

C-22

動画をもとにした自動作曲

Automatic Composition based on Video Clips

山本 敏生† 宝珍 輝尚† 野宮 浩揮†
Toshiki Yamamoto Teruhisa Hochin Hiroki Nomiya

1. はじめに

近年、youtubeをはじめとした動画共有サービスの台頭により、世界に向けて自作動画を手軽に公開できる環境が整った。それによって、個人による動画製作の機会が増加し、同時に、動画に使用するBGMへの需要が高まってきている。現在では、著作権フリーの音楽素材をダウンロードして使用するのが主流であるが、動画に合った音楽を探し出すのは容易ではない。また、著作権法で保護されている音楽を動画に付加してweb上に違法アップロードする者も多数存在し、大きな問題となっている。

本稿では、これらの問題への解決策として、動画に合うような楽曲を自動的に生成する自動作曲システムを提案する。これにより、動画ごとにオリジナルの楽曲を生み出し、BGMとして利用することが可能となる。

以降、2.で関連研究について述べ、3.から本研究についての説明を行う。3.では、まず、動画の色情報から音楽情報を生成する際に用いる音楽心理学について概説し、4.で音楽生成時に用いるスケールやコードについての用語解説を行う。次に5.で自動作曲システムの概要を示し、6.で動画処理について述べる。7.で音楽処理について詳述した後、8.で実行手順を示し、9.で自動作曲の結果を示す。最後に10.でまとめを行う。

2. 関連研究

自動作曲の手法に関しては、過去に様々な研究がなされている。遺伝的アルゴリズム[1]、ニューラルネットワーク[2]、重回帰分析[3]、マルコフ過程[4]を用いた自動作曲がこれまでに提案されている。これらの研究では、学習によって作曲家の作曲手法を獲得し、自動作曲を行っている。これに対して本研究では、音楽心理学に基づいて自動作曲を行っている。

動画をもとにした自動作曲に関しては、動画の再生と同時にリアルタイムで作曲演奏を行うシステムが既に提案されている[5][6]。これに対し、本研究で試作したシステムは、動画全体を先に調べてから音楽を生成する。これにより、既に提案されたシステムに比べて、本システムは曲の構成をより細かく指定できるというメリットを備えている。

3. 音楽心理学

音楽心理学[7]では、以下のように言われている。

- (1) 調と色彩との対応関係については、一般に、長調の曲を聴くと明るく楽しい気分になり、短調の曲を聴くと暗く悲しい気分になる、といった気分の変化との対応から色彩的な表象を描く場合が多い。

- (2) 音の高さと色彩との対応関係については、高音になるにつれて明るい色になり、低音になるにつれて暗い色になるという関係がほぼ成立する。

これらの性質をもとにして、色情報と音楽情報の対応付けを行い、音楽情報を決定する。

4. 音楽用語

音楽情報を決定する際、以下のようなスケールやコードを使用する。

- (1) ダイアトニックスケール、ダイアトニックコード
ダイアトニックスケールとは、1オクターブのうちの7音を並べた音階で、特に全音程5つと半音程2つから構成されるものを指す。ダイアトニックスケール上にできる3声または4声のコードをダイアトニックコードという。メロディに対してどのようなコードが使えるかを考える場合、ダイアトニックコードはその基本になる。

- (2) ペンタトニックスケール
1オクターブに5つの音が含まれる音階をペンタトニックスケールという。メジャーのペンタトニックは、ルート、長2度、長3度、完全5度、長6度の5音で構成される。このスケールを用いて和音を作ると、どの音に対しても協和する響きが得られる。

5. 自動作曲システムの概要

動画をもとにした自動作曲システムの全体的な流れを図1に示す。

初めに動画ファイルを読み込み、動画処理を行って色情報を得る。この色情報をもとに、コード進行・リズム・伴奏・メロディラインといった音楽情報を生成し、MIDIファイルとして出力する。

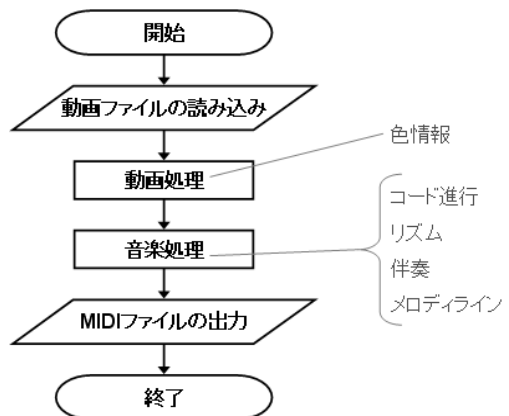


図1：自動作曲の全体的な流れ

6. 動画処理

動画を読み込んだ直後に、動画処理を開始する。読み込んだ動画から全フレームをビットマップとして切り出し、各フレームにおける全ピクセルに対して RGB 成分を抽出する。抽出した RGB 成分は 2 次元配列構造体に格納する。また、今回はより人間の知覚に基づいた HSB カラーモデルを、RGB カラーモデルと共に使用する。RGB 成分を HSB 成分に変換し、先程と同様、2 次元配列構造体に格納する。これらの色情報を用いて音楽処理を行う。

