The Simulation of the Human Visual System Model for Image Quality Evaluation

Jiří Kulhánek

Czech Technical University in Prague, Faculty of Electrical Engineering
Dept. of Radioelectronics, Technická 2, Prague, Czech Republic

Abstract. Image quality of the compressed pictures can be evaluated by objective testing method based on the lower level human visual system model. The simulation of the HVS model presented here is used either to show the region in a compressed picture where the distortion is visible or can produce a single image quality number which correlates well with subjective image quality evaluation. Sets of the pictures with various compression ratios and five methods of compression were tested by this model. Experimental results of objective testing are then compared with DSCQS subjective image quality evaluation method.

1 Introduction

Digital lossy compression of the picture is used in many multimedia applications. We used it because of reducing bit rate and storage capacity. It is important take into account image quality. The level of impairments can be specified either by an objective measure such as signal-to-noise ratio (SNR) or by a subjective measure such as mean opinion score (MOS). Still widely used traditional objective image quality measurements such as peak signal-to-noise ratio (PSNR) and mean square error (MSE) generally do not correlate well with viewers opinion (MOS) obtained from subjective image quality evaluation. On the other hand, subjective measurement [1] with human subjects is commonly used to obtain an accurate assessment of image quality. These subjective measurements have number of disadvantages: time consumption, specialized laboratories requirements, they are expensive and need a large number of subjects to obtain the required accuracy.

These problems have resulted in an extensive research into objective image quality metrics based on the human visual system modelling which correlate well with human perception. The development of objective metrics for assessing the quality of compressed images is currently an area of research. Such metrics are required to accurately and repeatably determine visual effects of the various impairments introduced by lossy digital compression algorithm. Human vision models based on early stages of vision show promise and have been successful in assessing the fidelity of an image. This is particularly useful in high quality applications where the small distortions are visible [2]. Simple vision models may not be powerful enough to predict picture quality in highly compressed images [2,3]. HVS based metrics for monochrome images presented in this paper specially take into account structure of the image and can accurately evaluated highly compressed pictures.

2 The Model of the Human Visual System

Figure 1. shows a block diagram of the HVS model derived from [2,3,7,10]. Matlab environments especially Image processing toolbox is used for the simulation of the HVS model. Original and compressed (distorted) images (in the luminance domain) are inputs of the model.

At first, both Fourier transformed images are filtered in the frequency domain by octave band-pass cosine logarithmic filter (Fig.2) [7] into frequency channels k (channel decomposition block).

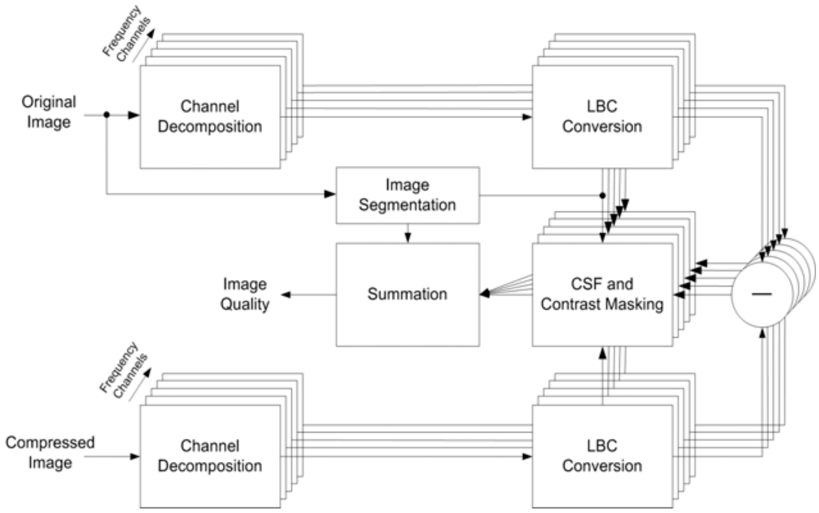

Fig. 1. Block diagram of the HVS model

A cosine log filter of 1-octave bandwidth centered at frequency 2^i cycles/picture is expressed as

$$G_i = \frac{1}{2}\left[1 + \cos\left(\pi \log_2 r - \pi i\right)\right] \qquad (1)$$

