ISSN: 2093-5374 / E-ISSN: 2093-5382

Design and Implementation of Gesture Music Composition System based on RealSense and Recurrent Neural Network

Gang Bao[1*]

[1*]PhD Studying, Department of Fine Arts, International College, Krirk University, Bangkok, Thailand. 564636867@163.com, https://orcid.org/0009-0002-0905-5000

Received: June 24, 2024; Revised: August 12, 2024; Accepted: September 06, 2024; Published: September 30, 2024

Abstract

Aiming at the problems of slow human-computer interaction and low accuracy of hand gesture recognition (HGR) in music composition systems such as human-computer interaction, gesture recognition, and algorithmic composition, this paper designs an HGR music composition system based on real-sense technology and recurrent neural network. The system uses an Intel RealSense 3D camera to shoot the user's hand and extract the information on the hand joint points with Smart Wearable Biosensors (SWBSs). The joint points are used to construct a three-dimensional model of the hand bone, and the joint points' equivalent distance and joint deflection are calculated by the joint point information to match the corresponding Curwen HGR. The matching music data is input into the recurrent neural network. GRU is combined with the Markov chain algorithm to complete the composition and avoid the phenomenon of gradient disappearance or gradient explosion. The HGR recognition function and composition algorithm are simulated and tested to verify the method's feasibility. Cohort verification investigations and efficacy evaluations of wearable biosensors are essential to support their clinical acceptability. The research results show that the HGR music composition system based on real-sense technology and recurrent neural networks can effectively identify the tester's HGRs and assist the tester in completing the composition using intelligent biosensors. The composition mode is close to the professional composition process.

Keywords: RealSense Technology, Recurrent Neural Network, Hand Gesture Recognition; Music Composition, Human–Computer Interaction, Chopin 's Nocturne, Biosensors.

Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA), volume: 15, number: 3 (September), pp. 501-520. DOI: *10.58346/JOWUA.2024.I3.032*

*Corresponding author: PhD Studying, Department of Fine Arts, International College, Krirk University, Bangkok, Thailand.

1 Introduction

The design and fabrication of Smart Wearable Biosensors (SWBSs) and their prospective applications in human health surveillance and customized treatment have garnered considerable attention. WBSs are transportable electronic gadgets that incorporate sensors into or onto the human body, manifesting as tattoos, gloves, clothes, or implants, facilitating in vivo detection, data storage, and computation via mobile or handheld devices. SWBSs facilitate reciprocal feedback between clinicians and patients. Recent advances and developments in materials research, mechanical technology, and wireless communication technology have contributed to the growth of multiple wearable gadgets (such as watches and bands) to handle and concurrently analyze biological indicators to enhance healthcare administration.

Traditional composition creation is a long and exquisite art form. Combining human creativity and music theory creates terrific and moving music works. However, with the rapid development of science and technology, digital music technology has gradually emerged, bringing unprecedented influence and challenges to traditional composition creation. Digital music technology has gradually become essential to modern music creation due to its convenience, innovation, and diversity. Composing requires professional music literacy and skills, and with the application of human-computer interaction to digital music technology, non-professionals can try to create music.

Such systems require users to interact with the human-machine interface in real-time, using input peripherals to control the interface's music components and composition parameters. Touch and space gestures are more natural than the traditional mouse, keyboard, control rod, etc. Real-time composition scenes often require quickness. Hand gestures (HGR) can quickly complete interactions for complex operations due to their semantic and spatial attributes. Therefore, using HGRs increases interactivity and dramatically simplifies the operation's complexity, resulting in more and more applications in real-time composition systems using SWBSs.

However, HGR interaction (Chen & Liu, 2022) puts forward higher requirements for HGR recognition technology. It is not only to identify the position, movement trend, speed, amplitude, and other motion characteristics of the hand but to identify the shape of the hand, as well as the dynamic HGRs with specific semantics, such as waving, shaking, and thumb up, which are commonly used in daily life. HGR technology, in the past decade, from computer vision, sensing, and EMG recognition to today's widespread 3D infrared detection, gradually evolved to mature. Microsoft (Ding & Chang, 2015), Apple PrimeSense (Deng et al., 2020), Intel (Jeeru et al., 2022), and many other companies have successively released their 3D infrared camera products, and the Microsoft Kinect4 series is widely used in research and commercial fields. The wearable sensors comprise an Inertial Measuring Unit (IMU) for detecting arm and finger motions, flexible strain gauges for monitoring bodily motion, and surface electromyography for capturing electrical impulses from contractions of muscles, among

others. Scientists have recently created wearable gadgets for the upper limb, including information gloves and wrist/armbands. Wrist or arm-mounted systems incorporate sensors that track numerous biosignals, such as muscle volume modifications, skin vibrations, arterial pressure, and ambient factors, including tiny arm motions detected by an IMU.

The advancement of HGR recognition technology continues to drive innovation in HGR interaction with SWBSs. With more and more innovative ways of interaction, real-time composition systems are becoming closer to the professional music creation mode. The mapping relationship between HGR and composition evolves more in line with the structure and principle of music creation. Hands (Tanaka, 2000) and BioMuse (Waisvisz, 1985) used HGRs to correspond to different scales to produce a melody from the beginning. Air Worms (Dixon & Goebl, 2005) and Hand Composer(Mandanici & Canazza, 2014) used HGRs to adjust the volume, rhythm, and tonality composition parameters. The turning point, fundamentally close to the actual composition mode, borrows the essential harmony theory in composition. Crossover uses HGRs to call multi-level chords and generate multi-part melody and orchestration. The significance of the turning point is that the result of the composition changes from a random scale melody to a piece of music with tonality, rhythm change, primary and secondary melody, and multi-part orchestration and structurally conform to the harmony theory, which makes it possible for the public without composition experience to create professional music employing the HGR of daily communication.

