

## AN ADAPTIVE PERCEPTION-BASED IMAGE PREPROCESSING METHOD

V. Bruni<sup>\*</sup>, I. Selesnick<sup>†</sup>, L. Tarchi<sup>‡</sup>, D. Vitulano<sup>‡</sup>

<sup>\*</sup> Rome University  
"La Sapienza"  
Dept. of SBAI, Italy

<sup>†</sup> New York University  
Dept. of Electrical and  
Computer Engineering, USA

<sup>‡</sup> National Research Council  
IAC  
"M. Picone", Rome, Italy

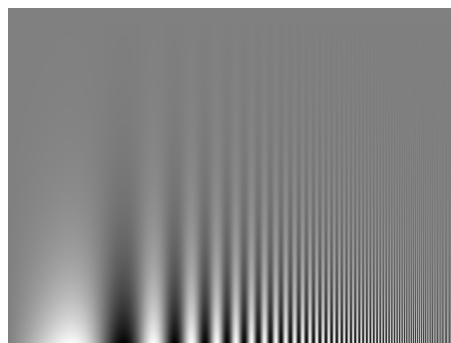
### ABSTRACT

The aim of this paper is to introduce an adaptive preprocessing procedure based on human perception in order to increase the performance of some standard image processing techniques. Specifically, image frequency content has been weighted by the corresponding value of the contrast sensitivity function, in agreement with the sensitiveness of human eye to the different image frequencies and contrasts. The 2D Rational dilation wavelet transform has been employed for representing image frequencies. In fact, it provides an adaptive and flexible multiresolution framework, enabling an easy and straightforward adaptation to the image frequency content. Preliminary experimental results show that the proposed preprocessing allows us to increase the performance of some standard image enhancement algorithms in terms of visual quality and often also in terms of PSNR.

**Index Terms**— Human Visual System, Contrast sensitivity function, Image enhancement, SSIM.

### 1. INTRODUCTION

In the last years there has been a growing interest in the use of Human Visual Perception (HVP) for solving some image and video processing problems [15]. The first important use of perception has been done in image coding where a different quantization is applied to the different image frequency content (that is expressed in a proper time-frequency transform) according to the sensitiveness of human eye to those frequencies [2–4]. Starting from these works, research community has dedicated a lot of effort in studying and embedding visual perception concepts in image/video processing applications reaching impressive results both in terms of the visual quality of the final image and in terms of adaptivity, flexibility and automaticity of the proposed tools. Some significant examples can be found in image quality assessment [8–10, 13], image representation [5], image restoration [1, 6] and target detection [11, 12]. In this paper we exploit these concepts to define a method for preprocessing images (degraded or not) in order to normalize their content according to the visual perception of the local contrast. This idea is well expressed by the Campbell Robson image in Fig. 1. The sinusoidal stimulus, having



**Fig. 1.** Campbell-Robson map: sinusoidal stimulus with increasing frequency from left to right and decreasing contrast from bottom to top.

changing frequency and contrast, is not perceived in the same manner in the whole image. For a fixed distance, it is more evident in the middle of the image while it is gradually less evident far away from it. A concave line gives the frontier between the perceived and not perceived part of this stimulus, namely the Contrast Sensitivity Function (CSF) [14]. By changing the observation distance, this curve changes its maximum point, its amplitude and it shrinks or dilates. The main idea of this work is to construct a CSF that is better adapted to the content of the analyzed image. Specifically, we are interested in finding the frequency at which the CSF attains its maximum for the specific image and in estimating the shape of the curve. Hence, we construct a multiscale transform that changes its frequency resolution according to the found shape: higher resolution at frequencies to which human eye is more sensitive and less resolution far from them. The image content in these bands can be then weighted according to the amplitude of the CSF function, as in [4]. To this aim, differently from the latter, a flexible time-frequency representation is required. In this case we need a transform that allows a tunable and variable Q factor: low Q for portions of the frequency axis that allow coarse resolution, high Q wherever a finer frequency resolution is required. The 2D RDWT [16] allows us the selection of different Q factors and,

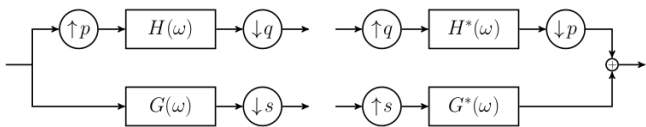


Fig. 2. Rational-dilation filter bank.

more importantly, the Q factor can change at each level of the transform, still guaranteeing perfect reconstruction. In addition the Q factor can, in principle, be selected arbitrarily small or large, making the 2D RDWT useful for our task.

