Subjective evaluation of compression artifacts for CMYK prepress images

Peter De Neve* ***Ignace Lemahieu***

Department of Electronics and Information Systems, ELIS-MEDISIP University of Gent, Belgium

****e-mail : deneve@elis.rug.ac.be***

Abstract

In todays digital prepress workflow images are most often stored in the CMYK color representation. Besides the fact that 32 bits per pixel are needed to store the color information, these images often have a large spatial resolution. This indicates that lossy compression can be considered to reduce disk capacity problems and to limit the transmission times needed to transmit images from the designer to the prepress shop.

Lossy compression techniques are mostly evaluated by calculating pixelwise distortion measures such as the root-mean-square-error (RMSE) or the peak-signal-to-noise-ratio (PSNR). Since we are aware of the weaknesses these distortion measures have, an experiment is designed in which the subjects are able to evaluate several lossy compression schemes in a subjective way.

Among the used compression methods are some popular algorithms such as the standard coder for lossy compression JPEG. They are evaluated on three typical CMYK images and for each compression method four different quality factors, ranging from no visual image degradation to clearly distinguishable artifacts, are used.

The method of evaluation is the paired comparison method. All subjects are presented a number of image pairs and their task is to decide which of the two images looks most pleasant to them.

From the results it is shown that for low compression ratios there is not very much difference in the performance of the different algorithms. For the highest compression ratios a wavelet coder and an algorithm that tends to reduce the blocking artifacts are considered best by the subjects. In general one will see similar results from the judgements between experts and non-experts although the experts are more determined in the results of their observations.

1. Introduction

Since the entire prepress workflow has been digitized the amount of data that has to be stored on different media or which has to be transmitted through local- and wide-area networks has increased tremendously. In general four color components, i.e. cyan (C), magenta (M), yellow (Y) and black (K) are used to represent the color information. This means that 32 bits per pixel have to be allocated in order to represent the continuous tone in images. Together with the large spatial resolution these images often have it is obvious that a certain amount of data reduction would be helpful.

At first, a lossless compression scheme might be considered in an application field such as the printing industry where high-quality is demanded. Since these algorithms only yield moderate compression ratios, i.e. up to two, lossy or near-lossless alternatives must be considered. The lossy alternatives will provide a large improvement in data reduction but their weakness is the difficulty to garantuee any image quality.

In general, lossy compression techniques are evaluated by calculating pixelwise distortion measures such as the root-mean-square-error (RMSE) or the peak-signal-to-noise-ratio (PSNR). Despite the low calculational complexity of these statistical distortion measures they have some severe shortcomings such as the prediction and detection of blocking artifacts that occur in block-based compression techniques such as JPEG.

In the literature several alternatives have been proposed to overcome this problem on color images. Some attempts have been made to include spatial structure into color spaces like CIEL*a*b*such S-CIELab [1] or PQS [2], but sometimes with limited succes when speaking about the ability to quantify blocking artifacts on some arbitrary images [3]. Therefore, psychovisual experiments are certainly the best way to evaluate the performance of compression techniques.

The paper is organized as follows. In section 2, we briefly propose the compression techniques that have been used in the experiment. In section 3 the psychovisual experiment is described. In section 4 the data-analysis and results are presented. Section 5 concludes with a summary.

musicians (1853 × 2103) | cafe (2048 × 2056) | scid0 (2048 × 2560)

Figure 1: *Grayscale thumbnails of the images that are used in this experiment. Between round brackets are the width×height of the images in pixels.*

2. Compression methods

2.1. JPEG

Among the class of block-based transform coders, JPEG[4] is well known as the standard coder for the lossy compression of still images. The main principle of the JPEG algorithm is dividing the image into separable blocks of 8×8 pixels that are transformed using a discrete cosine transform (DCT). The coefficients of each block are then quantized and coded independently of the other blocks. Because the base functions are only defined within the blocks, the quantization errors made during the quantization process will not be correlated between neighboring blocks. This mismatch can become visible near the block boundaries and can lead to blocking artifacts in the image which become very annoying especially at higher compression ratios. Blurring and banding are other possible artifacts, but are inherent to all lossy transform coders.

