TIME-VARYING NOISE COMPENSATION USING MULTIPLE KALMAN FILTERS

Nam Soo Kim

School of Electrical Engineering, Seoul National Univ.
Kwanak P.O.Box 34, Seoul 151-742, Korea
nkim@plaza.snu.ac.kr

ABSTRACT

The environmental conditions in which a speech recognition system should be operating are usually nonstationary. We present an approach to compensate for the effects of time-varying noise using a bank of Kalman filters. The presented method is based on the interacting multiple model (IMM) technique well-known in the area of multiple target tracking. Moreover, we propose a way to get fixed-interval smoothed estimates for the environmental parameters. The performances of the proposed approaches are evaluated in the continuous digit recognition experiments where not only the slowly evolving noise but also the rapidly varying noise sources are added to simulate the noisy environments.

1. INTRODUCTION

Several approaches to compensate for the time-varying environments have been attempted for robust speech recognition in adverse conditions. Since the background noise characteristics that exist in real world usually show nonstationary nature, these approaches are considered desirable for practical use of the recognition systems.

In [1], the sequential *expectation maximization* (EM) algorithm is used to estimate the environmental parameters in an on-line fashion. The time-varying noise mean vector is tracked by adopting the conventional exponential forgetting scheme. Further improvement on the sequential EM algorithm has been achieved with the application of the interacting multiple model (IMM) method in which both the estimates of the mean and variance can be simultaneously updated for each time [2], [3].

In this paper, we are based on the IMM method where a bank of Kalman filters is used to cope with time-varying noise characteristics. Clean feature vectors are statistically characterized by a mixture of Gaussian distributions, and each mixture component forms a Kalman filter. For mathematical tractability, the speech contamination rule expressed in a nonlinear function of the relevant vectors is linearly approximated for each mixture component. Parameter estimation is proceeded in both the forward and backward directions, and the estimate at a specific time is obtained in a similar way to the fixed-interval smoothing which is generally encountered in Kalman filtering techniques. Through a number of continuous-digit-recognition experiments, we can observe that the proposed method is effective not only in a slowly evolving environment but also in rapidly varying environments where sudden appearance or disappearance of the added noise exists.

2. ENVIRONMENT COMPENSATION BASED ON FUNCTION LINEARIZATION

Let $\mathbf{z} = [z_1, z_2, \cdots, z_N]'$ be a noisy feature vector with dimension N. Assume that $\mathbf{z}$ is related to the clean feature $\mathbf{x} = [x_1, x_2, \cdots, x_N]'$ and the noise $\mathbf{n} = [n_1, n_2, \cdots, n_N]'$ by

$$\mathbf{z} = \mathbf{f}(\mathbf{n}, \mathbf{x}) \tag{1}$$

and that all the vectors $\mathbf{z}$, $\mathbf{x}$ and $\mathbf{n}$ at a time are statistically independent of those at a different time. With environment compensation, we mean that given a noisy feature vector sequence $\mathbf{Z} = \{\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_T\}$, we estimate the clean feature vector sequence $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_T\}$. Here, as is usually adopted in various approaches, the probability density function (PDF) of the clean feature vector is given by a mixture of Gaussian distributions such that

$$p(\mathbf{x}) = \sum_{k=1}^{M} p(k)\mathcal{N}(\mathbf{x}; \mu_k, \Sigma_k) \tag{2}$$

where M is the total number of mixture components and $p(k)$, μ_k and Σ_k represent the given a priori probability, mean and covariance of the kth Gaussian distribution, respectively. As for the distribution of the noise, which is statistically independent of the clean feature, it is assumed to be a single Gaussian distribution $\mathcal{N}(\mathbf{n}; \mu_\mathbf{n}, \Sigma_\mathbf{n})$ where the mean vector$\mu_\mathbf{n}$and the covariance $\Sigma_\mathbf{n}$ are not known and should be estimated during the environment compensation procedure.

