Novel Progressive Region of Interest Image Coding Based on Matching Pursuits

Abbas Ebrahimi-Moghadam and Shahram Shirani

Department of Electrical and Computer Engineering, McMaster University

{ebrahia, shirani}@ mcmaster.ca

Abstract— **A progressive and scalable, region of interest (ROI) image coding scheme based on matching pursuits (MP) is presented. Matching pursuit is a multi-resolutional signal analysis tool and can be employed in order to progressively refine the quality of a set of selected regions of an image up to a specific grade. The computational complexity of this analysis method can be reduced by decreasing the size of MP dictionary. Thus, the proposed method provides a trade off between complexity, rate, and quality. By the suggested scheme, regions of an image with higher receiver's priority are refined in an interactive manner. The transmitter sends an initial coarse version of the image. Then, he receiver transmits its preferred ROI parameters. Afterwards, the reconstructed image is refined according to the ROI parameters, in a progressive way.**

I. INTRODUCTION

For a decent image browsing experience, efficient delivery of large high resolution images is essential. On the other hand, because of increasing number of Internet users and their transmitted data rate, the available bit-rate for each application has to be restricted. With a limited bit-rate, maintaining the original high visual quality for all parts of a large high resolution image is a time consuming process. To address this problem, a selection of regions of interest of an image (according to the receiver) can be transferred with higher quality, while the remainder of the image is sent at a lower quality.

Matching pursuit is a greedy signal analysis algorithm in which a signal is iteratively decomposed into a linear expansion of waveforms, chosen from an over-complete dictionary [1], [2]. Assume $\mathcal{D} = \{g_\gamma\}_{\gamma \in \Gamma}$ is an over-complete set with finite number of normalized elements ($\|g_\gamma\| = 1$) in $L^2(\mathcal{R})$ that spans a space of dimension N (Γ is a finite dictionary index set of N_D elements). Each vector f of the space can be projected on a dictionary element $g_\gamma \in \mathcal{D}$ to approximate the vector in that direction. Matching pursuit algorithm is based on the following residual vector update formula

$$R^n f = < R^n f, g_{\gamma_n} > g_{\gamma_n} + R^{n+1} f \qquad (1)$$

where $R^0 f = f$, $< R^n f, g_{\gamma_n} >$ is the inner product of $R^n f$ with dictionary element g_{γ_n}, and $R^{n+1} f$ is the residual vector of $R^n f$ in g_{γ_n} direction. Matching pursuit algorithm selects a dictionary element which yields minimum residual energy ($\|R^{n+1} f\|^2$). After m iteration of matching pursuit analysis we have :

$$f = \sum_{n=0}^{m} < R^n f, g_{\gamma_n} > g_{\gamma_n} + R^{m+1} f \qquad (2)$$

The most computationally expensive part of MP algorithm is the inner product calculation ($< R^n f, g_{\gamma_n} >$) in each iteration and for all dictionary elements. Taking an inner product with g_γ on both side of (1) yields the following updating equation:

$$< R^{n+1} f, g_\gamma > = < R^n f, g_\gamma > - < R^n f, g_{\gamma_n} > < g_{\gamma_n}, g_\gamma > \qquad (3)$$

In equation (3), $< R^n f, g_{\gamma_n} >$ and $< R^n f, g_\gamma >$ are computed at the previous iteration. Dictionary inner products $< g_{\gamma_n}, g_\gamma >$ may be calculated and stored in advance of MP analysis to be used like a look up table. With the above updating formula, the MP computational cost can be reduced drastically [1]. In this paper we call this approach Mallat's method.
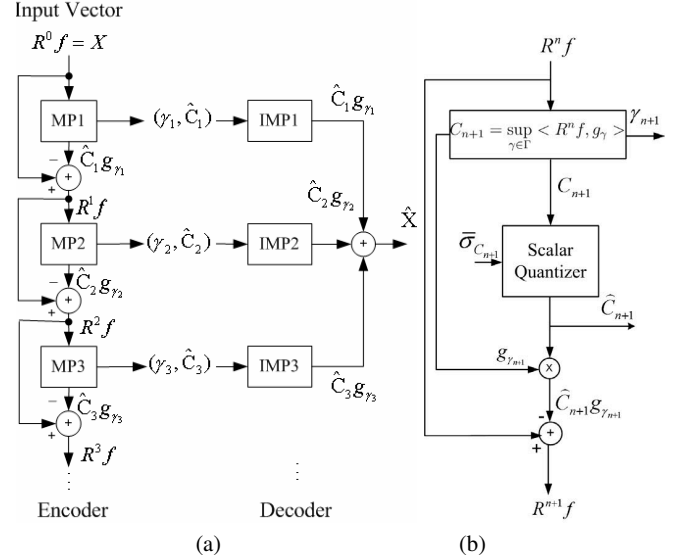

