QRD RLS Algorithm for Hand Gesture Recognition Applications

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

***Abstract* - The paper is focused on an algorithmic technique for detection of hand presence and distance from a hand to device transmitting ultrasound signals. The described method is based on a QRD Recursive Least Squares (RLS) algorithm with double precision arithmetic and exponential forgetting (EF). Modelling of a hand detection problem is based on linear Finite Impulse Response (FIR) based regression models and performed using MATLAB tools. The modelled system comprises an environment model, a hand model and an identification block. A series of experiments testing both time-invariant and time-variant environment models and time-variant hand models show the importance of a correct choice of the EF factor. The experiments have proven the accuracy of the algorithm and the possibility to calculate a distance from the hand to the device. The final version of the algorithm is supposed to be implemented on the embedded Xilinx Zynq device equipped with a microphone and ultrasound transducers.**

***Keywords*—QRD RLS algorithm, exponential forgetting, double precision arithmetic, regression model, hand detection, ultrasound**

I. Introduction

Adaptive Recursive Least Squares (RLS) algorithms [1] are widely used in digital signal processing applications for parameter estimation, echo suppression, beam-forming and etc. One of the first practically applicable algorithms from this group was so-called Levinson-Durbinov recursion [2]. Since then these algorithms and their variants were meticulously researched and many works were devoted to their description, analysis and development as well as their practical applications [1-5].

However, it appears that there is a certain difficulty to implement the algorithms on hardware due to their high computational complexity and problems with numerical stability [1]. To deal with computational complexity, the fast versions of the RLS algorithms were developed [1, 4, 6]. To solve the issue with numerical stability, a so-called QR decomposition of RLS algorithms was proposed [1, 7-9].

Based on the QRD RLS algorithm [1], this work attempts to provide an algorithm applicable for hand presence detection applications using ultrasound technology. The algorithm has to be able to compute a distance between the hand and the device to make the further stages of data processing for gesture recognition easier. It should be noted that there are several hand gesture recognition techniques already available on the market [10-11]. However, these technologies have certain limitations described in [12]. Therefore, there is an attempt to develop a new method for hand recognition, which will be based on ultrasound technology. The final ultrasound-based applications are supposed to be less power consuming, less expensive and more user-friendly. The domain of ultrasound applications includes, but not limited to automobile industry, smart home/building, wearables, etc.

II. Hand Detection Using QRD RLS Algorithm

A. General Principle

The QRD RLS algorithm described in this work serves only as a pre-processing stage for a hand detection problem. Therefore, it aims to preliminary detect the hand, reducing the noise from the environment, and, respectively, to compute the distance between the hand and the device. It will be a part of the more complex algorithm for gesture recognition based on beamforming technique and supposed to be used in final gesture recognition application. The described algorithmic technique continues studies presented in [13-14].

The basic concept of a hand recognition application is as follows: the device sends ultrasound impulses, which are reflected from the hand and return back to the device. Taking into account responses and their characteristics, the device is supposed to detect presence, position and distance of the hand from the device [13].

However, undesired responses can cause a great problem for a recognition process. The undesired responses can come from other objects in the environment and should be removed from the target signal. The algorithm presented in this work aims to eliminate the undesired reflections from the environment and provide more accurate information about the desired signal for further processing stages [13].

The method described in the work benefits from a noise cancellation method [1] shortly described in the next subsection.

The experiments showing a hand detection process and distance computation are fulfilled using MATLAB tools and discussed in details in further parts of the work.

B. Algorithm Description

A noise cancellation technique is well described in [1]. Briefly speaking, there are two signals: the desired signal and the reference ultrasound source signal. The desired signal consists of reflections both from the hand and from the environment. The algorithm has to remove/reduce the reflections from the

978-1-7281-3227-3/19/$31.00 ©2019 IEEE

environment out of the desired signal using the reference ultrasound source signal. These two signals are assumed to be uncorrelated.

In this work the noise cancellation problem is solved using the QRD RLS algorithm with double precision arithmetic and exponential forgetting (EF). The algorithm uses QR decomposition of the information matrix, which provides a numerically stable solution [1, 15-16].

