Journal of Theoretical and Applied Information Technology
15th February 2017. Vol.95. No.3
© 2005 – ongoing JATIT & LLS

ISSN: **1992-8645** www.jatit.org E-ISSN: **1817-3195**

SUPER-PIXEL BASED SALIENCY IN 3D-IMAGE OBJECT DETECTION USING CONTENT BASED IMAGE RETRIEVAL

[1]RAKESH Y, [2]DR. K. SRI RAMA KRISHNA
[1] SRK Institute of Technology, Vijayawada (Andhra Pradesh), Electronics and Communication Engineering
[2] Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada (Andhra Pradesh), Electronics and Communication Engineering
E-mail: [1]rakesh.yemineni@gmail.com , [2]srk_kalva@yahoo.com

ABSTRACT

Visual attention is an important factor in the Human Visual System (HVS) based method for the process of Information processing in visual media. Lots of visual information is provided by the HVS system, from which it would particularly process the wanted object or simply salient object by adding filtering techniques and also it lowers the level of complexity in the process of scene analysis. Content-based image retrieval is one of the major processes in visual processing technique. CBIR is used to identify the most similar images that are visually seen in response to the given set of query image from the huge contents of image database. Many techniques have been introduced so far in saliency detection of 3D images that are used by various multimedia and CBIR processing systems. In this paper, an efficient image retrieval algorithm is designed based on the Super-pixel based saliency detection. This aims at providing a successful method of saliency detection in 3D objects, which are the major process in CBIR applications. The Experimental results and verification of object tracking from this saliency detection method is visibly efficient than the all other methods as it provides super pixel generation in the objects and its contents.

Keywords: *Content Based Image Retrieval (CBIR), Human Visual System (HVS), Region of Interest (ROI), Simple Linear Iterative Clustering (SLIC).*

1. INTRODUCTION

The human visual system has an extraordinary capability to get the prominent regions speedily in static and dynamic scenes deprived of training [2]. Visual attention is a major feature in the Human Visual System to process the visual information. Using the huge volume of visual information, visual attention selectively process the vital visual information through filtering out others to decrease the complexity for scene analysis. These visual information is known as salient regions or Regions of Interest (ROIs) in natural images [3], [7]. As, digital cameras and other consumer electronics become popular, capturing image also becomes very widespread. Every day, millions of pictures are getting uploaded by the users of the online photo sharing web sites for, example, Facebook, Flickr, Google plus etc. Customer's pictures have turn into a significant ingredient elements on basis of the compelling content is produced for automated publishing purposes. Though, customer's pictures frequently hurt from comparatively poor composition compared to professional images [1].

At present time, research in the field of human visual attention mechanism to extract prominent objects in image has been attracting research scholar significantly. With the development of stereoscopic display, which is able to carry depth perception to the viewer, the necessity of planning computational saliency detection models for 3D multimedia usage increased. Different from that of 2D images, it is important to consider the depth factor, in Saliency detection for 3D images [9]. The objective of Image saliency detection is to locate the important objects in image and to extract the resultant Saliency map. It is a gray-scale image that shows how attractive diverse objects of image for human visual system. Using saliency map, we can rapidly locate and acquire these salient objects in the image. This is the reason behind the suitability of saliency detection for computer vision researches as an initial process. The expertise of image saliency identification has been extensively implemented in numerous computer vision issues, for instance,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

image retargeting, object recognition, image retrieval [4]. The features are frequently bounding to image features like gradient, color, context and boundaries. Exponential rate at which images and videos are being produced now-a-days, has generated a major requirement for effective content-based image retrieval (CBIR) methods, for example, minkowski, spearman, relative deviation etc., which permit one to rapidly characterize and find 3D-images in large collections on basis of the features of a given test image [5], [6]. Therefore, in our proposed method we used 3D images from Wang database to detect saliency, in which each input 3D-image is segregated into a set of super-pixels that generally has more regular and compact shape with superior boundary adherence compared with the usually segmented regions produced utilizing traditional image segmentation methods [8]. Along with saliency detection, the image CBIR techniques are utilized in order to automatically retrieve and process the image.

The edge or color based super-pixel segmentation method is used in this paper, which consist of a Simple Linear Iterative Clustering (SLIC). The SLIC technique used, involve in analyzing the meaningful feature or characteristics from the separated super-pixels. The main contribution of this method is to improve the efficiency and the performance of saliency detection. A high quality based key point selection needed for the effective detection is achieved in this paper by the use of hybrid descriptors as well. The performance in the form of precision, recall and accuracy of the traditional LBP and BRISK descriptors has been improved to a great extent with this super-pixel based segmentation method.

2. RELATED WORK

Fang *et al.* [10] in their research, described the saliency detection in a framework consist of compressed domain. All the attributes like intensity, color, and texture were extracted by applying a JPEG bit-stream of discrete cosine transform (DCT) coefficients. Then the saliency value present in each DCT block was analyzed on the basis of Hausdorff distance measure and the process of fusion of feature maps. The developed framework guides for the development of an adaptive image focusing algorithm of the domain that are compressed. The proposed image retargeting method algorithm, overall consist of multi-operator operation with blocks based on seam and image scaling methods for resizing the images.

