LSTM FOR HUMAN ACTIVITY RECOGNITION BASED ON FEATURE EXTRACTION METHOD USING CONFORMAL GEOMETRIC ALGEBRA

Nguyen Nang Hung Van¹, Pham Minh Tuan², Do Phuc Hao³, Kanta Tachibana⁴

¹,² The University of Da Nang - University of Science and Technology.
³Danang Architecture University, 566 Nui Thanh St., Da Nang 550000, Vietnam.
⁴Kogakuin University, 1-24-2 Nishi-Shinjuku, Shinjuku-ku, Tokyo 163-8677, Japan.
nguyenvan@dut.udn.vn, pmtuan@dut.udn.vn, haodp@dau.edu.vn, kanta@cc.kogakuin.ac.jp

ABSTRACT: Deep Learning (DL) is a new research trend in recent years for many applications, such as image processing, object detection, and remote control. DL has two main models: Convolutional Neural Network (CNN) used to feature extraction in image processing, and Recurrent Neural Network (RNN) used to handle sequence identification (sequence/time-series). However, the architecture of RNN is quite simple and the ability to remember information from long-distance data is not good. Therefore, the first information in the input sequence usually does not have much influence on the output sequence prediction results of the following steps. So the Long Short Term Memory (LSTM) is designed to overcome this problem. Furthermore, training data with high numbers can lead to memory overflow or low classification accuracy. To solve this problem, Some commonly used machine learning models, such as Principal Components Analysis (PCA), Principal Components Regression (PCR), and Linear Discriminant Analysis (LDA), were proposed to reduce dimensions of the data and computational complexity simultaneously for training models. However, these methods, which use linear transformations, should work well only when the data is distributed on a plane or hyper-plane. Therefore, in this paper, we propose to use Conformal Geometric Algebra (CGA) to feature extraction and reduce dimensions of the data. First, the action data is preprocessed to normalize the data. Next, use CGA to reduce dimensions of data and create feature vectors. Finally, use the LSTM for training and prediction. The experiment was conducted on the CMU dataset with 8 different actions and the results showed that the proposed method has higher results than the previous methods.

Keywords: Conformal Geometric Algebra, Deep Learning, Human Activity Recognition, Long Short Term Memory.

I. INTRODUCTION

Today, Human activity recognition is one of the most important areas of computer vision research. Its applications include intelligent security monitoring systems, health care systems, smart transportation systems, and a variety of systems that involve interactions between people and electronic devices such as human-computer interfaces. Action data generated from systems are getting increasingly large and complicated. Big data can lead to problems in machine learning, such as overfitting and degradation of accuracy. To address these issues, some methods like Principal Components Analysis (PCA) [1, 2], Linear Discriminant Analysis (LDA) [3], and Principal Component Regression (PCR) [4] were proposed to reduce dimensions of the data and computational complexity simultaneously. However, these algorithms use linear transformations to represent the data and assume the data is distributed on a plane. In case data is distributed on the sphere or hyper-sphere form of objects that move and rotate in space, then results from data processing are not accurate. Therefore, this paper proposes a feature extraction method using CGA combined with LSTM to Human activity recognition [5, 6]. CGA is extended from real m dimensional space by adding two base vectors and using transformations to convert vectors in real space into a set of points in CGA space. Complex data distributions are optimized by the hyper-plane or hyper-sphere data approximation method. A vector in CGA space is represented as a point, hyper-plane, hyper-sphere.

Deep Learning (DL) [7, 8, 9, 10] is a new research trend in recent years for many applications, such as image processing, object detection, and remote control. DL has two main models: Convolutional Neural Network (CNN) [11] used to feature extraction in image processing, and Recurrent Neural Network (RNN) used to handle sequence identification (sequence/time-series) recognition and classification tasks [10]. However, from the limitations of the RNN model is a recurrent connection causes the input’s influence to either decay or blow up exponentially, which is referred to as the vanishing gradient problem [12, 13]. LSTM [10, 14] is an RNN architecture that addresses the vanishing gradient problem. The LSTM hidden layer is composed of memory blocks, which are self-connected subnetworks containing multiple internal cells. Through multiplicative gates, the cell is capable of storing and accessing information over a length of time.

