

# 基于计算机图像风格迁移的音乐可视化智能设计研究

金思雨, 覃京燕

( 北京科技大学, 北京 100083 )

**摘要:** **目的** 基于 Processing 程序识别音乐特征和图像特征, 控制图像风格迁移的实时生成效果, 尝试将人对音乐和图像的联觉进行可视化表达。 **方法** 预处理阶段分别分析输入的音乐文件和图像文件, 提取音乐文件中左右声道的实时强度等特征, 识别图像文件中色彩空间及单位像素的亮度。数据交互阶段, 建立随机生成效果, 并以音乐特征和图像特征为因变量控制粒子生成的位置、时机等。画面生成阶段, 通过蠕虫效果对每个有效粒子进行动态视觉展示, 根据实时更新的有效特征值指导渲染效果, 主要控制渲染过程中新粒子的位置和数量, 以及粒子的形状、颜色和速度等。 **结果** 音乐文件与图像文件同时对图像风格迁移的可视化表达产生影响, 不同搭配方式生成视觉画面和风格属性是独特的唯一的; 同时, 图像重建过程动态可视, 屏幕从空白到完成人像效果呈现的完整过程与音乐播放进程实时同步输出。 **结论** 将传统的音乐可视化和图像风格迁移相结合, 为音乐可视化增加了具体内容表达, 为图像风格迁移增加了时序表达; 这种实时信息可视化是人工智能模仿人类通感的一种艺术表达形式, 同时是一种表演艺术。这种可视化效果可以帮助用户用音乐生成独一无二的个性肖像; 也可被广泛应用于艺术家表达音乐和视觉的关联性。

**关键词:** 通感; 图像风格迁移; 音乐可视化; 智能设计; 交互艺术

中图分类号: TB472 文献标识码: A 文章编号: 1001-3563(2020)16-0193-06

DOI: 10.19554/j.cnki.1001-3563.2020.16.029

## Music Visualization Intelligent Design Based on Image Style Migration

JIN Si-yu, QIN Jing-yan

(University of Science & Technology Beijing, Beijing 100083, China)

**ABSTRACT:** The work aims to identify music features and image features based on the Processing program to control the real-time generation effect of image style migration and try to visualize the synesthesia of music and image. In the pre-processing stage, the input music file and image file were analyzed respectively, and the real-time intensity of left and right channels in the music file was extracted to identify the color space and the brightness of a unit pixel in the image file. In the data interaction stage, the effect of random generation was established, and the position and timing of particle generation were controlled by music feature and image feature as dependent variables. In the image generation stage, each effective particle was displayed dynamically by worm effect, and the rendering effect was guided by the real-time updated effective eigenvalues, mainly controlling the position and number of new particles in the rendering process, as well as the shape, color, and speed of particles. Both music files and image files influenced the visual expression of image style migration. The different collocation methods were distinct and unique in generating visual images and style attributes. At the same time, the process of image reconstruction was dynamically visible, and the complete process from the screen to the completion of the portrait effect was synchronized with the music playing process in real-time. The combination of traditional music visualization and image style migration adds specific content expression for music visualization and time sequence expression for image style migration. This real-time information visualization is an art expression form of arti-

收稿日期: 2020-03-28

作者简介: 金思雨 (1996—), 女, 吉林人, 北京科技大学硕士生, 主攻信息与交互设计。

通信作者: 覃京燕 (1976—), 女, 四川人, 博士, 北京科技大学教授, 博士生导师, 主要研究方向为人工智能与创新设计、交互设计、信息设计、大数据的信息可视化、可持续设计、数字文化遗传、数字娱乐等。

ficial intelligence imitating human synesthesia, and also a performance art. This visualization effect can help users create the unique personalities of the music, and can also be widely applied to artists' expression of the relevance of music and vision.

