Complex Wavelets and SPIHTL for Edge Feature Enhancement and Optimum Feature Selection with Deep Learning Algorithm for Video Salient Object Detection

Suresh Babu D[1,*], Dr. K B Raja[2], Dr. Cyril Prasanna Raj[2]

[1] *Research Scholar, Dept. of ECE, UVCE, Bangalore, Karnataka, India*
[2] *Professor & Chairman, Dept. of ECE, UVCE, Bangalore, Karnataka, India*
[3] *Professor, Dept. of ECE, Cambridge Institute of Technology, Bangalore, Karnataka, India*
[*] [1] sureshbd1976@gmail.com [2] raja_kb@yahoo.com [3] cyrilyahoo@gmail.com

Abstract

Cognitive property of the human visual system to detect salient objects in image and video sequences to correlate with neighbouring objects for information retrieval and understanding is developed for computer vision application. Detecting salient objects with accurate boundaries and features is addressed in this work by using the directional selectivity property of complex wavelets. Computation complexity of processing complex wavelet sub band is addressed by companding the features into eight sub bands and optimum number of features are selected using SPIHTL (Set Partition In Hierarchical Trees List) algorithm. Five layered deep learning algorithm is designed with 2D FFNN (Feed Forward Neural Network) structure and reordering units to detect salient object features. The 3D DTCWT (Dual Tree Complex Wavelet Transform) model combined with 2D SPIHTL encoder and deep learning algorithm is modelled in MATLAB and is trained to identify 120 objects with PSNR (Peak Signal To Noise Ration) of 40dB. The designed model is demonstrated to detect generic objects.

Keywords: *Video saliency, DTCWT, Directional Features, SPIHTL, Deep Learning, 2D Neural Network*

1. Introduction

Salient object detection is the process of detecting salient objects in a given scene and segmenting the object considering the accurate boundary of the object of interest. Any model or algorithm that is used for salient object detection needs to meet three criteria's. Firstly the probability of false objects as salient regions should be low that in turn refers to good detection process. Secondly, the salient object detected should have high resolutions or full resolution so that the objects detected obtained retain the original information. Thirdly, the computation efficiency of the algorithms needs to compute the salient features faster [1-3]. Detection and recognition of objects, compression of image and video data, imaging process such as photo collage, cropping, thumb nailing, quality assessment of video and image data, image retrieval based on content, browsing of net based on image data, tracking of objects, human robot interaction and object discovery are few of the applications of salient object detection process [4-9]. With resurgence of neural networks and deep learning with advantages of independency on features and bias information salient object detection is carried out using Convolutional Neural Networks (CNN) [10]. In CNN model there are thousands of neurons with tuneable parameters

that can identify salient regions using receptive fields. Large receptive and small receptive fields provide global information and local information respectively therefore highlighting salient regions and further refining these salient objects along with their boundaries.

In salient object detection process based on deep learning algorithms, edge features of the object of interest is detected first or the CNN is trained to learn edge features by the first fundamental layers. The edge features in an object are created by boundaries of the object, highlights between different objects, shadows, textures in objects, occlusion and intensity variations. To enhance the edge features wavelet transform is used as pre-processing layer for salient object detection [11]. Combining wavelets with neural networks auto encoders have been designed comprising of three layers for classification of objects [12-14]. Combining SVM (Support Vector Machine) and KNN (K-Nearest Neighbours) with wavelet features is used for handwritten recognition [15]. Wavelets are used with CNNs as pre-processing layer for classification of images, detection of textures and for improving face resolution [16]. Wavelet sub bands are combined or fused prior to classification [17] or wavelet sub bands are used to compute feature vectors for classification [18] or significant features are extracted from wavelet feature for classification [19]. Liu et al. have presented algorithms for image restoration that combines CNN with multi-level wavelet sub bands. De Silva et al. in their work have presented mechanisms to enhance edge information in the wavelet domain and then perform classification process using CNNs. The high frequency or detail wavelet sub bands are only considered for edge enhancement that is carried out using gradient algorithms and modulus maxima methods. Inverse wavelet transform is carried out to reconstruct the image after edge enhancement without the approximation coefficients. Seven layer CNN or the AlexNet [20] is used to process the reconstructed image to perform object detection and classification.

Hand pose classification algorithm developed by Phan Ngoc Hoang and Bui Thi Thu Trang [21] have proposed four step process: hand pose location based on Viola-Jones method, feature detection using wavelet transform method, dimension reduction using PCA (Principal Component Analysis) and neural networks for classification of features. The viola-jones method [22] is effective in detecting objects in real time based on rectangle features for computing integral image [23]. Haar and Daubechies wavelets are used for decomposing the input image and the low frequency sub band is considered for feature detection. The dimensionality of features detected is reduced by using PCA components computed from the low pass sub bands [24]. The classification algorithm is demonstrated to improve the performances of object classification as compared with [25]. Salient object detection is carried out considering directional features in addition to all other features for which wavelet transforms are being widely used. With wavelet transforms three directional features (vertical, horizontal and diagonal) are computed from the sub bands and with these features combined with colour information, intensity, contrast and orientation [26] salient object detection process is improved [27]. It is also reported that multi-scale frequency analysis [28] is considered in HVS (human Visual Analysis) when objects are recognized by humans and automatically the objects are magnified to extract the details such as orientation, directions and boundaries [29]. Directional information is also considered for extraction of salient objects meeting the requirements of visual perception in [30]. Wavelet transform of image data decomposes the input image into multiple sub bands of low frequency components that capture the DC components or intensities of input image, and all other sub bands capture the high frequency components or the edge information in the input images. The high frequency sub bands localize the edge information along the vertical, lateral and diagonal axis providing flexibility in processing and edge enhancement process. Shift variance and directional selectivity are the limitations of DWT that is addressed by use of Dual Tree Complex Wavelet Transforms (DTCWT). Salient object detection using complex wavelet transforms combined with deep learning model is proposed in this work demonstrating superiority in object detection accuracy as compared with existing methods.

Section II presents introduction to DTCWT, Section III presents proposed algorithm, Section IV presents the new methods for salient object detection, section V discusses implementation details and results and section VI presents conclusion.

1. DTCWT

The DTCWT sub band computation is similar to DWT with DTCWT employing Hilbert Pair of DWT filters to produce complex wavelet sub bands comprising of real and imaginary parts. The shift invariance property of DTCWT has the advantage of overcoming aliasing loss as reported by Kingsbury [31] in DWT and it is reported by Simoncelli et al. [32] on the interoperability of sub band coefficients. The wavelet filters bandwidth for DTCWT are approximately one octave wide and hence the features such as edges or surfaces are localized in space within the 28 sub bands and are also found to be uniformly spaced at any scale. The optimal characteristics of DTCWT of directional selective property are hence suitable for feature selection and improve image registration process. Figure 1 compares the directional selective property of DWT and DTCWT. DTCWT obtains 6 directional sub bands and is twice higher than the directional sub bands generated by DWT. DWT sub bands LH, HL and HH capture edges oriented along 90^0, 180^0 and 45^0 only. DTCWT produces 12 wavelet sub bands with orientations of $\pm15^0$, $\pm45^0$ and $\pm75^0$.

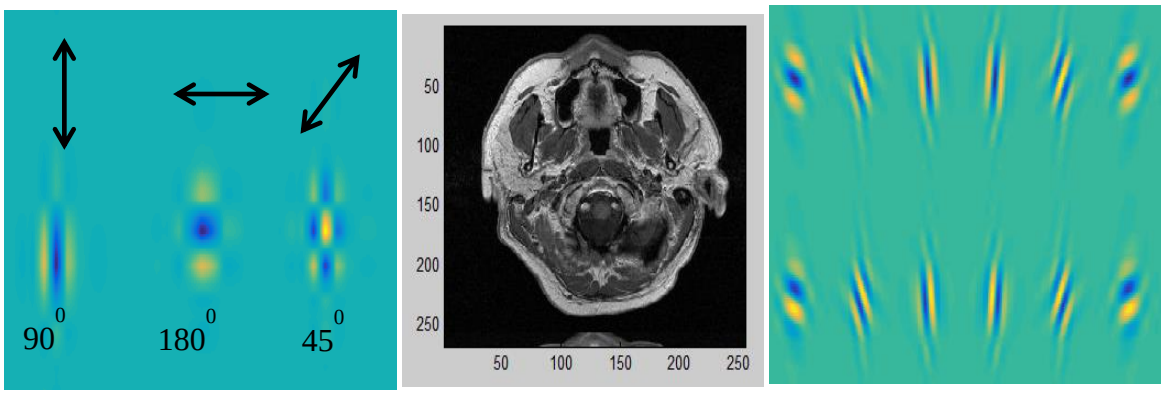


Figure.1 Directional sub bands of medical image in (b) (a) DWT (b) medical image (c) DTCWT

The directional sub band at every level for the 1st quadrant can be represented as presented in the mathematical Equation (1). ψ_a and ψ_b represent the tree a and tree b filters that are applied in all three dimensions of x, y and z [33]. The 2nd, 3rd and 4th quadrants the wavelet sub bands are represented as in Equation (2) – (4).

$$\begin{aligned}\Psi_1(x,y,z) &= [\Psi_a(x)+j\Psi_b(x)][\Psi_a(y)+j\Psi_b(y)][\Psi_a(z)+j\Psi_b(z)] \\ &= \begin{bmatrix}\psi_a(x)\psi_a(y)\psi_a(z) - \psi_b(x)\psi_b(y)\psi_a(z) \\ -\psi_a(x)\psi_b(y)\psi_b(z) - \psi_b(x)\psi_a(y)\psi_b(z)\end{bmatrix} \\ &\quad + j\begin{bmatrix}\psi_a(x)\psi_a(y)\psi_b(z) - \psi_b(x)\psi_b(y)\psi_b(z) \\ +\psi_a(x)\psi_b(y)\psi_a(z) + \psi_b(x)\psi_a(y)\psi_a(z)\end{bmatrix} \qquad (1)\end{aligned}$$

$$\Psi_2(x,y,z) = [\Psi_a(x)-j\Psi_b(x)][\Psi_a(y)+j\Psi_b(y)][\Psi_a(z)+j\Psi_b(z)] \qquad (2)$$

$$\Psi_3(x,y,z) = [\Psi_a(x)+j\Psi_b(x)][\Psi_a(y)-j\Psi_b(y)][\Psi_a(z)+j\Psi_b(z)] \qquad (3)$$

$$\Psi_4(x,y,z) = [\Psi_a(x)-j\Psi_b(x)][\Psi_a(y)-j\Psi_b(y)][\Psi_a(z)+j\Psi_b(z)] \qquad (4)$$

3D DTCWT sub bands of low pass ($C^1_{1a/b}$) and high pass ($D^m_{1a/b}$) are mathematically represented as in Equations (5) – (8), where A and B are real and imaginary filter coefficients respectively.

