2010 3rd International Conference on Computer and Electrical Engineering (ICCEE 2010)
IPCSIT vol. 53 (2012) © (2012) IACSIT Press, Singapore
DOI: 10.7763/IPCSIT.2012.V53.No.2.20

Dual Cubature Kalman Filter for Speech Enhancement

Xianfeng Tang[1], Binglei Guan[2] and Quanbo Ge[3+]

[1]Modern Education Technology Center, Zhejiang University

Hangzhou, P.R. China

[2]Telecom College, Ningbo University of Technology

Ningbo, P.R. China

[3]Institute of Information and Control, Hangzhou Dianzi University

Hangzhou, P.R. China

Abstract—The speech enhancement using dual cubature Kalman filter (DCKF) is developed in this paper. The speech model is nonlinear and modeled by a BP neural network. Subsequently, we adopt the cubature Kalman filter technology to estimate the speech states and weights in parallel. Finally, two experiments are carried out using the TIMIT speech database to evaluate the new algorithm. Compared with the spectral subtraction approach, the results show that the novel method can further improve the speech feature enhancement performance in terms of the signal to noise ratio (SNR) improvements.

Keywords- speech enhancement ; DCKF; BP neural network

1. Introduction

Speech enhancement is an active area for speech signal processing. It has many applications in voice communication, speech recognition, and hearing aids and so on. In the real world, speech signals are often distorted by background noise such as car engine, traffic or wind. Such degradation can lower the intelligibility or quality of speech. So, it is necessary to preprocess noisy signal using speech enhancement technology, which can retrain the noises and improve the performance in speech reorganization system.

Speech enhancement algorithms have been studied a lot for the past three decades. The spectral subtraction approach is commonly used as a standard algorithm. The advantage of this method is that it is straightforward and easy to implement. However, it will result in a series of annoying low-level tones called musical noises throughout the estimated signal [1]. The Wiener filtering algorithm was first proposed to be applied to speech enhancement by Lim, etc. [2]. It is optimal estimator in the sense of minimum mean square error (MMSE) while speech signal and noise signal are stationary. Unfortunately, these signals are non-stationary in practical application. The authors in [3] proposed the use of a Kalman Filter (KF) for the purpose of speech enhancement. Compared to the Wiener filtering method, the performance of this model-based approach was shown to be considerably better. Above methods depend on a linear prediction model of the speech signal. However there is strong theoretical and experimental evidence for the existence of important nonlinear aerodynamic phenomena during the speech production that cannot be accounted for by the linear model [4]. In recent years, extended Kalman filter (EKF) and Unscented Kalman filter (UKF) were applied to

[+] Corresponding author.
E-mail address: quanboge@163.com

speech enhancement rely on nonlinear auto-regression model, which can remove non-stationary noise[5,6]. However, the limitations of EKF and UKF are as follows. 1) Estimate accuracy of EKF is too low owing to the first-order Taylor series approximation for nonlinear functions. 2) The performance of UKF will decrease obviously when the system state dimension is high comparatively, which is called the curse of dimensionality. Recently, the Cubature Kalman Filter (CKF) was developed in [7]. It adopts the spherical cubature rule and radial rule to optimize the sigma points and weights. So, the ability to deal with high dimension state is greatly enhanced for the nonlinear estimation.

In view of this, it makes the nonlinear model for the speech signal using the BP network. Applying the CKF into the speech enhancement, a novel speech enhancement algorithm based on dual Cubature Kalman Filter (DCKF) is presented. Simulation results show that proposed method can effectively improve the SNR.

The rest of this paper is organized as follows. Section II presents BP neural network modeling of noisy speech in a state-space framework. In Section III, we describe how the DCKF can be applied in speech enhancement. Experimental results are reported in Section IV and conclusions are drawn in Section V.

2. A Neural State-Space Model for Speech Signal

In this paper, the clean speech signal is modeled as an order p nonlinear autoregressions with process noise [5]

$$x(k) = f\left(x(k-1), x(k-2), \cdots, x(k-p), \mathbf{w}(k)\right) + u(k-1) \tag{1}$$

where $x(k)$ is the clean speech signal; $u(k)$ is the process noise; $\boldsymbol{w}(k)$ is the weight vector of the BP neural network model. $f(\cdot)$ is the nonlinear function including parameter and past values of $x(k)$.

