Motion Feature Extraction and Stylization for Character Animation using Hilbert-Huang Transform

Ran Dong
Tokyo University of Technology
Tokyo, Japan
randong@stf.teu.ac.jp

Yangfei Lin*
University of Tsukuba
Ibaraki, Japan
linyangfei@osdp.cs.tsukuba.ac.jp

Qiong Chang
Tokyo Institute of Technology
Tokyo, Japan
q.chang@c.titech.ac.jp

Junpei Zhong
The Hong Kong Polytechnic University
Hongkong, China
joni.zhong@polyu.edu.hk

Dongsheng Cai
University of Tsukuba
Ibaraki, Japan
cai@cs.tsukuba.ac.jp

Soichiro Ikuno
Tokyo University of Technology
Tokyo, Japan
ikuno@stf.teu.ac.jp

ABSTRACT

This paper presents novel insights to feature extraction and stylization of character motion in the instantaneous frequency domain by proposing a method using the Hilbert-Huang transform (HHT). HHT decomposes human motion capture data in the frequency domain into several pseudo monochromatic signals, so-called intrinsic mode functions (IMFs). We propose an algorithm to reconstruct these IMFs and extract motion features automatically using the Fibonacci sequence in the link-dynamical structure of the human body. Our research revealed that these reconstructed motions could be mainly divided into three parts, a primary motion and a secondary motion, corresponding to the animation principles, and a basic motion consisting of posture and position. Our method help animators edit target motions by extracting and blending the primary or secondary motions extracted from a source motion. To demonstrate results, we applied our proposed method to general motions (jumping, punching, and walking motions) to achieve different stylizations.

CCS CONCEPTS

• **Computing methodologies → Motion capture**; **Motion processing**.

KEYWORDS

feature extraction, motion stylization, Hilbert-Huang transform, biomechanics, deep learning

ACM Reference Format:
Ran Dong, Yangfei Lin, Qiong Chang, Junpei Zhong, Dongsheng Cai, and Soichiro Ikuno. 2021. Motion Feature Extraction and Stylization for Character Animation using Hilbert-Huang Transform. In *2021 ACM International Conference on Intelligent Computing and its Emerging Applications (ACM ICEA '21), December 28–29, 2021, Jinan, China.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3491396.3506524

*Corresponding author

ACM ICEA '21, December 28–29, 2021, Jinan, China

ACM ISBN 978-1-4503-9160-3/21/12.
https://doi.org/10.1145/3491396.3506524

1 INTRODUCTION

In recent years, motion capture data have been used in various fields. Motion capture systems can obtain positions and rotations of human joints based on a hierarchical structure. Many types of research have been on motion features extraction and stylization using motion capture data based on the human hierarchical structure in the time domain. However, methods for editing these motion features in the frequency domain are desirable for complicated and detailed features that are difficult to extract and stylize in the time domain.

Fourier transform (FT) is known as that it can be applied to decompose and analyze signals in the frequency domain [2]. However, since FT linearly decomposes data into monochromatic waves, motion features can only be represented by a large number of decomposed monochromatic waves (modes). Although Wavelet transform (WT) presented a method to decompose signals in a smaller number of modes based on Gaussian function than FT [15], nonlinear data such as motion features can not be easy to edit as WT is still using window functions without considering dynamic meaning for the human hierarchical structure. Meanwhile, Huang et al. [11] provided the empirical mode decomposition (EMD) without using any window function to decompose data. Then, EMD can obtain decomposed nonlinear modes corresponding to nonlinear properties of data better than WT. EMD decomposes nonlinear and nonstationary signals into pseudo monochromatic waves known as intrinsic mode functions (IMFs), and a residual (no frequency components) called "trend." Furthermore, decomposed IMFs are pseudo monochromatic waves that are suitable for the Hilbert transform (HT) [2]. Then, we can obtain instantaneous frequencies and amplitudes of each IMF [10], which can be used to analyze motion features in the instantaneous frequency domain.