**KEY WORDS:** synesthesia; image style migration; music visualization; intelligent design; interactive art

音乐的旋律、节奏可以激起受众情感上的共鸣,随着数字时代的到来,人们体验和消费音乐的方式也随之丰富起来<sup>[1-2]</sup>。图像通常与情感和情绪相关联,并受到相关心理学理论的支持<sup>[3]</sup>。计算机生成技术的进步,使得音乐可视化和图像风格迁移成为计算机信息交互领域的重要组成部分,在体验经济中,提供更多个性化、定制化的服务和作品<sup>[4]</sup>。本文以研究音乐特征与图像表达之间的关联和表达方式为主要目的,将信息的视觉思维与数据结构精准匹配,进行音乐的信息可视化的智能设计研究。

## 1 音乐图像学

音乐图像学致力于研究与音乐相关的视觉资料<sup>[4]</sup>。音乐可视化是音乐图像学的表达方式之一,主要关注音乐特征提取、情感检测、图像处理等方面<sup>[5]</sup>。可视化是数据表达和传递的一种形式,类似于计算机模拟人们通过表情、肢体动作等可被看到的表达情感的方式。音乐可视化着重于听觉与视觉的交互、转化、表达。在可视化的设计过程中,存在多种映射模式和媒介,用于传达可能被思想和动作激发或表达感情的观念,因此如今的可视化作品表达通常受到很强的创作者的主观影响<sup>[6]</sup>,起到类似于其他美学创作中“情感交流”的作用<sup>[7]</sup>。有研究发现若能够在音乐演奏的同时提供视觉辅助,可以更好地帮助听众识别音乐表达的情感<sup>[8]</sup>,有助于让听众通过视觉的效果呈现加深对音乐的理解。现如今,越来越多的计算机学者和设计师们开始关注音乐与色彩、图像、语义、情感之间的关系,通过色彩明度、纯度、色相与可视化的色彩元素对应,图像的点线面构成与数据可视化的基本单元体线条风格对应,并通过信息可视化的方式进行设计表达。

### 1.1 音乐图像的通感

音乐和图像是现今网络上两种最流行的媒体,人类对音乐和图像的感知高度相关,两者都可以在适宜的情况下激发人的情感和力量<sup>[9-10]</sup>。因此研究音乐可

视化的同时,不能忽略图像风格化处理所获得的成就和对计算机视觉所带来的影响。音乐与图像之间的关系是密切相关的,有研究证明,颜色和音乐之间的关联确实存在合理程度的一致性<sup>[3]</sup>。这种一致性被认为是通感或联觉,是一种多样的现象。其中普通的刺激,如数字、音乐,可以作为联觉的诱导剂,并引发非同寻常的即时性的体验,如颜色、口味或空间表现形式。同时研究表明艺术工作者具有这种感受的比例更高,尤其是音色联觉者的个人感受更强烈<sup>[11]</sup>。一些艺术家正在尝试通过艺术表达的方式阐述他们所看到的两者之间的关联。曲绘是一种被认可的表达方式<sup>[6]</sup>,指根据音乐所表达的思想感情创作画作,达到增加音乐作品表现力的目的。如 Melissa McCracken,一个具有对音乐和图像通感的联觉艺术家,曾表达当音乐进入他的耳朵,他能将脑海中看到的色彩表达在画布上。Melissa McCracken 的联觉的画作案例见图 1, Melissa McCracken 用曲绘的信息可视化方式表达通感。

### 1.2 音乐可视化研究现状

音乐可视化将听觉与视觉相结合,增强音乐的可理解性和情感共鸣。对音乐可视化而言,音乐与图像之间的研究已有大量成果可供参考,目前的研究主要可被归纳为两个研究方向。

#### 1.2.1 固有标签匹配 1

匹配并应用各自的标签<sup>[12]</sup>。如: Subramaniam A 与 Barthet M 制作的 Mood Visualiser,是第一个提出将音频和生理学领域结合起来以生成音乐可视化的系统,该可视化可以实时响应音乐和听众的唤醒反应,提供了在声音和生理学领域获得的音乐和听觉体验的艺术视觉表示<sup>[6]</sup>。Wu X 等人通过计算机处理,自动估算图像和音乐之间的匹配程度,进一步为音乐生成背景视频<sup>[9]</sup>。Wu Xi-xuan 等人认为视觉内容必须捕捉歌词的语义含义或音乐的情感灵感,或两者兼而有之。这些观点在音乐与图像关系的研究中得到了证



