Vol.2020-CG-178 No.6
2020/6/25

ビデオデータベースを用いた照明演出動画生成

坂尾 南帆[1,a)]　土橋 宜典[1,b)]

概要：コンサートや音楽ライブの演出において照明は欠かせない要素である．また, インターネット配信を利用した仮想的なら興行も普及していることから, 個人レベルで照明の演出を必要とする機会が増加している．しかし照明演出は一般に照明について専門的な知識や経験が必要で, 初心者が初めから演出を作りあげるのは困難である．本研究ではこの問題に着目し, 効果的な演出照明の設計を支援するための演出照明動画を生成する手法を提案する．提案法では, 演出照明を撮影した短いビデオクリップからなるデータベースを用意しておく．そして, データベースから最適なビデオクリップを複数選択して結合することで, 入力として与えられる音楽波形に調和した演出照明動画を生成する．ビデオクリップの選択においては, 入力波形とデータベースの各ビデオクリップに対応する音楽波形の音響的類似度を考慮する．音響的類似度はメル周波数ケプストラム係数（Mol-Frequency Cepstrum Coefficients; MFCC ）を用いて計算する．ビデオクリップの選択問題を二部グラフを用いて定式化し, 最小費用流問題を解くことで最適なビデオクリップの選択を行う．本研究の有効性を確認するため, いくつかの日本の楽曲に適用した．

キーワード：照明演出動画，音響的類似度，ビデオ合成，メル周波数ケプストラム係数

Abstract: Lighting is an indispensable factor for music concerts and live shows. Moreover, virtual live shows using the internet have become popular and this allows us to perform a personal live show, which increases the demand for creating the impressive lighting effects. However, the stage lighting often requires advanced knowledge and skills on the lighting and it is often difficult for us to create nice lighting effects. We address this problem and propose a method for synthesizing stage lighting videos semi-automatically to help us design impressive lighting effects. Our method takes a music sound wave as input and synthesize a video that matches the input music wave. The video is created by concatenating multiple videos selected from our database containing existing short video clips of stage lighting. The method selects the optimal set of the video clips taking into account the acoustic similarities between the input sound wave and the sound waves corresponding to the video clips in the database. The similarity is computed by Mel-Frequency Cepstrum Coefficients (MFCC). We formulate the video selection problem using a bipartite graph and the optimal set is obtained by solving a minimum cost flow problem of the graph. We verify the effectiveness of our method by applying the method to several songs of the Japanese popular music.

Keywords: Stage lighting videos, acoustic similarity, video synthesis, Mel-Frequency Cepstrum Coefficients

1. まえがき

音楽ライブやコンサートは多くの人に親しまれているエンターテイメントの一つで, その市場規模は非常に大きい．また, 今日ではＣＧ技術を活用し, 個人が 3D モデルを用いたバーチャルライブ等を配信する機会も生まれている．このような音楽ライブにおいて照明の演出は雰囲気を作り上げるために欠かせない要素であるが, 誰もが簡単に効果的な照明演出を作り出すことは容易なことではない．そのため, 多くの照明演出はそれを職業としている専門家が行っている．知識や経験に基づく感覚が必要であるため, 照明についての知識のない初心者が一から演出を考える事は困難である．我々は, この問題に着目し, 照明演出について深い知識と経験を持たないユーザでも効果的な照明演出動画を生成する手法について研究している．本稿では, 対象とする音楽に既存の照明演出動画を組み合わせることで, 照明に関する知識のないユーザーに対して照明演出を補助する手法について提案する．

楽曲に合わせた照明演出を提案する先行研究として, 合志らの研究 [1] や的場らの研究 [2] が存在する．しかしこれら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．また, 通常, 照明の演出では複数の色を組み合わせて使用するが,

[1] 北海道大学
Hokkaido Uniersity
[a)] sakao@ime.ist.hokudai.ac.jp
[b)] doba@ime.ist.hokudai.ac.jp