In this paper, to help animators extract and stylize complicated motion features in the instantaneous frequency domain, we present a motion feature extraction and stylization method using HHT and Fibonacci sequence relationships in the human hierarchical structure. Our proposed framework decomposes the motion capture data into three different motion groups: (i) A primary motion consists of

lower frequency and stronger amplitude motions; (ii) A secondary motion consists of higher frequency and weaker amplitude motions; and (iii) A basic motion consists of posture and position. As a result, extracted primary and secondary motions can be added, replaced, exchanged, removed, and blended into other target basic motions to achieve stylization automatically. Thus, animators can use our proposed method to efficiently extract and stylize desired motions. In summary, the contributions of this study are summarized as follow:

- A algorithm based on the Fibonacci sequence relationship in the human joint velocity to extract motion features.
- A method to edit extracted motion features (primary and secondary motion) into the basic motion of other target motions for stylization.
- Training data sets in the instantaneous frequency domain for deep learning.

2 RELATED WORKS

Recently, more and more researches presented to extract features and generate motions using machine learning. For example, Aberman et al. [1] researched motion retargeting with the learning of character-agnostic movements. Furthermore, the deep learning method is getting more and more attention as a state-of-the-art technology, which can generate high-quality motion data better than traditional methods. Pavllo et al. [14] introduced a recurrent neural network that represents rotations using quaternions instead of Euler angle to generate motion in both long-term and short-term. Holden et al. [8] presented a one-dimensional convolutional autoencoder to find a motion manifold, which can be applied to fix corrupted human motion data. Furthermore, Holden et al. [7] also proposed a deep learning framework that generates locomotions using only high-level parameters based on the motion manifold. However, these deep learning methods have a disadvantage in that each proposal needs massive specific training data whose features are already known. For example, the previous study using deep learning like [7, 8] used massive training data to generate locomotions such as walking and jumping. The main weakness of these studies is painful to collect enough appropriate training data for complicated and advanced motions since it is difficult to label and classify these nuanced motion features manually.

Meanwhile, to deal with these complicated motion features, Dong et al. [4] proposed a framework to edit character dance motion using HHT in the instantaneous frequency domain. Moreover, general human motions also can be applied to HHT, giving in-depth biomedical knowledge in human movements [5]. However, although these previous researches showed that HHT could decompose human motions into several IMFs in the instantaneous frequency domain, they could not give dynamic meaning to these decomposed IMFs due to over-decomposed [9]. As a result, these IMFs need to be reconstructed and recombined manually, preventing the spread of the method. Although Dong et al. [6] revealed that decomposed nonlinear modes (IMFs) could be treated as motion training data, learning by a neural network to generate nuanced motion features for robots automatically, dynamic meanings of decomposed IMFs are still not addressed based on human hierarchical human models.

In this paper, to challenge this problem and provide training data with nuanced human motion features, we propose a novel method for character animations using the Hilbert-Huang transform and Fibonacci sequence relationships regarding the human body structure to extract and stylize motion features automatically. Using our proposed method, we can nonlinearly and separately extract motion features into primary, second, and basic motion that corresponds to their biomechanical properties [16] using the Fibonacci sequence. These extracted motion data can be edited into another motion data for achieving stylization. In addition, these decomposed motion features also can be thought of as training data to train neural networks.

3 PROPOSED METHOD

This section proposes a method to extract and stylize motions using HHT regarding Fibonacci sequences in the human hierarchical structure. First, to discuss the relationship among decomposed motions, IMFs, we decompose a jumping motion [3] to explain how primary, secondary motions, and a basic motion can be extracted based on the Fibonacci sequence. Second, we propose an algorithm reconstructing IMFs to primary, secondary, and basic motions.

3.1 Fibonacci sequence in the human hierarchical structure

In this study, we adopt hierarchical models (3 Euler angles for each joint: $\theta x, \theta y, \theta z$) as the input data according to the previous research [5]. If we apply MEMD to motion data and plot the data as a Hilbert-spectrum in the frequency domain, each instantaneous frequency of IMFs will correspond to human joints. We put a jumping motion to explain the Fibonacci sequence in human joint systems for simplicity. Figure 1 shows the Hilbert-spectrum of the jumping motion. In this jumping motion, we adopt θ_z to demonstrate the decomposed result. We can see that the jumping starts at 1.5 seconds and ends at 2.5 seconds. In this research, we calculate each $F_i(t)$ using WAFA [13]. Surprisingly, their average instantaneous frequencies f_i (blue and green on the right of the figure) are approximately in Fibonacci sequence order $f_{i+1} \sim f_i + f_{i-1}$ (the error is about $\pm 10\%$), separately.