It gets usually difficult to estimate directly the environmental parameters such as $\mu_\mathbf{n}$ and $\Sigma_\mathbf{n}$ due to the nonlinearity of the speech contamination rule $\mathbf{f}(\cdot, \cdot)$ in (1). One possible way to alleviate this difficulty is to piecewise linearly

approximate the given nonlinear function. This indicates that in the kth mixture component, $\mathbf{f}(\cdot,\cdot)$ is approximated by

$$\mathbf{z} = A_k\mathbf{x} + B_k\mathbf{n} + C_k \tag{3}$$

where $\{A_k(N\times N), B_k(N\times N), C_k(N\times 1)\}$ are constant matrices. What is remained is to obtain $\{A_k, B_k, C_k\}$ and we apply the statistical linear approximation (SLA) method proposed in [4] for that purpose.

For the kth mixture component, μ_k and the given initial value for $\mu_{\mathbf{n}}$ are used as the center of Taylor series expansion, and we can characterize $\mathbf{f}(\cdot,\cdot)$ by a mixture of linear functions in such a form as shown in (3). This piecewise linear modeling of the speech contamination process enables us to solve the problem of environmental parameter estimation. After the environmental parameter estimation, the environment compensation procedure is completed by computing the clean feature estimate according to the minimum mean square error (MMSE) criterion.

3. IMM-BASED ENVIRONMENT ESTIMATION

As in [2], we assume that the background noise evolves according to the following process.

$$\mathbf{n}_{t+1} = \mathbf{n}_t + \mathbf{w}_t \tag{4}$$

where $\mathbf{w}_t$ is a Gaussian process possessing the following statistical properties.

$$\left.\begin{array}{rcl} E[\mathbf{w}_t] &=& \mathbf{0} \\ E[\mathbf{w}_t\mathbf{w}_t'] &=& \mathcal{Q} \end{array}\right] \quad \text{for } t>0 \tag{5}$$

in which $\mathbf{0}$ represents a zero vector while $\mathcal{Q}$ is a fixed covariance matrix independent of the time, t. Based upon the environment evolution modeling (4) and the linearized feature relationship (3), we can construct a linear state space model for each mixture component. All the mixture components share the same state transition equation, in which $\mathbf{n}_t$ is treated as the state at time t even though they have separate observation models. Given the constructed multiple linear state space models, the environmental parameters, $\lambda = \{\mu_{\mathbf{n}}, \Sigma_{\mathbf{n}}\}$ are sequentially estimated using the IMM method.

In the following, we will briefly summarize the parameter estimation procedure which is divided into three steps [2]. The first step is the *Mixing* (or *Output Generation*) *Step* in which the parameter estimates are obtained by combining the corresponding estimates of all the mixture components. Let $\hat{\mu}_{\mathbf{n}}(t)$ and $\hat{\Sigma}_{\mathbf{n}}(t)$ respectively denote the combined estimates for $\mu_{\mathbf{n}}$ and $\Sigma_{\mathbf{n}}$ at time t given the noisy data sequence $\mathbf{Z}_t = \{\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_t\}$. Then, they are generated by

$$\begin{aligned} \hat{\mu}_{\mathbf{n}}(t) &= \sum_{j=1}^{M} \hat{\mu}_{\mathbf{n}}(t|j)\gamma_j(t) \\ \hat{\Sigma}_{\mathbf{n}}(t) &= \sum_{j=1}^{M} \gamma_j(t)\Big[\hat{\Sigma}_{\mathbf{n}}(t|j) + (\hat{\mu}_{\mathbf{n}}(t|j) - \hat{\mu}_{\mathbf{n}}(t)) \\ &\qquad \cdot(\hat{\mu}_{\mathbf{n}}(t|j) - \hat{\mu}_{\mathbf{n}}(t))'\Big] \end{aligned} \tag{6}$$

in which

$$\begin{aligned} \hat{\mu}_{\mathbf{n}}(t|j) &= E[\mathbf{n}_t|k_t=j,\mathbf{Z}_t] \\ \hat{\Sigma}_{\mathbf{n}}(t|j) &= Cov[\mathbf{n}_t|k_t=j,\mathbf{Z}_t] \\ \gamma_j(t) &= p(k_t=j|\mathbf{Z}_t) \end{aligned} \tag{7}$$

with k_t being the mixture component index at time t.

