IJCSN - International Journal of Computer Science and Network, Volume 9, Issue 4, August 2020
ISSN (Online) : 2277-5420
www.IJCSN.org
Impact Factor: 1.5

Design and Implementation of Visual Communication Systems for Images and Videos using Just-Noticeable Difference (JND) Visual Masking Model

[1] Valerian Agbasonu; [2] Amanze Ikwu; [3] Emmanuel Ekwonwune; [4] Ezenwa Nwawudu; [5] Ugochi Ikwu

[1] Computer Science Department, Imo State University,
Owerri, Nigeria

[2] Cardiology Department, University Hospitals Plymouth NHS Trust,
Derriford Road, PL6 5DH, United Kingdom

[3] Computer Science Department, Imo State University
Owerri, Nigeria

[4] Global Needs Institute
Saint Etienne, France

[5] Senior Enterprise Engineer, Allergan USA.
New Jersey, USA

Abstract - The major aim of designing visual communication systems is to use the least resources to achieve the highest visual quality with respect to certain constraints such as bit rate, complexity, and maximum delay. In most circumstances, the human visual system (HVS) makes final evaluation on the quality of images and video that are processed, transmitted, and displayed. Thus, it is essentially futile to spend significant effort on encoding those signals that are beyond the human perception by application of visual masking model that estimates the masking effect of the HVS. This study aims to explore the possibility of a computerised visual communication systems using Just Noticeable Distortion (JND), which accounts for the maximum distortion that the HVS does not perceive, which serves as a perceptual threshold to guide an image/video processing task. This goal was achieved by using b Model, Visual Attention Model and Weighted JND. This study also built the system in a robust manner so that it would utilize the masking model in determining the important level of each of the pixels, and this information is then in application specific processing.

Keywords - *Visual Masking, HVS, JND, Foveation, motion estimation, eye tracking*

1. Introduction

The main aim of designing visual communication systems is to use the least resources to achieve the highest visual quality with respect to certain constraints such as bit rate, complexity, and maximum delay. In most circumstances, the human visual system (HVS) makes final evaluations on the quality of images and video that are processed, transmitted, and displayed. Thus, it is essentially futile to spend significant effort on encoding those signals that are beyond the human perception. Just Noticeable Distortion (JND), which accounts for the maximum distortion that the HVS does not perceive, can serve as a perceptual threshold to guide an image/video processing task. In image compression schemes, JND can be used to optimize the quantizer [1]–[2] or to facilitate the rate-distortion control [2]. Information of higher perceptual significance is given more bits and preferentially encoded, so that the resultant image is more appealing. In video compression schemes, JND plays more diverse roles. As in image compression, JNDs for video can be used to improve quantizers [3] and bit allocation; moreover, motion estimation can be facilitated with the help of JND profiles [4]. For both image and video, objective quality evaluation based on the

characteristics of the HVS can be achieved by using the JND [5]–[6].

There are various applications where a visual masking model could be used to do efficient image/video processing; some examples are Image/Video filtering for display, Video compression, Watermarking,

Encryption/Steganography etc. These applications would utilize the masking model in determining the importance level of each of the pixels, and this information is then in application specific processing [7].

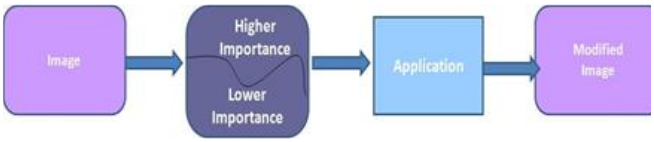

Figure 1. Program flow diagram

Our implementation of the visual masking mainly follows the functional block of [8] , the block diagram is displayed as below.

