A Study on Pitch Period Detection Algorithm Based on Rotation Transform of AMDF and Threshold

Hyun-Soo Seo*, Nam-Ho Kim*

Abstract

As a lot of researches on the speech signal processing are performed due to the recent rapid development of the information-communication technology, the pitch period is used as an important element to various speech signal application fields such as the speech recognition, speaker identification, speech analysis, or speech synthesis. A variety of algorithms for the time and the frequency domains related with such pitch period detection have been suggested. One of the pitch detection algorithms for the time domain, AMDF (average magnitude difference function) uses distance between two valley points as the calculated pitch period. However, it has a problem that the algorithm becomes complex in selecting the valley points for the pitch period detection. Therefore, in this paper we proposed the modified AMDF(M-AMDF) algorithm which recognizes the entire minimum valley points as the pitch period of the speech signal by using the rotation transform of AMDF. In addition, a threshold is set to the beginning portion of speech so that it can be used as the selection criteria for the pitch period. Moreover the proposed algorithm is compared with the conventional ones by means of the simulation, and presents better properties than others.

Keywords : Pitch period, AMDF, valley point

Ⅰ. Introduction

Recently, the speech signal processing has been getting a growing attention due to the rapid of the information-communication technology. In addition, the pitch period, the fundamental frequency of the speech signal is used as a parameter sensitive to the human sense of hearing for recognizing speakers of the speech signal, which large affects on the natural. Accordingly, extracting and reconstructing of the pitch in the speech synthesis and the speech coding plays a key role in the sound quality. Therefore, many algorithms for detecting the pitch in the time and the frequency domains have been suggested [1]-[7].

The pitch-detecting algorithms in the time domain use periodic property of the waveform for detecting pitch, which includes ACF(autocorrelation function), AMDF, and so on [2]-[5]. Such algorithms in the time domain do not need the domain transform for analyzing, but only use simple operations such as summation and subtraction, or comparison logics. However, when the speech exists in the transition region so that amplitude difference in the frames largely occurs, or noise is superposed with the speech, the algorithms have a drawback of being complex for detecting the pitch.

In addition, the pitch detection algorithms in the frequency domain measure the harmonic intervals of the speech spectrum for detecting the fundamental frequency of voiced sounds, which includes the harmonic analysis method [7]. In general, because a signal in a frame can be separated according to the frequency bands, these algorithms have an advantage when noise is superposed in a band adjacent to the speech frequency, or the speech exists in the transition region.

However, while processing the signal, complex calculations are necessary for transforming to the frequency domain, and the transform property becomes slower when the number of FFT points for detecting precise fundamental frequency increases. Therefore, AMDF is widely used in real-time systems because it has advantages of detecting precise pitch period and

* School of Electrical and Control Engineering, Pukyong National University
접수 일자 : 2006. 8. 4 수정 완료 : 2006. 10. 23
논문 번호 : 2006-4-1

less complexity, but selecting valley points for detecting the pitch period is not an easy problem [8].

Therefore, in this paper, a new algorithm for detecting the pitch period using the rotation transform of AMDF is proposed, and the threshold set for the beginning portion of speech is used as the selection criteria for the pitch period. And, the proposed modified AMDF(M-AMDF) algorithm is compared with representative the linear-weighted AMDF(W-AMDF) and the length-varied AMDF(LV-AMDF) algorithms in pitch period detection algorithms using AMDF [9]-[12]. This paper is organized as follows: in section 2, we briefly introduce basic-AMDF algorithm and the conventional AMDF algorithm. In section 3, the proposed M-AMDF algorithm is derived and section 4 presents experimental results. Finally in section 5, we summarize the conclusion.

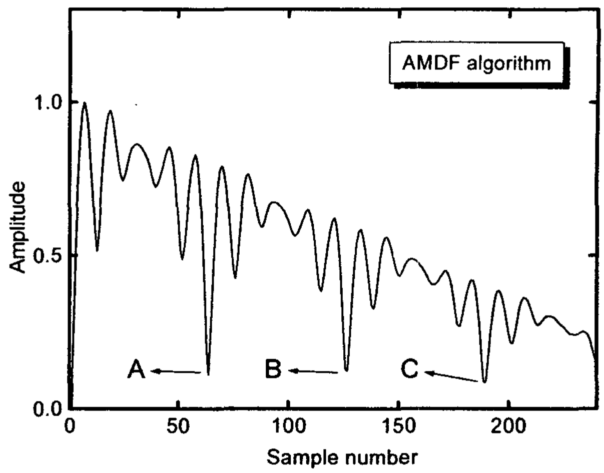


Fig. 1. AMDF in voiced speech.

