Block-based detection systems for visual artifact location

Citation for published version (APA):

Eerenberg, O., Kettenis, J., & With, de, P. H. N. (2013). Block-based detection systems for visual artifact location. *IEEE Transactions on Consumer Electronics*, 59(2), 376-384.
<https://doi.org/10.1109/TCE.2013.6531120>

DOI:

[10.1109/TCE.2013.6531120](https://doi.org/10.1109/TCE.2013.6531120)

Document status and date:

Published: 01/01/2013

Document Version:

Accepted manuscript including changes made at the peer-review stage

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Block-Based Detection Systems for Visual Artifact Location

Onno Eerenberg, Jeroen Kettenis and Peter H. N. de With

Abstract — *The core of many video coding standards is formed by the Discrete Cosine Transform (DCT) for decorrelating spatial video data. When quantizing DCT coefficients, artifacts may appear such as mosquito noise and ringing. Spatial artifact reduction requires artifact-location information, to control the filter process, thereby avoiding unnecessary blurring of non-visual artifact-contaminated regions. This location information can be derived either in the spatial or frequency domain. In this paper, we will present two systems: one for each domain. As coding artifacts are most annoying in flat or low-frequency regions, the objective of the detector is to localize these artifact-sensitive locations. The detection accuracy, coverage and sensitivity differ between the two different detection systems. Experiments have revealed that the spatial-domain system has a better performance on edge tracking and detection of small-sized flat region. However, the system degrades in separating originally present low-amplitude texture from artifact contamination, compared to the same capability of the frequency-domain system¹.*

Index Terms — Compression, DCT, Graphics, Locator information, Mosquito noise, Ringing.

I. INTRODUCTION

Image and video communication have largely benefited from the established standards (e.g. JPEG/MPEG) in compression techniques, achieved in the last decades. Many of the popular compression techniques deploy a 2D-DCT to decorrelate a block-based spatial region prior to quantization [4][5][6]. For cost-effective communication, not only irrelevant information, but also visually relevant information, typically in the form of high-frequency detail is lost due to quantization. The removal of such high-frequency information results in a variety of artifacts, such as lack of sharpness (blurring), and other visually noticeable artifacts, known as coding artifacts, which differ from analog television artifacts. Examples of typical coding artifacts are blockiness, ringing and mosquito noise [10][11]. State-of-the-art Liquid-Crystal or Plasma-based flat-panel displays show a nearly perfect and sharp picture, clearly revealing any video imperfections, as the

earlier mentioned coding artifacts [7] [8]. These modern digital televisions deploy Temporal Noise Reduction (TNR) or Motion Compensated Temporal Noise Reduction (MCTNR). The latter noise reduction removes dominant Gaussian noise and also influences the visibility of moving coding artifacts [1] [9] [13]. However, for the situation that a coding artifact is static, spatial filtering has to be applied to attenuate the disturbance [2][14]. Unfortunately, straightforward fixed spatial low-pass filtering attenuates also high-frequency information, thereby severely jeopardizing the overall picture quality. Therefore, we aim at enhancing the picture quality after decoding, with adaptive image processing, where the adaptivity concentrates on the filtering location and the applied filter characteristics.

Transform coding involving quantization, introduces artifacts, which are clearly noticeable around the transition between texture/edges and within flat/low-frequency regions [3], see Fig. 1(b-c). This observation forms a key feature for identifying locations where mosquito/ringing noise tends to be well visible.

Removal of static coding artifacts requires a two-step approach, consisting of an identification stage, followed by a reduction stage [10]. Coding artifacts differ in appearance, so is their detection. Detection of mosquito-noise contamination can be achieved by a variance calculation over a spatial region. For the situation that the variance meets a threshold constraint, that local spatial region is low-pass filtered [16]. Drawback of such a variance calculation is the risk of potentially indicating original texture as mosquito noise, which has to be compensated by limiting the applied low-pass filter strength, resulting in a sub-optimal design. In order to improve this method, information is required that reveals whether a pixel is contaminated with coding artifacts or part of a texture region. To improve the detector, neighboring edge information is added to the mosquito noise detector [17], thereby increasing the appearance probability of coding artifacts. Although this improves the discrimination between original texture and mosquito noise up to a certain extend, there is still a missing aspect.

Both variance-based and the edge-based enhanced coding artifact detection, do not consider the artifact visibility, which is essential to control the low-pass filtering strength. The detection of such artifact-contaminated locations requires an *activity* measurement, revealing the presence of a transition within a bounded region, followed by a spatial reasoning step, which results in a binary decision: *contaminated* versus *not-contaminated*. On the basis of this contamination information, a signal is constructed controlling the final low-pass filter strength.

¹ Onno Eerenberg is with Trident Microsystems, Laan van Diepenvoorde 23, 5582LA Waalre, The Netherlands. (e-mail: onno.eerenberg@tridentmicro.com).

Jeroen Kettenis is with Trident Microsystems, Laan van Diepenvoorde 23, 5582LA Waalre, The Netherlands. (e-mail: Jeroen.kettenis@tridentmicro.com).

Peter H. N. de With is with Eindhoven University of Technology, Den Dolech 2, 5600 MB, The Netherlands (e-mail: p.h.n.de.with@tue.nl).

