

DIVA : 画像の印象に合わせた音楽自動アレンジの一手法の提案

大山 喜冴 伊藤 貴之

お茶の水女子大学大学院

E-mail: {kisa, itot}@itolab.is.ocha.ac.jp

概要

映画やテレビCMの制作, およびマルチメディア技術において, 画像の印象に合った音楽を用いることは非常に重要である. そこで我々は, ユーザが任意の音楽と画像を入力した際に, 画像の印象に合わせて音楽を自動アレンジする手法の研究を進めている.

本論文ではその初期成果として, 画像の色分布および対象物からその印象を推測し, その印象に合ったリズムパターンで音楽を自動アレンジする手法を提案する. 本手法では前処理で, ユーザに多数のサンプル色, サンプル画像, 画像に写る対象物をキーワードとして提示し, この各々から連想されるリズムパターンを回答させる. この回答結果から各々のリズムパターンに対して, どのような画像から高い連想度を得られるかを推測する算出式を導出する. 続いて本処理では, 任意の入力画像に対して, 各々のリズムパターンの連想度を算出し, 最も連想度の高いリズムパターンを用いて自動アレンジした音楽をユーザに提示する.

DIVA: An Automatic Music Arrangement Technique Based on Impressions of Images

Kisa Ohyama Takayuki Itoh

Graduate School of Humanities and Sciences, Ochanomizu University

Abstract

It is very important to match images and music according to their impression. Our research focuses on automatic arrangement of music according to impressions of images. As the first result of our research, this paper reports a technique to automatically arrange music by inferring the impression of input images according to their color distribution and objects, and selecting a rhythm pattern according to the impression of images. As a preprocessing, the technique first provides users sample colors, sample images, and keywords of objects, and then collects the answer of users which are selection of rhythm patterns associated from the colors, images, and keywords. The technique then leads an equation to infer the association of rhythm pattern to the colors, images, and keywords. Finally, the technique automatically calculation the association values of all prepared rhythm patterns from arbitrary input images, and provides users the music arranged by the most associated rhythm pattern.

1. はじめに

映画やテレビCMの制作において, 画像と音楽は密接な関係にある. また, 個人でも「自分で制作したホームページの印象の合った音楽を載せたい」という感想は, 多くの人が一度は持つような感想だろう. しかし音楽に精通していない人は, 画像に印象

の合う音楽を選べない場合があり, 画像と音楽の相関性のない一見不釣り合いなホームページをつくる場合が多い.

一方, 音楽と映像は相補的な関係であり, 印象の合う音楽と映像を組み合わせることで, お互いの印象が向上されたり強調されたりすることが知られている[1]. このような現象の有効活用に関する研究は,

感性情報処理という分野で活発に議論されている[2].

音楽と映像を有効に組み合わせるための一手段として我々は、ユーザが任意の音楽と画像を入力した際に、画像の印象に合わせて音楽を自動アレンジする手法「DIVA: Digital Image Varies Arrangement」の研究を進めている[3,4]. DIVAは以下の2段階の処理により、任意画像の印象に合わせて音楽自動アレンジ手法を特定するものである。

[前処理] サンプルとなる視覚情報と音楽情報を被験者に提示し、その印象をユーザに回答させることで、そのユーザにおける画像と音楽の印象の関連性を数式化する。

[本処理] 任意の画像と音楽を入力した際、その画像に対してユーザがもつ印象を計算機が類推し、それに合うように音楽を自動アレンジする。

本手法では、サンプル画像や任意画像に実写画像を仮定し、その色分布から画像の印象を推定する。またあらかじめ、同じメロディに対して数種類のリズムパターンを用いてアレンジされた楽曲を用意する。この前提のもとで本手法の[本処理]では、以下のような手順によって、画像の印象に合わせて音楽アレンジ方法を特定する。

- ・ 画像に写る対象物(「海」「山」等)をキーワードとして提示し、その対象物から各々のリズムパターンの連想度を算出する。
- ・ 画像から色分布を算出し、その色分布から各々のリズムパターンの連想度を算出する。
- ・ 上記2種類の連想度の合計値が最も高いリズムパターンを用いて、音楽を自動アレンジする。

このような技術を確立するために本論文では、以下の4種類の前処理を実装・実験している。

[前処理 1] 風景画像に写る対象物(「山」「木」「海」等)を表すキーワードを被験者に提示し、このキーワードから連想されるリズムパターンを選んでもらい、これを対象物に対するリズムパターンの連想度とする。

[前処理 2] 所定のサンプル色及び、メロディは同じだがリズムパターンが違う音楽を用意する。被験者にサンプル色を提示して、サンプル色から連想されるリズムパターンを選んでもらう。この回答結果から、各々のサンプル色に対するリズムパターンの連想度を算出する。

[前処理 3] 数枚のサンプル画像および、メロディは同じだがリズムパターンが違う音楽を用意する。被

験者にサンプル画像を見てもらい、そのサンプル画像から連想されるリズムパターンを選んでもらう。さらに、サンプル画像における各サンプル色の重要度を算出する。この算出結果と回答結果から、各々のサンプル色に対するリズムパターンの連想度を算出する。

[前処理 4] [前処理 1]と同様に、風景画像に写る対象物の単語のみから連想されるリズムパターンを選んでもらい、これを対象物に対するリズムパターンの連想度とする。それに加え、[前処理 3]と同様にサンプル画像から連想されるリズムパターンを選んでもらい、サンプル画像における各サンプル色の重要度を算出する。この算出結果と回答結果から、各々のサンプル色に対するリズムパターンの連想度を算出する。

