symmetry

MDPI

Article

Creating a Computable Cognitive Model of Visual Aesthetics for Automatic Aesthetics Evaluation of Robotic Dance Poses

Hua Peng [1,2,3], Jing Li [4,*], Huosheng Hu [3], Keli Hu [1], Chao Tang [5] and Yulong Ding [6]

[1] Department of Computer Science and Engineering, Shaoxing University, Shaoxing 312000, China; penghua_47@163.com (H.P.); ancimoon@gmail.com (K.H.)
[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
[5] Department of Computer Science and Technology, Hefei University, Hefei 230601, China; tangchao@hfuu.edu.cn
[6] School of Automation, Beijing Institute of Technology, Beijing 100081, China; dingyvlong@163.com
* Correspondence: lijing_47@126.com; Tel.: +86-0575-8834-2988

Received: 21 November 2019; Accepted: 18 December 2019; Published: 20 December 2019

Abstract: Inspired by human dancers who can evaluate the aesthetics of their own dance poses through mirror observation, this paper presents a corresponding mechanism for robots to improve their cognitive and autonomous abilities. Essentially, the proposed mechanism is a brain-like intelligent system that is symmetrical to the visual cognitive nervous system of the human brain. Specifically, a computable cognitive model of visual aesthetics is developed using the two important aesthetic cognitive neural models of the human brain, which is then applied in the automatic aesthetics evaluation of robotic dance poses. Three kinds of features (color, shape and orientation) are extracted in a manner similar to the visual feature elements extracted by human brains. After applying machine learning methods in different feature combinations, machine aesthetics models are built for automatic evaluation of robotic dance poses. The simulation results show that our approach can process visual information effectively by cognitive computation, and achieved a very good evaluation performance of automatic aesthetics.

Keywords: robotic dance pose; automatic aesthetics estimation; visual aesthetic cognition; machine learning

1. Introduction

Robotic dance is a multi-disciplinary research area, involving the arts, cognitive science, robotics, etc. It can be classified into four categories, namely cooperative human–robot dance, imitation of human dance motions, synchronization for music, and robotic choreography creation [1]. For the category of robotic choreography creation, the aim is to develop the humanoid abilities of robots and understand dance objects (including dance poses, dance motion, and dance works) for robotic dance creation. Notably, only the robotic choreography creation has the aesthetic requirements for dance objects, and accordance with human aesthetics is a necessary feature of good robotic choreography [1].

Aesthetics is a high-level cognitive function of the human brain, and its task is to distinguish the beauty of objects. By using the methods of brain science and cognitive neuroscience, neuro-aesthetics investigates the relationship between the human brain and aesthetics mechanisms, and explores the aesthetic activity process and neural processing mechanism of the human brain [2]. As an important part of neuro-aesthetics, aesthetic cognitive neural models abstract and reflect the neural processing mechanism of the human brain [3–8]. If a computable aesthetic cognitive neural model can be introduced into robots, it will make robots achieve aesthetic cognition and further develop their cognitive abilities in a way similar to the human brain, which has been rarely studied so far.

In human dance activities, human dancers always observe their own dance poses in mirrors, and then evaluate the aesthetics of these dance poses by using their accumulated aesthetic experience. In this way, human dancers can further create their own dance works. The imitation of human behaviors is an effective way to develop the autonomous ability and intelligence of robots [9–11]. A robot should understand its own dance poses, which appear in a mirror, and make an automatic aesthetics evaluation of them. Some related literatures include human subjective aesthetics [12–14] and machine learning based methods [15–17]. However, the automatic aesthetic evaluation approach to robotic dance poses, which builds on the aesthetic cognitive neural model of the human brain, has not been investigated to date.

This paper proposes a new computable cognitive model of visual aesthetics to conduct a new automatic aesthetic evaluation of robotic dance poses. After a robotic dance pose is visually perceived from a mirror image, three kinds of features (color, shape, orientation) are extracted separately, and then selected and combined to form a suitable feature combination to completely describe a robotic dance pose. Finally, a machine aesthetics model of robotic dance poses is trained by machine learning, and the automatic judgments of robotic dance poses are achieved.

The main contributions of this paper are as follows:

- Based on the two influential aesthetic cognitive neural models [3–5], this paper proposes a new computable cognitive model of visual aesthetics, which can be used for aesthetic calculations of visual arts. The proposed model enables robots to process visual stimulus in a way similar to the human brain processing mechanism, which helps to improve the cognitive ability of robots;
- By applying the proposed computable cognitive model of visual aesthetics to the aesthetics problem of robotic dance poses, this paper proposes a new approach to achieve automatic aesthetics evaluations of robotic dance poses. The approach enables a robot to exhibit more autonomous and human-like behaviors;
- The feature combination, namely *color* + *shape* + *orientation*, can better portray robotic dance poses from vision and make the trained machine aesthetic model more accurately evaluate the aesthetics of robotic dance poses. Moreover, the machine learning method, *Random Forest*, has been verified to be effective;
- An 8-bit orientation encoder and the method of "the spatial distribution histogram of color block" are designed in this paper to have good effects on the orientation characterization of robotic dance poses. They provide a novel means to express the spatial relation of different parts of the same object in an image.

The rest of the paper is structured as follows: Section 2 outlines the related work, and Section 3 describes in detail the computable cognitive model of visual aesthetics. Based on the proposed model, the new approach of automatic aesthetics evaluation of robotic dance poses is presented in Section 4. Section 5 reports the simulation experiment being conducted, and Section 6 discusses the experimental results. Finally, a brief conclusion and future work are presented in Section 7.

2. Related Work

2.1. Aesthetic Cognitive Neural Model

Aesthetics is a complex mental process and involves many sub-processes (such as perception, memory, and judgment). It is accompanied by a variety of neurophysiological activities. There are only few existing aesthetic cognitive neural models, and the two models in [3–5] are important and influential. This is because neuro-aesthetics is a new discipline and has many unknown cognitive neural mechanisms to be explored.

Chatterjee proposed the first cognitive neural model of visual aesthetics in [3,4], in which the visual aesthetic neural mechanism was divided into three processing stages: early, middle and late. In the early processing stage, many kinds of visual elements (such as color, shape, and orientation) are extracted and analyzed in the different areas of the human brain. In the middle stage, these visual elements are combined or split automatically to form a unified characterization. In the late stage, some aspects of the aesthetic object are selectively processed further, and given meaning by activating memory. The corresponding emotional response is triggered and fed back to the aesthetic processing system by the attention mechanism. Aesthetic preference and judgment is made by the aesthetic subject.

Höfel and Jacobsen [5] proposed another aesthetic processing model with three stages, namely the receptive, central and productive processes. In the receptive process, many areas of the human brain related to perceptual processing are activated, and the aesthetic object is processed by perception. In the central process, the previous knowledge is recalled, the aesthetic value of object is calculated, and an aesthetic judgment is made. In the productive process, some behaviours are exhibited and aesthetic expressions can be achieved by different art activities (painting, music, poetry, dance, etc.).

Leder et al. [6] proposed an information-processing stage model of aesthetic processing, which has five stages: perception, explicit classification, implicit classification, cognitive mastering and evaluation. Each stage includes the corresponding sub-process procedure and influence factors. However, this model is difficult to implement by computation for two reasons: (i) it is built on the complex psychological process level; (ii) its sub-processes are difficult to divide into the corresponding neural parts.

Redies [7] proposed a model of visual aesthetic experience that combines formalist and contextual aspects of aesthetics. In the model, a bottom–up perceptual channel and a top–down cognitive channel process visual information parallel to each other, and work together to trigger an aesthetic experience. The model includes three stages: the acquisition and processing of an external visual information, information encoding, and aesthetic experience. The model considers the differences between aesthetic perception and cognition, and it can explain the aesthetic activities of human brain in response to artwork to a certain extent.

Koelsch et al. [8] proposed an integrative and neurofunctional model for human emotions. From the perspective of neural motion processing, the model includes four emotional core systems (the affect, effector, language, and cognitive appraisal systems) that constitute a quartet of human emotions. The model not only includes the neural mechanism of emotional perception and cognition, but also considers many factors, such as psychology, neurophysiology and language psychology.

2.2. Aesthetic Evaluation on Robotic Dance Object

The robotic choreography creation requires aesthetic evaluations of dance objects. Based on different research purposes, aesthetic evaluations are carried out in terms of three aspects: dance pose, dance motion, and dance works [17]. The existing aesthetics evaluation approach to robotic dance poses includes human subjective aesthetics [12–14] and machine-learning-based methods [15–17].