$$C^1_{1a/b}(Z_1Z_2Z_3) = (2^j\downarrow)\left[\left(A^j_a(Z_1)\pm iA^j_b(Z_1)\right)\left(A^j_a(Z_2)+iA^j_b(Z_2)\right)\left(A^j_a(Z_3)+iA^j_b(Z_3)\right)S(Z_1Z_2Z_3)\right] \qquad (5)$$

$$D^1_{1a/b}(Z_1Z_2Z_3) = (2^j\downarrow)\left[\left(A^j_a(Z_1)\pm iA^j_b(Z_1)\right)\left(B^j_a(Z_2)+iB^j_b(Z_2)\right)\left(A^j_a(Z_3)+iA^j_b(Z_3)\right)S(Z_1Z_2Z_3)\right] \qquad (6)$$

$$D^1_{2a/b}(Z_1Z_2Z_3) = (2^j\downarrow)\left[\left(B^j_a(Z_1)\pm iB^j_b(Z_1)\right)\left(A^j_a(Z_2)+iA^j_b(Z_2)\right)\left(B^j_a(Z_3)+iB^j_b(Z_3)\right)S(Z_1Z_2Z_3)\right] \qquad (7)$$

$$D^1_{3a/b}(Z_1Z_2Z_3) = (2^j\downarrow)\left[\left(B^j_a(Z_1)\pm iB^j_b(Z_1)\right)\left(B^j_a(Z_2)+iB^j_b(Z_2)\right)\left(B^j_a(Z_3)+iB^j_b(Z_3)\right)S(Z_1Z_2Z_3)\right] \qquad (8)$$

In DTCWT decomposing video sequences using 3D decomposition method generates eight octaves with each octave comprising of 7 sub bands capturing high frequency components in seven different orientations and low frequency components comprising of DC component. There will be real and imaginary components generated. Figure 2 presents the DTCWT sub bands represented in four quadrants (Q^+1, Q^+2, Q^+3, Q^+4) of real components. There will be four more quadrants represented as (Q^-1, Q^-2, Q^-3, Q^-4) of imaginary components. In the first quadrant there will be 7 sub bands representing high frequency components in 7 orientations. Similarly there will be 28 real and 28 imaginary sub bands.

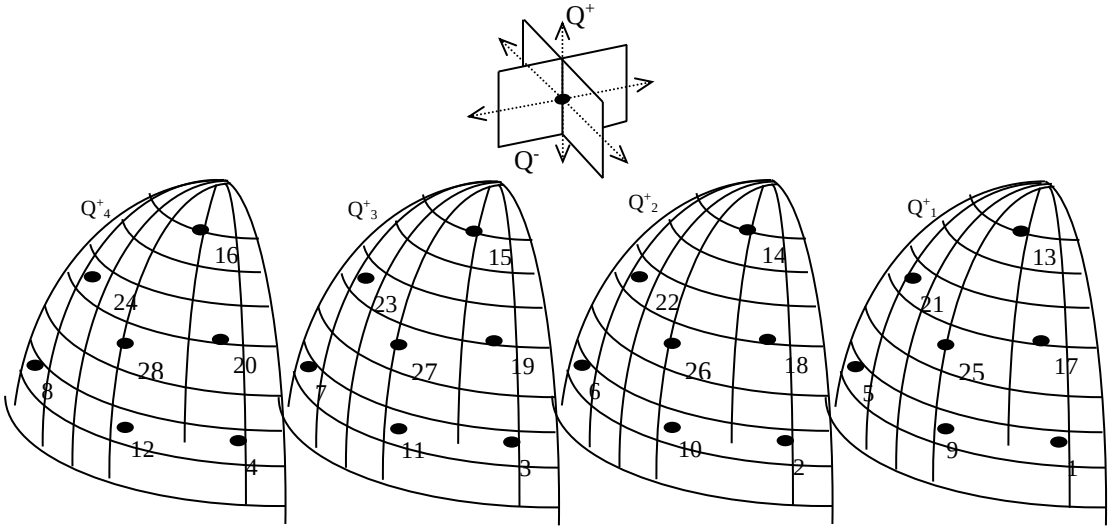


Figure. 2 Directional vectors in positive four quadrants

Table 1 presents the DTCWT sub bands and their representations after 3-level decomposition. With input image of size N x N x M, after three levels of decomposition, at level-3 there are 8 low pass sub bands and 28 high pass sub bands each of size N/8 x N/8 x M/8. At level-2 there are 28 high pass sub bands each of size N/4 x N/4 x M/4. At level-1 sub band there are 28 high pass sub bands each of size N/2 x N/2 x M/2.

Table 1 3D-DTCWT sub bands for three levels

Wavelet (High Pass) Sub-bands										
L\Q	**DV**	Q^+_1	Q^+_2	Q^+_3	Q^+_4	Q^-_1	Q^-_2	Q^-_3	Q^-_4	Q_{Avg}
1	1	w(1)(1)(1)(1)(1)	w(1)(1)(1)(2)(1)	w(1)(1)(2)(1)(1)	w(1)(1)(2)(2)(1)	w(1)(2)(1)(1)(1)	w(1)(2)(1)(2)(1)	w(1)(2)(2)(1)(1)	w(1)(2)(2)(2)(1)	W^1_1
	2	w(1)(1)(1)(1)(2)	w(1)(1)(1)(2)(2)	w(1)(1)(2)(1)(2)	w(1)(1)(2)(2)(2)	w(1)(2)(1)(1)(2)	w(1)(2)(1)(2)(2)	w(1)(2)(2)(1)(2)	w(1)(2)(2)(2)(2)	W^1_2
	3	w(1)(1)(1)(1)(3)	w(1)(1)(1)(2)(3)	w(1)(1)(2)(1)(3)	w(1)(1)(2)(2)(3)	w(1)(2)(1)(1)(3)	w(1)(2)(1)(2)(3)	w(1)(2)(2)(1)(3)	w(1)(2)(2)(2)(3)	W^1_3
	4	w(1)(1)(1)(1)(4)	w(1)(1)(1)(2)(4)	w(1)(1)(2)(1)(4)	w(1)(1)(2)(2)(4)	w(1)(2)(1)(1)(4)	w(1)(2)(1)(2)(4)	w(1)(2)(2)(1)(4)	w(1)(2)(2)(2)(4)	W^1_4
	5	w(1)(1)(1)(1)(5)	w(1)(1)(1)(2)(5)	w(1)(1)(2)(1)(5)	w(1)(1)(2)(2)(5)	w(1)(2)(1)(1)(5)	w(1)(2)(1)(2)(5)	w(1)(2)(2)(1)(5)	w(1)(2)(2)(2)(5)	W^1_5
	6	w(1)(1)(1)(1)(6)	w(1)(1)(1)(2)(6)	w(1)(1)(2)(1)(6)	w(1)(1)(2)(2)(6)	w(1)(2)(1)(1)(6)	w(1)(2)(1)(2)(6)	w(1)(2)(2)(1)(6)	w(1)(2)(2)(2)(6)	W^1_6
	7	w(1)(1)(1)(1)(7)	w(1)(1)(1)(2)(7)	w(1)(1)(2)(1)(7)	w(1)(1)(2)(2)(7)	w(1)(2)(1)(1)(7)	w(1)(2)(1)(2)(7)	w(1)(2)(2)(1)(7)	w(1)(2)(2)(2)(7)	W^1_7
		Wavelet (High Pass) Sub-bands								
2	1	w(2)(1)(1)(1)(1)	w(2)(1)(1)(2)(1)	w(2)(1)(2)(1)(1)	w(2)(1)(2)(2)(1)	w(2)(2)(1)(1)(1)	w(2)(2)(1)(2)(1)	w(2)(2)(2)(1)(1)	w(2)(2)(2)(2)(1)	W^2_1
	2	w(2)(1)(1)(1)(2)	w(2)(1)(1)(2)(2)	w(2)(1)(2)(1)(2)	w(2)(1)(2)(2)(2)	w(2)(2)(1)(1)(2)	w(2)(2)(1)(2)(2)	w(2)(2)(2)(1)(2)	w(2)(2)(2)(2)(2)	W^2_2
	3	w(2)(1)(1)(1)(3)	w(2)(1)(1)(2)(3)	w(2)(1)(2)(1)(3)	w(2)(1)(2)(2)(3)	w(2)(2)(1)(1)(3)	w(2)(2)(1)(2)(3)	w(2)(2)(2)(1)(3)	w(2)(2)(2)(2)(3)	W^2_3
	4	w(2)(1)(1)(1)(4)	w(2)(1)(1)(2)(4)	w(1)(1)(2)(1)(4)	w(2)(1)(2)(2)(4)	w(2)(2)(1)(1)(4)	w(2)(2)(1)(2)(4)	w(2)(2)(2)(1)(4)	w(2)(2)(2)(2)(4)	W^2_4
	5	w(2)(1)(1)(1)(5)	w(2)(1)(1)(2)(5)	w(2)(1)(2)(1)(5)	w(2)(1)(2)(2)(5)	w(2)(2)(1)(1)(5)	w(2)(2)(1)(2)(5)	w(2)(2)(2)(1)(5)	w(2)(2)(2)(2)(5)	W^2_5
	6	w(2)(1)(1)(1)(6)	w(2)(1)(1)(2)(6)	w(2)(1)(2)(1)(6)	w(2)(1)(2)(2)(6)	w(2)(2)(1)(1)(6)	w(2)(2)(1)(2)(6)	w(2)(2)(2)(1)(6)	w(2)(2)(2)(2)(6)	W^2_6
	7	w(2)(1)(1)(1)(7)	w(2)(1)(1)(2)(7)	w(2)(1)(2)(1)(7)	w(2)(1)(2)(2)(7)	w(2)(2)(1)(1)(7)	w(2)(2)(1)(2)(7)	w(2)(2)(2)(1)(7)	w(2)(2)(2)(2)(7)	W^2_7
		Wavelet (High Pass) Sub-bands								
3	1	w(3)(1)(1)(1)(1)	w(3)(1)(1)(2)(1)	w(3)(1)(2)(1)(1)	w(3)(1)(2)(2)(1)	w(3)(2)(1)(1)(1)	w(3)(2)(1)(2)(1)	w(3)(2)(2)(1)(1)	w(3)(2)(2)(2)(1)	W^3_1
	2	w(3)(1)(1)(1)(2)	w(3)(1)(1)(2)(2)	w(3)(1)(2)(1)(2)	w(3)(1)(2)(2)(2)	w(3)(2)(1)(1)(2)	w(3)(2)(1)(2)(2)	w(3)(2)(2)(1)(2)	w(3)(2)(2)(2)(2)	W^3_2
	3	w(3)(1)(1)(1)(3)	w(3)(1)(1)(2)(3)	w(3)(1)(2)(1)(3)	w(3)(1)(2)(2)(3)	w(3)(2)(1)(1)(3)	w(3)(2)(1)(2)(3)	w(3)(2)(2)(1)(3)	w(3)(2)(2)(2)(3)	W^3_3
	4	w(3)(1)(1)(1)(4)	w(3)(1)(1)(2)(4)	w(3)(1)(2)(1)(4)	w(3)(1)(2)(2)(4)	w(3)(2)(1)(1)(4)	w(3)(2)(1)(2)(4)	w(3)(2)(2)(1)(4)	w(3)(2)(2)(2)(4)	W^3_4
	5	w(3)(1)(1)(1)(5)	w(3)(1)(1)(2)(5)	w(3)(1)(2)(1)(5)	w(3)(1)(2)(2)(5)	w(3)(2)(1)(1)(5)	w(3)(2)(1)(2)(5)	w(3)(2)(2)(1)(5)	w(3)(2)(2)(2)(5)	W^1_5
	6	w(3)(1)(1)(1)(6)	w(3)(1)(1)(2)(6)	w(3)(1)(2)(1)(6)	w(3)(1)(2)(2)(6)	w(3)(2)(1)(1)(6)	w(3)(2)(1)(2)(6)	w(3)(2)(2)(1)(6)	w(3)(2)(2)(2)(6)	W^3_6
	7	w(3)(1)(1)(1)(7)	w(3)(1)(1)(2)(7)	w(3)(1)(2)(1)(7)	w(3)(1)(2)(2)(7)	w(3)(2)(1)(1)(7)	w(3)(2)(1)(2)(7)	w(3)(2)(2)(1)(7)	w(3)(2)(2)(2)(7)	W^3_7
		Low Pass Sub-bands								
4		w(4)(1)(1)(1)	w(4)(1)(2)(1)	w(4)(2)(1)(1)	w(4)(2)(2)(1)	w(4)(1)(1)(2)	w(4)(1)(2)(2)	w(4)(2)(1)(2)	w(4)(2)(2)(2)	W^3_8

