Video Compression Based on Orthonormal Matching Pursuits

Jian-Liang Lin†*, Wen-Liang Hwang†, and Soo-Chang Pei*
*Institute of Communication Engineering, National Taiwan University, Taiwan, R.O.C.
†Institute of Information Science, Academia Sinica, Taipei, Taiwan, R.O.C.

***Abstract*— Video coding schemes based on matching pursuits have been shown to have better coding efficiency and perceptual quality at low bit rates than DCT-based video coding schemes. Matching pursuit is a greedy algorithm that decomposes a signal into a linear combination of bases within an overcomplete dictionary. However, as the bases are not independent, the redundancy in linear combinations increases as the number of iterations of matching pursuit increases. In this paper, we introduce an orthonormal matching pursuit algorithm that reduces the redundancy between bases and represents a signal by a linear combination of bases more efficiently. Our proposed updated full search matching pursuit algorithm has been shown to be effective in both coding performance and computational complexity. Based on this algorithm, we present an orthonormal matching pursuit video coding scheme that achieves a more efficient coding performance. The coding efficiency and perceptual quality are evaluated and compared to traditional matching pursuit video coding schemes.**

I. Introduction

Efficiently encoding motion residuals is essential for low-delay video applications in which videos are encoded by hybrid motion compensation and a residual encoding structure. As well as nonredundant transformation, a frame-based technique, called matching pursuit (MP), has been proposed to encode motion residual images. Mallat and Zhang [5] were the first to propose a matching pursuit algorithm that decomposes a signal into a linear combination of bases within an overcomplete dictionary. In [6], Neff and Zakhor show that using a matching pursuit algorithm to encode motion residual images achieves a better performance than a discrete cosine transform (DCT) in terms of PSNR and perceptual quality at very low bit rates. The results in [3] also demonstrate that the fined-grained scalable (FGS) MP coding scheme performs better than MPEG-4 FGS at very low bit rates. Unlike a transform-based decoder, an MP decoder does not require an inverse transform; therefore, it is less complex. In a transform-based decoder, loop filtering and post processing are usually applied at very low bit rates to remove blocking and ringing artifacts, whereas an MP decoder can achieve comparable quality without such filtering and processing. Because the matching pursuit algorithm is a data-dependent frame-based representation, a matching pursuit video coding technique cannot be directly translated from conventional transform-based approaches. New matching pursuit video coding techniques have therefore been developed to deal with quantization noise in the matching pursuit algorithm [7], scalable bitstream generation [1], [8], and dictionary learning and adaptation.

Matching pursuit is a greedy algorithm that decomposes a signal into a linear combination of bases, but as the bases are not independent, there are many ways to represent a signal. The algorithm selects the maximum basis to represent a signal at each iteration; however, it can not guarantee optimal linear combinations. In fact, since the bases are redundant, as the number of iterations increases, the redundancy in linear combinations also increases. To solve this problem, an orthogonal matching pursuit algorithm has been proposed to reduce the redundancy between bases [2]. At each iteration, the projection of bases on the selected orthogonal basis is removed, but it does not normalize the orthogonal basis; therefore, the basis selection unfair. To ensure that all bases have the same norm, instead of selecting the maximum absolute inner product between a signal and the orthogonal basis, we propose an orthonormal matching pursuit algorithm that finds the maximum absolute inner product after normalization. By combining the orthonormal matching pursuit algorithm with an updated full search matching pursuit video coding scheme, we present an effective orthonormal matching pursuit video coding scheme.

The remainder of the paper is organized as follows. In Section 2, we propose an orthonormal matching pursuit algorithm that efficiently approximates a signal as a linear combination of orthonormal bases. In Section 3, we present the orthonormal matching pursuit video coding scheme based on an updated full search matching pursuit algorithm. In Section 4, we present an evaluation of the coding performance of the orthonormal matching pursuit video coding scheme and compare it to a traditional matching pursuit video coding scheme. Finally, we present our conclusions in Section 5.