The noisy speech signal with the additive noises can be formulated as follow

$$y(k) = x(k) + v(k) \tag{2}$$

where, $v(k)$ is the measurement noise. A state-space formulation of equation (1) and (2) is

$$\begin{cases} \boldsymbol{x}(k) = F\left(\boldsymbol{x}(k-1)\right) + \boldsymbol{G}u(k) \\ y(k) = \boldsymbol{H}\boldsymbol{x}(k) + v(k) \end{cases} \tag{3}$$

where,

$$\boldsymbol{x}(k) = \begin{bmatrix} x(k-p+1) \\ \vdots \\ x(k-1) \\ x(k) \end{bmatrix}; \boldsymbol{G} = \boldsymbol{H}^T = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix};$$

$$F(\boldsymbol{x}(k)) = \begin{bmatrix} x(k-p+2) \\ \vdots \\ x(k) \\ f\left(x(k), x(k-1), \cdots, x(k-p+1), \boldsymbol{w}(k)\right) \end{bmatrix}$$

It is assumed that $u(k)$ and $v(k)$ are uncorrelated Gaussian white noise sequences with zero means and variances σ_u^2 and σ_v^2.

Now, the problem is how to design optimal estimator given the noisy measurements $\boldsymbol{y}(k) = \{y(0), y(1), \cdots, y(k)\}$. That is to solve $E[\boldsymbol{x}(k) \mid \boldsymbol{y}(k)]$.

3. Dual Cubature Kalman Filter for Speech Enhancement

In the Bayesian framework, the nonlinear filter reduces to computing multi-dimensional integrals when all conditional densities are assumed to be Gaussian, whose integrands are all of the form nonlinear function multiply the Gaussian. Similar to UKF, CKF obtains a set of cubature points and its weights according to spherical-radial integration rule. For a linear model with known parameters, the KF algorithm can be readily used to estimate the states. When the model is nonlinear, KF cannot be applied directly, but CKF is still effective.

3.1. State Estimation

Assume the state estimate $\hat{\boldsymbol{x}}(k-1|k-1)$ and its error covariance matrix $\boldsymbol{P}(k-1|k-1)$ are available at time k, the CKF can be executed as follows

1) Time Update

- Factorize

$$\boldsymbol{P}(k-1|k-1)=\boldsymbol{S}(k-1|k-1)\boldsymbol{S}^T(k-1|k-1) \tag{6}$$

- Calculate the cubature points($i=1,2,\cdots,m_x$)

$$\boldsymbol{x}_i(k-1|k-1)=\boldsymbol{S}(k-1|k-1)\boldsymbol{\xi}_i+\hat{\boldsymbol{x}}(k-1|k-1) \tag{7}$$

where, $m_x=2p$; And the parameter $\boldsymbol{\xi}_i$ ($\boldsymbol{\varepsilon}_i$ is the *p*-order unit vector) is given by

$$\boldsymbol{\xi}_i=\begin{cases}\sqrt{m_x/2}\cdot\boldsymbol{\varepsilon}_i, & i=1,2,\cdots,p\\ -\sqrt{m_x/2}\cdot\boldsymbol{\varepsilon}_{i-p} & ,i=p+1,p+2,\cdots,2p\end{cases}$$

- Compute the propagated cubature points ($i=1,2,\cdots,m_x$)

$$\boldsymbol{x}_i^*(k|k-1)=F\left(\boldsymbol{x}_i(k-1|k-1)\right) \tag{8}$$

- Evaluate the predicted state and its error covariance

$$\begin{cases}\hat{\boldsymbol{x}}(k|k-1)=\dfrac{1}{m_x}\displaystyle\sum_{i=1}^{m_x}\boldsymbol{x}_i^*(k|k-1)\\ \boldsymbol{P}(k|k-1)=\dfrac{1}{m_x}\displaystyle\sum_{i=1}^{m_x}\boldsymbol{x}_i^*(k|k-1)[\boldsymbol{x}_i^*(k|k-1)]^T\\ \qquad -\hat{\boldsymbol{x}}(k|k-1)\hat{\boldsymbol{x}}^T(k|k-1)+\boldsymbol{R}_u(k)\end{cases} \tag{9}$$

