Note: This is a pre-print version subject to changes in formatting

Compression artifacts in modern video coding and state-of-the-art means of compensation

Andreas Unterweger
University of Salzburg, Austria

ABSTRACT

This chapter describes and explains common as well as less common distortions in modern video coding, ranging from artifacts appearing in MPEG-2 Video, MPEG-4 Part 2, H.264 and VC-1 to scalable and multi-view video coding based distortions, including the proposals for next generation video coding (NVC). In addition to a discussion about avoiding these artifacts through encoder-side measures, a state-of-the-art overview of their compensation at the decoder side is given. Finally, artifacts emerging from new sophisticated coding tools in current and upcoming video coding standards are discussed.

INTRODUCTION

As the coding tools used in modern video coding advanced in the last decades, new compression artifacts emerged, creating the need for sophisticated means of compensation. As the human eye is eventually the final recipient of the coded video, including distortions, artifact compensation based on human visual perception is an important research field, which is faced with new challenges due to new coding tools and the respective new artifacts induced by them.

It is important to be aware of these new artifacts and to analyze their sources in order to be able to compensate for them. As new coding tools are developed, most prominently represented by the current contributions to NVC, a basic understanding of the effects of the artifacts caused by these coding tools as well as their effect on the overall video quality is crucial. Although most of the current research is focused on the compensation of blocking, blurring and ringing artifacts and the development of new coding tools, this book chapter gives an overview of the artifacts caused by existing and new coding tools, focusing on mainstream block-based video coding represented by MPEG-2 Video, MPEG-4 Part 2, H.264, VC-1 and the amendments to H.264 for scalable and multi-view video coding. The interested reader may additionally find an overview of Wavelet-based compression artifacts appearing in Motion JPEG 2000 and others in Watson (1997) and Ramos (2001). Literature on non-mainstream video coding formats like Ogg Theora (Xiph.Org Foundation, 2011) is sparse (Crop, 2010) and therefore out of the scope of this book chapter.

The description of artifacts herein includes a discussion on the impact of new coding tools on artifacts in general and suggestions on how to minimize the appearance of these artifacts, thus eliminating the requirement for compensating them at the decoder side. After summarizing the properties and causes of commonly appearing artifacts such as blocking, blurring and ringing, including a number of artifacts originating from new coding tools, a short outlook on the perception of new artifacts and their connection to quality metrics concludes this chapter.

BACKGROUND

The origins of artifacts in block based transform video coding are, in most cases, directly or indirectly related to quantization errors in the transform domain, which are inevitable when lossily compressing images or sequences thereof. Since the first coding standards of this kind, e.g. JPEG for still image coding and H.261 for video coding, various related visual artifacts have been discussed throughout the literature.

Blocking artifacts

Perhaps *the* most "famous" and most widely studied artifacts in today's block based video coding are blocking artifacts which occur due to the division of frames into macroblocks of rectangular shape. All blocks are coded separately from one another despite a possibly existing spatial correlation between them, yielding visible edges at macroblock borders. Due to the equidistant distribution of macroblock borders in JPEG, MPEG-2 Video and MPEG-4 Part 2 which is caused by the constant transform size of 8x8 samples, blocking artifacts are, in most cases, easily spotted by the Human Visual System (HVS) as a regular structure which does not belong to the image (Wu, 2006).

Due to the intense research concerning blocking artifacts, a number of possibilities for their compensation is available, e.g. (Oosa, 1998) and (Triantafyllidis, 2002). As both MPEG-2 Video and MPEG-4 Part 2 do not have an integrated deblocking filter, the artifact compensation has to be performed at the decoder side. In order not to cause a drift between encoder and decoder, deblocking has to be performed as a form of post processing on the decoded pictures which are displayed, but must not be applied to reference pictures which are used for motion compensated prediction.

Simple forms of deblocking involve low pass filtering at or around all macroblock borders, which causes blurring artifacts at borders which do not expose blocking artifacts (see below), whereas advanced approaches use edge detection algorithms to identify visually prominent edges or adaptively adjust the filter strength and/or area of influence, i.e. the number of samples around the macroblock border, based on image properties, quantizers, coding modes etc. The latter approach is incorporated in both H.264 and VC-1 in the form of an in-loop deblocking filter which is applied to all coded pictures before storing them in the reference buffers, yielding filtered references which are used for motion compensation. As experiments have shown that an image or video with blurring artifacts arising from strong deblocking appears more pleasant to a typical viewer than the corresponding unfiltered image or video (Wiegand, 2003), this supports the decision to incorporate in-loop deblocking filters into both video coding standards to improve the perceived quality of the decoded pictures.

Blurring artifacts

As mentioned above, strong deblocking can expose blurring artifacts due to the loss of high frequencies caused by low pass filtering during the attempt to flatten block edges. However, blurring may also be a result of quantization, if all high-frequency components in the transform domain are quantized to zero, yielding a low-pass-like behavior of the transform and quantization process. Using coarse quantization, i.e. selecting a high quantization parameter, favors blurring as it increases the probability of high-frequency components to be quantized to zero. As the HVS notices the loss of high frequency components to a lower degree than the loss of low-frequency components, the quantization matrices defined by MPEG-2 Video, MPEG-4 Part 2, H.264 and VC-1 cause a coarser quantization of high-frequency components, yielding blurring artifacts for high quantization parameters (Wu, 2006).

All standards mentioned above have no built-in filter to compensate for blurring artifacts and therefore require decoder-side deblurring algorithms, if desired. As the high-frequency components have been quantized to zero at the encoder side, they cannot be restored at the decoder side. Therefore, it is necessary to introduce high frequency components similar to noise, based on the image properties and the number of coefficients quantized to zero. Although sharpening might be an option in some cases, a number of approaches rely on boundary conditions (Ng, 1999) or inverse filtering (Biemond, 2005), yielding oversharpening artifacts or introducing noise. It is important to note that motion blur causes similar effects in the transform domain but requires different forms of compensation, involving – amongst others – deconvolution (Ben-Ezra, 2004).

In chroma-subsampled images (Kerr, 2009), blurring is often also referred to as color bleeding, as one chroma sample may stretch across multiple luma samples. Thus, the blurred chroma sample(s) spread(s) across a wider area, i.e. multiple luma samples, around an edge or other areas of high frequencies in the chroma planes (Wu, 2006). Although chroma subsampling increases the perceived strength of color bleeding due to the wider area affected, color bleeding can also occur in pictures where there is no chroma subsampling.

Mosquito noise

At the borders of moving objects, an artifact called mosquito noise or mosquito effect appears, when a block is coded using inter frame prediction, but only a part of the predicted block contains the predicted moving object. The (static) rest of the block, therefore, differs strongly from the prediction, thus accounting for a major part of the total prediction error. As all video coding standards mentioned above use a form of prediction which operates in the transform domain, the quantization error together with the prediction error may yield a high concentration of error energy in the high frequency components due to the attempt to reduce the ringing from the prediction at the object border, thus yielding high frequency noise in the picture domain. The latter is referred to as mosquito noise if it is visible over a number of frames and the conditions described above apply.

Mosquito noise is visually prominent as the prediction error changes from frame to frame, yielding different high frequency noise patterns. It has to be noted that different coding of the same picture region across multiple pictures may also expose mosquito-noise-like artifacts. Another form of mosquito-like noise can be caused by encoder/decoder drift in MPEG-2 Video and MPEG-4 Part 2, which is due to the finite precision of the floating point operations involved in the transform and inverse transform process, yielding an imperfect reconstruction, differing between encoder and decoder, thus causing a drift between the two which propagates through prediction (Wiegand, 2003).

Ringing artifacts

Another common form of artifacts which manifest as "halos" around sharp edges is known as ringing (Wu, 2006). As steep edges in general contain a larger range of frequencies, the quantization of blocks with steep edges yields an insufficient reconstruction through the sum of basis functions, forming a less steep slope at the position of the original edge and both over- and undershooting at the samples around the original edge. In one-dimensional Fourier analysis, this is also known as Gibbs Phenomenon. Through the smoother slope of the edge it may appear blurry due to the loss of high frequency components whereas the over- and undershooting typically introduces the "halo"-like effect initially mentioned, creating a silhouette-like shade parallel to the original edge.