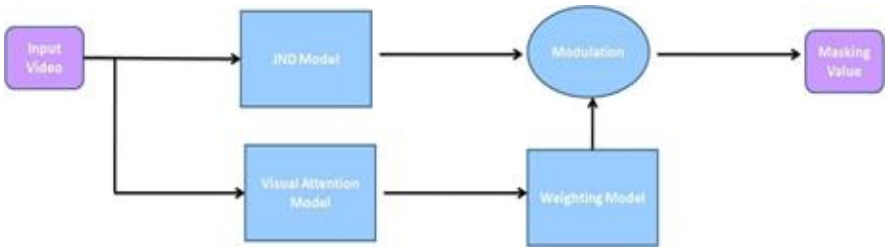

Figure 2. Diagram of JND model sub-blocks

The model consists of three main components JND (Just Noticeable Difference Model): Just Noticeable Difference is defined as the maximum distortion the human visual system cannot perceive.

Our implementation of JND is based on [9]. This implementation also considers temporal properties (eye tracking) in addition to the spatial properties. Visual Attention Model : The visual attention model we have considered in this project estimates the attention point of the eye in an image/video based on bottom up space based contrast stimuli (texture, luminance & motion) & top down object based features.

This is based on work done on paper [10]. Weighing Model : HVS has the highest spatial resolution & sensitivity at the point of fixation, the estimation of the fixation is incorporated into the overall model by modulating a weighing map generated using foveation method. Some components, such as multi-levelled equations, graphics, and tables are not prescribed, although the various table text styles are provided. The formatter will need to create these components, incorporating the applicable criteria that follows.

2. Basic JND Model

JND model considered here takes into account both temporal and spatial properties of the human visual system, the model incorporates a pixel domain edge detection using canny edge detection and then utilises the results to do a block type classification. In the next stage the model operates in the frequency domain by performing a discrete cosine transform (DCT) on the input image and incorporates spatial-temporal contrast sensitivity function, the influence of eye movement, luminance adaptation and contrast masking.

$$JND(n, i, j, t) = T(n, i, j, t)\alpha Lum^{(n, t)} \quad {}^{\alpha}intra^{(n, i, j, t)} \, {}^{\alpha}inter^{(n, t)} \quad \ldots. (1)$$

where ${}^{\alpha}Lum^{(n, t)}$, ${}^{\alpha}intra^{(n, i, j, t)}$, and ${}^{\alpha}inter^{(n, t)}$account for the effects of luminance adaptation, intra-band masking, and inter-band masking, respectively.

The JND model sub-blocks are explained in detail below as per how we have implemented it.

2.1 Edge Detection & Block Type Classification:

The edge detection we have employed is canny edge detection; the output of the edge detection is a binary image with the edge pixel identified. We divided this image into blocks (8x8) which can be used to calculate the edge density in each of the blocks. The blocks are further classified as plain blocks, edge blocks and texture blocks. The edge density is calculated as edge density = (number of edge pixels in block / number of pixels in the block). The classification thresholds are based on the equation (3) in paper [1]. The block type classification results were further filtered using a Majority filter in immediate neighbourhood blocks for Edge and Texture blocks. Plan blocks do not go through this filter since it is not possible to eliminate a lone Texture/Edge block from any arbitrary image without any contextual information.

2.2. Baseline Spatio-Temporal CSF Model:

The HVS is sensitive to contrast and can only sense a signal whose contrast is above a certain threshold with respect to a signal frequency. The reciprocal of this is the contrast sensitivity. The baseline spatio-temporal contrast sensitivity that we have implemented is based on equation (1) in the paper [9].This equation operates in the DCT sub-band domain and also incorporates the eye movement effect in the form of retinal image velocity which is explained below.

Eye Movement Effect (Eye Tracking): The eye movement is classified into three types : i) Smooth-Pursuit Eye Movement : tracks moving object and reduces retinal velocity. ii) Natural Drift Eye Movement: refers to very slow eye movement and is a used as measure for viewing static images. iii) Saccadic Eye Movement : refers to rapidly moving objects to which HVS has low sensitivity.

Because of the different eye movements, the perceived retinal velocity is different from the Image plane velocity. Retinal Image Velocity is defined, V= VI -VE. where VI is the Image plane object velocity and VE is the eye movement velocity. In our implementation predefined equation (6) in paper [9] for calculating VE.

The image plane velocity is implemented using two different methods for motion estimation.

