VIDEO PROCESSING WITH 3D-DWT FOR DEEP LEARNING

Suresh Babu D[1], Dr. K B Raja[2], Dr. K R Venugopal[4]

[1] *Research Scholar, Dept of EC, UVCE, Bangalore, Karnataka, India*
[2] *Professor, Dept of EC, UVCE, Bangalore, Karnataka, India*
[3] *Ex Vice Chancellor, Bangalore University, Karnataka, India*

ABSTRACT

This work explores the processing of video sequences by employing 3D-DWT for better performance in accuracy and speed leading to reduction in computation complexity for the deep learning scenario. The 3D-DWT is used for processing of video sequences involving various orientation of objects in the deep learning. Deep learning technique used here consists of seven layers that process the different sub-bands of wavelet obtained from 3D-DWT process for the detection of optimum number of features. Various metrics of evaluation are presented for measuring the performance of the proposed methods and comparison with traditional methods is performed.

Keyword : - *Video processing, 3D-DWT, Deep learning*

1. Introduction

Human visual system is a system that is very unique system. It can perform recognition of the objects of interest present in many images or sequences of videos. It can then start to perform the identification of the objects in the neighbor-hood of the objects to get the information of the finer details of the object that we are interested in for performing the extraction of the high level of information from the concerned scene. Scientists have carried out many studies to understand this cognitive property present in the human visual system that can be extended to perform computer vision that can lead to the application development say for example the scene understanding. Research studies performed in many scenarios so that we can perform the detecting of important regions and presence of various objects in an image data effectively. This task has been carried out by the scientists who specialize in the are of cognitive studies over the last few years that has lead to a boom in applications that are for computer vision applications. Detection and recognition of objects, compression of image and video data, imaging process such as photo collage [1], cropping, thumb nailing, quality assessment of video and image data, image retrieval based on content [2], browsing of net based on image data [4], tracking of objects [5], human robot interaction and object discovery [6] are few of the applications of salient object detection process. Salient object detection is generally classified into bottom-up or top-down approaches [3]. The important features of the objects are called the salient objects. They are of important in any given scene, objects that have undergone the recording, already, like the ones that are been recorded and that which appears in a given scene [7], objects that have undergone cluttering in the given scene, objects that are very interesting, surprise objects, aesthetic, attributes and scene context. Most of the salient models need to consider neurobiological aspect as the objects that will be considered will have motion associated in video sequences [8]. Some times there are objects for Detection of performing the salient objects which exhibit the motion, they require a good and deep understanding of the concepts in the field of the neuroscience. Studies performed for the task like that of the video saliency detection algorithms have been reported by Liu et al [9] for the applications that deals in multimedia in a big way. Presently the video saliency detection methods carry out the task of computing the salient objects that are present in each frame. This is performed in the spatial domain and salient features from temporal domain.

The resurgence of neural networks and its associated techniques along with the concepts of deep learning have the advantage of a sense of independency that are exhibited by the features present in the object along with the bias