In this work 3D DTCWT is carried out for 3-levels and processing all these sub bands for capturing salient features for detection of objects in a given input data increases processing complexity by 8 times the processing complexity of DWT. In order to reduce processing

complexity and also to capture all the orientation features 8 times more than DWT features new method is proposed. Figure 3 shows deriving DWT compatible sub bands from the DTCWT sub bands.

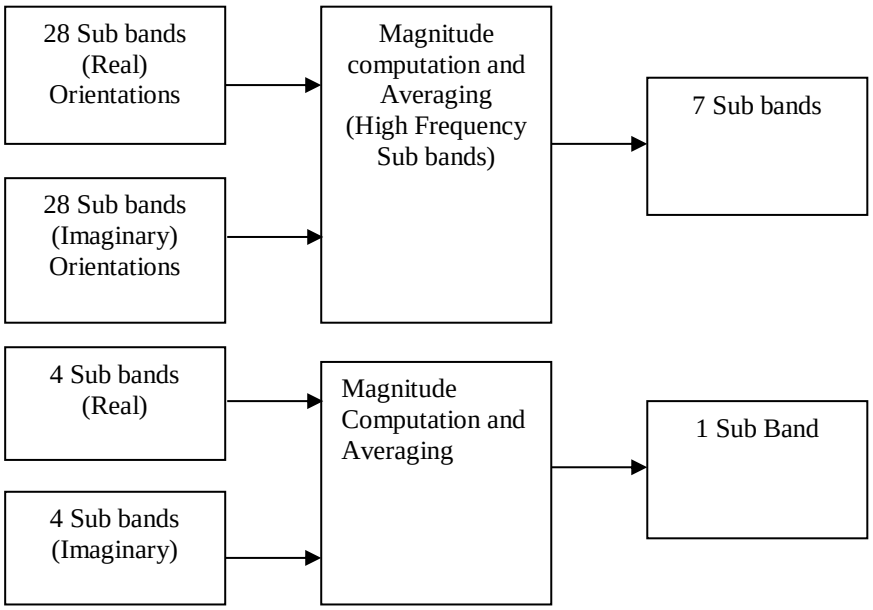


Figure. 3 Deriving DWT compatible sub bands from DTCWT sub bands

Equation (9) presents the mathematical operation of computing sub bands that are compatible (in terms of number of sub bands) from DTCWT sub bands. From the real and imaginary components the magnitude is computed for all the sub bands. From the eight sub bands for each orientation (7 orientations denoted as Directional Vectors (DV)) one sub band is generated in each level of DTCWT. Further the average of all four sub bands is computed as in Equation(10).

$$Q^1_{1,DV}=\sqrt{Q^+_1 + Q^-_1},\quad Q^1_{2,DV}=\sqrt{Q^+_2 + Q^-_2},\quad Q^1_{3,DV}=\sqrt{Q^+_3 + Q^-_3},\quad Q^1_{4,DV}=\sqrt{Q^+_4 + Q^-_4} \qquad (9)$$

DV= 1, 2, 3, . . . 7

$$Q^1_{Avg,DV} = W^1_{DV} = (Q^1_{1,DV} + Q^1_{1,DV} + Q^1_{1,DV} + Q^1_{1,DV})/4 \qquad (10)$$

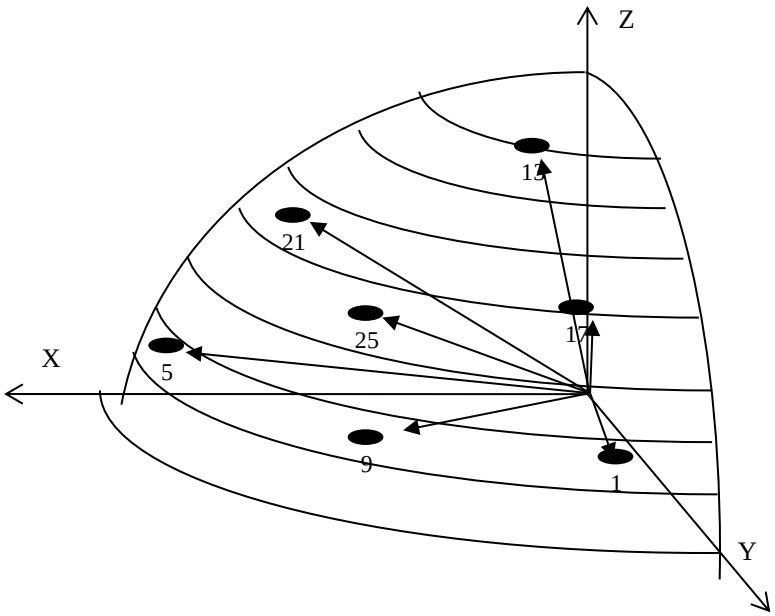


Figure. *4* Directional vectors combined to first quadrant

Considering the method presented above, the salient features in 7 directions are captured along with the low pass components as that of DWT sub bands. In Table 1 the last column represents the companded DTCWT sub bands that are similar in number as compared to DWT sub bands but they have captured more than 3 orientations as compared with DWT. Processing these features will provide additional information and limitations of DWT such as shift invariance and directional selectivity are addressed, further reducing computation complexity of processing all sub bands of DTCWT. Figure 4 gives the directional vectors combined to first quadrant, in this case.

One of the important aspects in 3D DWT decomposition is the redundancy in the multi-level sub bands. In this proposed work, all these sub bands generated by the 3-level 3D decomposition are not processed for salient object detection from the input video sequence, but only the significant wavelet coefficients that are contributing to salient objects are considered. To consider only the significant wavelet coefficients, modified SPIHT algorithm is considered in this work.

2.1 SPIHTL Algorithm

The SPIHT algorithm [34] stores the wavelet coefficients in three ordered list such as List of Insignificant Sets (LIS), List of Insignificant Pixels (LIP) and List of Significant Pixels (LSP). The value of each pixel coordinate is stored in the LIP and LSP and the LIS stores the approximation and detail coefficient. The four steps in SPIHT are initialization, sorting, refinement and quantization step update. After multiple iterations the LSP contains the coordinates of the pixels that are evaluated during the refinement pass. In order to identify the significant pixels that are required for salient object detection the SPIHT encoding algorithm is modified only to generate the three ordered list and the encoding of data 1's and 0's is not considered. The wavelet coefficients in the LIP list are set to zero intensity as it is observed that these pixels are not very significant towards information present in the wavelet pyramid. As the SPIHT algorithm is modified only to identify the location of significant pixels in the ordered tree and is not used for encoding process, the SPIHT algorithm is faster and also retains the pixel objects that are significant and contributing to the information of salient object in the image. The modified SPIHT algorithm is labelled as SPIHT List (SPIHTL) algorithm and is presented in Figure 5 for computing the most significant pixels that contribute towards computing salient object detection process. The SPIHTL algorithm computes the significance of set of coordinates with the Equation (11) to find the relationship between magnitude comparisons and message bits

$$.S_n(T)\begin{cases} 1, & \max_{(i,j)\epsilon T}\{|c_{i,j}|\} \geq 2^n \\ 0 & otherwise \end{cases} \qquad (11)$$

The notations that are used for SPIHTL algorithm representation is as follows:

- $\mathcal{O}(i,j)$: set of coordinates of all offspring of node (i,j)
- $\mathcal{D}(i,j)$: set of coordinates of all descendants of the node(i,j):
- $\mathcal{H}$: set of coordinates of all spatial orientation tree roots (nodes in the highest pyramid level)
- $\mathcal{L}(i,j) = \mathcal{D}(i,j) - \mathcal{O}(i,j)$

The set partition rules that are used for the modified algorithm as are follows:

1) The initial partition is formed with the sets {(i, j)} and $\mathcal{D}(i,j)$ for all (i, j) ε $\mathcal{H}(i,j)$
2) If $\mathcal{D}(i,j)$ is significant, then it is partitioned into $\mathcal{L}(i,j)$ plus the four single element sets with $(k\ l)$ ε $\mathcal{O}(i,j)$
3) If $\mathcal{L}(i,j)$ is significant, then partition into the four sets $\mathcal{D}(k\ l)$ with $(k\ l)$ ε $\mathcal{O}(i,j)$

1) **Initialization**: Output $n = \left\lfloor log_2(\max_{(i,j)}\{|c_{i,j}|\}) \right\rfloor$; set the LSP as an empty list and add the coordinates (i, j) ε $\mathcal{H}$ to the LIP and those with the descendants to LIS, as type A entries
2) **Sorting Pass** :
    2.1) For each entry (i, j) in the LIP do:
        2.1.1) output S_n(i, j);
        2.1.2) if S_n(i, j) = 1, then move (i, j) to the LSP
    2.2) for each entry (i, j) in the LIS do:
        2.2.1) if the entry is of type A then
            - Output $S_n(\mathcal{D}$ (i, j));
            - If S_n(i, j) = 1 then
                - For each (k, l) ε $\mathcal{O}$(i, j) do:
                    - Output S_n(k, l);
                    - If S_n(k, l) = 1 then add (k, l) to the end of the LIP and output the sign of $c_{k,l}$;
                    - If S_n(k, l) = 0 then add (k, l) to the end of the LIP;
                - If $\mathcal{L}$ (i, j) ≠ 0 then move (i, j) to the end of the LIS, as an entry of type B, and got to step 2.2.2; otherwise, remove entry (i, j) from the LIS;
        2.2.2) if the entry is of type B then
            - Output $S_n(\mathcal{L}$ (i, j));
            - If $S_n(\mathcal{L}$ (i, j)) = 1 then
                - Add each (k, l) ε $\mathcal{O}$ (i, j) to the end of the LIS as an entry of type A;
                - Remove (i, j) from the LIS.

