Hand Gesture Recognition Based on Ultrasound Technology: Pre-processing Stage

1st Raissa Likhonina
Department of Signal Processing
The Czech Academy of Sciences, Institute of Information Theory and Automation
Pod Vodárenskou věží 4, Prague 8, Czech Republic
likhonina@utia.cas.cz

***Abstract*—This paper describes an approach, which can be used as a pre-processing stage for a hand detection and gesture recognition problem. The approach is based on noise cancellation using QRD Recursive Least Squares (RLS) algorithm with double precision arithmetic and exponential forgetting (EF). The paper discusses algorithmic techniques and presents experiments showing how it is possible to calculate the distance between a hand and a device. A series of experiments were performed. During them a time-variant environment regression model and a time-variant hand model as well as different values of the EF factor were used.**

***Keywords*—noise cancellation, QRD RLS algorithm, exponential forgetting, gesture recognition, ultrasound, regression model**

I. Introduction

In [1] different approaches to hand gesture recognition were shortly described. It was stated that nowadays both approaches - wired gloves-based and camera-based technology - have certain limitations and disadvantages and are not widely available on the market [2-4]. In this work a different approach to a hand recognition problem is offered. It is based on ultrasound technology. The ultrasound technology is supposed to solve the problem with costs and power consumption of such kind of devices and make them widely spread in such areas as automotive industry, smart home/buildings, wearables, in applications for disabled people, etc. [1, 5].

The basic concept of a hand recognition device based on ultrasound technology is simple: the system will contain a network of ultrasound transducers and integrated pre-processing unit. The ultrasound impulses will be transmitted by the system, reflected from a hand and return back to the system. On the basis of responses and their characteristics the device will be able to detect the presence, position and distance of the hand. In final application an adaptive beamforming technique may be used for directional signal transmission or reception [1, 5]. Meanwhile, a pre-processing stage is essential before final data processing and gesture recognition. During this stage the accurate data about the hand appearance resp. about the hand distance from the device should been obtained. The final results of gesture detection will greatly depend on the pre-processing stage [1].

It is clear that the main problem is to detect undesired responses from obstacles other than the hand (see Fig. 1).

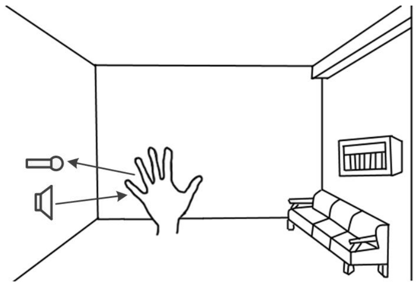

Fig. 1. Example of gesture recognition application [1]

These responses will be considered to be a static response, which makes the process of precise gesture identification difficult and, therefore, should be removed or at least suppressed from the signal going to the next stage of data processing [1, 6].

The present paper describes an attempt to use noise cancellation to eliminate static sound reflections from the environment. Noise cancellation is performed using QRD Recursive Least Squares (RLS) algorithm with exponential forgetting (EF). Apart from the description of removing environment reflections, it also describes the method of detection of the distance between the hand and the device. The experiments described in next sections were implemented and tested in MATLAB simulation environment.

II. Hand Detection Using QRD RLS Algorithm

This section describes algorithmic technique and gives some general information about experiments performed.

A. Algorithmic Technique

According to [7], a noise cancellation method assumes that there are two types of signals:

- the desired signal composed both from the temporary present short distance reflection signal and reflection signal mixed with the environment;
- the reference ultrasound source signal.

It is further assumed that these signals are uncorrelated. Taking it into account the static relation of the environment reflection signal and ultrasound source signal, the temporary present short distance reflection signal can be reconstructed as a prediction error of the adaptive RLS algorithm.

978-1-7281-1740-9/19/$31.00 ©2019 IEEE

The noise cancellation in the present work is based on QRD algorithm with double precision arithmetic and EF. The QRD algorithm is also called an information filter without square root operations and is based on QR decomposition of the input/output information matrix. The recursively updated QRD factorization of the information matrix allows avoiding the loss of positive definiteness of the matrix and provides a numerically stable solution [5]. More detailed information about QRD algorithm as well as about its variants is presented in [5, 7-11]. Here for better comprehension only a brief insight to the RLS algorithm used in the experiments is provided.