information. The process of salient object detection is carried out with the employment of the Convolutional Neural Networks (CNN) [10]. If we take any model of CNN, we can observe that there exists thousands of neurons that possess tenability in their parameters. This feature can lead to the identification of the salient regions that can be used of the receptive fields. In the models of deep learning systems and that are based on CNNs salient object detection can be categorized broadly into two main methods: classical CNNs and Fully Convolutional Networks (FCN). In CNN models the task of classification of the networks undergoes the processing of the segment information that has been obtained from a group of multi-layered perceptrons. These perceptrons can perform the extraction of the features up to several scales. The ongoing FCNs, other than considering of the sub images or patches present in the input image, similar to that in CNNs, every pixel undergoes the consideration, this is for even the pixels that are near to the boundaries of salient objects. With fully CNNs being capable of demonstrating that they are superior in many aspects say for example the performance over traditional methods for the task of salient object detection process. Currently, there are a lot of literature that is a result of research studies carried out in this direction [10]. If we observe that there are some transparent objects that are present in the scene that we are considering or if we observe that there is low contrast that is existing between differnet video frames, it becomes that the detection capability of FCN based algorithm is limited [11]. Different studies that account the [12,13] task of filtering performed on various images, when it is done prior to salient object detection, this will lead to the improvement in the reliability and accuracy. This is the direct result because there will be noise that is present during the stage of data acquisition, in scenarios where the images or video sequences are acquired in real time. Challenges that are posed in the task of salient object detection done exclusively for the video sequences are the presence of the constraints, say for example the existence of intensity changes or it can be the constraint of contrast variations between video frames or it can also be the motion that resulted in the time of acquisition of the objects between frames [14]. Guanqun Ding, Yuming Fang [15] in their work have accounted broadly on the use of the 3D convolutional networks. This is used for performing the task of extracting spatio-temporal features that are present in any video sequences. The succeeding section would be a 3D de-convolutional networks that performs the task of fusing of the spatio-temporal features extracted for salient object detection. In order to carryout the enhancement of the edge features obtained by the wavelet transform is used as pre-processing layer in many cases of task that perform the salient object detection [16]. Wavelet transform when used on an image data gives rise to the decomposition of the given input image into multiple sub bands that are off low frequency components. These sub-bands capture the DC components or intensities of the given input image. They also capture all other sub bands capture the high frequency components this leads to obtaining of the edge information that are present in the input images [17]. The high frequency sub bands does the task of localizing of the edge information that are present along the vertical, lateral and diagonal axis, thereby providing the flexibility to perform the processing and edge enhancement process [18]. Combining wavelets with, the most interesting concepts like, neural networks gives reise to the auto encoders. They have been used in the design that is comprising of as many as three layers that perform the task of classification of objects [19]. CNNs perform the process task on thses featured enhanced objects that can be directly extended for salient object detection; this however uses the both forward and inverse wavelet transform. This task leads to requiring additional processing time and as a result gives reise to complexity in arithmetic operation. Therefore we need to carry out the task of performin the salient object detection in wavelet domain. In this work, we have used the fully connected deep learning model. This model is used to process the data in the wavelet domain. This model is also proposed for performing the task of salient object detection in that are present in both image and video sequences.

2. Related work

Many studies carried out by various teams of people have given the reports and accounted this in their literature on how to judicially use the wavelet and obtain the sub bands that inturn can be used for training the deep neural networks. This can further be used for classification process and the task of object detection. Combining Support Vector Machine (SVM) and k-Nearest Neighbour (KNN) and blending them with wavelet features is an excellent method that can be used for handwritten recognition [20]. Wavelets when blended with CNNs are a result that are an class of pre-processing layer for classification of images. This concoction can also be used in the detection of textures. They can also be used for improving face resolution [21]. Wavelet sub bands undergoes the combining or in other words, fusing. This is done prior to classification [22]. The use of wavelet sub bands to compute feature vectors for the process of classification [23] in an important aspect. The significant features are extracted from wavelet feature so that the task of classification [24] can be carried out. Liu et al. have presented, in their work, that relates the algorithms for that can perform the image restoration. This task is then combined to the CNN which exists as a multi-level wavelet sub bands. De Silva et al [16], have worked on mechanisms that lead to the enhancement of the edge information that are presnt in the wavelet domain. They then perform the task of

classification process that are performed with the use of CNNs. The high frequency or the wavelet sub bands that contain the detailed information, are only considered for the task of edge enhancement. This preocss is carried out by the extensive using of the gradient algorithms and modulus maxima methods. The process of performing DWT and inverse DWT gives rise to more time consumption and as a result it also leads to computation complexity. The DWT sub bands help in capturing the information presnt in the low pass and high pass sub bands. Processing of thes these sub bands that are used to represent intensity information of the components of the high frequency components or it can be the edge information that is present as a result in the low pass and high pass sub bands respectively can lead to the reduction in the computation time by upto a maximum of 50%. Enhancement in the resulting of edge features present in the wavelet sub bands have been achieved to a great extent by using gradient operation and blending it with the maximum modulus methods as mentioned previously. These methods are also teadious and consuming a lot of time. They are also a result of also have limitations. Thses limitations are the terms of eliminating features. They are very closely co-correlated with the objects of interest and the associated edges of the objects. In order to overcome the above said set of limitations novel techniques is deviced. It leads to improvising of the existing methods that are of great use in the task of salient object detection.