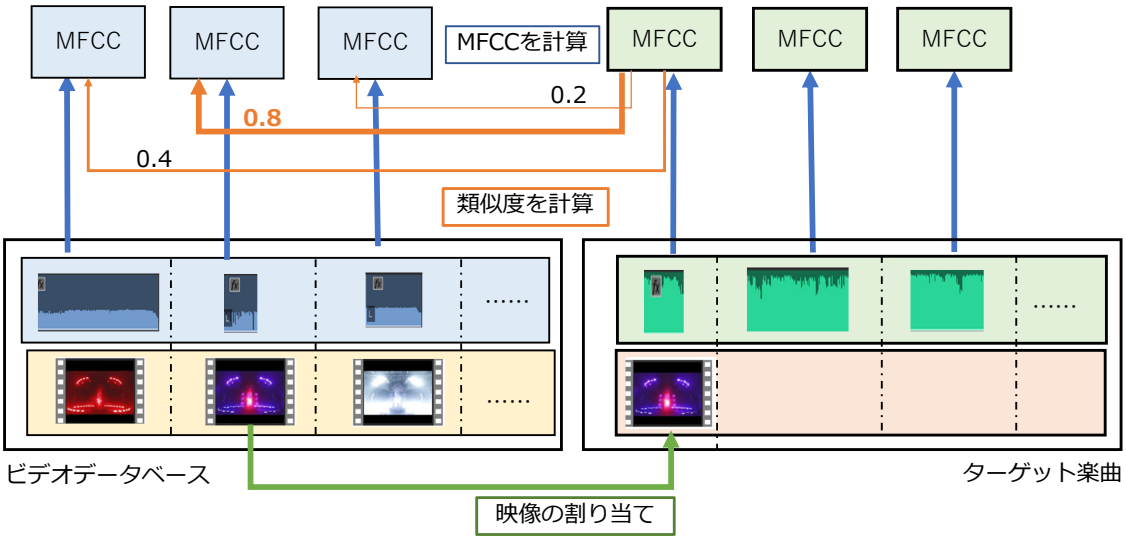


図 1: 提案手法の概要

単色での調査となっている．本研究では既存の照明演出動画を使用する事により，複雑な演出パターンを生成することが可能である．

提案法の基本的な考え方は以下のとおりである．まず，事前に，さまざまな楽曲に対して照明演出を施したビデオデータベースを用意しておく．そして，照明映像を付与したい楽曲が与えられると，データベースに保存されている楽曲との音響的な類似度を算出する．そして，類似した楽曲に対応するビデオ映像をユーザに提示する．ユーザは提示されたビデオ映像を組み合わせて最終的な照明演出動画を生成する．このとき，ビデオ映像の選択を最小費用流問題として定式化し，これを解くことで最適なビデオ映像の組み合わせを提案する．

2. 関連研究

本研究は音楽信号からビデオ映像を自動生成する研究に分類できる．この研究については，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 [3]．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像を生成する手法を提案している [4]．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．Ohya らは既存の音楽動画から音響特徴と動画特徴を抽出し，クラスタリングを行うことで，与えられた楽曲に対するビデオ映像を自動生成する手法を提案している [5]．Shin らは感情情報を用いて音楽信号からビデオを自動生成する手法を提案している [6]．また，最近では，YouTube を用いて音楽信号からビデオ映像を自動生成する手法も提案されている [7]．しかし，これらの研究は，いわゆるプロモーションビデオと呼ばれる動画を生成することが目的で，照明演出動画を対象としたものではない．

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 [1]．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．また，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 [2]．しかし，単色の照明しか考慮できない．小長谷らは SMF 形式で記述された楽曲からユーザの好みを考慮した照明パターンを自動的に生成するシステムを提案している [8]．しかし，SMF 形式で与えられた楽曲にしか適用できない．

3. 提案手法の概要

提案手法の概要を図 1 に示す．まず，事前に照明演出動画を収集した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．取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映像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はメル周波数ケプストラム係数（MFCC）を用いて計算する．ターゲット楽曲は音楽的な観点と照明演出の変化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映像を割り付ける．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．

