A NEW SPATIAL ACTIVITY METRIC FOR FILM CONTENTS

Xiaoan Lu, Jiefu Zhai, and Cristina Gomila

Corporate Research, Thomson Inc., Princeton, NJ 08540 U.S.A.

ABSTRACT

The masking property of human vision systems has been successfully applied in various image/video applications. To invoke the spatial masking effect, it is important to design a metric that effectively identifies the spatial activity of a region. This metric indicates which areas are more textured and more artifacts can be masked. We review three widely used metrics and evaluate their performance in context of film content. We observe that these metrics have strong dependencies on the brightness. More specifically, for smooth areas with film grain, these metrics usually assign greater degrees of texture to the bright regions and lower degrees to dark ones. This causes problems in the bright areas that are mistakenly identified as more textured than the dark areas. Utilizing the property of film grain, we explain the origin of this dependency and propose a new spatial activity metric that removes the dependency on the brightness. In our simulation, we use this new metric in the rate control algorithm of a MPEG-2 video encoder. The result shows more homogeneous film grain in the reconstructed pictures and improved visual quality.

Index Terms— Video signal processing, Noise

1. INTRODUCTION

The masking property of human vision systems has been successfully applied in the areas of image/video quality assessment, image/video compression and information hiding. The masking occurs because of the inability of the human perceptual mechanism to distinguish two signal components in the same spectral, temporal, or spatial locality. Invoking the masking effect at the right location requires accurate models of the human visual system [1]. Significant progress has been made in exploiting the masking effects. In a MPEG-2 reference software [2], the quantization stepsize of a macroblock (MB) is modulated by the spatial activity to obtain high visual quality. In [3], a MB is characterized according to homogeneity, flatness, various degrees of texture, and strength of edges. Then adaptive quantization is applied based on its activity class and the global scene complexity in a video encoder. In [4], the masking effect is utilized to decide where to insert digital watermarking.

In general, the spatial activity metric is used to indicate the degree of texture or flatness of an area and can be used to determine the amount of distortion or inserted information to be allowed in the area. In this paper, we focus on how to effectively measure the spatial activity of a block in the presence of film grain. The problem is to design a reliable metric that effectively identifies the degree of texture of a region in a picture. In Section 2, we first review three types of existing spatial activity metrics and observe that in film contents they all strongly depend on the brightness. To be more specific, these metrics assume greater (lower) degrees of texture to bright regions and lower (greater) degrees to dark ones even these areas share similar homogeneity or similar degrees of flatness. We use an example of a MPEG-2 encoder to explain that such metrics will introduce visible blockiness and loss of grainy appearance in the bright regions. It also caused inconsistent quality in similarly smooth areas with various brightness. We explore the origin of the dependency on the brightness using the property of film grain and explain why the existing metrics are not reliable for film contents.

Film grain can be regarded as an additive, signal dependent noise. It is a technical effect used by many cinematographers to transmit a certain mood and tone to the movie and it is clearly noticeable in many high-definition movies. Hence it is important to preserve them for the subjective quality. Based on the property of the film grain, we develop in Section 3 our new metric that removes this dependency. We simulate this metric and use it in the rate control algorithm of a MPEG-2 video encoder. The result shows more preserved film grain and therefore improved visual quality for pictures with film grain. We discuss our future work and conclude our paper in Section 4.

2. EXISTING METRICS

In this section, we review three widely used types of metrics: (1) variance-based; (2) gradient-based; and (3) DCT-based metrics [5]. The goal of these metrics is to distinguish flat and homogeneous regions, where distortion is more visible to human eyes, from busy and textured areas, where distortion is masked and less visible. All metrics are calculated on a 16x16 MB basis.

Variance-based metric This approach measures the spatial activity using the variance of luminance. A representative method is the one used in the rate control algorithm of the

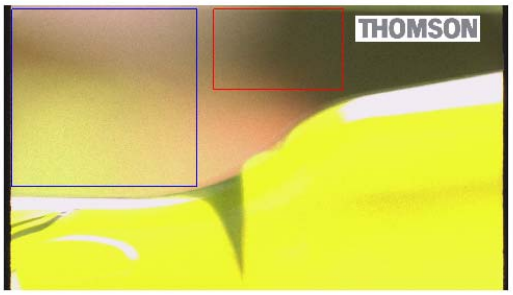

Fig. 1. Frame 416 from "Pouring Liquids" with two marked flat regions.

MPEG-2 reference software [2]:

$$ACT_{var} = 1 + \min_{i=1,2,3,4}(var_i), \tag{1}$$

where var_i is the variance for i^{th} 8x8 subblock.

