Muscles data compression in body sensor network using the principal component analysis in wavelet domain

Elmira Yekani Khoei¹, Reza Hassannejad²*, Behzad Mozaffari Tazehkand³

¹ Faculty of Computer, College of Engineering, East Azerbaijan Science and Research Branch, Islamic Azad University, Tabriz, Iran
² Faculty of Mechanical Engineering, University of Tabriz, Tabriz, Iran
³ Faculty of Electrical and Computer Engineering, University of Tabriz, Tabriz, Iran

Abstract

Introduction: Body sensor network is a key technology that is used for supervising the physiological information from a long distance that enables physicians to predict and diagnose effectively the different conditions. These networks include small sensors with the ability of sensing where there are some limitations in calculating and energy.

Methods: In the present research, a new compression method based on the analysis of principal components and wavelet transform is used to increase the coherence. In the present method, the first analysis of the main principles is to find the principal components of the data in order to increase the coherence for increasing the similarity between the data and compression rate. Then, according to the ability of wavelet transform, data are decomposed to different scales. In restoration process of data only special parts are restored and some parts of the data that include noise are omitted. By noise omission, the quality of the sent data increases and good compression could be obtained.

Results: Pilates practices were executed among twelve patients with various dysfunctions. The results showed 0.7210, 0.8898, 0.6548, 0.6009, 0.7435, 0.7651, 0.7623, 0.7736, 0.8596, 0.8856 and 0.7102 compression ratios in proposed method and 0.8256, 0.9315, 0.9340, 0.9509, 0.8998, 0.9556, 0.9732, 0.9580, 0.8046, 0.9448, 0.9573 and 0.9440 compression ratios in previous method (Tseng algorithm).

Conclusion: Comparing compression rates and prediction errors with the available results show the exactness of the proposed method.
Compressed data after compression with multiplex.

Materials and methods

Tseng algorithm

Tseng et al.11 represented a method for data compression using temporal and spatial correlation in 2011. In this research, a network of body wireless sensor network with n sensor nodes has been considered in which every sensor node is equipped with m axis. Considering that in the present research, tri-axial accelerometers have been used, m is regarded 3. The mentioned algorithm has two offline and online phases. In offline phase, received data from the sensors in body sensor network are collected and sent for processing to a fusion center analyzer. The main aim of the present phase is to achieve suitable order of the sensor node prediction. Block diagram of Fig. 1 shows the details of offline phase.

Online phase: the main aim of this phase is data compression according to produced compression tree in offline stage. There are two cases for each sensor. 1. \( V_i \) node which is the first level sensor node. 2. \( V_j \) node which is not first level sensor node. The stages shown in (Fig. 2) and (Fig. 3) are performed for case 1 and 2. Since physiological signals studied in physical engineering
Data compression

The amounts of $\epsilon(i,q)|\emptyset[t]$ errors produced in former block, for all amounts of $q=1,2,..,m$ are coded by Huffman coding encoder.

Data collection from three dimensions of the all sensors and saving in column vectors $R^{(1)}_{i}[t] \leftarrow$ th sample from axis i sensor i

Changing each column vector $R^{(1)}_{i}$ to $\Delta X^{(1)}_{i}$ vector

$x^{(1)}_{i}[t] \leftarrow \Delta x^{(1)}_{i}[t] \quad$ for $i=1,2,...,m$

Fir filter based on Kaiser Window is used for the mentioned aim

In order to use temporal correlation, each $\Delta x^{(1)}_{i}$ is change to $\Delta X^{(1)}_{i}$ a differential column vector.