HGR interaction has been gradually applied in real-time composition systems. However, this innovative way of interaction needs to be closer to nature. The new 3D infrared HGR recognition technology will make a breakthrough in recognizing natural HGRs with semantics. At the same time, by training the deep learning algorithm of the recurrent neural network model to predict the melody, the music composition system is pushed to a more professional height. There is room for improvement and innovation in the four aspects of interaction: HGR recognition technology, mapping relationship between HGR and composition, and generation algorithm. This system takes the HGR music composition system as the research object. It aims to explore a real-time composition system that is closer to nature in interaction and closer to professionals in composition mode.

This study aims to thoroughly examine recent innovative SWBS connections and methods for heart rate identification while identifying the problems that impede their practical application.

2 Literature Review

Elements of Music

Music is an art form organized by time and composed of melody, harmony, rhythm, and other elements. Some of these elements are emphasized or ignored depending on the style or type. The scope of music performance is comprehensive, including various vocal skills from singing to rap and performance

skills of different kinds of instruments. Music works include works with only instrumental parts, works with only vocal parts (e.g., songs without instrumental accompaniment), and works that combine vocal and instrumental music.

The characteristics of the music structure in the Romantic period are not only reflected in the morphological changes in how the structure is generated. Taking Chopin's Nocturne as an example, its detailed structure is closely related to the overall structure, and the two are mutually causal and influence each other. In the period of classicism, the method of music construction and development has always taken ' theme ' and ' tonality ' as the main elements of the structure of works.

In traditional Western music, music is considered to be composed of many musical elements (Yu & Kwak, 2011). The music sound is a stable periodic sound with duration attributes (Liu et al., 2021), pitch (Rong, 2012), loudness, and timbre. Notes used in music are more complex representation systems than music because they contain non-periodic properties, such as transients, trills, and envelope modulation. This paper needs to study music theory and its development history deeply, and it only describes some basic music concepts to to facilitate a more intuitive understanding of the research objectives of this paper.

Pitch (Jiang, 2022), known as tone, without confusion, is a perceptual feature that enables sound to be sorted according to frequency-related scales. More generally speaking, pitch is the nature of the sound that can be judged as ' high ' and ' low ' in the musical sense.

Sound length (Hashimoto, 2010), or time value, refers to the duration of notes, phrases, passages, and other elements or a specific time interval. According to Benward & Saker, (2003), the length is one of the characteristics of rhythms, a musical element, and the core of music time measurement and musical form.

The Oval (Kawamura & Yokota, 2005) is between two pitches in music theory. The interval can be reflected in a continuous musical note (such as two adjacent notes in the melody). Alternatively, two vertical musical notes (such as chords) can be emitted simultaneously.

In music theory, a scale is any notes ordered based on fundamental frequency or pitch. The scale sorted by the increase of pitch is called the ascending scale, while the scale sorted by the decrease of pitch is called the descending scale. The pitches in the up and down of some scales are not the same, such as the scale of the melody minor.

Intel RealSense Technology

Intel RealSense technology is a RealSense technology developed by Intel. The original name is Intel Perceptual Computing, which was named in 2012 and translated into perceptual computing. In 2013, Intel opened the SDK of Perceptual Computing Software Development Kit and attracted many perceptual computing developers to participate in the Intel Perceptual Computing Challenge held by

Intel with high bonuses. One year later, with the development of perceptual computing technology becoming increasingly mature, Intel decided to rename Perceptual Computing technology to RealSense, and then computer reality technology was officially born.

Intel's RealSense technology includes two parts: hardware and software. They are the 3D cameras that support Intel RealSense computing and the SDK (Software) - supporting Intel RealSense computing. Some of the Intel RealSense 3D camera models and parameter pairs are represented in Table 1.

Table 1: Analysis of the Specification Parameters of the Four Cameras

Camera model	D405	D415	D435	D455
RGB resolution	1280×720	1920×1080	1920×1080	1280×800
Depth precision	< 50% at 2 cm	< 2% at 2m.	< 2% at 2m.	< 4% at 2m.
RGB sensor field of view (H × V)	87°×58°	69°×42°	69°×42°	90°×65°
Minimum depth distance at maximum resolution (Min-Z)	~7cm	~45cm	~28cm	~52cm
Size (length × height × depth)	42mm×42mm×23mm	99mm×23mm×20mm	90mm×25mm×25.8mm	124mm×29mm×26mm

The 3D camera is divided into two types: a front-end 3D camera for close distance and high precision and a rear-end 3D camera for long distance and low accuracy. The front camera is a hardware structure that comprises an infrared depth sensor color sensor, an infrared laser emitter, and a RealSense image processing chip (see Fig. 1). The real-world vision technology of structured light is adopted. The emitter emits the laser, and the images taken by the infrared and color sensors are processed on the chip to complete the three-dimensional reconstruction. The rear camera uses binocular stereo technology, which simulates the principle of the eyes, emits infrared light into the environment, tracks the position of the light with the infrared sensors and the suitable sensors, and then uses the triangulation principle to calculate the depth information in the 3D image. This paper uses the Intel RealSense Depth D435 camera, as shown in Figure 1.

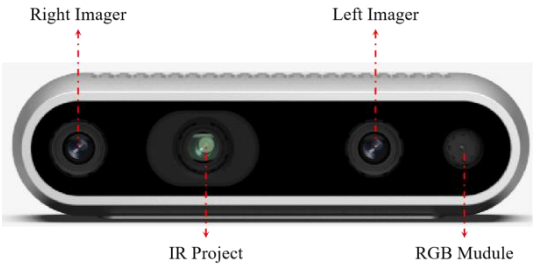

Figure 1: Intel RealSense Depth Camera D435

Intel RealSense SDK (Esper et al., 2022) is an Intel RealSense technology development kit, an algorithm library for image detection and recognition. The application that integrates the Intel RealSense SDK is located on the SDK multi-layer component, which is based on the SDK core. The core of the SDK includes two necessary modules: the I/O module and the algorithm module, which represent different input modes. These modules provide SDK functions for applications, unify programming interfaces, and support multiple languages, frameworks, and game engines. The IO module captures the input data through the camera device. The SDK unifies the programming interface of the I/O module and the algorithm module. The research can directly use the application to access the SDK framework without paying attention to the specific implementation of the underlying code.