II. Orthonormal Matching Pursuits

Since the dictionary of matching pursuit $\mathcal{D} = (g_\gamma)_{\gamma \in \Gamma}$ is redundant, the dictionary selected to represent a signal f does not have to be unique. To present f as a linear combination of g_γ efficiently, we propose an orthonormal matching pursuit algorithm that successively approximates f with orthogonal projections onto elements of $\mathcal{D}$. Let $g_{\gamma_0} \in \mathcal{D}$. At the first iteration, the signal f can be decomposed into

$$f = < f, g_{\gamma_0} > g_{\gamma_0} + Rf, \tag{1}$$

where Rf is the residual vector after approximating f in the direction of g_{γ_0}. To minimize $\|Rf\|$, we must choose $g_{\gamma_0} \in D$

such that $|< f, g_{\gamma_0} >|$ is the maximum, i.e.,

$$|< f, g_{\gamma_0} >| \geq \sup_{\gamma \in \mathbf{\Gamma}} |< f, g_\gamma >|. \tag{2}$$

We define $u_0 = g_{\gamma_0}$ as the first orthonormal basis selected to represent f. Clearly u_0 is orthogonal to Rf, hence

$$\|f\|^2 = |< f, u_0 >|^2 + \|Rf\|^2. \tag{3}$$

Since the projection of f on u_0 has been subtracted, we also remove the projection of other bases on u_0. Let $\xi'_\gamma = g_\gamma - < g_\gamma, u_0 > u_0$ denote the residual element by removing the projection of g_γ on the previously selected element u_0. Then, we update the dictionary $\mathcal{D} = (g_\gamma)_{\gamma \in \mathbf{\Gamma}}$ to $\mathcal{D}' = (g'_\gamma)_{\gamma \in \mathbf{\Gamma}}$ by normalizing the vector ξ'_γ:

$$g'_\gamma = \frac{\xi'_\gamma}{\|\xi'_\gamma\|} = \frac{g_\gamma - < g_\gamma, u_0 > u_0}{\|g_\gamma - < g_\gamma, u_0 > u_0\|}. \tag{4}$$

Similarly, u_0 is orthogonal to ξ'_γ, so the norm of $\|g_\gamma\|$ can be obtained by

$$\|g_\gamma\|^2 = |< g_\gamma, u_0 >|^2 + \|\xi'_\gamma\|^2. \tag{5}$$

We sub-decompose the residue Rf by projecting it onto the vector $g'_{\gamma_1} \in \mathcal{D}'$ that best matches Rf, as we did for f. The projection of Rf generates a second residue, $R^2 f$. We then define $u_1 = g'_{\gamma_1}$ and update the dictionary $\mathcal{D}'$ to $\mathcal{D}^2$. This procedure is repeated at each iteration. After n iterations, we have computed the nth order residue $R^n f$ and the dictionary has been updated to $\mathcal{D}^n = (g^n_\gamma)_{\gamma \in \mathbf{\Gamma}}$. We choose the element $g^n_{\gamma_n}$ in the dictionary $\mathcal{D}^n$ that most closely matches the residue $R^n f$, i.e,

$$|< R^n f, g^n_{\gamma_n} >| \geq \sup_{\gamma \in \mathbf{\Gamma}} |< R^n f, g^n_\gamma >|. \tag{6}$$

Let $u_n = g^n_{\gamma_n}$ denote the nth orthonormal basis selected. The residue $R^n f$ is then sub-decomposed into

$$R^n f =< R^n f, u_n > u_n + R^{n+1} f. \tag{7}$$

Since $R^{n+1} f$ is orthogonal to u_n, we have

$$\|R^n f\|^2 = |< R^n f, u_n >|^2 + \|R^{n+1} f\|^2. \tag{8}$$

Likewise, we update the dictionary $\mathcal{D}^n = (g^n_\gamma)_{\gamma \in \mathbf{\Gamma}}$ to $\mathcal{D}^{n+1} = (g^{n+1}_\gamma)_{\gamma \in \mathbf{\Gamma}}$ by computing

$$g^{n+1}_\gamma = \frac{\xi^{n+1}_\gamma}{\|\xi^{n+1}_\gamma\|}, \tag{9}$$

where

$$\xi^{n+1}_\gamma = g^n_\gamma - < g^n_\gamma, u_n > u_n \tag{10}$$

denotes the new element by removing the projection of g^n_γ on the previously selected element u_n. Similarly, ξ^{n+1}_γ is orthogonal to u_n.

$$\|g^n_\gamma\|^2 = |< g^n_\gamma, u_n >|^2 + \|\xi^{n+1}_\gamma\|^2. \tag{11}$$

The updated element g^{n+1}_γ becomes the residual element by deleting the projection of the original vector g_γ on the previously selected element $\{u_i\}_{0 \leq i \leq n}$. Let

$$p^{n+1}_\gamma = g_\gamma - \sum_{i=0}^{n} < g_\gamma, u_i > u_i. \tag{12}$$

Then, we have

$$g^{n+1}_\gamma = \frac{p^{n+1}_\gamma}{\|p^{n+1}_\gamma\|}, \tag{13}$$

and

$$\|p^{n+1}_\gamma\|^2 = \|g_\gamma\|^2 - \sum_{i=0}^{n} |< g_\gamma, u_i >|^2. \tag{14}$$

