Multiobjective Optimization of Multiple Scale Visual Quality Processing

Ulrich Engelke and Hans-Jürgen Zepernick

Blekinge Institute of Technology
PO Box 520, SE-372 25 Ronneby, Sweden
{ulrich.engelke, hans-jurgen.zepernick}@bth.se

Abstract— In this paper, multiobjective optimization is applied to determine the optimal feature weights for a multi-resolution image quality metric. The optimization is conducted with respect to two aims; maximization of quality prediction accuracy and generalisation ability to unknown images. A thresholding of the optimal weights is applied to further reduce computational complexity of the quality metric. A detailed discussion of the trade-off between prediction performance and complexity with respect to the threshold is provided. The metric design and evaluation is supported using mean opinion scores from subjective quality experiments that we conducted in two independent laboratories. Comparison to other state of the art quality metrics reveals the ability of the proposed metric to accurately predict subjective image quality perception.

I. Introduction

In wireless image communication, lossy source coding and error-prone channels induce a wide range of artifacts potentially resulting in a degradation of received image quality. Objectively measuring these artifacts and determining their impact on visual perception is thus a desirable, albeit difficult task. Metrics must not rely on the original image, since it is not available at the receiver. Hence, a metric needs to base its quality prediction either solely on the received image or additionally make use of extracted low-bandwidth features from the transmitted image. The widely established peak signal-to-noise ratio (PSNR) is not applicable here since it depends on a reference image to be available for the quality assessment. Furthermore, PSNR is based on a pixel-by-pixel comparison between two images whereas the human visual system (HVS) is adapted to extraction of structural information instead [1]. For these reasons, PSNR is not suitable for perceptual quality evaluation in wireless imaging.

A favourable image quality metric should hence be based on measures that are more closely related to characteristics of the HVS. It is well known that the HVS is adapted to visual information processing at different scales [2]. This evolutionary adaption arises from the nature of the environment around us, which contains objects of all different sizes. Objects may also be located at varying distances changing the size of their appearance. The multiple scale processing in the HVS suggests to also objectively analyse images at multiple scales in order not to miss information in the image due to single resolution analysis. In relation to image quality, one may suspect that distortions are also perceived differently at given scales. The Gaussian pyramid is a convenient and computationally effi-cient multi-resolution image representation that mirrors very well the multiple scales of processing in the HVS [3]. It has therefore been widely used before to perform image analysis at multiple scales. In our earlier work [4], we have utilised the Gaussian pyramid decomposition to significantly enhance prediction accuracy of a single resolution metric.

In contrast to our previous work, in which weights were obtained separately for each structural feature and Gaussian pyramid level, the aim of this paper is to determine simultaneously the optimal weights of all features at all scales using multiobjective optimization [5]. In this way, interdependencies between artifacts at different scales and their impact on perceptual image quality can be accounted for. It further allows to identify the weights that are of negligible importance at a given scale and can thus be discarded. Furthermore, rather than just considering a training set of images, as done before [4], we utilise a set of unknown images to validate the metrics prediction performance and to prevent overfitting of the metric on the training data. Therefore, the optimization is done with respect to two objectives; maximisation of quality prediction accuracy and generalisation to unknown images. The metric performance is evaluated and compared to other state of the art perceptual quality metrics. An exponential mapping may further be applied to increase quality prediction performance by accounting for non-linear quality processing in the HVS.

The paper is organised as follows. Section II reports about two subjective image quality experiments. Section III discusses multiple scale objective image quality. Section IV describes the multiobjective optimization method used to determine the optimal weights. Finally, numerical results are discussed in Section V and conclusions are drawn in Section VI.

II. Subjective Image Quality Experiments

Image quality can most precisely be judged by humans themselves. That is why mean opinion scores (MOS) obtained in subjective experiments are considered to be the most precise measures of perceptual quality [6]. Subjective quality assessment is usually not applicable though when real-time processing is required. However, MOS obtained in subjective experiments may be used to support the design of objective image quality metrics which can in turn be applied for automated quality assessment. For this reason, we conducted subjective image quality experiments in two independent laboratories. The first experiment was done at

Blekinge Institute of Technology (BIT) in Ronneby, Sweden, and the other at the Western Australian Telecommunications Research Institute (WATRI) in Perth, Australia [7]. In each experiment 30 non-expert viewers were asked to judge the quality of distorted images as compared to the corresponding reference images. The experiment procedures were designed according to ITU-R Rec. BT.500-11 [8].