In this paper we design and compare two block-based visual-artifact location-detection systems. The detection metric for finding visual artifacts is based on an *activity* measurement and is either computed in the spatial- or frequency-domain, deploying a detection kernel, which is implemented by a block-based window filter. This *activity* measurement, which reveals the amount of texture, turns out to be more effective in the frequency-domain compared to the spatial-domain detection system. As a consequence, the frequency-domain artifact detection system provides a better discrimination between the various textures and results in a binary decision for a group of pixels (filter-aperture) rather than a pixel-based decision. The aforementioned decision making is based on spatial reasoning, which is performed over a cluster of small-sized regions forming the detection aperture. In this paper we will show that this approach is effectively suppressing visible artifacts at clearly visible locations in the image, leading to a significant image quality improvement.

The paper is organized as follows. Section II elaborates on mosquito and ringing artifacts. Section III presents the system aspects for a visual-artifact location detector. Section IV discusses the spatial-domain based visual-artifact detector, while Section V presents the frequency-domain based visual-artifact detector. Section VI elaborates on the low-pass filter control signal. Section VII presents experimental results. Finally, conclusions are discussed in Section VIII.

II. MOSQUITO AND RINGING ARTIFACTS

Visual mosquito noise appears in flat or low-frequency regions near the object boundaries, when the block-based transform region contains both parts of the flat background region as well as part of the object [1] [3] [11][12]. For the situation that the transform block size is 8×8 pixels, the mosquito noise can occur anywhere within the 8×8 block, resulting in a distance of up to 7 pixels in horizontal, vertical and diagonal direction relative to the location of the object,

Fig. 1. Mosquito noise and ringing due to 8×8 transform coding $Q=40$. grid size is 8×8 pixels. (a) Original video fragment. (b) Decoded video. (c) Decoding error, error-gain factor 8. (d) Fragment from up-scaled SD to full-HD video (1920×1080), with the same grid size. (e) Decoding error for scaled SD to full-HD video fragment, error-gain factor 8.

Fig. 2. Factor-2 zoomed sharpened examples of mosquito noise and ringing due to MPEG 8×8 transform coding. (a) Original fragment. (b) Mosquito noise fragment. (c) Original fragment. (d) Ringing fragment.

see Fig. 1(b)(c). This distance is increased for the situation of an up-scaled video signal e.g. SD to full-HD, see Fig. 1(d)(e). Mosquito noise is also introduced due to motion compensation mismatch [11], where the spatial prediction residual data may contain pixels that have high difference values, resulting in a high number of DCT coefficients, which will be quantized to meet the compression constraints, resulting in reconstruction errors, finally visible as mosquito noise.

Ringing, "ghosts" near edges, only occurs when a line segment of an object crosses the transform block in either horizontal or vertical direction, see Fig. 1 (b), whereby the ringing location depends on the location of the horizontal or vertical object in the transform block. The ringing effect is introduced by the quantization of the DCT coefficients during compression, and is known as the Gibb's phenomenon [11]. Figure 2 indicates the appearance of mosquito noise, e.g. around the trees, see Fig. 2 (b), and ringing artifacts, e.g. at the vertical edge of the building, Fig 2. (d), introduced due to an 8×8 DCT transform with a constant quantization factor ($Q=20$).

III. SYSTEM ASPECTS FOR BLOCK-BASED VISUAL CODING ARTIFACT DETECTION

With the full block-based processing in video compression and the involved coding artifact generation, it is evident that the detection of associated artifacts should be block-based as well. This section presents the system aspects for block-based visual coding artifact detection.

- **Video Coding Standard.** The received video signal can be coded with MPEG-2 or a more advanced standard such as H.264/AVC. A major difference between these two standards is that H.264/AVC supports next to the 8×8 transform block size also the 4×4 transform block size. Detection of such a small-sized contaminated region may be challenging.
- **Video Resolution.** The received video signal can be either native resolution (SD or HD) or up-scaled (from SD to HD). A major difference between the two is that

the coding artifacts not only change in character, see Fig. 1(c) and 1(e), but are also located at a distance further away from the texture causing artifacts.

- **Influence of Deployed Block Size.** Whereas the frequency-domain activity detection-kernel consists of fixed-sized 4×4 blocks, the spatial-domain activity detection-kernel allows a construction with fixed-sized or unequally-sized blocks.
- **Separability of Detection and Spatial Reasoning.** Calculation of the *activity* measurement in the spatial- or frequency-domain and the associated spatial-reasoning phase can be separated. This enables the trade-off between the costs of embedded line-memories versus the costs of additional bandwidth in a unified-memory architecture-based SoC.
- **Low-Pass Filter Control.** The result of the artifact-location detection system is a control signal, which controls the final low-pass filtering strength. Abrupt changes of the low-pass filtering strength may introduce noticeable filtering artifacts. Such situation is avoided when applying a fade-in prior to and fade-out after the detected location, realizing a Gaussian-type filter kernel.

IV. DETECTION SYSTEM IN SPATIAL DOMAIN

Location detection of regions contaminated with mosquito noise or ringing-artifacts, involves an *activity* measurement, which reveals the presence of a transition in combination with the presence of a flat area or a low-frequency area for that region. In order to avoid the detection of individual edges, the *activity* metric in the spatial domain is based on a simple 2D high-pass filter, implemented as a 2D SAD according to

$$SAD = \sum_{y=j}^M \sum_{x=i}^N |P(x, y) - P(x+1, y)| + \sum_{y=j}^M \sum_{x=i}^N |P(x, y) - P(x, y+1)|. \quad (1)$$

Hereby, $P(x, y)$ denotes the spatial pixel location, while M and N denote the width and height of the sub-blocks constructing the detection kernel, see Fig. 3(a). For each pixel in the image,

Fig. 3. Spatial-domain detection system for visual artifacts. (a) Construction of detection kernel deploying overlapped blocks. (b) Positioning of thresholds in SAD range. (c) Basic block diagram of the artifact detection system.