3. 3D DWT algorithm

DWT is the result that is obtained by the decomposing of the the input image. This is done at into multiple sub bands. These sub-bands help to capture the low frequency and high frequency components that are present in the input image and it will be half of the resolution. In the work reported here, the Daubechies 4 and 9/7 filters are recommended for perform of the computing of 3D-DWT sub-bands. In order to compute the 3D-DWT, there are three modules that are employed. These modules process the given input image along the three directions x, y and z direction. The Figure 1 shown below depicts the building block of 3D-DWT (17). In turn, each block of the low pass and high pass filter is used to process the data that can be present in either in the x-direction or the rows or in the y-direction, say the columns, or the z-direction say the temporal. The input data or video frame of size NxNx64 is decomposed into eight sub-bands each of size N/2xN/2x32 sub-bands.

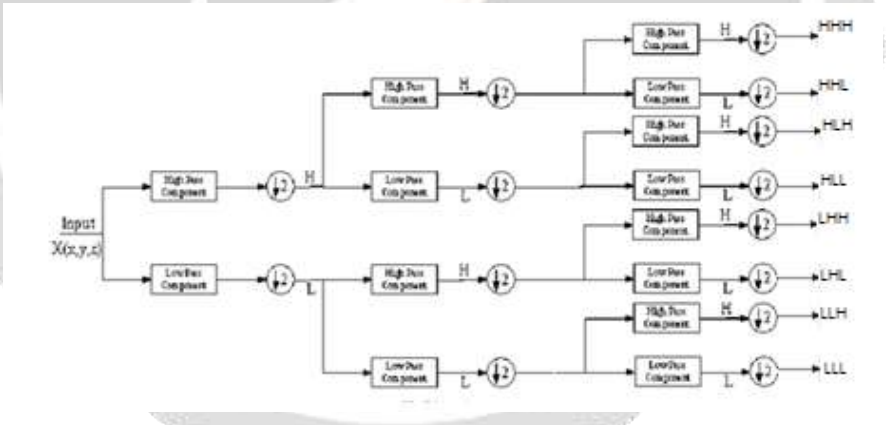

Fig.1 DWT processing of video frames

The first stage carries out the processing of the input data (that is each of the frame) along the row direction. This results in the decompose of each frame into 2 sub-bands that is either the L or H each of size NxN/2. The subsequent second stage carries out the task of processing the two sub bands column-wise. this is done with the two pairs of filters. As a result the size obtained now is N/2xN/2. After the two stages have performed the decompose of each frame into four sub-bands, in which the each of size is N/2xN/2. With 2D-DWT getting to operate on each of the 64 frames, it leads to the generation of 256 sub-bands whose size is N/2xN/2. The 3D-DWT method of performing the decomposition results in generation of eight different sub bands. These sub-bands can be denoted as LLL, LLH, LHL, LHH, HLL, HLH, HHL and HHH. The LLL sub-band is the sub-band that can capture the intensity component that is present in each of the frame. This band can also capture the intensity component that is present in the temporal direction. The LLH component is the component that is used to capture the intensity level that is present in each of the frame. It can also capture the changes that are present in these intensities that are evident in the temporal direction. The LLL sub-band undergoes the subsequent decompose. This give rise into second level DWT

sub-bands that are done with 3D-DWT. There is also a third level of DWT sub-bands. In this manner, this will result into 10 differnt sub-bands after the successful application of 3 level of decomposition on a NxNx8 input.

4. FFNN for object detection and classification

The reconstructed image aftet the entire processing is comprised of 32 frames per sub-band. The resulting frame is of size N/2xN/2. There are eight different sub-bands that are processed by the depicted deep learning structure in Figure 2. The wavelet sub bands undergoes the grouping into two sub-groups. One is of low pass sub-band (LLL_1) and the remaining sub-bands (LLH_1, LHL_1, LHH_1, HLL_1, HLH_1, HHL_1, HHH_1) are grouped. The LLL1 sub-band is the one that holds the intensity details of the concerned input data. the remaining sub-band holds the information of the direction. This is done along with the associated the motion vector in the temporal direction. Salient object detection is done using the LLL_1, in combination with any of the other sub-bands.

