International Journal of Electronics and Communication Engineering.
ISSN 0974-2166 Volume 4, Number 5 (2011), pp. 491-516

http://www.irphouse.com

Automatic Detection of Perceived Ringing Regions in Compressed Images

D. Minola Davids

Research Scholar, Singhania University, Rajasthan, India

Abstract

An efficient approach toward a no-reference ringing metric intrinsically exists of two steps: first detecting regions in an image where ringing might occur, and second quantifying the ringing annoyance in these regions. This paper presents a novel approach toward the first step: the automatic detection of regions visually impaired by ringing artifacts in compressed images. It is a no-reference approach, taking into account the specific physical structure of ringing artifacts combined with properties of the human visual system (HVS). To maintain low complexity for real-time applications, the proposed approach adopts a perceptually relevant edge detector to capture regions in the image susceptible to ringing, and a simple yet efficient model of visual masking to determine ringing visibility.

Index Terms: Luminance masking, per-ceptual edge, ringing artifact, texture masking.

Introduction

In current visual communication systems, the most essential task is to fit a large amount of visual information into the narrow bandwidth of transmission channels or into a limited storage space, while maintaining the best possible perceived quality for the viewer. A variety of compression algorithms, for example, such as JPEG[1] and MPEG/H.26xhave been widely adopted in image and video coding trying to achieve high compression efficiency at high quality. These lossy compression techniques, however, inevitably result in various coding artifacts, which by now are known and classified as blockiness, ringing, blur, etc.

The blocking effect is the discontinuities found at the boundaries of adjacent blocks in a reconstructed frame, since it is a common practice to measure and reduce the blocking effect by only taking into account the pixels at the block boundaries. Blurring manifests as a loss of spatial detail and a reduction in sharpness of edges in

cmoderate to high spatial activity regions of frames, such as in roughly textured areas or around scene object edges. The occurrence of the compression induced artifacts depends on the data source, target bit rate, and underlying compression scheme and their visibility can range from imperceptible to very annoying, thus affecting perceived quality.

Depending on the compression scheme used and target bit-rate, the visibility of these artifacts can range from imperceptible to extremely annoying, thus affecting the perceived quality of the digital video. To alleviate these effects, many algorithms have been designed to remove or reduce the visibility of these artifacts thereby increasing the video quality. The removal of the artifacts is best carried out at the receiving end prior to local storage or display to prevent propagation of new artifacts through the communication pipeline. Hence, to accurately detect regions with perceived ringing, two essential aspects need to be explicitly addressed: 1) an (strong) edge detector; and 2) a masking model of the HVS.

Compression Standards

JPEG (the Joint Photographic Experts Group) has been responsible for a number of standards for coding of still images. The most relevant of these are the JPEG standard and the JPEG2000 standard. Each share features with MPEG-4 Visual and/or H.264 and whilst they are intended for compression of still images, the JPEG standards have made some impact on the coding of moving images. The 'original' JPEG standard supports compression of still photographic images using the 8× 8 DCT followed by quantization, reordering, run-level coding and variable-length entropy coding and so has similarities to MPEG-4Visual (when used in Intra-coding mode). JPEG2000 was developed to provide a more efficient successor to the original JPEG. It uses the Discrete Wavelet Transform as its basic coding method and hence has similarities to the Still Texture coding tools of MPEG-4 Visual. JPEG2000 provides superior compression performance to JPEG and does not exhibit the characteristic blocking artifacts of a DCT-based compression method.

The Moving Picture Experts Group is a Working Group of the International Organization for Standardization (ISO) and the International Electro technical Commission (IEC). MPEG's remit is to develop standards for compression, processing and representation of moving pictures and audio. It has been responsible for a series of important standards starting with MPEG-1 (compression of video and audio for CD playback) and following on with the very successful MPEG-2 (storage and broadcasting of 'television-quality' video and audio). MPEG-4 (coding of audio-visual objects) is the latest standard that deals specifically with audio-visual coding. Two other standardization efforts (MPEG-7 and MPEG-21)are concerned with multimedia content representation and a generic 'multimedia framework'. However, MPEG is arguably best known for its contribution to audio and video compression.

H.261 was the first widely-used standard for videoconferencing, developed by the ITU-T to support video telephony and videoconferencing over ISDN circuit-switched networks. These networks operate at multiples of 64 Kbit/s and the standard was designed to offer computationally-simple video coding for these bitrates. The standard

uses the familiar hybrid DPCM/DCT model with integer-accuracy motion compensation. In an attempt to improve on the compression performance of H.261, the ITU-T working group developed H.263. This provides better compression than H.261, supporting basic video quality at bitrates of below30 Kbit/s, and is part of a suite of standards designed to operate over a wide range of circuit- and packet-switched networks.

Coding Artifacts

A comprehensive analysis and classification of the numerous coding artifacts which are introduced into the reconstructed video sequence through the use of the hybrid MC/DPCM/DCT video coding algorithm [2]. Artifacts such as the blocking effect, ringing, and the mosquito effect, MC mismatch, blurring, and color bleeding, will be comprehensively analyzed.

Two types of impairments in a psychophysical experiment to measure the overall annoyance and individual strength of three impairment features (fuzzy, blocky, and blurry) are used [3]. The impairments were generated by compressing the original videos with MPEG-2 at two different bitrates: 1.0 and 7.5 Mbps. The heavily compressed videos presented blurry and blocky impairments, while the lightly compressed videos presented 'fuzzy' impairments [4]. Two types of post processing algorithms based on image enhancement and restoration principles are reviewed. Finally, current bottlenecks and future research directions in this field are addressed [5].A scheme for artifact reduction of low bit rate discrete-cosine-transform-compressed (DCT-compressed) images is presented [6]. First, the DC coefficients are calibrated using gradient continuity constraints. Then, an improved Huber-Markov-random-field-based (HMRF-based) smoothing is applied. The constrained optimization is implemented by the iterative conditional mode (ICM). Final reconstructions of typical images with improvements in both visual quality and peak signal-to-noise ratio (PSNR) are also shown.

A new approach that can blindly measure blocking artifacts in images without reference to the originals [7].The key idea is to model the blocky image as a non-blocky image interfered with a pure blocky signal. The task of the blocking effect measurement algorithm is then to detect and evaluate the power of the blocky signal. A novel no-reference blockiness metric that provides a quantitative measure of blocking annoyance in block-based DCT coding is presented. [8]The metric incorporates properties of the human visual system (HVS) to improve its reliability, while the additional cost introduced by the HVS is minimized to ensure its use for real-time processing. A new method is proposed for coding artifact reduction of MPEG compressed video sequences [9]. The method makes use of a simple cost-effective technique that allows the block grid position and its visibility to be determined without the need for access to the coding parameters. A novel no-reference metric to quantify ringing annoyance in compressed images is presented [10]. Ringing annoyance is estimated first locally based on the number of visible ringing pixels in a ringing region, taking into account possible spurious ringing pixels .A novel no-reference metric that can automatically quantify ringing annoyance in

compressed images is presented [11]. In the first step a recently proposed ringing region detection method extracts the regions which are likely to be impaired by ringing artifacts.

A full- and no-reference blur metric as well as a full-reference ringing metric [12]. These metrics are based on an analysis of the edges and adjacent regions in an image and have very low computational complexity. As blur and ringing are typical artifacts of wavelet compression, the metrics are then applied to JPEG2000 coded images.

Ringing is an annoying artifact frequently encountered in low bit-rate transform and sub band decomposition based compression of different media such as image, intra frame video and graphics [13]. A mathematical morphological based post-processing algorithm is presented for image ringing artifact suppression. First, they used binary morphological operators to isolate the regions of an image where the ringing artifact is more prominent to human visual system(HVS) while preserving genuine edges and other fine details present in the image. Then, a gray-level morphological nonlinear smoothing filter is applied to the unmasked regions of the image under the filtering mask to eliminate ringing within this constraint region.

A computational approach to edge detection[14]. The success of the approach depends on the definition of a comprehensive set of goals for the computation of edge points. These goals must be precise enough to delimit the desired behavior of the detector while making minimal assumptions about the form of the solution.