Fu *et al.* [11] presented an algorithm based on cluster for co-saliency detection. Global correspondence between numerous images was indirectly learned at the time of clustering method. Three visual attention parameter: contrast, spatial, and corresponding, were planned to calculate the cluster saliency efficiently. The concluding co-saliency maps were produced by combining single image saliency & multi-image saliency. Quantitative and qualitative experimentation on different benchmark datasets showed the benefits of the proposed method compared with the other co-saliency methodologies. Moreover, the proposed methodology on single image beat most of the well-known saliency detection techniques.

Ren *et al.* [12] have made a framework of spatio-temporal saliency detection based on the regularized feature technique of reconstruction. The saliency detection in videos are both temporal and spatial, as per this research. The both charac6teristics are processed in this algorithm of saliency detection by the authors. In temporal saliency, the movement of target patches is moved using the patches from the neighbors in the reconstruction process. In addition, Laplacian smoothing term has been introduced in this model, which consists of the coherent motion details and their trajectories. Finally the combination of temporal and spatial saliency is made to provide the salient regions with more confidence in video saliency detection technique.

Liu *et al.* [13] presented a super spatiotemporal saliency model based on pixel, for finding saliency in videos. On basis of the super pixel representation of video frames, motion histograms and color histograms were extracted at the super pixel level as local features and frame level as global features. After that, super pixel level temporal saliency was calculated by incorporating motion distinctiveness of super pixels with a method of temporal saliency prediction and adjustment, and super pixel-level spatial saliency was calculated by assessing global contrast and spatial sparsity of super pixels. A pixel-level saliency derivation technique has been achieved in this research of saliency detection shown in the results.

Ren *et al.* [14] presented an efficient solution for saliency detection based on region. To discover semantically significant salient regions, they extracted super pixels on basis of an adaptive mean shift algorithm as a primary elements for saliency detection. Saliency of every super pixel was calculated by utilizing its spatial compactness

Journal of Theoretical and Applied Information Technology
15th February 2017. Vol.95. No.3
© 2005 – ongoing JATIT & LLS
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

that was calculated in relation to the outcomes of Gaussian mixture model (GMM) clustering. To transmit saliency between same clusters, they considered a modified Page Rank algorithm to improve the saliency map. Their method enhanced saliency detection by large salient region detection and noise tolerance in messy background and at the same time produced saliency maps with a precise object shape.

Fang *et al.* [15] presented a saliency detection model for video on basis of the feature contrast in compressed domain. Four kinds of characteristics comprising luminance, color, texture, and motion were extracted from the discrete cosine transform coefficients and motion vectors bitstream contained video. The unpredicted frames (I Frames) was calculated in this method with the consideration of luminance, color, and texture features of saliency detection. Additionally, predicted frames (P and B) are analyzed and the motion of saliency maps are founded. Hence a fusion based system design has been developed in this method that integrates static and motion saliency features from the video frame. The final results from the experiments represents that this method of video saliency detection is efficient than the other systems.

3. PROPOSED METHOD

In recent years, Saliency detection has continuously been analyzed and modified by the several researchers in computer vision field. As, it can be seen in current trends of image segmentation, object detection, image retrieval on basis of content, image classification, image semantic understanding, etc., like-wise research in saliency detection in 3D images is one of the key aspects that leads to performance enhancement of work in the related area. Usually, saliency detection has to meet the subsequent criteria.

1. The detection of salient object should to be accurately placed in 3D-images as the humans perception seem to be uniquely coherent with focus of object formed in cluster scenes.
2. The detected object is to be perfectly separated from the complex background at the time of reallocating the information object integrity.
3. Saliency detection method has to be performed within a particular timescale with the maintenance of low computational complexity.

Thus, super-pixel segmentation suits well for performing and achieving the above mentioned criteria. The evaluation of precision, recall and accuracy allows the proposed system to achieve its goal.

3.1 Edge or Color Based Super-Pixel Segmentation

Simple Linear Iterative Clustering (SLIC) is simple and precise to apply and understand. k is the initial parameter of the proposed algorithm, which approximately denotes the number of equally-sized super pixels. The CIELAB color space design is formed and the clustering procedure is initiated with the parameter k and the initial cluster centers are given by $\mathrm{A}_i = [l_i \ a_i \ b_i \ x_i \ y_i]^T$ which are arranged in a regular grid space manner at S pixels each other. The grid interval is given as $S = \sqrt{\frac{N}{k}}$ consist of equally sized super pixels roughly. The centers of the grid now moved to form a location in their corresponding lowest gradient position from 3 × 3 neighborhood. The procedure done above avoids the toughness in centering a super pixel present in the edges and also to lower the chance of seeding method with a noisy pixel.