The distorted images used in the experiments were obtained based on a set $\mathcal{I}_R$ of 7 reference monochrome images of dimensions 512×512 pixels which was chosen to account for different textures and complexities. The images were encoded into Joint Photographic Experts Group (JPEG) format. A simple simulation model of a wireless channel was used to create two sets $\mathcal{I}_B$ and $\mathcal{I}_W$ of 40 distorted images each, to be used in the BIT and WATRI experiments, respectively. In particular, blocking, blur, ringing, and intensity masking artifacts were observed in the images in different degrees of severity. The experiments at BIT and WATRI resulted in two respective sets of MOS, $\mathcal{M}_B$ and $\mathcal{M}_W$. For the metric design and validation we randomly created two sets of images, a training set $\mathcal{I}_T$ and a validation set $\mathcal{I}_V$. The training set contains 60 images, 30 from each $\mathcal{I}_B$ and $\mathcal{I}_W$. The validation set contains the residual 20 images. Similarly, we created corresponding MOS training set $\mathcal{M}_T$ and validation set $\mathcal{M}_V$.

III. Multiple Scale Objective Image Quality

A. Gaussian Pyramid Generation

The Gaussian pyramid is a convenient multi-resolution image representation that very well mirrors the multiple scales of processing in the HVS [3]. An efficient iterative algorithm for the pyramid generation is described in [9] and summarised here. The pyramid consists of L levels with the base level image g_0 being the original image in full resolution 512×512. The higher level images g_l, $l = 1, \cdots, L - 1$, are low-pass filtered and sub-sampled versions of the underlying images. The low-pass filtering is performed using a generating kernel $\phi(u, v)$ of size 5×5. The size has been chosen with respect to filtering performance and low computational cost. Sub-sampling is done by a factor of two. Therewith, each image g_l is obtained from its predecessor g_{l-1} as

$$g_l(x, y) = \sum_{u=-2}^{2} \sum_{v=-2}^{2} \phi(u, v) \cdot g_{l-1}(2x + u, 2y + v). \quad (1)$$

For simplicity, the generating kernel is made separable

$$\phi(u, v) = \phi(u) \cdot \phi(v). \quad (2)$$

Furthermore, the one-dimensional patterns $\phi(u)$ and $\phi(v)$ have to be normalised

$$\sum_{u=-2}^{2} \phi(u) = \sum_{v=-2}^{2} \phi(v) = 1 \quad (3)$$

and must be symmetric

$$\phi(u) = \phi(-u) \quad \text{and} \quad \phi(v) = \phi(-v). \quad (4)$$

The image pixel density is reduced by four from level g_{l-1} to level g_l. Hence, an additional constraint called equal contribution requires all pixels at a given level to contribute the same total weight of $1/4$. The above constraints are satisfied when

$$
\begin{aligned}
\phi(0) &= \alpha \\
\phi(1) &= \phi(-1) = \frac{1}{4} \\
\phi(2) &= \phi(-2) = \frac{1}{4} - \frac{\alpha}{2}
\end{aligned}
\quad (5)
$$

where $\alpha = 0.4$. It should be noted that the algorithm was slightly modified to fit our original image size of 512×512.

For the multi-resolution analysis we considered a maximum of six Gaussian pyramid levels. Taking the original image resolution and the sub-sampling of factor two into account, the highest level in the pyramid has a resolution of 16×16. Images of higher levels were not taken into account since the feature extraction algorithms do not work on such a small number of pixels. Figure 1 shows an illustrative example of the multi-resolution levels for a reference image, a distorted image, and a difference image. For better visualisation the downsampled images were expanded to original size using the pixel replication technique [10]. As one would expect, the Gaussian filter introduces some blur into higher level images.

