PERCEPTUAL VIDEO CODING: CHALLENGES AND APPROACHES

Zhenzhong Chen[1], Weisi Lin[2], and King Ngi Ngan[3]

[1]School of Electrical and Electronic Engineering, Nanyang Technological University
[2]School of Computer Engineering, Nanyang Technological University
[3]Department of Electronic Engineering, Chinese University of Hong Kong
Email: zzchen@ntu.edu.sg, wslin@ntu.edu.sg, knngan@ee.cuhk.edu.hk

ABSTRACT

Investigation on the human perception can play an important role in video signal processing. Recently, there has been great interest in incorporating the human perception in video coding systems to enhance the perceptual quality of the represented visual signal. However, the limited understanding of the human visual system and high complexity of computational models of human visual system make it a challenging task. Furthermore, the hybrid video coding structure brings difficulties to integrate computational models with coding components to fulfill the requirements. In this paper, we review the physiological characteristics of human perception and address the most relevant aspects to video coding applications. Moreover, we discuss the computational models and metrics which guide the design and implementation of the video coding system, as well as the recent advances in perceptual video coding. To introduce this overview with the latest technologies and most promising directions in perceptual video coding, we focus on three key areas. Specifically, we cover 1) visual attention and sensitivity modeling, with which we concentrate on the computational models of bottom-up and top-down attention, contrast sensitivity functions and masking effects, and fovea based manipulations; 2) perceptual quality optimization for constrained video coding, with which we discuss how to achieve maximum perceptual quality whilst satisfying various constraints; and 3) the impact of the human perception on advanced video applications, including emerging immersive multimedia services, and compression of high dynamic range video content and 3D video. For each aspect, we discuss the major challenges, highlight significant approaches, and outline future research directions.

Keywords— Perceptual video coding, human visual system, visual perception, quality optimization

1. INTRODUCTION

The last decade has witnessed the explosion of applications of video coding technologies in our daily life. Various video coding schemes have been developed to achieve high compression efficiency or flexible functionalities. In a typical video coding system, the main objective is to maximize the video quality whilst satisfying the bit budget constraint. How to quantify the video quality and use it for practical video coding system design is an important issue. Mean square error (MSE) and peak signal-noise ratio (PSNR) have been widely used in video coding systems due to their clear physical meaning and simple computation. However, MSE and PSNR cannot well quantify the video quality in a manner consistent with human visual perception. To address this problem, extensive research activities have been conducted to define the measurement of the video quality, either by subjective quality assessment, or perceptual video quality metrics (VQMs) [13][44].

Taking the human perception into consideration, the concept of perceptual image/video coding has been established for quite some time [15][23]. The objective of perceptual video coding is to achieve maximum visual quality of the decoded video. With the progress of VQMs and visual perception modeling [12][21], perceptual video coding is attracting more and more attention. However, perceptual models may not be easily integrated into the video coding system due to the hybrid structure, various coding units, and complex mode selection. Recently developed computational models for visual sensitivity, attention, and visibility [8][20], are incorporated into video coding systems and have led to video coding applications [5] [26] [35] [41][47].

This paper provides a review of perceptual video coding. We not only review the physiological characteristics of human perception and their computational modeling, but also their impacts in video coding. We classify perceptual video coding into three categories, namely, vision-model based approach, signal-driven approach, and hybrid approach. For each category, we introduce the latest finds and contributions, discuss the major challenges, and outline the future research directions. The remaining of this paper is organized as follows. We describe the human visual system and the computational modeling techniques in the next section. In Section 3, the definitions of video quality are reviewed. In addition, we introduce several public video quality assessment databases. A comprehensive review of

978-1-4244-7493-6/10/$26.00 ©2010 IEEE

perceptual video coding is then given in Section 4. In the last section, summary and discussions are provided.