本論文では、以上の処理によって実現された自動アレンジに対して、被験者の満足度を調査することで、本研究の有効性を検証する。

2. 関連研究

本論文の関連研究として考えられるものに、ユーザの嗜好や印象に基づいて音楽や提供する技術、音楽と画像を嗜好や印象に基づいて関連づける技術、ユーザの嗜好や印象に基づいて音響情報を合成する技術、あるいは音楽を自動作曲する技術、などがあげられる。

検索技術に基づいてユーザの求める音楽を提供する手法は、旧来から多く発表されている。音楽に関する感性を注釈語や特徴量として表現することの有益性は旧来から議論されており[5]、この注釈語や特徴量を検索キーなどに用いたシステムの研究も活発に議論されている[6-11]。近年ではこれに加えて、入力画像の印象に合うように音楽を検索する技術や、入力画像群と入力音楽群を感性に基づいてマッチングする技術も、活発に議論されている[12-16]。しかしこれらの技術でシステムが提供する楽曲は、あらかじめデータベースなどに登録されている楽曲に限定される。裏を返せば、ユーザが求めるような楽曲がシステムに登録されていないければ、ユーザを満足させる楽曲を提供することは難しい。

これとは別に、画像などの視覚情報に対応して音響情報を合成する技術[17,18]や、画像の内容に合わせて音楽を自動作曲する手法[19,20]などの研究が進んでおり、さらには画像の内容に合わせた自動作曲ソフトウェアも公開されている[21]。

以上の関連研究と違って我々は、ユーザの好みのメロディをあらかじめ指定し、これを画像の印象に基づいて自動アレンジする技術の研究を進めている。多くの音楽愛好者は、音楽の構成要素の中でもメロディに思い入れを持つ。それに対して次章以降でも論じるように、音楽の印象はメロディ以外の構成要素にも左右されるものである。よってメロディをユーザの指定したものに固定しつつ、メロディ以外の構成要素を差し替えることで、同一のメロディから自動アレンジによってさまざまな印象を提供する、という手法には一定の理があると考えられる。しかし我々が調べた限り、画像の印象に合わせた音楽自動アレンジに関する関連研究は見つかっていない。

なお我々のグループでは、音楽と画像を関連づける別の研究事例として、音楽の印象に合わせてアイコン画像を自動選出する研究[16]がある。この研究は音楽から画像を選出するものであり、入力画像に基づいて音楽を自動アレンジする本論文の提案手法とは処理手順を逆にしている。

3. 提案手法の概要

3.1 基本構想

本論文で提案する「画像の印象に合わせて音楽を自動アレンジする手法」は、1章で論じた[前処理][本処理]の2段階処理によって、音楽を自動アレンジする手法であると定義づけられる。

一方、音楽や画像には以下のような多様な構成要素がある。音楽や画像の印象は、これらの構成要素と大きな関連性がある。

[音楽の構成要素] 「調(長調/短調)」「テンポ」「旋律(上昇/下降)」「音高(高/低)」「和声(単純/複雑)」「リズム(固定/流動)」「歌詞の内容」など。

[画像の構成要素] 「色彩」「構図」「対象物」「場面設定」など。

音楽と画像の印象を計算機が類推するためには、これらの構成要素すべてを計算機が分析することが望ましいが、現実的にはそれは難しい。そこで現時点での実装では、上記の多様な構成要素の中から、人間がもつ印象を特に大きく左右する構成要素だけを用いている。

3.2 現時点の実装の概要

現時点での実装では、3.1節で紹介した音楽の構成要素のうち、「リズム」を用いている。また3.1節で紹介した画像の構成要素のうち、「画像に描かれている対象物」と「色彩」を用いている。以下に、これ

らの構成要素が最も重要であると判断した理由について論じる。

文献[1]には、音楽の印象と色彩の印象に相関性があることが論じられている。また文献[22]には、それを実証する実験結果が示されている。例えば色相と音楽の印象の相関性には、「赤には迫力のある音楽が似合う」「緑には明るい音楽が似合う」という回答が多いとの結果が出ている。また、明度や彩度と音楽の印象の相関性には、「明度が高くなるに連れて音楽を明るく感じる傾向にある」「彩度の高低と音楽の力強さの度合いが対応する傾向にある」という回答が多いとの結果が出ている。また文献[23]においても、画像の持つ印象は主に配色が重要である、という実験結果が報告されている。

このことから画像の持つ「色」が音楽に与える影響は大きいと考えられる。したがって現時点での実装では、画像の特徴のうち、主に色彩とその配色からその印象を類推することにした。しかし、それでも色単体のみから画像の印象を特定するのは不確実であると考えられる。例えば、赤い花と赤い夕日では、同じ色分布を有する画像であっても印象が大きく変わることが多い。そこで本研究では、画像に描かれている対象物を示すキーワードを与え、あらかじめ提示する手法を併用した。

次に、楽曲の構成要素について論じる。文献[1]では、音楽のさまざまな構成要素が、その印象に複雑に関与していることを示している。さらに文献[3]では、この構成要素の中でも「リズム」→「旋律」→「和声」→「音高」の順で印象に残りやすく、楽曲を特徴付ける大きな要素となっていると述べている。このことから現時点での実装では、音楽の構成要素のうちリズムパターンを差し替えることで自動アレンジを行う。他の構成要素は、現時点ではアレンジには用いていない。