Figure. 5 SPIHTL Algorithm

In SPIHTL algorithm is quantized by removing the LIP elements to zero, however quantizing all elements will lead to loss of information in the reconstructed image. To overcome these limitations adaptive SPIHTL algorithm is proposed in this work. The encoded data obtained from SPIHTL encoder is decoded and inverse transformation is carried out and the distortions checked considering PSNR parameter with reconstructed image sequence.

3. Proposed algorithm

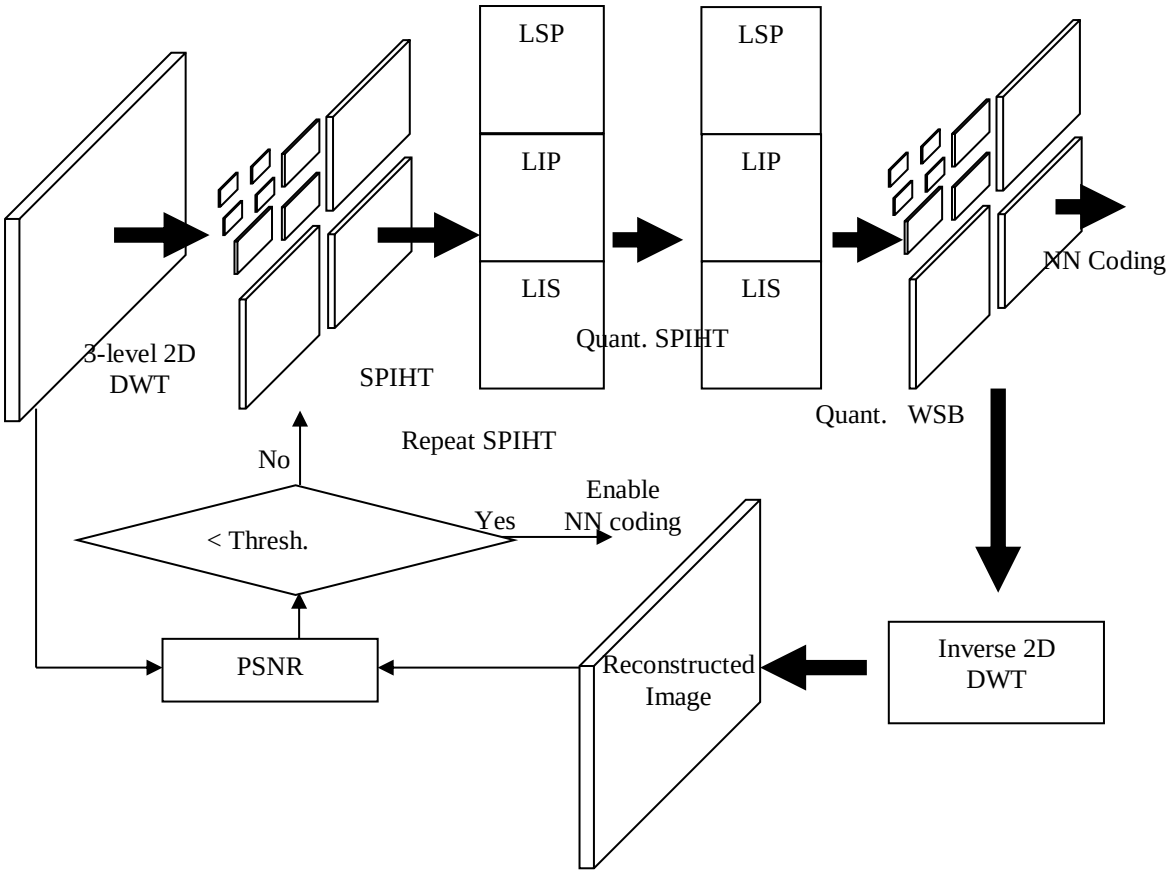


Figure. 6 Proposed SPIHT algorithm with novel method of encoding

If the distortion error is very high then the progressive encoding scheme is carried out to the next iteration, this process is continued until the distortion in the reconstructed image is within the set limits. Figure 6 illustrates the proposed algorithm. The number of iterations is set such that every wavelet coefficient is searched for its relevance in the parent-child pair and LIS is empty. The contents of LIP are quantized and made to zero. From the encoded SPIHTL ordered list the wavelet coefficients are rearranged into sub bands. The rearranged sub bands are further processed by the inverse DWT to reconstruct the image. The PSNR is computed considering the original image and reconstructed image, based on the PSNR results and wavelet sub bands are considered for further processing by the NN module for salient object detection. If the PSNR results are less than the desired threshold, SPHITL algorithm is revisited to encode the data with threshold level set to 2^{n-1}(16 in this example). If the PSNR are above than the desired results (1.5 time higher than the desired results), SPIHTL is further carried out by setting the threshold level to 2^{n+1} (64 in this example). The SPIHTL algorithm with adaptive logic proposed in this work identifies the pixels in the wavelet domain at higher level and the self-similarity pixels at the all the lower levels that constitute towards salient object detection. The three level decomposed images after quantization process (achieved after performing SPIHTL process) is further processed by the deep learning algorithm for salient object detection and classification.

From the quantized wavelet coefficients two-level inverse 3D DWT is carried out to generate the reconstructed data that generates 8 groups of sub bands each of size N/2 x N/2 x 32 as shown in Figure 7.

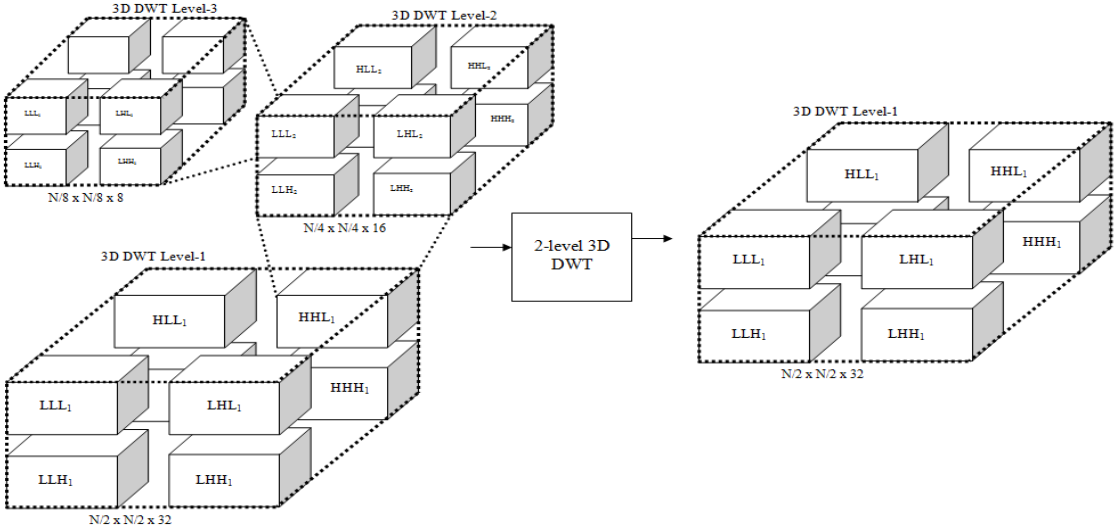


Figure.7 2-level 3D inverse DWT for reconstruction

After 2-level 3D DWT reconstruction there will be 8 sub bands each of size N/2 x N/2 x 32. The purpose of performing 2-level 3D inverse DWT is to ensure that the deep learning algorithm operates on wavelet coefficients to detect salient objects and classify salient objects. The advantages of this method is that there are eight sub bands of which one low pass sub band that holds intensity information denoted by (LLL_1) and there are 7 high frequency sub bands denoted by (LLH_1, LHL_1, LHH_1, HLL_1, HLH_1, HHH_1) that are independently processed to detect salient objects. Improvement in accuracy of salient object detection and classification is achieved by considering all the 8 sub bands or only the LLL sub band with additional sub bands selected from any of the seven high frequency sub bands.

4 Proposed method for salient object detection

The reconstructed image that comprises of 32 frames per sub band with each frame of size N/2 x N/2 and there are eight sub bands that are processed by deep learning structure as presented in Figure 8. The wavelet sub bands are grouped into two sub groups of low pass sub band (LLL_1) and all other sub bands (LLH_1, LHL_1, LHH_1, HLL_1, HLH_1, HHL_1, HHH_1). The LLL1 sub band holds the intensity or DC components of input data and all other sub band holds the directional information along with its motion vector along the temporal direction. For salient object detection the LLL_1 along with any of the other sub bands are considered for processing in the deep learning structure.

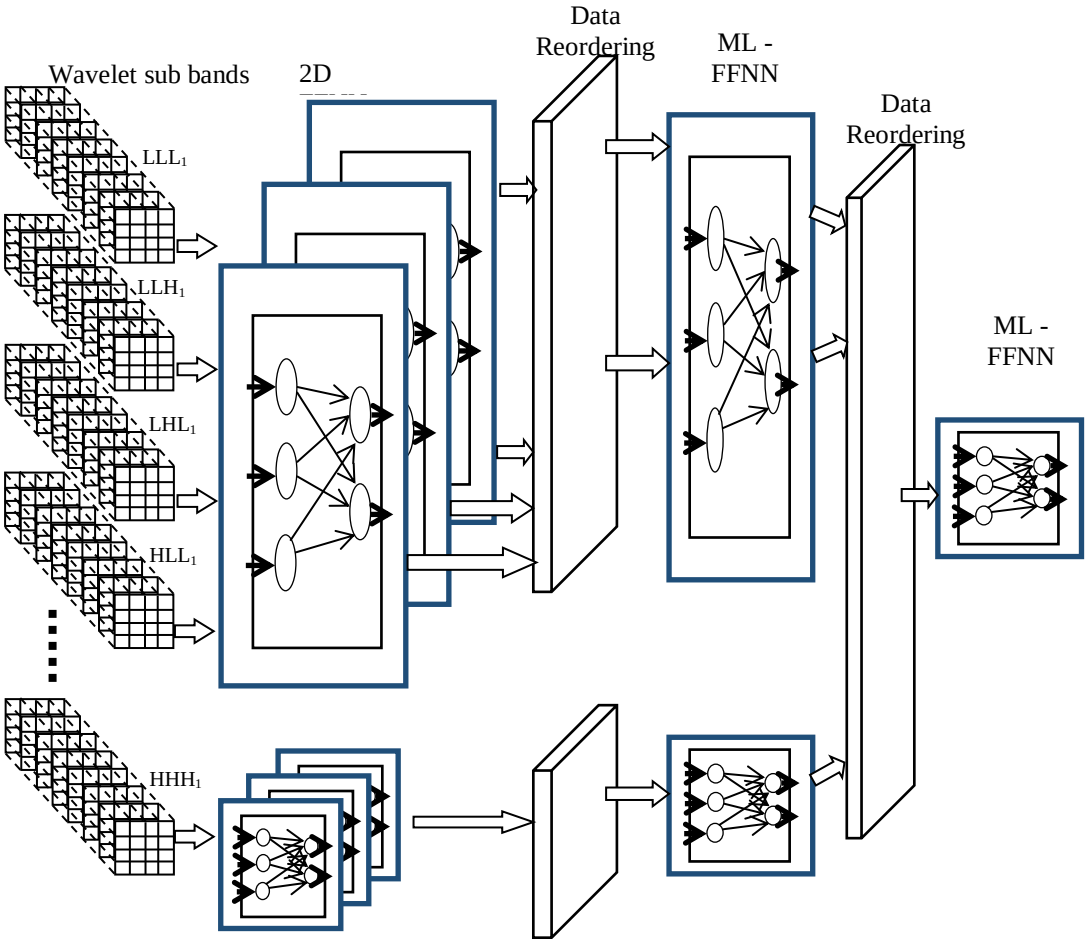


Figure. 8 Proposed deep learning structure for salient object detection and classification