2. HUMAN VISUAL PERCEPTION & MODELLING

2.1. Human visual system

Understanding the Human Visual System (HVS) [39] plays the key role in perceptual video coding. The HVS consists of two functional parts, the eye and the brain. The eye consists of three layers, the outermost layer composed of the cornea and sclera, the middle layer composed of the choroid, ciliary body, and iris, and the innermost layer retina. Light entering the eye is focused and inverted by the cornea and lens, and is projected onto the retina. The retina is a thin layer of neural cells which converts the light signal into a neural signal and transmits it to the brain through the optic nerve.

The retina contains two major classes of photoreceptors: the rods and cones. The rods work under low ambient light and extract luminance information. The cones handle color vision and only function in relatively bright light. The cones are able to perceive finer detail and more rapid changes in images since their response times to stimuli are faster than the rods. The cones concentrate at the fovea which is responsible for sharp central vision towards the object which attracts our focus of attention.

The eyes can track the moving object in the visual scene to keep the object of interest on the fovea and compensate the object motion to improve the visual acuity [9]. This is known as the Smooth Pursuit Eye Movement (SPEM) [31]. Vision is generated by photoreceptors in the retina, transmitted to the brain along the optic nerve, and get to the lateral geniculate nucleus (LGN). The LGN then relay the visual information to the primary visual cortex (V1) at the back of the brain, which is responsible for processing the visual information, together with other areas (V2, V3, V4 and V5/MT) [2].

2.2. Computational visual perception

Computational models have been developed to study unique characteristics of the HVS. In this subsection, we focus on the most relevant aspects to video coding and introduce the advanced modeling techniques.

2.2.1. Visual sensitivity modeling

The human being cannot perceive the fine-scale variation of visual signal due to the psychovisual properties of the HVS. The related psychophysical experiments show the contrast required to detect a flickering grating of different spatial and temporal frequencies. Masking effect refers to the perceptibility of one signal in the presence of another signal in its spatial, temporal, or spectral vicinity [15]. In addition, the sensitivity of the HVS depends on the background luminance and color of the stimuli. In general, the visual sensitivity can be measured by a spatio-temporal contrast sensitivity function (CSF) [17]. Ninassi et al. [27] proposed to model masking effects by both contrast masking and semi-local masking where semi-local masking considers the modification of the visibility threshold due to the semi-local complexity of an image, also known as entropy masking, or texture masking.

2.2.2. Visual attention modeling

Visual attention is a subconscious process of the HVS and is one of the most concrete cognitive processes of the human being. While attention is controlled in a voluntary manner, it is attracted in a "bottom-up" automatic and unconscious manner to locate the conspicuous (salient) visual area. This property is crucial to constitute a computational system to attempt to mimic the neuronal mechanisms of the human perception responsible for attracting our attention to salient objects in the visual scene. The feature integration theory (FIT), developed by A. Treisman [36], has been one of the most influential psychological models of human visual attention. In FIT, several primary visual features such as color, orientation, and intensity, are processed and represented with separate feature maps and then integrated in a saliency map [37]. To achieve this purpose, feature analysis and the search for coherence among spatially distributed features have therefore been thought to constitute a basic neuronal mechanism of underlying figure-ground discrimination in the human visual system [45]. There are two ways in which a pre-attentive process can be modeled to direct attention: bottom-up (stimulus-driven) and top-down (user-driven) [12].

Computational models of saliency-based visual attention, either bottom-up based or integrated, have achieved modest success in representing the conspicuity (saliency) by integrating low level and high level features to a saliency map in which an explicit two-dimensional map represents a scalar quantity at every location over the entire visual scene. As visual attention is often object-based [32], it is important to develop object-level visual attention model and corresponding video quality metrics for the development of perceptual video coding.

2.2.3. Visual visibility modeling

Based on visual sensitivity studies, the just-noticeable distortion (JND) provides cues for measuring the HVS visibility for the video signal [15][20]. For the human vision, JND refers to the maximum distortion which cannot be perceived by the HVS of most observers. As indicated by Weber's law [15], the perceptible luminance difference of a stimulus depends on the surrounding luminance level. By incorporating masking effect and luminance adaptation into the CSF, various computational JND models have been developed in spatial (pixel) domain, subband, DCT, and wavelet domains [16][20][42]. Since the retina in a human eye does not have a uniform density of photoreceptor cells, the human visual system is space-variant. As the density of the cone and retinal ganglion cells drops with the increased retinal eccentricity, the visual acuity decreases with increased eccentricity as well. Some recent researches focus on the space-variant property of the HVS and further enhance JND models to measure the global visibility to perceive videos [4].

