information

Article

Multiple Visual Feature Integration Based Automatic Aesthetics Evaluation of Robotic Dance Motions

Hua Peng [1,2], Jinghao Hu [1], Haitao Wang [1], Hui Ren [1], Cong Sun [1], Huosheng Hu [3] and Jing Li [4,*]

[1] Department of Computer Science and Engineering, Shaoxing University, Shaoxing 312000, China; penghua_47@163.com (H.P.); 18145104@usx.edu.cn (J.H.); 18145118@usx.edu.cn (H.W.); 18145315@usx.edu.cn (H.R.); 18145316@usx.edu.cn (C.S.)
[2] College of Information Science and Engineering, Jishou University, Jishou 416000, China
[3] School of Computer Science and Electronic Engineering, University of Essex, Colchester CO4 3SQ, UK; hhu@essex.ac.uk
[4] Academy of Arts, Shaoxing University, Shaoxing 312000, China
* Correspondence: lijing_47@126.com; Tel.: +86-0575-8834-2988

Abstract: Imitation of human behaviors is one of the effective ways to develop artificial intelligence. Human dancers, standing in front of a mirror, always achieve autonomous aesthetics evaluation on their own dance motions, which are observed from the mirror. Meanwhile, in the visual aesthetics cognition of human brains, space and shape are two important visual elements perceived from motions. Inspired by the above facts, this paper proposes a novel mechanism of automatic aesthetics evaluation of robotic dance motions based on multiple visual feature integration. In the mechanism, a video of robotic dance motion is firstly converted into several kinds of motion history images, and then a spatial feature (ripple space coding) and shape features (Zernike moment and curvature-based Fourier descriptors) are extracted from the optimized motion history images. Based on feature integration, a homogeneous ensemble classifier, which uses three different random forests, is deployed to build a machine aesthetics model, aiming to make the machine possess human aesthetic ability. The feasibility of the proposed mechanism has been verified by simulation experiments, and the experimental results show that our ensemble classifier can achieve a high correct ratio of aesthetics evaluation of 75%. The performance of our mechanism is superior to those of the existing approaches.

Keywords: robotic dance motion; machine aesthetics; visual understanding; motion history image; ensemble learning

Citation: Peng, H.; Hu, J.; Wang, H.; Ren, H.; Sun, C.; Hu, H.; Li, J. Multiple Visual Feature Integration Based Automatic Aesthetics Evaluation of Robotic Dance Motions. *Information* **2021**, *12*, 95. https://doi.org/10.3390/info12030095

Academic Editor: Gholamreza Anbarjafari

Received: 16 January 2021
Accepted: 19 February 2021
Published: 24 February 2021

1. Introduction

As a good breakthrough point of artificial intelligence research, robotic dance is widely used to explore and develop a robot's autonomous ability, interaction ability, imitation ability, and coordination ability with the environment [1–3]. Robotic dance motion, which expresses the movement path of a robotic body from two dimensions of space and time, is the fundamental part of robotic dance [4]. Meanwhile, a good robotic choreography requires that its internal dance objects (such as dance pose, dance motion, etc.) need to be in accordance with human aesthetics [3]. Thus, any robotic dance motion has its own aesthetic attribute and constraint.

Imitation of human behaviors is one of the effective ways to develop artificial intelligence [5–7]. Human dancers always present dance motions before a mirror, visually observe the mirror reflections of their own dance motions, and finally make aesthetic judgments about those motions. Similarly, if a robot perceives the aesthetics of its own dance motions just like this, it expresses more autonomous, humanoid behavior [3] and, to a certain extent, develops machine consciousness [8]. Therefore, it is meaningful to explore the self-aesthetics of robotic dance motions. This paper explores the following key problem: How can a robot achieve an automatic aesthetic evaluation of its own dance motions, using only its visual information?

In the field of robotic dance, many researchers have explored the aesthetic problems of robotic dance objects (such as dance pose, dance motion, and dance works). However, the aesthetic method of robotic dance motion, which draws lessons from the mature aesthetic experiences of human beings, is still rarely studied.

Introducing human subjective aesthetics directly to the aesthetic evaluation of robotic dance objects is a simple and efficient method, but it inevitably brings a heavy burden to human evaluators. Vircikova et al. [9–11] designed robotic choreography by using interactive evolutionary computation based on robotic dance pose and robotic dance motion. Shinozaki et al. [12] constructed a robot dance system of hip-hop, and invited ten people to evaluate the generated robotic dance motions from some specific items (such as dynamic, exciting, wonder, smooth, etc.). Moreover, Oliveira et al. [13] implemented a real-time robot dancing framework based on multimodal events, and each evaluator was asked to fulfill a Likert-scale questionnaire to make an empiric evaluation of robotic dance.

Furthermore, an automatic system for robotic dance creation based on a hidden Markov model (HMM) was proposed by Manfrè et al. [14], and then it was introduced into an improvisational robotic dance system based on human–robot interaction [15] and a computational creativity framework of robotic dance based on demonstration learning [16]. Finally, all the robotic dances, created by the above three ways, were judged aesthetically by human evaluators. Moreover, Qin et al. [17] proposed a humanoid robot dance system driven by musical structures and emotions, and asked twenty people to fill in questionnaires to make aesthetics evaluations of robotic dances.

Drawing on some principles in the theory of dance aesthetics seems to be an instructive way to construct a rule-based aesthetic evaluation method. However, aesthetic subjectivity and principle-to-rule mapping quantification bring great difficulties to its implementation. Referring to the "Performance Competence Evaluation Measure" [18], an aesthetic fitness function of robotic dance motions was built and regarded as the core of traditional evolutionary computation, which was used for the synthesis of humanoid robot dance [19]. The fitness function involved the sum of all movement values over all of the joints multiplied by the time that the robot remained standing [19].

From the perspective of developing machine intelligence, a machine learning-based method is a better choice. By dynamic supervised learning, a machine aesthetic model is trained to guide autonomous aesthetic evolution in a semi-interactive evolutionary computational system of robotic dance poses [20]. Moreover, based on multimodal information fusion, two different approaches of automatic aesthetics evaluation of robotic dance poses [21,22] were proposed, and several machine aesthetic models are trained to enable robots to understand and judge their own dance poses.

From the existing literature, human subjective aesthetics [9–13] is the only used aesthetic method of robotic dance motions, and machine learning is ignored, although it may be a more appropriate aesthetic method of robotic dance motions. Therefore, this paper proposes a novel approach of automatic aesthetics evaluation of robotic dance motions by using machine learning.

The main contributions of this paper are listed as follows:

(1) To imitate human dance behavior, a novel approach of automatic aesthetics evaluation of robotic dance motions is proposed.
(2) Inspired by cognitive neuroscience, the approach integrates multiple visual features, which come from the visual information channel of a robot.
(3) To describe the spatial features of robotic dance motion, a method named "ripple space coding" is designed.
(4) Verified by simulation experiments, the highest correct ratio of aesthetic evaluation is 75%.
(5) The approach is applicable to the classification problem based on action videos, such as human behavior recognition, etc.

