Chapter 6
Spatio-Temporal Just Noticeable Distortion Model Guided Video Watermarking

Yaqing Niu
Communication University of China, China

Sridhar Krishnan
Ryerson University, Canada

Qin Zhang
Communication University of China, China

ABSTRACT

Perceptual Watermarking should take full advantage of the results from human visual system (HVS) studies. Just noticeable distortion (JND), which refers to the maximum distortion that the HVS does not perceive, gives a way to model the HVS accurately. An effective Spatio-Temporal JND model guided video watermarking scheme in DCT domain is proposed in this paper. The watermarking scheme is based on the design of an additional accurate JND visual model which incorporates spatial Contrast Sensitivity Function (CSF), temporal modulation factor, retinal velocity, luminance adaptation and contrast masking. The proposed watermarking scheme, where the JND model is fully used to determine scene-adaptive upper bounds on watermark insertion, allows providing the maximum strength transparent watermark. Experimental results confirm the improved performance of the Spatio-Temporal JND model. The authors' Spatio-Temporal JND model is capable of yielding higher injected-watermark energy without introducing noticeable distortion to the original video sequences and outperforms the relevant existing visual models. Simulation results show that the proposed Spatio-Temporal JND model guided video watermarking scheme is more robust than other algorithms based on the relevant existing perceptual models while retaining the watermark transparency.

DOI: 10.4018/978-1-4666-1758-2.ch006

INTRODUCTION

The rapid growth of the Internet has created a need for techniques that can be used for copyright protection of digital images and videos. One approach is to introduce digital watermark into images or video sequences. For a well-designed watermark there are many requirements including imperceptibility, robustness and capacity. However, in order to maintain the image quality and at the same time increase the probability of the watermark detection, it is necessary to take the human visual system (HVS) into consideration when engaging in watermarking research (Wolfgang, Podilchuk, & Delp, 1999; Podilchuk & Zeng, 1998; Doerr & Dugelay, 2003).

HVS makes final evaluations on the quality of videos that are processed and displayed. Just noticeable distortion (JND), which refers to the maximum distortion that the HVS does not perceive gives us a way to model the HVS accurately and can serve as a perceptual visibility threshold to guide video watermarking. JND estimation for still images has been relatively well developed. An early perceptual threshold estimation in DCT domain was proposed by Ahumada and Peterson (1992), which gives the threshold for each DCT component by incorporating the spatial Contrast Sensitivity Function (CSF). This scheme was improved by Watson (1993) after the luminance adaptation effect had been added to the base threshold, and contrast masking (Legge, 1981) had been calculated as the elevation factor. In Zhang, Lin, and Xue (2005) an additional block classification based contrast masking and luminance adaptation was considered by Zhang for digital images. A spatial JND model proposed by Wei and Ngan (2008) incorporates new spatial CSF, luminance adaptation and contrast masking. Since motion is a specific feature of videos, temporal dimension needs to be taken into account for human perceptual visibility analysis. JND estimation for video sequences need to incorporate not only the spatial CSF, but the temporal CSF as well. A spatio-temporal CSF model was proposed by Kelly (1979) from experiments on visibility thresholds under stabilized viewing conditions. Daly (1998) extended Kelly's model to fit unconstrained natural viewing conditions with a consideration of eye movements. Based on Daly's model, Jia, Lin, and Kassim (2006) estimated the JND thresholds for videos by combining other visual effects such as the luminance adaptation and contrast masking. An improved temporal modulation factor proposed by Wei and Ngan (2008) incorporates not only temporal CSF, but the directionality of motion is also considered. In Niu, Zhang, Krishnan, and Zhang (2009) a video-driven JND profile which incorporates the temporal modulation factor, retinal velocity, luminance adaptation, and block classification was developed.

Previous Watermarking schemes have only partially used the results of the HVS studies (Wolfgang, Podilchuk, & Delp, 1999; Ling, Lu, Zou, & Li, 2006; Podlilchuk & Zeng, 1998; Huang, Shi, & Shi, 1998; Kankanhalli & Ramakrishnan, 1998). Many video watermarking algorithms utilize visual models for still images to increase the robustness and transparency. The perceptual adjustment of the watermark is mainly based on Watson's spatial JND model (Wolfgang, Podilchuk, & Delp, 1999; Ling, Lu, Zou, & Li, 2006; Podilchuk & Zeng, 1998). An image-adaptive watermarking procedure based on Watson's spatial JND model was proposed in Podilchuk and Zeng (1998). In Wolfgang, Podilchuk, and Delp (1999), the DCT-based watermarking approach uses Watson's spatial JND model in which the threshold consists of spatial frequency sensitivity, luminance sensitivity and contrast masking. An Energy Modulated Watermarking Algorithm Based on Watson's spatial JND model was proposed in Ling, Lu, Zou, and Li (2006). During the modulation, Watson's perceptual model is used to restrict the modified magnitude of DCT coefficients. The main drawback of utilizing visual models for still images in video watermarking to increase the robustness and transparency is that it