

# Performance evaluation of HD camcorders: measuring texture distortions using Gabor filters and spatio-velocity CSF

Kongfeng Zhu, Dietmar Saupe

Department of Computer and Information Science, University of Konstanz, Germany

## ABSTRACT

This paper presents a new method of measuring physical texture distortions (PhTD) to evaluate the performance of high definition (HD) camcorders w.r.t. motion and lossy compression. It is extended to measure perceptual texture distortions (PeTD) by taking into account the spatio-velocity contrast sensitivity function of the human visual system. The PhTD gives an objective (physical) distortion of texture structures, while the PeTD measures the perceptual distortion of textures. The dead leaves chart, invariant to scaling, translation, rotation, and contrast, was selected as a target texture. The PhTD/PeTD metrics of the target distorted by camcorders were measured based on a bank of Gabor filters with eight orientations and three scales. Experimental results for six HD camcorders from three vendors showed: 1) the PhTD value increases monotonically w.r.t. the motion speed, and decreases monotonically w.r.t. the lossy compression bitrate; 2) the PeTD value decreases monotonically w.r.t. the motion speed, but stays almost constant w.r.t. the lossy compression bitrate. The experiment gives a reasonable result even if the distortions are not radially symmetric. However, some subjective tests should be done in future work to validate the performance of the perceptual texture distortion metric.

**Keywords:** Texture distortion, dead leaves chart, Gabor filter bank, spatio-velocity contrast sensitivity function, perceptual video quality assessment, compression bitrate, motion

## 1. INTRODUCTION

Digital camcorders are camcorders producing digital video files as the output. A digital camcorder is normally composed of three main components: the lens system, the image sensor, and the recorder. All the three components of a camcorder can cause quality loss in recorded digital videos. Their influences on the final visual quality of the recorded video are often mixed in a complicated way, thus making it a challenging task to assess the performance of camcorders based on the final visual quality of recorded videos.

Among all the visual quality performance factors, blurring and noise are widely studied. These two factors are closely related. Improving one factor often leads to compromise of the other. Another related visual quality factor is about how well texture regions in a video are preserved. As important local structures in digital videos, they can be heavily blurred if denoising filtering is applied to them, which will normally cause a clearly visible deterioration of visual quality. Motion can also cause loss of texture due to the averaging effect of neighboring pixels in both spatial and temporal domains. Therefore, it is important to assess to what extent textures are preserved by a camcorder when we evaluate its overall performance in terms of producing high-quality videos. In this paper we propose an objective method to achieve this goal.

Compared with digital still cameras,<sup>1-7</sup> quantitative performance evaluation of digital camcorders is a less studied topic in the literature. This is not surprising because many techniques used for measuring digital still cameras can be applied to digital camcorders in exactly the same way by taking each video frame as a still picture. However, since object motion plays a key role in visual quality of recorded videos, a good measurement of the performance of a camcorder should relate the quality measurement to the level of motion. Generally speaking, a camcorder that can still maintain a high quality for digital videos with fast object motion is considered as a good camcorder. The need of including object motion in the loop calls for performance evaluation methods tailored to digital camcorders.

---

Contact information of the authors: Kongfeng Zhu: Kongfeng.Zhu@uni-konstanz.de; Dietmar Saupe: Dietmar.Saupe@uni-konstanz.de.

In some previous work,<sup>1,2,4</sup> several methods of evaluating the texture preservation capability of images produced were developed based on a special test target called the dead leaves chart, which was originally proposed by Matheron in 1968 for mathematical morphology.<sup>3,8</sup> The dead leaves chart was chosen because it simulates a scene with textures of different scales and it is invariant to many image processing operations. The main idea behind those methods is to measure the level of texture preservation by analyzing the spatial frequency response (SFR) of the images produced by digital still cameras. The SFR was computed simply by first computing the two-dimensional Fourier transform of an image of the dead leaves chart produced by the camera under study and then comparing its radial power spectral density (PSD) with that of the chart itself, assuming the distortion of the digital still camera system is radially symmetric. The dead leaves chart was generated by computer and then printed out with a calibrated printer, so its PSD can be considered as known. In more recent work,<sup>9</sup> the SFR-based method was extended to evaluate the performance of camcorder systems by taking into consideration motion and compression. A texture preservation ratio (TPR) was proposed to quantify to what extent texture structures are preserved in a video with respect to motion speed and compression bitrate.

For digital camcorder systems, often the texture distortion is not radially symmetric, so the SFR-based methods are not sufficient to fully capture the distortion. This work improves previous research on digital camcorders<sup>9</sup> to cover asymmetric texture distortion, for example the distortion introduced by motion. It is based on extracting Gabor energy features for 8 orientations and 3 scales, and considers not only object motion and lossy video compression, but also the human visual system (HVS). The dead leaves chart is generated, printed out, and a video of the chart is acquired by a camcorder system. Then,  $8 \times 3 = 24$  Gabor filters are applied to the original distortion-free test chart and also to the distorted target in the video. The Euclidean distance between the two results is then defined as a scalar metric of physical texture distortion (PhTD). Furthermore, to measure perceptual texture distortion (PeTD), both the original and distorted targets are weighted by the spatio-velocity contrast sensitivity function (SV-CSF) of the HVS in the frequency domain before Gabor energy features are extracted. Each of the PhTD and PeTD is a single non-negative value, and a higher value indicates more (physical or perceptual) distortions caused by the digital camcorder.

It is expected that a distorted video with a low compression bitrate or a fast moving object has high PhTD and PeTD measurements. This expectation of PhTD measurement was confirmed by experiments on six camcorders from three different vendors, as the PhTD value increases monotonically w.r.t. the motion speed, and decreases monotonically w.r.t. the lossy compression bitrate. The PeTD value increases monotonically w.r.t. the motion speed as we expected, but surprisingly stays almost constant w.r.t. the lossy compression bitrate, which is out of our expectation.

The rest of the paper is organized as follows. In the next section, related work is reviewed. Our method for measuring the (perceptual) texture distortion is proposed in Section 3. In Section 4, the experimental setup and the process of collecting recorded videos is described in detail. Section 5 reports our data analysis of the recorded videos and interpretation of the experimental results. Conclusions and future work are briefed in the last section.

## 2. RELATED WORK

This section surveys previous work in digital imaging system evaluation, such as the dead leaves chart and linearization, Gabor filters, and the spatio-velocity contrast sensitivity function of the HVS.