以上の前提によって構築された、現時点での実装の概要を、図1に示す。現時点での実装では、二種類の前処理のいずれかを用いて、画像を構成する色と対象物からのリズムパターン連想度を算出する。続いて本論文では、入力された任意の画像の色分布を算出し、次に画像に描かれている対象物(「海」「山」など)から連想されるリズムパターンを算出する。そして、色分布から連想されたリズムパターンと、画像に描かれている対象物から連想されるリズムパターンの二つを用いて、この画像からの連想度が高いと思われるリズムパターンを自動選択する。そしてこのリズムパターンを用いて、音楽を自動アレンジする。

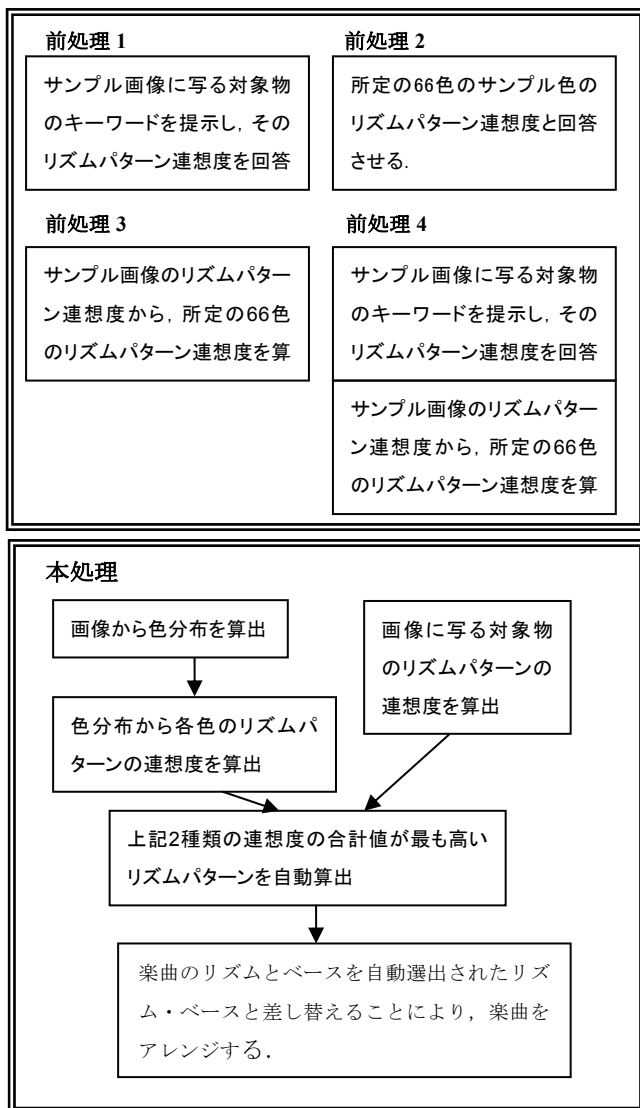


図 1：本手法の概要

4. 提案手法の実装

4.1 サンプル色の選定

我々の現時点での実装では、66色のサンプル色を規定し、この66色の各々に対するリズムパターンの連想度を算出する。我々の実装における66色の内訳は以下の通りである。

- ・ まず有彩色をHSV色空間に配置する。
- ・ 続いて、その色相を7段階に分割する。また、彩度および明度を「低」「中」「高」の3段階に分割する。また無彩色の明度も同様に3段階に分割する。この結果として、HSV空間を66個の部分空間に分割する。
- ・ 各々の部分空間内部に1個ずつサンプル色を選ぶ。

なお色相の分割数を決定する際には、太陽光が分光して見られる虹の色彩を、日本人は7色で認識する文化がある[24]、という事実を参考にした。

66色にポストライズした画像の例を図2に示す。著者らは、この66色によるポストライズ結果と、さらに少ない色数でポストライズした結果を主観比較した。その結果、66色では画像の印象をほぼ損なわないが、さらに減色すると画像の印象を損ないやすくなる、と判断した。以上の判断結果により著者らは、ポストライズの色数を66色と決定した。

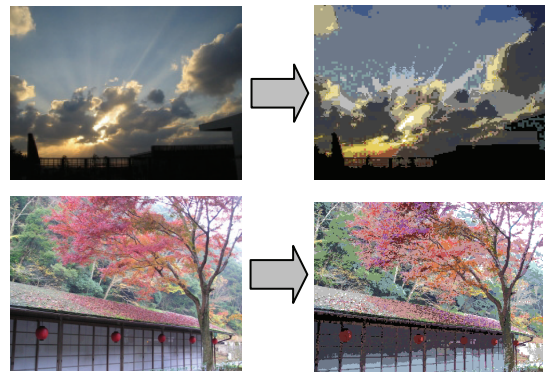


図 2：66色の妥当性の検証例

本研究では66色のサンプル色それぞれに対して、リズムパターンの連想度を算出する。以下、その理由を論じる。文献[25]では、色相が与える印象の平均的傾向が記されている。しかし一方で文献[3]では、色相が同じ色であっても、明度・彩度が異なる場合、心理的にそれぞれ違う影響も与えることも報告している。このことから本研究では、色相・彩度・明度の異なる66色のサンプル色それぞれに対して、リズムパターンの連想度を算出するべきだ、と考えた。

