4}$$

where $w1$-$w4$ are weights and $f1$-$f4$ are the feature matrices. Here we use three features to compute the saliency for the static input case:

$$f2 = Is = (R + G + B)/3, \tag{5}$$

$$f3 = rg = r - g, \tag{6}$$

$$f4 = by = b - y, \tag{7}$$

where R,G,B are the RED, GREEN and BLUE channels of an input colour picture and

$$r = R-(G+B)/2,\ g = G-(R+B)/2,\ b = B - (R + G)/2,$$

$$y = (R + G)/2 - |R - G|/2 - B.$$

These three component maps include the rival color space representation of the input image. The computation of the saliency map and algorithm to compute saliency model using HFT is shown in (19).

(C) Overview of the Proposed Model

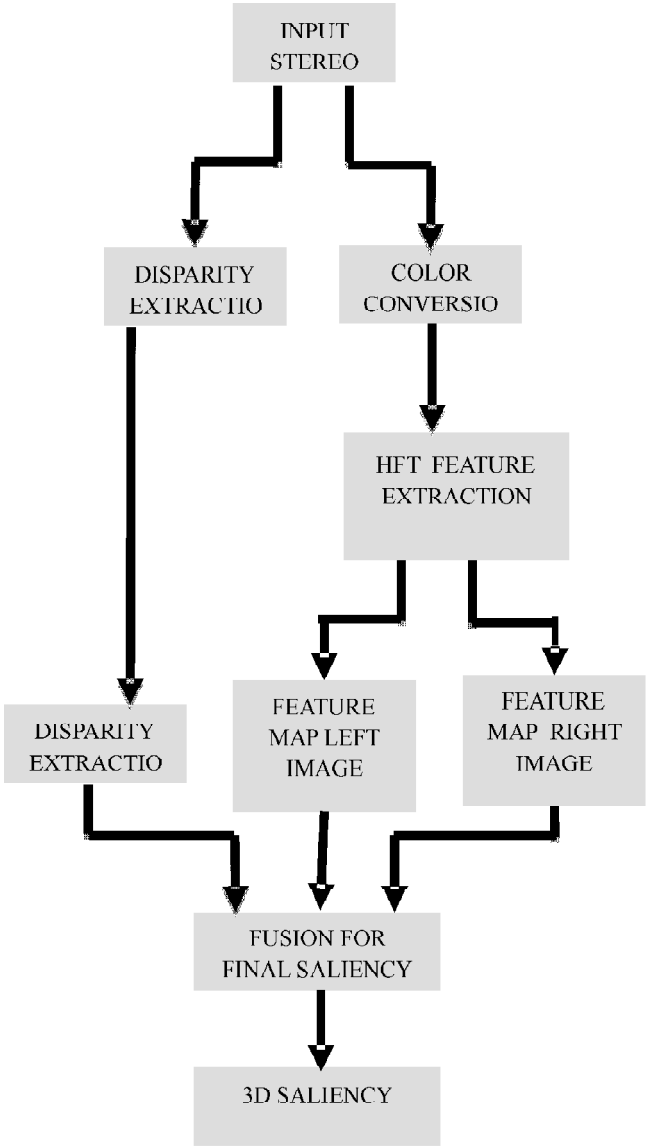

Figure 1: Block Diagram of Proposed model

The info RGB picture is changed over to YCbCr color space because of its visual quality. In YCbCr color space, the Y segment corresponds to the luminance data, while Cb and Cr are two shading adversary segments. Individually HFT is applied to the Left and Right image, accordingly feature maps are calculated . Finally the fusion for final saliency map is calculated and hence 3D saliency is extricated from the fusion of Depth/Disparity, Left and Right images.

(D) Feature Map Generation

1. Read the stereo pair of images as inputs.
2. Convert input image into YCbCr color space.
3. Apply Hypercomplex Fourier transform to each of the color component Y Cb Cr.
4. Adjust to a value in which decompostion can reach the coarsest level on x or y axis before this value and can stop.
5. Compute the Intensity, Chroma Blue and Chroma red Conspicuity Map, *C1, C2 and C3* respectively.
6. Accumulate features in several decomposition level for saliency detection in multi-channel :*DATA = [DATA C1(:) C2(:) C3(:)]*.
7. Combining conspicuity maps of each channel using max function : *CS = max(max(C1,C2),C3)*;
8. Local saliency across-scale addition : SL = SL + CS
9. Apply gaussian smoothing and enhance the Feature Map.

