PERCEPTUAL IMPACT OF TRANSFORM COEFFICIENTS QUANTIZATION FOR ADAPTIVE LIFTING SCHEMES

Marco Cagnazzo, Béatrice Pesquet-Popescu

[1]TELECOM-ParisTech, TSI department. Paris, FRANCE

ABSTRACT

Adaptive wavelet transforms are a very useful tool for image and video compression. However, their intrinsic non-linear nature makes it difficult to estimate the effect that quantization has on the reconstructed image, when this operation is performed (as usual) in the transform domain.

In a previous work, we showed how a simple, non-perceptual metric such as mean squared error can be almost perfectly estimated in the transform domain even for these non-linear operators, provided that suitable weights are used. In this paper, we propose a perceptual distortion metric inspired by the concept of saliency map. The new metric should allow to estimate the image quality in the transform domain even for non-linear transforms, allowing an effective resource allocation for image compression.

Our experiments confirm that the proposed approach can be profitably used to drive a resource allocation algorithm such that the perceptual quality of the decoded image is improved.

1. INTRODUCTION

Wavelet transform (WT) is a very useful tool for image processing and compression. In particular, the lifting scheme (LS) implementation of WT was originally introduced by Sweldens [1] to design wavelets on complex geometrical surfaces, but at the same time it offers a simple way to build up both classic wavelet transforms and new ones.

The elements composing the lifting scheme are shown in Fig. 1. We call x the input signal, and y_{ij} the wavelet subbands. In particular, the first index determines the decomposition level ($i = 0$ being the first one), and the second index discriminates the channel ($j = 0$ for the low-pass or approximation signal, $j = 1$ for the high pass or detail signal). The input signal x is split into its even and odd samples, respectively called the approximation and the detail signal. Then, a prediction operator P is used in order to predict the odd samples of x from a linear combination of even samples. The prediction is removed from the odd samples in order to reduce their correlation with the even ones. Finally for the third block, the update operator U is chosen in such a way that the approximation signal y_{00} satisfies certain constraints, such as

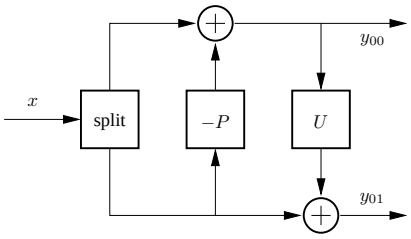

Fig. 1. Lifting scheme with a single lifting stage

preserving the average of the input or reducing aliasing. It is interesting to notice that, with a proper combination of lifting steps (prediction and update) it is possible to enhance a given transform by imposing new properties on the resulting decomposition (for example, more vanishing moments).

LS are very flexible while preserving the perfect reconstruction property, and this allows to replace linear filters by nonlinear ones. For example, LS with adaptive update [2] or adaptive prediction [3, 4] have been proposed in the literature, with the target of avoiding oversmoothing of object contours, and at the same time of exploiting the correlation of homogeneous regions by using long filters on them. The adaptivity makes it possible to use different filters over different parts of image. As a consequence, the resulting transform can be strongly non-isometric. This is a major problem for compression, since all most successful techniques rely on the distortion estimation in the transform domain, either explicitly like in EBCOT [5], or implicitly, like in the zero-tree based algorithms (EZW, SPIHT [6, 7]. Therefore, in order to efficiently use the adaptive lifting scheme for image compression, we need to estimate correctly the distortion from the transform coefficients. Usevitch showed that the energy of an uncorrelated signal (such as the quantization noise is supposed to be) can be estimated for generic linear wavelet filter banks [8]. We extended this approach to adaptive update LS (AULS) [9], and to adaptive prediction LS (APLS) [10] (in particular those inspired by the paper by Claypoole *et al.* [3]), obtaining satisfying results in term of distortion estimation and of rate-distortion (RD) performance improvement.

When non-isometric linear analysis is used, Usevitch [8] showed that the MSE in the original domain D is related to the MSE's D_{ij} of the wavelet subbands y_{ij} by the linear relation

$D = \sum_{ij} w_{ij} D_{ij}$. The weight w_{ij} is computed as norm of the reconstruction polyphase matrix columns for subband y_{ij}.