B. Structural Feature Extraction

To obtain information about structural degradations in the pyramid levels, we extract the following five features $\tilde{f}_i$:

$$
\begin{aligned}
\tilde{f}_1 &\triangleq \quad \text{Blocking [11]} \\
\tilde{f}_2 &\triangleq \quad \text{Blur [12]} \\
\tilde{f}_3 &\triangleq \quad \text{Edge-based image activity [13]} \\
\tilde{f}_4 &\triangleq \quad \text{Gradient-based image activity [13]} \\
\tilde{f}_5 &\triangleq \quad \text{Intensity masking}
\end{aligned}
$$

The features were selected with respect to the artifacts observed in the distorted images (see Section II). Here, the two image activity (IA) metrics were found to account well for measuring ringing artifacts. We define three feature matrices for the reference (R), training (T), and validation (V) set of images as follows

$$
\begin{aligned}
\widetilde{\mathbf{F}}_R &= [\tilde{f}_{i,j,l}^{(R)}]_{I \times J \times L} \\
\widetilde{\mathbf{F}}_T &= [\tilde{f}_{i,k,l}^{(T)}]_{I \times K \times L} \\
\widetilde{\mathbf{F}}_V &= [\tilde{f}_{i,m,l}^{(V)}]_{I \times M \times L}
\end{aligned}
\quad (6)
$$

where $\tilde{f}_{i,j,l}^{(R)}$, $\tilde{f}_{i,k,l}^{(T)}$, and $\tilde{f}_{i,m,l}^{(V)}$, respectively, denote the ith feature measure of the jth, k^{th}, and m^{th} image in $\mathcal{I}_R$, $\mathcal{I}_T$, and $\mathcal{I}_V$ at the lth level in the Gaussian pyramid. Also, the dimensions of these matrices relate to the number of features, $I = 5$, the number of reference images, $J = 7$, the number of training images, $K = 60$, the number of validation images, $M = 20$, and the number of pyramid levels $L = 6$. A partitioned matrix containing the features of the total of $N = (J + K + M) \times L = 609$ images may be introduced as

$$\widetilde{\mathbf{F}} = [\tilde{f}_{i,n,l}]_{I \times N \times L} = [\widetilde{\mathbf{F}}_R | \widetilde{\mathbf{F}}_T | \widetilde{\mathbf{F}}_V] \quad (7)$$
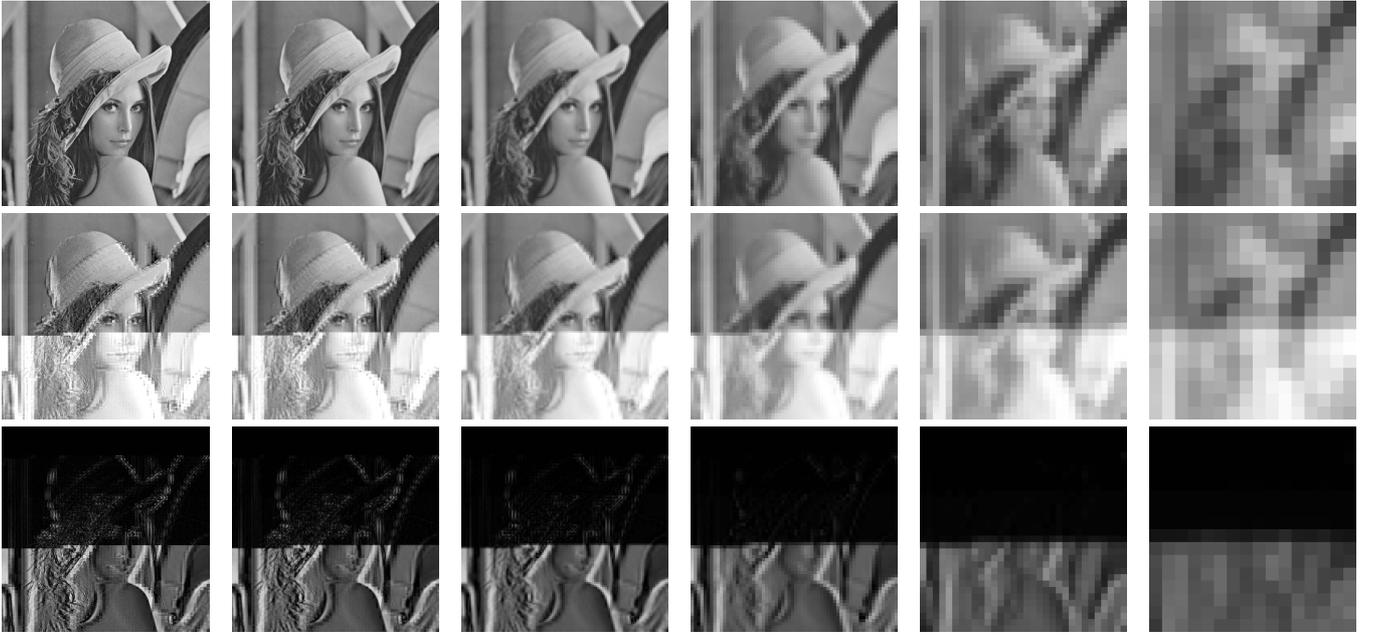
Fig. 1. First six levels of Gaussian pyramid of the image Lena (from top to bottom: reference image, distorted image, difference image; from left to right: g_0 (512×512), g_1 (256×256), g_2 (128×128), g_3 (64×64), g_4 (32×32), g_5 (16×16)).