Wearable Sensors

Smartwatches represent an emerging domain for hand gesture recognition technologies (Masoumian et al., 2023). Due to the limited size of wristwatch displays for numerous touch-based gestures, hand gestures for operations or input present a viable alternative.

The research employed electromyography and fiducial marker-based tracking to record the user's myoelectric stimulation while performing certain hand motions (Kwon et al., 2021). The technology proved capable of commanding a dexterous robotic hand system and converting gestures into the corresponding grip type for the robotic hand.

The research utilized B-mode ultrasound pictures to recognize four distinct motions in real time, achieving an accuracy of 77% (Li et al., 2022). B-mode ultrasound demonstrates exceptional precision in classification for many muscle conditions and activities, yet it is unsuitable for straightforward, cost-effective, and wearable applications. B-mode ultrasound imaging necessitates complex calculations and visual recognition of pertinent techniques for feature detection.

The research identified four basic hand gestures among four participants with a precision of 81.3% using three detectors. The gesture recognition tests were performed in a static setting in a separate investigation.

A time-of-flight sensor is a widely utilized optical sensor for distance measurement, formerly employed to ascertain the depth of data in a picture (Koerner et al., 2021). The bones and muscles are connected to the skin, resulting in skin distortion from hand motions.

The study introduced the notion of epidermal electronics, which laid the groundwork for contemporary, sophisticated uses of e-skin and e-tattoos in sensing technologies (Li et al., 2024). The study suggested a skin-integrated connection for tactile feedback that delivers targeted vibrating sensations (Jung et al., 2021). In contrast, the study offered a soft skin-stretch gadget for enhanced proprioceptive input (Abad et al., 2022). Progress in electronic flexibility facilitates the creation of soft circuits, resulting in a more user-friendly design. An initial study included energy-collecting devices utilizing daylight or epidermis triboelectric nanotechnology to supply electricity.

RNN

Algorithmic composition can be traced back to the 11th century. The Italian monk Gui-do d'Arezzo created the earliest pitch model, namely the six-tone scale system (Pelchat & Gelowitz, 2020). After a long development period, the algorithmic composition technology (Cai & Cai, 2019; Xia, 2022) has increased, and its application is becoming more and more extensive. The following table summarizes the commonly used composition algorithms.

Table 2: Commonly used Composition Algorithms

Algorithm	Core content
Chaos theory	A mapping cell is established between chaos theory and music
Automaton	Based on the mechanism of biological self-reproduction
Markov model	stochastic process
Genetic algorithm	Random search strategy based on natural system modeling
Rule-based algorithmic	Generated by defining rules
Neural network	Simulating the human brain to construct a neural network to generate music

The Recurrent Neural Network (RNN) model was proposed in the late 1980s (Chapel, 2004) (Gunawan et al., 2020). The fundamental characteristics of this system are internal feedback and feedforward connections among processing devices. From a systemic perspective, it constitutes a feedback dynamic system that embodies the dynamic properties of the procedure during computation, exhibiting more definitive dynamic behavior and computational capacity than the feedforward neural networks (Hannum, 2018). The RNN has emerged as a pivotal subject of investigation for neural network researchers globally, with its principles illustrated in Figure 2.

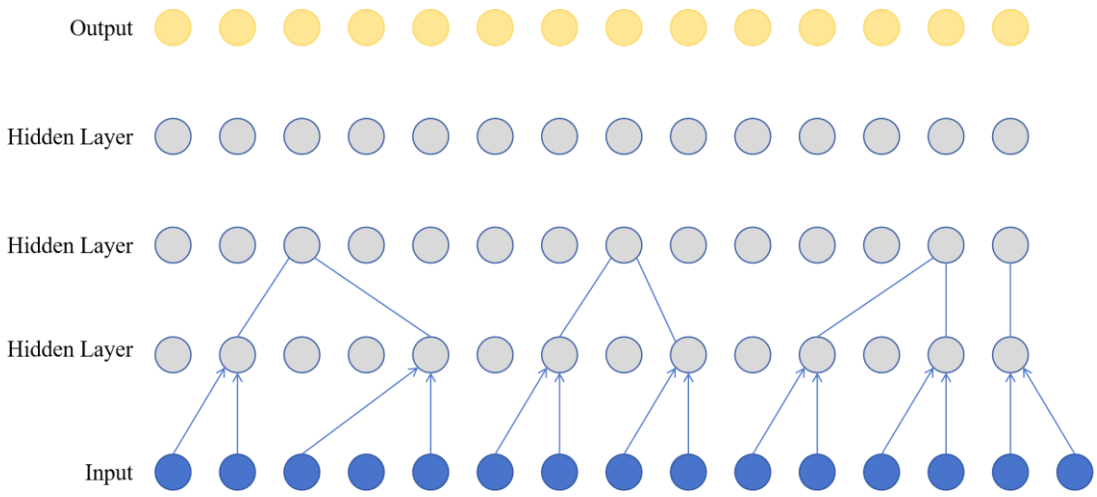

Figure 2: Principle of Convolutional Neural Networks

The RNN network structure includes an input, a concealed, and a resultant tier. The input set is $X=(x_1, x_2, ..., x_t)$, and the output set is expressed as $Y=(y_1, y_2, ..., y_t)$ (Li, 2020). The hidden layer is composed of multiple interconnected repeated nodes. When the input at time t enters the network, the t-th node is added to the hidden layer, and the memory of the first t-1 nodes is passed to the t-th node (Liang, 2022). Currently, the hidden layer memory state is updated to ht, where Φ is a nonlinear function, as shown in Formula 1.

$$h_t = \begin{cases} 0, & t = 0 \\ \Phi(h_{t-1}, x_t), & otherwise \end{cases} \quad (1)$$

$$Y_t = \varphi(Wh_t) \quad (2)$$

: The structural characteristics of the recurrent neural network show that the network structure is good at solving problems related to time series (Lu, 2020). HGR music composition must predict the n + 1st note through the first n note sequences. Therefore, the recurrent neural network is the basic unit for constructing the HGR music neural network model. The ring neural network can theoretically support sequences of any length, but in the actual training process, if the sequence is too long, there will be a long-term dependence problem.