Note that the "halo" effect also affects low quantization parameters, i.e. small quantization step sizes, as a higher number of non-zero high frequency basis functions does not necessarily improve the approximation of an edge, thus not always decreasing the amount of ringing around sharp edges. Therefore, ringing may also be present in videos coded with high bit rates.

Although there are multiple approaches available describing how to measure ringing effectively, like (Shen, 1998) and (Liu, 2010), there are currently only two approaches available for compensating ringing artifacts, disregarding approaches optimized for JPEG 2000 and the like (Chang, 2005). Despite a sophisticated approach based on projections onto convex sets (POCS) (Zakhor, 2002) which is also used in the context of compensating blocking artifacts, there is an approach based on edge detection and adaptive filtering (Park, 1999) optimized for MPEG-4 Part 2, but applicable to all DCT and quantization based video coding standards which use a transform size of 8x8 samples, including H.264 with its High profile up to a certain extent.

Stair case and basis pattern artifacts

Another visual artifact closely related to ringing is the so-called stair case artifact which refers to the incapability of horizontal and vertical basis functions (as building blocks of the DCT and its variations) to accurately represent diagonal edges (similar to steep edges), thus resulting in the visually prominent presence of horizontal or vertical basis functions (Wu, 2006). Across multiple coded macroblocks, the appearance of a diagonal edge may be similar to the pattern of a stair case rather than that of a smooth diagonal connection between two points. Through the influence of blocking, stair case artifacts become visually more prominent as the "stair case step size" equals the size of a macroblock.

High quantization may reduce the number of non-zero coefficients in a transformed block to one, yielding so-called basis pattern artifacts which are similar to stair case artifacts, but exhibit a single basis function with its prominent picture domain representation. Note that this artifact is not limited to the scenarios described for stair case artifacts, but applies in general, when high quantization parameters are used which increase the probability of reducing the number of non-zero coefficients to one. If only the DC coefficient remains non-zero after quantization, a smooth “non-texture” block is coded which exhibits strong blocking and blurring and together with adjacent blocks of equal appearance forms a visual distortion referred to as mosaïking.

Summary

As apparent from the commonly described artifacts above, the causes for most of them are related to distortions through quantization. Some of them share manifestation patterns in terms of the quantization step size range and/or the number of non-zero quantized coefficients, increasing the probability under certain circumstances for artifacts to appear together. Therefore, Table 1 summarizes the quantization-dependent characteristics of these artifacts, together with possible accompanying artifacts. Figure 1 visualizes some of the artifacts described in Table 1. Note that this overview only covers the most relevant artifacts described in the literature. Further artifacts which mostly resemble the artifacts described herein may be found in the subsequent sections and (Wu, 2006).

	Causes of appearance			**Coexisting artifacts**
Artifact	**Reason for appearance**	**Typical quantization step size (range)**	**Number of non-zero quantized coefficients**	**Possibly appearing together with**
Blocking	Independent coding of spatially correlated adjacent blocks	High, but also depending on quantization step size (difference) of adjacent blocks	-	Mosaïking if neighboring blocks are also affected
Blurring	Loss of high-frequency components	High	Low (or zero) for high-frequency coefficients	Ringing at sharp edges, color bleeding (chroma)
Ringing	Insufficient approximation of steep edges	-	-	Blurring
Stair cases	Insufficient approximation of diagonal edges	-	-	Basis patterns for low quantization step sizes
Basis patterns	Loss of all but one transform coefficients	Very high	1	Stair cases
Mosquito noise	Quantization of high-frequency components and prediction errors	-	High for high-frequency components with a notable amount of total error energy	-

Table 1: Common artifacts in video coding and their causes (hyphens denote that the given artifact does not depend directly or necessarily on a certain value or value range of the respective cause of appearance)

Figure 1: Common artifacts in video coding caused by high quantization and the H.264 deblocking filter (right image part only): blocking at the ceiling (example marked with 1), blurring through deblocking in the right image part (2), ringing at the window borders on the left (3), stair cases at the screens in the front (4), basis patterns between the screens in the middle (5), mosaïking on the white board in the back (6)

NEW CODING TOOLS, NEW ARTIFACTS: ISSUES, CONTROVERSIES, PROBLEMS

New transforms and transform sizes

Although most of the artifacts described above have been depicted and explained when MPEG-2 Video was state of the art or prior, they still appear when coding videos with MPEG-4 Part 2, H.264 and VC-1, albeit sometimes with different causes of appearance and probability as explained below. Concerning blocking artifacts, MPEG-4 Part 2 coded videos expose a similar behavior due to its transform size and function which is equal to the transform used in MPEG-2 Video, i.e. an 8x8 DCT (Richardson, 2003). By contrast, both H.264 and VC-1 support a smaller transform size of 4x4 besides 8x8 (VC-1 also supports 8x4 and 4x8, H.264 allows to switch between 4x4 and 8x8), reducing ringing due to the limited space for over- and undershooting within one transformed block (Wiegand, 2003). Smaller transform sizes also increase the probability of blocking as the number of block borders increases, although the transform size may be chosen adaptively. More information on the type of transform used in H.264 and VC-1 as opposed to the DCT may be found below.

Concerning the increase of the probability of blocking, both H.264 and VC-1 apply an in-loop deblocking filter which adaptively smoothens block borders in order to avoid blocking. The strength and area of application is determined by various parameters, like the type and quantization parameter of the involved blocks. As can be seen in the right half of Figure 1, high quantization parameters with in-loop deblocking in H.264 lead to blurring, effectively eliminating blocking artifacts. Although the filter can be turned off, this is not encouraged due to the increased presence of blocking artifacts for high quantization parameters as shown in the left half of Figure 1.

As opposed to MPEG-2 Video and MPEG-4 Part 2, which use a DCT of 8x8 size, both H.264 and VC-1 use an approximation thereof, allowing for implementations with additions, subtractions and logical (barrel) shifts only, thus improving performance on modern CPUs. Although H.264 and VC-1 use different approximations of the DCT, the approximations themselves are similar to one another as they are based on the same transform matrix and were derived through similar operations (Lee, 2008). Therefore, the subsequent paragraphs describe the integer transform used in H.264 for the sake of illustration.

Although H.264 supports both a 4x4 and an 8x8 integer transform since the amendment of the fidelity range extensions, allowing to switch between the two when using the High profile (International Telecommunication Union, 2010), actually two approximations of the DCT are used – one of size 4x4 and one of size 8x8. For the sake of simplicity and illustration, the subsequent paragraph will focus on the 4x4 integer transform whose basis functions are illustrated in Figure 2 and compared to the corresponding DCT basis functions. As can be seen, the integer transform clearly is an approximation of the DCT, with similar basis functions arising from this relationship. The detailed derivation and approximation process may be found in Malvar (2003). It has to be noted that the encoder/decoder drift described above is avoided due to the use of integer operations and the resulting absence of rounding errors.

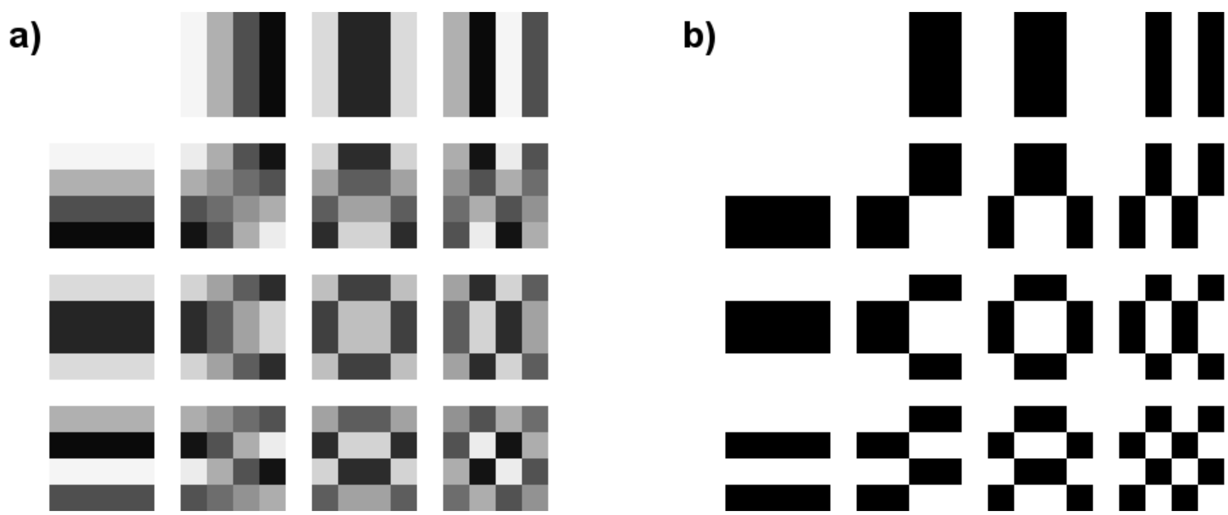

Figure 2: Differences between the basis functions of the DCT and the H.264 integer transform: a) 4x4 DCT basis functions, b) H.264 4x4 integer transform basis functions; both derived through inverse transform of single transform coefficients. Black denotes minimum values, white denotes maximum values; picture domain values are within [-128;127]