Fig. 1. (a) Schematic block diagram of matching pursuit encoder and decoder (b) More detailed block diagram of $(n + 1)^{th}$ stage of MP analyzer

II. MP BASED ROI IMAGE CODING

A schematic diagram of MP encoder and decoder is shown in fig. 1(a). An input signal X is applied to the first MP analysis stage (MP1). Here the signal is compared with all dictionary members g_γ and an element g_{γ_1}, resulting an inner product with the maximum magnitude, is chosen. Since we can only transmit analysis parameters with finite precision, the inner product coefficient C_1 is then quantized to $\hat{C}_1$. In order to reduce the quantization error effect of inner product coefficients on MP analysis performance, the quantized version of inner product coefficient $\hat{C}_1$ is then employed to calculate the residual signal $R^1 X = X - \hat{C}_1 g_{\gamma_1}$. The quantized coefficient $\hat{C}_1$ along with the index of the selected dictionary element γ_1 is sent to the decoder. Now, $R^1 X$ is the input to the second MP analysis stage (MP2). The resulting $\hat{C}_2$ and γ_2 are transmitted and $R^2 X = R^1 X - \hat{C}_2 g_{\gamma_2}$ is used for the next MP analysis stage. This procedure continues and the corresponding analysis parameters are sent. Fig. 1(b) illustrates the above process at the $(n + 1)^{th}$ MP analysis stage. As shown in this figure, the scalar quantizer of each stage is adapted based on the standard deviation of the inner product coefficients in that MP analysis stage. At the receiver side the signal X is reconstructed using the received information $\hat{C}_i$, and γ_i by the following linear combination formula:

$$\hat{X} = \sum_{i=1}^{m} \hat{C}_i g_{\gamma_i} \qquad (4)$$

where m is the number of MP analysis stages.

A. *Progressive ROI refinement*

Since matching pursuit algorithm is a multi-resolutional signal analysis method, it is a proper choice for progressive ROI image coding. The key idea for MP based ROI image coding is to apply more MP analysis stages and extract more structural features for the ROI rather than the background. To apply this intuitive idea, an image is divided to N_B square image blocks. Using the mean value of each image block, a coarse version of the image is sent to the receiver. The receiver then indicates his/her point(s) of interest (i.e. center(s) of ROI(s)) and transmits the corresponding coordinates of the point(s) to the transmitter. The level of interest for the jth image block is considered to be a function of d_j, the minimum distance of the center point of the block from the point(s) of interest, i.e.,

$$d_j = \min_i d_j^{(i)} \quad i = 1, 2, ..., N_{ROI} \quad j = 1, 2, ..., N_B \quad (5)$$

where $d_j^{(i)}$ is the distance of the jth image block from i^{th} point of interest and N_{ROI} is the number of points of interest. To refine the quality of the initial coarse image with emphasis on ROI in a progressive way, the encoder examines all the image blocks to see if the distances of their center points from the point of interest (d_j) is less than an initial distance value of R_1. If $d_j > R_1$, no refinement information is sent for j^{th} image block. If $d_j \leq R_1$, the residue of the jth image block with respect to the existing version of it at the receiver is approximated using the first stage of the MP analysis. To further improve the quality of the ROI, using the relation

$$R_{k+1} = \alpha \times R_k \qquad \alpha \geq 1 \quad , \quad k = 1, 2, \ldots, N_S \quad (6)$$