Fig. 4. Visualization of artifact location detection in the spatial domain. (a) Original image. (b) 2D SAD for 5x5 block size. (c) Artifact location without false positive protection. (d) Artifact location with false positive protection.

located at the centre of block H , the block-based metric is calculated using a group of overlapped blocks, forming a spatial kernel aperture, see Fig. 3(a). Basically, the spatial kernel aperture size depends on the size of block H , which contains the center pixel and therefore consists of an odd number of pixels, e.g. size 3×3 or 5×5 pixels. The *activity* detection kernel must be symmetric to avoid the detection to be biased in a certain direction. In order to detect small-sized flat regions, the block-size of centre block H must be kept small. The surrounding blocks in vertical direction (block C and M) automatically obtain the same width, whereas the blocks in horizontal direction (block F , G , I , and J) have equal height. The remaining blocks may have different sizes to create a 2D-filter, with a behavior other than a basic box-filter. The flatness of the region covered by all blocks is now evaluated for the detection of visible artifacts. To this end, the 2D SAD value of each sub-region is compared see Fig. 3(b), against a threshold Th and Tl , except sub-region H , which is compared against threshold Th_{mos} and Tl_{mos} . This specificity is used for the case that block H is situated in the transition between 'flat' and 'texture' or visa versa. As a result, for the centre block H , the SAD classification is either 'flat', 'texture' or 'mosquito', while for the other blocks this is 'flat' or 'texture'. Figure 3(c) depicts a basic block diagram for the spatial-domain artifact-location detector. For a uniform 5×5 SAD block size, Fig. 4 visualizes the main processing results of the spatial-domain artifact-location detection system. Fig. 4(b) depicts on a per-pixel basis, the 2D SAD calculated for the centre block H . Fig. 4(c) shows the spatial detection performance for maximal sensitivity, while Fig 4(d) shows the detection result with a lower sensitivity, to reduce false-positive detections (controlled by T_{flat} threshold value).

By applying spatial reasoning, involving a Constraint Satisfaction Problem (CSP), followed by a Boolean equation, the calculated 2D SAD values are reduced to a binary signal,

indicating if the center pixel of sub-block H is contaminated. The constraint satisfaction problem is defined by

$$\langle X, D, C \rangle, \quad (2)$$

where $X=\{A,B,C,...,M,N,O\}$ is a set of 15 values, representing the calculated 2D SAD values, D the domain of possible 2D SAD values and C is a set of constraints, implemented as thresholds, as visualized in Fig. 3(b). Each element of X is evaluated and mapped onto the possible 2D SAD range, thus

$$f: X \rightarrow D. \quad (3)$$

The evaluation involves a comparison of the calculated 2D SAD value and the individual thresholds. The satisfaction of each SAD value to the imposed constraints, results in a binary indicator, being true or false. Hence, the evaluation results in three 15-tuples sets L , F and T , respectively indicating 'low', 'flat', and 'texture', containing the binary values for each variable of set X . The binary values of these sets are written as

$$\begin{aligned} L &= (l_A, l_B, l_C, l_D, l_E, l_F, l_G, l_H, l_I, l_J, l_K, l_L, l_M, l_N, l_O), \\ F &= (f_A, f_B, f_C, f_D, f_E, f_F, f_G, f_H, f_I, f_J, f_K, f_L, f_M, f_N, f_O), \\ T &= (t_A, t_B, t_C, t_D, t_E, t_F, t_G, t_H, t_I, t_J, t_K, t_L, t_M, t_N, t_O). \end{aligned} \quad (4)$$

For example, if the SAD value for sub-block A is smaller than T_I , then the binary indicator l_A is set to true, otherwise set to false.

Let us now use the above results to determine the contamination status for the center pixel of sub-block H . This status is derived by employing a set of Boolean equations using the binary indicators in the binary datasets L , F and T . To this end, we define the following status parameters, describing block-based regions of equal status. These parameters are 'low', 'flat', 'texture' and 'centre'. The Boolean equations to compute these status parameters are:

$$\begin{aligned} Low &= (l_A \vee l_B \vee l_C \vee l_D \vee l_E \vee l_F \vee l_G \vee l_H \vee l_I \vee l_J \vee l_K \vee l_L \vee l_M \vee l_N \vee l_O), \\ Flat_top &= (f_A \wedge f_B \wedge f_C \wedge f_D \wedge f_E), \\ Flat_bottom &= (f_K \wedge f_L \wedge f_M \wedge f_N \wedge f_O), \\ Flat_left &= (f_A \wedge f_F \wedge f_K), \\ Flat_right &= (f_E \wedge f_J \wedge f_O), \\ Flat &= (Flat_top \vee Flat_bottom \vee Flat_left \vee Flat_right), \\ Tex &= (t_A \vee t_B \vee t_C \vee t_D \vee t_E \vee t_F \vee t_G \vee t_H \vee t_I \vee t_J \vee t_K \vee t_L \vee t_M \vee t_N \vee t_O), \\ Centre &= (\neg l_H \wedge \neg t_H). \end{aligned} \quad (5)$$

The above equations describe the spatial reasoning process associated to the right-hand block in Fig. 3(c). The result of this reasoning process is a pixel-based decision function, indicating whether the centre pixel of sub-block H is artifact contaminated. This pixel-based decision function is based on the following Boolean equation

$$Mosquito = (Low \wedge Tex \wedge Flat \wedge Centre). \quad (6)$$

V. DETECTION SYSTEM IN FREQUENCY DOMAIN

As an alternative to the previous spatial-domain based detection system, we now define an *activity* detector in the transform domain. This can be defined using a 2D-DCT according to $Y=AXA^T$. Matrix X has size 4×4 samples and A is a 4×4 integer DCT transform matrix. After transformation, Y

