The Rational Dilation Wavelet Transform: A Flexible Tool for Perception-inspired Signal and Image Processing

Vittoria Bruni
Dept. of SBAI
University of Rome La Sapienza
Rome, Italy
e-mail: vittoria.bruni@uniroma1.it

Domenico Vitulano
Institute for Calculus Applications (IAC)
CNR
Rome, Italy
e-mail: d.vitulano@iac.cnr.it

***Abstract*—The aim of this work is to investigate the properties of an adaptive multiscale transform for defining perception-based methods for signal and image processing. Particular attention is devoted to the possibility of partitioning the frequency plane in a flexible way, which depends on both human perception and task purpose. Examples concerning image enhancement and timbre recognition will be presented and discussed.**

Keywords-contrast sensitiviy; rational dilation wavelet transform; MEL cepstrum; adaptive scale selection.

I. INTRODUCTION

In the last years, there has been a huge research work concerning time-frequency transforms, since many problems of signal and image processing can be successfully solved by expanding the signal in a proper basis where signal features are emphasized. In general, this property is indicated with the term sparsity, i.e., the signal is represented as a series expansion in a proper basis where only few coefficients are non zero. Unfortunately, there is not a unique optimal basis for each kind of signal or for each kind of problem; moreover, sometimes there is not a unique optimal basis for a given signal, since it depends on its spatial/spectral components. That is why the family of transforms/bases is wide. They differ according to the class of functions they are able to compactly represent, the existence of a computable inverse transform and a fast algorithm for their discrete implementation.

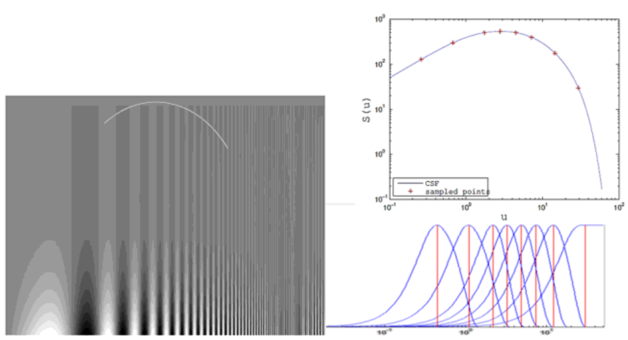

Figure 1. Left) Campbell-Robson map. The white curve is the contrast sensitivity function (CSF). Right) CSF sampling grid (top); corresponding frequency partition (bottom).

On the other hand, in several applications there is the need of reproducing the different spectral components of a signal or image which not always obey a logarithmic law. Let us consider, for example, visual perception. The Campbell Robson image in Figure 1 represents a sinusoidal stimulus having changing frequency and contrast --- contrast is constant at each row and decreases from bottom to top; frequency increases from left to right. However, the stimulus is not perceived in the same manner in the whole image. At a fixed distance, the sinusoidal shape is perceived just below the white line and it is more evident in the middle of the image. By changing the observation distance, the curve shifts, changes its amplitude, shrinks or dilates. The curve that separates the perceptually homogeneous region (top) from the non homogeneous part (bottom) is the Contrast Sensitivity Function (CSF) [1]. It would be then desirable to construct a CSF that is better adapted to the content of the analyzed image in order to simulate frequency axis partitioning to eye sensitivity, i.e., more dense close to CSF maximum. Hence, a multiscale transform that changes its frequency resolution according to CSF shape has to be defined: higher resolution is required at frequencies to which human eye is more sensitive while less resolution is allowed far from them. Similarly, in audio processing, Mel scale is the one that better simulates the ability of the auditory system to distinguish two similar sounds [2]. Even in this case, Mel scale does not correspond to a logarithmic partition of the frequency axis.

The paper is organized as follows: Section II briefly revises the rational-dilation wavelet transform (RADWT) [3] properties. Section III shows two representative examples, while the last section draws the conclusions and provides hints for future work.

II. THE RATIONAL-DILATION WAVELET TRANSFORM

RADWT [3] is an overcomplete discrete wavelet transform in which the dilation factor can be set between 1 and 2 to perform a more gradual scaling between consecutive subbands, as depicted in Figure 2. It is a powerful tool for signal analysis and our purposes for the following reasons:

- it allows a tunable scale factor (known as Q factor). It means that it gives a time-scale representation of the signal where the scale parameter changes according to a factor that is smaller than two. Hence, the high frequencies of the signal are analysed with a finer resolution than in the dyadic case;

- it is implemented through a filter bank by using just a couple of filters (low-pass and high pass) that satisfy the perfect reconstruction condition (Figure 2);
- it involves downsampling and upsampling operations (Figure 2) even though it has some redundancy with respect to the dyadic case. The redundancy depends on the Q factor, i.e., the frequency resolution chosen for the analysis;
- it allows a high flexibility in the construction of the involved filters since it is possible to select not only the Q factor and/or the redundancy, but also the filter decay in the transition band;
- it also has a straightforward extension to 2D, making it also suitable for image analysis.