2) Measurement Update

- Estimate the predicted measurement

$$\hat{y}(k|k-1)=\boldsymbol{H}\hat{\boldsymbol{x}}(k|k-1) \tag{10}$$

- Compute the gain matrix of the KF

$$\boldsymbol{K}(k)=\boldsymbol{P}(k|k-1)\boldsymbol{H}^T[\boldsymbol{H}\boldsymbol{P}(k|k-1)\boldsymbol{H}^T+\sigma_v^2]^{-1} \tag{11}$$

- Evaluate the updated states and its error covariance

$$\begin{cases}\hat{\boldsymbol{x}}(k|k)=\hat{\boldsymbol{x}}(k|k-1)+\boldsymbol{K}(k)[y(k)-\hat{y}(k|k-1)]\\ \boldsymbol{P}(k|k)=[\boldsymbol{I}-\boldsymbol{K}(k)\boldsymbol{H}]\boldsymbol{P}(k|k-1)\end{cases} \tag{12}$$

- Finally, calculate the enhanced speech signal

$$\hat{x}(k)=\boldsymbol{G}^T\hat{\boldsymbol{x}}(k|k) \tag{13}$$

3.2. Weight Estimation

Because the model parameter of the clear speech signal is unknown in practice, the above method can not be applied directly. Now, it needs to construct a separate state-space formulation for weights $\boldsymbol{w}(k)$

$$\begin{cases}\boldsymbol{w}(k)=\boldsymbol{w}(k-1)\\ y(k)=f\left(\boldsymbol{x}(k-1),\boldsymbol{w}(k)\right)+u(k-1)+v(k)\end{cases} \tag{14}$$

where $\boldsymbol{w}(k)$ is n_w-dimensional state vector; state transition matrix is the identity matrix; $f\left(\boldsymbol{x}(k-1),\boldsymbol{w}(k)\right)$ plays the role of a nonlinear measurement on $\boldsymbol{w}(k)$ .These state-space equations allow us to estimate $\boldsymbol{w}(k)$ with second CKF.

1) Time Update

- Evaluate the predicted weight and its error covariance

$$\begin{cases}\hat{\boldsymbol{w}}(k|k-1)=\hat{\boldsymbol{w}}(k-1|k-1)\\ \boldsymbol{P}_w(k|k-1)=\boldsymbol{P}_w(k-1|k-1)\end{cases} \tag{15}$$

2) Measurement Update

- Factorize

$$\boldsymbol{P}_w(k|k-1)=\boldsymbol{S}_w(k|k-1)\boldsymbol{S}_w^T(k|k-1) \tag{16}$$

- Calculate the cubature points($i=1,2,\cdots,m_w$)

$$\boldsymbol{w}_i(k|k-1)=\boldsymbol{S}_w(k|k-1)\boldsymbol{\eta}_i+\hat{\boldsymbol{w}}(k|k-1) \tag{17}$$

where, $m_w=2n_w$; And the parameter $\boldsymbol{\eta}_i$ ($\bar{\boldsymbol{\varepsilon}}_i$ is the n_w-order unit vector) is given by

$$\boldsymbol{\eta}_i = \begin{cases} \sqrt{m_w/2} \cdot \bar{\boldsymbol{\varepsilon}}_i, & i = 1,2,\cdots,n_w \\ -\sqrt{m_w/2} \cdot \bar{\boldsymbol{\varepsilon}}_{i-n_w} & ,i = n_w+1, n_w+2,\cdots,2n_w \end{cases}$$

- Compute the propagated cubature points ($i = 1,2,\cdots,m_w$)

$$y_i^*(k|k-1) = f\left(\boldsymbol{x}(k-1), \boldsymbol{w}_i(k|k-1)\right) \tag{18}$$

- Estimate the predicted measurement

$$\hat{y}(k|k-1) = \frac{1}{m_w}\sum_{i=1}^{m_w} y_i^*(k|k-1) \tag{19}$$

- Compute the innovation covariance matrix

$$\begin{aligned} \boldsymbol{P}_{yy}(k|k-1) = \frac{1}{m_w}\sum_{i=1}^{m_w} y_i^*(k|k-1)[y_i^*(k|k-1)]^T \\ - \hat{y}(k|k-1)\hat{y}^T(k|k-1) + \sigma_u^2 + \sigma_v^2 \end{aligned} \tag{20}$$

