INTEGER SUB-OPTIMAL KARHUNEN-LOEVE TRANSFORM FOR MULTI-CHANNEL LOSSLESS EEG COMPRESSION

Yodchanan Wongsawat, Soontorn Oraintara and K. R. Rao

Department of Electrical Engineering, University of Texas at Arlington,
416 Yates St., Arlington, TX, 76019-0016 USA
Email: yodchana@gauss.uta.edu, oraintar@uta.edu, and rao@uta.edu

ABSTRACT

In this paper, a method for approximating the Karhunen-Loeve Transform (KLT) for the purpose of multi-channel lossless electroencephalogram (EEG) compression is proposed. The approximation yields a near optimal transform for the case of Markov process, but significantly reduces the computational complexity. The sub-optimal KLT is further parameterized by a ladder factorization, rendering a reversible structure under quantization of coefficients called IntSKLT. The IntSKLT is then used to compress the multi-channel EEG signals. Lossless coding results show that the degradation in compression ratio using the IntSKLT is about 3% when the computational complexity is reduced by more than 60%.

1. INTRODUCTION

In some countries, legally recording of electroencephalogram (EEG) signals for diagnosis has to be done losslessly for the purpose of perfect transmission to the expert analysts. For this reason, there have been interests in lossless EEG signal compression. Generally, EEG signals are measured from the electrodes at different positions on the human scalp. A typical number of channels (N) of the signals can be as high as a few hundred (or may be a few thousand in the near future), and thus these neighboring channels can be highly correlated. In order to efficiently compress multi-channel signals, the inter-channel redundancy must be exploited. Although, these multi-channel signals seem to be correlated, the correlation models of the signals are unpredictable due to the unknown combination of the cortical potentials. Thus, data independent transforms such as DCT, DST, or DFT [1] usually fail to efficiently decorrelate these types of signals. This problem can be solved by employing the optimal transform to decorrelate these EEG signals by exploiting the eigenvectors of their correlation matrix. This optimal transform is known as Kahunen-Loeve Transform (KLT) [1]. However, calculating N-point KLT can be complicated especially for large N (number of electrodes or number of channels).

Several schemes for single-channel lossless EEG compression have been proposed [2], [3]. However, not many methods exploit the multi-channel case. An efficient approach to losslessly compress multi-channel EEG signals by employing KLT to reduce inter-channel redundancy is proposed in [4]. KLT is approximated with finite precision and reversible operations. The key to the approximation of these transforms is *ladder-type*, also known as lifting, factorization. The method for matrix factorization in [5], [6], [7] are used. The resulting approximation is called integer KLT (IntKLT). According to [7], it is known that, for a given matrix, the factorization is not unique. Each solution results in a different permutation and dynamic range of coefficients, which are of particular importance to lossless coding applications. To demonstrate this, let us consider an example.

Example 1, lifting factorization: Let $\mathbf{A}$ be an arbitrary matrix of size 3×3 such that $|\det \mathbf{A}| = 1$ as follows:

$$\mathbf{A} = \begin{bmatrix} 0.6 & -0.64 & 0.48 \\ -0.8 & -0.48 & 0.36 \\ 0 & 0.6 & 0.8 \end{bmatrix}.$$

Two lifting factorizations of $\mathbf{A}$ can be shown:

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0.33 & -0.5 & -1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0.27 & 1 & -0.8 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & -0.3 & 0.6 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -0.33 & 0.5 & 1 \end{bmatrix}\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \quad (1)$$

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3.75 & -3.42 & -1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 1.2 & 1 & -0.48 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & -3.33 & 0.8 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1.25 & 4.92 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2)$$

It is clear that both factorizations can be approximated with reversibility preserved. The magnitudes of the coefficients in (1) are all less than or equal to one whereas those in (2) are as high as 4.92. These large coefficients can significantly impact the dynamic range of the internal nodes in the transform and thus the lossless coding performance. The problem is more severe as N increases.