Aiming to improve human–robot interactions, Vircikova et al. [12–14] used interactive human subjective aesthetic evaluation to create robotic choreography. Robotic dance poses and robotic dance motions were evaluated by 20 people, and the evaluation results guided the direction of evolutionary computation. In [15], we designed an algorithm, namely semi-interactive evolutionary computation, to generate good robotic dance poses automatically. In this procedure, a trained machine aesthetic model was built dynamically and used to evaluate all the robotic dance poses. The best machine aesthetic model obtains a 71.6667% correct ratio on AD Tree, based on joint feature.

In order to train a suitable machine aesthetic model and solve the problem of the automatic aesthetics evaluation of robotic dance poses (mentioned in Section 1), we presented two different approaches in [16,17] based on multi-modal information fusion (visual and non-visual information of robots). Three kinds of features (joint feature, region shape feature, and contour shape feature) of robotic dance poses were extracted and fused to become a mixed feature, and then inputted into several machine-learning methods for training [16]. The best machine aesthetic model obtained an 81.6% correct ratio on AD Tree. In [17], four kinds of features (the kinematic feature, region shape feature, contour shape feature, and spatial distribution feature of color block) of robotic dance poses were extracted, and all the feature combinations were considered as the inputs of the machine-learning methods to train. It was confirmed that the best feature combination was kinematic + region + spatial distribution of color block, and the best machine aesthetic model obtained an 81.8% correct ratio on AD Tree.

Furthermore, Eaton [18] used the method of "Performance Competence Evaluation Measure" [19], as a specific implementation of fitness function in traditional evolutionary computation, to evaluate the aesthetics of robotic dance motions. A synthesis approach to humanoid robotic dance was then proposed. By mapping rules between robotic dance motions and multi-module events, Oliveira et al. [20] constructed a robot dancing framework to empirically evaluate aesthetics, and achieved positive results. Based on the Hidden Markov Model (HMM), Manfrè et al. [21] designed an automatic system for robotic choreography creation, and then used human subjective aesthetics to evaluate the creation results by matching robotic dance motions with music.

To explore live performances, Augello et al. [22] presented a cognitive architecture based on human–robot interactions. The architecture integrated HMM and genetic algorithms to create and perform improvisational robotic dance according to the perceived music. The positive feedback from the audience was used as the evaluation result of human subjective aesthetics. Moreover, as an extension of the creative HMM system for robot dancing, Manfrè et al. [23] built a computational creativity framework for a robot to learn dance motions from human demonstrations. The good feedback from the audience was acquired as the evaluation result of human subjective aesthetics.

Qin et al. [24] proposed a music-driven dance system for humanoid robots, which used HMM to match dance phrases with the perceived musical emotions. Twenty evaluators were invited to use questionnaires to evaluate the subjective aesthetics of the created robotic dance. They believed the system could create satisfactory robotic dances.

3. The Computable Cognitive Model of Visual Aesthetics

As mentioned above, two important aesthetic cognitive neural models [3–5] were proposed to describe the inner psychological or physiological activities of human beings. Based on these two models, this paper proposes a new computable cognitive model of visual aesthetics, aiming to lead the way in the aesthetic calculation of visual arts. To describe the computable cognitive model of visual aesthetics more clearly, we describe its function framework, computation framework, neural processing framework, and the corresponding relationship of components between the three frameworks separately.

3.1. Function Framework

Inspired by the three-stage model of aesthetic processing in [5], we propose the function framework in Figure 1, which consists of three stages: visual perception, neural processing decision, and behavior extension. In the visual perception stage, aesthetic objects are found from a visual stimulus and then separated from their background. Some important and suitable features are used to portray these aesthetic objects. The feature types color, shape, and orientation of visual aesthetic processing are still used in our visual perception stage.

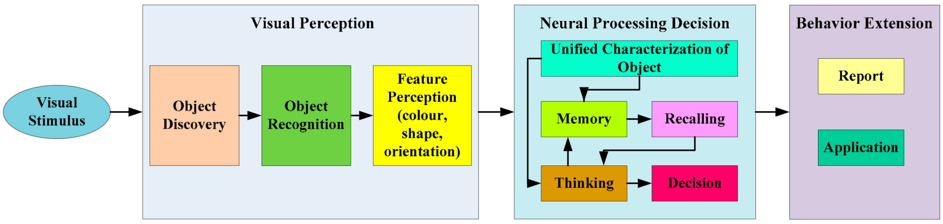


Figure 1. The function framework.

In the neural processing decision stage, the above features are combined or split automatically to form a unified characterization, which is same as the task of the middle processing stage of the visual aesthetic cognitive neural model in [3,4]. After this, aesthetic objects are characterized uniformly by a suitable feature combination, and the following nervous activities begin to work: memory, recalling, thinking, and decision making. Among these nervous activities, memory is used for knowledge storage, recalling is used for acquiring previous knowledge, thinking is used for the accumulation and usage of aesthetic experience, and decision making is used for aesthetic judgment.

Furthermore, there are two possible pathways after aesthetic objects are characterized uniformly. One pathway shows that some aesthetic objects are sent to the memory, which will be further extracted by recalling. Moreover, the results of thinking are stored in the memory, and will be used at a suitable time. Another pathway shows that some new aesthetic objects will be thought of directly using prior knowledge before a decision is made.

In the behavior extension stage, report and application are the two main means of aesthetic expression. Among them, report is used to orally express the results of aesthetic processing and application is used to exhibit some overt behaviors (such as painting, music, poetry and dance). The computable cognitive model of visual aesthetics proposed in this paper is based on the two aesthetic cognitive neural models [3–5], whose functions will be compared in Section 6.2.

3.2. Computation Framework

Based on the above functional framework, we can deploy computer vision and machine learning technology to make it computable. As shown in Figure 2, our computational framework has the same three stages as the functional framework: visual perception, neural processing decision and behavior extension. Each stage will be implemented by some specific algorithms from computer vision or machine-learning technology. In the visual perception stage, an aesthetic object (target) that appears in a visual image/video is located/tracked automatically. Then the target is segmented from its background, and three kinds of features (color, shape, and orientation) are extracted to portray the aesthetic object from different perspectives.

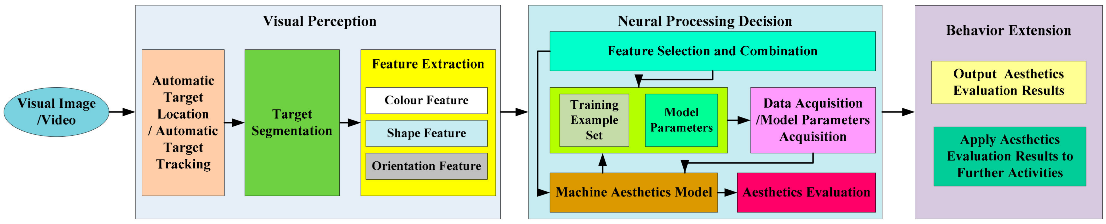

Visual Image /Video
Visual Perception
Automatic Target Location / Automatic Target Tracking
Target Segmentation
Feature Extraction
Colour Feature
Shape Feature
Orientation Feature
Neural Processing Decision
Feature Selection and Combination
Training Example Set
Model Parameters
Data Acquisition /Model Parameters Acquisition
Machine Aesthetics Model
Aesthetics Evaluation
Behavior Extension
Output Aesthetics Evaluation Results
Apply Aesthetics Evaluation Results to Further Activities

Figure 2. The computation framework.

In the neural processing decision stage, a suitable feature combination will be found based on the aforementioned three features (color, shape, and orientation), which is regarded as a more complete and accurate description of an aesthetic object. Considering the formation of human aesthetic experience is a procedure of supervised learning, a training example set of aesthetic objects is necessary for knowledge reserve, which is an important part of memory's implementation. By acquiring the training example set, a good machine aesthetics model will be trained, and the corresponding model parameters will be saved as another important part of the memory's implementation. When a new aesthetic object needs to be judged, the trained machine aesthetics model will be restored by acquiring the saved model parameters, and then an automatic aesthetics evaluation will be achieved.

In the behavior extension stage, the results of the aesthetics evaluation can be outputted, which is regarded as the implementation of report. Moreover, as the implementation of application, the results of aesthetics evaluation can also be used to direct further activities, such as creation and education.

3.3. Neural Processing Framework

As the neural basis of the function framework, the neural processing framework presents the specific locations where neural activities occur and the related visual aesthetic activities happen. As our model builds on the two aesthetic cognitive neural models [3–5]; all the function components and their corresponding neural bases in our model draw from the two models. Meanwhile, the knowledge of cognitive neuroscience [25] is also utilized.

Figure 3 shows our neural processing framework, in which the visual perception stage corresponds to the occipital cortex and frontal cortex, which are related to visual processing. Moreover, the neural processing decision stage corresponds to two parts: (i) the neural circuit of attention in the frontal–parietal cortex, which is related to the unified characterization of objects; (ii) the prefrontal cortex and cingulate cortex, which are related to working memory, emotional response and cognitive control. Finally, the behavior extension stage corresponds to the motor cortex, which is related to action/behavior control.