The output of first layer (2D FFNN) that comprises of two structures one for processing the DC components and the seocnd for processing the high freqnecy components, processing of both data is carried out by 2D FFNN structure. The 2D $FFNN_{DC}$ is designed to capture the intensity and motion vector information of salient objects in the video sequences. The 2D $FFNN_{HF}$ is designed to capture the directional orientations or the high frequency information of the salient objects. The structure of 2D FFNN structure is shown in Figure 9. F^1 and F^2 are the two consecutive frames that are considered for demonstrating the design of 2D FFNN structure. The sub images of F^1 and F^2 are denoted as $F^1{}_1$ and $F^2{}_1$ respectively. Each of these sub images and its wavelet coefficients are represented as {$F^1{}_1(0)$, $F^1{}_1(1)$, $F^1{}_1(2)$, ... $F^1{}_1(15)$ } and {$F^2{}_1(0)$, $F^2{}_1(1)$, $F^2{}_1(2)$, ... $F^2{}_1(15)$ }. These 32 coefficients are processed by the 2 x 2 neuron and the intermediatery outputs are denoted as {$n_1(0)$, $n_1(1)$, $n_1(2)$, $n_1(3)$} and the corresponding output of neurons after network activation function processing is denoted as {a_1, a_2, a_3, a_4}. The relation between the network output a and the network input F is given as in Equation(12) – (13), similarly all other outputs are represented.

$$n_1(0) = F_1^1(0)W_{1,0}^1 + F_1^1(1)W_{1,1}^1 + F_1^1(2)W_{1,2}^1 + F_1^1(3)W_{1,3}^1 + F_1^1(4)W_{1,4}^1 \ldots. + F_1^1(15)W_{1,15}^1 + F_1^2(0)W_{1,16}^1 + F_1^2(1)W_{1,17}^1 + F_1^2(2)W_{1,18}^1 + F_1^2(3)W_{1,19}^1 + F_1^2(4)W_{1,20}^1 \ldots. + F_1^2(15)W_{1,31}^1 + b_1(0) \quad (12)$$

$$a_1 = f(n_1(0)) \quad (13)$$

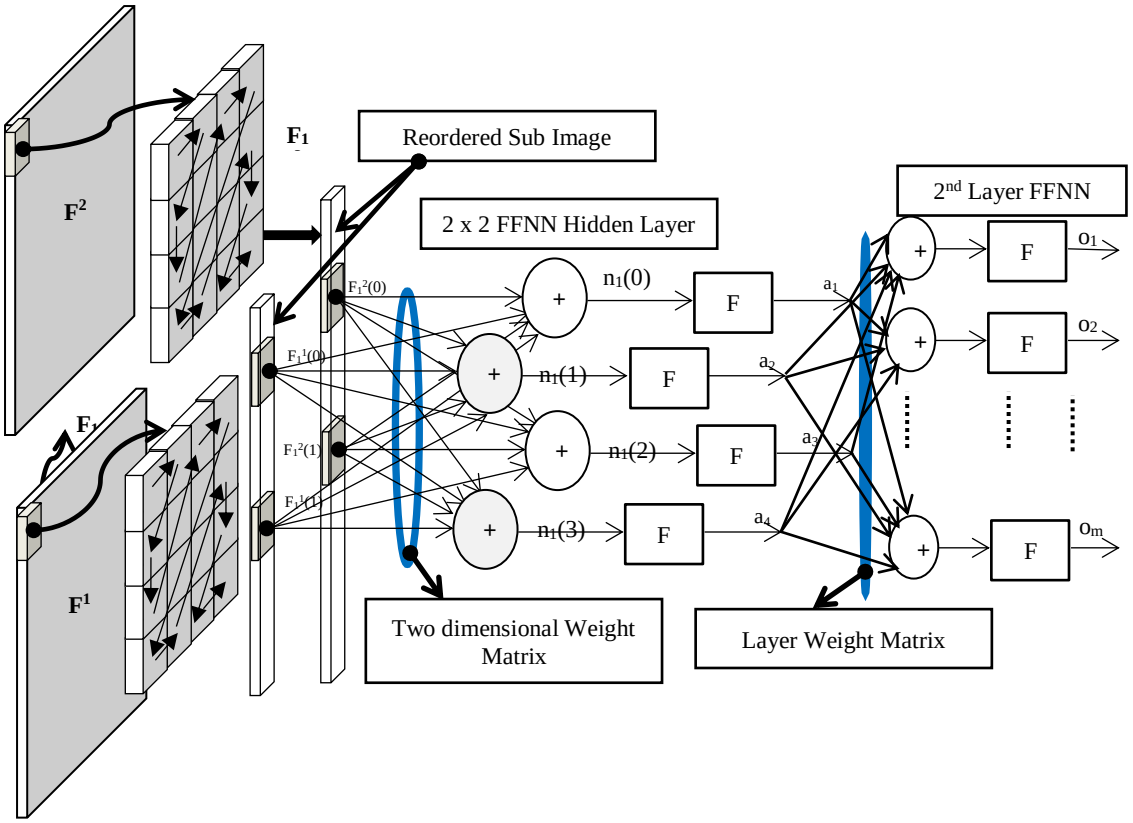


Figure 9 Data processing in the first layer of 2D FFNN structure

The two consecuting frames are processed by the 2D FFNN structure that processes 4 x 4 sub images of two consecutinve frames.The number of inputs in the input layer of 2D FFNN will be 32 input arranged into 2D matrix of 4x4 which will be processed by each neruon in the 2x2 model. The output of 2D FFNN will be 2x2 output. The 2D FFNNDC will process $N^2/64$ sub images per frame and will generate $2xN^2/64$ outputs per frame.The 2D FFNNHF will process $7N^2/64$ sub images per frame and generate $2xN^2/64$ outputs per frame. The FFNN structure of second and third are realized using the generic structure and the outputs of each of the layer is mathematically represented as in Equation(14) – (15).

$$d_k = \sum_{i=1}^{n} P_i \; W_{k,i} \; + b_k \; , a_k \; = f(d_k \;) \qquad (14)$$

$$c_m = \sum_{i=1}^{k} a_k \; W_{m,k} \; + b_m \; , o_m \; = f(c_m) \qquad (15)$$

The number of neurons in the 2nd and 3rd layer can be set according to the desing requriements. In this proposed design, the number of neurons in the hidden layer is set to 16 or 8 and the number of neurons in the output layer is set to 4 or 2. The designed FFNN structure reduces the dimensionality of the input data as well as detects the significant features that represent the salient objects. At each stage the purelin function is used for the hidden layer for training the NN and for the testing.

5. Implementation and Evaluation

The top level block diagram of the designed encoder and decoder that is based on 3D wavelet sub band is shown in Figure 10. The deep learning network designed is trained considering 30 different objects that are captured using 256 frames during motion. The captured video sequence is grouped into four GOFs each of 64 frames. The images are resized to 512 x 512 resolutions in

order to perform DWT and feature selection using modified SPIHT algorithm. By considering 64 frames per GOF, there are 120 GOFs that have 120 different objects captured with motion information. 3-level 3D DWT is carried out and the modified SPIHT algorithm is applied to retain the wavelet coefficients that are very significant towards salient object detection. 2-level 3D inverse DWT is computed and the wavelet sub bands are obtained for training the deep learning network. The inputs are set as the targets at the reconstruction network of deep learning structure. The number of neurons in the encoder module of deep learning network is set to either 4 or 8 or 16. The number of neurons in the hidden layer is adjustable and network activation functions set to both linear and nonlinear. Multiple iterations of training are carried out to evaluate to find the number of neurons for each hidden layer and appropriate network activation functions. The 2D FFNN structure and all the other FFNN structures are modeled in MATLAB environment including the reconstruction network. The training targets are set to achieve minimum gradient or MSE less than 10^{-5}, number of Epochs is set to 600 and number of iterations are set to 1000. The network is trained and the training results are presented in Figure 11 that is used to evaluate the network performances in terms of gradient or MSE achieved and regression number.

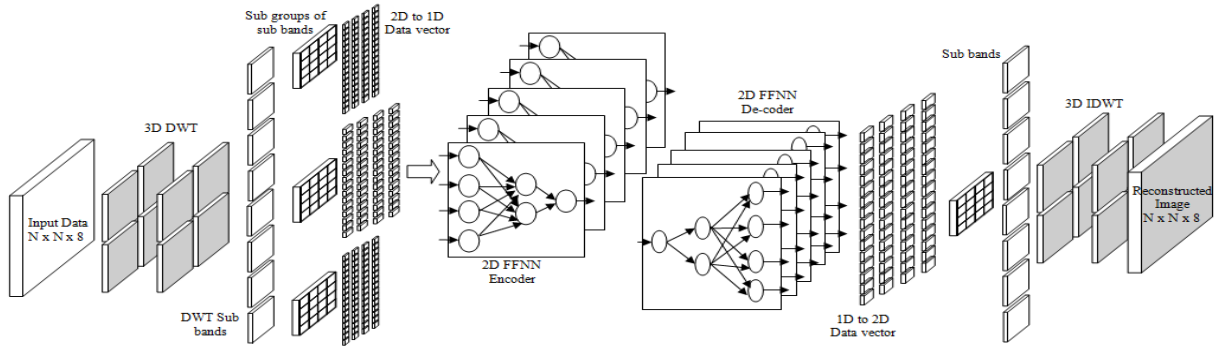
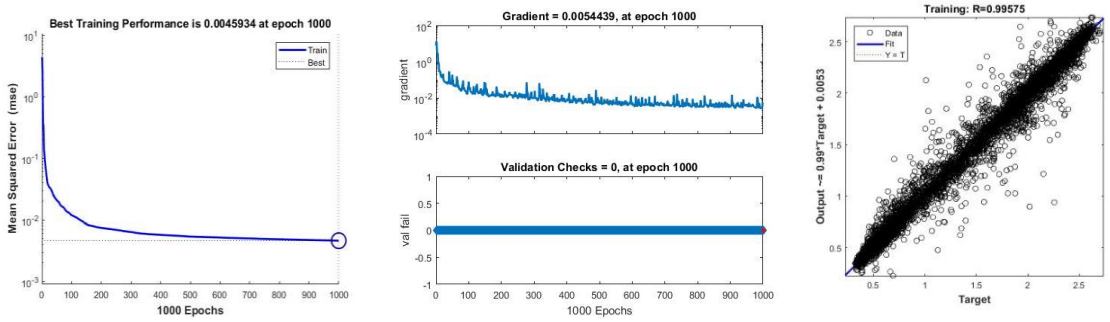