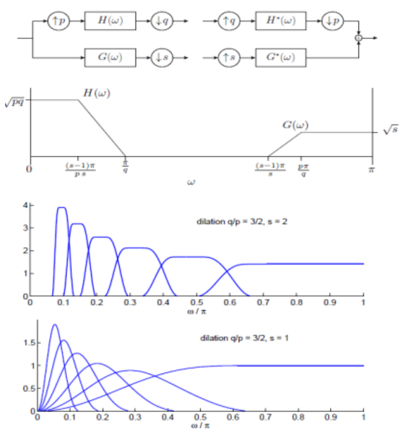

Figure 2. Top) RADWT implementation. Bottom) Frequency partition provided by two different Q factors.

Based on these properties, the dilation factor can be adjusted at each scale while preserving perfect reconstruction property. As a result, RADWT provides a useful tool for defining a multiscale transform that changes its frequency resolution according to human perception.

III. EXAMPLES

In the field of audio processing, the Mel-cepstrum transform [2] combines two elements: the logarithmically spaced Mel scale, modeled on the human auditory system; and the cepstrum transform, which allows separation between excitation and resonances. The aim of our research is to mimic the Mel scale using RADWT to obtain the Mel frequency cepstral coefficients (MFCC). The main idea is to find the parameters p_j, q_j, s_j, such that the support of the high pass filter G_j at level j is close to the support of the j-th Mel band. The result is shown in Figure 3, where the energy distribution in the adaptive RADWT of three different instruments is also shown. The energy distribution clearly characterizes the instrument and then it represents a feature that can be successfully used in timbre recognition.

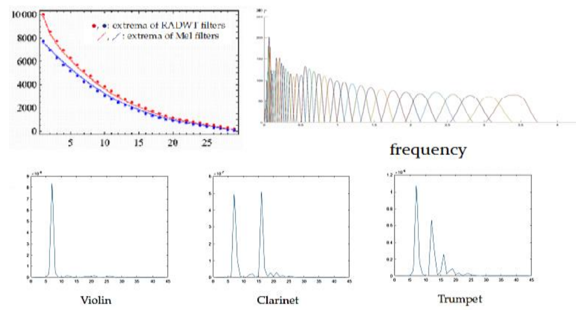

Figure 3. Top) RADWT-based frequency partition according to MEL filters. Bottom) RADWT energy distribution of three different instruments.

With regard to visual perception, the adaptive transform derives from an adaptive sampling of the contrast sensitivity function related to the analysed image. CSF sampling is based on the conjecture in the field of visibility, which asserts that CSF is the envelope of the contrast sensitivity of the cortical cells that take part to the visual perception process [1]. Hence, it is possible to design a filter bank covering all the frequency axis but having bandwidths that are adapted to the curve shape: tighter around the point of maximum visibility and wider elsewhere, as shown in Figure 1.right. The distance between consecutive points is fixed such that the interpolation error is less than a prefixed tolerance. Once the transform is fixed, expansion coefficients can be processed according to the task, as for example, denoising, deblurring, fusion, etc. For example, for degraded images, coefficients can be modified in order to map the CSF of the degraded image to the one of the ideal not degraded image, preserving or emphasizing frequencies close to the maximum of the curve. Preliminary results showed that whenever this mapping is used as a preprocessing step of conventional restoration methods, the final result improves in terms of objective measures, as Peak Signal to Noise Ratio (PSNR), and, especially in terms of image quality assessment measures, as Structural SIMilarity Index (SSIM).

TABLE I. DENOISING USING RADWT-BASED CSF MAPPING AS PREPROCESSING STEP: IMPROVEMENT WITH RESPECT TO STANDARD DWT

Metric	**Denoising results (Lena image)**		
	Noisy image	***Denoised after DWT-based CSF mapping***	***Denoised after RADWT-based CSF mapping***
PSNR	20.22	28.31	**28.40**
SSIM	0.599	0.788	**0.812**

IV. CONCLUSIONS

In this paper the properties of the rational-dilation wavelet transform (RADWT) have been revised and the advantages in using it in some perception-based models have been discussed. Specifically, contrast sensitivity function sampling and Mel-like frequency axis partition have been considered and the benefit given by RADWT has been demonstrated respectively in some classical image processing problems and for audio processing purposes. Future research will be devoted to define automatic and fast

 ISBN: 978-1-61208-724-5

methods for the definition of the best level-dependent dilation parameters for the analysed data and/or the task.

REFERENCES

[1] P. Barten, “Formula for the contrast sensitivity of the human eye,” Proc. of SPIE El. Imaging, vol. 5294, 2004.

[2] L. R. Rabiner and B. H. Juang, “Fundamentals of Speech Recognition,” Prentice Hall, 1993

[3] I. Bayram and I. W. Selesnick, “Frequency-domain design of overcomplete rational-dilation wavelet transforms,” IEEE Trans. Signal Process., vol. 57, pp. 2957–2972, Aug. 2009.