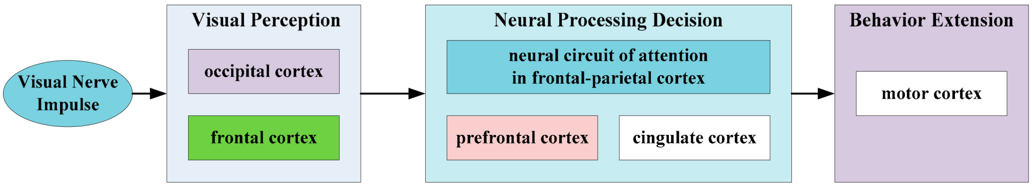


Figure 3. The neural processing framework.

3.4. Corresponding Relationship of Components

The function framework, the computation framework and the neural processing framework portray the same computable cognitive model of visual aesthetics from three different perspectives. The proposed computable cognitive model of visual aesthetics can be described by combining all of them. Table 1 shows a closely correlated relationship of the components in the three frameworks.

Table 1. The corresponding relationship of components among the three frameworks.

Function Framework	Computation Framework	Neural Processing Framework
visual stimulus	visual image/video	visual nerve impulse
object discovery	automatic target location/automatic target tracking	occipital cortex and frontal cortex
object recognition	target segmentation	
feature perception	feature extraction	
unified characterization of object	feature selection and combination	neural circuit of attention in frontal–parietal cortex
memory	training example set and model parameters	prefrontal cortex and cingulate cortex
recalling	data acquisition/model parameters acquisition	
thinking	machine aesthetics model	
decision	aesthetics evaluation	
report	output aesthetics evaluation results	motor cortex
application	apply aesthetics evaluation results to further activities	

4. Automatic Aesthetics Evaluation of Robotic Dance Poses

To imitate human dance behavior, the above computable cognitive model of visual aesthetics is deployed to evaluate robotic dance poses as a way that human dancers make autonomous aesthetics evaluations of their own dance poses as they appear in mirrors. Figure 4 shows our instantiated computation framework used for the automatic aesthetics evaluation of robotic dance poses.

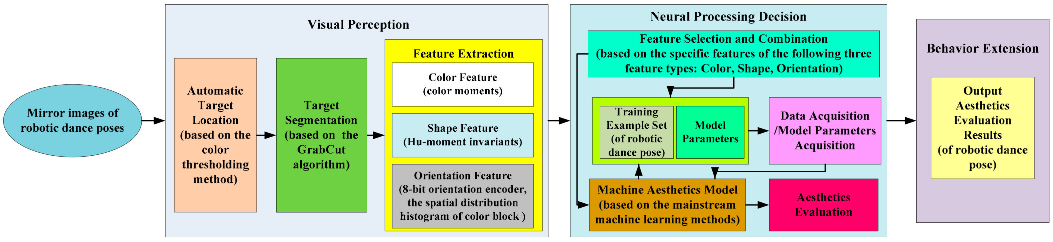

Mirror images of robotic dance poses
Visual Perception
Automatic Target Location (based on the color thresholding method)
Target Segmentation (based on the GrabCut algorithm)
Feature Extraction
Color Feature (color moments)
Shape Feature (Hu-moment invariants)
Orientation Feature (8-bit orientation encoder, the spatial distribution histogram of color block)
Neural Processing Decision
Feature Selection and Combination (based on the specific features of the following three feature types: Color, Shape, Orientation)
Training Example Set (of robotic dance pose)
Model Parameters
Data Acquisition /Model Parameters Acquisition
Machine Aesthetics Model (based on the mainstream machine learning methods)
Aesthetics Evaluation
Behavior Extension
Output Aesthetics Evaluation Results (of robotic dance pose)

Figure 4. The instantiated computation framework on the automatic aesthetics evaluation of robotic dance poses.

This instantiated computation framework includes the following two features:

- Accumulation of aesthetic experience. After abundant mirror images of robotic dance poses are perceived visually and characterized uniformly, a corresponding training example set is built to train a machine aesthetic model. Such a supervised learning procedure enables the robot to possess the aesthetic abilities of human dancers [15];
- Aesthetic judgment. When the training of the machine aesthetic model is finished, a new mirror image of a robotic dance pose can be evaluated automatically on its aesthetics.

The details of this instantiated computation framework are described in the following Sections.

4.1. *Automatic Target Location*

A dancing robot can use its cameras to observe and acquire its own dance poses from a mirror. These mirror images are generally RGB images. This paper uses the method of color thresholding [16] to implement automatic target location. As a prerequisite of the color thresholding method, a dancing robot should have unique color blocks on the important parts of its body, such as its head, shoulder, hand, foot and leg. The color blocks should differ from its embodied environment.

The color blocks on the robot are filtered from the original image, and all the other colors are converted to a background color (e.g., black). An approximate minimum enclosing rectangle (AMER) is then used to cover all the specific color pixels based on their spatial distribution. Finally, the AMER bounding box is used to identify the accurate position of the dancing robot. Figure 5a shows an original image of a robotic dance pose, and Figure 5b shows the result of automatic target location. Furthermore, Figure 5c shows the result of target segmentation; Figure 5d is the shape sub-image; and Figure 5e is the spatial distribution sub-image of color blocks.

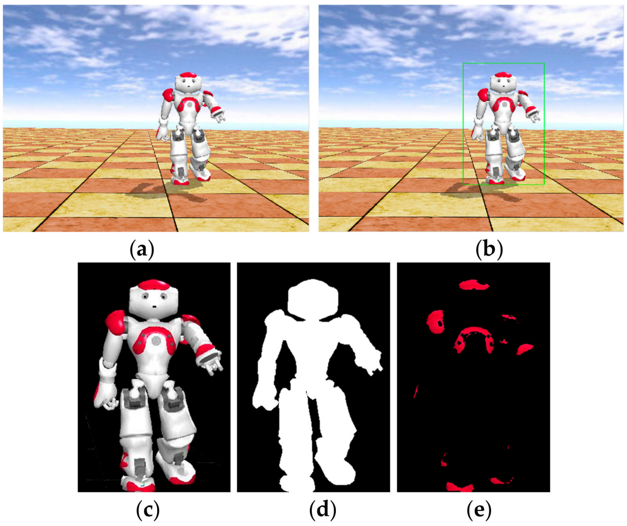


Figure 5. An example of visual perception: (**a**) the original image; (**b**) the result of automatic target location (with green rectangle); (**c**) the result of target segmentation; (**d**) the shape sub-image; (**e**) the spatial distribution sub-image of color blocks.

4.2. *Target Segmentation*

As an important part of visual perception, target segmentation separates a target from its background. By utilizing the accurate position of the dancing robot identified by AMER, we could separate a sub-image of the robot ontology from an original image of a robotic dance pose by deploying the GrabCut algorithm [26].

The GrabCut algorithm is an interactive foreground extraction algorithm using iterated graph cuts and requires users to mark the foreground object by drawing a rectangle around it. Therefore, AMER is inputted to the GrabCut algorithm as the replacement of user's interactive operation. Meanwhile, an energy function, E, is defined in the GrabCut algorithm so that its minimum value corresponds to a good segmentation. Figure 5c shows an example of the result of target segmentation.

4.3. *Feature Extraction*

Based on the sub-image of a robotic dance pose (the result of target segmentation), the following two sub-images are further generated: the shape sub-image in Figure 5d, and the spatial distribution sub-image of color blocks in Figure 5e. To acquire the shape sub-image, we use the shape extraction algorithm in [16]: an RGB color image is converted firstly to a black-and-white image, and the eight-connected breed filling algorithm is then used to fill holes and two further morphological operations (corrode and dilate) are conducted; finally, the shape sub-image is extracted. Moreover, the spatial distribution sub-image of color blocks is extracted by the color thresholding method [16] (mentioned in Section 4.1), and the sub-image of a robotic dance pose (the result of target segmentation) is regarded as the input.

Based on the three sub-images above, three kinds of feature type (color feature, shape feature, and orientation feature) are considered separately for extraction. To implement the related computation, the three feature types are also instantiated to be three specific features, which are described in the following Sections.

4.3.1. Color Feature

The method of color moments is a simple and effective color feature description method [27]. In this paper, we use it as the specific color feature of a robotic dance pose, and the sub-image of the robotic dance pose (the result of target segmentation) is used as the input of the proposed method. Any color distribution in an image can be described by moments that have very low dimensions.