Considerations for using the IntKLT for multi-channel coding include:

- Approximating a reversible structure, since the factorization of the KLT is not unique, finding the best solution in the sense of minimum dynamic range of coefficients is very difficult. An obvious approach is to compare all the possible factorizations. This, however, is impractical for a large number of channels since the number of solutions is of order $O(N!)$, where N is the number of channels.
- Since the KLT is statistically dependent transform, its parameters must be transmitted as side information which is of order $O(N^2)$. Hence, the side information should also be minimized.
- The calculation of the KLT and the implementation of the IntKLT are highly complex, especially for large N.

In this paper, an efficient approximation with reversibility preserved for large KLT is proposed. The signals are divided into groups using the 'divide-and-conquer' philosophy. For each group, the signals are decorrelated by its KLT, referred to as *marginal* KLT in [8]. Each of the small KLTs is realized by a lifting factorization, rendering a reversible transform with small rational lifting coefficients. In order to obtain the small lifting coefficients, appropriate permutation matrices have to be well selected. Finding the right permutation matrices for factorizing each marginal KLT instead of full-sized N-point KLT is more practical [7]. As a result, using the proposed simplified transform can significantly reduce computational complexity over using IntKLT. For convenience, this proposed transform is called integer sub-optimal KLT (IntSKLT). The performance of IntSKLT is presented via the decay of its sorted variances. Since, our focus is on the investigation of IntSKLT for decorrelating inter-channel redundancy of EEG data. IntSKLT will also be embedded to the inter-channel decorrelation part of the multi-channel EEG lossless scheme in [4] (Figure 6) instead of IntKLT to verify the merit of its coding performance.

This work was supported in part by NSF Grant ECS-0528964.

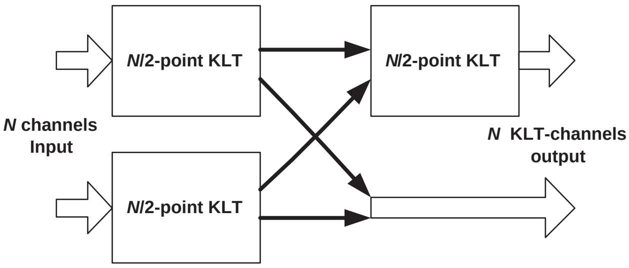

Figure 1: Structure of sub-optimal KLT for N-channel signals

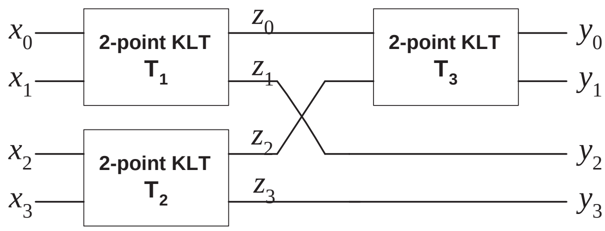

Figure 3: Approximated 4-point KLT.

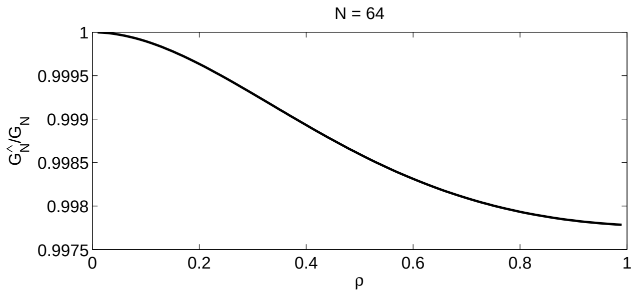

Figure 4: $\hat{G}_N/G_N$ for the case of $N = 64$.