### 2.1 The dead leaves chart

The dead leaves chart is generated by a stochastic image model simulating overlapped and occluded objects as commonly found in natural scenery. Following the model, a image is generated by placing shapes of the same kind (called leaves, but not necessarily looking like real leaves) with random radii and random grey scales at random positions onto a canvas until the entire canvas is fully covered.

The dead leaves chart has been widely used to evaluate the image quality of a digital camera<sup>1,2,4</sup> due to its invariance to scale, rotation, shift, and contrast (exposure). It is very close to images of nature in some elementary statistics, such as gradient distribution, wavelet coefficients, co-occurrence values, and power spectrum. In particular, its power spectral density follows a power law. It is generated by computer and then printed out on a calibrated printer, so its PSD can be considered as known.

## 2.2 Linearization

When evaluating the performance of an imaging system or image capture subsystem, it is necessary to transform the observed output image data into an equivalent input exposure. The transformation is usually done by measuring the input-output characteristic, called the opto-electronic conversion function (OECF).<sup>10</sup> The OECF is nonlinear in general, so the contrast will be amplified or reduced in a nonlinear way, which causes loss of gray tones. It is expected that the input of an imaging system will be distorted due to the nonlinear OECF, even if other parts of the system are perfect and lead to no extra distortion. Therefore, the inverted OECF is usually applied to the output of an imaging system to obtain linear data. This step is called *linearization*.

The flexibility of digital systems complicates the determination and presentation of the functional relationship between the cameras' optical input and digital output signal level, thus the OECF curve is usually unknown when evaluating a digital image/video acquisition system, and it can be changed in an unknown way due to the adaptive nature of image processing. To assure that the relationship between optical input and digital output of the digital image/video acquisition system is obtained accurately, a test chart with patches of different grey levels are necessary for OECF measurement. The patches of an OECF chart can be aligned in a straight line, a circle around the center, or in any other fashion, as long as the luminance values (relative to paper white) are known for each patch. In our case, we chose an OECF chart with 21 patches aligned in one line, and the grey levels are graduated linearly, since we are most interested in linearizing the midtones in the captured images.

## 2.3 The spatio-velocity contrast sensitivity function (SV-CSF)

The CSF describes the sensitivity of the HVS to recognize patterns as a function of their spatial frequency. The CSF which is measured with moving stimuli is called spatio-velocity CSF (SV-CSF). Kelly in 1979 proposed a SV-CSF model as a function of spatial and temporal frequency.<sup>11</sup> Daly in 1998 modified this model by taking into account the natural drift, smooth pursuit, and saccadic eye movements.<sup>12</sup> Daly's SV-CSF model, which is also called the unconstrained eye movement SV-CSF, is given in Eq. (1) to (3):

$$\text{CSF}(\rho, v_R) = k \cdot c_0 \cdot c_2 \cdot v_R \cdot (c_1 2\pi\rho)^2 \cdot \exp\left(-\frac{c_1 4\pi\rho}{\rho_{max}}\right) \quad (1a)$$

$$k = s_1 + s_2 \cdot \left| \log\left(\frac{c_2 v_R}{3}\right) \right|^3 \quad (1b)$$

$$\rho_{max} = \frac{p_1}{c_2 v_R + 2} \quad (1c)$$

where  $\rho$  is the spatial frequency of the visual stimulus in *cy/deg*. The variable  $v_R$  is retinal velocity in *deg/sec*.  $s_1 = 6.1, s_2 = 7.3, p_1 = 45.9$  are the original constants provided by Kelly. The scale factor  $k$  and  $\rho_{max}$  are responsible for the vertical shift and horizontal shift of the peak sensitivity, respectively. Constants  $c_0, c_1, c_2$  are modifications to the model that allow fine tuning.

For practical applications it is important to take into account the observer's eye movements during the presentation of moving objects in a video sequence. The model of the eye velocity  $v_E$  as a function of target velocity  $v_I$  is given in Eq. (2). The image plane velocities are converted to the retinal velocity in Eq. (3), such that the retinal spatio-velocity model can be used to determine the sensitivity to image plane spatio-velocities.

$$v_E = \min((g_{SP} \cdot v_I) + v_{MIN}, v_{MAX}) \quad (2)$$

$$v_R = v_I - v_E \quad (3)$$

where  $g_{SP}$  is the gain of the smooth pursuit eye movements,  $v_{MIN}$  is the minimum eye velocity, and  $v_{MAX}$  is the maximum eye velocity. The values of  $g_{SP}, v_{MIN}$  and  $v_{MAX}$  are set to 0.82, 0.15, and 80.0 respectively.

## 2.4 Gabor filters

A two dimensional Gabor function can be written as:<sup>13</sup>

$$g_{\lambda, \Theta, \gamma}(x, y) = \exp(-(x'^2 + \gamma^2 y'^2)/(2\sigma^2)) \cdot \exp\left(j2\pi \frac{x'}{\lambda}\right) \quad (4a)$$

$$x' = x \cos(\Theta) + y \sin(\Theta) \quad (4b)$$

$$y' = -x \sin(\Theta) + y \cos(\Theta) \quad (4c)$$

A bank of Gabor filters consisting of several filters of  $M$  scales and  $N$  orientations is generated from the 2D-filter by appropriate dilations and rotations,<sup>14</sup> i.e.,

$$\lambda = 2^{S_m} \quad (5a)$$

$$\Theta = n\pi/N \quad (5b)$$

where  $S_m$  is nonnegative for  $m = 0, \dots, M - 1$  and  $n = 0, \dots, N - 1$ . Typically, an input image  $I(x, y)$ ,  $(x, y) \in \Omega$  ( $\Omega$  is the set of image points), is convolved with a 2-D Gabor function  $g_{\lambda, \Theta, \gamma}(x, y)$ ,  $(x, y) \in \Omega$ , to obtain a Gabor feature image  $r_{\lambda, \Theta}(x, y)$  as follows:<sup>13</sup>

$$r_{\lambda, \Theta}(x, y) = \iint_{\Omega} I(\xi, \eta) g_{\lambda, \Theta, \gamma}(x - \xi, y - \eta) d\xi d\eta \quad (6)$$

The output of Gabor kernel filter  $r_{\lambda, \Theta}(x, y)$  in each image point can be combined in a single quantity that is called the *Gabor energy*.