As the color distribution information is mainly concentrated in low order moments, the following three moments are sufficient to represent the color distribution in an image: the first moment (mean) E_i, the second moment (variance) σ_i, and the third moment (skewness) s_i. The definitions of the three moments are as follows [27]

$$E_i = \frac{1}{N}\sum_{j=1}^{N} p_{ij} \tag{1}$$

$$\sigma_i = \left(\frac{1}{N}\sum_{j=1}^{N}\left(p_{ij} - E_i\right)^2\right)^{\frac{1}{2}} \tag{2}$$

$$s_i = \left(\frac{1}{N}\sum_{j=1}^{N}\left(p_{ij} - E_i\right)^3\right)^{\frac{1}{3}} \tag{3}$$

where p_{ij} is the value of the *i*-th color channel at the *j*-th image pixel, and *N* is the total number of pixels in an image.

Color moments are usually calculated in the RGB space, and each color channel (R/G/B) is involved. Thus, nine color moments (E_R, $\sigma_{R,}$ $s_{R,}E_G$, $\sigma_{G,}$ $s_{G,}$ E_B, $\sigma_{B,}$ s_B) in an inputted RGB image will be calculated to represent the color distribution of the image. If the inputted RGB image is divided into the *t* sub-image equally, the above nine color moments will be calculated on each sub-image, and the corresponding nine color moments of the *k*-th sub-image are expressed as ($E_{k,R}$, $\sigma_{k,R,}$ $s_{k,R,}E_{k,G}$, $\sigma_{k,G,}$ $s_{k,G,}$ $E_{k,B}$, $\sigma_{k,B,}$ $s_{k,B}$). There are *9*t* color moments in total for the inputted RGB image. Conveniently, the *9*t* color moment features (CMF) are expressed as (CMF_1, CMF_2, CMF_3, ... , CMF_{9*t-2}, CMF_{9*t-1}, CMF_{9*t}) (t = 1, 2, 3, ...).

4.3.2. Shape Feature

As an important feature description method of the region shape, Hu-moment Invariants use the nonlinear combinations of geometric moments to portray the statistical properties of an image, and are invariant against translation, rotation and scaling [28]. Therefore, we use it to calculate the specific shape features of a robotic dance pose, and the shape sub-image is regarded as the input of the method. There are seven invariant moments in total, and their definitions are as follows [28]:

$$\varphi_1 = \tau_{20} + \tau_{02} \tag{4}$$

$$\varphi_2 = (\tau_{20} - \tau_{02})^2 + 4\tau_{11}^2 \tag{5}$$

$$\varphi_3 = (\tau_{30} - 3\tau_{12})^2 + (\tau_{03} - 3\tau_{21})^2 \tag{6}$$

$$\varphi_4 = (\tau_{30} + \tau_{12})^2 + (\tau_{03} + \tau_{21})^2 \tag{7}$$

$$\begin{aligned}\varphi_5 = {} & (\tau_{30} - 3\tau_{12})(\tau_{30} + \tau_{12}) * \left[(\tau_{30} + \tau_{12})^2 - 3(\tau_{21} + \tau_{03})^2\right] \\ & +(\tau_{03} - 3\tau_{21})(\tau_{03} + \tau_{21})\left[(\tau_{03} + \tau_{21})^2 - 3(\tau_{30} + \tau_{12})^2\right]\end{aligned} \tag{8}$$

$$\begin{aligned}\varphi_6 = {} & (\tau_{20} - \tau_{02}) * \left[(\tau_{30} + \tau_{12})^2 - (\tau_{03} + \tau_{21})^2\right] \\ & +4\tau_{11}(\tau_{30} + \tau_{12})(\tau_{03} + \tau_{21})\end{aligned} \tag{9}$$

$$\begin{aligned}\varphi_7 = {} & (3\tau_{21} - \tau_{03})(\tau_{30} + \tau_{12})\left[(\tau_{30} + \tau_{12})^2 - 3(\tau_{03} + \tau_{21})^2\right] \\ & +(\tau_{30} - 3\tau_{12})(\tau_{03} + \tau_{21})\left[(\tau_{03} + \tau_{21})^2 - 3(\tau_{30} + \tau_{12})^2\right]\end{aligned} \tag{10}$$

where τ_{jk} is the normalized *(j + k)*-order central moment

$$\tau_{jk} = \frac{M_{jk}}{(M_{00})^r}, \quad r = \left[\frac{j+k}{2} + 1\right] \tag{11}$$

Moreover, M_{jk} is the *(j + k)*-order central moment based on the region shape *f(x,y)* in a discrete digital image

$$M_{jk} = \sum_{y=1}^{V}\sum_{x=1}^{U}(x-\bar{x})^{j}(y-\bar{y})^{k}f(x,y);\ j,k = 0,1,2,\ldots \tag{12}$$

where V and U refer respectively to the height and width of the image, and $(\bar{x}, \bar{y})$ refers to the barycenter of the region shape.

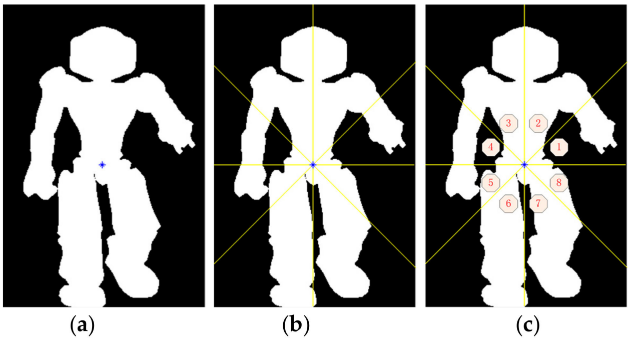
Thus, to describe the region shape feature of a robotic dance pose, Hu-moment Invariants can be calculated for the corresponding shape sub-image. Conveniently, Hu-moment Invariants are expressed as ($\mathrm{HuMIF}_1, \ldots \ldots, \mathrm{HuMIF}_q$) ($1 \le q \le 7$) in this paper, which should be further normalized before machine learning is implemented.

4.3.3. Orientation Feature

In an image, the orientation information of an object reflects its direction, position and motorial tendency, and can be acquired by analyzing the spatial relationship of the components of an object. Specifically, to acquire the orientation information of a robotic dance pose from an image, the spatial relationship of the body parts of a robotic dance pose needs to be analyzed. Thus, we have designed an orientation encoder and a new method of "the spatial distribution histogram of color block" to analyse the above spatial relationship. The orientation feature of a robotic dance pose will be extracted based on the shape's sub-image and the spatial distribution sub-image of color blocks.

As mentioned in Section 4.3, the spatial distribution sub-image of color blocks is acquired by color thresholding, with the prerequisite that a dancing robot should have unique color blocks on the important parts of its body (such as head, shoulder, hand, foot, and leg). Thus, the spatial relationship of the body parts of a robotic dance pose can be acquired from the analysis of the corresponding spatial distribution of color blocks, and the orientation feature of a robotic dance pose can then be extracted.

Figure 6a shows the centroid of the region shape of a robotic dance pose that is first calculated based on the shape's sub-image. Figure 6b,c show that the whole space of the image is divided into eight sequentially numbered quadrants by four lines, with the center as the centroid of the region shape. Specifically, the division method of eight quadrants uses the pixel coordinate system, which includes the following steps: (i) going through the centroid of the region shape, two straight lines parallel to the *x*-axis and *y*-axis respectively are drawn (see Equations (13) and (14), (ii) going through the centroid of the region shape, two straight lines with slopes of 1 and -1 are drawn (see Equations (15) and (16).



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

Figure 6. An example diagram of the division method of eight quadrants: (**a**) the centroid of the region shape of a robotic dance pose (the blue point) in the shape sub-image; (**b**) the division of eight quadrants based on the centroid; (**c**) the eight sequentially numbered quadrants ([1,8]).

Figure 7 shows a detailed diagram of the division method of eight quadrants, and the Equations (13)–(16) show the equations of the above four straight lines. The divided eight quadrants are sequentially numbered ([1,8]). Note that point *O (0,0)* in Figure 7 is the origin of the pixel coordinate system, and the point $G\ (x_0, y_0)$ is the centroid of the region shape, and c is the width of image, and r is the height of image.

$$y = y_0 \tag{13}$$

$$x = x_0 \tag{14}$$

$$y = x + (y_0 - x_0) \tag{15}$$

$$y = -x + (y_0 + x_0) \tag{16}$$

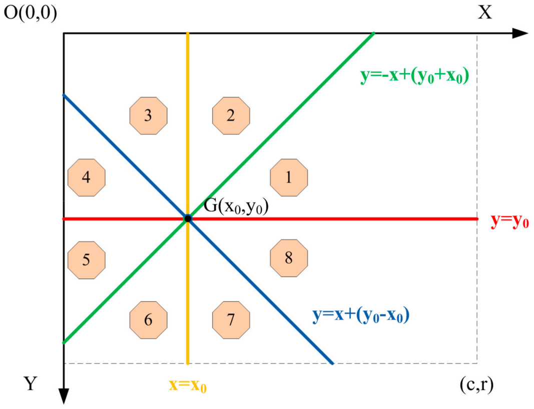


Figure 7. A detailed diagram of the division method of eight quadrants.