However APLSs (as well as AULSs) are nonlinear systems, therefore no polyphase representation of them exist. Our contribution in previous papers [9, 10] was to extended this approach to adaptive LS, and to show how to compute the weights w_{ij}. The error D is still obtained as a weighted sum of the subband errors, but now the weights depend on the input image, since the transform itself depends on it. In conclusion the proposed approach shows how to compute, in the transform domain, a metric which estimates the quantization noise MSE. This objective, non-perceptual distortion metric is then expressed as:

$$D_1 = \sum_{ij} w_{ij} d_{ij} \tag{1}$$

where d_{ij} is the MSE in the subband ij:

$$d_{ij} = \sum_{n,m} \left[y_{ij}(n,m) - \hat{y}_{ij}(n,m) \right]^2 . \tag{2}$$

2. PERCEPTUAL QUALITY EVALUATION

Even though in our previous work the MSE estimation was quite reliable, we did not take into account the perceptual quality of the compressed image. Actually, we provided just a tool for estimating the MSE between two images in the wavelet domain, which was not possible before for non-linear (*i.e.* adaptive) wavelet transforms. Now, it is well known that MSE is not satisfactory for perceptual quality evaluation. In this work we propose a method for estimating the perceptual quality of an image compressed with an adaptive LS (with particular focus on APLS since they have by far the best performance). We make use of saliency maps in order to evaluate the different contributions of wavelet coefficients affecting different areas of the image; moreover we use the weights proposed in our previous works [10] in order to correctly compare different subbands.

We take inspiration from the quality metrics based on the saliency of specific areas in an image or a video. Let x be the original image, $\hat{x}$ the distorted (or compressed) one, and n, m the spatial coordinates for pixels. The perceptual distortion is a weigthed sum of errors:

$$D_2 = \sum_{n,m} \mu(n,m) \left[x(n,m) - \hat{x}(n,m) \right]^p \tag{3}$$

where μ is a suitable saliency map. For example, in [11], it is proposed to take into account three phenomena: the image contrast (on a frame-by-frame basis), the global activity and the local activity (on a temporal basis). The image contrast mask, inspired on the work by Kutter and Winkler [12] uses a non-decimated WT of the image. Let W^{LL} be the LL band of undecimated wavelet transform of x. The contrast saliency map is defined as:

$$\alpha(n,m) = T[C_0(n,m)] \cdot W^{\mathrm{LL}}(n,m) \tag{4}$$

where

$$T[C_0] = \begin{cases} C_T & \text{if } C_0 < C_M \\ C_T \left(\frac{C_0}{C_M} \right)^{\epsilon} & \text{otherwise} \end{cases}$$

$$C_0(n,m) = \sqrt{2} \cdot \frac{\sqrt{|W^{\mathrm{HH}}(n,m)|^2 + |W^{\mathrm{HL}}(n,m)|^2 + |W^{\mathrm{LH}}(n,m)|^2}}{W^{\mathrm{LL}}(n,m)}$$

The parameters C_T, C_M, ϵ can be assigned according to the observations made in the paper [12]. This map does not take into account temporal effects in a video, and could be used if only fixed images are to be considered. However, one can make up for it by adding a further contribution accounting for global activity and based on motion vector norms.

In conclusion, we end up with a single masking function α which is higher where the observer is less sensible to errors, such that we can use $\mu = 1/\alpha$ in Eq. (3).

The problem is that this distortion metric should be computed in the spatial domain, which is a quite large impairment for compression algorithms, as already noted in the first section.

3. PROPOSED METRIC

Based on the previous work [9, 10] and inspired on the perceptual metrics used in [11, 12], we propose a new metric which would allow to evaluate the perceptual effect of quantization (and actually of any other degradation) performed in the transformed domain. In other words, we want to make it possible to evaluate the perceptual quality of a compressed image directly from its transformed coefficients, when *adaptive* and highly non-linear transforms are used.