Let $R^0 f = f$. If we decompose f by iterating the process m times, f will be decomposed into a telescoping sum as follows:

$$f = \sum_{n=0}^{m-1} \left(R^n f - R^{n+1} f\right) + R^m f. \tag{15}$$

Equation (7) then yields

$$f = \sum_{n=0}^{m-1} < R^n f, u_n > u_n + R^m f. \tag{16}$$

Since $\{u_n\}_{0 \leq n \leq N-1}$ and $R^N f$ are orthogonal, we have

$$\|f\|^2 = \sum_{n=0}^{m-1} |< R^n f, u_n >|^2 + \|R^m f\|^2. \tag{17}$$

Thus, the original vector f is decomposed into the sum of dictionary elements that best match f. G. Davis and S. Mallat [2] proved that the matching pursuit algorithm converges, even in infinite dimensional spaces. Similarly, we have proven that the orthonormal matching pursuit algorithm is also convergent.

III. Orthonormal Matching Pursuit Video Coding

We have proposed an updated full search algorithm to obtain better coding performance and reduce encoding time [4]. In this section, we describe an orthonormal matching pursuit video coding scheme based on the updated full search algorithm. At each iteration, after an optimal atom g_{γ_n} has been selected, it is transformed into an orthonormal basis u_n by using the Gram-Schmidt algorithm.

$$u_n = \sum_{p=0}^{n} b_{p,n} g_{\gamma_p}. \tag{18}$$

We define $p^{n+1}_{\gamma_{i,l}}$ as the orthogonal basis without normalization by removing the projection of $g_{\gamma_{i,l}}$ on the previous selected orthonormal bases $\{u_i\}, i = 0, 1, ..., n$. The inner products between the residual image and $p^{n+1}_{\gamma_{i,l}}$ are obtained by the following updated equation.

$$\begin{aligned} &< R^{n+1} f, p^{n+1}_{\gamma_{i,l}} >= \\ &\quad < R^n f, p^n_{\gamma_{i,l}} > - < R^n f, u_n >< u_n, g_{\gamma_{i,l}} >, \end{aligned}$$

where

$$< u_n, g_{\gamma_{i,l}} >= \sum_{p=0}^{n} b_{p,n} < g_{\gamma_p}, g_{\gamma_{i,l}} > . \qquad (19)$$

The norm of $p^{n+1}_{\gamma_{i,l}}$ after removing the projection of $g_{\gamma_{i,l}}$ on the new selected orthonormal basis u_n is

$$\|p^{n+1}_{\gamma_{i,l}}\|^2 = \|p^{n}_{\gamma_{i,l}}\|^2 - | < g_{\gamma_{i,l}}, u_n > |^2. \qquad (20)$$

Then, at the next iteration, another atom $g_{\gamma_{n+1}}$is selected to maximize the absolute inner product between the residual image and the orthonormal basis $p^{n+1}_{\gamma_{i,l}}/\|p^{n+1}_{\gamma_{i,l}}\|$.

$$\gamma_{n+1} = \max_{\gamma_{i,l}} \frac{| < R^{n+1}f, p^{n+1}_{\gamma_{i,l}} > |}{\|p^{n+1}_{\gamma_{i,l}}\|}. \qquad (21)$$

This algorithm is repeated until the bit budget is reached, or the error of the residual image is less than a given threshold. We now describe our orthonormal updated full search algorithm.

Orthonormal Updated Full Search Algorithm

1) **Initialization** : The residual image is first divided into blocks. For each block, if the energy is larger than a given threshold, i.e.,

$$Energy \geq \eta \times MaximumEnergy, \qquad (22)$$

where $0 < \eta \leq 1$, we calculate the inner product between f and each basis within the current block i. L atoms with the largest inner product values, $\{g_{\gamma_{i,l}}\}$, are then recorded. Let $< R^0 f, p^0_{\gamma_{i,l}} >=< f, g_{\gamma_{i,l}} >$ and $\|p^0_{\gamma_{i,l}}\| = 1$. $n = 0$.