图 1 Melissa McCracken 的联觉的画作案例  
Fig.1 Case study of Melissa McCracken's synesthesia

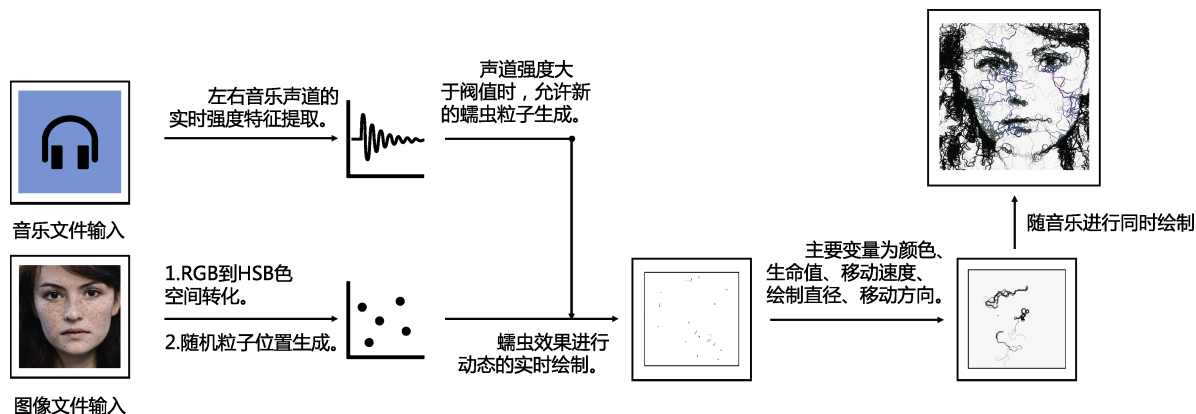


图 2 音乐特征指导的图像风格迁移运行流程

Fig.2 Flow of image style migration guided by music feature

明<sup>[13]</sup>。Voong 等人通过情绪与颜色的关联, 让用户通过对颜色情绪标签的感知, 实现音乐的分类收藏<sup>[3]</sup>。针对标签的匹配方法, 可以有效、精准地将具有相同标签的音乐文件或者图像文件匹配, 这些标签内容通常以设计者的经验理解或实验结论为依据, 用来对原始音乐文件或图像文件的整体风格进行语义描述, 帮助其他观众通过语义进行理解。

### 1.2.2 参数与固定算法匹配

通过音乐或图像中计算机可识别的参数, 结合设计者预先搭建的参数表达内容, 创作新的图像或音乐。如 Li W H 继承 Musicpainter (一种联网的图形组成环境), 并实现直接将视觉属性与音乐属性相关联, 用户可以通过在虚拟画布上绘画来创建新的音乐旋律<sup>[14]</sup>。Li W H 通过扫描识别规定范围内的儿童的图像位置, 将位置对应到 C 大调的两个八度上生成全新的音乐<sup>[12]</sup>。DiPaola S 与 Arya A 开发的 MusicFace, 通过节奏、响度和时间等结构使表达线索与情感色彩线索相结合, 根据已定义的关联将其转换为情感状态, 进而决定面部情绪和动作以创建视觉效果<sup>[8]</sup>。Reiko Miyazaki 等人通过开发一个称为 comp-i (Comprehensible MIDI Player-Interactive) 的系统, 为音乐家提供了 3D 交互式虚拟空间, comp-i 通过将速度、振幅和音高分别对应 3D 高度、直径和颜色的视觉属性进行映射, 将音乐感知和数据可视化技术相结合<sup>[15]</sup>。这种方式为音乐可视化提供了表达的自由度, 但依旧需要设计者设定固定的表达范式, 音乐文件生成画面, 或者图像文件生成声音都只允许单侧数据具有较高的自由度, 因此在一定程度上依旧无法满足音乐和图像之间的自由交互。