The next step is the *Kalman Step* in which the conventional Kalman update is carried out based on the parameter estimates computed at the previous time. Let $\mu^p_{\mathbf{n}}(t|j)$ and $\Sigma^p_{\mathbf{n}}(t|j)$ respectively be the mean and covariance of the one-step-ahead predictive state estimate in the jth mixture component at time t. Then, by using the usual time-update approach, we can derive

$$\begin{aligned} \mu^p_{\mathbf{n}}(t|j) &= \hat{\mu}_{\mathbf{n}}(t-1) \\ \Sigma^p_{\mathbf{n}}(t|j) &= \hat{\Sigma}_{\mathbf{n}}(t-1) + \mathcal{Q}\,. \end{aligned} \tag{8}$$

Due to the assumed linear state space model given by (3) and (4), the innovation $\mathbf{e}(t|j)$ in the jth Kalman filter at time t is defined by

$$\mathbf{e}(t|j) = \mathbf{z}_t - A_j\mu_j - B_j\mu^p_{\mathbf{n}}(t|j) - C_j \tag{9}$$

and further its covariance

$$R_{\mathbf{e}}(t|j) = B_j\Sigma^p_{\mathbf{n}}(t|j)B_j' + A_j\Sigma_j A_j'\,. \tag{10}$$

In addition, the Kalman gain, $K_f(t|j)$ is obtained as follows.

$$K_f(t|j) = \Sigma^p_{\mathbf{n}}(t|j)B_j'R^{-1}_{\mathbf{e}}(t|j)\,. \tag{11}$$

With $\mathbf{e}(t|j)$, $R_{\mathbf{e}}(t|j)$ and $K_f(t|j)$, we can compute $\hat{\mu}_{\mathbf{n}}(t|j)$ and $\hat{\Sigma}_{\mathbf{n}}(t|j)$ by the use of conventional measurement-update scheme shown below.

$$\begin{aligned} \hat{\mu}_{\mathbf{n}}(t|j) &= \mu^p_{\mathbf{n}}(t|j) + K_f(t|j)\mathbf{e}(t|j) \\ \hat{\Sigma}_{\mathbf{n}}(t|j) &= \Sigma^p_{\mathbf{n}}(t|j) - K_f(t|j)B_j\Sigma^p_{\mathbf{n}}(t|j)\,. \end{aligned} \tag{12}$$

From various experiments, we have observed that this Kalman filtering approach is not effective in improving recognition performance. Large deviation of the parameter estimates along the time axis has been found responsible for the phenomena. For that reason, we have introduced a modified

approach where the original Kalman gain is shrunk in order to avoid an abrupt change in parameter estimates [2]. The Kalman gain is modified such that

$$K_f^*(t|j) = \alpha K_f(t|j) \tag{13}$$

where K_f^* represents the shrunk Kalman gain and α referred to the shrinking factor is a positive scalar between $(0, 1)$.

After the *Kalman Step*, we conduct the *Probability Calculation Step* in which the posterior probability corresponding to each mixture component is updated. Since the mixture component is assumed to be independent of the previous observations, we have

$$\begin{aligned} \gamma_j(t) &= p(k_t = j|\mathbf{Z}_t) \\ &= p(k_t = j|\mathbf{z}_t, \mathbf{Z}_{t-1}) \\ &= \frac{p(\mathbf{z}_t|k_t = j, \mathbf{Z}_{t-1})p(k_t = j)}{p(\mathbf{z}_t|\mathbf{Z}_{t-1})} \end{aligned} \tag{14}$$

where $p(k_t = j)$ is the prior probability of the jth mixture component and $p(\mathbf{z}_t|\mathbf{Z}_{t-1})$ plays the role of a normalizing term such that the summation of $\gamma_j(t)$ over all j should be 1. Moreover, $p(\mathbf{z}_t|k_t = j, \mathbf{Z}_{t-1})$ represents the one-step-ahead predictive likelihood of the observation within the jth Kalman filter, and can be calculated during the *Kalman Step*.