The rest of the paper is organized as follows: Section 2 provides a detailed explanation of the entire mechanism, including the following five parts: the whole framework, extraction and optimization of motion history images, feature extraction, feature integration, and ensemble learning. Section 3 introduces the complete experimental process and presents the simulated experimental results. Based on these results, Section 4 discusses our mechanism in six aspects. Section 5 gives a brief conclusion and plans for future work.

2. Automatic Aesthetics Evaluation for Robotic Dance Motions

2.1. The Whole Framework

Vision is not only a main information source of human beings, as perceived from their embodied environment [23], but it is also an important channel of aesthetic cognition of human beings [24]. In dance creation activities, human dancers always use their eyes to observe their own dance motions from mirrors and then understand the aesthetics of their own dance motions, aiming to improve their dancing performance.

As a way to develop autonomous ability and cognitive ability, a humanoid robot uses a similar mechanism to achieve automatic aesthetics for its own dance motions. Specifically, a humanoid robot, placed before a mirror, uses its "eyes" (visual cameras) to observe its own dance motions in the mirror and finally comprehensively judges the aesthetics of its own dance motions. Therefore, we propose an automatic machine aesthetics mechanism for robotic dance motions based on multiple visual feature integration (see Figure 1 for the whole framework of the mechanism).

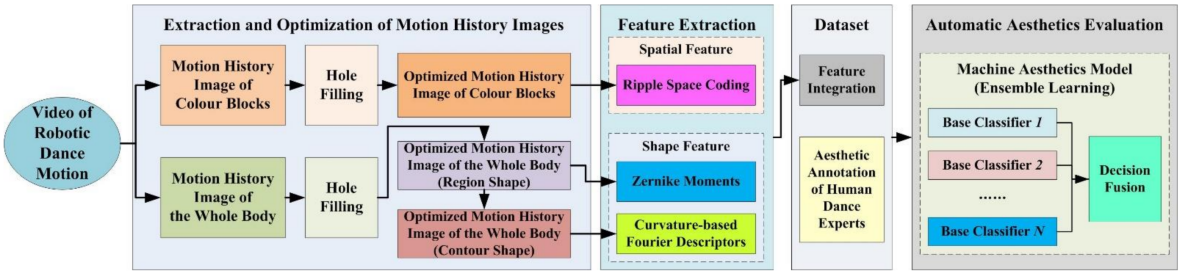

Figure 1. The proposed framework.

In our proposed mechanism, vision is regarded as the only information source. After demonstrating its own dance motions before a mirror, a humanoid robot uses its "eyes" (visual cameras) to capture the corresponding mirror videos.

Firstly, a video of robotic dance motion is converted into two kinds of motion history images (MHIs): MHI of color blocks, and MHI of the whole body. The former focuses on the historical movements of the color blocks on the robot's body. Because these color blocks are usually distributed on the robot's limbs, their historical movements reflect the extension and spatial distribution of robotic dance motions [22]. The latter focuses on the historical movement of the whole body of the robot, which mainly investigates the overall space activities of the robot.

Secondly, the two kinds of MHIs are processed respectively by a hole filling operation of mathematical morphology, and the optimized results are as followings: optimized MHI of color blocks, and optimized MHI of the whole body. Notably, they have no holes, and belong to regional images. Furthermore, based on the optimized MHI of the whole body (region shape), the corresponding contour image is generated by boundary detection, namely, the optimized MHI of the whole body (contour shape).

Next, based on the above three optimized MHIs, spatial feature (ripple space coding) and shape features (Zernike moments and curvature-based Fourier descriptors) are extracted, respectively, and then these features are fused together to form an integrated feature that characterizes the robotic dance motion completely. After human dance experts give their aesthetic annotations of robotic dance motions that they observed, each aesthetic annotation (label) and the corresponding integrated feature (instance) form an example of

a robotic dance motion. Thus, an example dataset of robotic dance motions can be built for the following machine learning.

Based on the example dataset of robotic dance motions, a homogeneous ensemble classifier, which fuses the decisions of several base classifiers, is trained to build a machine aesthetics model of robotic dance motions. Finally, when the humanoid robot presents a new robotic dance motion before a mirror, the trained machine aesthetics model automatically judges the aesthetics of the new robotic dance motion.

2.2. Extraction and Optimization of Motion History Images

Motion history image (MHI) is an effective method to represent target movement, and it transforms the description of target movement from a video to a single image [25]. Specifically, MHI contains the temporal and spatial information of target movement, which comes from the original video. By calculating the pixel changes at the same position in a time period, MHI shows the target movement in the form of image brightness.

After a video is converted into a sequence of grayed frames, a frame-difference method is used to acquire motion regions:

$$D(x, y, t) = |GF(x, y, t + \Delta) - GF(x, y, t)|; \tag{1}$$

where $GF(x, y, t)$ refers to the intensity value of the pixel (x, y) in the t-th grayed frame, Δ refers to the distance between two grayed frames, and $D(x, y, t)$ refers to the differential value of the pixel (x, y) in the t-th differential image.

Then, the differential image is binarized by the following processing:

$$\varphi(x, y, t) = \begin{cases} 1 \; if \; D(x, y, t) \geq \varepsilon; \\ 0 \; otherwise; \end{cases} \tag{2}$$

where ε refers to the difference threshold, and $\varphi(x, y, t)$ refers to the intensity value of the pixel (x, y) in the t-th binary image.

Based on the above time sequence of binary images, MHI can be updated by the following update function [25]:

$$H_\tau(x, y, t) = \begin{cases} \tau \; if \; \varphi(x, y, t) = 1; \\ max(0, H_\tau(x, y, t - 1) - \delta) \; otherwise; \end{cases} \tag{3}$$

where $H_\tau(x, y, t)$ refers to the intensity value of the pixel (x, y) of MHI at time t, τ refers to the duration of movement, and δ refers to the decay parameter.

Because MHIs often have holes, they are not conducive to the subsequent feature extraction. Thus, hole filling is necessary to optimize MHIs. Moreover, to generate a contour shape from a region shape, boundary detection is still necessary. Therefore, this paper uses the corresponding algorithms of mathematical morphology [26] to process the above two tasks.

After a video of robotic dance motion (as shown in Figure 2) is acquired, it is firstly converted into two kinds of MHIs: an MHI of color blocks (as shown in Figure 3a), and an MHI of the whole body (as shown in Figure 3b). Next, the two kinds of MHIs are optimized by hole filling, and the corresponding results are generated: the optimized MHI of color blocks (as shown in Figure 4a), and the optimized MHI of the whole body (region shape) (as shown in Figure 4b). Then, by using the boundary detection algorithm [26] on the optimized MHI of the whole body (region shape), the corresponding optimized MHI of the whole body (contour shape) is generated (as shown in Figure 4c).

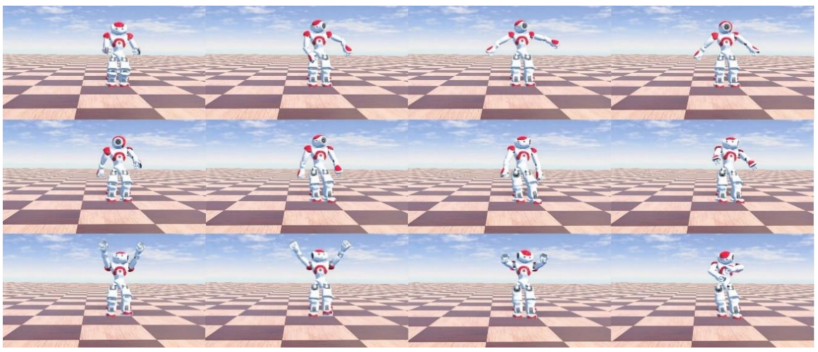

Figure 2. An example video of robotic dance motion.

