Adaptive Multiresolution Filter Banks for Image and Audio Processing

Alexey Lukin, Daria Kalinkina, Denis Kubasov
Graphics & Media Lab, State University of Moscow, Russia
{lukin, dkubasov, dakalinkina}@graphics.cs.msu.ru

Abstract

In this paper, we consider models of filter banks with variable time-frequency resolution, adapting to signal properties and human perception. The application of the proposed methods to image and audio denoising is demonstrated.

Keywords: *Adaptive, Multiresolution, Filterbanks, Image, Audio, Denoising, PCA, Principal Components.*

1. INTRODUCTION

Digital Signal Processing (DSP) plays an increasing role since multimedia capabilities of personal computers are becoming widespread. During the rise of basic DSP methods in 1950-s a huge attention was paid to effective implementations of developed DSP algorithms. Nowadays, when power of PCs has grown enough to handle many complicate signal processing tasks in real time, the issue of processing quality becomes prevailing over complexity issue. Often users are ready to sacrifice speed for getting better results.

Filter banks are transforms that split a signal into several frequency bands and can restore the signal back [1]. Examples of filter banks are Short-Time Fourier Transform (STFT) widely used in audio processing, and Discrete Wavelet Transform (DWT) which is a basis of many image processing algorithms. In this paper, we consider more complex filter banks for image and audio processing that allow achieving better processing quality by means of variable time-frequency resolution adapting to the human perception.

As will be shown in section 3, the suggested filter banks are able to improve many existing DSP algorithms, since they can be integrated into flowchart of different methods. In this paper we focus on their applications to noise reduction and give several examples of using our methods for other tasks.

2. DENOISING ALGORITHMS

Let's consider a problem of suppression of stationary additive noises uncorrelated with a useful signal. Algorithms for reduction of such noises are following a common scheme:

1. Signal transform performing energy compaction (energy compaction means that most useful signal energy is being placed in minimal possible number of transform coefficients);
2. Suppression of transform coefficients corresponding to noise;
3. Inverse transform restoring the denoised signal.

Discrete Wavelet Transforms (DWT) are most widely used as energy compaction transforms for image processing. They are computationally efficient and allow processing of image details in several scales [2].

In audio processing, most widely used filter banks are based on Short-Time Fourier Transform (STFT). They are the basis of virtually all noise reduction algorithms in audio [3] since they have good enough energy compaction for speech and tonal music.

3. SUGGESTED APPROACH

Abovementioned filter banks have a significant drawback – fixed time-frequency resolution. For example, STFT divides a time-frequency plane into equal rectangles, DWT divides it into octaves. At the same time, it is known that human perception systems vary their time-frequency resolution according to input signals. For example at the moments of rapid changes of audio signals ("transients" – onsets of musical instruments, beats) our temporal resolution increases and we can distinguish small differences in temporal energy profile of signals. And during stationary audio intervals the frequency resolution increases allowing us to determine precise pitch of audio signals [4]. Therefore using filter banks with fixed time-frequency resolution sometimes results in frequency resolution being below our perception capabilities (which leads to poor energy compaction and insufficient noise reduction), and sometimes – in insufficient temporal resolution (which leads to the noticeable Gibbs phenomenon).

3.1 General scheme

We suggest the following scheme for implementation of a filter bank with a variable time-frequency resolution (fig. 1). The same processing algorithm is running several instances with different time-frequency tradeoffs that work in parallel on the same data stream. Their resulting signals are combined by another filter bank with a fixed time-frequency resolution. The block where transform domain (filter bank) coefficients of different signals are mixed together will be called a "mixer of coefficients". The process of mixing can be controlled by some prior strategy (e.g. reflecting properties of human perception) or depending on signal properties at the present moment (e.g. on its stationarity).

Fig. 1 shows the example of parallel processing with just two different time-frequency tradeoffs, but in practice more parallel blocks can be used (depending on available computational power) to control time-frequency tradeoff more smoothly.

Since mixing of processed signals $x_1[t]$ and $x_2[t]$ is performed in the transform domain, the present method allows achieving of arbitrary time-frequency resolution in arbitrary areas of time-frequency plane.

