A Spectrally Adaptive Noise Filling Tool for Perceptual Transform Coding of Still Images

Christian R. Helmrich, Sebastian Bosse, Paul Keydel, Heiko Schwarz, Detlev Marpe, and Thomas Wiegand

Video Coding and Analytics (VCA) Department
Fraunhofer Heinrich-Hertz-Institut (HHI)
Einsteinufer 37, 10587 Berlin, Germany

***Abstract*—Modern perceptual image coders reach impressively high subjective quality even at low bit-rates but tend to denoise or "detexturize" the coded pictures. Traditionally, two independent parametric approaches, known as texture and film grain synthesis, have been applied in the spatial domain as pre and post-processors around the codec to counteract such effects. In this work, a unified alternative, operating directly within the spectral domain of conventional transform codecs with tight coupling to the transform coefficient quantizer, is proposed. Due to its design, this spectrally adaptive noise filling tool (SANFT) enables highly input adaptive realizations by reusing the coder's existing optimized spatial and spectral partitioning algorithms. Formal subjective evaluation in the context of a "main still picture" High Efficiency Video Coding (HEVC) implementation confirms the benefit of the proposal.**

***Keywords*—Film grain synthesis; still image coding; perceptual coding; texture synthesis; transform coding; video coding; HEVC.**

I. Introduction

PERCEPTUAL transform coding of digital still images has evolved remarkably over the last two decades. A quarter of a century after the completion of the JPEG image specification T.81 [1], formal objective and subjective evaluation of state-of-the-art still image coding approaches was presented at the 2016 International Conference on Image Processing (ICIP) [2]. In the visual assessments performed in the course of this evaluation, a "main still picture" coding profile of the H.265/High Efficiency Video Coding (HEVC) standard [3] was found to outperform, with statistical significance, most of the other coding techniques in visual quality across all tested images and bit-rates, without being exceeded in subjective quality by any other scheme [4].

Although the HEVC *main still picture* profile represents one of the most efficient picture coding designs available, like other methods it tends to introduce visible artifacts like blurriness or loss of textural detail in the reconstructed images. The authors observed that, in particular, picture or camera noise and quasi-random textures are softened by the coarse quantization of the "lossy" transform coder at very low bit-rates. To alleviate this effect, two independent procedures, carried out at the decoder side upon image reconstruction, have been devised recently.

The first method, *texture synthesis*, is intended to regenerate certain textural structures of foreground or background objects in images of natural scenes. This approach is generally applied in a parametric fashion by transmitting, to the decoder, auxiliary spectral and/or spatial information about the textural regions to allow *guided* synthesis [5]–[13]. Note that, on the one hand, a very compact description of the parametric side-information is critical particularly in low-rate coding, in order not to cancel the perceptual benefits of the texture synthesis by notably reducing the bit-budget of the underlying image coder. On the other hand, texture-specific artifacts due to insufficient parameter rates (see [5], [9]) or completely *unguided* synthesis shall be avoided.

The second method, traditionally called *film grain synthesis*, is employed to recreate the film medium or camera sensor noise introduced into a natural image during its acquisition especially in low lighting conditions [14], [15]. The basic principle behind this approach is an additive *image plus noise* model: the image to be coded is split into a noise-like and a denoised component, the latter is compressed by a conventional coding solution, and the former is reconstructed, also in a usually guided parametric fashion, by a dedicated synthesis algorithm. Then, the decoded denoised image is "renoised" by adding the synthesized noise-like component [14]–[20]. The noise parameters comprise data such as higher-order statistics (variance, skewness, or kurtosis) [14], [20] or autoregressive model data [16]–[19] and may be acquired in either the spatial domain [14]–[19] or some spectral domain [20]. Notice that, in the abovementioned work, a single *global* noise model is used for the entire image (or video), thus limiting the ability of adapting to *locally* varying statistics especially in multi-source images such as computer screenshots.

In this paper, a unification of the texture and film grain reconstruction paradigms is presented. The proposal, referred to as a *spectrally adaptive noise filling tool* (SANFT), bears three key advantages compared with the state of the art. First, it avoids an explicit binary classification of the coded image into textural vs. nontextural spatial areas [7], [10] or noisy vs. denoised portions by way of filtering [14], [16]–[19]. This reduces the possibility of limited visual coding quality due to false classifications and yields seamless scaling of the codec to subjective transparency, two aspects which will be investigated in Section II. Second, as shown in Section III, the proposal needs a low algorithmic complexity and provides highly localized control over the synthesis process by means of a precise adaptation to the "instantaneous" spectral and spatial properties of the input picture to be coded.