Bilateral filtering smooth's images while preserving edges, by means of a nonlinear combination of nearby image values [15]. The method is noniterative, local, and simple. It combines gray levels or colors based on both their geometric closeness and their photometric similarity, and prefers near values to distant values in both domain and range. In contrast with filters that operate on the three bands of a color image separately, a bilateral filter can enforce the perceptual metric underlying the CIE-Lab color space, and smooth colors and preserve edges in a way that is tuned to human perception. A novel approach towards automatic detection of perceived ringing regions is presented [16]. The algorithm takes into account both the physical structure and the human visual perception of the ringing artifacts. All perceived ringing regions are explicitly captured by means of a newly proposed edge detector, followed by an efficient analysis of ringing visibility around each detected edge segment.

To integrate quadtree (QT) decomposition with the block-shift filtering for deblocking by incorporating the QT decomposition, they easily find the locations of uniform regions and determine the corresponding suitable block sizes [17]. The variable block sizes generated by the QT decomposition facilitate the later block-shift filtering with low computational cost. In addition, large block based shift filtering can provide better deblocking results because the smoothing range of large blocks spans over the conventional 8 × 8 block size. Furthermore, they extend the QT based block-shifting algorithm for deringing JPEG2000 coded images.

Ringing

A compression artifact (or artifact) is a noticeable distortion of media [18]– an image,

audio or video– due to the application of an overly aggressive or inappropriate lossy data compression algorithm. These lossy data compression schemes discard some data to simplify the media sufficiently to store it in the desired space (data-rate) – if there is not enough data in the compressed version to reproduce the original with acceptable fidelity, artifacts will result. Alternatively, the compression algorithm may incorrectly determine certain distortions to be of little subjective importance, but they may in fact be objectionable to the viewer.

Compression artifacts occur in many common media such as DVDs common computer file formats such as JPEG, MP3 or MPEG files, and Sony's ATRAC compression algorithm. Uncompressed media (such as on Laserdiscs, Audio CDs, and WAV files) or losslessly compressed media (FLAC, PNG, etc.) do not suffer from compression artifacts. The minimization of artifacts is a key goal in implementation of lossy compression schemes. However, artifacts are occasionally intentionally produced for artistic purposes, a style known as glitch art. Technically speaking, a compression artifact is a particular class of data error that is usually the consequence of quantization in lossy data compression. Where transform coding is used, they typically assume the form of one of the basis functions of the coder's transform space.

In signal processing, particularly digital image processing, ringing artifacts are artifacts that appear as spurious signals near sharp transitions in a signal. Visually, they appear as bands or "ghosts" near edges; audibly, they appear as "echos" near transients, particularly sounds from percussion instruments; most noticeable are the pre-echos. As with other artifacts, their minimization is a criterion in filter design. The main cause of ringing artifacts is overshoot and oscillations in the step response of a filter.

The main cause of ringing artifacts is due to a signal being band limited (specifically, not having high frequencies) or passed through a low-pass filter; this is the frequency domain description. In terms of the time domain, the cause of this type of ringing is the ripples in the sinc function which is the impulse response (time domain representation) of a perfect low-pass filter. Mathematically, this is called the Gibbs phenomenon. One may distinguish overshoot (and undershoot), which occurs when transitions are accentuated – the output is higher than the input – from ringing, where after an overshoot, the signal overcorrects and is now below the target value; these phenomena often occur together, and are thus often conflated and jointly referred to as "ringing".

The term "ringing" is most often used for ripples in the time domain, though it is also sometimes used for frequency domain effects: windowing a filter in the time domain by a rectangular function causes ripples in the frequency domain for the same reason as a brick-wall low pass filter (rectangular function in the frequency domain) causes ripples in the time domain, in each case the Fourier transform of the rectangular function being the sinc function. There are related artifacts caused by other frequency domain effects, and similar due to unrelated causes.

The sinc function, the impulse response for an ideal low-pass filter, illustrating ringing for an impulse. The Gibbs phenomenon, illustrating ringing for a step function. By definition, ringing occurs when a non-oscillating input yields an oscillating output: formally, when an input signal which is monotonic on an interval

has output response which is not monotonic. This occurs most severely when the impulse response or step response of a filter has oscillations – less formally, if for a spike input, respectively a step input (a sharp transition), the output has bumps. Ringing most commonly refers to step ringing, and that will be the focus.

Ringing is closely related to overshoot and undershoot, which is when the output takes on values higher than the maximum (respectively, lower than the minimum) input value: one can have one without the other, but in important cases, such as a low-pass filter, one first has overshoot, then the response bounces back below the steady-state level, causing the first ring, and then oscillates back and forth above and below the steady-state level. Thus overshoot is the first step of the phenomenon, while ringing is the second and subsequent steps. Due to this close connection, the terms are often conflated, with "ringing" referring to both the initial overshoot and the subsequent rings.

If one has a linear time invariant (LTI) filter, then one can understand the filter and ringing in terms of the impulse response (the time domain view), or in terms of its Fourier transform, the frequency response (the frequency domain view). Ringing is a time domain artifact, and in filter design [19] is traded off with desired frequency domain characteristics: the desired frequency response may cause ringing, while reducing or eliminating ringing may worsen the frequency response.

Hence, to accurately detect regions with perceived ringing, two essential aspects need to be explicitly addressed:

1. An (strong) edge detector and
2. A masking model of the HVS.

Ringing Region Detection

The algorithm consists of two parts: 1) extraction of edges relevant for ringing, and 2) detection of visibility of ringing in the edge regions. The schematic overview of the proposed algorithm is illustrated in Fig.1

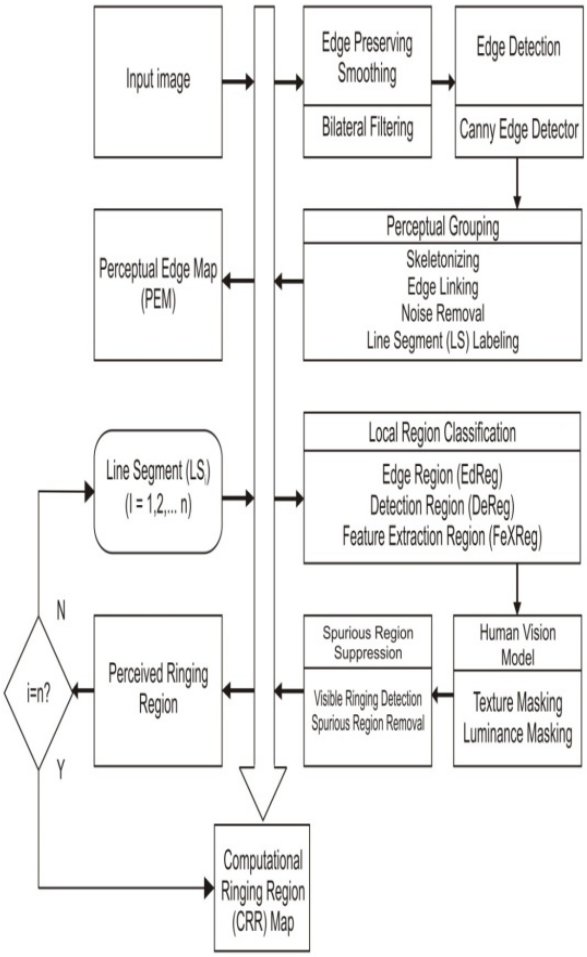

Figure 1: Schematic overview of the proposed algorithm, with at the top the part to detect edges relevant for ringing, and at the bottom the part to measure visibility of ringing around these edges

In the first part, an advanced edge detector is adopted, attempting to select the edges most relevant for ringing (i.e., contours of objects) in combination with the avoidance of the irrelevant edges (i.e., in textured areas).This results in a perceptual edge map (PEM), existing of a set of so-called line segments (LS). Edge detection is a fundamental tool used in most image processing applications to obtain information from the frames as a precursor step to feature extraction and object segmentation. This process detects outlines of an object and boundaries between objects and the background in the image. An edge-detection filter can also be used to improve the appearance of blurred or anti-aliased video streams. The basic edge-detection operator is a matrix area gradient operation that determines the level of variance between different pixels. The edge-detection operator is calculated by forming a matrix centered on a pixel chosen as the center of the matrix area. If the value of this matrix area is above a given threshold, then the middle pixel is classified as an edge.