Next, we example the reason for Fibonacci relation in each IMF. As shown in the Figure 2, One of the human joints is represented by i. Then, $i-1$ is defined as the parent joint of i, and $i+1$ is defined as the child joint of i. Because decomposed IMFs are only vibrations on the origin (usually T-pose) without the posture of motion, angular velocities (frequencies) of IMFs can be represented by ω_i, ω_i, and ω_{i+1}. Thus, frequencies of each IMF from the same motion are satisfied (1).

$$\mathrm{IMF}_{\omega_{i+1}} = \mathrm{IMF}_{\omega_i} + \mathrm{IMF}_{\omega_{i-1}} \tag{1}$$

Then, the frequency relationship of each IMF shown in Figure 2 can also be assumed from root to hands, foot, and head, according to human hierarchical structures. Consequently, when the hierarchical human body performs different motions, the average frequency of IMFs form distinct pseudo-Fibonacci sequences.

According to the Fibonacci sequence relationship, our research reveals that human motions are divided into three groups. One is a set of lower frequency IMFs that is the primary motion, and another

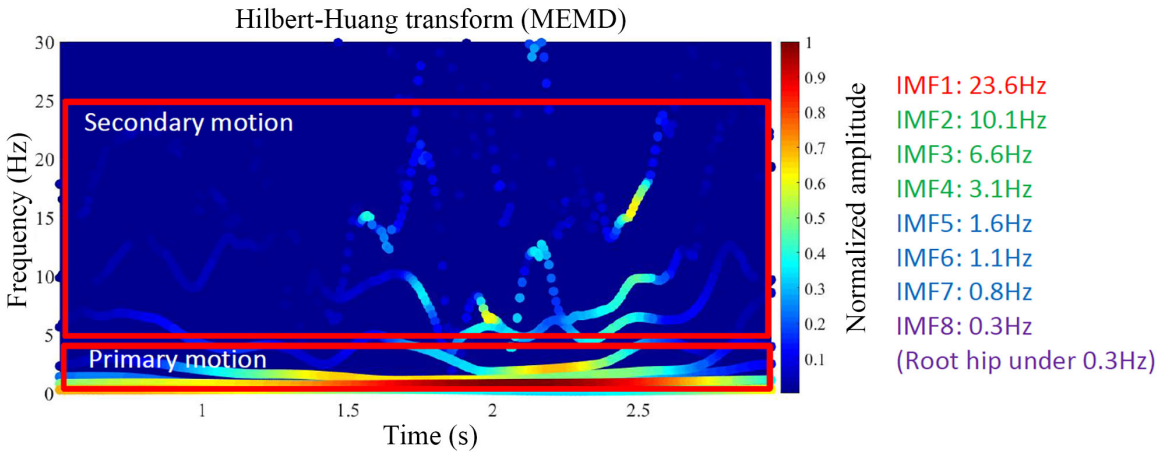

Figure 1: MEMD Hilbert spectrum of a jumping motion (hip θZ).

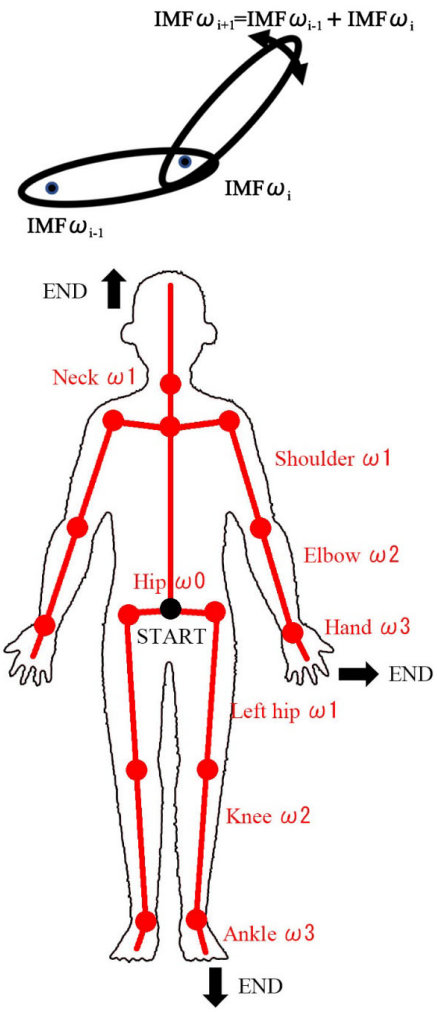

Figure 2: Fibonacci relation in human body link-dynamic system.