Each pixel i associated with the cluster center of nearest pixel, in which many overlapped regions are available as given in fig.1. By using this method, the speed up of this proposed algorithm is handled, as the size limit of search region reduces the distance calculation numbers and results in providing a good speed result than the other conventional K-Means method of clustering, in which each and every pixel needs comparison with all the cluster centers present. Since $S \times S$ shown in fig1, is considered as the size of the spatial extent formed by the super pixel then the size is modified as $2S \times 2S$ the pixel region found as a boundary of super pixel center, shown in fig.2.

After associating the pixels to the nearest neighbor centers, the cluster centers are adjusted to the mean value of $[l_i \ a_i \ b_i \ x_i \ y_i]^T$ vector. The $L2$ norm representation is now used to calculate the residual error E formed in between the new and previous cluster center locations. The steps are updated as mentioned above repeatedly in a set of iterations until the error is recovered and it is stated that 10 iterations is sufficient to clear the error. The

© 2005 – ongoing JATIT & LLS
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

final process is to report all the results acquired from the experiments of this paper using the above values. The entire algorithm is encountered in the single representation term as Algorithm 1.

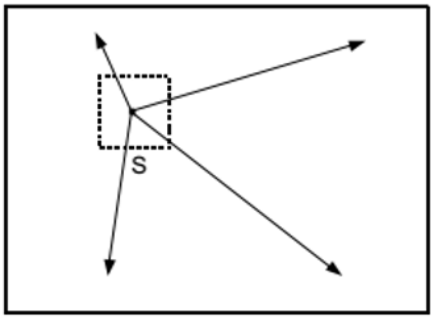

Figure 1: Spatial extent representation

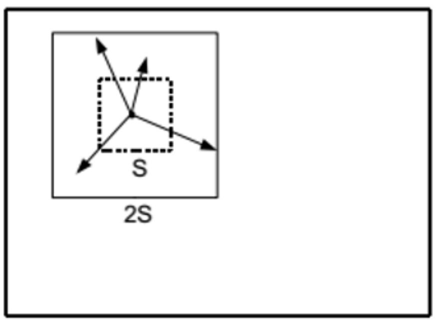

Figure 2: Super pixel representation

Algorithm 1 SLIC super pixel segmentation

/∗ Initialization ∗/

Initialize cluster centers $A_k = [l_k \; a_k \; b_k \; x_k \; y_k]^T$ by sampling pixels at regular grid steps S.

Move [cluster centers (low gradient) position to a window 3×3]

Assign label $la(i) = -1$ for i

Assign distance $ds(i) = \infty$ for i,

where $i = pixel$; A_k =cluster center;

1. repeat

2. */∗ Process ∗/*

3. **for** each A_k **do**

4. **for** each i in the region $2S \times 2S$ around A_k

5. do

6. Calculate the distance D from A_k to i.

7. **if** $D \langle ds(i)$ **then**

8. set $ds(i) = D$,

9. set $la(i) = k$,

10. end if

11. end for

12. end for

13. /∗ revaluate ∗/

14. Compute new cluster centers.

15. Compute residual error E.

16. **until** $E \le threshold$

The color present in each pixel is notified on the basis of CIELAB of color space $[z \, a \, b]^T$, from which the possible values and their ranges are known. The position of the pixel is represented as $[x \, y]^T$ that may consider a range of values in varying size of an image.

Two distances are combined into a separate single measure to normalize the color and spatial proximity with the help of their respective maximum distances found inside a cluster, M_S and M_c. Doing so, D' is written as followed by the equation 1, 2, 3,

$$ds_c = \sqrt{(z_j - z_i)^2 + (a_j - a_i)^2 + (b_j - b_i)^2} \quad (1)$$

$$ds_s = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} \quad (2)$$

$$D' = \sqrt{\left(\frac{ds_c}{M_c}\right)^2 + \left(\frac{ds_s}{M_s}\right)^2} \quad (3)$$

Here, $N_S = S$. Based on this distance measure we are able to find a cluster as well as cluster number. Cluster number k defines number of segmentation used in this research. This segmentation method used as semantic map for foreground and background extraction. Extracted regions are further process into feature extraction. Features are calculated using medium level descriptors like BRISK and LBP for foreground object identification.

3.2 LBP and BRISK

The traditional LBP method is mainly developed as a measure of complementary for representing the image contrast locally. A threshold center method is consisting of eight neighboring pixels are used in this technique. The threshold values of each weight formed by powers of two and adding the results is multiplied to produce a final LBP code. By definition, LBP is very simple to be calculated and they are invariant to all type of monotonic transformation from the gray scale.

The spatial support area found in the LBP feature extractor is extended by the use of operators containing different radii and sample counts to provide a combination results from this. By utilizing a set of N operators, we achieve N different LBP codes that are linked to provide an N single feature descriptor vector from the codes. The distance between the sample and the model are calculated at the time of inserting the marginal distributions present in the feature extractors in a linear fashion as mentioned below equation. 4,

$$L_N = \sum_{n=1}^{N} L(S^n, M^n) \qquad (4)$$

Where S^n, M^n are the sample and model distributions extracted by the nth operator.