4.2 色彩対比を利用した色重要度の算出

本手法では、画像の印象を類推するために「色重要度」という概念を導入している。色重要度とは、その色が画像の印象を支配する度合いを推定する数値である。本手法では色重要度を、

- ・ 画像中に占める面積
- ・ 隣り合う色との差分
- ・ 色自体の持つ印象

により算出する。

現時点の実装では、文献[12]に紹介される色彩対比のうち、以下の二項目を実装することで、「隣り合う色との差分」「色自体の持つ印象」を加味して色重要度を算出する。

[明度対比] 暗い色に囲まれた色の方が明るく見える現象。この現象を色重要度に反映するために、現時

点での実装では、面積が小さいにも関わらず、周囲と比較して極端に明度が高い領域において、その色彩の重要度をあげる。本手法では、この条件に該当するとき $b > 1$ 、それ以外のとき $b = 1$ となる変数 b を導入する。現段階では、画像中の明度の高い 22 色の総面積に比べ、明度の低い 22 色の総面積が大きい場合は b の値が大きくなり、画像中の明度の高い 22 色の総面積に比べ、明度の低い 22 色の総面積が小さい場合は b の値が小さくなるような単調増加関数を b に用いる。

[進出・後退] 暖色系が進出し手前に見え、寒色系の色が後退して見える現象。この現象を色重要度に反映するために、現時点での実装では、暖色は寒色に比べ進出して見えるため重要度を上げている。本手法では、この条件に該当するとき $p > 1$ 、それ以外のとき $p = 1$ となる変数 p を導入する。現段階では、66 色全てに変数 p の値をあらかじめ決定し、要素数 66 の配列にこれを保存している。

以上の二項目を用いて本手法では、66 色に減色された画像に対して、 i 番目のサンプル色の色重要度 C_i を、以下の式で算出する。

$$C_i = \sum_{i=1}^N b_i p_i f_i \quad (1)$$

ここで N は画像の総画素数、 b_i および p_i は i 番目の画素における b および p の値、 f_i は i 番目の画素の画素値が i 番目のサンプル色であれば $f_i = 1$ 、そうでなければ $f_i = 0$ となる二値変数である。

4.3 画像に描かれている対象物の印象

本手法では画像全体の色分布だけでなく、画像に表現されている対象物を表すキーワードも、リズムパターンの選出のために用いる。

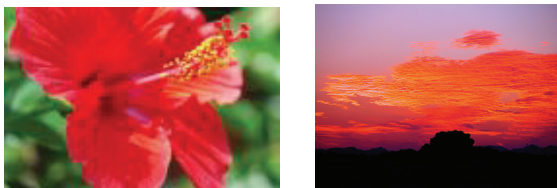


図 3：赤い画像の比較

以下、その理由を論じる。赤い対象物を含む 2 種類の画像を、図 3 に示す。この図を見てもわかるように、赤い花の画像と、赤い夕焼けの空の画像では、同じ赤でも被験者が受ける印象は大きく変わる可能性があると考えられる。そこには「花」または「夕焼け」という、個々の対象物が持つ印象が大きく関係していると考えられる。このことから本手法では、

これらの印象の違いを反映するには、画像に写る対象物からのリズムパターンの連想度を算出し、この連想度を反映してリズムパターンを選択するべきだと考えた。今回使用した対象物のキーワードは、花・夕日・もみじ・月・樹木・海・雲・夜・雪・空の 10 個を使用した。

4.4 前処理：サンプル色・対象物とリズムパターンの連想度の算出式の導出

本手法では前処理として、視覚情報を被験者に提示し、その視覚情報から連想されるリズムパターンを回答させる。この回答結果をもとに本手法では、サンプル色とリズムパターンの連想度の算出式をたてる。以下、本手法で実装している 4 種類の前処理について論じる。

前処理 1

前処理 1 では、画像に写る対象物をキーワードとして提示し、これに対するリズムパターンの連想度も、被験者に回答させる。ここで本論文では、 k 番目のキーワードに対する j 番目のリズムパターンの連想度を、 Q_{kj} と記述する。

前処理 2

前処理 2 では、所定の 66 色のサンプル色を提示し、同時に複数のリズムパターンを被験者に提示する。続いて、各サンプル色に対するリズムパターンの連想度を、被験者に回答させる。ここで本論文では、 i 番目のサンプル色に対する j 番目のリズムパターンの連想度(回答結果から得られる値)を、 R_{ij} と記述する。

前処理 3

前処理 3 では、複数のサンプル画像(我々の実験では 25 枚)を提示し、同時に複数のリズムパターンを被験者に提示する。続いて、各サンプル画像に対するリズムパターンの連想度を回答させる。続いて、この回答結果から、66 色の各サンプル色に対する各リズムパターンの連想度 R_{ij} を算出する。