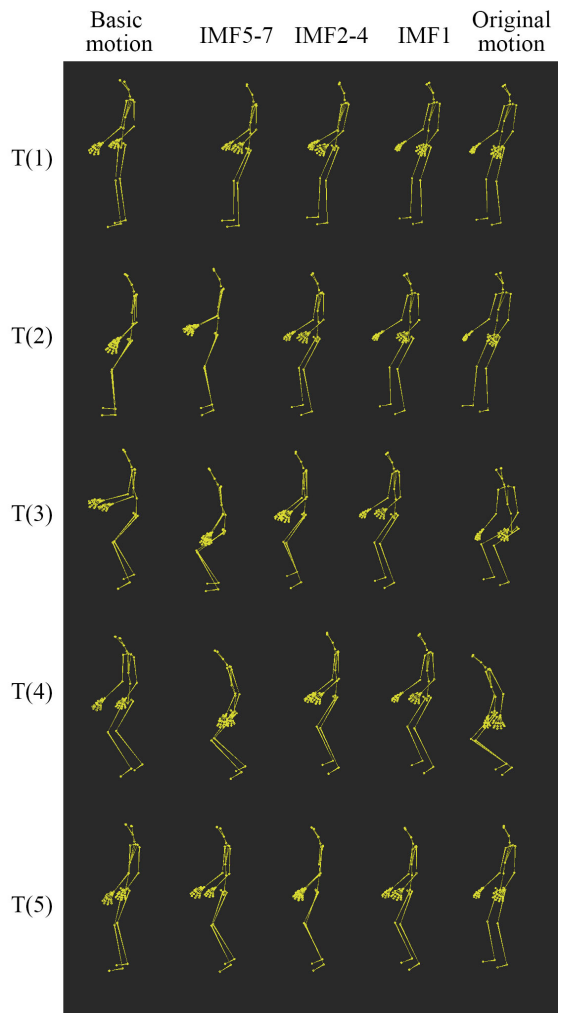

Figure 3: The jumping motion decomposed result using the Fibonacci sequence.

one is a set of higher frequency IMFs that is the secondary motion. The remaining IMFs with lower frequency than hip joints are combined as a basic motion. For example, in the Hilbert spectrum of the jumping motion shown in Figure 1, the Fibonacci sequence relationships are observed from IMF_2 to IMF_4, and from IMF_5 to IMF_7. Thus, this jumping motion can be reconstructed into primary motion (IMF_{5-7}), secondary motion (IMF_{2-4}), and basic motion (IMF_8 with trend). In addition, because these decomposed IMFs are only vibrations, we need to add trend (posture) motion into each IMFs group.

Figure 3 shows the result of each motion group of IMFs. IMF_1 is the noise during the motion capturing. IMF_{2-4} (secondary motion) is the reaction force generated by foot hitting the ground. IMF_{5-7} (primary motion) is the main jumping made by the human. This result indicates that the jumping motion mainly consisted of two different motions. Under hip joint motion speed, IMF_8 should be added into the trend as the basic motion consisting of the position of the root joint. Other primary motion or secondary motion can be blended into this basic motion for stylization. Thus, using primary and secondary motion based on the Fibonacci sequence, users can easily extract and blend motion features from a source motion and blend them into target basic motion for different stylizations.

