

G-003

## 距離尺度に Earth Mover's Distance を用いたハミングによる類似音楽検索手法 Similar Music Retrieval Based on Humming Query using the Earth Mover's Distance

大西 泰代†      獅々堀 正幹†  
Yasuyo Ohnishi    Masami Shishibori

柘植 寛†      北 研二†  
Satoru Tsuge      Kenji Kita

### 1. はじめに

近年、ハードディスクの低価格化に伴い、デジタル化された音楽データが一般家庭の計算機でも容易に利用可能となった。しかし、蓄積された膨大な音楽データから必要なデータのみを見つけ出すことが困難となり、音楽データに対する効率の良い検索手法が必要になっている。音楽データに対する検索としては、曲名、歌手名や歌詞等を入力とするキーワード型の検索方式が一般的であるが、より汎用的にメロディやリズム等を入力とするコンテンツ型の検索方式が注目されている[1], [2]。

我々は、コンテンツ型検索方式の中でも、特にハミングを入力として MIDI 形式の音楽データを検索するシステムを開発している。MIDI 音楽データを対象にした従来のコンテンツ型音楽検索手法としては、音楽データのリズムや音高の変化を特徴ベクトル化し、特徴ベクトル間の類似性をユークリッド距離や DP マッチングで判定する手法が提案されている[1], [2]。しかし、これらの手法では、リズムと音高変化の類似性を同じ範疇の距離尺度で判定していないため、リズムと音高の双方(全体の曲調)が類似した曲を検索するためには工夫が必要であった。

そこで本稿では、距離尺度に Earth Mover's Distance (EMD) [3]を用いた類似音楽検索手法を提案する。EMD は、画像検索の分野で用いられる距離尺度であり、領域分割した画像内の領域の広さ、重心点、平均色等を特徴量とすることで、画像全体の構図が類似した画像(同じ位置に類似した色が配色された画像)を検索できるといった特徴を持つ。本手法では、曲内の音符の長さ、開始時刻、音高等を特徴量とし、EMD を距離尺度として用いることにより、全体の曲調が類似した音楽データを高精度に検索する。

### 2. 距離尺度に EMD を用いた類似画像検索

EMD は、線形計画問題の1つである輸送問題の解に基づいて計算される。輸送問題とは、一定の供給量を持つ複数の供給地と同じく一定の需要量を必要とする需要地を設定し、各供給地から需要地までの輸送コストが与えられた際に、需要地の需要を満たすよう供給地から需要地へ最小の輸送コストで荷物を輸送する輸送方法を探す問題である。


距離尺度に EMD を用いた画像検索手法[3]では、色情報に基づいて画像を領域分割し、領域毎に領域の広さ、平均色、重心点、テクスチャ情報を抽出する。そして、画像内の領域数を供給地数(需要地数)、各領域の広さを各供給地(需要地)が持つ供給量(需要量)とみなす。

更に、各領域の平均色、重心点、テクスチャ情報を固定長の特徴ベクトルで表現し、各領域間の距離(ground distance)を特徴ベクトル間のユークリッド距離で定義し、

この値を輸送コストと見なす。そして、比較対象の2つの画像の一方を供給サイド、他方を需要サイドとみなして画像間の類似度を EMD によって計算する。

例として、領域分割画像とその特徴量を図1に示し、図1を入力として類似画像検索した結果を図2に示す。この例のように、EMD を適用した類似画像検索手法は、色合いのみでなく構図も考慮した画像を検索可能であり、高い検索精度が示されている。

原画像 分割画像



領域数: 10

領域番号	領域の広さ	中心点	LUV	RGB	Wavelet特徴量
0	114	(8, 5)	(32, -7, 18)	#0d1405 ■ #4.038966, 4.304441, 1.221109)	
1	147	(3, 12)	(10, 0, -3)	#030304 ■ #2.908031, 2.902046, 1.148521)	
2	279	(4, 50)	(17, 1, 4)	#360604 ■ #3.761041, 3.60339, 1.257681)	
3	341	(11, 22)	(61, -1, 34)	#494d22 ■ #3.220512, 2.824208, 1.013474)	
4	410	(15, 46)	(62, -2, 41)	#4a531c ■ #1.762217, 1.851554, 0.530159)	
5	1918	(44, 31)	(67, -7, 43)	#329723 ■ #1.821663, 1.77224, 0.773979)	
6	108	(22, 15)	(57, 18, 14)	#593a2a ■ #7.357546, 6.597779, 2.65028)	
7	529	(38, 23)	(72, 8, 13)	#7a6719 ■ #3.306662, 2.385396, 0.975359)	
8	148	(37, 40)	(85, 5, 6)	#7c7c7c ■ #4.33389, 3.616271, 1.583303)	
9	102	(33, 48)	(61, 21, 28)	#985228 ■ #4.540121, 4.130689, 1.482269)	