7. 音楽処理

7.1 曲の長さとのテンポ

7.1.1 動画の長さとの曲の長さの対応

動画に合うような曲を生成するには、動画と曲の再生時間を一致させる必要がある。動画の長さに応じて、MIDI データの長さを決定するために、テンポ情報から拍数を求める。テンポは 1 分間に刻む拍数を表しているため、これに動画の長さ(単位は「分」)を掛ければ、曲全体で刻まれる拍数が求まる。

例えば、動画の長さが 2 分 30 秒で、これと同じ長さの曲をテンポ $\text{♩} = 120$ で作成する場合、 $120 \times 2.5 = 300$ より、拍数は 300 となる。つまりこの曲は 4 分音符 300 個分の長さになることが分かる。

7.1.2 曲のテンポと動画のフレーム数の対応

フレームレートが n fps の動画に合わせたテンポ $\text{♩} = t$ の曲を作成する場合、1 拍当たりのフレーム数は $\frac{n \times 60}{t}$ (フレーム/拍)となる。この値を用いて、動画の変化に合わせて曲調を変化させることが可能になる。

7.2 場面転換とブロックの処理

本稿では、動画の中で画面が大きく変化するタイミングを「場面転換」と見なす。場面転換のタイミングは、フレーム間における HSB 各値の変化量の合計が閾値を上回ったときとする。また、場面転換してから、次に場面転換するまでの区間を、一つの「ブロック」とする。各ブロックに含まれるフレームの色情報をもとにして、音楽心理学を参照しながら、ブロック毎にコード進行、リズム等の音楽情報を決定する(図 2)。

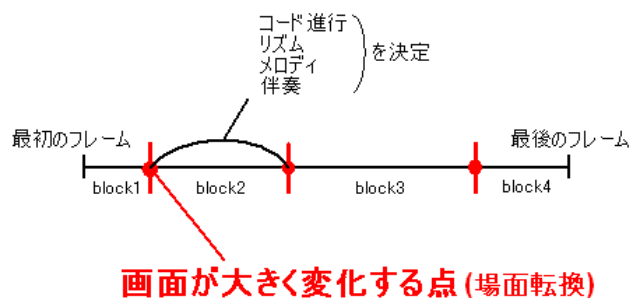


図 2: 場面転換とブロックの処理

7.3 音楽情報の決定

3.で概説した音楽心理学を利用して、色情報をもとに音楽情報を決定する。その際に、4.で解説したスケールやコードを使用する。

(1) RGB 成分からのコード進行の決定

基本的には循環コードを用いて作曲を行う。4 小節で 1 つのパターンを作り、それを繰り返して曲を構成していく。まず、RGB 成分の平均から、R 成分と B 成分の割合を求め、それをもとにメジャーコードとマイナーコードの割合を以下のように決定する。

- $7 < \frac{R}{B}$ の場合、メジャー : マイナー = 4 : 0
- $\frac{5}{3} < \frac{R}{B} \leq 7$ の場合、メジャー : マイナー = 3 : 1
- $\frac{3}{5} \leq \frac{R}{B} \leq \frac{5}{3}$ の場合、メジャー : マイナー = 2 : 2
- $\frac{1}{7} \leq \frac{R}{B} < \frac{3}{5}$ の場合、メジャー : マイナー = 1 : 3
- $\frac{R}{B} < \frac{1}{7}$ の場合、メジャー : マイナー = 0 : 4

このようにして、1 小節ごとにコードを割り当てていく。コードにはダイアトニックコードと、ペンタトニックスケールの音を用いて構成されるコードを使用する。コードの順番は乱数で決定する。

(2) HSB 変化量からのリズムの決定

画面の動きが激しいときは曲調も激しく、画面の動きが緩やかなときは曲調も緩やかにしたい。そのためにブロック内における HSB 変化量の平均を求め、曲のリズムに反映させる。変化量が多い場合はリズムを激しくし、変化量が少ない場合はリズムを緩やかにする。リズムを激しくするには 1 小節内における発音のタイミングを増やせばよい。反対に、リズムを緩やかにするには 1 小節内における発音のタイミングを減らせばよい。