It is known that the standard RLS algorithm given in [7-9, 11] has the following form:

$$\overline{\theta}(N+1|n) = \theta(N+1|n) + k\cdot(z(N+1) - Z^T(n)\cdot\theta(N+1|n)) =$$

$$= \theta(N+1|n) + [V(n)]^{-1}\cdot Z(n)\cdot(z(N+1) - Z^T(n)\cdot\theta(N+1|n)) \quad (1)$$

where $\overline{\theta}(N+1|n)$ are estimated parameters, $\theta(N+1|n)$ are known parameters from the previous step, k is a Kalman gain vector, $z(N+1)$ is an output, $Z(n)$ is a data vector, $[V(n)]^{-1} = [U(n)^T U(n)]^{-1}$ is an autocorrelation matrix of the filter input signal, N is the order of the system, n is an order of the system nested in N.

To obtain the stable variants of RLS algorithms, it is necessary to use so-called QR decomposition of the information matrix U [7].

$$U = Q\cdot R \quad (2)$$

where U is a long matrix of size Lxn, where L comprises all measured data, Q is an orthogonal matrix of size Lxn and R is an upper triangular matrix of size nxn with positive diagonal elements. It is valid [7]:

$$U^T\cdot U = R^T\cdot R \quad (3)$$

The compound matrix $U(N+1)$ has the following form [7]:

$$U(N+1)= \begin{bmatrix} Z_1(n) & \dots & \dots & Z_L(n) \\ z_1(N+1) & z_2(N+1) & \dots & z_L(N+1) \end{bmatrix} \quad (4)$$

Its decomposition is fulfilled in the following way [7]:

$$\left[\begin{array}{c|c} U_{L-1}(N+1) & \begin{matrix} Z_L(n) \\ z_L(N+1) \end{matrix} \end{array}\right] = \begin{bmatrix} Q|q \end{bmatrix} \cdot \begin{bmatrix} R & h \\ 0 & \zeta \end{bmatrix} \quad (5)$$

where the left side matrix has size $Lx(n+1)$, $[Q|q]$ is an orthogonal matrix of size $Lx(n+1)$ and the last term is a triangular matrix of size $(n+1)x(n+1)$ [7].

To avoid the product U^TU, the equation for a square-root information RLS (SRI RLS) is used [7]:

$$\begin{bmatrix} \overline{R} & \overline{h} \\ 0 & * \end{bmatrix} \leftarrow \underline{\overline{Q}}^{\,T} \cdot \begin{bmatrix} R & h \\ Z(n)^T & z(N+1) \end{bmatrix} \quad (6)$$

The equation for computing θ takes the following form [7]:

$$\theta = R^{-1}\cdot h \quad (7)$$

The next step of computation will look as follows [7]:

$$\overline{\theta} = \overline{R}^{\,-1}\cdot \overline{h} \quad (8)$$

It is obvious that the algorithm consists of two computational steps per time update: a triangular updating and a triangular back-substitution. It has a computation complexity $O(n^2)$ [7].

In this work, QRD algorithm free from square-root computation is used to prevent a computational bottleneck. It is very important when applying the final solution on small platforms with a small memory footprint.

As for the final application, the QRD version of the Lattice algorithm, which works only with EF, might be used. The Lattice algorithm requires only 18N operations per time update compared with $(3/2)N^2$ operations for QRD RLS algorithm, and, thus, it is more appropriate for systems with large orders. Therefore, in the present work EF is preferred to directional forgetting (DF).

It is also worth saying that one of the benefits of the algorithm used in the work consists in the fact that it allows calculating the error signal without explicitly computing estimates of regression model parameters at each time step [7].

B. Modelling

While performing the experiments, it is assumed that the hand appears for a short time period and the reflections, coming from it, present an additional short period disturbance, which should be detected and transmitted to another processing stage.

The block diagram of the whole process can be seen in Fig. 2.

It is clear from the diagram that the process of noise cancellation consists of three blocks:

- the environment model, which produces undesired noise and serves as a reference signal;
- the hand model, which appears for a short period and produces additional disturbance;
- the identification block, which serves for calculating parameters and errors.

