A LOW-COMPLEXITY METRIC FOR THE ESTIMATION OF PERCEIVED CHROMINANCE SUB-SAMPLING ERRORS IN SCREEN CONTENT IMAGES

Andreas Heindel[1], *Eugen Wige*[2], *Felix Fleckenstein*[1],
Benjamin Prestele[2], *Alexander Gehlert*[2], *and André Kaup*[1]

[1]Friedrich-Alexander University Erlangen-Nürnberg (FAU)
Multimedia Communications and Signal Processing
Cauerstr. 7, 91058 Erlangen, Germany

[2]LogMeIn Inc.
Theaterstr. 6, 01067, Dresden, Germany

ABSTRACT

In web conferencing and remote desktop applications screen sharing is a popular feature. Due to the limited data rate, image and video compression technologies are used to enable this functionality over the internet. However, the commonly used basic profiles of modern video codecs typically utilize the YCbCr color space involving 4:2:0 chrominance sub-sampling. This sub-sampling introduces visually disturbing artifacts for screen content, which may be avoided or reduced if they could be automatically detected. However, these artifacts are shown to be hardly captured by conventional image quality metrics. To address this problem, we developed the Perceived Chrominance Sub-sampling Error (PCSE) metric. Moreover, we created a subjectively evaluated screen content image data set, which is employed for the performance evaluation. The PCSE achieves absolute correlation coefficient values of almost 0.8. Thus, conventional quality metrics are significantly outperformed.

Index Terms— Chrominance sub-sampling, image quality assessment, PCSE, screen content

1. INTRODUCTION

Video coding standards and their encoders have significantly improved during the last decades. However, the basic profiles of common image and video coding standards for lossy compression in consumer applications like JPEG [1], High Efficiency Video Coding (HEVC) [2], or VP9 [3] are targeting the compression of 4:2:0 chrominance sub-sampled input data. Due to the human visual system being much more sensitive to brightness changes than to chrominance variations, the introduced loss and resulting error of the chrominance sub-sampling is usually hardly perceivable in the case of natural image content. The introduced artifacts are almost entirely hidden in the useful signal part. For screen content data, however, the remaining spatial resolution of the chrominance components is often not sufficient and can lead to visible artifacts. Especially the sharpness and color characteristics of fine-lined text and diagrams may be degraded significantly, up to illegibility. Fig. 1 illustrates examples of the occurring artifacts.

Those kinds of errors may be avoided by sticking to full chrominance resolution (4:4:4 YCbCr or RGB) coding, using the advanced profiles of the coding standards where available. However, the requirement of higher profiles drastically limits the amount of supported decoder implementations, as most implementations (especially hardware) only support the basic profiles. Even if software decoders for 4:4:4 YCbCr or RGB data may be available at the receiver side, they should be used carefully to keep the computational load low. Consequently, most of the screen content data may not be compressed using these kind of profiles and the use of chrominance sub-sampled YCbCr data may still be mandatory.

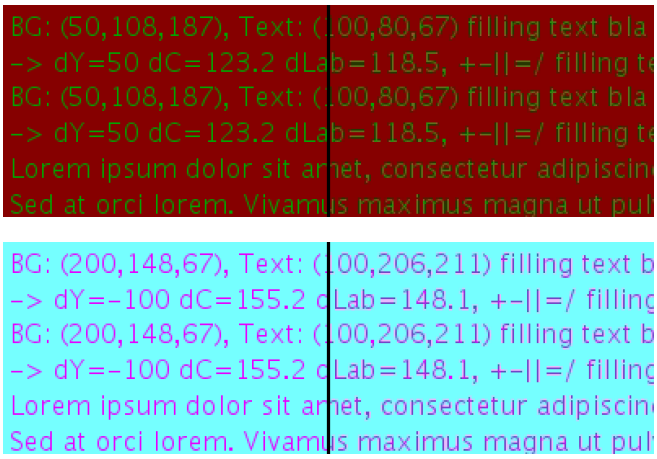

Fig. 1. Two examples from the data set explained in Section 4, split into original parts (left) and 4:2:0 sub-sampled and reconstructed parts (right).

