The Efficient Discrete Tchebichef Transform for Spectrum Analysis of Speech Recognition

Ferda Ernawan[1], Nur Azman Abu[2] and Nanna Suryana[2]
[1]Faculty of Information and Communication Technology, Universitas Dian Nuswantoro
Semarang, Indonesia
[2]Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka
Melaka, Malaysia
[1]e-mail: ferda1902@gmail.com

ABSTRACT

Spectrum analysis is an elementary operation in speech recognition. Fast Fourier Transform (FFT) is a famous technique to analyze frequency spectrum of the signal in speech recognition. The Discrete Tchebichef Transform (DTT) is proposed as possible alternative to the FFT. DTT has lower computational complexity and it does not require complex transform with imaginary numbers. This paper proposes an approach based on 256 discrete orthonormal Tchebichef polynomials for efficient to analyze a vowel and a consonant in spectral frequency of speech recognition. The comparison between 1024 discrete Tchebichef transform and 256 discrete Tchebichef transform has been done. The preliminary experimental results show that 256 DTT has the potential to be efficient to transform time domain into frequency domain for speech recognition. 256 DTT produces simpler output than 1024 DTT in frequency spectrum. The used of 256 Discrete Tchebichef Transform can produce concurrently four formants F_1, F_2, F_3 and F_4 for the consonant.

Keyword-Speech Recognition, Spectrum Analysis, Fast Fourier Transform and Discrete Tchebichef Transform.

1. INTRODUCTION

Speech recognition is heavy process that required large sample data which represent speech signal for each windowed. Each window consumes 1024 sample data. Speech signal methods using Fourier transform are widely used in speech recognition. One of the most widely used speech signal methods is the FFT. FFT is a traditional technique for digital signal processing applicable for spectrum analysis. 1024 sample data FFT computation is considered the main basic algorithm for several digital signals processing [1]. The FFT is often used to compute numerical approximations to continuous Fourier. However, a straightforward application of the FFT to computationally often requires a large FFT to be performed even though most of the input data to the FFT may be zero [2].

The DTT is another transform method based on discrete Tchebichef polynomials [3][4]. DTT has a lower computational complexity and it does not require complex transform unlike continuous orthonormal transforms [5]. DTT does not involve any numerical approximation. The Tchebichef polynomials have unit weight and algebraic recurrence relations involving real coefficients, which make them matches for transforming the signal from time domain into frequency domain for speech recognition. DTT has been applied in several computer vision and image processing application in previous work. For example, DTT is used in image analysis [6][7], texture segmentation [8], multispectral texture [9], pattern recognition [10], image watermarking [11], monitoring crowds [12], image reconstruction [3][13][14], image projection [15] and image compression [16]-[18].

The organization of the paper is as follows. The next section gives a brief description on FFT, DTT, speech signal windowed, coefficient of discrete Tchebichef transform and spectrum analysis. Section 3 presents the comparison of spectrum analysis and time taken between 1024 DTT and 256 DTT. Finally, section 4 concludes the comparison of spectrum analysis using 1024 DTT and 256 DTT in terms of speech recognition.

2. TRANSFORMATION DOMAIN

FFT is an efficient algorithm that can perform Discrete Fourier Transform (DFT). FFT is applied in order to convert time domain signals $x(j)$ into the frequency domain $X(k)$. The sequence of N complex numbers $x_0, \ldots, x_{N-1}$ represents a given time domain signal. The following equation defines the Fast Fourier Transform of $x(j)$:

International Conference on Graphic and Image Processing (ICGIP 2012),
edited by Zeng Zhu, Proc. of SPIE Vol. 8768, 87680O · © 2013 SPIE
CCC code: 0277-786X/13/$18 · doi: 10.1117/12.2010642

$$X(k) = \sum_{j=1}^{N} x(j) e^{\frac{-2\pi i}{N}(j-1)(k-1)} \tag{1}$$

where $k = 1, \dots, N$, $x(j)$ is the sample at time index j and i is the imaginary number $\sqrt{-1}$. $X(k)$ is a vector of N values at frequency index k corresponding to the magnitude of the sine waves resulting from the decomposition of the time indexed signal. The FFT takes advantage of the symmetry and periodicity properties of the Fourier Transform to reduce computation time. It reduced the time complexity from $O(n^2)$ to $O(n \log n)$. In this process, the transform is partitioned into a sequence of reduced-length transforms that is collectively performed with reduced computation [19]. The FFT technique also has performance limitation as the method. FFT is a complex field computationally with imaginary numbers.