To make a brief insight into the algorithm used in the work, it is worth saying that the standard RLS algorithm [1-3] is described by equation:

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

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

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

The stable version of RLS algorithms is obtained by making QR decomposition of the information matrix U [1-3]:

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

where U is a long matrix of size Lxn, 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. Further it is valid [1-3]:

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

The compound matrix $U(N+1)$ is described as follows [1-3]:

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

Its decomposition is performed as follows [1-3]:

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

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 $(n+1)x(n+1)$ [1-3].

To avoid the product $U^T U$, a square-root information RLS is used [1]:

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

Then computation of θ is fulfilled in the following way [1]:

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

The next step of computation will have the form [1]:

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

The algorithm has two computational steps per time update: a triangular updating and a triangular back-substitution. It results in computation complexity $O(n^2)$ [1-3].

To prevent a computational bottleneck and to avoid problems during hardware implementation of the algorithm, the QRD algorithm free from square-root computation is used [17].

It should be also noted that the algorithm used in the work is able to calculate the error signal without explicitly computing estimates of regression model parameters at each time step [1-3].

C. Modelling

The block diagram of the modelling process is shown in Fig. 1. It consists of three parts: an environment model, a hand model and an identification block.

Furthermore, it is assumed that the hand appears for a short period of time. In this case the reflections from the hand represent an additional short period disturbance.

Both models are based on linear finite impulse (FIR) models. They have common input u. Using input data, the environment and hand models produce outputs y_1 and y_2 respectively. These outputs are summed up and their summation is sent to the identification block, which estimates parameters of the hand model and computes prediction and filtration errors [13].

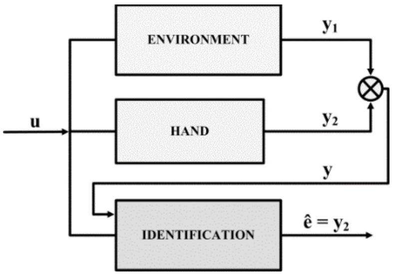

Figure 1. Block diagram

The hand reflection signal is reconstructed as a prediction error of the algorithm. Thus, it can be stated that the development of prediction error $\hat{e}$ estimates the development of output y_2.

III. Results

A series of experiments using both time-invariant and time-variant environment models and different values of the EF factor have been fulfilled. The results of the experiments are thoroughly discussed in this section.

The input signal in the form of pulses (see Fig. 2) is used for all experiments performed. The choice of the input signal is determined by the fact that such kind of a signal allows computing the distance between the hand and the device.

The signal is created from pulses with a period of 500 samples and a width of 50 samples.

To model the environment and the hand, regression models operating with uncorrelated additional output noise are used.

The environment model is a static regression model, i.e. has constant coefficients. The model is of the 1000th order. Its first 500 coefficients are set to zero to allow the system to learn to estimate parameters correctly. The rest are time-invariant random values.

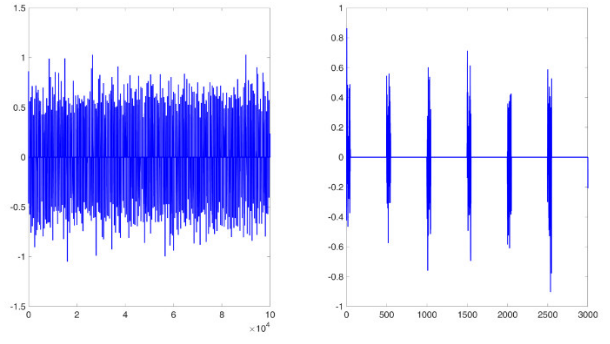

Figure 2. Input signal

The hand model is a time-variant regression model of the 500th order. The absence of the hand is modelled by setting coefficients in the columns of the matrix to zero values. The columns of the matrix correspond to time development of the system.

To simulate a short-term appearance of the hand, the coefficients of the hand model are set to non-zero values.