3.3 Binary Robust Invariant Scalable Key Points

BRISK Descriptor is a method introduced to achieve scale invariance for a high quality key point selection. In this BRISK detector algorithm, the scale space of the stack layers consist of n octaves B_i and n number of intra octaves C_i, where $i = \{0,1,.....,n-1\}$ in which n=4. Let the half-sampling base image be B_0, in the set of octaves. Then, each d_i intra octave is located in between 2 consecutive layers in octave. The base image is down sampled in order to achieve the first intra-octave d_0 with the help of a scaling factor 1.5 followed by repeating the down sapling of factor 2 in an iterative manner. If (t) represents scale value, then $t(B_i) = 2^i$ and $t(d_i) = 2^i.1.5$

In BRISK detector, the number of masks used is 9-16, in which it tries to provide at least 9 minimum points continuously in the 16-point circle which are centered in the candidate point. To find out the original scale of interest as a point, it is detected that 1D parabola is attached at the scale axis.

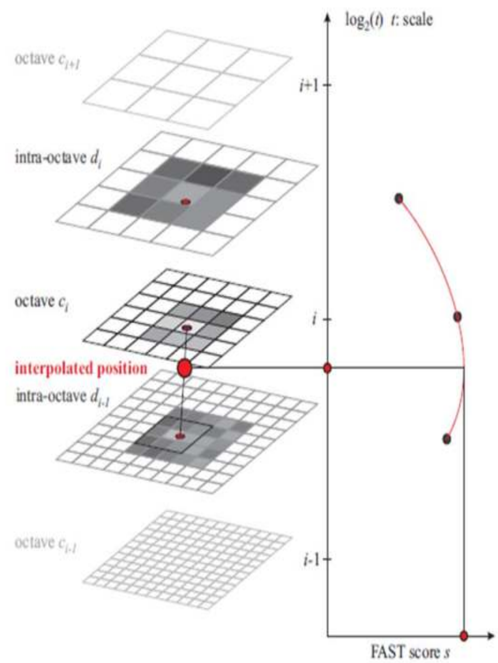

Figure 3: Representation of BRISK detector

The BRISK descriptor is created by the combination outputs from the intensity comparison test, from which the number of key points are provided. The orientation direction is calculated for each and each key point for the orientation-normalized descriptors. So, the pixel clarity and analysis is comparatively more than the other systems cited as literature.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

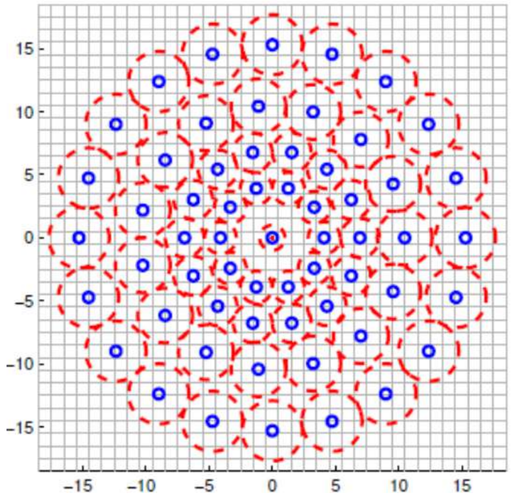

Figure 4: Partitioned view of the key points

Figure 4 shows the N distinct places from the equally partitioned circles centered in the key point. In order to avoid the aliasing effects, where the sampling intensity levels at a point p_i of image, Gaussian smoothing filter is used by adding standard deviation σ_i that is directly proportional to the distance among 2 points found in the respective circle. Consider the sampling point pairs to be (p_i, p_j) and $I(p_i, \sigma_j)$, which can be a version of Gaussian smoothed method in these two sampling points. With the help of sampling points, the local gradient $gr(p_i, p_j)$ are evaluated from the pairs and it is processed by the Gaussian smoothing method as mentioned below eq. 5 and eq. 6,

$$gr(p_i,p_j) = p_j - p_i.\frac{I(p_j,\sigma_i) - I(P_i,\sigma_i)}{\left\| p_j - p_i \right\|^2} \quad (5)$$

$$gr = \begin{pmatrix} gr_x \\ gr_y \end{pmatrix} = \frac{1}{L} \sum_{(p_i,p_j)\in L} gr(p_i, p_j)$$
(6)