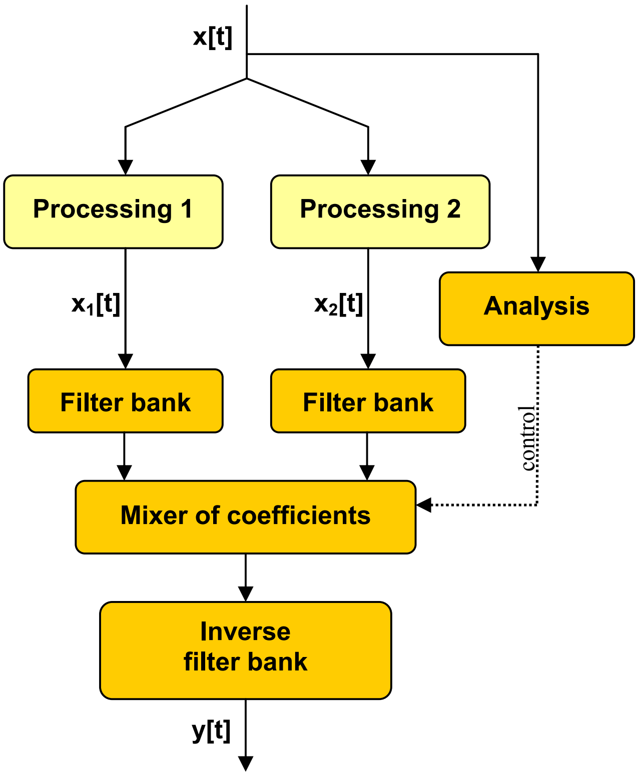

Fig. 1: Scheme of the proposed filter bank.

3.2 Denoising of images

Variants of a discrete wavelet transform are most popular as energy compaction transforms for image denoising [2]. They are computationally efficient and at the same time allow to analyze the image in several scales and compact several edge directions relatively well (usually – vertical and horizontal, in more complex variants – more directions).

The drawback of DWT is that its basis doesn't adapt to local image features. In paper [5] a Principal Component Analysis (PCA) is proposed for design of locally adaptive images basis, and this basis is used as energy compaction transform in denoising. The results of [5] in most cases are superior to wavelet denoising with respect to both visual quality and PSNR.

3.2.1 Application of the suggested model

The drawback of "Adaptive Principal Components" (APCA) method from [5] is a Gibbs phenomenon occurring around edges in the image, and also lack of suppression of low-frequency noises with a period higher than PCA block size. Both of these problems arise because of a fixed spatial resolution of PCA "filter bank". Indeed, all the basis vectors of PCA are supported by the same interval (image block). This means that high-frequency basis vectors contain many oscillations, which leads to the Gibbs phenomenon after modification of transform coefficients (see [6]).

Let's apply the model described in section 2 to the APCA method in order to reduce the support area (and the number of oscillations) of high-frequency basis vectors of APCA. In this case, the space-frequency resolution of the method is controlled by PCA block size. Let's process the input image with APCA algorithm with 2 different block sizes (we suggest 6x6 and 16x16 pixel blocks, but it can be selected depending on the image size and scale of details). As a result, we get two denoised images: one will have Gibbs phenomenon effectively suppressed, and another will have low-frequency noise removed well. To obtain the result we just need to combine these two images using a filter bank. We suggest using a non-decimated DWT [2] as a mixer filter bank. Both images are transformed with the DWT, and two upper levels of wavelet coefficients are taken from the first image, while the rest of (low-pass) coefficients are taken from the second image. In this way, we take a high-quality low frequency band from the second image, and Gibbs-free high frequency band from the first image. After performing the inverse DWT we get the resulting image (fig. 2b).

Let's note that in this method we didn't use all the capabilities of mixer of coefficients of our model. We didn't perform any image analysis and didn't adapt mixing rules to local image features. This is justified by the fact that APCA method itself adapts the shape of its basis vectors to local image features.

3.2.2 Additional modifications of the APCA method

In order to further improve the quality of denoising, we suggest the following modifications of APCA method.

1. Rotation of the local color space to line up with Color Direction Vector (CDV) [7] allows better image energy compaction near edges of different color hue. All the pixels of a central APCA block and training blocks are transformed into a new color space whose axes are principal components of a set of pixel colors of the training area. The noise reduction is performed in this space, and before writing the denoised block to the processed image the inverse transformation of color into the original color space is performed.
2. Selection of training image vectors using proximity of their content to the central processed block. In the original method of [5], training vectors are formed by image blocks that are spatially close to the central block. We suggest adding another criterion of content similarity. We calculate a pixel-by-pixel color difference of training blocks with a central block and select only a half of all training blocks – those whose sum of absolute differences with the central block is minimal.

