動画特徴量からの印象推定に基づく動画 BGM の自動生成

清水 柚里奈* 菅野 沙也* 伊藤 貴之* 嵯峨山 茂樹†

概要. 動画を撮影して SNS で公開する際に，BGM を付与して楽しむ人が増え，またそれを支援するアプリも増えてきた．本研究では，動画から一定時間ごとに抽出した動きや色の動画特徴量から動画の印象を推定し，その結果に基づいて楽曲生成を行うことで，動画の印象に合った楽曲を付与する手法を提案する．また，ユーザに予め印象を回答してもらったリズム・メロディ素材をマッシュアップすることで楽曲生成を行うことから，ユーザごとの印象の違いを考慮した楽曲生成が可能となる．これにより，印象に合った音楽を自分で探すことなく動画に付与することができる．

1 はじめに

近年，写真や動画を撮影する機会が増え，またその撮影したものを SNS サイトに投稿することで，多くの人々と共有して楽しむようになった．その際に，撮影映像に BGM を付与するなどの動画編集も行うようになってきたが，動画編集では一般的に，動画に合った音楽を自分で探したり，動画の長さに合うように音楽を調整したり，といった手間とスキルが必要となる．そこで本手法では，動画特徴量からの印象推定結果に基づいて楽曲生成を行う手法を提案する．また，ユーザの印象と動画特徴量，音楽特徴量の関係を学習させることで，動画・音楽の印象を推定することから，ユーザ 1 人 1 人の動画に対する印象に合った音楽を生成することが可能となる．

2 提案手法

2.1 動画特徴量

現時点の我々の実装は，色分布，動き分布の 2 種類の低レベルな特徴量と印象の関係を学習している．

2.1.1 色分布の特徴量抽出

まず動画から 5 秒ごとに静止画を抽出し，その静止画の各々に対して OpenCV を用いて 12 色(黒, 灰色，白，茶色，赤，オレンジ，黄色，緑，水色，青，ピンク，紫)の減色処理を施し，各色の画素数を集計することにより，カラーヒストグラムを得る．得られたそのヒストグラムの数値から各色の画素数の平均を求め，これを動画全体に対する平均の色の割合とみなし，12 次元の特徴量ベクトルとする．

2.1.2 動き分布の特徴量抽出

まず動画を時間で 4 分割し，各時間帯に対して OpenCV を用いてオプティカルフローを求める．次にそのオプティカルフローを構成するベクトル群の速度・角度を集計し，各々のヒストグラムを生成する．そして速度の平均・分散，速度のヒストグラム上で度数が最大となる階級値，角度の分散，角度のヒストグラム上で度数が最大となる階級値を求める．各特徴量の全体の平均を求め，これら計 5 つを動きの特徴量とみなす．

2.2 音楽特徴量

現時点での我々の実装では，メロディとリズムを別々の素材として用意し，それぞれ図 1 に示す音楽特徴量を文献[1]，文献[2]を参考に算出している．

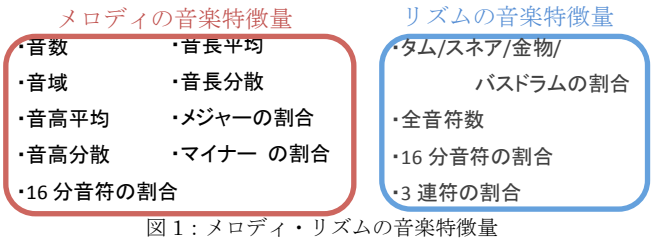

図 1：メロディ・リズムの音楽特徴量

2.3 学習

続いて本手法では，動画特徴量，リズム・メロディの音楽特徴量に対する各ユーザの印象の関係を学習する．

2.3.1 ユーザ印象評価

まず予め用意したサンプル動画，サンプルリズム・メロディを評価する際に使用する感性語対を決定する．本手法では心理学の観点から，また動画と音楽に共通して適用できそうな感性語対を選んだ．その中で動画の色・動きに関して適用する感性語，リズム・メロディに関して適用する感性語を，我々自身の主観に基づいて，図 2 のように定めた．