During the experiments three cases of the hand appearance are considered:

- A short-term appearance of the hand, which does not allow calculating the distance (see Fig. 3). It is the first hand appearance at time step 10 000 lasting for 500 samples.
- A short-term appearance of the hand modelled with a certain delay presented by zero coefficients in the rows of the matrix (see Fig. 3). It is the second hand appearance at time step 50 000 lasting for 500 samples.
- A longer-term appearance of the hand modelled with a certain delay presented by zero coefficients in the rows of the matrix (see Fig. 3). It is the third hand appearance at time step 80 000 lasting for 5 000 samples.

Due to delays modelled in two latter cases, it is possible to compute the distance between the hand and the device. In this case the delay corresponds to time needed for a signal to come back from an obstacle to sensors. The delay for the second hand appearance corresponds to 200 samples; while the delay for the third hand appearance corresponds to 350 samples (see Fig. 4, 6).

The experiments also show the influence of the EF factor on identification process. The results of the experiments with different EF factors are shown in Fig. 3.

The values of the EF factor used in the experiments are 0.995 (the upper graph in Fig. 3), 0.99995 (the middle graph in Fig. 3) and 0.99999995 (the bottom graph in Fig. 3).

The graphs show the development of prediction error (red curve) and output y_2 (green curve). As it was noted above the prediction error should have similar development as output y_2 has. It is due to the fact that it is the hand that causes a short-term disturbance and, thus, the increase of prediction error.

In the very beginning on all graphs there is a certain increase of prediction error. It can be explained by the fact that the system needs some time to learn to identify parameters. After it the system should be able to estimate the parameters and compute prediction errors more or less accurately.

It is also obvious from Fig. 3 that the accuracy of the results greatly depends on the chosen EF factor. On the upper graph the EF factor is small; therefore, the system learns fast. It results in considering both appearance and disappearance of the hand as a short-term disturbance. In its turn it causes the increases of prediction error.

The situation is better on the middle graph, where the EF factor is higher (0.99995). The process of learning and adapting to a new situation is slower. Therefore, the first and the second appearance of the hand are identified very precisely. However, the third appearance of the hand lasts a longer period of time: 5000 samples instead of 500 samples in the first two cases. This time is enough for the system to adapt to a new situation, when the hand is present, and it causes the small increases of prediction error after the hand has already disappeared.