Gradient-based metric This approach considers the gradients. One metric is described in [6]:

$$ACT_{gra} = \sum_{i=0}^{15}\sum_{j=0}^{15} \max_{n}(grad_{i,j,1}, \cdots, grad_{i,j,4}), \tag{2}$$

where $grad_{i,j,n}$ is a local gradient computed by one of four 5x5 directional high-pass filters at (i, j) [6].

DCT-based metric This type of approach uses the DCT coefficients of luminance values. One such metric normalizes the AC coefficients by the DC coefficient [5]:

$$ACT_{DCT} = \frac{1}{16 \times 16}\sum_{i=0}^{15}\sum_{j=0}^{15}\frac{F^2(i,j)}{F^2(0,0)} - 1. \tag{3}$$

where $F(i, j)$ is the DCT coefficient of frequency (i, j).

2.1. Metric Evaluation

In general, the spatial activity metric is used to indicate at which quality a MB should be compressed [2], [6], and more distortion or inserted information is allowed in the block with higher spatial activity because of the masking effect. For example, in a perceptual video encoder the smooth areas are expected to be compressed with finer quantization and the textured ones with coarser quantization. In the following, we evaluate how these metrics measure spatial activity of pictures with film grain. We further use ACT_{var} as an example to evaluate how it performs when it is used in the rate control algorithm of a MPEG-2 encoder. The problems we describe for this example also apply to other spatial activity metrics.

We take a public high-definition sequence [1], "Pouring Liquids" (1920 × 1080, 24 fps, progressive), for our illustration purpose. This sequence contains film grain and is overall flat. Fig. 1 shows Frame 416 in it. We plot for this frame ACT_{var}, ACT_{gra}, and ACT_{DCT} in Fig. 2 for the masked flat regions versus the average luminance of the corresponding MBs. We observe that the metrics strongly depend on the brightness and have big dynamic ranges despite the regions share similarly in visual flatness.

[1] While the advantage of this algorithm is more evident in other sequences with stronger film grain, we only present results in this paper for "Pouring Liquids" because of copyright restrictions.

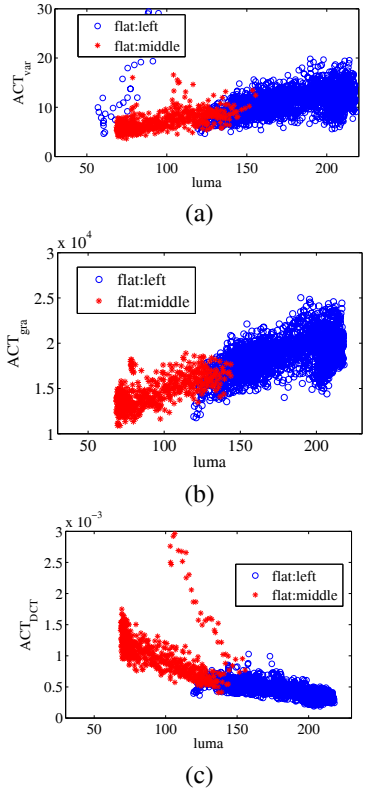

Fig. 2. (a) ACT_{var}, (b) ACT_{gra}, and (c) ACT_{DCT} for the flat regions.

Assuming we apply the MB-level quantization stepsize modulation used in the MPEG-2 reference software [2] as

$$w = \frac{2 \times ACT_{var} + \overline{ACT_{var}}}{ACT_{var} + 2 \times \overline{ACT_{var}}}, \tag{4}$$

where w is the weight on the quantization stepsize, and $\overline{ACT_{var}}$ is the average spatial activity measured by the variance. The particular picture quality and rate of the encoder is achieved by selecting a specific quantizer scale $Q_{i,m}$ for each MB m in picture i. This value is calculated for each MB by combining a picture global quantization scale Q_i with the perceptual weighting factor $w_{i,m}$:

$$Q_{i,m} = w_{i,m} \times Q_i, \tag{5}$$

where $w_{i,m}$ is the weight for MB m in picture i. This perceptual modulation factor aims to obtain the same visual quality for each encoded MB in a picture. For the picture we consider, ACT_{var} can be as small as 4 in the dark masked region and as big as 15 in the bright masked region. Using

(a) (b)

Fig. 3. (a) A block from the bright masked area from Frame 416 of "Pouring Liquids" and $Q_{i,m} = 8$. (b) A block from the dark masked area and $Q_{i,m} = 6$.

(4) and $\overline{ACT_{var}} = 7.8$, w for a MB with $ACT_{var} = 4$ and $ACT_{var} = 15$ will be 0.81 and 1.24, respectively. This great weighting factor gap is very likely to cause visible artifacts, such as blockiness or blurness, in the MBs where ACT_{var} are big.