3. PERCEPTUAL VISUAL QUALITY

To evaluate the performance in perceptual video coding, subjective experiments or objective video quality metrics (VQMs) are considered.

3.1. Subjective quality evaluation and database

In a subjective experiment, viewers are asked to watch the video, evaluate the quality assessment or impairment assessment, and rate the video quality by the Opinion Score (OS). The evaluation procedure may follow different protocols, such as simultaneous double stimulus for continuous evaluation (SDSCE), Double Stimulus Continuous Quality Scale (DSCQS), and Double Stimulus Impairment Scale (DSIS) [13]. For different protocol, different scale method, such as impairment scale, comparison scale, Mean Opinion Scale (MOS), or differential MOS, could be chosen according to the specific application. However, the subject experiment requires a large number of viewers and is usually time consuming.

It is worthy to mention that several subjective video quality databases have been released to public, such as VQEG Phase I database [38], EPFL-PoliMI video quality assessment database [7], A57 database [1], and LIVE video quality database [33]. These databases are invaluable and could be used as benchmark for the research and development of objective video quality metrics.

3.2. Video quality metrics

To overcome the shortcoming of MSE or PSNR but avoid the expensive subjective experiments, various advanced objective VQMs have been developed. They could be classified into two categories: vision modeling approach and engineering approach [43]. Vision-model based approach is based on human vision properties, such as CSF, masking effects, foveation filtering, and eye movement [44]. Engineering approach is to evaluate the video quality based on distorted visual signal features, structures, or artifacts. Recent successful VQMs in engineering approach include structural similarity (SSIM) index [40] which measures the distortion of distortion of structure information, Video Quality Model (VQM) [30] which measures distorted low-level features from spatial-temporal video blocks, and Visual Information Fidelity (VIF) [34]. As some of existing VQMs are directly extended from metrics for image quality assessment, more and more VQMs incorporate temporal cues such as flickering and jerkiness [27][46].

4. PERCEPTUAL VIDEO CODING

With the progress of the understanding of neurobiological properties of the human visual perception, the trend in video coding becomes to improve perceptual quality in constrained applications. VQMs play important roles in the evaluation of the performance of a perceptual video coding system and the guidance for the perceptual video coding design. We classify perceptual video coding into three categories, vision-model based approach, signal-driven approach, and hybrid approach. Vision-model based approach refers to the perceptual video coding based on human vision properties, e.g., space-variant vision, SPEM, and selective attention. Signal-driven approach refers to the algorithms which attempts to analyze visual features by signal processing. Some algorithms considered both human vision model and visual features, they are considered as hybrid approaches.

4.1. Vision-model based approaches

4.1.1. ROI based video coding

Region of Interests (ROI) based perceptual video coding has been established for a long time. The main idea is straightforward: enhance visual quality by improving fidelity of human interested regions. Earlier ROI base video coding schemes have limitations on how to define the ROI. Head-shoulder videos are typical applications in which face can be generally treated as ROI [47]. The foreground/background video coding [3] also belongs to this approach. Recent visual attention techniques make it possible to identify more flexible ROI, or region of attention (ROA). In a ROI video coding scheme, smaller quantization parameter is used to represent the ROI with lower distortion which could significantly contribute to the subjective quality of overall video. However, it may introduce artifacts on the boundary of the ROI if very different quantization parameters are used for ROI and regions outside ROI.

As some existing ROI video coding algorithms do not consider the fovea characteristics of the HVS and simply use fixed weights for quantization parameters to code ROI and background region, these ROI video coding schemes could not remove the perceptual redundancy sufficiently. By incorporating with foveation model, the ROI based video coding can provide better coding efficiency whilst utilizing the perceptual features of the HVS.

