Signal Processing: *Image Communication* 33 (2015) 1–13

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

journal homepage: www.elsevier.com/locate/image

Exploiting entropy masking in perceptual graphic rendering

Lu Dong [a], Yuming Fang [b], Weisi Lin [a], Chenwei Deng [c,*], Ce Zhu [d], Hock Soon Seah [a]

[a] School of Computer Engineering, Nanyang Technological University, Singapore
[b] School of Information Technology, Jiangxi University of Finance and Economics, China
[c] School of Information and Electronics, Beijing Institute of Technology, China
[d] School of Electronic Engineering, University of Electronic Science and Technology of China, China

ARTICLE INFO

Article history:
Received 4 March 2014
Received in revised form
1 February 2015
Accepted 2 February 2015
Available online 18 February 2015

Keywords:
Entropy masking
Unstructured visual signal
Graphic rendering

ABSTRACT

Since the human visual system (HVS) is the ultimate appreciator of most photorealistically rendered images, rendering process can be accelerated by exploiting the properties of the HVS. According to the concept of entropy masking, the HVS is not sensitive to visual distortions in unstructured visual signals. For structured regions, pixels are highly correlated, while the similarity among pixels in unstructured regions is low. In this paper, we detect unstructured regions by extracting local patches from each pixel and its neighboring pixels, and comparing the similarity between the local patches of the center pixel and the neighboring pixels. We further exploit entropy masking in perceptual rendering, and experimental results demonstrate that the proposed method can accelerate rendering, without degrading the perceived quality of resultant images.

1. Introduction

Physically based rendering produces photorealistic images by simulating the physical behavior of light interaction with surfaces and volumes, and this process is usually time-consuming [1]. Since the human visual system (HVS) is the ultimate appreciator of most photorealistic images, significant computation can be saved by exploiting properties of the HVS [2]. Among various properties, visual masking refers to the decrease in visibility to a visual signal in the presence of image background [3]. That means we can use fewer computational resources in rendering, since low-level artifacts are not visible in regions with strong masking effect.

Existing studies on visual masking include contrast masking and entropy masking [4]. Contrast masking refers to the variation of visibility of visual signals due to the background contrast [5]. Entropy masking refers to the decrease of visibility to a visual signal imposed on a mask signal which is unfamiliar or uncertain to human eyes [6]. For each pixel in an image, the entropy masking effect depends on the structure complexity of a neighborhood [4]. Unstructured regions are less predictable and more difficult to be learned than structured regions, and thus have strong entropy masking effect [7].

Existing studies on entropy masking [4,8,9] try to estimate the masking effect using Shannon entropy [10]. Given a pixel x_o, a luminance histogram is first constructed from pixels in the $m \times m$ local patch P_{x_o} centered at the pixel x_o. The entropy $H(P_{x_o})$ of patch P_{x_o} centered at the pixel x_o is calculated as

$$H(P_{x_o}) = -\sum_{i=0}^{M} p(l_i) \cdot \log p(l_i) \qquad (1)$$

* Corresponding author.
E-mail addresses: s090059@e.ntu.edu.sg (L. Dong), fa0001ng@e.ntu.edu.sg (Y. Fang), wslin@ntu.edu.sg (W. Lin), cwdeng@bit.edu.cn (C. Deng), eczhu@uestc.edu.cn (C. Zhu), ashsseah@ntu.edu.sg (H.S. Seah).

http://dx.doi.org/10.1016/j.image.2015.02.001
0923-5965/

where $l_0, \ldots, l_M$ are luminance values of pixels in patch P_{x_o}, and $p(l_i)$ is the occurrence rate of luminance value l_i in patch P_{x_o}. From Eq. (1), the entropy of an image patch mainly depends on the relative occurrence of pixel values in the patch regardless of the patch structure. With the Shannon entropy, the computed masking effect of an unstructured patch and a structured patch is the same, if these two patches share the same histogram.