As shown in Figure 8a–e, some image processing methods (including greying, binarization, corrosion, and dilation) are applied to the spatial distribution sub-image of color blocks (see Algorithm 1 in [17]). The centroids of the color blocks are calculated and shown in the corresponding processed spatial distribution sub-image, as shown in Figure 8f. Figure 8g,h show that the centroids are re-projected into the eight quadrants of the shape sub-image.

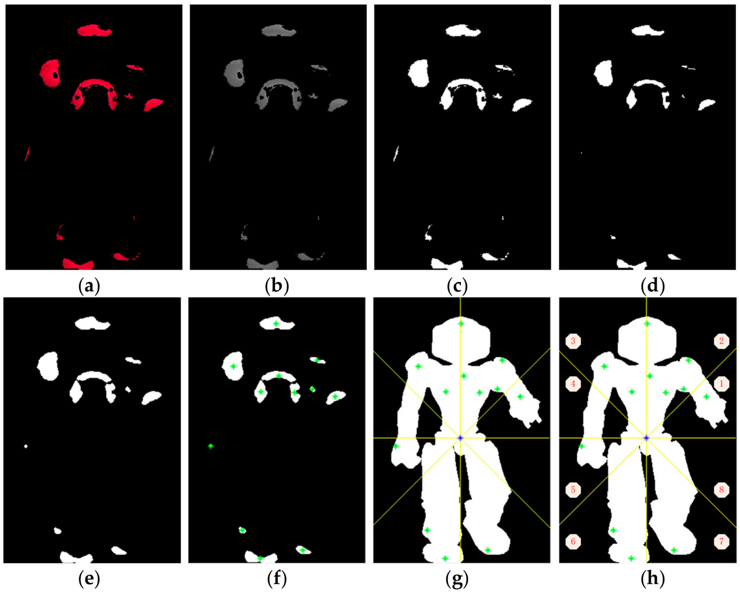


Figure 8. The acquisition procedure of the distribution of color blocks' centroids in the eight quadrants of the shape sub-image: (**a**) the spatial distribution sub-image of color blocks; (**b**) the single grey image; (**c**) the black-and-white image; (**d**) the corroded image; (**e**) the dilated image; (**f**) the centroids of the color blocks are shown in the corresponding processed spatial distribution sub-image (subfigure (**e**)); (**g**) the centroids are re-projected into the eight quadrants of the shape sub-image; (**h**) the centroids are re-projected into the eight, sequentially numbered quadrants of the shape sub-image.

The different robotic dance poses introduce different spatial distributions of the color blocks' centroids into the corresponding eight quadrants of shape sub-image, and vice versa. According to the spatial distributions of color blocks' centroids in the eight quadrants of the shape sub-image, we compare coordinate positions counting the following two pieces of information: (i) whether the centroid of a color block appears in a specific quadrant (ii) how many centroids of color blocks appear in a specific quadrant.

To find out the former, the related situation of each quadrant is counted respectively, and the whole situation of the eight quadrants is aggregated, which is described by an 8-bit orientation encoder (an 8-bit integer). In the 8-bit orientation encoder, each bit corresponds to a quadrant. Specifically, the *i*-th bit describes whether the centroid of a color block appears in the *i*-th-specific quadrant from right to left. If so, the *i*-th bit takes on the value of 1; otherwise, it takes the value of 0. Thus, for an image of a robotic dance pose, there is a whole distribution of the color blocks' centroids into eight quadrants, which is described by an 8-bit integer.

When a color blocks' centroids appear in every quadrant, it is the most complete situation, which has the 8-bit code 11,111,111 (corresponding to the decimal number 255). To generate an orientation feature for an image of a robotic dance pose, an 8-bit orientation code needs to be further normalized in the following way: the 8-bit orientation code (the 8-bit integer) is divided by 255. Conveniently, the above orientation feature is expressed as an appearance orientation feature (AOF). Regarding the example shown in Figure 8h, the corresponding 8-bit orientation code is 01,110,111 (corresponding to the decimal number 119), and the corresponding normalized result is 0.4667 (calculated by 119/255, retaining four figures after the decimal point).

Furthermore, we have designed a method of "the spatial distribution histogram of color block" to calculate how many centroids of color blocks appear in a specific quadrant. The method is detailed as follows. Supposing there are *W* color blocks, which have the same specific color, in a color image, and the whole space of the image is divided into *Z* quadrants (the division method can be designed according to the specific task), there are *W* centroids of color blocks.

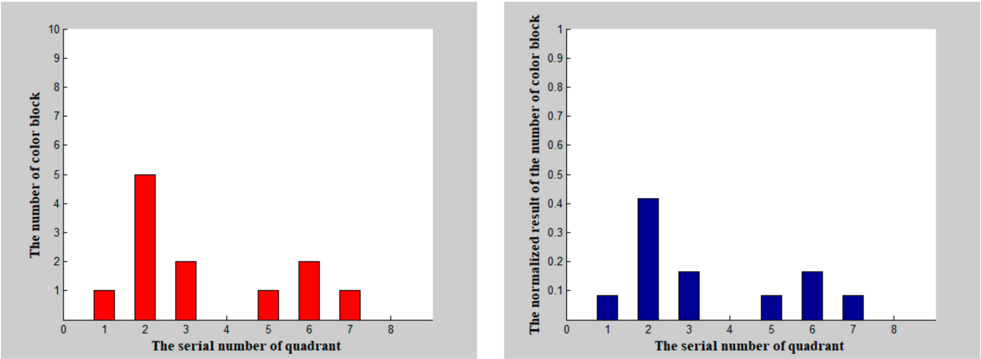
Meanwhile, supposing *T(i)* is the number of centroids of color blocks, which appear in the *i*-th quadrant of the image, then the original spatial distribution histogram of color block *H* is defined as follows

$$\mathrm{h_i} = \mathrm{T(i)},\ \mathrm{i} = 1, 2, \ldots, \mathrm{Z} \tag{17}$$

and the corresponding normalized spatial distribution histogram of color block *H′* is defined below:

$$\mathrm{h'_i} = \frac{h_i}{\mathrm{W}},\ \mathrm{i} = 1, 2, \ldots, \mathrm{Z} \tag{18}$$

Regarding the example shown in Figure 8h (*W* = 12, *Z* = 8), the original spatial distribution histogram of color block *H* is (1, 5, 2, 0, 1, 2, 1, 0), and the normalized spatial distribution histogram of color block *H′* is (0.0833, 0.4167, 0.1667, 0, 0.0833, 0.1667, 0.0833, 0), and the corresponding figures are shown in Figure 9.



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

Figure 9. The example spatial distribution histograms of the color block: (**a**) the original spatial distribution histogram of color block *H*; (**b**) the normalized spatial distribution histogram of color block *H′*.

In this paper, $\left(h_1', h_2', h_3', h_4', h_5', h_6', h_7', h_8'\right)$ $(Z = 8)$ are used to store the normalized quantitative statistics of the spatial distributions of color blocks' centroids, which appear in the eight quadrants of the image of a robotic dance pose. Therefore, the above normalized spatial distribution histogram of color block $\left(h_1', h_2', h_3', h_4', h_5', h_6', h_7', h_8'\right)$ $(Z = 8)$, is regarded as eight orientation features. Conveniently, they are expressed as (CHF_1, CHF_2, . . . , CHF_Z) (Z = 8). Generally, the orientation feature of a robotic dance pose is expressed as (AOF, CHF_1, CHF_2, . . . , CHF_Z) (Z = 8).

4.4. Feature Selection and Combination

As mentioned in Section 3.1, different kinds of features are combined or split automatically to form a unified characterization of an object. As a result, feature selection and combination need to find a suitable unified characterization of a robotic dance pose, based on three kinds of features (color feature, shape feature, and orientation feature) acquired from the feature extraction procedure. Each feature type portrays some properties of a robotic dance pose in a certain aspect, and three feature types bring a diversity of feature combinations [17]. Thus, there will be a higher possibility of discovering the most appropriate feature combination to characterize a robotic dance pose more completely.

Moreover, three kinds of features bring a total of seven feature combinations, and the most complete feature combination is (CMF_1, CMF_2, CMF_3, . . . , CMF_{9*t-2}, CMF_{9*t-1}, CMF_{9*t}, $HuMIF_1$, . . . , $HuMIF_q$, AOF, CHF_1, CHF_2, . . . , CHF_Z) (t = 1, 2, 3, ...; $1 \leq q \leq 7$; Z = 8). However, including more features in a feature combination does not make it better as there may be conflicts among the different feature types [17]. Therefore, the only way to find out the most appropriate feature combination is through experimental results, and the above seven feature combinations were tested in our experiments.