In principle, these recurrent neural networks can capture time dependencies on long-time scales. However, they are challenging to fit data through gradient descent because the gradient involves a higher-order cyclic weight W_{ht}, which will cause gradient disappearance or gradient explosion due to the largest eigenvalue of W_{ht}. The problem (Wang et al., 2021) can be avoided by regenerating the recurrent neural network.

3 Methodology

Figure 3 illustrates a revised SWBS technique within the HGR. This technique outlines the methodology for identifying human actions. The research starts with preparing the database to reduce noise and normalize data from various sources. The research partitions the data via the windowing approach. A deep learning network receives sample information to produce a set of forecasted labels using SWBSs. The resulting set of projected tags is further verified using standard HGR metrics.

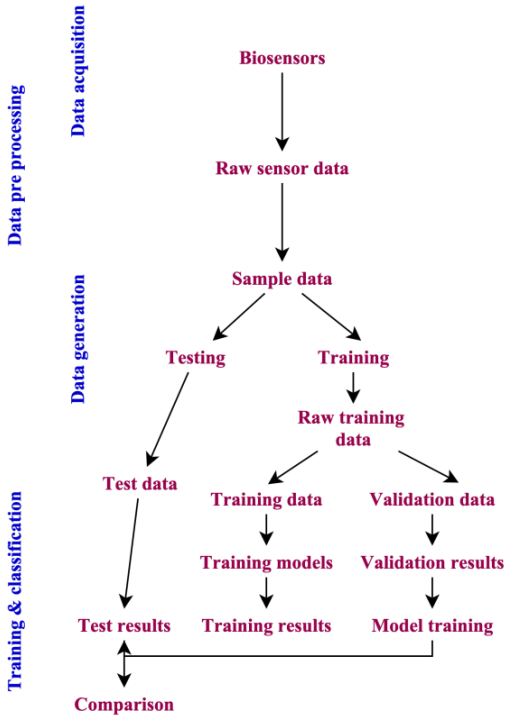

Figure 3: Workflow of the Proposed Model

This section provides an in-depth review of Deep SLR. The essential phases of Deep SLR are data collecting, data preparation, and continuous identification.

Data Acquisition

The research utilizes two armbands on the wrists to capture real-time sign signals from each hand. Each wristband has an AI biosensor and eight axes of sEMG detectors. The biosensor measures the angle of motion and speed of hand motions, while sEMG detectors capture the muscle activity associated with hand moves.

Data Preparation

This process involves cleansing data and extracting functions to normalize and remove noise in real-time data. Considering the disparate sample rates of the SWBS and detectors, the research initially employed spline interpolation to standardize the acquired data to a uniform length and remove spike noise.

Ongoing Acknowledgment

The research employs an RNN architecture to attain end-to-end continuous sign language recognition without disintegration, enhancing the precision of detection without segmentation. The research

employs grammar-based categorization algorithms and a laser technique to deduce the most probable sequencing from the likelihood matrices, which the research utilizes as the final text sentence.

Based on real-sense technology and a recurrent neural network, the HGR music composition system can be divided into three parts: HGR recognition, music composition algorithm, and music output from the functional point of view in SWBSs. The three parts are strictly executed following the timing relationship; the user inputs the HGR to the system in front of the 3D camera. The HGR recognition module captures the information and triggers the composition algorithm. The generated music data passes through the music output module and feeds the music back to the user. The following is a detailed description of each module.

Realization of HGR Recognition Module based on RealSense Technology

The reality camera can accurately extract the complete hand skeleton from the background (see Fig. 4) and track the position and direction of 22 joint points of the hand in real time. These 22 joint points include the fingertips of five fingers, interphalangeal joints, metacarpal joints, palmar joints, and wrist joints. These joints can be used to construct a real-world model proportional to the size of the hand using SWBSs. When the information about the nodes is lost, it indicates that it is beyond the interaction space of the RealSense camera. With the help of the flexion and extension parameters provided by the SDK, the folding state of the finger from the fist to the full extension can be analyzed. With the help of the opening and closing parameters, the curling state from the first to the whole pinch of the finger can be judged. Specific HGRs and movements can be determined by combining the joint point information, flexion and extension parameters, and hand opening and closing parameters using SWBSs.

