Saliency-Based Image Compression Using Walsh–Hadamard Transform (WHT)

A. Diana Andrushia and R. Thangarjan

Abstract Owing to the development of multimedia technology, it is mandatory to perform image compression, while transferring an image from one end to another. The proposed method directly highlights the salient region in WHT domain, which results in the saliency map with lesser computation. The WHT-based saliency map is directly used to guide the image compression. Initially, the important and less important regions are identified using WHT-based visual saliency model. It significantly reduces the entropy and also reserves perceptual fidelity. The main aim of the proposed method is to produce the high-quality compressed images with lesser computational effort and thereby achieving high compression ratio. Due to the simplicity and high speed of WHT, the proposed visual saliency-based image compression method is producing reliable results, in terms of peak signal-to-noise ratio (PSNR), compression ratio, and structural similarity (SSIM), compared to the state-of-the-art methods.

Keywords Saliency detection · Image compression · Walsh–Hadamard transform · PSNR

1   Introduction

Salient region detection drastically attracts the attention toward many of the computer vision and pattern recognition tasks such as image compression, object recognition, content-based image retrieval, image collection browsing, image editing, visual tracking, and human–robot interaction. It aims to detect the salient

A. Diana Andrushia (✉)
Department of ECE, Karunya Universtiy, Coimbatore, Tamil Nadu, India
e-mail: andrushia@gmail.com

R. Thangarjan
Department of CSE, Kongu Engineering College, Erode, Tamil Nadu, India
e-mail: thangs_68@yahoo.com

region of an image under biological plausibility. In this paper, Walsh–Hadamard transform (WHT)-based visual saliency detection is used for the image compression application.

Demand on data compression is increasing rapidly as the modern technologies are growing high. High storage capacity is required for uncompressed images. Many images are having the common characteristics as their neighboring pixels are highly correlated with redundant information [1]. Image compression-based techniques are aiming to reduce the redundant information by eliminating the spectral and spatial redundancies. It results in the reduction of consumption of expensive resources in the form of transmission bandwidth and hard disk space. Image compression techniques are generally classified into two types. One is spatial coding and another one is transform coding. In the transform coding type, discrete cosine transform (DCT), discrete Fourier transform (DFT), Walsh–Hadamard transform (WHT), etc., are used to perform natural image compression. Each transform is having its own advantages and disadvantages in the compression domain. Transform type coding is used in the proposed method.

WHT-based visual saliency detection and WHT-based image compression based on saliency map are the two main phases of this proposed method. The proposed WHT-based visual saliency detection is transform domain approach. So the frequency domain approaches are only considered for the performance comparison. As the WHT-based compression also transform domain approach so frequency domain methods only considered for fair performance comparisons in the compression phase.

The proposed work is experimented through MIT dataset. It is one of the benchmark datasets for visual saliency detection which consists of the indoor, outdoor images. The performance metric of receiver operating characteristics (ROC), area under the curve (AUC), precision, recall, and F-measure is obtained to analyze the proposed visual saliency detection. Peak signal-to-noise ratio (PSNR), structural similarity (SSIM), and compression ratio are obtained for saliency-based image compression. The performance metrics are yielding significant results while comparing with state-of-the-art methods.

WHT is used in the saliency detection as well as image compression. WHT is chosen because the number of computations in this transform is significantly less compared to the other transforms, and WHT is the key transform to provide energy compactness. Finding the salient information in image/audio/video will reduce the number of computations and lesser hardware in the compression techniques. The proposed visual saliency-based compression method achieves 90% compression ratio. It is due to the lesser computations of WHT transform. This is the most added advantage of the proposed method compared to the state-of-the-art methods.

The reminder part of this chapter constructed as follows: Sect. 2 clearly explains about the backgrounds of visual saliency detection, saliency-based image compression. Section 3 elaborates the proposed methodology. Section 4 explains the experimental results with the performance metrics. Finally, the conclusion of the chapter is given in the last section.

2 Backgrounds

2.1 Saliency Detection

The world is full of visual data. Humans selectively perceive the visual information that is getting in through their eyes. Visual attention is the process of selecting particular information from the plenty of raw data perceived. For example, while sincerely watching a cricket match in the play ground, a sudden change in the action of the umpire with red shirt picks the attention of the spectators in the gallery despite the colossal load of visual inputs such as actions of the batsmen, bowlers, and fielders is there. The eyes of the spectators gaze the umpire momentarily before shifting to other events in the visual scene. The phenomenon of drawing the focus of attention to certain regions in a given scene or image is called visual attention [2]. In the jargon of computer vision, these regions are known as *salient* regions.

Detection of salient regions finds application in a wide spectrum of processes such as automatic object detection, image retrieval, remote sensing, automatic target detection, image and video segmentation, robotics, scene understanding, computer–human interaction, driver assistance, action recognition, background subtraction, image and video compression, video summarization, medical imaging, and automatic cropping. The cognitive process that directs human to select highly relevant data from a scene/image is named as visual attention.

Recently, a number of computational models have been developed to highlight salient regions. As far as computational models are concerned, there are two types of models in the literature, namely bottom-up and top-down approaches. Bottom-up approach works from the low-level visual features and moves up to construct a saliency map. The top-down approach is goal-driven, and it uses prior knowledge to achieve the goal such as target detection, scene classification, and object recognition [3]. The top-down approach starts from a particular visual processing task.

The computational models of visual attention either bottom–up or top–down can be constructed in spatial domain or in frequency domain, in order to highlight the salient regions as saliency map. Spatial domain methods require more computation time to obtain the features compared to the frequency domain methods [4]. The computational complexities of these models are very high, and these models are not performed with multi-scale spatial and frequency analysis. Many models which come under frequency domain approaches have used only local features to identify the salient regions.

The very first method of saliency detection is developed by Itti et al. [5]. Local contrast information is used to develop the method. The local features are only used by Ma et al. [6], Harel et al. [7], and Goferman et al. [8] to obtain the visual saliency detection. The global features are considered in [9–12] to construct the visual saliency model. However, the accurate identification of salient regions should also involve the global features. In recent years, many researchers have shown more interest to build computational visual models in the transform domain. Mainly

Fourier transform (FT) and wavelet transform (WT) have also been extensively used to highlight the salient regions in the transform domain. Every approach has its own pros and cons.

