Perceptual Blur and Ringing Metrics: Application to JPEG2000

Pina Marziliano, [1] Frederic Dufaux, [2] Stefan Winkler, [3,*]
Touradj Ebrahimi [2]

Genista Corp., 4-23-8 Ebisu, Shibuya-ku, Tokyo 150-0013, Japan

Abstract

We present a full- and no-reference blur metric as well as a full-reference ringing metric. These metrics are based on an analysis of the edges and adjacent regions in an image and have very low computational complexity. As blur and ringing are typical artifacts of wavelet compression, the metrics are then applied to JPEG2000 coded images. Their perceptual significance is corroborated through a number of subjective experiments. The results show that the proposed metrics perform well over a wide range of image content and distortion levels. Potential applications include source coding optimization and network resource management.

Key words: blur, ringing, perceptual quality, subjective experiments, image compression, JPEG2000.

1 Introduction

Tremendous advances in computer and communication technologies have led to a proliferation of digital media content. However, digital images and video are still demanding in terms of processing power and bandwidth, and thus are often impaired by various types of artifacts such as noise, blockiness, blur, ringing, etc. [1]. In order to optimize imaging systems and to improve the

* Corresponding author, email: stefan.winkler@epfl.ch
[1] P. Marziliano is now with the Division of Information Engineering in the School of EEE at Nanyang Technological University, Singapore.
[2] F. Dufaux and T. Ebrahimi are with the Signal Processing Laboratory at the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland.
[3] S. Winkler is now with the Audiovisual Communications Laboratory at the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland.

perceptual quality of delivered content, metrics are needed to identify and measure these various artifacts.

Several perceptual metrics have already been developed for some of these artifacts. We can distinguish two categories of metrics: full-reference and no-reference. In the former case, a processed image is compared to a reference (e.g. the original). In the latter case, the metric is not relative to a reference image, but rather an absolute value associated to any given image. Much of the research up to now has been on metrics falling in the full-reference category. Quality assessment without a reference is intrinsically difficult, because the distinction between image features and artifacts is often ambiguous. Most existing no-reference metrics focus on blockiness, which is still relatively easy to detect due to its regular structure – see e.g. [2] for a comparison of three such metrics. More recently, we presented results on video quality assessment for Internet streaming [3] and mobile applications [4] using a no-reference quality metric.

In this paper, we are interested in two types of artifacts, namely blur and ringing. Blur is due to the attenuation of the high spatial frequencies in the image, and ringing is caused by the quantization of high frequency coefficients in transform coding. Blur is characterized by a smearing of edges and a general loss of detail, whereas ringing introduces ripples around sharp edges.

We present both full-reference (FR) and no-reference (NR) metrics that measure the perceptual amount of blur as well as a full-reference metric that measures the amount of ringing. The proposed metrics are defined in the spatial domain and are based on the analysis of the edges in an image. The blur metric measures the spread of edges, and the ringing metric measures oscillations around edges. No assumptions are made on the type of content or the particular blurring and ringing process. These objective measures correlate well with the perception of blur and ringing. The proposed metrics also have the advantage of a very low complexity and can therefore be used to analyze video quality in real-time [3,4].

Image coding aims at minimizing the distortion of a compressed image for a given bit rate (alternatively, one can minimize the bit rate for a given distortion level). This requires methods for accurately measuring the distortion or quality of a coded image. The distortion is often evaluated by simple fidelity metrics such as Mean Square Error (MSE) or Peak Signal-to-Noise Ratio (PSNR). Unfortunately, such metrics do not correlate well with human perception. Therefore, perceptual metrics are needed for a more relevant measurement of image quality, the ultimate goal being encoder optimization based on these metrics. However, different coding schemes are characterized by very different types of artifacts. For instance, the coding techniques based on the Discrete Cosine Transform (DCT), such as those used in JPEG and MPEG,

mostly bring about blockiness artifacts. Conversely, the new JPEG2000 standard [5,6], which is based on a wavelet transform, mostly introduces blur and ringing artifacts.

In this paper, the proposed blur and ringing metrics are applied to measure the quality in JPEG2000 coded images. Note that the use of no-reference metrics is especially interesting for the case of JPEG2000. Indeed, thanks to its scalable properties, a JPEG2000 bitstream can be decoded at multiple quality levels and/or resolutions. In the latter case, an original may not exist, which makes it impossible to use a full-reference metric.