For a given positive integer N (the vector size) and a value n in the range $[1,\ N-1]$, the N order orthonormal Tchebichef polynomials $t_k(n),\ n = 1, 2, \dots, N-1$ are defined using the following recurrence relation [13]:

$$t_0(n) = \frac{1}{\sqrt{N}}, \tag{2}$$

$$t_k(0) = \sqrt{\frac{N-k}{N+k}} \sqrt{\frac{2k+1}{2k-1}} t_{k-1}(0), \tag{3}$$

$$t_k(1) = \left\{1 + \frac{k(1+k)}{1-N}\right\} t_k(0), \tag{4}$$

$$t_k(n) = \gamma_1 t_k(n-1) + \gamma_2 t_k(n-2), \tag{5}$$

$$k = 1, 2, \dots, N-1,\ n = 2, 3, \dots, \left(\frac{N}{2} - 1\right), \tag{6}$$

where

$$\gamma_1 = \frac{-k(k+1) - (2n-1)(n-N-1) - n}{n(N-n)}, \tag{7}$$

$$\gamma_2 = \frac{(n+1)(n-N-1)}{n(N-n)}, \tag{8}$$

The forward Discrete Tchebichef Transform (DTT) of order N is defined as follow:

$$X(k) = \sum_{n=0}^{N-1} x(n) t_k(n), \tag{9}$$
$$k = 0, 1, \dots, N-1,$$

where $X(k)$ denotes the coefficient of orthonormal Tchebichef polynomials. $x(n)$ is the sample of speech signal at time index n. The orthonormal Tchebichef polynomials are proper especially when Tchebichef polynomials of large degree are required to be evaluated. The orthonormal Tchebichef polynomials matches for signal processing which have large sample data represents speech signal. The Tchebichef transform involves only algebraic expressions and it can be compute easily using a set of recurrence relations.

The voice used in this experiment is the vowel ‘O’ and the consonant ‘RA’ from the International Phonetic Alphabet [20]. A speech signal has a sampling rate frequency component of about 11 KHz. Speech signals are highly redundant and contain a variety of background noise. For this reason, the threshold is 0.1 to remove the silence part. This means that any zero-crossings that start and end within the range of t_a, where $-0.1 < t_a < 0.1$, are not included in the total number of zero-crossings in that window. The Next step is pre-emphasis technique. Pre-emphasis is used in speech processing to enhance high frequencies of the signal. It reduces the high spectral dynamic range. Therefore, by applying pre-emphasis, the spectrum is flattened, consisting of formants of similar heights.

2.1 Speech Signal Windowed

The sample of sound has 4096 sample data which representing speech signal. The sample of speech signal is windowed into several frames. On one hand, speech signal of the vowel 'O' is windowed into four frames. Each window consumes 1024 sample data as presented in Fig. 1. In this experiment, the fourth frame for 3073-4096 sample data is used to analysis and evaluates using 1024 DTT and FFT.