FT gives promising results for applications involving only stationary signals. The amplitude and phase spectrum of Fourier transform is used in [13]. Guo et al. [12] used the phase spectrum of quaternion Fourier transform (PQFT) to highlight the saliency, and it is also applied for efficient video compression. Hou et al. [14] used Fourier transform and log spectrum to construct the spectral residual approach for the saliency detection.

WT has the capacity to provide multi-scale spatial and frequency analysis because it codes the signal at different bands and bandwidths. WT can represent singularities in a much better manner than FT can. And moreover, WT can be applied for nonstationary signals also [15]. WT is used in [16] to find the salient object. WT-based orientation feature maps are obtained in different scale. The order map is also found by using Fourier analysis. The local, global information are used in the WT-based salient point detection [17]. WT-based salient detection is used in multi-scale image retrieval problem.

Wavelets are very good to represent point singularities, but when it comes to directional features they fall short [18]. The main reason is that wavelets are generated by isotropic elements. The quantity of decomposition level should be very large when approximating a curve using WT. The disadvantages of WT have been overcome by using multi-directional and multi-scale transforms.

The higher directional wavelet transforms of ridgelet, curvelet, shearlet are also used for the visual saliency detection. The directional features are captured effectively, and the potential salient regions are identified. Bao et al. [19] proposed visual saliency detection based on shearlet transform, in which the local and global contrasts are used to obtain local and global saliency map. Initially, the potential salient regions are identified in order to update the feature maps in shearlet domain.

Even though the transform based saliency detection methods are producing reliable results, suffered highly from computation complexity. Larger computations are required for these methods.

In general, the transform domain visual saliency detection methods are using the following steps.

- Transform the input image into transform domain,
- Obtain the feature maps for various features,
- Combine the various feature maps,
- Use the top-down features if required for particular application,
- Apply inverse transforms to get the saliency map.

The saliency map is the topographical map which shows the visual saliency in the visual scene. So the visual saliency detection methods are showing the outputs as visual saliency map. Saliency-based image segmentation, image compression, and image retrieval are the popular areas of research.

The Walsh–Hadamard transform (WHT) has lesser computations and extremely fast transform. It is computed only by addition and subtraction. Lesser hardware is required for the practical implementation [20]. The highly correlated pixels are captured by the WHT in the visual space. Hence, in this chapter WHT is used to detect the salient regions in transform domain.

2.2 Visual Saliency-Based Image Compression

Usually, the important regions of an image may be small and highly degraded at low bitrates. The standards of compression such as JPEG/JPEG-2000, MPEG 4 are not handling the salient regions well. Guo et al. [21] and Hadi et al. [22] investigated saliency-based compression techniques. To adhere the saliency values, the transmitted coefficients are modified. These methods cannot handle salient regions well and also suffer from complicated computations. Barua et al. [23] developed wavelet-based image compression technique for images and videos. The algorithm is designed to obtain the saliency values in wavelet domain and then corresponding image/video coefficients are transmitted. It preserves the important region of an image/video.

Nabil Ouerhani et al. [24] proposed adaptive color image compression based on biologically inspired visual attention. The initial stage perceptual salient regions of interest are identified automatically. The adaptive coding scheme allocates higher number of bits for the salient regions. The results are compatible with the JPEG standards.

Li et al. [25] performed video compression based on computational models of visual attention. The salient regions are encoded with higher quality compared with non-salient regions. The salient regions are awarded with higher priority rather than others. But it may generate visible artifacts in the non-salient part where the quality of image is poorer. The artifacts also sometimes draw the end-user's attention. In several cases, the high level of artifacts becomes salient and captures the viewer's attention. But the notable artifacts are not to be salient.

Hadizadeh [26] dissertation reveals the visual saliency methods for video compression and transmission. Saliency-based video coding is investigated. The main concept is that high salient regions are having higher ability to percept than lesser salient regions. The quality of image/video is handled toward the user most attended regions. This method effectively performs video coding expect in two major cases. If any region is richly salient, then its saliency will be increased after the compression, provided the quality of the image/video remains high. The reason is that the users are noticing the high salient regions in the scene. If the region is lesser salient, then its saliency will be decreasing after the compression task, because the lower saliency regions are ended with lesser quality.

Ho-Phuoc et al. [27] proposed the visual saliency-based data compression for image sensors. The adaptive image compression is presented in each block. First the saliency value is obtained, and then the Haar wavelet transform is applied for the compression. This framework gives lesser memory and compact operators. The data stored in the image sensors are very much reduced, and image quality is not altered.

Zundy et al. [28] proposed the content-based image compression using visual saliency methods. Initially, the saliency map is obtained from the video, automatically or by user input. The salient regions are performed with nonlinear image scaling. Salient image regions are given higher pixel count, and non-salient regions are given lesser pixel count. Existing compression techniques are utilized to compress the nonlinearly down-scaled images, and in the receiver end it is up-scaled. This method supports for anti-aliasing effect which reduces the aliasing in highly scaled regions.

In order to reduce redundant information in the dynamic scenes, the visual saliency in videos is proposed by Tu et al. [29]. Based on the video visual saliency map, the redundant information is removed. In this paper, video visual saliency is the catalyst of video compression technique. DCT is utilized to perform the video compression, and this technique is adopted for MPEG-1, MPEG-4, and H.265/HEVC standards.

Yu et al. [30] used visual saliency to guide image compression. At individual scales, the saliency is measured through Laplacian pyramid. The proposed compression algorithm decreases the entropy of the image with respect to the saliency map in each scale. Dhavale et al. [31] proposed visual computational model-based image compression. It successfully locates the regions of interest of the human and thereby applies for image compression.

Duan et al. [32] proposed the image compression technique based on saliency detection and independent component analysis. The input image is transformed first using ICA. The transformed coefficients are numbered with set zero coefficient percentage. The sparse nature of independent component analysis is used in this method. It is compared with DCT-based compression method.

Many of the state-of-the-art methods are failed to show the energy compactness. In the field of image compression, energy compactness is the key point and these methods also suffer from higher computational complexity. In order to revoke the key points, the WHT is used in the proposed image compression method. The redundant pixels are captured by the WHT with lesser computations.

2.3 Walsh–Hadamard Transform (WHT)

Discrete Fourier transform (DFT), discrete cosine transform (DCT), and Walsh–Hadamard transform (WHT) are widely used in the image processing applications. These linear image transforms are chosen in the image processing application because of their flexibility, energy compaction, and robustness. These transforms effectively extract the edges and also provide energy compaction in the

state-of-the-art methods. Among all these transforms, WHT is very gorgeous one because of its simplicity and its computational efficiency. The major properties of WHT are same as that of other image transforms. The basis vector components of WHT are orthogonal, and it is having binary values (±1) only.