Examples of gradient-based edge detectors are Roberts, Prewitt, and Sobel operators. All the gradient-based algorithms have kernel operators that calculate the strength of the slope in directions which are orthogonal to each other, commonly vertical and horizontal. Later, the contributions of the different components of the slopes are combined to give the total value of the edge strength. In the second part, each LS of the PEM is examined individually on the occurrence of visible ringing in its direct neighborhood, taking into account masking by the HVS. All regions with visible ringing are accumulated in a single binary map, which we refer to as the computational ringing region (CRR) map. Remind that the CRR map is used as input to the second step of the objective metric, in which the ringing artifact is reduced. Each part of the ringing region detection algorithm is further detailed in the following sections. Note that the entire metric is only based on the luminance channel of the images in order to further reduce the computational load.

In this paper, a perceptually more meaningful edge detection algorithm is proposed. To achieve this, propose the application of a canny edge detector [14] to an image, which first is nonlinearly smoothened. After some additional post-processing, this results in the PEM.

Bilateral filtering was introduced [15] as a simple and fast scheme for edge-preserving smoothing. It is a nonlinear operation that combines nearby image values based on both their geometric closeness and their photometric similarity, and prefers near values to distant values in both spatial domain and intensity range. In the Gaussian case, it can be expressed as

$$\mathrm{F}(\vec{x})\int_{-\infty}\int_{-\infty}\vec{I}(\vec{\varepsilon})\omega(\vec{\varepsilon},\vec{x})\mathrm{d}\vec{\varepsilon} \tag{1}$$

Where

$$\omega(\vec{\xi},\vec{x}) = \exp\left(\frac{-(\xi-\vec{x})^2}{2\sigma^2{}_d}\right)\exp\left(\frac{-(I(\vec{\xi})-I(\vec{x}))^2}{2\sigma^2{}_r}\right) \tag{2}$$

I and F denote the input and output images, and are space variables, and the standard deviations and characterize the domain and range filtering, respectively[20]. The advantage of using bilateral filtering instead of Gaussian filtering for the localization specific detection of perceptually strong edges is illustrated in Fig. 2. Subsequently, a Canny edge detector is applied to the bilaterally filtered image to obtain the perceptually more meaningful edges. After bilateral filtering [22], a Canny edge detector is applied to the image to yield an edge map. Since the original image is already filtered, the subsequent Canny algorithm is implemented without smoothing step, while keeping the other processing steps unchanged.

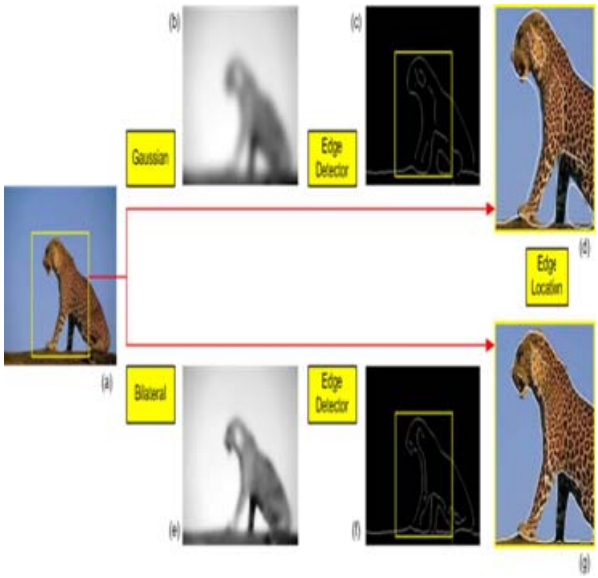

Figure 2: Bilateral filtering and Gaussian filtering for the detection of perceptually strong edges. (a) Original image. (b) Gaussian filtered image $\sigma_d = 15$, (c) Edge map of (b). (d) Superposition of (c) on (a). (e) Bilateral filtered image ($\sigma_d = 3, \sigma_r = 100$). (f) Edge map of (e). (g) Superposition of (f) on (a).

The high threshold in the Canny algorithm is set such that 85% of total pixels are cumulated in the magnitude histogram of the gradient image, and the low threshold is selected to be 0.4 of the computed high threshold[26]. These edge segments are extracted by (1) edge-linking: linking edge pixels into a set of edge segments of one pixel thick, each segment either containing two end-points or being a closed loop; and (2) noise removal: edge segments with the number of connected pixels below a certain threshold are discarded, which is done with the ringing detection accuracy and speed in mind [21].

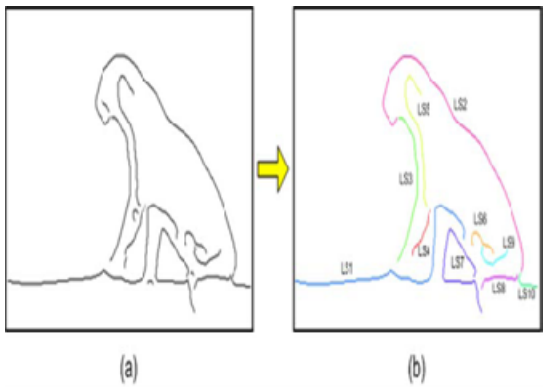

Figure 3: Construction of the perceptual edge map (PEM). (a) Canny edge map.(b)RelatedPEM with labeled line segments.

The following processing steps are implemented to define the LS in the PEM.

1. Skeletonizing: To guarantee that an edge is only one-pixel thick, a kernel of 4 × 4 pixels is slid over all pixels, and those pixel configurations that have a structure of [1 1; 0 1] or [1 0; 1 1] are replaced by [1 0; 0 1], and those with a structure of [1 1; 1 0] or [0 1; 1 1] are replaced by [0 1; 1 0].
2. Edge Linking: The algorithm links all the edge pixels into a set of elements; each element either contains two endpoints or is a closed loop. If an edge junction is encountered, the tracing procedure breaks, and a separate element is generated for each of the branches.
3. Noise Removal: The elements with the number of connected edge pixels below a certain threshold are discarded. This is done with the ringing detection accuracy and speed in mind [26].
4. Line Segment Labeling: The resulting elements of connected edge pixels are referred to as line segments (LS), and labeled.

Once this process is complete, we have the PEM [23]. Fig.3 illustrates the labeling of the LS in the PEM. Around the perceptually strong edges perceived ringing regions can now be located. Because of the properties of the underlying lossy compression scheme, ringing artifacts [28] spread out to a finite extent surrounding the edges. In addition, spatial masking as existing in the HVS, is highly relevant to the perception of ringing artifacts [32]. In this paper, detection regions are initially selected as all pixels around a detected edge segment, and then a model including luminance and texture masking is proposed to extract the perceived ringing regions. Each LS of the PEM is examined individually on the occurrence of visible ringing artifacts in their direct neighborhood, taking into account luminance and texture masking. The regions with visible ringing are then combined in a computational ringing region (CRR) map [29].

In order to characterize the visibility of ringing around LS [24], its surrounding is classified into three different zones. Assuming a single step edge with at its two adjacent sides smooth regions of a pixel intensity around the mid-gray level, the regions surrounding such an edge can be classified into (see Figure 4a): 1) Edge Region (EdReg): the original edge including the compression induced blur; 2) Detection Region (DeReg): the direct neighborhood of the EdReg, which potentially contains ringing artifacts; and 3) Feature Extraction Region (FeXReg): a region representative for the original local background, which is located outward from the corresponding DeReg. These regions are defined by thickening the LS with a different size for the structuring element of dilation/

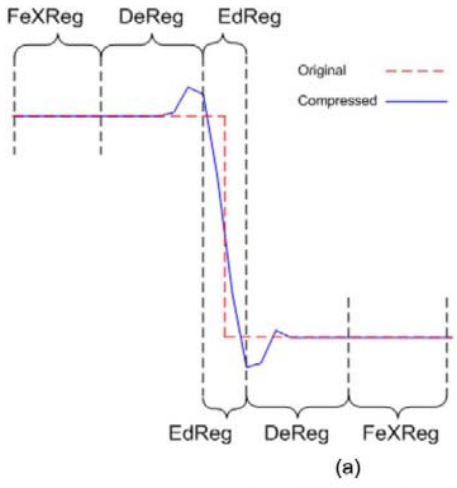

Figure 4: Illustration of local region classification. (a) Illustration of the three zones for a schematic step edge. (b) Illustration of how the zones are defined around an actual line segment as part of a natural image. In (b) the black line indicates the EdReg, the gray area defines the DeReg, and the white area refers to the FeXReg.