The paper is structured as follows. In Section 2, we illustrate the origins of blur and ringing artifacts. We then describe the perceptual blur metric, which was initially defined in [7], and use it in the design of a new ringing metric. In Section 3, we validate each of our perceptual metrics via subjective experiments and analyze the agreement between the metrics' predictions and the observer ratings. Finally, we draw some conclusions in Section 4.

2 Artifacts and Metrics

Blur in an image is due to the attenuation of the high spatial frequencies, which commonly occurs during filtering or visual data compression. While measuring the perceptual blur in an image or a video sequence has not yet been investigated, related research topics include blur identification [8], blur estimation [9,10], image deblurring [11] and blind deconvolution [12]. In practice most of these methods require iterative solving algorithms, which are computationally demanding.

Ringing in an image is also caused by the quantization or truncation of the high frequency transform coefficients resulting from DCT- or wavelet-based coding. In the spatial domain this causes ripples or oscillations around sharp edges or contours in the image. This is also known as the Gibbs phenomenon. The problem of removing ringing artifacts is considered in [13] and solved using a maximum-likelihood approach. A method for the detection of image regions that exhibit ringing is presented in [14] as part of a blockiness measurement technique.

In the lossy JPEG2000 compression scheme [5,6] for example, the standard filter used for the wavelet decomposition is the Daubechies $(9, 7)$. Since the decomposition is done in a separable manner, i.e. first on the rows and then on the columns, it suffices to show the effect of these filters in 1D. Figure 1 illustrates the effects of blur and ringing on a sharp edge.

Fig. 1. Effect of Daubechies $(9, 7)$ filter on a sharp edge.

Most of the papers cited above do not attempt to measure the perceptual impact of these artifacts. However, it is of great importance to be able to objectively quantify the *perceived* blur and ringing in an image. The goal is to establish metrics which correlate with the human visual experience by mapping the objective measurements onto subjective test results.

Our blur and ringing metrics are defined in the spatial domain. Both artifacts appear mostly along edges or in textured areas. The proposed blur metric thus attempts to measure the spread of the edges, whereas the ringing metric measures the ripples or oscillations around these edges.

For color images, blur and ringing are measured on the luminance component. While the algorithms consider primarily still images, it is straightforward to extend the techniques to digital video by measuring the artifacts in every frame [3,4]. The low algorithmic complexity is essential in this case in order to be able to measure the distortions in real time.

2.1 Blur Metric

Our technique for measuring blur is based on the smoothing or smearing effect of filtering or compression on sharp edges, and consequently attempts to measure the spread of the edges. The algorithm is summarized in Figure 2.

First we apply an edge detector (e.g. a Sobel filter) to the luminance component of the image. Noise and insignificant edges are removed by applying a threshold to the gradient image. We then scan each row of the processed image. For pixels corresponding to an edge location, the start and end positions of the edge are defined as the locations of the local luminance extrema closest to the edge. The spread of the edge is then given by the distance between the end and start positions, and is identified as the local blur measure for this edge location. The global blur measure for the whole image is obtained by averaging the local blur values over all edges found.

An example of a row in an image is illustrated in Figure 3. For the edge location $P1$, the local maximum $P2$ defines the start position, while the local minimum
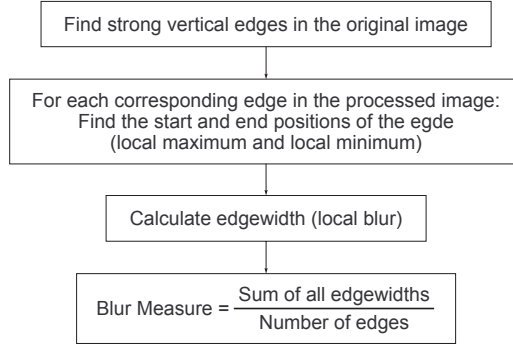
4