今回は、ブロック内における HSB 変化量の平均と 1 小節内における発音数の対応関係を表 1 のように決定した。

(3) 色相割合からの伴奏の決定

色相を、色相差が等間隔に見える主要な 5 種類の色相「赤 (R)、黄 (Y)、緑 (G)、青 (B)、紫 (P)」と、5 種類の間景色「黄赤 (YR)、黄緑 (GY)、青緑 (BG)、青紫 (PB)、赤紫 (RP)」の 10 種類に分割し、さらに彩度が 20 以下の色を白黒 (GRAY SCALE) と設定する。これら全 11 種類の色相に対応する音色(ピアノ、ギター等)をそれぞれ決めておき、画面内の色が属する色相に対応する音色を発音する。画面内の色が、各色相に属する割合をもとに伴奏を決定する。ある色相の割合が高い場合は、その色相に対応する音色で発音する和音の構成音数を多くし、低い場合は少なくする。

今回は、色相の割合と和音の構成音数の対応関係を以下のように決定した。

- 画面全体に対して、ある色相が 5% 未満の場合、その色相に対応する音色は鳴らさない(和音の構成音数は 0)

表 1 : ブロック内における HSB 変化量の平均と 1 小節内における発音数の対応関係

HSB の変化量 d		$0 \leq d < 30$	$30 \leq d < 60$	$60 \leq d < 90$	$90 \leq d < 120$	$120 \leq d$
発音数	ベース	2	3	4	5	6
	コード	2	3	4	5	6
	キックドラム	2	3	4	4	5
	クローズドハイハット	3	4	4	4	4
	オープンハイハット	0	1	2	3	4
	スネアドラム	1	2	2	3	4

- 画面全体に対して、ある色相が 5%以上 15%未満の場合、その色相に対応する音色で発音する和音の構成音数は 1
- 画面全体に対して、ある色相が 15%以上 25%未満の場合、その色相に対応する音色で発音する和音の構成音数は 2
- 画面全体に対して、ある色相が 25%以上 35%未満の場合、その色相に対応する音色で発音する和音の構成音数は 3
- 画面全体に対して、ある色相が 35%以上の場合、その色相に対応する音色で発音する和音の構成音数は 4

図 3 に一例を示す。中段と下段の音色は、場面転換の前後で和音の構成音数が変化している。これは、場面転換の前後でそれらの音色に対応する色相の割合が変化しているからである。

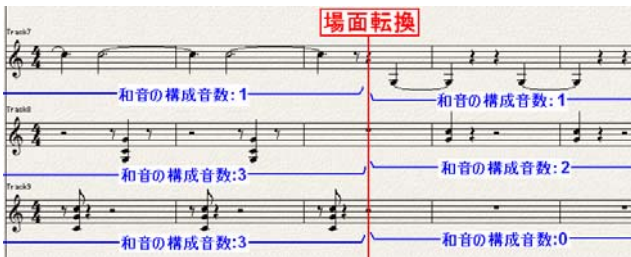


図 3 : 和音の構成音数の決定

(4) 明度からのメロディラインの決定

画面の明度が高い場合はメロディの音高を高く、明度が低い場合はメロディの音高を低くする。これを行うために、ノートナンバーを操作する。ノートナンバーとは 0~127 の数値で表される MIDI 楽器の音域のことである。拍に対応するフレーム内の明度の平均値を B とすると、ノートナンバー N は $B / 10 + 60 \rightarrow N$ で算出する。その後、メロディが単調になるのを防ぐために、 N の値に $-2 \sim +2$ の値をランダムで加える。その後、 N の値が C メジャーペンタトニックスケール上にあるなら、これを最終的なノートナンバーとし、そうでなければ再び $-2 \sim +2$ の値をランダムで加える。 N が C メジャーペンタトニックスケール上の値になるまで同様の処理を繰り返す。このようにしてメロディを決定していき、メロディラインを生成していく。