The rest of the paper is organized as follows. In Section 2, we first introduce related works about CGA and LSTM. Section 3 will present the proposed method. The empirical results of the proposed methods with the CMU dataset [15] are shown in Section 4. Finally, Section 5 concludes the paper.

II. RELATED WORKS

The human activity recognition model using DL is of interest to many researchers around the world. The basic training model is built from the following steps. The first is to collect data via sensors or images directly from the camera [16]. Next is pre-processing and using many machine learning methods to extract features of the object. Finally, using the training models to recognize activities. In this section, the paper will present an overview of CGA and LSTM in recognition models.
A. Conformal Geometric Algebra

Conformal Geometric Algebra [17, 18, 19, 20, 21] is a part of Geometric Algebra [17] and is also called Clifford Algebra. GA defines the signature \( p + q \) orthonormal basis vector \( O = \{e_1, \ldots, e_p, e_{p+1}, \ldots, e_{p+q}\} \), such as \( e_i^2 = 1, \forall i \in \{1, \ldots, p\} \) and \( e_i^2 = -1, \forall i \in \{p+1, \ldots, q\} \). GA defines \( O \) by \( G_{p,q} \). For example, \( m \)-dimensional Euclidean vector space \( \mathbb{R}^m \) is denoted by \( G_{m,0} \).

A CGA space is extended from the real Euclidean vector space \( \mathbb{R}^m \) by adding 2 orthonormal basis vectors. Thus, a CGA space is defined by \( m + 2 \) basis vectors \( O = \{e_1, \ldots, e_m, e_+, e_-\} \), where \( e_+ \) and \( e_- \) are defined as follows:

\[
e_+^2 = e_+ \cdot e_+ = 1, \\
e_-^2 = e_- \cdot e_- = -1, \\
e_+ \cdot e_- = e_+ \cdot e_i = e_- \cdot e_i = 0, \forall i \in \{1, \ldots, m\}.
\]

Thus, a CGA can be expressed by \( G_{m+1,1} \). In addition, CGA defined:

\[
e_0 = \frac{1}{2}(e_- - e_+), \quad e_\omega = (e_- + e_+).
\]

Given training set \( X = \{x_i | x_i \in \mathbb{R}^m\} \), \( i \in \{1 \ldots n\} \) represented in real \( m \)-dimensional space. This training set is re-represented by the set of points \( P \in G_{m+1,1} \) in CGA space [16, 17] as follows,

\[
P_i = x_i + \frac{1}{2} \|x_i\|^2 e_\omega + e_0 \in G_{m+1,1}
\]

A conformal vector \( S \) is generally written in the following:

\[
S = s + s_\omega e_\omega + s_0 e_0
\]

The process of estimating using least squares \( d^2(P_i, S) \). The error function \( E \) is as follows:

\[
E = \sum_{i=1}^{n} (x_i S - s_\omega - \frac{1}{2} \|x_i\|^2 s_0)^2
\]

This means that when minimizing the error \( E \) function, \( s_\omega \) can be limited by \( \|s\|^2 = 1 \). In this case, the optimization problem becomes as follows:

\[
\min \sum_{i=1}^{n} (x_i S - s_\omega - \frac{1}{2} \|x_i\|^2 s_0)^2
\]

with the condition is:

\[
\|s\|^2 = 1
\]

The optimal result can be solved by the output of Pham [22]. The decomposition of the Eigen solves the optimal problem.

\[
As = \lambda s
\]

where \( A \) is the variance matrix of the \( i^{th} \) training set in CGA space.

\[
A = \sum_{i=1}^{n} f(x_i) f^T(x_i)
\]

The function \( f(x_i) \) is defined as follows:

\[
f(x_i) = x - f_\omega - \|x\|^2 f_0 \in \mathbb{R}^m
\]

