楽曲マッチングにおけるテンポに不変な特徴量抽出法

EP05037 **加藤 光**　　**指導教授：阪田泉， 藤吉弘亘**

1. はじめに

近年，マルチメディア情報の流通が盛んになっており，大量データから特定のマルチメディア情報を検索・探索する技術が求められている．その中でも楽曲は，テンポなどを容易に変化させることが出来るため，同一のものを探索するのが困難である．そこで，本研究では画像から特徴点検出と特徴量記述を行うアルゴリズム”Scale-Invariant Feature Transform(SIFT)”[1] を応用し，楽曲のテンポの変化に不変な特徴量抽出法を提案し，その有効性を示す．

2. SIFT による特徴点の検出と特徴量の記述

SIFT は画像の勾配情報に基づく局所的なキーポイントを検出 · 記述する手法である．検出したキーポイントに対して，画像の回転 · スケール変化 · 照明変化に頑健な特徴量を記述するため，画像のマッチングや物体認識 · 検出に用いられている．以下に SIFT のアルゴリズムを示す．

Step1 スケールと特徴点 (キーポイント) の検出
Step2 キーポイントのローカライズ
Step3 オリエンテーション算出
Step4 特徴量の記述

Step1 では Difference-of-Gaussian(DoG) 処理により極値探索を行い，キーポイントの最適なスケールを検出する．Step2 では検出した特徴点から，特徴点として向かない点を削除する．Step3 では画像の勾配方向と大きさから重み付き方向ヒストグラムを作成する．Step4 ではオリエンテーション算出によって求めた勾配情報とスケールに基づいて特徴量を記述する．

3. 楽曲における SIFT 特徴量の抽出

本研究では，楽曲に対して SIFT を適用する．まず，DoG 処理を行うために，1 次元ガウス関数 $G(t,\sigma)$ を式 (1) に示す．

$$G(t,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{t^2}{2\sigma^2}\right) \tag{1}$$

ここで，t は窓の大きさ，σ^2 は分散を表す．式 (1) の 1 次元ガウス関数を楽曲に適用し，DoG 処理による極値探索を行う (図 1)． DoG 処理は，入力データに対して段階的

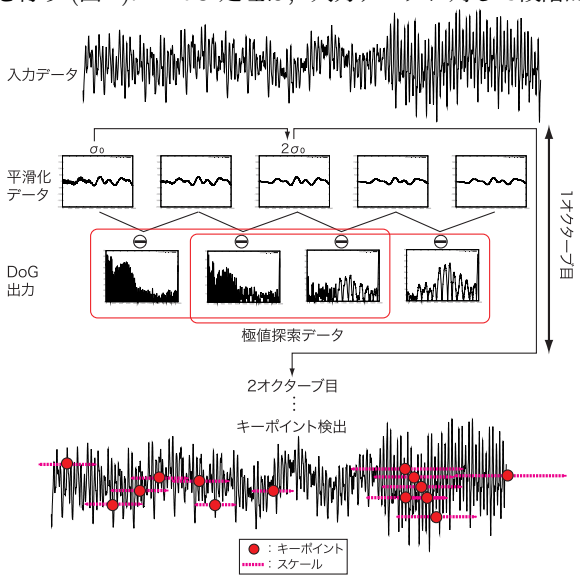

図 1：1 次元の DoG 処理による極値探索

に平滑化を施し，2 組の平滑化データの差分を行う．差分から得た DoG 出力を 3 つ 1 組とし，中心データ (極値探索データ) における注目点の DoG 出力値が，上下スケールを含む 8 近傍の DoG 出力値と比較して最大となる場合，その注目点をキーポイントとして検出する．次に，検出したキーポイントを中心としたスケールの範囲内から，勾配方向ヒストグラムを作成する．まず，スケール領域内の一辺を 4 ブロックに分割し，各ブロック毎に 19 方向 (-90 °～90 °) の勾配方向ヒストグラムを作成する．従って，4 ブロック× 19 方向=76 次元の特徴量が記述できる（図 2）．以上の処理を原楽曲と変更後の楽曲に対して行い， 2 つの特徴量を抽出する．