ここで、1 枚の画像に対する各リズムパターンの連想度の算出式を考える。1 枚の画像に対する j 番目のリズムパターンの連想度(回答結果から得られる値)を P_j とする。また本手法では、画像を 66 色に減色したときの、画像中の i 番目のサンプル色の重要度を、式(1)から算出した結果として C_i と記述する。このとき P_j と R_{ij} には、以下の式(2)が成立する。

$$P_j = \sum_{i=1}^{66} C_i R_{ij} \quad (2)$$

ここで R_{ij} を求めるために、式(2)を以下のように変形する。合計 m 枚のサンプル画像のうち、 k 番目の画

像の*i*番目のサンプル色の重要度を C_{ki} とする。また、*k*番目の画像に対する*j*番目のリズムパターンの連想度（回答結果から得られる値）を P_{kj} とする。このとき

$$\mathbf{P} = \begin{bmatrix} P_{1j} \\ \vdots \\ P_{mj} \end{bmatrix}, \mathbf{C} = \begin{bmatrix} C_{11} & \cdots & C_{1i} \\ \vdots & & \vdots \\ C_{m1} & \cdots & C_{mi} \end{bmatrix}, \mathbf{R} = \begin{bmatrix} R_{1j} \\ \vdots \\ R_{ij} \end{bmatrix}$$

とすると、

$$\mathbf{P} = \mathbf{C}\mathbf{R} \quad (3)$$

が成立する。このとき 66 枚のサンプル画像に対する回答結果があれば、連立方程式を解くことで R_{ij} の各々の値を得ることができる。しかし 66 枚のサンプル画像に対する回答を全て用意しなければならないとしたら、前処理での被験者の負担が大きくなる。そこで本手法では、式(3)を以下のように変形する。

$$\mathbf{R} = \mathbf{C}^{-1}\mathbf{P} \quad (4)$$

これに*m*枚の画像における \mathbf{P} および \mathbf{C} の値を代入することで、サンプル画像が 66 枚未満である場合にも、*j*番目のリズムパターンに対する R_{ij} 値の近似値を得ることができる。このような近似解算手法は、本手法の目的以外にも実績のある解法である。例えば 3 次元CGの代表的なレンダリング手法であるラジオシティ法においても、漸進的かつ高速な照度算出のために、これと等価な解法が用いられている。

前処理 4

前処理 4 では、前処理 1 と同様に、画像に写る対象物のキーワードに対するリズムパターンの連想度 Q_{kj} を用いる。それと同時に、同じく前処理 4 と同様に*i*番目のサンプル色に対する*j*番目のリズムパターンの連想度 R_{ij} を用いる。

4.5 本処理：任意の入力画像に対する音楽の自動アレンジ

4.5.1 任意の入力画像に印象の合うリズムパターンの自動選定

本処理では任意の画像を入力すると、まずその画像を 66 色に減色し、各サンプル色の色重要度 C_i を算出する。また R_{ij} および Q_{kj} は前処理で算出された値をそのまま用いる。画像に写る対象物のキーワードを*n*個であるとする、本処理では以下の式

$$P_j = a \sum_{i=1}^{66} C_i R_{ij} + b \sum_{k=1}^n Q_{kj} \quad (5)$$

により、入力画像に対する*j*番目のリズムパターンの連想度 P_j を導出する(ただし*a*, *b*は任意の正実数)。本

処理では、この値が最大であるリズムパターンを選ぶことで、入力画像に印象が合うと思われるリズムパターンを特定する。

4.5.2 音楽の自動アレンジ

我々の現時点での実装では、音楽情報に GM (General MIDI) を想定し、メロディ、和音、ベース、リズムの 4 パートが所定のチャンネルに記録された SMF (Standard MIDI File) 形式の楽曲ファイルを用いる。画像の印象に合うリズムパターンを特定すると、我々の実装は、それにしたがってリズムとベースを差し替えることにより、楽曲をアレンジする。

なお現時点では単純のために、アレンジ時にメロディ・和音のパートを一切変更しない。メロディや和音の発音のタイミングなどをリズムパターンに応じて操作する実装ができれば、より効果的な音楽アレンジが実現できると考えられる。

5. 実行結果

5.1 実験方法

我々は本手法の前処理および本処理を C 言語で実装した。そして、Windows XP 上にて Cygwin をインストールし、Cygwin 上の GNU gcc でコンパイルした実行ファイルを実験に用いた。我々の実装では、画像には 24 ビット非圧縮型の BMP 形式ファイルを、音楽情報には SMF 1.0 形式ファイルを採用した。BMP ファイルおよび SMF ファイルの入出力に汎用のライブラリ等は用いず、我々独自のライブラリを用いて入出力処理を実装した。

以下に記述する本実験では、8 人の被験者を依頼した。8 人の年齢は全員 20 代で、内訳は男性 3 人、女性 5 人、社会人 3 人、学生 5 人、音楽演奏熟達者 5 人、非熟達者 3 人であった。

本実験ではまず準備段階として、同じメロディに対して異なる 7 種類のリズムパターンを適用した 7 種類の音楽上法を、被験者に鑑賞させた。本実験では 4.5.2 節で説明した SMF 形式の音楽情報を用意し、ベース、リズム、の 2 パートだけを差し替えた 7 種類のファイルを用意した。SMF ファイルは再生環境により音質などが変化してしまうので、本実験では我々の環境（シーケンサーソフトウェア YAMAHA SOL 2.0 およびシンセサイザー Roland SC-8820）で再生したものを MP3 形式で録音し、これを被験者に鑑賞させた。

続いて以下の 4 種類の実験を用い、式(5)によりリズムパターンの連想度 P_j を求めた。なお、実験 1~4 で使用された楽曲は一貫して、著者の一人である伊

藤によって作曲されたものである。

[実験 1] 4.4 節の前処理 1 で示したように、所定の画像に写る対象物のキーワードを被験者に提示し、各々のキーワードから連想されたリズムパターンを回答させた。この回答結果からリズム連想度 Q_{kj} を得た。