以上两种音乐可视化的研究都是基于人对音乐或图像进行标签标注的, 或者算法应用的单侧转化效果, 在一定程度上实现了艺术领域中高度主观的创造表达试想, 却错失了音乐创作者和图像拍摄者。通过音乐或图像传递的本质的信息提取, 不再增加创作者的主观标签, 而是让计算机完成对初始音乐和图像文

件的识别与输出的两个环节, 不仅可以最大限度保留原始文件中的信息, 还可以创造更多的可能性与未知的表达。例如 Weerasinghe P K 和 Cohen M 通过实现计算机读取音乐节奏数据, 控制虚拟木偶 Alice 的舞动, 其过程是模拟人控制提线木偶的方式<sup>[2]</sup>。

综上, 本文将音乐文件和图像文件之间的数据进行双边处理, 同时识别音乐特征和图像特征以控制图像风格迁移的实时生成效果, 让计算机拿起画笔成为一个通感艺术家, 将人对音乐和图像的联觉进行可视化模拟。

## 2 基于音乐可视化的图像风格迁移方法

本案例使用的编程语言为来自美国麻省理工学院媒体实验室 Casey Reas 与 Ben Fry 开发的 Processing, 利用 Processing 语言针对音乐特征对图像风格迁移进行设计, 音乐特征指导的图像风格迁移运行流程见图 2。

### 2.1 音乐文件处理

音乐文件中包含的特征非常复杂, 包括频率、振幅、音色、音调、音高、音长、和弦、速度、响度、节拍、旋律等, 以及随着技术发展被新提出的特征<sup>[5]</sup>, 如音乐能量、频谱矩阵、频谱流、带宽、带周期、噪音、帧率等。这些数字化的音乐特征可以被用来解析和应用用于可视化表达, 但如果对每种特征都进行提取和运算是很耗费时间的, 应根据可视化需求提取必要的特征, 忽略或去除那些不必要的特征, 从而缩短程序运行时间, 达到实时性等方面的要求<sup>[5]</sup>。主要选用代码库中的 Minim 库解析音乐。信息可视化的视觉设计中, 音乐的频率、音色、音调、音高、音长分别对应可视化中色彩的明度、色相、纯度、线条的粗细和线条的长短, 通过人工智能建立精准匹配关系, 进行可视化设计。

美学也是视觉成功的重要方面。考虑到音乐意义的复杂性, 经典的可视化概念认为, 复杂的想法应该

以“清晰、准确、高效”的方式进行交流,以实现“出色的图形效果”,因此它们不能直接应用于音乐。因此在 Processing 的开源库中选择 Minim 库用于音乐特征识别,在对不同音乐特征进行测试后,以实现视觉效果为首要判断指标,决定选用音乐文件的左右音乐声道的实时强度特征作为图像风格转移效果的自变量,指导图像风格参数中出现新画面的时机和新画面的形态。

## 2.2 图像文件处理

将图像的色彩空间更改为色调饱和度空间,对应人类对物体的感知和对重要边缘信息的重视,在计算机识别过程中为提取图像中的人像,以单位像素的亮度值为变量,规定输入图像的合格像素亮度阈值为 0~200 范围内,并允许后期在原像素点上生成新的运动像素。为保证画面的美观性,经过测试确定每个像素位置上被新像素经过的次数超过五次。

## 2.3 粒子随机处理

利用粒子引导 (Seeker) 建立新画面出现的随机性,代表图像空间中的坐标像素,并定义相应的数据结构。它具有位置、速度和惯性等参数。当每个音符到来时,速度和轨迹大小将根据音频的强度进行更新,即时的声道强度大于阈值 0.8 时,允许新的蠕虫粒子生成,也是新的画面像素生成。然后可以根据对象的惯性推断当前粒子的位置。最后,获得平滑的运动轨迹。