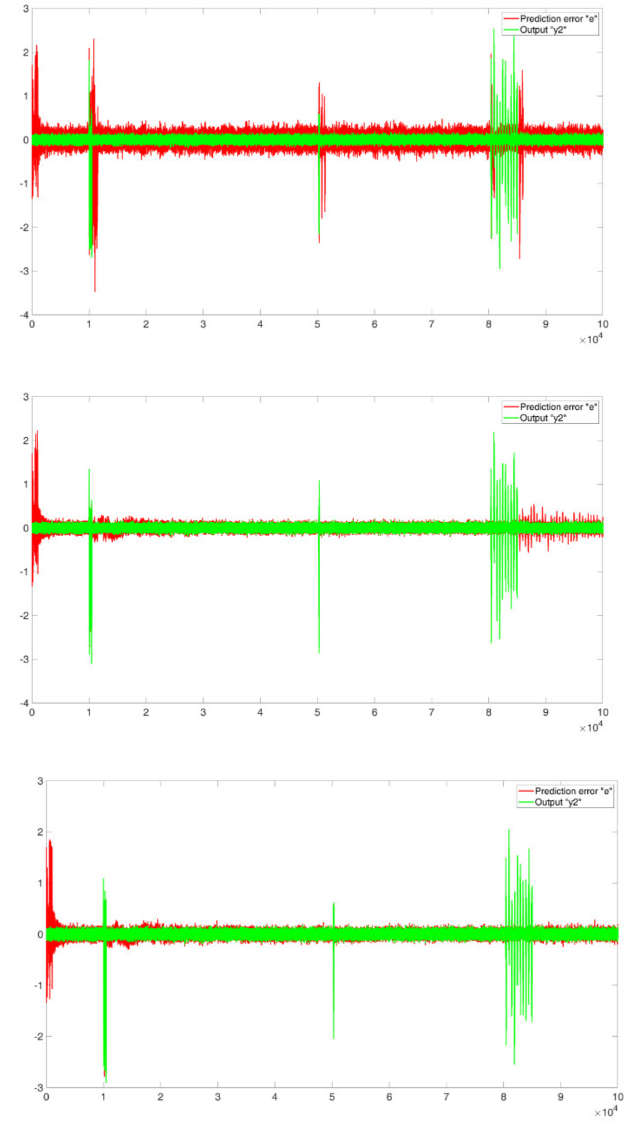

Figure 3. Identification results, time-invariant environment model, different EF factors (from up to down): φ=0.995, φ=0.99995, φ=0.99999995

The low graph in Fig. 3 shows the most precise results as the development of prediction error and output y_2 coincide. In this case the EF factor is very close to 1 (0.99999995). It is obvious that the hand detection process for all three hand appearance is very accurate. Even though in the third case the hand remains a longer period of time, the system is able to identify it correctly.

To show the accuracy of the results in more details, the fragments of the results for the third hand appearance for different EF factors are presented in Fig. 4.

Taking into account the results discussed above, it can seem that the higher value of the EF factor is the better the results can be obtained. However, it is not always true.

Let us suppose that there are some objects in the environment, which can move from time to time. To simulate this situation, a time-invariant environment model should be changed into a time-variant environment model, i.e. its coefficients will be changing with time. All other settings of the experiments remain the same.

Fig. 5 shows the outputs of the experiments given the time-variant environment model using different EF factors: φ=0.995 for the upper graph, φ=0.99995 for the middle graph, φ=0.99999995 for the bottom graph.

It is obvious that the results given the time-variant environment model are not as precise as they were for time-invariant environment model. Note that in the very beginning the system needs more time to learn to estimate parameters. Again the small value of the EF factor results in fast learning. As it was in the previous case, it causes the increase of prediction error both for hand appearance and hand disappearance. However, to make a value of the EF factor larger is not always safe while using the time-variant environment model. From the middle and especially from the bottom graph it is obvious that for greater values of the EF factor the increase of prediction error in the very beginning of identification process is so large that it hardens to detect the first hand appearance.