II. AMDF Algorithm

1. Basic-AMDF algorithm

If $s(n)$ is a speech signal, $s_w(n)$ is derived from $s(n)$ by means of a calculation with a window function , $w(n)$ having length of N. In such a case, the basic-AMDF, $\gamma(l)$ can be defined as follows.

$$\gamma(l) = \sum_{n=1}^{N-l+1} |s_w(n+l-1) - s_w(n)|, \quad l = 1, 2, \cdots, N \tag{1}$$

According to equation (1), when $s_w(n)$ is a voiced sound, $\gamma(l)$ shows the periodic property. In addition, when the delay element, l is same with the period, $\gamma(l)$ shows a minimum value close to 0 [8]. In general, the voiced sound shows periodic local minimum valley points by the AMDF algorithm as shown in Fig. 1, and the first valley point for the entire samples is used for calculating the pitch period.

Therefore, in Fig. 1, the time interval between the local minimum valley point A and origin point is regarded as the 'pitch point'. But, if the first local minimum valley point is not the global lowest valley point, it is not easy to select the first local minimum valley point to detect pitch period.

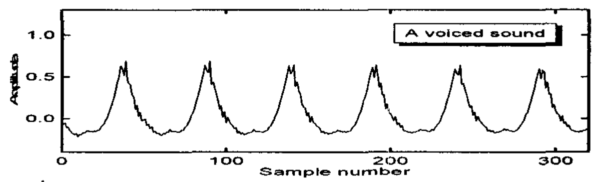
2. Conventional AMDF algorithm

After the basic-AMDF was made, some applications of basic-AMDF have been introduced. And there are W-AMDF and LV-AMDF respectively.

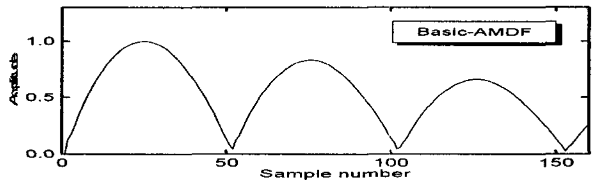
The W-AMDF can be defined by equation (2).

$$\gamma_W(l) = \frac{1}{N-l+1} \sum_{n=1}^{N-l+1} |s_w(n+l-1) - s_w(n)|, \quad l = 1, 2, \cdots, N \tag{2}$$

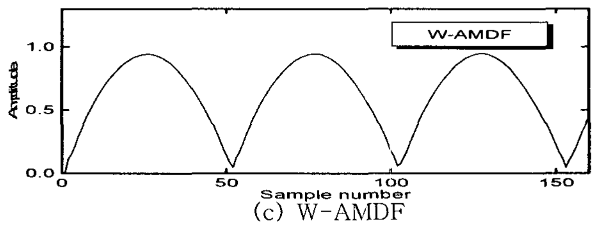
The W-AMDF is introduced to overcome pitch detection error which basic-AMDF has in low-pitched signal environments.



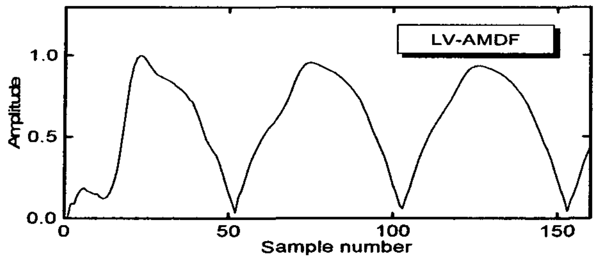
(a) A Speech signal



(b) Basic-AMDF



(c) W-AMDF



(d) LV-AMDF

Fig. 2. Several AMDF.

That is, according to the above equation, the scale factor, $(N-l+1)^{-1}$ can remove the decreasing property of the basic-AMDF.

In general, because the minimum of AMDF has a value close to 0, l given as the minimum of AMDF, i.e. the fundamental frequency should be properly selected.

