Vol.2013-MUS-100 No.30
2013/9/2

Hough 変換を用いた楽曲構造の境界抽出

三浦 亮[1,a)]　寺島 裕貴[1,b)]　喜田 拓也[1,c)]

概要：近年，膨大な数の楽曲を個人がインターネットを介して利用できるようになった．なおも増加し続ける楽曲データを効率よく管理し検索する上で，楽曲の構造理解は重要な役割を担っている．たとえば，ポピュラー音楽には，イントロ，コーラス (サビ)，ブリッジといった音楽的に意味の有る構造が存在しており，これを自動的に抽出することで高度な楽曲鑑賞や管理を行うことができる．本稿では，楽曲信号を入力として音楽的に意味の有る構造の境界を見つける問題について議論する．楽曲構造の解析は，楽曲信号に対する自己相似性行列上に現れるブロック構造を抽出することで行われる．本稿では，自己相似性行列に各種のフィルタによる画像処理を行った後，Hough 変換で直線構造を抽出することによってブロック構造を検出する手法について議論する．RWC 音楽データベースと AIST Annotation for RWC Music Database を用いた実験の結果，Sobel，Scharr フィルタを通して構造検出する手法はガウス分布を利用した既存のフィルタリング手法よりも正解データに近いブロックの境界を検出することができた．

キーワード：楽曲構造解析，Scharr フィルタ，Hough 変換，

1. はじめに

昨今，インターネット上で利用可能なデジタル音楽データの数が増加している．たとえば，iTunes Store では 2600 万を超える数の楽曲が配信されており[*1] 利用者はいつでも自分の好きな楽曲購入し楽しむことができる．

一方，ユーザ生成コンテンツ (UGC) の成長も著しい．音楽好きなアマチュアの中には，配信されている楽曲を購入するでけでなく，自らが制作した楽曲を YouTube[*2]などのコンテンツ共有サービスサイトへ投稿し，全世界へ公開する人も増えてきた．今日の音楽制作ソフトウェアの性能向上と普及は，こうしたアマチュアの創作活動を促進している．

こうした背景の中，楽曲が持つ構造を解析することは，オンライン上に存在するデータを分類し整理する上で重要な役割を果たしている．ここで楽曲が持つ構造というのは，たとえば，イントロ，ブリッジ，コーラス (サビ) と呼ばれるようなもののことである [11].

楽曲構造が分かると，楽曲試聴を行う上で，曲の開始時点のみからの再生以外にもサビからの再生などができると，試聴におけるバリエーションが増え便利である．実際，そのようにサビからの再生ができる試聴機は既に存在する．これは，各楽曲データに楽曲構造に関するメタデータを添加することによって行うことができる．このようなメタデータはまた，大規模な音楽データを管理するシステムのために有用である．たとえば，類似楽曲のクラスタリングなどの基本情報として用いることで，より適切な楽曲推薦を行ったり高速な検索を実現したりするシステムが構築できる [5].

プロ作家による楽曲に対しては供給元の企業によってメタデータが付与されるが，UGC 楽曲の場合には作家あるいは視聴者のボランティアによって付けられるため，人気のない楽曲などではメタデータが付与されないことのほうが多い．自動的な構造解析技術は，このような状況を改善できる可能性がある．

音楽構造の解析の良く知られた方法の一つは，**自己相似性行列** (SSM) [6] を用いることである．SSM は，各要素が音楽の一部分に対応する部分シーケンス間の類似度で定義される行列である．この行列の要素を 2 次元画像の輝度に対応付けると，楽曲中に繰り返し現れる類似構造を可視化することができる（図 1）．よって，SSM 上に現れる類似の矩形領域（ブロック）を見つけることで，楽曲の繰り返し構造を知ることができる．

本稿の解析では，楽曲の微小な区間ごとにクロマベクトルを計算し，そのクロマベクトルどうしの類似度を用いて

[1] 北海道大学大学院 情報科学研究科
Sapporo, Hokkaido 060.0814, Japan
[a)] ryo_miura@ist.hokudai.ac.jp
[b)] tera@ist.hokudai.ac.jp
[c)] kida@ist.hokudai.ac.jp
[*1] http://www.apple.com/jp/itunes/
[*2] http://www.youtube.com/