In order to improve the visual quality when employing 4:2:0-based screen content coding, there are basically three options:

1. Pre-processing of the 4:4:4 or 4:2:0 data before encoding to reduce the effect of the chrominance sub-sampling.
2. Special coding, e.g. by employing codecs supporting 4:4:4 YCbCr or RGB data, to avoid the artifacts. Applied, however, to the affected image regions only, in order to limit the computational complexity and the data rate.
3. Post-processing of the 4:2:0 data after decoding for artifact reduction.

Post-processing has the advantage of not requiring any additional information to be transmitted. However, there is the inevitable drawback of additional complexity being introduced on the decoder side, caused by artifact identification and removal. The additional complexity may be reduced by shifting the artifact detection to the encoder side and transmitting the necessary side information. A similar effect is given by the second approach, where special coding of selected regions may lead to a slightly higher decoding complexity and probably a little data rate increase. Eventually, pre-processing is the most elegant solution, as neither additional decoding complexity nor increased data rate is produced. Note that for all three variants,

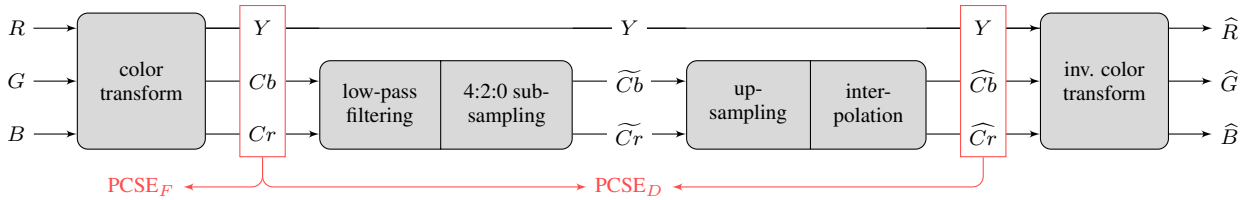

Fig. 2. Color space conversion steps for transmitting RGB data in the 4:2:0 sub-sampled YCbCr format $Y\widetilde{Cb}\widetilde{Cr}$. The signals used to compute the proposed metrics PCSE_F and PCSE_D are marked in red color. The transmission channel, including encoding and decoding, is omitted for simplicity.

pre-processing, special coding, and post-processing with reduced decoding complexity, the artifact detection has to be performed on the encoder side, where the original image data is available. Thus, we propose an objective metric for the detection of the chrominance sub-sampling artifacts before encoding. It may be applied directly on the original image data, being able to predict artifacts even before they occur. Moreover, it can be used as a full reference metric based on the original and the 4:2:0 sub-sampled image data. This measure is considered as a first step towards reducing those kinds of artifacts. Methods for pre- or post-processing are left out of consideration here and should be tackled in future works.

Many objective measures for quality assessment have been proposed for natural images. Besides the conventional PSNR, there are more recent methods like the Structural Similarity (SSIM) index [4], the Visual Information Fidelity (VIF) metric [5], or approaches based on machine learning like in [6]. Specialized metrics for screen content [7, 8, 9] have been published recently, as well. However, they are designed to correlate to more general types of distortions like Gaussian noise, blurring, or coding artifacts, and not to chrominance sub-sampling artifacts. The metrics [7, 8, 9] are all fitted to the subjectively evaluated Screen Image Quality Assessment Database (SIQAD) [10], where chrominance sub-sampling is only present in two of the seven distortion types. In [11] a subjectively evaluated data set called SCD is created and evaluated, which uses two types of distortions: HEVC compression including 4:2:0 chrominance sub-sampling, as well as HEVC compression with the Screen Content Coding (SCC) extension, where no chrominance sub-sampling is applied. There the evaluation of the relationship between the subjective voting and the considered objective metrics focuses on compression artifacts, rather than on distortions due to color space conversion.

In contrast to the previously described metrics, the Perceived Chrominance Sub-sampling Error (PCSE) proposed in this paper is not intended for general quality assessment of screen content. It rather focuses on the detection of chrominance sub-sampling artifacts only, providing an objective measure for this specific error case.