Fig. 2. Flow chart of the full-reference blur metric. In the no-reference case, the processed image replaces the original image in the first box.
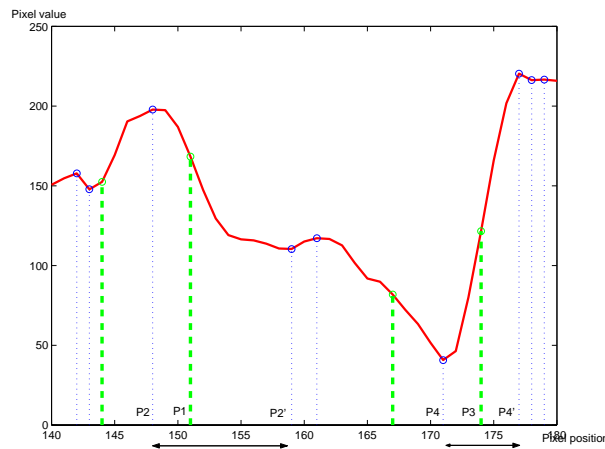

Fig. 3. One row of the blurred image. The detected edges are indicated by the dashed lines, and local minima and maxima around the edge by dotted lines. The edge width at $P1$ is $P2' - P2$.

$P2'$ corresponds to the end position. The spread of the edge is $P2' - P2 = 11$ pixels in this example. Similarly, for the edge $P3$, the local minimum $P4$ is the start position, the local maximum $P4'$ is the end position, and $P4' - P4 = 6$ pixels is again the spread.

In the algorithm described above, only vertical edges are considered. This is done mostly for performance reasons. It is obviously an approximation, as only the blur projected onto the horizontal direction is measured. The algorithm can easily be extended to the horizontal edges by filtering with a horizontal Sobel filter and then scanning each column. It is also possible to measure blur along the actual local edge gradients by taking into account the gradient orientation. However, our tests showed that this does not improve the measurements; using the vertical edges is sufficient in practice.

The algorithm described here lends itself to both a full-reference and a no-reference implementation. In the full-reference blur metric, we use the edges

of the original image to determine the edge locations. For the no-reference blur metric, the edges are obtained directly from the processed/compressed image [7]. While this affects the precision of edge detection to a certain extent (depending on the amount of compression or distortion), it is still possible to achieve good correlations with perceived blur, as will be shown in Section 3.

In addition to encoder optimization applications, the blur metric can also be used for autofocusing an image capturing device.

2.2 Ringing Metric

The ringing metric is based on and makes use of the blur metric described in the previous section. The algorithm is summarized in Figure 4.
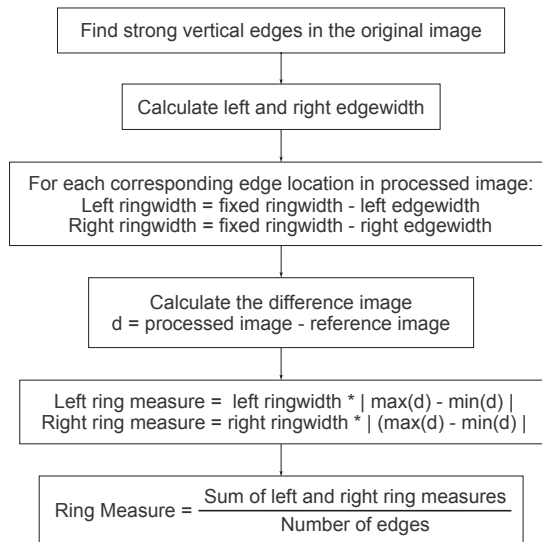
Fig. 4. Flow chart of the full-reference ringing metric.

Similar to the blur metric, the ringing metric is defined for each important vertical edge. It first finds the vertical edges in the original image (weak edges and noise are again discarded by means of thresholding) and calculates the difference between the processed image and the reference. It then scans each row in the processed image and measures the ringing around each edge.

We define a left and a right ring measurement. Furthermore, we define the ringing support as a fixed ringwidth (given a priori from the effects of the wavelet decomposition filters, cf. Figure 1) minus the edge width due to blur (as defined in the blur measurement in the previous section). Then we take the difference between the minimum and the maximum of the difference image inside this support and multiply by the ringing support width. We add the

6

left and right ring measures and take an average over all edges to obtain the global ringing measurement.