Finally, the bit stream vector d_k is to be calculated for all the short distance intensity value differences in the sample of the point pairs $(p^{\alpha}{}_i, p^{\alpha}{}_j) \in S$ where the components of binary string b is given by the following equation 7,

$$b = \{1,\ I(P_j^{\alpha}, \sigma_j) > I(P_i^{\alpha}, \sigma_i); 0\ otherwise\} \quad (7)$$

$$\forall (P_j^{\alpha}, \sigma_j) \in S$$

The two BRISK descriptor and their vectors are matched together by establishing a simple calculation from Hamming distance equation. Then the similarity between the descriptors is calculated by counting the total number of bits having same properties. To perform this, the bitwise XOR operation is introduced and the numbers of zeros are counted simultaneously. The combination of BRISK descriptors and similarity measurement using XOR provide a hybrid method of combining the properties, which helps in identifying the similarities and differences regarding the generated super-pixels. Thus, the super-pixel based segmentation using hybrid techniques are efficient than the other existing systems

4. EXPERIMENTAL RESULTS

The proposed method and its experiments are done on the WANG database, which is known to be a large universal database that provides a subset of 1000 images of Corel stock datasets. All the coral datasets are selected manually form a separate 10 classes consisting of 100 images in the class. The comparison of results are given as the final process by calculating the accuracy, precision and recall as the performance measure technique.

The distance between the feature vectors L1 and L2 along with the normalized L2 (NL2) are evaluated. The precision and recall for a query image is computed as mentioned in the below eq.8:

$$\Pr ecision = \frac{\#CIR}{\#IR} \,\&\, \mathrm{Re}\,call = \frac{\#CIR}{\#MID} \quad (8)$$

Where #CIR is the number of correct images retrieved, #IR is the number of images retrieved, and #MID is the number of matching images in the database.

The comparison of acquired results is provided in this section to justify the performance of the proposed method with the recent and state-of-the-art techniques such as Local Binary Pattern (LBP) and BRISK. Retrieved image from Edge or

Journal of Theoretical and Applied Information Technology
15th February 2017. Vol.95. No.3
© 2005 – ongoing JATIT & LLS

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Color based Super-pixel segmentation, process into CBIR. Similarity between close textures images are identified as a same class in retrieved section.

For Example while testing we given Bus query image but more relevant retrieval is Building why because of texture data. Similarly for Rose Query data Buses are retrieved because of color information. These redundancy is eliminated through the combination of textural and new level descriptor combination BRISK-LBP. The Comparative results are given below,

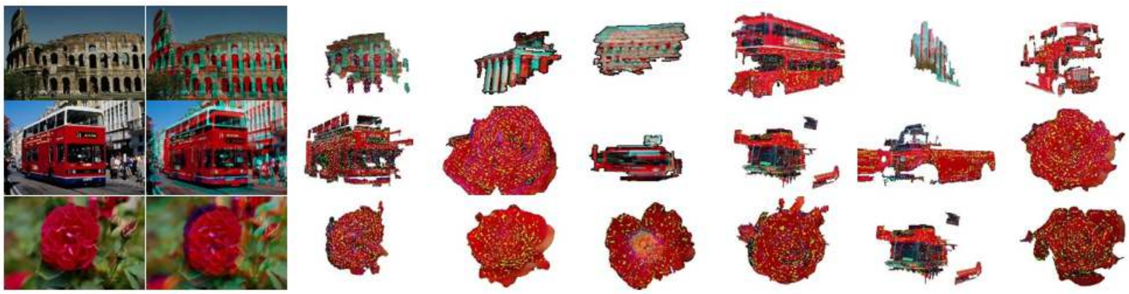

Figure 5: Input images and separated saliency objects

The retrieval result data form the fig.5 shows the first 5 rows of the sequence of retrieved images from database. Based on texture interference coming between building and bus. Similarly based on color bus and rose images are get affected. The below figures of graph represents the comparison of recall, precision and accuracy of the saliency regions.