Even though the basis functions of the 4x4 DCT and the integer transform are similar, they are not the same, thus yielding different transform coefficients and transform coefficient distributions for a number of input signals (disregarding simple cases like DC only blocks whose DC transform coefficient only differs in magnitude due to scaling). Figure 3 illustrates this using a simple input signal which yields eight transform coefficients when using the DCT, but six when using the integer transform used in H.264. Although the two additional coefficients do not contribute much to the total signal energy and are likely to be quantized to zero, inverse transform of the quantized coefficients will yield different reconstructed signals for both transforms, considering the small loss of signal energy described.

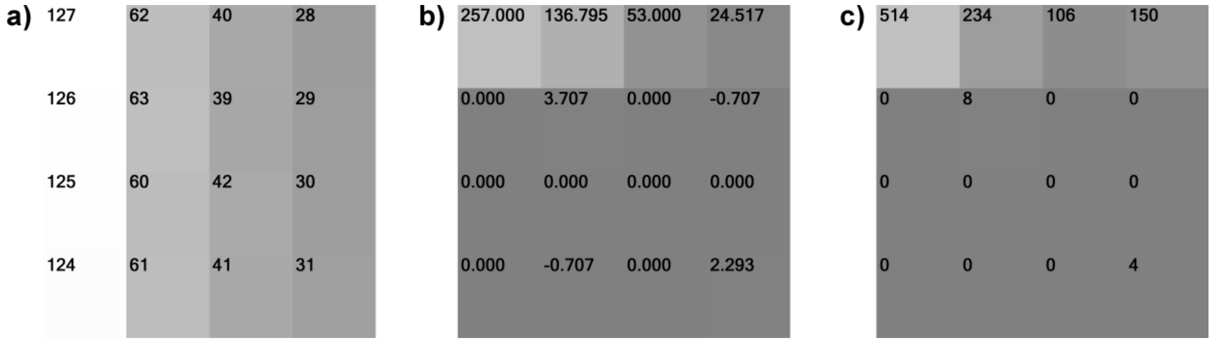

Figure 3: Transform differences between the DCT and the H.264 integer transform: a) Original signal, b) 4x4 DCT transform, c) H.264 4x4 integer transform. Black denotes minimum values, white denotes maximum values; picture domain values are within [-128;127]

Although experiments have been carried out to determine new quantization matrices for H.264 due to the new transform (Richardson, 2003), no research is currently focusing on the impact that the new transform may have on type and quantity of distortions. Due to its historical connection to the DCT, the integer transform used in H.264 shares many of its properties, but also yields different transform coefficient distributions due to its different basis functions, as shown above. Although the speedup through design approximations favoring the capabilities of state-of-the-art CPUs may be convenient, it is necessary to investigate the modified behavior regarding the appearance of artifacts as well as the possible creation of new artifacts which have not been described yet.

The effect of macroblock partitioning

Despite the change in transform size and type, various other new algorithms have been developed for MPEG-4 Part 2, H.264 and VC-1 in order to make coding more efficient or to improve visual quality. Such algorithms are usually referred to as coding tools. One of them is macroblock partitioning in inter-predictive coding, enabling to perform a separate motion search for each part of a macroblock, allowing finding better matches for each part. While all of the aforementioned standards allow to split a 16x16 inter-predicted block into 4 block partitions of size 8x8 each, H.264 additionally supports 16x8 and 8x16 partitions as well as sub-partitions for 8x8 partitions of sizes 8x4, 4x8 and 4x4, respectively (International Telecommunication Union, 2010).

Such partitioning possibilities not only increase the probability of blocking (without the in-loop deblocking filter) but also favor the appearance of an artifact referred to as motion compensation (MC) mismatch (Wu, 2006). MC mismatch describes an effect during motion compensation where the match found during motion estimation does not belong to the object currently being coded, thus appearing misplaced. When using coarse quantization, the difference to the currently coded macroblock cannot be appropriately compensated for, thus yielding additional blocking and blurring which makes the HVS sensible for the “misplaced” object block. While MC mismatch may be a result of the lack of chroma motion estimation (thus trying to find a matching luma block only), it is also possible that it is due to a purely mathematical error measurement like the sum of squared differences (SSD) or the mean squared error (MSE), yielding a motion estimation match with the minimal mathematical difference, but with a distinct perceptual difference as the match does not belong to the object currently being coded.

As described above, an increasing number of partitions and sub-partitions increases the probability for MC mismatches, which become visually more prominent when surrounded by perceptually adequate matches (Wu, 2006). When using low quantization parameters, this problem quasi disappears as the difference through the MC mismatch can be compensated for by predictive coding. Nonetheless, MC mismatches also favor the appearance of mosquito noise due to the borders of mismatching objects found during motion estimation. This noise may also be visible at higher bit rates, i.e. lower quantization parameters.

Besides the number and shape of partitions which both increase the probability of certain artifacts as described above, the number of available coding modes to choose from also yield new artifacts. Most prominently, an artifact named flickering or pumping, also known as stationary area fluctuations, appears when the chosen coding modes of one picture area changes over time, i.e. over subsequent frames. As the predicted residuals from intra and inter prediction differ strongly, the form of the coded residual after quantization is different, yielding different errors and thus flickering due to the change of error over time (Chun, 2006).

Although this artifact is often described as having similarities to mosquito noise, its origins are different. As applying temporal smoothing yields side effects when trying to compensate for this artifact during post processing (Wu, 2006), pumping artifacts can be avoided during the coding process by selecting similar modes for co-located regions in subsequent frames as described in Chun (2006). It has to be noted that a similar selection of partitions and sub-partitions is also helpful in order to achieve this, although not all inter predicted partitions have an equivalent intra predicted partition in terms of size. Furthermore, the prediction signals of inter and intra prediction vary strongly, as do the different intra prediction modes, thus requiring careful adaption of quantization parameters in addition to the coding mode selection in order to reduce pumping artifacts effectively.

Multi-view video coding

Another current field of research is multi-view video coding (MVC), i.e. the coding of multiple views of a scene in order to either produce a three dimensional rendering of said scene or a part of it, albeit often limited to the number of existing views and the interpolated views between them. The most prominent configuration is stereoscopic coding, i.e. the coding of two views – one for the left eye and one for the right – which enables a three dimensional effect when each view is exposed to the corresponding eye. There are currently multiple technologies (like polarized glasses or active shutter glasses) in order to achieve this (May, 2005). In terms of video coding, there are currently three basic approaches for multi-view video coding, which will be shortly described in the subsequent paragraphs, each together with the artifacts it induces or favors.

Depth map quantization artifacts

The first approach constitutes the coding of a two dimensional image or texture and a so-called depth map indicating the distance from the camera for each pixel. This depth map can either be provided in special cases or is otherwise estimated by the encoder when given one or multiple views (Smolic, 2007). Depth estimation is a research topic of great current interest due to the emerging three dimensional TV sets and the associated technologies (Ohm, 2010). The coding of depth maps is explicitly specified in MPEG-4 part 2. Using transform, quantization and residual coding, depth maps are compressed like textures, thus yielding similar artifacts (Richardson, 2003).

