Opponent Color, Human Vision and Wavelets for Image Compression

Marcus J. Nadenau and Julien Reichel
Signal Processing Laboratory, Swiss Federal Institute of Technology
Lausanne, Switzerland

Abstract

Modern image compression codecs produce embedded bitstreams with optimal rate-distortion properties. For natural imagery the distortion measure should by extended by a model of human color vision. The choice of the color space becomes here crucial since it has to be optimal in terms of compressibility and in terms of its visual distortion prediction properties. For three alternative color spaces, incorporated into a wavelet-codec (JPEG 2000), the quality-equivalent compression ratios are determined by subjective testing. For each color space the contrast sensitivity is measured and incorporated. Finally objectively measurable signal properties are aligned to the results to understand better the observed performance.

Introduction

In this paper the impact of the color space on visual compression performance is analyzed. In difference to similar analysis for the DCT based JPEG [1],[2], the here presented work focuses on a wavelet-based compression codec like the future still image compression standard JPEG 2000 [3]. Additionally, a stronger importance is laid on the combination of compression with a model of human low-level vision. Since this affects significantly the visual compression quality and interferes with the chosen color space, it is an aspect that cannot be treated separately. Therefore the analysis is limited to color spaces that describe the color by a luminance and two chrominance channels, in analogy to the human vision. Color spaces like RGB or CMYK are not considered.

Different to the common incorporation of the contrast sensitivity by weighting factors, a new CSF filtering technique [4] is applied. This allows a more accurate incorporation of the CSF.

The performed contrast sensitivity measurements were different than for instance in [5], because closer linked to the compression application. The color patterns were varied along the principal axis of each color space.

The compression performance of the different color spaces is compared by a new method of subjective quality evaluation to obtain a realistic and accurate performance ranking.

In the first section the paper presents the used compression scheme. In the following section the three compared color spaces are introduced. A large portion is assigned to the psycho-visual measurements of the contrast sensitivity functions (CSF) and the subjective image compression quality ranking in sections 3 and 4. In the last section some objective signal characteristics are analyzed for an better understanding of the observed performance.

Compression Scheme

Today, the most promising performance in compression quality measured in signal-to-noise-ratio is shown by wavelet-based codecs. An additional advantage is their dyadic structure that allows a good incorporation of a human vision model exploiting contrast sensitivity and contrast masking. In the large variety of wavelet-codecs the status-quo of the future standard for still image compression (JPEG 2000) is chosen [3].

The encoder generates an embedded bitstream up to the lossless encoded image. The bitstream is consequently ordered by a rate-distortion scheme [6]. It guarantees the highest rate-distortion quotient $\Delta D/\Delta R$ at any bitrate R. By default ΔD is measured as MSE (mean-square error) of the quantized wavelet coefficients. To achieve optimal visual quality this metric needs to be modified. This is achieved by the incorporation of a model of the contrast sensitivity function (CSF), described in the corresponding section. It predicts the visibility of compression artifacts depending on the spatial frequency. Additionally the model can be extended by contrast masking (CM), which is for reasons of simplicity not considered in these experiments.

The JPEG 2000 encoder is block-based for reasons of complexity. Therefore any knowledge about the entire image like coefficient mean values or distribution statistics are unknown. Even if it might be important for an optimization of the color transformation it cannot be exploited. The low-level coding engine itself is based on an arithmetic coder (AC) that exploits intensively correlations between neighbored coefficients, but no inter-color channel correlations.

To avoid any computing precision related influences all computations are done with floating point operation. In particular the color transformations.

Color Spaces

The proper choice of a suitable color space for a rate-distortion based compression is an important issue. On one hand it needs to predict properly any visual distortion and on the other hand it needs good compressibility features like energy compaction, small inter-channel correlation and so forth. Actually, the best would be to do the compression in one space and the distortion prediction in another space. In a first attempt it is considered to use only one color space for both.