WHT is orthogonal, non-sinusoidal transform which is used in image filtering, speech processing, and medical signal analysis. To be more specific the lunar images/signals are well processed, coded, and filtered by WHT. It is known well because of its simplicity and fast computation. WHT is the substitute of Fourier transform. It is computationally simpler because it requires no multiplication or division operations. Every computation is performed by simple addition and subtraction operation. WHT is one of the very fast transforms which can be implemented in $O(N\log_2 N)$ additions and subtractions. So the hardware implementation of WHT-based applications is also so simpler [20].

So it is beneficial in terms of energy consumption and lesser computation. WHT is real, orthogonal, and symmetric $H = H^* = H^\mathrm{T} = H^{-1}$. Walsh–Hadamard transform is represented in terms of Walsh–Hadamard transform matrix (WHTM).

It consists of set of N rows denoted by H_j for $j = 0, 1, 2 \ldots N - 1$. The properties of WHT matrix are:

i. H_j takes values as +1 and −1,
ii. The size of WHT matrix is usually the power of 2,
iii. $H_j [0] = 1$ for all j.

The size of the WHTM is generally the power of two. The second-order Hadamard matrix is given by, $H = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$.

The WHTM with the order of 4 is $\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{pmatrix}$.

Each row in the matrix is called as basis vector of WHTM [33]. Generally, the basis vectors are orthonormal and orthogonal. Orthonormal means dot product of each basis vector themself is one. Orthogonal means dot product of any two basis vectors is zero.

The computation of WHT involves very simple step only, when the image is projected into basis images, each pixel is multiplied by ±1, whereas the FFT needed complex multiplication. So WHT is more efficient than FFT in terms of computation complexity.

Consider the image C of size $N \times N$ with the pixels of $c(x, y)$, the 2D WHT is defined as

$$H(u, v) = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} c(x, y) g(x, y, u, v) \tag{1}$$

where $H(u, v)$ is the WHT-transformed image, $g(x, y, u, v)$ is the WHT kernel function. The beautiful property of WHT which is related to the energy conservation is given below

$$\sum_{x=0}^{N-1}\sum_{y=0}^{N-1}|c(x,y)|^2 = \sum_{u=0}^{N-1}\sum_{v=0}^{N-1}|H(u,v)|^2 \tag{2}$$

The energy conservation property exists between the spatial and Walsh–Hadamard domains. The average brightness of the image is calculated in terms of zero sequence term of WHT matrix

$$H(0,0) = \sum_{x=0}^{N-1}\sum_{y=0}^{N-1} c(x,y) \tag{3}$$

In this proposed method, Walsh–Hadamard transform plays vital role. WHT is used to detect the salient regions in the transform domain. Visual saliency-based natural image compression is proposed in this chapter. The redundant pixels are identified with the help of saliency map and WHT. The saliency map is also used to identify the correlation among the pixels effectively.

3 Proposed Method

The framework of the proposed work is given in Fig. 1. The proposed method involves two phases. The first phase is the saliency map computation based on WHT and entropy. The second phase consists of WHT-based image compression in which saliency map is acting as catalyst.
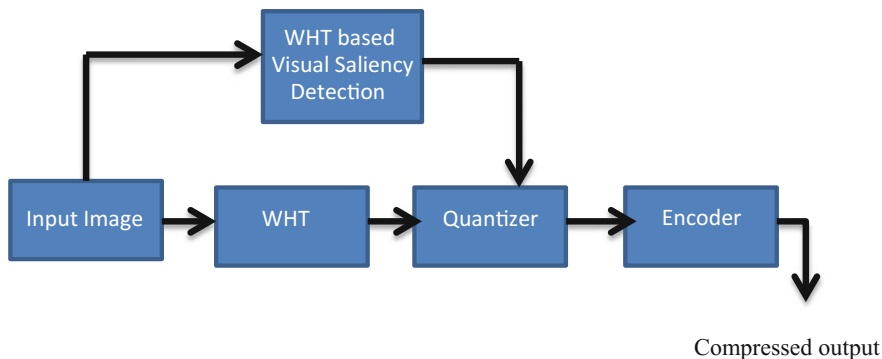

Compressed output

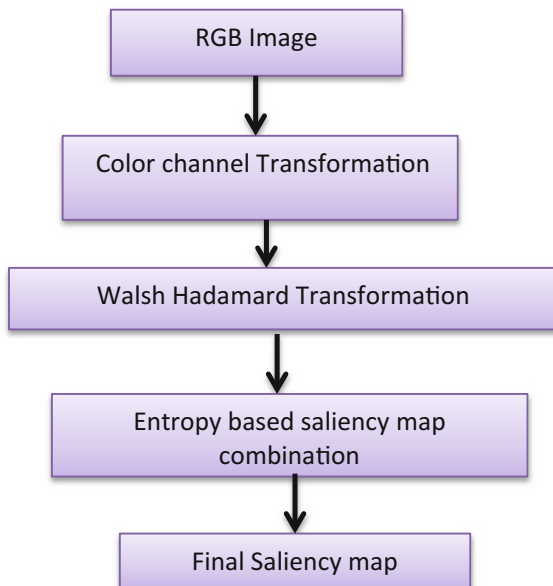**Fig. 1** Framework of proposed saliency-based image compression

Visual saliency detection is used to reduce the correlation between the pixels. Therefore, the redundant information in the pixels is identified. WHT-based quantization table is utilized. The quantization block is used to reduce the number bits needed to incorporate the transformed coefficients. In order to get the compressed image, entropy coding is used in the encoder block.

3.1 WHT-Based Saliency Map Computation

The human visual system is richly structured, and high redundant information is available in the sensory input. The predominant redundant information of human arisen from the input and the human visual system has the ability to reduce the redundant information of visual data [34, 35]. In this proposed method, WHT is used to seize the highly correlated features in the visual field and overwhelm them. So the salient visual features will be highlighted. Figure 2 is describing about the WHT-based saliency computation.