E-mail: {christian.helmrich, sebastian.bosse, paul.keydel, heiko.schwarz, detlev.marpe, thomas.wiegand}@hhi.fraunhofer.de.

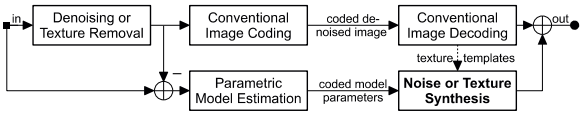

Fig. 1. Traditional parametric image coding with spatial-domain pre-processor and noise/texture synthesis as post-processor around a codec. (--) if applicable.

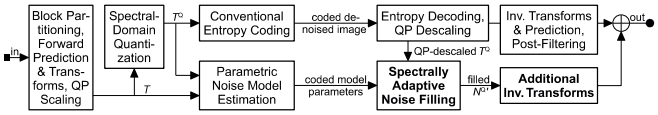

Fig. 2. Proposed modification to the diagram of Fig. 1 with transform-domain parametric analysis and synthesis in the context of a modern transform codec.

Third, straightforward implementations into conventional block transform image coding designs are possible. In fact, Section IV reports on a SANFT integration into HEVC's *main still picture* profile and on the preparation, execution, and outcome of a formal subjective evaluation to assess its visual benefit in a typical use case. Section V, lastly, summarizes and concludes the paper.

II. Noise or Texture Synthesis in Image Coding

The fundamental principle behind both texture and film grain regeneration, as outlined in Section I, is a segmentation into two components, of which one – the coarse structural part – is coded conventionally and the other – the fine textural detail – is parameterized by the transmitter and synthesized by the receiver. A block diagram of this procedure is illustrated in Fig. 1. For both synthesis schemes, segmentation can be done using local classification or denoising, followed by subtraction of the result from the input, yielding a residual to extract the model parameters.

Fig. 1 shows that the extra parametric analysis and synthesis steps act as pre- and post-processors, respectively, to the legacy codec (coder/decoder). This complicates both a low-complexity implementation of the parametric coding extension (as existing codec components cannot be reused) and a seamless transition towards visual transparency at very high rates (where texture or noise synthesis is unnecessary). Specifically, the only means to reduce the amount of parametric synthesis in the decoded image at high bit-rates is to decrease the denoising strength until it is disabled, or by forcing any segmentation as in, e. g., [7], [10] to classify the complete image as nontextural and/or noise-free.

In order to avoid these drawbacks, a structural modification of the approach of Fig. 1, shown in Fig. 2, is proposed. Instead of applying the analysis and synthesis operations *outside* of the legacy codec, their execution is moved *inside* the codec, i. e., around the codec's quantizer. Moreover, with a transform coder like HEVC, where the quantization is applied to DCT or DST spectral coefficients obtained from a spatial-domain prediction residual block [3], the parameter extraction and noise or texture synthesis can be realized directly in the frequency domain. This, in turn, allows for a noisy-vs-nonnoisy respectively textural-vs-nontextural partitioning into fully disjoint coefficient sets which overlap spatially but not spectrally. By varying the spectral size of each set until one set is empty (and all available coefficients are collected in the other set), fine scalability between low-rate fully parametric and high-rate waveform coding is achieved.

Moreover, in combination with a spatial adaptation of the set sizes depending on local image characteristics, the somewhat limiting binary image segmentation in the prior work becomes obsolete. In other words, as shown hereafter, all denoising and/or classification pre-processing can be replaced with the coder's block partitioning as well as the residual quantization step itself.