2.3 Motion Estimation Motion estimation:

This is the process of determining motion vectors that describe the transformation from one 2D image to another; usually from adjacent frames in a video sequence. It is an ill-posed problem as the motion is in three dimensions, but the images are a projection of the 3D scene onto a 2D plane. In our JND model, motion estimation is used to get the image velocity representing eye movement effect.

Motion estimation using Optical flow calculation: Optical flow is the distribution of apparent velocities of movement of brightness patterns in an image. Optical flow calculation is a very popular gradient-based image matching method. It can give important information about the spatial arrangement of objects viewed and the rate of change of this arrangement. We estimated the direction and speed of object motion from one image to another or from one video frame to another using the Horn-Schunck method. By assuming that the optical flow is smooth over the entire image, the Horn-Schunck method computes an estimate of the velocity field that minimizes this equation:

$$E = \iint (I_x u + I_y v + I_t)^2 \, dxdy + \alpha \iint \left\{ \left(\frac{\partial u}{\partial x} \right)^2 + \left(\frac{\partial u}{\partial y} \right)^2 + \left(\frac{\partial v}{\partial x} \right)^2 + \left(\frac{\partial v}{\partial y} \right)^2 \right\} dxdy$$

… (2)

Motion estimation using block matching method: Block Matching Algorithm is a way of locating matching blocks in a sequence of digital video frames for the purposes of motion estimation. The purpose of a block matching algorithm is to find a matching block from a frame i in some other frame j, which may appear before or after i.

2.4 Luminance Adaptation:

The JND model we have implemented also incorporates the luminance adaptation, which is a property of the HVS where the eye has higher visibility threshold for dark and light regions and is more sensitive to noise in medium gray regions. The average local intensity of a block is determined by the dc component of a DCT block we use an equation (13) from paper [9] which utilizes this property E. Contrast Masking: The extent of contrast masking depends on the local intensity activity of the image. We performed a DCT domain Intra & Inter-band

contrast masking. In this method a DCT block is divided into DC, low frequency (LF), medium frequency (MF) and High Frequency and calculate a Texture Energy and utilize the block type classification done earlier to implement equation (15) for paper [9].

3. Visual Attention Model

One major factor in masking is how the human eye directs attention to various parts of an image or image sequence. There are two kinds of attention features: bottom-up features, processed by the brain from neutral detail into areas of interest, and top-down features, automatically recognized as specialized qualities and transformed into evaluation of its details. In the project, we implemented bottom-up features of colour and texture variation to find a weighted map of visual attention. In some literature, top-down processing of faces, skin tones, common objects, and patterns and motifs are incorporated into a visual attention model. However, considering that top down recognition code is still in a highly developmental stage and in the interest of computational ability, we focused our project on just the bottom-up components. As shown below, the input video sequence is evaluated for color and texture contrast, using a k-means clustering algorithm, combined into one conglomerate map based on correlation between stimuli, and truncated according to limitations of human attention.

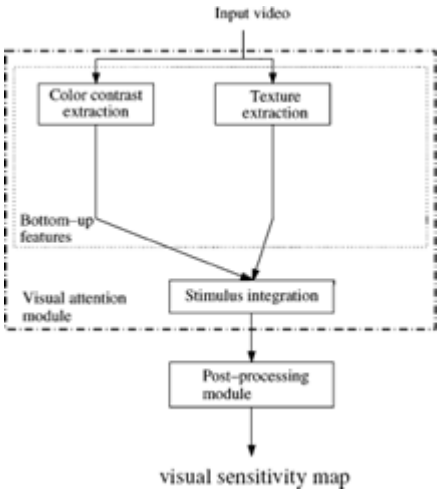

Figure 3. Diagram of visual attention process

For colour, we take each block and find the average RGB values inside the block and then apply the k-means method, which seeks to find arg min $\sum^{k}_{i=1}\sum_{xj\in Si} \| x_j - \mu_i \|^2$, for xj = ($R_j$, G_j , B_j), where there are k sets S_i and each has a mean value μi. Then if there is a cluster larger than some fraction of pixels, we designate that cluster as the background and compute relative distances, $d = \sqrt{(R - \overrightarrow{B})^2 + (G - \overrightarrow{G})^2 + (B - \overrightarrow{B})^2}$, and these values are discretized into ranges to create proper scaling. Otherwise, if the image is relatively uniform, default to using the centre of the image as the focal point. Texture variation is done similarly to colour contrast, except we count edge pixels in each block and use $d = | n_e - n_e |$.