© 2013 Information Processing Society of Japan

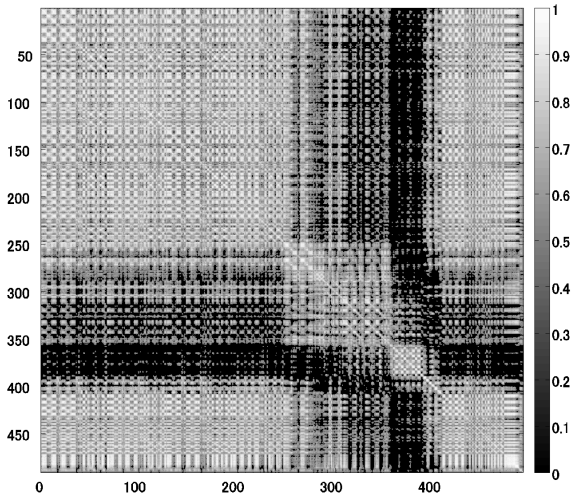

図 1 自己相似性行列: Franz Schubert, Piano Quintet in A major, Op.32, D. 667, "Die Forelle" (RWC-MDB-C-2001-C04, Tr.03)

SSM を構築する方法 [4] を取る．そのようにして求めた SSM 上のブロックの境界を見つけることで，楽曲の意味的な境界を知ることができる．

本稿では，上述の SSM に対して輪郭抽出画像フィルタリングを行い，さらに Hough 変換 [2] による画像内の直線構造を抽出することで楽曲構造の境界を求める手法の有効性について議論する．画像の輪郭部を強調する輪郭抽出フィルタリングには，Sobel フィルタや Scharr フィルタ等がある [18]．同様に Hough 変換を用いた楽曲構造解析手法には，Aucouturier と Sandler ら [1] によるものがある．この手法は SSM に存在する画像に対して斜めに走る直線構造を抽出し，解析に利用するものである．本稿で議論する手法は，SSM に対して縦方向に存在する構造を抽出するものである．

2. 準備

2.1 楽曲が持つ構造

一般的に，それぞれの楽曲は独自かつ独特の構造を有している．楽曲には**意味区間** (semantic segment) というものがあり，意味区間はある程度の長さを持つ構造である．たとえば，イントロ，ブリッジ，コーラスなどのような特定の音調や音楽的な意味を持っている部分が意味区間となる．一つの楽曲は，いくつかの意味区間が集まって成り立っていると考えることができる．一方で楽曲には小節という譜面上の別の区切り方がある．小節は特定の拍節数で定義された構造の区切り方であり，上述の意味区間は複数の小節からなることが多い．楽曲構造解析は，楽曲中の意味区間が繰りかえされる構造を発見することで行われる [10]．

後藤 [11] は，音楽音声信号を入力とした楽曲構造解析の手法を提案し，サビ区間自動抽出を行う音楽試聴インタフェースである "SmartMusicKIOSK" を示した．この研究は現在，"Songle" [9] というウェブサービスへと結実し，ウェブ上でより洗練されたインタフェースを使用することができる．Paulus と Klapuri ら [16] は，確率的適合指標を利用する別の楽曲の区間分け手法を提案した．彼らの手法では，楽曲構造解析を，クロマベクトル，メル周波数ケプストラム係数 (MFCC)，およびリズモグラムを利用し行う．上述の二つの手法はどちらも，楽曲の区間分けを行う際に SSM を用いている．

2.2 クロマベクトル

クロマベクトルは 12 次元の音声信号特徴量であり，極めて短い断片においてそれぞれの音階を周波数軸としたパワー分布から成り立っている．楽曲信号データをクロマベクトルの列に変換したものは，見方を変えると，楽曲の非可逆な圧縮を行っていると考えることができる．クロマベクトルの計算手順は，楽曲を微小な区間の列に分割し，各区間ごとにフーリエ変換を行い周波数スペクトルを計算する．計算されたスペクトルパワーを 12 半音階ごとに加算し累積し，12 次元のベクトルとする．

