A FAST MULTI-STEP ATOM SEARCHING IN NON-LOW SIGNAL ENERGY REGIONS FOR MATCHING PURSUITS

K. Imamura and H. Hashimoto***

Kanazawa University
Division of Electrical Engineering and Computer Science
Kakuma-machi, Kanazawa, Ishikawa 920-1192 Japan
{*imamura,**hasimoto}@ec.t.kanazawa-u.ac.jp

ABSTRACT

Recently, a high-efficiency video coding method using matching pursuits has been proposed. An immense number of operations is required for atom searching in matching pursuits, so speed-up of the searching procedure is indispensable. In the present paper, we propose a fast atom searching method for matching pursuits. An atom searching algorithm is based on both the correlation between the high-signal-energy regions and optimal matching points, and the correlation among highly efficient approximating points. As a result, improvement of the computational complexity problem is attained by reducing the number of searching points.

1. INTRODUCTION

Video coding standards such as H.26x and MPEG are based on a hybrid system of motion compensation (MC) and discrete cosine transform (DCT). On the other hand, another high-efficiency video coding method using a waveform coding technique called matching pursuits (MP) in place of DCT has been suggested in recent years[1, 2, 3]. In this coding method, the predicted image is generated by MC, and the prediction error image is then encoded by matching pursuits.(Fig.1)

Matching pursuits is a greedy algorithm by which a signal is iteratively decomposed into a linear expansion of waveforms from an over-complete dictionary[4]. Encoding of the matching pursuits algorithm approximates the signal so as to give high priority to dominant waveform components, and good coding performance is expected, especially for low bit rates.

The most computationally expensive part of matching pursuits is inner product calculation with all elements in the dictionary during the iteration procedure. This problem has been discussed as one of the difficult problems for real-time processing of matching pursuits.

Many approaches for developing fast encoding techniques have been suggested. Zakhor *et al.*[5] improved the encoding time by the hierarchical dictionary based on fundamental patterns. On the other hand, the prediction error image is decomposed into sub-bands and each low resolution sub-band image is encoded by matching pursuits in [6, 9]. The above methods attained to improve the coding performance and computational complexity.

However, from the viewpoint of the fast searching algorithm, searching for only restricted appropriate points is an indispensable approach because encoding with any dictionary needs to search for the optimal approximated signal from an image. Thus, searching at regular pixel intervals and searching in the neighboring area of the maximum signal energy block have been suggested[1, 2] in order to reduce the computational cost of matching pursuits. However, the former often happens to cause the optimal approximating point to be missed due to tree searching, and the latter also cannot avoid the decline of the reconstructed image quality.

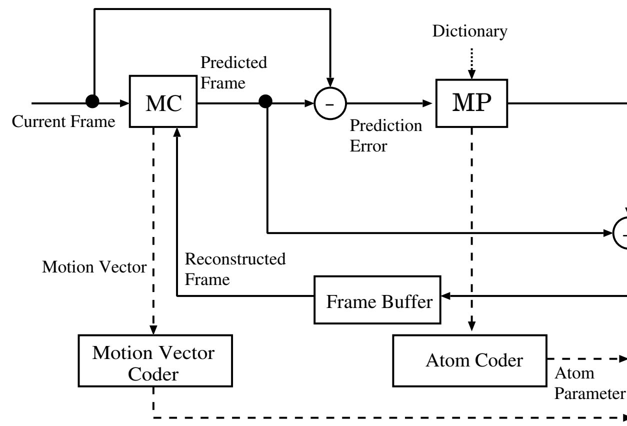


Figure 1: Video coding system using matching pursuits.

In the present paper, we propose a fast atom searching method of matching pursuits toward a high-efficiency video coding. We focus on the reduction of searching points. This approach is based on two important features of prediction error images. The first is that high-signal-energy regions include almost optimal approximating points, and the second is that the near points of the optimal approximating point provide highly efficient approximation. The proposed method excludes searching points in the low-signal-energy region as non-optimal approximating points, and the remaining points are searched by multi-step procedure. The proposed fast encoding algorithm is shown to be effective by reducing searching points while maintaining the high quality of the reconstructed image.