3.2 IMFs reconstruction algorithm using the Fibonacci sequence

To make decomposed motions with dynamic meaning for stylization, we propose an algorithm to reconstruct IMFs based on the Fibonacci sequence. Our proposed algorithm can rebuild and combine corresponding IMFs into three groups: (a) primary motion, (b) secondary motion, and (c) basic motion. Hence, the decomposed IMFs using MEMD are defined as (2).

$$\sum_{i=1}^{n} \mathrm{IMF}_i(t) + \mathrm{Trend}(t) = P(t) + S(t) + B(t) \tag{2}$$

Here, $\mathrm{IMF}_i(t)$ and $\mathrm{Trend}(t)$ are decomposed from the original motion using MEMD. $P(t)$ is a motion set reconstructed from several IMFs with average frequencies $F_i(t)$ according to the Fibonacci sequence relationship. $S(t)$ is a motion set reconstructed from residual IMFs that do not satisfy the Fibonacci sequence when extracting $P(t)$. $B(t)$ is the basic motion consisting of IMFs with frequency lower than root joint, representing positions and postures of original motion data.

Since the human body's hierarchical skeleton is usually defined as several joints, with each joint having 3 Euler angles, in our proposed method, we check the Fibonacci sequence relationship in each joint Euler angle, respectively. Then, each joint and each Euler angle has its own average frequencies $F_i(t)$. In order to edit all joints to obtain the desired motion, it is necessary to calculate all of them to decompose motion data based on the Fibonacci sequence. However, Because human motions concentrate on the hip joint, for simplicity, we only chose the root (hip joint) IMF to demonstrate the motion reconstruction as shown in Figure 4 and algorithm 1.

The same algorithm can be applied to all other joints to reconstruct the entire body joint movement. Since the extracted IMFs (the basic motion $B(t)$) under the average frequency of the hip include motion positions and postures, we can achieve stylization

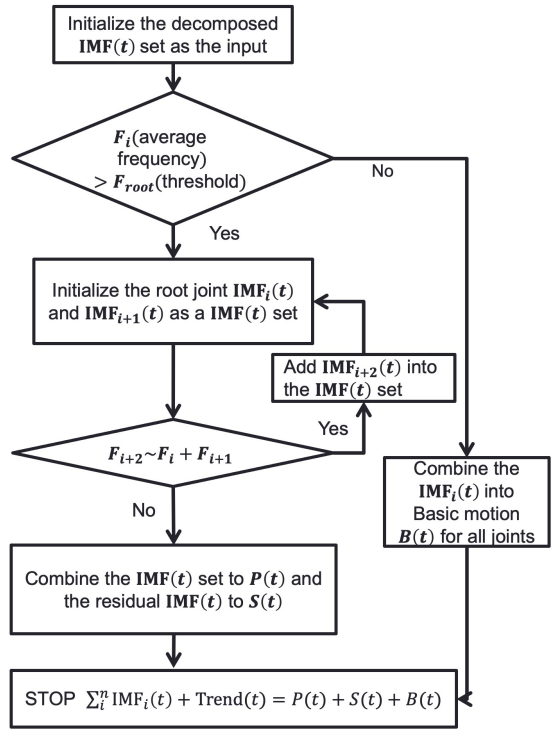

Figure 4: Proposed algorithm to reconstruct IMFs based on the Fibonacci sequence

Algorithm 1 IMFs reconstruction based on the Fibonacci sequence

1: Decompose motion data using MEMD to obtain $\mathrm{IMF}_n(t)$ of the hip joint.
2: Calculate average frequency of each decomposed $\mathrm{IMF}_n(t)$.
3: Obtain the basic motion $B(t)$ by reconstructing IMFs with average frequencies lower than the root joint.
4: Reconstruct the primary and secondary motions:
   (1) Initialize $\mathrm{IMF}_i(t)$ and IMF_{i+1} (t) from the lowest IMF average frequency as the first two Fibonacci sequence components and add them to $\mathrm{IMF}(t)$ set.
   (2) Add IMF_{i+2} to the $\mathrm{IMF}(t)$ set if the average frequencies of IMFs satisfy $f_{i+2} \sim f_i + f_{i+1}$ (error ±10%).
   (3) Repeat the above procedure until the average frequency do not satisfy $f_{i+2} \sim f_i + f_{i+1}$.
5: Combine the $\mathrm{IMF}(t)$ set as the primary motion $P(t)$, and combine the residual $\mathrm{IMF}(t)$ set as the secondary motion $S(t)$.

by blending and editing $P(t)$ and $S(t)$ as the primary motion and secondary motion into $B(t)$ of target motions.