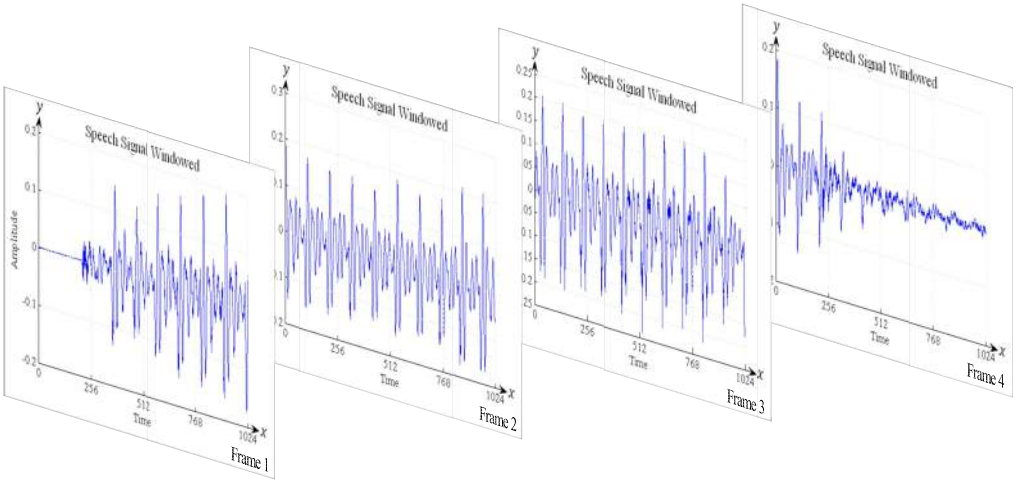

Figure 1. Speech signal windowed into four frames.

Next, the fourth frame of sample speech signal is computed with the DTT of order 1024 to transform time domain into frequency domain. The 1024 discrete orthonormal Tchebichef polynomials is shown in the Fig. 2.

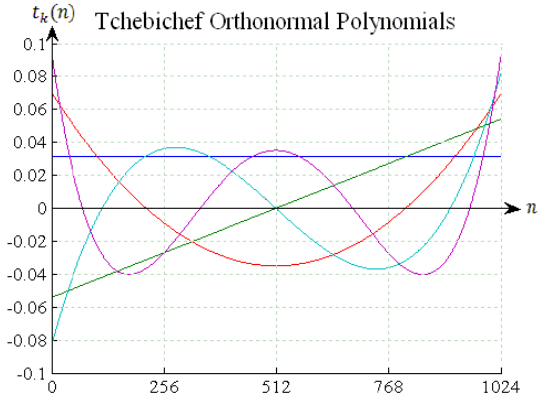

Figure 2. The First Five 1024 Discrete Orthonormal Tchebichef Polynomials for and .

On the other hand, the sample speech signal of the vowel 'O' is windowed into sixteen frames. Each window consists of 256 sample data which represents speech signals. The sample speech signal of the vowel 'O' is presented in Fig. 3.

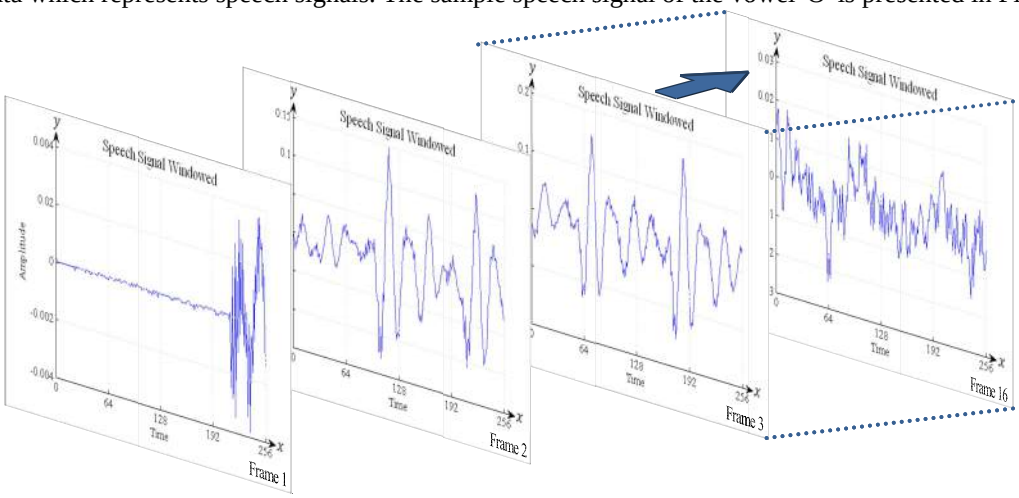

Figure 3. Speech signal windowed into sixteen frames.

In this study, the third frame for 513-768 sample data is used to compute using 256 discrete orthonormal Tchebichef polynomials. The 256 discrete Tchebichef transform is shown in Fig. 4.