III. Joint Parametric Partitioning and Quantization

Section I noted the removal of fine quasi-random texture and noise components of an image due to the coarse quantization in low-rate transform coding. Assuming, similar to the film grain synthesis, an additive "structure (edges) plus texture (surfaces)" model in texture coding, a common explanation considers two disjoint quantized transform coefficient sets, S^{Q} and N^{Q}, with

$$S^{\mathrm{Q}} \in T^{\mathrm{Q}},\ N^{\mathrm{Q}} \in T^{\mathrm{Q}},\ S^{\mathrm{Q}} \cap N^{\mathrm{Q}} = \emptyset, \tag{1}$$

obtained for some local rectangular block of the input image, or a prediction residual thereof [21], using a forward DCT or DST. The coefficients T of this transform are quantized to T^{Q} exhibiting lower entropy than T. Specifically, some coefficients in T^{Q} are mapped to index zero even though the respective coefficient magnitudes in T are nonzero. Let this "zero-quantized" subset of T^{Q} be denoted by N^{Q}, the *noise* part. The remaining subset of T^{Q}, containing all nonzero-indexed coefficients, shall be called S^{Q}, the *signal* part. It can be observed that, at low coding rates, $N^{\mathrm{Q}} \approx T^{\mathrm{Q}}$, $S^{\mathrm{Q}} \approx \emptyset$, while at very high rates, $S^{\mathrm{Q}} \approx T^{\mathrm{Q}}$, $N^{\mathrm{Q}} \approx \emptyset$. In addition, the authors observed that N^{Q} primarily comprises random high-frequency coefficients (i. e., S^{Q} has a low-pass shape) and that the variance of the coefficients in T corresponding to those in S^{Q} exceeds the variance of the remaining coefficients in T associated with those in N^{Q}. This lets us conclude that, at sufficient bit-rates, N^{Q} is primarily dominated by noise or texture content, thus explaining the denoising effect of lossy image transform coders [15]. As such, N^{Q} is efficiently representable by parametric models conveying, instead of coefficient indices, only a compactly coded description of statistical properties.

It is worth emphasis that an alternative countermeasure to the denoising effect is the use of a subtractively or nonsubtractively dithered quantizer [22]. In case of coarse quantization, however, this results in more coefficients of T^{Q} being nonzero, so S^{Q} and, thereby, the entropy of T^{Q} and the coding bit-rate increase.

In HEVC and similar codecs, the types and sizes of the local transforms, applied to all image components (luma or chroma), are chosen input-adaptively by the encoder and signaled to the decoder so that appropriate inverse transforms can be applied. The search for optimal transform block segmentation is usually carried out jointly with the search for optimal prediction in a rate-distortion (RD) loop, and for the sake of maximized coding efficiency, this architectural property shall be maintained. Note that, generally, the resulting *spatial partitioning* into transform units (TUs) also serves as a subjectively reasonable partitioning into local N^{Q} parameter units and can simply be reused, hence avoiding the need for additional algorithmic operations. Analogously, a low-complexity *spectral partitioning* of each T^{Q} into subsets S^{Q} and N^{Q} can be realized by just letting the transform

coefficient quantizer "do its job". Notice that the size of subset N^Q – and, thereby, the amount of denoising – can be influenced by traditional encoder-side optimizations such as static or input adaptive quantizer deadzone variation [23], [24]. This renders the codec-external pre-processing methods of Fig. 1 obsolete.

IV. Integration into HEVC and Subjective Evaluation

Since, as shown in Section III, both spatial and spectral partitioning for input-adaptive parametric noise/texture coding can be inherited from the underlying image transform codec, a low-complexity variant of the SANFT proposal was integrated into version 16.7 of the HEVC reference software [25] as follows.

A. Implementation of SANFT into HEVC Reference Software

In HEVC, the largest *spatial partition* size is the coding tree unit (CTU, 64 × 64 pels), which is split recursively into smaller transform units (TUs, minimum size 4 × 4 pels) by a quad-tree based on RD considerations. Exploiting this quad-tree, which is signaled in the bit-stream, a set of parametric model parameters can be explicitly assigned to (and transmitted for) each N^Q associated with a TU after the $D \times D$ pels belonging to that TU have been transformed and the resulting $D \times D$ transform coefficients T have been quantized. To reduce the parameter rate overhead, it was found beneficial to encode only one set of (averaged) N^Q parameters for each 8 × 8 image region split into 4 × 4 TUs.

With respect to the *spectral partitioning*, the overall bit-rate increase due to the SANFT parameters can be partially compensated for by adaptively increasing the quantizer's deadzone in a similar fashion as the optimized quantization approach in EVS audio coding [24]. More precisely, by successively zeroing out higher-frequency coefficients of T^Q initially quantized to index magnitude one, until a coefficient with a greater magnitude has been reached (beginning at the highest frequency and ending, at the latest, at half that frequency), the number of elements in N^Q can be increased. This, in turn, reduces S^Q and, thereby, the bit-rate required for entropy coding of T^Q, without causing visible blur (i. e., loss of structural detail) in the decoded image.