ある節回しが楽曲中で繰り返されるときには，メロディーラインや伴奏にいくらかの変化が伴うことがある．したがって，似た意味区間どうしであっても，まったく同一のクロマベクトル列が得られるとは限らない．しかしながら，全体的な音の響き（同時になっている音の構成）が類似しているならば，類似したクロマベクトル列になっていることが期待できる．たとえば，サビ部分の歌詞が 1 番と 2 番で異なっていても，メロディーが同様ならば，その部分のクロマベクトル列は類似していると考えられる．

2.3 自己相似性行列 (Self-similarity Matrix)

区間類似度は，系列をある長さで区切った区間どうしを比較して，それらがどれだけ類似しているかを表す尺度である．自己相似性行列 (Self-similarity Matrix; SSM) [6,12,15,17] は，系列中のすべての区間の組み合わせの区間類似度を並べた行列である．本稿での音楽構造解析においては，系列は対象となる楽曲から得られたクロマベクトル列であり，区間類似度はクロマベクトル間の類似度に対応する．

入力する楽曲の楽曲信号データを m 個の区間に分割する．以降では，それぞれの微小な区間を**フレーム**と呼ぶ．フレームの長さよりもフレームのずらし幅を小さくとると，近辺のフレームどうしは重なる部分を持つことになる．あるいはフレームごとに（たとえば拍節に合わせて）フレーム長やずらし幅を異なるようにとることもできる．

系列の i 番目のフレームのクロマベクトルをパワースペクトルについてノーマライズし，それを $\mathcal{X}_i = (\mathbf{x}_1^{(i)}, \mathbf{x}_2^{(i)}, \ldots, \mathbf{x}_{12}^{(i)})$ とすると，i 番目のフレームと j 番目のフレームの区間類似度は，

© 2013 Information Processing Society of Japan

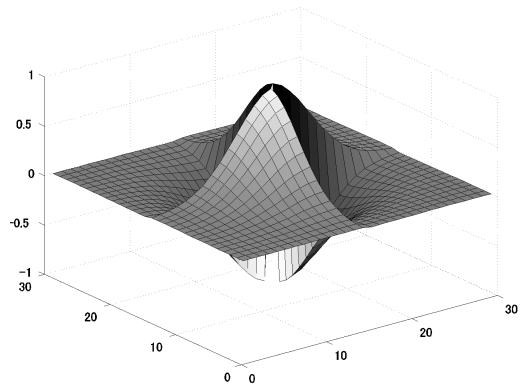

図 2 Gaussian filter used in [16]

$$s_{i,j} = \frac{\langle \mathcal{X}_i, \mathcal{X}_j \rangle}{|\mathcal{X}_i| \cdot |\mathcal{X}_j|}.$$

で求められ，0 から 1 までの値を取る．以降では，この $s_{i,j}$ をフレーム類似度と呼ぶ．

楽曲が m 個のフレームに分けられているとき，SSM は m 次正方行列状となり，SSM の i 行 j 列目の要素は i フレーム目と j フレーム目の類似度が挿入される．すなわち，SSM M は，$M(i,j) = s_{i,j}$ で定義される．SSM の各要素について，値の高低を画像の明暗で塗り分けることにより，その楽曲中で類似した部分が帯状となって現れる [12,14]．よって，楽曲の中で繰り返して使われているフレーズがどのあたりに存在し，そのフレーズがどのくらいの長さを持つのかといったことが視覚的に理解できるようになる．

SSM が得られたら，そのブロック構造の境界を列挙することで楽曲の意味区間の境界候補を見つけることができる．Paulus と Klapuri ら [16] の方法では，Foote [7] によって提案された新規性計算 (novelty calculation) を用いて境界候補が算出される．その手法で用いられる畳み込み行列（カーネル）は，ガウス分布の第 2 象限と第 3 象限の正負を反転させた図 2 のような形である．このフィルタを SSM の対角線に沿って動かしながら算出された系列は新規性ベクトル (novelty vector) と呼ばれる．この新規性ベクトル中のピークがある位置は，SSM のブロックの境界候補となっている．