In order to obtain a defined and finite feature space, the feature measures were normalised into an interval using an extreme value normalisation [14] as follows

$$f_{i,n,l} = \frac{\tilde{f}_{i,n,l} - \min\limits_{\substack{n=1,\cdots,N \\ l=1,\cdots,L}} \{\tilde{f}_{i,n,l}\}}{\delta_i}, \qquad i = 1,\cdots,I \quad (8)$$

where the denominator is computed as

$$\delta_i = \max\limits_{\substack{n=1,\cdots,N \\ l=1,\cdots,L}} \{\tilde{f}_{i,n,l}\} - \min\limits_{\substack{n=1,\cdots,N \\ l=1,\cdots,L}} \{\tilde{f}_{i,n,l}\} \quad (9)$$

and as a consequence, we have $\forall i, n, l : 0 \leq f_{i,n,l} \leq 1$.

The absolute difference between the normalised features of the distorted and reference image may then be used to quantify structural changes as follows

$$\begin{array}{rcl} \Delta f_{i,k,l}^{(T)} & = & |f_{i,k,l}^{(T)} - f_{i,j,l}^{(R)}| \\ \Delta f_{i,m,l}^{(V)} & = & |f_{i,m,l}^{(V)} - f_{i,j,l}^{(R)}| \end{array} \quad (10)$$

where $f_{i,j,l}^{(R)}$ are chosen according to the distorted image. Finally, we define delta-feature matrices for both training and validation set

$$\begin{array}{rcl} \Delta\mathbf{F}_T & = & [\Delta f_{i,k,l}^{(T)}]_{I \times K \times L} \\ \Delta\mathbf{F}_V & = & [\Delta f_{i,m,l}^{(V)}]_{I \times M \times L} \end{array} \quad (11)$$

C. Feature Pooling Across Pyramid Levels

Considering the above, a total of $I \times L = 5 \times 6 = 30$ features is available for pooling into the objective quality metric. We have shown earlier [4] that it is crucial to preserve the information about degradation for each feature in each level, rather than pooling the reference features and distorted features across all levels first, before computing an overall difference. Therefore, we define a simple pooling function as a weighted sum of all delta-features which defines the image quality metric as follows

$$\Theta = \sum_{i=1}^{I} \sum_{l=1}^{L} w_{i,l} \cdot \Delta f_{i,l} \quad (12)$$

Here, the weights $w_{i,l}$ correspond to the perceptual relevance of each of the feature metrics at a given Gaussian pyramid level. The aim of Section IV will be to determine the optimal weights $w_{i,l}$ for all delta-features $\Delta f_{i,l}$. The pooling function (12) has intentionally been kept simple to put no restrictions on the multiobjective optimization.

D. Exponential Mapping to Predicted MOS

Perception of visual quality in the HVS is not a linear function of the distortions observed in images. This relates to the phenomenon, that distinction between subjective quality levels is more pronounced in ranges of small distortions and less pronounced in ranges of strong distortions. We found [7] that an exponential function very well approximates this non-linear relationship. Therefore, we apply an exponential mapping of the metric value Θ onto predicted MOS MOS_Θ as follows

$$MOS_\Theta = a\,e^{b\,\Theta} \quad (13)$$

The parameters a and b are obtained through non-linear curve fitting between Θ and MOS $\mathcal{M}_T$. As will be shown in Section V, such an exponential mapping does not only enhance the prediction performance of the Θ metric proposed in this paper but can also be applied to other image quality metrics to improve quality prediction performance.