Figure 10 Encoder-decoder modules for salient object detection based on DWT and deep learning



(a) (b) (c)

Figure. 11 : Training performance of the Deep learning Network. (a) Performance plot (b) Gradient and validation check (c) Regression plot

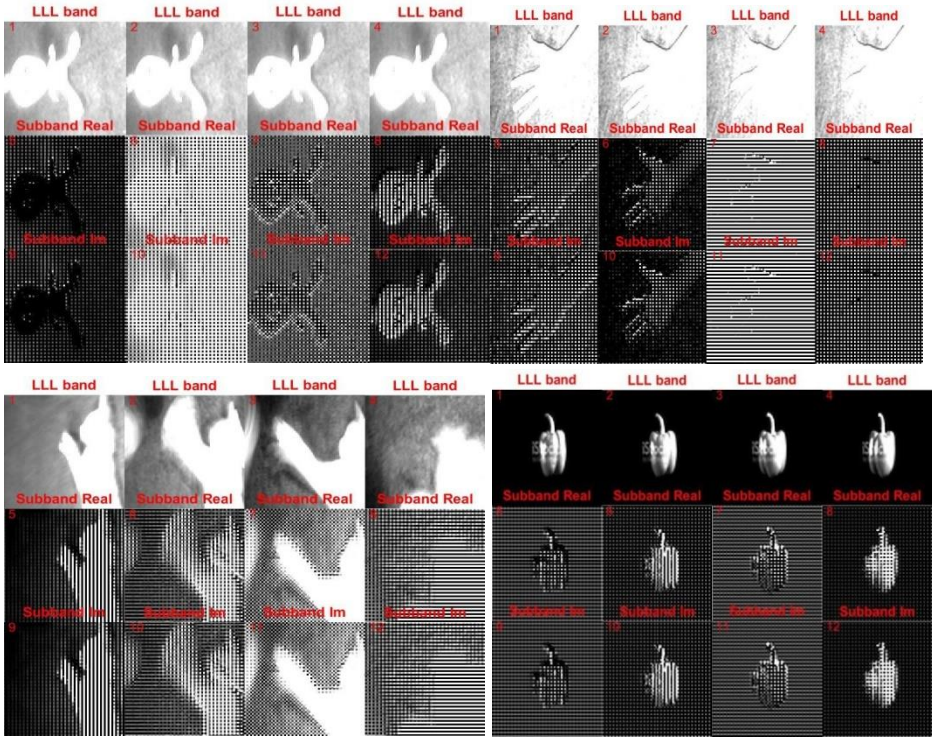
Figure 11 (a) presents the MSE obtained for reaching the best performance at 1000 Epoch which is found to be 0.0045934. From 800 Epoch the MSE decreases slowly and hence the network reaches its saturation point after 1000 Epoch. The plot of gradient is shown in the figure (b), it can be observed that as the number of epochs are increased to a maximum of 1000, the gradient keeps declining and a minimum gradient of 0.0054439 is reached at 1000 epoch. The regression plot is shown in figure (c), and obtained is the value of 0.99575, demonstrating that the network has reached global minima point and the weights and bias elements obtained are significant to detect all the 120 objects in the training input and reconstruct the original images with minimum

distortions. Also from figure (c), we can observe that the outliers in the data has a mismatch which is around 0.03%.

5.1 Results & Discussion

Figure 12 below shows images from some of the databases used for the training of the Neural networks. The images are from Tweety database, Hand2 Database, Cat2 database, Bell pepper database (courtesy istock images), Hand database (courtesy istock images) and the Hand1 database. The images are grouped in blocks of 12 and each set of 12 images are of one particular database. The images in each column is of the same frame number from the database. The database was created such that they all have the same uniform background. The images are taken after the process of DTCWT application to the database just before applying to the SPIHT algorithm. It shows the LLL band image of different frame numbers in the first row. The second row of the images are the images of one of the sub-bands containing the real coefficients. The third row of images show the imaginary coefficients of the same sub-band.

The input images undergo pre-processing such that the standard resolution of N x N x 64 is applied with 3-level DTCWT to generate the different wavelet coefficients for each sub-band.. The significant low pass and high pass coefficients are retained by applying the modified SPIHT algorithm and thereby the self-similarity components are removed. The reconstruction of the image is done by performing the 3-level inverse DTCWT. The salient object features are detected by the seven layers of the Deep learning model. The deep learning model and 3 level inverse DTCWT processes these features for reconstruction of input data. From the results obtained it is observed that, with a minimum amount of distortions, the original image is reconstructed from detected salient object features.



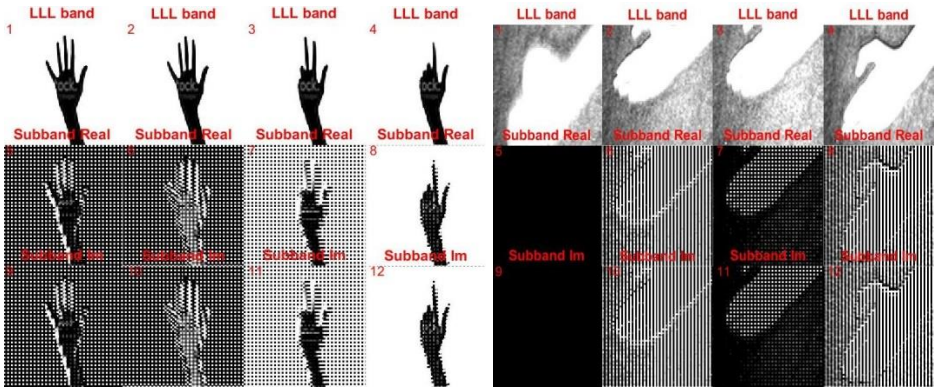


Figure 12: Database images

The hand1 database images has vertical, horizontal and diagonal edges that are reconstructed with least amount of distortion. The Tweety database images has features in the form of circles, vertical, horizontal and diagonal edges and these are also reconstructed with minimum distortions as visible from the results compared.

Table 2 presents the results in terms of MSE and PSNR of one of the sample databases obtained for the 32 different frames of a database. The input data undergoes encoding and decoding and for each of the frame MSE and PSNR are computed. The PSNR results for the three sample data sets are found to be more than 20dB and it is observed that the PSNR variation across all the 32 frames is within +/-5 dB, from the average value, demonstrating that the designed pair of encoder and decoder is capable of detecting the salient object features and reconstruct all the frames with motion vectors.

Table 2 MSE and PSNR results of Tweety database

	MSE			PSNR		
	LLL($x10^{-3}$)	SBR($x10^{-5}$)	SBI ($x10^{-5}$)	LLL	SBR	SBI
1	2.9	3.19	2.47	25.3692	44.9674	46.0719
2	2.5	3.02	2.12	26.1026	45.2011	46.7421
3	2.8	2.16	2.56	25.5661	46.6559	45.9228
4	3.9	2.87	2.29	24.0688	45.4185	46.3988
5	2.3	3.03	3.10	26.3496	45.1793	45.0889
6	2.8	2.45	2.95	25.5337	46.1093	45.2946
7	1.9	2.51	3.00	27.2904	45.9965	45.2223
8	2.7	2.45	1.90	25.6861	46.1151	47.2224
9	2.1	1.41	1.64	26.9691	48.5166	47.8586
10	2.8	2.02	2.28	25.4591	46.9545	46.4214
11	2.4	2.10	2.03	26.2075	46.7779	46.9262
12	6.5	2.78	2.44	21.8552	45.5546	46.1297
13	3.1	3.11	3.41	25.0622	45.0729	44.6772
14	2.9	3.27	4.00	25.3103	44.8481	43.9759
15	3.7	3.38	3.44	24.3294	44.7106	44.6399
16	2.5	4.02	3.08	25.9723	43.9531	45.1198
17	6.8	3.90	3.62	21.6755	44.0914	44.4073
18	2.7	3.50	3.32	25.7591	44.5585	44.7856
19	2.3	4.68	3.34	26.3347	43.2983	44.7656
20	2.6	3.62	3.48	25.7786	44.4171	44.5836
21	3.7	2.93	2.59	24.3751	45.3383	45.8676
22	2.2	2.67	3.06	26.5727	45.7411	45.1438

23	2.9	3.31	3.51	25.4172	44.8083	44.5463
24	3.8	2.87	2.58	24.2341	45.4246	45.8876
25	5.8	3.82	2.67	22.3533	44.1832	45.7406
26	2.9	2.25	2.33	25.3821	46.4753	46.3356
27	2.7	3.24	2.55	25.6226	44.8915	45.9267
28	1.8	4.51	3.53	27.5659	43.4536	44.5264
29	2.7	3.57	4.12	25.7647	44.4691	43.8556
30	4.6	3.64	3.82	23.4197	44.3947	44.1803
31	3.5	4.03	4.55	24.5242	43.9451	43.4163
32	6.3	3.51	2.98	22.0227	44.5429	45.2597

The PSNR for the Tweety database is observed to vary from 22 to 25 for the LLL band, this is a variation of 3 dB and the PSNR ranges from 44 to 46 dB for the sub-band for the real and imaginary coefficients. This is with +/- 5dB of the average value. The low PSNR in the LLL band indicates that there is loss in the original to the reconstructed frames. The high PSNR in the real sub-band and imaginary sub-band indicates very less loss. For example the variation in the PSNR between frame 29 and 30 in the LLL band accounts for changes in the salient features between the two frames in the LLL band and the corresponding sub-band has very less variation in salient features.