The proposed metric is based on subband energy weighting (to make it possible to use adaptive filters) and on the perceptual saliency described in the previous section. The weighting allows to compare wavelet subbands having different orientations and resolutions; the spatial masking allows to evaluate the impact of each WT coefficient according to the spatial region it will affect in the reconstructed image. However, since the different subbands have different resolutions, the mask α must be adapted to it. To this end, we define the mask value $\alpha_i(n,m)$ at resolution level i as the average of mask values in the positions associated to the coefficient (n,m):

$$\alpha_i(n,m) = \frac{1}{4^i} \sum_{k=2^i n}^{2^i n + 2^i - 1} \sum_{\ell=2^i m}^{2^i m + 2^i - 1} \alpha(k,\ell) \tag{5}$$

Now we can define the distortion evaluation in the transform domain. The new metric is similar to the one in Eq. (1):

$$D_3 = \sum_{ij} w_{ij} d'_{ij} \tag{6}$$

since the weights (computed as defined in [9, 10]) are necessary to compare the distortion in different subbands. The innovation stands in the term d'_{ij}, defined as follows:

$$d_{ij} = \sum_{n,m} \mu_i(n,m) \left[y_{ij}(n,m) - \hat{y}_{ij}(n,m) \right]^2 \tag{7}$$

This equation is similar to the perceptual metric in Eq. (3); however here we use $\mu_i = 1/\alpha_i$. In its turn, α_i is defined in Eq. (5), and any saliency mask can be used in principle, even though in a first moment we propose the one suggested in [11].

4. EXPERIMENTAL RESULTS

The proposed metric can be used to compare the perceptual quality of images, so one can easily conceive a battery of tests devoted to inspect the correlation between the proposed metric and a subjective measure.

However we introduced the metric in Eq. (6) in order to improve resource allocation for image and video coding. Therefore, a more significant set of experiments would consist in using Eq. (6) to drive any resource allocation algorithm, be it a simple uniform quantization of WT coefficients (the resource allocation would decide the quantization step for each subband) or more efficient techniques such as EBCOT.

A first set of experiments is conducted as follows: for a given image, the saliency map μ is computed as specified in the previous Section. Then the image is transformed using the adaptive wavelet transform proposed by Claypoole. The we considered three methods to allocate coding resources to coefficients coding blocks: a traditional method based on coefficient variances; a weighted method using MSE-minimizing weights proposed in [10], and a perceptual method using the weigths and the average value of the saliency map in the locations corresponding to the code block.

Then, the image was coded by simple uniform quantization and entropic coding, using the rates which in turn take into account:

1. only the variances;
2. variances and the normalizing weights;
3. the variances, the normalizing weights and the saliency map.

For each technique, all the images were coded at several coding rates, ranging from 0.1 to 2 bpp. Then we evaluated the quality of the reconstructed image using PSNR and SSIM. Finally, we computed the difference in PSNR and SSIM (indicated with $\Delta^{\mathrm{PSNR}}$and Δ^{SSIM}) between techniques 1) and 2) and between 2) and 3). The first set of differences measures the impact of correct subband weighting from an objective (Δ_1^{PSNR}) and subjective (Δ_1^{SSIM}) with respect to classical coding. The second set of measures indicates the objective (Δ_2^{PSNR}) and subjective (Δ_2^{SSIM}) impact of saliency maps.

Image	Δ_1^{SSIM}	Δ_2^{SSIM}
barbara	4.750	0.692
baboon	6.748	1.835
bottle	2.277	0.043
cameraman	5.459	0.095
couple	4.309	0.141
crowd	3.264	0.048
einst	5.443	0.111
house	2.697	0.009
lena	3.109	0.194
man	4.613	0.235
plane	2.835	0.174
spring	3.259	0.298
truck	2.398	0.233
woman1	3.858	0.184

Table 1. Average SSIM gains, percent values.

Analytical results are shown in Tables 1 and 2. In the first one, we show the quantities $\Delta_1^{\mathrm{SSIM}}$and Δ_2^{SSIM}. Looking at the Δ_1^{SSIM} column, we see that the correct weighting of the transform subband has a beneficial impact over subjective quality, allowing to improve the SSIM up to 6.7%. A further improvement, which is smaller but not negligible, is possible when the saliency information is used, since the values of the Δ_2^{SSIM} column are always positive. The performance gain of our contribution can globally be assessed as the sum of the two deltas.

The second tables shows us that correct weights do always improve the image PSNR. This is exactly what we expected, since this weighting was conceived to improve the objective quality of the image. We also notice that taking into account saliency does not always improve the PSNR[1]. This result is not surprising, since it is known that PSNR is not perfectly correlated to subjective quality.

We conclude that the proposed metric, used in the transformed domain, allows to improve the SSIM of decoded images, simply by altering the coding resource allocation between coding blocks. This is obtained in spite of the occasional reduction of PSNR.