$\Delta X^{(1)}_{i}[t] = \Delta x^{(1)}_{i}[t] - \Delta x^{(1)}_{i}[t-1]$

Restoration of available sensor node data on the path

$\Delta x^{(1)}_{i+1} = \alpha(i,q)|1 + \sum_{m=1}^M \beta(i,q)|(i,l) \Delta X^{(1)}_{i}$

$\Delta X^{(1)}_{i} = \alpha(k+1,q)|k + \sum_{m=1}^M \beta(k+1,q)|(k,l) \Delta X^{(1)}_{i}$

Fig. 2. Online phase of Tseng algorithm: $V_i$ is a level-1 node.

are often weak, they are always exposed to noise and interference. In this paper, in order to increase compression rate, two methods of signal processing as principal component analysis and wavelet transform have been used. Principal component analysis is used for separating blind signals and increasing correlation among the received data from sensors. In addition, wavelet transform is also used to reduce noise and interference, increasing the compression rate and reducing the error between the real and predicted amounts. As a result, prediction error between the real amount and prediction amount decreases and compression rate increases in comparison with Tseng method.

Proposed algorithm

Proposed algorithm diagram block is depicted in Fig. 4 and Fig. 5. As it is shown in Fig. 4 and Fig. 5, after data collection, principal component analysis is performed on the data. Then, the obtained result of this stage is sent to online and offline stages with applying wavelet transform (as Figs 1, 2, and 3).

Principal components analysis

In principal component analysis stage shown in Fig. 4 and 5, the raw data that have been collected and processed in data collection stage, would be saved in $R^{(1)}_{i}$ column vectors for each $i=1,2,..,n$. In this analysis, processed data should be normalized and their average should be zero. Thus, according to equation 1 the average of all numbers of column vector $R^{(1)}_{i}$ would be calculated for each $i=1,2,...,n$

Fig. 3. Online phase of Tseng algorithm: $V_i$ is non-level-1 node.

Data collection from three dimensions of the all sensors and saving in column vectors $R^{(1)}_{i}[t] \leftarrow$ th sample from axis i sensor i

Changing each column vector $R^{(1)}_{i}$ to $\Delta x^{(1)}_{i}$ vector

$x^{(1)}_{i}[t] \leftarrow \Delta x^{(1)}_{i}[t] \quad$ for $i=1,2,...,m$

Fir filter based on Kaiser Window is used for the mentioned aim

In order to use the temporal correlation, each $\Delta x^{(1)}_{i}$ is change to $\Delta X^{(1)}_{i}$ a differential column vector.

$\Delta X^{(1)}_{i}[t] = x^{(1)}_{i}[t] - x^{(1)}_{i}[t-1]$

Data transference without predicting and only through intranode correlation

$\Delta x^{(1)}_{i} = \alpha(i,q)|1 + \sum_{m=1}^M \beta(i,q)|(i,l) \Delta X^{(1)}_{i}$

$e(i,q)|\emptyset[t] = \Delta x^{(1)}_{i} - \alpha(i,q)|1 + \sum_{m=1}^M \beta(i,q)|(i,l) \Delta X^{(1)}_{i}$

The amounts of $e(i,q)|\emptyset$ errors produced in former block, for all amounts of $q=1,2,...,m$ are coded by Huffman coding encoder.
Where, $\delta_{ij}$ is unit impulse response.

According to equation IV, the matrix $E$ of eigenvectors are multiplied to $R^{(q)}$ and the data is saved in $PX^{(q)}$:

$$PX^{(q)} = E \ast R^{(q)}$$  \hspace{1cm} Eq. (IV)

The theory of principal component analysis increases the correlation between the sensors. Therefore, it causes an increase in compression rate. According to equation IV, eigenvector ($e_i$) forms a space with $m$ dimensions. Therefore, a second space is produced that is dimensionally smaller. Then, matrix of $E$ is defined as follows. The principal component analysis is defined as equation V:

$$E_s = [e_1, e_2, \ldots, e_m] \in \mathbb{R}^{m \times n}$$  \hspace{1cm} Eq. (V)

Wavelet transform

Fourier transform is extracting oscillatory components of frequency signals in time domain. The aim of wavelet transform is calculating components of signal that there are locally, like sudden changes of a noisy signal in a limited time frame. Wavelet transform could be stated as a number of mother wavelet transform functions. Using acts of contraction, expansion and translation in time domain, a family of wavelets based on mother wavelet could be created. Wavelet transform of one signal such as $x(t)$ could be written as equation VI. In the present equation, function $\psi(t)$ is the mother wavelet and plays the role of $e^{jwt}$ function in Fourier transform:

$$X(t) = \sum_{k=-\infty}^{\infty} C_k \phi(t-k) + \sum_{k=0}^{\infty} \sum_{m=0}^{\infty} d_{m,k} \psi_{j,k}(t)$$  \hspace{1cm} Eq. (VII)