Three candidates were compared: The YC_bC_r [7], the La^*b^* [8] and the pattern-color separable opponent space [9]. The YC_bC_r is defined as a linear transformation applied to rgb frame-buffer values. It does not suppose any primary colors and is applied directly to gamma corrected images. It is the most common color space used for compression purposes due to its good compressibility features and simplicity. The La^*b^* is designed to map perceived color differences into a Euclidean color distance metric. It is designed for large uniform color patches (0 cycles/degree). The pattern-color separable opponent space (in the following just opponent space) was designed as the best linear transformation that allows to separate the issue of color sensitivity from the sensitivity dependent on the spatial frequency. It is a linear transformation applied to a color vector given by:

$$\left(\frac{L-L_B}{L_B}, \frac{M-M_B}{M_B}, \frac{S-S_B}{S_B}\right) \qquad (1)$$

where L,M and S stand for 1.0-normalized Smith-Pokorny cone fundamentals [10] and (L_B, M_B, S_B) describes the reference background color in the same coordinates system. The background color dependency is a specific characteristic of the opponent space.

Actually, the separability of color- and pattern-sensitivity has to be assumed for all three color space if they shall be used in a compression scheme that considers the sensitivity over frequency. While YC_bC_r and La^*b^* fulfill this requirement within certain limits the opponent space has been designed under this criterion. For this reason the opponent space should be a preferred candidate from distortion prediction point of view. On the other hand it has probably bad compression properties.

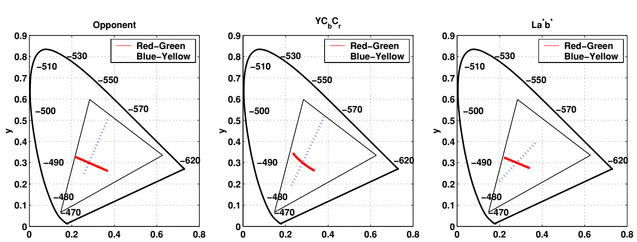

Figure 1: CIE 1931 chromaticity plots of the three color spaces

Figure 1 shows the three chromaticity plots. The chrominance axes of all three space are oriented in a similar way. Actually, the red-green of opponent and La^*b^* are nearly identical. The La^*b^* is the only non-linear transformation in itself. All three space are nevertheless treated by the non-linear gamma correction.

In the following it is referenced to the different color channels always as luminance (L), red-green (RG) and blue-yellow (BY) channel. That means the RG-channel of YC_bC_r is the third channel, while it is for opponent and La^*b^* the second channel. The YC_bC_r transformation is directly applied to the gamma-corrected rgb-values, while the opponent and La^*b^* are applied to images with linear gamma (1.0). For all original input images the not known implicit gamma correction is defined to 1.0. Since the output medium is a continuous tone printer no correction is necessary.

The compression scheme expects three input images quantized to a zero symmetric dynamic range. Given the rgb-gamut, the maximum output range can be computed and the transformation can be normalized to exploit the whole dynamic range. Nevertheless, not the whole dynamic range is used, to keep the now asymmetric chrominance range centered around zero. Otherwise a quantization to zero would result in a constant color shift in achromatic regions.

Contrast Sensitivity Curves

The human sensitivity to perceive small contrast differences is decreasing with higher spatial frequencies. The contrast sensitivity curves (CSF) are the quantitative description of this phenomena. For compression this means that quantization errors in high frequency contents cause less visual degradation than in low-frequency contents. Therefore the CSF can be applied to equalize the visual distortion (caused by compression) over frequency. It is implemented as a prefiltering operation before scalar quantization with a fixed minimal step-size. As shown at the left side of Figure 2, a 5 level

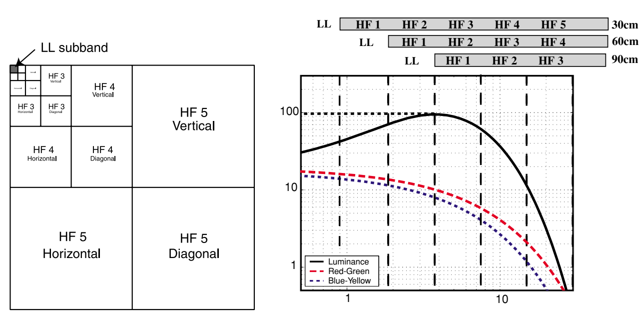

Figure 2: Relation between the CSF and the wavelet subband levels.