### 3. PROPOSED METHOD FOR MEASURING TEXTURE DISTORTION

The SFR-based methods are based on the radial power spectral density assuming that the distortion is radially symmetric. However, the texture distortion may not be radially symmetric due to motion and lossy compression in digital camcorders. Therefore our proposed method aims to improve the SFR-based methods by applying a bank of Gabor filters of  $M$  scales and  $N$  orientations. The Gabor features of the original distortion-free test chart and the distorted target in the video are extracted, and the square of their Euclidean distance is defined as a scalar metric of physical texture distortion (PhTD). It is further extended to measure perceptual texture distortion (PeTD) by taking the HVS into consideration. Namely, both the original and distorted targets are weighted by the SV-CSF of the HVS in the frequency domain before Gabor energy features are extracted. Each of these measurements of physical and perceptual texture distortion result in a single non-negative value. Larger values correspond to larger (physical or perceptual) distortion caused by the digital camcorder. The proposed PhTD and its extension PeTD are detailed in the rest of this section.

#### 3.1 Physical texture distortion (PhTD)

Let  $I(x, y; v, b)$  be a video frame of the target object moving at speed  $v$ , compressed with bitrate  $b$ .  $x$  and  $y$  are horizontal and vertical spatial coordinates. The Gabor energy image in the frequency domain is defined as:

$$R_{\lambda, \Theta}(\xi, \eta; v, b) = O(\xi, \eta; v, b) \cdot G_{\lambda, \Theta, \gamma}(\xi, \eta) \quad (7)$$

where  $O(\xi, \eta; v, b)$ ,  $G_{\lambda, \Theta, \gamma}(\xi, \eta)$  are the PSD of  $I(x, y; v, b)$ ,  $g_{\lambda, \Theta, \gamma}(x, y)$  respectively. To merge the Gabor energy image  $R_{\lambda, \Theta}(\xi, \eta; v, b)$  into a single value, the Gabor energy feature is defined as

$$E_{v, b}(m, n) = \log_{10} \left( \sum_{\xi} \sum_{\eta} R_{\lambda, \Theta}(\xi, \eta; v, b) \right) \quad (8)$$

where  $m = 0, \dots, M - 1$ ,  $n = 0, \dots, N - 1$  are indices for the  $M$  scales and  $N$  orientations in Eq. (5). These numbers are collected in a matrix of Gabor features,

$$\mathbf{E}_{v, b} = \begin{pmatrix} E_{v, b}(0, 0) & E_{v, b}(0, 1) & \cdots & E_{v, b}(0, N - 1) \\ E_{v, b}(1, 0) & E_{v, b}(1, 1) & \cdots & E_{v, b}(1, N - 1) \\ \vdots & \vdots & \ddots & \vdots \\ E_{v, b}(M - 1, 0) & E_{v, b}(M - 1, 1) & \cdots & E_{v, b}(M - 1, N - 1) \end{pmatrix}_{M \times N} \quad (9)$$

The physical texture distortion (PhTD) is finally defined as the squared Euclidean distance between Gabor energy features of the undistorted image\* and the distorted camera image,

$$\text{PhTD}_{v, b} = \sum_m \sum_n (E_{v, 0}(m, n) - E_{v, b}(m, n))^2 \quad (10)$$

\*Note that the undistorted image in our experiment is the ideal dead leaves chart generated by computer without motion.

where  $E_{0,0}(m, n)$  is the Gabor energy feature of the undistorted image at orientation  $n$  and scale  $S_m$ . It is a single and non-negative value. The larger the distortion, the higher the value.

### 3.2 Perceptual texture distortion (PeTD)

To take the HVS into consideration, the unconstrained eye movement SV-CSF model is introduced to filter the two-dimensional PSD of the object,

$$O'(\xi, \eta; v, b) = O(\xi, \eta; v, b) \cdot \text{CSF}_{filter}(\xi, \eta; v) \quad (11)$$

where  $O'(\xi, \eta; v, b)$  is named the *perceptual PSD* of the object image moving at speed  $v$  in the video with compression bitrate  $b$ . The  $\text{CSF}_{filter}(\xi, \eta; v)$  is a 2D filter generated from  $\text{CSF}(\rho, v)$  in Eq. (1) with  $\rho = \sqrt{\xi^2 + \eta^2}$ , i.e.:

$$\text{CSF}_{filter}(\xi, \eta; v) = \text{CSF}(\sqrt{\xi^2 + \eta^2}, v). \quad (12)$$

The PSD  $O(\xi, \eta; v, b)$  in Eq. (7) is replaced by  $O'(\xi, \eta; v, b)$  to get the perceptual texture distortion, thus Eqs. (7) to (10) are modified to:

$$R'_{\lambda, \Theta}(\xi, \eta; v, b) = O'(\xi, \eta; v, b) \cdot G_{\lambda, \Theta, \gamma}(\xi, \eta) \quad (13)$$

$$E'_{v,b}(m, n) = \log_{10} \left( \sum_{\xi} \sum_{\eta} R'_{\lambda, \Theta}(\xi, \eta; v, b) \right) \quad (14)$$

$$\mathbf{E}'_{v,b} = \begin{pmatrix} E'_{v,b}(0, 0) & E'_{v,b}(0, 1) & \cdots & E'_{v,b}(0, N-1) \\ E'_{v,b}(1, 0) & E'_{v,b}(1, 1) & \cdots & E'_{v,b}(1, N-1) \\ \vdots & \vdots & \ddots & \vdots \\ E'_{v,b}(M-1, 0) & E'_{v,b}(M-1, 1) & \cdots & E'_{v,b}(M-1, N-1) \end{pmatrix}_{M \times N} \quad (15)$$

$$\text{PeTD}_{v,b} = \sum_m \sum_n (E'_{0,0}(m, n) - E'_{v,b}(m, n))^2 \quad (16)$$

where  $E'_{0,0}(m, n)$  is the perceptual Gabor energy feature of the undistorted image at orientation  $n$  and scale  $S_m$ .  $\text{PeTD}_{v,b}$  is a single and non-negative value, stands for the perceptual texture distortion of the object at speed  $v$  in the video with compression bitrate  $b$ . The value increases with perceptual distortion.

## 4. EXPERIMENTAL SETUP

### 4.1 Dead leaves chart generation

The disks for the dead leaves chart were generated following [4], initially for a large square image of linear size  $N = 2^{15} = 32768$ . This large image was lowpass filtered using a box filter of size  $s \times s$  and subsampled to yield an anti-aliased dead leaves image of size  $L = N/s$ . The lower bound for the radii of the disks in the dead leaves image was set such that the resulting disks, printed on paper and recorded by a digital camcorder, are smaller than one pixel of the camcorder image sensor array. Here,  $r_{\min} = N/4096$ . The maximal radius is set such that the corresponding maximal disks are about the size of the entire image. We followed the choice of J. McElvain et al.<sup>1</sup> to choose  $r_{\max} = 497r_{\min}$ .

