DOI: 10.15625/vap.2021.0057

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

Nguyen Nang Hung Van[1], Pham Minh Tuan[2], Do Phuc Hao[3], Kanta Tachibana[4]

[1,2] The University of Da Nang - University of Science and Technology.
[3]Danang Architecture University, 566 Nui Thanh St., Da Nang 550000, Vietnam.
[4]Kogakuin 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 Geomatric 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 $\mathcal{O}=\{e_1,\dots,e_p,e_{p+1},\dots,e_{p+q}\}$, such as $e_i^2=1,\forall i\in\{1,\dots,p\}$ and $e_i^2=-1,\forall i\in\{p+1,\dots,q\}$. GA denotes $\mathcal{O}$ by $\mathcal{G}_{p,q}$. For example, *m*-dimensional Euclidean vector space $\mathcal{R}^m$ is denoted by $\mathcal{G}_{m,0}$.

A CGA space is extended from the real Euclidean vector space $\mathcal{R}^m$ by adding 2 orthonormal basis vectors. Thus, a CGA space is defined by $m+2$ basis vectors $\mathcal{O}=\{e_1,\dots,e_m,\boldsymbol{e}_+,\boldsymbol{e}_-\}$, where $\boldsymbol{e}_+$ and $\boldsymbol{e}_-$ are defined as follows:

$$\begin{gathered}\boldsymbol{e}_+^2=\boldsymbol{e}_+\cdot\boldsymbol{e}_+=1,\\ \boldsymbol{e}_-^2=\boldsymbol{e}_-\cdot\boldsymbol{e}_-=-1,\\ \boldsymbol{e}_+\cdot\boldsymbol{e}_-=\boldsymbol{e}_+\cdot\boldsymbol{e}_i=\boldsymbol{e}_-\cdot\boldsymbol{e}_i=0,\forall i\in\{1,\dots,m\}.\end{gathered} \tag{1}$$

Thus, a CGA can be expressed by $\mathcal{G}_{m+1,1}$. In addition, CGA defined:

$$\boldsymbol{e}_0=\frac{1}{2}(\boldsymbol{e}_--\boldsymbol{e}_+),\ \boldsymbol{e}_\infty=(\boldsymbol{e}_-+\boldsymbol{e}_+). \tag{2}$$

Given training set $\mathbf{X}=\{\boldsymbol{x}_i|\boldsymbol{x}_i\in\mathcal{R}^m\}, i\in\{1\dots n\}$ represented in real *m*-dimensional space. This training set is re-represented by the set of points $\boldsymbol{P}\in\boldsymbol{\mathcal{G}}_{m+1,1}$ in CGA space [16, 17] as follows,

$$\boldsymbol{P}_i=\boldsymbol{x}_i+\frac{1}{2}\|\boldsymbol{x}_i\|^2\boldsymbol{e}_\infty+\boldsymbol{e}_0\in\mathcal{G}_{\mathrm{m+1,1}} \tag{3}$$

A conformal vector S is generally written in the following:

$$\boldsymbol{S}=\boldsymbol{s}+s_\infty\boldsymbol{e}_\infty+s_0\boldsymbol{e}_0 \tag{4}$$

The process of estimating using least squares $d^2(\boldsymbol{P}_i,\boldsymbol{S})$. The error function E is as follows:

$$E=\sum_{i=1}^{n}(\boldsymbol{x}_i\boldsymbol{s}-s_\infty-\frac{1}{2}\|\boldsymbol{x}_i\|^2s_0)^2 \tag{5}$$

This means that when minimizing the error E function, s can be limited by $\|\boldsymbol{s}\|^2=1$. In this case, the optimization problem becomes as follows:

$$\min\sum_{i=1}^{n}(\boldsymbol{x}_i\boldsymbol{s}-s_\infty-\frac{1}{2}\|\boldsymbol{x}_i\|^2s_0)^2 \tag{6}$$

with the condition is:

$$\|\boldsymbol{s}\|^2=1 \tag{7}$$

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

$$A\boldsymbol{s}=\lambda\boldsymbol{s} \tag{8}$$

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

$$A=\sum_{i=1}^{n}f(\boldsymbol{x}_i)f^T(\boldsymbol{x}_i) \tag{9}$$

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

$$f(\boldsymbol{x}_i)=\boldsymbol{x}-f_\infty-\|\boldsymbol{x}\|^2f_0\in\mathcal{R}^m \tag{10}$$