where N_S is the maximum number of the available MP stages, the encoder updates the R_k with R_{k+1}. The above routine is repeated for all image blocks using R_{k+1} as the new distance measure. This process results in generation of successive MP refinement layers of the image with more emphasize on the region of interest. The updating multiplier α, in equation (6), specifies how gradual the image quality is reduced from ROI toward the background. This parameter plays an important role in emphasizing the ROI visual information. An appropriate value has to be assigned to α in order to meet the image transmission bit budget and, at the same time, give a desired level of reconstruction fidelity to the ROI compared to the background. If the ROI has an absolute transmission priority and the bit budget is small, α can be set to 1. This implies that the refinement bits are sent merely for the region of interest. If a larger value is assigned to α the distance metric R_k grows more rapidly and more image blocks become eligible for the refinement. The virtual circle with radius R_1 and centered at the point of interest is the initial region which its image blocks are enhanced with the highest priority. Since a bigger R_1 means more initial image blocks with this privilege, there would be no ROI if R_1 exceeds the size of the image. Practical values for R_1 can be in the range of $1/10$ to $1/2$ of image size. Thus, in case of low bit budget or low bit-rate transmission, α is set close to 1 and R_1 is selected in the above mentioned range and a higher priority of transmission will be given to the ROI image blocks.

The complete bit stream generated by above algorithm conveys refinement information for all image blocks with a final quality dictated by the maximum MP analysis stages. The bit stream can be truncated whenever a desirable image quality is provided at the receiver. In this case, the reconstructed image has better quality at region(s) of interest.

The proposed MP-based ROI image coding method allows changing the position of point of interest without losing the previously
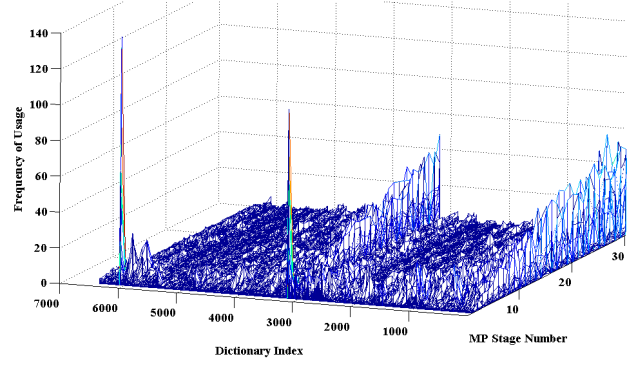

Fig. 2. Dictionary elements usage frequency for different MP stages when MP applied on a set of 8 different images

received refinement data. In order for the transmitter to handle the ROI change, the number of MP stages that each image block has been analyzed with, should be recorded at the encoder. Receiver sends the location of the new point of interest. After this interaction, the transmitter sends the refinement data for image blocks around the newly defined point of interest without the need to retransmit the already sent information. More specifically, let $P_j^{(-1)}$ denote the number of MP stages by which the jth block is analyzed. Let d_j represent the distance of the jth image block from the new point of interest. If $d_j < R_k$, encoder checks $P_j^{(-1)}$ and if $P_j^{(-1)} \geq k$ no transmission for j^{th} image block is required, otherwise, the residue of the image block is approximated using the kth stage of the matching pursuits and the index of the chosen dictionary element and the corresponding inner product coefficients are transmitted and $P_j^{(-1)}$ is updated.

B. *Reduction of computational complexity in MP-based ROI image coding*

The size of MP dictionary has a direct effect on the computational complexity of the MP algorithm. A larger dictionary requires more bits to represent the dictionary index than that for a smaller one. On the other hand, a properly designed larger dictionary would better represent signal structures. Thus, when we choose a dictionary size for MP analysis, there are trade-offs between rate, quality, and complexity. Here, a method for reducing the computational complexity of MP-based ROI image coding is proposed. A huge initial dictionary is used for MP analysis of a large number of test images. According to the residual patterns of image blocks, in each MP stage, different subsets of the dictionary are more often used. We can notice this fact by observing usage frequency of dictionary elements. Fig. 2 shows the usage frequency of dictionary elements as a function of matching pursuit stage number when MP is applied to a set of 8 different images for a Gabor dictionary [3] with 6400 element of 8×8 pixels. As it can be seen from this figure, the usage frequency of dictionary elements is largely skewed and a significant number of elements are rarely used in each stage of MP analysis. Using this prior knowledge, the computational complexity of MP-based ROI image coding can be reduced. The idea is to sort the dictionary for each MP stage based on the frequency of usage of dictionary elements for that stage. The original and sorted dictionaries have exactly the same elements, but elements of the latter are sorted by their frequency of usage in a descending manner. According to our rate-quality-complexity choice of working point, an N_d element subset of the original dictionary can be selected in each MP stage simply by taking N_d first elements