4. FIXED-INTERVAL SMOOTHING

In this section, a new method to do fixed-interval smoothing under the IMM structure is proposed. As in [1] and [2], the sequential parameter estimation is proceeded in both the forward and backward directions, and the two separate estimates are combined to produce the smoothed estimate at each time. Let us define

$$\begin{aligned} \mu_t^f &= E\left[\mathbf{n}_t|\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_t\right] \\ \Sigma_t^f &= Cov\left[\mathbf{n}_t|\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_t\right] \\ \mu_t^{b,p} &= E\left[\mathbf{n}_t|\mathbf{z}_{t+1}, \mathbf{z}_{t+2}, \cdots, \mathbf{z}_T\right] \\ \Sigma_t^{b,p} &= Cov\left[\mathbf{n}_t|\mathbf{z}_{t+1}, \mathbf{z}_{t+2}, \cdots, \mathbf{z}_T\right] \end{aligned} \tag{15}$$

Then, $\left\{\mu_t^f, \Sigma_t^f\right\}$ can be computed in the forward pass of estimation while $\left\{\mu_t^{b,p}, \Sigma_t^{b,p}\right\}$ can be obtained in the backward pass with a slight modification to the IMM method. Our approach is based on the assumption that the two sets of estimates are derived from a single Kalman filter model [5]. If the smoothed estimates are $\{\mu_t^s, \Sigma_t^s\}$, they are given by

$$\begin{aligned} \mu_t^s &= \Sigma_t^s\left[\left(\Sigma_t^f\right)^{-1}\mu_t^f + \left(\Sigma_t^{b,p}\right)^{-1}\mu_t^{b,p}\right] \\ \Sigma_t^s &= \left[\left(\Sigma_t^f\right)^{-1} + \left(\Sigma_t^{b,p}\right)^{-1}\right]^{-1}. \end{aligned} \tag{16}$$

5. CONTINUOUS DIGIT RECOGNITION EXPERIMENTS

Performances of the IMM method were evaluated with a number of speaker-independent continuous digit recognition experiments. Utterances from 93 speakers constructed the training data and those from the other 47 speakers were used for evaluation. A 19th-order mel-scaled log filterbank energy vector was extracted for each frame of 10 ms with the sampling rate of 8 kHz. By applying discrete cosine transform (DCT), a 12th-order cepstral coefficient vector was derived for each frame. Derived cepstrum vectors and their first-order differences, delta-cepstrum vectors were used for recognition. Each digit was modeled by a five-state semi-continuous hidden Markov model (HMM) where the codebook size was 256 for both the cepstrum and delta-cepstrum.

Environment compensation procedures were carried out in the log spectral domain. For each frame of input signal, the noisy feature vector which was represented by 19 mel-scaled log filterbank energies was transformed to a clean feature estimate. Clean speech features were modeled by a mixture of 128 Gaussian distributions with diagonal covariance matrices. We took the second-order SLA method to approximate the speech contamination rule by the linear model of (3) [4].

Three kinds of typical noise sources were applied to simulate the noisy environments. Two of them are the white and babble noises from the NOISEX-92 database, and the other is a highly nonstationary noise collected from the recording of consecutive impulsive sounds. Noise samples from these three sources were added to the pure speech waveform by varying the signal-to-noise ratio (SNR). From the time-frequency analyses, it was found that the white noise is almost stationary while the characteristic of the babble noise slowly changes. As for the impulsive noise, it was generated when various solid materials were irregularly beaten. For that reason, if this noise is added to a speech waveform, some parts of the data are affected by the existence of impulses while the other parts can remain without any noise.

All the experimental results of the IMM method shown in this paper were obtained with the Kalman gain shrinking factor of 0.3, which could yield the best results. Recognition performances of the proposed approach are shown in Figs. 1, 2 and 3 for the white, babble and impulsive noise, respectively. For all the three types of noise, the IMM method could achieve remarkable improvements in recognition performance compared to those obtained without such a processing of noise. What is noticeable is that the IMM-based approach is effective in compensating not only the stationary noise but also the highly nonstationary noise.