The resulting  $L \times L$  image was printed out on a calibrated high-resolution printer (Epson Stylus Photo R3000 Inkjet) to make the test chart with a physical size of  $21\text{cm} \times 21\text{cm}$ . The printing resolution is 300 DPI (dots per inch). The printer was calibrated to behave linearly in the range of the gray scale selected to produce the dead leaves image. A grey scale consisting of 21 patches with linear exposure levels (labeled 0-20) is also printed out for the purpose of printer calibration and OECF measurement. The grey scale of length  $21\text{cm}$  is put next to the dead leaves chart. Thus, the whole chart contains two sub-charts: the dead leaves chart and the grey scale chart.

## 4.2 Camcorders tested

Six camcorder models from three vendors were tested with interlacing mode in our experiments:

- Four camcorder models of Panasonic: HDC SD800, HDC SDX1, and HDC TM80 with four recording (quality) modes (all with full HD resolution  $1920 \times 1080$ ) HA 1920 (17 Mbps VBR), HG 1920 (13 Mbps VBR), HX 1920 (9 Mbps VBR), and HE 1920 (5 Mbps VBR); HC-V100EG with three recording (quality) modes (all with full HD resolution  $1920 \times 1080$ ) HA 1920 (17 Mbps VBR), HG 1920 (13 Mbps VBR), and HX 1920 (9 Mbps VBR).
- One camcorder model of Sony: HDR-CX250E with four recording (quality) modes FX ( $1920 \times 1080$ , 24 Mbps VBR), FH ( $1920 \times 1080$ , 17 Mbps VBR), HQ ( $1440 \times 1080$ , 9Mbps VBR), and LP ( $1440 \times 1080$ , 5Mbps VBR).
- One camcorder model of Canon: HFM46, with five recording (quality) modes MXP ( $1920 \times 1080$ , 24 Mbps VBR), FXP ( $1920 \times 1080$ , 17 Mbps VBR), XP ( $1440 \times 1080$ , 12 Mbps VBR), SP ( $1440 \times 1080$ , 7 Mbps VBR), and LP ( $1440 \times 1080$ , 5 Mbps VBR).

The video compression format of all tested camcorders is MPEG-4 AVC/H.264.

## 4.3 Video recording

To study how the PhTD and PeTD change with respect to both motion speed and compression bitrate, we designed our experiments to record a number of videos by fixing all other settings of the camcorders under study, only changing the object motion speed and the compression bitrate. We chose to move the camcorders to create uniform global motion. This was achieved by fixing the camcorder on a tripod and rotating the camcorder by an electronic motor with a speed control. The compression bitrate was changed by switching to different quality modes supported by the camcorders. To make object detection and tracking in our data analysis phase easy and fast, the dead leaves chart was fixed on a white wall in a stable lighting condition.

In our experiments, the camcorders and the motor were both mounted on a tripod in a shooting distance of about 1.5 meters from the printed test chart, and the camcorders were focused on the dead leaves chart. The illumination on the chart was 168K LUX. The videos were made so that the test chart occupies no more than 30% of the whole video field to reduce possible negative influence of geometric non-uniformity of the camcorders. The motor does not have an accurate speed control, but only a continuous voltage adjuster. Therefore, the actual speed was estimated in the video analysis stage from the distance the pattern moved from frame to frame. To have sufficient videos to get the relationship between PhTD and PeTD with respect to motion speed, we recorded 30 to 40 videos for each compression bitrate (recording/quality mode) with each camcorder model. All recording/quality modes supported by the six camcorders were tested.

To minimize the influence of automatic components to the results, we switched off all such functions including auto focus, auto white balancing, auto iris control, auto image stabilizer, wind cut (for reducing the noise caused by wind), auto brightness adjustment, and backlight compensation. The shutter time was also fixed to have a uniform exposure time so that the comparison among different settings is in order.

For each tested camcorder in a stable lighting condition, the video recording process proceeded as follows:

- Step 1: Set the lighting condition stable.
- Step 2: Capture a video of the printed dead leaves chart under each recording mode when the camcorder is still.
- Step 3: Switch on the motor to let the camcorder rotate with a particular speed on the tripod.
- Step 4: Capture a video under each recording mode by starting the recording when the dead leaves chart appears in the field of view of the lens until it starts moving out of the scene. A remote control was used to avoid shake when the recording process was started. For camcorders without remote control, the video started a few seconds earlier to make sure the motion is stable when the chart appears in the field of view.

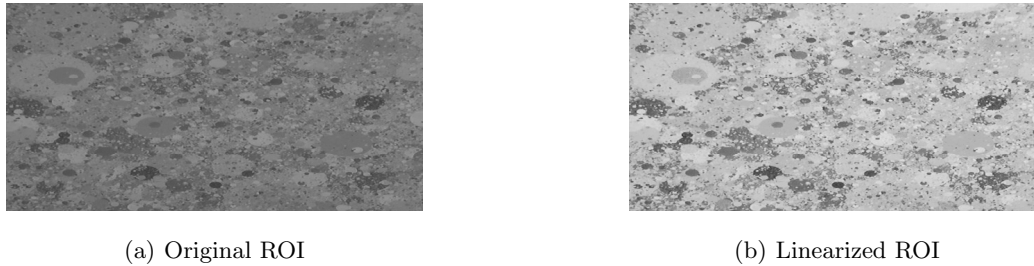


Figure 1. An example of linearization.

- Step 5: Adjust the voltage of the motor to change the rotation speed of the camcorder and repeat step 3 until videos are made for all desired rotation speeds.
- Step 6: Change the recording mode of the camcorder and repeat Steps 4 and 5 until all recording modes are tested.

#### 4.4 Video analysis

The OECF curve of each recording mode in each lighting condition was obtained from the still video, which was collected in Step 2 (Section 4.3). The average grey values of all patches were computed. The measured OECF curve is the plot of these average grey values with respect to the corresponding exposure levels. Figure 1 illustrates the region of interest (ROI) before and after linearization adopting the measured OECF curve.

Since all videos were recorded in interlacing mode, the two fields in each frame were not recorded at the same time. Therefore, if we take the whole frame as a single shot to calculate the PhTD and PeTD, the mismatch between the two fields can cause inaccuracy. As a result, we decided to handle the two fields separately as if they were independently recorded. As the texture in each field of the video sequence is our focus, the RGB images of all fields were transformed to grey tone images before any further analysis. The process of obtaining the PhTD and PeTD of a distorted video can be described by the following steps:

- Step 1: Object extraction

An efficient method of object extraction in our case is simple thresholding. The grey level distribution is clearly bimodal allowing the segmentation of the object (i.e., the dead leaves chart) from the background (i.e., the white wall). The Matlab function *kmeans* was used to determine the threshold from the histogram of the current field.