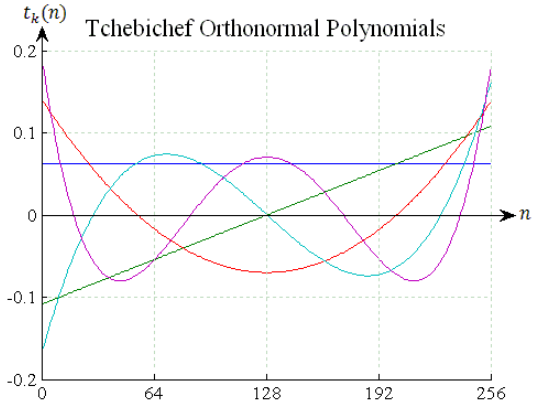

Figure 4. The First Five 256 Discrete Orthonormal Tchebichef Polynomials $t_k(n)$ for $k = 0, 1, 2, 3$ and 4.

2.2 Coefficient of Discrete Tchebichef Transform

Coefficient of DTT of order 256 sample data are given as follow formula:

$$C\,T = S \tag{10}$$

$$\begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} \begin{bmatrix} t_0(0) & t_0(1) & t_0(2) & \cdots & t_0(n) \\ t_1(0) & t_1(1) & t_1(2) & \cdots & t_1(n) \\ t_2(0) & t_2(1) & t_2(2) & \cdots & t_2(n) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t_k(0) & t_k(1) & t_k(2) & \cdots & t_k(n) \end{bmatrix} = \begin{bmatrix} x(0) \\ x(1) \\ x(2) \\ \vdots \\ x(n) \end{bmatrix}$$

where n is $0 \leq n \leq N-1$, C is the coefficient of discrete Tchebichef transform, which represents $c_0, c_1, c_2, \ldots, c_n$. T is matrix computation of discrete orthonormal Tchebichef polynomials $t_k(n)$ for $k = 0, 1, 2, \ldots, N-1$. S is the sample of speech signal at time index n, which represents $x(0), x(1), x(2), \ldots, x(n)$. The coefficient of DTT is given in as follow equation:

$$c(n) = \frac{x(n)}{t_k(n)} \tag{11}$$

where $x(n)$ is sample speech signal of order 256 sample data and $t_k(n)$ is orthonormal Tchebichef polynomials of order 256.

2.3 Spectrum Analysis

Spectrum analysis is the absolute square value of the speech signal, so the values are never negative. The spectrum analysis using DTT can be defined in the following equation:

$$p(k) = |c(n)|^2 \tag{12}$$

where $c(n)$ is coefficient of discrete Tchebichef transform. The spectrum analysis using DTT of the vowel 'O' and the consonant 'RA' [20] for 1024 sample data is shown on the left of Fig. 5 and Fig. 6. Next, the spectrum analysis using DTT of the vowel 'O' and the consonant 'RA' for 256 sample data is shown on the right of Fig. 5 and Fig. 6. The frequency formants of the vowel 'O' and the consonant 'RA' using 256 DTT is shown in Fig 7.

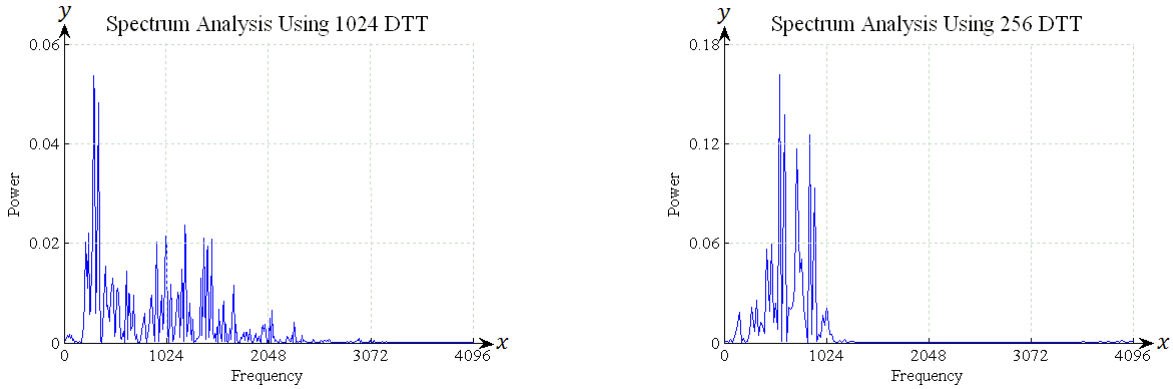

Figure 5. Coefficient of DTT for 1024 sample data (left) and coefficient of DTT for 256 sample data (right) for spectrum analysis of vowel ‘O’.

