# Image Quality of 4:2:2 and 4:2:0 Chroma Subsampling Formats

**Emil Dumic, Mario Mustra, Sonja Grgic, Goran Gvozden**
University of Zagreb, Faculty of Electrical Engineering and Computing,
Department of Wireless Communications
Unska 3/XII, HR-10000 Zagreb, Croatia
E-mail: *emil.dumic@fer.hr*

**Abstract** – *There is a large number of different video compression formats being used for various purposes today. These formats differ between each other in possibility of achieving suitable bitrates while maintaining acceptable image quality. Current video compression standards offers the possibility of different sampling rate for the luminant and the chrominant components of video signal. Because HVS is much more sensitive to changes in luminance than changes in chrominance of an image, compression can be achieved by subsampling the chroma components. This paper presents the measurement of the influence of two different chroma subsampling formats (4:2:2 and 4:2:0) on image quality after MPEG-2 compression. The obtained results show that 4:2:0 subsampling format gives slightly better results for bitrates up to 15 Mbit/s.*

**Keywords –** *RGB, YUV, MPEG-2, chroma subsampling*

## 1. INTRODUCTION

There are many different color models used in digital video and imaging. The *RGB* color model is an additive color model in which red, green, and blue lights (primary colors) are added together in various ways to reproduce a broad array of colors [1]. Each color can be specified by its spectrum or by its tristimulus values. This color property is known as a metameric match. Metamerism defines the possibility that two spectrally different stimuli have the same tristimulus values. This means that the combination of two lights results in a certain color that has different spectral characteristics than one obtained with combining two or more lights. Digital cameras capture images by producing values proportional to radiance that approximate red, green, and blue (*RGB*) tristimulus values. In most imaging systems, *RGB* tristimulus values are subject to a nonlinear transfer function, gamma correction, which mimics the perceptual response. For that purpose is used the *R'G'B'* notation which denotes the nonlinearity. *RGB* is not an absolute color space. It is a device-dependent color space, which means that different devices detect or reproduce a given *RGB* value differently. Therefore an *RGB* value does not define the same color across devices from different manufacturers or production lines without some kind of color management. The result of color mixing in the RGB is specified as relative to the primary colors. When the exact chromaticities of the red, green, and blue primaries are defined, the color model becomes an absolute color space, such as *sRGB*. Fig. 1 shows *sRGB* triangle in *xyz* space. It can be seen that *sRGB* model cannot describe all colors that we see, but only those inside triangle.

*YUV* presents another common color space. It encodes a color image or video taking human perception into account. Human visual system (HVS) has considerably less ability to sense detail in color information than in lightness [2]. In this HVS imperfection lays a good possibility for information reduction keeping the image quality acceptable.

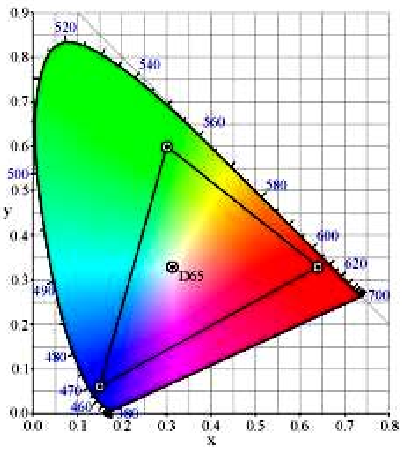
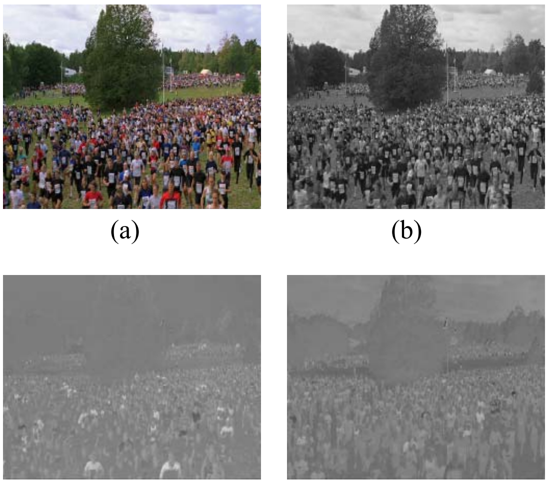