The remainder of this paper is structured as follows. Section 2 introduces the different color representations used and their notations. Afterwards, the proposed error metric is presented in Section 3, followed by the description of the chrominance sub-sampling artifacts data set creation in Section 4. Finally, Section 5 presents the experimental results and Section 6 concludes the paper.

2. COLOR REPRESENTATIONS

A flow chart depicting the different color representations involved when transmitting RGB data in the 4:2:0 sub-sampled YCbCr format is shown in Fig. 2. Processing starts with the image in the RGB color space, represented by the signals R, G, and B. Decorrelation of these color components is achieved by the transformation to the YCbCr 4:4:4 color space, represented by the components Y, Cb, and Cr. The subsequent low-pass filtering and 4:2:0 sub-sampling of the Cb and the Cr planes halve the spatial resolution of these color planes in each dimension. This $Y\widetilde{Cb}\widetilde{Cr}$ signal would be the input signal of a codec for image or video compression supporting 4:2:0 sub-sampled data only. Up-sampling and interpolation of the color planes $\widetilde{Cb}$ and $\widetilde{Cr}$ yield the reconstructed color planes $\widehat{Cb}$ and $\widehat{Cr}$, exhibiting the same spatial resolution as Y again. By the inverse color transformation back to the RGB domain, the degradation introduced by the chrominance sub-sampling is distributed to all components $\widehat{R}$, $\widehat{G}$, and $\widehat{B}$.

3. ESTIMATION OF THE PERCEIVED ERROR

Artifacts in the 4:2:0 sub-sampled image occur because the spatial resolution of the chrominance components is decreased compared to the spatial resolution of the luminance component. In other words, the high frequency components responsible for the sharpness of the chrominance components get lost, with visually disturbing artifacts appearing in areas of high sharpness.

Different features can be used as indicators of the local sharpness for a certain pixel, for example the variance of the local neighborhood, frequency analysis of the local neighborhood, or other linear or non-linear filtering operations. We employ the well-known Sobel operators for computing the gradient in x direction $G_x[x, y]$ and the gradient in y direction $G_y[x, y]$. The pixel-wise sharpness for a certain color channel can then be modeled as the local gradient magnitude, i.e.,

$$S[\boldsymbol{i}] = S[x, y] = \sqrt{G_x[x, y]^2 + G_y[x, y]^2}. \tag{1}$$

The two-dimensional index $\boldsymbol{i} = (x, y)$ abbreviates the spatial position. This sharpness may be computed for each individual color plane Y, Cb, and Cr of an image, resulting in the sharpness values $S_Y[\boldsymbol{i}]$, $S_{Cb}[\boldsymbol{i}]$, and $S_{Cr}[\boldsymbol{i}]$, respectively.

3.1. Pixel-Based Artifacts Forecast Using the 4:4:4 Image

An initial estimate of how noticeable the introduced artifacts will be can be obtained by computing the sharpness of the original 4:4:4 image in the YCbCr domain. More specifically, as the luminance channel is preserved, a higher sharpness in the luminance channel reduces the artifact visibility. On the other hand, a higher sharpness in the chrominance channels generally increases the degree of artifact visibility, as more sharpness is lost. Consequently, the PCSE forecast $\text{PCSE}_F[\boldsymbol{i}]$ is based on the original 4:4:4 $YCbCr$ image only, as depicted in Fig. 2. It can be computed as

$$\text{PCSE}_F[\boldsymbol{i}] = 1 - \frac{S_Y[\boldsymbol{i}]^2}{S_Y[\boldsymbol{i}]^2 + S_{Cb}[\boldsymbol{i}]^2 + S_{Cr}[\boldsymbol{i}]^2}. \tag{2}$$

We selected to use the squared sharpness values, as we found that this choice yields a high correlation to the subjective voting in the data set introduced in Section 4. Moreover, this eliminates the square root in the sharpness computation and is thus computationally efficient. For locations $\boldsymbol{i}$ with no sharpness in the $YCbCr$ image and the denominator being zero, the chrominance sub-sampling does not introduce any distortion and thus $\text{PCSE}_F[\boldsymbol{i}] = 0$ is set for these cases. This results in $0 \leq \text{PCSE}_F[\boldsymbol{i}] \leq 1$.