After extracting and localizing the object in the field, the center region of the object with size of  $300 \times 600$  was taken as the ROI. Those ROIs near the center of the corresponding fields were selected as valid ROIs for further analysis. Figure 1 (a) shows an example of the original ROI. The original ROI was extracted from one field of the video without motion taken by Panasonic SD800 with recording mode HA1920. Note that the vertical resolution was halved.

- Step 2: Linearization

The original ROI was linearized by mapping each pixel value to the exposure level in the range  $[0, 20]$  according to the measured OECF curve. Next, the exposure level was mapped linearly from  $[0, 20]$  to  $[0, 255]$  to obtain the linearized ROI. Figure 1 (b) is an example of the linearized ROI.

- Step 3: SV-CSF filtering

We set  $c_0 = 1.14$ ,  $c_1 = 0.67$ ,  $c_2 = 1.92$  in Eq. (1) which yields a peak sensitivity of the SV-CSF of about 250 at  $\rho = 3.05$  cy/deg,  $v_R = 5.18$  deg/sec. It has been plotted as two-dimensional surface in Figure 2. The SV-CSF function was first normalized to the range of  $[0, 1]$  by dividing the peak sensitivity. The normalized SV-CSF function at a certain velocity of image plane,  $v_I$ , is a one-dimensional CSF curve (Figure 3).

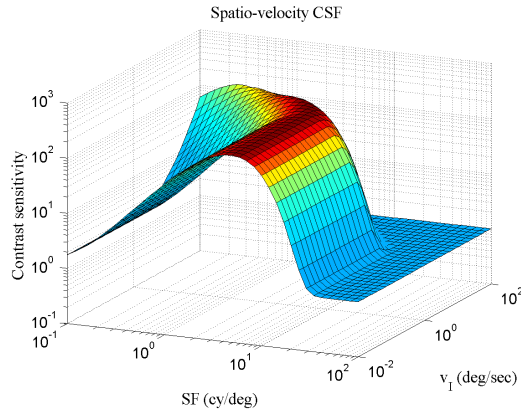


Figure 2. An example of Daly's SV-CSF model

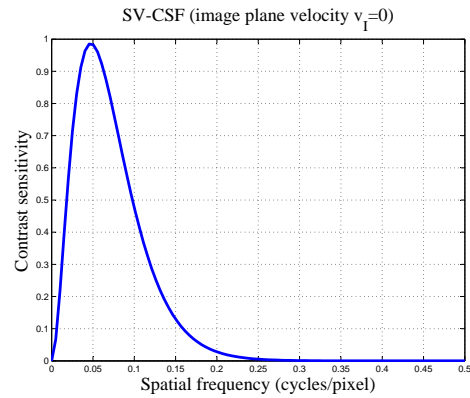


Figure 3. 1D SV-CSF with respect to spatial frequency (cycles/pixel) at image plane velocity 0 pixels/second.

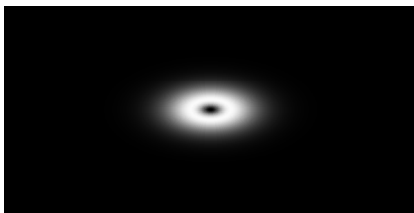


Figure 4. 2D CSF filter in frequency domain

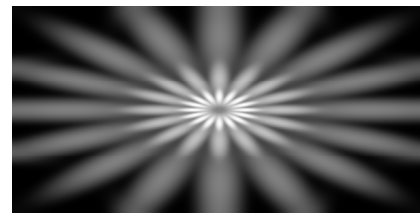


Figure 5. Gabor filters of 8 orientations and 3 scales.

The velocity of the image plane,  $v_I$ , in units of *pixels/sec* was obtained from the location of the ROI in each field. It was then transformed from *pixels/sec* to *deg/sec*. In our experiment, the number of vertical lines in the image is 1920, and the viewing distance is 3 picture heights. Thus, one degree corresponds to 100.5 pixels.

At last, for each given velocity  $v_I$ , a two-dimensional CSF filter with circular symmetry as shown in Figure 4 was generated from the corresponding one-dimensional CSF following Eq. (12) to filter the PSD of the linearized ROI in the frequency domain. The result of this step is called the perceptual PSD of the linearized ROI.

- Step 4: Gabor energy features extraction

The Gabor filter bank of 8 orientations and 3 scales ( $S_m = 1.50, 2.75, 4.00$ ) was generated from Eq. (4) and Eq. (5). The PSDs of all these  $8 \times 3$  filters are shown in one plot in Figure 5. The PSD and perceptual PSD of linearized ROI were both filtered by each of the Gabor filters in frequency domain to obtain the Gabor energy features and the perceptual Gabor energy features, as described in Eq. (8) and Eq. (9), Eq. (14) and Eq. (15), respectively.

When the (physical or perceptual) Gabor features of all valid ROIs, which located near the center of the corresponding fields, in the video sequence were obtained, their arithmetic average was taken as the (physical or perceptual) Gabor features of the whole video sequence, and listed in a  $8 \times 3$  matrix.

- Step 5: PhTD and PeTD computation

The squared Euclidean distance between (physical or perceptual) Gabor features of the distorted videos and the undistorted dead leaves chart was computed by following Eq. (10) and Eq. (16).  $\mathbf{E}_{0,0}$  and  $\mathbf{E}'_{0,0}$  were computed by following Step 3 and 4 but taking the distortion-free dead leaves chart as the linearized ROI and setting the velocity of the image plane to 0.



