4T-07

画像処理と音声信号を用いたギターの自動採譜システムの開発

小峰 拓也†　　田村 仁†

日本工業大学 工学部 創造システム工学科†

1. 研究背景

ポピュラー音楽ではギターは欠かせない楽器であり，演奏者の人口も多い．しかし楽譜の作成は不慣れな者にとっては多大な労力と時間がかかるため，演奏はできるが読譜や採譜に苦手意識がある演奏者は多く作曲等において大きな妨げとなる．そこで本研究では，ギターの演奏を自動で採譜するシステムを考案する．

MIDI ギターを使用した自動採譜[1]や音声ファイルからの自動採譜に関する研究[2]は今までも行われてきたが，これらは運指情報を取得することができない．ギターは複数のポジションで同じ音程の音を出すことができ，同じフレーズでも音色や弾きやすさなどによって様々なパターンがあるが，これらの手法ではそれを記録することができない．触弦認識システム[3]があるがギターに専用装置の装着が必要となる．そこで本研究では音声と画像を利用することで音高と運指を記録するシステムを構築する．

2. 目的

本研究では画像と音声信号を用いてエレクトリックギターを対象とした単音演奏の自動採譜システムの構築を行う．システムの概要図を図 1 に示す．本研究では簡便に使用できるシステムを目標とするため特別な機器を必要とせず，比較的容易に入手できる一般的な web カメラとオーディオインターフェースを使用し，画像と音声信号を取得する．音声信号と合わせて画像を用いることによって音声信号だけでは得ることができない指のポジションを取得し，演奏者の意図通りの TAB 譜を作成することを最終目的とし，今回は音高解析，ギターの検出，手の検出を行った．

3. システム概要

本研究は主にギターから出力された音声信号の音高解析，画像でのギターと手の検出の 2 つで構成される．システム概要図を図 1 に示す．

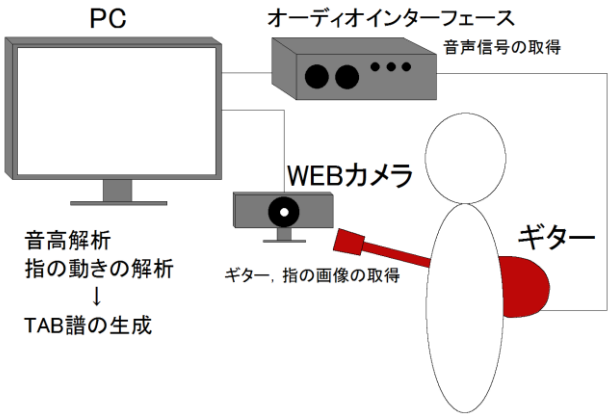

図 1 システム概要図

3. 1 音高解析

高速フーリエ変換を用いた音高解析プログラムを作成し単一音のそれぞれ 1 秒間の音高解析の認識率を測定した．A ならば 1 番低い音から A1，1 オクターブ上がるごとに A2, A3, A4 と表し，今回の実験では E1 から G4 までの音を測定した．正しい認識率が 95%以下だったものを表 1 に示す．正しく認識されなかった 9 音のうち 7 音がオクターブ下の音と誤認識されていたが，一般的なギターの音域は 4 オクターブ程度であり，演奏で想定されるポジションは概ね 12 箇所に限られることから，画像を用いて指のポジションを取得することにより補完することができると考える．

表 1 誤認識された音

入力した音名	認識率	誤認識した音名	理由
F1	30%	C2	+4 音
G1	43%	G#1	+半音
A2	0%	A1	−7 音
C3	47%	C2	−7 音
G3	67%	G2	−7 音
B4	2%	B3	−7 音
C4	0%	C3	−7 音
D4	0%	D3	−7 音
E4	0%	E3	−7 音

3. 2 ギターの検出

ギターのネックの位置を検出する．まず AKAZE

Development of automatic guitar transcription system using image processing and audio signal
†Takuya Komine, Hitoshi Tamura, Nippon Institute of Technology faculty of innovative system engineering.

Copyright ©2019 Information Processing Society of Japan.
All Rights Reserved.