2. MATCHING PURSUITS

2.1 Basic Principle

Matching pursuits reconstructs a signal $\hat{f}(t)$ (an approximated version of $f(t)$) using a linear combination of basis function $g_\gamma(t)$ included in the over-complete dictionary $\mathscr{D}$, as follows:

$$\hat{f}(t) = \sum_{k=1}^{m} p_k \cdot g_{\gamma_k}(t - \tau_k) \tag{1}$$

where τ_k, γ_k, and p_k are parameters of the basis function and denote its position, type, and scale, respectively. A waveform with this parameter set obtained from the basis function is referred to as an atom.

In matching pursuits, atoms are selected by orthogonally projecting the signal $f(t)$ on basis function $g_\gamma(t)$ sequentially. First, the signal $f(t)$ is projected on the basis function $g_{\gamma_1}(t-\tau_1)$, and the signal $f(t)$ is denoted as

$$f(t) = p_1 \cdot g_{\gamma_1}(t-\tau_1) + Rf(t), \tag{2}$$

$$p_1 = < f(t), g_{\gamma_1}(t-\tau_1) > . \tag{3}$$

Here, p_1 is the inner product value between $f(t)$ and the basis function $g_{\gamma_1}(t-\tau_1)$. $Rf(t)$ is a residual signal that is orthogonal to $g_{\gamma_1}(t-\tau_1)$ and is projected on the basis function $g_{\gamma_2}(t-\tau_2)$.

If the norm of basis function $g_\gamma(t)$ is normalized to unit value, then the following equation can be obtained:

$$||f(t)||^2 = p_1^2 + ||Rf(t)||^2. \tag{4}$$

Namely, we can obtain the required parameters by maximizing the absolute value of the inner product, and the atom that minimizes the residual signal energy $\Delta e = ||f(t)||^2 - p_1^2$ is obtained. Since the residual signal $Rf(t)$ is also expanded

$$Rf(t) = p_2 \cdot g_{\gamma_2}(t-\tau_2) + R^2 f(t), \tag{5}$$

the signal $f(t)$ expanded by m atoms is denoted as follows:

$$f(t) = \sum_{k=1}^{m} p_k \cdot g_{\gamma_k}(t-\tau_k) + R^m f(t). \tag{6}$$

Therefore, signal $f(t)$ can be approximately reconstructed by m atoms.

2.2 Matching Pursuits Dictionary

A dictionary of matching pursuits is constructed from the Gabor function in Eq. 7, and two dimensional separable dictionary $G(i,j)$ is composed of the Cartesian product of one-dimensional Gabor functions as shown in Eq. 8 below:

$$g_{\vec{\gamma}}(n) = K_{\vec{\gamma}} \mathrm{g}\left(\frac{n-\frac{N}{2}+1}{s}\right) cos\left(\frac{2\pi\xi(n-\frac{N}{2}+1)}{N}+\phi\right) \tag{7}$$

$$G(i,j) = g_{\vec{\alpha}(i)} \otimes g_{\vec{\beta}(j)} \tag{8}$$

$\otimes$: Cartesian product

where $\mathrm{g}(\cdot)$ denotes a gaussian window and $K_{\vec{\gamma}}$ denotes a normalization factor. $\vec{\gamma} = (s, \xi, \phi)$ decides the Gabor function. $\vec{\alpha}$ and $\vec{\beta}$ denote a set of parameters of horizontal and vertical components, respectively.

Table 1 shows the parameters, and Figure 2 shows the dictionary constructed from Eqs. 7 and 8[1]. In the present paper, this dictionary, which includes 16×16 pixels images, is called the basic dictionary.

Table 1: Parameters of Basic Dictionary.

$\vec{\alpha},\vec{\beta}$	s	ξ	ϕ	$\vec{\alpha},\vec{\beta}$	s	ξ	ϕ
0	2	0	0	8	1	1	$\pi/2$
1	3	0	0	9	5	1	$\pi/2$
2	4	0	0	10	11	2	$\pi/2$
3	5	0	0	11	10	3	0
4	6	0	0	12	8	2	0
5	8	0	0	13	4	2	0
6	10	0	0	14	4	2	$\pi/4$
7	11	0	0	15	6	4	$\pi/4$

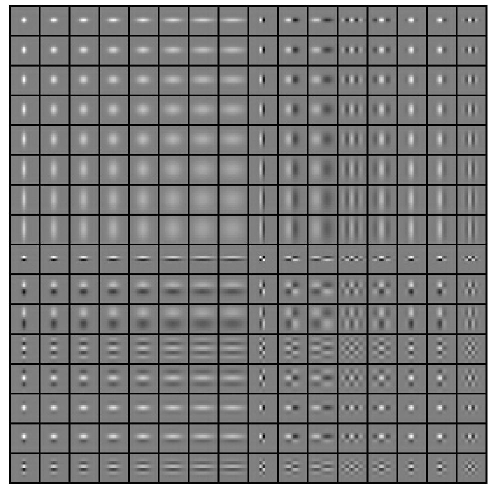


Figure 2: Basic Dictionary for Matching Pursuits (N = 16).