In such a case, since position of the fundamental frequency on the frequency axis, f_0 depends on the sampled frequency, F_S, l can be obtained from the relation of F_S.

$$\gamma_{LV}(l)=\frac{1}{l}\sum_{n=1}^{l}|s_w(n+l-1)-s_w(n)|,\quad l=1,2,\cdots,\frac{N}{2} \tag{3}$$

As shown in equation (3), the LV-AMDF shows same periodicity as the existing AMDF. In addition, because complexity for summation is determined by a variable l, it has less complexity compared with the basic-AMDF and the W-AMDF. Fig. 2 is a result that basic-AMDF and conventional AMDF is applied to a voiced sound. However, the conventional AMDF still has an error in detecting the first local minimum valley point for general voiced sounds. Therefore existing AMDF needs following additional algorithm shown in Fig. 3 [8].

Step 1. Detect local minimum valley points satisfied with Th1.

Step 2. Remove values which are not satisfied with time interval Th2.

Step 3. Choose the leftest value satisfied with Th3, determining the deepness of valley points.

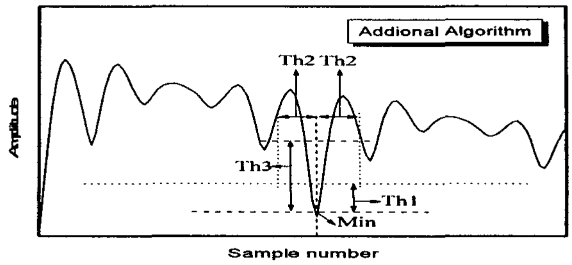
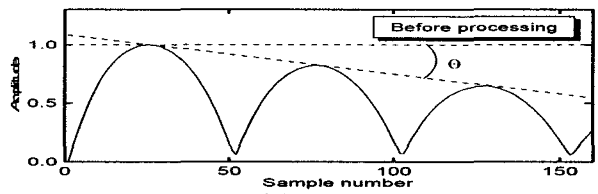


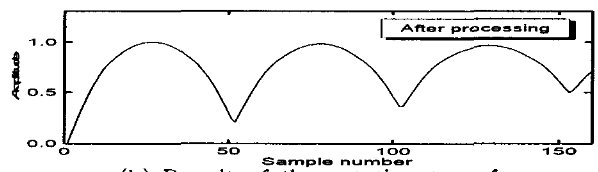
Fig. 3. Null picking of AMDF.

1. Ⅲ. Proposed M-AMDF

According to this paper, by analyzing the decreasing characteristic of the basic-AMDF $\gamma(l)$, angle Θ to the horizontal line, formed by a line connecting the first peak point of the AMDF and the corresponding point, is set as shown in Fig. 4(a) and (b).



(a) Angle Θ



(b) Result of the rotation transform

Fig. 4. M-AMDF.

Then, in order to make the angle Θ to 0°, rotate the AMDF in the positive direction. Following equation (4) is a derived equation for setting the angle Θ and this angle Θ is established adaptively. Here, $R_{\max}=\max\{\gamma(l)\}$, N indicates the signal length for analyzing $\gamma(l)$ and $l_{R\max}$ is index of $R_{\max}$.

$$\Theta = \arctan\frac{R_{\max}}{N-l_{R\max}} \tag{4}$$

Following equation (5) represents M-AMDF by rotating the basic-AMDF with the angle Θ.

$$\gamma_M(l) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \cdot \gamma(l) \qquad (5)$$

M-AMDF proposed in equation (5) makes it easy to detect the first valley point of the pitch period by transforming the basic-AMDF $\gamma(l)$ with the angle θ.

In addition, M-AMDF additionally includes a parameter α as shown in following equation (6) so as to have the flexibility for the rotation angle. Therefore, it is easier to detect the global minimum valley point by selecting the proper value, α. Its effect is shown in Fig. 5 and in this figure, α is a parameter for flexibility of angle.