The ringing along an image row with two sharp edges is illustrated in Figure 5. The left edgewidth as defined previously is $|P3 - P1|$. The left ringwidth is $|P3' - P3|$, where $P3' = P1 + \text{fixed ringwidth}$. The left ring measurement is calculated as $|\max(L1 - L2) - \min(L1 - L2)| * |P3' - P3|$, where the maximum and minimum of the difference between reference and processed are computed over the left ring support between $P3'$ and $P3$. The same quantities are computed on the right side of the edge.
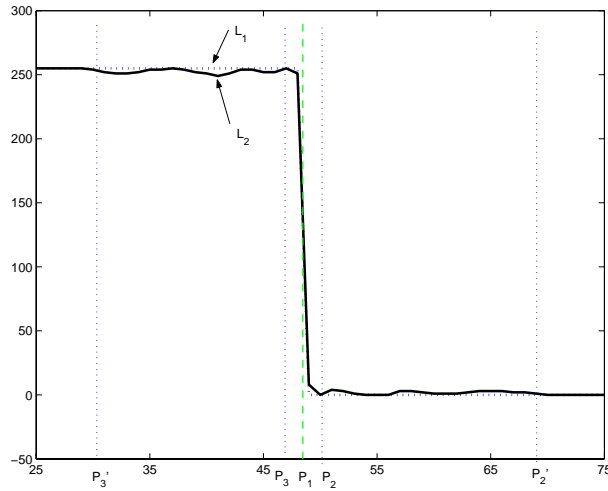

Fig. 5. The dotted line $L1$ is one row of the original image. The solid line $L2$ is the same row of the JPEG2000 coded image. Ringing can be observed around the edge at $P1$ between $P3$ and $P3'$ as well as $P2$ and $P2'$.

3 Experiments and Results

To corroborate the perceptual relevance of our metrics, we carried out two sets of subjective experiments. We asked ten expert viewers to evaluate – in separate sessions – the blur and the ringing perceived in a set of test images. The images are shown in random order, and the observers are asked to quantify the amount of the respective distortion for each image on a scale from 0 (no distortion visible) to 10 (maximum distortion). The average observer ratings are then compared to the predictions of the blur and ringing metrics. Finally, in Section 3.3 we use our metrics to predict the overall quality of JPEG2000-coded images and evaluate the prediction performance with the help of the LIVE Image Quality Assessment Database [15].

3.1 Perceived Blur

We consider the five 24-bit color images of size 768×512 shown in Figures 6 and 7(a). Blur is induced in two ways:

- The images are compressed in JPEG2000 with five different compression ratios $C_R \in \{40, 80, 120, 160, 200\}$, yielding 25 test images.
- The images are filtered with a Gaussian filter with five different standard deviations $\sigma \in \{0.4, 0.8, 1.2, 1.6, 2\}$ pixels, yielding another 25 test images.

We thus obtain a total of 55 test images (including the originals). Figures 7(b,c) show examples of maximum JPEG2000 compression and maximum Gaussian blur, respectively.
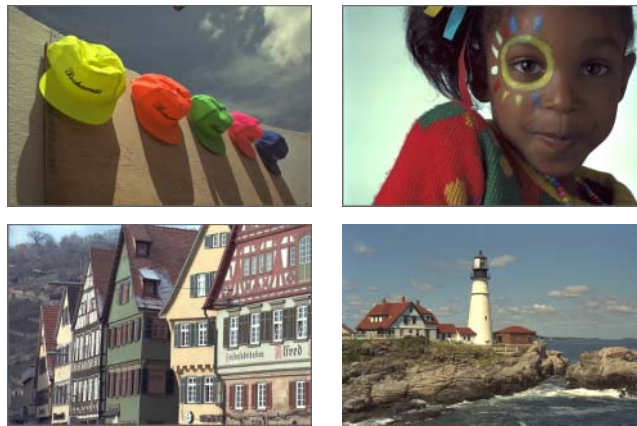

Fig. 6. Test images.

(a) Original image.

(b) Maximum JPEG2000 compression ($C_R = 200$).

(c) Maximum Gaussian blur ($\sigma = 2$).

Fig. 7. Motocross test image demonstrating the maximum distortion levels.

Figure 8 illustrates the behavior of the blur metric across distortion levels. The strong linear relationship is consistent for all the test images.