Assuming quasi-lossless compression of textures, the quantization of depth maps yields a number of different artifacts which are related to their counterparts in regular image and video coding, although their appearance to the viewer may be different. One example is so-called depth ringing where ringing artifacts emerge from depth map compression, yielding distortions of the depth map and therefore the perceived depth (Boev, 2008). Figure 4 a) depicts the effects of depth ringing, also referred to as depth bleeding. As its image distortion counterpart, depth ringing is most prominent at steep edges (of the depth map), i.e. the region between the ball and the checkerboard background in Figure 4 a). In general, fluctuations in depth may be perceived easily in some scenes, making MSE, PSNR and similar metrics unsuitable for the quality estimation of multi-view videos which rely on depth map quantization.

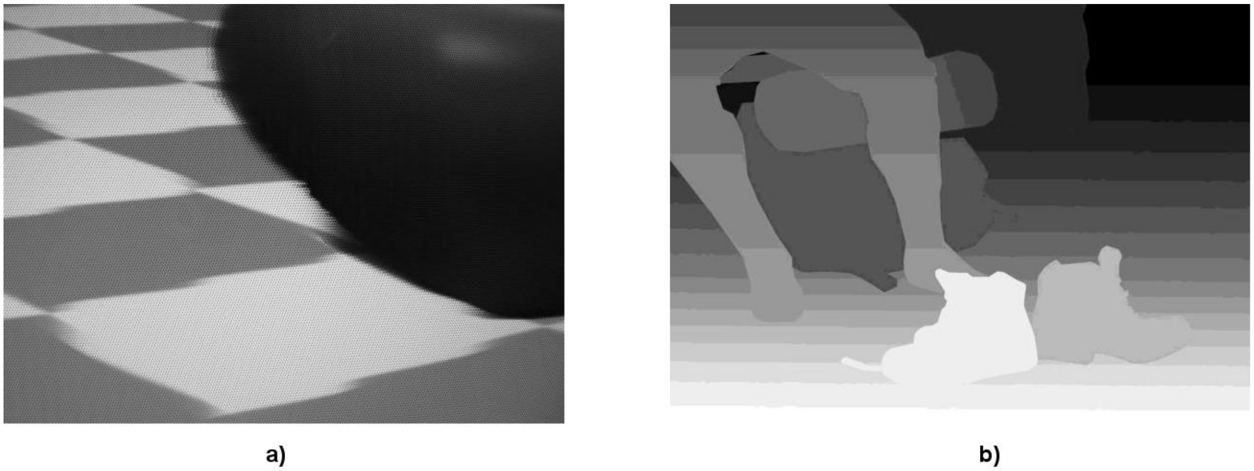

Figure 4: Depth map compression artifacts: a) depth ringing. © 2008, Mobile3DTV (Boev, 2008). Used with permission. b) card board effect. Lighter colors in the depth map indicate greater depth.

Although depth blocking, blurring and similar artifacts may appear when coding depth maps, they have not been described in the literature. Most likely, this is due to the fact that current research efforts are focused on depth estimation and other coding techniques for multi-view video coding. Nonetheless, both depth estimation and harsh quantization may yield an artifact which has been described as card board or puppet theater effect, depicted in Figure 4 b). This refers to a layer-like depth map, similar to the layers of objects in a puppet theater, creating the perception of a number of two dimensional layers instead of smooth depth transitions.

Combining the artifacts of depth map and texture coding, a superposition of them may appear. Depending on the severity of artifacts, they may mask each other, making one so visually prominent that the other one is not visible any more (Wu, 2006). This is also true for all other artifacts described herein, although there is currently no research focused on the human visual perception of jointly appearing artifacts. This may change with the number of emerging coding technologies, currently represented by multi-view and scalable video coding, increasing the number of artifacts and therefore the probability of their joint appearance.

Frame packing artifacts

The second approach allowing for stereoscopic video coding only is an extension specifically available for H.264, called frame packing (Vetro, 2011). It uses a supplemental enhancement information (SEI) message to signal the frame packing of the pictures in the coded video. Frame packing refers to the coding of both views – left and right – in one single view, with both core encoding and decoding algorithms being possibly unaware of the existence of two separate views, thus maintaining compatibility to the H.264 standard as the core coding tools do not need to be changed. Combining and separating the views before and after coding, respectively, must be performed by the encoder (or a preprocessor) and the decoder (or a postprocessor), respectively as the combination of the two views for coding, the insertion of the SEI message and the separation of the two views for display must be performed.

The latest revision of the H.264 standard (International Telecommunication Union, 2010) specifies a number of frame packing arrangement types depicted in Figure 5. Assuming two views of a size of eight times eight macroblocks each, both views are either horizontally or vertically subsampled, depending on the arrangement type, and then rearranged in order to form a picture of eight times eight macroblocks containing both views. However, one arrangement type – frame alternation, depicted in Figure 5 f) – does not require spatial subsampling as each view is represented by all even and odd pictures, respectively, i.e. it is temporally subsampled. If subsampling is used, upsampling after decoding is necessary to restore the original picture of each view, thus introducing similar artifacts as upsampling in scalable video coding described below.

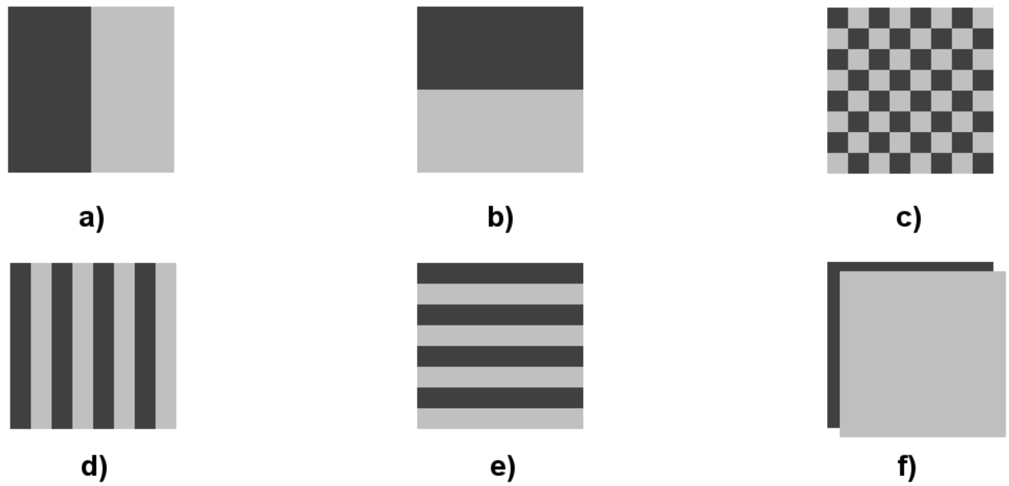

Figure 5: H.264 frame packing arrangement types: a) side-by-side (horizontal), b) top-bottom (vertical), c) checkerboard, d) column alternation, e) row alternation, f) frame alternation. Light and dark gray depict macroblocks of the left and the right view, respectively

In addition to the artifacts caused by subsampling and upsampling – reduced by quincunx sampling of the original samples of both views – crosstalk of artifacts is possible due to the interleaved coding of the two views. Although side-by-side and top-bottom frame packing are only likely to expose these artifacts at the borders between the two views, column and row alternation as well as checkerboard arrangements are expected to introduce crosstalk. As described in Vetro (2011), color bleeding is very likely to propagate across views, mostly when using checkerboards arrangements.

The probability of appearance of mosquito noise, pumping and MC mismatch is also increased due to the interleaving of views, albeit of limited influence to the overall coding performance in terms of PSNR. MC mismatch is favored due to the macroblock size spaced interleaving in most arrangements, causing the motion estimation algorithm to find a match in a macroblock of the other view as the motion estimation search range in most current H.264 encoders is around 16 samples (Jurkiewicz, 2011; Lee, 2008; Richardson, 2003) which equals the size of one macroblock for progressive input.

Using frame alternation arrangements, MC mismatches are even more likely to occur as motion compensation can only be performed on frames which have already been coded, thus disallowing motion compensation in the frame currently being encoded (International Telecommunication Union, 2010). In addition, due to the limited size of the reference lists and practical considerations which limit the number of references used for motion estimation, the number of frames from the other view searched during the motion estimation phase is likely to be greater than the number of frames from the same view that the picture currently being coded belongs to.