In general, it can be observed that the $\text{PCSE}_F[\boldsymbol{i}]$ will be equal to zero for the case that no sharpness in the chrominance planes has been detected, i.e., for constant chrominance planes. On the other hand, for the case of a constant luminance channel, the $\text{PCSE}_F[\boldsymbol{i}]$ is completely dominated by the sharpness of the chrominance channels. Both observations confirm the intuitive expectations.

3.2. Pixel-Based Full-Reference Artifact Detection

The quality of the perceived error estimate may be improved when the reconstructed image $Y\widehat{Cb}\widehat{Cr}$ is also available. This is usually the case, or at least it can be easily computed from the $Y\widetilde{Cb}\widetilde{Cr}$ image, when the image is converted on the encoder side. Then, the sharpness of the $Y\widehat{Cb}\widehat{Cr}$ image can be related to the sharpness of the original $YCbCr$ image (cf. Fig. 2), and the detected perceived chrominance sub-sampling error $\text{PCSE}_D[\boldsymbol{i}]$ becomes

$$\text{PCSE}_D[\boldsymbol{i}] = 1 - \frac{S_Y[\boldsymbol{i}]^2 + S_{\widehat{Cb}}[\boldsymbol{i}]^2 + S_{\widehat{Cr}}[\boldsymbol{i}]^2}{S_Y[\boldsymbol{i}]^2 + S_{Cb}[\boldsymbol{i}]^2 + S_{Cr}[\boldsymbol{i}]^2}. \tag{3}$$

Analogously to (2), the special case of no sharpness in the $YCbCr$ image has to be considered, where we set $\text{PCSE}_D[\boldsymbol{i}] = 0$. Moreover, it is now also possible that the sharpness at certain locations $\boldsymbol{i}$ is increased, which would lead to negative values of $\text{PCSE}_D[\boldsymbol{i}]$. This is prevented by setting zero as the lower bound of $\text{PCSE}_D[\boldsymbol{i}]$, also limiting $\text{PCSE}_D[\boldsymbol{i}]$ to the interval $[0, 1]$.

Note that there is no direct relationship between the sharpness in the original chrominance planes $S_{Cb}[\boldsymbol{i}]$ or $S_{Cr}[\boldsymbol{i}]$ and the respective sharpness in the reconstructed planes $S_{\widehat{Cb}}[\boldsymbol{i}]$ or $S_{\widehat{Cr}}[\boldsymbol{i}]$. This is dependent on how sharp edges are aligned with the 2×2 pixel raster used for the sub-sampling.

3.3. From Pixel-Based to Image PCSE

The error metrics introduced in (2) and (3) are both pixel-wise metrics, i.e., the chrominance sub-sampling artifacts are estimated for each single pixel. To model the error at picture level, we propose to convert the $\text{PCSE}_F[\boldsymbol{i}]$ image to a single value $\overline{\text{PCSE}}_F$ by computing

$$\overline{\text{PCSE}}_F = \frac{1}{\|\text{PCSE}_F[\boldsymbol{i}]\|_0} \sum_{\boldsymbol{i}} \text{PCSE}_F[\boldsymbol{i}], \tag{4}$$

where $\|\text{PCSE}_F[\boldsymbol{i}]\|_0$ is the L_0 norm of $\text{PCSE}_F[\boldsymbol{i}]$, representing the number of non-zero elements in $\text{PCSE}_F[\boldsymbol{i}]$. Naturally, we set $\overline{\text{PCSE}}_F = 0$ for the case that $\|\text{PCSE}_F[\boldsymbol{i}]\|_0 = 0$. Analogously, also the $\text{PCSE}_D[\boldsymbol{i}]$ image may be transformed to one single value $\overline{\text{PCSE}}_D$.