4.1.2. Foveation and attention based video coding

The HVS is highly space-variant and the spatial resolution is highest at the fovea. So the object projected on the fovea significantly contributes to the subjective quality. Foveated video compression algorithms have been proposed based on this knowledge. Such algorithms aim to match the non-uniform sampling of the human retina. Wang et al. [41] proposed a perceptually scalable video coding framework based on face detection based automatic fixation region selection. An adaptive frame prediction algorithm is employed to avoid significant degradations of compression performance for peripheral regions in motion compensation due to the foveation model.

With the progress of computational techniques for visual attention [12], exploring human attention have been considered in recent perceptual video coding. Itti [11] uses the bottom-up visual attention model to detect ROA and employs foveation filter to pre-process the video before coding. As the high frequency components of the background are removed, the coding efficiency is improved. Attention-based bit allocation for video object based coding has been proposed [5]. Integrated low-level and high-level feature fusion is used to generate the saliency map to define visual attention values for video objects and the object of attention (OOA) are encoded with higher quality by finer quantization parameters.

In some work, the eye movement is considered as human eyes track salient moving object. Nguyen et al. [27] presented an approach based on the SPEM, to control the rate of the video data generated by the encoder such that high motion activities should be finely quantized. They developed a SPEM based rate control mechanism to improve the quality of moving scenes in the video sequence by using the motion information provided by the motion analysis and SPEM.

The foveation and visual attention based video coding methods generally represent perceptual-friendly videos. But the computational complexity for automatic fixation region selection and visual attention modeling is high and the performance relies on the accuracy of the fixation selection and visual attention models. They would be more useful in interactive video communications in which the fixation region or attention region can be controlled through user interaction.

4.2. Signal-driven approaches

4.2.1. Sensitivity based video coding

Spatio-temporal CSF defines the video sensitivity for the HVS. It could be used to measure the visibility by integrate with masking effects, luminance adaptation, and other factors. To control the distortion in perceptual video coding, JND models are used to determine the quantization steps and pre-process motion compensation residual for perceptual quality enhancement [6][47]. Based on the sensitivity of the video signal, the typical coding unit, such as an MB, could be classified into different levels, e.g., busy and textured areas, and flat and low-detail regions and unstructured texture, and different quantization class could be applied accordingly. Scene level classification based on perceptual properties of the HVS can also considered in video coding. These video coding approaches attempt to control the distortion of blocks therefore the perceptual feelings of different blocks are consistent, i.e., blocks which can tolerate higher distortion will be coarsely coded.

However, the perceptual redundancy cannot be fully discovered without the consideration of fovea features of the HVS. It is noted that perceptually lossless video coding can be achieved by JND based visual sensitivity analysis. It provides better compression efficiency when compared to lossless video coding. It plays important role for high end video compression applications such as digital cinema and immersive video entertainment.

4.2.2. Content based video coding

Content based video coding is extensively studied recently. Unlike conventional methods, current content based video coding takes the advantages of some properties in the HVS, e.g., spatio-temporal masking effect, and uses texture analysis and synthesis in video coding [26]. Based on the advanced techniques developed in computer vision area, some regions in video can be skipped at the encoder and be restored by texture synthesis at the decoder [48]. For this reason, it is sometime called as vision-based video coding. Since the texture analysis is based on texture and edge features, we consider it as a signal-driven approach.

In addition to above signal-driven approaches, Malo et al. [22] proposed multigrid motion compensation based on HVS contrast discrimination models. We consider it as a signal-driven approach since it is based on signal processing. The adopted HVS model determines the maximum perceptual distortion generated by perceptually matched quantization and the perceptually weighted motion estimation. Inter-frame replenishment can be controlled by the HVS model to reduce bit rate and complexity. Moreover, post-processing tools, such as deblocking-filter, have been integrated in some hybrid video coding schemes (e.g., H.264/AVC [14]) and have demonstrated their performance in terms of better subjective quality.