All these models are based on linear finite impulse response (FIR) models [7] and have a common input u. Using the input data, the environment and hand models calculate outputs y_1 and y_2 respectively, which then are summed up. After it output y is sent to the identification block. The identification block estimates parameters of the hand model and calculates prediction and filtration errors [1].

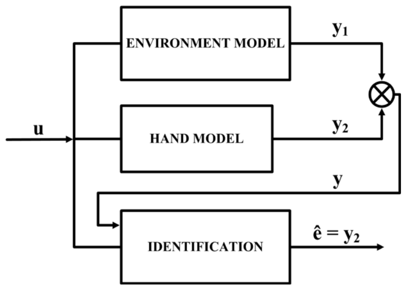

Fig. 2. Block diagram of gesture recognition problem

During simulation, several experiments have been performed to evaluate the identification process. Their results were close to each other; therefore, only two experiments are described in details in this work. In the first experiment the EF factor was set to 0.999, while the second experiment presents the results when the EF factor is close to 1 (0.9999999999). In both experiments the hand appears two times and it is shown how the distance to the hand can be detected. The results of the experiments are described in details in next sections.

III. Results

This section presents the results of modelling and simulation for a hand recognition problem and discusses the outputs of the experiments.

In the experiments a time scale is set to 100 000 steps.

The input signal u is in the form of pulses with the period of 500 samples and the width of a pulse equal to 50 samples. It will allow estimation of the distance between the ultrasound source signal peaks and peaks of the reconstructed near-distance signal, i.e. the distance to the hand. The graph of the input signal can be seen in Fig. 3.

To simulate the static environment model, a regression model of the 1000^{th} order is used. It has slowly time-varying coefficients to represent the situation, when objects in the environment slowly move. The first 500 coefficients are set to zero. This allows the identification process to learn to estimate correct values of parameters. Other parameters are set to slowly changing random values.

The hand model is also a regression model with time-varying coefficients. It has 500 parameters, which are random values.

Both the environment model and the hand model operate with uncorrelated additional output noise to make the situation correspond to the reality.

The absence of a hand is simulated in a way that the coefficients of the hand model are put to zero. The short term appearance of the hand is modelled by setting coefficients of the hand simulation model to nonzero values.

During the experiments, the hand appears two times: firstly at time step 10 000 with delay of 350 samples, then at time step 50 000 with delay of 200 samples. The delay is represented by zero coefficients in the model. Both the first and the second appearance last 500 samples.

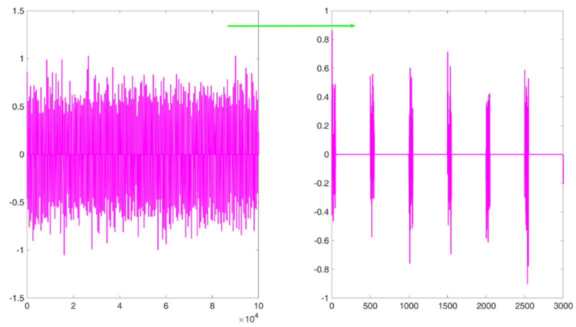

Fig. 3. Input signal

During the experiments the EF factor is set to $\varphi = 0.999$.

The results of identification given $\varphi = 0.999$ are shown in Fig. 4.

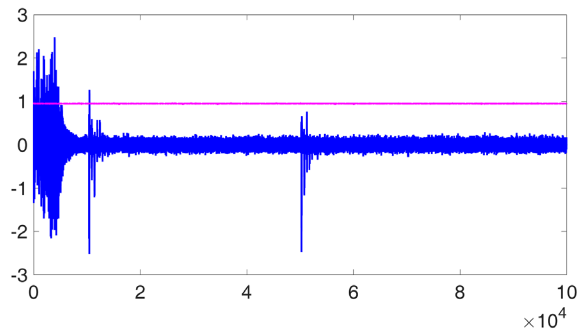

Fig. 4. Identification results: $\varphi = 0.999$

On the graph the input signal is shown in magenta, while the prediction error is depicted in blue. It is clear from Fig. 4 that in the beginning of identification process there are large values of prediction error. It is caused by the fact that the system needs some time to start estimating parameters describing the static environment correctly. Gradually the estimation process converges to correct values.