The outline of the paper is the following. Next section gives a brief introduction to the 2D RDWT. Section 3 shows how it can be used for defining a perception-based preprocessing method. Preliminary experimental results are in Section 4 while Section 5 draws the conclusions.

## 2. 2D RATIONAL DILATION WAVELET TRANSFORM

In this work we utilize a two-dimensional version of the rational-dilation wavelet transform (RDWT) [16]. The RDWT is an overcomplete discrete wavelet transform in which the dilation factor can be adjusted by the user. In particular, the dilation factor can be set between 1 and 2 to perform a more gradual scaling between consecutive subbands as compared to the dyadic wavelet transform (DWT) [17]. Like the DWT, the RDWT is implemented using an iterated perfect reconstruction digital filter bank. It likewise also possesses the Parseval energy preservation property. However, the RDWT filter bank (Fig. 2) has a different configuration of up- and down-samplers. Also, unlike the usual DWT, the RDWT is best implemented in the frequency domain. (This is because the perfect reconstruction conditions are not easily satisfied by FIR filters or even by recursive IIR filters.) That is to say, the filters  $H$  and  $G$  are defined directly by their frequency responses and they are implemented by point-wise FFT multiplication (in accordance with the DFT convolution property). The integer sampling factors ( $p$ ,  $q$ ,  $s$ ) control the dilation factor of the RDWT and the Q-factor of the wavelet. In this work, we set  $s = 1$  as we are interested in low Q-factor wavelet transforms (i.e., not highly oscillatory). One should use  $s > 1$  for high Q-factor transforms which are suitable for the analysis of oscillatory signals (e.g., speech). We furthermore vary the sampling factors  $p$  and  $q$  from scale to scale, so as to tailor the wavelet analysis to the signal. In order to extend the RDWT to two dimensions, we adopt the methodology and architecture of the steerable pyramid (SPY) [18]. The SPY is implemented using an iterated 2D filter bank with circularly symmetric frequency responses which, in turn, are implemented using the FFT. The 2D filters in the SPY are defined in terms of a perfect reconstruction 1D filter bank. To implement the 2D RDWT, up- and down-

sampling (Fig. 2) are applied separably (to rows and columns) of the image. Additionally, the low-pass and high-pass 1D frequency responses of the 1D RDWT filter bank are used to form circularly symmetric 2D frequency responses. The filters are implemented using the 2D FFT. Hence, the 1D rational-dilation filter bank, on which the RDWT is based, can be substituted into the SPY to obtain a transform for which the dilation factor can be set by the user. Furthermore, the dilation factor can be adjusted at each scale while preserving the perfect reconstruction property of the transform.

## 3. CSF-BASED PREPROCESSING

Campbell-Robson map shows how the human visual system is more sensitive to medium spatial frequencies. In fact, while in the lower portion of the image it is possible to see clearly the different columns of gray, it is not so in the upper part of the image. The set of points giving this visual threshold outlines the CSF and represents precisely the threshold beyond which no longer one has the perception of the contrast as a function of the spatial frequencies. In general, the maximum of this function is attained for values of the frequency comprised between 2 and 5 cpd (cycles per degree) and depends on the image content and the viewing distance. Hence, in order to perform adaptive operations, it would be desirable to get the CSF for the specific analysed image and use it for balancing its frequency content: preserving (or emphasizing) frequencies close to the maximum of the curve. The main steps of the proposed method are:

1. Find the maximum of the CSF for the analysed image
2. Sample the CSF in a non uniform way according to its shape
3. Find the sequence of Q factors (dilations of 2D RDWT) which implements the non uniform sampling
4. Use the CSF (normalized with respect to its maximum) to weight the coefficients of 2D RDWT
5. Invert the 2D RDWT.