We encode the sequence at 14 Mbps using this encoder. We select a patch from each masked area of Frame 416, where $Q_{i,m} = 6$ for the dark patch and $Q_{i,m} = 8$ for the bright one, for illustration purpose in Fig. 3. We observe that the dark patch appears closer to the original and the bright patch sees blockiness patterns[2]. Since the film grain construction is very important for the subjective picture quality, especially for the movie industry [7], a more intelligent encoder should have provided similar quality in both patches since they appear to be similarly flat. This imposes a challenge on the spatial activity metric that it should assign similar values to regions with similarly visual smoothness. In the next, we explore the origin of the dependency on the brightness using the property of film grain and explain why the existing metrics are not reliable for film contents.

2.2. Explanation from Film Grain Models

We observe in the above that the existing metrics strongly depend on the brightness in the context of film contents. One reason of this strong dependency is the presence of film grain, a random texture generated during the process of film development. Film grain could be regarded as an additive, signal dependent noise, which differs in size, shape and intensity depending on the film stock, lightening condition and development process. Studies in [8, 9] show that the intensities of film grain are highly correlated to pixel intensities. In [9], it shows that film grain can be modeled as:

$$g(i,j) = f(i,j) + f(i,j)^{\gamma} * n(i,j), \tag{6}$$

where $g(i,j)$ and $f(i,j)$ are the observed and noise-free pixel value at location (i,j), respectively, γ is a constant given the film stock and shooting condition, and $n(i,j)$ is a zero mean normal distributed noise. Usually γ is between 0.3 - 0.7, and in most cases around 0.5. For a smooth $M \times N$ block where $f(i,j)$ is close within a block, we approximate (6) as:

$$g(i,j) = f(i,j) + \overline{f}^{\gamma} \times n(i,j), \tag{7}$$

where $\overline{f} = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} f(i,j) \approx f(i,j)$. In the flat areas, $f(i,j)$ is almost a constant and the strength in ACT_{var} is mainly from the film grain, which strongly correlates with the brightness decided by $f(i,j)$. Because existing metrics do not consider the effect of film grain, they all show strong correlation with the brightness and are not effective for film contents. This matches our observation in Fig. 2.

3. A NEW SPATIAL ACTIVITY METRIC

In Section 2, three types of spatial activity metrics are evaluated. We observe that they strongly correlates to the brightness. We provide an example where ACT_{var} is used in the rate control algorithm of a MPEG-2 encoder and visible artifacts exist in the reconstructed pictures. We explain the origin of the dependence using the property of film grain. We then develop a new metric that considers the effect of film grain. Using ACT_{var} as an example, we explain how we eliminate the dependency. The methodology can be applied to other metrics.

Assuming $n(i,j)$ is independent of $f(i,j)$ in (6), the variance of $g(i,j)$ becomes:

$$\sigma_g^2 = \sigma_f^2 + \overline{f}^{2\gamma} \times \sigma_n^2, \tag{8}$$

where σ_g^2, σ_f^2 and σ_n^2 are the variance for $g(i,j)$, $f(i,j)$ and $n(i,j)$, respectively. For the grain-free content, σ_n^2 is negligible and σ_g^2 represents the variation within a MB and can be a good measure for the spatial activity. For the grainy content, we need remove the effect of the film grain when computing the spatial activity metric, as the human vision system does, and calculate the variance of noise-free signal in order to obtain a good measure:

$$\sigma_f^2 = \sigma_g^2 - \overline{f}^{2\gamma} \times \sigma_n^2. \tag{9}$$

Note that $\overline{f}^{2\gamma}$ describes the dependency of the film grain on the signal and σ_n^2 describes the noise strength. We denote $\sigma_{grain}^2 = \overline{f}^{2\gamma} \times \sigma_n^2$.

3.1. Parameter Estimation

In the film content, especially for high resolution video, there usually exists a large amount of blocks that are flat or almost flat. The variances of these blocks are mainly contributed by film grain, i.e, $\sigma_f^2 \approx 0$, $\sigma_g^2 \approx \sigma_n^2$. For the flat blocks with similar brightness, the characteristic of the grain are almost the same, which result in very close σ_g^2 that is smaller than the variance of textured MBs. As a consequence, the histogram of the variance usually has a peak at the beginning.

[2]The difference may not be clear in the printout. It will be more evident when we watch this on a computer monitor or a 1080p TV screen.

