

playlist2scapeseq: 楽曲プレイリストに基づくシーケンス景観の生成

杉本 蒼 奥 健太

龍谷大学 理工学部 情報メディア学科

t140468@mail.ryukoku.ac.jp okukenta@rins.ryukoku.ac.jp

概要 入力された楽曲プレイリストに合ったシーケンス景観をスライドショーとして提示するシステム、playlist2scapeseq を提案する。シーケンス景観とは、ドライブ中に景観が遷り変わるといったように、視点を移動させながら連続的に遷り変わっていく景観のシーケンスである。一方で、視点を固定したときの景観をシーン景観とよぶ。我々はこれまでにシーン景観について、四つの景観要素（田園系，山林系，水辺系，都市系）で表現される景観ベクトルを定義した。本研究では、景観ベクトルで表現されるシーン景観を時系列的に連続させることでシーケンス景観を構成する。併せて、楽曲については、クラウドソーシングにより得られた楽曲-景観関係データに基づき景観ベクトル化を行う。playlist2scapeseq は、入力された楽曲プレイリスト内の各楽曲の景観ベクトルと類似するシーン景観を探索し、時系列的につなぎ合わせることでシーケンス景観を生成する。

キーワード 楽曲プレイリスト，スライドショー生成，シーケンス景観

1 はじめに

ドライブ中に音楽を聴くことはドライブ体験をより楽しいものに変えてくれる。このような需要の下、ドライブ時のコンテキストに合った楽曲を推薦する、コンテキストウェアドライブ楽曲推薦システムに関する研究 [1][2] が行われている。我々は、ドライブ時の景観をコンテキストの一つとして扱った景観ウェアドライブ楽曲推薦システムの実現を目指し、これまでに景観ウェア楽曲推薦システム [3] を提案してきた。このシステムは、景観に合った楽曲を推薦するシステムであり、景観が入力となり楽曲が出力となる。

一方で、本稿では、このシステムの入出力を逆転させ、入力として楽曲（プレイリスト）が与えられたとき、その楽曲に合った景観が出力されるシステムを検討する。これは、ユーザがドライブ時に自分の好きな楽曲プレイリストをセットしたとき、そのプレイリストに合ったルートを紹介し誘導してくれるようなシステムへの拡張を想定している。このようなシステムを実現することで、新たな娯楽性をもったドライブ体験を創出することを目指している。

本稿では、この考えの下、入力された楽曲プレイリストに合ったシーケンス景観をスライドショーとして提示するシステム、playlist2scapeseq を提案する。シーケンス景観とは、ドライブ中に景観が遷り変わるといったように、視点を移動させながら連続的に遷り変わっていく景観のシーケンスである。一方で、視点を固定したとき、1枚の写真に収めるような景観をシーン景観とよぶ。我々はこれまでにシーン景観について、四つの景観要素

（田園系，山林系，水辺系，都市系）で表現される景観ベクトルを定義した [4]。本稿では、景観ベクトルで表現されるシーン景観を時系列的に連続させることでシーケンス景観を構成する。併せて、楽曲については、クラウドソーシングにより得られた楽曲-景観関係データに基づき景観ベクトル化を行う。playlist2scapeseq は、入力された楽曲プレイリスト内の各楽曲の景観ベクトルと類似するシーン景観を探索し、時系列的につなぎ合わせることでシーケンス景観を生成する。

2 問題定義

本章では、景観ベクトルおよび楽曲，景観，道路ネットワークの定義を示した後、本研究の問題定義を述べる。

定義 1：景観ベクトル。 景観ベクトルは、奥らの先行研究 [4] で定義した 4 種類の景観要素（田園系 (r)，山林系 (m)，水辺系 (w)，都市系 (u)）から構成される 4 次元の確率ベクトルとして定義される。ベクトルの各要素はその景観要素が含まれる確率を表す。したがって、全要素の総和は 1 となる。