Where:

$$s_\infty=\boldsymbol{f}_\infty.\boldsymbol{s},\frac{1}{2}s_0=\boldsymbol{f}_0.\boldsymbol{s} \tag{11}$$

$$\boldsymbol{f}_\infty=\frac{-\sum_4\sum_{i=1}^n\boldsymbol{x}_i+\sum_2\sum_{i=1}^n\|\boldsymbol{x}_i\|^2\boldsymbol{x}_i}{(\sum_2)^2-n\sum_4} \tag{12}$$

$$\boldsymbol{f}_0=\frac{\sum_2\sum_{i=1}^n\boldsymbol{x}_i-n\sum_{i=1}^n\|\boldsymbol{x}_i\|^2\boldsymbol{x}_i}{(\sum_2)^2-n\sum_4} \tag{13}$$

and the sum of squares $\sum_2=\sum_{i=1}^n\|\boldsymbol{x}_i\|^2$ and the sum of the four powers $\sum_4=\sum_{i=1}^n\|\boldsymbol{x}_i\|^4$

An eigenvector $\boldsymbol{s}$ is an Eigen conformal vector of a subset $\boldsymbol{X}$ defined in hyper-plane or hyper-sphere $S=\boldsymbol{s}+\boldsymbol{s}_\infty\boldsymbol{e}_\infty+\boldsymbol{s}_0\boldsymbol{e}_0$ and eigenvalues λ 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,

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

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

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

Activation value i_t is also calculated as follows:

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

An Output gate $\mathrm{o_t}$ to control the output of the cell,

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

$$h_t = o_t \times tanh(C_t) \quad (18)$$

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

$$C_t = f_t \times c_{t-1} + i_t \times \tilde{C}_t \quad (19)$$

where σ 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.

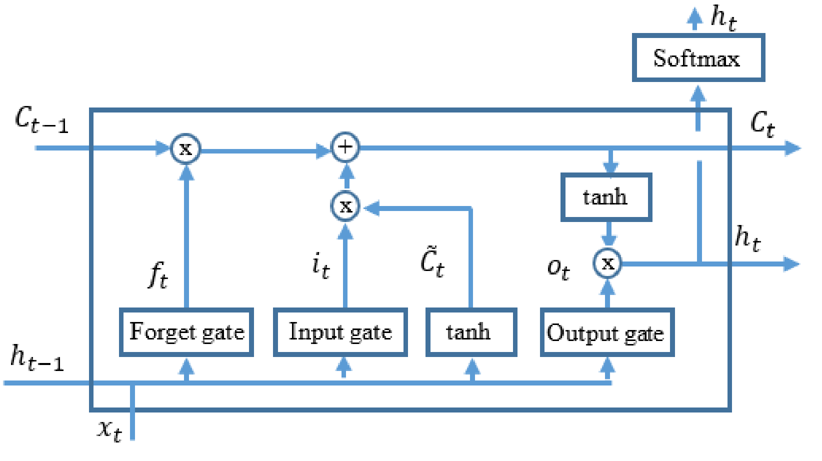

Figure 1. LSTM model of the memory cell

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.