Other increases of prediction error are explained by the hand appearance. Let us remind that during simulation the hand appears two times. Respectively on the graph there are two increases of prediction error. Thus, it can be concluded that the identification process detects the hand presence correctly.

As it was mentioned above, the main interest lays in prediction error development, which should be identical to the output of the hand model y_2, because it is the hand appearance, which causes the error. If the development of prediction error $\hat{e}$ and output y_2 is compared for this particular case, it is obvious that the difference between the shapes exists (see Fig. 5-7).

Output y_2 is shown in a magenta colour. A blue curve stands for prediction error development. The difference between y_2 and $\hat{e}$ as well as the increase of prediction error in places, where there is no hand, is caused by a small value of the EF factor. In this case the system learns very quickly, so that after a short time period it does not take the presence of the hand as a disturbance anymore. In contrary it takes its disappearance as a new situation and, therefore, the prediction error increases again.

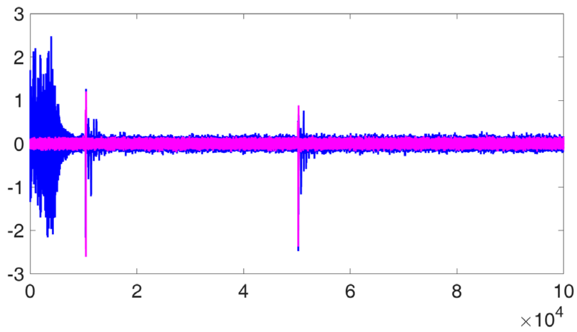

Fig. 5. Comparison of y_2 and $\hat{e}$

The detailed development of prediction error during the first and the second appearance of the hand is shown in Fig. 6 and 7 resp.

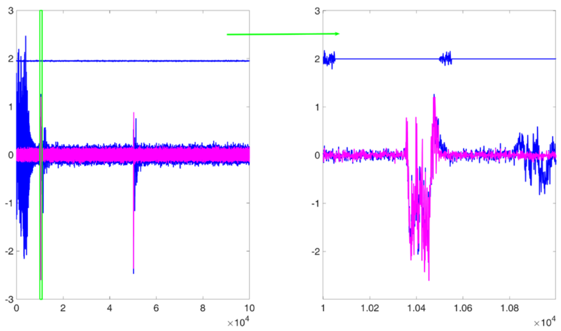

Fig. 6. Comparison of y_2 and $\hat{e}$: first hand appearance in details

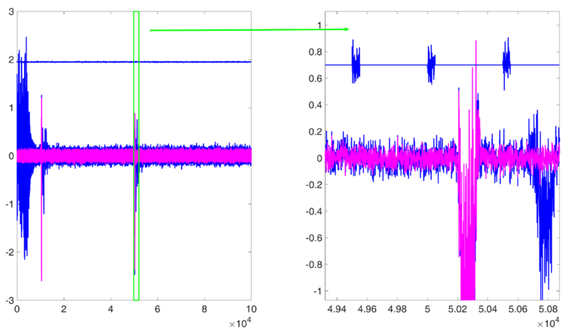

Fig. 7. Comparison of y_2 and $\hat{e}$: second hand appearance in details

The upper blue curve is input signal u, the blue curve in the middle is perdition error $\hat{e}$, and the magenta curve is output y_2.

The first appearance of the hand should be at time step 10 000 samples. On the right graph (Fig. 6) it is clearly shown that the hand appears at time step 10 350 samples. It is caused by delay of 350 samples set in the hand model.

The second appearance of the hand (Fig. 7) should be at time step 50 000, while on the graph it is at time step 50 200. Again it is caused by the delay considered in the hand model.

These delays can be used to determine the distance between the hand and the device. It corresponds to time needed for a signal to come from the hand back to the device and on its basis the distance can be calculated. The experiments show that this concept functions accurately.

To see how a correct choice of EF is important, the EF factor in the further experiments is set to a value close to 1 (see Fig. 8).

The input signal is coloured in magenta, while the prediction error is depicted in blue.

It is shown on the graph that there are two appearances of the hand, which is correct according to the experiment settings. But the first appearance is almost undetectable, because the noise in the beginning is too large. It is caused by a large value of the EF factor, which means that the system learns slowly to identify parameters correctly.

