# 1
# Introduction

In the year 2016, global internet traffic reached the rate of 1.1 zettabytes (nearly 12 billion gigabytes) per year, and by the end of the year 2021, it is projected to increase about threefold, i.e. 3.3 zettabytes (nearly 36 billion gigabytes) per year [networking Index, 2016]. Globally, image and video traffic over the internet are projected to be 82% of all overall Internet traffic by 2021, rising from 73% in 2016. Image and video coding standards developed long back had been doing the needful successfully. However, due to the growing internet traffic, and to fulfil the expected quality of experience (QoE), there is a demand to improve compression coding algorithms.

To validate the performance of any new compression algorithm, it needs to be compared with the existing one with respect to the quality of the reconstructed image at the same bitrate. Various image quality assessment (IQA) methods have been developed for such purposes. An efficient IQA method is expected to provide a quantitative measure that is consistent with the human visual system (HVS). IQA techniques are classified into three categories subjected to the availability of the reference image at the receiver end. These categories are: $(i)$ Full-Reference method (FR-IQA), in case the reference image is available, $(ii)$ Reduced-Reference method (RR-IQA), in case only some partial information about the reference-image is known, and $(iii)$ No-Reference method (NR-IQA) or blind IQA, when no information of reference-image is present at the receiver side.

In the last few years, there is high demand for a new set of images called screen-content images (SCI). The images are used to share the live screen of the display unit to the end user who is present at a remote location. The properties of SCIs are different from the camera-content images (CCIs) and will be discussed later in this chapter. SCIs include text, graphics, and pictures together are used in multi-client communication systems, such as video conferencing, virtual screen sharing, mobile gaming, information sharing between computers and smart-phones, cloud computing, remote education, etc. Due to the distinguishing properties of SCIs from CCIs, many advanced coding techniques for CCIs have extended the codec for SCIs [Xu *et al.*, 2016]. Moreover, it has been observed that even the top IQA methods for CCIs failed to perform satisfactorily on SCIs. This motivated many research groups to work on the development of the IQA methods for SCIs in order to fulfil the expected quality-of-experience (QoE).

This chapter introduces the area of Image compression and its quality assessment. It also defines the problem statement and discuses the motivation and literature survey for this thesis. It first introduces the basics of image compression and image quality assessment, and then the state-of-the-art methods in each context are studied and the motivation for the thesis is explained. Finally, the thesis overview and contributions are discussed.

## 1.1 MOTIVATION AND LITERATURE SURVEY

This section is organized as follows. Section 1.1.1 discusses about the CCI compression. It also explains some state-of-the-art lossy compression techniques and explains the need for saliency enabled image compression. Section 1.1.2 introduces a new set of images called screen-content

images (SCIs). It explains the distinguishable properties of SCIs compared to CCIs and the need for a separate compression structure for SCIs. Finally, Section 1.1.3 reviews the state-of-the-art quality assessment methods for CCIs and SCIs. This section also provide motivation towards designing a reduced-reference SCI quality assessment method which can perform adequately under the distortion caused by compression methods.

### 1.1.1 Camera Content Image Compression

Usage of image data through the Internet has exponentially increased among the users [Sayood, 2012]. Compression is essentially required to manage this high data-rate of images without degrading the quality to an unacceptable level. The necessity of accessing high definition images with quality as of paramount importance has become the major issue in designing such algorithms to operate in real time. Image compression can broadly be categorized into two categories: lossy and lossless. Lossless image compression mainly focuses on the identification and removal of redundancy that can be recovered on the decoder side [Weinberger *et al.*, 2000]. It is preferred for archival purposes and often for medical imaging. Lossy compression techniques usually remove the irrelevant information from the image space. Such methods are suitable for camera-content or natural images like photographs, where a small loss in fidelity (sometimes perceptually lossless) is acceptable to achieve a high compression-ratio or low bitrate. The lossy compression which yields imperceptible differences is also called visually lossless [Wang *et al.*, 2015]. The lossy image compression methods can be further enumerated into two categories:

The methods that fall into the first category are called direct methods [Sayood, 2012], which act directly on the image samples in the spatial domain. Block truncation coding (BTC) [Kurita and Otsu, 1993; Yang *et al.*, 1994; Dhara and Chanda, 2007] and vector quantization [Feng and Nasrabadi, 1991; Lee and Chan, 1994] based methods are widely used under this category. BTC based approach uses a one-bit adaptive non-parametric quantizer over the local regions of an image. For this, the image is first divided into small non-overlapping blocks, and then the first and second moments of each block are evaluated as $\bar{X}$ and $\bar{\sigma}$, respectively. A bit plane is then constructed by assigning each pixel location as 0 or 1 depending on whether the corresponding pixel is smaller than the first moment $(\bar{X})$. The receiver receives the bit plane, $\bar{X}$, and $\bar{\sigma}$ and reconstructs the image blocks by preserving the $\bar{X}$, and $\bar{\sigma}$.

The optimal vector quantization from empirical data was first proposed in [Linde *et al.*, 1980] and then extended in [Lloyd, 1982] using K-means clustering algorithm which is also referred to as LBG or generalized Lloyd algorithm [Hartigan, 1975]. The basic idea of vector quantization is to map an input data $(X)$ of $K$-dimensional Euclidean space $(R^K)$ into a reduced finite subset $Y$ of $R^K$. For colour images, [Feng and Nasrabadi, 1991] exploited the inter block and inter-colour correlations with vector quantization in order to reduce the bit-rate.

The methods under the second category are called transform methods [Sayood, 2012], where the image is transformed into frequency domain. Principle Component Analysis (PCA) [Abadpour and Kasaei, 2008], Discrete Cosine Transform (DCT) [Wallace, 1992; Douak *et al.*, 2011; Dhara and Chanda, 2007; Messaoudi and Srairi, 2016], Discrete Wavelet Transform (DWT) [Skodras *et al.*, 2001; Boucetta and Melkemi, 2012; Barua *et al.*, 2015] are the most popular transformations used for this purpose. Such transformations concentrate the energy of the image in a few number of coefficients, making it suitable for removing perceptual redundancies.

Among all transform based image compression methods, DWT achieves the best energy compaction. The comparative study between DCT and DWT-based image and video coding techniques [Xiong *et al.*, 1999] suggests that DWT based methods yield slightly better quality of reconstructed image after compression i.e. higher peak-signal-to-noise ratio (PSNR) value (< 0.7db) compared to DCT, for the same compression ratio. However, the DCT based coder has significant

lower complexity than its DWT counterpart. For this reason, the state-of-art image and video coding standards, and multimedia devices prefer DCT over DWT [Sullivan *et al.*, 2012].

Due to the computationally efficient encoding and decoding structures, JPEG baseline [Wallace, 1992] is a well-accepted standard used for lossy image compression. It is widely used in digital cameras and other photographic image capturing devices for storing and transmitting images on World Wide Web (WWW). During the encoding process, JPEG divides the image into blocks of size $8 \times 8$ and applies 2D-DCT on each of them. To achieve compression, the DCT coefficients of the blocks are quantized by fixed quantization parameters, irrespective of the Region-of-Interest (RoI) of the block. The JPEG encoding framework is disucssed in detail in section 1.2.1. In the case of high compression requirements, scaled up quantization parameters are used with the DCT coefficients. We observed that the reconstructed blocks with higher variance (i.e., the second central moment of the pixel intensities within a block) are highly degraded, compared to the blocks with lower variance [Yang *et al.*, 2016; Lam and Goodman, 2000]. This happens due to the structure of JPEG baseline quantization table designed to quantize higher frequency regions more heavily than the lower frequency ones. A non-homogeneous compression is observed causing compression artifacts in the reconstructed image. RoI independent quantization of DCT coefficients causes degradation of the overall perceptual quality of the reconstructed image particularly at high compression-ratio (CR) (i.e. at low bit rate).

There are some regions in images where more information is anticipated which can be delineated as ROIs or salient regions. The amount of attention steered among all the regions of an image is non-identical as per the human visual and cognitive systems [Borji and Itti, 2013]. A saliency-guided compression method is ideally suited to preserve perceptually important regions. These methods can intelligently compress the salient regions lightly and non-salient ones heavily, to ensure as small perceptual loss as possible, for required CR.

Efforts have been made towards saliency-based image and video compression techniques [Xia *et al.*, 2012; Yang *et al.*, 2005; Hadizadeh and Bajic, 2014; Guo and Zhang, 2010; Christopoulos *et al.*, 2000; Barua *et al.*, 2015]. Mostly, these approaches segment the image into two regions [Hadizadeh and Bajic, 2014; Guo and Zhang, 2010; Christopoulos *et al.*, 2000; Barua *et al.*, 2015]: salient and non-salient. On the segmented regions, different compression algorithms are applied in order to obtain a good combination of reconstructed image quality and CR. However, our cognitive system does not always classify images into salient and non-salient. Human visual system (HVS) bestows multi-level attention on different regions. This leads to the requirement of segmentation of images into saliency driven multiple regions.