(**a**) (**b**)

Figure 3. Two kinds of motion history images (MHIs): (**a**) MHI of color blocks; (**b**) MHI of the whole body.

(**a**) (**b**) (**c**)

Figure 4. Three kinds of optimized MHIs: (**a**) optimized MHI of color blocks; (**b**) optimized MHI of the whole body (region shape); (**c**) optimized MHI of the whole body (contour shape).

2.3. Feature Extraction

Feature extraction refers to the conversion of the primitive features into a group of physical or statistical features. Based on the above three kinds of optimized MHIs, feature extraction acquires suitable features that characterize a robotic dance motion, which include spatial feature and shape features (region, contour).

2.3.1. Spatial Feature

In an image, the components of a single target (or multiple targets) are often scattered in it. Their locations or distributions can present a unique topological relationship [27], which can be used as a spatial feature to identify a target (or multiple targets). To a biped humanoid robot (such as NAO robot [28], HRP-2 robot [29], Robonova robot [30], etc.), its important body parts (such as head, shoulder, hand, foot, leg, etc.) are often attached with color blocks. When the robot performs a dance motion, its color blocks still present a specific spatial distribution in the corresponding action space. The robot can be regarded as a single target, and these color blocks can be regarded as the components of the target. Based on the optimized MHI of color blocks, we design a method named "ripple space coding", which is used to describe the spatial distribution of color blocks as the spatial feature of robotic dance motion.

In the method of ripple space coding (RSC), a specific point in an image is firstly determined as the origin, and then several concentric circles are generated with the origin

as the center. The radii of these concentric circles form an arithmetic sequence. For example, the radius (r_i) of the i-th concentric circle (C_i) is $i * r$ ($i = 1, 2, \ldots, n$), where r is a constant value of radius. Moreover, taking the origin as the intersection point, two mutually perpendicular number axes (X and Y) are generated. Thus, by intersecting these concentric circles and the two number axes, the whole image is divided into many subareas, and it is also divided into four quadrants (Q1, Q2, Q3, and Q4). Figure 5 shows a schematic diagram of the region division in this method.

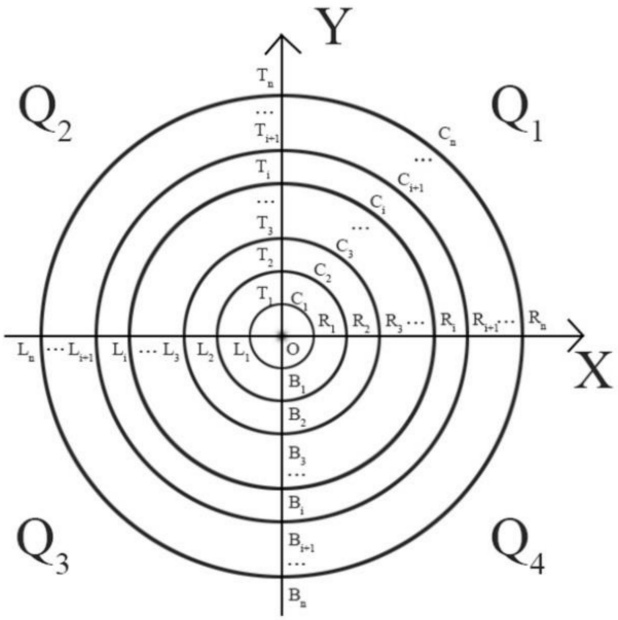

Figure 5. The schematic diagram of the region division in ripple space coding. The point O: the origin; C_i ($i = 1, 2, \ldots, n$): the i-th concentric circle; L_i ($i = 1, 2, \ldots, n$): the intersection of the i-th concentric circle and the negative half axis of X; R_i ($i = 1, 2, \ldots, n$): the intersection of the i-th concentric circle and the positive half axis of X; T_i ($i = 1, 2, \ldots, n$): the intersection of the i-th concentric circle and the positive half axis of Y; B_i ($i = 1, 2, \ldots, n$): the intersection of the i-th concentric circle and the negative half axis of Y; Q_j ($j = 1, 2, 3, 4$): the j-th quadrant; the divided subareas: $R_kT_kT_{k+1}R_{k+1}$, $T_kL_kL_{k+1}T_{k+1}$, $L_kB_kB_{k+1}L_{k+1}$, $B_kR_kR_{k+1}B_{k+1}$, ($k = 1, 2, \ldots, n-1$), OR_1T_1, OT_1L_1, OL_1B_1, OB_1R_1.

For the components of a single target (or multiple targets) that are scattered in an image, each one is represented by its centroid, which must also fall in one of the above subareas. In a specific quadrant, the number of centroids appearing in each subarea is counted, and several binary bits are used to constitute a code to store the corresponding number of centroids for each subarea. Then, the binary codes corresponding to each subarea are arranged from the inside out in order to form a longer one, which corresponds to the specific quadrant. Finally, the binary code of the quadrant is converted to the corresponding decimal code to obtain a spatial feature. Thus, four spatial features (RSC_j, $j = 1, 2, 3, 4$) can be generated from four quadrants (Q_j, $j = 1, 2, 3, 4$), and they describe the robotic dance motion together.

Figure 6 shows an example that uses the method of ripple space coding to extract spatial features of a robotic dance motion. Firstly, based on the optimized MHI of the whole body (region shape), the centroid of region shape is calculated, and it is projected into the same position in the optimized MHI of color blocks, so it is regarded as the origin (ripple center). Secondly, six concentric circles are generated with the origin as the center, and two mutually perpendicular number axes (X and Y) are generated. Next, the centroid of each color block is calculated, and it must fall in a certain subarea. Notably, each subarea uses 2 binary bits to store the number of centroids in it; thus, "00" means that the number of centroids is 0, and "11" means that the number of centroids is 3. In the fourth quadrant (Q4) of Figure 6, the binary bits of each subarea from the inside out are as follows: "00",

"00", "01", "01", "00", "00". Thus, the corresponding binary code of the fourth quadrant (Q4) is "000001010000", and its decimal code after conversion is 80, which is regarded as a spatial feature RSC_4 of the robotic dance motion.

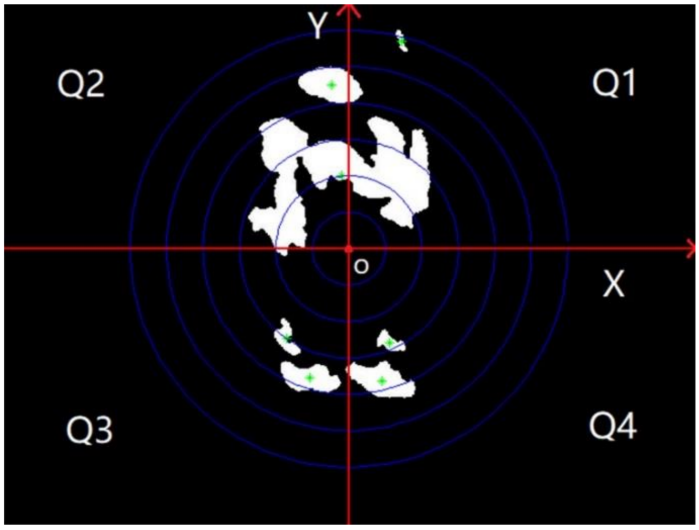

Figure 6. An example of ripple space coding for a robotic dance motion. The green points: the centroids of color blocks.