of the corresponding sorted dictionary. Only the corresponding sub-dictionary is searched for the most similar element to the residual signal at each MP stage. Since saving multiple sorted versions of the original huge dictionary for each MP stage is inefficient, in terms of memory requirement, for each stage of matching pursuit, a transition vector is used to map the index from the original dictionary to the sub-dictionary for that stage. This transition vector reduces the storage requirement for sorted dictionary and dictionary inner products (used in Mallat's fast algorithm) to only the original ones. The ith element of the transition vector for the kth MP stage is a pointer to an element in the original dictionary which is identical to the ith element in the sorted dictionary for the kth MP stage. In other words, in order to point to the ith element of the sorted dictionary in k^{th} MP stage the index in the original dictionary (i_{orig}) can be found using the following relation:

$$i_{orig} = trans^k(i) \qquad (7)$$

In (7), $trans^k(\cdot)$ represents the transition vector. The formula can also be used when the elements of the dictionary inner product matrix are required for the Mallat's fast matching pursuit algorithm. To point to the element (i, j) of the inner product matrix of the sorted dictionary at k^{th} MP stage, we simply find the element (i_{orig}, j_{orig}) of inner product matrix associated to the original dictionary, without the necessity of storing another version of the original huge $N_\mathcal{D} \times N_\mathcal{D}$ matrix ($N_\mathcal{D}$ is the number of elements in the original dictionary). This task can be done using

$$(i_{orig}, j_{orig}) = (trans^k(i), trans^k(j)) \qquad (8)$$

III. EXPERIMENTAL RESULTS

In this section we present the simulation results for our proposed MP based ROI image coding technique. Here, we have used a 2D Gabor dictionary with 6400 elements. The dictionary is similar to one used in [3].

Quantization of the inner product coefficients is an integral part of the proposed MP based ROI image coding. Uniform, N_Q level, mid-tread scalar quantizers [4], [5] with step sizes adapted to the standard deviation of inner products in each MP stage are employed. The statistical data for this adaption is extracted from MP analysis of a set of 8 different images ("news" image is not one of those images). The granular region [5] of the quantizer is from $-3\sigma_n$ to $3\sigma_n$, where σ_n is the standard deviation of inner product coefficients at n^{th} MP analysis stage. Fig. 3 demonstrate the progressive image enhancement using MP-based ROI image coding and the fact that the proposed scheme can handle ROI changes. To apply our method of ROI image coding, the image is divided into 8×8 non-overlapping image blocks. The MP dictionary contains 6400 2-D Gabor elements with the size of 8×8. Here the MP analysis is restricted to maximum 5 stages. MP inner products are quantized with a 4-bit ($N_Q = 16$) adaptive uniform scalar quantizer. Fig. 3(a) shows the original monochrome 8 bpp "news" image with indicated regions of interest. The left hand side region is the initial ROI. Fig. 3(b) is the coarse version of "news" image formed by mean value of each image block at 0.0625 bpp. The receiver reconstructs this version of image and indicates the region of interest by sending the coordinates of the point of interest to the transmitter. At this stage, the transmitter starts sending refinement bits generated by the proposed MP based ROI image coding. In figure 3(c), the reconstructed image at the time of ROI change is shown. The analysis parameters are $R_1 = 0.125$ (the unit here is the width of the image), and $\alpha = 1.4$. The image is progressively reconstructed at 0.1424 bpp. As time passes, more refinement bits participate in
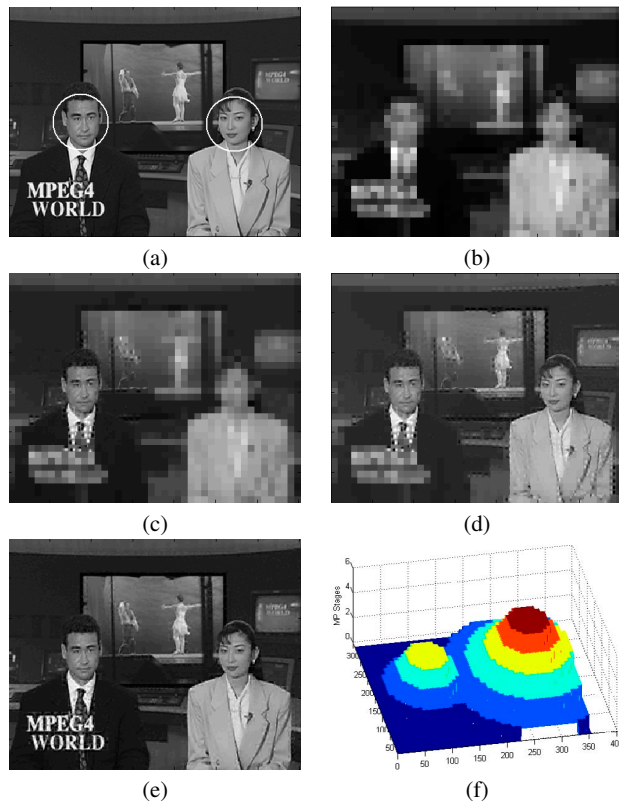