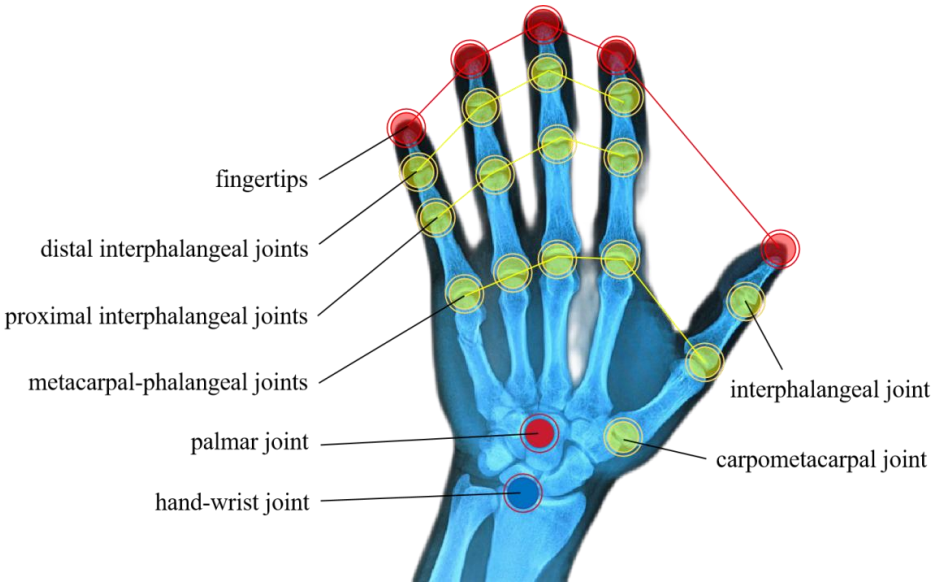

Figure 4: The Extracted Hand Skeleton and Joint Point Distribution Map

The purpose of HGR recognition is to screen out meaningful HGRs. During the interaction with the camera, users use various HGRs commonly used in daily communication, but only part of the HGRs can be recognized in the system. Intel RealSense SDK provides 13 widely used natural HGRs. The system uses all the built-in HGR sets and matches the corresponding operations. They are open finger, V HGR, click, wave, double finger pinch open, complete pinch, fist, thumb up, thumb down, upward swipe, downward sweep, left swipe, and right swipe. According to the characteristics of the HGR music composition system, the Curwen music HGR (Gentry, 2016; Quintero & Roa, 2022) (see Fig. 5) is selected as the screening target; that is, the data that conforms to the Curwen HGR feature is selected from the HGRs used by the user consciously or unconsciously.

Figure 5: Curwen HGR Map

John Curwen first founded the Curwen music teaching method in 1870. Curwen music HGR is an essential content in music teaching. Various high and low position changes are carried out in front of the body using seven different HGR changes to express seven different singing names. The high and low relationship between notes is vividly and effectively reflected in front of people's eyes so that the line of sight cannot capture the pitch that the human body cannot touch into the relationship between the visible pitch changes with SWBSs.

According to the position information of 22 joint points, the hand can be layered and divided into 21 hand areas. A part is discarded from the position of the thumb to the palm. The Euclidean distance of any two joint points can be known through the hierarchical network model and the position coordinate information of 22 joint points.

According to the acquisition of three-dimensional coordinates of joint point position information, the coordinates of joint points 1-22 are expressed as ($a_{i,x}$, $a_{i,y}$, $a_{i,z}$). Let all joint point information be represented as matrix M, then M can be expressed as the following formula:

$$M=\begin{bmatrix} a_{1,x} & a_{1,y} & a_{1,z} \\ \vdots & a_{i,y} & \vdots \\ a_{22,x} & a_{22,y} & a_{22,z} \end{bmatrix} \quad 1\le i\le 22 \qquad (3)$$

The Euclidean distance formula D from all joint points to the No.1 joint point can be denoted as given below:

$$D_{i,j}=\sqrt{(a_{i,x}-a_{j,x})^2+(a_{i,y}-a_{j,y})^2+(a_{i,z}-a_{j,z})^2} \qquad (4)$$

In the above formula, i and j are expressed between 1 and 22, respectively, and $D_{i \text{ and } j}$ are described as the Euclidean distance from the hand joint point i to the joint point j. From the above formula, the Euclidean distance between any two joint points can be calculated. The experimenter's character, hand shape, and size are different due to the uncertainty. The Euclidean distance $D_{i,j}$ is proposed to be standardized to calculate its equivalent distance. The calculation formula is as follows:

$$E_{i,j}=\frac{5\times D_{i,j}}{D_{3,x}+D_{8,x}+D_{12,x}+D_{16,x}+D_{20,x}} \qquad (5)$$

According to the joint deflection angle of the finger, the bending degree of 0 to 100 can be calculated. 0 indicates that the finger is completely bent, 100 indicates that the finger is completely straightened, and gradually Straightens from 0 to 100. Curwen music HGRs can be divided into three categories. The degree of straightening between 0 and 20 indicates that the finger is bent, between 80 and 100 indicates that the finger is straightened, and between 20 and 80 indicates that the finger is not entirely straightened, which is an uncertain state. From this, the following equation can be obtained:

$$Z_i=\begin{cases} 0 & 0\le Q_i\le 20 \\ 1 & 80\le Q_i\le 100 \\ 2 & 20\le Q_i\le 80 \end{cases}, \ 0\le i\le 4 \qquad (6)$$

Z is introduced to denote the finger extension state, Z_i denotes the extension state of i th finger, Q denotes the extension degree value of the finger, Q_k denotes the extension degree value of the k th finger, k is equal to 0, denotes the big finger, k is equal to 1 denotes the index finger, and so on, k is equal to 0 to 4 denotes five fingers respectively. When Z_k is set to 0, the finger is bent; when Z_k is set to 1, the finger is straight; when Z_k is set to 2, it means the finger is in a semi-straight state.

HGRs can be classified from the obtained joint point equivalent distance and joint deflection angle of SWBSs. The HGR recognition process depended on the RealSense model, represented in Figure 6.

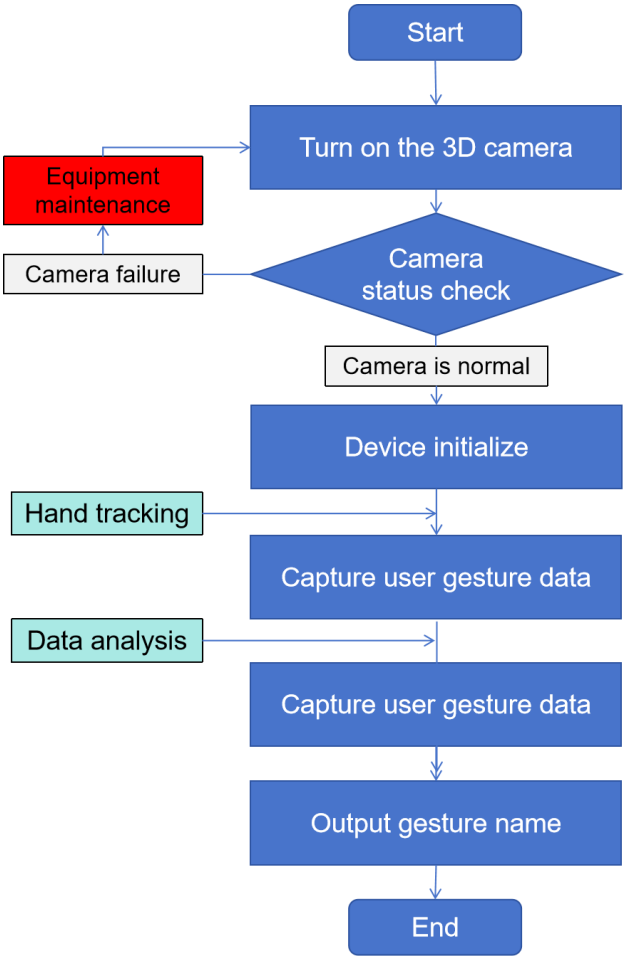

Figure 6: HGR Recognition Flow Chart

A Music Composition Algorithm based on Recurrent Neural Network

Music is a sequence composed of pitch and duration without considering other complex conditions. Based on this, this paper proposes a composition algorithm model combining GRU and Markov chains.