From the above table, we can compute the improvement factor for the database considered, for example, the 10^{th} frame with the PSNR of 25.4591 in the LLL band has a PSNR of 46.9545, this gives an improvement factor of 45.78 %. SBR is the sub bad real and SBI is the sub band Imaginary

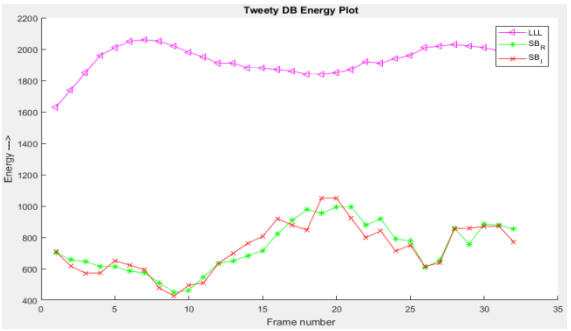
Table 3 MSE and PSNR of the reconstructed Tweety database

	PSNR Using the Modified SPIHTL	MSE Using the Modified SPIHTL	PSNR Without the Modified SPIHTL	MSE Without the Modified SPIHTL
1	38.5941	0.0311	35.3314	0.0403
2	38.4101	0.0323	34.5643	0.0412
3	38.9854	0.0327	36.4531	0.0698
4	39.9141	0.0317	36.8871	0.0363
5	42.8182	0.0654	39.1213	0.0475
6	41.6663	0.0364	38.0123	0.0342
7	43.1835	0.0407	44.1011	0.0379
8	41.9346	0.0362	39.0231	0.0586
9	41.4407	0.0709	39.1613	0.0493
10	40.8169	0.0356	39.3521	0.0423
11	41.0644	0.0635	42.1132	0.0427
12	41.6605	0.0391	39.1212	0.0433
13	42.8809	0.0403	39.8123	0.0435
14	43.8019	0.0348	40.7642	0.0480
15	44.6237	0.0417	43.5231	0.0713
16	44.8367	0.0406	39.6472	0.0428
17	43.3296	0.0331	39.5876	0.0644
18	43.6359	0.0535	40.2132	0.0330
19	42.7153	0.0423	38.4356	0.0377
20	41.3303	0.0372	37.3452	0.0360
21	41.5066	0.0295	40.2341	0.0322
22	41.4115	0.0322	38.4356	0.0430
23	42.1034	0.0336	38.4567	0.0387

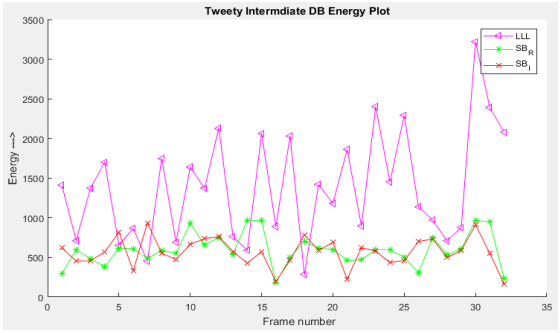
24	41.9986	0.0321	39.0678	0.0359
25	42.4835	0.0637	38.2564	0.0542
26	41.6221	0.0321	38.4522	0.0484
27	42.4501	0.0411	39.1573	0.0315
28	41.3026	0.0366	38.4562	0.0415
29	41.2513	0.0372	38.3214	0.0463
30	40.7593	0.0418	39.3563	0.0482
31	38.5124	0.0412	37.4256	0.0645
32	40.2192	0.0350	39.4562	0.0575

Table 3 shows the PSNR and the MSE for the sample Tweety database after the reconstruction from figure 10. The table also shows the MSE and PSNR using the modified SPIHT algorithm and without the modified SPIHTL algorithm. It can be observed from the results, the proposed method performs significantly better.

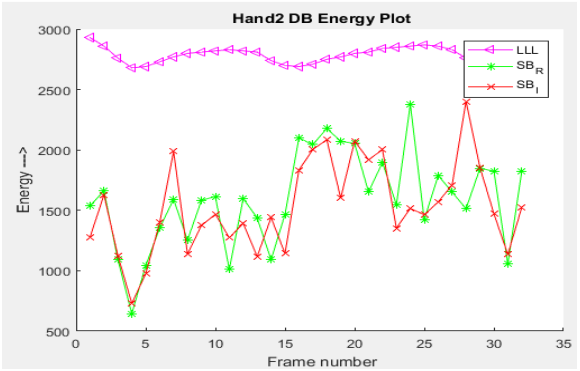
The energy content of the frames of the other example database considered, it is observed that the maximum of the energy of the frame is available in the LLL band and the sub band chosen consists of the maximum of the energy of the sub bands. The output of the hidden layer consists of the reduced energy of the corresponding bands. The figure 14 depicts the plot of the energy content of the different frames taken for sample.



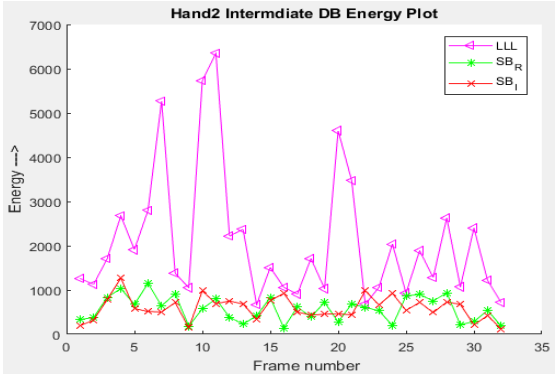
(a)



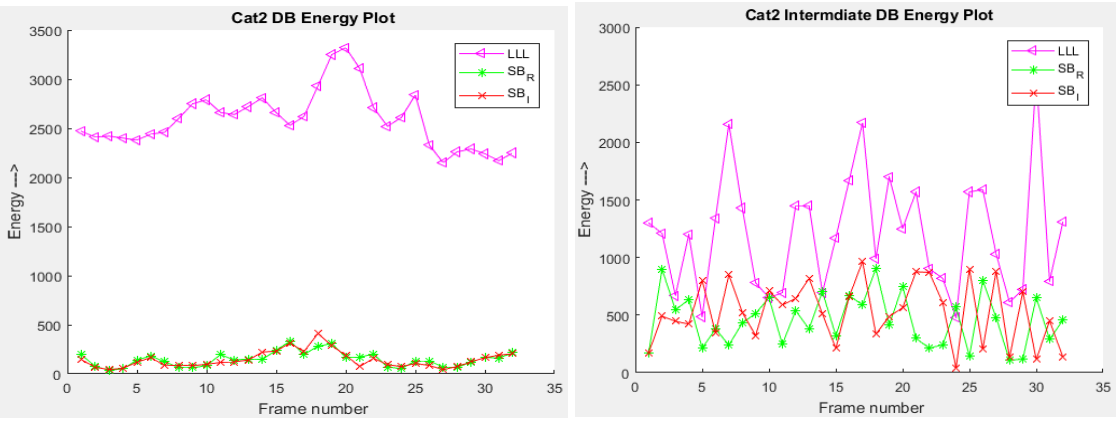
(b)



(c)



(d)



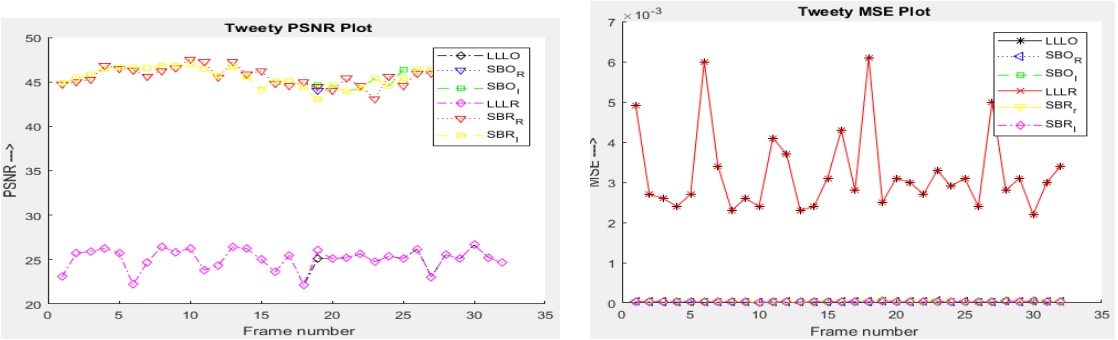
(e) (f)

Figure 14: Energy plot of the sample database for the LLL and chosen sub band

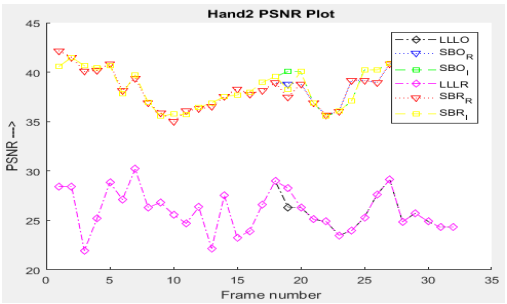
The plot shows the variation in the energy content for different frames across different databases considered. Where the distinction of the energy is good in some cases and overlaps in few cases. The detection of salient features is carried out successfully even with this overlap.

The designed model has the capability to reconstruct the salient object from the features without much distortion as visible from the result. Both the spatial and motion vector are extracted by the proposed in the wavelet domain and the deep learning module is trained for the detection of the salient features that represent the objects. The decoder is trained to reconstruct the original image from optimum number of features and as the DTCWT is used the limitations of the DWT are overcome. The model designed has the ability for detection of salient object features from more than 120 data sets and is a generic object detection design. Deep learning networks are widely used in salient object detection as they have demonstrated better performance compared to conventional methods. The deep learning methods uses more than 5 layers of data processing as a result the complexity increases this is addressed by reduction in dimension of data sets to a size of 32 x 32.

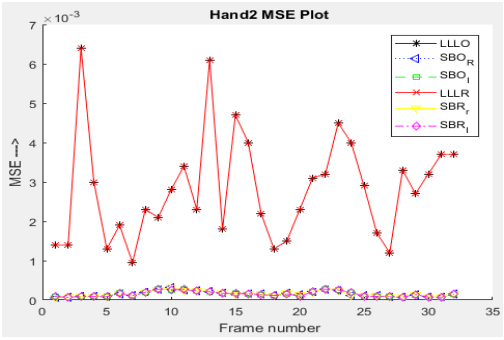
Figure 15 shows the PSNR vs the frame number plots and the MSE vs the frame number plots of the different sample databases.



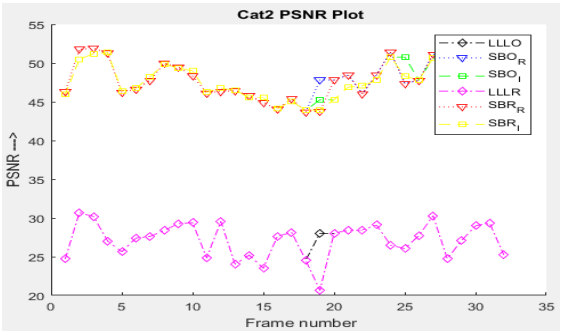
(a) Tweety Database PSNR plot (b) Tweety Database MSE plot