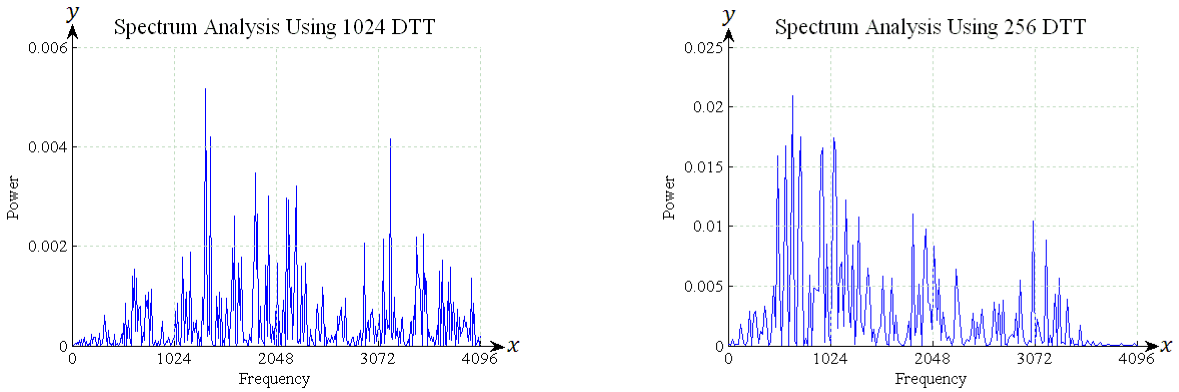

Figure 6. Coefficient of DTT for 1024 sample data (left) and coefficient of DTT for 256 sample data (right) for spectrum analysis of consonant ‘RA’.

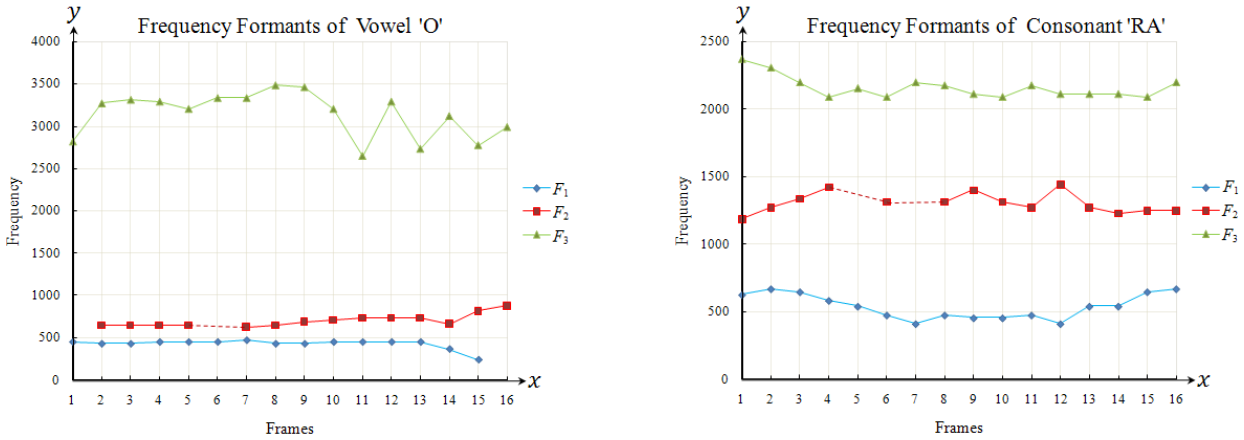

Figure 7. Frequency Formants of Vowel 'O' (left) and Consonant 'RA' (right) using 256 DTT.

The frequency formants of the vowel ‘O’ and the consonant ‘RA’ [20] using 256 DTT, 1024 DTT and 1024 sample data FFT computation are shown in Table 1.