In general, natural images can be classified into salient and non-salient regions. The uncommon regions of an image are identified as salient regions. The uncommon regions of an image are generally caused by the preattentive features such as color, orientation, texture, and shape. In order to identify the salient regions, the suitable color channels must be used. Luo et al. [36] revealed that the salient regions are identified easily in specific color channels. Under different conditions, the images may have different contrasts in the diverse color channels. Usually YCbCr

Fig. 2 Framework of WHT-based saliency map computation

RGB Image

Color channel Transformation

Walsh Hadamard Transformation

Entropy based saliency map combination

Final Saliency map

color channels, Lab color channels are used in the saliency detection literature. In the human visual system, the color space of images is separated into various decorrelated color channels. RGB is one of the highly correlated color spaces, whereas lab color space is a well decorrelated color space with uniformity [3].

Recently, RGBYI color channels are used in the saliency detection [37]. In this proposed method also, RGBYI color channel is employed because this color channel mimics the visual cortex of human system by involving color channel pairs such as yellow/blue, blue/yellow, red/green, and green/red. It also includes color difference channels and four broadly tuned color channels. Let r, g, and b be the red, green, and blue color channels of an input image. The grayscale image is obtained by Eq. (4). Four broadly color channels and their pairs are obtained by Eqs. (5), (6), and (7)

$$\text{Intensity} = \frac{r + g + b}{3} \tag{4}$$

$$\text{RE} = r - \frac{g + b}{2}, \quad \text{GR} = g - \frac{r + b}{2}, \quad \text{BL} = b - \frac{r + g}{2} \tag{5}$$

$$\text{YELLOW} = \frac{r + g}{2} - \frac{|r - g|}{2} - b \tag{6}$$

$$\text{REGR} = R - G \text{ and BLYE} = B - Y \tag{7}$$

The input image is represented through r, g, b, RE, GR, BL, REGR, BLYE, intensity color channels. The components of WHT reflect the global features in the visual field. So the highly correlated features are seized by WHT components. The 2D WHT is performed by each row-wise and column-wise. So the transformation of WHT is very much simpler compared to the fast algorithms, and so, the computational complexity of WHT is extremely lower than other transforms. Let consider I^C be the color image. Each channels of I^C are applied with Walsh–Hadamard transform, and the transformed image is represented as

$$X = \text{sign}\left(\text{WHT}\, I^C\right) \tag{8}$$

Sign (.) represents signum function. WHT represents 2D Walsh–Hadamard transform. It holds the sign information of WHT components and omitting the other information such as amplitude, phase. Binary information (1 and −1) is only considered. So WHT is very compact by assigning single bit for each component. The WHT components are normalized by setting 1 for all positive coefficients and −1 for negative components. Sign (.) function usually boosts the image features as binary decision process. It reduces the highly correlated features in the visual field by normalizing WHT coefficients, thereby yielding visual saliency in the transform

domain. The inverse Walsh–Hadamard transform is applied to get the visual saliency map in the transform domain.

$$Y = \text{abs}(\text{IWHT}(X)) \tag{9}$$

All the nine color channels are performed with these steps. The final saliency map is obtained by combining the saliency maps at each color channel. The concept of entropy is vested to combine the saliency maps. The salient region and non-salient regions are differentiated from each other form in the saliency map at certain level. The salient regions may produce higher values, whereas non-salient regions produce lesser values. The weight of the saliency map is calculated by taking reciprocal of the entropy.

$$\text{SMAP} = \sum_{i=1}^{9} \frac{Y_i}{M\{H(Y_i)\}}. \tag{10}$$

'M' is for the normalization operator [0 1], where H (.) is the entropy. It is defined as

$$H(y) = -\sum_{i=1}^{n} p_i \log p_i. \tag{11}$$

p_i is the probability of different values of y. It varies from 0 to 255. 'n' is the possible values of 'y'. If the saliency map is having higher entropy, then the weight will be a lower one. The weight of the saliency map and its entropy are inversely proportional to each other. To smooth the saliency map, the Gaussian low-pass filter is applied before the updation of entropy.

The major steps in the WHT-based saliency detection are given below:

i. In the initial stage, input image is converted into RGBYI color channels with suitable scale,
ii. Apply Walsh–Hadamard transform (WHT) for each color channel and obtain the saliency maps for each channel,
iii. The final saliency map is obtained by combining the saliency maps using entropy concept.

3.2 Saliency-Based Image Compression

The multimedia data of image, audio, and video require extensive capacity to store. If these data are compressed, then quite a considerable amount of space is only

needed for transmission or storage. In the lower capacity, space is not sufficient to store large amount of data. So it is mandatory to perform compression while transferring the large amount of data from one end to other end. The reduction of redundant information in the multimedia data is the main goal of the compression techniques.

The proposed method highlights how the visual saliency-based methods are used to reduce the redundant information in the input and thereby performs compression. The framework of the proposed method is shown in Fig. 1. In this proposed method, there are three major blocks. Initially, the input image is divided into subblocks in order to apply Walsh–Hadamard transform. It is followed by quantizer unit. The quantitation table depends on the output from visual saliency detection and the coefficients of WHT. The modified quantitation table is used in the encoder unit. Entropy-based encoding technique is vested in the proposed method.

3.2.1 Preprocessing Stage

In order to design the quantitation table for quantizer, it is necessary to perform preprocessing steps in the initial stage. Human visual system is more sensitive to luminance than chrominance. RGB color space is highly correlated. The inter-component correlation among Red, Green, and Blue color channels is high. In order to obtain good image compression performance, it is necessary to reduce the correlation among the color channels [38].

YCbCr color space is one of the decorrelated color spaces. So RGB color space is not well suited for image compression. In this proposed method, input image is converted from RGB color space to YCbCr color space. Each color channel of YCbCr is subsampled into 8×8 nonoverlapping blocks.

3.2.2 Transform Domain

HT is non-sinusoidal orthogonal transform. It has wide application in the field of image processing. It is so popular among the research community because of its decorrelating capability. It also requires simple hardware implementation. The kernel function of WHT based on square and rectangular pulses with the highest magnitudes of ± 1. For two-dimensional images, the WHT and IWHT kernels are identical. It is due to the symmetric matrix in the kernels. Orthogonal rows and columns are in the symmetric matrix. So the forward and inverse WHT kernels are identical. The 2×2 smallest orthogonal Hadamard matrix is given by,

$$H_2 = \frac{1}{\sqrt{2}} \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix}$$

The WHT coefficients of an image I^C with the size of $N \times N$ are computed by

$$H(x,y) = \frac{1}{N} \sum_{r=0}^{N-1} \sum_{c=0}^{N-1} I^C(r,c)(-1) \sum_{i=0}^{n-1} [b_i(r)p_i(x) + b_i(c)p_i(y)] \quad (12)$$

Each 8×8 nonoverlapping block is applied with WHT, where $H(x,y)$ is the result of Hadamard transform. $I^C(r,c)$ is the image pixel value in spatial domain at (r,c) position. At frequency domain, the index is represented as (x,y).