Noise parameter extraction and coding: The authors follow the notion that natural texture, like film grain, often exhibits a noise-like character [8] and make use of the parametric model applied for spectral-domain noise filling in recent MPEG audio coders [26], [27]. Following the QP-dependent scaling (normalization) of T, quantization to T^Q, and definition of $N^Q \in T^Q$, a noise *level* l is derived in each TU t from the coefficients of T at the same frequencies as their respective counterparts in N^Q:

$$l_t = \frac{1}{|N^Q|} \sum_{c \in C} |c|, \; C = \{c \in T : c^Q \in N^Q\}, \tag{2}$$

with c being the scaled coefficients of T and $|N^Q|$ the size of C. Note that, for better stability of the analysis, it was beneficial to exclude the coefficient with the lowest possible frequency (first coefficient in T, i. e., DC offset in DCT-II) if that is included in N^Q, unless the employed transform is an identity one. With only moderate deadzone adaptations assumed, $0 \le l_t \le 0.5$ due to the normalization of T and, thus, l_t can be easily quantized, e. g., via

$$l_t^Q = 0 \text{ if } l_t \le \tfrac{1}{32}, \; l_t^Q = \lfloor 10 + 2 \log_2 l_t \rfloor \text{ otherwise.} \tag{3}$$

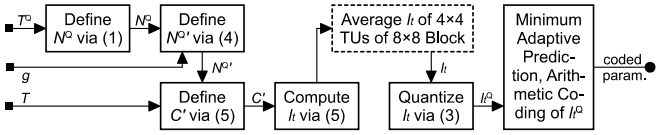

Fig. 3. Extraction of SANFT level parameter l in the HEVC transform domain. Note: the lowest-frequency coefficient is excluded from analysis. (- -) optional.

In addition, a noise shape or *granularity* parameter g is defined for the given picture. This $g \le 1$ represents the average spectral tilt (or low-pass characteristic) of all coefficients in T associated with a $c \in C$. For this work, g is chosen based on the dimensions of the input image: for high-resolution (e. g., 3840 × 2160) content, where textures generally appear larger and coarser, $g = \sqrt{0.5}$ while for smaller and finer input (e. g., 1920 × 1080), $g = 1$. For $g < 1$, it makes sense to define a lowpass subset of N^Q as follows:

$$N^{Q\prime} = \{c^Q[x, y] \in N^Q : x < \lfloor w\, g \rfloor, \; y < \lfloor h\, g \rfloor\}, \tag{4}$$

where w and h are the width and height, respectively, of the TU t in luma pel units and x and y are the horizontal and vertical coordinates, respectively, of c^Q. Using $N^{Q\prime}$ of (4), which defines the frequency region of T^Q to be noise filled at the decoder side, more accurate l_t can be derived by replacing N^Q with $N^{Q\prime}$ in (2):

$$l_t = \frac{1}{|N^{Q\prime}|} \sum_{c \in C'} |c|, \; C' = \{c \in T : c^Q \in N^{Q\prime}\}. \tag{5}$$

Note that g does not need to be transmitted in the bit-stream. The spatiospectrally localized SANFT parameterization via l^Q, hence, suffices to convey to the decoder the essential statistical properties of the film grain or textures to be reconstructed. The l^Q are coded arithmetically and differentially using Martucci's minimum adaptive prediction from spatially neighboring decoded l^Q values [28]. Fig. 3 summarizes the encoding process.

Noise parameter decoding and parametric synthesis: At the receiver, the transmitted T^Q, l^Q, and QP indices are subjected to entropy decoding and, if applicable, inverse spatial prediction. l_t is then reconstructed to $0 \le l_t' \le 0.5$ using the decoded index:

$$l_t' = 0 \text{ if } l_t^Q = 0, \; l_t' = 2^{(l_t Q - 10)/2} \text{ otherwise.} \tag{6}$$

The actual SANFT application, i. e., *noise synthesis*, in each TU is performed separately from the inverse transform of the associated T^Q (whose operation is left unchanged) in four steps.