2. SUB-OPTIMAL KLT

In order to solve the complexity issue caused by the large dimension of the KLT, the N channels of EEG signals are equally divided into two groups of $N/2$ channels along the arbitrary scan order (in this paper, the scanning orders in Figure 7 are used with $N = 64$). Each group is decorrelated by its local marginal KLT, and the outputs are sorted according to the variances from large to small. To further reduce the dependency, the largest $N/4$ channels from the two groups are combined, and decorrelated by their $N/2$-point KLT. Under the assumption that, the EEG signals are highly decorrelated, the remaining $N/2$ channels of small signals are exported directly to the outputs. Figure 1 shows a block diagram for the case of $N = 64$. A self-similar structure is used to further reduce the complexity where the same structure is repeated for each of the $N/2$ KLT. It is noted that every marginal KLT is updated over the period of m second EEG data (in this paper, $m = 8$ is used).

Figure 1 shows structure of the proposed sub-optimal KLT, where the details of each $N/2 = 32$-point sub-optimal KLT is illustrated in Figure 2 for the case of $N = 64$. Figure 2 also demonstrates that the iteration of dividing the channels stops at *8-point KLT* which is a feasible size to optimize for the best parameters. From the implementation point of view, the $N = 64$ channels are clustered into groups of eight processed locally (Figure 7(b)). Optimality of the proposed approximation of the KLT depends on the statistics of the signals and how they are clustered.

2.1 Markov process

In this section, let us consider a special case of Markov process. Assume that the correlation between the i-th and j-th channels is given by:

$$[\mathbf{R}_x]_{ij} = E\{x_i x_j\} = \rho^{|i-j|}, \quad i, j = 0,\ 1, \cdots,\ N-1.$$

It can be shown that the maximum achievable coding gain [9] for this case is

$$G_N = \frac{\frac{1}{N}tr(\mathbf{R}_x)}{det(\mathbf{R}_x)^{1/N}} = (1-\rho^2)^{-(1-\frac{1}{N})}.$$

2.1.1 Four-channel case

Consider a simple case of $N = 4$ where the signals are decorrelated using the structure in Figure 3. It is easy to see that the marginal KLTs $\mathbf{T}_1$ and $\mathbf{T}_2$ are

$$\mathbf{T}_i = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \tag{3}$$

and the corresponding eigenvalues are $1 \pm \rho$. The coding gain after the first stage (z_i) is $(1-\rho^2)^{-\frac{1}{2}}$. Applying a 2-point KLT to the larger components z_0 and z_2 results in a sub-optimal coding gain of

$$\hat{G}_4 = (1-\rho^2)^{-\frac{1}{2}}\left[1 - \frac{1}{4}\rho^2(1+\rho)^2\right]^{-\frac{1}{4}}.$$

It can be shown that $0.9036 \approx \left(\frac{2}{3}\right)^{1/4} \leq \frac{\hat{G}_4}{G_4} \leq 1$ for $0 \leq \rho \leq 1$.

2.1.2 2^p-channel case

In extension to the case of $N = 2^p$, assume that the three $N/2$-point KLTs used in the approximation in Figure 1 are obtained by the eigenvectors of the correlation matrices of the corresponding inputs, i.e. the three KLTs are the marginal KLTs. Figure 4 plots the ratio $\frac{\hat{G}_N}{G_N}$ as a function of ρ for the case of $N = 64$. It is clear that when these sub-KLTs are optimal, the degradation in coding gain is very insignificant (less than 0.3%).

In order to apply the proposed sub-optimal KLT to multi-channel EEG compression, two points should be noted. First, in the proposed recursive structure, the $N/2$-point KLTs are further approximated, and thus the difference in coding gain accumulates. Second, and perhaps more importantly, the EEG signals, although highly correlated, may not be a Markov process. In fact, there exists some correlation among the neighboring channels. Hence, the choice in clustering the inputs into groups (Figure 7(b)) also has an impact on the coding performance.