N is the dimension of an image $N = 2^n$. The binary representation of r at ith bit is denoted as $b_i(r)$. $p_i(x)$ in terms of $b_i(x)$ is given below:

$$p_0(x) = b_{n-1}(x) \quad (13)$$

$$p_1(x) = b_{n-1}(x) + b_{n-2}(x) \quad (14)$$

$$p_2(x) = b_{n-2}(x) + b_{n-3}(x) \quad (15)$$

$$\text{simillarly } p_{n-1}(x) = b_1(x) + b_0(x) \quad (16)$$

The transform coefficients are used in the construction of quantization matrix.

3.2.3 Quantitation

Quantitation step is used to reduce the number of bits needed to store the transformed coefficients. WHT coefficients are representing the frequency components in the range of low, high, and middle. The saliency mask is generated to highlight the high-frequency components. It obtains the coefficient depending on the saliency map.

$$\text{SMAP}'(x,y) = \begin{cases} 0, & \text{if } x \leq 0.08 \text{ and } y \leq 0.08 \\ 1, & \text{otherwise} \end{cases} \quad (17)$$

Visual saliency-based modified Hadamard matrix is obtained by multiplying the WHT coefficients with the human visual saliency map.

$$H'(x,y) = H(x,y) * \text{SMAP}'(x,y) \quad (18)$$

In this proposed method, quantitation is performed for each block with respect to the saliency maps which are obtained in Sect. 3.1. Now, the quantitation matrix contributes the perceptual quality of human attention. The quantization matrix can be obtained by

$$Q(x,y) = \frac{q}{H'(x,y)} \quad (19)$$

where $Q(x, y)$ is the quantization matrix; q is the step size of the uniform quantizer; $H'(x, y)$ is visual saliency-based modified Hadamard matrix. The quantization matrix is rounded off and is represented as.

$$Q'(x, y) = \text{round}[Q(x, y)] \tag{20}$$

3.2.4 Encoding

According to human visual perception, the removal of high-frequency components is not at all achieving considerable impact on the input. The reduction of high-frequency contents yields the output of better compression [39]. Mainly, the human fixations felt into salient regions of an image. Huffman coding and arithmetic coding are needed with additional steps to encode the transform coefficients. Due to these extra steps, there are certain losses in the information. However, entropy coding is a lossless technique. In this proposed method, entropy-based encoding is used.

In this proposed method, saliency map-based quantitation matrix is obtained which reduces the correlation between the two adjacent pixels of an input image. If the correlation between pixels is reduced, then automatically the performance will be increased.

   i. Convert RGB color space to YCbCr color space,
  ii. Divide the input image into 8×8 blocks,
 iii. Each block is applied with Hadamard transform,
 iv. Obtain saliency map-based quantitation table,
  v. Perform the quantitation and rounding.

The redundant pixels are captured by the WHT with the assistance of second-order correlation statistics.

4 Experimental Results

According to the proposed method which is mentioned in Sect. 3, there are two major parts. First part is Walsh–Hadamard-based saliency map generation. Second part is saliency map-based image compression. Following subsections clearly intimate the results of WHT-based saliency detection and saliency-based image compression.

4.1 Results of WHT-Based Saliency Detection

In this section, the objective and subjective results of WHT-based saliency detection are discussed. The proposed method is compared with the seven state-of-the-art saliency detection approaches. However, the transform domain approaches are only taken for considerations. To validate the results of the proposed method, 1003 images are used from MIT dataset [40]. It consists of indoor, outdoor images, and portraits. All these images are captured from 15 viewers under free-viewing task.

The selected extant methods are IT in Itti et al. [5], VS in Li et al. [13], SR in Hou et al. [14], SE in Murray et al. [41], WT in Imamoglu et al. [3], ST in Lei Bao et al. [19], and FT in Achanta et al. [9]. It is the very first computational model in the field of visual saliency detection. The other six models proposed are in various transform domains. Fourier transform-based methods [9, 13, 14], wavelet transform-based methods [3, 41], and shearlet transform-based methods [19] are considered for the fair comparison. In order to enumerate the consistency of the results, two major evaluation metrics are obtained.

Receiver operating characteristics (ROC) and area under curve (AUC) are the major metrics. It is found in many of the major saliency detection methods for objective evaluation. Receiver operating characteristics (ROC) are plotted between true-positive rate (TPR) and false-positive rate (FPR). In order to do a fair comparison, all saliency maps are obtained by the state-of-the-art models, and the proposed method is normalized in the same range [0, 255] of original images. The better visual image saliency detection method is tried to have huge AUC values.

Figure 3 shows the subjective comparison of saliency maps of the proposed method and seven existing methods on MIT benchmark dataset. The proposed WHT method can highlight the salient region even though the image background is cluttered (row 5 in Fig. 3). The images with complex background are also effectively handled by the proposed method. The performances of the proposed method are evaluated, with the parameters of precision P, recall R, and F-measure F_α which are considered as first test criteria. These parameters are defined as:

$$P = \frac{\sum_x \sum_y (g(x,y) \times s(x,y))}{\sum_x \sum_y s(x,y)} \qquad (21)$$

$$R = \frac{\sum_x \sum_y (g(x,y) \times s(x,y))}{\sum_x \sum_y g(x,y)} \qquad (22)$$

$$F_\alpha = \frac{(1+\alpha) \times P \times R}{\alpha \times P + R} \qquad (23)$$