We also report some more detailed results for a couple of images. In Fig. 2 we show the SSIM as a function of the cod-

[1]If the quantization noise was a perfectly uncorrelated random process, the expected value of PSNR obtained with the weights would be larger than any other. However the facts that the quantization noise is not white and that we can only compute the average PSNR and not its expected value, makes it possible for some positive Δ_2^{PSNR}'s to appear.

Image	Δ_1^{PSNR}	Δ_2^{PSNR}
barbara	4.265	0.222
baboon	3.474	0.126
bottle	4.567	-0.182
cameraman	3.810	-0.186
couple	3.237	-0.105
crowd	2.356	0.033
einst	4.180	0.001
house	4.179	0.068
lena	3.583	-0.103
man	3.115	0.028
plane	2.995	-0.037
spring	3.187	0.080
truck	2.649	0.058
woman1	3.168	0.071

Table 2. Average PSNR gains, in dB.

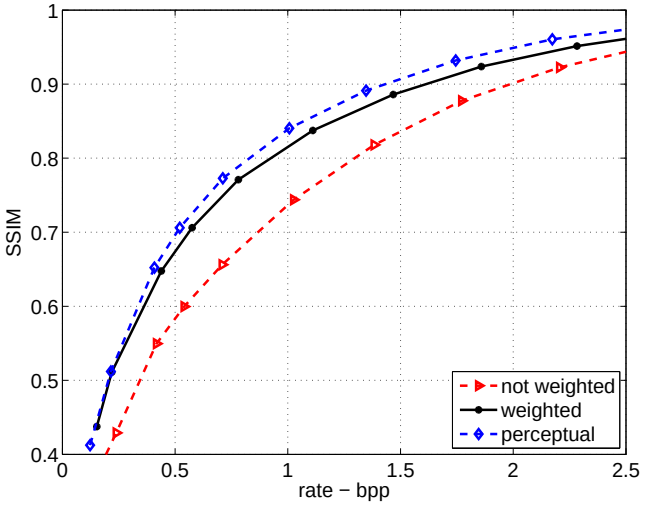

Fig. 2. SSIM for the image "baboon" at several coding rates

ing rate for the three considered techniques and for the test image "baboon". We see that the improvement with respect to the basic technique (red curve) is consistent for all the coding rates. Moreover, in Fig. 3 we show the Δ_2^{SSIM} for this image (using interpolated values for computing the SSIM difference). We observe that SSIM improvements are relevant above all at the medium coding rates. It is worth nothing that for coding rates below 0.5 bpp the quality of the decoded image is not satisfactory, whatever the coding technique is. In Fig. 4 we report the SSIM behavior for another test image, "barbara". Similar conclusions (with respect to the previous case) can be drawn.

Finally, in Fig. 5, 6, and 7 we show the decoded "baboon" images for the three techniques. The coding rates are approximately the same (0.7 bits per pixel), but the visual quality are far different. In the first image, neither the weights nor

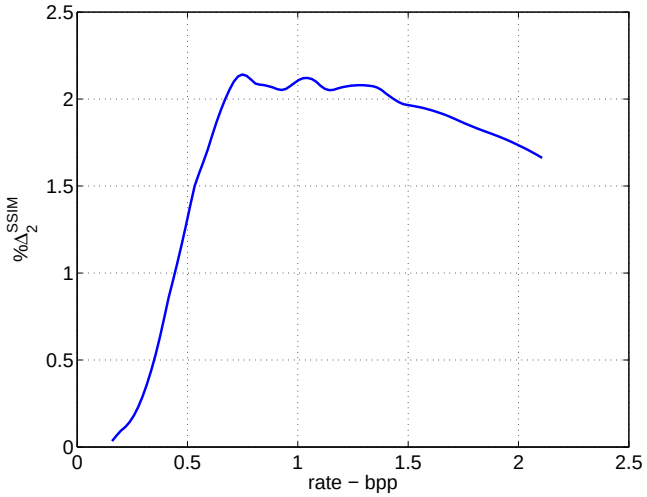

Fig. 3. SSIM differences Δ_2^{SSIM} for the image "baboon"

the saliency map have been taken into account. This explains the poor visual and objective (PSNR) quality of the decoded image. In the second image, relative subband importance has been take into account, in order to maximise the PSNR. This results in an improved visual quality with respect to the non-weighted case. However, the best perceptual quality (measured by SSIM) appears to be in the third image, where, at the cost of a very small loss in PSNR, we have an improved SSIM (Δ_2^{SSIM}= 2.3 %) and we are able to keep some fine details the we would loose with the MSE-oriented technique. For example, we can remark that the nose contours are sharper, and that some detail (like the baboon's hairs in the highlighted box) are kept only when using the perceptual approach.