定義 2：楽曲。 一つの音楽の曲を 1 楽曲 t_i と定義する。ここで、楽曲 t_i は景観ベクトルで表現され、 $t_i = (t_i^r, t_i^m, t_i^w, t_i^u)$ と表す。

楽曲プレイリストを楽曲 t_i のシーケンスとみなし、 $\mathcal{P} = \langle t_1, t_2, \dots, t_n \rangle$ と表現する。これは、楽曲 t_1, t_2, \dots, t_n の順に時系列に楽曲が登録されていることを表す。

定義 3：景観。 一つの景観の単位を l_j で表す。ここで、楽曲と同様に、景観 t_i は景観ベクトルで表現され、 $l_i = (l_i^r, l_i^m, l_i^w, l_i^u)$ と表す。

シーケンス景観を景観 l_j のシーケンスとみなし、 $\mathcal{L} = \langle l_1, l_2, \dots, l_m \rangle$ と表現する。これは、景観 l_1, l_2, \dots, l_m の順に時系列に景観が遷移することを表す。

定義 4: 道路ネットワーク. 道路ネットワークは有向グラフ $G^R = (V^R, E^R)$ で表現される。ここで、 V^R は道路ノードの集合であり、 E^R は道路リンクの集合である。各道路ノードは地理座標系における位置座標 (経度 x , 緯度 y) をもつ。各道路リンク $r_i \in E^R$ は端点 $r_{i.s}$ および $r_{i.e}$ をもつ。ここで、 $r_{i.s} \in V^R$ は r_i の始点、 $r_{i.e} \in V^R$ はその終点である。それぞれの位置座標は $(r_{i.s.x}, r_{i.s.y})$ および $(r_{i.e.x}, r_{i.e.y})$ で表す。

本研究における問題は、与えられた楽曲プレイリスト P において、そのプレイリストに合ったシーケンス景観 \mathcal{L} を生成することである。

3 システム概要

本稿では、playlist2scapeseq を提案する。playlist2scapeseq は、ユーザが登録した楽曲プレイリストに合ったシーケンス景観を生成し、そのシーケンス景観を景観画像スライドショーの形式で提示する。

このシステムは以下の要素から構成される：

- (a) 楽曲の景観ベクトル化 (4 章)。
- (b) 景観画像の取得および景観ベクトル化 (5 章)。
- (c) 楽曲プレイリストに基づくシーケンス景観の生成 (6 章)。

以下、それぞれについて各章で説明する。

4 楽曲の景観ベクトル化

どの楽曲がどの景観にマッチしているかという判断は、人間の感性に委ねる部分が多い。実際にはその感性は個人に依存することも考えられるが、本研究では、先行研究 [3] にならって、まず一般的な感性の獲得を試み、クラウドソーシングを用いる。クラウドソーシングにより得られた楽曲-景観関係データに基づき、楽曲の景観ベクトル化を行う。

任意の楽曲には MP3 形式の楽曲ファイル (MP3 形式) が含まれる。事前に全楽曲について楽曲長の中心から前後 30 秒間を切り出し、評価用ファイルとしておく。楽曲 $t_i \in T$ について、以下の手順によりクラウドソーシングを用いて楽曲-景観関係データを収集する。

- (1) ワーカは評価用ファイルを 30 秒間聴取する。

- (2) ワーカは聴取した楽曲と最もマッチしていると感じる景観イメージとして、「田園景観」、「山林景観」、「水辺景観」、「都市景観」、「いずれにも該当しない」の選択肢の中から一つを選択する。何らかの不具合により楽曲が再生されなかった場合は「楽曲が再生されない」を選択する。

1 楽曲あたり複数名のワーカが回答する。1 名のワーカによる 1 楽曲に対する回答を 1 タプルとして workers-tracks-landscape テーブルに登録する。このテーブルは次の関係スキーマで定義される。

```
workers-tracks-landscape(worker_id,
                           track_id, landscape)
```