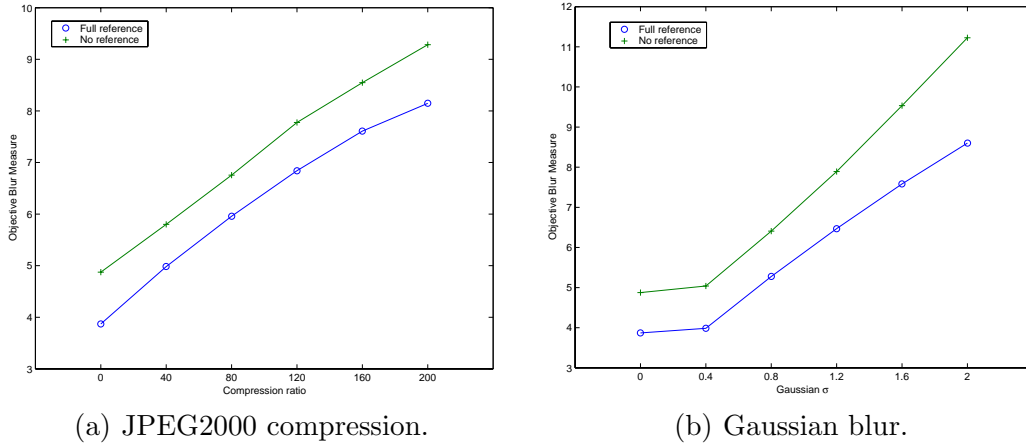(a) JPEG2000 compression.   (b) Gaussian blur.

Fig. 8. Behavior of blur metric for the motocross test image. (a) Blur measurement versus compression ratio; (b) blur measurement versus standard deviation of the Gaussian blurring filter.

Figure 9 illustrates the correlation between the subjective blur ratings and the proposed full-reference and no-reference blur metrics. We obtain 87% linear correlation and 85% rank-order correlation between our full-reference blur metric and perceived blur. For the no-reference blur metric, the correlations decrease to 73% and 81%, respectively (see also Table 1 below). This is mainly due to the problem of reliably detecting the edges in the processed image: as blur increases, the number of edges found by the Sobel filters goes down, which reduces the number of local blur measurements. This is one of the weak points of the NR metric, and using a more advanced edge detection method would certainly make it more robust. However, low complexity was one of our prime objectives in the design of these metrics.

In general, the difficulty lies mainly in predicting the perceived blur for two distinct blur sources (Gaussian filtering and JPEG2000 compression) with a common metric, as can be seen from the plots. If we analyze the metrics' predictions for these two sets separately, we can obtain correlations as high as 98%. The additional artifacts introduced by JPEG2000 compression change the observers' perception of blur with respect to the Gaussian case and thus affect the overall prediction performance of our metrics.

3.2   Perceived Ringing

Here we extend the original test set from the previous section to include the additional images shown in Figure 10 for a total of 9 original images. They are compressed in JPEG2000 with the same five compression ratios as in
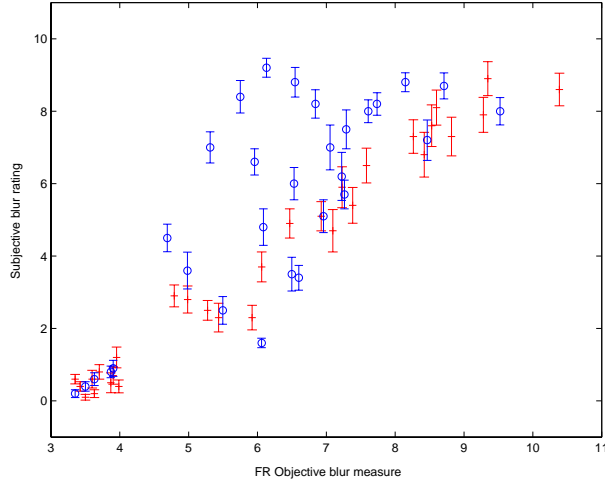
(a) Full-reference blur metric.
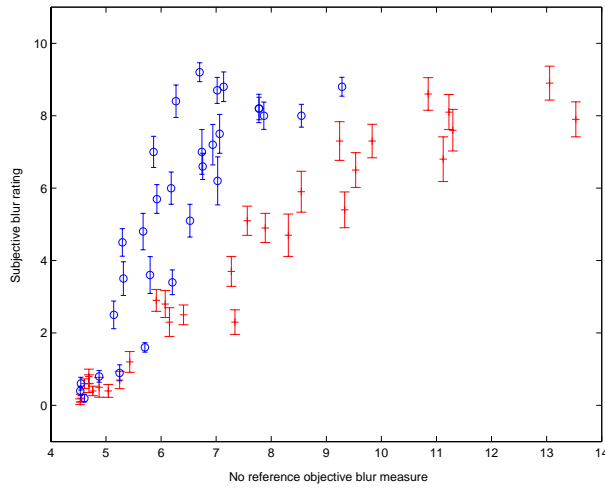
(b) No-reference blur metric.