Fig. 3. Changing the point of interest during the course of transmission. (a) Original monochrome 288×352 pixels "news" image represented by 8 bpp. The region of interest is indicated by a circle. (b) A coarse version of image, generated by mean value of each image block, is sent to the receiver with 0.0625 bpp. (c) The refinement information according to the first ROI (left hand side ROI) is transmitted. At this point the receiver changes the ROI choice to the right hand side ROI. MP-based ROI analysis parameters are: $\alpha = 1.4$, and $R_1 = 0.125$. The image is reconstructed at 0.1424 bpp. (e) Completely refined image by 5 stages of matching pursuit analysis bit-stream at 1.3628 bpp. (f) Spatial status of MP stages for part (d)

image enhancement with the emphasis on the second ROI. Fig. 3(d) shows the reconstructed image after receiving some refinement bits according to the second choice of ROI at 0.4544 bpp. In fig. 3(e), we receive the complete bit-stream at 1.3628 bpp. In this figure, all image blocks are enhanced with 5 stages of MP refinement bits. Fig. 3(f) demonstrates the number of MP analysis stages that have been used in different locations of the image. This MP stage status presentation is associated with fig. 3(d).

The proposed MP based ROI image coding is capable of providing compromises among bit rate, quality of the reconstructed image, and the computational complexity of MP analysis. It is essential to know the inter-relation among rate, quality, and complexity in order to choose the best MP analysis set-up. Following are experimental results which provide this essential knowledge which is brought by MP analysis and synthesis of a set of 8 different images. The analysis parameters, based on the proposed MP based ROI image coding, are $\alpha = 1.4$, and $R_1 = 0.125$. For different stages of MP algorithm, the original dictionary (i.e., 6400-element 2-D (8×8) Gabor dictionary) is sorted and truncated to a dictionary of size $N_D = 2^{B_D}$ elements of 8×8, 2-D Gabor functions.

Fig 4 shows the quality variation of the reconstructed ROI coded images for ROI region and for whole image as a function of number
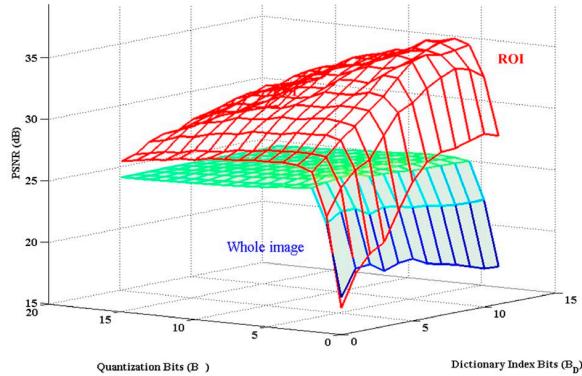
Fig. 4. PSNR versus dictionary size and quantization step number for a fixed rate of 0.3 bpp
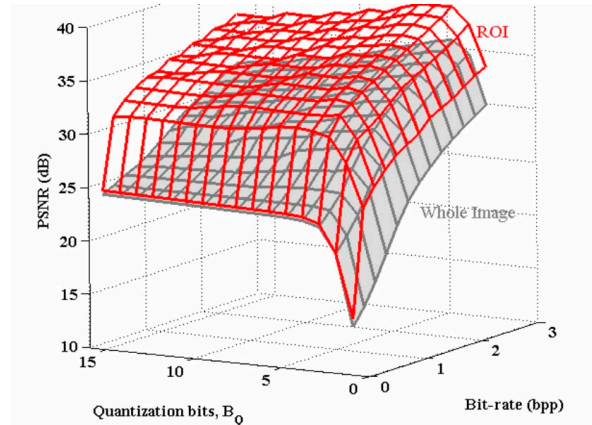