### 3.1. CSF model

CSF depends on a large number of physical parameters, such as quality of the eye lens, photoreceptors sensitivity, several neural characteristics, external noise and so on. There exist several equations that take into account all these dependencies; however, by fixing some of them, it is possible to simplify the formal expression of the CSF. The following equation describes the CSF as a function of the frequency space for a standard observer between 20 and 30 years with a good vision [19]

$$S(u) = \frac{5200/\sqrt{2} e^{-0.0016u^2(1+\frac{100}{L})^{0.08}}}{\sqrt{\left(1 + \frac{155}{X_0^2} + 0.64u^2\right) \left(\frac{63}{L^{0.83} + \frac{1}{1-e^{-0.02u^2}}}\right)}} \quad (1)$$

where  $u$  is the frequency (in cpd),  $L$  is the total luminance of the image,  $X_0$  is the angular dimension of the object expressed in degree, i.e.  $X_0 = 2 \arctan\left(\frac{d}{2v}\right)$  with  $d$  and  $v$  respectively the image dimension and the viewing distance.

The point at which the CSF attains its maximum is seen as the optimum level of resolution at which seeing the image, i.e. the one which gives the just noticeable contrast in the image. This level can be automatically found by computing the contrast between two successive low pass filtered versions of the original image. To this aim we consider a set of  $N$  pairs of natural numbers  $(p_i, q_i)$ ,  $i = 1, \dots, N$  that satisfy the conditions for generating a low-pass filter  $H_{p_i, q_i}$  for the 2D RDWT. By selecting the rule  $q_i = p_i + 1$ , the sequence of filters is such that the filter support grows in the frequency domain as  $p_i$  increases<sup>1</sup>. This sequence of filters is used for defining the Peli's contrast  $C$  of the image  $I$  as follows

$$C(p_i, q_i) = \sum_{(x,y) \in \Omega} \frac{|I * h_{p_i, q_i}(x, y) - I * h_{p_{i-1}, q_{i-1}}(x, y)|}{|\Omega| |I * h_{p_{i-1}, q_{i-1}}(x, y)|}$$

where  $\Omega$  is the image domain and  $h_{p_i, q_i}$  is the impulse response function of  $H_{p_i, q_i}$ . The couple of parameters giving the best level of resolution is the one measuring the minimum perceivable contrast  $\tau$  between two successive blurred images, i.e.  $(\bar{p}, \bar{q}) = \operatorname{argmin}_{(p_i, q_i)} \{C(p_i, q_i) \leq \tau\}$ . The contrast curve versus the level of resolution is depicted in Fig. 3.left. It is a decreasing function and the optimal point coincides with the maximum inflection of the curve. It has been empirically observed that this inflection point is attained when  $\tau = 0.02$ , that coincides with the Weber's contrast threshold.

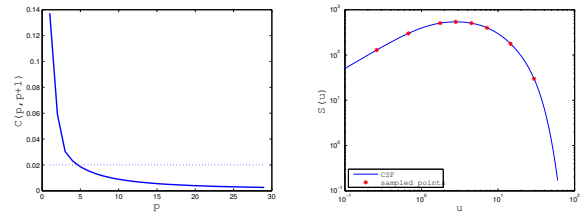
By denoting with  $f_{max}$  the maximum perceived frequency, the maximum of the CSF can be fixed at  $\bar{u} = \frac{\bar{p}}{\bar{q}} \frac{f_S}{2}$ , where  $f_S = 2f_{max}$  is the sampling frequency that gives the level of resolution under which a sinusoidal stimulus whose maximum frequency is  $f_{max}$  is not perceived. Since  $f_S$  can be expressed also in terms of the viewing distance  $v$  (measured in meters) and the image resolution  $r$  (measured in dpi), i.e.  $f_S = 2v r \tan\left(\frac{\pi}{360}\right)$ , by imposing that  $S(u)$  attains its maximum value at  $\bar{u}$ , we can get the optimal viewing distance for the considered image<sup>2</sup>. The value of the viewing distance completely defines the CSF of the analysed image, as the one depicted in Fig. 3.right.

### 3.2. CSF sampling

CSF sampling is based on the conjecture in the field of visibility, which asserts that the human visual system is based on multiple channels and the CSF is the envelope of the contrast sensitivity of the cortical cells that take part to the visual perception process [14]. Hence, it is possible to design a series of

<sup>1</sup>This rule has been selected for simplicity. A different rule which provides filters with increasing support in the frequency domain can be selected.