2.3.2. Region Shape Feature

For a region image, region shape is viewed as a whole, and its feature is relatively less affected by noise and shape changes. The region shape feature is to use all the pixels in the region to obtain the parameters describing the properties of the region surrounded by the target contour, such as area, Euler number, eccentricity, geometric moment invariants, Zernike moment, rotation moment, etc. Taking into consideration that the Zernike moment is an effective region feature with the advantages of low information redundancy and noise insensitivity [31], we select the Zernike moment as the region shape feature of a robotic dance motion. Specifically, several Zernike moments are extracted from the optimized MHI of the whole body (region shape), and they characterize a robotic dance motion together from the region shape.

Zernike moments are orthogonalization functions based on Zernike polynomials. The set of orthogonal polynomials that is used forms a complete positive intersection in a unit circle ($x^2 + y^2 = 1$). A Zernike polynomial with m order and n repeatability is defined as follows:

$$V_{mn}(x, y) = V_{mn}(\rho, \theta) = R_{mn}(\rho)e^{jn\theta} \tag{4}$$

where ρ is the vector length between the origin and the point (x, y); θ is the angle between the vector ρ and the counterclockwise direction of the x axis; $R_{mn}(\rho)$ is an orthogonal radial polynomial defined as follows:

$$R_{mn}(\rho) = \sum_{t=0}^{(m-|n|)/2} (-1)^t \frac{(m-t)!}{t!\left(\frac{m+|n|}{2}-t\right)!\left(\frac{m-|n|}{2}-t\right)!} \rho^{m-2t} \tag{5}$$

where m is a positive integer or zero; $(m - |n|)$ is an even number; $|n|$ is less than or equal to m; and $j = \sqrt{-1}$.

Taking into consideration that a region image is a discrete digital image, the corresponding definition of the Zernike moment based on a discrete digital image is as follows:

$$Z_{mn} = \frac{m+1}{\pi} \sum_x \sum_y f(x,y) V_{mn}^*(x,y),\ x^2 + y^2 \leq 1 \tag{6}$$

where "*" expresses the complex conjugate.

Any high-order Zernike moments in a discrete digital image are easily constructed to represent region features. However, low-order Zernike moments always describe the overall shape of an image; high-order Zernike moments always describe the specific details of an image [31,32]. Thus, we use 0~4-order Zernike moments to describe the overall shape of a robotic dance motion, which is presented in the optimized MHI of the whole body (region shape). There are nine Zernike moments: Z_{00}, Z_{11}, Z_{20}, Z_{22}, Z_{31}, Z_{33}, Z_{40}, Z_{42}, Z_{44}. After the natural logarithms for the moduli of these Zernike moments are computed, normalization is implemented further. Consequently, nine Zernike moment features of a robotic dance motion (ZMF_1, ZMF_2, ... , ZMF_9), which describe the robotic dance motion together from the perspective of overall shape, are obtained.

2.3.3. Contour Shape Feature

Contour shape refers to a set of pixels that constitutes the boundary of a region. By characterizing the geometrical distribution of a regional boundary, the contour shape feature is described with descriptors. The Fourier transform coefficients of an object shape boundary curve are Fourier descriptors—a classical shape description method in the transform domain. As verified in the existing literature, Fourier descriptors, based on the coordinate sequence of an object contour, perform the best among the various typical methods of 2-D shape recognition [33]. Therefore, Fourier descriptors, extracted from the optimized MHI of the whole body (contour shape), are regarded as the contour shape features of a robotic dance motion.

In a contour image, T sampling points are acquired by equidistant sampling along the boundary of contour. Any sampling point $BS(i)$ is expressed in the following form:

$$BS(i) = (x_i, y_i),\ i = 0, 1, 2, \ldots\ldots, T-1 \tag{7}$$

where (x_i, y_i) are the coordinates of $BS(i)$ in the XOY plane.

With reference to the coordinates of the T sampling points, the curvature of the contour is computed. Mathematically, the curvature of a discrete curve is calculated as follows:

$$Cur_i = |Der_i''| / \left(1 + Der_i'^2\right)^{3/2} \tag{8}$$

where, at sampling point $BS(i)$, the first and second derivatives of the curve function are Der_i' and Der_i'', respectively, defined as follows:

$$Der_i' = (y_{i+1} - y_i) / (x_{i+1} - x_i) \tag{9}$$

$$Der_i'' = \left(Der_{i+1}' - Der_i'\right) / (x_{i+1} - x_i) \tag{10}$$

The discrete Fourier coefficients of a one-dimensional sequence are defined as follows:

$$f(u) = \frac{1}{T} \sum_{i=0}^{T-1} Cur_i exp\left(\frac{-j2\pi ui}{T}\right), u = 0, 1, 2, \ldots, T-1 \tag{11}$$

These discrete Fourier coefficients are Fourier descriptors, which must be further normalized. In this paper, we use the min–max normalization method to generate normalized curvature-based Fourier descriptors, defined as follows:

$$CFD_{i+1} = \frac{\| f(i) \| - Minf}{Maxf - Minf},\ i = 0, 1, 2, \ldots, T-1 \tag{12}$$

where Minf (Maxf) is the minimum (maximum) value of $\| f(u) \|$, (u $= 0, 1, 2, \ldots, T-1$).

The normalized Fourier descriptors have the invariance characteristics of rotation, translation, scale, and the position of the starting point. Moreover, the low-frequency components of the normalized Fourier descriptors always describe the contour better than their high-frequency components; therefore, some low-frequency components of the normalized curvature-based Fourier descriptors (CFD_1, CFD_2, ... , CFD_q) ($q \leq [T/4]$) are selected as the contour shape features of a robotic dance motion. Specifically, in this paper, we use thirty normalized curvature-based Fourier descriptors ($q = 30$, $T = 200$) as the contour shape features of a robotic dance motion.

2.4. Feature Integration

To characterize a robotic dance motion, the spatial feature and shape features (region, contour) are extracted respectively from the vision information channel. Each kind of feature characterizes a nature of a robotic dance motion from a certain aspect. We believe the spatial feature portrays the spatial geometric distribution of robotic body parts; the region feature portrays the overall silhouette for a robotic dance motion; the contour feature portrays the overall peripheral shape. Therefore, to describe a robotic dance motion more completely, the above three kinds of features are fused together to form an integrated feature. In this paper, the integrated feature is expressed by (RSC_1, RSC_2, RSC_3, RSC_4, ZMF_1, ZMF_2, ... , ZMF_9, CFD_1, CFD_2, ... , CFD_{30}).