**Fig. 1.** s*RGB* color triangle in *xyz* space.

Color detail can be reduced by subsampling, which is a form of filtering (or averaging). Commonly used terms like *Y'UV*, *YUV*, $YC_bC_r$, $YP_bP_r$ (analog version of $YC_bC_r$), etc., are sometimes ambiguous and overlapping. *YUV* and *Y'UV* are used for a specific analog encoding of color information in TV systems, where *Y'* stands for luma and it is formed as a suitably weighted sum of *R'G'B'*. $YC_bC_r$ is used for digital encoding of color information suited for video and still-image compression and transmission such as MPEG and JPEG. $Y'C_bC_r$ is a family of color spaces used in video and digital photography systems. Component video systems convey image data as a luma component *Y'*, approximating lightness, and two color difference components, $C_b$ and $C_r$ in the digital domain or $P_b$ and $P_r$ in analog that represents color disregarding lightness. $Y'C_bC_r$ is not an absolute color space, it is only a way of encoding *RGB* information. The actual color displayed depends on the actual *RGB* colorants used to display the signal [1]. Fig. 2 shows *YUV* signal and corresponding signal components.



(a) (b) (c) (d)

**Fig. 2.** *YUV* components of video signal: (a) *YUV*, (b) *Y*, (c) *U* and (d) *V* component.

MPEG-2 standard allows different chroma subsampling formats. This paper presents the comparison in image quality between 4:2:0 and 4:2:2 formats. By storing less chroma information, compression ratio will definitely be higher as presented in [3]. There is no absolute measure which can define the actual quality of digital video signal and because of that we use combination of PSNR, SSIM and VQM measures. This paper presents influence on the image quality of two different chroma subsampling formats in MPEG-2 standard using three different quality measures.

The rest of the paper is organized as follows. In Section 2 color model transform from *RGB* to $YC_bC_r$ and chroma subsampling are described. In Section 3 the video quality measures used in experiments are described. Section 4 presents the experimental results. Section 5 draws the conclusion.

## 2. COLOR TRANSFORM AND CHROMA SUBSAMPLING

$Y'C_bC_r$ is the digital version of $YP_bP_r$ which is created from the corresponding gamma-adjusted *RGB* (red, green and blue) source using two defined constants $K_b$ and $K_r$:

$$\begin{aligned} Y' &= K_r \cdot R' + (1 - K_r - K_b) \cdot G' + K_b \cdot B', \in [0,1] \\ P_b &= 0.5 \cdot \frac{B' - Y'}{1 - K_b}, \in [-0.5, 0.5] \\ P_r &= 0.5 \cdot \frac{R' - Y'}{1 - K_r}, \in [-0.5, 0.5], \end{aligned} \tag{1}$$

where $K_b$ and $K_r$ are ordinarily derived from the definition of the corresponding *RGB* space. The form of $Y'C_bC_r$ that was defined for standard-definition television (SDTV) used in the ITU-R BT.601 (formerly CCIR 601) standard for use with digital component video is obtained from the corresponding *RGB* space as follows:

$$\begin{aligned} K_b &= 0.114 \\ K_r &= 0.299 \end{aligned} \tag{2}$$

The primaries of modern displays for high-definition television (HDTV) are standardized in Rec. ITU-R BT.709. Weights computed from these primaries are appropriate to compute relative luminance from red, green, and blue tristimulus values for computer graphics, modern video cameras and modern CRT displays in both STDV and HDTV [1]. For ITU-R BT.709 the constants are:

$$\begin{aligned} K_b &= 0.0722 \\ K_r &= 0.2126 \end{aligned} \tag{3}$$

In component video color information is transmitted separately for all three components. Luma detail is being maintained while color information is reduced, again using the relatively poor color acuity of vision property. First, luma and chroma are formed from *RGB* signal. Afterwards, subsampling (filtering) reduces detail in chroma components. In Fig. 3, the left-hand column shows a 2×2 array of *R'G'B'* pixels denoted 4:4:4 *R'G'B'*. With 8 bits per sample, this array occupies a total of 12 bytes. Each *R'G'B'* triplet (pixel) can be transformed into $Y'C_bC_r$, this is denoted 4:4:4 $Y'C_bC_r$. $Y'C_bC_r$ studio digital video according to Rec. 601 uses 4:2:2 sampling which means that $C_b$ and $C_r$ components are each subsampled by a factor of 2 horizontally. In this way the 12 bytes of *R'G'B'* are reduced to 8, resulting in 1.5:1 lossy compression. Some digital video systems use 4:1:1 sampling, where $C_b$ and $C_r$ components are each subsampled by a factor of 4 horizontally, and cosited with every fourth luma sample. The 12 bytes of *R'G'B'* are reduced to 6, resulting in 2:1 compression. In 4:2:0 scheme $C_b$ and $C_r$ are each subsampled by a factor of 2 horizontally and a factor of 2 vertically. $C_b$ and $C_r$ are effectively centered vertically halfway between image rows. There are two variants of 4:2:0, having different horizontal placing as shown in Fig. 3 [1].