In order to reduce the complexity, the mentioned transform is calculated in discrete form for each amount of $a$ and $b$. Discrete wavelet transform is written according to equation VII:

$$T_x(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \psi^* \left(\frac{t-b}{a}\right) dt$$  \hspace{1cm} Eq. (VI)

In equation II, matrix $D$ is a diagonal matrix and $\lambda_i$ is the eigenvalue of covariance matrix $C_{xx}$ and $e_i$ is the eigenvector corresponding to eigenvalue $\lambda_i$. Columns of matrix $E$ is formed by eigenvectors. $E$ is an orthogonal matrix. $k$ Vectors are orthogonal, thus:

$$e_i^T e_j = \delta_{ij}$$  \hspace{1cm} Eq. (III)

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$  \hspace{1cm} Eq. (III)
Data compression

BioImpacts, 2015, 5(1), 9-16

Fig. 7. A sample of decomposition of the collected data: According to discrete wavelet transform: (A) displays the collected raw data; (B) displays the low-pass part of the decomposed signal; and (C) displays the high-pass part of the signal.

Results

Simulation process

A wireless body sensor network of two nodes, in which, each node is equipped with three axes, is considered. Twelve different test conditions for patients with low back pain (LBP, test.a), weak muscles (test.b), calf strain (test.c), chondromalacia of patella (test.d), disc herniation (test.e), spinal canal stenosis (test.f), DJD of spine (test.g), hamstring shortness (test.h), medial meniscal tearing (test.i), DJD of knee (test.j), DJD of hip (test.k) and anterior collateral ligament tearing (test.m) are considered. As an example, Supplementary 1 shows seven models of these tests. Both of the sensor nodes are installed on the body of the patients in order to receive and monitor the movements. Each sensor is equipped with a LPC1768 Cortex-M3 microprocessor, one RFM70 module and an ADXL335 tri-axial accelerometer. In order to make a comparison between Tseng algorithm and proposed method, the data are collected from sensors. Then, collected data is processed in base station and two considered compression methods are applied on the processed data in the base station. Every test is carried out on the considered patient and the saved results in the nodes are transferred to base station.
station through a trustable mechanism. Some parts of the processed data are considered as a set of learning sequence and remaining part is considered as test data. For every test case, two methods of data compression are compared:

1. Data compression method using the temporal-spatial correlation (Tseng algorithm\(^1\)): in this method only temporal-spatial correlation is used to compress received data.
2. Compression method using wavelet transform, principal component analysis and temporal-spatial correlation: in the proposed method, in addition to temporal-spatial correlation, wavelet transform and principal component analysis are also used to increase the compression rate and reduce the error between the real amount and the prediction amount.

**Discussion**

**Evaluation of results**

For each compression method, a compression model would be created using the learning sequence in offline phase and then online compression rate would be measured using the data test set. For each method, the obtained results from offline phase include transference order tree, \(\alpha_s\) and \(\beta_s\) correlation coefficient and Huffman code word for each sensor. Fig. 8 shows a sample of prediction result. Where the black signal represents the raw collected data and the signals with dotted line displays the predicted data. As, it is shown in the figure, the presented prediction method, creates a prediction data which is much closer to the raw data with a low error. This has leaded the reduction of prediction error; therefore the proposed method increases the compression rate and optimized bandwidth usage.  

Fig. 9 shows the compression rate for both methods (proposed and Tseng methods). Compression rate is defined as the ratio of the size of compressed data to that of the non-compressed data, and it is used as main criterion of the evaluation. In fact, size of the compressed data is always smaller than size of the non-compressed data. Therefore, by decreasing compression rate, data would be more compressed. In fact, non-compressed data is the same with obtained data in diff-coding stage. Compressed data is amount of the obtained error in online stage. As it was mentioned in the online section, the stages of obtaining prediction error depend on the level of considering node. In Fig. 9, it is observed that compression rate in the proposed method (gray column) has been increased and it has led to optimum use of the bandwidth and reduction of the conflict between the sent data by the sensors.  