2.5. Ensemble Learning

The stage of automatic aesthetics evaluation has two tasks: (1) train a machine aesthetics model so that the machine possesses human aesthetic ability and (2) autonomously judge the aesthetics of a robotic dance motion. By feature extraction and integration, each robotic dance motion is expressed as an instance of the integrated feature. When enough robotic dance motions are processed, a corresponding dataset is generated. For a machine to possess human aesthetic ability in robotic dance motion, supervised learning is necessary [20,22]. Thus, a human dance expert, after observing all the robotic dance motions, should provide the corresponding aesthetic annotations (good/bad) for these motions. Although there is an option to use machine learning (including deep learning) for video annotation, it is not suitable for the aesthetic annotation of robotic dance motion. This is because of the lack of objective evaluation criteria in artistic aesthetic cognition [20] and the small dataset.

For machine learning, an example of each robotic dance motion consists of two parts: an instance of the integrated feature and the corresponding aesthetic label (good/bad). Therefore, an example dataset of robotic dance motions can be built. That dataset then becomes the basis for further training a machine aesthetics model.

Notably, as an effective method of machine learning, ensemble learning trains multiple learners to solve the same problem [34]. In this paper, ensemble learning is used in the stage of automatic aesthetics evaluation. Specifically, based on the above example dataset of robotic dance motions, a homogeneous ensemble classifier, which fuses the decisions of several base classifiers, is trained to build a machine aesthetics model of robotic dance motions. After the model is built, a humanoid robot can automatically evaluate the aesthetics by observing its own dance motions. Thus, it is possible to further autonomously create robotic choreography.

3. Experiments

In our experiments, Chinese Tibetan tap is selected as the robotic dance form, and a NAO robot is selected as the dance carrier. We tested our proposed mechanism in an experimental simulated environment, which included: Webots R2019a simulator, Matlab R2014a, Dev-C++ 5.11, and PyCharm 2019.1.1. Notably, we used the NAO robot module in Webots, and the sklearn library in Pycharm. Moreover, Matlab and Dev-C++ were just used routinely.

In the "Simulation View" area of the Webots simulator, a simulated NAO robot displayed a dance motion. The video shown in "Simulation View" was treated as the visual information source in which the robot observes its own dance motion in the "mirror". By designing the corresponding processing programs in Matlab, motion history images were extracted and optimized, and the spatial and shape features were extracted. In Dev-C++, robotic dance motions were generated randomly, and data file formats were transformed. Moreover, ensemble learning was implemented in PyCharm.

Based on the dance expressive space of a humanoid robot [4] and three dance element sets [4], 5000 robotic dance poses of Chinese Tibetan tap were generated randomly. Then, every 6–10 robotic dance poses were randomly selected to form a robotic dance motion, which is a sequence of robotic dance poses. In this way, 120 robotic dance motions of Chinese Tibetan tap were generated randomly and used for our experiments. Notably, there are two constraints in the generation of robotic dance poses and motions: (1) they combine the innovativeness and preservation of human dance characteristics [20]; (2) they exclude the dance poses and motions that make a robot fall.

For supervised learning, a Chinese folk dance expert, with extensive experience in stage performance and teaching and a high capacity for aesthetic appreciation and evaluation, was invited to label the aesthetic categories of the 120 robotic dance motions as good or bad. To each robotic dance motion, four ripple space codings of quadrants (RSC_j, j = 1, 2, 3, 4) were taken as spatial features, and nine Zernike moments (0~4-order Zernike moments) were taken as region features, and 30 low-frequency components of curvature-based Fourier descriptors were taken as contour features, where the total number of sampling points on the boundary was 200. Thus, the integrated feature that was used to characterize a robotic dance motion was expressed by (RSC_1, RSC_2, RSC_3, RSC_4, ZMF_1, ZMF_2, . . . , ZMF_9, CFD_1, CFD_2, . . . , CFD_{30}).

By combining each integrated feature (instance) and the corresponding aesthetic annotation (label) to form an example of a robotic dance motion, an example dataset of robotic dance motions was built. To verify the effectiveness of the model, the example dataset was randomly divided into a training dataset and a test dataset. Furthermore, the size of the training dataset and the test dataset was 80% and 20% of the example dataset, respectively.

In the aesthetic cognition of art, each kind of art has its own particularity, and often lacks objective evaluation criteria [20]. This also leads to the correlations between machine learning methods for aesthetic evaluation and art forms. Random forest is a suitable machine learning method for aesthetic evaluation in the field of robotic dance, which has been verified on robotic dance poses [35]. Therefore, random forest was selected as the specific implementation method of the base classifier in ensemble learning, and three random forests were formed to be a homogeneous ensemble classifier. Specifically, the parameters of our ensemble classifier were set as follows: (1) the three random forests had 7, 10, and 13 decision trees, respectively; (2) the initial seeds of random numbers were different; (3) Gini impurity was used as the split criterion; (4) the values of the remaining parameters used default settings. By fusing the independent decisions that came from the three random forests, the ensemble classifier made the final decision. Moreover, the method of decision fusion was voting, and the final decision was the category with the most votes.

For comparison, this paper also used several mainstream machine learning methods to train machine aesthetic models. Specifically, because the sklearn library in Python provides the implementation of these machine learning methods, we called the corresponding functions and built the corresponding machine aesthetic models, aiming to obtain their performance data to compare with our ensemble classifier. Moreover, these machine learning methods used default parameter values when the machine aesthetic models were built. The detailed results are shown in Table 1.

Table 1. The performance comparison of different machine learning models.

Machine Learning Method	Correct Ratio (Accuracy)
KNN	66.7%
Logistic Regression	45.8%
GBDT	54.2%
AdaBoost	58.3%
Naive Bayesian	54.2%
MNB	50%
QDA	50%
SVM	50%
Decision Tree	66.7%
Random Forest	70.8%
Our Ensemble Classifier	**75%**

The comparison results show that our ensemble classifier has achieved the high correct ratio of aesthetics evaluation at 75%. Notably, the correct ratio refers to the index of accuracy in machine learning, and it is defined as the proportion of correctly classified examples to the total number of examples. From the confusion matrix of classification in Table 2, our ensemble classifier can effectively identify good/bad robotic dance motions. Although there are some wrong classification results, we think it should be related to the lack of objective evaluation criteria in the aesthetic cognition of art [20]. The other performance indexes of our ensemble classifier are as follows: the precision is 80%, the F-score is 72.7%, and the AUC is 75%.

Table 2. The confusion matrix of classification.

		Actual Value	
		Bad	Good
Prediction Outcome	Bad	10	4
	Good	2	8

4. Discussion

4.1. Visual Feature Integration

In the aesthetic cognition of human beings, visual information plays an important role [24]. Meanwhile, for the human brain, all visual processing determines what objects in the field of vision are and where they are located [35]. Therefore, spatial and shape features are helpful to characterize the target in the aesthetic cognition of the human brain.

Furthermore, multiple feature integration is always better than single feature. Viewed from the perspective of cognitive neuroscience, many cells in the superior colliculus of the human brain fuse the information emanating from different sensory channels, and the cells show multisensory properties [36]. The response of the cell is stronger when there are inputs from multiple senses compared to when the input is from a single modality [37,38]. Thus, this paper integrates two kinds of visual features (spatial and shape features) to characterize a robotic dance motion.