4.5. Machine Aesthetics Model

A machine aesthetics model aims to make machines achieve automatic aesthetics evaluations on new aesthetic objects, as a human dancer does. It works in the following ways: accumulation of aesthetic experience, and aesthetic judgment, as described in Section 4.

To train a machine aesthetics model of robotic dance poses, we prepared a corresponding training example set and adopted supervised learning [17]. Human dance experts were invited to label every robotic dance pose they observed, with an aesthetic evaluation (good/bad). Thus, each example of robotic dance pose is expressed using an instance of a specific feature combination and an aesthetic label (good/bad).

Meanwhile, good effects of the machine aesthetics model depend on an appropriate feature combination and a suitable machine-learning method [17]. Therefore, a diversity of feature combinations and machine-learnings method could produce different effects from a machine aesthetics model. The best solution should be found through experiments in which several mainstream machine learning methods will be used to verify the different effects of machine aesthetics model.

Notably, the training example set provides knowledge of human's aesthetic understanding of robotic dance poses, and machine-learning teaches a robot to learn and handle the above knowledge, so the robot can possess the aesthetic ability of human dance experts to a certain extent. Specifically, the effect of the machine aesthetics model reflects the accordant extent of aesthetic ability between robots and human dance experts. Essentially, machine-learning realizes the end-to-end transmission of knowledge from human dance experts to robots in an implicit way, and the knowledge has the characteristics of unity and continuity.

5. Experiments

To verify the effectiveness of the proposed computable cognitive model of visual aesthetics, we conducted simulation experiments. The experimental tools used include the Webots 7.4.1 simulator used to simulate the required original "mirror" images of robotic dance poses, and Weka 3.6, used for machine-learning.

When a simulated NAO robot presented a dance pose in Webots, the picture shown in the "Simulation View" area of Webots was treated as the original "mirror" image of the presented robotic dance pose and observed from a "mirror" by the robot. Based on the constrained random generation method [15,17,29], 500 robotic dance poses of Chinese Tibetan Tap were generated randomly, and then the corresponding original "mirror" images of these robotic dance poses were acquired from Webots. These randomly generated dance poses have two constraints [17]: (i) they combine the innovativeness and preservation of dance characteristics of humans; (ii) dance poses causing a robot to fall are excluded.

Furthermore, to prepare a corresponding training example set for supervised learning, a Chinese folk dance expert was invited to label the 500 robotic dance poses with an aesthetic label (good/bad). To acquire the color features from the original images of robotic dance poses, each image was divided into 25 (t = 25) sub-images equally, and nine color moments were calculated on each sub-image.

Then, the color feature of each robotic dance pose was expressed as (CMF_1, CMF_2, CMF_3, … , CMF_{9*t-2}, CMF_{9*t-1}, CMF_{9*t}) (t = 25). Moreover, seven Hu-moment Invariants were used to describe the shape feature of each robotic dance pose, which was expressed as (HuMIF_1, … , HuMIF_q) (q = 7). Additionally, the orientation feature of each robotic dance pose was acquired and was expressed as (AOF, CHF_1, CHF_2, … , CHF_Z) (Z = 8) using the eight quadrants division on the shape sub-image.

To portray a robotic dance pose, there were three kinds of feature: color (CMF_1, CMF_2, CMF_3, … , CMF_{9*t-2}, CMF_{9*t-1}, CMF_{9*t}) (t = 25), shape (HuMIF_1, … , HuMIF_q) (q = 7), and orientation (AOF, CHF_1, CHF_2, … , CHF_Z) (Z = 8). The three feature types have seven feature combinations. Moreover, ten mainstream machine learning methods were used to train the machine aesthetics models based on each feature combination, and 10-fold cross validation was used for performance evaluation.

Table 2 shows the detailed results of the machine aesthetics. Each percentage datum refers to the correct ratio of machine aesthetics evaluation, which was acquired by applying a specific machine-learning method to a specific feature combination. The highest correct ratio of machine aesthetics evaluation (81.6%) came from the Random Forest method, based on the most complete feature combination *color* + *shape* + *orientation*.

Table 2. The effect comparison of different machine learning methods based on different feature combinations.

Machine Learning Method	Feature Combination						
	Color	Shape	Orientation	Color + Shape	Shape + Orientation	Color + Orientation	Color + Shape + Orientation
Naïve Bayes	66.2%	66.6%	73.2%	65%	67.6%	67.8%	67%
Bayesian Logistic Regression	73.8%	73.8%	73.6%	74%	75.4%	73.8%	74.8%
SVM	73.8%	73.8%	73.8%	73.8%	73.8%	73.8%	73.8%
RBF Network	73.8%	73.8%	74.4%	73.8%	75%	73.8%	73.8%
AD Tree	76.4%	75.6%	76.6%	76.4%	76%	76.6%	76.6%
Random Forest	78.4%	70%	71%	80.8%	73.2%	80.6%	81.6%
Voted Perceptron	73.8%	73.6%	75%	73.8%	74.4%	75.2%	75.8%
Bagging	79.2%	75.2%	76.6%	78.6%	74.4%	80.2%	80.2%
Rotation Forest	79.4%	76%	76.4%	81%	76.8%	77.8%	78.8%
LWL	74%	72.2%	76.8%	73.8%	76.8%	75.2%	75.2%

6. Discussion

6.1. Feature Selection and Combination

In this paper, three kinds of feature type (color, shape, and orientation) can be selected to describe a robotic dance pose. Each feature type portrays some properties of a robotic dance pose in a certain aspect, and the three feature types bring a diversity of feature combinations. In theory, the feature combination that mixes the most feature types can describe a robotic dance more completely and accurately, which is verified in Table 2. The best feature combination (*color* + *shape* + *orientation*) was found, and its effect was better than the feature combinations with a single/double feature type. Meanwhile, the effectiveness of each feature type was also demonstrated, which characterizes a robotic dance pose in a certain aspect.

Note that there were some exceptions. For example, although the feature combination *shape* + *orientation* mixed two feature types, its correct ratio of machine aesthetics evaluation from *Bagging* (*74.4%*) was lower than that of the feature combination with the single feature types *shape* (*75.2%*) or *orientation* (*76.6%*). We believe these exceptions are caused by the feature conflicts between the mixed feature types.

Moreover, for feature combinations that mixed two or more feature types, the severity of the feature conflict varies with different machine-learning methods. For example, the feature combination *shape* + *orientation's* correct ratio of machine aesthetics evaluation from *Rotation Forest* (*76.8%*) was higher than that of the feature combination with the single feature type *shape* (*76%*) or *orientation* (*76.4%*). For the feature combination *shape* + *orientation*, the severity of its feature conflict in *Rotation Forest* was notably weaker than that of *Bagging*. We believe that the severity of feature conflict is related to machine-learning methods, and a good machine-learning method can better restrain the adverse effects of feature conflict.

However, regarding the visual characterization of a robotic dance pose, the following problems about feature conflict have yet to be resolved: (i) What are the root causes of feature conflict? (ii) How can feature conflicts be eliminated in the feature fusion? (iii) Which components work to restrain the adverse effects of feature conflict in a machine-learning method?

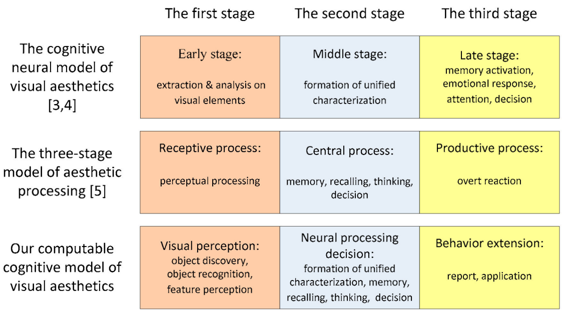
Additionally, the highest correct ratio of machine aesthetics evaluation (*81.6%*) came from the most complete feature combination *color* + *shape* + *orientation*. In our opinion, the reasons why a feature combination that includes more feature types is better are as follows: (1) It is consistent with the visual processing mechanism of the human brain [25], which is the production of natural selection in the long process of human evolution. Extensive physiological evidence shows that different types of visual information (such as color, shape, and orientation) are processed by different visual areas of the human brain, and then all this information is integrated to form recognizable percepts [25]. (2) Many cells in the superior colliculus show a phenomenon of multisensory integration. Specifically, the response of the cell is stronger when there are inputs from multiple senses compared to when the input is from a single modality [30]. Similar to this, we believe there are some mutual promotions among the different feature types of a feature combination.