3. 画像処理

3.1 輪郭抽出フィルタ

デジタル画像処理におけるフィルタリング技術のひとつに，領域境界抽出と呼ばれるものがある [18]．領域境界抽出とは，対象となる画像について，画素の明度が急激に変化する点を抽出する処理である．明度が急激に変化する箇所は，画像中に存在する構造の輪郭に対応している．したがって，領域境界抽出は画像に含まれている構造の輪郭を抽出することとも言える．

明度の変化は，微分計算を利用することで算出することができる．Scharr フィルタと呼ばれるフィルタは，領域境界抽出に使用される一次微分で輝度勾配を計算するフィルタである．画像の縦方向に潜在する輪郭を抽出するためのカーネルは以下のようになる．

$$\begin{bmatrix} -3 & 0 & 3 \\ -10 & 0 & 10 \\ -3 & 0 & 3 \end{bmatrix}.$$

入力画像のすべての画素について，このカーネルを使用して近傍領域との差分を計算すると領域境界を抽出することができる．入力として与えるデジタル画像を I，出力されるデジタル画像を R とする．このとき，出力画像のある画素 $R(x,y)$ は入力画像の画素 $I(x,y)$ から $R(x,y) = -3(I(x-1,y-1)+I(x-1,y+1)) - 10(I(x-1,y)) + 10(I(x+1,y)) + 3(I(x+1,y-1)+I(x+1,y+1))$ で求められる．このカーネルを転置して計算を行うと，画像に対して横方向の領域境界を抽出するようになる．

また，本稿における実験では，Scharr フィルタと同様に画像の輪郭抽出を行うための画像フィルタとして，Sobel フィルタと Canny フィルタを比較に用いている．Sobel フィルタは，Scharr フィルタと同様に空間 1 次微分を計算し輪郭抽出を行うフィルタであり，そのカーネルは

$$\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$

で表される．Canny フィルタは，輪郭抽出の前に画像全体にガウスフィルタをかけ画像全体を平滑化し，その後輪郭抽出を行う 1 次微分フィルタをかけたものである．

3.2 Hough 変換

Hough 変換 [2] は，デジタル画像処理における特徴抽出手法の一つである．パラメータ空間への投票と多数決に基づく特徴抽出手法であり，特に画像内に潜在する直線構造を検出し抽出する．

二次元平面上のある点 P を考える．点 P を通る平面状の直線は無数に存在する．この直線の集合の中で，画像に含まれている構造の輪郭に当たる部分を指す点を通過するような直線を探索する．

Hough 変換では，二次元平面上の 1 つの直線を，原点 O から直線に下ろした法線の長さ r とその法線の角度 θ の組で表す．任意の直線に対する (r,θ) の組み合わせは一意に定めることができることに注意する．このとき，直線が点 $\mathrm{P}(x,y)$ を通っているとすると，$r = x\cos\theta + y\sin\theta$ が成り立つ．ここで，r と θ で張られる平面を Hough 空間と呼ぶ (**図 3**)．

解析を行う画像について，その画像の特徴点と思われる点を通過しているすべての直線について，(r,θ) を計算する．２次元平面上の 1 点を通る直線の集合は，Hough 空間

© 2013 Information Processing Society of Japan

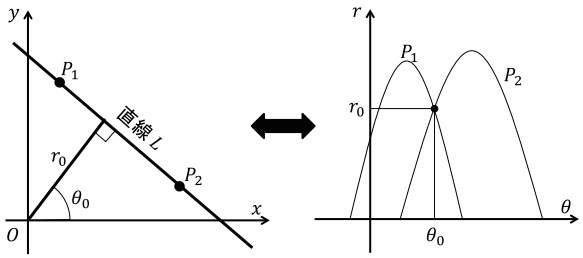

図 3 2点を通る xy 平面の直線と Hough 空間の関係.