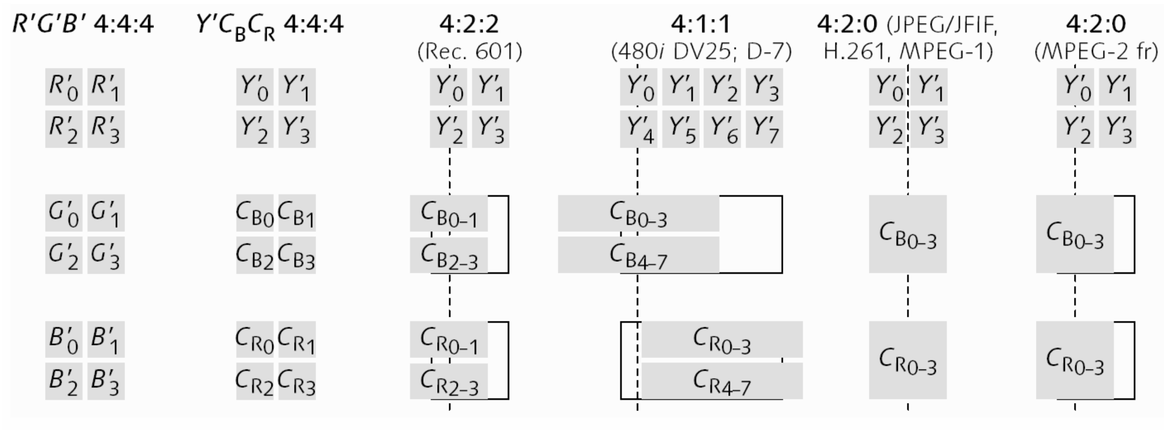



**Fig. 3.** Chroma subsampling for different subsampling schemes.

## 3. OBJECTIVE QUALITY MEASURES

To be able to compare original and compressed sequences, we used 3 objective image quality measures:

- *PSNR* (Peak Signal to Noise Ratio) [5];
- *SSIM* (Structural Similarity Index) [6];
- *VQM* (Video Quality Measure) [7].

*PSNR* is the ratio between the maximum possible signal power and the noise power. It is usually expressed in terms of the logarithmic decibels. In (5) $a_{i,j}$ and $b_{i,j}$ are pixels from original and compressed image, $x$ and $y$ describe height and width of an image and *MSE* stands for Mean Square Error.

$$PSNR = 10\log_{10}\frac{255^2}{MSE}$$
$$MSE = \frac{\sum_i \sum_j (a_{i,j} - b_{i,j})^2}{x \cdot y} \quad (5)$$

The Structural Similarity (*SSIM*) is a novel method for measuring the similarity between two images [6]. It is computed from 3 image measurement comparisons: luminance, contrast and structure. Each of these measures is calculated over the 8×8 local square window moved pixel-by-pixel over the entire image. At each step, the local statistics and *SSIM* index are calculated within the local window. Because resulting *SSIM* index map often exhibits undesirable "blocking" artifacts, each window is filtered with Gaussian weighting function (11×11 pixels). In practice, one usually requires a single overall quality measure of the entire image, so Mean *SSIM* (*MSSIM*) index is computed to evaluate the overall image quality. The *SSIM* represents a quality measure of one of the images being compared, while the other image is regarded as being perfect. *SSIM* gives results between 0 and 1, where 1 means excellent quality and 0 means poor quality.

Fig. 4 shows an overview of the flowchart of *VQM* [7].