Fig. 5. Rate-distortion performance of the MP based ROI image coding as a function of quantization step number of the MP inner product coefficients when the dictionary size is fixed to $N_D = 512$
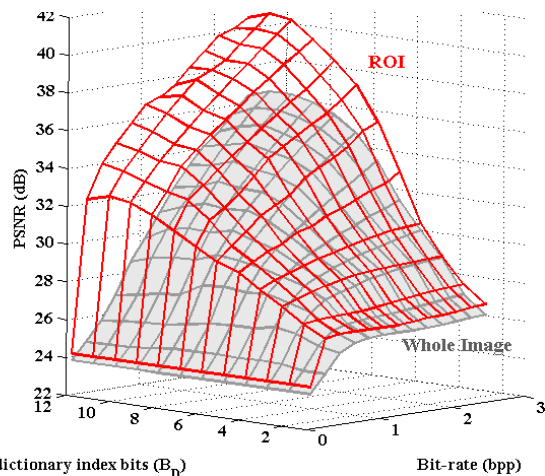

Fig. 6. Rate-distortion behavior of the MP-based ROI image coding scheme as a function of MP dictionary size

of quantization steps ($N_Q = 2^{B_Q}$) and MP dictionary size ($N_D = 2^{B_D}$) for a given fixed rate of 0.3 bpp. According to this figure, quantization of inner product coefficients with 3 or 4 bits results in better quality performance when a fixed rate, i.e. 0.3 bpp, is targeted. As it can be seen from the graphs, for a subset of original dictionary with just 4 elements, the quality performance is far worse than the other dictionary sizes. In this case, the dictionary is under-complete and not a good representative of all features of image blocks. According to this figure, the quality performance for $N_D = 64, 256, 1024,$ and 4096 are comparable. Although bigger dictionary, i.e. more complexity, almost always yields better quality performance, the computational cost sometimes is too much, since adding one more bit to the dictionary index bits means doubling the computational burden. Figures. 5 illustrates the quality of the output image of the proposed scheme as a function of bit-rate and the inner product coefficients quantization step number (B_Q). Here the dictionary size is set to $N_D = 2^9 = 512$. According to this figure, the worst quality is obtained when the coarse version of image is sent with minimum number of quantization steps, i.e. $B_Q = 1$. Having only the coarse version of image at the receiver, the qualities at ROI and the whole image are almost the same. Refinement information separates the quality surfaces and alleviates ROI quality faster than the whole image. These two surfaces approach each other when the whole image is refined with maximum number of MP stages. According to this figure, the best choice of quantization step number is $N_Q = 2^4 = 16$. Figures 6 demonstrate the rate-distortion behavior of the proposed ROI image coding scheme as a function of dictionary size. For this figure, the number of quantization steps is fixed to $N_Q = 2^4 = 16$. According to this figure, MP dictionary of very small size yields little quality improvement with refinement bits. On the other hand, a large MP dictionary results in very rapid quality improvement by receiving refinement bits especially for the ROI. According to this figure, the rate-distortion performances for large MP dictionaries, e.g. $B_D =8$ to 12, are close. When the computational burden is strictly limited and the bit-budget is fixed, decrementing B_D by one bit means reducing the computational cost to half.

IV. CONCLUSION

In this work we introduced a new ROI image coding approach based on matching pursuits. The method interactively allows the transmitter to send MP analysis data with the emphasis on ROI(s). The ROI parameters include the radius of initial virtual circle (R_1) and the updating multiplier (α). Depending on the bit-budget, the degree of importance of the ROI(s), and the computational capability of the transmitter, a proper ROI parameter set and MP dictionary size can be selected.

The proposed MP based ROI image coding is computationally unbalanced and the computation cost at the decoder is much lower than that at the encoder. For the case that the receiver is not very well equipped, the proposed scheme concentrate the complexity into the transmitter and leaves the receiver very simple.

REFERENCES

[1] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Processing*, vol. 41, pp. 3397-3415, Dec. 1993.

[2] S. Mallat, *A Wavelet Tour of Signal Processing*. Second edition, Academic Press 1999.

[3] R. Neff, A. Zakhor, "Matching pursuit video coding at very low bit rates," *IEEE Trans. Circuits Syst. Video Technology*, pp 158-171, Feb. 1997.

[4] A. Gersho and R. Gray, *Vector Quantization and Signal compression*, Kluwer academic publishers, Boston, 1992.

[5] K. Sayood, *Introduction to Data Compression*. Second edition, Kaufmann publishers 2000.