上の正弦曲線に写像される．したがって，複数の点を通る直線は，Hough 空間上の正弦曲線の交点として現れる．厳密に Hough 空間上の交点を求めるのは計算コストが膨大となるため，Hough 空間を離散的に分割（量子化）して計算コストを削減する．Hough 空間上の量子化された小領域（ビン）に対し，正弦曲線が通る箇所に「投票」を行っていく．すべての投票を終えた結果，数多くの交差が確認されたビンを上位から並べ，それらビンに対応する２次元平面上の直線を元画像の特徴構造として出力する．

以上のようにして，元の画像に潜在する特徴的な直線構造を検出することができる．

4. 提案手法

この節では，本稿の楽曲構造解析手法について説明する．

先にも述べたが，手法の概要は以下のとおりである．最初に，構造解析を行う楽曲の楽曲信号データを入力し，クロマベクトルを計算する．次に，計算されたクロマベクトル列から SSM を計算する．さらに，計算された SSM に対して画像フィルタをかけ，その後に Hough 変換で垂直な直線構造を抽出する．

図 4 の例は，RWC-MDB-P-2001 No.9『慟哭』[*3]に対する SSM である．図の左上が楽曲データの開始地点，右下が終了地点を表し，楽曲は右下へ向かって進行する．SSM を構成する各要素はフレーム間の音声の類似度を表し，類似度が高い箇所ほど明るい色で表示されている．類似度が高い箇所がある程度の長さを持って現れるような楽曲では，SSM を計算した際に明度の高いブロックのような正方形構造が現れる．これが楽曲構造である．楽曲におけるサビなどの区間では，その区間内で鳴っている音は似ていることが多いため，SSM で明度の高いブロック構造が現れてきた箇所を楽曲の意味区間として考えることができる．

図 4 の SSM に対して Scharr フィルタを施したものが図 5 である．縦の輪郭を抽出するような Scharr フィルタで処理を行うことにより，SSM におけるブロック構造の境界部分が縦の線となって抽出される．

この画像にさらに Hough 変換を通すと図 6 のようになる．本稿の実験では，Hough 変換については MATLAB の Hough 変換の専用関数を用いた．この関数を実行した後，

[*3] 4 分 37 秒，全 559 フレーム

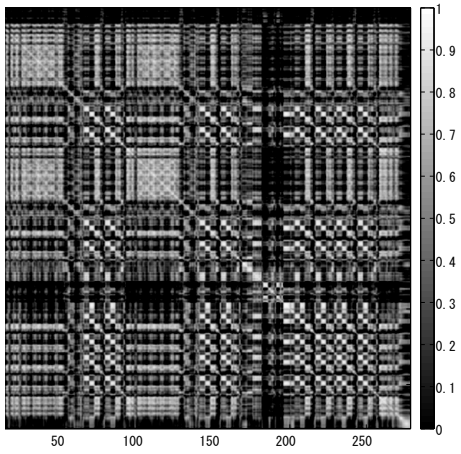

図 4 自己相似性行列 (SSM) の例.

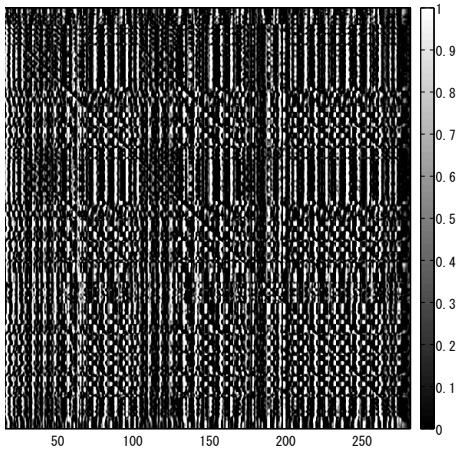

図 5 図 4 に対して Scharr フィルタによる処理を行ったもの.