2.2. Exploiting tonal decorrelation using block-based transform coders

Another problem occurs when compressing four-dimensional CMYK images with JPEG. The only correct way to compress them using the current implementations is to process the four different color channels independent of each other, but then the tonal correlation that is present in any natural scene is not taken into account. Transforms like CMYK → YUV or CMYK → YCbCr are not tolerated because the inverse transform is irreversible.

Recently a method was proposed to exploit the tonal correlation while staying in a four-dimensional color space[5, 6]. This technique consists of two steps. In a first step the CMYK color space is decorrelated into a new four-dimensional color space. Therefore the image is divided into blocks of usually 32×32 pixels since there will be a higher tonal correlation between pixels that are not too far from each other. The tonal decorrelation is performed on each block using a singular value decomposition (SVD) of the image data. From this approach one gets a 4×4 unitary transformation matrix from which six rotation angles, used to rotate the original color space into the decorrelated one, can be achieved. These angles are then quantized, usually to 256 values, and have to be stored per block. The color data is recalculated according to the quantized angles. This will cause the decorrelation between the different components to be no longer zero but only sub-optimally. But it also indicates that the color decorrelation step until now is still lossless (except for rounding errors). The actual gain of this technique will be in the fact that due to the decorrelation of the image data, the most dominant color will be along the principal axis of the new color space. The other components perpendicular to that axis will be much smaller and can be coded with less bits.

In a second step the spatial redundancy between neighboring pixels in the decorrelated color data has to be exploited. This can be done by applying any lossy compression technique. Because the color decorrelation mechanism itself is lossless all the distortion will be concentrated in the quantizer of the compression technique. It should also be noted that the order in which the two steps are executed can be switched. In this paper we used the SVD technique in combination with JPEG and with the "lapped biorthogonal transform" (LBT) coder[7] in order to reduce the spatial redundancy. This last algorithm performs best when it is executed before the tonal decorrelation step. The LBT coder works with overlapping blocks. This means that the base functions in the different blocks are not independent of each other so there will be some correlation between the neighboring blocks. This will eventually lead

Table 1: *Description of the stimuli used in the experiment ("CR" denotes compression ratio)*

Stimuli	method	scid0 & cafe	musicians
1	JPEG, CR=	5	10
2	JPEG, CR=	10	17
3	JPEG, CR=	15	24
4	JPEG, CR=	20	30
5	SVD+JPEG, CR=	5	10
6	SVD+JPEG, CR=	10	17
7	SVD+JPEG, CR=	15	24
8	SVD+JPEG, CR=	20	30
9	LBT+SVD, CR=	5	10
10	LBT+SVD, CR=	10	17
11	LBT+SVD, CR=	15	24
12	LBT+SVD, CR=	20	30
13	SPIHT, CR=	5	10
14	SPIHT, CR=	10	17
15	SPIHT, CR=	15	24
16	SPIHT, CR=	20	30

to a compressed image with less blocking artifacts compared to JPEG for the same data reduction.

2.3. Wavelets

Unlike in block-based transform coders, wavelet transforms operate on the entire image. An immediate advance of this approach is that wavelet compressed images do not suffer from blocking artifacts. Since they have become very popular in compression technology during the last years the performant SPIHT algorithm[8], designed by Said&Pearlman, was included in this experiment as well.

The algorithm is based on the well-known zerotree coding principle in which the aim is to exploit the spatial similarity between wavelet subbands of the same orientation. A tree structure is used to define a spatial relationship in the wavelet pyramid. The method consists of a series of sorting and refinement passes in which the wavelet coefficients of the different levels in the pyramid are labeled as significant/insignificant towards a certain treshold. The entire tree will be scanned according to this procedure. An additonal advantage of the technique is the generation of a fully embedded bit stream, making us able to stop the bit stream at any point to reconstruct a temporary "best" image.