Although the artifacts appearing in pictures using frame packing are similar to the artifacts previously described, both subsampling and upsampling have to be considered in terms of the range of artifacts as well as in terms of artifact superposition. Currently, there is only very little research performed in this area, although frame packing is already used intensively by broadcasting and other companies due to its compatibility to H.264 and the frame packing capabilities of the video signal transmission standard HDMI 1.4a.

Due to its increasing popularity in the years 2010 and 2011, more and more H.264 encoders begin to implement support for frame packing (Jurkiewicz, 2011), thus requiring research focused on further coding artifacts due to frame packed coding, including the consideration of the human visual perception of these artifacts when the stereoscopic video is decoded and displayed on a device capable of three-dimensional display using different technologies. As crosstalk and uneven visual quality of the two views are known to influence the perception of depth (Boev, 2008), frame packing requires further investigation regarding these issues.

Artifacts in H.264 MVC

The third approach for coding multiple views is the recent amendment of H.264 named multi-view video coding or MVC for short (Vetro, 2011), extending H.264 in a backwards compatible way. While the first view is coded like a regular H.264 bit stream, all other views use special signaling so that they are ignored by decoders which do not support MVC. In order to improve coding efficiency, all other views may be predicted from the first one or another view of the same access unit, i.e. a view of the same point in time as the view currently being coded (International Telecommunication Union, 2010).

Due to the current widespread use of stereoscopic coding, H.264 MVC defines new Stereo profiles for two views, enabling coding performance similar to the frame packing approach described above, albeit without subsampling and the restrictions imposed thereby (Vetro, 2011). MVC makes use of the similarity of multiple views at any given time instant, referred to as inter-view prediction. As can be seen in Figure 6, the similarity of the left and the right view eases prediction, although predictions from other views may yield a higher amount of MC mismatch artifacts as described above.

Figure 6: Similarity of the left and the right view of a stereoscopic view. Differences are clearly visible in the form of illumination differences in the left part of both images as well as in differences of perspective at the borders on the left and the right

Despite the similarity of both views, there are differences due to the different perspective of the depicted objects, mostly notable at the left and right borders. In addition, the illumination of the two views may be slightly different as shown in the left part of both views in Figure 6. Although illumination compensation, i.e. an algorithm to compensate for luminance differences between the coded view and the view used for prediction, had been part of the working drafts of H.264 MVC, it was not included in the final version of the standard as it would require changes of the low level syntax and the corresponding coding tools. However, notable efforts regarding illumination compensation in terms of the improvement of coding efficiency have been made (Park, 2008) which may be incorporated into future video coding standards (Vetro, 2011).

As the artifacts arising from illumination compensation have not been studied in depth, this leaves room for future research. Other artifacts like crosstalk and different forms of MC mismatches depending on the difference in perspective between the depicted objects in the views used for motion compensation also need a more detailed inspection. Although MPEG-2 Video already included means to encode stereoscopic videos and entailed perceptual considerations of its coding mechanisms (Tseng, 1995), the approach used in MPEG-2 Video and, therefore, its artifacts differ from the approach used in H.264 MVC as the former makes use of the scalability features of MPEG-2, using one view as the base layer and the second one as an enhancement layer, yielding a special form of inter-view prediction based on inter-layer prediction.

Scalable video coding

Scalable video coding (SVC) is available in MPEG-2 Video and MPEG-4 Part 2 and has been introduced as an amendment of H.264 in 2007 (Schwarz, 2007). It generally refers to the ability to decode specified parts of the bit stream, yielding a smaller frame rate, spatial resolution or quality than when decoding the whole bit stream. Scalability relies on layers defining temporal scalability in terms of frame rate, spatial scalability in terms of spatial resolution and quality or SNR scalability in terms of fidelity as well as combinations thereof. In general, the overhead of having multiple layers is small, thus allowing for coding performance similar to a stream containing only the highest frame rate, resolution and quality.

Temporal-scalability-related artifacts

H.264 SVC implements temporal scalability similar to prior standards by using backwards compatible B pictures and special signaling. It makes use of hierarchical prediction structures with B pictures using previously coded B pictures within a group of pictures (GOP) for prediction, each B picture hierarchy being equal to one temporal layer. As only the existing concept of B pictures and special signaling is used, no new artifacts originate besides the ones introduced by B pictures themselves, such as mosquito noise, MC mismatch and others (Wu, 2006). In order to avoid pumping artifacts, (Schwarz, 2007) recommends increasing the quantization parameter in higher temporal layers in a predefined pattern when using hierarchical B pictures to achieve temporal scalability, although this results in PSNR fluctuations inside a GOP.

Spatial-scalability-related artifacts

The implementation of spatial scalability differs between MPEG-2 Video, MPEG-4 Part 2 and H.264 SVC, although the basic concept is similar. Higher spatial layers use the coded information of the spatial layer below by upsampling it and using it for prediction, whereas the lowest spatial layer, also called base layer, is coded regularly, i.e. without the use of scalable video coding tools. The upsampled signal from the lower layer can be based on a reconstructed, i.e. decoded, signal (MPEG-2 Video and MPEG-4 Part 2 as well as inter-layer intra prediction in H.264 SVC) or on transform coefficients of a residual signal (H.264 SVC only, referred to as inter-layer inter prediction).

In H.264 SVC, upsampling also requires the block partitions of the lower layer to be upsampled accordingly, i.e. by a factor of two in each direction when using dyadic differences in resolution between two spatial layers, thereby possibly upsampling blocking artifacts and favoring MC mismatches as motion vectors are scaled accordingly and reference list indices are reused, yielding the same prediction area for motion compensation. In addition, mosquito noise may be upsampled, making it more visible when using higher quantization parameters in the enhancement layer.

The upsampling of transform coefficients in H.264 SVC is performed using a bilinear filter in order to avoid additional signal components due to the in-loop deblocking filter. Inter-layer intra prediction uses a 4-tap FIR filter for the luma samples and a bilinear filter for the chroma samples, albeit based on reconstructed samples instead of transform coefficients. Similar to the required upsampling process in MVC, inter-layer intra prediction upsampling introduces blurring to the prediction signal due to bilinear filtering (Krylov, 2008), thus favoring the appearance of mosquito noise due to the lack of high frequency signal components.

Quality-scalability-related artifacts

Quality scalability is available in a number of different forms throughout the standards mentioned herein. Coarse grain quality scalability in H.264 SVC relies on the same mechanisms as are used for inter-layer inter prediction described above apart from the upsampling operations (the two layers have the same spatial resolution), thus yielding similar artifacts. The difference to be coded is based on the difference between the coarser quantized coefficients in the base layer and the original signal, resulting in finer quantized coefficients in the enhancement layer (Schwarz, 2007). Medium grain scalability uses the same basic concept as coarse grain scalability, whereas fine grain scalability which has already been supported in MPEG-2 Video and MPEG-4 Part 2 allows for more sophisticated prediction structures between base and enhancement layers.

Although enhancement layers in MPEG-4 Part 2 can only be predicted from the corresponding base layers, MPEG-2 Video allows the prediction of base layers from previously coded enhancement layers, thus introducing a drift between encoder and decoder if enhancement layers are discarded. This has been avoided by a special key picture concept in H.264 SVC which is described in detail in International Telecommunication Union (2010) and Schwarz (2007). Although the aforementioned drift differs from the drift of MPEG-2 Video encoders and decoders described above, no research has been conducted so far in order to explore the visual effects of drift-based distortions.

SOLUTIONS AND RECOMMENDATIONS

Encoder-side artifact awareness

As can be seen from the descriptions in the previous section, new coding tools tend to introduce new forms of artifacts or to modify the presence or number of existing artifacts. Therefore, it is important to know the sources of these new artifacts and to develop strategies to avoid or compensate for them. Taking MC mismatch as a prominent example from the previous section, avoiding the resulting artifacts might be a better idea than trying to compensate for them as both detection and compensation may be difficult. This may be the reason why currently neither is described in the literature.

Modifying the encoder may therefore be a better option in order to avoid or at least reduce MC mismatch significantly. Chroma motion estimation, for example, may assist regular luma motion estimation by supplying additional information to consider when calculation the SSD or MSE in order to avoid MC mismatches, albeit more time consuming than luma-only motion estimation. Another option would be to replace the mathematical functions for difference and error measurement by functions which reflect the properties of the HVS better. This would allow avoiding MC mismatches as well as other artifacts up to a certain extent, incorporating properties of the HVS during coding.