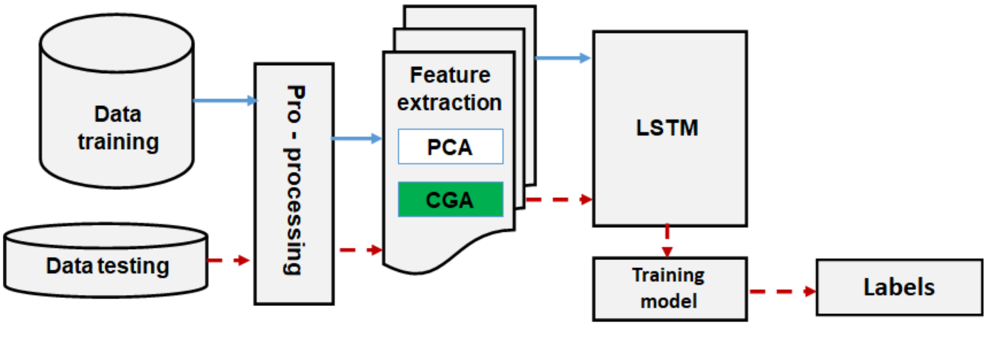

Figure 2. Proposed model using CGA based on LSTM to human activity recognition

Given a dataset:

$$\mathbf{X} = \{\boldsymbol{x}_{ij} | \boldsymbol{x}_{ij} \in \mathcal{R}^{j\times m\times 3}\}, \qquad i \in \{1, \dots, n\} \tag{20}$$

Where:

- $j \in (1, F)$ is the number of frames of i^{th} activity; m is the number of makers; n is the number of activities; F is the sum of frames for the dataset.
- $\boldsymbol{x}_{ij} = [\delta_{ij}^T, \dots, \delta_{ij}^T]^T \in \mathcal{R}^{j\times m\times 3}$ is the feature vector of i^{th} activity.
- $\delta_{ij} = [[\theta_{ij1_1}, \theta_{ij1_2}, \theta_{ij1_3}], \dots, [\theta_{ijm_1}, \theta_{ijm_2}, \theta_{ijm_3}]]^T \in \mathcal{R}^{m\times 3}$ are all coordinates of makers of i^{th} activity at j^{th} frame. $\theta_{ijm_{k(1,2,3)}}$are coordinate in the *x*, *y*, *z* axes makers of i^{th} activity at j^{th} frame, respective.

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 data projected on that axis is greatest. Similar to the PCA algorithm, feature extraction uses CGA by projecting data from P onto conformal vector S to determine the maximum variance.

The first, transfer the data **X** from the real space R^m that will be re-expressed in CGA space with the set of points $P \in \boldsymbol{\mathcal{G}}_{m+1,1}$:

$$P_{ij} = \boldsymbol{x}_{ij} + \frac{1}{2}\|\boldsymbol{x}_{ij}\|^2 \boldsymbol{e}_\infty + \boldsymbol{e}_0 \in \mathcal{G}_{m\times d+1,1} \tag{21}$$

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

$$E = \sum_{i=1}^{n}\sum_{j=1}^{F} d^2(P_{ij}, S) = \sum_{i=1}^{n}\sum_{j=1}^{F} (\boldsymbol{x}_{ij}\boldsymbol{s} - s_\infty - \frac{1}{2}\|\boldsymbol{x}_{ij}\|^2 s_0)^2 \tag{22}$$

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

$$L(s, \lambda) = \frac{1}{\sum_{i=1}^{n} t(i)} \sum_{i=1}^{n}\sum_{j=1}^{F} (\boldsymbol{x}_{ij}\boldsymbol{s} - s_\infty - \frac{1}{2}\|\boldsymbol{x}_{ij}\|^2 s_0)^2 - \lambda(\|s\|^2 - 1) \tag{23}$$

From Eq. (5) to Eq.(10), The function $f(\boldsymbol{x}_i)$ is defined as follows:

$$f(\boldsymbol{x}_{ij}) = \boldsymbol{x}_{ij} - f_\infty - \|\boldsymbol{x}_{ij}\|^2 f_0 \in \mathcal{R}^m \tag{24}$$

The optimal result can be solved using the Eigen problem:

$$As = \lambda s \tag{25}$$

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

$$A = \sum_{i=1}^{n}\sum_{j=1}^{F} f(\boldsymbol{x}_{ij}) f^T(\boldsymbol{x}_{ij}) \tag{26}$$

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

$$f_{CGA}(\boldsymbol{x}) = ((P_{11}, S_1), \dots, (P_{ij}, S_k))^T \tag{27}$$

Now, we use the transform $f_{CGA}(\boldsymbol{x})$ to apply the learning model by converting the dataset $\mathbf{T} = \{f_{CGA}(\boldsymbol{x}), h_t | \boldsymbol{x} \in \mathcal{R}^{k\times 3}, h_t \in \{1, \dots, c\}\}$, where $f_{cga}(\boldsymbol{x})$ 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:

$$f_t = \sigma(U_f \times \boldsymbol{x} + W_f \times h_{t-1} + h_f) \tag{28}$$

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

$$h_t = o_t \times tanh(C_t) \tag{29}$$

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.