- Evaluate the cross-covariance matrix

$$\begin{aligned} \boldsymbol{P}_{wy}(k|k-1) = \frac{1}{m_w}\sum_{i=1}^{m_w} \boldsymbol{w}_i(k|k-1) y_i^*(k|k-1) \\ - \hat{\boldsymbol{w}}(k|k-1)\hat{y}^T(k|k-1) \end{aligned} \tag{21}$$

- Compute the gain matrix of the KF

$$\boldsymbol{K}_w(k) = \boldsymbol{P}_{wy}(k|k-1)\boldsymbol{P}_{yy}^{-1}(k|k-1) \tag{22}$$

- Estimate the updated weight and its error covariance

$$\begin{cases} \hat{\boldsymbol{w}}(k|k) = \hat{\boldsymbol{w}}(k|k-1) + \boldsymbol{K}_w(k)[y(k) - \hat{y}(k|k-1)] \\ \boldsymbol{P}_w(k|k) = \boldsymbol{P}_w(k|k-1) - \boldsymbol{K}_w(k)\boldsymbol{P}_{yy}(k|k-1)\boldsymbol{K}_w^T(k) \end{cases} \tag{23}$$

In a word, at each time step, we use the CKF to estimate both the states $\boldsymbol{x}(k)$ and weights $\boldsymbol{w}(k)$. This structure is called dual cubature Kalman filter (DCKF), shown in Figure 1, run in parallel.

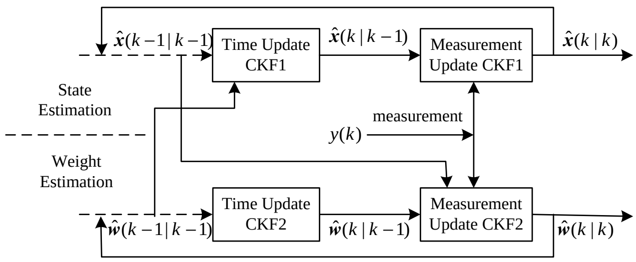

Fig.1. The Dual Cubature Kalman Filter (DCKF)

4. Computer simulation

Application of the DCKF to speech processing involves segmenting the signal into short windows over which the signal is approximately stationary. We use successive 25ms windows of the signal, with a new window starting every 10ms. Hamming window was employed to emphasize the estimates in the center of each window when combining the overlapping segments. For the experiments presented in this paper, BP neural networks of 10 inputs, 4 hidden units, and 1 output were used.

In order to compare the performance between spectral subtraction approach (SS) and proposed algorithm (DCKF). Clean speech “Maybe they will take us.” is obtained from the TIMIT speech database. It was uttered by a female speaker and digitized at 16-kHz sampling frequency with 16-bit quantization. Additive white Gaussian noise from Noisex92 noise database is used to contaminate the speech. The quality of the enhanced speech is evaluated by SNR improvements. The input SNR and output SNR are given by:

$$SNR_{in} = 10\lg \frac{\frac{1}{N}\sum_{k=1}^{N} x^2(k)}{\frac{1}{N}\sum_{k=1}^{N}\left[y(k) - x(k)\right]^2} \tag{24}$$

$$SNR_{out} = 10\lg \frac{\frac{1}{N}\sum_{k=1}^{N} x^2(k)}{\frac{1}{N}\sum_{k=1}^{N}\left[\hat{x}(k) - x(k)\right]^2} \tag{25}$$

where N was the length of speech. SNR_{in} and SNR_{out} represent input SNR and output SNR, respectively. The SNR improvement SNR_{imp} is then $SNR_{in} - SNR_{out}$.

Noise sequences are added to speeches yielding 0 and 5 dB SNR. The simulation results are given by Figure 2 to Figure 4, synchronously, the SNR improvement results of two algorithms are given by Table 1.