Finally the final saliency map is calculated using weighted average of the two saliency maps and the disparity map extracted.

IV. EXPERIMENTAL RESULTS

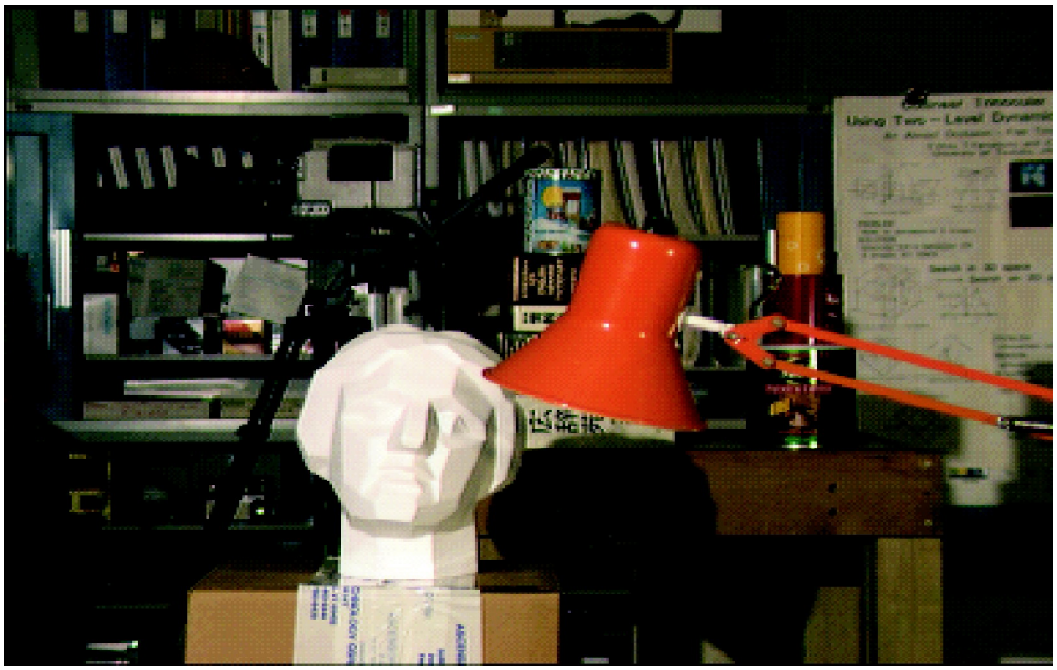

Figure 2: Input stereo Left image

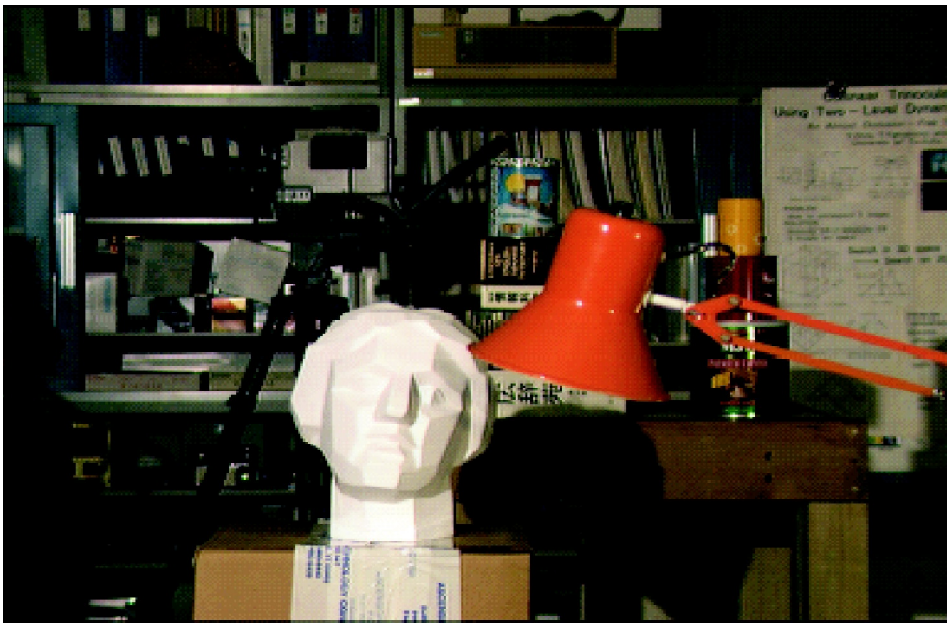

Figure 3: Input stereo Right image

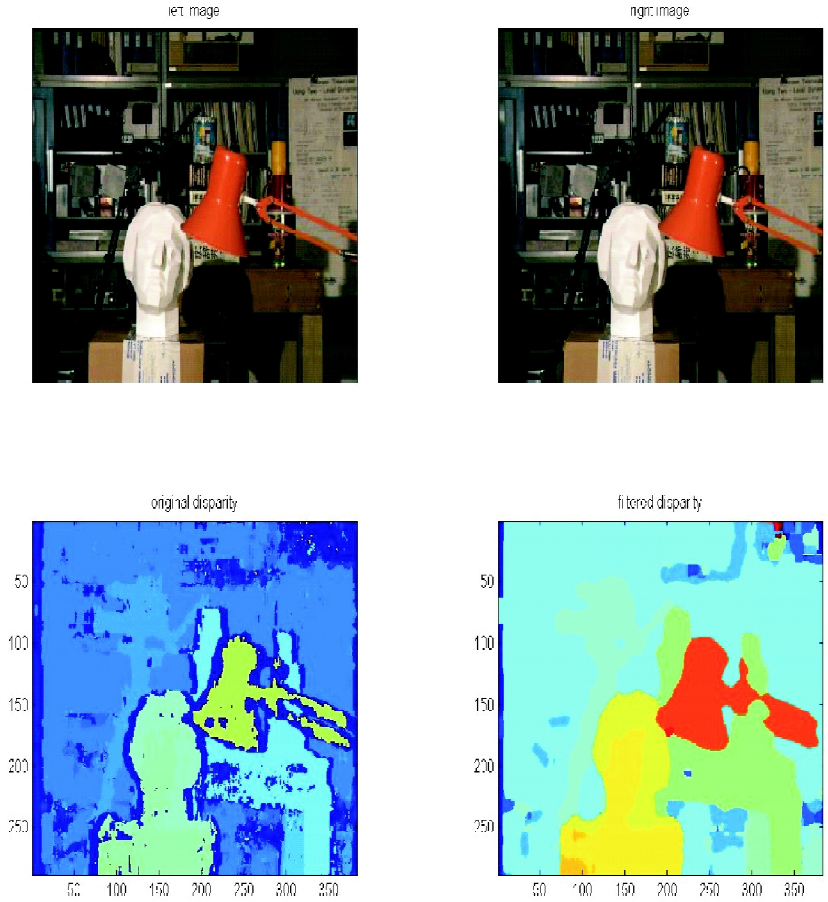