[実験 2] 4.4 節の前処理 2 で示したように、所定の 66 色を被験者に提示し、各々の色から連想されたリズムパターンを被験者に回答させた。この回答結果からリズム連想度 R_{ij} を得た。

[実験 3] 4.4 節の前処理 3 で示したように、サンプル画像 25 枚を被験者に提示し、各々の画像から連想されたリズムパターンを被験者に選択させた。この回答結果から式(4)を用い、リズム連想度 R_{ij} を得た。

[実験 4] 4.4 節の前処理 4 で示したように、[実験 1]と[実験 3]を同時に行い、この回答結果からリズム連想度 Q_{kj} とリズム連想度 R_{ij} を得た。

[実験 1]～[実験 4]のいずれにおいても、本処理では任意の画像を入力した際に、式(5)を用いてその画像から各リズムパターンへの連想度 P_j を算出し、この値が最大であるリズムパターンを用いて楽曲をアレンジした。ただし[実験 1]では、式(5)において $a=0$ とし、[実験 2][実験 3]では、式(5)において $b=0$ とした。我々の実験では本処理において、様々な風景が写った、サンプル画像とは異なる 25 枚の画像を用いた。

以上の実験結果の一部は、以下の URL に掲載されている。 <http://itolab.is.ocha.ac.jp/~kisa/dival/>

5.2 実験結果の分析

本研究では、提案手法の本処理によるリズムパターン選択結果の妥当性を検証するために、本処理で用いた画像を被験者にも提示し、被験者が自分で選択したリズムパターンと比較し、両者の一致率を算出した。この結果を以下に記述する。

実験 1,2,3 の一致率

前処理に[実験 1][実験 2][実験 3]を採用した際に、本処理の結果と被験者の回答との一致率は以下の通りであった。

- 実験 1 ... 11%
- 実験 2 ... 28%
- 実験 3 ... 48%

これらの結果から、リズムパターンの自動選出には、キーワードやサンプル色などを提示するよりも、サンプル画像を提示することが好ましい、ということが示唆される。ただしいずれの実験結果も、一致率

は 50%未満であり、十分な結果ではないといえる。

実験 4 におけるキーワードおよび色分布の相関性

続いて[実験 4]を行い、その過程において被験者ごとに回答を観察した。その結果として我々は被験者を、

[被験者グループ 1] 対象物とリズムパターン連想度の相関性より、色とリズムパターン連想度の相関性の方が大きい被験者

[被験者グループ 2] 色とリズムパターン連想度の相関性より、対象物とリズムパターン連想度の相関性の方が大きい被験者

の 2 グループに分類した (分類方法については付録参照)。そこで[実験 4]では、式(5)の係数 a , b について、[傾向 1]の被験者に対しては a が b の 2 倍、[傾向 2]の被験者に対しては b が a の 2 倍となるように、 a , b の値を使い分けた。その結果、一致率は 71%に上昇した。以上の結果により、リズムパターンの自動選出には、画像の色分布だけでなく、画像に写る対象物の種類を表すキーワードも参照したほうが望ましいこと、また色分布とキーワードの重み付けに相当する係数 a, b の値を被験者ごとに使い分けたほうが好ましいこと、がわかった。

リズムパターンの特徴と実験結果の関係

続いて、リズムパターンに対する被験者間の回答の相関性について観察した。その結果、各リズムパターンの選択される回数に、被験者間の偏りの大きいものと小さいものが見られた。1 人の被験者を例にとると、9 枚の画像に対して選ばれたリズムパターンもあれば、2 枚の画像に対してしか選ばれなかったリズムパターンもあった。選ばれる回数の少ないリズムパターンは、もともと入力されたメロディに対してアレンジしにくいリズムパターンである、ということも想定される。この点について統計的に有意性のある結果を出すためには、さらに多くのリズムパターンを用意した上での再実験が必要であると考えられる。

画像の色分布と実験結果の関係

続いて、画像の色分布に対する被験者間の回答の相関性について観察した。ここで 66 色のサンプル色のうち有彩色から、以下の 2 つの色グループを抽出した。

[色グループ 1] 彩度または明度の最低一方が「高」であり、他方が「中」または「高」である色

[色グループ 2] 彩度または明度の最低一方が「低」であり、他方が「中」または「低」である色

続いて各画像に対して、

- ・ 色グループ 1 に属する色の重要度 C_i の累算値
 - ・ 色グループ 2 に属する色の重要度 C_i の累算値
- の各々を算出した。そして実験に用いた画像群を、色グループ 1 のほうが重要度累算値の高い画像と、色グループ 2 のほうが重要度累算値の高い画像に分類した。

その結果、[実験 2] [実験 3]において、色グループ 2 の重要度累算値のほうが高い画像群では、本処理の結果と被験者の回答との一致率は 77%であった。反対に色グループ 1 の重要度累算値のほうが高い画像群では、一致率は 53%であった。

このことより、明度、彩度の高い色は、色自体に強い印象があるため、[実験 2]に関しては、画像中の物体や情景の持つ意味と被験者の色印象が整合しない画像では失敗しやすい、ということがわかる。同様に、[実験 3]に関しては、画像中の物体や情景の持つ意味と、求められた色のリズム連想度 R_{ij} の色印象が整合しない画像では失敗しやすい、ということが推察される。つまり言い換えれば、4.3 節の図 3 に例示したような問題点は、明度、彩度の高い色が重要な存在をもつ画像において起こりやすく、このような画像にこそキーワードの存在が重要であると考えられる。またこれらの問題点を改善する一方向として、4.2 節で示した各色の重要度算出方法の再検討があげられる。