4. SUBJECTIVELY EVALUATED DATA SET

For the evaluation of the proposed metrics we created a subjectively evaluated data set. This data set, explicitly designed to exhibit the investigated chrominance sub-sampling artifacts, is necessary for reliably measuring the performance of the proposed PCSE metric. Naturally, the PCSE metric also works for all other kinds of screen content images. Example images with constant background colors and

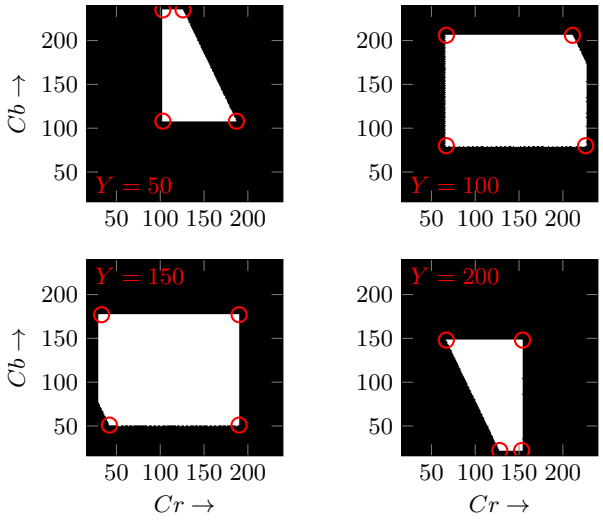

Fig. 3. Visualization of valid chrominance values (white areas) for $Y = \{50, 100, 150, 200\}$. The Cb/Cr combinations used for creating the data set are marked by red circles.

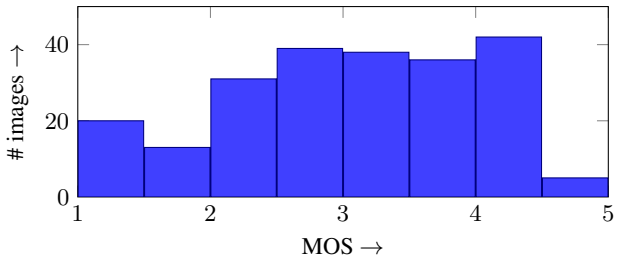

Fig. 4. Histogram of MOS values of the data set, where higher values indicate a higher quality. For the histogram computation, the continuous MOS scale is partitioned into bins of width 0.5 and bin centers at $\{1.25, 1.75., \ldots\}$.

certain text colors were created with Matlab's *insertText(...)* function in the RGB domain. For each image a different color combination was used. They were picked to represent special colors from the YCbCr color space, as described in the following: Six luminance values $Y = \{16, 50, 100, 150, 200, 235\}$ were selected, where 16 and 235 depict black and white, respectively, and the other values yield different color planes as subsets of the complete YCbCr color space. Note that even though the chrominance values Cb and Cr may originate from the interval $[16, 240]$ in general, only a certain subset of Cb/Cr combinations can be transformed to valid RGB colors. Fig. 3 visualizes the valid chrominance values for the selected luminance values $Y = \{50, 100, 150, 200\}$. For each of these four chrominance planes, four Cb/Cr points from the corners (approximated, if not distinct) are selected, leading to 18 different YCbCr colors (including black and white).

Using all possible foreground/background color combinations yields 306 original images in total. As some images suffered from rendering artifacts, we eliminated images from the set that were not labeled as "good" or "excellent" by two experts according to the five-level Absolute Category Rating (ACR) scale as suggested in [12]. The 224 remaining original $YCbCr$ images were finally converted to $Y\widehat{Cb}\widehat{Cr}$ images according to Fig. 2. The sub-sampling was per-