## 2.4 动态画面生成

画面出现的视觉效果由蠕虫效果控制,主要变量为颜色、生命值、移动速度、绘制直径、移动方向,结合粒子随机选取的像素点作为起点进行动态的实时绘制。

## 2.5 交互处理

通过鼠标点按可以实现,右键切换图片,左键切换音乐。

根据以上技术分析可知,在动态图像风格迁移和重建过程中,算法会把握音乐的节奏,并在绘制运动像素的基础上调整绘制艺术效果的速度和内容。同时,对运动点进行滤波以确保生成的艺术图像保留原始图像的重要边缘信息。它掌握了样式迁移的变化和核心信息的“不变性”。同时,它具有从零开始,从无序信息到低熵有序信息的动态过程,生成具有生命力的视听效果。

## 3 案例应用

根据以上程序,进行测试。图片和音源来源于网络,音乐文件为 MP3 格式,分别是歌手 Avril Lavigne 的单曲 Smile、歌手 OMFG 的单曲 Hello; 图像文件为两个 jpg 格式的人像。并对生成的阶段进行了单帧保存。由于画面中出现的像素点位的随机性设计,当音乐文件固定,改变图像文件时,画面中新的蠕虫粒子的分布与图像文件密切相关,见图 3—4; 当图像文件固定,改变音乐文件时,画面的动态生成效果和风格与音乐文件相关,见图 4—5。当同时改变音乐和图像文件时,画面风格和生成效果均呈现出较强的视觉效果,两首歌曲的声音强度差异,在音乐可视化后造成了图 4 和图 5 色彩和线条密集程度的差异。

由实验可知,视觉接收的来自新画面的频率、形态、颜色等信号和对音乐特征的感知有较明显的相关性。当音乐节奏快时,新像素的产生也就更多。声音强度大时,添加到图片的笔触会更夸张。这个程序能较完整地呈现图像风格迁移效果随音乐特征变化的过程,呈现过程生动有趣,呈现的最终效果相比于原始图像文件也更加抽象,最终提取人物肖像轮廓,



图 3 音乐文件: Smile.mp3 图像文件: 人像 1  
Fig.3 Music files: Smile.mp3 Image file: Portrait 1

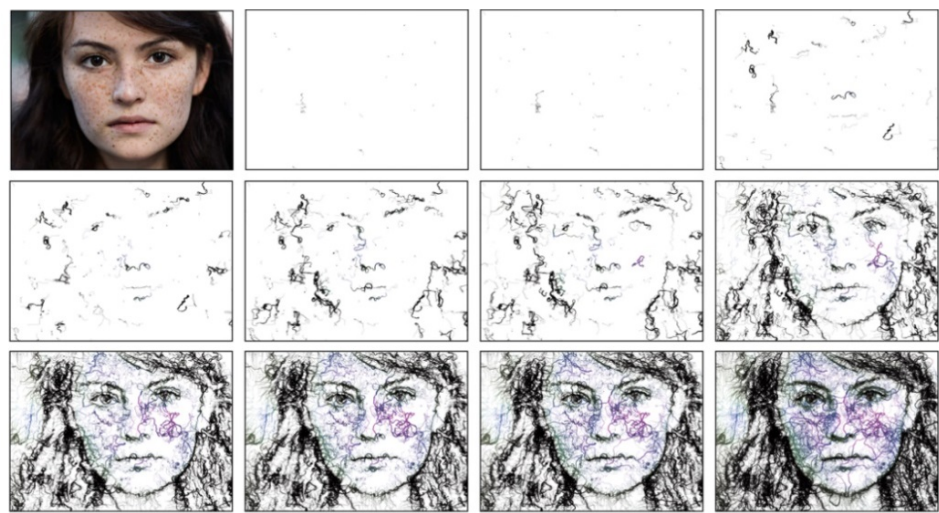


图 4 音乐文件：Smile.mp3 图像文件：人像 2  
Fig.4 Music files: Smile.mp3 Image file: Portrait 2