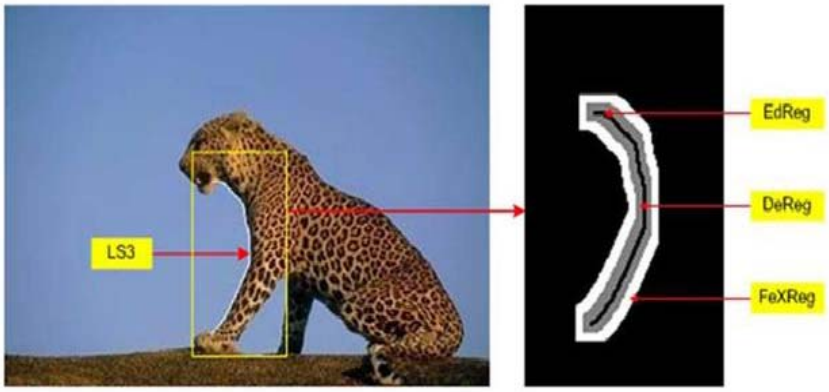

Figure 5: gives an example,in which for one LS the EdReg,DeReg and FeXReg obtained with a structuring element of 2,9 and 17 pixel width,respectively,is shown.

Ringing intrinsically appears around strong edges, though can be visually masked by image content. This is modeled by applying texture masking and luminance masking to each detected edge segment. As a result, invisible ringing regions are removed, and the retained regions of DeReg are considered as perceived ringing regions. Whether ringing is actually visible in the DeReg strongly depends (because of masking in the HVS) on the content of the original background, here represented by the FeXReg. Hence, the visibility of ringing is evaluated for each LS by applying a model for texture and luminance masking, using the texture and luminance

characteristics of the FeXReg[30]. As a result, DeReg regions, in which ringing is visually masked are eliminated, and only the perceptually prominent DeReg ringing regions remaining.

The visibility of ringing is significantly affected by the spatial activity in its local background [27], i.e., ringing is visually masked when located in a textured region, while it is perceptually prominent against a smooth background [13], [16]. In this paper, texture masking is modeled classifying the FeXReg of each LS into "smooth" and "textured" objects, depending on the local background characteristics. The DeReg is segmented accordingly,and those DeReg regions of which the corresponding FeXReg is clustered as "textured" are removed. This approach intrinsically avoids explicit modeling of the HVS, and formulates texture masking as a simple yet efficient local pixel clustering procedure [31]. The proposed scheme to implement this is illustrated in Fig. 5.5(b). It generally involves the following steps.

Calculating the local activity of the image content covered by the FeXReg by applying a global threshold to the gradient in pixel intensity to create a local binary map (LBM) of the FeXReg. This yields a profile of local pixel activities, and is formulated as in (3) and (4),

$$\text{LBM (i,j)} = \begin{cases} 0 \; LA\;(i,j) < Thr\_txt \\ 1 \; otherwise \end{cases}, i,j \in FeXReg \tag{3}$$

$$\begin{aligned}\text{LA (i,j)} = &|[\text{I(i-1,j-1)+2xI(i-1,j)+I(i-1,j+1)}]-[\text{I(i+1,j-1)+2xI(i+1,j)} + \text{(I(i+1,j+1)}]| \\ &+|[\text{I(i-1,j+1)+2xI(i,j+1)+I(i+1,j+1)}]-[\text{I(i-1,j-1)+2xI(i,j-1)+(I(i+1,j-1)}]|\end{aligned} \tag{4}$$

where the local activity LA(i,j) at location (i,j) is approximated by the gradient of the image intensity using a gradient operator (e.g., a Sobel operator). The 3×3 pseudoconvolution template used to calculate the gradient magnitude of a pixel at location(i,j) is shown in Fig.5 (I(i,j) corresponds to the pixel intensity at location(i,j)). The threshold Thr_txt is related to the magnitude histogram of the gradient image, and thus, image content dependent.

Dilating the LBM using a morphological operator, and labeling (e.g., by 8-connectivity) them into a set of connected components, which are referred to as texture components. This step intrinsically transfers pixel activities to a higher level structure of region activities, motivated by the fact that the human eye is not sensitive to variations at pixel level.

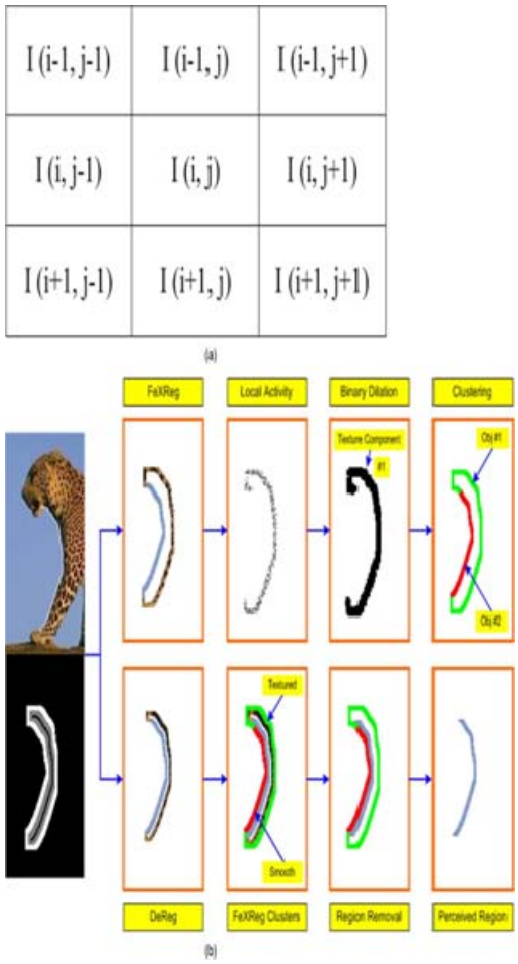

Figure 6. Implementation of texture

Removing the regions of DeReg that belong to the "texture objects" of FeXReg, since in these regions ringing is supposed to be masked by texture, and discarding the resulting regions of DeReg with their size under a certain threshold. The maintained regions of DeReg are considered as perceived ringing regions.

The visibility of variations in luminance depends on the local mean luminance. As a result, the visibility of ringing is largely reduced in extremely dark or bright surroundings. Here masking is implemented by simply calculating the local averaged luminance for each region of De-Reg remaining after the application of texture masking, and by subsequently removing regions, in which ringing is expected invisible due to luminance masking. For reasons of simplicity, the relationship between the region visibility (i.e.RV) and the local mean luminance (i.e. LML) is determined by two pre-defined threshold values[25]. This functional behavior as shown in Figure 4 is an approximation considered to be good enough (T_low=25 and T_high=220 in our experiments with 8bit gray-scale images). Ultimately, only the regions of DeReg that contain perceptually visible ringing artifacts remain. The

proposed human vision model results in a binary image, which we refer to as computational ringing region (CRR) map, indicating the detected perceived ringing regions for the corresponding image. The implementation of luminance masking is the same as for texture masking, but to guarantee efficiency, it is only applied to those regions of the DeReg remaining after the application of texture masking. The procedure for luminance masking is similarly formulated as a local pixel clustering model, and it mainly contains the following steps.

Calculating the local averaged luminance, over a 3×3 template, centered on each pixel that is part of a "smooth object" of the FeXReg

$$\mathrm{LML}(i,j) = 1/9 \sum_{k=i-1}^{i+1} \sum_{l=j-1}^{j+1} I(k,l) \tag{5}$$

Where $\mathrm{I}(i,j)$ denotes the pixel intensity at location (i,j) ,and $\mathrm{LML}(i,j)$ denotes the local mean luminance. The visibility of ringing due to luminance masking is determined according to the functional behavior shown in Fig. 9 [12],and a local binary map (LBM) is generated by applying a predefined threshold to the visibility coefficient (VC)

$$\mathrm{LBM}(\mathrm{i,j}) = \begin{cases} 0 & \mathrm{VC}(\mathrm{i,j}) > Thr\_lum \\ 1 & \text{otherwise} \end{cases} \tag{6}$$

Where $\mathrm{LBM}(i,j) = 0$ indicates a visible pixel location,and $\mathrm{LBM}(i,j) = 1$ indicates a nonvisible pixel location.This generates a profile of local visibility due to luminance masking.