3.3 Advantages and limitations

As we discussed in the introduction, FT decompose data $x(t)$ linearly as $x(t) = \mathrm{Re}\left[\sum_{j=1}^{n} a_j \exp(i\omega_j t)\right]$, where n is decomposed mode number, amplitude a_j and frequency ω_j are constant. On

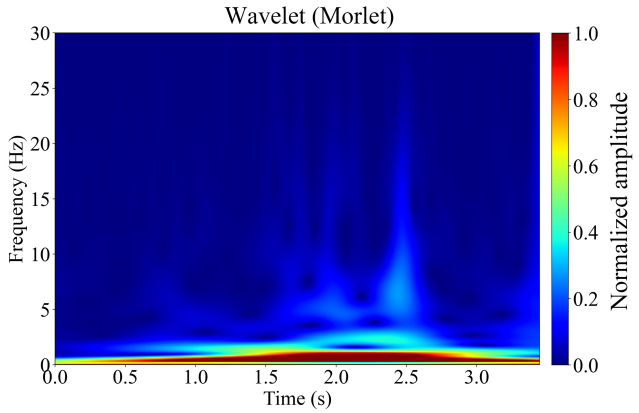

Figure 5: Wavelet spectrum of the jumping motion (hip θZ).

the contrary, since EMD decompose data $x(t)$ nonlinearly, we can obtain $x(t) = \mathrm{Re}\left\{\sum_{j=1}^{n} a_j(t) \exp\left[i \int \omega_j(t)dt\right]\right\}$ by applying HT to each IMF, where amplitude a_j and frequency ω_j are variables. Obviously, HHT provides better performance than FT in motion data feature analysis.

Another helpful analysis method in the frequency domain is WT. FT assumes that decomposed monochromatic waves are infinity in the time series while WT adopts window function based on Gaussian function to achieve better results of catching nonlinear properties. Previous research demonstrated that the Morlet wavelet showed high performance in feature extractions for audios [12]. For comparison, we performed a continuous Morlet wavelet (ω_0 = 0.8) to the jumping motion. As shown in figure 5, WT can also catch the detailed features of the jumping motion. Higher frequency components (0 - 10 Hz) are detected during 1.5 seconds (jump-start) to 2.5 seconds (jump-end). However, since WT also employs a window function based on a Gaussian function to represent data, it is hard to decompose nonlinear motion data into several nonlinear modes and calculate Fibonacci sequence relationships among human joints like HHT.

For the limitations, there is no guarantee that the motions can be decomposed mathematically correct as the empirical mode decomposition is not mathematically proven. Thus, our method needs to be confirmed and tested manually when editing motions.

4 RESULTS

In this section, we applied our proposed method to two types of general motion [3] for archiving stylization using Fibonacci sequence relationships in the human hierarchical structure. Our method extracted different features from the primary motion and the secondary motion. Then, we selected two different motions to show the stylization results using primary or secondary motions, respectively.

4.1 A walking motion stylization using the primary action

Walking motion is one of the simplest motions among general motions. All type of walking motions almost has their features in the primary motion. This example extracted the primary motion from a foot injured walking motion and stylized the features into a regular walking. Our algorithm can decompose the feet injured walking into three groups. A secondary motion (IMF_{1-2}), a primary motion (IMF_{3-6}), and a basic motion (Trend). This walking motion has the injured feature in the primary motion.

Then, we blended the injured motion feature (primary motion) into the regular walking for stylization. Figure 6 (a) shows the result of the injured walking stylized motion. After we blended primary motion (IMF_{3-6}) into the basic motion extract from the regular walking motion, we obtained a new injured walking stylized by the source motion. The stylized motion has the same posture and position as the target motion, which cannot be achieved by directly blending the Euler rotation angles into the original motion.

4.2 A punching motion stylization using the secondary action

Next, for more complicated motion such as a punching motion performed by a professional boxer, the boxer-style feature is in the secondary motion. We adopted our method to a punching motion to extract the boxer-style features based on the Fibonacci sequence. IMF_{5-7} is the main feature of the punching extracted in the primary motion. IMF_{2-4} is the motion features of the professional boxer punching style extracted in the secondary motion.