Mallat wavelet decomposition results in 16 subbands. Each level (HF1-5) contains subbands with the horizontally, vertically and diagonally oriented high-frequency details. The remaining low-frequencies are found in the LL subband. Each subband is filtered with the corresponding sub-part of the CSF, then quantized and at the decoding side filtered with the inverse filter function before it is reconstructed [4]. Which subband is filtered with which sub-part of the CSF depends on the assumed viewing conditions. The shown color bars on the top of the graphic show for three different viewing distances and a printing resolution of 270 dpi, where the subband levels of the decomposition are mapped to. It is evident that the maximum luminance sensitivity is linked for larger viewing distances to lower subband levels. Therefore the low-frequency dip cannot be exploited and the sensitivity has to be assumed also there as maximal. If the CSF parameters are now implemented for a minimal viewing distance of 15 cm (slightly below minimal focusing distance), the scheme delivers optimal quality for closest viewing and applies a rather conservative filtering for large viewing distances. But this is adequate for bitrates close to visually lossless compression.

There are two methods to measure the CSF; the threshold detection (TDE) or color matching experiments (CME). The TDE have the advantage that the experimental task is easy (either the observer sees the pattern or not), which results in stable data. They allow also the determination of absolute contrast thresholds, which gives an inter-channel weighting. The main drawback is that the contrasts at detection threshold are at least one order of magnitude smaller than the smallest contrast differences caused by quantization. Usually, the final sensitivity curve is achieved by a linear extrapolation from the threshold data.

Different than in [4] here TDE's were performed. The observers looked at 2.7 degree, sinusoidal gratings on a mean background with CIE 1931 Yxy coordinates (50cd, 0.28, 0.30). The viewing distance was 72cm and the task a Yes/No experiment with an automatic time forced setup (fadein/out each 500ms; 1s entirely visible). The measurements were done for each channel of each color space with 5 - 12 observers. The final results were fitted with a least-square fit to the logarithmic sensitivity. The luminance model is given by,

$$S_L = a_1 x^2 \mathbf{e}^{(b_1 x^{c_1})} + a_2 \mathbf{e}^{(b_2 x^{c_2})} \tag{2}$$

the chrominance by

$$S_C = a_3 \mathbf{e}^{(b_3 x^{c_3})} \tag{3}$$

The measurements were done at spatial frequencies 1, 2, 3, 4, 6, 7, 8, 10 cpd. Therefore the shape of the curves for frequencies smaller than 1 cpd is an extrapolation. Figure 3 shows the sensitivity curves after the fitting. The ordinate of these plots represents $1/C_T$ where the contrast $C_T = 1$ corresponds to a $\Delta E_{Lab} = 10$ difference to the mean background along the principal color space axis.

Some preliminary experiments were done by applying these absolute contrast curves to the task of compression. It resulted in significantly improved performance for the luminance channel, but at the same time some color artifacts (in particular for the La^*b^*, probably due to aspects of non-linearity) were observed. This was not very surprising since the inter-channel weighting gave up to 7 times less importance to the chrominance channels than recommended in La^*b^*. TDE actually give *an* inter-channel-weighting, but it has to be considered carefully: the relative shape of the measured curves varied very few between the observers, but the absolute sensitivity (equal to an entire shift of the curve in y-direction) differed by factor 2. The curves should deliver equal sensitivity at 0 cycles per degree for all color channels if all assumptions were correct. Because the La^*b^* metric is designed to have equal color distance for the same contrast (difference) at 0 cpd. This is not the case even if the curve shape below 1cpd is an extrapolation. The strongest argument why the absolute weighting needs to be corrected is the extrapolation from detection contrast to the contrast range finally used for the quantization prediction. Because similar curve shapes were observed in CME, it is recommended to modify the model regarding the inter-channel weighting. Additional experiments with combined matching and detection experiments need to be done to clarify this aspect.