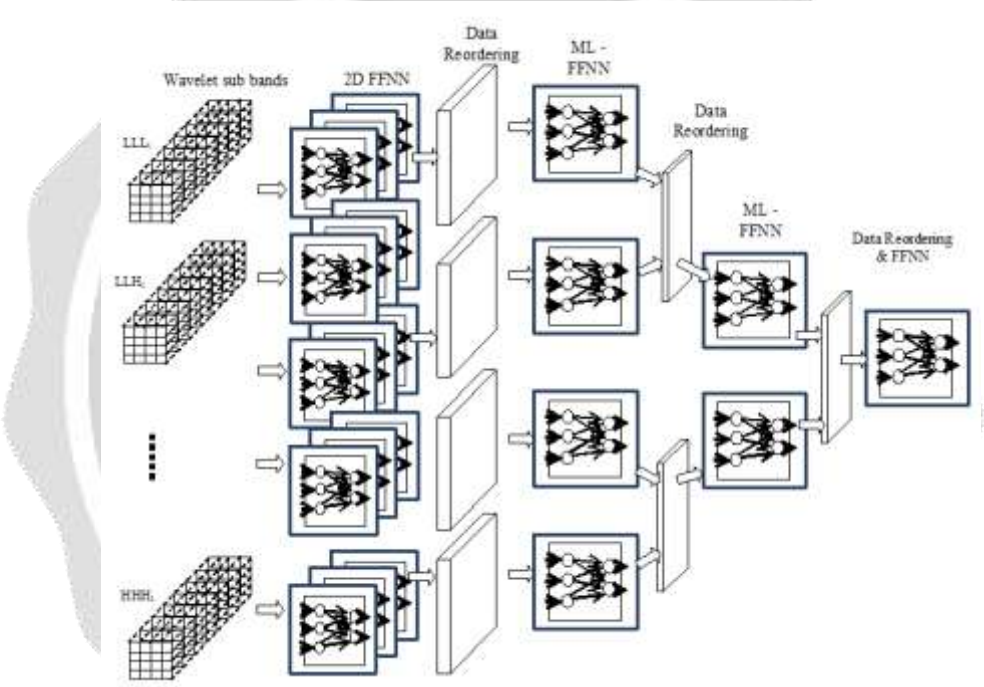

Fig. 2 Proposed deep learning structure for salient object detection and classification

The deep learning structure depicted in the figure comprises of four multi layered Feed Forward Neural Network (FFNN) structure. They also possess three layers that perfomr the reordering. The first stage present in the FFNN is the 2D-FFNN structure, the 2nd, 3rd and 4th FFNN is made of 1D structure. The three structure for reorder are used so that the input elements are rearranged by zig-zag scanning method. Figure 2 presents the internal structure. This is the 2D-FFNN proposed for performing the processing one of the sub images.

5. Results & Discussion

Various datametrics are analysed and presented here between the video processing for deep learning that is applied between the original and the reconstructed data set. The Standard Deviation is a datametric that denotes the measure of dispersion with respect to the mean value. An SD of low value indicates the clustering of the data around the mean of the dataset and a high value of SD indicates the high level of dispersion of the dataset from the mean value. Figure 3 represents the SD of the proposed work.