Where:

\[
s_\omega = f_\omega \cdot s, \quad \frac{1}{2} s_0 = f_0 \cdot s
\]

\[
f_\omega = -\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \|x_i\|^2 x_i
\]

\[
f_0 = \sum_{i=1}^{n} \|x_i\|^2 - n \sum_{i=1}^{n} \|x_i\|^2 x_i
\]

and the sum of squares \( \sum_2 = \sum_{i=1}^{n} \|x_i\|^2 \) and the sum of the four powers \( \sum_4 = \sum_{i=1}^{n} \|x_i\|^4 \).

An eigenvector \( s \) is an Eigen conformal vector of a subset \( X \) defined in hyper-plane or hyper-sphere \( S = s + s_\omega e_\omega + s_0 e_0 \) and eigenvalues \( \lambda \) are variance.

B. Long Short Term Memory

The LSTM [14] model is an RNN architecture that addresses the vanishing gradient problem. Figure 1 shows the memory cell structure in the LSTM model. The input gate, output gate, and forget gate in the module are nonlinear summary units containing excitation functions.
LSTM cells have layers called “gates,” which will allow information to be “forgotten” or “perpetuated” to the next steps/cells and the determination of the output values $C_t, h_t$ are determined by the following steps:

- **Input**: $C_{t-1}, h_{t-1}, x_t$ where $x_t$ is the input vector at time $t$ of the model. $C_{t-1}, h_{t-1}$ is the output of the previous layer.
- **Output**: $C_t, h_t$ are called cell states and hidden states.

There is a forget gate $f_t$ to forget information no longer necessary in previous Cell State $C_{t-1}$. The function $f_t$ is calculated based on the input value $x_t$ at the time $t$ has the same value of $h_{t-1}$ and bias $h_f$ forget gate,

$$f_t = \sigma(U_f \times x_t + W_f \times h_{t-1} + h_f)$$  \hspace{1cm} (14)

An Input gate to save new information will necessary. For the potential Candidate value $\tilde{C}_t$,

$$\tilde{C}_t = \tanh(U_c \times x_t + W_c \times h_{t-1} + h_c)$$  \hspace{1cm} (15)

Activation value $i_t$ is also calculated as follows:

$$i_t = \sigma(U_i \times x_t + W_i \times h_{t-1} + h_i)$$  \hspace{1cm} (16)

An Output gate $o_t$ to control the output of the cell,

$$o_t = \sigma(U_o \times x_t + W_o \times h_{t-1} + h_o)$$  \hspace{1cm} (17)

$$h_t = o_t \times \tanh(C_t)$$  \hspace{1cm} (18)

The new $t^{th}$ values are updated using this equation,

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t$$  \hspace{1cm} (19)

where $\sigma$ is the *sigmoid* activation function; $W_i, W_o, W_f$ and $W_c$ are the input gate, output gate, forget gate, and memory letter, respectively. The weight matrix of the element; $b_i, b_o, b_f$ and $b_c$ are the offsets of the input gate, the output gate, the forgetting gate, and the memory cell. Fig.1. is the LSTM architecture.

![Figure 1. LSTM model of the memory cell](image)

**III. PROPOSED METHOD**

The proposed method is to represent moving objects (markers) on the body from which to recognition actions. Specifically, the proposed use of the CMU motion capture dataset consists of 08 different actions, each action consisting of multiple files and each file consisting of corresponding frames. In each frame there are 41 markers (41 joints), each marker is each coordinate are represented in 3D space. Figure 2 is the proposed model LSTM for human activity recognition based on the feature extraction method using Conformal Geometric Algebra.

