PERCEPTUALLY RELEVANT RINGING REGION DETECTION METHOD

Hantao Liu [1]*, Nick Klomp* [1] *and Ingrid Heynderickx* [1, 2]

[1] Department of Mediamatics, Delft University of Technology
P. O. Box 5031, 2628 CD, Delft, The Netherlands
phone: + (31) 015 2784566, fax: + (31) 015 2787141, email: {Hantao.Liu, N.C.R.Klomp}@tudelft.nl
[2] Group Visual Experiences, Philips Research Laboratories
Prof. Holstlaan 4, 5656 AA, Eindhoven, The Netherlands
phone: + (31) 040 2742917, fax: + (31) 040 2744660, email: Ingrid.Heynderickx@philips.com

ABSTRACT

A novel approach towards automatic detection of perceived ringing regions is presented. 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. Determining visibility is based on luminance masking and texture masking as typical for the human visual system. The proposed detection method is validated by comparing its performance with the ringing regions resulting from a psychovisual experiment.

1. INTRODUCTION

In current visual communication systems, the receiving end, for example a TV-set, typically adopts various video enhancement algorithms to reduce compression artifacts, such as blocking, ringing and blur, so to improve overall quality [1]. In such a scenario, objective metrics, which determine the quality degradation caused by each individual artifact, and adapt the processing chain for artifact prioritization and reduction accordingly, are highly needed. In order to successfully measure individual coding artifacts, one must be able to identify where they occur in a given image. This implies that artifact detection is highly beneficial to an objective metric. Since the human visual system (HVS) is the ultimate assessor of most visual information, objective detection of artifacts in agreement with human visual perception can be assured by including properties of the HVS in the design of a detection algorithm.

Ringing is one of the most annoying coding artifacts introduced by lossy compression [1]. It results from the high frequency detail loss due to quantization, and manifests itself, in the spatial domain, as ripples or oscillations around high contrast edges [2]. The visibility of ringing is related to both the compression ratio and the image content [2-5]. Until recently, only a limited amount of research effort has been devoted to measure perceived ringing. The essential task behind the existing methods generally includes the following two steps: (1) a detection phase of regions, where ringing artifacts potentially occur; and (2) a ringing artifact metric, which provides a quantitative measure of ringing annoyance within the detected regions. The methods in [2] and [3] both simply assume that ringing appears unconditionally in regions surrounding strong edges. This, however, does not always reflect human visual perception of ringing, because of the absence of spatial masking as present in the HVS. The approach in [4] is based on the edge map of an image, in which the potential ringing regions are isolated using morphological techniques, and then only visually prominent regions according to HVS masking properties are retained. In [5], a global analysis is performed to classify the smooth regions in an image into objects. The visible ringing regions are then determined depending on the activity of the objects around edges (i.e. texture masking effect is included). The perceived ringing regions are then obtained, considering additionally HVS luminance masking.

From a practical point of view, it is highly desirable to reduce the complexity of a HVS based objective metric without compromising its performance. The computational complexity is reduced in [4] and [5] by applying the HVS model to the strong edges only. This yields accurate results only in case the relevant edges are detected. The approach described in [4, 5], however, detects strong edges based on the gradient magnitude only. Depending on the choice of the threshold, this has the risk of omitting obvious ringing regions near non-detected edges or of increasing the computational power by modelling the HVS near irrelevant edges. Therefore, we propose in this paper a perceptually more relevant edge detector. Additionally, the complexity of the HVS model is reduced with respect to the approaches used in [4] and [5]. In our approach an efficient local analysis based on luminance and texture masking of the HVS is applied to each detected edge segment. The entire procedure is built upon the luminance component only in order to further reduce the computational load. To evaluate this approach, a psychovisual experiment to obtain perceived ringing regions in images is carried out. It is checked whether the ringing regions predicted by the computational model are in agreement with the perceived regions.