3. INTEGER KLT

In order to design the reversible structure for each 8-point KLT for the purpose of constructing a reversible approximation of N-point sub-optimal KLT in section 2, a lifting factorization in [7] together with the slight modification in [4] is exploited. The N-point KLT matrix, $\mathbf{K}_{N\times N}$, is factorized using single-row non-zero off-diagonal factorization as,

$$\mathbf{K}_{N\times N} = \mathbf{P}_L\mathbf{S}_N\mathbf{S}_{N-1}\cdots\mathbf{S}_1\mathbf{S}_0\mathbf{P}_R, \tag{4}$$

where $\mathbf{P}_L$ and $\mathbf{P}_R$ are the permutation matrices that exchange the rows and the columns, $\mathbf{S}_n$ (n = 1, 2 …, N) are single-row non-zero off-diagonal matrices, $\mathbf{S}_0$ is a lower triangular matrix, and N is the number of inputs. To clarify (4), an example of this factorization for the case of 4-channel is shown in Figure 5. According to Figure 5, the magnitudes of the rational lifting coefficients, s_{ij}, depend directly on the selection of the permutation matrices. It should be noted that, for some choices of $\mathbf{P}_L$ and $\mathbf{P}_R$, the coefficients obtained in $\mathbf{S}_i$ can be very large. This can result in degradations of lossless coding performance. In order to make the problem feasible for selecting good $\mathbf{P}_L$ and $\mathbf{P}_R$, the size of smallest KLT is limited