Fig. 9. Error-bar plots with 95% confidence intervals of subjective blur ratings versus objective blur measurements for Gaussian filtered images (small dots) and JPEG2000 coded images (open circles).

Section 3.1, namely $C_R \in \{40, 80, 120, 160, 200\}$. We thus obtain a total of 54 test images, including the originals.

Figure 11 illustrates the subjective ring ratings versus the perceptual full-reference ringing metric with correlations of approximately 85%. The lower correlations (compared to the blur metric) can partly be explained by the fact that the viewers found it more difficult to evaluate the ringing artifacts in the JPEG2000 coded images. Furthermore, the effects of ringing are not always as well-behaved as in Figure 5, which affects the ringing measurements, and also thwarted our efforts to use the metric in the no-reference case.

The correlations between the subjective blur/ringing ratings and the proposed blur/ringing metrics are summarized in Table 1.

11

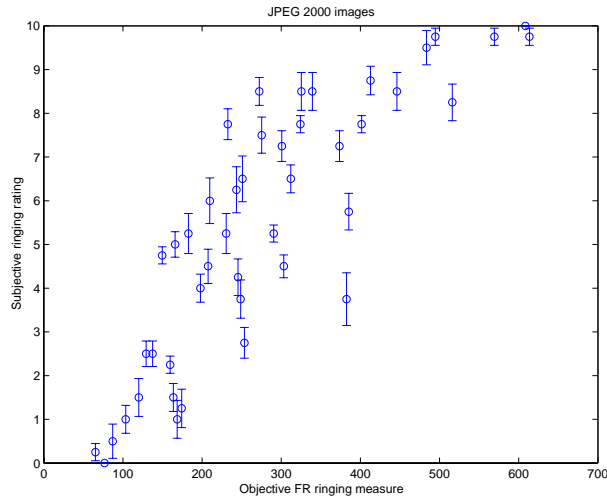Fig. 10. Additional test images for the ringing experiment.

Fig. 11. Error-bar plot with 95% confidence intervals of subjective ring ratings versus the full-reference perceptual ringing measurement for JPEG2000 coded images.

Correlations:	Linear	Rank-order
Full-reference blur	87%	85%
No-reference blur	73%	81%
Full-reference ringing	85%	86%

Table 1
Correlation between average observer ratings and the proposed perceptual blur and ringing metrics.

3.3 Perceived Quality of JPEG2000 Images

In addition to the artifact-specific experiments for blur and ringing described above, we also test the performance of our metrics as a predictor of overall perceived image quality. For this we use the subjectively rated JPEG2000-coded images from LIVE Image Quality Assessment Database [15], which was

made available recently by the University of Texas at Austin.

The test images in this database were created by compressing 29 RGB color images (typically of size 768×512 pixels) using Kakadu's JPEG2000 encoder. Compression ratios range from 7.5 to 800, yielding a total of 169 compressed images. The subjective experiments were conducted in two separate sessions with 29 and 25 observers, respectively; the original uncompressed images were included in both. Observers provided their quality ratings on a continuous linear scale from 1 (lowest quality) to 100 (highest quality), which was marked with the adjectives "Bad", "Poor", "Fair", "Good" and "Excellent". Refer to [15] for more information about the experiments.

We screened the subjective ratings for outliers according to ITU-R Rec. BT.500 [16]. For our analysis, we combined the data from the two test sessions and computed the mean opinion scores (MOS) and the corresponding 95% confidence intervals. Thanks to the large number of observers, the average confidence interval size is only 4.2 (on the 1-100 scale).