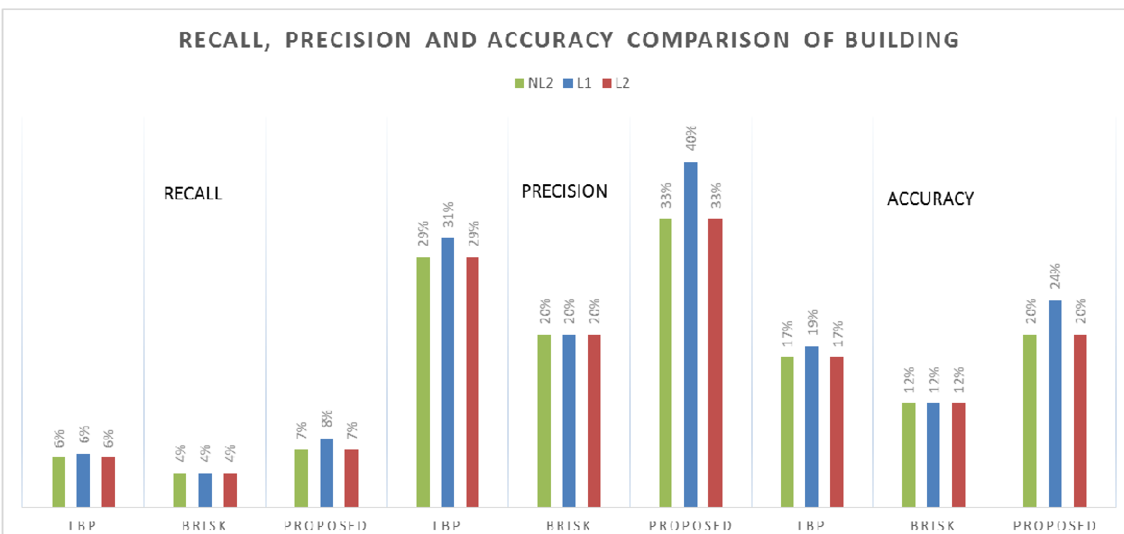

Figure 6: Recall, Precision, and Accuracy representation of buildings

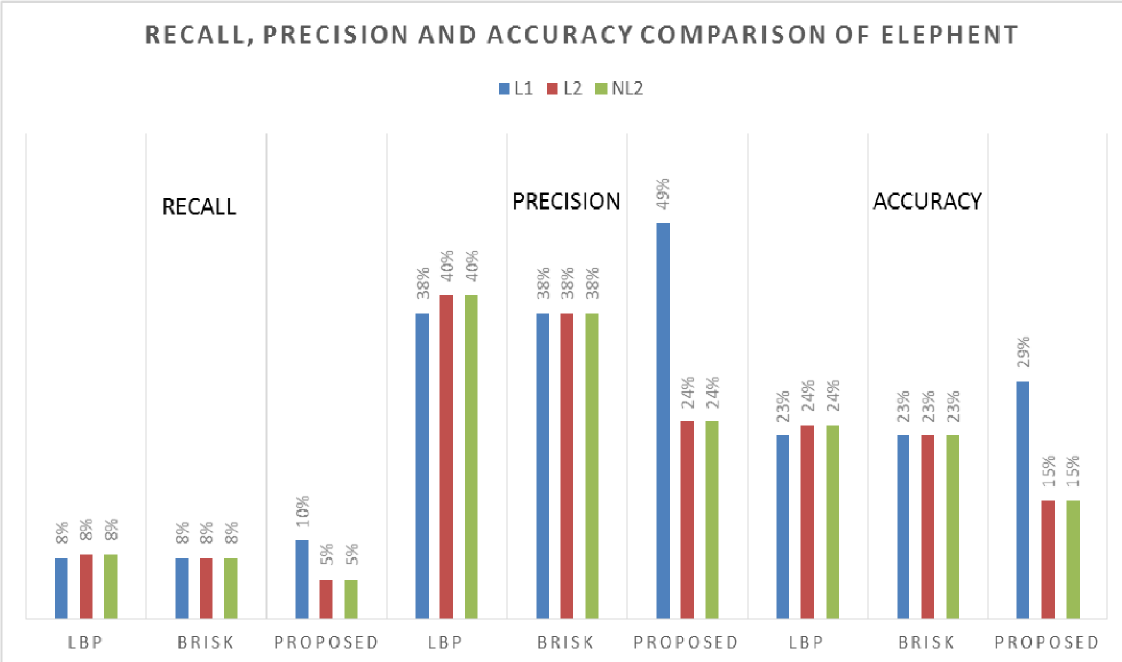

Figure 7: Recall, Precision, and Accuracy representation of elephants

The fig. 6, 7, 8, 9, 10 gives the comparison of results from the acquired saliency objects like Building, elephant, bus, rose and dinosaur. The acquired objects are compared on the basis of the calculated recall, precision and the accuracy values from the images using the eq.8