Fig. 5. Transform-domain filter kernel. (a) Transform-block locations. (b) Feature ranking.

holds a set of 4×4 DCT coefficients describing the local video signal. The transform-domain *activity* filter kernel consists of the same number of sub-blocks as the spatial-domain kernel, see Fig. 5(a), with the difference that the sub-blocks must be equally sized and non-overlapping, because of the involved transformation. In order to reduce the wide variety of energy distributions of each 4×4 region, the energy is matched with four video features as indicated by inequality Equations (7)–(10) below and visualized in Fig. 6(a-d).

$$E_{Un} = \begin{cases} \text{if } \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} Y(u,v)^2 \leq Tun & \text{where } (u,v) \neq (0,0), \\ E_{MAX} & \text{otherwise.} \end{cases} \quad (7)$$

$$E_{Vu} = \begin{cases} \text{if } \sum_{u=0}^{N-1} \sum_{v=1}^{N-1} Y(u,v)^2 \leq Tv_h, \\ E_{MAX} & \text{otherwise.} \end{cases} \quad (8)$$

$$E_{Hu} = \begin{cases} \text{if } \sum_{u=1}^{N-1} \sum_{v=0}^{N-1} Y(u,v)^2 \leq Tv_h, \\ E_{MAX} & \text{otherwise.} \end{cases} \quad (9)$$

$$E_{Mos} = \begin{cases} 0, & \forall |Y(u,v)| = b < Tb \wedge \forall |Y(u,v)| = c < Tc, \\ E_{MAX} & \text{otherwise.} \end{cases} \quad (10)$$

The above inequalities are measurements of the AC energy in the DCT blocks as indicated in Fig. 6, where the X locations are excluded from the energy calculation. For the situation that a feature does not apply, that feature is assigned a maximum energy level E_{MAX} . When all supported features are assigned E_{MAX} , none of the supported video features apply and the 4×4 block is classified as 'texture'. Prior to matching the supported video features, the 2D DCT coefficients can be quantized to influence the video feature matching. To illustrate this, Figure 7 depicts a basic block diagram for visual-artifact location detection in the DCT domain. Let $F=\{E_{Un}, E_{Vu}, E_{Hu}, E_{Mos}\}$ be the set of calculated video features for a single DCT block, see Fig. 6 and $X=\{A,B,C,...,M,N,O\}$ the set containing the finally selected video feature for each DCT block constructing the artifact detection kernel. The

Fig. 6. Video features used for activity measurement in the transform domain. (a) Uniform area. (b) Vertical uniform. (c) Horizontal uniform. (d) Mosquito contaminated. Parameters a , b , c in the blocks are magnitudes of the DCT coefficients, hence $|Y(u,v)|$, and are used in the feature calculation.

Fig. 7. Basic block-diagram for frequency-domain based visual-artifact location detection system, constructed from a feature calculation of 15 DCT blocks as the context for contamination detection of block H .

resulting set of selected features in X is based on feature ranking, see Fig. 5(b), whereby the uniform feature has the highest position and remaining features follow the order of Fig. 6(a-d) according to inequality

$$f(X_i) = \begin{cases} Un & \text{if } E_{Un} < E_{Max} \\ Vu & \text{if } E_{Un} = E_{Max} \wedge E_{Vu} \neq E_{Max} \\ Hu & \text{if } E_{Un} = E_{Max} \wedge E_{Vu} = E_{Max} \wedge E_{Hu} \neq E_{Max} \\ Mos & \text{if } \forall f_i \in F \text{ with } f_i = E_{Max} \wedge E_{Mos} \neq E_{Max} \\ Tex & \text{where } \forall f_i \in F \text{ satisfy } f_i = E_{Max} \end{cases} \quad (11)$$

with X_i being a finally selected feature of set X .

As a next step for deciding whether the centre block H is contaminated, we apply spatial reasoning to the set of finally selected features X by employing a set of Boolean equations. In this way, set X is reduced to a binary decision signal directly indicating the presence of contamination.

This spatial reasoning involves developing a set of rules empirically determined rules, based on the visibility of mosquito noise and ringing, occurring in the presence of flat or low-frequency regions in the vicinity of dominating texture. The empirical rules are based on numerous coding experiments with a broad set of representative video sequences and subsequent visual inspection of the resulting picture quality.

Let us now explain the plausibility of the developed rules. Texture, denoted as T_{xtr} , is defined as one of the DCT blocks is classified as 'texture', see Eqn. (12). With the presence of 'texture' according to Eqn. (12), we verify the presence of coding disturbance in centre block H , as specified by Eqn. (13). The presence of coding artifacts in centre block H , is only visible if there is a sufficiently flat region surrounding this block. Since the amount of flat-region possibilities surrounding block H is very large, we have adopted a set of classification rules to describe such a region that results in visibility. This classification rule evaluates whether a 'flat' classified block has a direct neighbor with the same classification. For example, if block A is 'flat', then we test if block B , G or F are also 'flat', as in Eqn. (14). A similar process is deployed for the remaining positions and columns of the detection kernel window, as given in Eqns. (15) and (16). The column-based results are combined by Eqn. (17), whereas the final 'flat' detection result is given by Eqn. (18). In order to deploy the 'horizontal uniform' and 'vertical uniform' video features, Eqns. (19)–(22) are defined, which are combined into Eqn. (23), providing the detectors ringing

component. The final artifact-location detection involves the combination of the individual Boolean parameter contributions, leading to Eqn. (24).