3. Quality assessment

As mentioned in the previous section four algorithms are used in the experiment. For each compression technique

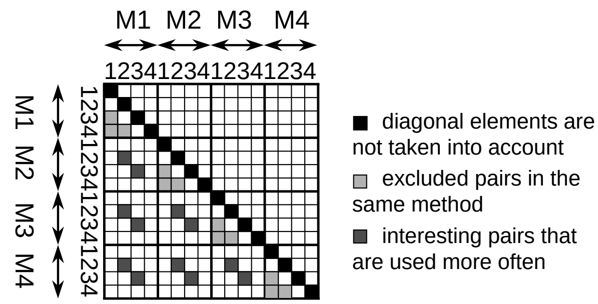

Figure 2: *Block design in which all stimulus pairs are shown. M1, M2, M3 and M4 represent the four compression algorithms that are used in the experiment. A sequence "1234" indicates the four different quality factors or compression ratios with "1" being the lowest and "4" the highest compression ratio.*

four quality factors are considered which brings the total number of stimuli to 16. The stimuli are presented and described in table 1. Three representative prepress images were included in the test. A thumbnail of these images and some of their properties are shown in figure 1. The images were carefully choosen such that different aspects like details, vignettes, fleshtones, ... are presented to the subjects in this experiment.

The method of evaluation is the pairwise comparison method. This means that all subjects are shown a number of image pairs with a left and right printed image. The images were printed on a BARCO GRAPHICS DIGIPRESS digital printer. The experiment is of the two-alternative-forced-choice (TAFC) type, i.e. all subjects have to choose a best image between the left and right presented stimulus. A neutral choice is invalid. The images are judged and compaired based on their pleasantness. The subjects have no knowledge about the image processing operations that were applied on the images, except that no identical image pairs, this means pairs in which the left and right images are equal, are present in the test.

Based on four the compression algorithms with four different quality factors each, the total number of image pairs is $(16 \times 15)/2 = 120$. The block design in which all stimulus pairs are presented is shown in figure 2. It should be noted that only the lower triangular part of the block design is considered. The maximum duration of the test is set to 30 minutes and the subjects are supposed to perform two comparisons per minute. Since this results in 20 pairs per image from a set of 120 pairs per image a problem seems to arise. It can be overcome by introducing incomplete block designs so that not all subjects have to judge all pairs. Having 120 pairs and 20 evaluations per subject per image this means that six different incomplete sets have to be made. From figure 2 it follows that some pairs are excluded while other interesting pairs are taken

Table 2: *Origin of the subjects.*

Subjects	Number
University personnel	61 (37%)
University students	44 (27%)
Specialists	43 (26%)
Others	16 (10%)
	164

into account more often. All image pairs are randomized in order to eliminate any systematic errors.

The total number of subjects that participated in the psychovisual experiment is 164. Nine people had to be rejected because they failed for a test on color vision deficiencies. The origin of the subjects is given in table 2. The youngest subject is 19 years old while the oldest one is 61 years old. 68% of the subjects have their age between 20 and 30 years old.

4. Analysis and results

In this experiment one has to make an analysis for paired comparion results where subjects had to judge a one-dimensional attribute, such as image quality. It is assumed that a basic Thurstonian model[9] can be applied, i.e. the strength of each stimulus attribute x_i (i=1,...,16 in this experiment) is measured on a psychometric interval scale Ψ. Due to internal noise this strength is considered to have a Gaussian distribution

$$x_i \sim N(S_i, \sigma_i^2) \tag{1}$$

where S_i and σ_i respectively denote the position and variance of stimulus i on the Ψ-scale. In case V of this Thurstonian model the noise spread parameter is assumed to be constant for stimuli, or $\sigma_i = \sigma, \forall i$. Another assumption is that the strengths of the different stimuli are uncorrelated, or $\rho_{ij} = 0, \forall i = j$ where ρ_{ij} is the correlation coefficient between the perceived attributes of stimulus pair (i, j). By making these assumptions the distribution of the perceived difference between scale values of a stimulus pair (i, j) becomes