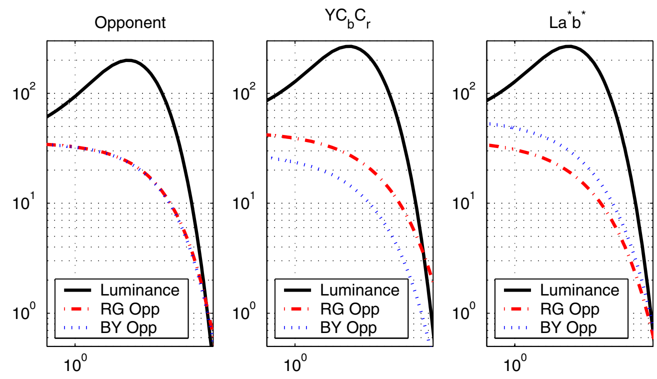

Figure 3: Fitted not normalized CSF curves

In the modified model it is assumed that the extrapolation to the relevant contrast range from the detection threshold changes the inter-channel importance in a non linear way, but preserves the relative curve shape. This way the sensitivity regarding color is described by the well established La^*b^* model and the sensitivity regarding the spatial frequency by the 1.0 normalized CSF curves. The inter-channel weighting of the opponent and YC_bC_r space were modified in a way that a color difference of 1 along any principal color space axis introduces the same ΔE_{Lab} value. This results for the YC_bC_r in the factors (1.00, 3.46, 2.34). The new modified model showed in all cases better chrominance and no visible loss in luminance performance.

Subjective Quality Evaluation

It is a well known problem that there does not exist an objective quality measure for images that reflects perfectly the subjective impression of a human observer. Due to this problem the peak-signal-to-noise-ratio (PSNR) is used. But it indicates only roughly the image quality and it is not properly defined for color-images. Briefly spoken, a subjective test to evaluate the real image quality cannot be avoided.

The goal of this evaluation is to determine the compression performance of three different color spaces. The La^*b^* and opponent space are always combined with the corresponding CSF filtering stage, the YC_bC_r space is tested once with CSF and once without CSF filtering (references as YCC). The YCC can be considered as default of JPEG 2000 and allows a judgement about the impact of the CSF filtering.

The observers are asked to judge the image quality based on images printed on a continuous tone printer (Fujix Pictrography 3000) with a resolution of 267 dpi. The print quality is as good that no printer related degradation (MTF) has to be considered. Each observer receives a pile of about 25 prints, representing the same image compressed in the different color spaces and at different bitrates. Given a print of the original the observer is asked to separate the pile into three categories:

1. **Perfect**: even in a side-by side comparison the compressed image cannot be distinguished from the original
2. **Good**: the image is degraded by compression, but it is only evident by direct comparison with the original
3. **Refused**: the image is obviously degraded by compression; the quality is no longer acceptable

This task is done for three different images (Bike 2560×2048, Bottles 1655×1313 and Fruits 1655×1325), which are 24 bit rgb-images. For each image and bitrate the average pile number is evaluated and in a second step the data is fitted to a Weibull function that models the smooth turnover from one category to the next, as shown in Fig.4. At least 6 observers attributed to the final averaged result. Given a specific com-

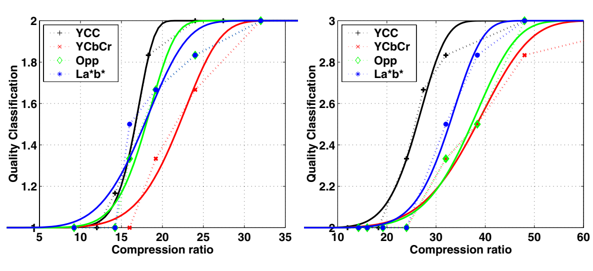

Figure 4: Result of fitting for transition from quality level 1 to 2 on the left; for level 2 to 3 on the right (for Fruits)

pression ratio the graph allows to predict in which category the image quality will be grouped by an average human observer. In Fig. 5 the corresponding compression ratios for the turnover points between the refused-accepted and accepted-perfect quality are resumed in a bar-graph.