$$T_{xtr} = \begin{cases} 1 & \exists X_i \in X \text{ where } X_i = Tex, \\ 0 & \text{otherwise.} \end{cases} \quad (12)$$

$$C_{ntr} = \begin{cases} 1 & X_H = Un \vee X_H = Mos, \\ 0 & \text{otherwise.} \end{cases} \quad (13)$$

$$F_{col1A} = \begin{cases} 1 & (X_A = Un \wedge X_B = Un) \vee (X_A = Un \wedge X_G = Un) \vee \\ & (X_A = Un \wedge X_F = Un), \\ 0 & \text{otherwise.} \end{cases} \quad (14)$$

$$F_{col1F} = \begin{cases} 1 & (X_F = Un \wedge X_B = Un) \vee (X_F = Un \wedge X_G = Un) \vee \\ & (X_F = Un \wedge X_L = Un) \vee (X_F = Un \wedge X_K = Un), \\ 0 & \text{otherwise.} \end{cases} \quad (15)$$

$$F_{col1K} = \begin{cases} 1 & (X_K = Un \wedge X_G = Un) \vee (X_K = Un \wedge X_L = Un), \\ 0 & \text{otherwise.} \end{cases} \quad (16)$$

$$F_{col1} = F_{col1A} \vee F_{col1F} \vee F_{col1K} \quad (17)$$

$$F_{lat} = F_{col1} \vee F_{col2} \vee F_{col3} \vee F_{col4} \quad (18)$$

$$R_t = \begin{cases} 1 & (X_M = Un \wedge X_H = Un) \wedge (X_C = Vu \vee X_C = Hu), \\ 0 & \text{otherwise.} \end{cases} \quad (19)$$

$$R_l = \begin{cases} 1 & (X_I = Un \wedge X_H = Un) \wedge (X_G = Vu \vee X_G = Hu), \\ 0 & \text{otherwise.} \end{cases} \quad (20)$$

$$R_r = \begin{cases} 1 & (X_G = Un \wedge X_H = Un) \wedge (X_I = Vu \vee X_I = Hu), \\ 0 & \text{otherwise.} \end{cases} \quad (21)$$

$$R_b = \begin{cases} 1 & (X_C = Un \wedge X_H = Un) \wedge (X_M = Vu \vee X_M = Hu), \\ 0 & \text{otherwise.} \end{cases} \quad (22)$$

$$R_{ing} = R_t \vee R_l \vee R_r \vee R_b \quad (23)$$

$$M_{osq} = (T_{xtr} \wedge C_{ntr} \wedge F_{lat}) \vee R_{ing} \quad (24)$$

Figure 8 visualises the artifact-location detection in the frequency domain. Figure 8(b) depicts the raw detected video features without any spatial reasoning for the situation that the detector threshold in Eqns. (7)–(10) are set to zero. From

Fig. 8. Visualization of frequency-domain based artifact detection. (a) Original image. (b) Raw-feature detection, thresholds and quantization are set to zero. (c) Raw-feature detection with thresholds set to non-zero and $Q=4$. (d) Detected visual-artifact location.

TABLE I
VALIDATION RESULTS FOR SPATIAL-DOMAIN BASED VISUAL ARTIFACT-LOCATION DETECTION ON SD/HD VIDEO RESOLUTION

Detection kernel size (Pixels,Lines)	Kernel block size	Contour tracking	Detection performance	Horizontal expansion	Vertical expansion
11×7	fixed	good	reasonable	limited	limited
15×7	unequal	good	acceptable	good	limited
19×7	unequal	good	good	good	limited
21×13	fixed	sufficient	good	good	good

TABLE II
VALIDATION RESULTS FOR FREQUENCY-DOMAIN BASED VISUAL ARTIFACT-LOCATION DETECTION ON SD/HS VIDEO RESOLUTION

Detection kernel size (Pixels,Lines)	Kernel block size	Contour tracking	Detection performance	Horizontal expansion	Vertical expansion
20×12	fixed	sufficient	good	good	good

Fig. 8(b) it becomes clear that without proper thresholding and quantization, the features have a more random behaviour due to the presence of noise. When applying some coefficient quantization, thereby lowering the noise level, and setting appropriate thresholds, the video features become suitable for spatial reasoning, see Fig. 8(c). Also in this figure, the spatial reasoning is absent. Applying spatial reasoning to the raw video features, see Fig. 8(c), results in detected artifact-locations as depicted in Fig. 8(d). For natural video, the video features 'uniform' and 'texture' are the two essential features required for artifact-location detection. The contribution of the video features 'horizontal uniform' and 'vertical uniform' is typically limited. The latter two features perform good in artificial video, which typically contains firm vertical and horizontal transitions. The video feature 'mosquito' enables the detection of potential noise regions to be concentrated along the textured region.

VI. LOW-PASS FILTER (LPF) STRENGTH CONTROL

The purpose of artifact location finding, is to remove the coding artifacts at the appropriate places, where they are typically visible. At those places, the noise should be removed with a low-pass filter. In order to avoid visible switching artifacts introduced by the filter operation, we perform a smooth filter transition for both on- and off-switching of the filter. This smoothness is obtained by a transient function that gradually increases or decreases the filtering strength. An advantage of this approach is to improve the final filtering coverage, as the detection is only block-based thus will leave open small areas with coding artifacts in the transition regions between flat and textured transitions. The Gaussian spread of the smooth filtering operation thereby also covers the open small regions not found by the detector.