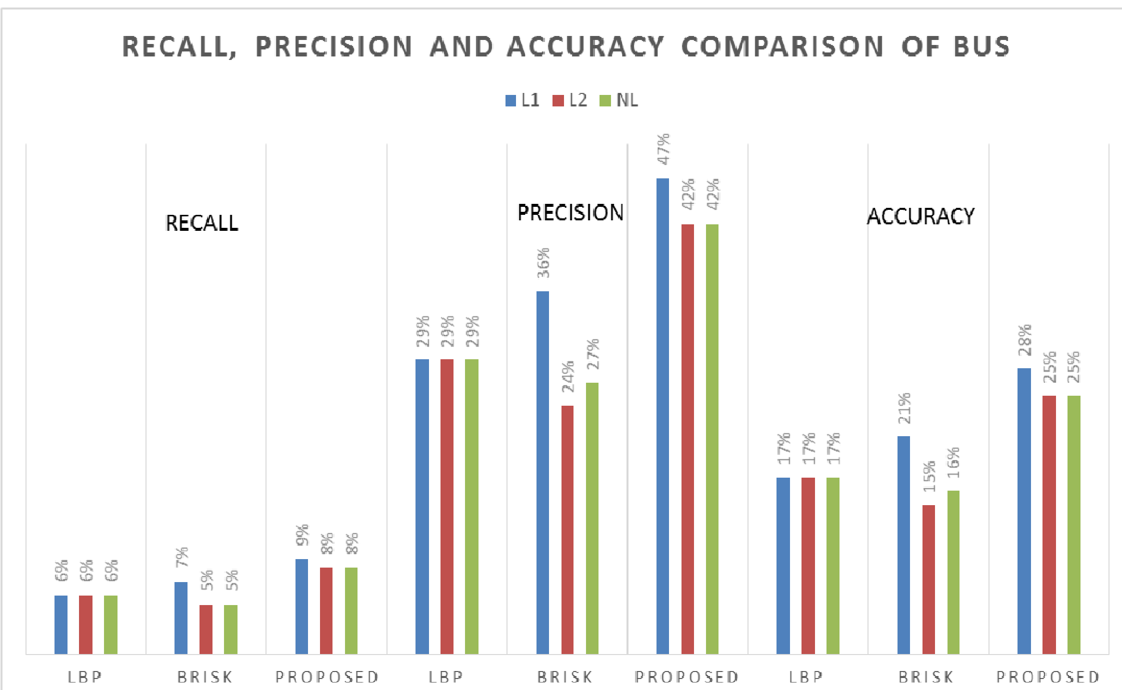

Figure 8: Recall, Precision, and Accuracy representation of bus

All the precision, recall and accuracy are discussed based on the three methods LBP, BRISK and proposed super-pixel method. L1, L2 and the normalized L2 (NL2) are measured,

© 2005 – ongoing JATIT & LLS

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

which is represented as bars in the figure notation.

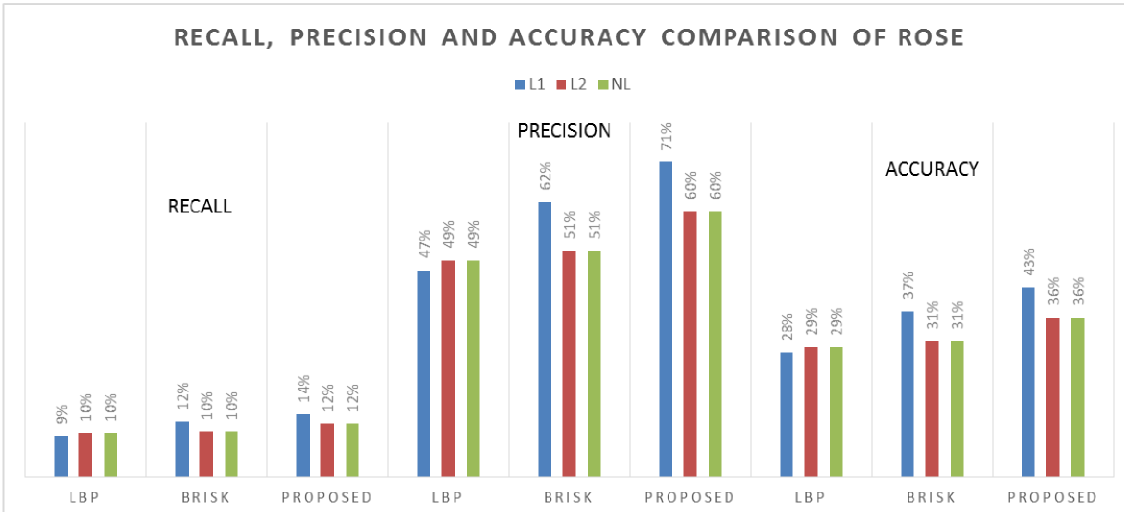

Figure 9 :Recall, Precision, and Accuracy representation of rose

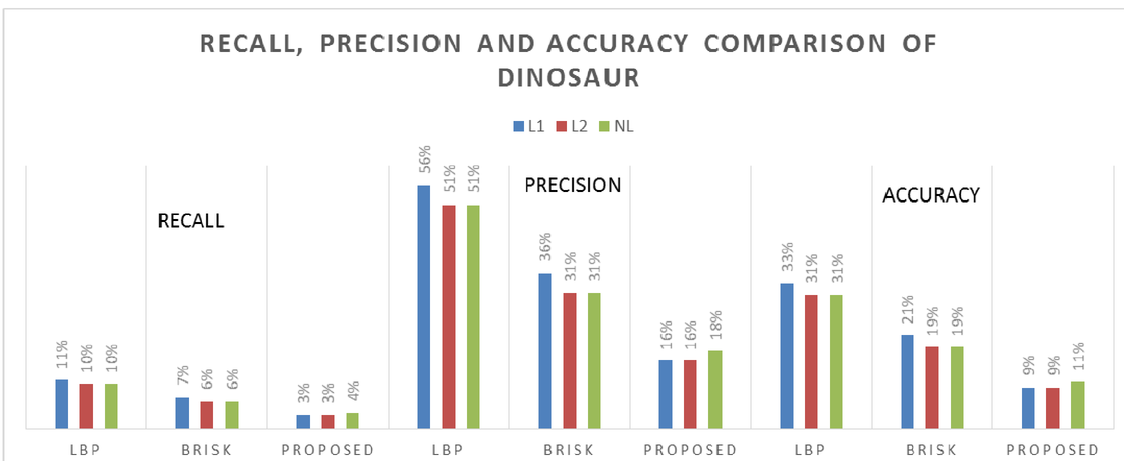

Figure 10: Recall, Precision, and Accuracy representation of dinosaur

Thus, it is clear that the proposed Super-pixel method of saliency detection produce consistent L1,L2, and NL2 features with an average simulation of the recall, precision, and accuracy when compared to the other methods such as LBP and BRISK.