Furthermore, in our instantiated computation framework for the automatic aesthetics evaluation of robotic dance poses (see Figure 4), three feature types (color, shape, and orientation) were also instantiated using the following specific features: color moments, Hu-moment Invariants, 8-bit orientation encoder, and the spatial distribution histogram of a color block. Certainly, other specific features can be used to highlight a different instance of the proposed computation framework under the three feature types.

For example, color correlogram can be treated as a specific implementation of color feature, and wavelet descriptors can be regarded as a specific implementation of shape feature, and the spatial relationship of color blocks [17] can be regarded as a specific implementation of the orientation feature. Therefore, the diversity of the specific feature highlighted the instantiated computation framework, so the proposed computation framework has a certain degree of universality.

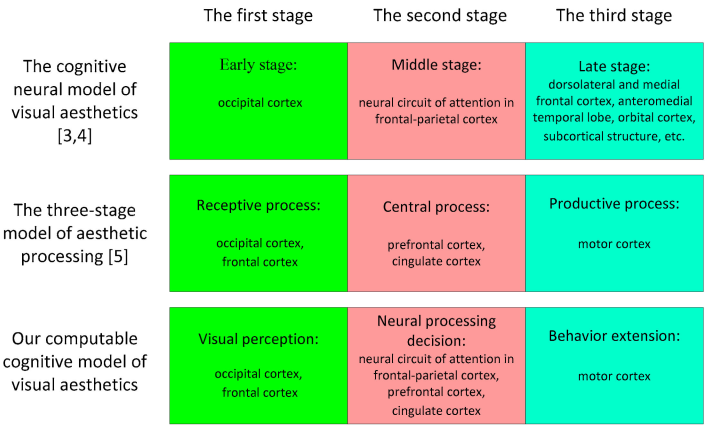
6.2. Comparison with the Related Aesthetic Cognitive Neural Models

Based on the two important aesthetic cognitive neural models [3–5], this paper proposes a new computable cognitive model of visual aesthetics. The above three models have similarities and differences. Figure 10 shows the detailed function comparison and Figure 11 shows the detailed neural basis comparison.



	The first stage	The second stage	The third stage
The cognitive neural model of visual aesthetics [3,4]	Early stage: extraction & analysis on visual elements	Middle stage: formation of unified characterization	Late stage: memory activation, emotional response, attention, decision
The three-stage model of aesthetic processing [5]	Receptive process: perceptual processing	Central process: memory, recalling, thinking, decision	Productive process: overt reaction
Our computable cognitive model of visual aesthetics	Visual perception: object discovery, object recognition, feature perception	Neural processing decision: formation of unified characterization, memory, recalling, thinking, decision	Behavior extension: report, application

Figure 10. The function comparison between the related aesthetic cognitive neural models and our model.



	The first stage	The second stage	The third stage
The cognitive neural model of visual aesthetics [3,4]	Early stage: occipital cortex	Middle stage: neural circuit of attention in frontal-parietal cortex	Late stage: dorsolateral and medial frontal cortex, anteromedial temporal lobe, orbital cortex, subcortical structure, etc.
The three-stage model of aesthetic processing [5]	Receptive process: occipital cortex, frontal cortex	Central process: prefrontal cortex, cingulate cortex	Productive process: motor cortex
Our computable cognitive model of visual aesthetics	Visual perception: occipital cortex, frontal cortex	Neural processing decision: neural circuit of attention in frontal-parietal cortex, prefrontal cortex, cingulate cortex	Behavior extension: motor cortex

Figure 11. The neural basis comparison between the related aesthetic cognitive neural models and our model.

Clearly, our model is close to the three-stage model of aesthetic processing [5], and some important components in the cognitive neural model of visual aesthetics [3,4] (such as the formation of unified characterization), are merged into our model. On a neural basis, they have a similar hypothesis or deduction, integrating the knowledge of cognitive neuroscience [25]. As our model is proposed based on the two models [3–5], the neural basis of our model draws lessons from both.

Additionally, they are different in the following seven aspects: stage task, function component, neural basis, discipline, perceptual channel, computability, and the technology involved. Regarding the function equivalence relation of stage tasks in the three models (shown in Table 3), only the cognitive neural model of visual aesthetics [3,4] lacks a stage task of overt reaction. In our model, the function components of aesthetic processing are refined, specified and integrated (shown in Figure 10), based on the fusion of those in the other two models.

Table 3. The function equivalence relation of stage tasks in the three models.

The Cognitive Neural Model of Visual Aesthetics [3,4]	**The Three-Stage Model of Aesthetic Processing [5]**	**Our Computable Cognitive Model of Visual Aesthetics**
early stage	receptive process	visual perception
middle stage and late stage	central process	neural processing decision
	productive process	behaviour extension

Moreover, the locations where neural activities occur, corresponding to the function components in our model, are also adjusted to specific stages (show in Figure 11). Furthermore, our model focuses on visual stimuli, and is the only computable model among the three models. It can work in the areas of neuro-aesthetics and artificial intelligence (shown in Table 4).

Table 4. The difference between the related aesthetic cognitive neural models and our model.

	The Cognitive Neural Model of Visual Aesthetics [3,4]	**The Three-Stage Model of Aesthetic Processing [5]**	**Our Computable Cognitive Model of Visual Aesthetics**
Discipline	neuroaesthetics	neuroaesthetics	neuroaesthetic, artificial intelligence
Perceptual Channel	visual	visual, auditory, etc.	visual
Computability	no	no	yes
Technology Involved	aesthetics, psychology, cognitive neuroscience, brain science	aesthetics, psychology, cognitive neuroscience, brain science	aesthetics, psychology, cognitive neuroscience, brain science, computer vision, machine learning

6.3. Comparison with the Existing Aesthetics Evaluation Approaches of Robotic Dance Poses

Regarding the aesthetics evaluation of robotic dance poses, the existing literatures focuses on the following two ways of doing so: human subjective aesthetics [12–14] and the machine-learning-based method [15–17]. The latter should be advocated, as it is helpful to develop artificial intelligence for dancing robots. In general, our approach still belongs to the machine-learning-based method. Table 5 compares the existing approaches with our approach.

Table 5. The comparison between the existing approaches and our approach.

	The Approach in [12–14]	**The Approach in [15]**	**The Approach in [16]**	**The Approach in [17]**	**Our Approach**
Information Channel	non-visual	non-visual	non-visual and visual	non-visual and visual	visual
Aesthetic Cognitive Model Based	No	No	No	No	Yes
Feature Type Involved	kinematic	kinematic	kinematic; shape (region and contour)	kinematic; shape (region and contour); spatial distribution of color block	color; shape (region); orientation
Specific Feature	joint	joint	joint; eccentricity, density, rectangularity, aspect ratio, Hu-moment Invariants; complex coordinate-based Fourier descriptors	joint; eccentricity, density, rectangularity, aspect ratio; centroid distance-based Fourier descriptors; basic spatial distribution, topology relationship, direction relationship, distance relationship	color moments; Hu-moment Invariants; 8-bit orientation encoder, spatial distribution histogram of color block
Feature Fusion	No	No	Yes	Yes	Yes
Aesthetic Manner	human subjective aesthetics	machine learning based method	machine learning based method	machine learning based method	machine learning based method
Machine Learning Method Involved	N/A	SVM, RBF Network, AD Tree	Naive Bayes, Bayesian Logistic Regression, SVM, RBF Network, AD Tree, Random Forest, Voted Perceptron, KStar, DTNB, Bagging	Naive Bayes, Bayesian Logistic Regression, SVM, RBF Network, AD Tree, Random Forest, Voted Perceptron, Bagging	Naive Bayes, Bayesian Logistic Regression, SVM, RBF Network, AD Tree, Random Forest, Voted Perceptron, Bagging, Rotation Forest, LWL
Highest Correct Ratio on Visual Feature Combination	N/A	N/A	76%	75.8%	81.6%
Highest Correct Ratio	N/A	71.6667%	81.6%	81.8%	81.6%
Best Feature Combination	kinematic	kinematic	kinematic + shape (region & contour)	kinematic + shape (region) + spatial distribution of color block	color + shape (region) + orientation
Best Machine Learning Method	N/A	AD Tree	AD Tree	AD Tree	Random Forest

Differing from the existing approaches, our approach is built based on the proposed computable cognitive model of visual aesthetics. Thus, our approach enables a dancing robot to behave like a human dancer (the autonomous aesthetics evaluation of self-dance poses based on mirror observations) by imitating human visual aesthetic cognition. Moreover, two feature types (color and orientation) are introduced in our approach, and the corresponding specific features (color moments, 8-bit orientation encoder, and spatial distribution histogram of color block) are designed and applied.