<sup>2</sup>It is computed by numerically solving the non linear equation  $\frac{dS}{du} |_{u=\bar{u}} = 0$ .



**Fig. 3. Left:** Peli Contrast curve provided by a sequence of rational filters having increasing support in the time domain. **Right:** Loglog plot of the CSF of Lena image corrupted by gaussian noise with standard deviation  $\sigma = 20$ .

filter banks covering all the frequency axis but having bandwidths that are adapted to the curve shape: more tight around the point of maximum visibility and wider elsewhere. In other words, we determine a non uniform grid  $u_1, u_2, \dots, u_N$  in the interval  $[0, f_{max}]$  that gives the limits of the bandwidths of the filter bank to construct, i.e.  $u_k = \frac{p_1 p_2 \dots p_k}{q_1 q_2 \dots q_k} f_{max}$ . The distance between consecutive points is fixed such that the interpolation error is less than a prefixed tolerance.

CSF fitting can be done in different ways. In this paper we propose to consider three contiguous intervals in  $[0, f_{max}]$ :  $D_1, D_2, D_3$ .  $D_2$  contains the maximum of the CSF. A different Q factor (couple of  $(p_i, q_i)$ ,  $i = 1, 2, 3$ ) is selected for each interval as the one providing the least allowed fitting error. Taking into account the shape and the regularity of the CSF, 8 points have been selected: 4 equally spaced (in the logarithmic scale) in the central interval and 2 in each of the remaining intervals. It means that we have a grid more dense in correspondence to the frequency band containing the maximum of the CSF, and a less dense elsewhere (see Fig. 3.right). It is worth observing that requiring equally spaced points in the logarithmic scale means that their distance is equal to  $\log(q_i) - \log(p_i)$ .

### 3.3. CSF based weighting

The analysis in the previous sections provides the sequence  $\{(p_1, q_1), (p_2, q_2), (p_3, q_3)\}$  of dilation factors that determines the 2D RDWT to be applied to the analysed image. Eight frequency bands (each corresponding to a sample  $u_k$  of the CSF) have been then determined and weighted using the following weights

$$w(u_k) = \frac{S(u_k)}{S(\bar{u})}, \quad k = 1, 2, \dots, 8.$$

The inversion of the transform using the modified coefficients provides the preprocessed image  $I_P$ .

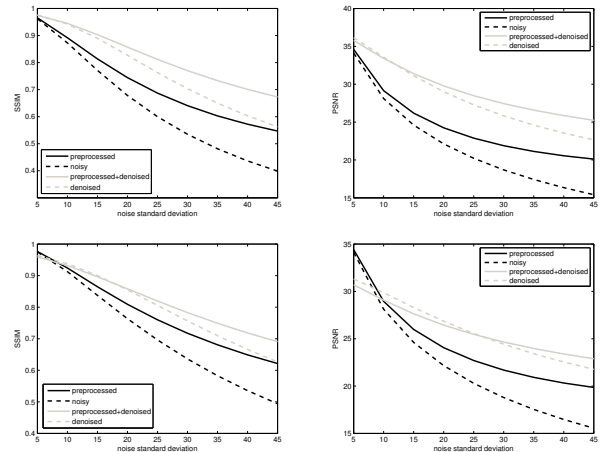
## 4. EXPERIMENTAL RESULTS

The proposed preprocessing procedure has been tested on several test images and some common image distortions,

**Table 1.** Gain in terms of SSIM and PSNR (in the brackets) provided by the proposed preprocessing strategy in denoising Boat image and images in TID2008 database (average values have been provided) affected by noise with different standard deviations.

Image	Method	Noise StD $\sigma$			
		10	15	20	40
Boat	$I_P/I$	0.014 (0.95)	0.03 (1.49)	0.05 (2.02)	0.14 (2.55)
	$I_{P,d}/I_d$	-0.002 (-0.46)	0.003 (-0.23)	0.014 (0.13)	0.076 (1.55)
TID2008	$I_P/I$	0.007 (0.46)	0.016 (0.77)	0.032 (1.20)	—
	$I_{P,d}/I_d$	-0.003 (-0.36)	0.007 (0.20)	0.011 (0.44)	—