Dilating the LBM to obtain a set of connected components, which are referred to as invisible components.

Classifying the "smooth objects" of FeXReg further into "visible objects" and "invisible objects" depending on the invisible components. This step combined with the one mentioned above intrinsically yields the structures of region visibility.

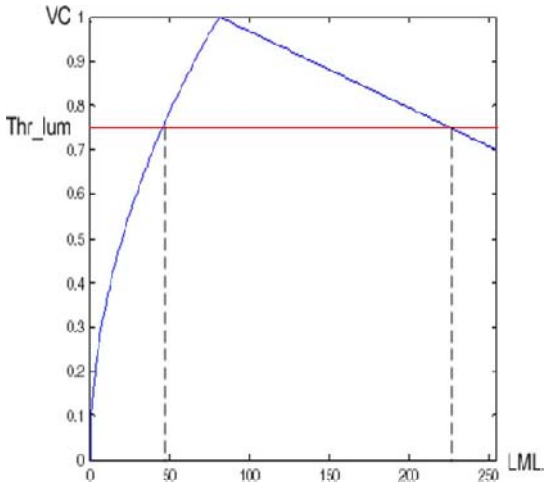

Figure 7: Implementation of luminance masking via the relation between the local mean luminance (LML) and the artifact visibility coefficient (VC). Thr_lum refers to the threshold used in the implementation.

Removing the DeReg that correspond to “invisible objects,”i.e., where ringing is not supposed to be visible against a very low or very high intensity background. Ultimately, only the regions of DeReg that yield visible ringing remain. These regions are combined in the CRR map,of which an example is given in Fig.4.7. The computational ringing region (CRR) map indicates the spatial location of perceived ringing, but it does not give any information yet on how annoying the ringing artifacts in these regions are.

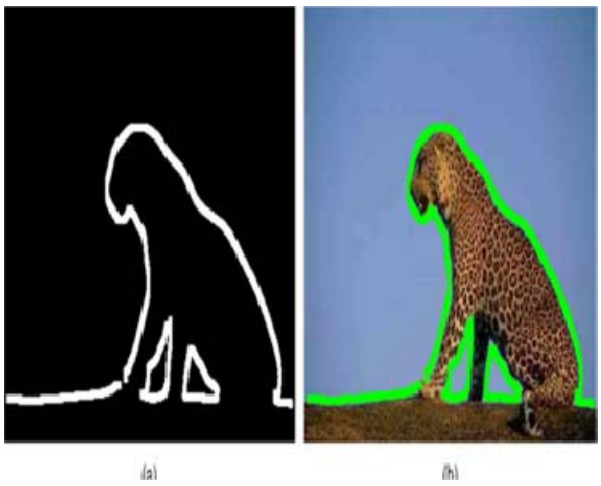

Figure 8. Example of (a) a computational ringing region (CRR) map corresponding to (b) a JPEG compressed image.

The resulting CRR map, however, still includes obvious spurious ringing regions, containing either “unimpaired” or “noisy” pixels misinterpreted as ringing pixels.

“Unimpaired pixels” indicate pixels in the detected regions of the CRR map, which are actually not impaired by ringing. An obvious example of the occurrence of “unimpaired” pixels is in an uncompressed image. The ringing region detection algorithm described so far will find the regions that might be impaired with visible ringing, independent of the compression level. But in an uncompressed image, these regions do not contain visible ringing, and hence, should be removed from the CRR map. Note that without removal of these regions the overall objective ringing metric including the step of quantification of ringing annoyance would not be less accurate, but less efficient.

“Noisy pixels” are pixels in the detected regions of the CRR map, that actually belong to an edge or texture. They are accidentally misclassified to a ringing region as a consequence of the dilation operation used in the human vision model.

To remove the spurious ringing regions, each detected ringing region (RR) is further examined by calculating its amount of visible ringing pixels. Those RRs with their number of visible ringing pixels below a certain threshold are considered as spurious,and consequently removed from the CRR map. The spurious ringing pixels are suppressed by applying two thresholds to the LV, a low threshold (Thr_v _low)

and a high threshold *(*Thr_v _high). Since unimpaired pixels exhibit no or very small intensity variance in their neighborhood, a pixel with its LV value below or equal to Thr _v_low is considered as an unimpaired pixel. In the same way, a pixel with its LV value above or equal to Thr _v_high is considered as a "noisy pixel." This can be formulated as

$$VC_n(i,j) = \begin{cases} 1 & \text{Thr\_}v\text{\_low} < LV(i,j) < Thr\_v\text{\_high} \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

Where

$$LV(i,j) = \frac{1}{9}\sum_{k=i-1}^{i+1}\sum_{t=j-1}^{j+1}\left[I(k,l) - \frac{1}{9}\sum_{k=i-1}^{i+1}\sum_{t=j-1}^{j+1} I(k,l)\right]^2, i, j \varepsilon RR_n \quad (8)$$

$$Thr\_\upsilon\_high = \alpha \,.\, MAX[LV(I,j)],\ i, j \in LS_n \quad (9)$$

where $VC_n(i,j)$ indicates the visibility of a ringing pixel at the ringing region (i.e., RR) with its associated line segment (i.e., LS), and indicates the local variance computed over a 3× 3 template, centered at a pixel intensity .The value of Thr_ v_ low is chosen to be zero, and the value of Thr _v _high is chosen to scale with the strength of corresponding edge. Thus, the ringing region RR is removed if

$$\frac{SUM(VC_n)}{SIZE(RR_n)} < R \quad (10)$$

Where SUM (VC_n) indicates the number of visible ringing pixels, SIZE (RR_n) indicates the size of the given RR, and indicates the predefined ratio of visible ringing pixels over the detected ringing region.

Performance Evaluation

The ringing region detection method is validated with respect to the results of the psychovisual experiment, and its performance is compared to existing alternatives in literature. For this performance comparison, implemented three ringing region detection algorithms recently proposed: 1) region clustering based ringing artifact measure (referred to as RCRM) 2) morphological filtering based ringing artifact measure (referred to as MFRM); and 3) no-reference ringing artifact measure (referred to as NRRM). In literature, all three methods are proved to be promising in terms of ringing region detection.

To evaluate the performance of various ringing region detection algorithms we compared the CRR map as calculated for each of the ringing region detection algorithms to the SRR map derived from the psychovisual experiment. These two binary images (i.e., the CRR and SRR map) were compared visually and via a quantitative correlation. For the visual assessment we produced a comparison map, which is an RGB color image generated by

$$\mathrm{Mc} = \begin{cases} \mathrm{Mc}(:,:,1) = \mathrm{M_{CRR}}\ [\mathrm{xor}(\mathrm{M_{CRR}}, \mathrm{M_{SRR}}] \\ \mathrm{Mc}(:,:,2) = \mathrm{M_{CRR}}\ \&\ \mathrm{M_{SRR}} \\ \mathrm{Mc}\ (:,:,3) = \mathrm{M_{SRR}}\ \&\ [\mathrm{xor}(\mathrm{M_{CRR}}, \mathrm{M_{SRR}})] \end{cases} \quad (11)$$

The G (green) channel is assigned to the logical operator AND of the two binary maps, and so, represents the correlated ringing regions. The R (red) and B (blue) channels are assigned to edges occurring only in the CRR map and the SRR map, respectively, and so, represent the uncorrelated ringing regions between both maps. Black regions represent the absence of visible ringing on both maps.