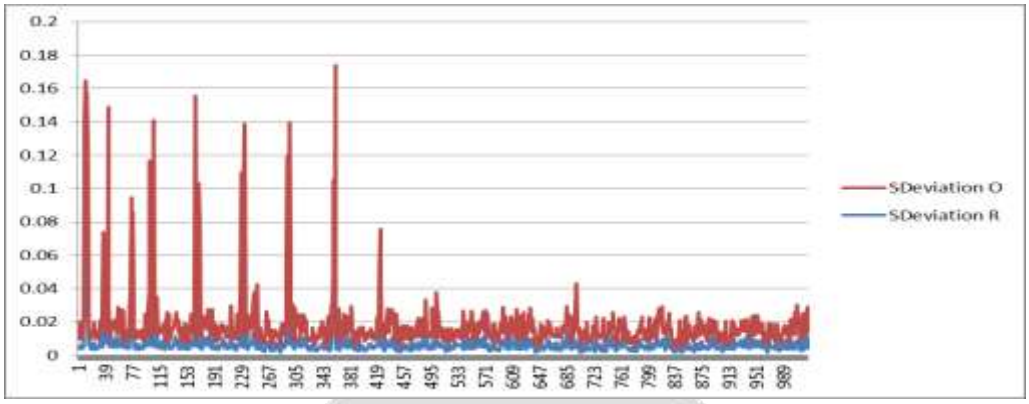

Fig.3 Standard deviation of the result

Figure 4 represents the energy plot of the result

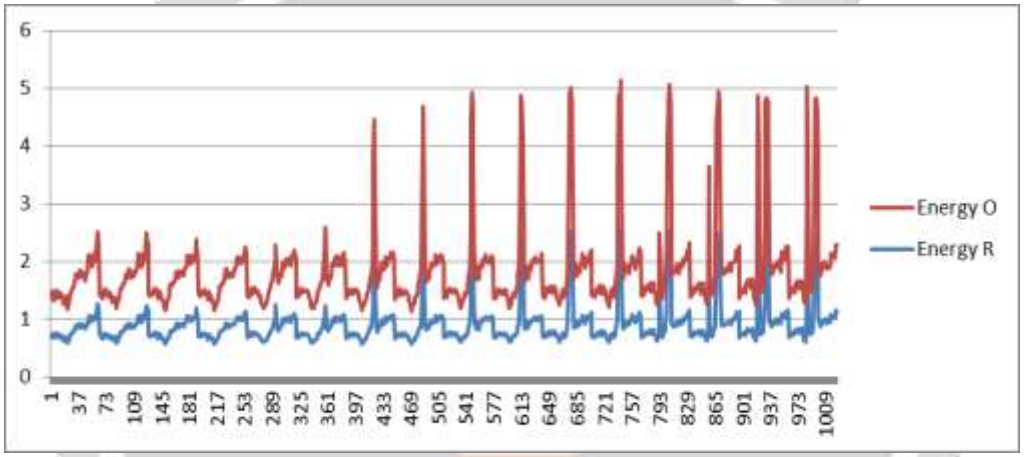

Fig.4 Energy Plot

Inter Quartile range is a measure of the spreading nature of the middle half of the data set. The figure 5 I is the plot of the IQR datametrics of the proposed work.

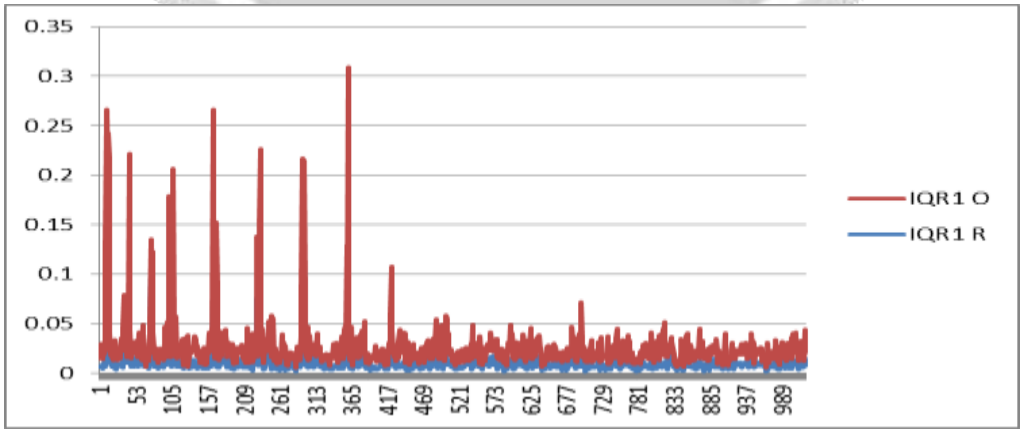

Fig.5 IQR plot

Kurtosis is a measure of the tailedness, the figure 6 shows the Kurtosis plot.