where $g(x,y)$ is the ground truth, $s(x,y)$ is the saliency map of the proposed method. α is taken as 0.3 in this work [3]. The performance metrics are calculated for proposed and state-of-the-art methods. The accurate assignment of salient pixels is highlighted by precision P. Recall R relates the correct detection of salient
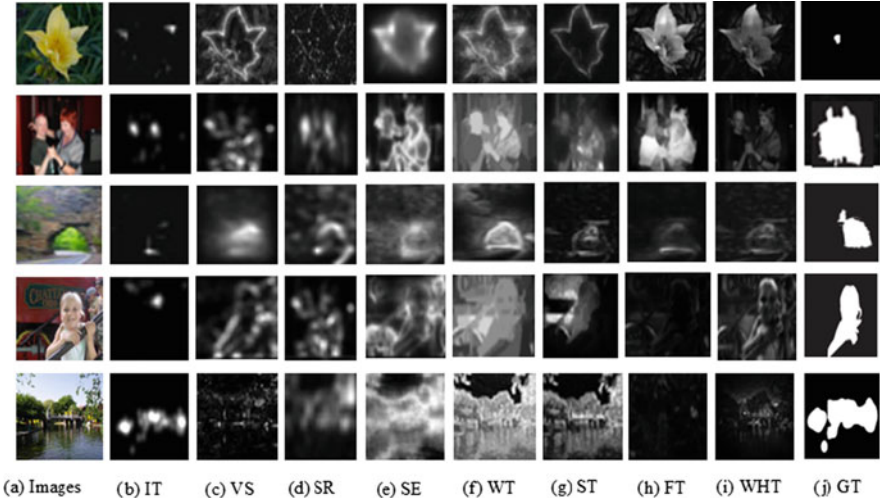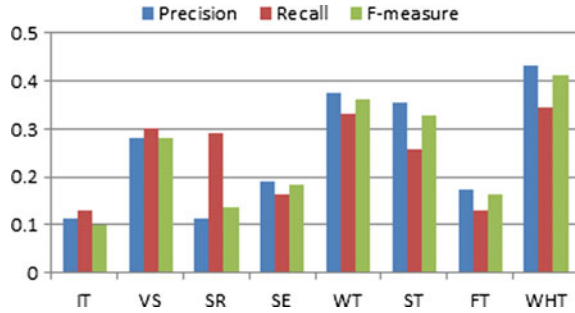
Fig. 3 Examples of saliency maps over MIT-1003 dataset **a** input images (image 1–image 5) **i** proposed WHT method **j** ground truths and **b–h** saliency maps of other existing methods

Fig. 4 Precision, recall, *F*-measure of the proposed WHT method and state-of-the-art methods on MIT-1003 dataset

regions in accordance with ground truth. *F*-measure F_α is the harmonic mean of *P* and *R*. Binary images of the saliency map are obtained by considering the mean value of saliency map and Otsu automatic threshold algorithm [42].

Each saliency map is the combination of salient regions and non-salient regions. Imamoglu et al. [3] explained true-positive rate (TPR) is the amount of salient pixels in the ground truth $g(x, y)$ intersecting with the amount of salient pixels in the saliency map $s(x, y)$. False-positive rate (FPR) is the amount of non-salient pixels from ground truth $g(x, y)$ intersecting with the salient pixels in the saliency map $s(x, y)$. ROC curve is the plot which is drawn between TPR and FPR. The relationship between the false positives and false negatives is linked in each point of ROC curve. The ROC curves and PR charts are given in Figs. 4 and 5, respectively. From Figs. 4 and 5, it is understood that proposed method shows higher performance with respect to the state-of-the-art methods.

Fig. 5 ROC of the proposed method and state-of-the-art method on MIT-1003 dataset
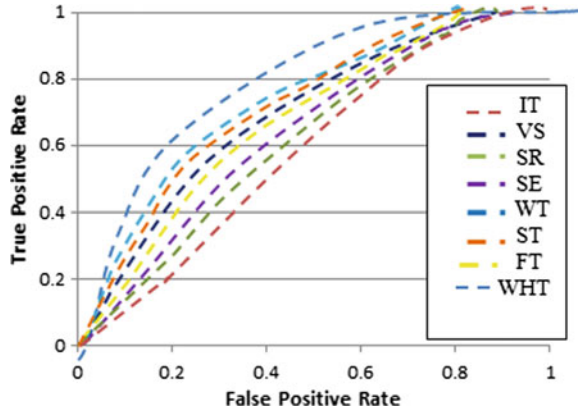

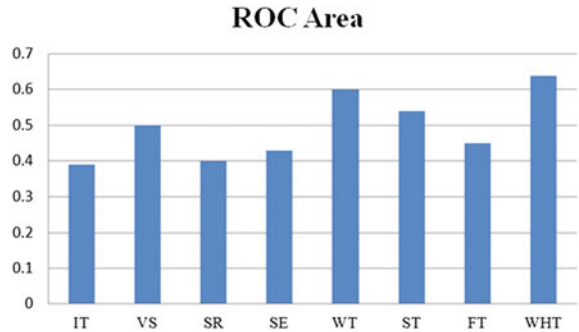**Fig. 6** AUC of the proposed WHT method and other existing methods

Figure 6 shows the ROC area of MIT benchmark datasets wherein the proposed method outperforms the state-of-the-art methods.

4.2 Results of Visual Saliency-Based Image Compression

The evaluation of visual saliency-based compression is performed for MIT benchmark dataset. Section 4.1 clearly shows the performance analysis of saliency detection of MIT dataset. To evaluate the consistency of the proposed method, the major performance metrics of peak signal-to-noise ratio (PSNR), compression ratio (CR), structural similarity index (SSIM), and mean square error (MSE) are obtained. The efficiency of the proposed method is assessed with certain state-of-the-art methods. These methods are performed to find image compression based on saliency analysis and DCT, DWT transforms. The methods are abbreviated as SM_DCT [32] and SM_DWT [23]. The proposed method is also compared

with JPEG [43] and DWT [44] methods. The proposed method is named as SM_WHT.

The input images (image 1–image 5) of Fig. 3 are used for the evaluation. The experiment is performed on a machine with Intel i3 2.4 GHz CPU and 4 GB RAM. The proposed method and all the comparative methods are implemented in Matlab R2012b. The average running time taken to process 400×300 image is 0.31 s. The image quality of the proposed method and state-of-the-art methods are analyzed by the performance metric of PSNR and SSIM. The compression ratio is achieved through the performance of metric of Bpp.