The objective comparison of the CRR map to the SRR map is quantitatively measured by two correlation coefficients, namely and , defined as follows:

$$\rho_1 = \frac{\Sigma[\mathrm{M_{CRR}}\ \&\ \mathrm{M_{SRR}}}{\Sigma \mathrm{M_{SRR}}} \quad (12)$$

$$\rho_1 = \frac{\Sigma\{\mathrm{M_{CRR}}\ \&[\mathrm{xor}\ (\mathrm{M_{CRR}}, \mathrm{M_{SRR}})]\}}{\Sigma[\sim \mathrm{M_{SRR}}]} \quad (13)$$

The numerator of ρ_1 indicates the total number of correlated pixels between the CRR map and SRR map, while the denominator indicates the size of the ringing regions in the SRR map. Thus, ρ_1 quantifies to what extent the subjective ringing regions are detected by the computational models. However, this coefficient by itself is obviously not enough to reflect the detection accuracy of a computational model. A model might be capable of capturing all subjective ringing regions, just by capturing all edges, also those that do not contain visible ringing. These falsely detected ringing regions consequently degrade particularly the efficiency of a subsequent ringing annoyance measurement. The degree of false detections is quantified by ρ_2 . Its numerator indicates the size of regions falsely detected by the computational models, and its denominator indicates the size of regions in the SRR map not detected by the human subjects. Evidently, a higher value of ρ_1 combined with a lower value of ρ_2 implies a good detection model.

The set of parameters includes the standard deviations (i.e., and) for the bilateral filter to control the extent of the smoothing effect, and the hysteresis thresholding (i.e., Thr_high and Thr_low) of the Canny edge detector to trace strong edges while preventing breaking of continuous edges. For the bilateral filter the selection of and has been intensively discussed for natural images, and they were set accordingly to and in our experiment. For the edge detector Canny sets the Thr_high such that a certain percentage of the total amount of pixels is cumulated in the magnitude histogram of the gradient image, and the Thr_low as a fixed fraction (i.e.,0.4) of the Thr_high. In implementation, used a relatively lowvalue of Thr_high in order to prevent loosing relevant edges. This may result in irrelevant LSs in the PEM, but these LSs are later discarded by applying the HVS model. In other words, the choice for the thresholds of the Canny edge detector affect the efficiency of the model rather than its accuracy. Finally, the threshold for the noise removal in the PEM formation was set to 20 pixels.

This set of parameters determines the width of the EdReg, DeReg, and FeXReg regions. The EdReg representing edge blur is chosen to be equal to the one-pixel thick

LS. In case this value is too small, blur pixels can easily be detected as spurious pixels in a ringing region. The width of the DeReg is set as a single-sided support dimension of four pixels, which approximates the maximal extent of ringing that spreads out to a region surrounding an edge in JPEG compression. The actual width of the DeReg may vary depending on the underlying properties of the coding technique. The width of the FeXReg is empirically selected to be the same as for the DeReg. The FeXReg may cross an object boundary or reach another edge, which consequently results in spurious pixels in a detected ringing region.

This set of parameters includes two essential thresholds, i.e., Thr_txt for texture masking and Thr_lum for luminance masking. The performance of our algorithm is fairly insensitive to variations of these thresholds within the range of [0.6, 0.95] and [0, 0.8] for Thr_txt and Thr_lum, respectively. Varying these thresholds within their respective range results in a variation of and over [85%, 95%] and [1%, 3%], respectively. For the final performance evaluation of this model, set Thr_txt=0.9 and Thr_lum=0.75.

This set of parameters contains three threshold values (i.e., Thr_ v_ low,Thr_v_high) to further eliminate undesired regions in the CRR map. It should be admitted that this processing step is a fine-tuned optimization to largely remove, for example, the "unimpaired regions" in the CRR map of an uncompressed (or high bitrate compressed)image. The parameters are determined as Thr_v_low=0,$\propto$= 0.5 and R=0.3. Thr_v_low and $\propto$ are set according to experiments and observations, while R is empirically chosen. R is mainly used to speed up the algorithm rather than to improve its accuracy. The inclusion of the detection of spurious ringing pixels hardly affects the overall performance of this model: including or omitting the detection of spurious ringing pixels corresponds to a deviation in and over a range of [-0.5%, +0.5 %] . It should, however, be noted that the concept of removing spurious ringing pixels is mainly important for the ringing annoyance estimation, and hence, these parameters might need to be calibrated again for the removal.

The output of the algorithm serves as input for the second step. In this respect it is relevant to realize that a good performance of the ringing region detection algorithm mainly contributes to the efficiency of the second step, rather than to the final accuracy of the prediction in ringing annoyance. So far, our algorithm is only tested for JPEG compressed image material. More research is needed to also evaluate its performance for different compression techniques. The algorithm is evaluated for two compression levels, and the corresponding CRR maps are highly comparable.

Results

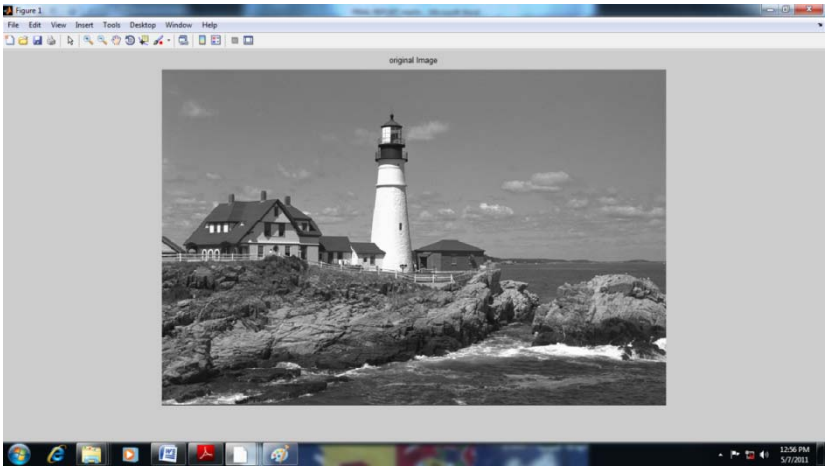

Figure 9: Original image.

The original image is shown in fig 9.It undergo compression i.e.,Jpeg compression results in coding artifacts (ringing).In the first part an advanced edge detector is adopted to select the edges most relevant for ringing in combination with the avoidance of irrelevant edges.After applying the edge detector to an image, use the gaussian filtering,It smoothens out noise and texture but also blur edges.The edge map of gaussian filtering is shown in the figure11.To overcome this introducing the bilateral filtering based on their similarity.The edge map of bilateral filtering is shown in the fig12.Then Canny edge detector is applied to the bilaterally filtered image,the detected edge pixels are combined into perceptual elements reffered as linesegment.To form the perceptual edge map the edge is skeletonizing,linking, noise removal and line segment labelling.

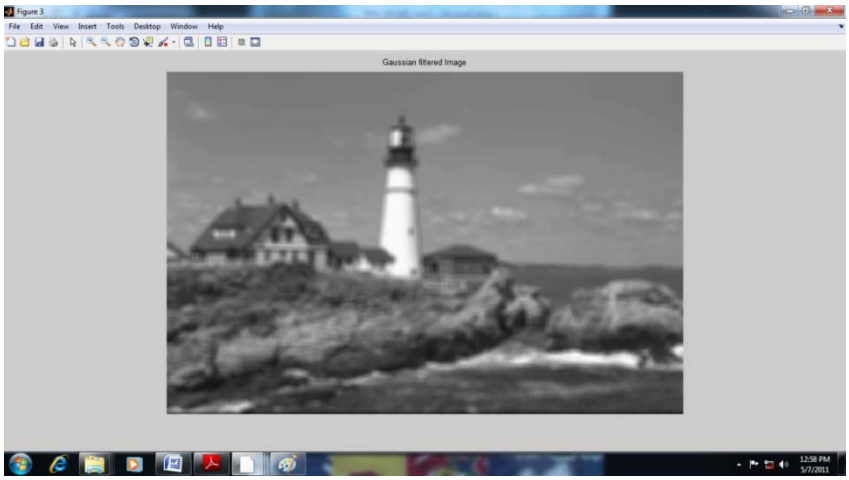

Figure 10: JPEG compressed image

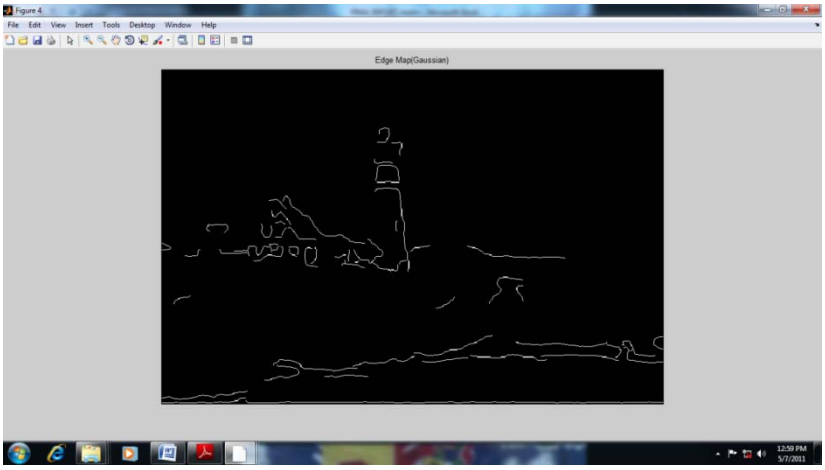

Figure 11: Edge map (Gaussian)