![Figure 2. Proposed model using CGA based on LSTM to human activity recognition](image)
Given a dataset:
\[ X = \{ x_{ij} | x_{ij} \in \mathbb{R}^{i \times m \times 3} \}, \quad i \in \{1, \ldots, n\} \]  

(20)

Where:
- \( j \in (1, F) \) is the number of frames of the \( i^{th} \) activity; \( m \) is the number of markers; \( n \) is the number of activities; \( F \)
  is the sum of frames for the dataset.
- \( x_{ij} = [\delta_{ij}^1, \ldots, \delta_{ij}^F] \in \mathbb{R}^{i \times m \times 3} \) is the feature vector of the \( i^{th} \) activity.
- \( \delta_{ij} = [\theta_{ij1}, \theta_{ij2}, \theta_{ij3}] \ldots [\theta_{ijm1}, \theta_{ijm2}, \theta_{ijm3}] \in \mathbb{R}^{m \times 3} \) are all coordinates of markers of the \( i^{th} \) activity at the \( j^{th} \)
  frame. \( \theta_{ijm(1,2,3)} \) are coordinate in the \( x, y, z \) axes makers of the \( i^{th} \) activity at the \( j^{th} \) frame, respectively.

The PCA algorithm uses orthogonal transformations to convert the data set from a multi-dimensional space to a new space with less dimension. This transformation is based on finding the axis of the new space so that the method of

from the minimization of (22):

\[ E = \sum_{i=1}^{n} \sum_{j=1}^{F} d^2(P_{ij}, S) = \sum_{i=1}^{n} \sum_{j=1}^{F} (x_{ij} - s, s - s_o - \frac{1}{2} \| x_{ij} \|^2 s_0)^2 \]

Therefore, we might be tempted to express the previous problem using a non-negative Lagrange multiplier \( \lambda \) as the minimization of (22):

\[ L(s, \lambda) = \frac{1}{\sum_{i=1}^{n} \sum_{j=1}^{F} (x_{ij} - s, s - s_o - \frac{1}{2} \| x_{ij} \|^2 s_0)^2 - \lambda \| s \|^2 - 1) \]

(23)

From Eq. (5) to Eq.(10), The function \( f(x_i) \) is defined as follows:

\[ f(x_{ij}) = x_{ij} - f_o - \| x_{ij} \|^2 f_0 \in \mathbb{R}^m \]

(24)

The optimal result can be solved using the Eigen problem:

\[ A s = \lambda s \]

(25)

where \( A \) is the variance matrix of the \( i^{th} \) training set in CGA space:

\[ A = \sum_{i=1}^{n} \sum_{j=1}^{F} f(x_{ij}) f(x_{ij})^T \]

(26)

CGA uses a decrease in the number of dimensions of the data using the first \( k \) eigenvectors \( 1 \leq k \leq m \times 3 \). The feature \( f_{CGA}(x) \) can be extracted from vector \( x \) using the first \( k \) eigenvector as follows:

\[ f_{CGA}(x) = ((P_{11}, S_1), \ldots, (P_{1k}, S_k))^T \]

(27)

Now, we use the transform \( f_{CGA}(x) \) to apply the learning model by converting the dataset \( T = (f_{CGA}(x), h_t|x \in \mathbb{R}^{kx3}, h_t \in \{1, \ldots, c\}) \), where \( f_{CGA}(x) \) and \( h_t \) are label and feature vector after applying CGA.

Then, we use the data set \( T \) to initialize the input data for the LSTM model. From Eq.(11), the formula is rewritten as follows:

\[ f_t = \sigma(U_f \times x + W_f \times x_{t-1} + h_f) \]

(28)

Because there is only one output value, \( h_t \) can be determined through the activation function is \( \text{tanh} \) and Eq.(15-18) is rewritten:

\[ h_t = o_t \times \text{tanh}(C_t) \]

(29)

This model is implemented on CGA space, i.e., data in real space is transferred to CGA space. With the feature of CGA, it is possible to represent objects in space and geometric relationships very well. So movements with complex distributions like human markers use CGA very reasonably. Fig. 3 is the training model in the LSTM network.
IV. EXPERIMENTAL RESULTS

A. Experimental data

The database of motions of Carnegie Mellon University (CMU), USA is free for all uses. Motions are captured in a working volume of approximately 3 m × 8 m. In this model, humans wear a black jumpsuit having 41 markers. The Vicon cameras see the markers in infrared. The images picked up from the various cameras are triangulated to get 3D data. The experimental process was carried out on the CMU dataset with 08 activities (dance, jump, kicking, placing tea, putt, run, swing, walk) and 19,869 frames, dividing the number of frames of each activity into two parts as table 1. experimental data.