4. ビデオデータベースの準備

本研究で使用する照明演出動画のデータベースについて説明する．本研究では, German Light Products(GLP) 社が YouTube にて公開している照明演出動画（Light Show）を使用する．GLP 社は照明器具販売会社であり，照明演出の様々な動画を公開している．本稿では, 2017 年と 2018 年に公開された二つの動画を使用する．これらの動画は 10 分弱の長さで, メドレー形式で演奏されるさまざまな楽曲に対して, 照明演出を行った様子を記録した映像となっている．曲調やリズムに合わせた多様な照明演出が行われている．これらの動画を観察し, 照明と曲調が変化する箇所で手動により分割してデータベースを作成した．本研究では, 73 個の短時間のビデオクリップからなるデータベースを作成した．図 2 にデータベース内のビデオクリップから抜き出した静止画を示す．

図 2: データベース内のビデオクリップ

5. 照明演出動画の生成手法

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手法について説明する．ユーザは, いわゆる A メロ・B メロなどといった音楽的特徴や, 照明を変化させたい点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ものとする．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詳しく説明する．

5.1 ビデオクリップの割り当て問題

いま, ユーザの指定したターゲット楽曲の i 番目のセグメントの音楽波形を $s_i^u(t)(i = 1, \cdots n; 0 \leq t < \Delta_i^u)$ とする． ただし, n はセグメント数, Δ_i^u はその長さである． また, データベース内のビデオクリップ j に対応する音楽波形および動画像を, それぞれ, $s_j^d(t)$ および $v_j^d(t)(j = 1, \cdots m; 0 \leq t_j < \Delta_j^d)$ で表す．ただし, m はデータベース内のビデオクリップの数, Δ_j^d はその長さとする．提案手法では, 以下に示す最大化問題を解くことで, ターゲット楽曲の各セグメント i 割り付けるデータベース内のビデオクリップ番号 $\sigma(i)$ を算出する．

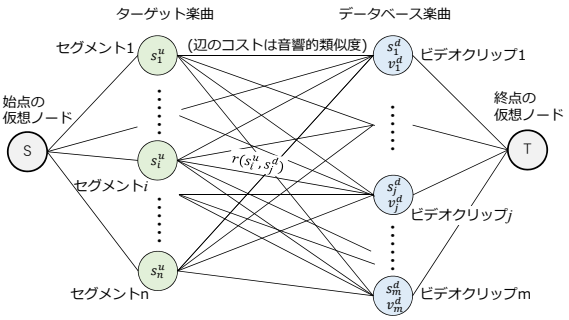


図 3: 二部グラフとビデオ割り当て問題

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

ここで, r は二つの音楽波形の類似度を表す関数である．この問題は二部マッチング問題として解釈することができる．図 3 に示すように, ターゲット楽曲の n 個の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 m 個のノード集合の完全二部グラフを考える．各辺には類似度 $r_{ij} = r(s_i^u, s_j^d)$ を重みとして割り付ける．このとき, 上述の最大化問題は, このグラフにおいて最大二部マッチングを求めることと等価である．これは最小費用流問題に帰着して解くことができる．

5.2 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 $r(s_i^u, s_j^d)$ はメル周波数ケプストラム係数（Mel-Frequency Cepstrum Coefficients;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 [9].

まず, 波形 s_i^u および s_j^d それぞれについて, 一定の微小時間間隔で分割し, 各区間ごとに MFCC を求める．MFCC は各区間の波形のサンプル数と同じ数だけ求まる．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 [10]．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．2 つの時系列の各点の誤差を総当たりで求め, 全て求めた上で 2 つの時系列のユークリッド距離が最短となる対応関係を見つける．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も類似度を求めることができる．本実験では Python のライブラリである librosa を使用し, 算出された累積距離を 0 から 1 の範囲に正規化した値を類似度として用いる．

5.3 最大二部マッチングによる映像の割り当て

5.1 節で述べた最大二部マッチング問題の解法について説明する. 前述したとおり, 最小費用流問題に帰着する. 図 3 に示すように仮想的なノード S と T を追加し, S はターゲット楽曲の各セグメントを表す全てのノードに接続し, T はデータベースの各ビデオクリップを表す全てのノードに接続する. 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, 仮想ノード S および T に接続する辺のコストは 0 とする. このグラフについて, ノード S からノード T への最小費用流を求め, 仮想ノード S と T を取り除けば, 最適な割り当てが求まる [11].