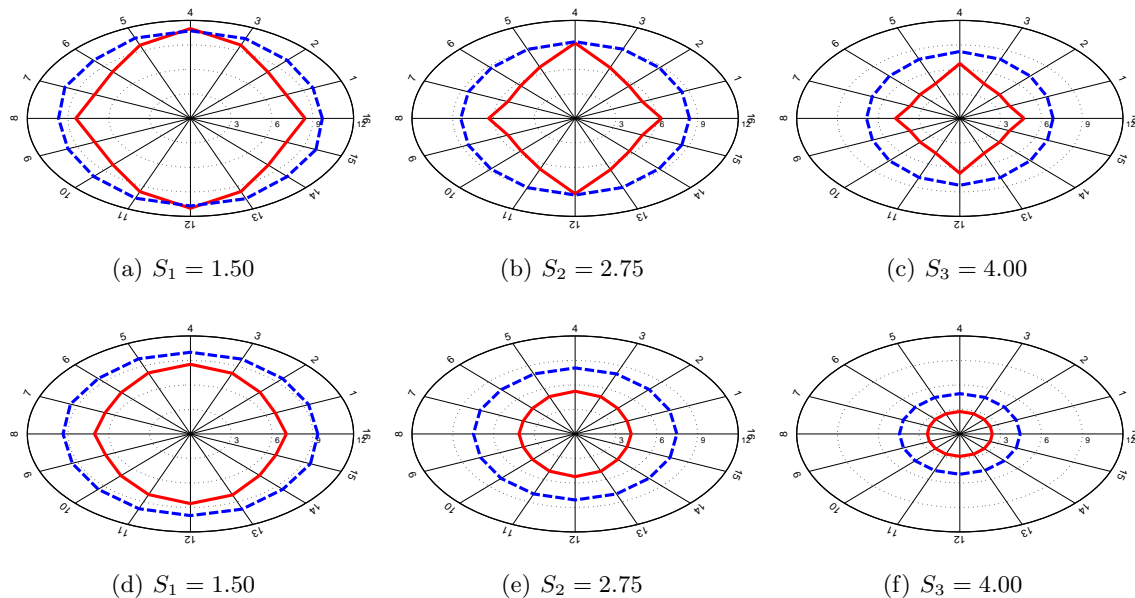


Figure 6. Gabor energy features of the dead leaves chart (dashed blue) and the distorted video (solid red). (a)-(c) illustrate the physical and (d)-(e) illustrate the perceptual Gabor energy features at 3 scales.

## 5. EXPERIMENTAL RESULTS

Figure 6 shows the star plots of Gabor energy features of 8 orientations and 3 scales. Features of the dead leaves chart and the distorted video are plotted in dashed blue lines and solid red lines, respectively. The distorted video was taken by the Panasonic SD800 with recording mode HA 1920 at speed 1459.8 pixels/sec. Each circle is symmetric about the origin and stands for features at all 8 orientations of one scale. The scales  $S_1$ ,  $S_2$ , and  $S_3$  correspond to low, medium and high frequencies, respectively.

The Gabor energy features in Figure 6(a) and (b) show that the texture loss of distorted video appears at horizontal and diagonal directions, and there is almost no vertical texture distortion. This indicates for this example that the optical system and the compression do not degrade texture quality at low and medium frequency; only the horizontal motion of camcorder system causes some loss of texture quality.<sup>†</sup>

In Figure 6(c) corresponding to high frequency, the loss of texture is observed in all orientations, and more horizontal texture is lost than vertical texture. Thus, we attribute the loss of vertical texture at high frequency to the optical system and signal processing (including lossy compression) of the camcorder system; the loss of horizontal texture is the joint contribution of motion, signal processing and optical system. Summarizing, the loss of physical texture quality at low and medium frequencies is caused mainly by motion; the texture distortion introduced by the optical system and compression mainly appears at high frequencies.

For perceptual energy features in Figure 6(d)-(f), the feature distance between the dead leaves chart and the distorted video is observed in all orientations at all frequency bands, while the horizontal distance is even larger than the vertical distance. It indicates that the perceptual distortion appears at all frequencies, and there is even more perceptual distortion in the horizontal direction, due to the horizontal motion of the camcorder system.

Figure 7(a)-(f) comprise all PhTD measurements of all six camcorders. Firstly, we see that the PhTD measurement increases monotonically when the motion speed goes up so there is more texture distortion at

<sup>†</sup>Note that the distortion introduced by optical system is known as radially symmetric. We assume the distortion introduced by signal processing including lossy video compression is also radially symmetric at the moment, though it is still unknown. The horizontal motion is assumed to have no influence on vertical texture. Thus the texture distortion introduced by motion is not radially symmetric. Furthermore, the camcorder system is considered as a linear time-invariant system after linearization. Therefore, the three portions of distortion can be analyzed in separation.

larger speeds. The result confirms our expectation that high motion speed causes more texture distortion.

Secondly, the PhTD measurement decreases monotonically with respect to the compression bitrate for all tested camcorders. For the first four camcorders (Figure 7(a)-(d)), the differences between the fitting curves at all recording modes (corresponding to compression bitrate) are very clear. For the last two camcorders (Figure 7(e)-(f)), the differences between fitting curves are also observed at high speed, though they are overlapping at low speed. Thus, the expectation that compression to low bitrates causes more texture distortion is confirmed as well.

Also, for all six camcorders the differences between fitting curves tend to increase when speed increases to the maximum, i.e., the fitting curves are divergent with respect to speed; the PhTD difference associated with recording modes is small compared to the difference associated with different motion speeds in the range of speeds considered. Therefore, the lossy compression has smaller impact on texture quality than motion speed. The impact tends to increase slightly when the motion speed grows large.

Figure 8 (a)-(f) are the PeTD measurements for all six camcorders. The PeTD measurements in all figures increase monotonically when motion speed increases. We observe the same pattern as for the PhTD measurements with respect to speed and in agreement with our expectations. But surprisingly, the perceptual texture distortion seems to be almost independent of the compression bitrate, as the fitting curves at all recording modes are almost on top of each other at all motion speeds. So object motion plays the main role in the PeTD, and lossy compression has almost no influence on perceptual texture quality.

## 6. CONCLUSION

This paper presents a method to evaluate the performance of camcorders in terms of texture distortion. The texture distortion is measured with respect to motion speed and lossy compression bitrate. The physical texture distortion (PhTD) metric is proposed to quantify the physical texture distortion. It is defined as the squared Euclidean distance between Gabor features of a distortion-free test chart and the distorted target in the video. A bank of Gabor filters of 8 orientations and 3 scales are designed to extract Gabor features, and the dead leaves chart with invariance of scaling, translation, rotation, and contrast (exposure) adjustment is generated and printed as the distortion-free chart. The PhTD metric is further extended to the perceptual texture distortion (PeTD) metric in a way that the distortion-free chart and the distorted target are weighted by the spatio-velocity contrast sensitivity function (SV-CSF) of the human visual system (HVS) in frequency domain before their Gabor energy features are extracted.

Experimental results on six camcorders from three vendors show that (1) the PhTD value decreases monotonically with respect to the motion speed, and increases monotonically with respect to the lossy compression bitrate; (2) the PeTD value decreases monotonically with respect to the motion speed, but almost stays constant with respect to the lossy compression bitrate; (3) the difference between fitting curves tends to increase when speed grows large.