As a refinement for the Gaussian on/off-switching of the filtering operation, we tune the shape of the Gaussian to adaptively match the nature of the surrounding video characteristics. For example, when the surrounding part of the detection signal is textured, the Gaussian filtering effectuation should not blur the content of the texture. This Gaussian control of the low-pass filter further depends on the resolution of the video picture i.e. SD, HD or up-scaled SD.

A. Influence of Video Resolution on Detection Aperture

Video coding-artifacts present in decoded Standard-Definition (SD) video are expanded when up-scaled to full HD, see Fig. 2(d-e). Hereby, this up-scaling ratio differs for horizontal and vertical direction, depending on whether the original video originates from conventional PAL or NTSC formats, e.g. generated by DVD players. For those formats there is also a difference in vertical up-scaling factor. For both formats, the up-scaling in horizontal direction is a factor 2.6, while for the vertical direction this up-scaling factor is either 1.9 for PAL-based video or 2.3 for NTSC-based video. From Fig. 2 (b) it becomes clear that for non-up-scaled formats, coding artifacts can occur at a distance of up to 7 pixels, with respect to the location from which the coding artifact originates. Such a situation occurs e.g. when an 8×8 region has a uniform behavior except for one corner pixel.

For artifact-location detection in SD or HD resolution on a per pixel basis, this results in 7 pixels in both horizontal (left, right) directions and 7 pixels in both vertical (up, down) directions, resulting in a final detection-kernel aperture of 15×15 pixels.

Sufficient detection coverage of artifact locations in up-scaled video requires the LP filter kernel to have an aperture, which equals 7 pixels times the up-scaling factor, resulting in 39×29 pixels for PAL and 39×33 pixels for NTSC.

B. None-Pixel Accurate Detection

The two block-based visual-artifact detection systems described in Section IV and V generate a low-pass filter control-signal for either a single pixel or a group of pixels. Visual-artifact location detection in either SD or HD resolution or up-scaled video, typically results in a gap between the detected locations and the actual 'texture' region see e.g. Fig. 4(d) and Fig. 8(d). This previously discussed gap, which has a typical width of a few pixels, can be bridged via the Gaussian control of the detected visual-artifact location.

C. Gaussian Aperture Control of Low-Pass Filter-Strength

The Gaussian control window is limited to blocks that are directly adjacent to the blocks, indicated by the artifact-location detection system. This enables a smooth filter-strength operation. A smooth control particularly holds for the

Fig. 9. Basic block-diagram of entropy-based local-adaptive LP filter.

situation that the neighboring blocks reveal a ‘flat’ region, while a steep filter-strength decline holds for the situation that the neighboring blocks reveal a ‘texture’ region. The open areas in the detection, potentially contaminated with artifacts, are covered by this increased filter aperture, which is a function of the block-size deployed. For implementation efficiency, we have included the filter aperture control in the spatial reasoning block of Fig. 7.

VII. EXPERIMENTAL RESULTS

The spatial-domain based detection system has been validated using an artifact-location detection kernel utilizing equally- and unequally-sized blocks. The frequency-domain based visual-artifact location detection system is validated for a fixed block-size of 4×4 . Both systems have been tested on SD/HD and up-scaled video. The final Gaussian LPF strength control is validated regarding filtering performance using an entropy-based locally-adaptive LPF [15] as depicted in Fig. 9.

A. Coding-Artifact Location Detection in SD/HD Video

The spatial-domain detection system has been tested for 4 differently constructed spatial-domain artifact-location detection kernels, as listed in Table I.

A spatial-domain artifact-location detection kernel with an aperture of 11×7 pixels is based on a 3×3 fixed block-size. For SD/HD video, this detection kernel results in a reasonable detection performance. However, it has a very good tracking performance to follow arbitrary including ‘small-sized’, objects. Moreover, it has a poor performance on discriminating intended texture from artifact-contaminated regions. Furthermore, due to the small footprint of the artifact-location detection kernel, not all contaminated pixels are detected. These ‘far away’ located pixels are not covered by the Gaussian LPF aperture control, resulting in reasonable

detection and filter performance, as depicted in Fig. 10(a).

A spatial-domain artifact-location detection kernel with an aperture of 21×13 pixels is based on a fixed block-size of 5×5 . For SD/HD video, this detection kernel results in a good detection performance and a good discrimination of intended texture versus artifact-contaminated locations. Unfortunately, it has a coarse tracking performance, which degrades for small-sized objects. Furthermore, due to the large footprint of the artifact-location detection kernel, all contaminated pixels including the ‘far away pixels’ located in large flat regions are detected, see Fig. 10(c), although this may not always be ensured for small-sized flat regions.

A spatial-domain artifact-location detection kernel with an aperture of 15×7 or 19×7 pixels shows for SD/HD video a detection improvement compared to the 11×7 -sized kernel, except for the detection in small-sized flat regions and the expansion of the LPF strength information in the vertical direction, see Fig. 10(b). This vertical control signal degradation is caused by the limited vertical size of the deployed block-size in the artifact-detection system.

The spatial-domain based detection system has four thresholds, which influence the detection and a single control signal, T_{flat} , to influence the detection-system sensitivity, simplifying the final detection control. The frequency-domain based detection system is validated using a frequency-domain artifact-location detection kernel with an aperture of 20×12 pixels, which is based on non-overlapped 4×4 fixed block sizes. For SD/HD video, this detection kernel has similar detection performance regarding discrimination and tracking of small-sized regions as the spatial artifact-location detection, with kernel aperture of 21×13 pixels, see Fig. 10(c) and (d).

The frequency-domain based detection system has various control signals to influence the detection-system sensitivity, flat-region detection in blurred as well as in-focus situations. This wide range of control enables various suppression concepts ranging from low, medium up to high contamination suppression.