2.3 Atom Searching in Matching Pursuits

Searching for a pair of the optimal signal and the waveform (atom) in dictionary $\mathscr{D}$ for approximation in the encoding process of matching pursuits is called *atom searching*. The searching process by inner product calculation requires large computational costs in matching pursuits. In order to reduce the computational costs of encoding, a fast algorithm based on reduction of searching points is generally used. The fast algorithm stores memory with the information of already searched points until the signal is renewed at the points, so as to reduce the computational cost of re-evaluation. In the following, typical current fast algorithms based on the reduction of searching points are described.

2.3.1 N-pixel Interval Search

The fast algorithm searches the points at every N pixels in an entire image. The process of this algorithm is simple and the computational cost is reduced drastically, but the quality of the reconstructed image declines significantly due to missing the optimal approximating points. In addition, it is not efficient to select searching points without utilizing stocastic features of images. In the present paper, we refer to this searching as *N-pixel interval search*.

2.3.2 Neighboring Maximum Energy Search[2]

The fast algorithm partitions an image into blocks and calculates the signal energy of each block. The points in a fixed square region around the maximum signal energy block

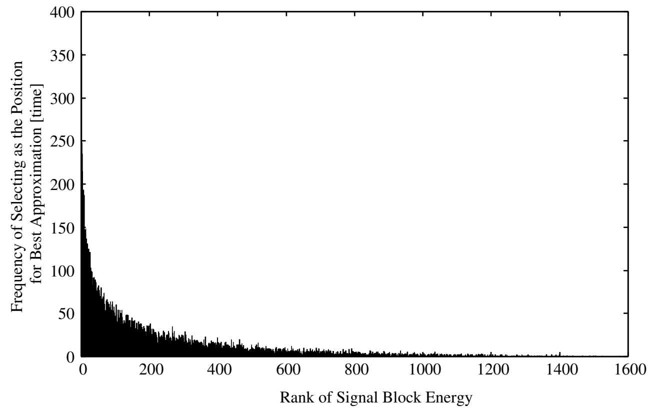


Figure 3: Correlation between signal block energy and optimal searching points (carphone).

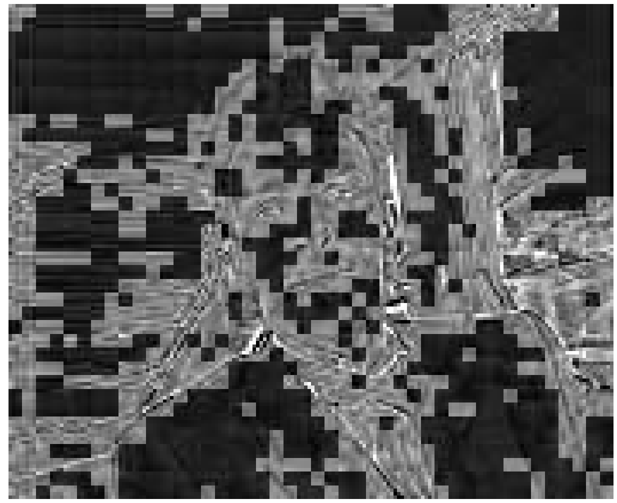


Figure 4: Search region of "car phone" sequence.

are searched. In the present paper, we refer to this searching as *neighboring maximum energy search*. This approach is based on the assumption that the optimal approximating point is often around the maximum energy block of an image. The computational cost reduction associated with the reduction of searching points is far larger than the overhead associated with the calculation of the signal energy. The total computational cost of the algorithm must be reduced.