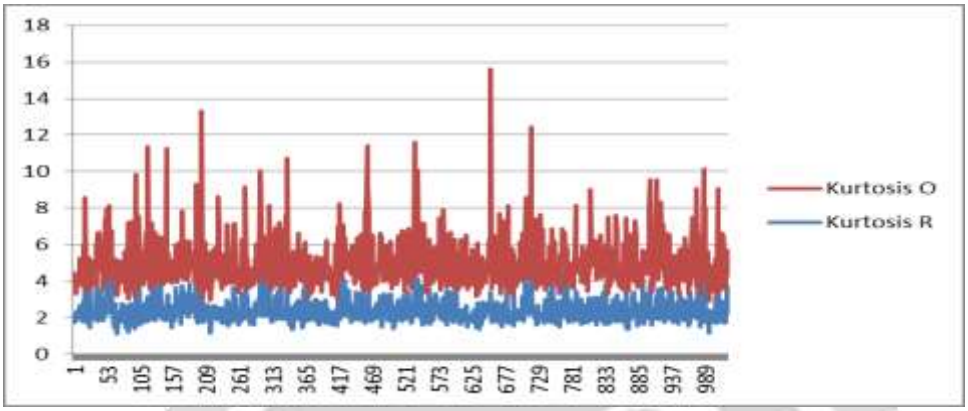

Fig.6 Kurtosis plot

Figure 7 gives the Qunatile plot of the proposed method.

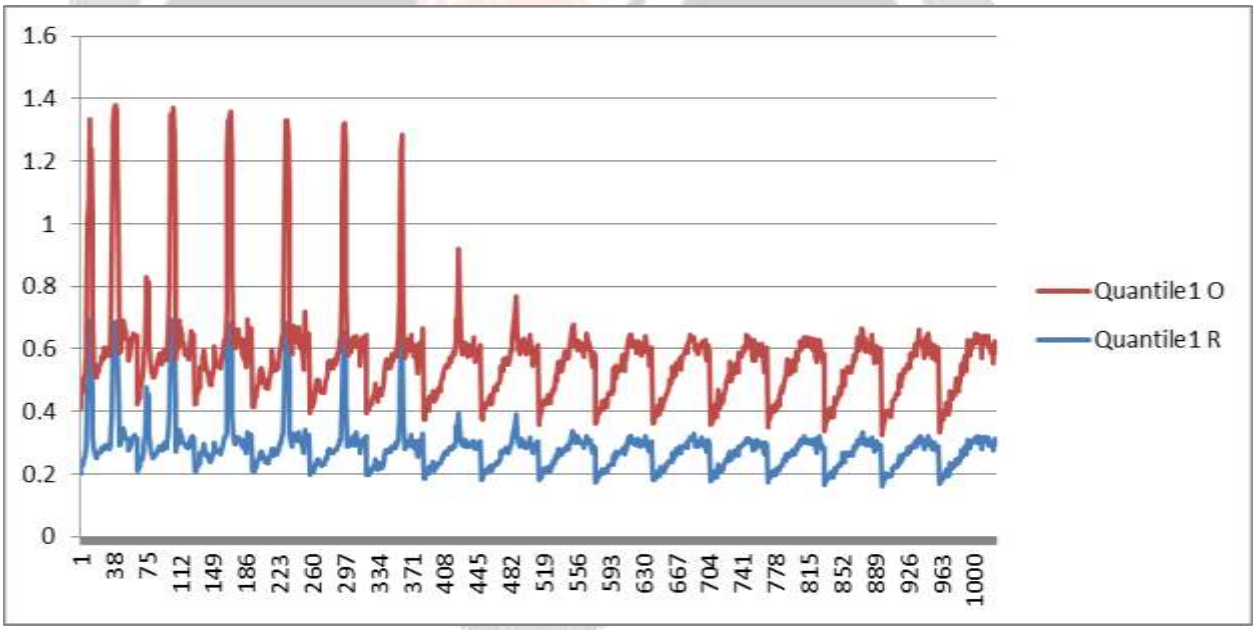

Fig.7 Quantile plot

Figure 8 depicts the percentile plot of the proposed work.