Table 1. Frequency Formants of Vowel 'O' and Consonant 'RA'

Vowel 'O'	256 DTT	1024 DTT	1024 FFT	Consonant 'RA'	256 DTT	1024 DTT	1024 FFT
F_1	430	441	527	F_1	645	624	661
F_2	645	710	764	F_2	1335	1248	1301
F_3	3316	3186	3219	F_3	2196	2131	2160

The Time taken of DTT coefficient for 256 sample data, 1024 sample data and Fast Fourier Transform for 1024 sample data of vowel 'O' and consonant 'RA' are presented in Table 2. Next, the time taken of speech recognition performance using DTT and FFT is shown in Table 3.

Table 2. Time Taken of DTT Coefficient and FFT

Vowel and Consonant	DTT		FFT
	256	1024	1024
Vowel 'O'	0.001564 sec	0.011889 sec	0.095789 sec
Consonant 'RA'	0.001605 sec	0.027216 sec	0.102376 sec

Table 3. Time Taken of Speech Recognition Performance

Vowel and Consonant	DTT		FFT
	256	1024	1024
Vowel 'O'	0.557000 sec	0.903968 sec	0.587628 sec
Consonant 'RA'	0.557410 sec	0.979827 sec	0.634997 sec

3. Comparative Analysis

Spectrum analysis of the vowel 'O' and consonant 'RA' using 1024 DTT on the left of Fig. 5 and Fig. 6 produce a lower power spectrum than 256 DTT. Next, spectrum analysis of vowel 'O' and consonant 'RA' using 256 DTT on the right of Fig. 5 and Fig. 6 produce simpler output than 1024 DTT. According observation as presented in the Fig. 7, frequency formants for sixteen frames show that identically similar output among each frame. Frequency formants as presented in Table 1 show those frequency formants using 256 DTT produce similar output with frequency formants using 1024 DTT. The discrete Tchebichef transform that developed in this paper is smaller computations. As proposed a 256 forward discrete Tchebichef transform can be used in spectrum analysis in terms of speech recognition. The faster and efficient of 256 discrete Tchebichef transform is much higher than 1024 discrete Tchebichef transform in terms of speech recognition performance. The time taken of DTT coefficient, FFT and speech recognition performance as represented in the Table 2 and Table 3 respectively show that the spectrum analysis using 256 DTT is faster and computationally efficient. The experiment result have shown that the propose 256 discrete Tchebichef transform algorithm efficiently reduces the time taken to transform time domain into frequency domain.

FFT algorithm produces the time complexity $O(n \log n)$. Next, the computation time of DTT produce time complexity $O(n^2)$. For the future research, DTT can be upgraded to reduce the time complexity from $O(n^2)$ to be $O(n \log n)$ using convolution algorithm. DTT can increase the speech recognition performance and get the similarity frequency formants in terms of speech recognition.

4. CONCLUSION

As a discrete orthonormal transform, 256 discrete Tchebichef transform has potential faster and computationally more efficient than 1024 discrete Tchebichef transform. A 256 DTT produces simpler output in spectral frequency than DTT which required 1024 sample data. On one hand, FFT is computationally complex especially with imaginary numbers. On the other hand, DTT consumes simpler and faster computation with involves only algebraic expressions and it can be compute easily using a set of recurrence relations. Spectrum analysis using 256 DTT produces four formants F_1, F_2, F_3 and F_4 concurrently in spectrum analysis for consonant. The frequency formants using 1024 DTT and 256 DTT are compared. They have produced frequency formants relatively identical outputs in terms of speech recognitions.

ACKNOWLEDGMENTS

The authors would like to express a very special thanks to Universitas Dian Nuswantoro (UDINUS), Semarang, Indonesia for giving the financial support funding this research project.

REFERENCES

[1] Vite-Frias, J.A., Romero-Troncoso, Rd.J. and Ordaz-Moreno., "A. VHDL Core for 1024-point radix-4 FFT Computation," *International Conference on Reconfigurable Computing and FPGAs*, pp. 20-24 (2005).

[2] Bailey, D.H. and Swarztrauber, P.N., "A Fast Method for Numerical Evaluation of Continuous Fourier and Laplace Transform," *Journal on Scientific Computing*, Vol. 15, No. 5, pp. 1105-1110 (1994).

[3] Mukundan, R., "Improving Image Reconstruction Accuracy Using Discrete Orthonormal Moments," *Proceedings of International Conference on Imaging Systems, Science and Technology*, pp. 287-293 (2003).