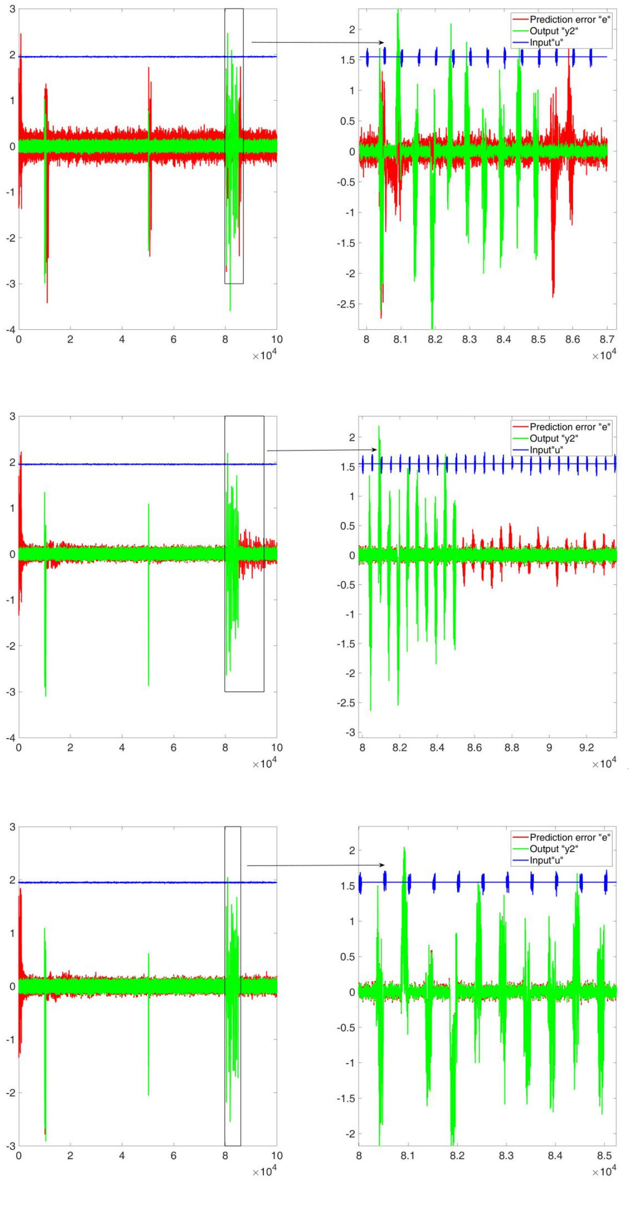

Fig. 4. Detection of the third hand appearance, different EF factors (from up to bottom): φ=0.995, φ=0.99995, φ=0.99999995

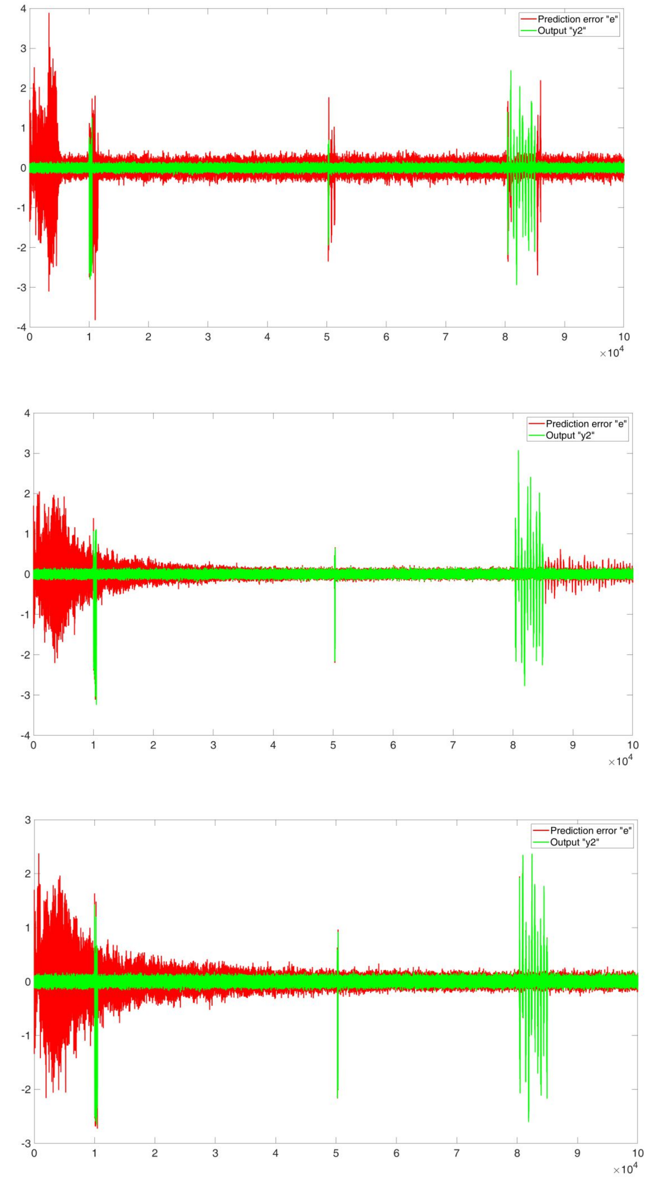

Fig. 5. Identification results, time-variant environment model, different EF factors (from up to down): φ=0.995, φ=0.99995, φ=0.99999995

Finally, Fig. 6 explains the method for computing distance between the hand and the device. For these purposes, the fragment with the second hand appearance given the time-invariant environment model is chosen.

As it has already been observed in previous sections, the second and the third hand appearance are modelled with a certain delay. The delay corresponds to time needed for a signal to return back to the device after being reflected from an obstacle.