4.3 Performance Analysis

All the compression methods are having the main objective to obtain best quality of images with lesser utilization of bit. PSNR is one of the important parameters to evaluate the quality of an image. If the PSNR value is high, then the image quality is good. If it is low, then the quality of an image is low. It is also a type of objective measurement which is based on the mean square error (MSE) and it is given by,

$$PSNR = 10 \times \log_{10}\left(\frac{255^2}{MSE}\right) \qquad (24)$$

The mean squared error of an image is defined as

$$MSE = \frac{1}{M \times N} \times \left[\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}\left(A_{ij} - B_{ij}\right)^2\right] \qquad (25)$$

where $M \times N$ denotes the size of the input image. A_{ij} is the input original image, and B_{ij} is the compressed image. The images of MIT dataset are resized as 200×200. The proposed method is experimented on different images of MIT datasets. The image quality of the proposed method is obtained by the PSNR which is shown in Fig. 7.

SSIM is another performance metric which is also used to evaluate the image quality. It is used to highlight the structural similarity of input and output images based on the human visual system characteristics. It quantifies the structural similarity other than error visibility of two images. SSIM gives better interpretation of image quality compared to the PSNR metric. It is defined as

$$SSIM(u, v) = \frac{(2m_u m_v + d_1)(2\sigma_{uv} + d_2)}{\left(m_u^2 + m_v^2 + d_1\right)\left(\sigma_u^2 + \sigma_v^2 + d_2\right)} \qquad (26)$$
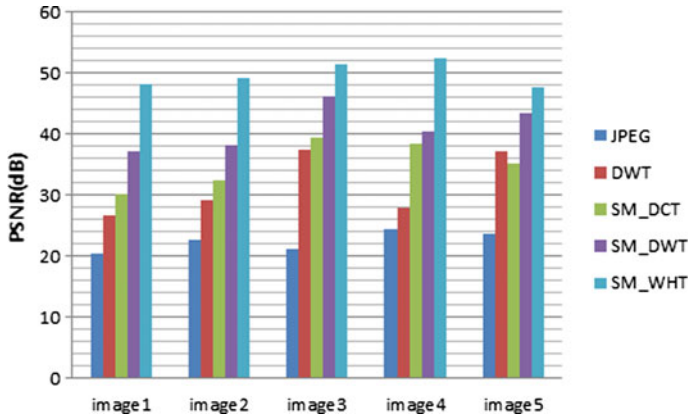
Fig. 7 PSNR (dB) of proposed and existing compression methods

where u and v are the block size of the images. $m_u m_v$ are the mean intensity values of u, v. σ_u^2, σ_v^2 are the standard deviation of u, v. d_1, d_2 are the constants.

From Fig. 8, it is understood that the proposed saliency-based WHT compression method produces better results than existing state-of-the-art methods. It is reaching nearly one. The proposed method uses the saliency-based quantitation table for image compression.

The compression ratio is the important performance metric in the scenario of image compression. If the compression ratio is higher, then a lot of memory is needed to store the data. If it is less, then lesser memory is only needed to store the output image. The compression ratio is defined as the total number of bits in the input image to the compressed image. It is used to estimate how much amount of compression has taken place. In the image, some regions are most important and others are less important. The most important regions are called salient regions. The compression ratio of salient regions is lesser compared to other regions. Whereas other regions are highly compressed, the compression ratios obtained from proposed method and state-of-the-art methods are updated in Table 1.

$$\text{Compression Ratio (CR)} = \frac{\text{No of bits in the compressed image}}{\text{No of bits in the original image}} \quad (27)$$

Table 1 clearly shows the compression ratios of the proposed method and state-of-the-art methods. The average compression ratio of the proposed method is 90%. The proposed SM_WHT image compression method extremely performs well compared to the saliency-based DWT and DCT methods. It is mainly due to the lesser number of computations in WHT.

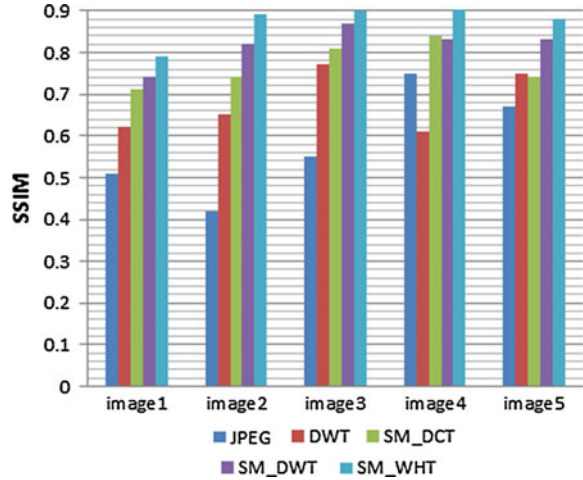**Fig. 8** SSIM of proposed and existing compression methods

Table 1 Compression ratio of the proposed method and state-of–the-art methods

Images	Compression methods				
	JPEG	DWT	SM_DCT	SM_DWT	SM_WHT
Image 1	0.75	0.77	0.82	0.83	0.88
Image 2	0.79	0.84	0.89	0.88	0.91
Image 3	0.84	0.86	0.88	0.90	0.92
Image 4	0.77	0.88	0.86	0.92	0.90
Image 5	0.80	0.84	0.90	0.82	0.89

5 Conclusion

Visual saliency-based image compression is exhausted in this chapter. Initially, the most important regions of an image are highlighted by the saliency map. Then the saliency-based quantitation table is used in the compression method. The performance metric of PSNR, SSIM, and compression ratio is obtained in order to estimate the performance of the proposed method. The proposed visual saliency-based compression method yields the average compression ratio of 90% compared to the state-of-the-art compression methods. The redundant pixels are captured by the WHT with the assistance of second-order correlation. Lesser computations in the WHT transform are the most added advantage in the proposed compression technique. The proposed method can be extended by considering the motion images as the inputs. The hardware implementation of the proposed method can be performed as the future work.