ここで、worker_id はワーカ ID、track_id は楽曲 ID、landscape は選択された景観ラベルをそれぞれ示す。

workers-tracks-landscape テーブルを基に、楽曲-景観関係行列 $R = [r_i^r, r_i^m, r_i^w, r_i^u]$ を作成する。ここで、 r_i^k は、楽曲 t_i に対し景観ラベル k を選択したワーカ数を表す。ただし、 $\sum_k r_i^k = 1$ となるように正規化する。

5 景観画像の取得と景観ベクトル化

5.1 景観画像の取得

本研究は将来的に楽曲プレイリストに合ったドライブルート推薦への拡張を想定している。そのため、景観画像として任意の道路リンク上を視点とした画像を用いる。

まず、以下の処理により任意の件数の道路リンクを取得する：

- (1) OpenStreetMap¹ から日本全国の道路ネットワークデータを取得する。本研究ではドライブルートに想定しているため、歩行者向け道路や自転車向け道路などは除外し、自動車向け道路に該当する道路リンクのみを選択する。
- (2) 取得した道路ネットワークデータから任意の件数の道路リンク $r_i \in E^R$ を抽出する。

つづいて、道路リンク r_i について、以下の処理により景観画像 l_i を取得する。

- (1) 道路リンク r_i の始点 $r_{i.s}$ と終点 $r_{i.e}$ から中点 (m_x, m_y) および進行方向 θ (真北方向を 0° とする) を算出する。具体的には、それぞれ次式で算出される：

$$m_x = \frac{d_x}{2}, m_y = \frac{d_y}{2}, \quad (1)$$

¹<http://download.geofabrik.de/>

$$\theta = \begin{cases} 90^\circ & d_x > 0 \text{ かつ} \\ & d_y = 0 \text{ のとき} \\ 270^\circ & d_x < 0 \text{ かつ} \\ & d_y = 0 \text{ のとき} \\ 180^\circ + \frac{180^\circ}{\pi} \arctan \frac{d_x}{d_y} & d_y < 0 \text{ のとき} \\ \frac{180^\circ}{\pi} \arctan \frac{d_x}{d_y} & \text{それ以外のとき} \end{cases}, \quad (2)$$

ただし, $d_x = r_i \cdot e_x - r_i \cdot s_x,$
 $d_y = r_i \cdot e_y - r_i \cdot s_y.$

(2) Google Street View Image API²を用いて, 中点 (m_x, m_y) を視点とした進行方向 θ のストリートビュー画像を取得する. 具体的には下記の URL を作成し, Google Street View Image API に問い合わせる.

```
https://maps.googleapis.com/maps/api/streetview?pitch=0&size=600x400&location=[m.y],[m.x]&heading=[\theta]
```

ここで, [] 内は変数であり, 式 (1) および式 (2) で求めた各値がそれぞれ代入される.

5.2 景観ベクトル化

5.1 節で取得した景観画像 l_i の景観ベクトルは, 景観画像 l_i に対応する道路リンク r_i に対して推定された景観ベクトルを用いる.

道路リンク r_i の景観ベクトル化は奥らの先行研究 [4] の成果を用いる. 文献 [4] では, JAXA が公開している土地被覆図³ を基に道路リンクの景観を推定する手法を提案してきた. 各景観要素をクラスとしたクラス確率推定問題として扱い, 与えられた道路リンクについて要素ごとのクラス確率を推定した. 推定されたクラス確率を基に道路リンクの景観ベクトルを作成した.

6 楽曲プレイリストに基づくシーケンス景観の生成

与えられた楽曲プレイリスト \mathcal{P} において, \mathcal{P} に合ったシーケンス景観 \mathcal{L} を生成する. 基本的な方式は, 楽曲の景観ベクトルと景観画像の景観ベクトルとの類似度に基づいてマッチングすることになる. 本稿では, 以下に示す 3 種類の方式を提案する. 以下, 各方式について説明する.

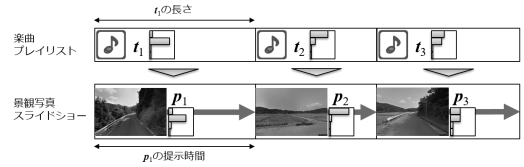


図 1 one-to-one 方式.