Figure 3 shows the rank of the signal block energy and the frequency of selection as the position for optimal approximating points for "carphone" sequence. Figure 3 provides evidence of the correlation between the amplitude of the signal energy and the optimal approximating points. However, we consider the maximum signal energy blocks, and only a small percentage of these blocks were selected as optimal approximating points. As a result of the pre-experiment, the blocks in the square $\pm$8-pixel region around the maximum signal energy blocks were selected as less than 20% of the total number of blocks. Therefore, the neighboring maximum energy search cannot avoid a decline in the reconstructed image quality.

3. MULTI-STEP SEARCH IN NON-LOW-SIGNAL-ENERGY REGIONS

In this section, we describe the proposed searching algorithm, which utilizes the spatial correlation between the high-signal-energy regions and the optimal approximating points. In the proposed method, all of the points in the low-signal-energy blocks are excluded as candidate non-optimal approximating points, and the remaining points are searched by multi-step search. In the present paper, we refer to this searching method as *multi-step search in non-low-signal-energy regions*.

3.1 Selecting Search Region

From Figure 3, most of the blocks in the low-signal-energy region do not contain the optimal approximating points. Therefore, the low-signal-energy blocks should be excluded from the search region.

The selection procedure of the search region is as follows: First, a prediction error image is partitioned into blocks, and the signal energy of each block is calculated. Next, the blocks are excluded from the search region by following two conditions based on the amount of signal energy.

The blocks are sorted in ascending order according to the signal energy, and the block is judged as to whether it is contained in the search region.

- Decision by the sum of the excluded signal energy
  The first stop condition excluding blocks from the search region is based on the sum of the excluded signal energy. The procedure is stopped when the sum of the excluded signal energy exceeds a fixed threshold value.
  In the present paper, the threshold value was set to 7% of the signal energy of the entire image. The pre-experimental results show that the threshold value of 7% corresponds to the exclusion of 40$\sim$50% of the blocks, which causes a large decline in the quality of the reconstructed image.
- Decision by the signal energy of the block
  Another stop condition occurs when the signal energy of an excluded block exceeds a fixed threshold value. The purpose of this condition is to prevent a high-signal-energy block being excluded from the search region by the first condition.
  In the present paper, the threshold value was set to 0.02% of the signal energy of the entire image. The threshold value was decided experimentally based on the relationship among the signal energy of a block, its frequency selected as the optimal approximating block, and the quality of the reconstructed image.

Figure 4 shows the selected search region with the above threshold values in the prediction error image of the "car phone" sequence. The black area denotes the excluded blocks from the search region. The search region occupies only 48% of the entire image.

3.2 Multi-step Search

These procedures provide significant regions for searching, but the computational cost required to search all of the points in the remaining regions remains large. Better approximating points often exist within high-signal-energy regions of the prediction error image, and there is strong spatial corre-

Table 2: Region size of 2nd step in multi-step search and PSNR (200 atoms, 2-step search, initial step = 4).

Seq.	Region size of 2nd step				
	±1	±2	±3	±4	±5
akiyo	37.98	38.55	38.67	38.71	38.68
mother	35.86	36.09	36.18	36.20	36.21
carphone	33.29	33.58	33.66	33.69	33.72
foreman	30.82	31.04	31.11	31.13	31.15

lation among highly efficient approximating points. Moreover, the high-signal-energy regions including the optimal approximating points are distributed over the entire image, that is, the distribution is multi-modal. Therefore, we introduce multi-step search method as searching from selected region considering the spatial distribution of high efficiency points for approximation.

The number of step is set to two in the multi-step search considering searching performance.

In the first step of the multi-step search, the best approximating candidate is searched from the points in the remaining regions at every four pixels considering the multi-modal distribution. We set the interval of first step to four pixels because this parameter shows good performance in N-pixel interval search.

Next, the optimal approximating point is searched from the square ± 3-pixel region around the point obtained in the first step, considering the spatial correlation among the highly efficient approximating points. Table 2 shows the search region of the second step and the PSNR of the reconstructed image in multi-step searching for whole images. The PSNR listed in the table are average values for each sequence. From Table 2, the PSNR is significantly declined for less than a $\pm 3 \sim 4$ pixel-wide search region in the second step. Based on these results, we assume that the highly efficient approximating points are distributed in a $\pm 3 \sim 4$ pixel-wide region around a candidate point. In the present paper, the search region in the second step is set to ± 3 pixels from the computational complexity.