图 5 音乐文件：Hello.mp3 图像文件：人像 2  
Fig.5 Music files: Hello.mp3 Image file: Portrait 2

形成人工智能和信息可视化的信息美学特征和艺术风格。在音乐的渲染下，计算机好像一个自由洒脱的画家，向观众描述着它听到音乐后的感受，并对这时间的推移，逐渐积累出具有背景音乐情绪的可重新绘制的背景图像，从而形成音乐听觉和视觉的通感联觉，并通过智能设计进行可视化表达。

总体而言，这种艺术表达效果体现出高度的自由，这种全新的艺术表达效果不同于传统的艺术流派，不仅具有印象派画作中点彩的细腻表达，而且具有抽象派画作的视觉冲击，还兼具未来主义的梦幻色彩，是一种通过人工智能创造的全新的艺术表达形式。

由此可知，信息可视化设计在方法上的改进，更加关注信息在视觉表现单元体和群体大数据集的视觉表征的关联。单元个体的选择不仅需要设计师的审美判断，而且需要通过用户体验的构想性、美感与愉悦性，几方面结合进行设计评判。信息可视化设计在整体轮廓上遵循设计的本体，而在设计细节方面，则

采用不同的视觉元素，通过不同的排列组合、穿插融合、解构重构的方式，进行设计表达。同时信息可视化所需的交互设计，包含数据自交互的人工智能驱动的变化，还包含设计师和用户在观看信息可视化影像之后的人机交互界面的变化，通过可视化基本单元体的造型选择、密集排列的数量控制、排列组合的方式选择，进行大规模客户的定制设计，达成非确定性的交互变化的多样化设计结果。

#### 4 结语

音乐可以在心理和生理层面对听众产生重大影响<sup>[6,16]</sup>，同时视觉图像已经成为人们沟通交流的一种手段<sup>[17]</sup>。在传统的视觉传达设计中，通过形色之间的交替、重复和所隔空间距离来表达节奏的快慢，只是瞬间的或者片段的音乐可视化表达，无法以时间进程衡量音乐变化<sup>[1]</sup>。本文基于计算机读取和处理音

乐的频率、振幅、音色,图像的像素和色彩填充等,在实时播放中实现一种音乐特征数据指导下的图像风格转移方法,采用已有的音乐文件和图像文件,处理并尝试重新搭建两者之间在听觉与视觉方面的通感关系,生成动态、音乐声音可视、具有实时变化的可视化作品。以人像为实验案例,但同时允许用户自定义输入音乐和图像文件,这种新的可视化方式可为用户提供大规模定制(Mass Customization)的个性化服务,并根据个人喜好生成个性化肖像。同时,在音乐的播放过程中显示实时的动态画面,增强用户的体验,象征着人工智能与人类情感之间的共鸣。

在许多西方表演艺术传统中,观众被赋予了不具有创造性角色的接收者位置<sup>[6]</sup>。然而在当今万物互联的时代,对艺术家来说,观众的感受、体验是关键因素,因此观众参与艺术的方式逐渐发生着根本性的改变,不论是创作者还是欣赏者都不断给每一个艺术作品赋予着新的生命。计算机技术支持音乐、图像与人的多元交互,有效推动了 HCI(体验设计)的发展<sup>[18]</sup>,让人们都有机会感受声色联觉者的感受,并表现出更高的艺术活动参与度<sup>[11]</sup>。未来,在科学、技术和艺术的跨学科研究中,可以帮助人们创造出更有感情和感知力的作品。