As shown in Figure 12, PSNR is already an excellent predictor of perceived quality for this database: the correlation with MOS is about 91%. These good results can be attributed largely to the fact that the database contains exclusively images created with a single type of encoder (JPEG2000) and thus only varying degrees of the same distortions. Note the saturation of the scatter plot towards high PSNR – this indicates that the database includes a number of compressed images in which subjects were unable to discern any quality degradation.
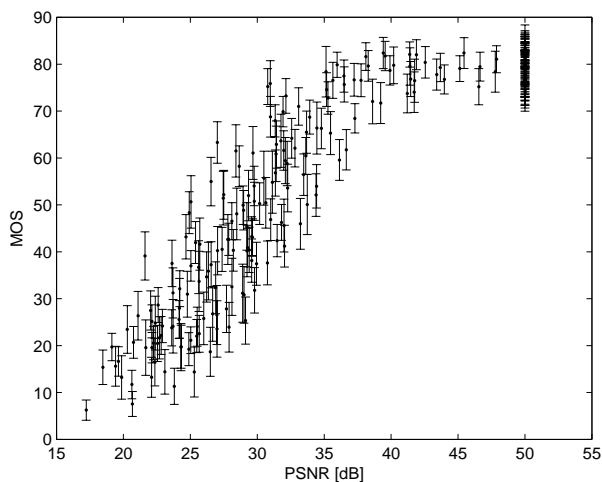

Fig. 12. Subjective MOS versus PSNR. The error bars indicate the 95% confidence intervals of the subjective ratings.

Combining our full-reference metrics for blur and ringing to a full-reference quality metric, we achieve a slight outperformance of PSNR (see Table 2 below). However, given the good prediction performance of PSNR in this exam-

ple, which is very close to the average correlation between individual subjects and MOS, it would be difficult to justify using any kind of more complex FR metric. We therefore focus on a no-reference solution based on the NR blur metric introduced in Section 2.1 above. More specifically, its MOS prediction is a simple non-linear transformation of the measured blur.

To evaluate its prediction performance, we separate the test images into a training set and a test set, using 100 different random divisions of the dataset. Figure 13 shows the results for our no-reference quality metric with the parameters obtained in the training. The saturation in the high quality regime is very similar to the behavior of PSNR. It achieves correlations of around 85% with MOS on the test sets, which is quite a good prediction performance for an NR metric.
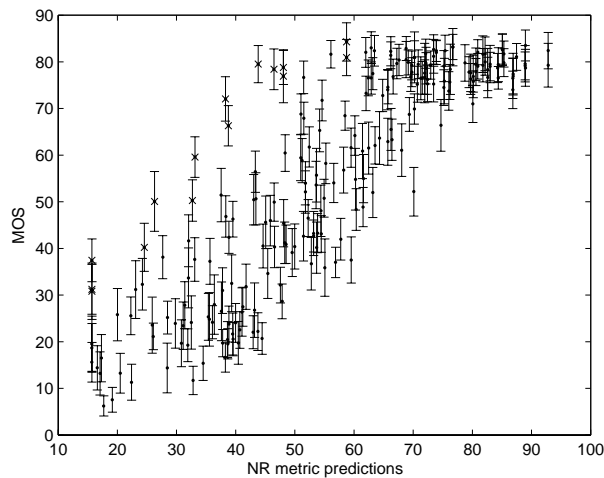

Fig. 13. Subjective MOS versus NR quality metric. The error bars indicate the 95% confidence intervals of the subjective ratings. Crosses denote images with very small depth of field (see text).

The most significant outliers are due to two specific pictures, namely one close-up and one macro shot with very small depths of field (they are marked with crosses in Figure 13). Since our blur metric does not distinguish between blur as a compression artifact and any other blur in the image, its MOS predictions for these images are too low in comparison to the observers' ratings, who do not consider this type of blur a degradation of quality. In one form or another, this problem is intrinsic to any no-reference metric. An added detector for distinguishing central objects from the potentially blurred background could help alleviate this problem when using our metric for the assessment of compression artifacts. In fact, when these two images are removed from the test set, the prediction performance of our NR metric approaches that of PSNR. All these results are summarized in Table 2.

We can also compare these results with an NR quality metric for JPEG2000-coded images described in [17], which is based on a statistical model for

	Linear correlation	Rank-order correlation	Prediction error
PSNR	91%	92%	9.7
Full-reference metric	94%	93%	9.5
No-reference metric	86%	84%	12.1
NR metric w/o outliers	90%	88%	10.1

Table 2

Prediction performance of the proposed quality metrics. The bottom row refers to the exclusion of the images with very small depth of field (see text).

wavelet coefficients and their quantization. Its design allows it to analyze the JPEG2000 image quality without decoding. This metric was evaluated using the same database, albeit with a slightly different computation of the mean ratings [17]. Its predictions have an RMSE of 9.8; its correlations are not reported, unfortunately. Since this metric only looks for compression artifacts, it does not suffer from the problem with images with a small depth of field. On the other hand, it cannot be used for images with blur coming from other sources than JPEG2000 compression.