Specifically, an original video of robotic dance motion is firstly obtained from the vision channel of the robot, and then the video is converted into three kinds of optimized MHIs, and finally the corresponding spatial and shape features are extracted and integrated to characterize the robotic dance motion. Essentially, the whole procedure above reflects (1) the possibility that a robot, through its visual channels, understands the beauty of its own dance motions, (2) that spatial and shape features are useful in the automatic machine aesthetics of robotic dance motions, and (3) that in visual aesthetic processing, our mechanism can process the spatial and shape information in a form similar to human brains.

Although our proposed model has achieved a high correct ratio of aesthetics evaluation of 75%, the machine aesthetic effect based on the integrated feature is insufficient

because of the following three possible reasons: (1) a humanoid robot shows its dance motions in three-dimensional space. However, the video of a robotic dance motion, captured by robotic cameras, is converted into two-dimensional space, with the loss of one-dimensional spatial (depth) data. (2) The adopted spatial feature (ripple space coding) and shape features (Zernike moments and curvature-based Fourier descriptors) are insufficient to characterize a robotic dance motion. Additionally, the more powerful spatial and shape feature descriptors might be lacking. (3) It may be a detour to acquire the movement characteristics of robotic dance motions by using MHIs.

To improve the machine aesthetic effects of robotic dance motions, based on the integrated features, the following three measures are considered: (1) by using the depth sensor to capture the video of a real humanoid robot in a mirror, the RGB and depth videos of the robotic dance motion are acquired simultaneously. Thus, the missing one-dimensional spatial (depth) data are retrieved. The extracted method of spatial and shape features, based on the RGB and depth videos of the robotic dance motion, must be designed. (2) By trying other spatial features (such as basic spatial distribution and spatial relationship [22], etc.) and shape features (such as wavelet descriptor, scale space, Hu moment invariants, autoregressive, etc.), more powerful, or suitable, spatial and shape feature descriptors are found. (3) By constructing a deep neural network, the movement characteristics of robotic dance motions are automatically extracted. At this time, each video of robotic dance motion is input into the network as a time sequence of frames, each of which in turn indicates an image of a robotic dance pose.

4.2. Selection of Visual Features

To understand the content from MHIs, various visual features can be selected to describe the target in MHIs. Considering that an MHI does not contain the information of color and texture, the available visual feature types are mainly focused on spatiality and shape. Obviously, combinations of the above two feature types should more comprehensively describe the target in MHIs. Notably, each feature type still includes many specific features, e.g., the type of shape feature includes wavelet descriptor, shape context, etc. Thus, many combinations of specific features exist, which may involve one or more feature types. The feature combination that best describes the target in MHIs varies with different problems.

In order to characterize a robotic dance motion, two feature types of spatiality and shape are selected in this paper. Under the spatial feature type, we design ripple space coding as the specific spatial feature. Moreover, under the shape feature type, there are two sub-types, region and contour, so we select two specific features from the above two sub-types: Zernike moments (region feature) and curvature-based Fourier descriptors (contour feature). Experimental results show that there is still much room for improving the effect of automatic aesthetics on robotic dance motions. More concretely, on the premise of not causing feature conflict [22], feature combinations that introduce different feature types and specific features, as much as possible, it may be possible to describe a robotic dance motion more completely.

4.3. Influence on Robotic Dance

Good dance motions are important for improving the quality of robotic dance. According to the dance expressive space of a humanoid robot [4], a robotic dance consists of a dance motion sequence, where each dance motion is a basic component of the robotic dance. Moreover, good robotic dance, in turn, also requires good dance motions. Good robotic choreography has three features: (1) preservation of the characteristics of human dance; (2) innovativeness of the dance; and (3) accordance with human aesthetics [3]. Notably, the above three features of good robotic choreography must be deconstructed in dance motions as the requirements for good dance motions. In our experiments, the first two requirements were met by using the random generation method [4]. The machine aesthetic model achieved the last requirement, thereby enabling a robot to possess the aesthetic

ability of human dance experts. Once an abundance of good dance motions is selected, the probability is greater that a good robotic dance will be created.

4.4. Practical Application

To provide a solution of the core problem of autonomous robotic choreography, this paper proposed the mechanism of automatic aesthetics evaluation of robotic dance motions based on multiple visual feature integration. In addition to this purpose, the method can also be used in other practical applications, such as human behavior recognition, etc.

Essentially, the automatic aesthetics evaluation of robotic dance motions is a classification problem. It takes action videos as input, and then builds and trains a machine learning model, and finally classifies the results. Human behavior recognition is also such a process [39]. Therefore, the proposed method can be used in the classification problem that takes action videos as input, especially in human behavior recognition.

4.5. Limitation of the Proposed Approach

To imitate human dance behavior, this approach selects a biped humanoid robot as the dance carrier, and requires that color blocks are attached to the important body parts of the humanoid robot (such as head, shoulder, hand, foot, leg, etc.). Some biped humanoid robots meet this requirement, such as NAO robot, HRP-2 robot, Robonova robot, etc., but some do not, such as Hubo, ASIMO, QRIO, DARwIn-OP, etc. For these biped humanoid robots without color blocks, it is feasible to manually transform them by pasting color stickers on their important body parts. Meanwhile, in order to make this approach work effectively, the color blocks on a robot's body are required to have a unique color, which is different from the robot's embodied environment. If this requirement cannot be met, a unique color that differs from the embodied environment of the robot should be determined, and then colored stickers with the unique color should be pasted on the important body parts of the robot [22]. In this way, the robot can be recognized from the video of robotic dance motion, and the feature of ripple space coding can be further extracted.

In addition, as the task of data preprocessing, the extraction and optimization of a motion history image expends many computing resources, which brings difficulties when building a real-time system. If there is a high-performance computing server and a suitable parallel computing method, this will help to enhance the real-time performance of the above process.

In this paper, the dataset of robotic dance motion is not big, and the procedure of automatic aesthetic evaluation is realized by feature extraction and ensemble learning. However, some potential alternatives remain to be found. The latest research in image understanding [40] and image segmentation [41,42] will be helpful to further improve the performance of machine aesthetic models. Moreover, the technology of deep learning [43–46] should also be considered to build machine aesthetic models. Furthermore, some advanced motion planning algorithms of humanoid robots [47,48] will be helpful to generate robotic dance motions quickly, avoid robots falling to a certain extent, and finally enrich the construction of the dataset of robotic dance motions. These open questions will be explored in the future.

4.6. Comparison with the Existing Approaches

Aimed at the aesthetics evaluation of robotic dance motions, the existing literature reports on research from different aspects and provides several solutions. However, human subjective aesthetics [9–13] is the only aesthetic method of robotic dance motions. Machine learning-based methods are ignored. Table 3 shows the comparison between the existing approaches and our approach.

Table 3. The comparison between the existing approaches and our approach.