画像のキーワードと実験結果の関係

続いて、画像の色分布に対する被験者間の回答の相関性について観察した。我々は本実験に用いたキーワードを、我々自身の主観に基づいて、[キーワードグループ 1] 「海」「月」のように、特定の色が連想されやすいキーワード
[キーワードグループ 2] 「花」のように、多種多様な色を連想される可能性のあるキーワード
の 2 種類に分類した。

この結果、[キーワードグループ 1]に属するキーワードを持った画像では、式(5)の係数 b を大きくすると、一致率が 84%に向上した。反対に[キーワードグループ 2]に属するキーワードのみをもつ画像では、一致率は 63%にとどまっており、式(5)の係数 b を調節しても一致率は向上しなかった。

以上の結果により、特定の色が連想されやすいキーワードをもつ画像においては、式(5)の係数 a , b を $a < b$ となるように設定することで、より好ましい結果が得られることがわかった。

6. まとめ

本論文では、画像の色分布と、画像に写る対象物の種類から連想されるリズムパターンを用いることで、与えられた画像に印象が合うように音楽の自動アレンジ手法を特定する手法を提案した。また我々の実験においては、本手法の前処理では被験者に色を提示するより実写画像を提示するほうが良好な結果が得られること、また画像の色分布だけでなく画像に写る対象物の種類も考慮したほうが良好な結果が得られること、がわかった。

今後の課題として、以下のようなことを検討中である。

- ・ 4.2 節で示した色彩対比に関する式(1), およびその係数である b 値や p 値について再考したい。これらは、明度や彩度の高い画像において望ましくない結果が出やすい一因として考えられる。
- ・ 統計的な有意性の高さを出す方向に、実験方法を改良したい。特に本実験ではリズムパターンを 7 種類しか用意できなかったことから、さらに多くのリズムパターンを用意して再実験を行いたい。
- ・ リズムパターンだけでなく、それ以外の音楽の構成要素（メロディ、和音、音色）を変化させる音楽の自動アレンジについて検討したい。
- ・ 前処理におけるリズムパターン算出式の導出過程に、学習アルゴリズムを導入したい。これにより、ユーザの満足度の向上、前処理の負担の軽減、などの効果が期待できる。

謝辞

本研究の被験者の方々に感謝の意を示します。

参考文献

- [1] 岩宮, 音楽と映像のマルチモーダル・コミュニケーション, 九州大学出版会, ISBN4-87378-647-9, 2000.
- [2] 井口編, 長田他, 感性情報処理, 電子情報通信学会ヒューマンコミュニケーション工学シリーズ, オーム社, ISBN 4-274-07774-8, 1994.
- [3] 大山, 伊藤, DIVA: 画像の印象に合わせた音楽自動アレンジの一手法, 第 22 回 NICOGRAPH 論文コンテスト, 2006.
- [4] Ohyama K., Itoh T., DIVA: An Automatic Music Arrangement Technique According to Impressions of Images, 7th International Symposium on Smart Graphics, 2007 (accepted).
- [5] 上田, 感性キーワードの発展とその限界, Library

and Information Science, No.41, pp. 17-25, 1999.

[6] Sato A. Ogawa J. Kitakami H., An Impression Based Retrieval System of Music Collection, Knowledge Based Intelligent Engineering Systems and Allied Technologies, Vol. 2, pp. 856-859, 2000.

[7] 池添, 梶川, 野村, 音楽感性空間を用いた感性語による音楽データベース検索システム, 情報処理学会論文誌, Vol. 42, No. 12, pp.3201-3212, 2001.

[8] 熊本, 太田, 印象に基づく楽曲検索システムの設計・構築・公開, 人工知能学会論文誌, Vol.21, No.3, pp.310-318, 2006.

[9] Sugihara, T. Morimoto, K. Kurokawa, T., An Improved Kansei-based Music Retrieval System with a New Distance in a Kansei Space, 13th IEEE International Workshop on Robot and Human Interactive Communication, pp. 141-146, 2004.

[10] Feng Y., Zhejiang Y., Pan Y., Popular music retrieval by detecting mood, 26th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 375-376, 2003.

[11] Vignoli F., Pauws S., A Music Retrieval System Based on User Driven Similarity and Its Evaluation, 6th International Conference on Music Information Retrieval, pp. 272-279, 2005.

[12] 古賀, 下塩, 画像に合った音楽の選定技術, ヒューマンコミュニケーション基礎研究会技報, 平 11-9, 1999.

[13] 北島, 土居, 相互検索のための画像と音楽の印象語を介した関連付け手法, 第 20 回ファジィシステムシンポジウム後援論文概要集, 3E3-3, 2004.

[14] 北島, 土居, 画像と音楽の分類と相互検索, 第 47 回システム制御情報学会研究発表講演会講演論文集, pp. 157-158, 2003.

[15] Nakanishi T., Kitagawa T., An Implementation Method of a Heterogeneous Associative Media Data Search for Music Data and Image Data, 24th IASTED International Conference on Database and Applications, pp. 143-152, 2006.