B. Coding-Artifact Location Detection in Up-scaled Video

Artifact-location detection in up-scaled video requires a larger artifact-location detection kernel. As discussed in Section VI, up-scaled SD video theoretically requires a detection aperture of 39×33 pixels. For up-scaled video, such a detection kernel becomes expensive in terms of embedded memory. System costs can be positively influenced when separating the detection stage from the spatial reasoning stage. Separation allows the embedded video memory to be reduced to a minimum, which depends only on the largest vertical block height used in the artifact-location detection kernel.

For the situation that separation is not possible, the artifact-location detection kernel as depicted in Fig. 3(a) and Fig. 5(a) is expanded in the vertical direction, preserving horizontal and vertical symmetry and avoiding biasing in a certain direction. This results in a detection kernel consisting of 25 sub-blocks. On the basis of the detection results, depicted in Table I and II, the final aperture size for up-scaled video is 20×20 pixels for the frequency-based detection system and 19×19 pixels for the spatial-based detection system. These aperture sizes are a

Fig. 10. Artifact-location detection and sharpened final LPF result for SD/HD resolution. (a) Difference signal and final LPF result for spatial detection system with kernel-size 11×7 . (b) Difference signal and final LPF result for spatial detection system with kernel-size 19×7 . (c) Difference signal and final LPF result for spatial detection system with kernel-size 21×13 . (d) Difference signal and LPF result for frequency detection system with kernel-size 20×12 .

(a)

(b)

(c)

Fig. 11. Artifact-location detection in up-scaled video. (a) Up-scaled SD to HD video frame. (b) Artifact location detection using frequency-domain detection system with kernel size 20×20 . (c) Artifact location detection using spatial-domain detection system with kernel size 19×19 .

trade-off between hardware cost and artifact-location detection performance.

Figure 11 depicts the artifact-location detection and associated Gaussian filter control for both frequency- and spatial-based artifact-location detection systems. On average, the detection performance is equal, with some subtle differences, which have a resemblance with the native video detection performance.

Fig. 11 reveals a clear performance difference between the two detection systems. The frequency-based detection and corresponding filtering, clearly outperforms the spatial detection system. This is clearly visible in the background details of the image, which contain low-amplitude information. Due to the lack of discrimination, the spatial detection system, results in a filter control signal that leads to strong LPF, so that the background details become visibly

blurred. The frequency detection system preserves the background details.

Although the locator information indicates where to apply LPF the LPF, see Fig. (9), is also controlled by an entropy calculation within a 3×3 region, which acts as a second guard for severe low-pass filtering. On the basis of the calculated entropy, the filter coefficients as well as the filter aperture are adjusted. In this way, the introduction of new artifacts due to severe low-pass filtering is limited, as depicted in Fig. 10(a)(b)(c)(d).

VIII. CONCLUSIONS

We have presented two efficient artifact-location detection systems, operating either in the spatial- or in the frequency-domain. The two detection systems are capable of detecting locations, which are potentially contaminated by visible mosquito noise or ringing. The concept deployed by both detection systems involves a feature detection measurement followed by spatial reasoning.

Both detection systems allow separation of the feature measurement and spatial reasoning stages, so that the involved embedded memory is significantly reduced. Furthermore, cost effective Boolean-based arithmetic is used for spatial reasoning, resulting in a low-complexity implementation and suitability for real-time operation. Separability is beneficial when detecting artifact-locations in up-scaled video, which theoretically requires a spatial reasoning aperture involving 33 lines.

The two detection systems yield a comparable performance, with several differences.

The spatial-domain based detection system has a good edge-detection performance and a high-quality tracking performance of the detected edge. However, the discrimination between intended texture and visual-artifact contamination is highly dependent on the selected block sizes. Limiting the vertical memory size of the artifact-location detection kernel, would reduce costs, but influences the final local-adaptive low-pass filtering, resulting in noticeable remaining artifacts. Good detection performance for SD/HD video is obtained for an artifact-location detection kernel with an aperture of 19×13 pixels and 19×19 pixels for up-scaled video, based on a fixed sub-block size of 5×5 pixels.

The frequency-domain based detection system has more coarse edge detection and tracking performance. However, the discrimination between intended texture and visual-artifact contamination is better compared to the spatial domain solution. Good detection performance for SD/HD video is obtained for an activity kernel with an aperture of 20×12 pixels and 20×20 pixels for up-scaled video, based on a fixed sub-block size of 4×4 pixels. The frequency-based solution is preferred over the time-domain system, because it offers far more control over the various forms of detection, distinguishing between a broad range of video details in foreground and background. This is enabled by the quantization stage, prior to the feature detection.

Both detection systems suffer from non-pixel accurate artifact-location detection with respect to the texture from which the artifact originates, so that a gap in detected artifact location occurs. We have proposed a Gaussian-shaped control for the subsequent low-pass filtering, to cover such gap areas. This smart filter control is adaptive to the local video features, thereby optimizing the smooth filtering performance without loss of texture details.