The work done in multi-level saliency-based compression techniques [Christopoulos *et al.*, 2000; Barua *et al.*, 2015; Xia *et al.*, 2012; Bruckstein *et al.*, 2003] exhibits an improved trade-off between CR and perceptual quality than using only two-level saliency. Texture-based methods [Xia *et al.*, 2012; Bruckstein *et al.*, 2003] classifies the image into edges, textures, and flat regions with the aim to save the edge, and important texture information of the image after compression. JPEG 2000 standard [Christopoulos *et al.*, 2000], incorporates both two-level and multi-level models of ROI encoding using maximum shift or MAXSHIFT and general scaling based method (GSBM), respectively. The major challenge in multi-level saliency-based compression technique is the requirement of sending the overhead for the shape of salient regions and their ranks used to grade the saliency. The overhead is proportional to the complexity of the shape and the number of ranks. This is because the complex shaped regions will require a larger number of model parameters which will increase the overhead. Also, the overhead for ranks information will increase with an increase in the number of salient regions i.e. for the $R$ number of ranks, $\lceil log_2 R \rceil$ bits- per-rank will be needed. If the RoI mask is generated for an arbitrary shaped RoI, the decoder needs to reproduce the RoI mask [Christopoulos *et al.*, 2000], making the decoder computationally complex and

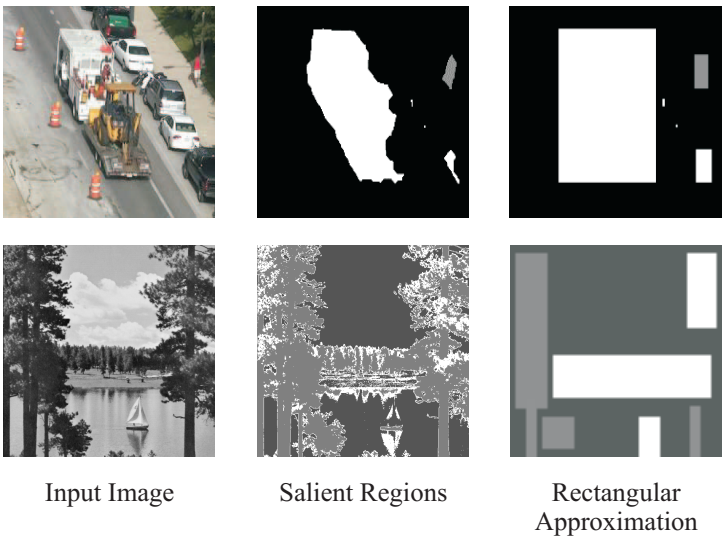



**Figure 1.1 :** Rectangular approximation in multi-level saliency based compression techniques

increased memory requirement on the decoder side. To reduce the requirement of the overhead information and to make the decoder simple, the RoI shape is approximated as a rectangular box [Christopoulos *et al.*, 2000; Barua *et al.*, 2015], as shown in Figure 1.1. The coordinates of the opposite vertices of the rectangular boxes and their rank information are sent to the decoder. This approach saves the overhead information to a good extent but the CR is compromised, as the actual RoI has been approximated by a rectangular bounding box.

Due to high computational efficiency of DCT based methods [Xiong *et al.*, 1999], with a small loss in performance as compared to DWT based such methods, our focus has been on the former one. Our efforts were directed towards developing a saliency enabled image compression framework which can easily be plugged in with the JPEG baseline in order to upgrade its performance. In order to achieve this, we developed a multi-level saliency enabled compression method which is discussed in Chapter 2.

### 1.1.2 Screen Content Image Compression

Screen-content-images (SCIs) include the combination of text, graphics, and pictures, and are used in multi-client communication systems, such as video conferencing, virtual screen sharing, mobile gaming, information sharing between computers and smart-phones, cloud computing, remote education, etc [Shen *et al.*, 2009; Lu *et al.*, 2011; Chang and Li, 2011]. Figure 1.2 shows some example of SCIs taken from two publicly available datasets [Yang *et al.*, 2015; Wang *et al.*, 2016]. The demand for graphically rich services has increased where the display of a remote system is shared for the purpose of accessing the computational resource or remote data via the Internet. Recently, there has been a noticeable growth in number of applications which display more than just camera-content images.

Such applications include displays that combine camera-captured and computer graphics, tablets, automotive displays, wireless displays, screen sharing, and so on [Yu *et al.*, 2014]. The major difference between camera-captured content or natural image and SCI is that SCI contains no sensor noise. Due to this difference SCIs have large uniform regions, highly saturated or a limited number of different colors, repeated patterns, and numerically similar blocks or regions among a sequence of images. SCIs also contain dominant textual information as graphs, codes, or

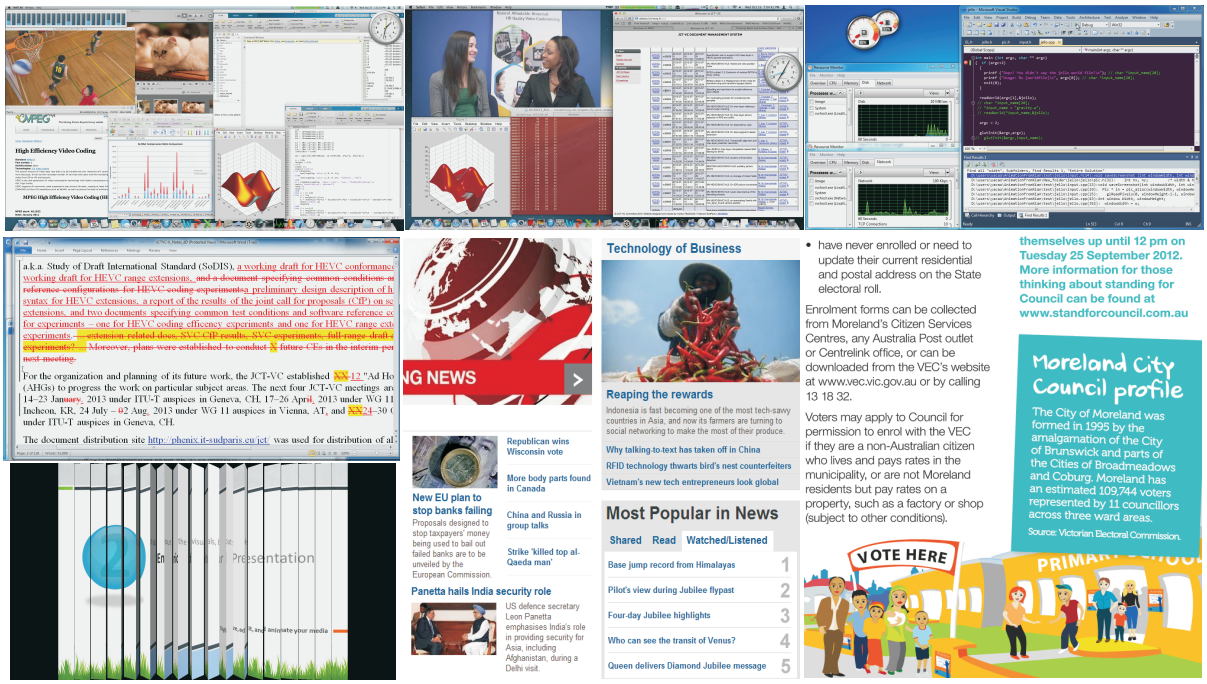



**Figure 1.2 :** Few examples of Screen-Content Images (SCIs)

descriptions.

Over the past few years, the usage of SCIs through the Internet has exponentially increased, especially for sharing screens over applications such as Teamviewer, Zoom, Skype, etc. Efficient compression is essentially required, to manage high data-rate of images without degrading the quality to an unacceptable level. The distinguishing properties of SCIs from natural images has motivated many advanced coding techniques for a Screen-Content-Coding (SCC) extension [Yu *et al.*, 2014; Xu *et al.*, 2016]. The quality-of-experience (QoE) as of utmost significance has become the major issue in designing such algorithms to operate in real time. The QoE is majorly dependent on two important factors, such as visual quality of the reconstructed image and the CR. In order to achieve high CR, lossy compression methods are majorly used. Lossy compression techniques developed over the years, can be broadly classified into two categories as direct methods, and transform based methods as discussed in Section 1.1.1. It has been observed that the performance of transform based compression methods is superior than the direct methods due to its ability of efficient energy compaction. Due to this, the state-of-the-art image and video coding standards, and multimedia devices prefer DCT based coder over DWT based coder [Xu *et al.*, 2016; Sullivan *et al.*, 2012].

JPEG baseline [Wallace, 1992] is the most widely used DCT based compression technique for natural and screen-content images, due to its computationally efficient encoding and decoding structures. It is widely used in digital cameras and other multimedia devices for storing and transmitting natural or screen-content images over the internet. To achieve low bit-rate requirements, the visual quality of the blocks with higher frequency get more degraded after reconstruction compared to the blocks with lower frequency. These methods are based on the fact that the HVS are more sensitive towards any change in low frequency regions than in the high frequency ones.

The statistical features of CCIs are different from SCIs, especially in a region where text

informations are present [Yang *et al.*, 2015]. These statistical differences between CCIs and SCIs can be measured in terms of naturalness and frequency level. The expression to find out the naturalness value of an $m \times n$ resolution image $I(x,y)$ was provided by [Mittal *et al.*, 2013] and given in (1.1). The naturalness of an image $(N(x,y))$ is evaluated in terms of the ratio between the zero mean intensity values and the standard deviation. As the CCI's follow central limit theorem, their naturalness is expected to follow a Normal distribution.

$$N(x,y) = \frac{I(x,y) - \mu(x,y)}{\sigma(x,y) + 1} \tag{1.1}$$

where $\mu(x,y)$, $\sigma(x,y)$ are the local mean and deviation of the image $I(x,y)$ as given in (1.2), and (1.3) respectively.

$$\mu(x,y) = \sum_{i=-3}^{3} \sum_{j=-3}^{3} \omega_{3,3} I(x+i, y+j) \tag{1.2}$$

$$\sigma(x,y) = \sqrt{\sum_{i=-3}^{3} \sum_{j=-3}^{3} \omega_{3,3} [I(x+i, y+j) - \mu(x,y)]^2} \tag{1.3}$$

where $\omega_{3,3}$ is a 2D circularly-symmetric Gaussian weighting function. The naturalness value of the image is computed in terms of the distribution of the coefficients $N(x,y)$. Figure 1.3 (a), and (b) shows the naturalness distribution of CCIs, whereas, Figure 1.3 (c), and (d) shows the naturalness of the SCIs. It is observed that the CCIs follow a Gaussian distribution i.e. the naturalness of CCIs is high which is also demonstrated in [Mittal *et al.*, 2013]. On the other hand, the distribution of SCIs is different than the CCIs. A sharp spike in the distribution of SCIs proves their less naturalness compared to CCIs.

To further examine the properties of SCIs, block-wise activity was analyzed by using the reported image activity analysis method in [Yang *et al.*, 2012]. The activity analysis reflects the degree of pixel variation in the local region of the image. It has been observed that the activity measure of textural blocks is larger than those from the pictorial blocks which is also the reason for the spike in the distribution curve. This also proves that the textural regions of an SCI have more frequency than the non-textural regions.

The distinguishing properties of SCIs and CCIs as discussed above, is the main reason for a comparatively weak performance of compression algorithms such as JPEG or HEVC, which are specifically designed for CCIs, on SCIs. These DCT based methods are designed to quantize the high frequency regions more than that of the low frequency regions. As the textual regions in an SCI has high frequency, it gets more distorted and even illegible after applying such compression techniques.