How the two stimuli combine is dependent on their correlation. According to literature, color and texture are slightly correlated features, so a value of around 0.25 is a good choice, thus we have $S_{combined} = s_c + s_t - 0.25 \min (s_c, s_t)$. Post processing is based on a thresholded decaying exponential

$k = e^{-\left(\frac{e-Scombined-1}{Scombined+1}\right)^2}$ if the radius is within a standard deviation, otherwise we take k = 1. The maximal attention values after convolution with the kernel are taken. Finally, the map is scaled and limited by the maximal attention capacity, predefined to equal the block area.

4. Weighted JND

Foveation is the tendency of the human eye to have highest resolution at points of highest attention and exponentially decreasing sensitivity with increasing eccentricity away from the focal points. Our model incorporates the k = 10 most attention-weighted fixation points to use as the focal points. We then calculate $W = e^{-\frac{\alpha \int \xi}{e0} \min_k \left\{ arctan \sqrt{\frac{(x-xfk)2+(y-yfk)2}{V}} \right\}}$, where f is the local frequency, and xfk and yfk are the coordinates of the kth fixation points and V is the distance from the observer to the screen. The final interpretation of the model is in the frequency domain of the DCT, with weighting from foveation based on frequency eccentricity and block type classification combined with frequency analysis, all dependent on block location and DCT coefficient values. Each block is a region where we could introduce a certain amount of noise or quantization error, the frequency of which corresponds to the values of the weights in the block.

5. Implementation Result

The Implementation was done in Matlab. Here we present the results of individual stages of the algorithm.

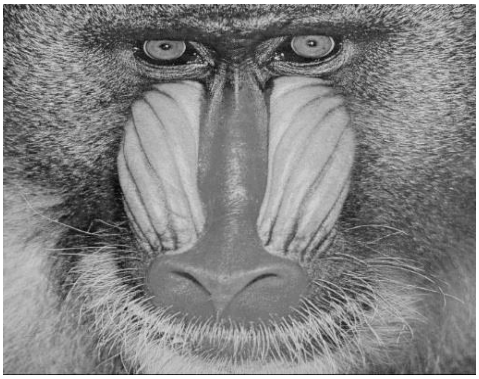

Figure 4. Input image

Figure 5. Edge Detection Output

Figure 6. Block Type Classification

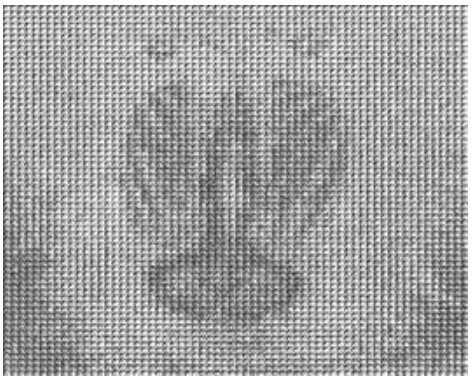

Figure 7. JND Model Output

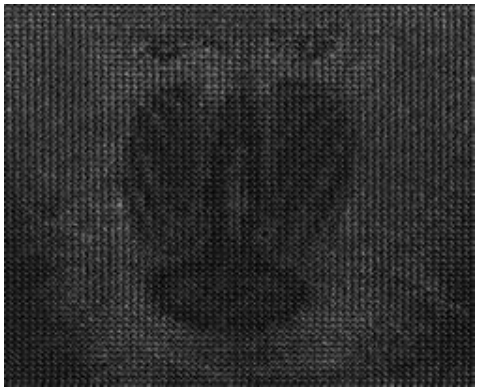

Figure 8. Visual Mask with Visual Attention in the Centre

6. Conclusion

This work explored the possibility of a computerised visual communication systems using Just Noticeable Distortion (JND). We demonstrated our experiment using b Model, Visual Attention Model and Weighted JND from a combination of various research papers that will give a relevance map of the image in terms of 8x8 pixel blocks. We also built the system in a robust manner so that it would utilize the masking model in determining the importance level of each of the pixels, and this information is then in application specific processing

References

[1] B. Watson, "DCTune: A Technique for Visual Optimization of DCT Quantization Matrices for Individual Images," in Soc. Inf. Display Dig. Tech. Papers XXIV, 1993, pp. 946–949.