REFERENCES

- [1] C. Mantel, P. Ladret and T. Kunlin, "Temporal mosquito noise corrector", International Workshop Quality of Multimedia Experience, QoMEX 2009, pp. 29-31 July 2009.
- [2] I. Kirenko, S. Ling and A. Nakonechny, "Quality Enhancement of Compressed Video Signals", Proceedings of ICCE 2008, pp. 1-2, 9-13 Jan. 2008.
- [3] S.J.P. Westen, R.L. Lagendijk, and J. Biemond, "Adaptive spatial noise shaping for dct based image compression," Acoustics, Speech, and Signal Processing, IEEE International Conference on, vol. 4, pp. 2124-2127, 7-10 May 1996.
- [4] ISO/IEC 13818-2: "Information Technology - Generic Coding of Moving Pictures and Associated Audio Recommendation H.262 (video)".
- [5] ISO/IEC 14496-2: "Information Technology Coding of audio-visual objects - Part 2: Visual"
- [6] ISO/IEC 14496-10: "Information Technology Coding of audio-visual objects - Part 2: Advanced Video Coding"
- [7] J. Hutchison, "Plasma display panels: the colorful history of an Illinois technology", ECE alumni news, university of Illinois, vol. 36(1), 2002.
- [8] W. C. O'Mara, Liquid crystal flat panel display: manufacturing science and technology, Van Nostrand Reinhold, 1993, ISBN 0442014287.
- [9] K. Jostschulte, "A new cascaded spatio-temporal noise reduction scheme for interlaced video", Proceedings of ICIP 1998, pp. 493-497, 4-7 Oct 1998.
- [10] C. Fenimore, J. Libert and P. Roitman, "Mosquito noise in MPEG compressed video: test patterns and metrics", Proc. Of SPIE: Human Vision and Electronic Imaging, SPIE Vol. 3959, San Jose, CA, pp. 604-612, 24-27 Jan., 2000.
- [11] M. Yuen and H.R. Wu, "A survey of hybrid MC/DPCM/DCT video coding distortions", Signal Processing Special issue on image and video quality metrics, Vol. 70, Issue 3, pp. 247-278, Nov. 1998
- [12] F.X. Coudoux, M.G. Gazalet, and P. Corlay, "A postprocessor for reducing temporal busyness in low-bit-rate video applications," Signal Processing: Image Communication, vol. 18, pp. 455-463, July 2003.
- [13] D.T. Vo, T.Q. Nguyen, S. Yea and A. Vetro, "Adaptive Fuzzy Filtering for Artifact Reduction in Compressed Images and Videos", IEEE Transaction on Image Processing, Vol. 18, No. 6, pp. 1166-1178, June 2009.
- [14] N. Yao, A. Vetro, S. Huifeng and K.E. Barner, "Coding artifacts reduction using edge map guided adaptive and fuzzy filtering", IEEE International Conference Multimedia and Expo, Vol. 2, pp. 1135-1138, 30-30 June 2004.
- [15] L. Shao and I. Kirenko, "Coding Artifact Reduction Based on Local Entropy Analysis", IEEE Transaction on Consumer Electronics, Vol. 53, No. 2, pp. 691-696, May 2007.
- [16] Patent US7949051, "Mosquito Noise Detection and Reduction".
- [17] Patent US2008085059A1, "Image Processing Method and Device for Performing Mosquito Noise Reduction".

Onno Eerenberg was born in Zwolle, the Netherlands, in 1966. He graduated from the Polytechnical College in Amsterdam in 1992. He joined Philips Research Laboratories Eindhoven, The Netherlands where he worked in the Magnetic Recording Systems department on digital video and data recording systems. He was involved in

several European research projects in this area and was involved in the implementation of e.g. video compression systems. He received an MSc degree in engineering product design in 1998 from the University of Wolverhampton, UK. He is currently a PhD Candidate at the Eindhoven University of Technology. He holds twenty patent applications in the field of digital recording, digital reception and digital television. Furthermore, he author of several publications and two book chapters in aforementioned fields.

applications in the field of digital video and is co-author of several publications in the field of digital television.

Jeroen Kettenis received a masters degree in Electrical Engineering from Delft University of technology (1985). In 1985 he joint Philips Consumer Electronics where he was involved in the development of storage and television products. In 1999 he joint Philips Semiconductors, where he continued to work on digital television. Jeroen is currently working as a picture quality architect for Sigma Designs. He holds several patent

Peter. H.N de With graduated in Electrical Engineering (MSc., ir.) from the University of Technology in Eindhoven. In 1992, he received his Ph.D. degree from the University of Technology Delft, The Netherlands, for his work on video compression for digital video recording applications. He joined Philips Research Labs Eindhoven in 1984, where he became a member of the Magnetic Recording Systems Department and set-up the first DCT-based

compression systems. From 1985 to 1993, he was involved in several European research projects on SDTV and HDTV recording. He was the leading video compression expert for the DV camcorder standard from 1989-1993. In 1994, he became a member of the TV Systems group at Philips Research Eindhoven, where he was leading the design of advanced programmable video architectures. In 1996, he became senior TV systems architect and in 1997, he was appointed as full professor at the University of Mannheim, Germany, at the faculty Computer Engineering. In Mannheim he was heading the chair on Digital Circuitry and Simulation with the emphasis on video systems. Between 2000 and 2007, he was with LogicaCMG in Eindhoven as a principal consultant and distinguished business consultant and simultaneously, he is professor at the University of Technology Eindhoven, at the faculty of Electrical Engineering, heading the chair on Video Coding and Architectures as part of the Dept. on Signal Processing Systems. In the period 2008 – 2010, he was vice president Video Technology at CycloMedia Technology, Waardenburg, The Netherlands, establishing image analysis applications. Early 2011, he was appointed scientific director of the Centre for Care & Cure Technology at the University of Technology Eindhoven and theme leader on smart diagnosis for the University. He has written and (co-) authored over 300 papers on video coding, analysis, architectures and their realization. He is the (co-)recipient of multiple papers awards of the CES, VCIP and transactions papers. Mr. de With is Fellow of the IEEE, program committee member of the IEEE CES and ICIP, board member of various communities and scientific boards and R&D advisor to multiple companies.