As per the human visual and cognitive systems [Borji and Itti, 2013], the amount of attention steered among all the regions of an image is non-identical. A saliency-guided compression method is ideally suited to preserve perceptually important regions. Moreover, these methods are capable to compress the salient and non-salient regions judiciously in order to ensure smallest perceptual loss at a required CR. It has been observed that in case of SCIs, the textual region anticipate more

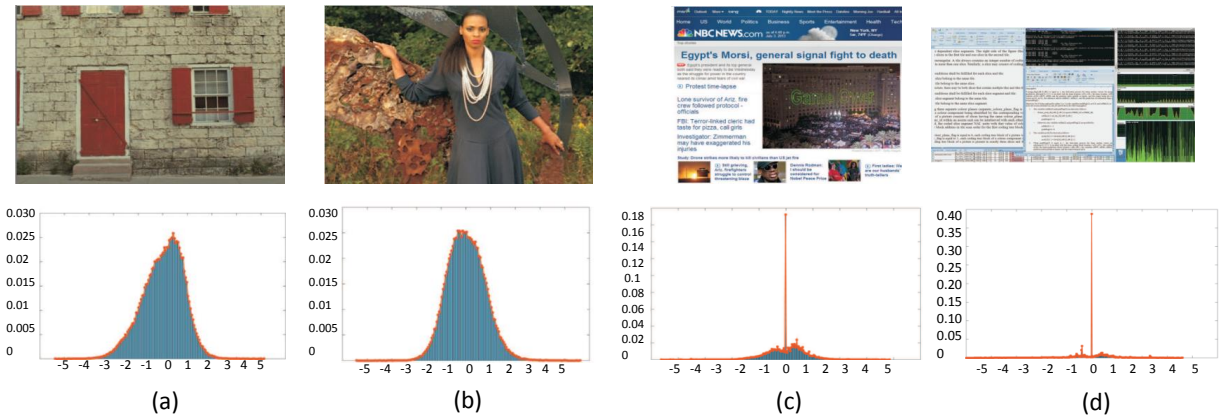



**Figure 1.3 :** Analysis of Naturalness $N(x,y)$ Histograms for: (a - b) Camera-content Images, (c - d) Screen-content Images.

information and can be delineated as ROI or salient region. The energy distribution of the text blocks among DCT coefficients are significantly higher compared to the non-text blocks and due to this the text regions tend to get highly distorted after applying DCT based codecs [Wallace, 1992; Xu *et al.*, 2016]. This exhibits the need to develop a saliency based compression method for SCIs.

Although, there was no work found on saliency based SCI compression, there have been efforts towards developing saliency-enabled compression techniques for natural or camera-content images [Guo and Zhang, 2010; Barua *et al.*, 2015; Kaur *et al.*, 2006; Christopoulos *et al.*, 2000; Rahul and Tiwari, 2018]. The methods in [Hadizadeh and Bajic, 2014; Christopoulos *et al.*, 2000] segment the image into two regions as salient and non-salient. A saliency map is then obtained based on the importance of the segemented regions. On the segmented regions, different compression algorithms are applied to obtain a good combination of reconstructed image quality and compression-ratio (CR).

We devised a two-level saliency-based SCI compression algorithm that provides an optimal trade-off between the overhead, perceptual quality, and CR. The developed method first classifies the SCI into salient and non-salient region by identifying the textual information. The aim is to enable the JPEG standard to judiciously retain high-frequency text regions in the image, particularly in the case of high compression requirement.

### 1.1.3 Camera and Screen Content Image Quality Assessment

Image Quality Assessment (IQA) is a method to evaluate the quality of an acquired image or a processed image obtained through systems such as image acquisition, de-noising, de-blurring, enhancement, transmission, reconstruction, and compression. An efficient IQA technique is expected to provide a objective quantitative measure that is consistent with the subjective human observations. The meaning of consistent here is that the objective evaluation of the algorithm should be within close agreement with HVS, irrespective of the type, content or strength of the distortion present in the image. IQA techniques can broadly be classified into three categories as $(i)$ Full-Reference method (FR-IQA) in case the reference image is available, $(ii)$ Reduced-Reference method (RR-IQA) in case the availability of the partial information about the reference image, $(iii)$ No-Reference method (NR-IQA). There are number of researchers who have contributed significantly towards designing image quality assessment algorithms over the years.

In recent years, the demand for graphically rich services has exponentially increased where the display of a remote system is shared in order to access the remote computational resources or data via the world-wide-web (WWW). Screen-content-images (SCIs) includes text, graphics, and pictures together and are used in multi-client communication systems, such as video conferencing, virtual screen sharing, mobile gaming, information sharing between computers and smartphones, cloud computing, remote education, etc [Shen *et al.*, 2009]. Unlike CCIs, SCIs do not contain any sensor noise, and such images mostly have large uniform regions, patterns in a repeated form, mostly saturated or a few number of different colors, and contain numerically similar blocks or regions among a sequence of images.

The distinguishing property of SCIs from CCIs as discussed in Section 1.1.2, has motivated the High-Efficiency-Video-Coding (HEVC) [Xu *et al.*, 2016] research group to include screen-content-coding (SCC) extension into their current video compression framework. This in fact motivated many researchers for Quality assessment for SCIs through systems such as transmission, reconstruction, and compression to fulfil the expected quality-of-experience (QoE) of the interactive remote system. It has been observed that even the top IQA methods for CCIs failed to perform satisfactorily on SCIs. Due to this, many IQA techniques have been proposed over past three years which are specifically designed for SCIs.

Full access of reference image is required in case of FR-IQA. The traditional FR-IQA, such as peak-signal-to-noise-ratio (PSNR) [Wang and Bovik, 2002] and the mean-squared error (MSE) provide the intensity difference between the reference image and the processed image. These index are computed by taking the average of the squared intensity differences between the corresponding pixels of distorted and reference image. These methods are simple in order to evaluate, have proper physical meanings, and are mathematically convenient to implement. However, such methods lack the correlation with the subjective visual quality. The widely used FR-IQA techniques for CCIs are visual signal-to-noise ratio (VSNR) [Chandler and Hemami, 2007], local-tuned-global (LTG) model [Gu *et al.*, 2014], structural similarity (SSIM) index [Wang *et al.*, 2004], gradient-similarity (GSIM) [Liu *et al.*, 2012], visual saliency index (VSI) [Zhang *et al.*, 2014], and feature-similarity (FSIM) index [Zhang *et al.*, 2011].

VSNR [Chandler and Hemami, 2007] is a wavelet-based IQA method which aims to quantify the perceptual loss of the distorted images based on the psychophysical findings as presented by [Ramos and Hemami, 2001; Chandler and Hemami, 2003a,b; Chandler *et al.*, 2006]. Later, it was observed that the HVS evaluates the overall distortion of a scene by combining the sensations of salient local artefacts and global degradation. To exploit this, LTG [Gu *et al.*, 2014] was proposed to first evaluate the gradient magnitude (GM) of an image. It then independently uses the global and local mean pooling to evaluate the HVS perception.

In SSIM [Wang *et al.*, 2004], it was observed that the HVS is more sensitive towards change in structure, luminance, and contrast of an image than the change in intensity. SSIM proposed a similarity index value using the weighted sum of the nomalized distortion in luminance, contrast, and structure. The weights of these three components are adjusted as per their relative importance. The framework of SSIM is shown in Figure 1.4.

GSIM [Liu *et al.*, 2012] was proposed later on the similar lines of SSIM [Wang *et al.*, 2004]. The major contribution in GSIM was that it used gradient information in order to capture the contrast and structure of the image which led to a more accurate quality assessment of the image. GSIM also propose an adaptive approach to integrate the gradient or contrast-structure and luminance components. Figure 1.5 shows the GSIM process.

The recent studies in the area of functional magnetic resonance imaging (fMRI) in neurobiology [Henriksson *et al.*, 2009] suggest that the HVS is more sensitive towards changes

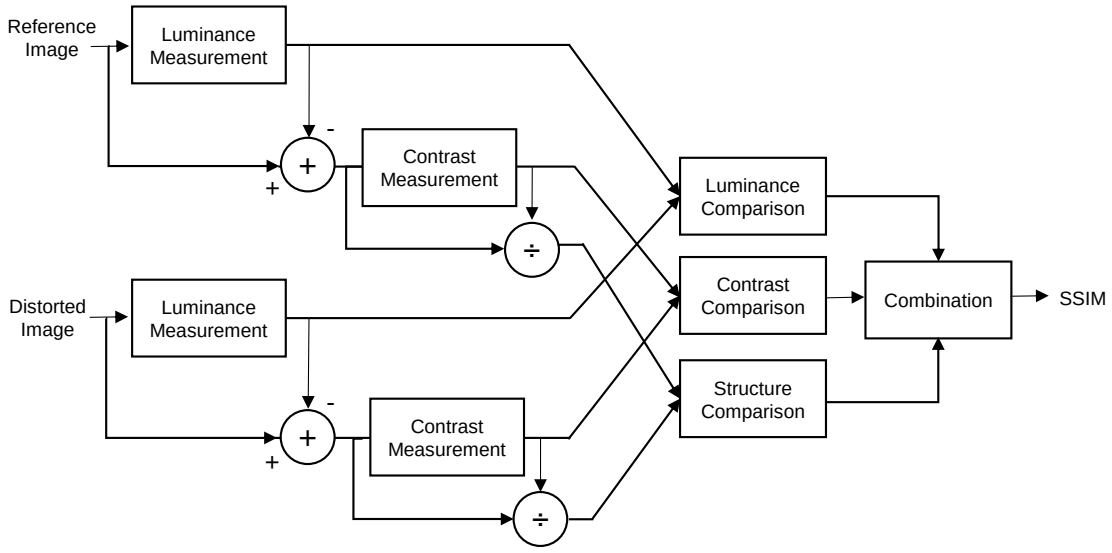



**Figure 1.4 :** Structural similarity index (SSIM) framework

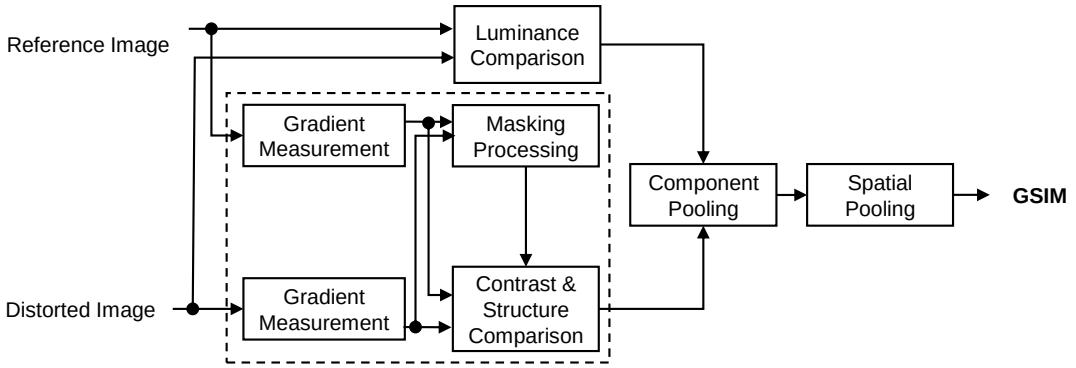



**Figure 1.5 :** Gradient similarity index (GSIM) framework

in the low-level feature of an image like edges and zero-crossings. In simplified words, our brain stores the information of a scene through the salient low-level features. FSIM [Zhang *et al.*, 2011] used this information to devise an IQA metric by comparing the low-level feature sets between the reference and distorted image. In order to efficiently extract the low-level feature, FSIM extracts the feature at points where high phase congruency (PC) is present. This is due to the fact that the visually distinctive features coincide with such points, where congruent phases are present in the Fourier waves at different frequencies [Concetta Morrone and Burr, 1988; Morrone *et al.*, 1986;

Morrone and Owens, 1987; Kovesi, 1999]. Moreover, taking into account that the PC is considered contrast invariant, and HVS is contrast-change sensitive, the image gradient magnitude (GM) is evaluated as the secondary feature in order to encode the contrast information. PC and GM are reciprocal and they reflect distinctive parts of the HVS, and produces a local similarity map of an image. In order to derive a single quality score, PC is utilized again after the evaluation of the local similarity map.