Most current video encoders make use of rate distortion optimization (RDO), i.e. testing multiple possible modes and selecting the mode with the smallest cost. The smallest cost is defined as the mode with the best rate-distortion tradeoff, subject to a predefined relation between rate and distortion in order to calculate these costs for all modes. As distortion is mostly measured in a mathematical sense in these calculations as described below, it may be favorable to replace it with distortion measures which are aware of some important properties of the HVS, considering the most common artifacts and their effect on the human visual perception.

Development and application of new quality metrics

Although a number of multiple similar distortion measurements have been discussed in the literature, each with its own strengths and weaknesses, the structural similarity (SSIM) index developed by Wang (2004) has proven to be a distortion measure that correlates well with the human visual perception subject to certain restrictions (Sheikh, 2006). By measuring the structural similarity in terms of variance and covariance of two images on a per-block (8x8) basis, the structural similarity index is capable of considering blurring, ringing, basis patterns and stair cases as well as other forms of image distortions, albeit unable to detect blocking artifacts with blocks of the size of the SSIM block size (8x8).

A first approach to incorporate SSIM indices as distortion measures has been proposed in Mai (2005). Therein, a subset of an H.264 encoder has been modified to use SSIM indices as a distortion measure for RDO during intra mode decision. As may be anticipated from a classical (peak signal to noise ratio, PSNR) point of view, the PSNR performance of the encoded pictures decreases compared to an unmodified encoder, while the perceptual quality increases. If this approach was extended to all RDO-amenable decisions involving new coding tools, the impact of new artifacts or the (re)appearance of classical artifacts might be reduced significantly, favoring coding modes with a reduced presence of artifacts. For motion estimation and similar coding tools, SSIM block matching may not only help to avoid MC mismatches, but also to find structurally more similar blocks than existing approaches, making the arising differences perceptually easier to encode.

The awareness of the perceptual influence of decisions during coding (mode decision, motion estimation etc.) is crucial. Therefore, it is necessary to include facilities into encoders which are aware of the perceptual impact of these decisions, helping to improve the perceived quality by design. As PSNR and other purely mathematical measures of difference give a general hint of the degree of quality degradation, they have proven insufficient when masking effects of the HVS and small or imperceptible differences come into play (Wang, 2004).

Although the computational complexity and ease of comparability of PSNR and the like is convenient for the purposes of state-of-the-art video coding, it is not in terms of the correlation between this metric and a typical human viewer rating video quality. Instead of sacrificing computational power (and therefore time) for new coding tools which improve the PSNR of a given configuration by a small amount, it is conceivable to sacrifice this time to design an HVS-aware quality metric for use within the encoder (perhaps even in form of a new coding tool) in order to improve the overall quality of the encoded pictures, thus also enabling perceptually aware coding control units which can distribute more bit rate to perceptually critical areas of a picture, thereby reducing the number and strength of perceived artifacts at the same bit rate.

Approaches to switch to a different metric for the measurement of differences and errors have also already been proposed by others, e.g. Ohm (2010), although there is currently no concrete direction observable in terms of a concrete metric to choose. SSIM may be an intermediate approach on the way to a new metric, albeit imperfect as it does not cover all important aspects of the human visual perception and does not correlate well with the HVS at low bit rates (Hofbauer, 2010). Despite its incapability to detect certain types of artifacts as discussed above, its high correlation throughout a wide range of bit rates with the HVS would make it a good choice to replace PSNR in the short or medium term, leaving potential for design optimizations in form of a new or different metric in the long term.

FUTURE RESEARCH DIRECTIONS

Next generation video coding

As the future of video coding and its arising artifacts is closely related to the new coding tools designed, the current development of next generation video coding (NVC) based on H.264's coding tools, also referred to as high efficiency video coding (HEVC), gives an insight into the new coding tools and artifacts that will have to be dealt with in the future. At the time of this writing, a preliminary version (1.0) of the future reference software "HM" (HEVC testing model) has been made available to the public (https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/), implementing most of the new coding tools selected for detailed evaluation after their approval in the call for proposals for NVC.

As the number of new coding tools compared to the latest revision of the H.264 standard (International Telecommunication Union, 2010) increased significantly and the release date of the preliminary reference software did not allow for thorough testing at the time of writing of this book chapter, those of the new coding tools which will probably have the strongest impact on artifacts will be discussed, considering that the current version of the reference software is not the final one and the continuing evaluation process may exclude coding tools described herein as well as include new ones.

A major change in terms of video coding is the macroblock size which is now 64x64 luma samples and referred to as a coding unit (CU) with accompanying concepts for prediction units (PU) and transform units (TU), allowing partitioning and sub-partitioning over four hierarchy levels (down to 4x4) as opposed to the two hierarchy levels in H.264 inter prediction (McCann, 2010). Although the number of partitions does not necessarily change the probability of the appearance of artifacts (the smallest size is still the same as in H.264), the introduction of a 16x16 integer transform might lead to a more significant appearance of ringing artifacts compared to the 4x4 and 8x8 transform sizes in H.264 due to the increased number of coefficients and samples available for over- and undershooting as described in the previous section. Transform sizes of 32x32 and 64x64, which are also being evaluated, yet increase the probability of ringing artifacts.

Besides the change in transform size, which also requires thorough inspection as described above for integer transforms in general, the interpolation filter for subsamples in the motion estimation and compensation process may be changed, too. The proposed improvements describe the use of a 6-tap directional filter or a 12-tap DCT based interpolation filter, replacing the Wiener and bilinear filter used in H.264 for subsample interpolation. As both approaches change the signal characteristics of the interpolated subsamples and therefore the likeliness to expose artifacts, future research will have to show how this affects the perceptual quality and artifact propagation.

In addition to the coding tools described, an extension of the number of available intra prediction modes has been proposed and modified (Min, 2010), introducing angular intra prediction in contrast to the nine 4x4 and four 16x16 prediction modes in H.264 making use of a limited number of horizontal, vertical and diagonal prediction. With a total of 17 modes for 4x4 PUs, 34 for 8x8, 16x16 and 32x32 PUs and 5 for 64x64 PUs, requiring the interpolation of predicted samples up to an accuracy of 1/32 of a sample, the new intra prediction modes will increase the probability of pumping artifacts further, apart from increasing the number of modes for RDO and therefore computational complexity significantly.

Analysis of existing artifacts

Despite the fact that future research related to HEVC which will have to wait until the reference software and the HEVC specification are finalized, the evaluation of artifacts arising from the emerging SVC and MVC standards will offer a number of opportunities for artifact research. As described above, there are multiple coding tools whose effects on existing and new artifacts have not yet been examined in depth, thus requiring further inspection and analysis.

Furthermore, the superposition of different artifacts and their effect on the HVS becomes more relevant as the number of known artifacts is already high and yet keeps increasing through the introduction of new video coding standards and amendments thereof. Studying which artifacts are visually prominent when appearing in certain constellations with other artifacts might not only provide a clearer perspective on the perception of artifact superpositions, but also on masking effects originating from the HVS in general.

Artifact-aware encoder design

Overall, the consideration of the human visual perception in video encoder design is an important issue to take into account. If the artifact-related properties of the HVS are already considered in the design process of new coding tools and in the encoder, future research can focus on artifact avoidance instead of compensation. Furthermore, the encoder-side awareness of the presence of artifacts could be used to apply artifact compensation algorithms on both, the encoder and the decoder side, more selectively, reducing the post-processing complexity on the latter side through encoder-generated artifact signaling.

In addition, new metrics can be developed which represent the human visual perception better than existing ones, allowing for improved encoder decisions. Such metrics can also be used for the difference and distortion measurements in general, making artifact detection easier. If, in addition, artifact propagation is analyzed in the encoder, video quality can be estimated more precisely, eventually enabling artifact-aware video coding.

CONCLUSION

Despite the increasing rate of improvement in terms of compression efficiency in modern video coding, the avoidance and compensation of coding artifacts are currently not getting the attention they deserves. The development of new coding tools decreases bit rates compared to previous standards at the cost of increased computational complexity and a lack of awareness of the impact of these new coding tools on the appearance of known or new artifacts. It is important to be aware of the artifacts arising from improvements in video coding algorithms, enabling a broader understanding of the human visual perception besides the classical artifacts, such as blocking, ringing, blurring and the like.