4.3. Hybrid approaches

Some perceptual video coding algorithms integrate the visual features analysis and the vision model to achieve better perceptual optimization. Tang et al. [35] have proposed a bit allocation technique using a visual distortion sensitivity model for H.264/AVC coder. The distortion sensitivity analysis process considers both the motion attention and the texture structures of video and evaluates the perceptual distortion sensitivity on a macroblock basis. The macroblocks that can tolerate higher perceptual distortion will be allocated fewer bits to save bit rate.

In [47], skin color based face detection is used to locate the ROI, and weighting factors derived from the JND model are used for adapting the rate in the different regions. The performance of compression can be further improved by exploiting effects of retinal eccentricity on visual acuity. Chen et al. [4] present a foveated just-noticeable-distortion (FJND) model which incorporates visual sensitivity, attention, and fovea features of the HVS. Since the perceptual acuity decreases with increased eccentricity, the visibility threshold of the pixel increases accordingly. The obtained FJND model could better explore the global perceptual redundancy and be used in H.264/AVC video coding to improve the perceptual optimization.

Hybrid approaches better exploit the perceptual redundancy in video by taking the advantages of vision-model based and signal-driven approaches but result in higher computational complexity.

5. SUMMARY AND DISCUSSION

In this paper, we have reviewed the advanced technologies in modeling visual perception, building video quality metrics, and designing perceptual video coding. Since a VQM is normally complex and modeling of visual perception is computationally expensive, perceptual video coding needs to balance the tradeoff between the algorithm complexity and the coding performance. With the development of visual perception models and advanced applications such as 3D video, high dynamic range video representation, computer graphics, and video animation, perceptual video coding will play more and more important roles.

The research in 3D video quality assessment and perceptual coding is still limited [10][18]. Unlike to conventional VQMs developed for 2D video, the VQM for 3D video involve more complicated factors in various aspects, from multiple camera calibration and configurations, 3D video signal capture and delivery, to 3D video display and rendering. Some distortions, which do not present in 2D video signal representation, play important roles in quality evaluation, such as keystone distortion, magnification and miniaturization effects, shear distortion, cross-talk, picket fence effect and image flipping. Those factors make the defining of a 3D VQM difficult and challenging [25]. However, the existing research has yet to provide us with a comprehensive and robust means to automatically assess the perceptual quality of 3D video and solutions for 3D video representation.

High Dynamic Range (HDR) video allows better representation for visual scene in emerging areas such as computer graphics, video game, and computational photography. With rapid progress in its applications, how to understand the human perception to HDR video and how to achieve efficient represent the HDR video are important. This requires modeling of the human perception to HDR scenario with various effects such as glare, day and night vision, contrast sensitivity, and visual adaptation. Mantiuk et al. [24] have presented a perceptual framework for contrast processing of HDR video and also provided a perceptual coding solution for HDR video based on MPEG-2. Further research could be conducted to improve the compression efficiency by advanced coding techniques.

Nowadays, people have much higher expectation in video applications, as large and high quality display system and broadband network connection have been popular. For example, immersive multimedia entertainment and education are receiving more and more attention and also bring new requirements and challenges in multimedia services. Improving user experience by perceptual video processing is one of the tasks.

Psychological, biological, and theoretical approaches have been used to study perception. Nonetheless, there are still many aspects in the human perception and their impacts in video representation need to be discovered. Moreover, existing VQMs could not be easily integrated into the hybrid video coding so it is difficult to optimize the system performance by largely removing perceptual redundancy. Further research addressing the multi-disciplinary problems in perception is expected to achieve new breakthrough and make great impact.

6. REFERENCES

[1] A57 dataset: foulard.ece.cornell.edu/dmc27/vsnr/vsnr.html

[2] R. Blake and R. Sekuler, Perception, McGraw-Hill, 2006.

[3] D. Chai and K. N. Ngan, “Foreground/background video coding scheme,” ISCAS, 1997.

[4] Z. Chen and C. Guillemot, “Perception-oriented video coding based on foveated JND Model,” Picture Coding Symposium, 2009.

[5] Z. Chen, J. Han, and K. N. Ngan, “Dynamic bit allocation for multiple video objects coding,” IEEE Trans. Multimedia, pp. 1117–1124, Dec. 2006.