References

1. Guo C, Zhang L (2010) A novel multi resolution spatiotemporal saliency detection model and its applications in image and video compression. IEEE Trans Image Process 19, 185–198
2. Arya R, Singh N, Agrawal RK (2015) A novel hybrid approach for salient object detection using local and global saliency in frequency domain. Multimed Tools Appl. doi:10.1007/s11042-015-2750-y
3. Imamoglu N, Lin WS, Fang YM (2013) A saliency detection model using low-level features based on wavelet transform. IEEE Trans Multimedia 15:96–105
4. Lin RJ, Lin WS (2014) Computational visual saliency model based on statistics and machine learning. J Vision 14(9), 1–18
5. Itti L, Koch C, Niebur E (1998) A model of saliency-based visual attention for rapid scene analysis. IEEE Trans Pattern Anal Mach Intell 20:1254–1259
6. Ma YF, Zhang HJ (2003) Contrast-based image attention analysis by using fuzzy growing. In: ACM International conference on multimedia, pp 374–381, Berkeley (2003)
7. Harel J, Koch C, Perona P (2006) Graph-based visual saliency. Neural Inf Process Syst 545–552
8. Goferman S, Zelnik Manor L, Tal A Context-aware saliency detection. In: IEEE conference on computer vision and pattern recognition (CVPR), pp 2376–2383
9. Achanta R, Hemami S, Estrada F, Susstrunk (2009) Frequency-tuned salient region detection. IEEE conference on computer vision and pattern recognition, pp 1597–1604
10. Cheng MM, Zhang GX, Mitra NJ, Huang X, Hu SM (2015) Global contrast based salient region detection. IEEE Trans Pattern Anal Mach Intell 37(3):569–582
11. Perazzi F, Krahenbuhl P, Pritch Y, Hornung A (2012) Saliency filters: contrast based filtering for salient region detection. In IEEE conference on computer vision and pattern recognition (CVPR), pp 733–740
12. Guo C, Ma Q, Zhang L (2008) Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform. In: IEEE conference on computer vision and pattern recognition (CVPR), pp 1–8
13. Li J, Levine MD, An X, Xu X, He H (2013) Visual saliency based on scale-space analysis in the frequency domain. IEEE Trans Pattern Anal Mach Intell 35:996–1010
14. Hou X, Zhang L (2007) Saliency detection: a spectral residual approach. In: IEEE conference on computer vision and pattern recognition (CVPR), pp. 1–8
15. Merry RJE (2005) Wavelet theory and application-a literature study. Eindhoven University of Technology, The Netherlands
16. Li ZQ, Fang T, Huo H (2010) A saliency model based on wavelet transform and visual attention. Sci China Inf Sci 53(4):738–751
17. Tian Q, Sebe N, Lew MS, Loupias E, Huang TS (2001) Image retrieval using wavelet-based salient points. Electron Imag 10(4):835–849
18. Candes E, Donoho D (2004) New tight frames of curvelets and optimal representations of objects with piecewise singularities. Commun Pure Appl Math 57:219–266
19. Bao L, Lu J, Li Y, Shi Y (2014) A saliency detection model using shearlet transform. Multimed Tools Appl. doi:10.1007/s11042-014-2043-x(2014)
20. Yu Y, Yang J (2016) Visual saliency using binary spectrum of Walsh–Hadamard transform and its applications to ship detection in multispectral imagery. Neural Process Lett. doi:10.1007/s11063-016-9507-0
21. Guo C, Zhang L (2010) A novel multi resolution spatiotemporal saliency detection model and its applications in image and video compression. IEEE Trans Image Process 19(1):185–198
22. Hadizadeh H, Bajic I (2014) Saliency-aware video compression. IEEE Trans Image Process 23(1):19–33
23. Barua S, Mitra K, Veeraraghavan A (2015) Saliency guided Wavelet compression for low-bitrate Image and Video coding. In: IEEE global conference on signal and information processing (2015)

24. Ouerhani N, Bracamonte J, Hugli H, Ansorge M, Pellandini F (2001) Adaptive color image compression based on visual attention. International conference on image analysis and processing, pp 26–28
25. Li Z, Qin S, Itti L (2011) Visual attention guided bit allocation in video compression. Image Vision Comp 29(1):1–14
26. Hadizadeh H (2013) visual saliency in video compression and transmission. Thesis dissertation (2013)
27. Ho-Phuoc T, Dupret A, Alacoque L (2012) Saliency-based data compression for image sensors. Sensors, IEEE
28. Zundy F, Pritch Y, Sorkine-Hornung A, Mangold S, Gross T (2013) Content-aware compression using saliency-driven image retargeting. In: 20th IEEE international conference on image processing (ICIP) (2013)
29. Tu Q, Mena A, Jiang Z, Ye F, Xu J (2015) Video saliency detection incorporating temporal information in compressed domain. Signal Process: Image Commun 38, 32–44
30. Yu SX, Lisin DA (2009) Image compression based on visual saliency at individual scales. In International symposium on visual computing, USA
31. Dhavale N, Itti L (2003) Saliency-based multi-foveated MPEG compression. In: Proceedings of signal processing and its applications pp 229–232 (2003)
32. Duan L, Ke C (2012) A natural image compression approach based on independent component analysis and visual saliency detection. Adv Sci Lett 5, 1–4
33. Lakshmi Priya GG, Domnic S (2014) Walsh-Hadamard transform kernel-based feature vector for shot boundary detection. IEEE Trans Image Process 23(12):5187–5197
34. Petrov Y, Li Z (2003) Local correlations, information redundancy, and sufficient pixel depth in natural images. J Opt Soc Am A 20(1):56–66
35. Li Z, Atick JJ (1994) Toward a theory of the striate cortex. Neural Comput 6(1):127–146
36. Luo W, Li H, Liu G, Ngan KN (2012) Global salient information maximization for saliency detection. Sig Process: Image Commun 27(3):238–248
37. Ma X, Xie X, Lam K-M, Zhong Y (2015) Efficient saliency analysis based on wavelet transform and entropy theory. J Vis Commun Image R 30:201–207
38. Acharya T, Tsai PS (2005) JPEG2000 standard for image compression: concepts, algorithms and VLSI architecture. Wiley 60(2005)
39. Gupta R, Khanna MT, Chaudhur S (2013) Visual saliency guided video compression algorithm. Signal Process: Image Commun 28:1006–1022
40. Judd T, Ehinger K, Durand F, Torralba A (2009) Learning to predict where humans look. In: IEEE conference on computer vision and pattern recognition (CVPR)
41. Murray N, Vanrell M, Otazu X, Parraga CA (2011) Saliency estimation using a non-parametric low-level vision model. In: IEEE conference on computer vision and pattern recognition (CVPR), pp 433–440
42. Gonzalez RC, Woods RE, Eddins SL (2004) Digital signal procesing using Matlab. Prentice Hall, Englewood Cliffs, NJ
43. Wallace GK (1991) The JPEG still picture compression standard. Commun ACM 34:31–44
44. Chowdhury MMH, Khatun A (2012) Image compression using discrete wavelet transform. Int J Comput Sci 9(4)