Previous SFR-based methods for digital camera systems were designed to measure radially symmetric texture distortion. Our method generalizes these SFR-based methods to cover asymmetric texture distortion. Furthermore, the texture distortion is measured with respect to motion speed and lossy compression bitrate, such that the performance of camcorders is fully evaluated at various levels of motion and compression. Taking the HVS into account, the proposed PeTD metric aims at measuring the perceptual texture distortion objectively and economically in contrast to the time-consuming and expensive assessment provided by subjective evaluation. The experiments provide reasonable results also for distortions that are not radially symmetric. However, some subjective tests should be done in the future work to confirm and validate the performance of the proposed objective metrics.

The lighting condition also affects the performance of camcorder systems significantly. Thus, in our future work, we plan to measure texture distortion also with respect to varying illumination of the target.

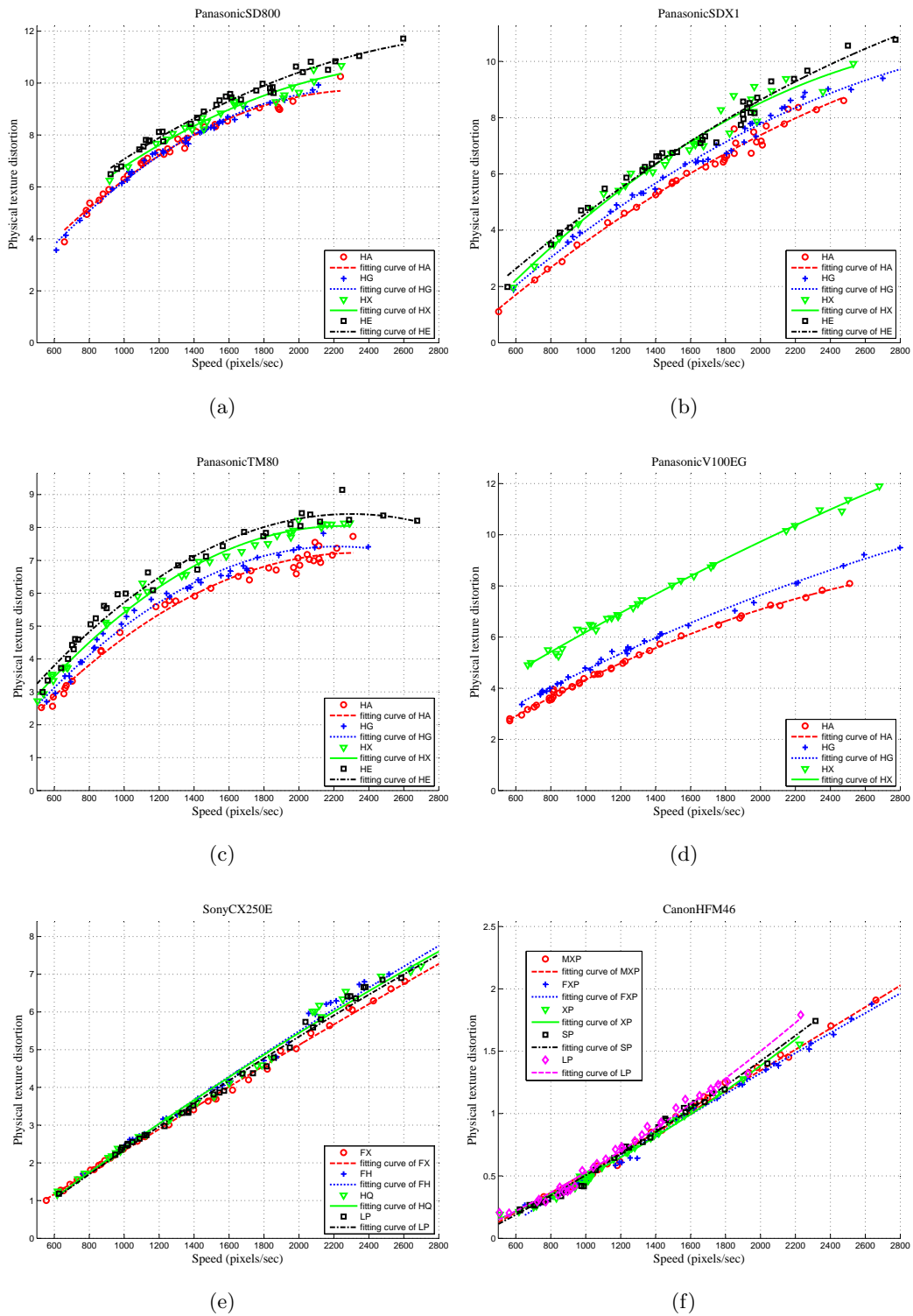
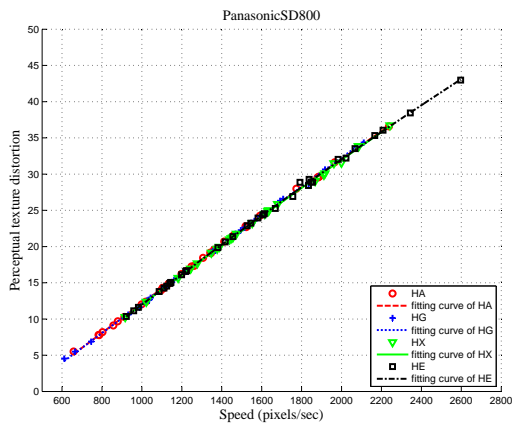
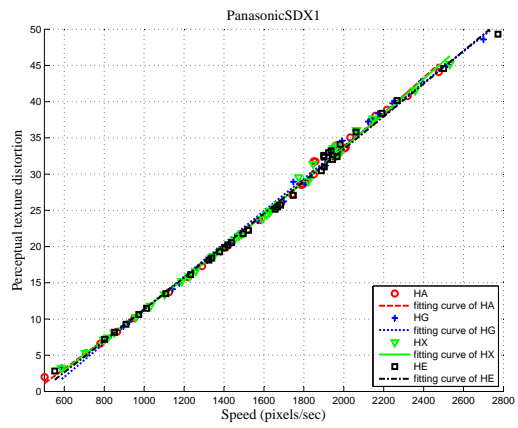


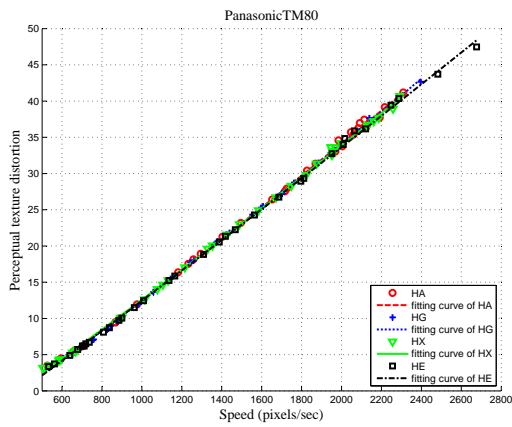
Figure 7. Physical texture distortion (PhTD) measurement.