2. PERCEPTUAL EDGE EXTRACTION

Existing methods for ringing detection [2-5] adopt an ordinary edge detection algorithm (e.g. Sobel), in which the significance of an edge is determined by simply applying a threshold to its gradient magnitude. This threshold-based approach neglects the spatial information, and so, does not fully reflect the way human beings perceive edges. As a result, perceptually salient edges e.g. at object boundaries may be discarded, while texture edges remain. This may heavily degrade the detection accuracy of ringing artifacts. In this paper, a perceptually more meaningful edge detection algorithm is proposed.

2.1 Bilateral Filtering

It is known that for natural images the human visual system tends to respond to differences between homogeneous regions rather than to structure within these homogeneous regions [6]. In this paper, finding the perceptually meaningful edges is based on this observation: texture existing in homogeneous regions is neglected as if viewed from a long distance. This is implemented by smoothing the original image until textural details are sufficiently reduced so that only perceptually relevant edges remain (i.e. the parameters used are given below). The subsequent application of an edge detector then allows obtaining the perceptually more meaningful edges. Gaussian filtering can be used to smooth out image noise and details. However, it also blurs edges, and consequently, changes their location in the resulting edge map. Since ringing detection intrinsically requires precise localization of edges, edge-preserving smoothing is needed. Bilateral filtering has been recently proposed as a simpler and faster alternative to anisotropic diffusion for edge-preserved filtering [7]. Bilateral filtering is a nonlinear filter that combines gray levels based on both their geometric closeness and their photometric similarity. In the Gaussian case, it can be expressed as:

$$\vec{F}(\vec{x}) = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \vec{I}(\vec{\xi})\,\omega(\vec{\xi},\vec{x})\,d\vec{\xi}}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \omega(\vec{\xi},\vec{x})\,d\vec{\xi}} \tag{1}$$

where

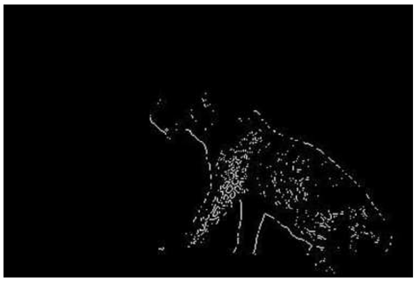
$$\omega(\vec{\xi},\vec{x}) = \exp\left(\frac{-(\vec{\xi}-\vec{x})^2}{2\sigma_d^2}\right)\exp\left(\frac{-(I(\vec{\xi})-I(\vec{x}))^2}{2\sigma_r^2}\right) \tag{2}$$

$\vec{I}$ and $\vec{F}$ denote the input and output images, $\vec{x}$ and $\vec{\xi}$ are space variables, and the parameters σ_d and σ_r (σ_d =3 and σ_r =100 in our experiments) characterize the domain and range filtering, respectively.

2.2 Perceptual Edge Extraction

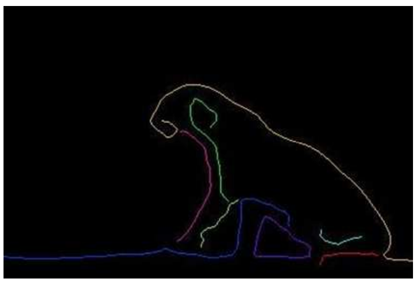
After bilateral filtering, a Canny edge detector [8] is applied to the image $\vec{F}$ to yield an edge map. Since the original image is already filtered, the subsequent Canny algorithm is implemented without smoothing step, while keeping the other processing steps unchanged. The high threshold in the 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. This implies that the thresholds are image content dependent. From the detected edge pixels perceptually meaningful edge segments are constructed. They are defined as an element of connected edge pixels, and are used as the basis for ringing detection. 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. An example of the proposed edge detection is illustrated in Figure 1. The extracted edge segments are shown in Figure 1d with random colors in the resulting edge map. As shown in Figure 1, our approach is able to capture strong edges in a way that is tuned to human perception, as compared to the result of an ordinary edge detector.