[2] I. Hontsch and L. J. Karam, "Adaptive Image Coding with Perceptual Distortion Control," IEEE Trans. Image Process., vol. 11, no. 3, pp. 213–222, Mar. 2002.

[3] C.-H. Chou and Y.-C. Li, "A Perceptually Optimized 3-D Subband Codec for Video Communication over Wireless Channels," IEEE Trans. Circuits Syst. Video Technol., vol. 6, no. 2, pp. 143–156, Apr. 1996.

[4] X. K. Yang, W. Lin, Z. K. Lu, E. P. Ong, and S. S. Yao, "Just Noticeable Distortion Model and its Applications in Video Coding," Signal Process.: Image Commun., vol. 20, no. 7, pp. 662–680, Aug. 2005

[5] Sarnoff Corp., Princeton, NJ, Sarnoff JND Vision Model Algorithm Description and Testing VQEG, Aug. 1997.

[6] W. Lin, L. Dong, and P. Xue, "Visual Distortion Gauge Based on Discrimination of Noticeable Contrast Changes," IEEE Trans. Circuits Syst. Video Technol., vol. 15, no. 7, Jul. 2005.

[7] Yuhong Wang, Chi Zhang, Sukesh Kaithaapuzha: "Visual Masking Model Implementation for Images & Video." EE368 spring final paper 2009/2010.

[8] Zhongkang Lu, Weisi Lin, Xiaokang Yang, EePing Ong, Susu Yao, Modeling Visual Attention's Modulatory After effect, IEEE Transactions on Image Processing, Vol. 14, No. 11, November 2005, pp. 1928-1942.

[9] Anmin Liu, Maansi Verma and Weisi Lin,;"Modeling the Masking Effect of the Human Visual System with Visual Attention Model", Nanyang Technological University, Singapore.

[10] Yuting Jia, Weisi Lin, and Ashraf A. Kassim "Estimating Just-Noticeable Distortion for Video"

First Author: *Dr Valerian Agbasonu* has a PhD in Computer Science and a Senior Lecturer, Computer Science Department of Imo State University, Owerri, Nigeria.

Second Author: *Dr Amanze Ikwu* is a trained Physician who works at University Hospitals Plymouth NHS Trust, United Kingdom. He has published 10 research work in reputable international journals. He is an astute scholar with interest in cardiovascular diseases, emerging infectious diseases, digital technology application in medicine, telemedicine advancement in Africa, arrhythmia, geriatric medicine and E-Health. He holds an MBBS certificate and Fellowship of the Medical College of Physicians in Internal Medicine/Cardiology.

Third Author: Dr Emmanuel Ekwonwune has a PhD in Computer Science and a Senior Lecturer, Department of Computer Science, Imo State University, Owerri, Nigeria.

Fourth Author: Ezenwa Nwawudu is a highly skilled IT professional and entrepreneur with the capacity for leadership and championing business growth. He is an astute scholar with interest in Internet of Things (IoT), Smart Technologies, Biometric Systems, Digital Inclusion, Telemedicine and E-Health, etc.
He holds Master degree in Science, Technologies and Health with specialisation in Optics, Image, Vision and Multimedia from Université Paris-Est, Creteil (UPEC), France.

Fifth Author: Engineer Ugochi Ikwu has certifications in CCNA Datacenter, CISCO WLAN Design Specialist, CCIE Data Center Written. He is a Senior Enterprise Engineer at Allergan USA, with over 22 years' experience in network designs, implementation and Data Center buildup.