5. CONCLUSION

The saliency detection model grabs the most attractive interest in Human Visual System for analysis of the regions. Different from all existing methods, we propose new stereoscopic saliency detection in this paper to detect and separate the salient region or important object from the whole feature image. The detection of salient region and their feature parts are more successful than the other systems. Experimental results show the promising performance of the proposed saliency detection model for stereoscopic images based on the object tracking or saliency detection in CBIR. Wang database is used for the experimental verification, in which the performance of descriptors for CBIR system over a particular database using Average Retrieval Precision (ARP) and Average Retrieval Rate (ARR) metrics are evaluated in this paper. Further, the retrieval results from the algorithm are compared with recent and state-of-the-art descriptors such as Local Binary Pattern (LBP) and BRISK model for 3D images.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The precision, recall and accuracy values show that the proposed system performs well in detecting the salient object than the other algorithms.

REFRENCES:

[1] H. Tang, "Object-Aware Saliency Detection For Consumer Images," *Proceedings of 19th IEEE International Conference on Image Processing*, 2012.

[2] D. Vaquero, M. Turk, K. Pulli, M. Tico and N. Gelfand, "A survey of image retargeting techniques," *Proceedings of Applications of Digital Image Processing*, vol. XXXIII, 2010.

[3] Y. Fang, J. Wang, M. Narwaria, P. L. Callet and W. Lin, "Saliency Detection For Stereoscopic Images," *Proceedings of Visual Communications and Image Processing (VCIP)*, 2013.

[4] X. Sun, Z. Liu and W. Guo, "A Background Prior based Saliency Detection for JPEG Image," *Proceedings of Seventh International Symposium on Computational Intelligence and Design*, 2014.

[5] Y. Fang, W. Lin, Z. Chen, C.M. Tsai, and C.W. Lin, "A Video Saliency Detection Model in Compressed Domain," *IEEE Transactions On Circuits And Systems For Video Technology*, vol. 24, no.1, 2014.

[6] Xi. Wu, H. Wang and W. Chen, "Saliency Detection Based on Graph and Independent Component Analysis with Reference," *Proceedings of 13th International Conference on Control, Automation, Robotics & Vision*, 2014.

[7] H.Y. Gao and K.M. Lam, "Segmentation-Enhanced Saliency Detection Model Based On Distance Transform And Center Bias," *Proceedings of IEEE International Conference on Acoustic, Speech and Signal Processing*, 2014.

[8] J. Xu, X. Guo, Q. Tu, C. Li and A. Men, "A Novel Video Saliency Map Detection Model In Compressed Domain," *Proceedings of IEEE Military Communications Conference*, MILCOM, 2015.

[9] K. Xue, X. Wang, G. Ma, H. Wang and D. Nam, "A Video Saliency Detection Method Based On Spatial And Motion Information," *Proceedings of IEEE International Conference on Image Processing (ICIP)*, 2015.

[10] Y. Fang, Z. Chen, W. Lin and C. W. Lin, "Saliency Detection in the Compressed Domain for Adaptive Image Retargeting," *IEEE Transactions on Image Processing*, vol. 21, no. 9, 2012.

[11] H. Fu, X. Cao, and Z. Tu, "Cluster-Based Co-Saliency Detection," *IEEE Transactions on Image Processing*, vol. 22, no.10, 2013.

[12] Z. Ren, S. Gao, L.T. Chia, and D. Rajan, "Regularized Feature Reconstruction for Spatio-temporal Saliency Detection," *IEEE Transactions on Image Processing*, vol.22, no.8, 2013.

[13] Z. Liu, X. Zhang, S. Luo, and O. L. Meur, "Super pixel-Based Spatiotemporal Saliency Detection," *IEEE Transactions On Circuits And Systems For Video Technology*, vol. 24, no. 9,2014.

[14] Z. Ren, S. Gao, L.T. Chia, and I.W.H. Tsang, "Region-Based Saliency Detection and Its Application in Object Recognition," *IEEE Transactions on Circuits And Systems For Video Technology*, vol. 24, no. 5, 2014.

[15] Y. Fang, W. Lin, Z. Chen, C.M. Tsai, and C.W. Lin, "A Video Saliency Detection Model in Compressed Domain," *IEEE Transactions On Circuits And Systems For Video Technology*, vol. 24, no. 1, 2014.