4. SIMULATION AND RESULTS

The multi-step search in non-low-signal-energy regions in mutching pursuits video coding was examined by computer simulations. The "akiyo", "carphone", "foreman" and "mother & daughter" (mother) (QCIF, 10 fps, 50 frames, grayscale) were used as test sequences.

We first compared the proposed method with the existing search method. Table 3 shows the PSNRs of the reconstructed image and the relative processing time for various search methods. The relative processing times were obtained and compared to the processing time of the full search method. In the experiments, the interval N was set to four pixels in the N-pixel interval search, and the search region was an ± 8 pixel-wide region in the neighboring maximum energy search. The partitioned block size was 4×4 pixels in both the neighboring maximum energy search and the proposed method.

The computational complexity of the proposed method was reduced to approximately 10% of that of the full search method with an approximately 0.2 dB decline in the reconstructed image quality. The PSNRs and the processing times

Table 3: Comparisons of PSNR and ratio of processing time (200 atoms).

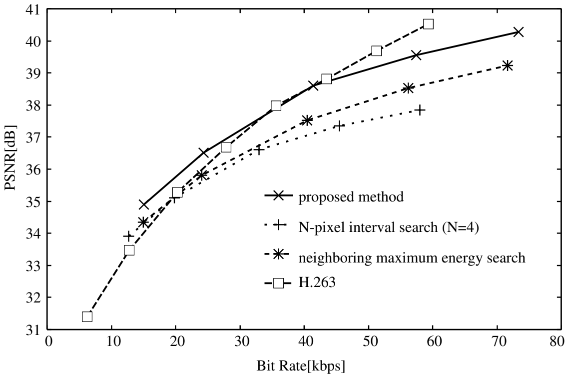
Method	Sequence	PSNR [dB]	Relative processing time
Full Search	akiyo	38.84	1.000
	mother	36.32	1.000
	carphone	33.84	1.000
	foreman	31.20	1.000
N-pixel Interval Search (N = 4)	akiyo	36.60	0.080
	mother	35.18	0.080
	carphone	32.50	0.080
	foreman	30.15	0.080
Neighboring Maximum Energy Searh (± 8)	akiyo	37.48	0.242
	mother	35.68	0.242
	carphone	32.65	0.242
	foreman	30.42	0.242
Multi-step Search	akiyo	38.67	0.122
	mother	36.18	0.122
	carphone	33.66	0.122
	foreman	31.11	0.122
Multi-step Search Non-low-Energy Region (proposed)	akiyo	38.61	0.101
	mother	36.15	0.101
	carphone	33.61	0.093
	foreman	31.11	0.096

of the proposed method were better than those of the N-pixel interval search and the neighboring maximum energy search. The computational cost of the proposed method was 20% larger than that of the N-pixel interval search. However, the PSNR of the proposed method is far superior. We have confirmed that the exclusion of the search region based on the signal energy can reduce the computational cost and that the multi-step search is effective for maintaining the quality of the reconstructed image.

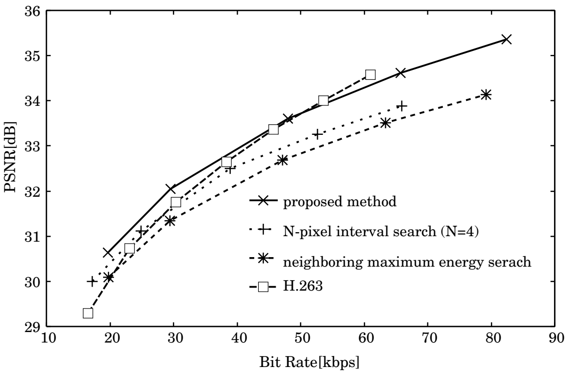
Figure 5 shows the coding performances of the proposed method, the N-pixel interval search, and the neighboring maximum search. The coding performance of H.263[11] is also shown for comparison, especially for low bit rates. Figure 5 indicates that the proposed method achieves better performance than H.263 at lower bit rates, despite the use of a poor basic dictionary.