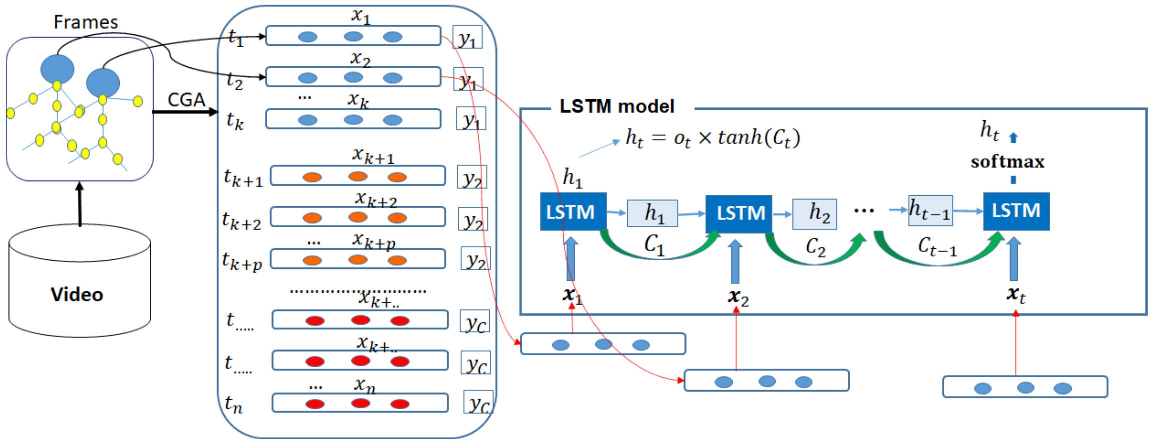

Figure 3. Human action data representation in the LSTM network

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 1. Experimental data

Action	Sample	Number of frames		
		Training	Testing	Total
Dance	12	3,305	1,577	4,882
Jump	5	1,198	846	2,044
Kick	7	1,605	1,163	2,768
Placing Tee	6	1,487	1,096	2,583
Putt	6	1,534	974	2,508
Run	2	452	322	774
Swing	6	1,324	977	2,301
Walk	5	1,074	928	2,002
Total	**49**	**11,979**	**7,883**	**19,862**

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.

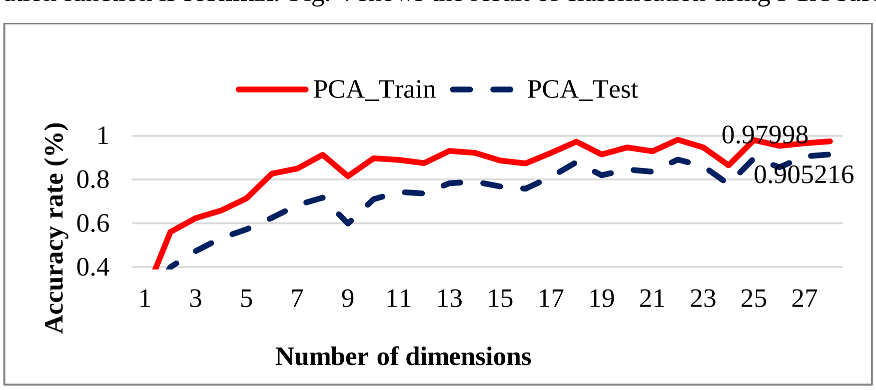

Figure 4. Classification results of PCA based LSTM

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%.