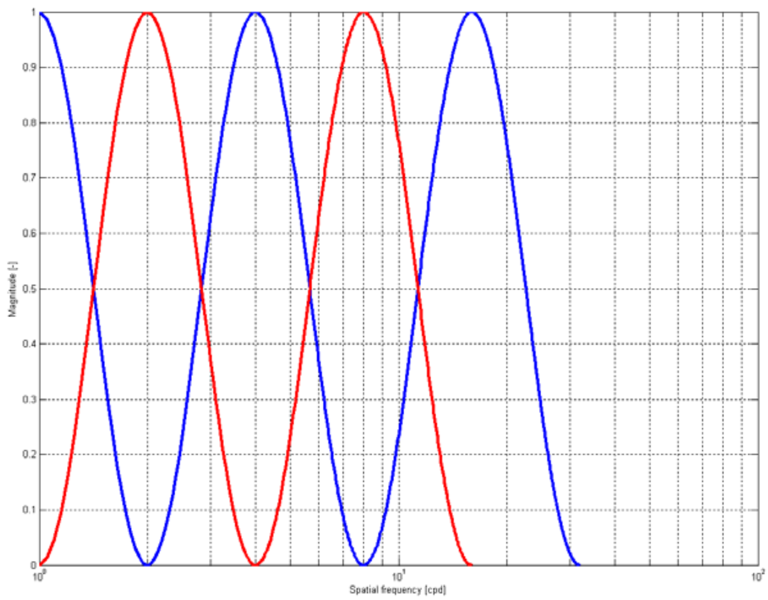

Fig.2. Filter bank of 1-octave-wide cosine log filters . Note also the symmetry of the cosine log filters on a logarithmic scale.

where $r = \sqrt{x^2 + y^2}$ represents the radial spatial frequency.

Secondly, each spatial frequency band for both images is then transformed back into the spatial domain via an inverse Fourier transform (IDFT).

In the next step, in LBC conversion block the luminance values for original and distorted images are converted to band-limited contrast (LBC) using Peli's LBC algorithm [7] :

$$LBC_k(x, y) = \frac{a_k(x, y)}{l_k(x, y)} \qquad (2)$$

where $LBC_k(x, y)$ represents local contrast in particular frequency band k, $a_k(x, y)$ is the band-pass filtered image and $l_k(x, y)$ is the local luminance mean (low–pass filtered image).

The errors between the two signals in each channel are calculated and weighted by a Contrast Sensitivity Function (CSF). The sensitivity of the human vision to contrast (CSF=1/Contrast threshold) is different in various spatial frequencies (frequency bands) as shows Fig.3. It is necessary set up contrast detection threshold (from CSF) for respective frequency band k.

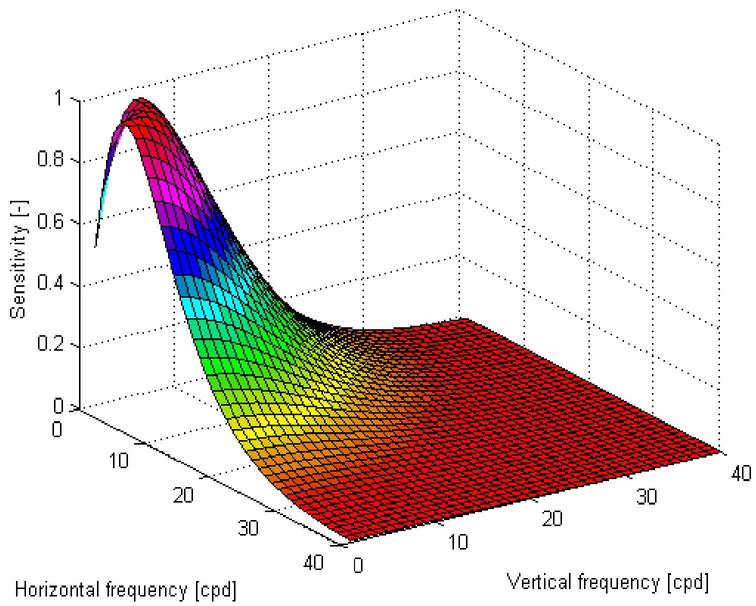

Fig. 3. Typical shape of contrast sensitivity function [8] for Luminance channel approximated from detection threshold experiment [9]. Points below the CSF are visible to the observer (those are the values that have even higher contrasts than the threshold level).

The weighted error signals are then adjusted by a visual masking effect model, which reflects the reduced visibility of errors presented on the background signal (stimulus that is visible by itself cannot be detected due to presence of another). This important phenomenon in vision is called spatial masking. The masking effect is performed by image segmentation. Image segmentation block uses the Sobel edge detector. The strategy in these segmentation algorithms is to classify the luminance component of each picture into three mutually exclusive context: plane, edge and texture regions. Spatial masking elevates contrast threshold level more for textured regions than along edges. This is modeled as a contrast threshold elevation :

$$CT_{new} = \begin{cases} CT_{base} & if\ CT_{mask} > CT_{base} \\ CT_{base}\left(\dfrac{CT_{mask}}{CT_{base}}\right)^{\varepsilon} & otherwise \end{cases} \quad (3)$$

where CT_{new} is contrast threshold in the presence of a masker, CT_{base} is the base threshold for the frequency band k (from CSF) , CT_{mask} is contrast of the masker (original image), $\varepsilon = 0.7$ for edge areas, and $\varepsilon = 1$ for textured areas.

New contrast thresholds from (2) are used to determine for each frequency band k and at each point of image, whether the errors between the original and coded images are visible.