Next, we blend the extracted professional boxer punching feature (secondary motion) into an untrained person punching. As we can see from figure 6 (b), two target motions, 1 and 2, are the punching motions performed by an untrained person. We cannot feel any powerful punching like the professional boxer showed. Then, after we blended extracted secondary motion (IMF_{2-4}), we can see two different target punching motions have been stylized as the same powerful taste as the professional boxer. Additionally, our method also can be used for motion sparsity by removing the secondary motion (e.g., IMF_{2-4}) from the professional boxer to achieve destylization.

5 CONCLUSION

This paper has proposed a framework for character motion feature extraction and stylization using Hilbert-Huang transform regarding the Fibonacci sequence in the human body structure. Our research has revealed that the human link-dynamical structure in character motions forms Fibonacci sequence relationships in joint velocity. By using our proposed framework based on this Fibonacci sequence relationship, animators can extract and stylize character motions. The results have shown that two general and advanced motions were edited self-consistently using these features.

Although motion features can be decomposed into primary and secondary motions using our method, animators still need to confirm the extracted features before stylizing them. Using deep learning to automatically edit and stylize these primary and secondary motions is the future work.

ACKNOWLEDGMENTS

This work was supported by JSPS KAKENHI Grant Number JP20K23352, JP21K17833, JP21K17868.