特徴量[4]を用いて特徴点の抽出を行う．抽出した特徴点はギター以外の特徴点も含んでいるため，ネックの特徴色検出を行い二値のマスク画像を作成し，特徴点画像にマスク処理を行うことでギターの特徴点を抽出する．出力された画像を camshift[5]を用いて検出，追跡する．処理の流れを図 2 に示す．

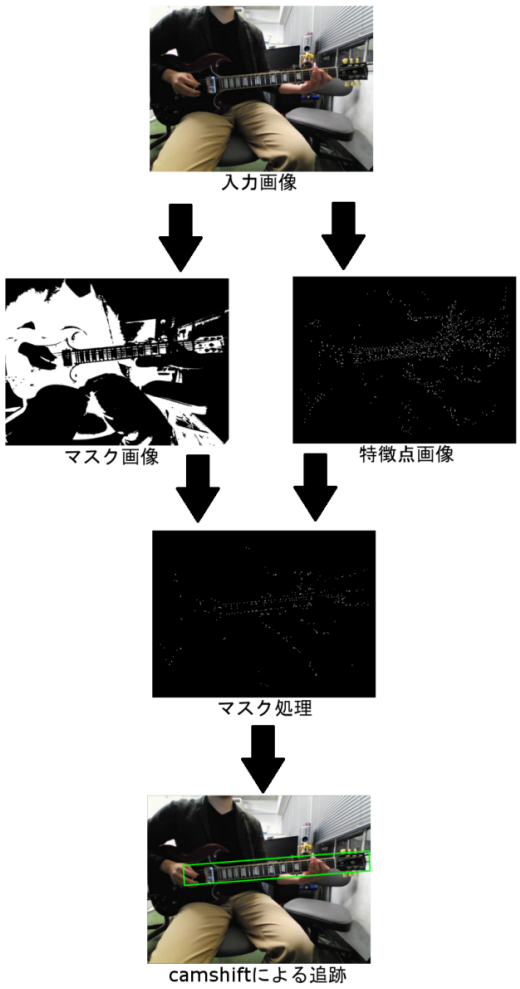

図 2 ギター検出の流れ

1 フレームごとにギターのネックを正しく検出しているフレームを測定し，検出率の測定実験を行った．総フレーム 1184 枚のうち，正しく検出できたフレームは 1170 枚，正しい検出率は 98.81756%であった．正誤を問わず検出していないフレームはなかった．

3.3 手の検出

手の検出には PoseNet を用いる．PoseNet とは，Google が公開している機械学習モデルであり人間の姿勢を推定し各部位の座標を取得することができ，手の位置を取得することによりおおまかな押弦の位置を予想することができる．PoseNet の動作例を図 3 に示す．

ギターを持って静止状態でのローポジション

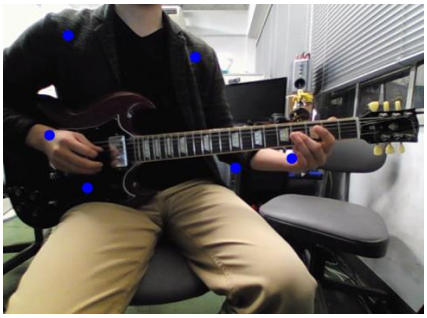

図 3 PoseNet の動作例

とハイポジションでそれぞれクラシックフォームとシェイクハンドフォームでの左手の座標の振れ幅を測定した．クラシックフォームは親指をネックの裏にそえて他の指がフレットと平行になるフォーム，シェイクハンドフォームはネックを握り込むようなフォームである．実験結果を表 2 に示す．

今回実験で使用したギターは 628mm スケールである．実験で使用した画像の最も間隔の狭いフレットは x 方向に約 7px であり，特にハイポジションでの座標の振れ幅が上回っているが，音声解析の結果を用いることにより補完することが可能であると考える．

表 2 左手の x 座標の振れ幅

	最大振れ幅[px]
ローポジション・クラシックフォーム	6.2
ローポジション・シェイクハンドフォーム	8.0
ハイポジション・クラシックフォーム	16.1
ハイポジション・シェイクハンドフォーム	12.6

4. まとめ

本研究では音声と画像の 2 つの情報を用いて自動採譜システムの開発を試みた．採譜をするための音声の音高解析，ギターのネックの位置の検出，手の位置の検出をすることができ，採譜をするための情報を取得することができた．しかし押さえている指板の位置やテンポの情報を取得できていないため，今後はこれらの情報を取得するプログラムを作成し，この情報から TAB 譜を作成するシステムを構築する必要がある．

参考文献

[1]菊地 淑晃ら（1996）“ベースギターの自動採譜システム” 情報処理学会 第 52 回全国大会講演論文集 52 巻 1 号 pp.1.459-1.460
[2]大塚匡紀，北原 鉄朗（2015）“MIDI ギターと NMF を用いた音響信号処理によるギター演奏の自動採譜”情報処理学会 第 77 回全国大会講演論文集 2015 巻 1 号 pp.419-420
[3]飛世速光ら（2014）“ギターのための触弦認識システムの構築”日本ソフトウェア科学会 31 巻 2 号 pp.2_57-2_66
[4] Pablo Alcantarilla（2013）“Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Spaces” BMVA Press Proceedings of the British Machine Vision Conference
[5]G. R. Bradski（1998）. “Computer Vision Face Tracking For Use in a Perceptual User Interface”, Intel Technology Journal, No.Q2, p15,

Copyright ©2019 Information Processing Society of Japan.
All Rights Reserved.