First, the spectral region of T^Q to be noise filled is determined by (4), with g defined identically to that in the encoder based on the image size. As at the encoder side, for $g < 1$, $N^{Q\prime}$ represents a low-frequency subset of N^Q, and the lowest-frequency "DC" coefficient in T^Q is excluded from $N^{Q\prime}$ if it was quantized to zero and the employed forward transform was a DCT or DST. The $N^{Q\prime}$ coefficients are now substituted by pseudo-random values, e. g., randomly signed unity values as in [27] or signed outputs of a unity-variance linear congruential pseudo-random number generator [29], which are then scaled (i. e., multiplied) by l_t'. In this work, the signs of 32-bit pseudo-random integers generated according to [29] were used as randomly signed unity values.

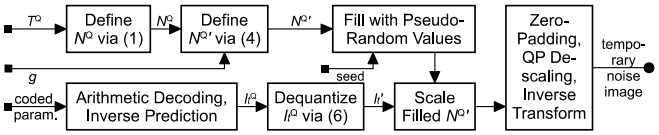

Fig. 4. SANFT transform-domain noise synthesis using transmitted T^Q and l', followed by scaled inverse transforms to obtain the spatial-domain noise image.

As the second step, $N^{Q\prime}$ is padded with zeros at the locations of the remaining coefficients in T^Q not included in $N^{Q\prime}$ and is subjected to the same inverse transform as that applied to T^Q, utilizing the same QP value and fixed-point representation for scaling. This ensures that the synthesized waveform coded and parametric parts are spectrally disjoint and compatibly scaled, as initially desired. The inverse transforms are applied for each t and CTU covering the given picture and, thus, result in a temporary noise image that is, as the third step, added to the legacy decoded denoised image for the final "renoised" reconstruction.

Generating the temporary noise image instead of applying a single joint inverse transform to the TU-wise mix of T^Q and $N^{Q\prime}$ bears two advantages. On the one hand, it allows for HEVC's spatial predictive coding to be undone using only the waveform coded, but not the noise filled, parametric parts. Given that the random noise components would lower the prediction gain, this architectural design guarantees that the core transform coding efficiency is maintained. On the other hand, a *delayed renoising* after HEVC's deblocking [30] and SAO deringing [31] can be realized, thus ensuring uncompromised post-filtering efficiency and preventing unwanted image denoising after the renoising.

Having synthesized the noise component, the fourth and final step in the decoder-side SANFT algorithm is the addition of the temporary noise picture to the traditionally decoded, denoised, and post-filtered image to obtain the output picture. Note that a coded image is typically split into luma (Y) and chroma (U, V) components. To limit both the algorithmic complexity and side information overhead without reducing the subjective merit too much, a SANFT realization operating only on the luma channel was used for the subjective evaluation, which is discussed in the next subsection. Fig. 4 visualizes the SANFT decoding process.

B. Assessment of Objective Aspects and Visual Performance

The SANFT integration into *main still picture* HEVC coding was evaluated both objectively and subjectively. To this end, the first frames of 11 full high definition (FHD) and ultra-high definition (UHD) 8-bit YUV video sequences used during the HEVC standardization [21], [30]–[33] (classes A and B) were coded with and without the SANFT coding and processing. The global (average) QP index was set to 35, for moderate quality, and 30, for higher quality, RD optimized quantization was used in each CTU, and all other coding parameters were either set to the default values of the reference software HM 16.7 [32] or, in case of activated SANFT, configured as discussed above.

For *objective evaluation*, the average bit-stream increase or, in other words, side-information rate consumed by the encoded l^Q indices was measured. This overhead was found to be nearly constant across the two QP operating points and, thus, causes a greater relative bit-stream growth at low rates (4–5 % at QP 35) than at higher rates (2.5–3 % at QP 30). The decoder complexity increases by roughly 25 % due to the noise synthesis, primarily because of the extra TU-wise inverse transforms required.

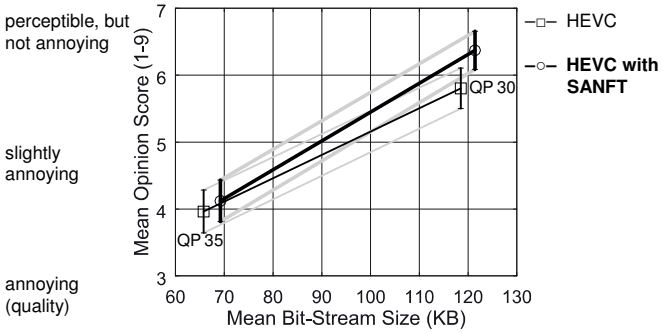

Fig. 5. Zoomed result (with 95 % CIs) of the subjective DCR test for two rates.