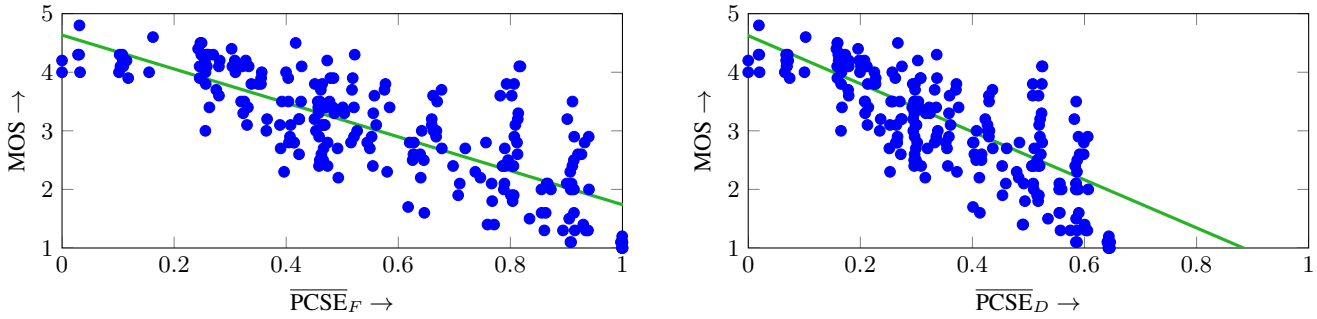

Fig. 5. Scatter plot of the $\overline{\text{PCSE}}_F$ and the $\overline{\text{PCSE}}_D$ against the MOS values with the least squares regression lines in green.

formed by simply averaging the respective four chrominance samples and a zero-order hold filter was used for the interpolation when up-sampling to the original resolution. This procedure is quite common due to its low complexity. We also tested other filters and found that more sophisticated approaches do not deliver a significantly superior reconstruction quality. Example images from the created data set are shown in Fig. 1.

For the subjective tests, the ACR according to [12] was used again. The images were presented in random order. The tests were conducted in a normal laboratory environment [13]. The subjects were instructed to focus on the quality of the sample text and to ignore the fact whether the text color is well distinguishable from the background color or not. In total, 10 subjects (partly experts) participated in the test. Their average scores are denoted by the Mean Opinion Score (MOS). As can be seen by the histogram of the MOS values in Fig. 4, the presented images exhibit a broad variety of quality levels, which is important for a meaningful data set. The data set including the images and the MOS values is publicly available on our website [14].

5. EXPERIMENTAL RESULTS

For visualizing the relationship between the proposed metrics and the subjective scores, Fig. 5 shows scatter plots of the $\overline{\text{PCSE}}_F$ and the $\overline{\text{PCSE}}_D$ against the MOS values, where one data point corresponds to one image of the data set. Also the least squares regression lines are shown. It can be observed that in both cases the data points are well clustered around the regression lines, indicating a rather high correlation between the error models and the subjective evaluation results. However, while the $\overline{\text{PCSE}}_F$ values are spread over the whole possible interval from zero to one, the $\overline{\text{PCSE}}_D$ results seem to saturate and no values higher than 0.7 are observed.

To analyze how well the metrics perform in modeling the subjective scores, Table 1 lists the Pearson Linear Correlation Coefficient (PLCC), the Spearman Rank-Order Correlation Coefficient (SROCC), and the Root Mean Squared Error (RMSE) for the proposed metrics as well as for the other well-established metrics MSE, PSNR, and SSIM. The RMSE values are based on least squares regression, e.g. given by the green lines shown in Fig. 5. The MSE and the SSIM index are computed in the RGB domain for each individual component and are subsequently averaged. The PSNR is derived from this averaged MSE.[1] Note that the PLCC and SROCC values are positive for the quality metrics PSNR and SSIM, while they are negative for the error metrics MSE, $\overline{\text{PCSE}}_F$, and $\overline{\text{PCSE}}_D$. In general, high absolute values are desired for the correlation coefficients PLCC and SROCC, and small values are favored for the RMSE.

[1]Two grayscale images without any degradation have been removed for the PSNR computation, as their PSNR values of infinity make the calculation of the PLCC, the SROCC, and the RMSE impossible.

Table 1. Performance analysis of different metrics.