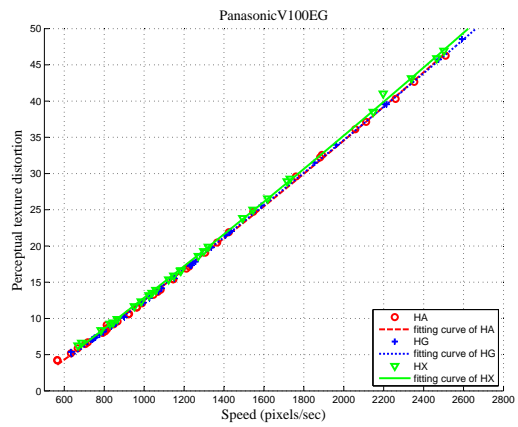
(a)



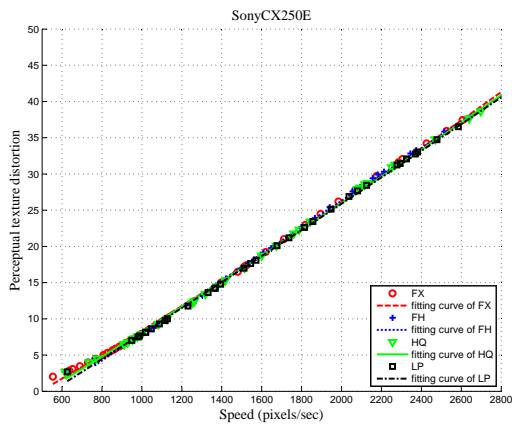
(b)



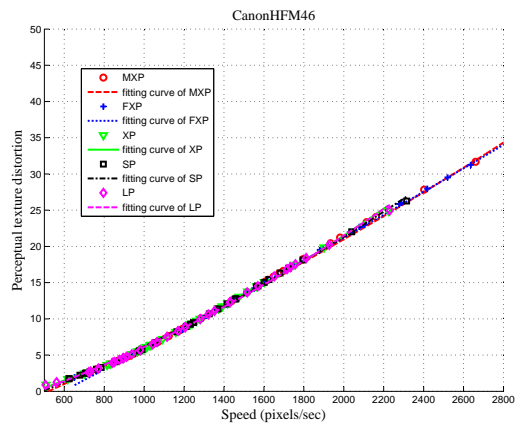
(c)



(d)



(e)



(f)

Figure 8. Perceptual texture distortion (PeTD) measurement.

## ACKNOWLEDGMENTS

The authors thank Shujun Li of the University of Surrey, UK, Martin Biebel from Aktiv Verlag & Medienservice GmbH, Reichenau, Germany, and Malte Neumann from nor-Elektronik, Stuttgart, Germany, for many fruitful discussions. Particularly, we thank Martin Biebel for providing the camcorders and the motor to make the experiments possible. We also thank Keita Hirai of Chiba University, Japan, for his help discussing with us spatial-velocity contrast sensitivity functions.

## REFERENCES

- [1] J. McElvain, S. Campbell, and J. Miller, "Texture-based measurement of spatial frequency response using the dead leaves target: extensions, and application to real camera systems," in *Digital Photography VI, Proceedings of SPIE* **7537**, p. Article Number 75370D, 2010.
- [2] F. Cao, F. Guichard, and H. Hornung, "Measuring texture sharpness of a digital camera," in *Digital Photography V, Proceedings of SPIE* **7250**, p. Article Number 72500H, 2009.
- [3] G. Matheron, *Random Sets and Integral Geometry*, John Wiley & Sons, Inc., Hoboken, New Jersey, USA, 1975.
- [4] F. Cao, F. Guichard, and H. Hornung, "Dead leaves model for measuring texture quality on a digital camera," in *Digital Photography VI, Proceedings of SPIE* **7537**, p. Article Number 75370E, 2010.
- [5] F. Cao, F. Guichard, and H. Hornung, "Information capacity: a measure of potential image quality of a digital camera," in *Digital Photography VI, Proceedings of SPIE* **7537**, p. Article Number 75370E, 2010.
- [6] F. Cao, F. Guichard, and H. Hornung, "Sensor information capacity and spectral sensitivities," in *Digital Photography V, Proceedings of SPIE* **7250**, p. Article Number 725006, 2009.
- [7] U. Artmann and D. Wueller, "Differences of digital camera resolution metrology to describe noise reduction artifacts," in *Image Quality and System Performance VII, Proceedings of SPIE* **7529**, p. Article Number 75290L, 2010.
- [8] G. Matheron, "Modèle séquentiel de partition aléatoire." Technical report, CMM, 1968.
- [9] K. Zhu, S. Li, and D. Saupe, "An objective method of measuring texture preservation for camcorder performance evaluation," in *Image Quality and System Performance IX, Proceedings of SPIE* **8293**, p. Article Number 829304, 2012.
- [10] P. D. Burns, "Tone-transfer (OECF) characteristics and spatial frequency response measurements for digital cameras and scanners," in *Electronic Imaging Symposium, Proceedings of SPIE* **5668**, pp. Article Number 123–128, 2005.
- [11] D. H. Kellt, "Motion and vision. ii. stabilized spatio-temporal threshold surface," *Journal of the Optical Society of America* **69**(10), pp. 1340–1349, 1979.
- [12] S. Daly, "Engineering observations from spatiovelocity and spatiotemporal visual models," in *Human Vision and Electronic Imaging III, Proceedings of SPIE* **3299**, pp. 180–191, 1998.
- [13] T. Grigorescu, N. Petkov, and P. Kruizinga, "Comparison of texture features based on gabor filters," *IEEE Trans. on Image Processing* **11**(10), pp. 1160–1167, 2002.
- [14] B. S. Manjunath and W. Y. Ma, "Texture features for browsing and retrieval of image data," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **18**(8), pp. 837–846, 1996.