[4] Mukundan, R., Ong, S.H. and Lee, P.A., "Image Analysis by Tchebichef Moments," *IEEE Transactions on Image Processing*, Vol. 10, No. 9, pp. 1357–1364 (2001).

[5] Ernawan, F., Abu, N.A. and Suryana, N., "Spectrum Analysis of Speech Recognition via Discrete Tchebichef Transform," *Proceedings of International Conference on Graphic and Image Processing (ICGIP 2011)*, Proceeding of the SPIE, Vol. 8285, No. 1, pp. 82856L-82856L-8 (2011).

[6] Teh, C.-H. and Chin, R.T., "On Image Analysis by the Methods of Moments," *IEEE Transactions on Pattern Analysis Machine Intelligence*, Vol. 10, No. 4, pp. 496-513 (1988).

[7] Abu, N.A., Lang, W.S. and Sahib, S., "Image Super-Resolution via Discrete Tchebichef Moment," *Proceedings of International Conference on Computer Technology and Development (ICCTD 2009)*, Vol. 2, pp. 315–319 (2009).

[8] Tuceryan, M., "Moment Based Texture Segmentation," *Pattern Recognition Letters*, Vol. 15, pp. 659-668 (1994).

[9] Wang, L. and Healey, G., "Using Zernike Moments for the Illumination and Geometry Invariant Classification of Multispectral Texture," *IEEE Transactions on Image Processing*, Vol. 7, No. 2, pp. 196-203 (1998).

[10] Zhang, L., Qian, G.B., Xiao, W.W. and Ji, Z., "Geometric Invariant Blind Image Watermarking by Invariant Tchebichef Moments," *Optics Express Journal*, Vol. 15, No. 5, pp. 2251-2261 (2007).

[11] Zhu, H., Shu, H., Xia, T., Luo, L. and Coatrieux, J.L., "Translation and Scale Invariants of Tchebichef Moments," *Journal of Pattern Recognition Society*, Vol. 40, No. 9, pp. 2530-2542 (2007).

[12] Rahmalan, H., Suryana, N. and Abu, N. A., "A General Approach for Measuring Crowd Movement," *Malaysian Technical Universities Conference and Exhibition on Engineering and Technology (MUCEET2009)*, pp. 098-103 (2009).

[13] Mukundan, R., "Some Computational Aspects of Discrete Orthonormal Moments," *IEEE Transactions on Image Processing*, Vol. 13, No. 8, pp. 1055-1059 (2004).

[14] Abu, N.A., Suryana, N. and Mukundan, R. Perfect Image Reconstruction Using Discrete Orthogonal Moments. *Proceedings of the 4th IASTED International Conference on Visualization, Imaging, and Image Processing (VIIP2004)*, Sep. 2004, pp. 903-907.

[15] Abu, N.A., Lang, W.S. and Sahib, S., "Image Projection Over the Edge," *International Conference on Industrial and Intelligent Information (ICIII 2010)*, Proceedings 2nd International Conference on Computer and Network Technology (ICCNT2010), pp. 344-348 (2010).

[16] Mukundan, R. and Hunt, O., "A Comparison of Discrete Orthogonal Basis Functions for Image Compression," *Proceedings Conference on Image and Vision Computing New Zealand (IVCNZ 04)*, pp. 53-58 (2004).

[17] Lang, W.S., Abu, N.A. and Rahmalan, H., "Fast 4x4 Tchebichef Moment Image Compression," *Proceedings International Conference of Soft Computing and Pattern Recognition (SoCPaR2009)*, pp. 295 300 (2009).

[18] Abu, N.A., Lang, W.S., Suryana, N. and Mukundan, R., "An Efficient Compact Tchebichef moment for Image Compression," *10th International Conference on Information Science, Signal Processing and their applications (ISSPA2010)*, pp. 448-451 (2010).

[19] Rapuano, S. and Harris, F., "An Introduction to FFT and Time Domain Windows," *IEEE Instrumentation and Measurement Society*, Vol. 10, No. 6, pp. 32-44 (2007).

[20] Esling, J.H. and O'Grady, G.N., "The International Phonetic Alphabet," *Linguistics Phonetics Research*, Department of Linguistics, University of Victoria, Canada (1996).