IV. MULTIOBJECTIVE WEIGHT OPTIMIZATION

A. Definition of Multiple Objectives

Optimization in general is concerned with minimisation of an objective function, subject to a set of decision variables. However, system performance cannot always be quantified by a single number. Thus, multiobjective optimization (MOO) is concerned with optimization of multiple, often conflicting objectives [5]. Two objectives are said to be conflicting when a decrease in one objective leads to an increase in the other. A MOO problem can be transformed into a single objective optimization, for instance by defining an objective as a weighted sum of multiple objectives. However, it is recommended to preserve the full dimensionality of the MOO problem [15]. The aim is then to find the optimal compromise between the objectives where system design aspects need to be taken into account to decide for the best trade-off solution [5].

Considering the above, we conduct a MOO based on a decision matrix $\mathbf{w} \in \mathbb{W} \subset \mathbb{R}^{I \times L}$ containing the feature weights $w_{i,l}$. The range of the weights in the decision space $\mathbb{W}$ is constrained to $w_{i,l} \in [0, 1]$. The MOO is conducted with respect to two objectives: a) maximising image quality prediction accuracy O_A and b) maximising generalization performance O_G. Objective O_A defines the metrics ability to predict MOS with minimal error and is measured as the Pearson linear correlation between metric Θ and MOS $\mathcal{M}$ on the training set

$$\rho_P = \frac{\sum_k (\Theta_k - \overline{\Theta})(\mathcal{M}_k - \overline{\mathcal{M}})}{\sqrt{\sum_k (\Theta_k - \overline{\Theta})^2} \sqrt{\sum_k (\mathcal{M}_k - \overline{\mathcal{M}})^2}} \quad (14)$$

where $\overline{\Theta}$ and $\overline{\mathcal{M}}$, respectively, denote the mean values of Θ and $\mathcal{M}$. Optimizing the weights using only objective O_A would likely overtrain the metric, meaning, it would work very well on the training set but not on a set of unknown images. Therefore, objective O_G defines the metrics ability to perform quality prediction on a set of unknown images. We compute it as the absolute difference of ρ_P on training and validation set as follows

$$\Delta \rho_P = |\rho_{P,T} - \rho_{P,V}| \quad (15)$$

We thus define the objective vector as

$$\mathbf{O}(w) = \begin{pmatrix} O_A(w) \\ O_G(w) \end{pmatrix} = \begin{pmatrix} -|\rho_{P,T}| \\ \Delta \rho_P \end{pmatrix} \quad (16)$$

The decision matrix $\mathbf{w}$ is evaluated by assigning it an objective vector $\mathbf{O}$ in the objective space $\mathbb{O}$: $\mathbb{W} \rightarrow \mathbb{O} \subset \mathbb{R}^2$. An overview of the MOO system is provided in Fig. 2.

B. Goal Attainment Method

We determine the optimal solution using the goal attainment method [16]. Here, goals $\mathbf{O}^* = (O_A^* \quad O_G^*)^T$ are specified, which can be interpreted as the desired level of the corresponding objectives. This requires sufficient intuitive understanding of the problem to know what values one would like to attain for each of the objectives. We define the MOO problem as

$$P_G : \begin{cases} min & z \\ s.t. & \mathbf{O}(w) - \boldsymbol{\lambda} \cdot z \leq \mathbf{O}^* \end{cases} \quad (17)$$
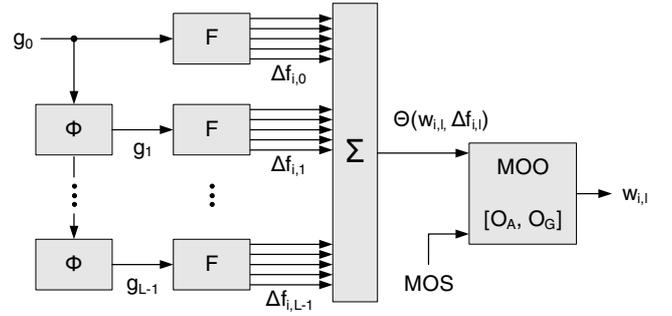
Fig. 2. System overview including Gaussian pyramid decomposition (Φ), feature extraction (F), pooling (Σ), and multiobjective optimization (MOO) with respect to two objectives O_A and O_G resulting in optimal weights $w_{i,l}$.