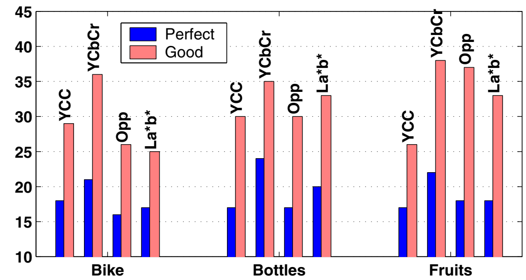

Figure 5: Result of fitting for transition from quality level 1 to 2 for each color space the left bar and for level 2 to 3 the right bar

It can be observed that the YC_bC_r space performs in all cases the best. The gain due to the CSF filtering can be determined by the comparison with the YCC space. It is consistent and in average around 29 %, which is a considerable improvement. Mapped into visual information it can be said that with the very conservative CSF filtering only the visible part of the high-frequency coefficients is encoded and the saved bits improve mostly the image quality in low-amplitude texture regions. Additional implementation of CM would here increase the compression ratios for the same quality even further. The opponent space, even when considered as a good choice from point of view prediction and separability capacities seems to suffer from a bad compressibility. Only for the Fruits-image, which is an image with very saturated colors and rather few luminance information it outperforms YCC and La^*b^*. In the other cases it performs worse or equal to the YCC that does not exploit the CSF. The La^*b^* performs in particular for the Bottles-image close to YC_bC_r, but for Bike and Fruits the performance is worse than YC_bC_r and opponent. This is probably due to some color related artifacts which concern in particular the red-green channel. In very saturated red-regions or achromatic regions close to a saturated red-region disturbing ringing artifacts appear.

Objective Signal Properties

No known objective measure allows a reliable prediction of subjectively perceived image quality. Nevertheless, the analysis of objective measures like signal energy, entropy, inter-channel and inter-coefficient correlation can help to understand some color space specific aspects. All presented measures are determined based on the wavelet coefficients from bitstreams decoded up to 1.75 bpp. This makes sure that visually irrelevant information, like it is encoded up to lossless compression, is not considered. All measures are averaged over a set of 8 large images, including the test images.

The signal energy E and the zero-order entropy H for a subband A is given by Eq. 4, where x and y indicate the coefficient position inside the subband. N is the considered quantization bitdepth and p_i the probability that the amplitude of coefficient A_{xy} is element of the i-th histogram bin.

$$E = \sum_{x,y} A_{xy}^2 \qquad H = -\sum_{i=0}^{2^N} p_i log_2(p_i) \tag{4}$$

In the following figures the horizontally and vertically oriented high frequencies (HF) are represented by its common mean value. The diagonal orientation is kept separate. The numbers from 1 to 5 on the x-axis indicate the subband level, where the highest number corresponds to the highest frequencies. The main energy contribution comes from the small LL-subband. Its encoding costs are negligible. Therefore the energy data is in the following expressed as percentage of entire HF energy.

Additionally, the inter-channel correlation is computed by Eq. 5. Only the absolute values of the coefficients are considered, because only the magnitude and not the sign have influence on the coding decision in the used scheme. Since the coder does not exploit any inter-color channel correlation it should be ideally very small.

$$\gamma_{IC} = \frac{\sum_m \sum_n |A_{mn}||B_{mn}|}{\sqrt{\sum_m \sum_n A_{mn}^2 B_{mn}^2}} \tag{5}$$

The analysis of the correlation between neighbored coefficients is not presented here. It showed that the chosen color space has no significant impact on it. But an improvement in the order of 5-7% is observed for the three spaces using the CSF filtering. This increased correlation allows a better exploitation of the coding contexts.