	The Approach in [9–11]	The Approach in [12]	The Approach in [13]	Our Approach
Information Channel	visual and non-visual	visual	visual	visual
Feature Type Involved	kinematic	N/A	N/A	spatial feature, shape features (region and contour)
Specific Feature	Pose sequence-based chromosome feature	dynamic, exciting, wonder, smooth, etc.	musical synchrony, variety of movements, human characterization, flexibility of control, etc.	ripple space coding, Zernike moments, curvature-based Fourier descriptors
Aesthetic Manner	human subjective aesthetics	human subjective aesthetics	human subjective aesthetics	machine learning-based method
Machine Learning Method Involved	N/A	N/A	N/A	KNN, logistic regression, GBDT, AdaBoost, naive Bayesian, MNB, QDA, SVM, decision tree, random forest, our ensemble classifier
Highest Correct Ratio	N/A	N/A	N/A	75%
Best Feature	N/A	N/A	N/A	ripple space coding, Zernike moments, curvature-based Fourier descriptors
Best Machine Learning Method	N/A	N/A	N/A	our ensemble classifier (including three base classifiers: three random forests)

Different from all existing approaches, we explore, in this paper, the automatic aesthetic evaluation of robotic dance motions by using only the robot's own visual information. Automatic aesthetic evaluation, which provides some new visual features (ripple space coding, Zernike moments, and curvature-based Fourier descriptors) to characterize a robotic dance motion, has achieved a relatively good result (75%) in machine aesthetic evaluation of the integrated visual feature. Notably, ensemble learning also provides great help to improve the performance of machine aesthetics models.

5. Conclusions

Using the technologies of computer vision and ensemble learning, we presented, for robotic dance motions, an automatic machine aesthetics mechanism based on multiple visual feature integration. Verified by simulation experiments, our mechanism can achieve the high correct ratio of aesthetics evaluation of 75%. Experimental results show that a robot can (1) perceive and understand its own dance motions by multiple visual feature integration, and (2) automatically judge the aesthetics of its own dance motions. Thus, in a dance activity, a robot behaves more like a human being, and the autonomy and cognitive ability of the robot are promoted to a certain extent. Moreover, it was proved that spatial and shape features are useful for evaluating the aesthetic feeling of robotic dance motions. Additionally, a homogeneous ensemble classifier that contains three random forests is verified as an effective, suitable machine learning method for robotic dance motion aesthetics.

Future work will focus on the following three aspects: (1) using the proposed mechanism, a real NAO robot, placed before a mirror, will autonomously complete an aesthetic evaluation of its own dance motions; (2) more useful spatial and shape features to characterize a robotic dance motion will be found; (3) some advanced neural networks [43–46] will be used for training machine aesthetics models.

Author Contributions: Conceptualization, H.P. and J.L.; methodology, H.P.; software, J.H.; validation, J.H., H.W. and H.R.; formal analysis, H.H.; investigation, H.R.; resources, C.S.; data curation, H.W.; writing—original draft preparation, H.P.; writing—review and editing, J.L.; visualization, C.S.; supervision, H.H.; project administration, J.H.; funding acquisition, H.P. and J.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (Grant No. 61662025, 61871289), and Zhejiang Provincial Natural Science Foundation of China (Grant No. LY20F030006, LY20F020011), and National Innovation Training Project for College Students (Grant No. 201910349026).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Aucouturier, J.J. Cheek to chip: Dancing robots and AI's future. *Intell. Syst.* **2008**, *23*, 74–84. [CrossRef]
2. Or, J. Towards the development of emotional dancing humanoid robots. *Int. J. Soc. Robot.* **2009**, *1*, 367–382. [CrossRef]
3. Peng, H.; Zhou, C.; Hu, H.; Chao, F.; Li, J. Robotic dance in social robotics—A taxonomy. *IEEE Trans. Hum.-Mach. Syst.* **2015**, *45*, 281–293. [CrossRef]
4. Peng, H.; Li, J.; Hu, H.; Zhou, C.; Ding, Y. Robotic choreography inspired by the method of human dance creation. *Information* **2018**, *9*, 250. [CrossRef]
5. Schaal, S. Is imitation learning the route to humanoid robots? *Trends Cogn. Sci.* **1999**, *3*, 233–242. [CrossRef]
6. Andry, P.; Gaussier, P.; Moga, S.; Banquet, J.P.; Nadel, J. Learning and communication via imitation: An autonomous robot perspective. *IEEE Trans. Syst. Man Cybern. Part A Syst. Hum.* **2001**, *31*, 431–442. [CrossRef]
7. Breazeal, C.; Scassellati, B. Robots that imitate humans. *Trends Cogn. Sci.* **2002**, *6*, 481–487. [CrossRef]
8. Chen, S.; Zhou, C.; Li, J.; Peng, H. Asynchronous introspection theory: The underpinnings of phenomenal consciousness in temporal illusion. *Mind. Mach.* **2017**, *27*, 315–330. [CrossRef]
9. Vircikova, M.; Sincak, P. Dance Choreography Design of Humanoid Robots using Interactive Evolutionary Computation. In Proceedings of the 3rd Workshop for Young Researchers on Human-Friendly Robotics (HFR 2010), Tübingen, Germany, 28–29 October 2010.
10. Vircikova, M.; Sincak, P. *Artificial Intelligence in Humanoid Systems*; FEI TU of Kosice: Košice, Slovakia, 2010.
11. Vircikova, M.; Sincak, P. Discovering art in robotic motion: From imitation to innovation via interactive evolution. In Proceedings of the Ubiquitous Computing and Multimedia Applications, Daejeon, Korea, 13–15 April 2011; pp. 183–190.
12. Shinozaki, K.; Iwatani, A.; Nakatsu, R. Concept and construction of a robot dance system. In Proceedings of the 2007 International Conference on Mechatronics and Information Technology: Mechatronics, MEMS, and Smart Materials (ICMIT 2007), Gifu, Japan, 5–6 December 2007.
13. Oliveira, J.L.; Reis, L.P.; Faria, B.M. An empiric evaluation of a real-time robot dancing framework based on multi-modal events. *TELKOMNIKA Indones. J. Electr. Eng.* **2012**, *10*, 1917–1928. [CrossRef]
14. Manfrè, A.; Infantino, I.; Vella, F.; Gaglio, S. An automatic system for humanoid dance creation. *Biol. Inspired Cogn. Archit.* **2016**, *15*, 1–9. [CrossRef]
15. Augello, A.; Infantino, I.; Manfrè, A.; Pilato, G.; Vella, F.; Chella, A. Creation and cognition for humanoid live dancing. *Rob. Auton. Syst.* **2016**, *86*, 128–137. [CrossRef]
16. Manfré, A.; Infantino, I.; Augello, A.; Pilato, G.; Vella, F. Learning by demonstration for a dancing robot within a computational creativity framework. In Proceedings of the 1st IEEE International Conference on Robotic Computing (IRC 2017), Taichung, Taiwan, 10–12 April 2017.
17. Qin, R.; Zhou, C.; Zhu, H.; Shi, M.; Chao, F.; Li, N. A music-driven dance system of humanoid robots. *Int. J. Hum. Robot.* **2018**, *15*, 1850023. [CrossRef]
18. Krasnow, D.; Chatfield, S.J. Development of the 'performance competence evaluation measure' assessing qualitative aspects of dance performance. *J. Danc. Med. Sci.* **2009**, *13*, 101–107.
19. Eaton, M. An approach to the synthesis of humanoid robot dance using non-interactive evolutionary techniques. In Proceedings of the 2013 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Manchester, UK, 13–16 October 2013.
20. Peng, H.; Hu, H.; Chao, F.; Zhou, C.; Li, J. Autonomous robotic choreography creation via semi-interactive evolutionary computation. *Int. J. Soc. Robot.* **2016**, *8*, 649–661. [CrossRef]
21. Li, J.; Peng, H.; Hu, H.; Luo, Z.; Tang, C. Multimodal Information Fusion for Automatic Aesthetics Evaluation of Robotic Dance Poses. *Int. J. Soc. Robot.* **2020**, *12*, 5–20. [CrossRef]