VSI [Zhang *et al.*, 2014] was proposed on the similar lines of FSIM [Zhang *et al.*, 2011]. The study on the relationship between visual saliency (VS) and HVS suggests that a VS based IQA model can achieve better prediction performance than the other feature-based methods [Zhang *et al.*, 2014]. VSI [Zhang *et al.*, 2014] analyzes the changes in the visual saliency map of the distorted image with respect to the reference image. Basically, VSI uses the saliency map of an image as a feature map in order to characterize the local region quality of an image.

The IQA methods which have been discussed above are specifically designed for CCIs. Due to the distinguishing properties of SCI, such methods are unable to perform satisfactory on these images. This led many researchers to start working on designing a SCI-IQA metric. Over the past few years there have been much works done towards the development of FR-IQA for SCI's [Gu *et al.*, 2016b, 2018; Yang *et al.*, 2015; Ni *et al.*, 2016].

Yang *et al.* [2015] was the first attempt towards in-depth study for object and subjective quality assessment of SCIs. They proposed a large scale SCI dataset named screen content quality assessment database (SIQAD). They also proposed a SCI-IQA method known as SCI Perceptual Quality Assessment (SPQA) in order to evaluate the visual quality of distorted SCIs. SPQA was proposed to handle the textual and pictorial regions in the SCI separately as the same distortion in different regions may have different visual perception according to the HVS. The framework of SPQA is shown in Figure 1.6

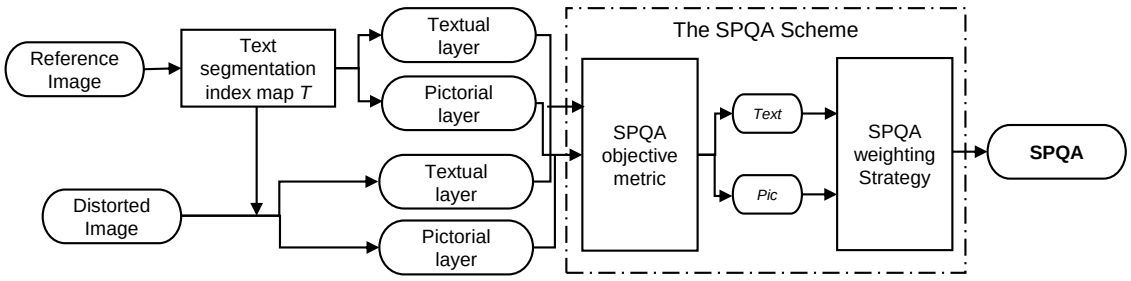



**Figure 1.6 :** SCI perceptual quality assessment (SPQA) framework

Gu *et al.* [2016b] proposed a saliency enabled quality measure of SCIs (SQMS). In order to assess the structural degradation, SQMS uses GM via Scharr operator [Jähne *et al.*, 1999]. The flowchart for SQMS is shown in Figure 1.7. On similar lines, the authors in [Ni *et al.*, 2016] proposed an IQA metric by using the similarity of gradient direction of SCIs. The gradient direction

is computed by finding the direction where the gradient magnitude field has highest variation. Recently, a structural variation based SCI-IQA method (SVQI) has been proposed in [Gu *et al.*, 2018]. SVQI evaluates the similarity between the reference and distorted SCI by identifying the variation in local and global structure.

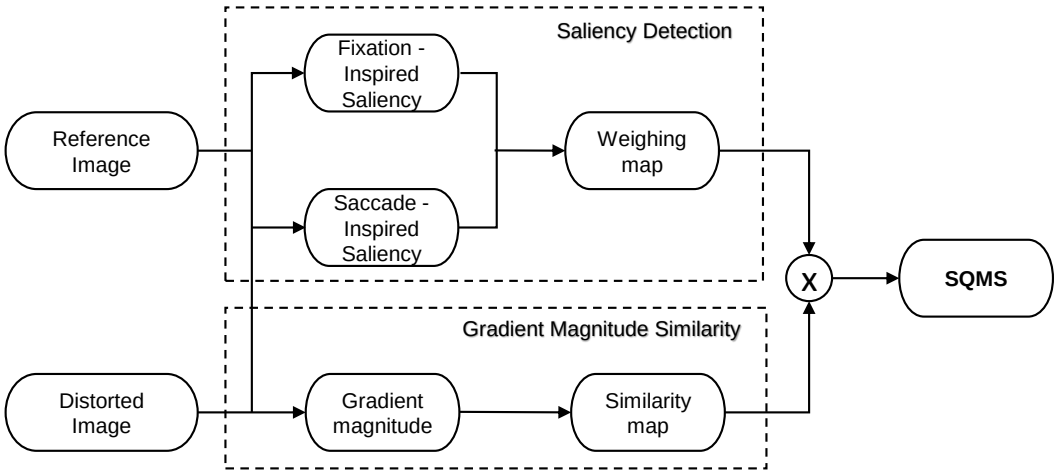



**Figure 1.7 :** Flowchart for saliency enabled quality measure (SQMS)

Although FR-IQA methods can efficiently predict the quality of the distorted image (DI) better than others methods, it is not suitable in case of SCIs where the receiver is connected to a shared remote display. Providing the reference-image (RI) information through the network increases the bit-rate requirements.

The NR-IQA methods do not allow any access to the RI and these methods are mainly suited for CCIs. The major challenge in designing an NR-IQA method is to first identify the distortion type. Most of the NR-IQA methods are distortion centric [Saad *et al.*, 2012], which work best on a particular set of distortions. However, a support vector machine (SVM) based NR-IQA method (NR-SVM) in [Wu *et al.*, 2016] claims to identify the distortion type before evaluating the quality assessment. The distortion type is predicted using SVM classifier by evaluating the distorted image's probability of belonging to each distortion type. The framework of NR-SVM [Wu *et al.*, 2016] is shown in Figure 1.8.

At the acquisition time of CCIs by using physical camera sensors, which may involve multiple distortions. Moreover, this may cause the unavailability of images with perfect quality. On the other hand, SCIs are directly generated from the sender computer and are presumed to have perfect quality. Due to this the NR-IQA methods are also not much useful for SCIs. However, some work has been done towards developing NR-IQA methods for SCIs [Gu *et al.*, 2016a; Qian *et al.*, 2017].

The blind quality measure for SCIs (BQMS)reported in [Gu *et al.*, 2016a] is a learning based model. BQMS first establishes the screen content statistics (SCS) model using free energy measure and the information of structural degradation in order to specify the auto-regressive (AR) model. The framework for BQMS is shown in Figure 1.9. Authors in [Qian *et al.*, 2017] claim to measure the distortion in the SCI by evaluating the loss of edge information using edge-preserving filters.

