Abstract--The core of many video coding standards is formed by the Discrete Cosine Transform (DCT) for de-correlating spatial video data. When quantizing DCT sub-bands, 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 artifact-free regions. This location information can be derived, either in the time- or frequency 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 possible detection domains. The time-domain solution has a 10-97% location detection performance, whereas the frequency domain results in 70-100% detection performance.

I. INTRODUCTION

Image and video communication have been benefiting from advances 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. Although this is an efficient method for removing irrelevant information from a video signal, there is also a strong drawback. In order to achieve sufficient compression ratio, also relevant information in the form of high-frequency information is removed by means of quantization. Removal of high-frequency information not only results in lack of sharpness, but also introduces coding artifacts. Examples of typical coding artifacts are blockiness, ringing and mosquito noise. Modern digital televisions perform Temporal Noise Reduction (TNR), which not only removes Gaussian noise, but to a certain extent also reduces coding artifacts, provided that these artifacts are not static [1]. However, for the situation that the artifact is static, spatial filtering has to be applied to attenuate the disturbance [2]. Visual artifact-location information is crucial for controlling the spatial filtering strength, in order to avoid loss of detail.

This paper describes two block-based visual artifact-location detection systems, suitable to detect regions in a DTV decoded image, which have a high probability to be contaminated with coding artifacts. Although the detection criteria are equal, two experimental detectors each operating in a different domain, are compared regarding the detection performance.

II. DETECTION OF CODING ARTIFACT-PRONE LOCATIONS

Transform coding introduces artifacts, which are clearly noticeable around the transition between texture/edges and flat/low-frequency regions [3], see Fig. 1. This observation is a key feature for locating mosquito/ringing prone locations. The detection of such artifact-contaminated locations requires an activity measurement, revealing the presence of edges within a bounded region, followed by a spatial reasoning step, which results in a binary decision: contaminated versus not-contaminated. We investigate two different approaches.

A. Detection system in time domain

The activity metric in the time domain is based on a simple 2D high-pass filter, implemented as a 2D SAD according to

\[
SAD = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} |P(x,y) - P(x+1,y)| + \sum_{y=0}^{N-1} \sum_{x=0}^{M-1} |P(x,y) - P(x,y+1)|
\]

For each pixel in the image, located at the centre of block H, the block-based metric is calculated using overlapped blocks, constructing a spatial kernel aperture, see Fig. 2(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 3x3 or 5x5 pixels. The surrounding blocks in vertical direction automatically obtain the same width, whereas the blocks in horizontal direction 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 2D-SAD value of each sub-region is compared, see Fig. 2(b), against a threshold \( Th \) and \( Tl \), except sub-region H, which is compared against threshold \( Th_{mos} \) and \( Tl_{mos} \). For the centre block H, the SAD classification is either ‘flat’, ‘texture’ or ‘mosquito’, while for
the other blocks this is 'flat' or 'texture'. Using spatial reasoning, the results of the threshold operations are reduced to a binary signal, indicating if the center pixel of sub-block $H$ is contaminated.

### B. Detection system in transform domain

The activity detector in the transform domain is based on a 2D-DCT according to $Y=AXAT$. Matrix $X$ has size 4x4 samples and $A$ is a 4x4 transform matrix. After transformation, $Y$ holds a set of 4x4 DCT coefficients describing the local video feature. The transform-domain activity filter-kernel consists of the same number of equally sized non-overlapped sub-blocks as the time-domain kernel, see Fig. 3(e).

In order to reduce the wide variety of energy distributions of each 4x4 region, the energy is matched with five video features. Figure 4 depicts the four video features, while the fifth video characteristic is 'texture', which applies if none of the four video features match. Prior to matching the supported video features, the 2D DCT sub-bands can be quantized to influence the video feature matching. The video matching process investigates the energy on locations indicated by a, see Fig. 3(a-c). For each video feature, the locations are squared, summed up and compared against a threshold $Tun$ and $Tvh$, for the uniform and horizontal/vertical uniform video feature respectively. For the mosquito video feature, the sub-bands at location $b$ and $c$ are compared against a threshold $Tb$ and $Tc$. The final result is a spatial region which is classified by maximal five video features. The final classification is based on feature ranking, whereby the uniform feature has the highest position and remaining features follow the order of Fig. 3(a-d). On the basis of spatial reasoning, the transform-domain activity filter-kernel reduces this to a binary signal, indicating if the center block $H$ is contaminated.

### III. EXPERIMENTAL RESULTS

We have tested the detection systems for a broad range of TV images and specific test images for which the detection performance can be carefully evaluated. Due to space limitations, we present here some results on the basis of a few test images. For the test image in Fig. 1(a), the artifact locations are determined using a block grid of 4x4, using the definition that artifacts are most visible in the vicinity of the transition 'flat' to 'texture', or visa versa. Hereby the pixels contained by the 4x4 blocks, form the reference pixels for validating the time-domain-based detector. On the basis of this reference set, the performances of the two artifact-location systems are validated. The coverage results are depicted in Table 1 and visualized in Fig. 4. Figure 4(c-d) indicate the detection performance on a region containing a static logo.

### IV. CONCLUSIONS

We have studied two block-based visual artifact-location detection systems, suitable to detect static regions in a DTV image. Visual artifact-location detection is successfully achieved, with a detection performance of 70-100% for the frequency domain and 10-97% for the time domain. The frequency-domain detector shows a higher selectivity and consistency, due to knowledge of the underlying video features. The time-domain detector locates on the average, the majority of the artifact-prone locations, but lacks selectivity.

### REFERENCES