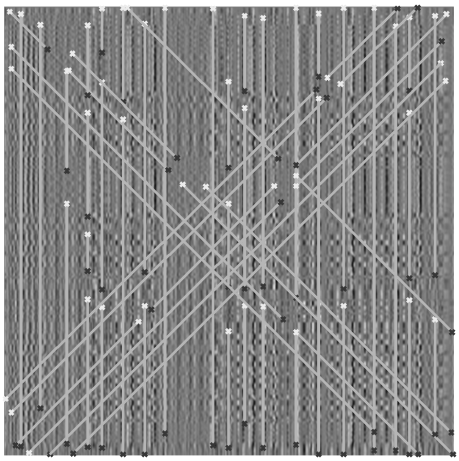

図 6 図 5 に対して Hough 変換を行ったもの.

画像に対して正しく垂直になっている線分のみを選び出し，その座標を記録した．直線の x 座標の値は楽曲の再生時間に対応しており，これをまとめた結果が楽曲の意味区間の境界候補となる．

© 2013 Information Processing Society of Japan

また本稿の実験では，楽曲中の転調を考慮した SSM も用いた．転調を考慮した SSM は，類似度を計算する際に，片方のフレームのクロマベクトルを 1 要素ずつ回転させながら 12 回分を計算し，最も類似度の高い値をそのフレーム間の類似度とする．

5. 実験

提案手法の楽曲構造の境界判定能力の比較実験を行った．実験には，RWC 研究用音楽データベースのポピュラー音楽全 100 曲[*4]を使用した．また楽曲構造の「正解」のデータには，RWC 音楽データベースのための AIST Annotation データから，ポピュラー音楽データベースの Chorus Section アノテーションデータ [8] を使用した．このデータは，RWC 研究用音楽データベースに収録されている楽曲について，楽曲のサビ区間や A メロ，あるいは前奏，間奏などの区間が，それぞれどこからどこまでであるのかということを記録した楽曲構造解析のための補足資料である．

比較実験の内容は，楽曲構造解析の結果得られた意味区間の境界と，AIST Annotation データで記されている正解とする意味区間の境界との一致度を調べるものである．クロマベクトルの計算には，MATLAB の関数集合である “Chroma toolbox” [13] を用いた．Chroma toolbox を用いると，クロマベクトルを求める際に正規化を行いノイズのような成分を除去することができる．今回の実験では，楽曲を解析するフレーム長は 0.5 秒とし，フレーム間の重複はないものを用いた．

区間境界の一致度を測るため，楽曲全体のフレームを境界によって先頭から順にクラスタリングすることを考える．たとえば，楽曲が 8 つのフレームに分割されているとする．これを解析した結果として 3 番目と 5 番目に意味区間の境界があるとしよう．このとき，各フレームに対して先頭から順番に番号付けを行い，$\begin{pmatrix}1 & 1 & 2 & 2 & 3 & 3 & 3 & 3\end{pmatrix}$ とラベリングすることで 3 つのクラスタができる．このラベル付けによって，境界で区切られた区間がフレーム列上でどの範囲にあるかが分かる．

次に，楽曲の区間分けの精度の評価の仕方について述べる．精度の評価指標として，調整ランド指数 (adjusted rand index) [3, 19] を用いた．調整ランド指数は，2 つのクラスタリングの類似度を見るために用いられる指標である．指数の値は −1 から 1 の間をとり，クラスタのラベリングが完全に一致していれば 1 となる．算出された境界候補によるフレームのクラスタリングが，正解データによるフレームのクラスタリングとどの程度近いのかを見ることで，境界候補の正しさを知ることができる．より理解しやすい F 尺度などの代わりに，このような評価を用いた理由は，出力された境界位置と正解データの境界位置との誤差がいくらまでなら正解とするかを考慮しなくて済むためである．多少のずれがあっても，そのずれが小さいうちは調整ランド指数への影響は小さい．また，F 尺度と同様に，クラスタの分割数が正解データと極端に異なる場合には調整ランド指数の値は低くなる．