以上により, 最適な割り当てを求めることができる. この方法では, ターゲット楽曲の各セグメントには必ず異なるビデオ映像が割り当てられる. しかし, 例えば, サビや A メロ・B メロなど, 楽曲中に繰り返し現れるパートには同じ照明演出を割り当てるのが自然である. そこで, 本研究では, 同一の照明を割り当てたいセグメントはあらかじめ手動でクラスタリングしておくことでこの問題を解決する.

5.4 ビデオ映像の編集

以上述べた手法により, 各セグメントに適したビデオクリップが割り当てられ, その結果がユーザに提示される. ユーザは提示された割り当てに基づき, ビデオ編集ソフトで照明演出動画を生成する. 本稿では, 各セグメントについて, 以下の三種類の編集処理を施して照明演出動画を生成した. 一つ目は, テンポ調整のためのビデオ映像の時間的な伸縮処理である. 二つ目は, セグメントの長さより短いビデオクリップが割り当てられた場合に, ビデオクリップを繰り返し用いる処理である. 三つ目は, 必要に応じて, セグメント間でビデオクリップをなめらかに切り替えるためのクロスディゾルブ処理を施した.

6. 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. 提案法を用いてビデオクリップの割り当てを行った場合とランダムに割り当てを行った場合とで比較を行った. 以下, 実験内容と実験結果を述べる.

6.1 実験内容

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.

曲調およびジャンルの異なる 5 曲のポピュラーミュージックについて実験を行った. 表 1 にターゲット楽曲に使用した曲の一覧を示す.

楽曲	アーティスト
Wasted Nights	ONE OK ROCK
紅蓮華	LiSA
THE DAY	ポルノグラフィティ
白日	King Gnu
マリーゴールド	あいみょん

表 1: 実験対象楽曲

また, データベースの作成には, 表 2 に示す楽曲に対して照明演出を施している動画を用いている. 前述したように, 照明変化に応じて 73 個の短時間のビデオクリップに分割してデータベースに格納されている.

楽曲	アーティスト
Eat The Rich (Album Version)	Aerosmith
Something Just Like This	The Chainsmokers & Coldplay
Heavy Is the Head	Zac Brown Band
Move	Saint Motel
Life In Color	OneRepublic
It Ain't Me	Kygo, Selena Gomez
m.A.A.d city (Explicit Version)	Kendrick Lamar
All Time Low	Jon Bellion
Can't Stop	Red Hot Chili Peppers
GDFR (feat. Sage the Gemini & Lookas)	Flo Rida
Skyfall	Adele
King Is Born	Aloe Blacc
HUMBLE.	Kendrick Lamar
HUMBLE. (SKRILLEX REMIX)	Skrillex, Kendrick Lamar
Love in An Elevator	Aerosmith
For You (Fifty Shades Freed)-1385	Liam Payne—Rita Ora
Hot Damn!	The Shadowboxers
'Til It's Over	Anderson .Paak
Sax	Fleur East

表 2: データベース動画で使用されている楽曲

6.2 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, 提案法を用いた場合とランダムで割り当てた場合について比較する. 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 つが, 楽曲セグメントにおける雰囲気と一致している数を調査した. 表 3 に, 実験

楽曲 (分割数)	提案手法			ランダム生成		
	光色	点滅	両方	光色	点滅	両方
Wasted Nights(5)	4	4	3	3	1	1
紅蓮華 (9)	7	7	7	7	3	3
THE DAY(8)	7	5	5	5	1	1
白日 (5)	4	3	3	4	1	1
マリーゴールド (6)	3	5	2	4	2	2

表 3: 実験結果

結果を示す.

光色と点滅という 2 つの要素が共に楽曲の雰囲気に合致していると感じられるセグメント数は, 5 曲中 4 曲で提案手法の結果が上回った. 点滅パターンについては全ての楽曲において提案手法で合致した結果が多かった一方, 光色については提案手法による結果がランダムで割り当てた場合と同数, もしくはランダムで割り当てた結果が上回る場合も存在した.

図 4 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 4(a) は提案法を用いた場合, 図 4(b) はランダムな割り当てを用いた場合である. いずれも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のみの静かな曲調で, 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が, ランダムに割り当てた場合は, 静かな曲調であっても激しい点滅を伴う不自然な照明演出となってしまっている.