(a) Original image



(b) Edge map of (a) using Sobel and thresholding

(c) Filtered image of (a) using bilateral filtering



(d) Perceptual edge segments extracted from (c)

Figure 1 – Illustration of the perceptual edge detection

3. RINGING REGION DETECTION

Around the perceptually strong edges perceived ringing regions can now be located. Because of the properties of the underlying lossy compression scheme, ringing artifacts spread out to a finite extent surrounding the edges [5]. In addition, spatial masking as existing in the HVS, is highly relevant to the perception of ringing artifacts. 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.

3.1 Local Region Classification

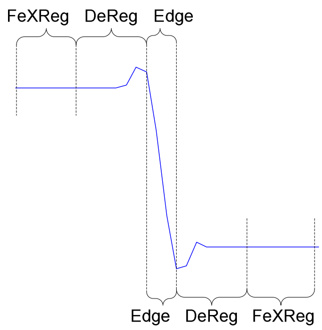
Assuming a single step edge with at its two adjacent sides smooth regions of a pixel intensity around the mid-gray level, the regions surrounding such an edge can be classified into (see Figure 2a): (1) a Detection Region (i.e. DeReg) which is close to the edge and potentially contains perceived ringing artifacts; and (2) a Feature Extraction Region (i.e. FeXReg), which is located outwards from the corresponding DeReg, indicating the local background taken into account for the possible visibility of ringing. This region classification is in line with the physical structure of the ringing artifact, and can be implemented, in the spatial image domain, using morphological operators, such as dilation over each detected edge segment. The size of the dilation operator is linearly scaled with the image size. Figure 2b shows the proposed local region classification around a detected edge segment in a real JPEG compressed image of 256x384 pixels. In this particular case, the width of the dilation operator was selected to be 9 and 17 pixels for DeReg and FeXReg, respectively.

3.2 The Human Vision Model

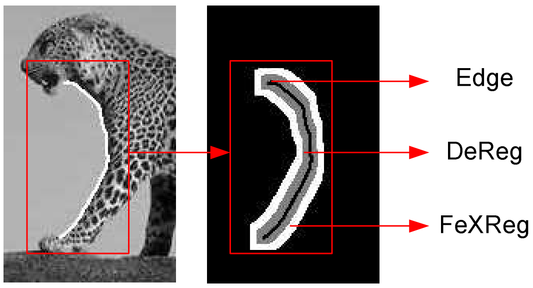
Ringing intrinsically appears around strong edges, though can be visually masked by image content. This is modelled by applying texture masking and luminance masking to each detected edge segment. As a result, invisible ringing regions are removed, and the retained regions of DeReg are considered as perceived ringing regions.

3.2.1 Texture Masking

The visibility of ringing is significantly affected by the spatial activity in the local background, i.e. ringing is masked when located in a textured region, while it is most visible against a smooth background [4, 5]. Texture masking is modeled classifying the FeXReg of each detected edge segment into "smooth" and "textured" parts. The DeReg is segmented accordingly, and only the regions of which the corresponding FeXReg is clustered as "smooth" are retained. The proposed scheme to implement this is illustrated in Figure 3. It generally involves the following steps: (1) calculating the gradient (e.g. using the Sobel operator) of image intensity along the pixels in the FeXReg, and applying a threshold to create a local edge map, indicating the activity (i.e. smooth or textured) within the FeXReg; (2) clustering the local edge points using morphological dilation and pixel connectivity (8-connectivity in our experiments), resulting in a set of connected components, which are referred to as texture objects; the noisy objects (e.g. according to their size and mean activity) are removed; (3) selecting the textured regions of FeXReg as being those that belong to a texture object; the remaining part of FeXReg is considered smooth; and (4) removing the corresponding texture regions in DeReg. Hence, the remaining regions of DeReg are only smooth regions around the detected strong edges, where ringing will be visually prominent.