GRU enables the recurrent unit to capture the timing dependence at different times adaptively. The structure of GRU is shown in Figure 7. This model combines the input and forgetting gates into a separate update gate, represented by z in the figure. The reset gate is added, represented by r in the figure. At the same time, the output gate is omitted, and the updated memory information h is directly output. GRU calculates r and z according to the input xt at time t and the condition of the concealed tier. The new memory information h ~ passes through the update gate, and the new memory unit h is obtained from the previous hidden layer state, which is output directly from the module.

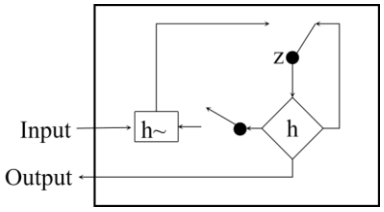

Figure 7: GRU Model Structure

It is assumed that the sequence at time x_t is j, and the sequence at period t-1 is i. According to the definition of the Markov model, the condition j at period x_t is only related to the condition at period xt-1. The transition likelihood condition is mathematically described as follows:

$$p_{ij} = p(i \to j) = p(x_t = j | x_{t-1} = i) \qquad (7)$$

Multiple transition probabilities can form a matrix of transition probabilities, and the p-state transition matrix is expressed as follows: Here, n denotes the total states, $p_{i,j}$(i,j=1,2,3,….,n) denotes the transitional from the current condition I to the following condition. The probability of j should satisfy the following conditions:

$$0 < p_{ij} < 1 \qquad (9)$$

$$\sum_{n}^{j} p_{ij} = 1 \qquad (10)$$

4 Results

The test environment is Intel RealSense camera D435 and HUAWEI MateBook X Pro, and the specific configuration is shown in the following table.

Table 2: Hardware Configuration

Item	Configuration
Operating system	Windows 10
Hardware configuration	The 13th generation Intel ® Core TM i7-1360P processor
	Intel ® Torch ® Xe graphics card
	RAM: 32 GB
	Hard disk:1TB
Software configuration	Vmware
	D435 camera driver
	Realsense Technology SDK

The main content of the test is the test of the previous module implementation content. The test is carried out according to the module using SWBSs. The main steps of the test include the serial number, test items, and test results. The test results are given below.

(1) HGR Recognition Component

The test of the HGR recognition component is whether the system can recognize the Curwen music HGR. In this experiment, the experimenter is seated in front of the D435 depth camera to make an HGR representing so. The HGR recognition result is shown in Figure 8. The three-dimensional coordinates of the 22 joint points corresponding to the HGR are shown in Table 3.

（a）Enter the HGR image　　（b）Joint point recognition results

Figure 8: HGR Recognition Test

Table 3: Position Coordinate Information of Hand Joint Points

Numbering	Name of joint	X	Y	Z
1	Fingertip of the thumb (T)	8.07338	11.3454	0
2	The interphalangeal (IP) link of the T	12.2503	11.7965	0.5
3	The metacarpal-phalangeal (MP) link of the T	16.7479	11.0169	0.5
4	Carpometacarpal joint	20.9606	8.56591	0.5
5	Fingertip of index finger (F)	-0.183599	8.2884	0
6	The distal IP link of the index F	2.31582	8.5426	0.615275
7	The proximal IP link of the index F	5.00489	8.55671	0.681822
8	The MP link of the index finger	8.64167	8.61597	0.94783
9	Fingertips of the middle F	-2.23079	6.53739	0
10	Distal IP link of the middle F	0.700602	6.30994	0.615275
11	Proximal IP link of the middle F	3.46455	6.06841	0.681822
12	The MP link of the middle F	7.79221	5.99751	0.94783
13	Fingertip of the ring F	-0.883407	4.28186	0
14	Distal IP link of ring F	2.08198	4.01626	0.615275
15	The proximal IP link of the ring F	4.89799	3.75656	0.681822
16	The MP link of the ring F	8.48336	3.5222	0.94783
17	Fingertip of the little T	2.51524	2.24725	0
18	Distal IP link of the little T	5.29119	1.59447	0.615275
19	Proximal IP link of the little T	7.46368	1.3443	0.681822
20	The MP link of the little T	9.80368	1.25689	0.94783
21	Palmar link	15.2457	5.33099	1.12625
22	Hand-wrist join	23.1983	5.23573	0.8

The data in the table shows the spatial three-dimensional coordinates of No.1 joint point, No.5 joint point, No.9 joint point, No.13 joint point, and No.17 joint point, and the Euclidean distance from each fingertip to the palm joint No.21 can be calculated. As shown in Figure 9, it represents the Euclidean distance from each fingertip of the so HGR to the palm.

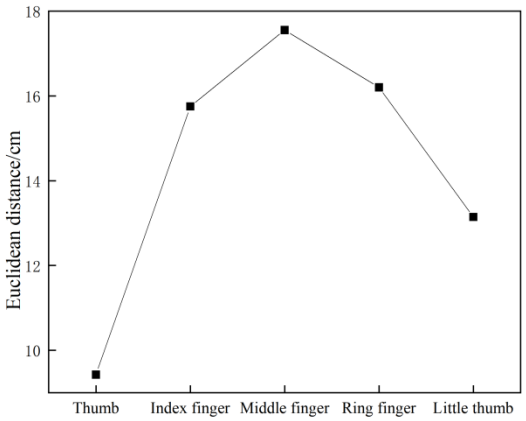

Figure 9: The Euclidean Distance between the Fingertips of the so HGR and the Palm Node

Invite 50 professional computer students to test the experiment; all the experimenters made do, re, mi, fa, so, la, si, do ' eight kinds of HGRs, each HGR left and right hand once, and finally get the experimental results, as shown in Table 4.