6.1 one-to-one 方式

one-to-one 方式は, 入力された楽曲プレイリスト \mathcal{P} 内の i 番目の楽曲 t_i について, 景観ベクトルの類似度が最も高い景観画像をシーケンス景観の i 番目の画像 p_i とする. 景観ベクトルの類似度はコサイン類似度により算出する. 楽曲 t_i と景観画像 l_i とのコサイン類似度は下記で定義される.

$$\text{Cosine}(t_i, l_i) = \frac{\sum_j t_{ij} l_{ij}}{\sqrt{\sum_j t_{ij}^2} \sqrt{\sum_j l_{ij}^2}} \quad (3)$$

one-to-one 方式のイメージを図 1 に示す. あらかじめ類似度に基づき, 楽曲プレイリスト \mathcal{P} 内の各楽曲 t_i に対応する景観画像 l_i を用意しておく. 楽曲 t_i が再生されている間, 景観画像 l_i が提示される.

6.2 事前ソート方式

one-to-one 方式では, 楽曲プレイリスト内の楽曲の順序はユーザにより指定された順序としている. そのため, 楽曲の順序によっては, 山林景観 → 都市景観 → 田園景観と遷移するなど, スライドショーにおける景観の遷り変わりが激しく感じられるという可能性もある. そこで, 事前ソート方式では, 隣り合う楽曲間において景観ベクトルの変化量が小さくなるように, あらかじめ楽曲プレイリスト \mathcal{P} 内の楽曲の順序をソートしておく.

事前ソート方式の処理は下記のとおりとなる (図 2).

- (1) 4 種類の景観要素のうちランダムに一つの景観要素を選択する.
- (2) one-hot 表現により, 選択された景観要素を 1 とし, それ以外の要素を 0 とした特徴ベクトルを作成する. これをシード特徴ベクトル s とする. 例えば, 山林景観が選択された場合, シード特徴ベクトルは $s = (0, 1, 0, 0)$ となる.
- (3) 楽曲プレイリスト \mathcal{P} 内の各楽曲 t_i について, シード特徴ベクトル s とのコサイン類似度を算出する.
- (4) コサイン類似度の降順に楽曲プレイリスト \mathcal{P} 内の楽曲をソートする. ソート後の楽曲プレイリストを \mathcal{P}' とする.
- (5) 楽曲プレイリスト \mathcal{P}' を基に one-to-one 方式を適用する.

²<https://developers.google.com/maps/documentation/streetview/?hl=ja>

³<http://www.eorc.jaxa.jp/ALOS/lulc/jlulc-jpn.htm>



図2 事前ソート方式.



図3 中間補完方式.

6.3 中間補完方式

one-to-one 方式では、一つの楽曲につき一つの景観画像を対応付けていた。中間補完方式では、隣り合う楽曲同士を滑らかにつなげるように、両楽曲の中間的な景観画像を補完する。

中間補完方式の処理は下記のとおりとなる (図3)。

- (1) 隣り合う楽曲として楽曲 t_i と楽曲 t_{i+1} を考える。このとき、両楽曲の景観ベクトルの平均ベクトル $t_{i,i+1}$ を求める。
- (2) 楽曲 t_i の楽曲長 25%–75% の範囲を景観ベクトル t_i に対応する範囲とする。楽曲 t_i の楽曲長 75% の位置から楽曲 t_i の楽曲長 25% の位置までの範囲を平均ベクトル $t_{i,i+1}$ の範囲とする。楽曲 t_{i+1} の楽曲長 25%–75% の範囲を景観ベクトル t_{i+1} に対応する範囲とする。ただし、楽曲 t_i が楽曲プレイリスト P 内の 1 曲目の楽曲の場合は、楽曲長 0%–75% の範囲を景観ベクトル t_i に対応する範囲とする。楽曲 t_i が楽曲プレイリスト P 内の最後の楽曲の場合は、楽曲長 25%–100% の範囲を景観ベクトル t_i に対応する範囲とする。
- (3) 各範囲において、それぞれの景観ベクトルとの類似度が高い景観画像を対応付ける。