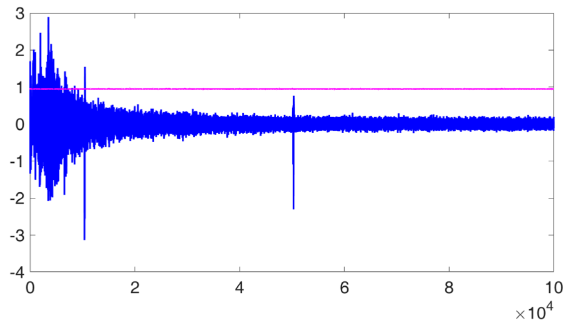

Fig. 8. Identification results: $\varphi = 0.9999999999$

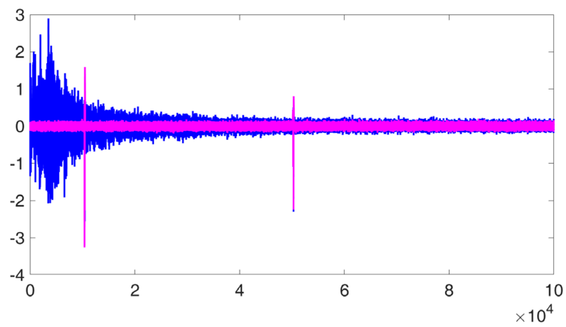

Fig. 9. Comparison of y_2 and $\hat{e}$

If the shapes of the curves of prediction error $\hat{e}$ and output signal y_2 are compared, it is also clear that they start coinciding only after some time needed by the system for learning (see Fig. 9-11).

It should be noted that with EF close to 1 the system learns slower and the hand disappearance is not taken as a new disturbance; therefore, the identification process is more precise and allows recognizing hand appearance more accurately, but only after certain time needed for learning.

This means that the EF factor should be set very carefully. As far as a large value of the EF factor in this experiment causes a lot of noise in the beginning of the identification process, the first hand appearance is hardly identifiable.

The detailed segments with the hand appearances are shown on the next graphs (see Fig. 10-11).

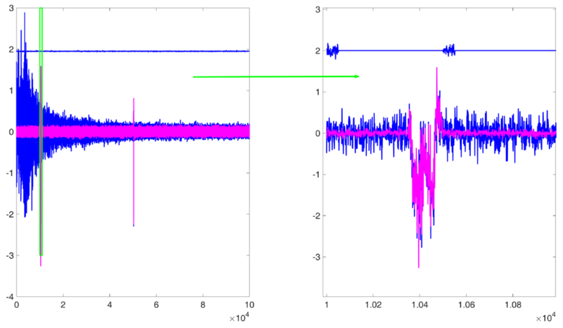

Fig. 10. Comparison of y_2 and $\hat{e}$: first hand appearance in details

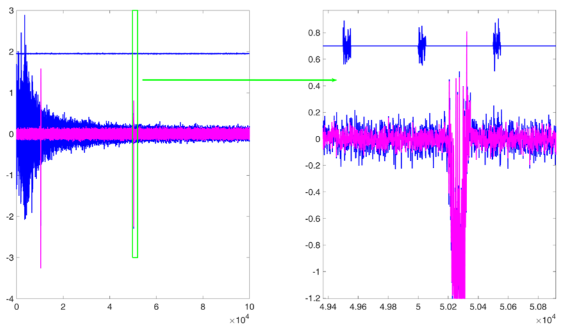

Fig. 11. Comparison of y_2 and $\hat{e}$: second hand appearance in details

It is clear that the signals are delayed for 350 samples in the first case (Fig. 10) and for 200 samples in the second case (Fig. 11). It can be used to calculate the distance to the hand.

It is worth recapitulating that the prediction error is of an essential interest for us, because it will be used as the input data for beamforming and further information processing. Therefore, it is very important to receive the prediction error corresponding to the real situation.

IV. Discussion

The experiments prove that the described approach using noise cancellation with QRD RLS algorithm is promising. Given experimental conditions this approach turns out to be sufficiently precise even for a time-variant environment model.