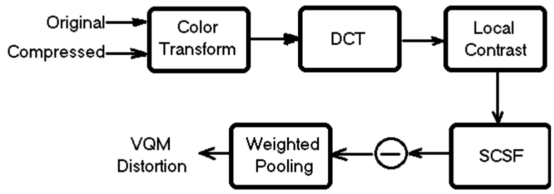



**Fig. 4.** VQM measuring algorithm.

The first step is color transform. Both MPEG and H.263 use the *YUV* color space, so the raw data can be used directly. After that original and compressed image are transformed using DCT transform. This step separates incoming images into different spatial frequency components. Third step is converting each DCT coefficient to local contrast (LC). After this step, most values lie inside [-1, 1]. Fourth step converts LC to just-noticeable differences (jnds). The DCT coefficients are converted to just-noticeable differences by multiplying each DCT coefficient with its corresponding entry in the SCSF (Spatial Contrast Sensitivity Function) matrix. For the static SCSF matrix, MPEG default quantization matrix is used. For the dynamic matrix each entry in the static SCSF matrix is raised to the power decided by the frame rate of video sequences. The final step is weighted pooling of the mean and the maximum distortion. Here the two sequences are subtracted. At this step *VQM* also incorporates contrast masking into a simple maximum operation and then weights it with the pooling mean distortion. This reflects the facts that a large distortion in one region will suppress sensitivity to other small distortion, because weighted maximum distortion into pooled distortion is much better than pooled distortion alone [7]. VQM measure can obtain results between few grades and zero, where values which are near zero mean almost identical tested and original video sequence.

## 4. RESULTS

Test sequences used in this comparison were:

- Crowdrun sequence;
- Duckstakeoff sequence;
- Intotree sequence;
- Parkjoy sequence;

which source material can be downloaded from [8].

Fig. 5 (a)-(d) shows the used Crowdrun, Duckstakeoff, Intotree and Parkjoy test sequences respectively.

(a)

(b)

(c)

(d)

**Fig. 5.** (a) Crowdrun sequence; (b) Duckstakeoff sequence; (c) Intotree sequence; (d) Parkjoy sequence.

Sequences in resolution 576i were first converted from .sgi to different .yuv formats (4:4:4, 4:2:2 and 4:2:0) using sgi2yuv program, [9]. Afterwards these uncompressed .yuv sequences were converted to 24 bit RGB format and then compared using PSNR, SSIM and VQM quality measures. Results are shown in Table 1. It can be seen that although PSNR gives 1.5-3 dB better results for the 4:2:2 format than for the 4:2:0 format, SSIM and VQM measures give very good results for both subsampling methods. It can be concluded that color subsampling does not affect the quality of tested video sequences too much.

**Table 1.** Comparison of 4:4:4 with 4:2:2 and 4:4:4 with 4:2:0 sequences; (a) Crowdrun, (b) Duckstakeoff, (c) Intotree, (d) Parkjoy.

| **Crowdrun** | **PSNR** | **SSIM** | **VQM** |
|---|---|---|---|
| 4:4:4 – 4:2:2 | 47.4289 | 0.9956 | 0.3562 |
| 4:4:4 – 4:2:0 | 46.7371 | 0.9947 | 0.3976 |

(a)

| **Duckstakeoff** | **PSNR** | **SSIM** | **VQM** |
|---|---|---|---|
| 4:4:4 – 4:2:2 | 49.0089 | 0.9973 | 0.2868 |
| 4:4:4 – 4:2:0 | 46.4407 | 0.9947 | 0.3752 |

(b)

| **Intotree** | **PSNR** | **SSIM** | **VQM** |
|---|---|---|---|
| 4:4:4 – 4:2:2 | 50.9208 | 0.9964 | 0.2610 |
| 4:4:4 – 4:2:0 | 49.9949 | 0.9955 | 0.2956 |

(c)

| **Parkjoy** | **PSNR** | **SSIM** | **VQM** |
|---|---|---|---|
| 4:4:4 – 4:2:2 | 47.8961 | 0.9958 | 0.3402 |
| 4:4:4 – 4:2:0 | 46.6294 | 0.9943 | 0.4022 |

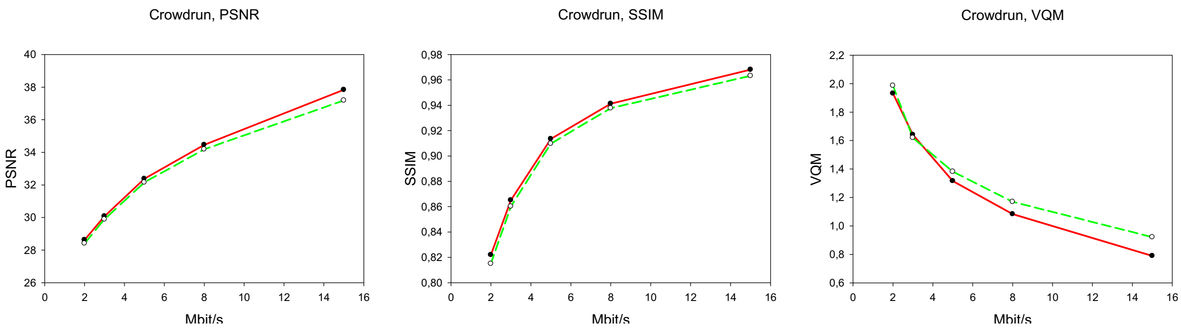
(d)

In the second comparison, we compressed 24 bit RGB format (from 4:2:0 and 4:2:2 .yuv sequences) to .mpg program stream (PS) [10].

