Dissertation on

"Automated Dance Choreography and Direction Using Music Classification and Contextual Factors"

Submitted in partial fulfilment of the requirements for the award of degree of

Bachelor of Technology
in
Computer Science & Engineering

UE20CS390B – Capstone Project Phase - 2

Submitted by:

Rayudu Srishti	**PES1UG20CS329**
Vibha Murthy	**PES1UG20CS495**
Vidisha Chandra	**PES1UG20CS498**
Vishakha Hegde	**PES1UG20CS506**

Under the guidance of

Prof. Kathargуppe Srinivas
Professor
PES University

August - December 2023

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
FACULTY OF ENGINEERING
PES UNIVERSITY
(Established under Karnataka Act No. 16 of 2013)
100ft Ring Road, Bengaluru – 560 085, Karnataka, India

System Design

Segmentation of Music based on Emotion:

Several music features such as length, danceability, accousticness, energy, instrumentalness, liveliness, valence, loudness, speechiness and tempo. A Neural Network is trained to classify the input music into 6 emotions namely, Energetic, Aggressive, Sad, Happy, Calm and Dark. The input music is broken down into 2-second segments and these segments are classified into one of the above mentioned emotions.

Seed Motion Selection based on Music Features:

Audio features are mapped to human motion features of a certain dance style that is most appropriate for a particular emotion. The Full Attention Cross Model Transformer (FACT) is the baseline model to generate the dance. A Support Vector Machine is used to generate motion features for the given audio features. This algorithm finds a hyperplane that divides the dataset into different classes while maximizing the margins between them. The most appropriate seed motion is chosen after comparing the predicted motion features with the existing motion features. The seed motion is injected in the beginning of each emotion segment and the relevant dance is generated per emotion.

Genre Detection:

K-Nearest Neighbours (KNN), a non-parametric classification technique is used to classify the genre of the input audio. Each element in the dataset is represented by its mean matrix, covariance matrix, and its respective class label. The class labels correspond to different music genres, providing a labeled training set for the KNN algorithm. Mel-Frequency Cepstral Coefficients (MFCCs) are extracted using the librosa library. These coefficients capture the spectral characteristics of the audio signal and serve as the features for the classification. The Mahalanobis distance metric calculates the dissimilarity between instances in the feature space.

Costume Generation:

A large language model (LLM) is used to create a prompt that encapsulates keywords of the outfits and accessories required for the dance, taking into consideration the emotion and genre of the song. Subsequently, this curated prompt is seamlessly integrated into a stable diffusion model specifically engineered for text-to-image generation, resulting in an appropriate costume.

Background Generation:

The Cohere LLM is used to create the prompt describing an appropriate dance setting fit for the mood and tonality of the song. The prompt is then passed into the Stability stable diffusion model which generates an image of a possible background the user could incorporate into their dance routine.

User Interface:

A user-friendly interface that displays the dynamically generated dance choreography, costume designs, and background settings tailored to the user-provided song.