The delays are modelled by putting coefficients in the rows of the matrix of the hand model to zero values. Thus, the second hand appearance is delayed for 200 samples: the first 200 coefficients are zeros. The hand appears at time step 50 200 instead of time step 50 000 as it was predefined in settings (see Fig. 6). Similarly, the third hand appearance is delayed for 350 samples and is at time step 80 350 instead of time step 80 000 (see Fig. 3).

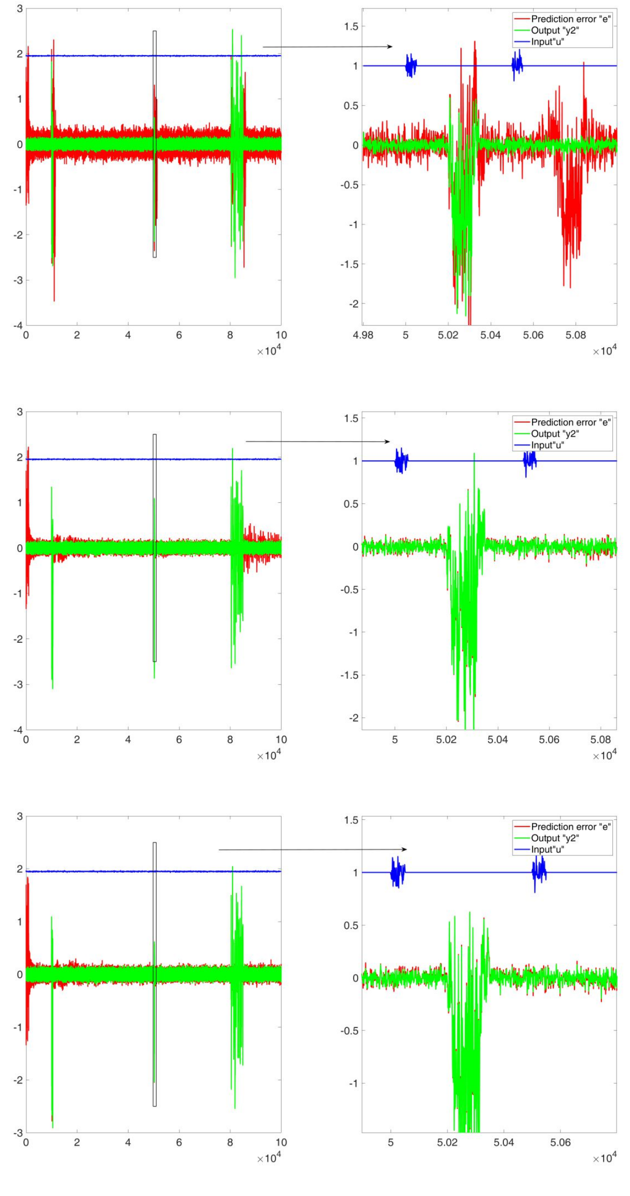

Fig. 6. Example of a signal delay, different EF factors (from up to bottom): φ=0.995, φ=0.99995, φ=0.99999995

The experiments show that this approach to distance computation functions accurately and can be used for further development of the algorithm.

IV. Discussion

The experiments prove that the proposed approach to the system identification using QRD RLS algorithm is promising and provides precise outputs both for time-invariant and time-variant models.

The experiments also show the importance of the EF factor, which strongly influences the results of identification and computation of prediction errors. A small value of the EF factor ensures the fastness of learning process. However, fast learning causes some complications when the hand remains for a longer period of time, so both its appearance and disappearance are considered to be a disturbance. It results in increase of prediction error. On the other hand, large values of the EF factor force the system to learn slower and detection results for a longer hand presence are more precise. However, for time-variant environment models large values of the EF factor caused a great increase of prediction error in the beginning of identification process, which complicated detection of the first hand appearance in the experiments described in the work. It can be concluded that the EF factor should be carefully chosen depending on a particular situation.