Basic settings for MPEG-2 Encoder are:

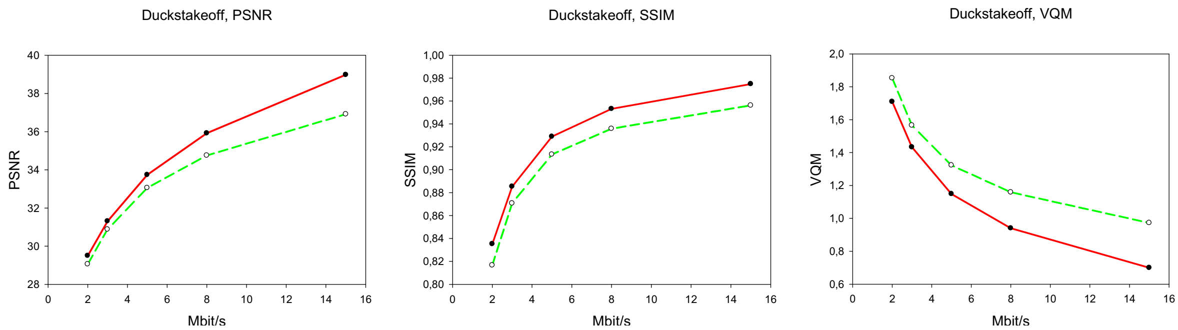
- MP@ML for 4:2:0 sequences; 4:2:2P@ML for 4:2:2 sequences,
- maximum bitrate: 15 Mbit/s,
- Group of pictures (GOP): 12 frames (IBBPBBPBBPBB),
- MPEG Program Stream,
- variable bitrate, average bitrate values: 2, 3, 5, 8, 15 Mbit/s.

Original RGB and compressed sequences were compared using PSNR, SSIM and VQM quality measures. Figs. 6-9 (a)-(c) show different objective quality measures in relation to bit rates, for 4:2:0 and 4:2:2 formats.



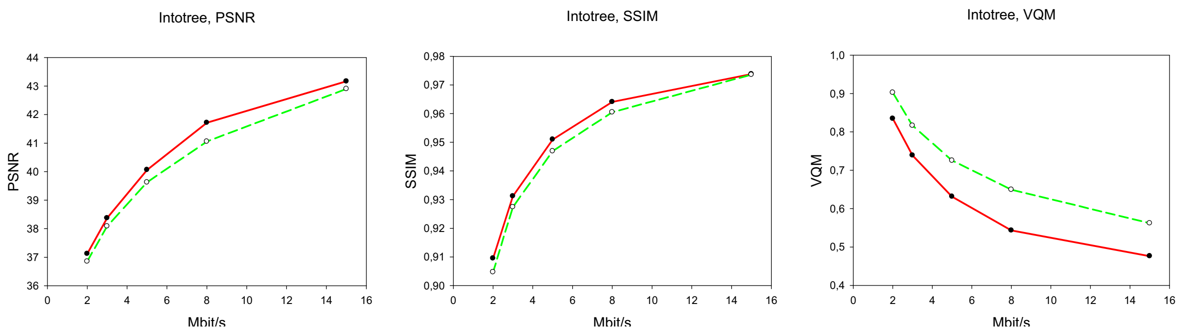

(a) (b) (c)

**Fig. 6.** Different video quality measures relative to bitrate; Crowdrun sequence; solid – 4:2:0 subsampling, dashed – 4:2:2 subsampling: (a) PSNR measure, (b) SSIM measure, (c) VQM measure.