$$d_{i,j} \sim N(S_i - Sj, 2\sigma^2). \tag{2}$$

It is projected on a ratio scale Ψ', i.e. an interval scale with a natural origin, which in our case indicates a zero difference. For the calculation of the scale values we use the software implementation DIFSCAL[10]. The input values for the program are the frequency distributions per category for each stimulus pair. In this experiment only two categories have to be considered since the subjects have to make a decision between which one of the two presented stimuli looks best. From these frequency distributions DIFSCAL calculates stimulus scale values in σ-units. Furthermore it is not necessary that all pairs are present since the algorithm is capable of dealing with a limited number of stimulus pairs that are absent.

In figure 3 the vertical axes indicate the stimulus scale values. Differences ΔD between points on these axes provide information on how different two stimuli are considered. These differences are related through a probability scale. For instance, when ΔD between two stimuli equals 0.5, the probability that they are considered differently is 64%. When ΔD equals 1 or 2 this corresponds respectively to a probability of 76% and 92%. So in order to be sure that two stimuli are interpreted different the ΔD value between them should be at least 2.

In figure 3(a) and (b) rate-distortion curves are shown for the musicians image. The four compression algorithms, as defined in table 1, are presented and the quality scale is pointed downwards. From the results for all subjects it can be seen that JPEG and LBT+SVD behave as expected while for SVD+JPEG and SPIHT somewhat odd relationships are found. This is not the case for the specialists however who observe the greatest range in scale values, almost two, in the SVD+JPEG algorithm. Stimulus 15 shows a strange behaviour and should in fact be rejected from the experiment because it was not printed color constantly regarding to the other stimuli. The yellow colors in the printed sample are more saturated, what causes the image to have a more natural look and probably explains why it is preferred over the other stimuli even those with lower compression ratio. Removing stimulus 15 from the experiment also shows that all subjects evaluate the wavelet compressed images having almost the same quality, considering the errorbars on the results. Other conclusions are that for the lowest compression ratios differences between the four algorithms are very small which indicates no difference in image quality is observed. The best coder for the highest compression ratio, i.e. 30, seems to be SPIHT closely followed by LBT+SVD.

In figure 3(c) and (d) the same analysis is made for the cafe image. At first sight it is clear that the ΔD values are much larger than for the musicians image which indicates that people can observe the differences between the stimuli with greater certainty for the cafe image. Almost all rate-distortion curves behave as expected. From stimuli 1 to 4 it follows that all subjects, including the specialists, observe big differences in the quality of the JPEG encoded images. This means that for JPEG observers clearly associate lower quality with higher data reduction. On the other hand for the SVD+JPEG algorithm differences are invisible except for the highest compressed image (stimulus 8). For the highest compression ratio, JPEG performs worst while the