いま，n 個の要素からなる集合 S と，2 つの比較したいクラスタ $X = X_1, X_2, \cdots, X_r$, $Y = Y_1, Y_2, ..., Y_s$ を考える．これらの要素 X_i, X_j について，$n_{ij} = |X_i \cap Y_j|$ をすべての組み合わせについて計算し，この値について行，列ごとに総和を取る．ここで，行ごとの総和を $a_1, a_2, ..., a_r$, 列ごとの総和を $b_1, b_2, ..., b_s$ とすると，調整ランド指数は，

$$ARI(X,Y) = \frac{\sum_{ij}\binom{n_{ij}}{2} - [\sum_i \binom{a_i}{2}\sum_j \binom{b_j}{2}]/\binom{n}{2}}{\frac{1}{2}[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2}] - [\sum_i \binom{a_i}{2}\sum_j \binom{b_j}{2}]/\binom{n}{2}}$$

と定義される．

上記の方法で，RWC 研究用音楽データベースのポップス楽曲全 100 曲を用いて実験を行った．本稿では，それぞれ以下の処理を行った上で調整ランド指数を計算した．

(1) SSM に Hough 変換を施して ARI 算出 (Normal)
(2) 転調を考慮した SSM に Hough 変換を施して ARI 算出 (Nomal-Modu)
(3) SSM に Canny フィルタをかけ Hough 変換を施して ARI 算出 (Canny)
(4) 転調を考慮した SSM に Canny フィルタをかけ Hough 変換を施して ARI 算出 (Canny-Modu)
(5) SSM に Scharr フィルタをかけ Hough 変換を施して ARI 算出 (Scharr)
(6) 転調を考慮した SSM に Scharr フィルタをかけ Hough 変換を施して ARI 算出 (Scharr-Modu)
(7) SSM に Sobel フィルタをかけ Hough 変換を施して ARI 算出 (Sobel)
(8) 転調を考慮した SSM に Sobel フィルタをかけ Hough 変換を施して ARI 算出 (Sobel-Modu)
(9) [16] で用いられているガウシアンフィルタを応用したフィルタリング手法の一部 (既存手法) から ARI 算出 (Gaussian)

計算結果を表 1 に示す．比較の結果，Hough 変換を用いた解析手法は概ね Gaussian と比較して良好な値を示していることがわかる．また転調を考慮した SSM を用いた方が転調を考慮しない SSM より良い結果が得られた．

6. まとめと今後の課題

本稿では楽曲構造の解析を行うために，クロマベクトルの類似度から計算された自己相似性行列と画像の輪郭抽出処理の手法とを組み合わせた手法について議論した．調整ランド指数というクラスタリング結果比較を行う指標を用いて解析結果を比較したところ，[16] で用いられているガ

[*4] RWC-MDB-P-2001 No.01-100

© 2013 Information Processing Society of Japan

表 1 各手法によるクラスタリングと正解データによるクラスタリングとの間の ARI 値.

手法	平均	分散
(1)Normal	0.50306	0.00708
(2)Normal-Modu	0.53041	0.00597
(3)Canny	0.44488	0.01825
(4)Canny-Modu	0.46331	0.01177
(5)Scharr	0.56217	0.00226
(6)Scharr-Modu	**0.57777**	0.00365
(7)Sobel	0.56436	0.00292
(8)Sobel-Modu	0.57643	0.00404
(9)Gaussian	0.47808	0.00891

ウシアンフィルタよりも良好な数値が得られた.

本稿における実験では，サイズの大きなノイズを特徴的な構造と判断し，その結果，正解データと全く合致しないところに特徴的構造を発見してしまう場合があった．また，今回の画像フィルタリングによる処理では，ブロック構造の境界部分だけを抽出できているわけではなく，多数の無駄な境界が出力される場合もあった.

楽曲中の意味区間の境界を抽出する既存手法は数多くあり，それらとの網羅的な比較検討を通じて，よりよい楽曲の構造解析が今後の課題である.

参考文献

[1] Aucouturier, J.-J. and Sandler, M.: Finding Repeating Patterns in Acoustic Musical Signals: Applications for Audio Thumbnailing, *AES 22th International Conference on Virtual, Synthetic and Entertainment Audio* (2002).