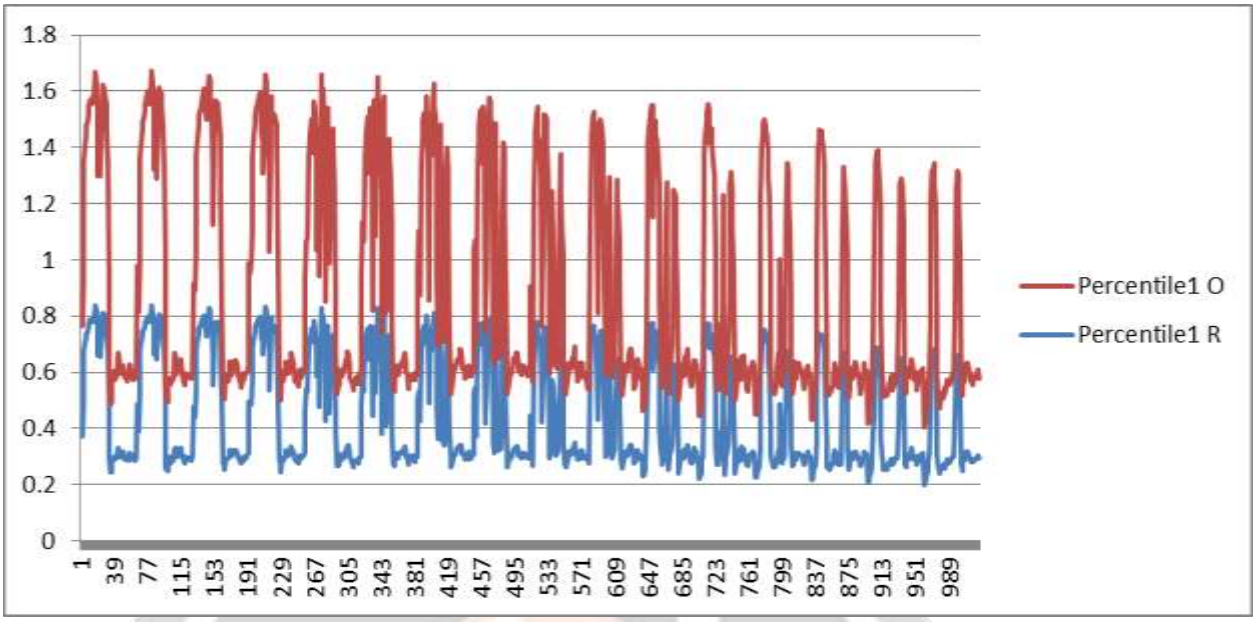

Fig8 Percentile plot

6. Conclusion

The 3D-DWT model is used to perform the processing of a dataset of size 512x512x64. This dataset undergoes an m-level sub bands processing as a result of processing that is performed by the SPIHTL algorithm. This step identifies the features that exhibit the self-similarity. The remaining wavelet coefficients that get retained by the use of this process are the most significant coefficients. These are the coefficients that assist the salient object feature detection. The SPIHTL algorithm is a special method that is designed to process the multi-band wavelet sub bands. This is done by considering frames individually. The quantization process is used so that it results in elimination of the wavelet coefficients. The quantized frames are rearranged and m-1 level inverse 3D DWT is carried out to reconstruct the image frames. Deep learning algorithm used here employs the 2D-FFNN, 1D-FFNN layers. They also use the reordering layers, additionally. These algorithms are designed in order to perform the extraction of salient object that are present in the video sequences. This results in optimum number of significant features. The decoder module is designed in a special way so that it effectively reconstructs the video sequences. This is done from the salient object features. Various data-metrics are presented that demonstrates the improvement of the algorithm.

7. References

[1]. C. Goldberg, T. Chen, F.-L. Zhang, A. Shamir, and S.-M. Hu, "Data-driven object manipulation in images," Computer Graphics Forum, vol. 31, pp. 265–274, 2012

[2]. Y.-S. Chia, S. Zhuo, R. K. Gupta, Y.-W. Tai, S.-Y. Cho, P. Tan, and S. Lin, "Semantic colorization with internet images," ACM TOG, vol. 30, no. 6, p. 156, 2011

[3]. U. Rutishauser, D. Walther, C. Koch, and P. Perona, "Is bottom-up attention useful for object recognition?" in CVPR, 2004

[4]. Kanan and G. Cottrell, "Robust classification of objects, faces, and flowers using natural image statistics," in CVPR, 2010, pp. 2472–2479

[5]. Moosmann, D. Larlus, and F. Jurie, "Learning saliency maps for object categorization," in ECCV Workshop, 2006