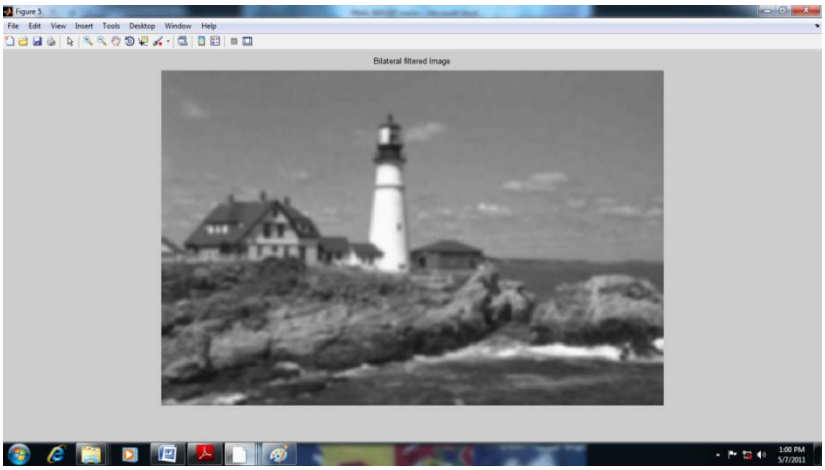

Figure 12: Gaussian filtered image

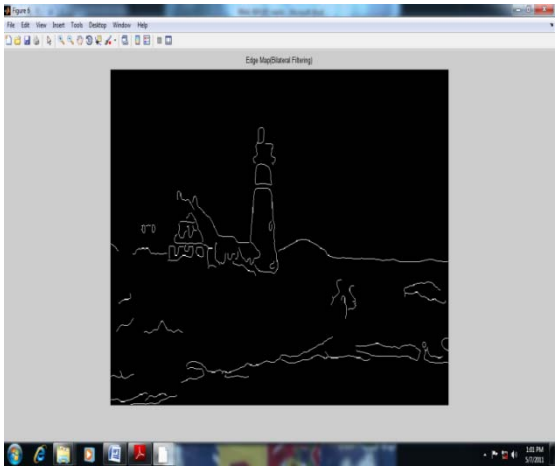

Figure 13: Bilateral filtered image

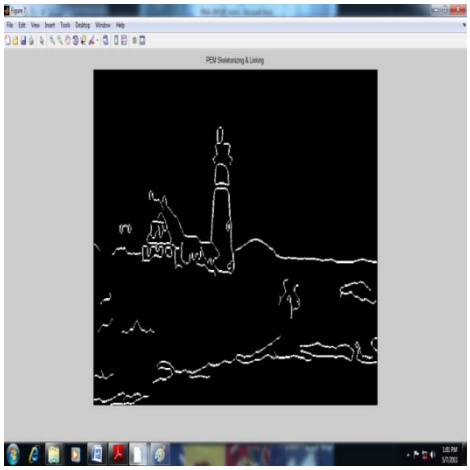

Figure 14: Edge Map (Bilateral)