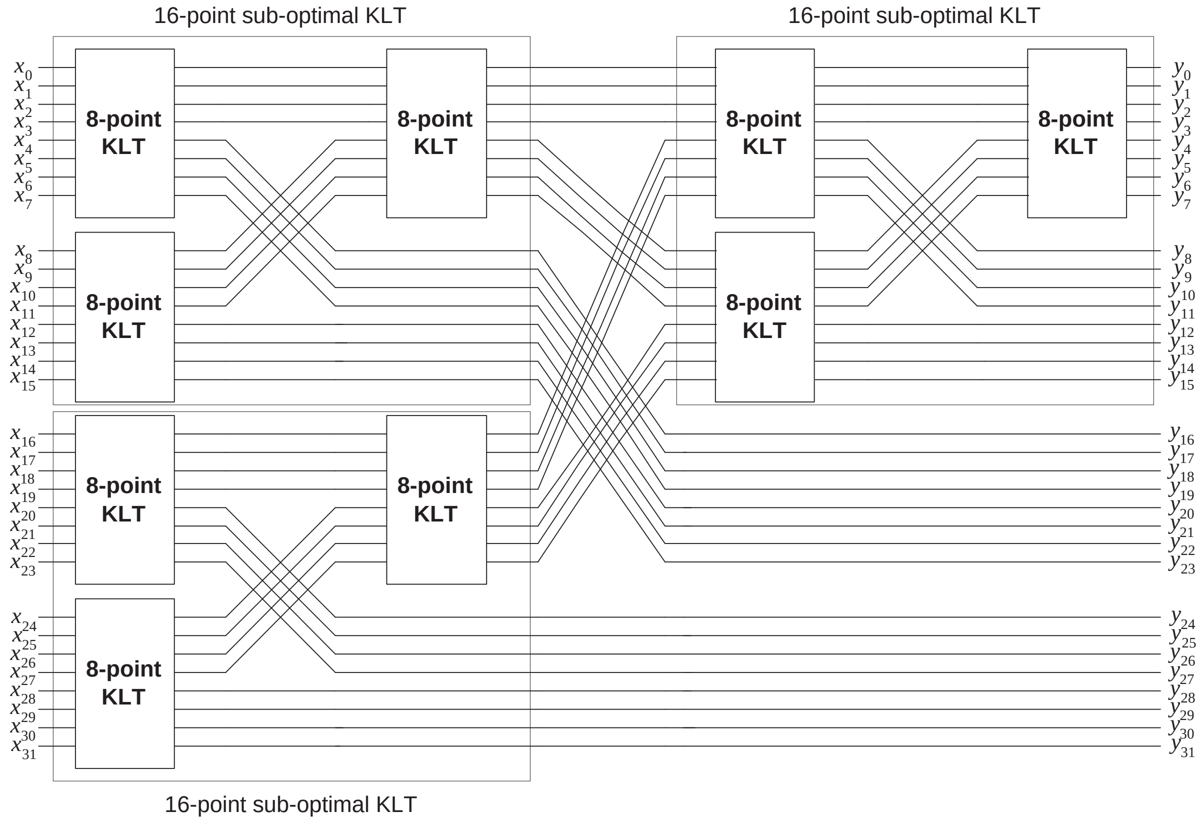

Figure 2: Structure of sub-optimal KLT for 32-channel signals which is composed of 16-point sub-optimal KLTs and 8-point KLTs

to N=8. Therefore, the reversible approximation of the sub-optimal KLT mentioned in Section 2 can be constructed using the structure of 8-point IntKLTs.

On the computational complexity, for the case of N-point IntKLT, the number of lifting coefficients used in the structure is $\frac{3}{2}N(N-1)$. For the case of N-point IntSKLT, the number of lifting coefficients is $\frac{3}{2}8(8-1)\times 3^{\log_2(N/8)} = \frac{28}{9}N^{log_2 3} \approx 3.1N^{1.585}$. It is clear that the ratio between the numbers of coefficients from both cases grows exponentially for large N.

4. MULTI-CHANNEL EEG LOSSLESS CODER

The performance of the proposed IntSKLT is evaluated in the multi-channel EEG lossless coder proposed in [4]. The block diagram of this coder is shown in Figure 6. Different pulse code modulation (forward difference) is used to remove DC bias in each EEG channel. Inter-channel decorrelation is then performed using the IntKLT. The resolution at all internal nodes is fixed to sixteen bits. The IntDCT-IV [10] is applied to reduce temporal redundancy. Finally, the statistical redundancy is removed by using the Huffman coding. As mentioned, using IntKLT may introduce undue computational complexity to the coder for large number of channels. The problem is solved by replacing the IntKLT with the IntSKLT. For the details of the coder, the reader is referred to [4].

5. SIMULATION RESULTS

In the simulation, eight seconds of sixty-four channels of EEG signals sampled at 1.024 kHz and digitized to sixteen bits are used. In order to achieve efficient channel partitioning, two types of channel scanning (spiral and clustering) shown in Figure 7 are applied. Figure 8 compares the (sorted) variances at the outputs when IntKLT, IntSKLT and IntDCT are used to decorrelate the inter-channel correlation. Since the IntDCT is signal independent, the majority of the variances are much higher. On the other hand, the output variances in the case of IntSKLT are similar to those obtained from the IntKLT. Figures 8(a) and (b) show the results for two different cases of spiral and clustering scanning. It is noticeable that the output variances obtained by the IntSKLT in the case of clustering are closer those of the IntKLT than in the case of spiral scanning.

The lossless coder described in Section 4 is applied to the EEG data. Table 1 summarizes the compression ratios obtained from the cases of IntKLT and IntSKLT with different channel scanning. When there is no inter-channel decorrelation, i.e. the IntKLT block in Figure 6 is removed, the compression ratio is 2.53. Maximum compression ratio of 2.84 is achieved when the full 64-point IntKLT is applied. When the IntSKLT, consisting of twenty-seven eight-point IntKLTs, is used, compression ratios of 2.80 and 2.82 are obtained for spiral and clustering scanning. Note that when the signals are randomly grouped, the average compression ratio is 2.78. This shows that the order of the EEG channels has some impact on the coding performance. Although, using IntSKLT causes about 3% coding degradation compared with using IntKLT, computational cost is dramatically reduced by more than 60%. Table 2 compares coding result of the coder in [4] (except that IntKLT is replaced with IntSKLT) with the existing algorithm Shorten (Lossless linear prediction based coder of order 6) [11], lossless JPEG2000 [12] and GZIP [13]. At the higher computational cost over the existing lossless algorithms, coding gain is improved by more than 20%. The proposed coder yields the best compression ratio at 2.82, while Shorten, lossless JPEG2000 and GZIP yield compression ratios of 2.16, 1.97 and 1.44, respectively.