7. 考察

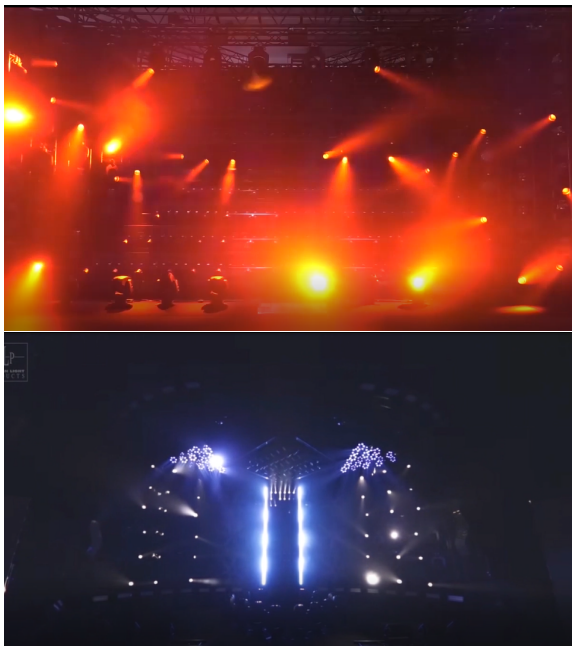
光色, 点滅パターンという 2 つの要素においてランダムに映像を割り当てた場合に比べて楽曲の雰囲気への一致する数が高かったことから, MFCC を用いて照明演出映像を提案する本手法は有効であると考えられる.

一方で光色に関して有効性があまり見られないことや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.

光色に関しては, 大きく色合いが外れていない場合は点滅のパターンの違いに比べて受ける印象の差が少ないこと, またデータベース内で似た光色の映像が多く含まれていることなどが原因として考えられる. 1 点目については, 個人の感覚による影響も存在すると予想されるため, 多人数でのユーザーテストを行うことで検証したい. 2 点目について, 別々のターゲット楽曲に対しても選ばれやすいデータベース内のビデオクリップがあったことからも, データベース内での類似度や映像の種類から, データベースの種類の偏りについて検証を行う予定である.

楽曲によっては雰囲気に合わない映像が選択されてしまうことに関しては, 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, 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.

(a) 提案法による結果



(b) ランダムに割り当てた結果

図 4: 楽曲「紅蓮華」に対する結果

8. 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の形式である 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ユーザー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, データベース内の楽曲の偏りの検証を行い, データベースの拡充も行っていきたい.

参考文献

[1] 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
[2] 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
[3] Xian-Sheng HUA, Lie LU, Hong-Jiang ZHANG, ”Automatic Music Video Generation basedon Temporal Pattern Analysis,” Proceedings of the 12th annual ACM international conference on Multimedia, 472-475 (2004).
[4] 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.
[5] H. Ohya and S. Morishima, ”Automatic Mash Up Music Video Generation System by Remixing Existing Video Content,” 2013 International Conference on Culture and Computing, Kyoto, 2013, pp. 157-158, doi: 10.1109/CultureComputing.2013.44.
[6] Ki-Ho Shin, Hye-Rin Kim and I. Lee, ”Automated music video generation using emotion synchronization,” 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Budapest, 2016, pp. 002594-002597, doi: 10.1109/SMC.2016.7844629.
[7] Sarah Gross, Xingxing Wei, Jun Zhu, ”Automatic Realistic Music Video Generation from Segments of Youtube Videos,” arXiv:1905.12245 (2019).
[8] 小長谷実希, 水上嘉樹, 松田憲, SMF 形式の楽曲に基づく照明パターン自動生成システムの開発, 画像電子学会研究会講演予稿, 08-05(0), 41-47 (2009).
[9] Peter Knees and Markus Schedl, ”Music Similarity and Retrieval,” Berlin Heidelberg:Springer-Verlag (2016).
[10] Donald J. Berndt and James Clifford, ”Using Dynamic Time Warping to FindPatterns in Time Series,” Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining, pp. 359-370 (1994).
[11] B. コルテ, J. フィーゲン著 ; 浅野孝夫, 平田富夫, 小野孝男, 浅野泰仁訳, 組合せ最適化―理論とアルゴリズム（第 2 版）－ 11 章　重み付きマッチングー,　丸善出版 (2012)