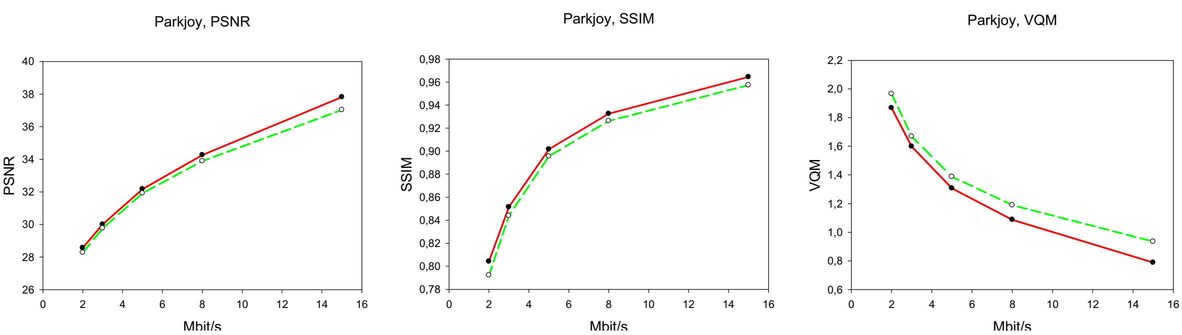

(a) (b) (c)

**Fig. 7.** Different video quality measures relative to bitrate; Duckstakeoff sequence; solid – 4:2:0 subsampling, dashed – 4:2:2 subsampling: (a) PSNR measure, (b) SSIM measure, (c) VQM measure.




(a) (b) (c)

**Fig. 8.** Different video quality measures relative to bitrate; Intotree sequence; solid – 4:2:0 subsampling, dashed – 4:2:2 subsampling: (a) PSNR measure, (b) SSIM measure, (c) VQM measure.




(a) (b) (c)

**Fig. 9.** Different video quality measures relative to bitrate; Parkjoy sequence; solid – 4:2:0 subsampling, dashed – 4:2:2 subsampling: (a) PSNR measure, (b) SSIM measure, (c) VQM measure.

## 5. CONCLUSION

In this paper we have compared MPEG-2 compressed 4:2:0 and 4:2:2 subsampling formats using different video quality measures. From the results it can be seen that objective measures give higher grades for 4:2:0 subsampling/main profile than for 4:2:2 subsampling/4:2:2 profile for bitrate values up to 15 Mbit/s, for all tested video sequences. Also, when comparing 4:2:0 with 4:2:2 uncompressed format, 4:2:0 subsampling does not affect the quality of tested video sequences too much. VQM and SSIM give very good results for both subsampling formats, although PSNR can be 1-2.5 dBs lower for 4:2:0 subsampling format. It can generally be concluded that for resolution 576i and MPEG-2 compression it is better to use 4:2:0 subsampling format for bitrates up to 15 Mbit/s and usual GOP structure. In the future, tests can be extended on H.264/AVC compression also, in 4:2:0 and 4:2:2 subsampling formats.

## ACKNOWLEDGEMENT

The work described in this paper was conducted under the research projects: "Picture Quality Management in Digital Video Broadcasting" (036-0361630-1635) and "Intelligent Image Features Extraction in Knowledge Discovery Systems" (036-0982560-1643), supported by the Ministry of Science, Education and Sports of the Republic of Croatia.

## REFERENCES

[1] C. Poynton, "Digital Video and HDTV Algorithms and Interfaces", Morgan Kaufmann Publishers, San Francisco, 2003

[2] D. H. Hubel, "Eye, Brain and Vision", Scientific American Library, New York, 1988

[3] K. Jongmyon, D.S. Wills, "Evaluating a 16-bit YCbCr (6:5:5) color representation for low memory, embedded video processing", International Conference on consumer Electronics, pp. 181-182, January 2005

[4] S. Winkler, P. Mohandas, "The Evolution of Video Quality Measurement: From PSNR to Hybrid Metrics", *IEEE Trans. on Broadcasting*, Vol. 54, No. 3, pp. 660-668, September 2008

[5] S. Grgic, M. Grgic, M. Mrak, "Reliability of Objective Picture Quality Measures", *Journal of Elect. Engineering*, Vol. 55, No. 1-2, pp. 3-10, January 2004

[6] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity", *IEEE Trans. on Image Proc.*, Vol. 13, No. 4, pp. 600-612, April 2004

[7] F. Xiao, "DCT-based Video Quality Evaluation", Final Project EE392J, winter 2000

[8] source material available at: ftp://vqeg.its.bldrdoc.gov/HDTV/SVT_MultiFormat/

[9] sgi2yuv program available at: http://www.ldv.ei.tum.de/Members/tobias/videotools/sgi2yuv.zip

[10] ISO/IEC 13818-1:2007 - Information technology - Generic coding of moving pictures and associated audio information: Systems (also called MPEG-2)