

音響的類似度を用いた照明演出動画生成システム

A System for Synthesizing Stage Lighting Videos Using Acoustic Similarity

坂尾 南帆[†]土橋 宜典[‡]

Naho Sakao Yoshinori Dobashi

1. はじめに

コンサートやライブの演出において照明演出は欠かせない要素になっている。また、Vtuber 等の広がりからパッチャル上での興行も個人レベルに普及し、照明の演出が必要とされる機会は増加している。しかし、照明演出は照明について専門的な知識や経験が必要で、初心者が演出を作りあげるのは困難である。そこで、本研究では、ビデオデータベースを利用した照明演出動画の生成システムを提案する。与えられた楽曲とデータベースとの音響的類似度を計算し、類似した箇所の動画を組み合わせる。音響的類似度には MFCC を使用し、動画の選択には、最大二部マッチングを使用する。

2. 関連研究

本研究は音響信号からビデオ映像を合成する研究に分類でき、様々な手法が提案されている。Hua らは楽曲には類似したパターンが繰り返し出現することに着目し、個人のホームビデオを用いて動画を自動生成する手法を提案している[1]。Cai らはウェブ上の画像を利用して与えられた楽曲に適した動画を生成する手法を提案している[2]。この方法では、音楽信号から自動的に抽出した歌詞を用いてウェブ上から画像を検索する。

音楽情報から照明演出を生成する研究も存在する。合志らは音楽情報から自動的に照明演出を行うことを目指した研究を行っている[3]。この研究では照明光の色を決定するために、対象楽曲の印象についてアンケート調査を実施する必要がある。的場らは、波形情報から特徴量を抽出して照明演出を自動生成する手法を提案している[4]。しかし、単色の照明しか考慮できない。

3. 提案手法

提案手法の概要を図 1 に示す。まず、事前に照明演出動画からビデオデータベースを作成する。各ビデオはある楽曲に対して照明演出を施した動画である。収集したビデオは音楽的な区切りと照明演出の変化を手動で検出して、短時間のビデオクリップに分割しておく。次に、ユーザにより与えられた照明演出を付与する楽曲（ターゲット楽曲）とデータベースの楽曲との類似度を計算し、一致度の高いビデオクリップを割り当てることで照明演出動画を生成する。データベース楽曲とターゲット楽曲の類似度計算にはメル周波数ケプストラム係数 (MFCC) を用いる。ターゲット楽曲は音楽的な観点と照明演出を変化させたい区切りの観点から複数のパートに手動で分割し、それぞれのパートごとにビデオクリップを割り付ける。このとき、同一の映像が重複して選択されることのないよう、また、全体として類似度が高くなるように割り当てを行う。この割り当て問題は最大二部マッチング問題と解釈でき、最適解を算出することができる。

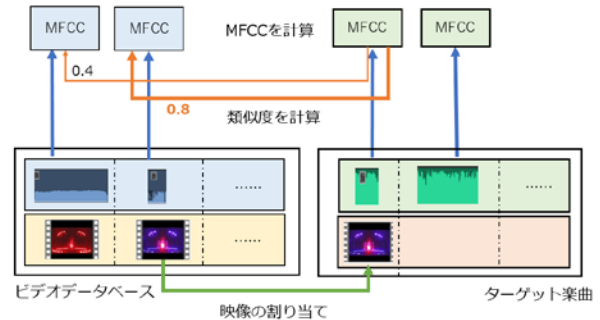


図 1 : 提案手法の概要

3.1 照明演出動画の生成

前節で述べたビデオデータベースを用いて、与えられたターゲット楽曲に調和する照明演出動画を生成する。前述の通りユーザは、いわゆる A メロ・B メロなどといった音楽の特徴や照明演出を変化させたい部分を考慮してターゲット楽曲を複数のセグメントに分割しておく。各セグメントについて、データベースから音楽的に類似した特徴を持ったビデオクリップを割り当てる。以下、ビデオクリップの割り当て問題の定式化と解法について説明する。