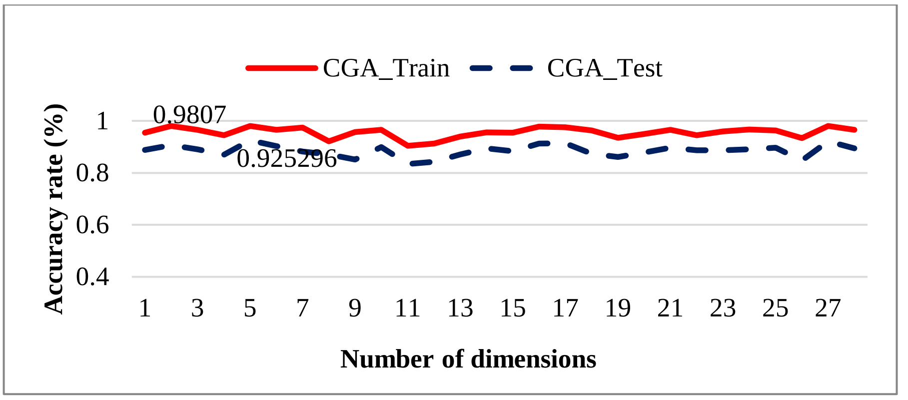

Figure 5. Classification results of CGA based LSTM

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 2. Compares the results of the methods

Method	Results	
	Train	Test
PCA_RNN	85,24%	72,83%
CGA_RNN	95,59%	88,50%
PCA_LSTM	98,00%	90,50%
CGA_LSTM	98,10%	**92,52%**

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

[1] I. T. Jolliffe, Principal Component Analysis, New York: 2nd Edn. Springer-Verlag, 2002.

[2] Lindsay I Smith, A tutorial on Principal Components Analysis, February 26, 2002.

[3] J. I, Alan, Linear Discriminant Analysis, Springer, 2012.

[4] I. T. Joliffe, “A note on the use of principal components in regression, Applied Statistics”, 1982.

[5] L. Jun, G. Wang, Li. Y. Duan, “Skeleton-based human action recognition with global context-aware attention LSTM networks, *IEEE Transactions on Image Processing*, vol. 27, no. 4, pp. 1586-1599, 2017.

[6] H. Liu and T. Taniguchi, “Feature extraction and pattern recognition for human motion by a deep sparse autoencoder”, in *IEEE International Conference on Computer and Information Technology*, 2014.

[7] A. Rosebrock, Deep Learning for Computer Vision, PyimageSearch, September 2017.

[8] K. Arthishwari, M. Anand, “Design of LSTM-RNN on a Sensor Based HAR using Android Phones”, *International Journal of Recent Technology and Engineering (IJRTE)*, vol. 8, no. 5, pp. 2277-3878, 2020.

[9] Anil K. Jain, “Artificial Neural network: A tutorial” *in IEEE Computer*, vol. 29, no. 3, pp. 31-44., March 1996.

[10] Ian Goodfellow, Yoshua Bengio, and Aaron Courville, Deep Learning, MIT Press, 2016.

[11] Y. Du, W. Wang, and L. Wang, “Hierarchical recurrent neural network for skeleton based action recognition”, in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015.

[12] Graves, A., Liwicki, M., Fernandez, S., Bertolami, R., Bunke, H., and Schmidhuber, J., "A novel connectionist system for unconstrained handwriting recognition", in *IEEE transactions on pattern analysis and machine intelligence*, vol. 31, no. 5, p. 855–868, 2009.

[13] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks", in *International Conference on Machine Learning*, 2013.

[14] S. Hochreiter and J. Schmidhuber, " Long short-term memory", in *Neural Computation*, vol. 9, no. 8, p. 1735–1780, 1997.

[15] The Carnegie Mellon University, "The Carnegie Mellon University Motion Capture Database", http://mocap.cs.cmu.edu.

[16] Hachaj, T., M. R. Ogiela, and M. Piekarczyk, "Dependence of Kinect sensors number and position on gestures recognition with Gesture Description Language semantic classifier", in *Computer Science and Information Systems (FedCSIS)*, Federated Conference, 2013.