[6] C.-H. Chou and Y.-C. Li, “A perceptually tuned subband image coder based on the measure of just-noticeable-distortion profile,” IEEE Trans. Circuits Syst. Video Technol., pp. 467–476, Dec. 1995.

[7] F. De Simone, M. Naccari, M. Tagliasacchi, F. Dufaux, S. Tubaro, and T. Ebrahimi, “Subjective assessment of H.264/AVC video sequences transmitted over a noisy channel", International Workshop on Quality of Multimedia Experience, July 2009.

[8] W. S. Geisler and J. S. Perry, “A real-time foveated multisolution system for low-bandwidth video communication,” SPIE Proceedings, vol. 3299, pp. 294–305, July 1998.

[9] B. Girod, “Eye movements and coding of video sequences,” in SPIE Visual Communications and Image Processing, 1988.

[10] C.T.E.R. Hewage, S. Worrall, S. Dogan, and A.M. Kondoz, “Quality evaluation of colour plus depth map based stereoscopic video”, IEEE Journal of Selected Topics in Signal Processing, pp. 304-318, April 2009.

[11] L. Itti, “Automatic foveation for video compression using a neurobiological model of visual attention,” IEEE Trans. Image Process., pp. 1304–1318, Oct. 2004.

[12] L. Itti and C. Koch, “Computational modeling of visual attention,” Nature Reviews Neuro., pp. 194–203, Mar. 2001.

[13] ITU-R BT.500-11, “Methodology for the subjective assessment of the quality of television pictures,” 2002.

[14] ITU-T Rec. H.264 and ISO/IEC 14496-10, “Advanced Video Coding for Generic Audiovisual Services,” May 2003.

[15] N. Jayant, J. Johnston, and R. Safranek, “Signal compression based on models of human perception,” Proc. IEEE, pp. 1385–1422, Oct. 1993.

[16] Y. Jia, W. Lin, and A. A. Kassim, “Estimating just-noticeable distortion for video,” IEEE Trans. Circuits Syst. Video Tech., pp. 820–829, July 2006.

[17] D. H. Kelly, “Motion and vision II stabilized spatio-temporal surface,” J. Opt. Soc. Amer., pp. 1340–1349, Oct. 1979.

[18] W.-S. Kim, A. Ortega, P. Lai, D. Tian, and C. Gomila, “Depth map distortion analysis for view rendering and depth coding,” ICIP, 2009.

[19] G. Krawczyk, K. Myszkowski, H.-P. Seidel, “Perceptual effects in real-time tone mapping,” Proceedings of the 21st spring conference on computer graphics, 2005.

[20] W. Lin, “Computational models for just-noticeable difference,” in Digital Video Image Quality and Perceptual Coding, H. R. Wu and K. R. Rao, Eds., CRC Press, 2006.

[21] Z. Lu, W. Lin, X. Yang, E. Ong, and S. Yao, “Modeling visual attention’s modulatory aftereffects on visual sensitivity and quality evaluation,” IEEE Trans. Image Process., pp. 1928 – 1942, Nov. 2005.

[22] J. Malo, J. Gutierrez, I. Epifanio, F. Ferri, and J. M. Artigas, “Perceptual feed-back in multigrid motion estimation using an improved DCT quantization,” IEEE Trans. on Image Process., pp. 1411–1427, Oct. 2001.

[23] J. L. Mannos and D. J. Sakrison, “The effects of a visual fidelity criterion on the encoding of images,” IEEE Trans. Information Theory, vol. IT-20, pp. 525–536, July 1974.

[24] R. Mantiuk, K. Myszkowski, H.-P. Seidel, “A perceptual framework for contrast processing of high dynamic range images,” ACM Transactions on Applied Perception, pp. 286-308, 2006.

[25] L. M. J. Meesters, W. A. Ijsselsteijn, and P. J. H. Seuntiëns, “A survey of perceptual evaluations and requirements of three-dimensional TV,” IEEE Trans. Circuits Syst. Video Tech., pp. 381-391, 2004.