like additive noise and blurring. In order to evaluate to what extent the proposed preprocessing strategy can be useful in image denoising, we consider a noisy image  $I$  and we compare the quality of the preprocessed images  $I_P$ , the denoised image using a conventional Wiener filter  $I_d$  and the denoised  $I_{P,d}$  preprocessed image using the same denoising method, in terms of SSIM (Structural SIMilarity index [9]) and PSNR (Peak Signal to Noise Ratio). Despite the limitations of both measures [20], they have been used in this paper since the former is a classical and widely used objective measure, the latter is defined based on simple visibility concepts, like luminance adaptation and contrast masking, that are also the basic concepts of the presented work. Fig. 4 depicts the results achieved on  $512 \times 512 \times 8$ bits Lena and Barbara images corrupted by additive gaussian noise with increasing variance. As it can be observed, the gain in terms of SSIM is considerable when comparing  $I$  and  $I_P$  (increasing for increasing  $\sigma$ : from 0.001 to 0.11 for Lena image; from 0.002 to 0.13 for Barbara) and also when comparing  $I_d$  and  $I_{P,d}$  (from 0.003 to 0.15 for Lena; from -0.0043 to 0.06 for Barbara). It is worth observing that also the PSNR of  $I_P$  and  $I_{P,d}$  increases. In particular, for Lena image the gain is from 0.57 db to 4.70 db for the preprocessed image  $I_P$  and from -0.36 db to 2.58 db for  $I_{P,d}$  for the preprocessed and denoised image; on the contrary, for Barbara image the gain is a bit lower: from 0.29 to 4.30 for  $I_P$ , from -0.61 to 1.12 for  $I_{P,d}$ . For these images, the couples of selected dilation factors of the 2D RDWT are (3, 5), (1, 2), (2, 5). To evaluate the visual quality of the images,  $I$ ,  $I_P$ ,  $I_d$  and  $I_{P,d}$  related to Lena image with  $\sigma = 25$  have been shown in Fig. 5. It is worth observing that if a constant  $Q$  factor is selected, for example  $(p, q) = (1, 2)$ , the gain in terms of SSIM and PSNR is smaller, respectively of about 0.04 and 0.9. This confirms the fact that the proposed preprocessing method allows for a considerable increment of the quality of the final denoised images both in terms of



**Fig. 4.**  $512 \times 512 \times 8$  bits Lena (top) and Barbara images (bottom): SSIM (left) and PSNR (right) results for increasing noise standard deviation.

SSIM and PSNR. The same considerations are valid if we consider a larger set of test images. As shown in Table 1, it is convenient to use the preprocessing procedure especially for higher noise levels. In general, for standard deviations greater than 15, the proposed denoising method provides a gain (in terms of SSIM and/or PSNR) in the final image.

We also tested the preprocessing method for blurred images. In this case we still observe an increment of the visual quality of the image, even though it is less evident as in denoising (SSIM gain for the final preprocessed and deblurred image is 0.02 on average). As a matter of fact, the proposed weighting strategy is very simple and involves all the coefficients of the frequency band without discriminating them with respect to their local contrast. In the case of blurring an adaptive weighting method would be desirable.

Finally, the additional cost required by the proposed procedure is very moderate, making it useful for fast and real time applications.

## 5. CONCLUSIONS

In this paper an image preprocessing method based on human perception has been presented. The contrast sensitivity function for the analysed image has been estimated and non uniformly sampled in order to get a visually significant partition of the frequency axis. The 2D RDWT has been used for expanding the image in the found frequency bands. The value of the contrast sensitivity function in these bands has been used for properly weighting their content: a higher weight has been assigned to those frequencies to which human eye is more sensitive and is able to capture the least contrasts; a lower weight is assigned elsewhere. Preliminary experimental results show that the proposed method allows to increase the visual quality of some standard image enhancement al-



**Fig. 5.** From left to right top to bottom: noisy  $I$  ( $\sigma = 25$ , PSNR = 20.22, SSIM = 0.599), preprocessed  $I_P$  (PSNR = 22.88, SSIM = 0.687), denoised  $I_d$  (PSNR = 27.27, SSIM = 0.763) and denoised preprocessed  $I_{P,d}$  (PSNR = 28.49, SSIM = 0.812) Lena image.

gorithms in terms of SSIM and often in terms of PSNR. The achieved results are promising and promote future research that will surely be devoted to the use of more perceptually based quality measures in the evaluation of images and in producing more extensive results (different images, different denoising algorithms). Future research will also deal with the refinement of the proposed strategy making it more adaptive to the local image visual contrast and to the image distortion kind.