[2] Ballard, D. H.: Generalizing the Hough transform to detect arbitrary shapes, *Pattern Recognition*, Vol. 13, No. 2, pp. 111 – 122 (1981).

[3] Barrington, L., Chan, A. B. and Lanckriet, G.: Modeling Music as a Dynamic Texture, *IEEE Transactions on Audio. Speech, and Language Processing* (2010).

[4] Bartsch, M. and Wakefield, G.: To catch a chorus: using chroma-based representations for audio thumbnailing, *2001 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 15–18 (2001).

[5] Bartsch, M. A. and Wakefield, G. H.: Audio thumbnailing of popular music using chroma-based representations, *Multimedia, IEEE Transactions on*, Vol. 7, No. 1, pp. 96–104 (2005).

[6] Cooper, M. and Foote, J.: Summarizing popular music via structural similarity analysis, *Applications of Signal Processing to Audio and Acoustics, 2003 IEEE Workshop on.*, pp. 127–130 (2003).

[7] Foote, J.: Automatic audio segmentation using a measure of audio novelty, *IEEE International Conference on Multimedia and Expo 2000 (ICME 2000)*, Vol. 1, pp. 452–455 vol.1 (2000).

[8] Goto, M.: AIST Annotation for the RWC Music Database, *the 7th International Conference on Music Information Retrieval*, pp. 359–360 (2006).

[9] Goto, M., Ogata, J., Yoshii, K., Fujihara, H., Mauch, M. and Nakano, T.: PodCastle and Songle: Crowdsourcing-Based Web Services for Retrieval and Browsing of Speech and Music Content, *CrowdSearch*, pp. 36–41 (2012).

[10] Lu, L., Wang, M. and Zhang, H. J.: Repeating Pattern Discovery and Structure Analysis from Acoustic Music Data, *ACM SIGMM international workshop on Multimedia information retrieval* (2004).

[11] Masataka, G.: A Chorus-Section Detecting Method for Musical Audio Signals, *Proceedings of the 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. V–437–440 (2003).

[12] Müller, M. and Clausen, M.: Transposition-Invariant Self-Similarity Matrices, *Proceedings of the 8th International Conference on Music Information Retrieval*, Vienna, Austria, pp. 47–50 (2007).

[13] Muller, M. and Ewert, S.: Chroma Toolbox: Matlab Implementations for Extracting Variants of Chroma-Based Audio Features, *Proceedings of the 12th International Society for Music Information Retrieval Conference*, pp. 215–220 (2011).

[14] Müller, M., Grosche, P. and Jiang, N.: A Segment-Based Fitness Measure for Capturing Repetitive Structures of Music Recordings, *Proceedings of 12th International Society for Music Information Retrieval Conference (ISMIR 2011)*, Miami, USA, ISMIR, pp. 615–620 (2011).

[15] Müller, M. and Kurth, F.: Towards structural analysis of audio recordings in the presence of musical variations, *EURASIP J. Appl. Signal Process.*, Vol. 2007, No. 1, pp. 163–163 (2007).

[16] Paulus, J. and Klapuri, A.: Music structure analysis using a probabilistic fitness measure and a greedy search algorithm, *IEEE Transactions on Audio, Speech, and Laguage Processing*, Vol. 17, No. 6, pp. 1159–1170 (2009).

[17] Paulus, J., Müller, M. and Klapuri, A.: State of the Art Report: Audio-Based Music Structure Analysis, *International Conference on Music Information Retrieval 2010* (2010).

[18] Szeliski, R.: *Computer Vision: Algorithms and Applications*, Springer-Verlag New York, Inc., New York, NY, USA, 1st edition (2010).

[19] Vinh, N. X., Epps, J. and Bailey, J.: Information theoretic measures for clusterings comparison: is a correction for chance necessary?, *Proceedings of the 26th Annual International Conference on Machine Learning*, ICML '09, New York, NY, USA, ACM, pp. 1073–1080 (2009).

© 2013 Information Processing Society of Japan