Fig. 10 shows the examples of error between the real and predicted amounts. The error between the real and predicted amount is achieved through the subtraction of real data from the predicted one. In order to increase the correlation among the received data from the nodes and reduce the noise, the theory of principal component analysis and wavelet transform is used. As it was observed in Fig. 10, error in the proposed method in each axis node is close to zero and shows more reduction because of using the analysis of principal components and wavelet transform (part B, Fig. 10) relative to Tseng algorithm (part A, Fig. 10).  

Since produced error is a datum that should be sent to the base station instead of raw data, reducing the error improves the bandwidth that is considered as the main aim of the present research. Fig. 11 shows the restored data that is calculated from the sum of predicted amount
Fig. 10. The examples of error between the real and predicted amounts in testes g, h, i, j, k and m in both compression methods: part (A) displays the error in Tseng method and part (B) displays the error in proposed method.

Fig. 11. An example of the restored data that displays a good coincidence between the raw data and predicted one.

with the produced error according to equation IX.

\[ \Delta X_i^{(j)} = \alpha(i, j) 1 + \sum_{p=1}^{m} \beta(i, j)(k, p)\Delta X_i^{(p)} \]

\[ + \sum_{q \neq i} \beta(i, j)(j, l)\Delta X_i^{(q)} + \varepsilon(i, j) 1 \]

Eq. (IX)

In this figure signals are displayed with black colour which shows the principle data, and signals displayed with dotted line shows restored data obtained by the proposed method. From Fig. 11 it could be clearly observed that the restored data coincides with raw data. It demonstrates the correctness of data prediction method. Vertical lines in Figs. 8, 10, and 11 separate offline and online stages. The signal in the right side of the dotted line shows the produced signal in offline stage and right side represents the produced signal in online stage.

Conclusion

A compression method based on the theory of principal component analysis, wavelet transform and temporal-spatial correlation is presented. The theory of principal component analysis has been used to increase compression rate and the correlation among the received data from different nodes. Wavelet transform is also used for better reduction of the noise and interferences in raw data received from sensors. The proposed method reduces: 1- the data volume sent by the sensors to the base station 2- interference and conflict among the sent...
data and 3- error between the real amount and predicted amount in comparison with the method available in the literature. Furthermore, it leads to the optimum use of wireless channel’s band width. Since transference of one bit data is equal to executing 1000 code in CPU order, thus this reduction in volume of sent data through data compression causes the optimum use of energy in sensor batteries and increase battery life. The importance of this problem could be observed clearly where body wireless sensor network especially planted ones, are needed to be used for a long time like diabetic patients. The amount of compression rate in tests a to m for 30000 data using Tseng method is 0.8256, 0.9315, 0.9340, 0.9509, 0.8998, 0.9556, 0.9732, 0.9580, 0.8046, 0.9448, 0.9573, 0.9440 and for proposed method is 0.7210, 0.8898, 0.6548, 0.6765, 0.6009, 0.7435, 0.7651, 0.7623, 0.7736, 0.8596, 0.8856 and 0.7102 which shows an increase in compression rate in proportion to Tseng method (see Fig. 9). The results display the usefulness of proposed algorithm in comparison to Tseng algorithm. The proposed method makes it possible to receive correct data by physicians and physiotherapists from the body sensor networks for better treatment and diagnosis of illness.

Acknowledgments
We thank Abdolvahab Moghadamm, specialist in physical medicine and rehabilitation, because of patient referral for experimental exercises, and vibration and modal analysis, laboratory of Mechanical Engineering Department, University of Tabriz for their sincere collaboration.

Ethical issues
There is none to be declared.

Competing interests
The authors declare no conflict of interests.

Supplementary materials
Supplementary file contains the supplementary 1.

References