<table>
<thead>
<tr>
<th>Action</th>
<th>Sample</th>
<th>Number of frames</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Training</td>
</tr>
<tr>
<td>Dance</td>
<td>12</td>
<td>3,305</td>
</tr>
<tr>
<td>Jump</td>
<td>5</td>
<td>1,198</td>
</tr>
<tr>
<td>Kick</td>
<td>7</td>
<td>1,605</td>
</tr>
<tr>
<td>Placing Tee</td>
<td>6</td>
<td>1,487</td>
</tr>
<tr>
<td>Putt</td>
<td>6</td>
<td>1,534</td>
</tr>
<tr>
<td>Run</td>
<td>2</td>
<td>452</td>
</tr>
<tr>
<td>Swing</td>
<td>6</td>
<td>1,324</td>
</tr>
<tr>
<td>Walk</td>
<td>5</td>
<td>1,074</td>
</tr>
<tr>
<td>Total</td>
<td>49</td>
<td>11,979</td>
</tr>
</tbody>
</table>

B. Experiment results

This experiment compared PCA based LSTM and CGA based LSTM. The parameters of the RNN network are the number of neural = 20, epochs = 20, and classes = 8 (8 kinds of human action), number of dimensions = 123 (41 marker × 3), and activation function is softmax. Fig. 4 shows the result of classification using PCA based LSTM.

Fig. 4 shows that when the number of dimensions is 85, the best recognition result is 90.52% and Fig. 5 shows the result of classification using CGA based LSTM. Fig. 5 shows that the result when using CGA will converge most
when receiving the full attributes of the object. At the same time, the results clearly show that if you remove some key attributes, the result will decrease. The best recognition result is 92.52%.

![Figure 5. Classification results of CGA based LSTM](image)

Table 2 is a comparison of the results of the methods. The results for the feature extraction method using CGA are always better than the feature extraction method using PCA.

<table>
<thead>
<tr>
<th>Method</th>
<th>Train</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA_RNN</td>
<td>85.24%</td>
<td>72.83%</td>
</tr>
<tr>
<td>CGA_RNN</td>
<td>95.59%</td>
<td>88.50%</td>
</tr>
<tr>
<td>PCA_LSTM</td>
<td>98.00%</td>
<td>90.50%</td>
</tr>
<tr>
<td>CGA_LSTM</td>
<td>98.10%</td>
<td>92.52%</td>
</tr>
</tbody>
</table>

The combination of the CGA feature extraction method with LSTM gives very good results for data with a complex distribution such as on spheres or hyperspheres.

V. CONCLUSION

In this paper, we propose method feature extraction using CGA to reduce the number of dimensions and create input data for the LSTM network. Experimental results show that the proposed method CGA_LSTM has 92.52% higher results than 90.50% of PCA_LSTM. However, the research needs to continue to apply the proposed model in practice. We plan to extend our approach to Human Activity Recognition and test its applicability in real-time applications in future work. Furthermore, we also explored the impact of some hyper-parameters on model performance such as the number of filters, the type of optimizers, and batch size. Finally, the optimal hyper-parameters for the final design were selected to train the model. To sum up, compared with the methods proposed in other literature, the CGA_LSTM model shows consistent superior performance and has good generalization. It can not only avoid complex feature extraction but also has high recognition accuracy under the premise of a few model parameters.

REFERENCES

LSTM FOR HUMAN ACTIVITY RECOGNITION BASED ON FEATURE EXTRACTION METHOD USING CONFORMAL GEOMETRIC ALGEBRA

Nguyen Nang Hung Van, Pham Minh Tuan, Do Phuc Hao, Kanta Tachibana