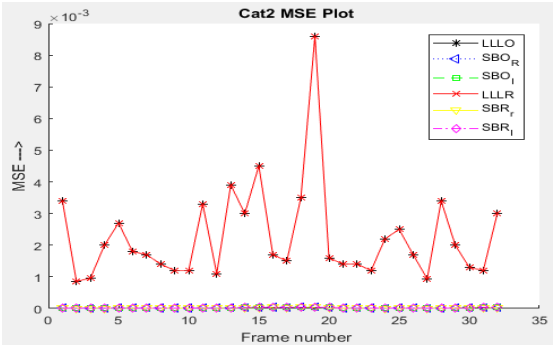
(c) Hand2 Database PSNR plot



(d) Hand2 database MSE plot



(e) Cat2 database PSNR plot



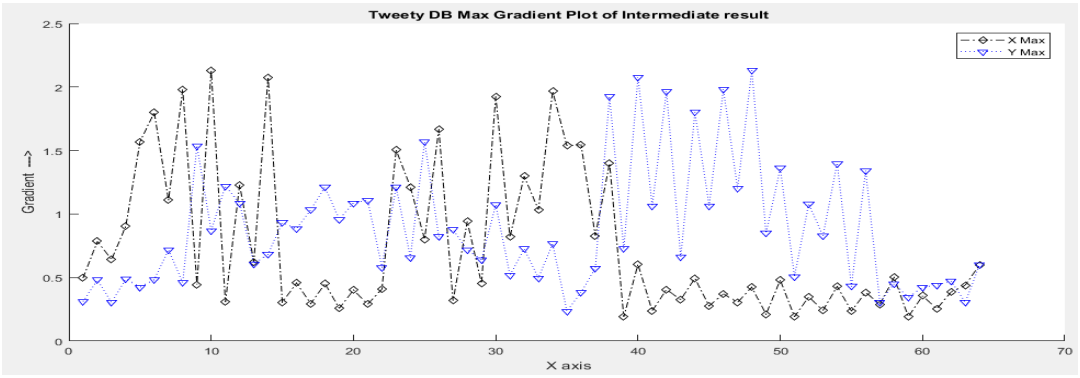
(f) Cat2 database MSE plot

Figure 15 PSNR and MSE plots for different databases

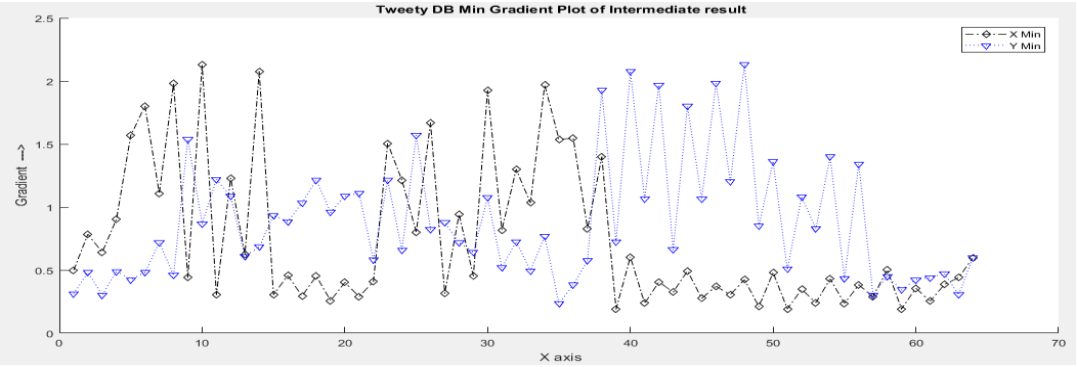
It can be seen that the PSNR for the Hand2 database varies from 20 to 25 for the LLL band, this is a variation of 5 dB and the PSNR range for the sub-band of the real and imaginary coefficients is from 36 to 40 dB and it is observed that the variation is with in the tolerance range.

We observe that the PSNR for the Cat2 database varies from 23 to 30 for the LLL band, this is a variation of 7 dB and the PSNR range for the sub-band of the real and imaginary coefficients is from 36 to 40 dB and we observe that the variation is within the tolerance range.

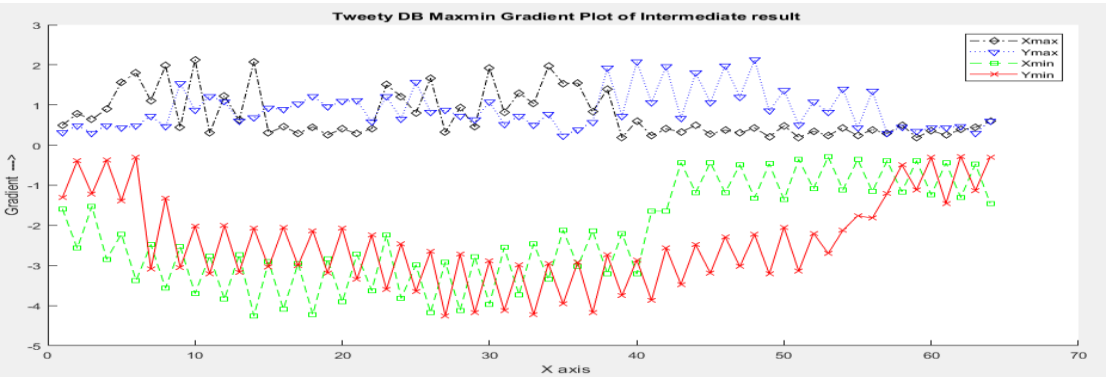
The figure 16 shows the gradient plot of the example database for maximum and minimum energy content of the database.



(a)



(b)



(c)

Figure 16: Gradient plot of the sample database.

Depicted above is the gradient plot of the sample database. Figure 16(a) is the plot of the variation of the maximum gradient along the X-axis and Y-axis. The figure 16(b) is the plot of the minimum gradient variation along the X- axis and Y – axis. The figure 16(c) is the plot of the variation of the gradient with maximum and minimum values at the hidden layer, for the database considered.

From the above results, it is evident that the salient features are detected, even if there is a loss of significant energy.

Conclusion

The 3D DTCWT model processes 512 x 512 x 64 into m-level sub bands. These sub-bands are then processed by the proposed algorithm of SPIHTL for the identification and removal of the self-similarity wavelet coefficients and to retain the coefficients that are most significant that constitute towards salient object features. The SPIHTL algorithm is designed to process the 3D DTCWT wavelet sub bands by considering individual frames. During the process of quantization, the elimination of the wavelet coefficients, from the SPIHTL ordered list, takes place. The quantized frames are rearranged and m-1 level inverse 3D DTCWT is carried out to reconstruct the image frames. Deep learning algorithm is designed to extract salient object from the video sequence with the help of the 2D FFNN, 1D FFNN layers and reordering layers resulting in optimum number of significant features. The video sequence is reconstructed form the salient object features using the designed decoder module. The performance of the algorithm is evaluated through the PSNR and MSE measurements for more than 120 different data sets. The developed algorithm has the advantage of identifying the salient object with greater accuracy and overcomes the drawback of the DWT method.

References

1. *Mishra, Y. Aloimonos, L. F. Cheong, and A. Kassim, "Active visual segmentation," IEEE TPAMI, vol. 34, 2012*
2. *Y. Li, X. Hou, C. Koch, J. M. Rehg, and A. L. Yuille, "The secrets of salient object segmentation," in CVPR, 2014.*
3. *Borji, "What is a salient object? a dataset and a baseline model for salient object detection," in IEEE TIP, 2014*
4. *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*
5. *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*
6. *Kanan and G. Cottrell, "Robust classification of objects, faces, and flowers using natural image statistics," in CVPR, 2010, pp. 2472–2479*
7. *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*
8. *Li, X. She, and Q. Sun, "Color image quality assessment combining saliency and fsim," in ICDIP, vol. 8878, 2013*
9. *Meger, P.-E. Forss´en, K. Lai, S. Helmer, S. McCann, T. Southey, M. Baumann, J. J. Little, and D. G. Lowe, "Curious george: An attentive semantic robot," Robotics and Autonomous Systems, vol. 56, no. 6, pp. 503–511, 2008*
10. *J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In CVPR, pages 3431–3440, 2015*
11. *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*
12. *Amar, C.B., Jemai, O., Wavelet networks approach for image compression. ICGST International Journal on Graphics, Vision and Image Processing pp. 37-45 (2007)*
13. *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).,*
14. *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)*
15. *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 (Sept 2013). https://doi.org/10.1109/ICDIM.2013.6693992*
16. *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*
17. *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)*

18. *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)*
19. *Fujieda, S., Takayama, K., Hachisuka,Wavelet convolutional neural networks for texture classification. CoRR abs/1707.07394 (2017)*
20. *Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q. (eds.) Advances in Neural Information Processing Systems 25, pp. 1097-1105. Curran Associates, Inc. (2012),*
21. *Phan Ngoc Hoang and Bui Thi Thu Trang, Context-aware hand poses classifying on images and video-sequences using a combination of wavelet transforms, PCA and neural networks, doi: 10.4108/eai.6-7-2017.152758*
22. *Viola P., Jones M.J. (2001) Rapid object detection using a boosted cascade of simple features // IEEE Conf. on Computer Vision and Pattern Recognition. Kauai, Hawaii, USA, V. 1, pp. 511–518*
23. *Gonzalez R. C., Woods R. E. (2001) Digital image processing. Reading MA // Addison-Wesley*
24. *Phan N.H., Bui T.T.T., Spitsyn V. G., Bolotova Y. A. (2016) Using a Haar wavelet transform, principal component analysis and neural networks for OCR in the presence of impulse noise // Journal Computer Optics, T 40, No 2, pp. 249–257*
25. *Phan N.H., Bui T.T.T., Spitsyn V. G. Face and Hand Gesture Recognition based on Wavelet Transforms and Principal Component Analysis // 7th International Forum on Strategic Technology IFOST: Proceedings of IFOST 2012, IEEE, (2012)*
26. *Y. F. Ma and H. J. Zhang, "Contrast-based image attention analysis by using fuzzy growing," in Proc. 11th ACM Int. Conf. Multimedia, New York, NY, USA, Nov. 2003, pp. 374–381*
27. *Muwei Jian, Kin-Man Lam, Junyu Dong, and Linlin Shen, Visual-Patch-Attention-Aware Saliency Detection, IEEE TRANSACTIONS ON CYBERNETICS*
28. *I. Daubechies, "Orthonormal bases of compactly supported wavelets," Commun. Pure Appl. Math., vol. 41, no. 7, pp. 909–996, 1988*
29. *N. Sebe, Q. Tian, E. Loupias, M. S. Lew, and T. S. Huang, "Content based retrieval using salient point techniques," in Proc. IEEE Conf. Computer. Vis. Pattern Recognition. (CVPR), Kauai, HI, USA, Dec. 2001*
30. *M. W. Jian, J. Y. Dong, and J. Ma, "Image retrieval using wavelet-based salient regions," Imaging Sci. J., vol. 59, no. 4, pp. 219–231, Aug. 2011*
31. *N. G. Kingsbury," The dual-tree complex wavelet transform: a new technique for shift invariance and directional filters", In the Proceedings of the IEEE Digital Signal Processing Workshop, 1998.*
32. *Eero P Simoncelli, William T Freeman, Edward H Adelson, and David J Heeger. Shiftable multi-scale transforms. IEEE Trans. Information Theory, 38(2):587–607, March 1992. Special Issue on Wavelets*
33. *N.G. Kingsbury, "A dual-tree complex wavelet transform with improved orthogonality and symmetry properties," in Proc. IEEE Int. Conf. Image Process., Sep. 2000, pp. 375–378.*
34. *J. Zho and S. Lawson, "Improvements of the SPIHT for image coding by wavelet transform", in Proc. IEE Sem. Time-Scale & Time Freq. Anal. Appl., 2000, pp. 24/1–24/5*

35. *Hermanus Vermaak, Philibert Nsengiyumva, and Nicolaas Luwes, "Using the Dual-Tree Complex Wavelet Transform for Improved Fabric Defect Detection" Hindawi Publishing Corporation, Journal of Sensors, Volume 2016, Article ID 9794723, 8 pages http://dx.doi.org/10.1155/2016/9794*
36. *Lowis, Hendra and Lavinia, "The Use of Dual-Tree Complex Wavelet Transform (DTCWT) Based Feature for Mammogram Classification ", International Journal of Signal Processing, Image Processing and Pattern Recognition Vol.8, No.3 (2015), pp.87-96 http://dx.doi.org/10.14257/ijsip.2015.8.3.08*