Furthermore, from the viewpoint of the highest correct ratio of machine aesthetics evaluation, the approach in [17] gets 81.8%, the approach in [16] gets 81.6%, and our approach gets 81.6%.

The other approaches [16,17] fuse visual and non-visual information, and our approach only uses visual information. The increase of information can bring more possibilities to improve the performance of the machine aesthetic model. Nevertheless, the addition of non-visual information does not bring a significant performance improvement.

Additionally, viewing the highest correct ratio of the visual feature combination, the approach in [17] gets 75.8%, the approach in [16] gets 76%, and our approach gets 81.6%. Therefore, our approach is superior to the others in visual information processing. We believe that this approach benefits not only from a good feature combination and a suitable machine-learning method, but also the proposed computable cognitive model of visual aesthetics.

6.4. Applicability of the Proposed Computation Framework

In our computation framework for the automatic aesthetics evaluation of robotic dance poses (see Section 4), there are two important prerequisites: a dancing robot should have unique color blocks on the important parts of its body (such as its head, shoulder, hand, foot, and leg), and the color blocks should differ from the embodied environment [16]. Based on these prerequisites, the following processing can be carried out: automatic target location and the extraction of the orientation feature.

In this paper, an NAO robot is used as a prototype, which has unique color blocks on the important parts of its body. Similarly, other kinds of biped humanoid robot can also be used as alternative research carriers in the area of robotic dance, such as HRP-2, and the Robonova robot. If a biped humanoid robot does not have unique color blocks on the important parts of its body (such as Hubo, ASIMO, QRIO, and DARwIn-OP), colored stickers can be pasted on its important body parts to make it meet the first prerequisite. Furthermore, the second prerequisite requires that the color blocks on a dancing robot differ from its embodied environment. If this can't be met, different color stickers should be pasted on the robot's important body parts.

7. Conclusions

During the long evolution process of human beings, the human brain played an important role in human's daily behaviors. Discovering similar mechanisms, by machine learning or by mimicking the human brain, may prove crucial for future artificial systems with human-like cognitive abilities [31]. Therefore, a brain-like intelligent system, which is symmetrical to the cognitive nervous system of the human brain, should be integrated into the humanoid behaviors of robots, aiming to improve their intelligence [32,33]. Based on two important aesthetic cognitive neural models of the human brain [3–5], this paper proposes a new computable cognitive model of visual aesthetics, which can be used for the aesthetic calculation of visual arts.

Another contribution of this paper is the application of the above model in the aesthetics problem of robotic dance poses to achieve automatic aesthetics evaluation of robotic dance poses, which develops the cognitive and autonomous abilities of robots. Simulation experiments have demonstrated that three kinds of feature types (color, shape, and orientation) are effective in portraying the nature of robotic dance poses. The best feature combination (*color* + *shape* + *orientation*) is identified. Based on the best feature combination, a suitable machine learning method (*Random Forest*) has a good effect on the machine aesthetics model (*81.6%*).

Our future work will focus on the following four aspects: (1) applying the proposed approach of automatic aesthetics evaluation of robotic dance poses in a real robot environment; (2) eliminating feature conflicts in the feature fusion that characterizes a robotic dance pose; (3) using some advanced neural networks [34] to train machine aesthetics models; (4) revealing the aesthetic relationship between robots and human beings from the viewpoint of aesthetic nature.

Author Contributions: H.P. designed the methodology and the experiments; J.L. conceived the research idea; H.H. and K.H. supervised the methodology and the experiments; C.T. provided comparative analysis; Y.D. performed the experiments; H.P. and J.L. wrote the paper together. All authors have read and agreed to the published version of the manuscript.

Funding: This work was financially supported by National Natural Science Foundation of China (Grant No. 61662025, 61871289), and the Research Foundation of Philosophy and Social Science of Hunan Province (Grant No. 16YBX042), and the Natural Science Foundation of Zhejiang Province (Grant No. LY20F030006, LY20F020011), and the Research Foundation of Education Bureau of Hunan Province (Grant No. 16C1311).

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

References

1. 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]
2. Zeki, S.; Nash, J. *Inner Vision: An Exploration of Art and the Brain*; Oxford University Press: Oxford, UK, 1999.

3. Chatterjee, A. Prospects for a cognitive neuroscience of visual aesthetics. *Bull. Psychol. Arts* **2004**, *4*, 55–60.
4. Chatterjee, A.; Vartanian, O. Neuroscience of aesthetics. *Ann. N.Y. Acad. Sci.* **2016**, *1369*, 172–194. [CrossRef] [PubMed]
5. Höfel, L.; Jacobsen, T. Electrophysiological indices of processing symmetry and aesthetics: A result of judgment categorization or judgment report? *J. Psychophysiol.* **2007**, *21*, 9–21. [CrossRef]
6. Leder, H.; Belke, B.; Oeberst, A.; Augustin, D. A model of aesthetic appreciation and aesthetic judgments. *Br. J. Psychol.* **2004**, *95*, 489–508. [CrossRef] [PubMed]
7. Redies, C. Combining universal beauty and cultural context in a unifying model of visual aesthetic experience. *Front. Hum. Neurosci.* **2015**, *9*, 218. [CrossRef]
8. Koelsch, S.; Jacobs, A.M.; Menninghaus, W. The quartet theory of human emotions: An integrative and neurofunctional model. *Phys. Life Rev.* **2015**, *13*, 1–27. [CrossRef]
9. Schaal, S. Is imitation learning the route to humanoid robots? *Trends Cogn. Sci.* **1999**, *3*, 233–242. [CrossRef]
10. 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]
11. Breazeal, C.; Scassellati, B. Robots that imitate humans. *Trends Cogn. Sci.* **2002**, *6*, 481–487. [CrossRef]
12. 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.
13. Vircikova, M.; Sincak, P. Artificial Intelligence in Humanoid Systems. Academia.edu. 2010. Available online: https://www.academia.edu/756300/Artificial_Intelligence_in_Humanoid_Systems (accessed on 19 December 2019).
14. 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.
15. 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]
16. 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.* **2019**. [CrossRef]
17. Peng, H.; Li, J.; Hu, H.; Zhao, L.; Feng, S.; Hu, K. Feature fusion based automatic aesthetics evaluation of robotic dance poses. *Robot. Auton. Syst.* **2019**, *111*, 99–109. [CrossRef]
18. 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.
19. Krasnow, D.; Chatfield, S.J. Development of the 'performance competence evaluation measure' assessing qualitative aspects of dance performance. *J. Dance Med. Sci.* **2009**, *13*, 101–107. [PubMed]
20. 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]
21. Manfrè, A.; Infantino, I.; Vella, F.; Gaglio, S. An automatic system for humanoid dance creation. *Biol. Inspired Cogn. Archit.* **2016**, *15*, 1–9. [CrossRef]
22. Augello, A.; Infantino, I.; Manfrè, A.; Pilato, G.; Vella, F.; Chella, A. Creation and cognition for humanoid live dancing. *Robot Auton. Syst.* **2016**, *86*, 128–137. [CrossRef]
23. 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.
24. Qin, R.; Zhou, C.; Zhu, H.; Shi, M.; Chao, F.; Li, N. A music-driven dance system of humanoid robots. *Int. J. Humanoid. Robot.* **2018**, *15*, 1850023. [CrossRef]
25. 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.
26. Rother, C.; Kolmogorov, V.; Blake, A. 'GrabCut': Interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.* **2004**, *23*, 309–314. [CrossRef]
27. Stricker, M.A.; Orengo, M. Similarity of color images. In Proceedings of the SPIE Conference on Storage and Retrieval for Image and Video Databases III, San Jose, CA, USA, 23 March 1995.

28. Hu, M.K. Visual pattern recognition by moment invariants. *IRE Trans. Inf. Theory* **1962**, *8*, 179–187.
29. 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]
30. Stein, B.E.; Stanford, T.R.; Wallage, 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.
31. Ullman, S. Using neuroscience to develop artificial intelligence. *Science* **2019**, *363*, 692–693. [CrossRef] [PubMed]
32. Hoffmann, M.; Marques, H.; Arieta, A.; Sumioka, H.; Lungarella, M.; Pfeifer, R. Body schema in robotics: A review. *IEEE Trans. Auton. Ment. Dev.* **2010**, *2*, 304–324. [CrossRef]
33. Asada, M.; Hosoda, K.; Kuniyoshi, Y.; Ishiguro, H.; Inui, T.; Yoshikawa, Y.; Ogino, M.; Yoshida, C. Cognitive developmental robotics: A survey. *IEEE Trans. Auton. Ment. Dev.* **2009**, *1*, 12–34. [CrossRef]
34. Xu, B.; Liu, Q.; Huang, T. A discrete-time projection neural network for sparse signal reconstruction with application to face recognition. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *30*, 151–162. [CrossRef]