図1 EMD を適用する画像特徴量の例

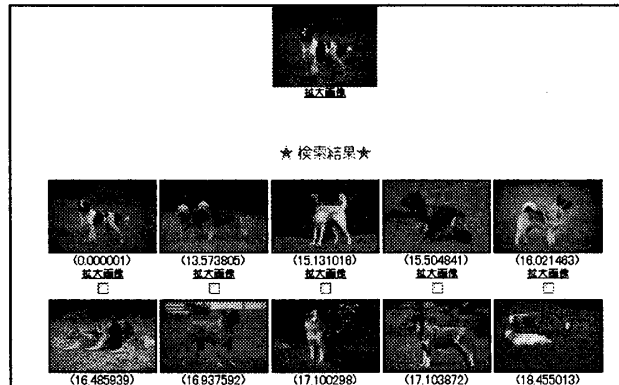


図2 EMD を用いた類似画像検索の例

### 3. 距離尺度に EMD を用いた類似音楽検索手法

従来手法として、リズムベクトルは音符数により曲全体を分割し、メロディ片を生成するため、メロディ片の長さは出現する音符数により変化する。一方、音高推移ベクトルは、一定のメロディ長によってメロディ片を生成するため、メロディ片の長さは一定であるが、音符数が増える。そのため、従来手法ではリズムと音高推移を同じ範疇の距離尺度で類似度を判定することが困難であった。

EMD を用いた類似画像検索が各領域の広さと各領域の画像的特徴(平均色、重心点)を同じ範疇の距離尺度で類似性を判定できたように、類似音楽検索の距離尺度に EMD を適用すると、音符の長さとは各音符の音楽的特徴

† 徳島大学 大学院工学研究科

‡ 徳島大学 高度情報化基盤センター

(Note-on 開始時刻, 音高) を同じ範疇の距離尺度で類似性を判定でき, その結果, 人間の感覚と同じように, 全体の曲調が類似した曲を精度よく検索できるのではないかと我々は考えた。

EMD を類似音楽検索手法に適用するにあたり, スライディングウインドウ方式[2]により分割されたメロディ片毎に特徴量を作成し, ground distance を求めるための特徴量は, 各音符の Note-on 開始時刻, 前音との音高差とした。図3に8分音符を1拍とした特徴量の例を示す。開始時間, 音長は拍数で表現し, 音高差は半音高い音を1, 半音低い音を-1とする。1行目にメロディ片内の音符の数を表記し, 2行目以降は, 出現音符の特徴量であり, 第1音は(音長, 開始時刻, 音高差) = (1.5, 0.0, 0)となる。

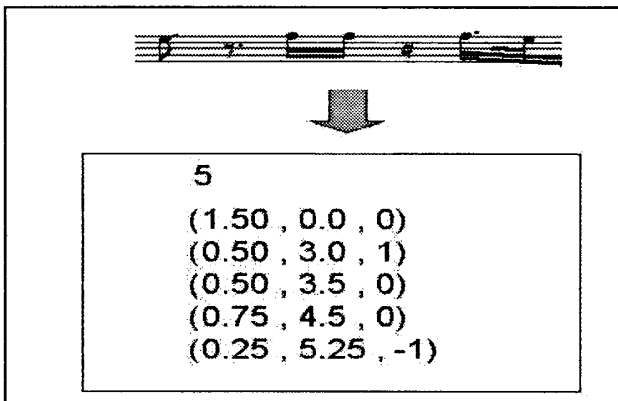


図3 EMD に適用する音楽特徴量の例

## 4. 評価

### 4.1 実験条件

検索対象データベースとして, 童謡, J-pop, 演歌等のジャンルが含まれるカラオケ用 MIDI 音楽データ 483 曲を使用した。検索入力には, 男女6名が歌ったハミングを採用し, 市販の採譜ソフト[4]で MIDI に変換した 40 曲を用いた。ハミングの長さは8拍を越える程度とし, ハミングの際には, 正確なハミングは要求せず, その曲を知っている人が聞いて分かるレベルの入力とした。そのため, 入力ハミングにはテンポ, 音程のずれが生じた。