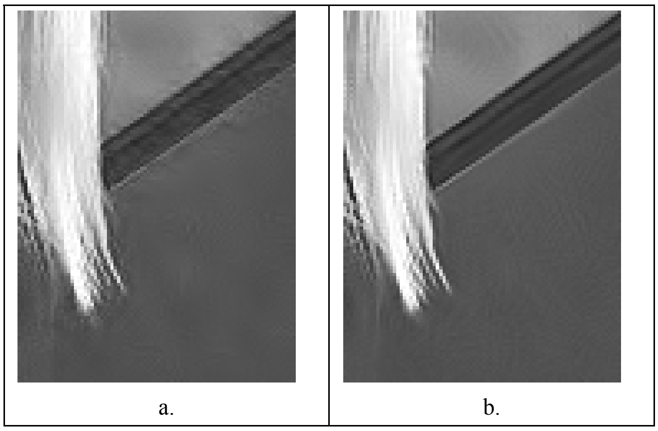

Fig. 2: **a** – result of the wavelet denoising, **b** – result of the proposed method.

These modifications have additionally improved the quality of denoising – the sharpness of edges has improved and Gibbs phenomenon has been further reduced.

3.2.3 Speed optimizations

The processing of image with APCA method with a block size 16x16 is a very time consuming operation. But we can note that only low frequencies are taken from the resulting processed image. This means that the noise reduction can be performed on the downsampled image copy, and then – the result can be upsampled to the original size. As a result of such operation we lose high frequencies of the image, but the useful low frequencies are not harmed. In our work, we decimate the image by a factor of 2 (with anti-aliasing filtering), process it with APCA denoising (the equivalent PCA block size is 8x8 since the scale of details is two times smaller after decimation), and upsample the result back to the original size using bicubic interpolation. After that, according to the section 3.2.1, we substitute high frequencies from another image processed with 6x6 block size.

Another possible optimization leading to some quality compromise is processing of smooth image areas with simpler methods (e.g. wavelet denoising) and use of the suggested PCA-based method only in detailed image areas.

3.3 Denoising of audio

Let's consider the problem of removing of stationary noise from audio signals. The standard and most successful method for this is spectral subtraction. It subtracts (in some generalized sense) the spectrum of noise from a short-time spectrum of a noisy signal [3].

One of drawbacks of spectral subtraction method is a fixed time-frequency resolution of a filter bank. All known noise reduction systems for audio are using STFT-based filter banks. This results in a constant frequency resolution throughout the whole frequency range and constant temporal resolution during the whole sound.

Let's consider effects of a filter bank time-frequency resolution on the resulting sound after denoising. High frequency resolution enables to effectively filter out tonal noises (interference) and to effectively suppress noise when the useful signal is dominantly tonal and stationary. But when transients (sharp attacks) are occurring in the signal, high frequency resolution and low temporal resolution result in slow reaction of the denoiser to changes of local signal-to-noise ratio. As a result, noisy pre-echoes and post-echoes can occur around transients, and transients themselves can smear in time (Gibbs phenomenon) or even get excessively suppressed (because of temporal "inertness" of the denoiser).

When frequency resolution is reduced, the effects of temporal smearing are reduced, but the noise suppression for tonal signals and suppression of tonal noises becomes less effective because of worse energy compaction by a filter bank. When frequency resolution becomes too low, and width of frequency bands comes close to width of human hearing critical bands [4], noise suppression starts producing unpleasant audible effects because of strong variations of signal level in critical bands of human hearing system.

Let's apply the model described in section 2 to the spectral subtraction method. We will process the signal with 3 parallel STFT-based denoisers with different window sizes (e.g. 200 ms, 50 ms, and 12 ms). We get 3 resulting signals processed with different time-frequency resolution. The next step is to combine these 3 signals in a time-frequency domain to obtain the final result. The mixing is performed in transform space of STFT filter bank with a single common time-frequency resolution (we suggest setting STFT window size to 12 ms). The mixer is controlled by some prior strategy reflecting time-frequency properties of a human hearing, and analyzer of signal transience. As a prior strategy we suggest using higher frequency resolution below 4 kHz. The frequency of 4 kHz has been selected as crossover frequency between mostly tonal part of music and speech signals (below 4 kHz) and mostly noisy part (above 4 kHz). For example, vocal and speech contain most harmonics below 4 kHz (including formants [4]), while above 4 kHz there are mostly sibilant noises and transients.

The second part controlling the mixer of coefficients is analyzer of signal transience that estimates transience of the input signal at each time moment. This estimate biases the work of mixer towards better temporal resolution during transients and better frequency resolution during stationary parts.