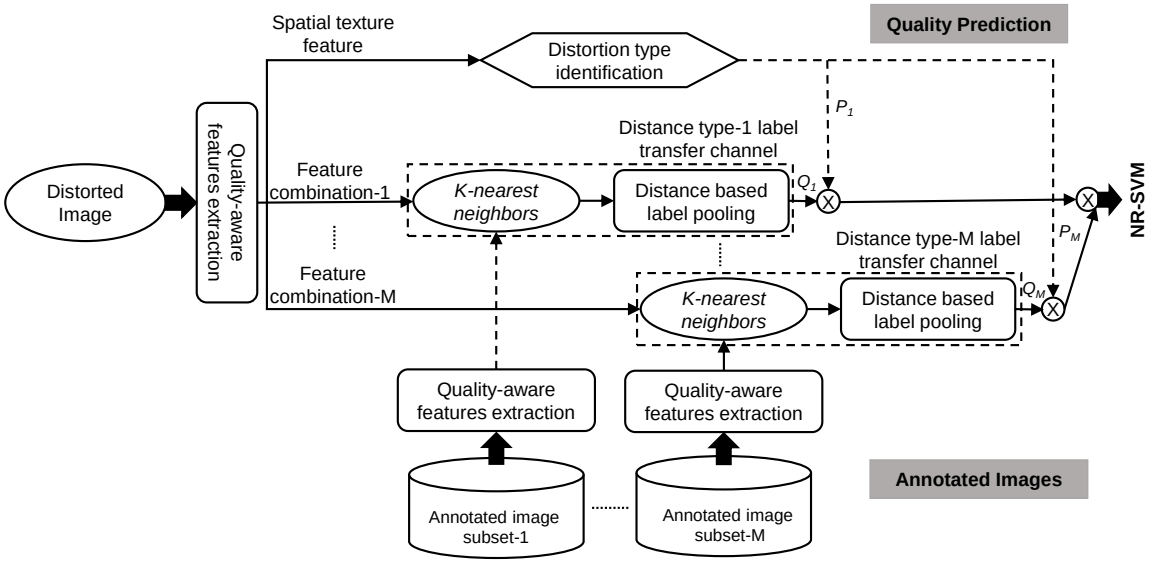



**Figure 1.8 :** Framework of no-reference support vector machine (NR-SVM) method

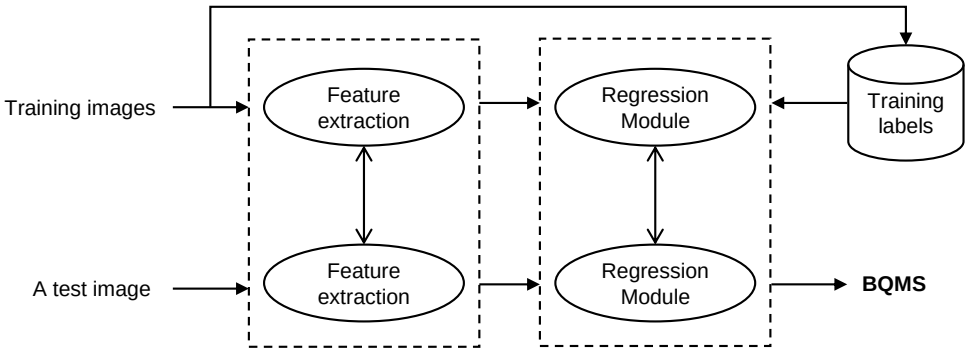



**Figure 1.9 :** Framework of blind quality measure for SCIs (BQMS)

The RR-IQA provides a compromise between FR-IQA and NR-IQA methods where only sparsely sampled meaningful features capable to reflect the visual quality of the RI are provided at the receiver end. The key criteria for any efficient RR-IQA method is to judiciously select the features of reference image such that its property can be captured adequately. Due to this, numerous RR-IQA techniques have been proposed for natural images such as wavelet-domain natural image statistic model (WNISM) [Wang and Simoncelli, 2005], Fourier transform-based scalable image quality measure (FTB) [Narwaria *et al.*, 2012], divisive normalization-based image

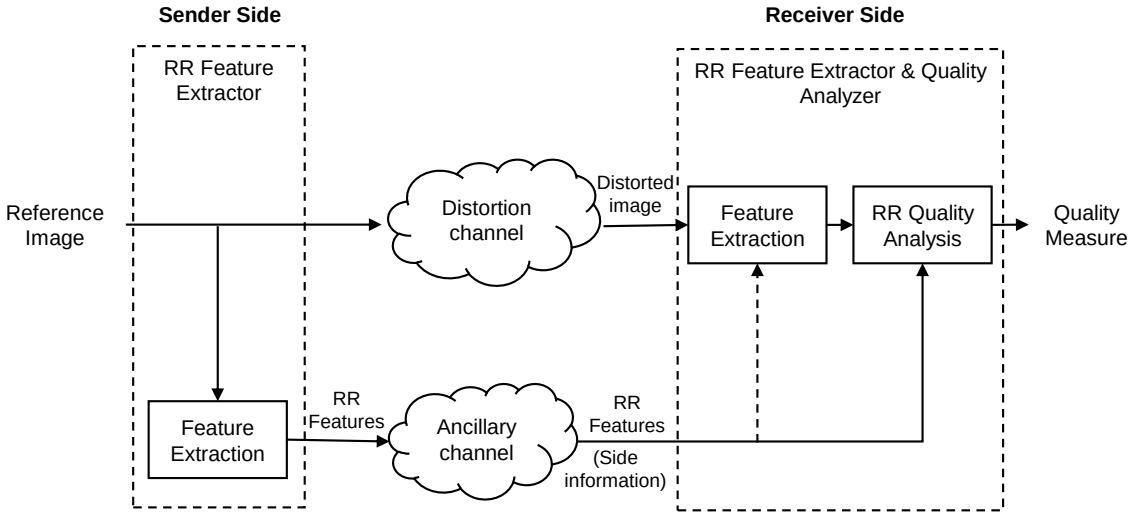



**Figure 1.10 :** General framework of reduced-reference image quality assessment systems

representation (DNT-RR) [Li and Wang, 2009], structural degradation model (SDM) [Gu *et al.*, 2013], edge pattern map (EPM) [Min Zhang, 2011], and reduced-reference image quality metric for contrast change (RIQMC) [Gu *et al.*, 2016c]. A general framework for RR-IQA systems is shown in Figure 1.10

WNISM [Wang and Simoncelli, 2005] is based on natural image statistics model in the wavelet transform domain. The idea of WNISM is to compare the naturalness of the reference and distorted images in order to measure the quality. DNT-RR [Li and Wang, 2009] is based the divisive normalization transform (DNT)[Wainwright and Simoncelli, 2000]. It has been observed that unlike DCT or DWT transform, which can only work on first-order correlation [Schwartz and Simoncelli, 2001], DNT is able to reduce higher order correlation. This helps DNT in order to reflect the neuronal responses in biological visual systems [Heeger, 1992; Simoncelli and Heeger, 1998].

Authors in [Min Zhang, 2011] proposed an IQA metrics called EPM which is based on the statistics of edge discrimination. EPM extracts low-level features from reference and distorted image by creating a binary edge map with the help of the multi-scale wavelet transform modulus. It uses the histogram of edge pattern map, which is generated by applying the gradient operator on the binary edge map, to measure the quality of the distorted image. On the other hand FTB [Narwaria *et al.*, 2012] proposes a regression-based IQA model in frequency domain, which combines the quality score from phase and magnitude. The motivation behind FTB is to provide more importance to the distortion in low frequency regions than in the high frequency ones.

SDM [Gu *et al.*, 2013] proposes an IQA model on the similar lines of SSIM [Wang *et al.*, 2004]. SDM exploits the structural degradation in an image by applying SSIM multiple times and combining them using SVM based integration. The authors in [Gu *et al.*, 2016c] propose a RR-IQA method called RIQMC to measure the distortion due to contrast change. They also proposed RIQMC based optimal histogram mapping (ROHIM) in order to enable the receiver to do a automatic contrast enhancement. It was observed that the third and fourth order statistics (skewness and kurtosis) are connected to human's feeling of comfort. RIQMC fused this information in order to measure the distortion in the contrast. The framework of RIQMC is shown

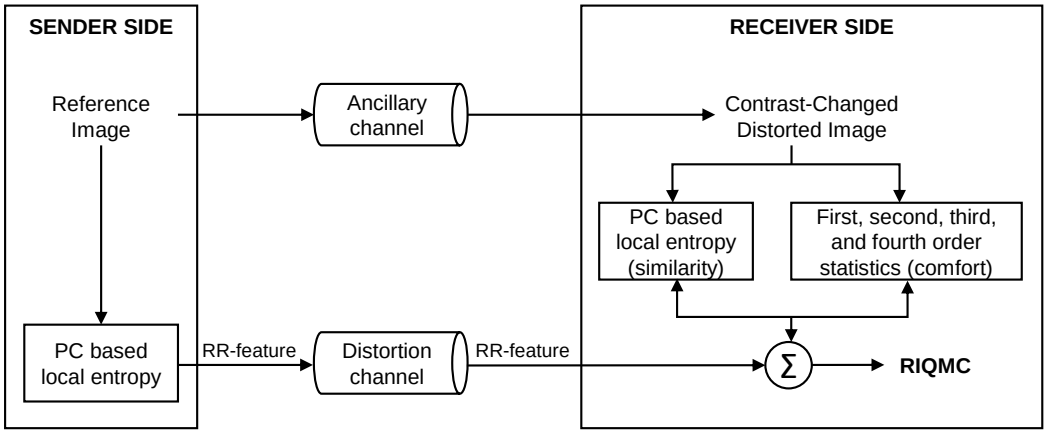



**Figure 1.11 :** IQA framework of reduced-reference image quality metric for contrast change RIQMC

in Figure 1.11.

Few RR-IQA methods for SCIs have also been proposed in recent years such as quality assessment of compressed SCIs (QACS) [Wang *et al.*, 2016], reduced-reference quality index (RQI) [Che *et al.*, 2017], reduced-reference quality assessment (RRQA) [Wang *et al.*, 2018]. The authors in [Wang *et al.*, 2016] develop an IQA model named QACS, which works in two stages. In the first stage, a wavelet domain based feature extraction method is proposed by using a scale-space orientation decomposition. For this, they use the multi-level pyramid structure in the directions of horizontal, vertical, and diagonal (HL, LH, and HH) by decomposing the input SCI in four scales, and merged the sub-bands HL, LH as they share the same statistics. In the second stage, they proposed a quality prediction model based on support vector regressor (SVR) training as proposed in [Schölkopf *et al.*, 2000].

In RQI [Che *et al.*, 2017], the authors first proposed a layer based segmentation model in order to extract the textual and pictorial layer from an SCI. Then RQI evaluates respective quality metrics for textual and pictorial layers, and judiciously combine them to evaluate the quality measure of the distorted SCI. In RRQA [Wang *et al.*, 2018], in order to find the quality of the distorted SCI, the statistical features are first extracted, that takes into account both primary HVS information and unpredictable uncertainty. In order to measure the quality of the distorted image, a significance histogram is generated with the help of GM, and the low pass filter. The flowchart for the RRQA model is shown in Figure 1.12.

Detectable image degradation will lead to change or loss in the image features, and an efficient IQA metric can be established by evaluating these changes. Many efforts have been made in developing feature-based IQA techniques [Zhang *et al.*, 2011; Wang *et al.*, 2016] as also discussed above. Scale-invariant feature transform (SIFT) [Lowe, 2004] is a widely used feature extraction method, which is extensively used in different image processing areas such as object identification, image matching, quality assessment, and many others. The feature extraction process of SIFT is discussed in detail in Section 1.2.2. The work done on SIFT-based IQA for CCIs have shown promising results in [Temel and AlRegib, 2016; Wen *et al.*, 2014; Sun *et al.*, 2014; Chen

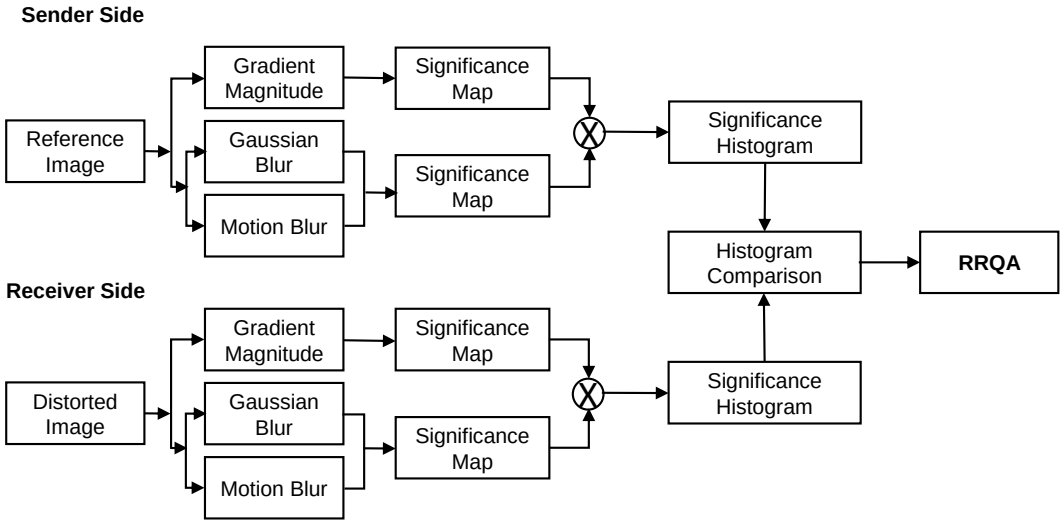



**Figure 1.12 :** Flowchart for RRQA

and Coulombe, 2013; Decombas *et al.*, 2012]. These methods measure the quality of the distorted SCI by comparing the features of reference and distorted images, where the features are extracted using SIFT.