7 関連研究

Hua ら [5] は、付随音楽と写真の特徴に基づく写真スライドショーシステムを提案している。写真の遷移位置を決定するために、楽曲の立ち上がり (オンセット) 位置を検出しておく。オンセット位置に基づき、楽曲を複数の部分楽曲に分割する。部分楽曲の切り替わり時点に合わせて写真を遷移させることでスライドショーを提示する。このシステムでは、楽曲の内容は写真を遷移させるタイミングを検出するために用いられており、楽曲の内容と写真の内容とのマッチングは行っていない。

Li ら [6] は、感情に基づく印象派絵画のスライドショーシステムを提案している。このシステムは感情をベースに伴奏音楽付きの絵画スライドショーを生成する。絵画

の色特徴およびテキスト特徴とアノテーションされた感情の対を学習データとし、これら特徴と感情との相関ルールを抽出する。抽出された相関ルールに基づき絵画に対する感情を推定し、この感情に基づき絵画をクラスタリングする。楽曲についても同様に楽曲の特徴 (メロディ、リズム、テンポ) と感情との相関ルールを抽出し、感情に基づきクラスタリングする。絵画と楽曲それぞれに対応付けられた感情の類似度に基づき、絵画と楽曲とをマッチングさせ、スライドショーを生成している。Cheng らの研究と本研究との違いは、本研究では楽曲と写真の特徴として景観ベクトルを用いている点である。

8 おわりに

本稿では、楽曲プレイリストに合ったシーケンス景観をスライドショーとして提示するシステム、playlist2scapeseq を提案した。playlist2scapeseq は、(a) 楽曲の景観ベクトル化、(b) 景観画像の取得および景観ベクトル化、(c) 楽曲プレイリストに基づくシーケンス景観の生成の要素から構成される。シーケンス景観の生成方式として、(A)one-to-one 方式、(B) 事前ソート方式、(C) 中間補完方式の三つの方式を提案した。

今後は、playlist2scapeseq の評価を行う。さらに、playlist2scapeseq を拡張した楽曲プレイリストに合ったドライブルート推薦システムを検討する。

謝辞

本研究は JSPS 科研費 15K12151 の助成を受けたものです。ここに記して謝意を表します。また、本研究で利用した楽曲データセットは、『クラシック名曲サウンドライブラリー』 (<http://classical-sound.seesaa.net/>) のサイト作成者のご厚意により使用許諾を頂きました。ここに謝意を表します。

参考文献

- [1] Aleksandar Stupar and Sebastian Michel. PICASSO To Sing you must Close Your Eyes and Draw. In *ACM SIGIR 2011*, pp. 715–724, 2011.
- [2] Linas Baltrunas, Marius Kaminskas, Bernd Ludwig, Omar Moling, Francesco Ricci, and Aykan Aydin. InCarMusic : Context-Aware Music Recommendations in a Car. In *EC-Web 2011*, pp. 89–100, 2011.
- [3] 糸井勇貴, 奥健太, 山西良典. 楽曲の風景特徴化に基づく風景アウェア楽曲推薦システム. In *DEIM Forum 2017*, A8–3, 2017.
- [4] 奥健太, 山西良典. 土地被覆図からの景観要素抽出に基づく道路リンクの景観ベクトル化. 情報処理学会研究報告, 2017-DBS-1, pp. 1–6, 2017.
- [5] Xian-Sheng Hua, Lie Lu, and Hong-Jiang Zhang. Content based photograph slide show with incidental music. In *IEEE ISCAS 2003*, 2003.
- [6] Cheng-Te Li and Man-Kwan Shan. Emotion-based impressionism slideshow with automatic music accompaniment. In *ACM MM 2007*, pp. 1–5, 2007.