$$\theta = \arctan\frac{R_{\max}}{N - l_{R\max}} + \alpha \qquad (6)$$

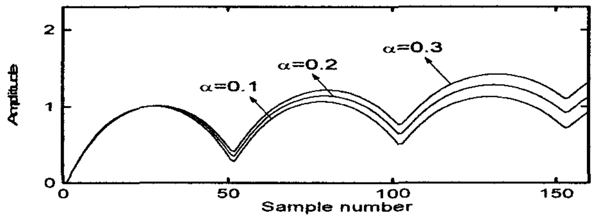
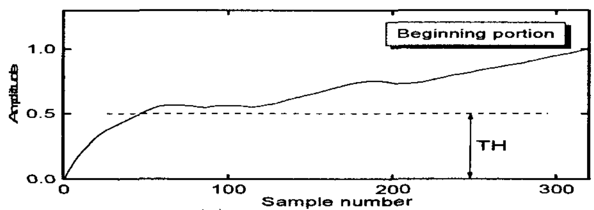
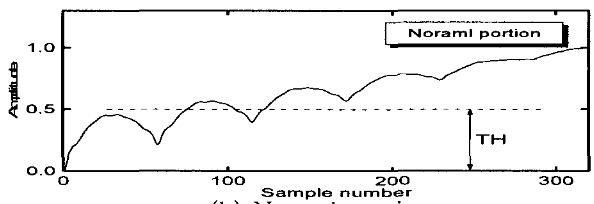


Fig. 5. AMDF for various angles.

In addition, because the first local minimum valley point for detecting the pitch period is not clear for the beginning portion of speech as shown in Fig. 6(a), an error may occur in detecting the pitch period.



(a) Beginning portion



(b) Normal portion

Fig. 6. Threshold TH of M-AMDF.

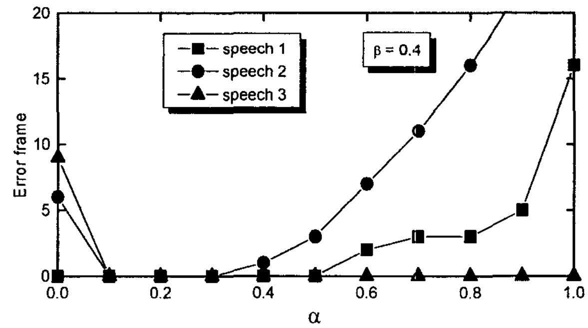
Therefore, by applying threshold TH to each frame, only valley points having smaller amplitude than TH are selected as the pitch period. Here, TH can be established by the following equation (7), and β is a parameter for determining threshold level.

$$TH = \beta \cdot \gamma_M(last) \qquad (7)$$

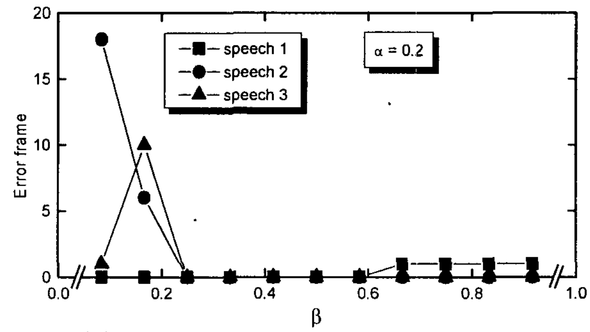
Ⅳ. Simulation and Results

In order to simulate the pitch period detection algorithm of the speech signal using M-AMDF proposed in this paper, the sound card of P4i4SGV model is used, where the resolution is 16-bit, and the sampling frequency is set to 8 [kHz]. In addition, length of one frame is 10 [ms] (80 samples), length of the analysis frame is 4 frames (320 samples). And frames are overlapped one by one for detecting the pitch period.

Fig. 7 is the result of error frames about three different speeches with α, β. And speech 1, speech 2 and speech 3 are speeches pronounced as '/Hwan/ /gol/ /tal/ /tae/', '/Il/ /sa/ /cheon/ /li/' and '/Cheon/ /go/ /ma/ /bi/' respectively.