The features contained in an SCI generally have non-identical importance i.e. there are some features which may provide more or less information than others. The work done on feature-based IQA techniques so far for CCIs as-well-as SCIs have provided equal importance to every feature in the reference image [Zhang *et al.*, 2011; Temel and AlRegib, 2016; Wen *et al.*, 2014; Sun *et al.*, 2014; Chen and Coulombe, 2013; Decombas *et al.*, 2012]. To achieve a better IQA metric, the features of the image should be weighted according to their importance, and it should be consistent with the HVS.

## 1.2 RELATED WORKS

This section discusses a few methods which will be used in the chapters to be followed throughout the thesis. Moreover, the steps which are relevant in the later part of the thesis are discussed in details here in order to provide a better insight.

### 1.2.1 JPEG Baseline

The encoding and decoding process of JPEG baseline is shown in Figure 1.13. There are three basic steps in JPEG compression [Wallace, 1992], such as color sub-sampling, block-DCT coefficient quantization, and entropy coding. The required bit-rate after encoding by JPEG baseline is mainly controlled in the quantization phase. The DCT coefficients of all the blocks are quantized by fixed quantization parameter of its quantization table (T, i.e. a quantization matrix which provides step sizes). As per the JPEG standard, the quantization table can be configured as per the bit-rate requirement [Yang *et al.*, 2016]. There have been series of several quantization tables developed and are widely used for the requirement of higher compression ratio or improved reconstructed image quality. The JPEG recommended a series of quantization tables $(T_F)$ is given in (1.4).

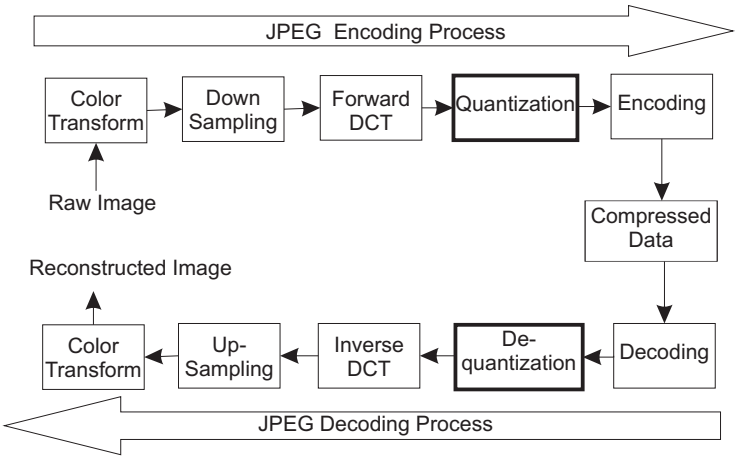



**Figure 1.13 :** Encoding and Decoding Process of JPEG Baseline

$$T_F = \begin{cases} \lfloor T_{50} \times \frac{50}{F} + \frac{1}{2} \rfloor, & 1 \leq F < 50 \\ \lfloor T_{50} \times (2 - \frac{F}{50}) + \frac{1}{2} \rfloor, & 50 \leq F \leq 100 \end{cases} \tag{1.4}$$

The value of quality factor $(F)$ ranges between 1 to 100. The small value of $F$ corresponds to large quantization step-size, and thus results in high CR. $T_{50}$ is the standard quantization table [Wallace, 1992] used for luminance component of an image is given in (1.5). JPEG uses separate quantization table for the chrominance component of the color images (1.6).

$$T_{50} = \begin{bmatrix} 16 & 11 & 10 & 16 & 24 & 40 & 51 & 61 \\ 12 & 12 & 14 & 19 & 26 & 58 & 60 & 55 \\ 14 & 13 & 16 & 24 & 40 & 57 & 69 & 56 \\ 14 & 17 & 22 & 29 & 51 & 87 & 80 & 62 \\ 18 & 22 & 37 & 56 & 68 & 109 & 103 & 77 \\ 24 & 35 & 55 & 64 & 81 & 104 & 113 & 92 \\ 49 & 64 & 78 & 87 & 103 & 121 & 120 & 101 \\ 72 & 92 & 95 & 98 & 112 & 100 & 103 & 99 \end{bmatrix} \tag{1.5}$$

$$T_{50}^C = \begin{bmatrix} 17 & 18 & 24 & 47 & 99 & 99 & 99 & 99 \\ 12 & 12 & 14 & 19 & 99 & 99 & 99 & 99 \\ 14 & 13 & 16 & 19 & 99 & 99 & 99 & 99 \\ 14 & 17 & 19 & 19 & 99 & 99 & 99 & 99 \\ 19 & 19 & 19 & 19 & 99 & 99 & 99 & 99 \\ 19 & 19 & 19 & 19 & 99 & 99 & 99 & 99 \\ 19 & 19 & 19 & 19 & 99 & 99 & 99 & 99 \\ 19 & 19 & 19 & 19 & 99 & 99 & 99 & 99 \end{bmatrix} \tag{1.6}$$

For illustration purpose, few other chrominance quantization tables at $F = 100, 60, 20, 1$ are given in (1.7-1.10), respectively.

$$T_{100}^{C} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \tag{1.7}$$

$$T_{60}^{C} = \begin{bmatrix} 13 & 9 & 8 & 13 & 19 & 32 & 41 & 49 \\ 10 & 10 & 11 & 15 & 21 & 46 & 48 & 44 \\ 11 & 10 & 13 & 19 & 32 & 46 & 55 & 45 \\ 11 & 14 & 18 & 23 & 41 & 70 & 64 & 50 \\ 14 & 18 & 30 & 45 & 54 & 87 & 82 & 62 \\ 19 & 28 & 44 & 51 & 65 & 83 & 90 & 74 \\ 39 & 51 & 62 & 70 & 82 & 97 & 96 & 81 \\ 58 & 74 & 76 & 78 & 90 & 80 & 82 & 79 \end{bmatrix} \tag{1.8}$$

$$T_{20}^{C} = \begin{bmatrix} 40 & 28 & 25 & 40 & 60 & 100 & 128 & 153 \\ 30 & 30 & 35 & 48 & 65 & 145 & 150 & 138 \\ 35 & 33 & 40 & 60 & 100 & 143 & 173 & 140 \\ 35 & 43 & 55 & 73 & 128 & 218 & 200 & 155 \\ 45 & 55 & 93 & 140 & 170 & 273 & 258 & 193 \\ 60 & 88 & 138 & 160 & 203 & 260 & 283 & 230 \\ 123 & 160 & 195 & 218 & 258 & 303 & 300 & 253 \\ 180 & 230 & 238 & 245 & 280 & 250 & 258 & 248 \end{bmatrix} \tag{1.9}$$

$$T_{1}^{C} = \begin{bmatrix} 800 & 550 & 500 & 800 & 1200 & 2000 & 2550 & 3050 \\ 600 & 600 & 700 & 950 & 1300 & 2900 & 3000 & 2750 \\ 700 & 650 & 800 & 1200 & 2000 & 2850 & 3450 & 2800 \\ 700 & 850 & 1100 & 1450 & 2550 & 4350 & 4000 & 3100 \\ 900 & 1100 & 1850 & 2800 & 3400 & 5450 & 5150 & 3850 \\ 1200 & 1750 & 2750 & 3200 & 4050 & 5200 & 5650 & 4600 \\ 2450 & 3200 & 3900 & 4350 & 5150 & 6050 & 6000 & 5050 \\ 3600 & 4600 & 4750 & 4900 & 5600 & 5000 & 5150 & 4950 \end{bmatrix} \tag{1.10}$$

To achieve low bit-rate requirements, the visual quality of the blocks with higher variance (or frequency) get more degraded after reconstruction compared to the blocks with lower variance. To illustrate this, we compressed an image at different bit-rates using JPEG baseline and highlighted a high variance region to analyze the distortion occurred, as shown in Figure 1.14. Figure 1.14 (a) is the original $(512 \times 512)$ dimension color image at 24 bits-per-pixel (bpp) and Figure 1.14 (b-d) are the reconstructed images after applying JPEG compression with quantization tables $T_{50}$, $T_{10}$, and $T_5$, respectively. It can be observed that the distortion in the highlighted region increases at lower bit-rate, causing degradation in the perceptual quality of the overall image. A similar effect is also observed in Figure 1.15 for Lena image with quantization tables $T_{20}$, $T_{10}$, and $T_5$, respectively.