TABLE I. SPEECH ENHANCEMENT RESULTS

Algorithm	**SNR (dB)**			
	SNR_{in}	SNR_{imp}	SNR_{in}	SNR_{imp}
SS	0	4.9390	5	6.4115
DCKF	0	9.0784	5	12.1706

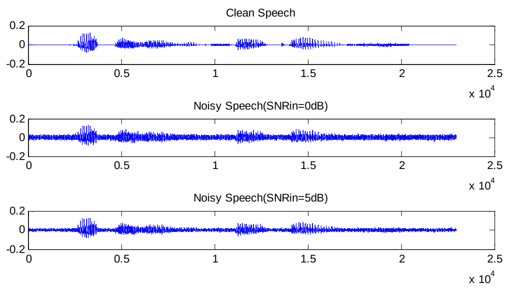

Fig.2. Clean speech and noisy speeches

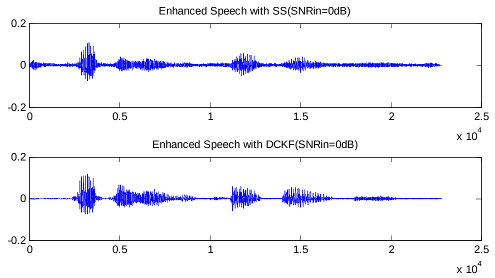

Fig.3. Enhanced speeches of two methods($SNR_{in} = 0$ dB)

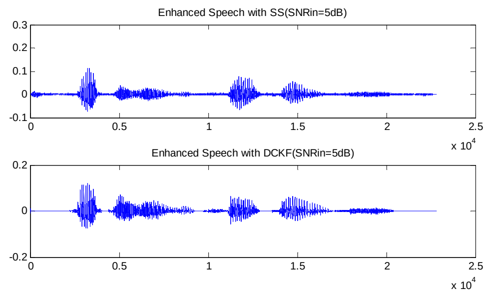

Fig.4. Enhanced speeches of two methods($SNR_{in} = 5$ dB)

From Table 1, Figure 2, Figure 3 and Figure 4, it is easy to obtain the following conclusions:

1) From Figure 3 and Figure 4, it indicates that both methods can remove noises from noisy speeches.

2) From Table 1, Figure 3 and Figure 4, it also shows that the SNR improvement of DCKF is more than SS. Comparing to SS, the DCKF increases by 4.1394dB ($SNR_{in} = 0$ dB) and 5.7591dB ($SNR_{in} = 5$ dB), respectively.

5. Conclusion

The speech enhancement based on dual cubature Kalman filter is proposed in this paper. We use the BP neural network for modeling the nonlinear process of speech signal. Subsequently, we adopt the CKF technology to estimate the speech states and weights in parallel. Finally, the effectiveness of the developed method is demonstrated through some simple experiments.

There are several other interesting directions that deserve further investigation. For example, a possible extension is to design algorithm based colored measurement noises. On the other hand, in implementation of the DCKF, it's assumed that both the variances of $u(k)$ and $v(k)$ are known. In practical application, however, these parameters must be estimated form the noisy speech. Besides, some masking properties of human auditory systems could be applied to speech enhancement in order to keep the speech distortion small. Therefore, further work is needed in these directions. All the conclusions above are conductive to further perfection and development in the field of speech enhancement.

6. Acknowledgment

This material is based upon work supported by the National Science Foundation under Grant No. 60804064.

7. References

[1] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Transactions on Acoustic Speech and Signal Processing, 1979, vol.27, no.2, pp. 113-120.

[2] J. Lim, A. Oppenheim, "All-pole modeling of degraded speech," IEEE Transactions on Acoustics, Speech and Signal Processing, 1978, vol.26, no.3, pp.197-210.

[3] K. Paliwal, A. Basu, "A speech enhancement method based on Kalman filtering," IEEE International Conference on Acoustics, Speech and Signal Processing, Dallas, Texas, USA, 1987, vol.12, 177-180.

[4] P. Maragos, A. G. Dimakis and I. Kokkinos, "Some Advances in Nonlinear Speech Modeling Using Modulations, Fractals, and Chaos,"Proc. 14th International Conference on Digital Signal Processing, DSP2002, Vol. 1, 1-3, July 2002, pp. 325-332.

[5] E. A. Wan, A. T. Nelson, "Neural dual extended Kalman filtering: applications in speech enhancement and monaural blind signal separation," Neural Networks for Signal Processing VU: Proceedings of the 1997 IEEE Workshop, ed. Principe,Morgan, Giles, Wilson, 1997.

[6] Xi Yang, Xiaoping Fan, Xiong Luo, "Unscented-Kalman-filtering based approach for single channel speech enhancement," Proceedings of 2005 Chinese Control and Decision Conference, 2005, pp.702-208.

[7] A. Ienkaran, H. Simon, "Cubature Kalman Filters," IEEE transactions on automatic control, 2009, vol.54, no.6, pp. 1254-1269.