6. CONCLUSIONS

This paper presents a method for efficiently approximating the KLT for the purpose of lossless multi-channel EEG compression. The approximation is done by dividing the signals into small groups

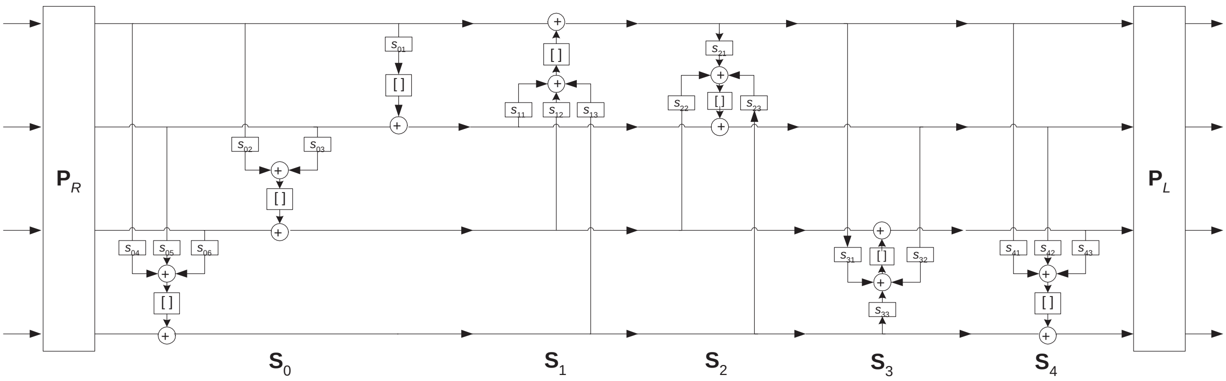

Figure 5: Example of the structure of factorization used in [4] for the case of 4×4 matrix

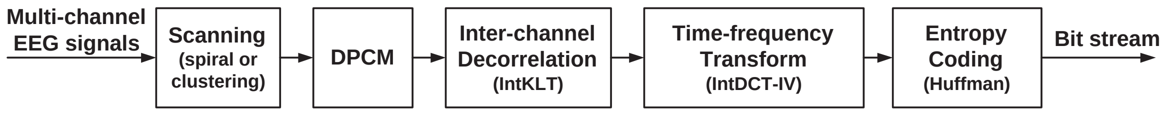

Figure 6: Multi-channel EEG lossless coder

Table 1: Compression ratio (CR) of the 64-channel EEG lossless compression coder (Figure 6) using IntKLT and IntSKLT for channel decorrelation

64-channel EEG signals	CR
No channel decorrelation	2.53
64-point IntKLT	2.84
64-point IntSKLT (spiral scan data)	2.80
64-point IntSKLT (clustered data)	2.82
64-point IntSKLT (random)	2.78

Table 2: Encoding 16 bit per sample (bps) of 64-channel EEG signals with various lossless coding algorithms: Compression percentage is defined as $(16 - \text{output bit rate}) \times 100/16$

64 channels	IntSKLT	Shorten	JPEG2000	GZIP
Compression Percentage	64.56	55.37	49.33	30.56
CR	2.82	2.24	1.97	1.44
bps	5.67	7.14	8.11	11.11

and using a number of small-size KLTs. A special case of Markov process is discussed. The sub-optimal KLT is then further approximated by using lifting factorization, resulting in the integer-to-integer mapping IntSKLT. It is shown that the number of lifting coefficients of the IntSKLT is of order $O(N^{1.585})$ instead of $O(N^2)$ for the case of IntKLT. Not only is the computational cost reduced, but also the amount of side information that needs to be transmitted with the compressed data is also reduced by using the IntSKLT. Despite the complexity reduction, the new transform yields near optimal results in multi-channel EEG lossless compression performance. Optimal channel ordering/mapping requires further study.