where the magnitude of $\boldsymbol{\lambda} = (\lambda_A \quad \lambda_G)^T$ determines how close the objectives $(O_A(w) \quad O_G(w))^T$ are to the goals $(O_A^* \quad O_G^*)^T$. It is typically set to the absolute value of the goals

$$\lambda_A = |O_A^*| \quad \text{and} \quad \lambda_G = |O_G^*| \quad (18)$$

The quantity $\boldsymbol{\lambda} \cdot z$ corresponds to the degree of under- or overattainment of the goals $\mathbf{O}^*$. With regards to results from our earlier metric design [17], we define the goals as $O_A^* = -0.9$ and $O_G^* = 0.0001$. If we relax the generalization goal O_G^* we can significantly increase the prediction accuracy $\rho_{P,T}$ on the training set. However, this would be done at the cost of prediction accuracy $\rho_{P,V}$ on the validation set. Since we want to avoid such overfitting of the metric to the training data, the generalisation goal O_G^* is chosen to be very small.

V. DISCUSSION OF NUMERICAL RESULTS

A. Optimal Weights

The optimal weights obtained using the procedure outlined in Section IV are shown in Fig. 3. One can see that many of the 30 weights are either zero or very close to zero. This might relate to the fact that too many features would cause overfitting of the data to the training set. In general, a good approximation is that the number of images in the training set should be no less than some multiple (about 5) of the number of adaptive model parameters [18]. One can also see that in each level there seem to be one or two features dominating over the other features. This provides insight into which feature has the strongest impact on perceived quality at a given scale.

B. Thresholding for Feature Elimination

The small magnitudes of some of the weights in Fig. 3 suggests, that there are weights of small importance that might be eliminated. This in turn results in computation of less features and therewith savings in valuable computational complexity. Therefore, we introduce a thresholding as follows

$$w_{i,l} = 0 \quad \text{for} \quad w_{i,l} \leq \tau \quad (19)$$

The compromise between prediction performance and computational complexity of the metric with regards to different thresholds τ is discussed in the following section.

Threshold τ	Number of discarded features	Number of discarded levels	Θ		MOS_Θ			
			$\rho_{P,T}$	$\rho_{P,V}$	$\rho_{P,T}$	$\rho_{P,V}$	$\rho_{S,T}$	$\rho_{S,V}$
0	4	0	0.893	0.893	0.926	0.933	0.914	0.947
0.05	13	0	0.893	0.894	0.926	0.933	0.916	0.952
0.10	15	0	0.892	0.899	0.926	0.936	0.912	0.952
0.15	16	0	0.892	0.900	0.926	0.936	0.912	0.952
0.20	18	0	0.892	0.903	0.924	0.936	0.913	0.952
0.25	21	1	0.890	0.902	0.921	0.934	0.908	0.953
0.30	22	1	0.878	0.901	0.917	0.921	0.898	0.959
0.35	23	1	0.878	0.898	0.917	0.919	0.898	0.959
0.65	24	1	0.879	0.896	0.917	0.917	0.896	0.955
0.75	25	1	0.861	0.904	0.902	0.901	0.883	0.944
0.80	26	1	0.858	0.883	0.901	0.890	0.879	0.926
0.85	27	1	0.598	0.685	0.692	0.783	0.759	0.741

C. Quality Prediction Performance of Θ

In this section, the prediction performance of metric Θ (12) and the predicted MOS MOS_Θ are discussed. As recommended in [19], the prediction accuracy is measured in terms of Pearson linear correlation coefficient (14) and the prediction monotonicity using Spearman rank order coefficient

$$\rho_S = \frac{\sum_k (\Psi_k - \overline{\Psi})(\Upsilon_k - \overline{\Upsilon})}{\sqrt{\sum_k (\Psi_k - \overline{\Psi})^2}\sqrt{\sum_k (\Upsilon_k - \overline{\Upsilon})^2}} \qquad (20)$$

where Ψ and Υ are the ranks of an objective metric and MOS, respectively, and $\overline{\Psi}$ and $\overline{\Upsilon}$ denote the corresponding mean values. It should be noted, that ρ_S is the same for Θ and MOS_Θ in case of a monotonic decreasing prediction function, as is the case for an exponential function. The numerical results are presented in Table I for different thresholds τ. All values are provided for both training and validation set. Furthermore, for each threshold τ the number of discarded features and levels is shown. A discarded feature corresponds to a weight that has been set to zero by either MOO or thresholding. A level is discarded if all features within the level are discarded and if there is no higher level in the pyramid.