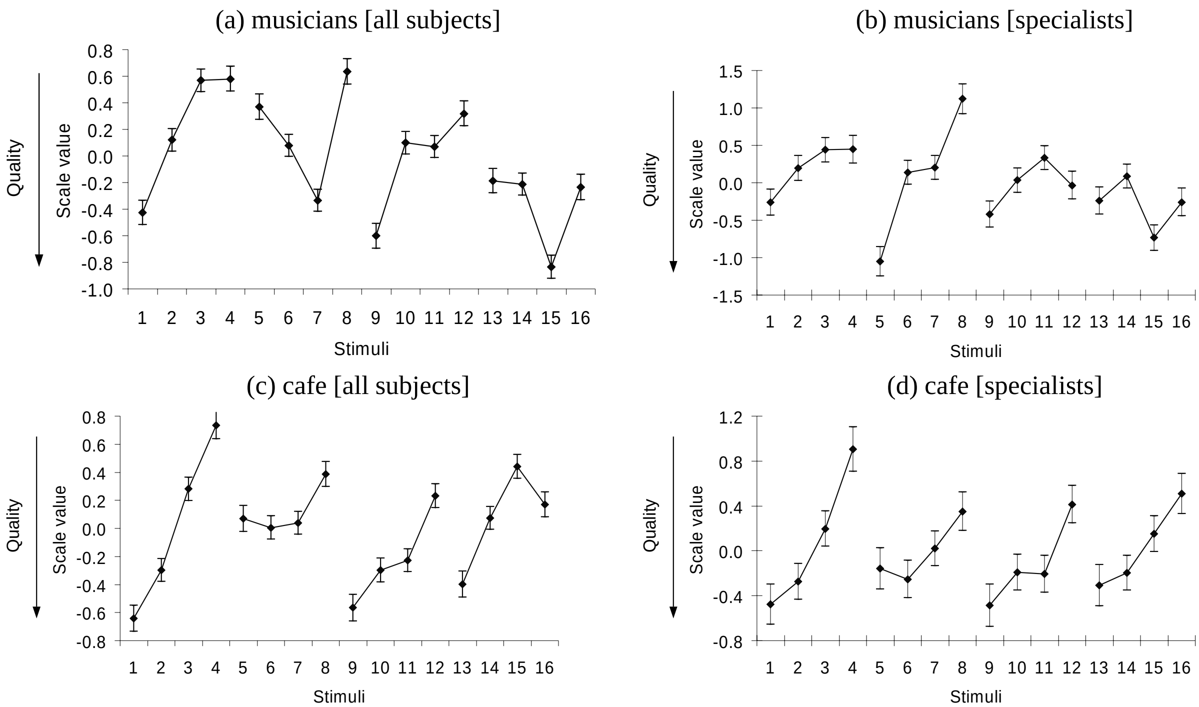

Figure 3: *Rate-distortion curves are shown for the musicians (a) and (b) and cafe (c) and (d) image. The four compression algorithms, as defined in table 1, are presented and the quality scale is pointed downwards.*

other algorithms perform almost equal.

In figure 4 results are shown for the scid0 image. From the alternating behaviour of the rate-distortion curves it is clear that all subjects had a very difficult task judging these image pairs. This can be further understood from the fact that the standard deviations on the resulting scale values are substantially larger than with the previous two test images and also the range in scale values is much smaller. All the observations indicate that all subjects, including the specialists, cannot observe any difference in the algorithms. The only structure is in the evaluation of the SPIHT coder by the specialists although the differences are small.

5. Conclusions

In this paper we have presented a psychovisual experiment that was designed to evaluate printed copies of some typical prepress images that were processed by some lossy compression schemes. The used compression ratios are situated in a range between 5 and 30. Among the used compression schemes are some popular coders such as JPEG and the SPIHT wavelet coder. We have also included a compression algorithm that, next to the spatial redundancy, exploits the tonal correlation between the four color components in the CMYK images. This scheme works in a four-dimensional color space and prevents an unwanted transform to a three dimensional color space because of the irreversibility of the inverse transform that is needed at the decompression side.

In general it is found that a psychovisual experiment on paper is more difficult to perform for the subjects than a similar experiment on a CRT monitor. It seems that artifacts will be observed earlier on the monitor than they are on paper. It can be made a rule that the printing process does "absorb" compression artifacts up to a certain level.

For the musicians and cafe image one finds that the rate-distortion curves behave more or less as expected, especially for the cafe image. It is found that the wavelet coder and LBT+SVD technique perform best at the highest compression ratios. In general the opinion of the experts and the non-experts can be considered the same although the range in scale values is higher for the experts. This indicates that when the experts observe a difference between the two images, they are more determined in their judgement compaired to the non-experts.

For the scid0 image, which contains a lot of fleshtones, it is found that the subjects do not notice any difference between the different compression algorithms. From the results it is clear that they are forced to make a guess between the images most of the time. Another conclusion is that the image can be compressed quite well before any artifacts are observed.