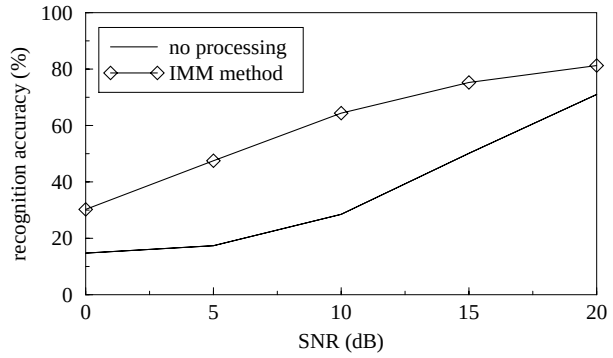

Figure 1: Recognition performance with additive white noise.

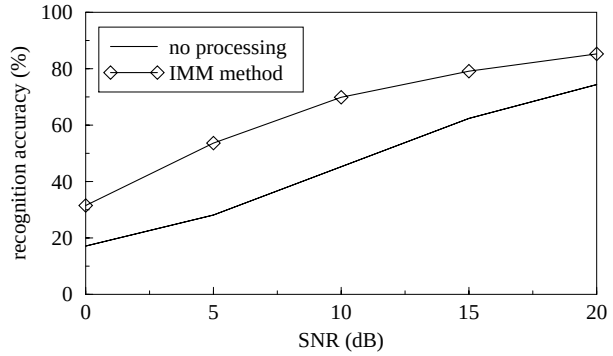

Figure 2: Recognition performance with additive babble noise.

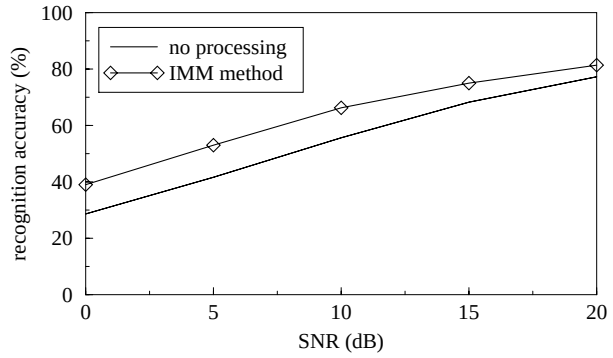

Figure 3: Recognition performance with additive impulsive noise.

6. CONCLUSIONS

In this paper, we applied the IMM approach for the compensation of time-varying noises. The environmental parameters associated with the noise characteristics were estimated in an on-line fashion based on the Kalman filtering scheme. In order to obtain smoothed estimates, the sequential estimation procedure was proceeded both in the forward and backward directions, We could discover the effectiveness of the proposed method with a series of speech recognition experiments under various background environments.

7. REFERENCES

[1] N. S. Kim, "Nonstationary environment compensation based on sequential estimation," *IEEE Signal Processing Letters*, Vol. 5, No. 3, pp. 57-59, March 1998.

[2] N. S. Kim, "IMM-based estimation for slowly evolving environments," *IEEE Signal Processing Letters*, Vol. 5, No. 6, pp. 146-149, March 1998.

[3] A. Averbuch, S. Itzikowitz and T. Kapon, "Radar target tracking-Viterbi versus IMM," *IEEE Trans. Aerospace, Electronic Systems*, Vol. 27, No. 3, pp. 550-563, May 1991.

[4] N. S. Kim, "Statistical linear approximation for environment compensation," *IEEE Signal Processing Letters*, Vol. 5, No. 1, pp. 8-10, Jan. 1998.

[5] R. E. Helmick, W. D. Blair and S. A. Hoffman, "Fixed-interval smoothing for Markovian switching systems," *IEEE Trans. Inform. Theory*, Vol. 41, No. 6, pp. 1845-1855, Nov. 1995.