From Table I it can be seen that only 4 features are zero if no thresholding is used. As the threshold τ increases, more features are discarded from the metric computation. It should be observed that up to $\tau = 0.25$, the loss in prediction performance is minimal as compared to the number of discarded features. In particular, for $\tau = 0.25$ only 9 out of 30 features need to be computed. We also do not need to compute any feature in level g_5 and can thus discard this level. This means for $\tau = 0.25$, savings in computational complexity are large as compared to loss in prediction performance. However, if we further increase τ the loss in prediction performance as compared to the number of discarded features increases. Considering the above, a threshold of $\tau = 0.25$ seems to be the best compromise. Table II provides an overview of the residual weights for a threshold of $\tau = 0.25$. One can see that we can disregard the pyramid level g_5 and also the intensity masking feature $f_{5,l}$. Figure 4 shows the exponential curve fitting of

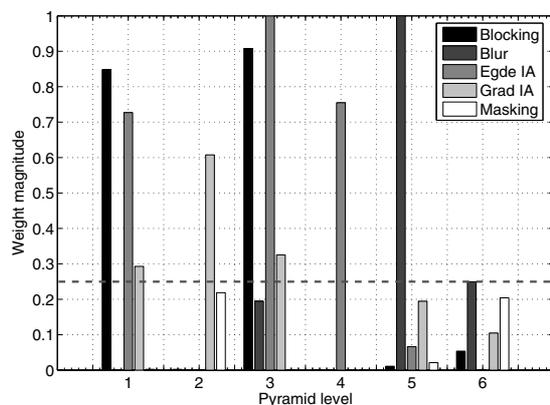
Fig. 3. Optimal feature weights for six Gaussian pyramid levels (the dashed line indicates a threshold $\tau = 0.25$).

Θ and MOS for $\tau = 0.25$ resulting in the prediction function $MOS_\Theta = 91.07\,e^{-1.9\,\Theta}$. The linear curve fitting in the same figure shows that MOS_Θ very well predicts the MOS.

D. Comparison to Other Image Quality Metrics

The results of this paper are in alignment with the results of our earlier work on multi-resolution image quality assessment [4] where we found that level g_5 of the pyramid is of low importance for high quality prediction accuracy. However, previously we only considered a training set of 40 images on which we obtained $\rho_{P,T} = 0.947$ for the predicted MOS. This was achieved though by computing a total of 25 features, most likely overfitting the metric to the training set [18]. In the current work, we achieve almost as good prediction accuracy on the training set while assuring generalization to unknown images, by computing only 9 features. Thus, the MOO achieves a reduction of feature computations by 64%.

For further comparison, we consider three perceptual image quality metrics, the structural similarity index (SSIM) [1], the visual information fidelity measure (VIF) [20], and a reduced-reference image quality assessment (RRIQA) [21].
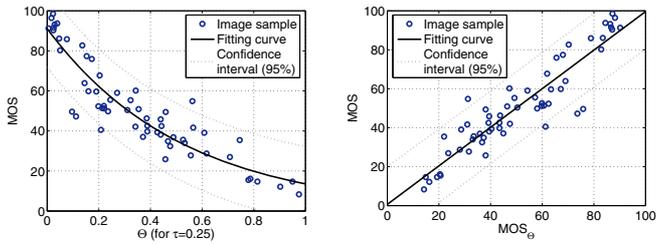
Fig. 4. Left: Exponential curve fitting of MOS vs Θ for $\tau = 0.25$; Right: Linear curve fitting of MOS vs MOS_Θ.

The two former metrics are full reference metrics and are thus not applicable for assessing image communication quality. However, these metrics are widely used and may serve as a benchmark. We also compare our metric to the well known PSNR. For all these metrics we performed non-linear curve fittings to obtain exponential prediction functions. The metrics and the corresponding predicted MOS are shown in Table III for both training and validation sets. For better comparison we provide Θ again for a threshold of $\tau = 0.25$.