本手法では各ユーザにサンプル動画，メロディ・リズムを閲覧してもらい，上に挙げた感性語への適応度を 6 段階評価で回答してもらう．以後，この適

* お茶の水女子大学, † 明治大学

合度を印象値と称する．このようにして，各ユーザの印象値を収集する．

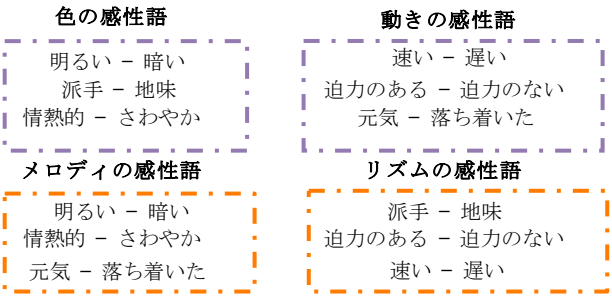

図 2：動画の色・動き,リズム・メロディに関する感性語

2.3.2 色分布からの印象学習

2.1.1 項で示した色分布の特徴量から印象値を推定する．v_{ki} は k 番目の動画における i 番目の色の頻度とする．また 3.3.1 項のユーザ印象評価で得られた 6 段階評価の値を[-1,1]の範囲で 6 等分した値とみなし，j 番目の印象語に対する k 番目の動画の評価に対応する数値を印象値 a_{kj} とする．そして i 番目の特徴量と j 番目の印象語に対する評価の値との関係 c_{ij} を以下の式(1)を用いて求める．以上の処理によってサンプル動画を用いた学習を終えた後，以下の式(2)を用いて，ユーザ評価結果の与えられていない動画の j 番目の印象語に対する印象値 a_j を算出する．ただし v_i は新しい動画における i 番目の色の頻度とする．

$$\sum_{k=1} a_{kj} v_{ki} \quad (1) \qquad a_j = \frac{\sum_{i=1} c_{ij} v_i}{\sqrt{\sum_{i=1} {c_{ij}}^2}} \quad (2)$$

2.3.3 動き分布，音楽特徴量からの印象学習

2.3.1 項のユーザ印象評価で得られた 6 段階の値と 2.1.2 項で示した動き分布の特徴量から重回帰分析を用いて計算式を求め，ユーザ評価結果の与えられていない動画に対して，動き分布の印象値を推定する．音楽特徴量についても同様に印象値を推定する．

2.4 楽曲生成

次に楽曲の素材となるメロディとリズムを選出し，合成する．2.3.2 項と 2.3.3 項で算出した動画の印象値，メロディ・リズムの印象値を比較して，ユークリッド空間上で最も距離の近いメロディ・リズムを選出し，それらを組み合わせて楽曲を生成する．続いて生成した楽曲にコード進行を加える．さらに，動画の再生時間に合うように小節数やテンポを設定する．以上によって生成された楽曲と動画を合成することで，動画に BGM を付与する．

3 実行結果と考察

本手法で使用するメロディには自動作曲システム Orpheus[3]を利用して作成した 30 パターンを用意し，リズムには文献[2]で使われていた 21 パターンを用意した．このうちメロディ 15 種類，リズム 10 種類を学習用のサンプルメロディ・サンプルリズムとした．また動画は 1 分以内の 11 種類の動画をサンプルビデオとして用意した．本実験ではユーザ A とユーザ B の各々に対してユーザ印象評価を依頼し，この結果をもとにしていくつかの異なるジャンルの動画に対して楽曲生成を行った．以下の 2 種類の動画に対して楽曲を付与した結果を表 1 に示す．

動画 1：人がいない夕暮れの海辺の様子
動画 2：犬が草むらを元気に走っている様子

表 1：動画 1,2 の楽曲生成を行った結果

	ユーザ A	ユーザ B
動画 1	melody22.mid rhythm6.mid	melody29.mid rhythm9.mid
動画 2	melody23.mid rhythm20.mid	melody17.mid rhythm20.mid

ユーザ A とユーザ B では異なる楽曲素材が選ばれており，学習段階の影響によりユーザの印象の違いを考慮した楽曲が生成されていることが分かる．しかし動画 2 の明るく元気な動画であるのに対し、ユーザ A とユーザ B でゆったりとした落ち着いた楽曲が生成されてしまった．このことから，例えば，ユーザ印象評価の改善や，動画および楽曲の特徴量の見直しなどが必要である．

4 まとめと今後の課題

本報告では動画から一定時間ごとに抽出した動きや色の動画特徴量から動画の印象を推定し，その結果に基づいて楽曲生成を行うことで，動画の印象に合った楽曲を付与する手法を提案した．今後の課題として，学習段階におけるユーザ印象評価，動画および音楽の特徴量，印象値の推定方法などを再検討することが挙げられる．また現段階では単純な音形で付与しているコードの弾き方を，リズムや曲調に合わせて変えることも検討する．

参考文献

[1] 中山達喜，吉田真一，”音楽分類における特徴量の検討”，ファジィシステムシンポジウム講演論文集，Vol. 26, pp. 1256-1261, 2010.

[2] 菅野沙也，伊藤貴之，”入力文書の印象と感情に基づく楽曲提供の一手法”，情報処理学会音楽情報科学研究会，Vol. 2014-MUS-103, 2014.

[3] 東京大学 大学院情報理工学系研究科 システム情報学専攻，自動作曲システム Orpheus, http://www.orpheus-music.org/v3/