On a final note, the bitrate of the encoded images alone is just as good an estimate of MOS as PSNR for the given database, and could thus be used for no-reference quality prediction here as well.

4   Conclusions

We presented a full-reference and a no-reference metric for perceived blur as well as a full-reference metric for perceived ringing. The metrics are of very low computational complexity and are shown to be in good agreement with observer ratings obtained in subjective experiments. Potential applications of such metrics include source coding optimization and network resource management. Future research will focus on the measurement of ringing without a reference, the consideration of color, and other types of perceptual artifacts.

References

[1] M. Yuen, H. R. Wu, A survey of hybrid MC/DPCM/DCT video coding distortions, Signal Processing 70 (3) (1998) 247–278.

[2] S. Winkler, A. Sharma, D. McNally, Perceptual video quality and blockiness metrics for multimedia streaming applications, in: Proceedings of the International Symposium on Wireless Personal Multimedia Communications, Aalborg, Denmark, 2001, pp. 547–552.

[3] S. Winkler, R. Campos, Video quality evaluation for Internet streaming applications, in: Proceedings of SPIE Human Vision and Electronic Imaging, Vol. 5007, Santa Clara, CA, 2003, pp. 104–115.

[4] S. Winkler, F. Dufaux, Video quality evaluation for mobile applications, in: Proceedings of SPIE Visual Communications and Image Processing, Vol. 5150, Lugano, Switzerland, 2003.

[5] M. Rabbani, R. Joshi, An overview of the JPEG2000 still image compression standard, Signal Processing: Image Communication 17 (1) (2002) 3–48.

[6] D. S. Taubman, M. W. Marcellin, JPEG2000: Image Compression Fundamentals, Standards and Practice, Kluwer Academic Publishers, 2002.

[7] P. Marziliano, F. Dufaux, S. Winkler, T. Ebrahimi, A no-reference perceptual blur metric, in: Proceedings of the International Conference on Image Processing, Vol. 3, Rochester, NY, 2002, pp. 57–60.

[8] R. L. Lagendijk, J. Biemond, Basic methods for image restoration and identification, in: A. Bovik (Ed.), Handbook of Image and Video Processing, Academic Press, 2000, Ch. 3.5, pp. 125–139.

[9] V. Kayargadde, J.-B. Martens, Perceptual characterization of images degraded by blur and noise: Model, Journal of the Optical Society of America A 13 (6) (1996) 1178–1188.

[10] J. H. Elder, S. W. Zucker, Local scale control for edge detection and blur estimation, IEEE Trans. Pattern Analysis and Machine Intelligence 20 (7) (1998) 699–716.

[11] A. S. Carasso, Linear and nonlinear image deblurring: A documented study, SIAM Journal on Numerical Analysis 36 (6) (1999) 1659–1689.

[12] D. Kundur, D. Hatzinakos, Blind image deconvolution, IEEE Signal Processing Magazine 13 (1996) 43–64.

[13] S. Yang, Y. H. Hu, T. Q. Nguyen, D. L. Tull, Maximum likelihood parameter estimation for image ringing-artifact removal, IEEE Trans. Circuits and Systems for Video Technology 11 (8) (2001) 963–973.

[14] Z. Yu, H. R. Wu, S. Winkler, T. Chen, Vision-model-based impairment metric to evaluate blocking artifacts in digital video, Proceedings of the IEEE 90 (1) (2002) 154–169.

[15] H. R. Sheikh, A. C. Bovik, L. Cormack, Z. Wang, LIVE image quality assessment database, http://live.ece.utexas.edu/research/quality (2003).

[16] ITU-R Recommendation BT.500-11, Methodology for the subjective assessment of the quality of television pictures, International Telecommunication Union, Geneva, Switzerland (2002).

[17] H. R. Sheikh, Z. Wang, L. Cormack, A. C. Bovik, Blind quality assessment for JPEG2000 compressed images, in: Proceedings of the Asilomar Conference on Signals, Systems and Computers, Vol. 2, Pacific Grove, CA, 2002, pp. 1735–1739.