5. CONCLUSIONS AND FUTURE WORK

In this work we proposed a visual quality metric to be used when coding images with non iso-metric, adaptive wavelet transforms. This metric is based on a weighted average of the transform-domain quantized error and then can be used to drive a resource allocation algorithm without a decoding loop.

The metric takes into account two aspects: the different weights of the transform subbands (due to the non-isometric transform) and the visual saliency of different image locations. Any saliency map can be used to perform this task, and in this preliminary work we rely on the map proposed in state-of-the-art papers [11, 12].

Using this metric results in an improved quality of the decoded images. This is proved by measuring the SSIM and by direct visual inspection of images. We provide some decoded images to support our conclusion. In conclusion, in this paper we have shown that a judicious rate allocation is necessary when one wants to compress images with adaptive,

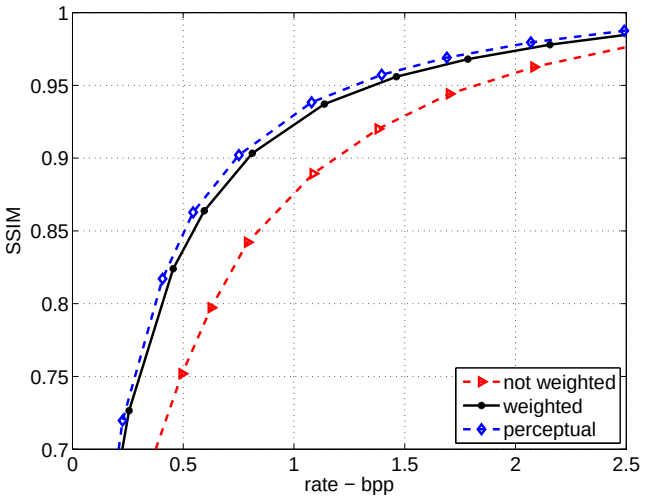

Fig. 4. SSIM for the image "barbara" at several coding rates

Fig. 5. Decoded image, no weighting, Rate 0.71bpp PSNR 22.53 dB SSIM 0.656

non-linear and non-isometric transforms such as the adaptive lifting schemes. We also propose a technique for performing a distortion analysis (be it objective or subjective) which improves the quality of the decoded image.

Being the result of an exploratory work, many improvements can be expected for the proposed method. We intend to explore the effect of more sophisticated saliency maps. In particular, we would like to explore the application to video, where the saliency map can take into account the effect of motion on the importance of the regions of interest.