2) **Maximum atom extraction**: At the nth iteration, from the blocks that have been processed, we find the atom with the maximum inner product, g_{γ_n}.

$$\gamma_n = \max_{\gamma_{i,l}} \frac{| < R^n f, p^n_{\gamma_{i,l}} > |}{\|p^n_{\gamma_{i,l}}\|}. \qquad (23)$$

The orthonormal basis u_n can be obtained by

$$\begin{aligned} u_n &= \frac{g_{\gamma_n} - \sum_{i=0}^{n-1} < g_{\gamma_n}, u_i > u_i}{\|p^n_{\gamma_n}\|} \\ &= \frac{g_{\gamma_n} - \sum_{i=0}^{n-1}\sum_{p=0}^{i} b_{p,i} < g_{\gamma_n}, u_i > g_{\gamma_p}}{\|p^n_{\gamma_n}\|} \\ &= \sum_{p=0}^{n} b_{p,n} g_{\gamma_p}. \end{aligned}$$

The residual image

$$R^{n+1}f = R^n f - < R^n f, u_n > u_n. \qquad (24)$$

3) **Inner product update**: Next, we update the inner products of L atom candidates for each block, i.e.,

$$\begin{aligned} &< R^{n+1}f, p^{n+1}_{\gamma_{i,l}} >= \\ &\quad < R^n f, p^n_{\gamma_{i,l}} > - < R^n f, u_n >< u_n, g_{\gamma_{i,l}} >, \end{aligned}$$

where

$$< u_n, g_{\gamma_{i,l}} >= \sum_{p=0}^{n} b_{p,n} < g_{\gamma_p}, g_{\gamma_{i,l}} > . \qquad (25)$$

The norm of $p^{n+1}_{\gamma_{i,l}}$ is

$$\|p^{n+1}_{\gamma_{i,l}}\|^2 = \|p^{n}_{\gamma_{i,l}}\|^2 - | < g_{\gamma_{i,l}}, u_n > |^2. \qquad (26)$$

4) **Search within new blocks**: After an atom has been extracted, the energy of some blocks may be affected. If the energy of a block that has not been searched is larger than the current maximum energy multiplied by η, we calculate the inner product between $R^{k+1}f$ and each basis within this block. L atoms with the largest inner product values, $\{g_{\gamma_{i,l}}\}$, are then recorded. Let $< R^{n+1}f, p^{n+1}_{\gamma_{i,l}} >=< R^{n+1}f, g_{\gamma_{i,l}} >$ and $\|p^{n+1}_{\gamma_{i,l}}\| = 1$.

5) **Next iteration**: $n = n + 1$, go to item 2.

IV. Performance Evaluation

We now evaluate the coding performance of the orthonormal matching pursuit video coding scheme and compare it to the traditional matching pursuit (non-orthonormal matching pursuit) video coding scheme. Both schemes are based on the updated full search matching pursuit video codec proposed in [4]. The video codec is a hybrid motion compensation and matching pursuit residual coding system. The first frame of a video sequence is an intra-frame (I-frame) encoded by DCT; all other frames are inter-frames (P-frames). The sequences are in QCIF format and the test frame rate is ten frames per second. The difference between the performance of the two schemes is based on the number of atoms encoded in each frame. If only a few atoms are encoded in one frame and they are dispersed over different positions, there is a high probability that the inner product between any two atoms will be zero. In this scenario, the atoms are orthogonal initially, so there is no difference between the performance of the two codecs. Figure 1 shows the performance difference based on the number of iterations of matching pursuit in one frame. When the number of iterations is small, the performance of the two codecs is almost the same. However, as the number increases, the difference becomes obvious. After more than 400 iterations, the improvement of orthonormal matching pursuit is approximately 0.5 dB better than the non-orthonormal matching pursuit scheme.

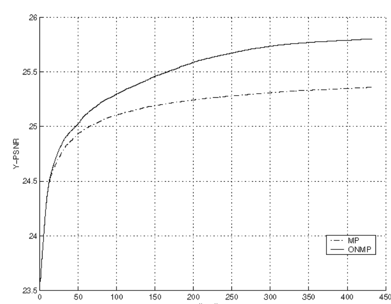
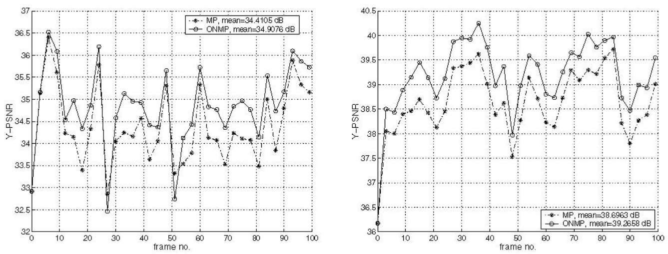


Fig. 1. The performance difference versus the iterations of the matching pursuit algorithm in the first P-frame of the QCIF Stefan sequence. MP denotes the traditional matching pursuit video coding scheme, and ONMP denotes the orthonormal matching pursuit video coding scheme.