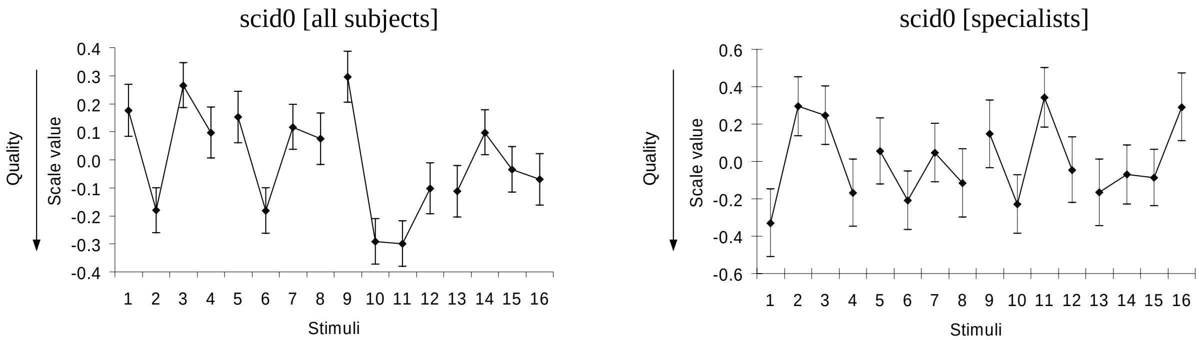

Figure 4: *Rate-distortion curves are shown for the scid0 image. The four compression algorithms, as defined in table 1, are presented and the quality scale is pointed downwards.*

6. Acknowledgements

This work was financially supported by the Fund for Scientific Research - Flanders (Belgium) and the Flemish Institute for the Advancement of Scientific-Technological Research in Industry (IWT). The authors would like to thank Jean-Bernard Martens (IPO, Eindhoven) for his contribution and Martin Boschman (IPO, Eindhoven) for making his software package DIFSCAL available. The authors would also like to thank Hans De Stecker from BARCO GRAPHICS for his support.

References

[1] X. Zhang and B. Wandell, "A spatial extension to CIELab for digital color image reproduction," in *SID Symposium Technical Digest*, vol. 27, pp. 731–734, 1996.

[2] K. Kotani, Q. Gan, M. Miyahara, and V. Algazi, "Objective picture quality scale for color image," in *Proceedings of the IEEE International Conference on Image Processing - ICIP'95*, vol. 3, pp. 133–136, IEEE, Oct 1995.

[3] X. Zhang, E. Setiawan, and B. Wandell, "Image distortion maps," in *The Fifth Color Imaging Conference: Color Science, Systems and Applications*, (Scottsdale, Arizona), pp. 120–125, IS&T/SID, Nov. 1997.

[4] W. Pennebaker and J. Mitchell, *JPEG Still Image Data Compression Standard*. Van Nostrand Reinhold, 1993.

[5] P. De Neve, W. Philips, K. Denecker, and I. Lemahieu, "Introducing a decorrelated color space in the lossy compression of pre-press applications," in *The Fifth Color Imaging Conference: Color Science, Systems and Applications*, (Scottsdale, Arizona), pp. 88–91, IS&T/SID, Nov. 1997.

[6] P. De Neve, K. Denecker, W. Philips, and I. Lemahieu, "Compressing CMYK images by removing color redundancy: A perceptual based quality evaluation," in *International Symposium on Electonic Image Capture and Publishing*, vol. 3409, (Zürich, Switzerland), pp. 267–276, SPIE, may 1998.

[7] H. Malvar, "Lapped biorthogonal transforms for transform coding with reduced blocking and ringing artifacts," in *Proceedings of the 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. III, (München, Duitsland), pp. 2421–2424, IEEE Computer Society Press, Apr. 1997.

[8] A. Said and W. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits ans Systems for Video Technology*, vol. 6, no. 3, pp. 243–250, 1996.

[9] L. Thurstone, "A law of comparative judgement," *Psychological Review*, pp. 273–286, 1927.

[10] M. Boschman, "DIFSCAL: A program for the analysis of difference scaling results," tech. rep., IPO, Center for Research on User-System Interaction, Eindhoven, 1997.