REFERENCES

[1] K. R. Rao and P. C. Yip, *The Transform and Data Compression Handbook*, Boca Raton: CRC Press, 2001.

[2] G. Antoniol and P. Tonella, "EEG data compression techniques," *IEEE Trans. Biomed. Eng.*, vol.44, pp.105-114, Feb. 1997.

[3] N. Memon, X. Kong, and J. Cinkler "Context-based lossless and near-lossless compression of EEG signals," *IEEE Trans. Biomed. Eng.*, vol.3, pp.231-238, Sep. 1999.

[4] Y. Wongsawat, S. Oraintara, T. Tanaka and K. R. Rao, "Lossless multi-channel EEG compression," *Accepted to present at IEEE ISCAS*, 2006.

[5] P. Hao and Q. Shi, "Reversible integer KLT for progressive-to-lossless compression of multiple component images," *IEEE ICIP*, vol.1, pp.633-636, Sept. 2003.

[6] L. Galli and S. Salzo, "Lossless hyperspectral compression using KLT," *IEEE IGARSS*, vol.1, pp.313-316, Sept. 2004.

[7] P. Hao and Q. Shi, "Matrix factorizations for reversible integer mapping," *IEEE Trans. Signal Processing*, vol.49, pp.2314-2324, Oct. 2001.

[8] M. Gastpar, P.-L. Dragotti and M. Vetterli, "The distributed, partial, and conditional Karhunen-Loeve transforms" *IEEE DCC*, pp.283-292, Mar. 2003.

[9] P. P. Vaidyanathan, *Multirate Systems and Filterbanks*, Englewood cliffs, NJ: Prentice Hall, 1993.

[10] R. Geiger, Y. Yokotani, G. Schuller and J. Herre, "Improved integer transforms using multi-dimensional lifting," *IEEE ICASSP*, vol.2, pp.1005-1008, May 2004.