[6]. H. Shen, S. Li, C. Zhu, H. Chang, and J. Zhang, "Moving object detection in aerial video based on spatiotemporal saliency," Chinese Journal of Aeronautics, 2013
[7]. Kim, H., Kim, Y., Sim, J. Y., Kim, C. S.: Spatiotemporal saliency detection for video sequences based on random walk with restart. IEEE Transactions on Image Processing. 24(8), 2552-2564 (2015)
[8]. Fang, Y., Lin, W., Chen, Z., Tsai, C. M., Lin, C. W.: A video saliency detection model in compressed domain. IEEE Transactions on Circuits and Systems for Video Technology. 24(1), 27-38 (2014)
[9]. Liu, Z., Li, J., Ye, L., Sun, G., Shen, L.: Saliency detection for unconstrained videos using superpixel-level graph and spatiotemporal propagation. IEEE Transactions on Circuits and Systems for Video Technology. PP(99), 1-1 (2016)
[10]. J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In CVPR, pages 3431–3440, 2015
[11]. Q. Hou, M.-M. Cheng, X.-W. Hu, A. Borji, Z. Tu, and P. Torr. Deeply supervised salient object detection with short connections. In CVPR, 2017
[12]. M.-M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S.- M. Hu. Global contrast based salient region detection. IEEE TPAMI, 37(3):569–582, 2015.
[13]. H. Liu, L. Zhang, and H. Huang. Web-image driven best views of 3d shapes. The Visual Computer, 2012
[14]. Guanbin Li, Yuan Xie, Tianhao Wei, Keze Wang, and Liang Lin. Flow guided recurrent neural encoder for video salient object detection. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2018
[15]. Guanqun Ding, Yuming Fang, Video Saliency Detection by 3D Convolutional Neural Networks
[16]. D. D. N. De Silva, S. Fernando, I. T. S. Piyatilake, and A. V. S. Karunarathne, Wavelet based edge feature enhancement for convolutional neural networks
[17]. Amar, C.B., Jemai, O., et al.: Wavelet networks approach for image compression. ICGSTInternational Journal on Graphics, Vision and Image Processing pp. 37-45 (2007)
[18]. Said, S., Jemai, O., Hassairi, S., Ejbali, R., Zaied, M., Amar, C.B.: Deep wavelet network for image classification. In: 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC). pp. 000922-000927 (Oct 2016).,
[19]. Szu, H.H., Telfer, B.A., Kadambe, S.L.: Neural network adaptive wavelets for signal representation and classification. Optical Engineering 31(9), 1907{1917 (1992)
[20]. Akhtar, M.S., Qureshi, H.A.: Handwritten digit recognition through wavelet decomposition and wavelet packet decomposition. In: Eighth International Conference on Digital Information Management (ICDIM 2013).pp.143-148(Sept2013). https://doi.org/10.1109/ICDIM.2013.6693992
[21]. Huang, H., He, R., Sun, Z., Tan, T.: Wavelet-SRnet: A wavelet-based CNN for multiscale face super resolution. In: 2017 IEEE International Conference on Computer Vision (ICCV). pp. 1698-1706 (Oct 2017). https://doi.org/10.1109/ICCV.2017.187
[22]. Williams, T., Li, R.: Advanced image classification using wavelets and convolutional neural networks. In: 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA). pp. 233-239 (Dec 2016)
[23]. Mohsen, H., El-Dahshan, E.S.A., El-Horbaty, E.S.M., Salem, A.B.M.: Classification using deep learning neural networks for brain tumors. Future Computing and Informatics Journal (2017)
[24]. Fujieda, S., Takayama, K., Hachisuka, T.: Wavelet convolutional neural networks for texture classi_cation. CoRR abs/1707.07394 (2017)
[25]. Shuihua WANG, Yi CHEN, Yudong ZHANG, Zhengchao DONG, Elizabeth LEE, Preetha PHILLIPS: 3D-DWT Improves Prediction of AD and MCI, FirstInternational Conference on Information Science and Electronic Technology (ISET 2015), pages 60 – 63
[26]. Jaison Bennet, Chilambuchelvan Arul Ganaprakasam, and Kannan Arputharaj : A Discrete Wavelet Based Feature Extraction and Hybrid Classification Technique for Microarray Data Analysis, Hindawi Publishing Corporation Scientific World Journal, Volume 2014, Article ID 195470, 9 pages, http://dx.doi.org/10.1155/2014/195470
[27]. S. Boopathiraja1, P. Kalavathi : A Near Lossless Multispectral Image Compression using 3D-DWT with application to LANDSAT Images, 2018 International Journal of Computer Sciences and Engineering Vol-6, Special Issue-4, May 2018