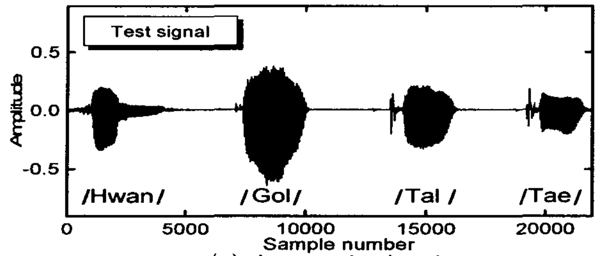


(a) Comparision of error frames with α

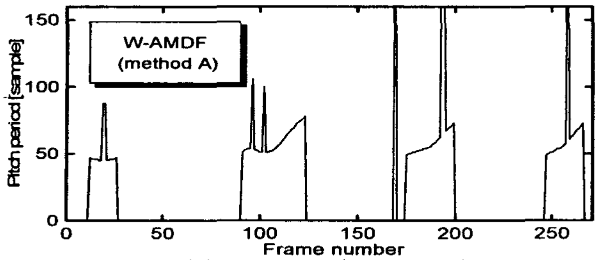


(b) Comparision of error frames with β

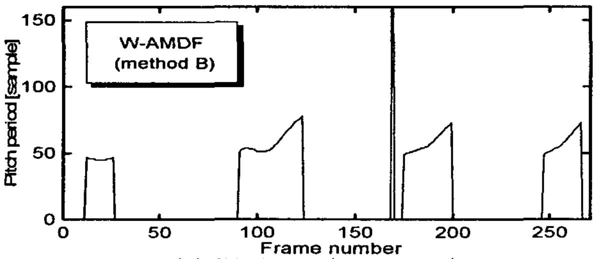
Fig. 7. Simulation results for setting α, β.



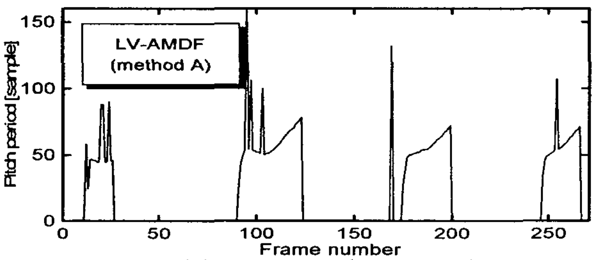
(a) A speech signal



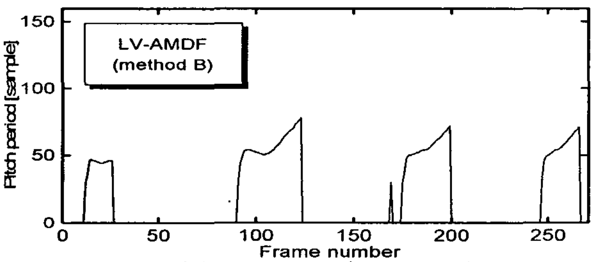
(b) W-AMDF(method A)



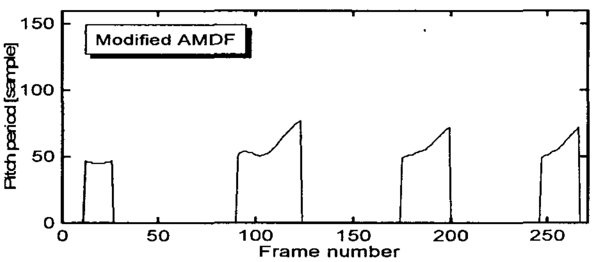
(c) W-AMDF(method B)



(d) LV-AMDF(method A)



(e) LV-AMDF(method B)



(f) M-AMDF ($\alpha = 0.2$, $\beta = 0.4$)

Fig. 8. Simulation results.

From the Fig. 7(a), in order to choose a proper α, error frames were compared with each other by changing the value of α, after fixing β. In $\alpha = 0.2$, the result proved that the fewest error appeared. Also, in Fig. 7(b), β could be chosen in the lowest error frame after α was fixed. So, when $\alpha = 0.2$ and $\beta = 0.4$, the proposed algorithm presented good pitch detection properties.

Fig. 8 shows the simulation results for a speech signal pronounced as '/Hwan/ /gol/ /tal/ /tae/'.

Fig. 8(b) represents the detected pitch period by the W-AMDF using the method which is to recognize the global minimum valley point as pitch period(method A). And Fig. 8(c) is the result by the W-AMDF using the additional algorithm with thresholds(method B : $Th\ 1 = 0.2$, $Th\ 2 = 8$, $Th\ 3 = 0.5$).

Fig. 8(d) represents the detected pitch period by the LV-AMDF using the method which is to recognize the global minimum valley point as pitch period(method A). And Fig. 8(e) is the result by the LV-AMDF using the additional algorithm with thresholds of same values as W-AMDF.

Fig. 8(f) shows the detected pitch period by the M-AMDF.