[16] Oda M., Itoh T., MIST2: An Automatic Image Selection Technique According to Impression of Music, NICOGRAPH INTERNATIONAL 2007 (In Process).

[17] Hermann T., Ritter H., Listen to your Data: Model-Based Sonification for Data Analysis, International Institute for Advanced Studies in System research and cybernetics, pp. 189-194, 1999

[18] Kaper H. G., Tipei S, Wiebel E., Data Sonification and Sound Visualization, Computing in Science and

Engineering, Vol. 1, No. 4, 48-58, 1999.

[19] Schwende H., Auditory Emotional Access to Visual Information, Lecture Notes in Computer Science, Vol. 2398, p. 445, 2002.

[20] Giannakis K., Smith M., Auditory-Visual Associations for Music Compositional Processes: a Survey, International Computer Music Conference, pp. 12-15, 2000.

[21] 佐藤, 画像から音楽を自動演奏, ピクチャーメロディー, vol. 2, <http://www.forest.impress.co.jp/article/2003/04/21/okiniiri.html>

[22] 安達, 岩宮, 色彩と音楽が互いに及ぼす影響--シヨパンのエチュードを手がかりに, 第 5 回学生のための研究発表会講演論文集(日本音響学会九州支部), pp. 13-16, 2003.

[23] 北島, 土居, 画像の構成を手掛かりとした音楽の検索, 平成 15 年度情報処理学会関西支部大会講演論文集, pp. 39-42, 2003.

[24] 大林, 銀河の道 虹の架け橋, 小学館, 1999.

[25] 色について <http://park11.wakwak.com/~d-o-b/color/>

付録：実験 4 における被験者の分類

画像a, 画像bにおけるi番目の色の重要度を C_{ai} , C_{bi} としたとき, 色分布に基づく画像ab間のユークリッド距離 $D_{c(a,b)}$ を以下のように定義する.

$$D_{c(a,b)} = \sqrt{\sum_{i=1}^{66} (C_{ai} - C_{bi})^2} \quad (6)$$

また画像a, 画像bにおけるj番目のキーワードの連想度を Q_{aj} , Q_{bj} とし, 実験に用いたキーワードの個数をnとしたとき, キーワードに基づく画像ab間のユークリッド距離 $D_{w(a,b)}$ を以下のように定義する.

$$D_{w(a,b)} = \sqrt{\sum_{j=1}^n (C_{aj} - C_{bj})^2} \quad (7)$$

このとき, 実験に用いたN枚の画像から全ての組み合わせで 2 枚を選んでユークリッド距離を算出した場合に, $D_{c(a,b)}$ および $D_{w(a,b)}$ の平均値は以下の通りである.

$$\overline{D_c} = \frac{2}{N(N-1)} \sum_{j=i+1}^N \sum_{i=1}^N D_{c(i,j)} \quad (8)$$

$$\overline{D_w} = \frac{2}{N(N-1)} \sum_{j=i+1}^N \sum_{i=1}^N D_{w(i,j)} \quad (9)$$

一方、実験に用いたリズムパターンをM個とし、k番目のリズムパターンが選ばれている画像の個数を N_k とする。また、k番目のリズムパターンが選ばれている画像どうしのユークリッド距離を $D_{c(k,a,b)}$ および $D_{w(k,a,b)}$ とする。このとき、 $D_{c(k,a,b)}$ および $D_{w(k,a,b)}$ の平均値は以下の通りである。

$$\overline{D_{ck}} = \sum_{k=1}^M \frac{2}{N_k(N_k - 1)} \sum_{j=i+1}^{N_k} \sum_{i=1}^{N_k} D_{c(k,i,j)} \quad (10)$$

$$\overline{D_{wk}} = \sum_{k=1}^M \frac{2}{N_k(N_k - 1)} \sum_{j=i+1}^{N_k} \sum_{i=1}^{N_k} D_{w(k,i,j)} \quad (11)$$

ここで式(10)(11)の算出値が小さいほど、同一のリズムパターンが選ばれる画像間のユークリッド距離が小さい、つまり被験者のリズムパターン選択との相関性が高い、と考えることができる。よって我々は、

$$\frac{\overline{D_{ck}}}{\overline{D_c}} < \frac{\overline{D_{wk}}}{\overline{D_w}} \quad (12)$$

が成立するような実験結果を導く被験者は、対象物とリズムパターン連想度の相関性より、色とリズムパターン連想度の相関性の方が大きい被験者である、と判断した。さもなければ、色とリズムパターン連想度の相関性より、対象物とリズムパターン連想度の相関性の方が大きい被験者である、と判断した。

著者紹介



大山 喜冴

2006年お茶の水女子大学理学部情報科学科卒業。現在お茶の水女子大学大学院人間文化研究科数理・情報科学専攻在学中。情報処理学会会員。



伊藤 貴之

1990年早稲田大学理工学部電子通信学科卒業。1992年早稲田大学大学院理工学研究科電気工学専攻修士課程修了。同年日本アイ・ビー・エム(株)入社。1997年博士(工学)。2000年米国カーネギーメロン大学客員研究員。2003年から2005年まで京都大学大学院情報学研究科COE研究員(客員助教授相当)。2005年日本アイ・ビー・エム(株)退職。2005年よりお茶の水

女子大学理学部情報科学科助教授。ACM, IEEE Computer Society, 情報処理学会, 芸術科学会, 画像電子学会, 他会員。