Although it might not be the final solution, the consideration of different quality metrics like SSIM for difference and error measurement as well as RDO can be a step towards the awareness of certain artifacts during the encoding process. Be it in form of a new coding tool as an integral part or as an addition to the core coding tools of a video encoder, the consideration of the human visual perception during the coding process can help to improve the perceptual quality of encoded videos in current and future video coding standards. It may also help to gain a better understanding of the influence of new coding tools regarding their vulnerability to induce artifacts.
In doing so, the need for decoder side artifact detection and compensation would also diminish, thus requiring less attention than currently, allowing future research to concentrate on the development of new metrics for quality measurement on the encoder side rather than sophisticated artifact compensation algorithms on the decoder side. Therefore, it is indispensable to focus future research efforts on artifact avoidance at the encoder side or (even before) in the design process of new coding tools. In the end, it is the casual user, unaware of the mere existence of the most sophisticated coding tools, who judges the visual quality and the visibility of coding artifacts.

REFERENCES

Ben-Ezra, M. & Nayar, S. K. (2004) Motion-Based Motion Deblurring. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(6), 689–698.

Biemond, J., Lagendijk, R. L. & Mersereau, R. M. (1990) Iterative methods for image deblurring. *Proceedings of the IEEE*, 78(5), 856–883.

Boev, A., Hollosi, D. & Gotchev, A. (2008) *Classification of stereoscopic artifacts*. Retrieved February 1, 2011 from http://sp.cs.tut.fi/mobile3dtv/results/tech/D5.1_Mobile3DTV_v1.0.pdf.

Chang, Y.-W. & Chen, Y.-Y. (2005) *Alleviating-Ringing-Artifact Filter Using Voting Scheme*. Paper presented at the ICGST International Journal on Graphics, Vision and Image Processing, Cairo, Egypt.

Chun, S. S., Kim, J. R. & Sull, S. (2006) Intra prediction mode selection for flicker reduction in H.264/AVC. *IEEE Transactions on Consumer Electronics*, 52(4), 1303–1310.

Crop, J., Erwig, A. & Selvaraj, V. (2010) *Ogg Video Coding*. Retrieved September 21, 2011 from http://people.oregonstate.edu/~cropj/uploads/Classes/577finalreport.pdf.

Hofbauer, H. & Uhl, A. (2010) *Visual Quality Indices and Low Quality Images*. Paper presented at the *IEEE 2nd European Workshop on Visual Information Processing*, Paris, France.

International Telecommunication Union (2010) *Recommendation ITU-T H.264 – Advanced video coding for generic audiovisual services (03/2010)*. Geneva, Switzerland: International Telecommunication Union.

Jurkiewicz, A. et al. (2011) *X264 Settings*. Retrieved February 1, 2011 from http://mewiki.project357.com/wiki/X264_Settings.

Kerr, D. A. (2009) *Chrominance Subsampling in Digital Images*. Retrieved February 1, 2011 from http://dougkerr.net/pumpkin/articles/Subsampling.pdf.

Krylov, A. & Nasonov, A. (2008) Adaptive total variation deringing method for image interpolation. *Proceedings of the 15th IEEE International Conference on Image Processing 2008*, 2608–2611.

Lee, J.-B. & Kalva, H. (2008) *The VC-1 and H.264 Video Compression Standards for Broadband Video Services*. New York, New York: Springer Science+Business Media LLC.

Liu, H., Klomp, N. & Heynderickx, I. (2010) *A no-reference metric for perceived ringing*. Paper presented at the Fourth International Workshop on Video Processing and Quality Metrics for Consumer Electronics, Scottsdale, Arizona.

Mai, Z.-Y., Yang, C. L. & Xie, S. L. (2005) Improved best prediction mode(s) selection methods based on structural similarity in H.264 I-frame encoder. *IEEE International Conference on Systems, Man and Cybernetics*, 3, 2673–2678.

Malvar, H. S., Hallapuro, A., Karczewicz, M. & Louis Kerofsky (2003) Low-Complexity Transform and Quantization in H.264/AVC. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(7), 598–603.

May, P. (2005) *A Survey of 3D Display Technologies*. Retrieved February 1, 2011 from http://www.ocuity.co.uk/Ocuity_white_paper_Survey_of_3D_display_technologies.pdf.

McCann, K., Han, W.-J. & Kim, I. K. (2010) *Samsung's Response to the Call for Proposals on Video Compression Technology (JCTVC-A124)*. Retrieved February 1, 2011 from http://wftp3.itu.int/av-arch/jctvc-site/2010_04_A_Dresden/JCTVC-A124.zip.

Min, J.-H., Lee, S., Kim, I.-K., Han, W.-J., Lainema, J. & Ugur, K. (2010) *Unification of the Directional Intra Prediction Methods in TMuC (JCTVC-B100)*. Retrieved February 1, 2011 from http://wftp3.itu.int/av-arch/jctvc-site/2010_07_B_Geneva/JCTVC-B100.zip.

Ng, M. K., Chan, R. H. & Tang, W.-C. (1999) A Fast Algorithm For Deblurring Models With Neumann Boundary Conditions. *Siam J. Sci. Comput.*, 21(3), 851–866.

Ohm, J.-R. (2008) *Recent, Current and Future Developments in Video Coding*. Retrieved February 1, 2011 from http://wiamis2008.itec.uni-klu.ac.at/keynotes/ohm.pdf.

Oosa, K. (1998) *A New Deblocking Filter for Digital Image Compression*. Retrieved February 1, 2011 from http://www.nsc.co.jp/en/tech/report/pdf/7716.pdf.

Park, H. W. & Lee, Y. L. (1999) A Postprocessing Method for Reducing Quantization Effects in Low Bit-Rate Moving Picture Coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 9(1), 161–171.

Park, G. H., Park, M. W., Lim, S. C., Shim, W. S & Lee, Y. L. (2008) Deblocking Filtering for Illumination Compensation in Multiview Video Coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 18(10), 1457–1461.

Ramos, M. G. & Hemami, S. S. (2001) Suprathreshold wavelet coefficient quantization in complex stimuli: psychophysical evaluation and analysis. *Journal of the Optical Society of America A, Optics, Image Science and Vision*, 18(10), 2385–2397.

Richardson, I.E.G. (2003) *H.264 and MPEG-4 Video Compression*. Chichester, England: John Wiley & Sons Ltd.

Schwarz, H., Marpe, D. & Wiegand, T. (2007) Overview of the Scalable Video Coding Extension of the H.264/AVC Standard. IEEE Transactions on Circuits and Systems for Video Technology, 17(9), 1103–1120.

Sheikh, H. R., Sabir, M.F. & Bovik, A. C. (2006) A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms. *IEEE Transactions on Image Processing*, 15(11), 3440–3451.

Shen, M. Y. & Kuo, C.-C. J. (1998) Review of Postprocessing Techniques for Compression Artifact Removal. Journal of Visual Communication and Image Representation, 9(1), 2–14.

Smolic, A., Müller, K., Stefanoski, N., Ostermann, J., Gotchev, A., Akar, G. B., Triantafyllidis, G. & Koz, A. (2007) Coding Algorithms for 3DTV—A Survey. *IEEE Transactions on Circuits and Systems for Video Technology*, 17(11), 1606–1621.

Triantafyllidis, G. A., Tzovaras, D. & Strintzis, M. G. (2002) Blocking Artifact Detection and Reduction in Compressed Data. *IEEE Transactions on Circuits and Systems for Video Technology*, 12(10), 877–890.

Tseng, B. L. & Anastassiou, D. (1995) *Perceptual Adaptive Quantization Of Stereoscopic Video Coding Using MPEG-2's Temporal Scalability Structure*. Paper presented at the International Workshop on Stereoscopic and Three Dimensional Imaging 1995, Santorini, Greece.

Vetro, A., Wiegand, T. & Sullivan, G. J. (2011) Overview of the Stereo and Multiview Video Coding Extensions of the H.264/AVC Standard. *Proceedings of IEEE, Special Issue on "3D Media and Displays"*, 99(4), 626–642.