ターゲット楽曲の i 番目のセグメントの波形を $s_i^u(t)$ ($i = 1, \dots, n; 0 \leq t < \Delta_i^u$) とするただし n はターゲット楽曲のセグメント数 Δ_i^u はその長さを表す。また、データベースのビデオクリップ j に対する音楽波形および動画をそれぞれ $s_j^d(t)$ および $v_j^d(t)$ ($j = 1, \dots, m; 0 \leq t_j < \Delta_j^d$) で表す。ただし、 m はデータベース内のビデオクリップ数 Δ_j^d は各クリップの長さとする。

提案手法では、以下に示す最大化問題を解き、ターゲット楽曲の各セグメント i に割り当てるデータベースのビデオクリップ番号 $\sigma(i)$ を算出する。

$$\max_{\sigma(1), \dots, \sigma(n)} \sum_{i=1}^n r(s_i^u - s_{\sigma(i)}^d) \quad (1)$$

ここで、 r は二つの音楽波形の類似度を表す。この問題は二部マッチングの最小重みマッチング問題として解釈することができ、最小費用流問題に帰着して解くことが出来る。

3.1.1 MFCC を用いた類似度の計算

ターゲット楽曲のセグメント i の波形とデータベースのビデオクリップ j の音楽波形の間の類似度 $r(s_i^u, s_j^d)$ は MFCC を用いて計算する。MFCC とは人間の聴覚特性を考慮した音声特徴量の一つであり、音響的類似度の算出にしばしば用いられている尺度である。長さの異なる s_i^u と s_j^d の類似度を求めるために、提案法では動的時間短縮法 (Dynamic Time Warping; DTW) を使用する。DTW とは長さの異なる時系列データ同士の類似度を測る際に用いる手法である。対応する点を選ぶ際にその時点までに選択済みの

点も選択可能であり、重複を許すため、時系列同士の長さや周期が違ってても類似度を求めることができる。

3.1.2 ビデオクリップの割り当て

前述したとおり、ビデオクリップの割り当ては重み付き最大二部マッチングを最小費用流問題に帰着して解くことができる。ターゲット楽曲セグメントを表すノード集合とデータベースの各ビデオクリップを表すノード集合の完全二部グラフを考える。始点と終点に仮想的なノード追加する。そして、各辺にはコストとして $1 - r_{ij}$ を割り当て、辺の最大容量を 1 とする。ただし、始点、終点に接続する辺のコストは 0 とする。このグラフについて、始点から終点へターゲット楽曲のセグメント数をフローとして流す際の最小費用流を求めて始点と終点を取り除けば、最適な割り当てが求まる。

以上により、重複なく尚かつ類似度を最大にするような最適な割り当てを求めることができる。

4. 実験

提案手法によるビデオ映像の割り当ての有効性を確認するための実験を行った。以下、実験内容と実験結果を述べる。

照明の演出においては、照明の光色、色数、点滅等の色変化のパターンが重要な要素となる。一般に穏やかで楽器数が少ない場合においては照明の点滅は少なく色変化は穏やかであり、曲調が激しい場合や楽器数が多い場合は照明の点滅が激しくなる等の傾向が見られる。本実験では、ターゲット楽曲の各セグメントにおける印象と生成した照明演出動画の色変化のパターンによる印象が一致しているか主観評価実験を行った。

4.1 実験結果

ターゲット楽曲の各セグメントに割り当てた映像について、照明の光色と色数（以下纏めて光色と表記する）と照明の点滅パターンの 2 点においてターゲット楽曲セグメントの雰囲気との一致を調査した。

図 2 に生成した照明演出動画からのスナップショットを示す。LiSA の紅蓮華に対する照明演出動画を作成した例で、図 2 は提案法を用いた場合の、楽曲開始から最初の二つのセグメントから抜き出した画像である。最初のセグメントはボーカルだけの静かな曲調で、その後二つ目のセグメントで一気にビートを伴う激しい曲調へと変化する。提案法では、この変化を適切にとらえた照明演出が実現できている。