[17] D. Hestenes and G. Sobczyk, Clifford Algebra to Geometric Calculus: A unified language for mathematics and physics, Reidel, 1984.

[18] D. Hildenbrand and E. Hitzer, "Analysis of point clouds using conformal geometric algebra", in *3rd International conference on computer graphics theory and applications*, Funchal, Madeira, Portugal, 2008.

[19] Eckhard Hitzer, Tohru Nitta and Yasuaki Kuroe, "Applications of Clifford's Geometric Algebra", in *arXiv:1305.5663v1*, 24 May 2013.

[20] L. Dorst, D. Fontijne, and S. Mann, "An Object-oriented Approach to Geometry Morgan Kaufmann Series in Computer Graphics", in *Geometric Algebra for Computer Science*, 2007.

[21] M. T. Pham, K. Tachibana, T. Yoshikawa, and T. Furuhashi, "A clustering method for geometric data based on approximation using conformal geometric algebra," in *2011 IEEE International Conference on Fuzzy Systems*, 2011.

[22] M.T. Pham, K. Tachibana, E. M. S. Hitzer, S. Buchholz, T. Yoshikawa, and T. Furuhashi, "Feature Extractions with Geometric Algebra for Classification of Objects," in *IEEE World Congress on Computational Intelligence*, Hongkong, 2008.

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

TÓM TẮT: *Trong những năm gần đây, Học sâu là một trong những nghiên cứu có tính thời sự trong lĩnh vực xử lý ngôn ngữ tự nhiên, xử lý ảnh và nhận dạng hành động. Học sâu có hai mô hình chính là Mạng nơron tích chập được áp dụng để trích chọn đặc trưng trong xử lý hình ảnh và Mạng Nơron hồi quy được áp dụng trong nhận dạng chuỗi thời gian. Tuy nhiên, kiến trúc của mạng nơron hồi quy khá đơn giản và khả năng ghi nhớ những thông tin từ xa không tốt. Nên những thông tin trong chuỗi đầu vào thường không có nhiều ảnh hưởng đến kết quả dự đoán chuỗi đầu ra của các bước sau. Do đó, Mạng nhớ ngắn hạn hồi tiếp (LSTM) được thiết kế để khắc phục những hạn chế của mạng nơron hồi quy. Hơn nữa, dữ liệu đầu vào trong các mô hình huấn luyện thường có số chiều cao nên chi phí tính toán lớn và dẫn đến độ chính xác nhận dạng không cao. Để giải quyết vấn đề này, một số nghiên cứu trước đây như Phân tích thành phần chính (PCA), Hồi quy thành phần chính (PCR) và Phân tích biệt thức tuyến tính (LDA) được đề xuất để giảm số chiều của dữ liệu và độ phức tạp cho các mô hình học máy. Tuy nhiên, các phương pháp này, đã sử dụng các phép biến đổi tuyến tính và giả sử dữ liệu được phân phối trên một mặt phẳng hoặc siêu mặt phẳng đặc biệt nào đó. Nên dẫn đến những khó khăn nhất định đối với dữ liệu phân bố trên hình cầu hay siêu cầu, chẳng hạn như các đối tượng chuyển động quay trong không gian nhiều chiều. Vì vậy, nghiên cứu này đã đề xuất phương pháp trích chọn đặc trưng sử dụng Đại số hình học bảo giác (CGA) để giảm số chiều dữ liệu cho LSTM huấn luyện và nhận dạng. Đầu tiên, dữ liệu hành động được tiến hành xử lý để chuẩn hóa dữ liệu. Tiếp theo, đề xuất phương pháp trích chọn đặc trưng sử dụng CGA để giảm số chiều dữ liệu và tạo vectơ đặc trưng. Cuối cùng, sử dụng LSTM để huấn luyện và nhận dạng hành động. Thực nghiệm được tiến hành trên bộ dữ liệu CMU với 8 hành động khác nhau và kết quả cho thấy phương pháp đề xuất có kết quả cao hơn các phương pháp trước đó.*