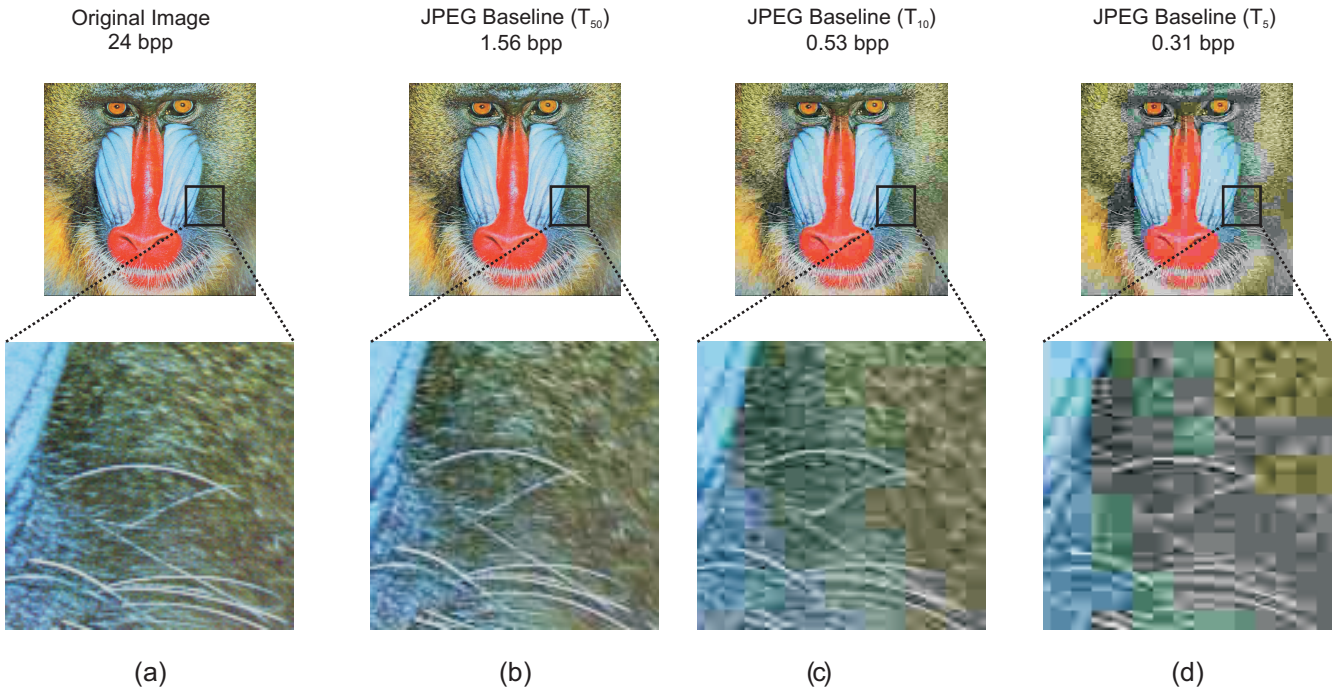



**Figure 1.14 :** Effect of JPEG compression at different scale values for Baboon image

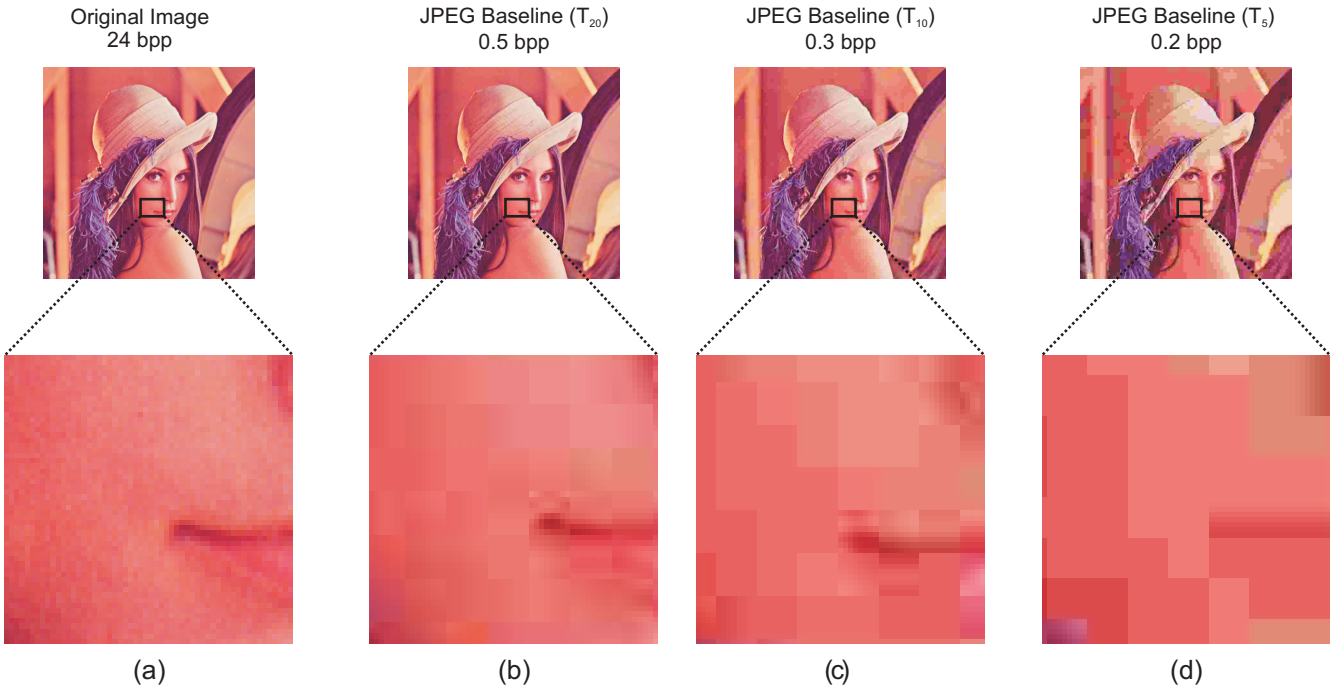



**Figure 1.15 :** Effect of JPEG compression at different scale values for Lena image

### 1.2.2 Scale-Invariant Feature Transform (SIFT)

Scale-invariant feature transform (SIFT) generates the set of image features in four major stages [Lowe, 2004]:

#### *Initial Feature Points Selection*

The candidate feature locations in the image are identified by using cascade filtering approach. The idea is to locate the feature points which are invariant to scale change of the image, using a continuous function of scale, known as scale-space [Lowe, 2004]. The function $S(x,y,\sigma)$ defines the scale-space of an image, which is produced by convolving variable-scale Gaussian function, $G(x,y,\sigma)$ with the image $I(x,y)$:

$$S(x,y,\sigma) = G(x,y,\sigma) * I(x,y) \tag{1.11}$$

where $*$ is the convolution operator in x, and y, and

$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \tag{1.12}$$

The difference-of-Gaussian (DOG) images, $D(x,y,\sigma)$ are then evaluated from the difference between two nearby scales which are separated by a constant multiply factor $k$ as shown in Figure 4.4. The DOG images are defined as follows:

$$D(x,y,\sigma) = (G(x,y,k\sigma) - G(x,y,\sigma)) * I(x,y) = S(x,y,k\sigma) - S(x,y,\sigma) \tag{1.13}$$

The scale-space pyramids as shown in Figure 4.4 are generated for $O$ number of octaves. Every octave of scale-space is segmented into $N$ number of intervals, so $k = 2^{1/N}$. Then number of scale-space $N$ in each octave is proposed to be 3 [Lowe, 2004]. To cover a complete octave, $N+3$ blurred images are produced in each octave. The number of octaves, $O$, is determined according to the size of image. The scale coefficient $C$ of $l^{th}$ scale layer in the $o^{th}$ octave in the scale-space pyramid is computed as follows:

$$\begin{aligned} C_{o,l} &= \sigma_1 2^{(o-1+(l/N))} = \sigma_1 k^{N(o-1)+l} \\ o = 1,2,...,O; \quad l = 1,2,...,N; \quad k &= 2^{1/N} \\ \sigma_1 &= 1.6 \end{aligned} \tag{1.14}$$

In order to find the maxima and minima of DOG pyramid, each pixel is compared to its eight neighbors at the same scale layer, and nine neighbors in the scale above and below (i.e. 26 in total), as shown in Figure 4.5. The point is selected only if it is higher or smaller than all of its neighbors.

#### *Refining Feature Points*

To reject the unstable feature points detected in the first step, the point that have low contrast or are poorly localized along an edge are removed. In order to do that, the subpixel location, the scale, and ratio of principle curvature of the DOG function $D(x,y,\sigma)$ are calculated. A 3D surface is fitted to the feature points using second order Taylor expansion of the DOG function $D(x,y,\sigma)$. These fitted-points are then shifted such that the origin is at the sample point.

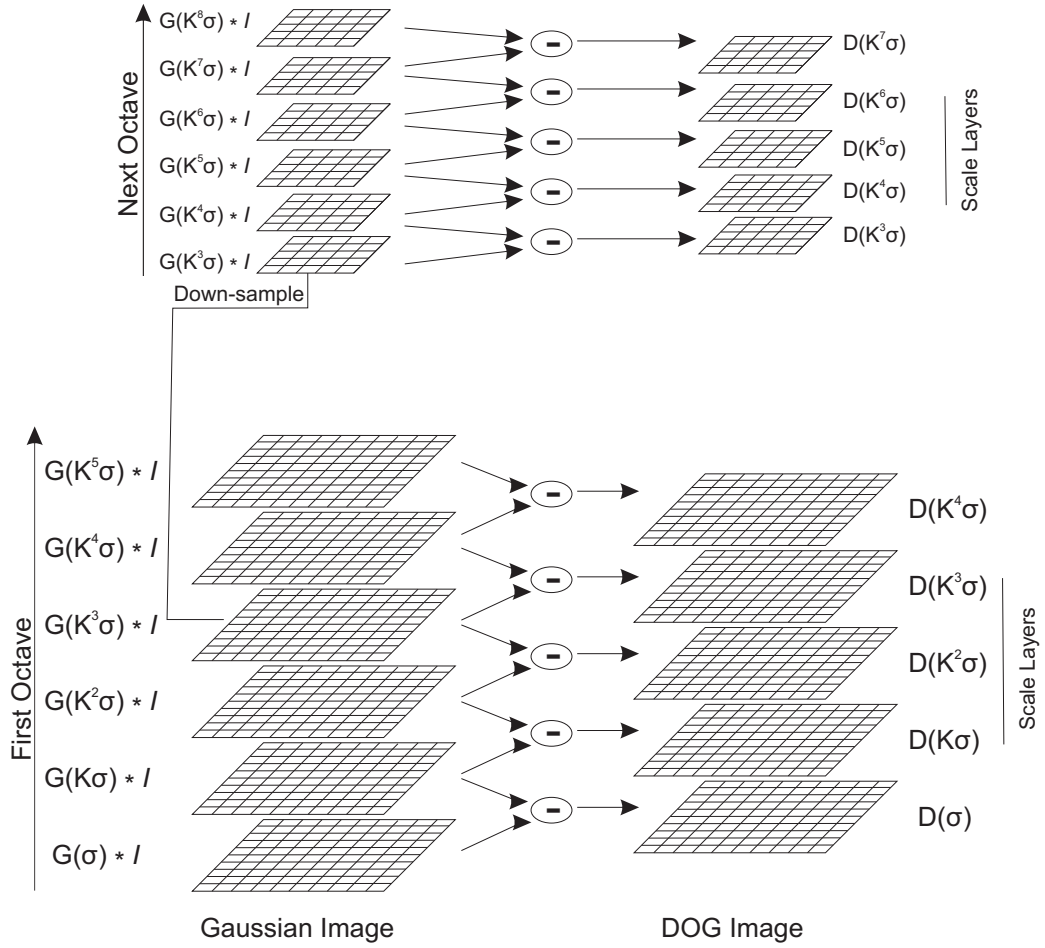



**Figure 1.16 :** Gaussian images (left), and difference-of-Gaussian images (right) are shown for each octave of scale-space. The Gaussian image is down-sampled by a factor of 2 for the next octave.