Fig.6 shows the averaged HF energy distribution over frequency. It represents somehow the "significance" that is assigned by the color space to the specific signal portion. It can be observed that the opponent space assigns a large energy portion to the blue-yellow channel, which is due to a strong coupling between the luminance and blue axis in the opponent transformation. Otherwise it behaves similar to the YC_bCr. The La^*b^* assigns very few energy to the chrominance channels, which might be an explanation for its sometimes poor red-green performance. The missing CSF filtering for YCC, results in a larger energy portion for luminance level 5 and almost all chrominance levels. Assuming that the CSF filtering is correct, this energy portion equals wasted bitrate, since it is not visible. It can also be observed that the energy percentage in the luminance subbands stays more or less constant over the first 4 levels, while it decreases constantly for the chrominance channels. This is on one hand due to the spectrum of natural images that decreases with high frequencies; on the other hand the rate-distortion scheme allocated rather more bitrate to the lower frequencies, since the $\frac{\Delta D}{\Delta R}$ was higher.

On the one hand entropy is related to energy, since more decoded energy costs more bitrate (entropy). On the other hand the entropy is a measure for image information. While

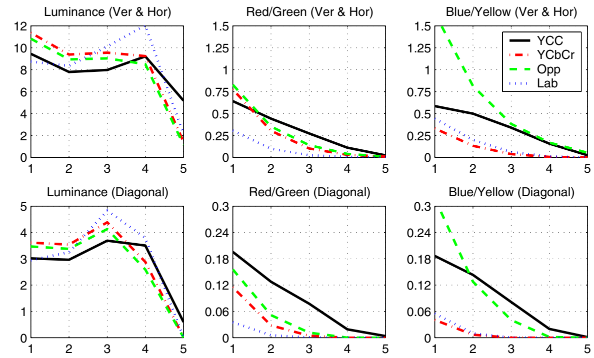

Figure 6: Percentage of energy in the high-frequency subbands, where the sum of all HF energy is 100%

the low-frequencies attribute 99% of the entire image energy it costs just 1% to encode. The HF attribute very few energy, but they cost a lot of entropy to be encoded. Nevertheless, this very small portion of energy is important, because it represents probably the difference between a human face with realistic skin texture or a blurred out face. The subjective ranking would be dramatically different.

Fig.7 shows the entropy plot. Again it is evident that the opponent space treats a portion of the image information as blue-yellow content that the other color spaces treat as luminance content. The La^*b^* assigns significantly less entropy to the red-green channel in diagonal orientation. This fits with the observed red-green ringing artifacts, since they appeared in particular along diagonally oriented edges.

Interesting is the allocation of entropy over frequency. The coefficients amplitude decreases in average by factor 2 that means the energy by factor 4 from level to level. The subband size increases by factor 4. This results in approximately constant energy per subband, ignoring the decreasing spectrum energy of natural images for HF. The entropy is increasing with the subband size by factor 4 from one level to the next. Both can be observed for the YCC in the vertical and horizontal luminance graph. Actually these curves represent the final "saturated" status. If the bitstream is decoded to smaller bitrates, the rate-distortion scheme allocates for the best $\Delta D/\Delta R$ the most bits. Therefore first in the bitstream the LL subband and lower HF subbands will be quantized to a value different than zero (coefficients become significant). This means that *in average* first the lower-frequency subbands get significant and than later on the HF subbands. This represents an *implicit* HF filtering. In Fig.7 can be observed that in particular the chrominance channels are not "saturated" for all frequency levels. From level 3 on, the spent entropy decreases, because the afforded bitrate was to high in relation to the gained energy. For the three color spaces with CSF this phenomena is more visible, since the visual importance of HF energy is lowered by the CSF filtering. The described behavior leads to a strong discrimination of visually important low-amplitude, high-frequency texture. It could be compensated by a distortion measure that considers CM and is probably itself a function of entropy.