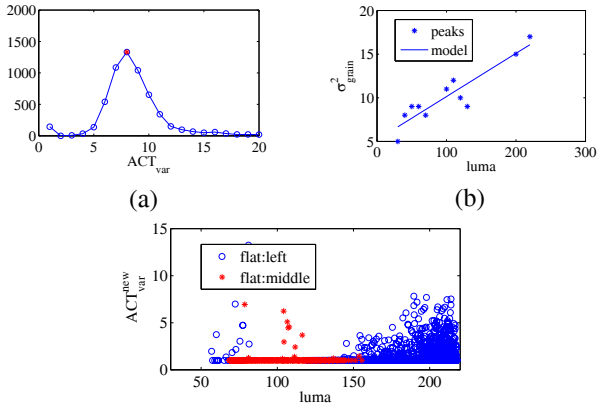

Fig. 4. (a) The histogram of ACT_{var} for a luma range of 70-79; (b) curve fitting for σ^2_{grain}; and (c) ACT^{new}_{var} for Frame 416 from "Pouring Liquids".

Fig. 5. (a) A block from the bright masked area from Frame 416 of "Pouring Liquids" and $Q_{i,m} = 7$. (b) A block from the dark masked area and $Q_{i,m} = 7$.

Therefore we introduce a histogram-based method to estimate the grain intensity. The blocks are first classified into m groups according to its brightness range. For each group, we calculate the histogram of variances and identify the first peak $\sigma^2_{peak,i}$, as in Fig. 4(a). In Fig. 4(b), using $\sigma^2_{peak,i}$ from all brightness ranges and assuming $\gamma = 0.5$ in (8), we derive σ^2_{grain} as a linear function of the brightness using the linear regression method. To eliminate the effect of film grain, we deduct it from the existing metric:

$$ACT^{new}_{var} = ACT_{var} - m(\sigma^2_{grain}). \qquad (10)$$

In our approach, we set $m(\sigma^2_{grain}) = \sigma^2_{grain} - 1$ for simplicity.

3.2. Simulation Results

We run the same simulation as we did in Section 2 and the resulted reconstructed patches are shown in Fig. 5. We observe from Fig. 4(c) that the new metric ACT^{new}_{var} provides a smaller dynamic range for the smooth areas. Under this new metric, $ACT^{new}_{var} = 1$ and $Q_{i,m} = 7$ for both patches. Because of the same quantization scale, both patches are encoded at similar quality and have well preserved the film grain.

4. DISCUSSIONS

In this paper, we review existing spatial activity metrics and observe that for film contents they all show strong dependency on the brightness. More specifically, these metrics assume greater (lower) degrees of texture to bright regions and lower (greater) degrees to dark ones even these areas have similar degrees of flatness. We use an example of a MPEG-2 encoder to illustrate that such metrics introduce visible blockiness and loss of grainy appearance in the bright regions. We also show that the encoder using such metrics causes inconsistent quality in similarly smooth areas with various brightness. Using the concept of film grain, We explore the origin of this dependency and propose a new spatial activity metric that removes the dependency on the brightness. Our simulation results show that the new metric provides close values to the areas with similar degrees of flatness despite the variation in the brightness. The encoding results show a more homogeneous and improved visual quality for the same content. Therefore this new spatial activity metric allows us to successfully exploit the spatial masking effect in contents with film grain.

In this paper, we assume a linear relation between brightness and grain intensity. However, in practice, we observe that in some cases the grain intensity would drop after brightness reaches a considerable high level. It is our future work to use a more accurate model to derive the spatial activity metric.

5. REFERENCES

[1] N. Jayant, J. Johnston, and R. Safranek, "Signal compression based on models of human perception," *Proc. of the IEEE*, vol. 81, pp. 1385–1422, October 1993.

[2] ISO/IEC JTC1/SC29/WG11 N0400, "Test model 5," April 1993.

[3] A. Puri and R. Aravind, "Motion-compensated video coding with adaptive perceptual quantization," *IEEE Trans. on CSVT*, vol. 1, pp. 351–361, December 1991.

[4] I. Cox, M. Miller, and J. Bloom, *Digital Watermarking: Principles and Practice*, Morgan Kaufmann, San Mateo, CA.

[5] W. J. Kim, J. W. Yi, and S. D. Kim, "A bit allocation method based on picture activity for still image coding," *IEEE Trans. on Image Proc.*, vol. 8, pp. 974–977, July 1999.

[6] X. Yang, W. Lin, Z. Lu, X. Lin, S. Rahardja, E. Ong, and S. Yao, "Rate control for videophone using local perceptual cues," *IEEE Trans. on CSVT*, vol. 15, pp. 496–507, 2005.

[7] T. Wedi, Y. Kashiwagi, and T. Takahashi, "H.264/AVC for next generation optical disc: a proposal on profiles," *JVT-K025*, March 2004.

[8] H. H. Arsenault and M. Denis, "Integral expression for transforming signal-dependent noise into signal-independent noise," *Optics Letters*, vol. 6, pp. 210–212, October 1981.

[9] J. C. K. Yan and D. Hatzinakos, "Signal-dependent film grain noise removal and generation based on higher-order statistics," in *SPW-HOS*, 1997, pp. 77–81.