The study [11] proposes a method to predict pattern complexity. In the proposed method, unstructured and structured patterns correspond to irregular and regular patterns mentioned in the study [11], respectively. The method [11] is designed based on a gradient count. If the difference between the colors of the pixel and its adjacent pixels exceeds a predefined threshold, the gradient count is one, otherwise, the gradient count is zero. A higher value of pattern complexity is assigned to texture with larger sum of gradient counts. According to the method [11], the gradient counts are zero for pixels in smooth patterns, and the gradient counts are one for pixels in dense, mixed color patterns, and edges. Therefore, the method [11] is able to distinguish smooth patterns from dense and mixed color patterns; however, it cannot distinguish structured patterns from unstructured ones, since both structured and unstructured patterns include edges and the sum of gradient counts from edge pixels are high.

To distinguish structured regions from unstructured regions, we design a method by taking into account the spatial correlation among neighboring pixels. In structured regions, pixels are highly correlated with their neighboring pixels and similar local patterns can be detected in a neighborhood. For each pixel, we extract the local patches centered at the pixel as well as its neighboring pixels, and compute the similarity between the patches of the center pixel and the neighboring pixels. Then the value of the center pixel is predicted from the neighboring pixels with similar local patches to the center pixel. The predicted pixel value is treated as the structured signal, and the difference between the original pixel and the predicted one is treated as the unstructured signal, which indicates the masking effect of the original pixel.

Contrast masking has been exploited in perceptual graphic rendering [12–15] to improve the efficiency. However, entropy masking has not been exploited in perceptual graphic rendering. In this work, we explore the entropy masking properties by proposing a perceptually adaptive sampling scheme. The proposed method distributes samples according to the level of noise as well as the entropy masking effect, and it substantially decreases the rendering cost.

The contribution of the proposed method is twofold. First, we propose a new formulation of entropy masking by detecting unstructured regions. Second, we exploit entropy masking in perceptual graphic rendering by adaptively distributing samples.

2. Related work on perceptual graphic rendering

Perceptual graphic rendering tries to improve rendering quality or decrease rendering computation by modeling properties of the visual perception. Bolin and Meyer [12] built a perceptual model for adaptive sampling by predicting the visible error between two intermediate rendering results, and iteratively distributing new samples to locations with the largest visible error. Ferwerda et al. [16] made an early attempt to take advantage of visual masking in computer graphics. Given a reference image and a distorted image, the method applies spatial frequency tuning and orientation tuning to each image, and computes the responses to different spatial frequency and orientation bands. The computed responses are modulated by the masking functions to account for the masking interactions among the spatial frequency components. The modulated responses of the reference image and the distorted image are compared to determine whether the two images are visually different. The method estimates the visual difference between a reference image and a distorted image but cannot estimate the masking effect of each pixel in the reference image. In the proposed method, we need to estimate the masking effect of each pixel and use the results to determine the number of pixels distributed to each pixel in adaptive sampling scheme.

Ramasubramanian et al. [17] proposed a computationally efficient vision model based on characteristics of the HVS, including luminance adaptation, spatiotemporal contrast sensitivity and visual masking. Given two intermediate rendered images, the vision model computes a threshold map with different visibility on the first image. The difference between the two images is also computed and compared against the visibility threshold map. In the next iteration, more computation is allocated to regions where difference is larger than the threshold. Myszkowski et al. [18] extended the use of human vision model to animation rendering. They developed the animation quality metric (AQM) by extending an existing vision model called visible difference predictor (VDP) [3]. Later they applied AQM to walk-through animation rendering [19] and global illumination computation for animation rendering [20]. A more detailed introduction of perceptual image rendering can be found in [2,21].

All perceptual rendering algorithms mentioned above include the evaluation of a full human vision model during image synthesis. Computational complexity is a major concern in these perceptual rendering methods, since the vision models used are generally sophisticated and it is time-consuming to continuously update the vision models. Therefore, perceptual rendering methods concentrating on visual masking have been proposed by detecting regions with strong masking effect to decrease rendering computation in these regions. They aim to benefit from the perceptual properties of image content with less computational complexity.

The method [14] computes visual masking properties of image content by using components of JPEG compression standard. The assumption in [14] is that at moderate compression rates, the quantization errors in JPEG compression are largely unnoticeable, i.e. the error between the original and JPEG decoded images is less than a masking threshold of the image content. Visual masking effect is estimated as the difference between original and JPEG decoded images.

Qu and Meyer [15] proposed to compute the visual masking properties of image content based on a computational model of human vision. The method does not use the vision model to predict the difference between two intermediate rendering