Figure 4: Output of the disparity algorithm in order for left to right and top to bottom. Input Left Image, Input Right Image, Original Disparity and Filtered Disparity

Figure 5: Saliency Map for Left Image using HFT

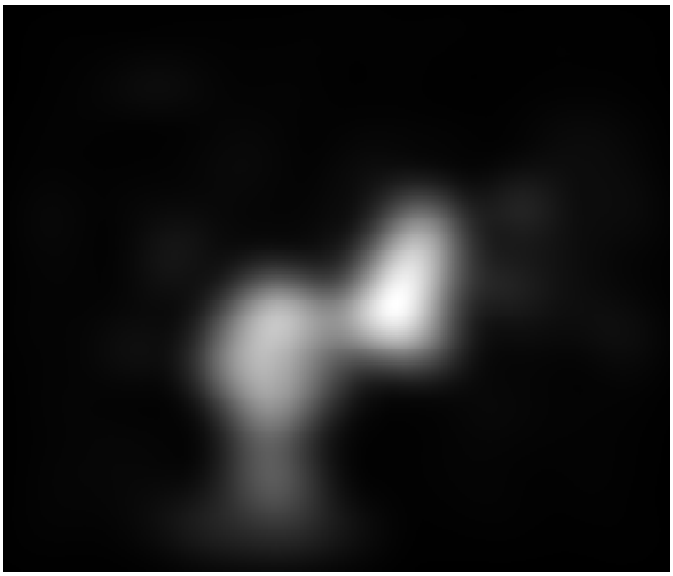

Figure 6: Saliency Map for Right Image using HFT

Figure 7: Final 3D saliency map

Table 1
Performance Comparison models between the proposed and other existing ones