Table 1 is the result of error frames about three speeches used in Fig. 7. From the table 1, we know proposed M-AMDF has better capability than existing AMDF. And effects of α and β on M-AMDF presented in this table. As shown in the table, when α and β were 0.2 and 0.4 respectively, it showed excellent detection results for the pitch period.

Table 1. Comparison of error frames about three speeches

	Error frames		
	Speech 1 (94 frame)	Speech 2 (93 frame)	Speech 3 (113 frame)
W-AMDF (method A)	8	14	8
W-AMDF (method B)	1	0	0
LV-AMDF (method A)	12	18	12
LV-AMDF (method B)	6	4	4
M-AMDF ($\alpha = 0.2$, $\beta = 0.3$)	0	6	10
M-AMDF ($\alpha = 0.6$, $\beta = 0.4$)	2	7	0
M-AMDF ($\alpha = 0.2$, $\beta = 0.4$)	0	0	0

V. Conclusion

In this paper, in order to detect the pitch period of speech, M-AMDF algorithm using the rotation transform of the basic-AMDF is proposed, and it shows excellent results when compared with the conventional W-AMDF and LV-AMDF by means of the simulation.

The proposed method is a simple algorithm for selecting the entire minimum valley points as the pitch period. In particular, it removes error for the beginning portion of speech by setting threshold TH. Therefore, the proposed algorithm for the pitch detection is thought be used for various application fields in the speech signal processing such as the speech synthesis, speech recognition, or speaker identification.

References

[1] L. R. Rabiner et al., "A comparative performance study of several pitch detection algorithms", IEEE Trans. Acoust., Speech, Signal Processing, 24(5), pp. 399-417, 1976.

[2] M. J. Ross et al., "Average magnitude difference function pitch extractor", IEEE Trans. Acoust., Speech, Signal Processing, 22(5), pp. 353-362, 1974.

[3] B. Gold and L. R. Rabiner, "Parallel processing techniques for estimating pitch periods of speech in the time domain", J. Acoust. Soc. Amer., vol. 46, pp. 443-448, Aug, 1976.

[4] A. M. Noll., "Cepstrum pitch determination", J. Acoust. Soc. Amer., 41(2): 293-309, 1967.

[5] L. R. Rabiner, "On the use of autocorrelation analysis for pitch detection", IEEE Trans. ASSP, ASSP-25: 24-33, 1977.

[6] S. Kadambe and G. F. Boudreaux-Bartels, "Application of the wavelet transform for pitch detection of speech signals", IEEE Trans. Information Theory, 38(2), pp. 917-924, 1992.

[7] Sneff, "Real time harmonic pitch detection", IEEE Trans. Acoust., Speech, Signal Processing, Vol. ASSP-26, 1978.

[8] W. Hess, "Pitch determination of speech signals: algorithms and devices", Berlin: Springer-Verlag, 1983.

[9] C. Manfredi, "Adaptive noise energy estimation in pathological speech signals", IEEE Trans. Biomedical Engineering, 47(11), pp. 1538-1543, 2000.

[10] X. mei, J. Pan and S. Sun, "Efficient algorithms for speech pitch estimation", Proc. of ISIMVSP-2001, pp. 421-424, 2001.

[11] G. S. Ying, L. H. Jamieson and C. D. "Michell. A probabilistic approach to AMDF pitch detection", IEEE ICSLP-96. pp. 1201-1204, 1996.

[12] C. Kwanun and S. C. Ying. "A pitch extraction algorithm based on LPC inverse filtering and AMDF", IEEE Trans. Acoust., Speech, Signal Processing. ASSP-25(6): 565-572, 1977.

Hyun-Soo Seo

He received the BS degree in Control and Instrumentation engineering from Pukyong National University, Busan, Korea in 2005. Since 2005 to now, He has been MS course student in Department of Control and Instrumentation engineering, Pukyong National University, Busan, Korea. His reserch interests include digital signal processing, speech signal processing and circuit systems.

Nam-Ho Kim

Corresponding author

He received the BS, MS and PhD degree in electronics engineering from Yeungnam University, Daegu, Korea in 1984, 1986 and 1991, respectively. He is currently a Professor in School of Electrical and Control Engineering, Pukyong National University. His research interests include circuits and measurement systems, theory for digital signal compression, restoration and noise removal, non-linear filtering, signal processing for communications, and wavelet transforms and applications.