(a) Region classification around a step edge



(b) Region classification around an edge segment detected in a JPEG compressed image

Figure 2 – Illustration of the local region classification

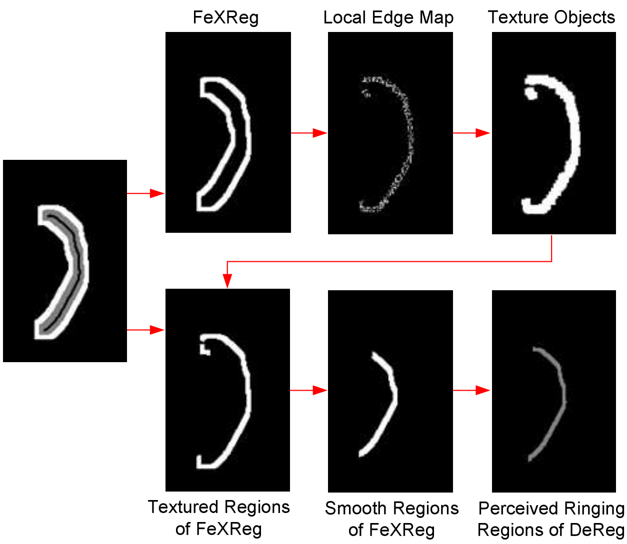


Figure 3 – Implementation of the texture masking

3.2.2 Luminance Masking

It is demonstrated in psychovisual experiments that the HVS sensitivity to variations in luminance depends on the local mean luminance [9]. The visibility of ringing is largely reduced in an extremely dark or bright surrounding, while the distortion is observed most easily on a background with an averaged luminance value in the mid-gray levels [9]. In this paper, luminance masking is implemented by simply calculating the local averaged luminance for each region of DeReg 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. This functional behaviour 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.

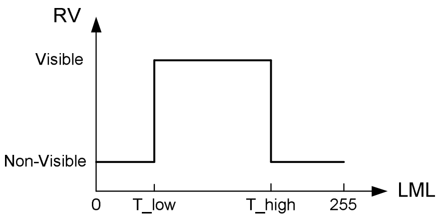


Figure 4 – Implementation of the luminance masking

4. SUBJECTIVE RINGING REGION

In order to evaluate the reliability of the detection algorithm for ringing regions, a subjective experiment is performed. For this experiment, visible ringing regions in a color image are marked by human subjects, and transformed into a subjective ringing region (SRR) map.

4.1 Subjective Experiment

The subjective experiment was carried out with a set of eight source images taken from the Kodak Image Suite [10]. They were full color images of size 512x768 (height x width) as shown in Figure 5. Ringing was induced by compressing these images using JPEG at two different compression ratios (CR=25 and 50). This yielded a set of 16 test stimuli. Eight subjects (Master students from the Delft University of Technology, 6 male and 2 female) participated in the experiment. A training session was conducted to make the subjects acquainted with ringing, and to teach them to distinguish ringing from other types of coding artifacts. The test images were presented on a 17 inch liyama Pro Lite E431S monitor in random order to each subject in a separate session. The subjects were asked to mark the regions in the image, in which they perceived ringing. The spatial location of the markings was recorded for each image and each subject in real time.

4.2 Subjective Ringing Region Map

The recorded data of the subjective experiment were transformed to a set of binary maps, where a white area indicated the marked region (i.e. the perceived ringing region) and a black area referred to absence of visible ringing. This resulted for each stimulus in an individual ringing region (IRR) map (i.e. per subject). We then computed a mean ringing region (MRR) map (i.e. M_{MRR}) as:

$$M_{MRR} = \frac{1}{n}\sum_{s=1}^{n} M_{IRR}(s) \tag{3}$$

where $M_{IRR}(s)$ denotes the IRR map for subject s (s=1..n), and n denotes the total number of subjects.