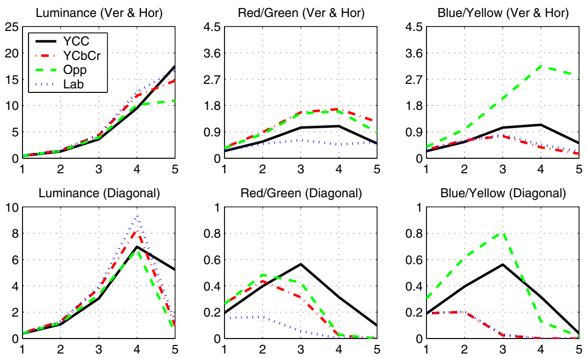

Figure 7: Percentage of entropy in the high-frequency subbands, where the sum of all entropy is 100%

Any inter-channel correlation (ICC) can be understood as not exploited information redundancy, since the encoder does not take it into account. Fig.8 shows the inter-channel correlation γ_{IC} for the color spaces. It can be observed that the correlation decreases for all spaces with increasing frequency. That means to extend the coding algorithm by considering inter-channel correlations would be of minor impact for the expensive HF subbands. The La^*b^* has the lowest ICC, what means that its lower compression performance is not due to ICC. The opponent space has a significantly increased ICC for all channel combinations, which is a strong indication for lost coding performance.

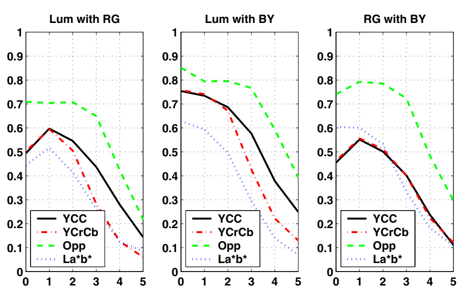

Figure 8: Inter-Channel correlation over increasing frequency

Conclusions

The YC_bC_r space performed for all images the best, even if the hypothesis of color-pattern separability is not entirely valid. But its better compressibility features seem to be more important. Together with the CSF filtering gain of 30%, the YC_bC_r has at time to be considered as best compromise in terms of prediction and compressibility properties. Further analysis should allow to design an equally simple color space with even better performance than the YC_bC_r. The separation of prediction and compression spaces might be even more promising, but it contains other difficulties and is significantly more complex. The presented analysis should be extended by the incorporation of CM, because a main criterion for bad visual quality ranking was lost low-amplitude texture.

The presented work attributes with the measurement of CSF for specific color space, the subjective compression performance evaluation and an analysis of objective signal properties to the gap between the domains of: Color, Human Vision and Image Compression.

References

1. A. Drukarev, "Compression-related properties of color spaces," *Proceedings of SPIE* **3024**, pp. 855–863, February 1997.
2. N. Moroney and M. Fairchild, "Color space selection for JPEG image compression," in *Journal of Electronic Imaging*, vol. 4, pp. 373–381, 1995.
3. "Coding of still pictures - JPEG 2000 verification model 4.1," in *ISO/IEC JTC 1/SC 29/WG 1 N1286*, 1999.
4. M. Nadenau and J. Reichel, "Compression of color images with wavelets under consideration of the HVS," in *Proc. SPIE Human Vision and Electronic Imaging*, vol. 3644, SPIE, (San Jose, CA), January 24–29, 1999.
5. K. T. Mullen, "The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings," *Journal of Physiology* **359**, pp. 381–400, 1985.
6. D. Taubman, "High performance scalable image compression with EBCOT," *submitted to IEEE Transactions on Image Processing* , 1999.
7. "Encoding parameters of digital television for studios," in *CCIR Recommendations 601-2*, 1990.
8. R. W. G. Hunt, *Measuring Colour*, Ellis Horwood Limited, Market Cross House , Cooper Street, Chichester, West Sussex, PO19 1EB, England, 1987.
9. A. B. Poirson and B. A. Wandell, "Appearance of colored patterns: pattern-color separability," *Optics and Image Science* **10**(12), pp. 2458–2470, 1993.
10. V. C. Smith and J. Pokorny, "Spectral sensitivity of the foveal cone photopigments between 400 and 500nm," *Vision Research* **15**(2), pp. 161–171, 1975.