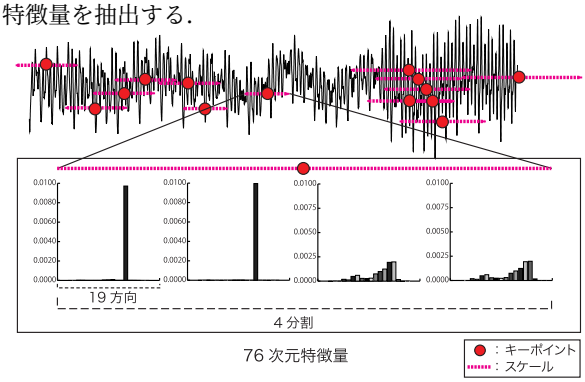

図 2：ブロック毎の特徴量記述

4. キーポイントのマッチング

次に，原楽曲と変更後の楽曲において対応をとることができるか調べるため，キーポイントのマッチング処理を行う．第 2 章において得られた原楽曲と変更後の楽曲の特徴量をそれぞれ $S^{R1} = (s_1^{R1}, s_2^{R1}, ..., s_{76}^{R1})^T$， $S^{R2} = (s_1^{R2}, s_2^{R2}, ..., s_{76}^{R2})^T$ としたとき，キーポイント間の距離を式 (2) により求める．このとき，最小となる距離を d_1，2 番目に小さい距離を d_2 とする場合，式 (3) の判別式を用いて，信頼性の高いキーポイントを対応点とする．

$$d(S^{R1}, S^{R2}) = \sum_{i=1}^{76} \sqrt{(s_i^{R1} - s_i^{R2})^2} \tag{2}$$

$$10 \times d_1 < m \times d_2 \quad (0 < m \leq 10) \tag{3}$$

5. マッチング実験

実験では，15 秒程度の楽曲データ X_1 と，テンポを 2 倍にした楽曲データ X_2 の特徴量を抽出し，マッチングを行う．結果を図 3 に示す．ただし，図 3 の波形は全体の波形ではなく，一部分の波形を抜粋している．赤色の点がマッチングしたキーポイント，紫色の線が特徴量を抽出する範囲 (スケール)，緑色の線は特徴量が似ているキーポイント同士を繋いだ線である．

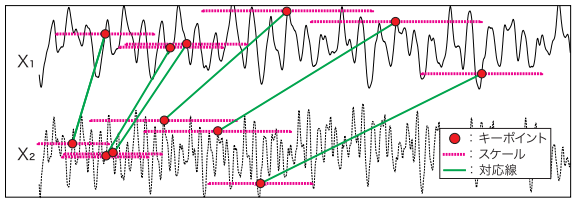

図 3：テンポの異なる楽曲のマッチング例

X_1 のキーポイントと，X_1 を 2 分の 1 にダウンサンプリングしたデータ X_2 のキーポイントの位置は，1 : 2 の関係を持っている．従って，原楽曲 X_1 と原楽曲のテンポを 2 倍にした X_2 の位置を比較すると，1 : 2 の関係になっており，正しくマッチングできていることがわかる．これは，原楽曲に対して，変更後の楽曲のスケールが小さいため，異なるテンポの楽曲でも，類似している特徴量を抽出することができたからである．

6. おわりに

本研究では，テンポに不変な特徴量抽出法を提案し，マッチングが可能であることを確認した．今後は，マッチング精度の向上を行い，楽曲の検出について検討していく予定である．

参考文献

[1] D. G. Lowe, ”Distinctive Image Features from Scale-Invariant Keypoints”, International Journal of Computer Vision, 60(2), pp. 91-110 (2004).