S. No.	*Precision*	*Technique / Reference*
1.	.582	Model in [15]
2.	.546	Model in [16]
3.	.418	Model in [17]
4.	.692	Proposed Method

CONCLUSION

In this paper, a stereoscopic saliency discovery model for 3D pictures is presented. The components of colour, and depth are removed from HFT coefficients to speak to the vitality for little picture patches. A novel base up calculation model of visual consideration is proposed to acquire the saliency map for pictures in view of Fourier coefficients. With a specific end goal to combine multi-dimensional feature maps, we utilize the HFT to substitute the standard Fourier Transform for range scale-space investigation. Here we used the HFT based saliency extraction technique to acquire the 3D saliency map for stereo pair of images. The comparative analysis proves that out method outperforms the enlisted referred methods.

REFERENCES

[1] A. Yarbus, *Eye movements and vision*. Plenum press, 1967.

[2] U. Neisser, *Cognitive psychology*. Appleton-Century-Crofts New York, 1967.

[3] D. L. Fugal, Conceptual Wavelets in Digital Signal Processing: An In-depth Practical Approach for the Non-Matematician. San Diego, CA: Space & Signals Technical Publishing, 2009, pp. 1–78.

[4] C. Christopoulos, A. Skodras, and T. Ebrahimi. The JPEG2000 still image coding system: an overview. IEEE Trans. Consumer Elec., 46(4):1103–1127, 2002.

[5] T. Chen, M.-M. Cheng, P. Tan, A. Shamir, and S.-M. Hu. Sketch2photo: Internet image montage. ACM Trans. Graph., 28(5):124:1–10, 2009.

[6] M. Ding and R.-F. Tong. Content-aware copying and pasting in images. The Viusal Computer, 26:721–729, 2010.

[7] J. Han, K. Ngan, M. Li, and H. Zhang. Unsupervised extraction of visual attention objects in color images. IEEE TCSV, 16(1):141–145, 2006.

[8] B. Ko and J. Nam. Object-of-interest image segmentation based on human attention and semantic region clustering. J Opt Soc Am, 23(10):2462, 2006.

[9] C. Koch and S. Ullman. Shifts in selective visual attention: towards the underlying neural circuitry. Human Neurbiology, 4:219–227, 1985.

[10] A. M. Triesman and G. Gelade. A feature-integration theory of attention. Cognitive Psychology, 12(1):97–136, 1980.

[11] U. Rutishauser, D. Walther, C. Koch, and P. Perona. Is bottom-up attention useful for object recognition? In CVPR, pages 37–44, 2004.

[12] Y.-S. Wang, C.-L. Tai, O. Sorkine, and T.-Y. Lee. Optimized scale-and-stretch for image resizing. ACM Trans. Graph., 27(5):118:1–8, 2008.

[13] Q. Huynh-Thu, M. Barkowsky, and P. Le Callet, "The importance of visual attention in improving the 3D-TV viewing experience: Overview and new perspectives," IEEE Trans. Broadcast., vol. 57, no. 2, pp. 421–431, Jun. 2011.

[14] J. Wang, M. Perreira Da Silva, P. Le Callet, and V. Ricordel, "Computational model of stereoscopic 3D visual saliency," IEEE Trans. Image Process., vol. 22, no. 6, pp. 2151–2165, Jun. 2013.

[15] L. Itti, C. Koch, and E. Niebur, "Model of saliency-based visual attention for rapid scene analysis," IEEE Trans. Pattern Anal. Mach. Intell., vol. 20, no. 11, pp. 1254–1259, Nov. 1998.

[16] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency tuned salient region detection," in Proc. IEEE Int. Conf. Comput. Vision and Pattern Recognition, 2009, pp. 1597–1604

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

[18] T. Ell, "Quaternion-fourier transforms for analysis of twodimensional linear time-invariant partial differential systems," in *IEEE Conf. Decision and Control*, 2002.

[19] Jian Li, Martin D. Levine, Xiangjing "Visual Saliency Based on Scale-Space Analysis in the Frequency Domain", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 6, No. 1, January 2007.