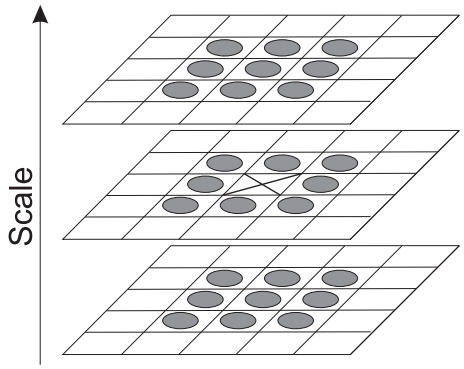
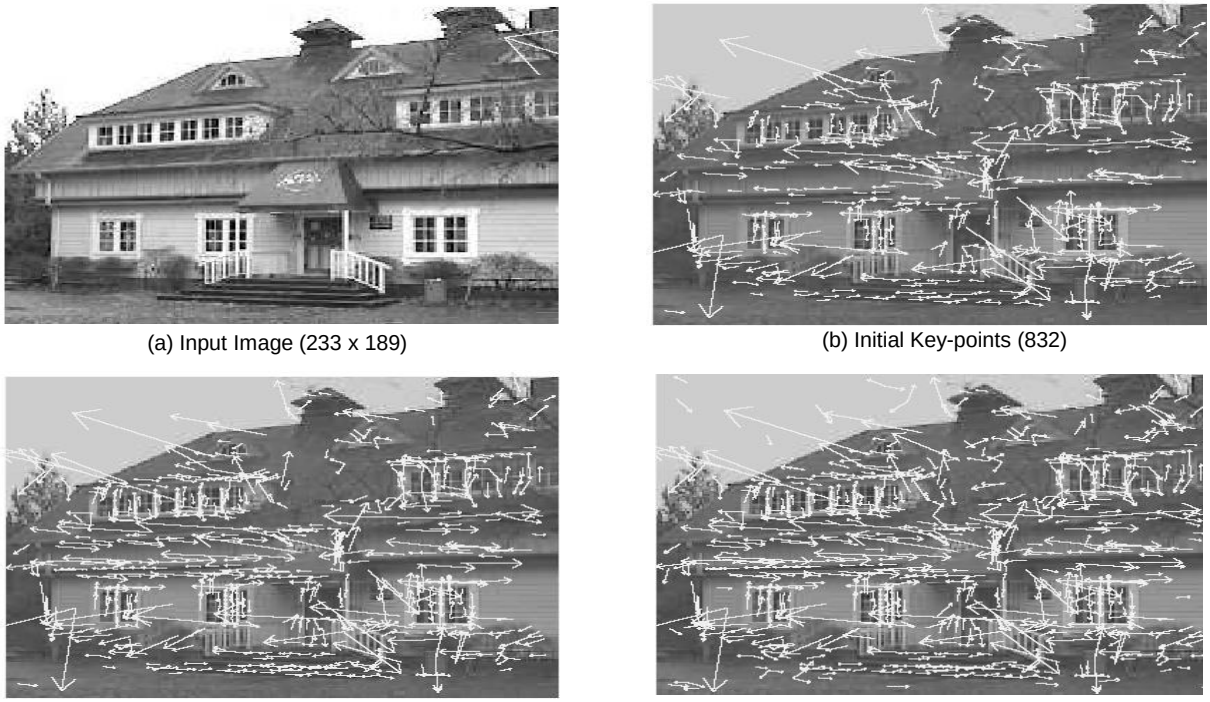



**Figure 1.17 :** The pixel marked with $X$ is compared to its 26 neighbors at the current and adjacent scales (marked with circles) in order to detect the maxima and minima of the DOG



(a) Input Image (233 x 189)

(b) Initial Key-points (832)

(c) Key-points after removing low contrast points (721)

(d) Key-points after removing points on edges (503)

**Figure 1.18 :** Effect of pruning feature points. (a) Input Image of dimension $233 \times 189$, (b) Shows the initial key-points as discussed in Chapter 1.2.2, (c) Shows the key-points after removing low contrast points, (d) Shows the remaining key-points after removing the points which are on edges

The effect of pruning feature points is shown in Figure 1.18. For an input Image of dimension $233 \times 189$, Figure 1.18 (b) shows 832 key-points after initial feature points selection, as discussed in Chapter 1.2.2. The key-points after removing low contrast points are shown in Figure 1.18 (c). It was observed that about 14% key-points were discarded as low contrast points. Figure 1.18 (d) Shows the remaining key-points after removing the points which are on edges. This further discarded about 31% of remaining key-points after removing the low contrast points.

$$\frac{Tr(H)}{Det(H)} \geq 12.1 \tag{1.15}$$

***Orientation Assignment***

The gradient magnitude $M(x,y)$, and the orientation $\theta(x,y)$, is computed using pixel differences as shown in (4.12), and (4.13) for every sample image $L(x,y)$, at a scale.

$$M(x,y) = \sqrt{[(L(x+1,y) - L(x-1,y)]^2 + [L(x,y+1) - L(x,y-1)]^2} \tag{1.16}$$

$$\theta(x,y) = \tan^{-1}\left[\frac{(L(x+1,y) - L(x-1,y))}{(L(x,y+1) - L(x,y-1))}\right] \tag{1.17}$$

The gradient orientation within a region around a feature point is then used to create an orientation histogram. This histogram consists of 36 bins corresponds to 360 degree range of orientations. The samples added to the histogram are weighted by using Gaussian-weighted circular window with a scaling factor 1.5 times higher than the scale of the feature point. A feature point is only created for the highest peak in the histogram or any other peak which is within 80% of the highest peak with the corresponding orientation. This provides multiple feature points at the same location and scale but a different orientation.

#### *Local Feature Discriptors*

The feature point descriptors are created by computing the gradient magnitude and orientation of sample points around the feature point location with a $16 \times 16$ window size. These are weighted by a Gaussian window with scaling factor as one half the width of descriptor window. An orientation histogram is then formed summarizing the contents over $4 \times 4$ subregions into 8 directions. This leads to a descriptor vector for each feature of length $4 \times 4 \times 8 = 128$.

## 1.3 THESIS OVERVIEW AND CONTRIBUTIONS

This thesis contains four contributions in the area of image compression and its quality assessment. The first contribution is towards developing multilevel saliency enabled compression method for camera-content images (CCIs) as discussed in Chapter 2. Because of the computational efficiency of DCT based methods [Xiong *et al.*, 1999], with a small loss in performance as compared to DWT based such methods, our focus has been on the former one. We developed a multi-level saliency-based image compression algorithm that provides a statistically optimal trade-off between the overhead, perceptual quality, and CR. The developed method chooses variance as a basis to classify and rank the image into an optimal number of classes [Rahul and Tiwari, 2018]. The aim is to enable JPEG standard to judiciously retain high-frequency regions in the image that provides perceptual homogeneity in compression, particularly in the case of high compression requirement.

In order to do this, the given image is segmented into multiple salient regions in such a way that the between-class variances are maximized [Otsu, 1979; Huang and Wang, 2009]. The different regions are ranked according to their weighted-variances i.e. within-class variances [Huang and Wang, 2009]. A region with high variance is given more importance and vice versa. We also devised an adaptive method to compute the number of different salient regions in an image. This is done based on the goodness of segmentation (GoS) [Otsu, 1979; Lagarias *et al.*, 1998], using the ratio of between-class variance and total variance. After identifying the multi-level salient regions and their rank information, the image is divided into non-overlapping blocks of size $8 \times 8$. If a block falls into more than one regions, then probability bound is used to identify its rank. It is observed that ranks of neighbouring blocks are, generally, highly correlated and hence delta encoding [Schindler, 1970] method is employed to reduce the overhead information for the ranks. 2D-DCT coefficients are obtained for each of the blocks and these coefficients are adaptively quantized based on the rank of the block.

After designing a compression framework for CCIs we proceeded towards developing a similar compression algorithm for screen-content images (SCIs) as discussed in chapter 3 [Rahul and Tiwari, 2019b]. The JPEG quantizer here was made intelligent to perform a judicious quantization. The key difference in this method is that instead of a multi-level saliency map, a binary saliency map is provided to the quantizer to identify text or non-text regions. Moreover, as JPEG framework transforms the image blocks into DCT at the initial state, we decided to use this DCT information for marking every block as textual or non-textual to create a saliency map with low computation cost.

This method provides an optimal trade-off between the overhead, perceptual quality, and CR. The developed method first classifies the SCI into the salient and non-salient region by identifying the textual information. The aim is to enable the JPEG standard to judiciously retain high-frequency text regions in the image, particularly in the case of high compression requirements.

To accurately identify the performance of the proposed compression methods, two reduced-reference Image-Quality-Assessment (RR-IQA) methods for camera and screen content images are also proposed in this thesis. These methods are based on the fact that Human Visual System (HVS) is more sensitive towards change in features than intensity or structure.

In the final contribution, two feature-based reduced-reference IQA methods are designed for CCIs and SCIs [Rahul and Tiwari, 2019a] in Chapter 4. For this, the feature points are extracted from the reference image (RI) and distorted image (DI). A low dimensional feature descriptor structure has been developed to reduce the computation complexity in the feature matching process. A feature matching technique is also developed by calculating the descriptor distance between each feature of the reference image and the feature present in their vicinity in the distorted image. This makes the proposed method computationally faster than other feature based IQA methods. The developed method also normalizes the features as per their importance to get a better IQA result with a larger dynamic range compared to the current state-of-the-art IQA techniques.

SCIs are usually high dimensional and more informative compared to natural images, and due to this, it contains more feature points. The proposed descriptor structure also helps in achieving low bit-rate for sending the RI information to the receiver.

…