The *subjective evaluation* was carried out using formal visual comparative tests following the ITU BT.500 methodology [34]. Specifically, a degradation category rating (DCR) experiment with sequence-wise simultaneous 10-second presentation of the uncoded input, on the left, and the coded version, on the right, assessed on a nine-grade numerical degradation scale according to Annex C and Appendix V, respectively, of ITU-T P.910 [35] was conducted. Each input/coded image pair was, if necessary, cropped to FHD resolution and depicted side-by-side on top of a gray background on a color-calibrated 65-inch LG E6 OLED television. An uncoded/uncoded *hidden reference* pair was also shown for each sequence at random points during the tests for post-screening purposes. Fourteen subjects (incl. two females) with normal color perception (as verified using an Ishihara test), aged between 24 and 35, participated. The voting period after each image pair presentation was unlimited to allow the viewers to take voluntary short breaks. The viewing distance for all participants was restricted to the range $3H$–$3.5H$, with $H = 0.4$ m for this study, i. e., half the height of the E6's UHD panel.

The mean opinion score (MOS) and associated 95 % confidence interval (CI, derived assuming a Student's t distribution) for each codec configuration and QP operating point is depicted in Fig. 5. Note that all illustrated values were derived using data from only 12 of the 14 viewers since the scores of two subjects had to be post-screened (one due to an excessively low MOS of 5.36 for the hidden reference and one because of an excessively high MOS of 7.54 when averaged across all stimuli). For these 12 viewers, the hidden reference attains a MOS of 8.22 and a CI of ±0.18, indicating sufficiently stable and accurate grading (MOS 8 is regarded the threshold of artifact perception [35]).

When linearly connecting corresponding codec conditions at the two QP points, as in Fig. 5, and interpolating the MOS and associated CI values along these lines, it is found that, around the QP 30 rate, the SANFT approach increases the visual coding quality by a small but statistically significant ($p < 0.05$) margin. At this QP, the two sequences scoring worst with legacy HEVC, *HomelessSleeping* (MOS 5) and *ParkScene* (MOS 4.5), benefit most from the noise filling (their MOS values improve by 1 and 1.9, respectively). The SANFT, thus, allows for more balanced and input independent reconstruction quality. A demonstration of these two sequences for personal viewing is provided in Figs.

6 and 7 and, using a slightly improved bugfixed version, at [36]. At very low rates near (or below) the QP 35 point, the SANFT synthesis does not lead to an overall subjective quality improvement because of the added bit-stream overhead. Moreover, due to the coarse quantization there, N^{Q} tends to comprise undesired structural image components (e. g., edges) alongside the noise and texture components, so the assumption of Sec. III does not hold anymore and the parametric model becomes inaccurate.

V. Conclusion

This paper examined the issue of reduced film grain or texture detail in images after "lossy" compression using modern transform coders like HEVC. As an alternative to, and unification of, traditional parametric film grain and texture synthesis methods, a *spectrally adaptive noise filling tool*, operating closely around the underlying coder's transform domain quantizer, was presented and found to significantly improve the visual coding quality on some input images with only a moderate increase in algorithmic decoding complexity. A demonstration is available at [36], and further subjective evaluation using better-known still image test-sets, such as Kodak's Color Image Suite [37], is planned.

References

[1] Int. Telecomm. Union (ITU), CCITT, Specification T.81, "Information Technology – Digital Compression and Coding of Continuous-Tone Still Images – Requirements and Guidelines," Geneva, Switzerland, Sep. 1992.

[2] T. Ebrahimi *et al*., "Overview and Benchmarking Summary for the ICIP 2016 Compression Challenge," in *Proc. IEEE Int. Conf. Image Process.*, Phoenix, USA, Sep. 2016.

[3] ITU-T, Recommendation H.265 and ISO/IEC, Int. Standard 23008-2, "High efficiency video coding," Geneva, Switzerland, Jan. 2017, online: http://www.itu.int/rec/T-REC-H.265

[4] T. Ebrahimi et al., "Summary and Next Steps for the Image Compression Challenge," in *Proc. IEEE Int. Conf. Image Process.*, Phoenix, USA, Sep. 2016.