5. CONCLUSION

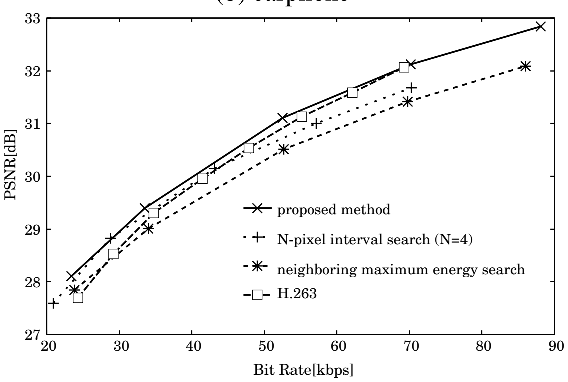
We have presented a fast coding method for matching pursuits by improveing atom searching. The proposed method reduces the number of searching points by taking advantage of both the correlation between the high-signal-energy regions and optimal matching points, and the correlation among highly efficient approximating points. The simulation results revealed that the proposed method can reduce the computational complexity. Compared with the full search, the computational complexity was reduced to about 10%, while maintaining the high quality of the reconstructed image. We are continuing our research on the combination of the proposed method with other fast methods, such as an efficient dictionary structure.



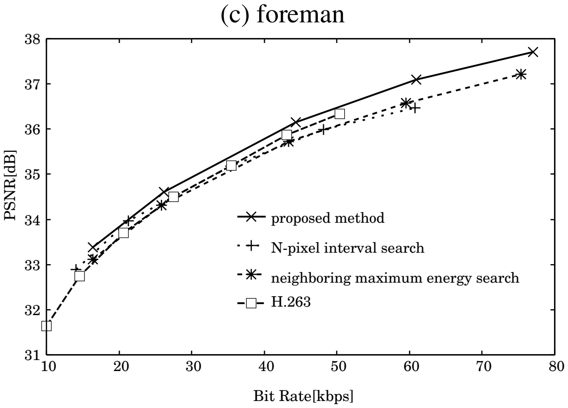
(a) akiyo



(b) carphone



(c) foreman



(d) mother and daughter

Figure 5: Comparison of coding performance.

REFERENCES

[1] R. Banham and C. Brailean, “A Selective Update Approach to Matching Pursuits Video Coding,” *IEEE Trans. Circuits & Syst. Video Technol.*, vol. 7, no. 1, pp. 119-129, 1997.

[2] R. Neff and A. Zakhor, “Very Low Bit-Rate Video Coding Based on Matching Pursuits,” *IEEE Trans. Circuits & Syst. Video Technol.*, vol. 7, no. 1, pp. 158-171, 1997.

[3] Osama K. Al-Shaykh, et al.: “Video Compression Using Matching Pursuits,” *IEEE Trans. Circuits & Syst. Video Technol.*, vol. 9, no. 1, pp. 123-143, 1999.

[4] S. Mallat and Z. Zhang, “Matching Pursuits With Time-Frequency Dictionaries,” *IEEE Trans. Signal Process.*, vol, 41, no. 12, pp. 3397-3415, 1993.

[5] R. Neff and A. Zakhor, “Matching Pursuits Video Coding. I. Dictionary Approximation,” *IEEE Trans. Circuits & Syst. Video Technol.*, vol. 12, no. 1, pp. 13-26, 2002.

[6] C.De. Vleeschouwer and B. Macq, “Subband Dictionaries for Low-Cost Matching Pursuits of Video Residues,” *IEEE Trans. Circuits & Syst. Video Technol.*, vol. 9, no. 7, pp. 984-993, 1999.

[7] P. S.-Saugeon, A. Zakhor: “Dictionary Design for Matching Pursuits and Application to Motion-Comppensated Video Coding”, *IEEE Trans. Circuits & Syst. Video Technol.*, vol. 14, no. 6, pp.880-886, 2004.

[8] F. Moshetti, L. Granai, P.Vandergheynst and P. Frossard: “New Dictionary and Fast Atom Searching Method for Matching Pursuit Representation of Displaced Frame Difference”, IEEE ICIP 22-25, Vol.3, September 2002, pp.685-688.

[9] K. Imamura, Y. Koba and H. Hashimoto, “Fast Matching Pursuits Algorithm Using Sub-band Decomposition for Video Coding,” *Journal of ITE*, vol. 60, 7, pp. 1059-1067, 2006.

[10] K. Imamura, T. Nagami and H. Hashimoto, “Fast Matching Pursuits Using Multiple-Step Search in Non-Low Signal Energy Regions,” *Journal of ITE*, vol. 61, no.8, pp. 1231-1235, 2007.

[11] ITU-T Study Group 15, “Draft Recommendation H.263,” 1995.