Finally, probability Minkowski summation [11] computed over all channels is used for the determination of visibility errors at each pixel in JND (Just Noticeable Distortions). This produces Distortion Map (DM). The summation of the errors in DM at each pixel of image gives single value Image Quality (IQ). IQ represents overall picture quality of compressed image.

3 Experimental Results

The results of using the model to predict a single valued IQ for JPEG2000 compressed images are shown in Figure 4. Five images with widely varying characteristics were used: Portrait: picture of two women faces with varying contents (some texture, edges and flat regions), Fruit: close-up image of a fruit (mainly flat regions with some edges; not much texture), Square, Posters and Garden with very varying texture, edge and flat areas.

Each picture was compressed by JPEG (DCT), JPEG2000, Fractal and Wavelet coders with various compression ratios.

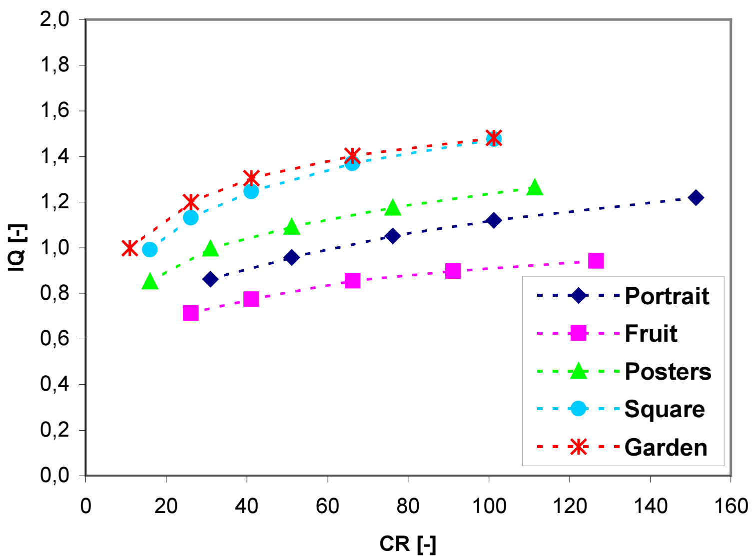

Fig. 4. The output of the model for five JPEG2000 compressed images at different compression ratios. Higher IQ values correspond with higher image distortions. IQ=0 is for original image.

4 Conclusion

Simulation of the model in a Matlab environment is very convenient and operative for work with digital images. The testing results verified the model ability to accurately predict an image quality of compressed pictures. Experiments show that for the highly compressed pictures the model correlates better with human judgements than the conventional PSNR or MSE metrics. Comparison with subjective DSCQS results gives linear correlation coefficient 0,894 for JPEG2000 compression method. For other compression methods, they are a little bit worse. For correct image quality evaluation is very important set up model input parameters from the subjective testing (viewing distance, maximum luminance and gamma of CRT).

5 Acknowledgement

This work has been supported by grant No. 102/02/0133 "Qualitative Aspects of Image Compression Methods in Multimedia Systems" of the Grant Agency of Czech Republic.

6 References

[1] Rec.BT.500-10 (03/00) : Methodology for the Subjective Assessment of the Quality of Television Pictures.

[2] S. Daly, The Visible Difference Predictor: An Algorithm for the Assessment of Image Fidelity. Digital Images and Human Vision, pages 179–206, 1993.

[3] E. Peli, Vision Models for Target Detection and Recognition.World Scientific Publishing, 1995.

[4] B.A.Wandell, Foundations of Vision, Sunderland MA : Sinauer Associates Inc., 1995.

[5] M.A.Webster, E.Miyahara, Contrast Adaption and the Spatial Structure of Natural Images, J. Opt. Soc. Am. A, 14(9): 2355-2366, Sep. 1997

[6] A.B. Watson, R. Borthwick and M.Taylor, Image Quality and Entropy Masking, Proceedings SPIE 3016, San Jose, pp. 330-337, February 1996

[7] E.Peli, Contrast in Complex Images, J.Opt.Soc.Am. A, Vol.7, No 1, pp. 1-14, 1993.

[8] H. Rushmeier, G. Ward, C. Piatko, P. Sanders, B. Rust, ``Comparing Real and Synthetic Images: Some Ideas About Metrics", Sixth Eurographics Workshop on Rendering, Dublin, Ireland, pp. 82-91, (1995).

[9] E.Peli, L.E. Arend, G.M. Young, Contrast Sensitivity to Patch Stimuli, Spatial Vision, Vol.7, No.1, pp.1-14, 1993 .

[10] W.Osberger, Perceptual Vision Models for Picture Quality Assessment and Compression Aplications, Queensland University of Technology, Brisbane, Australia, 1999.

[11] Huib de Ridder, Minkowski-metrics as a Combination Rule for Digital-image-coding Impairments, In Proceedings of SPIE Human Vision, Visual Processing and Digital Display, vol 1666, pp.16-26, San Jose, CA, February 9-14, 1992, ISBN 0-8194-0820-4.

Contact information : kulhanj@fel.cvut.cz