The subjective ringing region (SRR) map (i.e. M_{SRR}) was then derived using a threshold Thr (Thr=1/3 in our experiments) to the MRR map. Thus:

$$M_{SRR}(i,j) = \begin{cases} 0 & \text{if } M_{SRR}(i,j) < Thr \\ 1 & \text{otherwise} \end{cases} \tag{4}$$

where $M_{SRR}(i, j)$ denotes the intensity value of a pixel at location (i, j) in an image. In our case with a Thr = 1/3, it means that the SRR map contains the ringing regions in the MRR map where more than a third of the subjects perceived ringing.

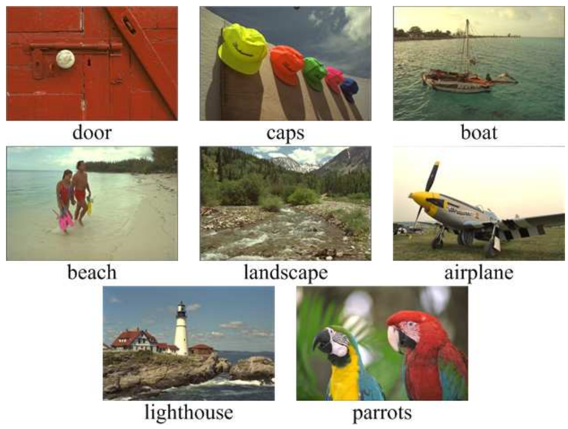


Figure 5 – Source images

5. PERFORMANCE EVALUATION

To evaluate the performance of our proposed approach the CRR map predicted by the ringing region detection algorithm is compared to the SRR map derived from the subjective experiment. To get a first impression the correlation between the subjective map and the computational map is here only evaluated visually [10]. This means that the similarity of both maps is visually represented as e.g. in Figure 6d. It shows a comparison map, which is an RGB color image with its blue channel assigned to the CRR map and its red channel assigned to the SRR map. On the color image: (1) black regions represent the absence of visible ringing on both maps, (2) red or blue regions indicate the uncorrelated ringing regions, and (3) magenta regions indicate the correlated ringing regions of both maps. Figure 6 shows as an example the experimental results for the test images “lighthouse” (CR=25), “beach” (CR=25) and “landscape” (CR=50).

In these cases the ringing region detection of the computational model and the subjective results from the psychovisual experiment exhibit a satisfactory correlation. The model is able to locate almost all ringing regions that are perceived in the experiment. But, it is also clear in all three examples that the model additionally detects ringing regions that are not observed subjectively. In other words, the model is still more sensitive to ringing regions than the participants of the experiment. This is not surprising, since our detection method so far only exposes regions which are likely to be impaired by ringing artifacts. Further analysis of the activity in these regions is expected to eliminate spurious regions, and to address the impact of compression ratio.