#### 参考文献:

- [1] 黄洪澜. 音乐之美在视觉传达设计中的运用[J]. 包装工程, 2010, 31(24): 106-108.  
HONG Hong-lan. Application of the Beauty of Music in Visual Transmission Design[J]. Packaging Engineering, 2010, 31(24): 106-108.
- [2] WEERASINGHE P K, COHEN M. Rhythm of Music Animating Virtual Environment Models[C]//Proceedings of the 2012 Joint International Conference on Human-Centered Computer Environments, 2012.
- [3] VOONG M, RUSSELL B. Music Organization Using Color Synaesthesia[J]. CHI'07 Extended Abstracts on Human Factors in Computing Systems, 2007.
- [4] 许世虎, 宋方. 基于视觉思维的信息可视化设计[J]. 包装工程, 2011, 32(16): 11-14.  
XU Shi-hu, SONG Fang. Information Visualization Design Based on Visual Thinking[J]. Packaging Engineering, 2011, 32(16): 11-14.
- [5] 屈天喜, 黄东军, 童卡娜. 音乐可视化研究综述[J]. 计算机科学, 2007(9): 20-26.  
QU Tian-xi, HUANG Dong-jun, TONG Ka-na. A Survey on Music Visualization[J]. Computer Science, 2007(9): 20-26.
- [6] SUBRAMANIAM A, BARTHET M. Mood Visualiser: Augmented Music Visualization Gauging Audience Arousal[C]//Proceedings of the 12th International Audio Mostly Conference on Augmented and Participatory Sound and Music Experiences, 2017.
- [7] 李静, 李世国. 从交互设计的视角探索人与产品的情感交流[J]. 包装工程, 2008(9): 151-153.  
LI Jing, LI Shi-guo. Exploration the Emotional Communications between Human and Products from the Perspective of Interaction[J]. Packaging Engineering, 2008(9): 151-153.
- [8] DIPOLA S, ARYA A. Emotional Remapping of Music to Facial Animation[C]//Proceedings of the 2006 ACM SIGGRAPH Symposium on Videogames, 2006.
- [9] WU X, XU B, QIAO Y, et al. Automatic Music Video Generation: Cross Matching of Music And Image[C]//Proceedings of the 20th ACM International Conference on Multimedia, 2012.
- [10] WU X, QIAO Y, WANG X, et al. Cross Matching of Music and Image[C]//Proceedings of the 20th ACM International Conference on Multimedia, 2012.
- [11] LUNKE, KATRIN, MEIER B. Creativity and Involvement in Art in Different Types of Synaesthesia[J]. British Journal of Psychology, 2019, 110(4): 727-744.
- [12] LI W H. Musical Box: Draw It Yourself[C]//International Conference on Interaction Design & Children. ACM, 2010.
- [13] WU Xi-xuan. Cross Matching of Music and Image[C]//Proceedings of the 20th ACM International Conference on Multimedia, 2012.
- [14] LI W H. Musicpainter: A Collaborative Composing Environment[D]. Cambridge: MIT Media Lab, 2008.
- [15] MIYAZAKI R, FUJISHIRO I, HIRAGA R. Comp-i: A System for Visual Exploration and Editing of MIDI Datasets[J]. 2004.
- [16] TROCHIDIS K, SEARS D, DIÊU-LY TRÂN, et al. Psychophysiological Measures of Emotional Response to Romantic Orchestral Music and Their Musical and Acoustic Correlates[C]//International Symposium on Computer Music Modeling and Retrieval, 2012.
- [17] 付莎莎, 潘红莲. 图像时代下视觉图像的生态化设计道路[J]. 包装工程, 2014, 35(22): 98-101.  
FU Sha-sha, PAN Hong-lian. Ecological Design Development of Visual Image in Image Era[J]. Packaging Engineering, 2014, 35(22): 98-101.
- [18] EDMONDS E A, BENFORD S D, BILDA Z. Digital Arts: Did You Feel That?[J]. CHI'13 Extended Abstracts on Human Factors in Computing Systems, 2013: 2439-2446.