	MSE	PSNR	SSIM	$\overline{\text{PCSE}}_F$	$\overline{\text{PCSE}}_D$
PLCC	-0.3995	0.4642	0.4512	-0.7937	-0.7943
SROCC	-0.5202	0.5116	0.4533	-0.7852	-0.7857
RMSE	0.8730	0.8422	0.8499	0.5793	0.5786

It can be observed that the proposed artifact forecast $\overline{\text{PCSE}}_F$ and the full-reference detection $\overline{\text{PCSE}}_D$ perform very similar. The $\overline{\text{PCSE}}_F$ is only slightly inferior. Thus, contrary to expectations, the additional information provided by the $Y\widehat{Cb}\widehat{Cr}$ image, which is used for the $\overline{\text{PCSE}}_D$ calculation, hardly improves the quality of the error model. Consequently, there is no real need to compute the $Y\widehat{Cb}\widehat{Cr}$ image on the encoder side, but the $\overline{\text{PCSE}}_F$ may be used directly. The other evaluated metrics are clearly outperformed by these two, such that they can be considered as not being well-suited for detecting chrominance sub-sampling errors.

Alternative filter kernels for the gradient computation as well as different weightings of the luminance and the chrominance components in (2) and (3) have been tested. However, this lead to no significant changes in the performances of the proposed metrics. Consequently, both proposed error models are considered to be rather stable.

6. CONCLUSION

The paper proposes two metrics for measuring the perceived artifact level of chrominance sub-sampling errors in screen content images. While $\overline{\text{PCSE}}_F$ is based on the original $YCbCr$ image only, $\overline{\text{PCSE}}_D$ works as a full-reference metric based on both the $YCbCr$ and the $Y\widehat{Cb}\widehat{Cr}$ images. The evaluation shows that both are highly correlated to the subjective quality ratings in the gathered data set, yielding absolute correlation coefficients values of almost 0.8. Other well-known quality metrics are significantly outperformed.

The proposed metrics allow for a reliable detection of image regions affected by chrominance sub-sampling artifacts. Consequently, specific coding or pre- and post-processing techniques may be applied to reduce the artifacts on the output display.

7. REFERENCES

[1] ITU-T (formerly CCITT) and ISO/IEC JTC1, "Digital compression and coding of continuous-tone still images," ISO/IEC 10918-1 - ITU-T Recommendation T.81 (JPEG), Sep. 1992.

[2] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.

[3] D. Mukherjee, J. Bankoski, A. Grange, J. Han, J. Koleszar, P. Wilkins, Y. Xu, and R. Bultje, "The latest open-source video codec VP9 – an overview and preliminary results," in *Proc. Picure Coding Symposium (PCS)*, San José, CA, USA, Dec. 2013, pp. 390–393.

[4] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[5] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.

[6] S. Bosse, D. Maniry, K.-R. Müller, T. Wiegand, and W. Samek, "Neural network-based full-reference image quality assessment," in *Proc. Picure Coding Symposium (PCS)*, Nürnberg, Germany, Dec. 2016.

[7] H. Yang, Y. Fang, and W. Lin, "Perceptual quality assessment of screen content images," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4408–4421, Nov. 2015.

[8] S. Wang, K. Gu, K. Zeng, Z. Wang, and W. Lin, "Perceptual screen content image quality assessment and compression," in *Proc. IEEE International Conference on Image Processing (ICIP)*, Sep. 2015, pp. 1434–1438.

[9] Z. Ni, L. Ma, H. Zeng, C. Cai, and K. K. Ma, "Screen content image quality assessment using edge model," in *Proc. IEEE International Conference on Image Processing (ICIP)*, Sep. 2016, pp. 81–85.

[10] H. Yang, Y. Fang, W. Lin, and Z. Wang, "Subjective quality assessment of screen content images," in *Sixth International Workshop on Quality of Multimedia Experience (QoMEX)*, Sep. 2014, pp. 257–262.

[11] S. Shi, X. Zhang, S. Wang, R. Xiong, and S. Ma, "Study on subjective quality assessment of screen content images," in *Proc. Picure Coding Symposium (PCS)*, May 2015, pp. 75–79.

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

[13] "ITU-R Recommendation BT.500-13: Methodology for the subjective assessment of the quality of television pictures," Jan. 2012.

[14] A. Heindel, E. Wige, F. Fleckenstein, B. Prestele, A. Gehlert, and A. Kaup. Data set for the evaluation of the perceived chrominance sub-sampling error (PCSE) metric. [Online]. Available: http://lms.lnt.de/files/peoples/extra/pcse_data_set.zip