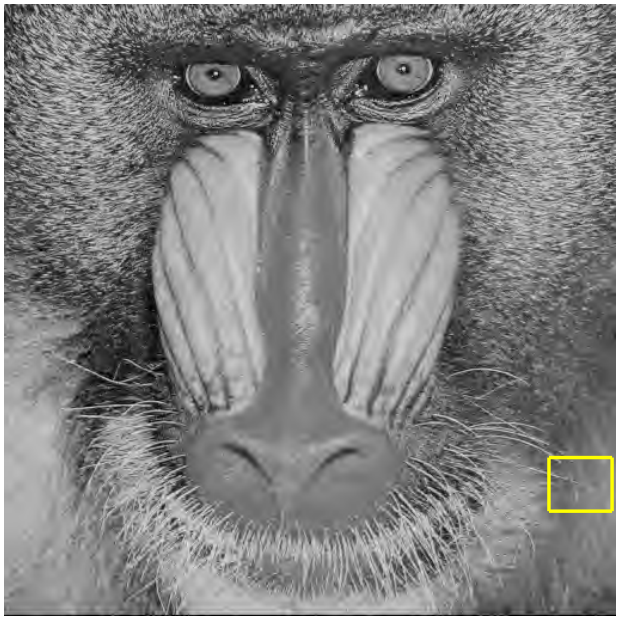

Fig. 6. Decoded image, weights, Rate 0.72bpp PSNR 25.88 dB SSIM 0.750

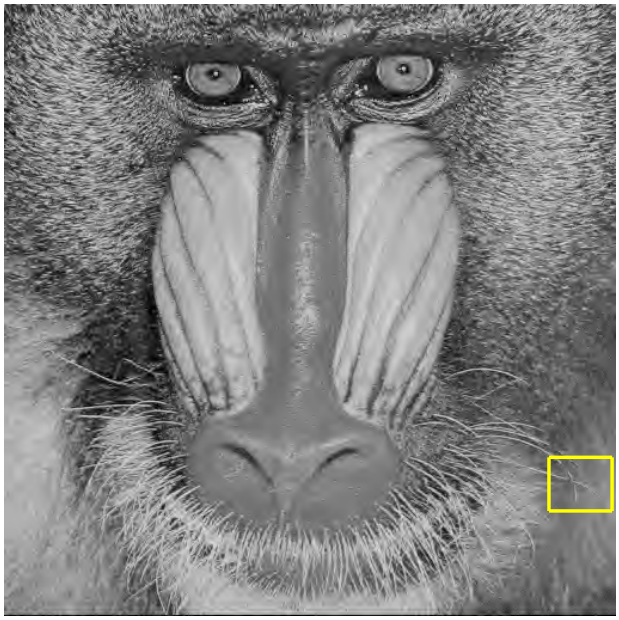

Fig. 7. Decoded image, perceptual coding, Rate 0.71 PSNR 25.85 SSIM 0.773

6. REFERENCES

[1] Wim Sweldens, "The lifting scheme: A custom-design construction of biorthogonal wavelets," *Appl. Comput. Harmon. Anal.*, vol. 3, no. 2, pp. 186–200, 1996.

[2] Gemma Piella, Béatrice Pesquet-Popescu, and Henk J. A. M. Heijmans, "Gradient-driven update lifting for

adaptive wavelets," *Signal Proc.: Image Comm. (Elsevier Science)*, vol. 20, no. 9-10, pp. 813–831, Oct.-Nov. 2005.

[3] R. L. Claypoole, G. M. Davis, W. Sweldens, and R. G. Baraniuk, "Nonlinear wavelet transforms for image coding via lifting," *IEEE Trans. Image Processing*, vol. 12, no. 12, pp. 1449–1459, Dec. 2003.

[4] Nagita Mehrseresht and David Taubman, "Spatially continuous orientation adaptive discrete packet wavelet decomposition for image compression," in *Proceed. of IEEE Intern. Conf. Image Proc.*, Atlanta, GA (USA), Oct. 2006, pp. 1593–1596.

[5] D. Taubman, "High performance scalable image compression with EBCOT," *IEEE Trans. Image Processing*, vol. 9, no. 7, pp. 1158–1170, July 2000.

[6] J. M. Shapiro, "Embedded image coding using zerotrees of wavelets coefficients," *IEEE Trans. Signal Processing*, vol. 41, pp. 3445–3462, Dec. 1993.

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

[8] B. Usevitch, "Optimal bit allocation for biorthogonal wavelet coding," in *Proceed. of Data Comp. Conf*, Snowbird, USA, Mar. 1996, pp. 387–395.

[9] S. Parrilli, M. Cagnazzo, and B. Pesquet-Popescu, "Distortion evaluation in transform domain for adaptive lifting schemes," in *Proceed. of IEEE Worksh. Multim. Sign. Proc.*, Cairns, Australia, 2008, pp. 200–205.

[10] S. Parrilli, M. Cagnazzo, and B. Pesquet-Popescu, "Estimation of quantization noise for adaptive-prediction lifting schemes," in *Proceed. of IEEE Worksh. Multim. Sign. Proc.*, Rio de Janeiro, Brazil, Oct. 2009.

[11] R. Caldelli, A. De Rosa, P. Campisi, M. Carli, and A. Neri, "Perceptual aspect exploitation in video data hiding," in *Proceed. of Intern. Worksh. on Video Proc. and Quality Metrics*, Scottsdale, AZ, U.S.A., Jan. 2006.

[12] M. Kutter and S. Winkler, "A vision-based masking model for spread-spectrum image watermarking," *IEEE Trans. Image Processing*, vol. 11, no. 1, pp. 16–25, Jan. 2002.