## REFERENCES

- [1] V. Bruni, E. Rossi, D. Vitulano, On the Equivalence Between Jensen-Shannon Divergence and Michelson Contrast, *IEEE Trans. on Information Theory*, vol. 58, no. 7, pp. 4278-4288, 2012.
- [2] A. B. Watson, "DCTune: A technique for visual optimization of DCT quantization matrices for individual images," *Soc. Inf. Display Dig. Tech. Papers*, vol. XXIV, pp. 946-949, 1993.
- [3] I. S. Hontsch, L. J. Karam, Adaptive image coding with perceptual distortion control, *IEEE Trans. on Image Proc.*, vol. 11, no. 3, pp. 213-222, 2002.
- [4] M.J. Nadenau, J. Reichel, M. Kunt, Wavelet-based color image compression: exploiting the contrast sensitivity function, *IEEE Trans. on Image Processing*, vol. 12, no. 1, 2003.
- [5] L. Itti, C. Koch, E. Niebur, A Model of Saliency-Based Visual Attention for Rapid Scene Analysis, *IEEE PAMI*, 1998.
- [6] V. Bruni, D. Vitulano, A generalized model for scratch detection, *IEEE Trans. on Image Proc.*, vol. 13, no. 1, 2004.
- [7] A. Kokaram, D. Vitulano, D. Corrigan, V. Bruni, F. Pitti, A.J. Crawford, *Advances in Automated Restoration of Archived Video, Digital Imaging for Cultural Heritage Preservation*, CRC Press, 2011
- [8] H. R. Sheikh, A. C. Bovik, Image information and visual quality, *IEEE Trans. on Image Proc.*, vol. 15, no. 2, pp. 430- 444, 2006
- [9] Z. Wang, A. Bovik, H. Sheikh, E. Simoncelli, Image quality assessment: from error visibility to structural similarity, *IEEE Trans. on Image Proc.*, vol. 13, no. 4, pp. 600-612, 2004
- [10] D.M. Chandler, S.S. Hemami, VSNR: A Wavelet-Based Visual Signal-to-Noise Ratio for Natural Images, *IEEE Transactions on Image Processing*, vol. 16, no. 9, pp. 2284-2298, 2007.
- [11] T. Arnou, A. Bovik, Foveated visual search for corners, *IEEE Trans Image Proc.*, vol. 16, no. 3, pp. 813-823, Mar. 2007.
- [12] A. Tavassoli, I. van der Linde, A. C. Bovik, L. K. Cormack, An efficient technique for revealing visual search strategies with classification images, *Perception & Psychophysics*, vol. 69, no. 1, 2007.
- [13] V. Bruni, E. Rossi, D. Vitulano, Jensen-Shannon divergence for visual quality assessment, *Signal Image and Video Processing, Signal Image and Video Proc.*, Vol. 7, no. 3, 2013.
- [14] S. Winkler, *Digital Video Quality-Vision Models and Metrics*, J. Wiley and Sons, 2005.
- [15] Special issue on Human Vision and Information Theory, *Signal Image and Video Processing*, Guest Editors V. Bruni, D. Vitulano, Z. Wang, Vol. 7, no. 3, May 2013.
- [16] I. Bayram, I.W. Selesnick, Frequency-Domain Design of Overcomplete Rational-Dilation Wavelet Transforms, *IEEE Trans. on Signal Proc.*, vol. 57, no. 8, 2009.
- [17] I. Daubechies, *Ten Lectures On Wavelets*, SIAM, 1992.
- [18] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, D. J. Heeger, Shiftable multi-scale transforms, *IEEE Trans. on Inform. Theory*, vol. 38, no. 2, pp. 587-607, 1992.
- [19] P. Barten, Formula for the contrast sensitivity of the human eye, *Proc. of SPIE El. Imaging*, vol. 5294, 2004.
- [20] K. Zeng, Z. Wang, Perceptual evaluation of image denoising algorithms, *Proc. of the 47th IEEE Asilomar Conf. on Sig., Syst. and Comp.*, 2013.