Furthermore, the experiments show that the concept of calculating distance using a delay in regression models functions accurately and can be used in further investigations.

The further work is supposed to use the real ultrasound data from the sensors to test the algorithm. Besides, it is supposed to supplement the algorithm with hypotheses testing [18] in a way that there will be different identification models using different EF factor and/or different orders of the model corresponding to this or that situation. The algorithm will detect, which identification model suits better for the data obtained.

V. Conclusion

Summarizing the results, it should be noted that the present work presents an approach to a hand detection technique based on ultrasound technology. The approach describes a pre-processing stage only using a hand detection algorithm. The algorithm used in the work is the QRD RLS algorithm with double precision arithmetic and EF. The algorithm is supposed to be a part of the more complex algorithm based on beamforming technique and supposed to be used in final gesture recognition application. The algorithm aims at detecting the hand presence and, thus, reducing noise from the environment; thus, helping the final gesture recognition algorithm to obtain highly accurate results.

A series of experiments were performed including the experiments with time-variant and time-invariant FIR filter based environment models and time-variant FIR filter based hand models. During the experiments different values of the EF factor were used.

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

During the experiments the method for computation of a distance between the hand and the device was proposed. It may be useful for further processing stages including beam-forming. The final goal is the application for hand detection, where the hand appears for a short period of time at a certain distance from the ultrasound source.

The final version of the algorithm is supposed to be implemented on Xilinx Zynq devices operating in real time with a microphone and ultrasound transducers. The final implementation is supposed to benefit from FPGA structure and pipeling technique to accelerate the computation process.

ACKNOWLEDGMENT

This work has been supported by the ECSEL project FitOptiVis "(From the cloud to the edge – smart IntegraTion and OPtimization Technologies for highly efficient Image and VIdeo processing Systems)" Project No.: ECSEL 783162 and MSMT 8A18013 (Ministry of Education Youth and Sports of the Czech Republic).

REFERENCES

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

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

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

[4] J. Cioffi and T. Kailath, "Fast recursive least squares transversal filters for adaptive filtering," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 32, no. 2, pp. 304 337, 1984.

[5] D. Lee, M. Morf and B. Fridlander, "Recursive least-squares ladder estimation algorithms,"IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 29, no. 3, pp. 627-641, 1981

[6] P. Farbre and C. Guegen, "Improvement of the fast recursive least squares algorithms via normalization: A comparative study," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 34, April 1986.

[7] G. E. Bottomley and S. T. Alexander, "A novel approach for stabilizing recursive least squares filters," IEEE Transactions on Signal Processing, vol. 39, pp. 1770–1779, 1991

[8] D. T. M. Slock and T. Kailath, "Numerically stable fast transversal filters for recursive least squares adaptive filtering," IEEE Transactions on Signal Processing, vol. 39, no. 1, pp. 92 114, 1991

[9] P. A. Regalia, "Numerical stability properties of a QR based fast least squares algorithm," IEEE Transactions on Signal Processing, vol. 41, no. 6, pp. 2096–2109, 1993

[10] N. Shah and J. Patel (2017). Gesture recognition technique: A review. International Journal of Recent Trends in Engineering and Research [Online].Available: https://pdfs.semanticscholar.org/ad2e/8e0975ceb5b4e369a473ad4675a89289de19.pdf

[11] P. Premaratne, "Human Computer Interaction Using Hand Gestures," Cognitive Science and Technology, Springer Science Business Media Singapore 2014.

[12] R. Pradipa and S. Kavitha (2014). "Hand gesture recognition—analysis of various techniques, methods and their algorithms," International Conference on Innovations in Engineering and Technology (ICIET'14) [Online]. Available: http://www.rroij.com/open access/hand gesture recognition analysis ofvarious techniques–methods and theiralgorithms.pdf

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

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

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

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

[17] S. F. Hsieh, K. J. R. Liu, and K. Yao, "A unified approach for QRD based recursive leastsquares estimation without square roots," IEEE Transactions on Signal Processing, vol. 41, no. 3, pp. 1405 1409, 1993.

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