Table 4: Statistics of HGR Recognition Results

Hand	HGR type	Recognition times	Rate of success
Left	do	49	98%
	re	50	100%
	mi	50	100%
	fa	44	88%
	so	50	100%
	la	39	78%
	si	42	84%
	do’	48	96%
Right	do	47	94%
	re	50	100%
	mi	49	98%
	fa	45	90%
	so	50	100%
	la	40	80%
	si	38	76%
	do’	49	98%

(2) Music Composition Algorithm Test

To test the feasibility of the music composition method, the research invited a graduate student of the Conservatory of Music to use the system to conduct an HGR music composition test and make a subjective evaluation of the generated music. The system generates music, as shown in Figure 10.

The testers believe that the melody automatically generated by the algorithm in the system is closer to the fundamental composition state than the generation mode of the professional composition software, and the generated melody is not like the stiff and direct melody generated by the traditional algorithm, which has a certain fluency and professional ratio. Although there is no way to be the same as the composer's custom composition results, the composer's commonly used creative skills and routines are heard in the generated melody output.

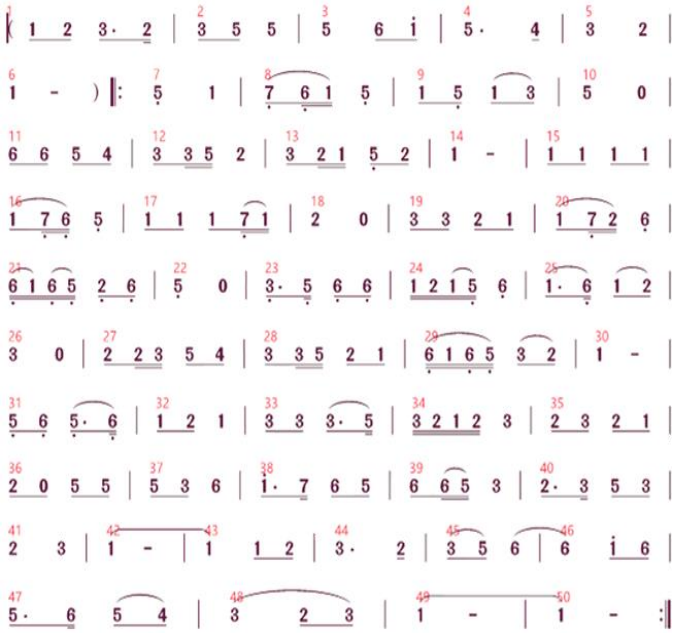

Figure 10: HGR Music Composition System Test to Generate Music Scores

5 Conclusion

Wearable systems for hand gesture detection have significantly contributed to the VR/AR interface and demonstrated significant promise in recovery, prosthetic control, sign language identification, and other domains of human-computer communication. This paper aims to reduce the obstacles that prevent non-professionals from creating music. At the same time, the human-computer interaction mode is smoother, and the composition results are closer to the professional requirements using SWBSs. An HGR music composition system based on RealSense technology and recurrent neural network is proposed as follows:

1. The Intel RealSense 3D camera captures the user's hand, and the information on the hand joint points is extracted. The hand bone's three-dimensional model is constructed using the joint points. The joint point information calculates the equivalent distance of the joint points and the joint deflection to match the corresponding Curwen HGR using SWBSs.
2. Input the matched music data into the recurrent neural network. To avoid the phenomenon of gradient disappearance or gradient explosions, the GRU is combined with the Markov chain algorithm to complete the composition.
3. the HGR recognition function and composition algorithm are simulated and tested to verify the system's effectiveness using SWBSs. The research results show that the HGR music composition system based on real-sense technology and recurrent neural networks can effectively identify the tester's HGRs and assist the tester in completing the composition. The composition mode is close to the professional composition process.

Hurdles persist in the development and subsequent marketing of these SWBS devices. The primary challenge in manufacturing these gadgets is utilizing materials and techniques that are eco-friendly, biocompatible, cost-effective, and scalable. Despite considerable advancements in biocompatible systems and technologies for detectors, several factors continue to restrict biocompatibility.

Prospective avenues for advancement encompassing expanded gesture sets, enhanced robustness, and soft technologies were examined. This study offers users a comprehensive grasp of SWBS for recognizing hand gestures.

References

[1] Abad, A. C., Reid, D., & Ranasinghe, A. (2022). A novel untethered hand wearable with fine-grained cutaneous haptic feedback. *Sensors*, *22*(5), 1924. https://doi.org/10.3390/s22051924

[2] Cai, L., & Cai, Q. (2019). Music creation and emotional recognition using neural network analysis. *Journal of Ambient Intelligence and Humanized Computing*, 1-10. https://doi.org/10.1007/s12652-019-01614-6

[3] Chen, J., & Liu, X. (2022). https://patents.google.com/patent/CN114489331A/zh

[4] Deng, X., Huang, P., Luo, J., Wang, J., Yi, L., Yang, G., & Yang, G. (2020). The consistency of an optical body surface scanning method compared with computed tomography: a validation study. *Journal of Pediatric Surgery*, *55*(8), 1448-1452. https://doi.org/10.1016/j.jpedsurg.2019.07.015

[5] Ding, J., & Chang, C. W. (2015). An eigenspace-based method with a user adaptation scheme for human gesture recognition by using Kinect 3D data. *Applied Mathematical Modelling*, *39*(19), 5769-5777. https://doi.org/10.1016/j.apm.2014.12.054

[6] Dixon, S., Goebl, W., & Widmer, G. (2005, September). The" Air Worm": an Interface for Real-Time manipulation of Expressive Music Performance. In *ICMC*, 5, 614-617.