Wang, Z., Bovik, A. C., Sheikh, H. R. & Simoncelli, E. P. (2004) Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, 13(4), 600–612.

Watson, A., Yang, G. Y., Solomon, J. & Villasenor, J. (1997) Visibility of Wavelet Quantization Noise. *IEEE Transactions on Image Processing*, 6(8), 1164–1175.

Wiegand, T., Sullivan, G. J., Bjøntegaard, G. & Luthra, A. (2003) Overview of the H.264/AVC Video Coding Standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(7), 560–576.

Wu, H. R. & Rao, K. R. (Eds.) (2006) *Digital Video Image Quality and Perceptual Coding*. Boca Raton, FL: CRC/Taylor & Francis.

Xiph.Org Foundation (2011) *Theora Specification*. Retrieved September 21, 2011 from http://theora.org/doc/Theora.pdf.

Zakhor, A. (2002) Iterative procedures for reduction of blocking effects in transform image coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 2(1), 91–95.

ADDITIONAL READING SECTION

Boujut, H., Benois-Pineau, J., Hadar, O., Ahmed, T. & Bonnet, P. (2011) Weighted-MSE based on saliency map for assessing video quality of H.264 video streams. *Proceedings of the SPIE – "Image Quality and System Performance VIII"*, 7867, 78670X–78670X-8.

Brooks, A. C., Zhao, X. & Pappas, T. N. (2008) Structural similarity quality metrics in a coding context: exploring the space of realistic distortions. *IEEE Transactions on Image Processing*, 17(8), 1261–1273.

Carballeira, P., Tech, G., Cabrera, J., Müller, K., Jaureguizar, F., Wiegand, T. & García, N. (2010) *Block based Rate-Distortion analysis for quality improvement of synthesized views*. Paper presented at the 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video 2010, Tampere, Finland.

Farrugia, R. A. & Debono, C. J. (2010) A Hybrid Error Control and Artifact Detection Mechanism for Robust Decoding of H.264/AVC Video Sequences. *IEEE Transactions on Circuits and Systems for Video Technology,* 20(5), 756–762.

Goldmann, L., De Simone, F., Dufaux, F., Ebrahimi, T., Tanner, R. & Lattuada, M. (2010) *Impact of video transcoding artifacts on the subjective quality*. Paper presented at the Second International Workshop on Quality of Multimedia Experience 2010, Trondheim, Norway.

Hewage, C. T. E. R., Worrall, S. T., Dogan, S. & Kondoz, A. M. (2008) Prediction of stereoscopic video quality using objective quality models of 2-D video. *Electronics Letters*, 44(16), 963–965.

Hwang, Y. & Park, J. (2003) *Reduction of compression Artifacts (Blocking Artifacts, Ringing Artifacts) Using POSC*. Retrieved February 1, 2011 from http://homepages.cae.wisc.edu/~ece533/project/f04/hwang_park.pdf.

Hoffmann, H., Itagaki, T., Wood, D., Hinz, T. & Wiegand, T. (2008) A Novel Method for Subjective Picture Quality Assessment and Further Studies of HDTV Formats. *IEEE Transactions on Broadcasting*, 54(1), 1–13.

Kim, W. S., Ortega, A., Lai, P. L., Tian, D. & Gomila, C. (2009) *Depth map distortion analysis for view rendering and depth coding*. Paper presented at the 16th IEEE International Conference on Image Processing 2009, Cairo, Egypt.

Le Meura, O., Ninassib, A., Le Calletc, P. & Barbac, D. (2010) Do video coding impairments disturb the visual attention deployment? *Signal Processing: Image Communication*, 25(8), 597–609.

Lee, J. & Park, H. W. (2011) A New Distortion Measure for Motion Estimation in Motion-Compensated Hybrid Video Coding. *Signal Processing: Image Communication*, 26(2), 75–84.

Liu, T., Wang, Y., Boyce, J. M., Yang, H. &Wu, Z. (2009) A Novel Video Quality Metric for Low Bit-Rate Video Considering Both Coding and Packet-Loss Artifacts. *IEEE Journal of Selected Topics in Signal Processing*, 3(2), 280–293.

Merkle, P., Morvan, Y., Smolic, A., Farin, D., Müller, K., de With P. H. N. & Wiegand, T. (2008) The effects of multiview depth video compression on multiview rendering. *Signal Processing: Image Communication*, 24(1), 73–88.

Shao, L. & Kirenko, I. (2007) Coding Artifact Reduction Based on Local Entropy Analysis. *IEEE Transactions on Consumer Electronics*, 53(2), 691–696.

Shao, L., Zhang, H. & Liu, Y. (2011) A Generalized Coding Artifacts and Noise Removal Algorithm for Digitally Compressed Video Signals. *Lecture Notes in Computer Science*, 6523/2011, 1–9.

Sugimoto, O., Naito, S., Sakazawa, S. & Koike, A. (2009) *Objective perceptual video quality measurement method based on hybrid no reference framework*. Paper presented at the 16th IEEE International Conference on Image Processing 2009, Cairo, Egypt.

Suk, J.-Y., Lee, G. W. & Lee, K.-I. (2005) Effective Blocking Artifact Reduction Using Classification of Block Boundary Area. *Lecture Notes in Computer Science*, 3768/2005, 606–616.

Unterweger, A. & Thoma, H. (2007) *The Influence of Bit Rate Allocation to Scalability Layers on Video Quality in H.264 SVC*. Paper presented at the Picture Coding Symposium 2007, Lisboa, Portugal.

van den Branden Lambrecht, C. J. & Verscheure, O. (1996) Perceptual Quality Measure using a Spatio-Temporal Model of the Human Visual System. *Digital Video Compression: Algorithms and Technologies*, 2668, 450–461.

Vo, D. T. & Nguyen, T. Q. (2009) *Optimal motion compensated spatio-temporal filter for quality enhancement of H.264/AVC compressed video sequences*. Paper presented at the 16th IEEE International Conference on Image Processing 2009, Cairo, Egypt.

Wang, Z., Bovik, A. C. & Lu, L. (2002) Why is Image Quality Assessment So Difficult? Paper presented at the *IEEE International Conference on Acoustics, Speech, & Signal Processing 2002*, Orlando, Florida.

Winkler, S. & Mohandas, P. (2008) The Evolution of Video Quality Measurement: From PSNR to Hybrid Metrics. *IEEE Transactions on Broadcasting*, 54(3), 660–668.

Wu, H. R. & Yuen, M. (1997) A generalized block-edge impairment metric for video coding. *IEEE Signal Processing Letters*, 4(11), 317–320.

Yang, J. & Wu, H. (2010) Robust Filtering Technique for Reduction of Temporal Fluctuation in H.264 Video Sequences. *IEEE Transactions on Circuits and Systems for Video Technology*, 20(3), 458–462.

Zhang, Z. L., Shi, H. & Wan, S. (2009) *A Novel Blind Measurement of Blocking Artifacts for H.264/AVC Video*. Paper presented at the Fifth International Conference on Image and Graphics 2009, Xi'an, China.

KEY TERMS & DEFINITIONS

Artifact	Image distortion induced by side effects of a coding tool and/or quantization
Block	Rectangular unit of image samples grouped for coding
Blocking	Artifact caused by independent coding of neighboring blocks
Blurring	Artifact caused by loss of high frequency components, making the block appear fuzzy
Coding tool	Distinct set of algorithms within a video encoder to improve compression or picture quality
Macroblock	Synonym for a block or a group thereof (depending on the context)
Ringing	Artifact related to Gibbs Phenomenon in Fourier analysis, creating a “halo” consisting of over- and undershooting samples as well as blurring parallel to steep edges due to the insufficient approximation of the original edge by the quantized coefficients in the transform domain

BIOGRAPHY

Andreas Unterweger received his Master's-equivalent Diploma degree in Information Technology and Systems Management (with distinction) from the Salzburg University of Applied Sciences in 2008 and his Master's degree in Computer Science (with distinction) from the University of Salzburg in 2011. He is currently pursuing his Ph.D. degree in Computer Science at the University of Salzburg where he specializes on selective video encryption. In addition, he is an external lecturer at the Salzburg University of Applied Sciences, teaching Microcontroller Programming and Applied Mathematics in the Bachelor's degree program. His current research interests include real-time video coding and selective video encryption.