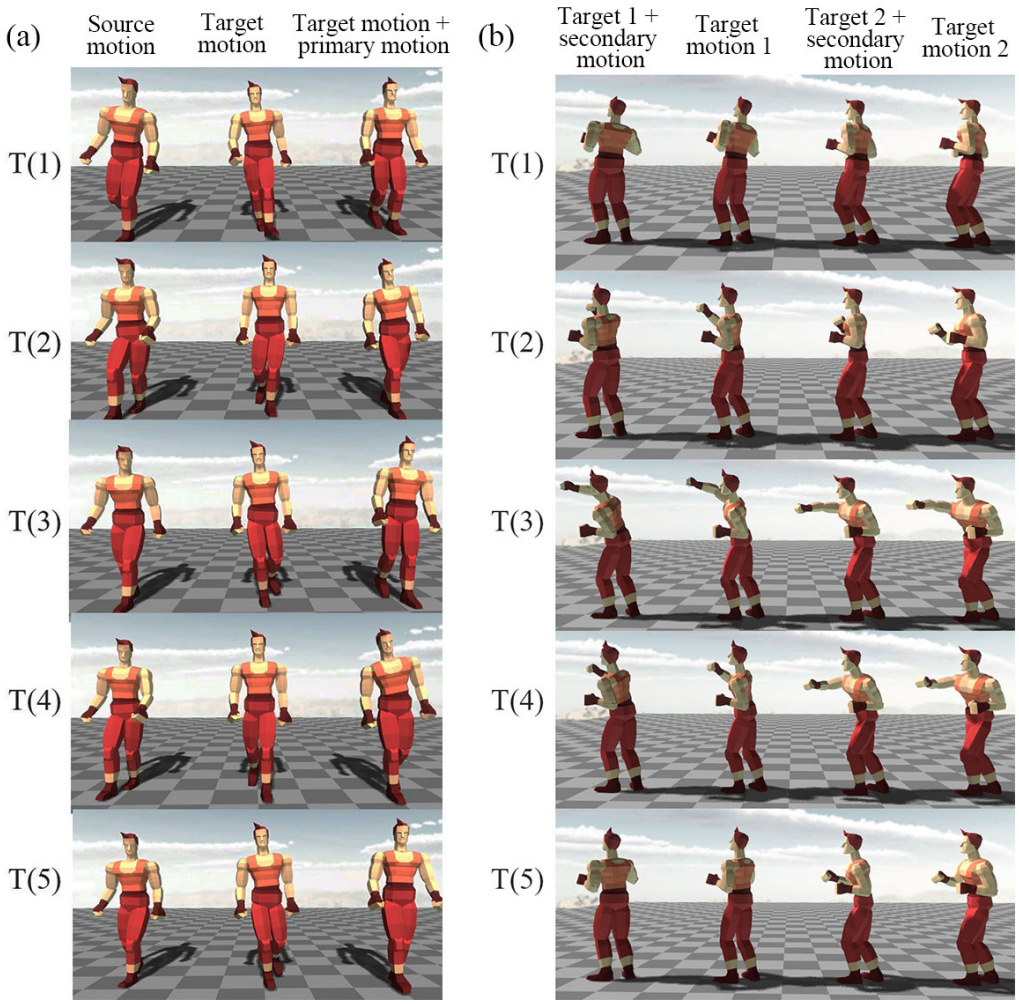

Figure 6: (a) Stylization result of blending walking with injured foot motion feature (primary motion) into a regular walking motion. (b) Blending professional boxer motion feature (secondary motion) into two different punching motions captured from an untrained person.

REFERENCES

[1] Kfir Aberman, Rundi Wu, Dani Lischinski, Baoquan Chen, and Daniel Cohen-Or. 2019. Learning character-agnostic motion for motion retargeting in 2d. *arXiv preprint arXiv:1905.01680* (2019).

[2] Ronald Newbold Bracewell and Ronald N Bracewell. 1986. *The Fourier transform and its applications*. Vol. 31999. McGraw-Hill New York.

[3] CMU. [n. d.]. BVH conversions of the 2500-motion Carnegie-Mellon motion capture dataset. ([n. d.]). Retrieved May 20, 2019 from https://sites.google.com/a/cgspeed.com/cgspeed/motion-capture

[4] Ran Dong, Dongsheng Cai, and Nobuyoshi Asai. 2017. Nonlinear dance motion analysis and motion editing using Hilbert-Huang transform. In *Proceedings of the computer graphics international conference*. 1–6.

[5] Ran Dong, Dongsheng Cai, and Soichiro Ikuno. 2020. Motion capture data analysis in the instantaneous frequency-domain using hilbert-huang transform. *Sensors* 20, 22 (2020), 6534.

[6] Ran Dong, Qiong Chang, and Soichiro Ikuno. 2021. A deep learning framework for realistic robot motion generation. *Neural Computing and Applications* (2021), 1–14.

[7] Daniel Holden, Jun Saito, and Taku Komura. 2016. A deep learning framework for character motion synthesis and editing. *ACM Transactions on Graphics (TOG)* 35, 4 (2016), 1–11.

[8] Daniel Holden, Jun Saito, Taku Komura, and Thomas Joyce. 2015. Learning motion manifolds with convolutional autoencoders. In *SIGGRAPH Asia 2015 Technical Briefs*. 1–4.

[9] Jianzhao Huang, Jian Xie, Feng Li, and Liang Li. 2013. A threshold denoising method based on EMD. *Journal of Theoretical and Applied Information Technology* 47, 1 (2013), 419–424.

[10] Norden Eh Huang. 2014. *Hilbert-Huang transform and its applications*. Vol. 16. World Scientific.

[11] Norden E Huang, Zheng Shen, Steven R Long, Manli C Wu, Hsing H Shih, Quanan Zheng, Nai-Chyuan Yen, Chi Chao Tung, and Henry H Liu. 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences* 454, 1971 (1998), 903–995.

[12] Jing Lin and Liangsheng Qu. 2000. Feature extraction based on Morlet wavelet and its application for mechanical fault diagnosis. *Journal of sound and vibration* 234, 1 (2000), 135–148.

[13] J Niu, Y Liu, W Jiang, X Li, and G Kuang. 2012. Weighted average frequency algorithm for Hilbert-Huang spectrum and its application to micro-Doppler estimation. *IET Radar, Sonar & Navigation* 6, 7 (2012), 595–602.

[14] Dario Pavllo, David Grangier, and Michael Auli. 2018. Quaternet: A quaternion-based recurrent model for human motion. *arXiv preprint arXiv:1805.06485* (2018).

[15] Christopher Torrence and Gilbert P Compo. 1998. A practical guide to wavelet analysis. *Bulletin of the American Meteorological society* 79, 1 (1998), 61–78.

[16] David A Winter. 2009. *Biomechanics and motor control of human movement*. John Wiley & Sons.