5. 考察

光色、点滅パターンという 2 つの要素において楽曲の雰囲気へ一致する動画が生成され、MFCC を用いて照明演出映像を提案する本手法は有効であると考えられる。

一方で一部楽曲について雰囲気にそぐわない映像が選ばれやすいという問題点も存在した。ターゲット楽曲とデータベース内の楽曲との雰囲気が大きく異なる場合に生じる問題であると予想される。この問題に対しては、ターゲット楽曲にジャンルの違う他の複数の楽曲を用いて更に実験を行いたいと考えている。

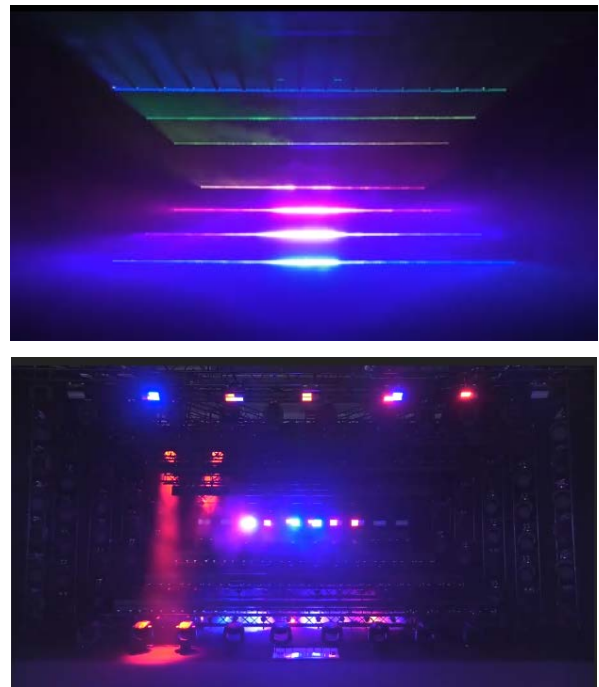


図 2 : 楽曲「紅蓮華」に対する結果

6. まとめと今後の課題

本論文では楽曲間の MFCC の類似度を元に、楽曲に既存の照明演出動画を割り当てる手法を提案した。本システムを使用することで、与えられた楽曲に対しての照明演出動画を簡単に作成することが可能である。また、MIDI ファイル等のデータを必要とせず、一般的な音楽ファイル形式である wav データを使用することが出来る。

今後の課題として、照明演出動画の作成時のテンポの自動調整、各セグメントが楽曲内で演出上どのような意味合いを持たせたいかを考慮し、映像の割り当て時に光色などにユーザの意図を反映することが挙げられる。また幅広い種類の楽曲に対応するため、データベースの拡充も行っていきたい。

参考文献

- [1] Xian-Sheng HUA, Lie LU, Hong-Jiang ZHANG, "Automatic Music Video Generation based on Temporal Pattern Analysis," Proceedings of the 12th annual ACM international conference on Multimedia, 472-475 (2004).
- [2] R. Cai, L. Zhang, F. Jing, W. Lai and W. Ma, "Automated Music Video Generation using WEB Image Resource," 2007 IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP '07, Honolulu, HI, 2007, pp. II-737-II-740, doi: 10.1109/ICASSP. 2007. 366341.
- [3] 合志和洋, 清田公保, 三好正純, 古賀広昭, "音楽の印象に合わせた照明表現システムの研究開発", 2008, 熊本電波工業高等専門学校研究紀要
- [4] 的場達矢, "音楽音響信号の音楽的な情報に基づく照明パターン自動生成手法の提案", 2012, 同志社大学理工学部インテリジェント情報工学科知的システムデザイン研究室 2011 年度卒業論文発表会