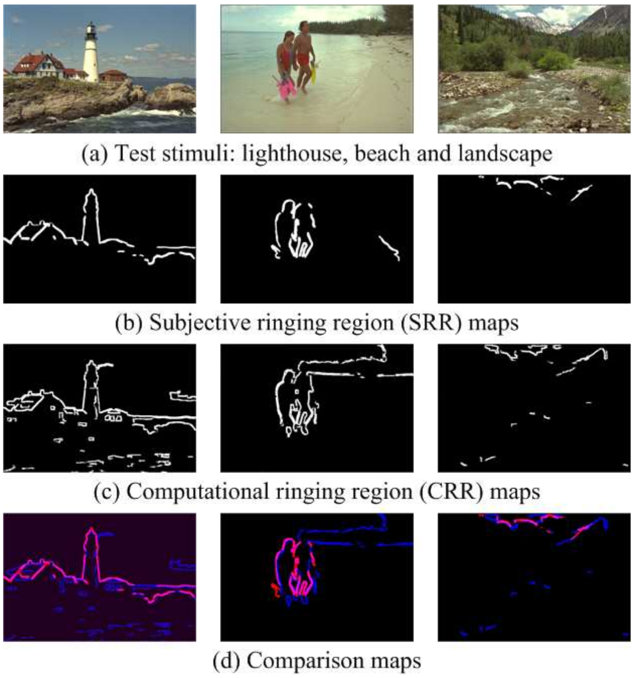


Figure 6 – Illustration of the performance evaluation

For the test image “lighthouse” the CRR map captures most of the visible ringing regions around high contrast edges, which is in agreement with the experimental results. This is mainly due to the proposed edge extraction method, which preserves only perceptually relevant edges for subsequent ringing region detection. The use of an ordinary edge detector (as in [2-5]) may fail in this case, depending on the threshold used; for a high threshold some visually salient edges may not be detected, while for a low threshold canto many edges may be preserved, such that applying the HVS model becomes computationally very expensive. Figure 6 also shows good results for the test image “landscape”. This suggests that the model is well able to predict perceived ringing regions in a highly textured image too. This is mainly achieved by including the spatial masking properties of the HVS.

The limited number of subjects and test images does not allow yet full assessment of the correlation between the computational and perceived ringing region maps. A more reliable subjective map is needed for an accurate comparison; nonetheless, the preliminary experimental results tend to validate the proposed ringing region detection algorithm.

6. CONCLUSIONS

A novel ringing region detection algorithm is presented. The proposed method includes a detector for perceptually meaningful edges as well as spatial masking properties of the HVS, such as luminance and texture masking. The performance of the algorithm is validated by comparing it to subjective results from a psychovisual experiment. The preliminary results show a strong correlation. However, more subjective data are needed in order to further substantiate the parameters used, to accurately assess the performance of the proposed model, and to compare it with alternatives existing in literature. We expect that the application of our detection algorithm in a ringing metric will result in a promising performance as well.

REFERENCES

[1] C.C. Koh, S.K. Mitra, J.M. Foley and I. Heynderickx, “Annoyance of Individual Artifacts in MPEG-2 Compressed Video and Their Relation to Overall Annoyance,” Proc. SPIE, vol. 5666, pp. 595-606, Jan. 2005.

[2] P. Marziliano, F. Dufax, S. Winkler and T. Ebrahimi, “Perceptual blur and ringing metrics: Application to JPEG2000,” Signal Processing: Image Communication, vol. 19, pp. 163-172, 2004.

[3] R. Barland and A. Saadane, “Reference Free Quality Metric for JPEG-2000 Compressed Images,” Proc. ISSPA, vol. 1, pp. 351-354, August 2005.

[4] S.H. Oguz, Y.H. Hu and T.Q. Nguyen, “Image Coding Ringing Artifact Reduction Using Morphological Post-filtering,” Proc. IEEE MMSP, pp. 628-633, 1998.

[5] X. Feng and J.P. Allebach, “Measurement of Ringing Artifacts in JPEG Images,” Proc. SPIE, vol. 6076, pp. 74-83, Feb. 2006.

[6] R. Lachman, J. Lachman and E. C. Butterfield, “Cognitive Psychology and Information Processing: An Introduction,” The American Journal of Psychology, vol. 92, no. 4, Dec. 1979.

[7] C. Tomasi and R. Manduchi, “Bilateral filtering for gray and color images,” Proc. ICCV, India, Jan. 1998.

[8] J. Canny, “A Computational Approach to Edge Detection,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 8, no. 6, pp. 679-698, Nov. 1986.

[9] T. N. Pappas and R. J. Safranek, “Perceptual criteria for image quality evaluation,” Handbook of Image and Video Processing, Academic Press, San Diego, May 2000.

[10] Kodak Lossless True Color Image Suite http://www.r0k.us/graphics/kodak/

[11] N. Ouerhani, R. Wartburg, H. Hügli and R. Müri, “Empirical Validation of the Saliency-based Model of Visual Attention,” ELCVIA, no. 1, pp. 13-24, 2004.