The overall suggested strategy of mixing coefficients is as follows:

$$Y[f,t] = \begin{cases} (1-\alpha)X_1[f,t] + \alpha X_2[f,t], & f \le 4kHz \\ (1-\alpha)X_2[f,t] + \alpha X_3[f,t], & f > 4kHz \end{cases}$$

Here $X_1[f,t]$, $X_2[f,t]$ and $X_3[f,t]$ are STFT coefficients of signals processed with different time-frequency resolution (starting from high frequency resolution and ending with low frequency resolution), $Y[f,t]$ are STFT coefficients of the resulting mixed signal, α is the output value of transience estimator (from 0 to 1, higher values meaning stronger transience).

The transience analyzer we used is a modification of the algorithm from [8]. This analyzer detects sharp onsets of signal energy in several frequency bands of audio signal. It's easy to see how this algorithm can be used for independent estimation of transience in several frequency bands. This data can be used for finer control of mixer of coefficients (i.e. control of time-frequency resolution of the resulting filter bank).

3.4 Other applications

The suggested model of the adaptive filter bank has been used in several other problems of audio processing. One of them is suppression of a central channel in a stereo recording. The traditional approach to this problem includes suppression of STFT coefficients that are close in magnitude and phase across two stereo channels. Application of the adaptive filter bank, as described in section 3.3, has improved the preservation of transients by reducing the Gibbs phenomenon, and increased the vocal reduction in mid-range frequencies by using higher frequency resolution.

Another application of the suggested model has been found for audio time stretching algorithms based on vocoders [8]. Use of our algorithm has allowed adaptation of time-frequency resolution of a vocoder to properties of audio signals, which again resulted in reduction of Gibbs phenomenon and improvement of time scaling on stationary signal segments.

Also we plan to apply our model of multi-scale APCA denoising to the problem of image sharpness enhancement. This operation can be performed simultaneously with denoising, similarly to amplification of certain wavelet coefficients during wavelet denoising.

4. RESULTS

The suggested model of adaptive filter bank significantly improves quality of different image and audio processing algorithms. The common feature of improvements is reduction of Gibbs phenomenon and increase of frequency resolution during stationary signal parts.

In image denoising, the model has been used in combination with APCA method [5] resulting in reduction of low-frequency noise and suppression of noisy contours around image edges. Both visual assesments [6] and PSNR measurements (table 1) register the quality improvement.

Method	**PSNR, dB**
Noisy image	32.73
Non-decimated DWT	40.04
Modified APCA	40.12
Proposed method	40.29

Table 1: Average PSNR values for a set test images.

The application of the proposed model to problems of audio processing has enabled to reduce time smearing of transients and to improve frequency resolution during processing of stationary parts (in particular – getting stronger noise suppression).

5. REFERENCES

[1] P.P. Vaidyanathan “Multirate Systems and Filter Banks”. *Prentice Hall, 1993, ISBN 0-13-605718-7.*

[2] S. Grace Chang, B. Yu, M. Vetterli “Spatially Adaptive Wavelet Thresholding with Context Modeling for Image Denoising”. *IEEE Trans. Image Processing, vol. 9, no. 9, pp. 1522-1531, Sept. 2000.*

[3] J. Thiemann “Acoustic Noise Suppression for Speech Signals Using Auditory Masking Effects”. *Ph.D. thesis, Department of Electrical & Computer Engineering, McGill University, Montreal, Canada, July 2001.*

[4] I.A. Aldoshina “Basics of Psychoacoustics” (in Russian). *«Audio Producer», vol. 6, 1999.*

[5] D.D. Muresan, T.W. Parks "Adaptive Principal Components and Image Denoising". *IEEE International Conference on Image Processing, September, 2003.*

[6] A. Lukin et al. “Noise Reduction for Digital Images”. *Demo web page* http://audio.rightmark.org/lukin/graphics/denoising.htm

[7] A. Lukin, D. Kubasov “An Improved Demosaicing Algorithm”. *Graphicon-2004 Conference Proceedings.*

[8] J. Bonada “Audio Time-Scale Modification in the Context of Professional Audio Post-production”, *Ph.D. thesis, Universitat Pompeu Fabra, Barcelona, 2002.*

About the authors

Alexey Lukin is a Ph.D. student and Daria Kalinkina and Denis Kubasov are students of Computer Science faculty of the State University of Moscow. See our e-mails on the first page.

Authors would like to thank their scientific supervisor, head of Graphics & Media Lab, Dr. Y.M. Bayskovski, Dr. A.S. Krylov for useful discussions on the subject, Drs. D.S. Vatolin and A.V. Pereberin for introduction to the topic.