[7] Gentry, K. (2016). *Music teachers' reported use of kinesthetic approaches to teach pitch matching and reading to elementary music students*, Master's thesis, Southern Methodist University.

[8] Gunawan, A. A. S., Iman, A. P., & Suhartono, D. (2020). Automatic music generator using recurrent neural network. *International Journal of Computational Intelligence Systems*, *13*(1), 645-654.

[9] Hannum, A. (2018). *RNN-Based Generation of Polyphonic Music and Jazz Improvisation*, Master's thesis, University of Denver.

[10] Hashimoto, J. (2010). https://patents.google.com/patent/JP2010175870A

[11] Hinojosa Chapel, R. (2004). Some projects and reflections on algorithmic Music. In *Computer Music Modeling and Retrieval: International Symposium, CMMR 2003*, 1-12.

[12] Jeeru, S., Sivapuram, A. K., León, D. G., Gröli, J., Yeduri, S. R., & Cenkeramaddi, L. R. (2022). Depth camera based dataset of hand gestures. *Data in Brief*, *45*, 108659. https://doi.org/10.1016/j.dib.2022.108659

[13] Jiang, J. (2022). Using pitch feature matching to design a music tutoring system based on deep learning. *Computational Intelligence and Neuroscience*, *2022*(1), 4520953. https://doi.org/10.1155/2022/4520953

[14] Jung, Y. H., Kim, J. H., & Rogers, J. A. (2021). Skin-integrated vibrohaptic interfaces for virtual and augmented reality. *Advanced Functional Materials*, *31*(39), 2008805. https://doi.org/10.1002/adfm.202008805

[15] Kawamura, M., & Yokota, M. (2005). https://patents.google.com/patent/JP2005321514A.

[16] Koerner, L. J. (2021). Models of direct time-of-flight sensor precision that enable optimal design and dynamic configuration. *IEEE Transactions on Instrumentation and Measurement*, *70*, 1-9. https://doi.org/10.1109/TIM.2021.3073684

[17] Kwon, Y., Dwivedi, A., McDaid, A. J., & Liarokapis, M. (2021). Electromyography-based decoding of dexterous, in-hand manipulation of objects: Comparing task execution in real world and virtual reality. *IEEE Access*, *9*, 37297-37310.

[18] Li, H. (2020). Piano automatic computer composition by deep learning and blockchain technology. *IEEE Access*, *8*, 188951-188958. https://doi.org/10.1109/ACCESS.2020.3031155

[19] Li, H., Tan, P., Rao, Y., Bhattacharya, S., Wang, Z., Kim, S., ... & Lu, N. (2024). E-Tattoos: Toward Functional but Imperceptible Interfacing with Human Skin. *Chemical Reviews*, *124*(6), 3220-3283.

[20] Li, J., Zhu, K., & Pan, L. (2022). Wrist and finger motion recognition via M-mode ultrasound signal: A feasibility study. *Biomedical Signal Processing and Control*, *71*, 103112. https://doi.org/10.1016/j.bspc.2021.103112

[21] Liang, M. (2022). Liang, M. (2022). An improved music composing technique based on neural network model. *Mobile Information Systems*, *2022*(1), 7618045. https://doi.org/10.1155/2022/7618045

[22] Liu, A., Guo, J., Han, B., & Xiao, J. (2021). https://patents.google.com/patent/CN113010730A.

[23] Lu, J. (2020). *An evaluation of generated lyrics*, Master's thesis, San Jose State University.

[24] Mandanici, M., & Canazza, S. (2014). The" hand composer": Gesture-driven music composition machines. In *Proc. of 13th Intl. Conf. on Intelligent Autonomous Systems*, 15-19.

[25] Masoumian Hosseini, M., Masoumian Hosseini, S. T., Qayumi, K., Hosseinzadeh, S., & Sajadi Tabar, S. S. (2023). Smartwatches in healthcare medicine: assistance and monitoring; a scoping review. *BMC Medical Informatics and Decision Making*, *23*(1), 248. https://doi.org/10.1186/s12911-023-02350-w

[26] Medeiros Esper, I., Cordova-Lopez, L. E., Romanov, D., Alvseike, O., From, P. J., & Mason, A. (2022). Pigs: A stepwise RGB-D novel pig carcass cutting dataset. *Data in Brief*, *41*, 107945.

[27] Pelchat, N., & Gelowitz, C. M. (2020). Neural network music genre classification. *Canadian Journal of Electrical and Computer Engineering*, *43*(3), 170-173.

[28] Quintero, C., & Roa, D. (2022). Musical approach through gestural interpretation. *TecnoLógicas*, *25* (53), e2131. https://doi.org/10.22430/22565337.2131

[29] Rong, H. (2012). https://patents.google.com/patent/CN102332279A

[30] Siradjev, D., Gurin, I., & Kim, Y. T. (2006). Scalable DiffServ-over-MPLS traffic engineering with per-flow traffic policing. In *Management of Convergence Networks and Services: 9th Asia-Pacific Network Operations and Management Symposium, APNOMS 2006,* 509-512.

[31] Tanaka, A. (2000). Musical performance practice on sensor-based instruments. *Trends in gestural control of music*, *13*(389-405), 284.

[32] Waisvisz, M. (1985). The Hands: A Set of Remote MIDI-Controllers. https://quod.lib.umich.edu/i/icmc/bbp2372.1985.049/1

[33] Wang, N., Xu, H., Xu, F., & Cheng, L. (2021). The algorithmic composition for music copyright protection under deep learning and blockchain. *Applied Soft Computing*, *112*, 107763. https://doi.org/10.1016/j.asoc.2021.107763

[34] Xia, J. (2022). Construction and implementation of music recommendation model utilising deep learning artificial neural network and mobile edge computing. *International journal of grid and utility computing*, *13*(2-3), 183-194.

[35] Yu, Y., & Kwak, B. M. (2011). Design sensitivity analysis of acoustical damping and its application to design of musical bells. *Structural and Multidisciplinary Optimization*, *44*, 421-430.