スライディングウインドウの条件として8分音符を1拍とし, ウインドウ長8拍, スライド長2拍とし音楽データを分割し, メロディ片を生成した。特徴量としては, 各音符に対して8分音符の長さを1とした音長, 開始時刻, 音高差を用いた。また, 従来手法として, 同様の条件で拍粒度を8分音符とした音高推移ベクトルを用いた類似音楽検索を行った。

### 4.2 実験結果と考察

図4に実験結果を示す。グラフは提案手法, 従来手法の検索精度(各順位までの適合率)を示す。実験結果より距離尺度に EMD を用いることにより, 従来手法より高い検索精度が得られることが分かる。図5に提案手法により正解率が向上した検索例(提案手法において1位に検索された例)として, 上図に1ウインドウ分の入力ハミング, 下図に対応する正解メロディ片の楽譜を示す。

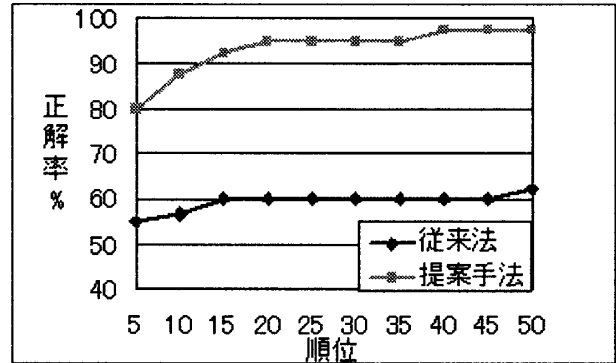


図4 検索精度の比較実験結果



図5 入力ハミングと正解データの例

ハミング入力の場合, 一定のリズムでハミングを行うことは困難であり, リズムの変動が生じる。このリズムの変動が音符の長さや開始時刻に影響を与え, 過剰な音符や音調の異なる音符を生成してしまう。この問題点に対して, 従来法では音符ベースではなく, 拍ベースで音高を表すベクトル値を作成するため, ユークリッド距離が大きくなり, 正解が検索できなかったと考えられる。一方, 提案手法では, 入力ハミングと正解音楽片との音符数は異なるが, EMD の特長の一つである輸送コストを最小とする点において, 一つの音符が複数の音符に分配されることにより効果的に距離計算されているといえる。

## 5. まとめ

本稿では, 距離尺度に EMD を用いたハミングによる類似音楽検索手法を提案した。提案手法では, 音長, 音高差, 開始時刻を特徴量として EMD で類似度計算を行う。これにより, 曲調が類似した曲が検索可能であるうえ, 音符数の異なるデータに対しても高精度の検索が可能である。今後の課題としては, 今回採用した特徴量以外の属性について, その有効性を評価したい。

## 謝辞

本研究の一部は科学研究費補助金基盤研究(B)17300036, 基盤研究(C)17500644 の補助を受けて行った。

## 参考文献

- [1] 池谷直紀, 服部正典, 大須賀昭彦: “リズム入力による音楽検索方式「タタタタップ」”, 第3回情報科学技術フォーラム, 2004
- [2] 小杉尚子, 小島明, 片岡良治, 串間和彦: “大規模音楽データベースのハミング検索システム”, 情報処理学会論文誌 Vol. 43 No.2, 2002
- [3] Y. Rubner, L. J. Guibas, and C. Tomasi. The Earth Mover's Distance, Multi-Dimensional Scaling, and Color-Based Image Retrieval. In Proc. Of the ARPA Image Understanding Workshop, pp. 661-668, May 1997
- [4] 株式会社メディア・ナビゲーション: “鼻歌ミュージシャン2”