[11] Available Shorten software: http://www.etree.org/shncom.html/

[12] Available JPEG2000 software: http://www.kakadusoftware.com/

[13] Available GZIP software: http://www.gzip.org/

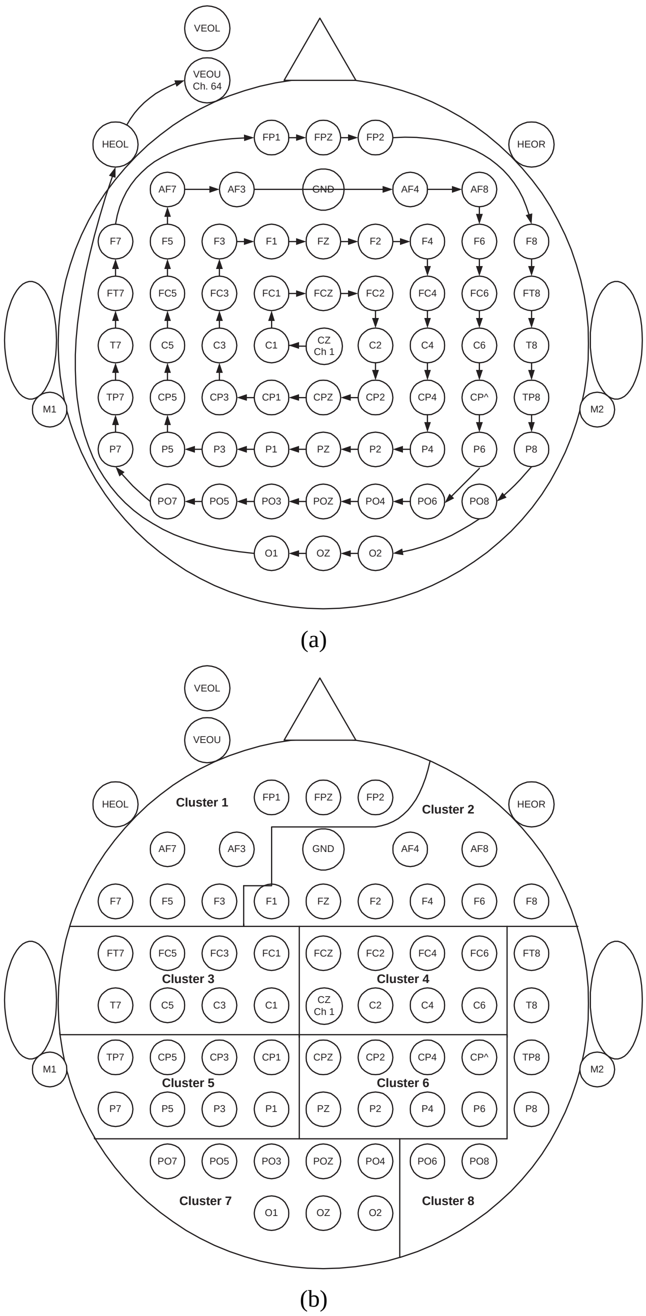

Figure 7: Pattern of scanning scheme: (a) spiral and (b) clustering (VEOU and HEOL are included in cluster 8).

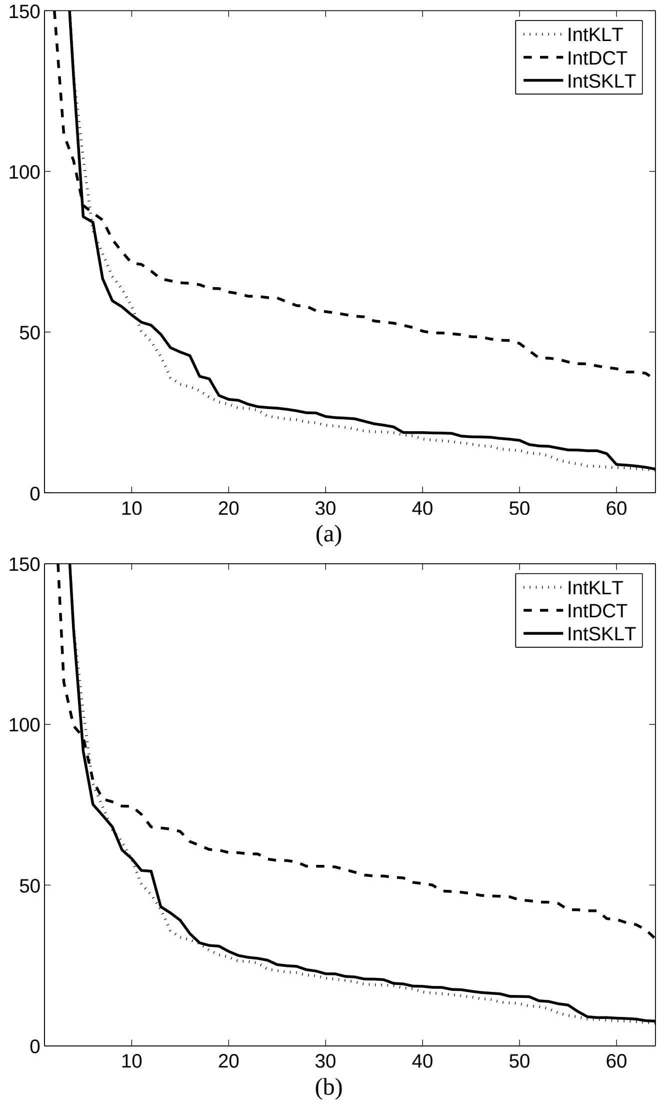

Figure 8: Performance analysis of 64-point KLT using: (a) spiral scan data and (b) clustering scan data, where *x*-axis represents the transform coefficients and *y*-axis represents their (sorted) variances.