[26] P. Ndjiki-Nya, D. Bull, and T. Wiegand, “Perception-oriented video coding based on texture analysis and synthesis,” ICIP, 2009.

[27] A. G. Nguyen and J.-N. Hwang, “A novel hybrid HVPC/mathematical model rate control for low bit-rate streaming video,” Signal Process.: Image Commun., pp. 423–440, May 2002.

[28] A. Ninassi, O. L. Meur, P. L. Callet, and D. Barba, “Which semi-local masking model for wavelet base image quality metric,” ICIP, 2008.

[29] A. Ninassi, O. L. Meur, P. L. Callet, and D. Barba, “Considering temporal variations of spatial visual distortions in video quality assessment,” IEEE Journal of Selected Topics in Signal Processing, pp. 253–265, April 2009.

[30] M. Pinson and S. Wolf, “A new standardized method for objectively measuring video quality,” IEEE Trans. on Broadcasting, pp. 312-322, Sept., 2004.

[31] D. A. Robinson, “The mechanics of human smooth pursuit eye movement,” J. Physiol., vol. 180, pp. 569–591, Oct. 1965.

[32] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio, “Object recognition with cortex-like mechanisms,” IEEE Trans. PAMI, pp. 411–426, 2007.

[33] K. Seshadrinathan, R. Soundararajan, A. C. Bovik and L. K. Cormack, "A Subjective Study to Evaluate Video Quality Assessment Algorithms", SPIE Proceedings Human Vision and Electronic Imaging, Jan. 2010.

[34] H.R. Sheikh and A.C. Bovik, “Image information and visual quality,” IEEE Trans. Image Process., pp. 430–444, 2006.

[35] C.-W. Tang, C.-H. Chen, Y.-H. Yu, and C.-J. Tsai, “Visual sensitivity guided bit allocation for video coding,” IEEE Trans. Multimedia, pp. 11–18, Feb. 2006.

[36] A. M. Treisman and G. Gelade, “A feature-integration theory of attention,” ACM Computing Surveys Cognitive Psychology, pp. 97–136, 1980.

[37] A. M. Triesman and R. Paterson, “Emergent features, attention and object perception,” J. Exp. Psychol: Human Perception and Performance, pp. 12–31, 1984.

[38] VQEG (Video Quality Experts Group) Phase-I Database. Available: http://www.vqeg.org

[39] B. Wandell, Foundations of Vision, Sinauer Associates, Inc., Sunderland, MA, 1995.

[40] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” IEEE Trans. on Image Process., pp. 600–612, Apr. 2004.

[41] Z. Wang, L. Lu, and A. C. Bovik, “Foveation scalable video coding with automatic fixation selection,” IEEE Trans. Image Process., pp. 243–254, Feb. 2003.

[42] Z. Wei and K.N. Ngan, “Spatio-temporal just noticeable distortion profile for grey scale image/video in DCT domain,” IEEE Trans. Circuits Syst. Video Tech., pp. 337–346, Mar. 2009.

[43] S. Winkler and P. Mohandas, “The evolution of video quality measurement: From PSNR to hybrid metrics,” IEEE Transactions on Broadcasting, pp. 660-668, Sept. 2008.

[44] S. Winkler, Digital Video Quality: Vision Models and Metrics. West Sussex, England: John Wiley & Sons Ltd., 2005.

[45] J. M. Wolfe, “Visual search,” in Attention, H. Pashler, Ed. London, UK: Psychology Press, pp. 13–74, 1998.

[46] K. C. Yang, C. C. Guest, K. El-Maleh, and P. K. Das, “Perceptual temporal quality metric for compressed video,” IEEE Trans. on Multimedia, pp.1528-1535, Nov 2007.

[47] X. Yang, W. Lin, Z. Lu, X. Lin, S. Rahardja, E. Ong, and S. Yao, “Rate control for videophone using perceptual sensitivity cues,” IEEE Trans. Circuits Syst. Video Tech., pp. 496–507, April 2005.

[48] C. Zhu, X. Sun, F. Wu, and H. Li, “Video coding with spatio-temporal texture synthesis and edge-based inpainting,” ICME, 2008.