The coding performances of the Silent and Miss America sequences encoded at 100 Kbits per second are shown in



(a) (b)

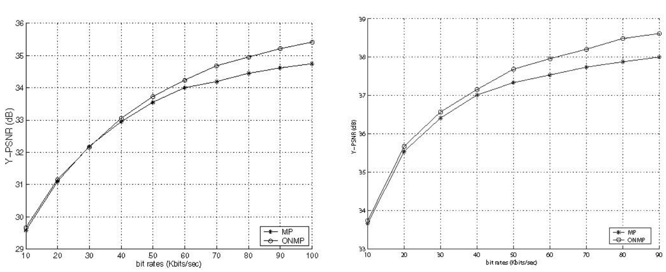
Fig. 2. The coding performance of the matching pursuit video coding scheme (MP) and the orthonormal matching pursuit video coding scheme (ONMP). The test sequence is in QCIF format and the bit rate is 100 Kbits/sec. (a) Silent; (b) Miss America.

(a) (b)

Fig. 3. Frame 78 of the QCIF Silent sequence encoded at 100 Kbits/sec, 10 frames/sec. (a) the matching pursuit video coding scheme, and (b) the orthonormal matching pursuit video coding scheme.

(a) (b)

Fig. 4. Frame 93 of the QCIF News sequence encoded at 100 Kbits/sec, 10 frames/sec. (a) the matching pursuit video coding scheme, and (b) the orthonormal matching pursuit video coding scheme.



(a) (b)

Fig. 5. The coding performance of the matching pursuit video coding scheme (MP) and the orthonormal matching pursuit video coding scheme (ONMP) based on various bit rates. The test sequence is in QCIF format and the testing time is 3.3 seconds. (a) News, and (b) Claire.

Figure 2. The testing time was 3.3 seconds. Compared to the matching pursuit video coding scheme, the average Y-PSNRs of the orthonormal matching pursuit video coding scheme are better by $0.5-0.65$ dB. Figures 3 and 4 illustrate the subjective performance of reconstructed frames obtained by the matching pursuit video coding scheme and the orthonormal matching pursuit video coding scheme respectively. The textures in video sequences can be encoded more clearly by the orthonormal matching pursuit algorithm. Note that, the matching pursuit algorithm produces a more blurred effect. Moreover, most blocking artifacts generated by motion compensation can not be removed efficiently by the matching pursuit algorithm.

The performance improvement is dependent on the number of bit rates. Figure 5 shows the Y-PSNRs encoded by the two video coding schemes at various bit rates. When the number of bit rates is high, more atoms will be encoded in each frame, and the orthonormal effect will be more pronounced. The coding efficiency can be increased by 0.7 dB as the bit rate increases.

V. Conclusion

We have presented an orthonormal matching pursuit algorithm and applied it to a hybrid motion compensation and matching pursuit video coding system. Using the algorithm, a signal can be efficiently approximated with a linear combination of selected bases and the redundancy between the selected bases can be successfully removed. By combining the orthonormal matching pursuit algorithm with the updated full search matching pursuit video coding scheme, we have devised an effective orthonormal matching pursuit video coding scheme. The experiment results for coding efficiency show that the proposed scheme can achieve a more efficient coding performance than a non-orthonormal matching pursuit video coding scheme. The perceptual quality also shows that proposed scheme can efficiently remove blocking artifacts generated by motion compensation.

References

[1] O. Al-Shaykh, E. Miloslavsky, T. Nomura, R. Neff, and A. Zakhor, "Video compression using matching pursuits", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 1, pp. 123–143, Feb. 1999.

[2] G. Davis, "Adaptive Nonlinear Approximations", Ph.D. dissertation, Department of Mathematics, New York University, 1994.

[3] J.L. Lin, W.L. Hwang, and S.C. Pei, "SNR scalability based on bitplane coding of matching pursuit atoms at low bit rates: fine-grained and two-layer", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 1, pp. 3–14, January 2005.

[4] J.L. Lin, W.L. Hwang, and S.C. Pei, "Multiple blocks matching pursuit update for low bit rate video coding," submitted to *IEEE Trans. Circuits Syst. Video Technol.*

[5] G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries", *IEEE Trans. Signal Processing*,Vol. 41, pp. 3397–3415, December 1993.

[6] R. Neff and A. Zakhor, "Very low bit-rate video coding based on matching pursuits", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, pp. 158–171, Feb. 1997.

[7] R. Neff and A. Zakhor, "Modulus Quantization for Matching Pursuit Video Coding", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 6, pp. 895–912, Sept. 2000.

[8] C. De Vleeschouwer and B. Macq, "SNR scalability based on matching pursuits", *IEEE Trans. on Multimedia,* vol. 2, No. 4, 2000, pp. 198–208.