メロディライン生成の過程を図 4 に示す。

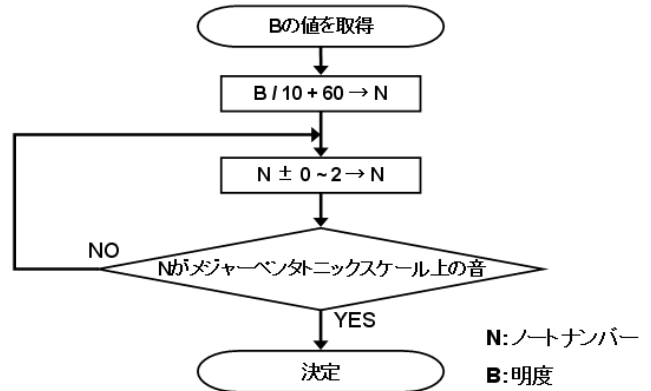


図 4 : メロディラインの生成

8. 実行

自動作曲システムの実行中の様子を図 5 に示す。メニューから「ファイル」→「ファイルを開く」を選択するとダイアログボックスが開かれる。そこから動画ファイルを選択すると、動画が読み込まれて各フレームの色情報が調べられる。全フレームについて処理が終了すると、抽出した色情報をもとにして音楽情報が自動生成され、MIDI ファイルとして出力される。なお、曲の BPM(テンポ)と各色相に対応する音色は「オプション」から変更することができる(図 6)。



図 5 : 実行例

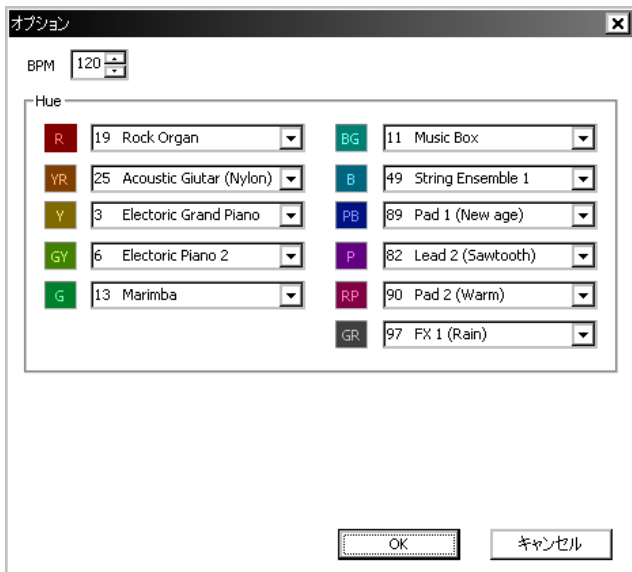


図 6 : テンポと音色の変更画面

9. 出力結果

自動作曲システムによって出力された MIDI ファイルの一部を楽譜として図 7 に示す。場面転換処理の前後で、フレーズパターンが変更されていることが確認できる。

図 7 : 出力結果例

10. おわりに

今回は、全フレームをビットマップとして抽出し、フレームの色情報をもとに作曲を行う自動作曲システムを試作した。初めに動画の全フレームを調べることで、動画を複数のブロックに分割し、曲の変化をよりはっきりと付けることが可能となった。

今後の課題としては、ユーザーの嗜好を考慮したシステムの構築が挙げられる。現時点では、音楽生成の際に、音楽心理学をもとにして動画の色彩と音楽要素の対応付けを行い、音楽情報を決定しているが、これによって生成された楽曲が動画に合致するとユーザーが感じるか、また、ユーザーの嗜好に合うかどうかは定かではない。今後は、ユーザーが動画と音楽の対応関係のある程度決められるよう、柔軟性を持たせるなどして、改善を図る必要がある。

参考文献

- [1] 沼尾正行, 高木将一, 中村啓佑 : “ユーザの感性に合わせた自動編曲及び作曲”, 情報処理学会 研究報告 2001-MUS-41 (9), pp.49~54 (2001)
- [2] 小寺慶生 : “ユーザに専門知識を要求しない自動作曲システムの研究”, 情報処理学会 研究報告 2005-MUS-59 (1), pp.1~6 (2005)
- [3] 伊藤丈一, 伊藤直樹, 西本一志 : “音楽的特徴量と作曲者の主観評価の関連性を用いたフレーズ作成支援システムの構築”, 情報処理学会 研究報告 2008-MUS-74 (26), pp.145~150 (2008)
- [4] 安容燮, 松永健 : “人間の感情価を変数としたマルコフ過程による作曲法”, 情報処理学会 研究報告 1998-MUS-28 (6), pp.27~32 (1998)
- [5] 勝田 哲司 : “帰ってきた自動作曲研究所 第 30 回 映像から音楽をつくる (1)”, DTM MAGAZINE, Vol.99, pp.72~73 (2002)
- [6] 勝田 哲司 : “帰ってきた自動作曲研究所 第 31 回 映像から音楽をつくる (2)”, DTM MAGAZINE, Vol.100, pp.82~83 (2002)
- [7] 谷口 高士 : “音は心の中で音楽になる 音楽心理学への招待”, 北大路書房, (2000)
- [8] 勝田 哲司 : “音楽研究所 研究テーマ->自動作曲”, <http://www.asahi-net.or.jp/~HB9T-KTD/music/Japan/Research/AlgoComp/index.html>