Also by simulating the delay in the hand model the experiments clearly show that the distance from the hand can be easily calculated as far as the input signal is presented in the form of pulses.

Moreover, it was proved that the EF factor has a great role in the process of parameter identification and prediction error computation. Using the EF factor equal to 0.999, the learning process is fast and, therefore, both appearance and disappearance of the hand is considered to be a disturbance. In its turn it causes the increase of prediction error.

However, if the EF factor is close to 1, the system needs more time to learn to identify parameters correctly. It gives more accurate results as far as the hand disappearance is not taken as disturbance and there is no increase of prediction error. But on the other hand the system needs more time to learn and it results in large values of prediction error in the beginning of the estimation process.

Taking these issues into account, in further work it is supposed to implement hypotheses testing. There will be several identification models with different values of the EF factor. Each model will correspond to a particular situation, e.g. if the environment has slowly moving objects or if it is completely static. On the basis of the hypotheses, the system will be able to choose the most appropriate model and perform more accurate identification process.

V. Conclusion

Hand detection techniques presented in the paper are based on noise cancellation approach. This approach uses QRD RLS algorithm to identify the parameters of the FIR filter and recursively calculate the prediction error.

During simulation several experiments were fulfilled. The experiments include a time-variant FIR filter based-environment model and a time-variant FIR filter based-hand model using two values of a constant EF factor.

The experiments show the importance of a correct choice of the EF factor. This choice depends on a specific situation and should be carefully considered.

It is worth pointing out that the algorithm serves as a pre-processing stage for measurement of distance between the hand and the device and possibly the beam-former based detection of the hand position. The final objective is the detection of hand gestures expressed by a short presence of the hand at a certain distance from the ultrasound source.

Presented simulation results serve as a reference model before the implementation of the recursive QRD or QRD Lattice identification on the embedded Xilinx Zynq device, operating in real time with a microphone and an ultrasound transducer.

Acknowledgment

This work has been supported by the ECSEL JU project SILENSE "(Ultra)Sound Interfaces and Low Energy iNtegrated Sensors" Project No.: ECSEL JU 737487-8 and MSMT 8A17006 (Ministry of Education Youth and Sports of the Czech Republic).

References

[1] R. Likhonina, "Hand gesture recognition: noise cancellation based approach using QRD RLS algorithm", unpublished.

[2] N. Shah and J. Patel, "Gesture recognition technique: a review", International Journal of Recent Trends in Engineering and Research,

ISSN (online) 2455-1457, 2017. Available at https://pdfs.semanticscholar.org/ad2e/8e0975ceb5b4e369a473ad4675a89289de19.pdf

[3] R. Pradipa and S. Kavitha, "Hand gesture recognition – analysis of various techniques, methods and their algorithms", International Conference on Innovations in Engineering and Technology (ICIET' 14). Available at http://www.rroij.com/open-access/hand-gesture-recognition--analysis-ofvarious-techniques-methods-and-theiralgorithms.pdf

[4] P. Premaratne, "Human computer interaction using hand gestures, cognitive science and technology", DOI 10.1007/978-981-4585-69-9_2, Springer Science+Business Media Singapore, 2014.

[5] R. Likhonina, "Fast Bayesian algorithms for FPGA platform: studies for dissertation work", CVUT, 2018.

[6] R. Likhonina and J. Kadlec, "Noise cancellation using QRD RLS algorithms (application note)", unpublished.

[7] M. Moonen, "Introduction to adaptive signal processing", K. U. Leuven, Leuven, Belgium, 1999.

[8] J. Kadlec, "Continuous probabilistic identification of autoregression model with unknown order", in: Analýza, syntéza a rozpoznávání řeči, ČSVTS, Prague, 1985.

[9] J. Kadlec, "Probabilistic identification of regression model in fixed point", Ph.D. thesis, UTIA CAS, Bolivar, Tenn, USA, September 1986.

[10] J. Kadlec and R. Likhonina, "Adaptive RLS algorithms reference implementations (application note)", UTIA, 2016. Available at http://sp.utia.cz/index.php?ids=projects/silense

[11] V. Peterka, "Bayesian approach to system identification", in: Eykhoff, P. (Ed.), Trends and Progress in System Identificaion. Pergamon Press, Oxford, pp. 239-304, 1981.