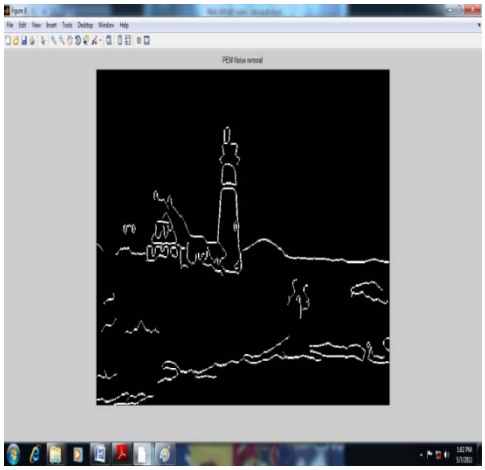

Figure 15: PEM Skeletonizing and linking

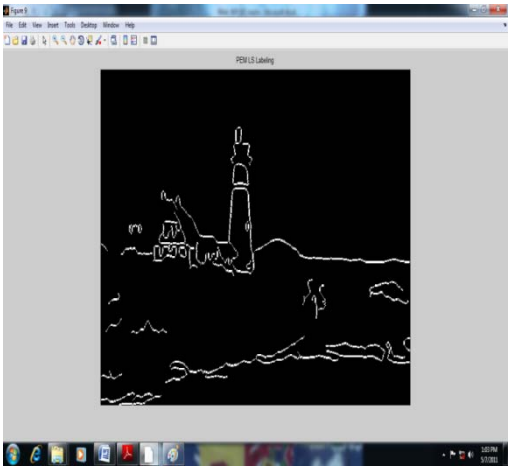

Figure 16: PEM Noise removals

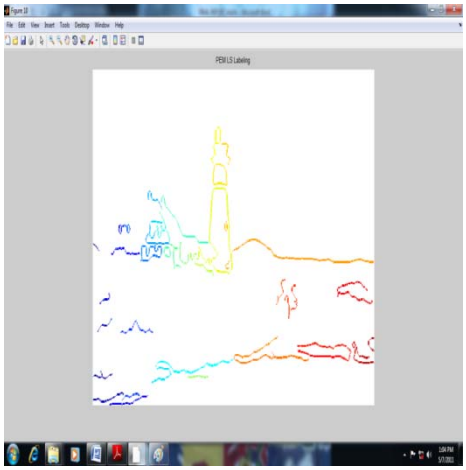

Figure 17: PEM LS Labelling

Figure 18: LBM

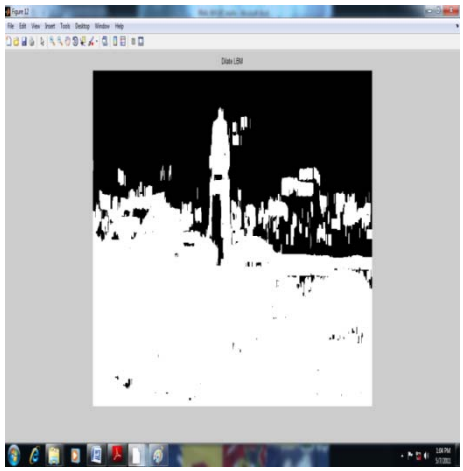

Figure 19: Dilate LBM

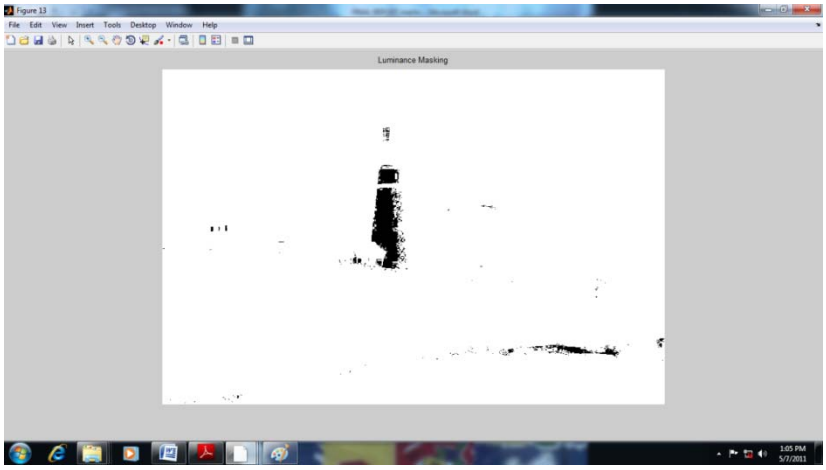

Figure 20: Luminance Masking

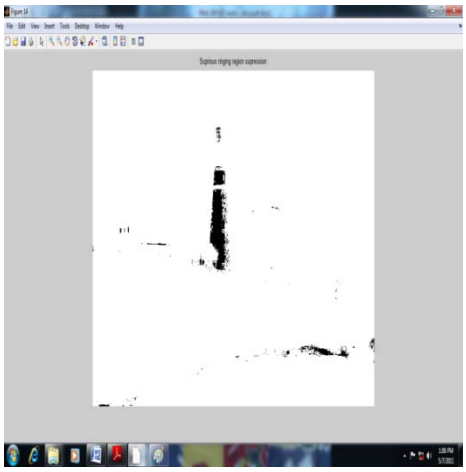

Figure 21: Spurious ringing region suppression

Figure 22: DeRing

Conclusion

A novel approach toward the detection of perceived ringing regions in compressed images is presented.The algorithm relies on the compressed image only,which is promising for its applicability in a real-time video chain,e.g.,to enhance the quality of artifacts impaired video.It adopts a perceptually more meaningful edge detection method for the purpose of ringing region location. This intrinsically avoids the drawback of applying an ordinary edge detector, which has the risk of omitting obvious ringing artifacts near nondetected edges or of increasing the computational cost by measuring ringing visibility near irrelevant edges. The objective detection in agreement with human visual perception of ringing artifacts is ensured by taking into account typical properties of the human visual system, such as texture masking and luminance masking. The human vision model is implemented, based on the local image characteristics around detected edges, to expose only the perceptually prominent ringing regions in an image. The proposed ringing region detection method is meanwhile extended by removing the ringing artifacts.

References

[1] Barland.R and Saadane.A, (2005), "Reference free quality metric for jpeg-2000 compressed images," in Proc. Int. Symp. Signal Processing and Its Applications, vol. 1, pp. 351–354

[2] Canny. J, (1986), "A computational approach to edge detection," IEEE Trans. Pattern Anal. Mach. Intell., vol. 8, no. 6, pp. 679–698.

[3] Farias. M. C. Q, Moore. M. S, Foley. J. M, and Mitra. S. K, (2004), "Perceptual contributions of blocking, blurring, and fuzzy impairments to overall annoyance," in Proc. SPIE, Human Vision and Electron. Imaging IX, vol. 5292, pp. 109–120.

[4] Feng.X and Allebach.J, (2006), "Measurement of ringing artifacts in jpeg images," in Proc. SPIE, vol. 6076, pp. 74–83.

[5] Franzen. R, Kodak Lossless True Color Image Suite

[6] Ghanbari.M, Standard Codecs, (2003): "Image Compression to Advanced Video Coding",London, U.K.: IEE Press.

[7] Guangtao Zhai, Weisi Jianfei Cai, Xiaokany Yang, Wejun Zhang, (2003), Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment Lin, " Efficient quadtree based block – shift filtering for deblocking and deringing".

[8] I.-R. R. , BT.500-11, (2002)"Methodology for the Subjective Assessment of the Quality of Television Pictures" Geneva, Switzerland: Int. Telecommun. Union.

[9] Kirenko. I. O, Muijs. R, and Shao. L, (2006), "Coding artifact reduction using non-reference block grid visibility measure," in Proc. IEEE Int. Conf. Multimedia and Expo pp. 469–472.

[10] Koh. C. C, Mitra. S. K, Foley. J. M, and Heynderickx. I, (2005) "Annoyance of individual artifacts in mpeg-2 compressed video and their relation to overall

annoyance," in Proc. SPIE, Human Vision and Electronic Imaging X, vol. 5666, pp. 595–606.

[11] Liu. H, Klomp. N, and Heynderickx. I, (2008), "Perceptually relevant ringing region detection method," in Proc. 16th Eur. Signal Processing Conf.

[12] Liu. H, Klomp. N, and Heynderickx. I, (2010), "A no-reference metric for perceived ringing artifacts in images," IEEE Trans. Circuits Syst. Video Technol., vol. 20, no. 4 to be published.

[13] Liu. H, Klomp. N, and Heynderickx.I, (2009) ,"A no-reference metric for perceived ringing," in Proc. 4th Int. Workshop on Video Processing and Quality Metrics for Consumer Electronics.

[14] Liu.C, Szeliski.R, Kang. S. B, Zitnick. C. L, and Freeman. W. T, (2008) "Automatic estimation and removal of noise from a single image," IEEE Trans. Pattern Anal. Machine Intell., vol. 30, no. 2, pp. 299–314.

[15] Liu.H. and Heynderickx.I,(2006), "A perceptually relevant no-reference blockiness metric based on local image characteristics," EURASIP J. Adv. Signal Process., vol. 2009, 2009. on Video Processing and Quality Metrics for Consumer Electronics.

[16] Luo. J, Chen. C, Parker. K, and Huang. T. S, (1996), "Artifact reduction in low bit rate dct-based image compression," IEEE Trans. Image Processing, vol. 5, pp. 1363–1368.

[17] Marziliano. P, Dufax. F, Winkler. S, and Ebrahimi. T,(2004) "Perceptual blur and ringing metrics: Application to jpeg2000," Signal Processing: Image Commun., vol. 19, pp. 163–172

[18] Muijs.R and Tegenbosch.J, (2006) "Quality-adaptive sharpness enhancement based on a no-reference blockiness metric," in Proc. 2nd Int. Workshop on Video Processing and Quality Metrics for Consumer Electronics,

[19] Oguz. S. H, Hu. Y. H, and Nguyen. T, (1998), "Image coding ringing artifact reduction using morphological post-filtering," in Proc. IEEE 2ndWorkshop on Multimedia Signal Processing, pp. 628–633.

[20] Oguz.S.H, "Morphological post-filtering of ringing and lost data concealment in generalized lapped orthogonal transform based image and video coding," Ph.D. dissertation, Univ.Wisconsin, Madison.

[21] Pappas. T. N and Safranek. R. J, (2000)," Perceptual Criteria for Image Quality Evaluation. Handbook of Image and Video Processing", New York: Academic.

[22] Paris. S and Durand. F, (2009) "A fast approximation of the bilateral filter using a signal processing approach," Int. J. Comput. Vis., vol. 81, pp. 24–52.

[23] Richardson. I,(2003) "H.264 and MPEG-4 Video Compression", Video Coding for Next-generation Multimedia. Hoboken, NJ: Wiley.

[24] Shen.M and Kuo.C.J,(1998), "Review of post processing techniques for compression artifact removal," J. Vis. Commun. Image Rep., vol. 9, no. 1, pp. 2–14.

[25] Sternberg. J, Cognitive Psychology, 4th ed. Florence, KY: Wadsworth, 2006.

[26] Tomasi. C and Manduchi. R,(1998), "Bilateral filtering for gray and color images", in Proc. IEEE Int. Conf. Computer Vision, pp. 836–846.

[27] Wandell. B. A, (1995), "Foundations of Vision. Sunderland", MA: Sinauer,.
[28] Wang. Z, Bovik. A. C, and Evans.B.L, (2000), "Blind measurement of blocking artifacts in images," in Proc. IEEE Int. Conf. Image Processing, vol. 3, pp. 981–984.
[29] Wang.Z and Bovik.A. C, (2006), "Modern Image Quality Assessment. Synthesis Lectures on Image. Video & Multimedia Processing"San Rafael, CA: Morgan and Claypool.
[30] Watson. A.B, (1993), Digital Image and Human Vision. Cambridge, MA: MIT Press,.
[31] Winkler.S, (2000) "Vision models and quality metrics for image processing applications," Ph.D.dissertation, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland.
[32] Xia. J, Shi. Y, Teunissen. K, and Heynderickx. I, (2009), "Perceivable artifacts in compressed video and their relation to video quality," Signal Process. Image Commun., vol. 24, no. 7, pp. 548–556.
[33] Yuen. M and Wu. H. R, (1998) "A survey of hybrid MC/DPCM/DCT video coding distortions," Signal Process., vol. 70, no. 3, pp. 247–278.
[34] Zon. K and Ali. W, (2001) "Automated video chain optimization," IEEE Trans. Consum. Electron., vol. 47, pp. 593–603.