One can see that the prediction accuracy and monotonicity of Θ outperforms the other metrics on both training and validation set. This is valid for both the actual metric and the predicted MOS. It is interesting to point out that Θ outperforms the other metrics even if we apply a threshold of $\tau = 0.8$, leaving only 4 residual features (see Table I). If we further increase the threshold to $\tau = 0.85$, the prediction performance of Θ drastically drops. This suggests that the blocking measure in base level g_0 (which is discarded here) has strong impact on the perceptual relevance of the quality metric. Finally, we observe that the exponential mapping also significantly improves prediction accuracy of both SSIM and VIF and can thus be applied to other quality metrics than Θ.

VI. CONCLUSIONS

We designed an image quality metric, based on multi-resolution extraction of structural features. Multiobjective optimization was applied to simultaneously determine optimal weights for all features at all resolutions. The optimal weights provide insight into the perceptual relevance of a feature at a given scale. A thresholding further reduced computational complexity while maintaining quality prediction performance. The metric was trained and validated using MOS from two independent subjective quality experiments. Comparison to other state of the art image quality metrics revealed the ability of the proposed metric to accurately predict subjective quality.

REFERENCES

[1] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. on Image Processing*, pp. 600–612, April 2004.
[2] B. A. Wandell, *Foundations of Vision*. Sinauer Associates, Inc., 1995.
[3] E. H. Adelson, C. H. Anderson, J. R. Bergen, P. J. Burt, and J. M. Ogden, "Pyramid methods in image processing," *RCA Engineer*, vol. 29, no. 6, pp. 33–41, 1984.
[4] U. Engelke and H. J. Zepernick, "Multi-resolution structural degradation metrics for perceptual image quality assessment," in *Proc. of Picture Coding Symposium*, Nov. 2007.
[5] C. M. Fonseca and P. J. Fleming, "Multiobjective optimization," in *Evolutionary Computation 2: Advanced Algorithms and Operations*, T. Back, Ed. Taylor & Francis, 2000, ch. 5, pp. 25–37.
[6] H. R. Wu and K. R. Rao (Ed.), *Digital Video Image Quality and Perceptual Coding*. CRC Press, 2006.
[7] T. M. Kusuma, H.-J. Zepernick, and M. Caldera, "On the development of a reduced-reference perceptual image quality metric," in *Proc. of Int. Conf. on Multimedia Communications Systems*, Aug. 2005, pp. 178–184.
[8] ITU, "Methodology for the subjective assessment of the quality of television pictures," Rec. BT.500-11, 2002.
[9] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. on Communications*, vol. 31, no. 4, Apr. 1983.
[10] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 2nd ed. Prentice Hall, 2002.
[11] Z. Wang, H. R. Sheikh, and A. C. Bovik, "No-reference perceptual quality assessment of JPEG compressed images," in *Proc. of IEEE Int. Conf. on Image Processing*, Sept. 2002, pp. 477–480.
[12] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "A no-reference perceptual blur metric," in *Proc. of IEEE Int. Conf. on Image Processing*, vol. 3, Sept. 2002, pp. 57–60.
[13] S. Saha and R. Vemuri, "An analysis on the effect of image features on lossy coding performance," *IEEE Signal Processing Letters*, pp. 104–107, May 2000.
[14] J.-R. Ohm, *Multimedia Communication Technology*. Springer, 2004.
[15] L. A. Zadeh, "Optimality and non-scalar-valued performance criteria," *IEEE Trans. on Automatic Control*, vol. 8, no. 1, pp. 59–60, Jan. 1963.
[16] F. W. Gembicki and Y. Y. Haimes, "Approach to performance and sensitivity multiobjective optimization: The goal attainment method," *IEEE Trans. on Automatic Control*, vol. 20, no. 6, pp. 769–771, Dec. 1975.
[17] U. Engelke and H. J. Zepernick, "Quality evaluation in wireless imaging using feature-based objective metrics," in *Proc. of IEEE Int. Symp. on Wireless Pervasive Computing*, Feb. 2007, pp. 367–372.
[18] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
[19] Video Quality Experts Group, "Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment, phase II," VQEG, Aug. 2003.
[20] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. on Image Processing*, vol. 15, no. 2, pp. 430–444, Feb. 2006.
[21] Z. Wang and E. P. Simoncelli, "Reduced-reference image quality assessment using a wavelet-domain natural image statistic model," in *Proc. of SPIE Human Vision and Electronic Imaging*, vol. 5666, Mar. 2005, pp. 149–159.