22. Peng, H.; Li, J.; Hu, H.; Zhao, L.; Feng, S.; Hu, K. Feature Fusion based Automatic Aesthetics Evaluation of Robotic Dance Poses. *Rob. Auton. Syst.* **2019**, *111*, 99–109. [CrossRef]
23. Farah, M.J. *The Cognitive Neuroscience of Vision*; Blackwell Publishing: Hoboken, NJ, USA, 2000.
24. Chatterjee, A. Prospects for a cognitive neuroscience of visual aesthetics. *Bull. Psychol. Arts.* **2004**, *4*, 55–60.
25. Bobick, A.F.; Davis, J.W. The recognition of human movement using temporal templates. *IEEE Trans. Pattern Anal. Mach. Intell.* **2001**, *23*, 257–267. [CrossRef]
26. Gonzalez, R.C.; Woods, R.E. *Digital Image Processing*, 3rd ed.; Prentice-Hall: Upper Saddle River, NJ, USA, 2007.
27. Liu, T.; Du, Q.; Yan, H. Spatial Similarity assessment of point clusters. *Geomat. Inf. Sci. Wuhan Univ.* **2011**, *36*, 1149–1153.
28. Xia, G.; Tay, J.; Dannenberg, R.; Veloso, M. Autonomous robot dancing driven by beats and emotions of music. In Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2012), Valencia, Spain, 4–8 June 2012.
29. Kudoh, S.; Shiratori, T.; Nakaoka, S.; Nakazawa, A.; Kanehiro, F.; Ikeuchi, K. Entertainment robot: Learning from observation paradigm for humanoid robot dancing. In Proceedings of the 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2008) Workshop: Art and Robots, Nice, France, 22–26 September 2008.
30. Grunberg, D.; Ellenberg, R.; Kim, Y.; Oh, P. Creating an autonomous dancing robot. In Proceedings of the 2009 International Conference on Hybrid Information Technology (ICHIT 2009), Daejeon, Korea, 27–29 August 2009.
31. Kim, W.Y.; Kim, Y.S. A region-based shape descriptor using Zernike moments. *Signal Process. Image Commun.* **2000**, *16*, 95–102. [CrossRef]
32. Teh, C.H.; Chin, R.T. On image analysis by the methods of moments. *IEEE Trans. Pattern Anal. Mach. Intell.* **1988**, *10*, 496–513. [CrossRef]
33. Kauppinen, H.; Seppanen, T.; Pietikainen, M. An experimental comparison of autoregressive and Fourier-based descriptors in 2-D shape classification. *IEEE Trans. Pattern Anal. Mach. Intell.* **1995**, *17*, 201–207. [CrossRef]
34. Zhou, Z.H. *Ensemble Methods: Foundations and Algorithms*; Chapman and Hall/CRC: London, UK; Boca Raton, FL, USA, 2012.
35. Peng, H.; Li, J.; Hu, H.; Hu, K.; Tang, C.; Ding, Y. Creating a Computable Cognitive Model of Visual Aesthetics for Automatic Aesthetics Evaluation of Robotic Dance Poses. *Symmetry* **2020**, *12*, 23. [CrossRef]
36. Gazzaniga, M.S.; Ivry, R.B.; Mangun, G.R. *Cognitive Neuroscience: The Biology of the Mind*, 4th ed.; W. W. Norton & Company: New York, NY, USA, 2013.
37. Stein, B.E.; Stanford, T.R.; Wallace, M.T.; Vaughan, J.W.; Jiang, W. Crossmodal spatial interactions in subcortical and cortical circuits. In *Crossmodal Space and Crossmodal Attention*; Spence, C., Driver, J., Eds.; Oxford University Press: Oxford, UK, 2004; pp. 25–50.
38. Holmes, N.P.; Spence, C. Multisensory integration: Space, time and superadditivity. *Curr. Biol.* **2005**, *15*, R762–R764. [CrossRef]
39. Tang, C.; Hu, H.; Wang, W.; Li, W.; Peng, H.; Wang, X. Using a Multilearner to Fuse Multimodal Features for Human Action Recognition. *Math. Probl. Eng.* **2020**, 4358728. [CrossRef]
40. Ju, Z.; Gun, L.; Hussain, A.; Mahmud, M.; Ieracitano, C. A Novel Approach to Shadow Boundary Detection Based on an Adaptive Direction-Tracking Filter for Brain-Machine Interface Applications. *Appl. Sci.* **2020**, *10*, 6761. [CrossRef]
41. Dey, N.; Rajinikanth, V.; Fong, S.J.; Kaiser, M.S.; Mahmud, M. Social Group Optimization–Assisted Kapur's Entropy and Morphological Segmentation for Automated Detection of COVID-19 Infection from Computed Tomography Images. *Cogn. Comput.* **2020**, *12*, 1011–1023. [CrossRef] [PubMed]
42. Ali, H.M.; Kaiser, M.S.; Mahmud, M. Application of Convolutional Neural Network in Segmenting Brain Regions from MRI Data. In Proceedings of the 12th International Conference on Brain Informatics. Lecture Notes in Computer Science, Haikou, China, 13–15 December 2019.
43. Mahmud, M.; Kaiser, M.S.; McGinnity, T.M.; Hussain, A. Deep Learning in Mining Biological Data. *Cogn. Comput.* **2021**, *13*, 1–33. [CrossRef] [PubMed]
44. Noor, M.B.T.; Zenia, N.Z.; Kaiser, M.S.; Mamun, S.A.; Mahmud, M. Application of deep learning in detecting neurological disorders from magnetic resonance images: A survey on the detection of Alzheimer's disease, Parkinson's disease and schizophrenia. *Brain Inf.* **2020**, *7*, 11. [CrossRef]
45. Kuang, Q.; Jin, X.; Zhao, Q.; Zhou, B. Deep Multimodality Learning for UAV Video Aesthetic Quality Assessment. *IEEE Trans. Multimed.* **2020**, *22*, 2623–2634. [CrossRef]
46. Xiao, L.; Li, S.; Li, K.; Jin, L.; Liao, B. Co-Design of Finite-Time Convergence and Noise Suppression: A Unified Neural Model for Time Varying Linear Equations with Robotic Applications. *IEEE Trans. Syst. Man. Cybern. Syst.* **2020**, *50*, 5233–5243. [CrossRef]
47. Muni, M.; Parhi, D.; Kumar, P. Improved Motion Planning of Humanoid Robots Using Bacterial Foraging Optimization. *Robotica* **2021**, *39*, 123–136. [CrossRef]
48. Devaraja, R.R.; Maskeliūnas, R.; Damaševičius, R. Design and Evaluation of Anthropomorphic Robotic Hand for Object Grasping and Shape Recognition. *Computers* **2021**, *10*, 1. [CrossRef]