[5] P. Campisi, D. Hatzinakos, and A. Neri, "A Perceptually Lossless, Model-Based, Texture Compression," *IEEE Trans. Image Process.*, vol. 9, no. 8, pp 1325–1336, Aug. 2000.

[6] P. Ndjiki-Nya, B. Makai, G. Blättermann, A. Smolic, H. Schwarz, and T. Wiegand, "Improved H.264/AVC Coding Using Texture Analysis and Synthesis," in *Proc. IEEE Int. Conf. Image Process.*, vol. 3, pp. 849–852, Sep. 2003.

[7] A. Dumitraş and B. G. Haskell, "An Encoder-Decoder Texture Replacement Method With Application to Content-Based Movie Coding," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 14, no. 6, pp. 825–840, June 2004.

[8] J. Ballé, B. Jurczyk, and A. Stojanovic, "Component-Based Image Coding Using Non-local Means Filtering and an Autoregressive Texture Model," in *Proc. IEEE Int. Conf. Image Process.*, Cairo, Egypt, pp. 1937–1940, Nov. 2009.

[9] B. T. Oh, Y. Su, C. A. Segall, and C.-C. J. Kuo, "Synthesis-Based Texture Video Coding with Side Information," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 21, no. 5, pp. 647–659, May 2011.

[10] J. Ballé, A. Stojanovic, and J.-R. Ohm, "Models for Static and Dynamic Texture Synthesis in Image and Video Compression," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 7, pp. 1353–1365, Nov. 2011.

[11] B. Zhou, F. Zhang, and L. Peng, "Compact Representation for Dynamic Texture Video Coding Using Tensor Method," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 23, no. 2, pp. 280–288, Feb. 2013.

[12] H. Xu, H. Zha, and M. A. Davenpot, "Manifold Based Dynamic Texture Synthesis from Extremely Few Samples," in *Proc. IEEE Conf. Computer Vision, Pattern Recognition*, Columbus, USA, pp. 3019–3026, June 2014.

[13] K. Naser, V. Ricordel, and P. Le Callet, "Local Texture Synthesis: A Static Texture Coding Algorithm Fully Compatible with HEVC," in *Proc. IEEE Int. Conf. Syst., Signals, Image Process.*, London, UK, pp. 37–40, 2015.

[14] J. C. K. Yan, P. Campisi, and D. Hatzinakos, "Film Grain Noise Removal and Generation for Color Images," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Seattle, USA, vol. 5, pp. 2957–2960, May 1998.

[15] O. K. Al-Shaykh and R. M. Mersereau, "Lossy Compression of Noisy Images," *IEEE Trans. Image Process.*, vol. 7, no. 12, pp. 1641–1652, Dec. 1998.

[16] C. Gomila and A. Kobilansky, "SEI message for film grain noise," *Proc. 8th Meeting Joint Video Team*, document JVT-H022, Geneva, Switzerland, pp. 1–14, May 2003.

[17] B. T. Oh, C.-C. J. Kuo, S. Sun, and S.-M. Lei, "Film Grain Noise Modeling in Advanced Video Coding," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 19, no. 12, pp. 1717–1729, Dec. 2009.

[18] J. Dai, O. C. Au, C. Pang, W. Yang, and F. Zou, "Film Grain Noise Removal and Synthesis in Video Coding," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Dallas, USA, pp. 890–893, Mar. 2010.

[19] I. Hwang, J. Jeong, J. Choi, and Y. Choe, "Enhanced Film Grain Noise Removal for High Fidelity Video Coding," in *Proc. IEEE Int. Conf. Inf. Science Cloud Comput. Comp.*, Guangzhou, China, pp. 668–674, 2013.

[20] T. Sun, L. Liang, K. H. Chiu, P. Wan, and O. C. Au, "DCT Coefficients Generation Model for Film Grain Noise and Its Application in Super-Resolution," in *Proc. IEEE Int. Conf. Image Process.*, Paris, France, pp. 3857–3861, Oct. 2014.

[21] J. Lainema, F. Bossen, W.-J. Han, J. Min, and K. Ugur, "Intra Coding of the HEVC Standard," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 22, no. 12, pp. 1792–1801, Dec. 2012.

[22] S. P. Lipshitz, R. A. Wannamaker, and J. Vanderkooy, "Quantization and Dither: A Theoretical Survey," *J. Audio Engineering Society*, vol. 40, no. 5, pp. 355–375, May 1992.

[23] G. J. Sullivan, "Efficient Scalar Quantization of Exponential and Laplacian Random Variables," *IEEE Trans. Inf. Theory*, vol. 42, no. 5, pp. 1365–1374, Sep. 1996.

[24] G. Fuchs, C. R. Helmrich, G. Marković, M. Neusinger, E. Ravelli, and T. Moriya, "Low Delay LPC and MDCT-Based Audio Coding in the EVS Codec," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Brisbane, Australia, pp. 5723–5727, Apr. 2015.

[25] C. Rosewarne *et al.*, "High Efficiency Video Coding (HEVC) Test Model 16 (HM16) Improved Encoder Description Update 3," document JCTVC-U1002 and ISO/IEC N15436, Warsaw, Poland, June 2015.

[26] ISO/IEC, Int. Standard 23003-3, "Information technology – MPEG audio technologies – Part 3: Unified speech and audio coding," Geneva, Switzerland, Jan. 2012.

[27] C. Helmrich, A. Niedermeier, S. Bayer, and B. Edler, "Low-Complexity Semi-Parametric Joint-Stereo Audio Transform Coding," in *Proc. EURASIP Europ. Signal Process. Conf.*, Nice, France, pp. 794–798, Sep. 2015.

[28] S. A. Martucci, "Reversible Compression of HDTV Images using Median Adaptive Prediction and Arithmetic Coding," in *Proc. IEEE Int. Symp. Circuits Syst.*, New Orleans, USA, pp. 1310–1313, May 1990.

[29] D. E. Knuth, *The Art of Computer Programming, Vol. 2: Seminumerical Algorithms*, 3rd edition, Addison-Wesley, 1997.

[30] A. Norkin, G. Bjøntegaard, A. Fuldseth, M. Narroschke, M. Ikeda, K. Andersson, M. Zhou, and G. v. d. Auwera, "HEVC Deblocking Filter," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 22, no. 12, pp. 1746–1754, Dec. 2012.

[31] C.-M. Fu, E. Alshina, A. Alshin, Y. Huang, C. Chen, C. Tsai, C. Hsu, S. Lei, J. Park, and W.-J. Han, "Sample Adaptive Offset in the HEVC Standard," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 22, no. 12, pp. 1755–1764, Dec. 2012.

[32] F. Bossen, "Common test conditions and software reference configurations," document JCTVC-L1100 and ISO/IEC m28412, Geneva, Switzerland, Apr. 2013.

[33] V. Sze, M. Budagavi, and G. J. Sullivan, *High Efficiency Video Coding (HEVC): Algorithms and Architectures*, 1st edition, Springer Intl., 2014.

[34] ITU-R, Recommendation BT.500, "Methodology for the subjective assessment of the quality of television pictures," Geneva, Switzerland, Jan. 2012.

[35] ITU-T, Recommendation P.910, "Subjective video quality assessment methods for multimedia applications," Geneva, Switzerland, Apr. 2008.

[36] C. R. Helmrich, "ecodis SANFT Demonstration Page," May 2018, online: http://www.ecodis.de/image.htm

[37] R. Franzen, "Kodak Lossless True Color Image Suite," Jan. 2013, online: http://r0k.us/graphics/kodak/

(a) uncoded original 8-bit image (UHD, losslessly coded size: 7842 KB)

(b) "main still picture" 8-bit HEVC (HM 16.7, UHD, coded size: 27.1 KB)

(c) "main still picture" 8-bit HEVC + SANFT (UHD, coded size: 28.4 KB)

Fig. 6. SANFT applied to HEVC coded *HomelessSleeping* still image, QP 30. The image dimensions are 2560×1600, here cropped to 768×512 for visibility.

(a) uncoded original 8-bit image (FHD, losslessly coded size: 4796 KB)

(b) "main still picture" 8-bit HEVC (HM 16.7, FHD, coded size: 81.9 KB)

(c) "main still picture" 8-bit HEVC + SANFT (FHD, coded size: 84.7 KB)

Fig. 7. SANFT demonstration on HEVC coded *ParkScene* still image, QP 30. The image dimensions are 1920×1080, here cropped to 768×512 for visibility.