

画像認識 AI を用いたエレキベース演奏映像における運指推定

Estimation of Electric Bass Fingering with Image Recognition AI

宇治田 良輔†
Ryosuke Ujita

黒木 修隆†
Nobutaka Kuroki

沼 昌宏†
Masahiro Numa

1. はじめに

エレキベースの演奏動作のうち、左手で行う運指動作は同一の旋律に対して複数のパターンを考えることができ、スムーズな演奏を行うためには、適切な運指を選択する必要がある。しかし、運指が指定されていない楽譜は多く、初心者がこれを自力で決定することは難しい。

既存の運指推定手法 [1][2] では、楽器を固定したり、マーカやセンサ類を用いたりする必要があり、web 上にアップロードされた動画等、一般的な演奏動画に対する汎用性が低いという課題があった。そこで、エレキベースの演奏動画から自動で運指を推定し、提示できれば、初心者が演奏動画を通じて運指動作を容易に学習できるようになる。

2. 提案手法

2.1 提案手法の概要

提案手法では、エレキベース各部の検出に物体検出 AI である YOLOv9 を、左手指先の検出に手の姿勢推定 AI である MediaPipe Hands を用い、以下の手順で運指推定の処理を行う。提案手法の概要を図 1 に示す。

- (1) YOLOv9 を用いて、エレキベース各部(ブリッジ, 12 番フレット, ナット)の物体検出を行う。
- (2) 検出結果をもとに、射影変換によるエレキベース本体の遠近補正を行う。
- (3) 射影変換後のブリッジ, 12 番フレット, ナットの座標から、各フレットの位置を算出する。
- (4) MediaPipe Hands を用いて、左手指先の検出を行う。
- (5) 各フレットと指先の位置、動画内で演奏される楽譜の MIDI データを合わせて、運指推定を行う。

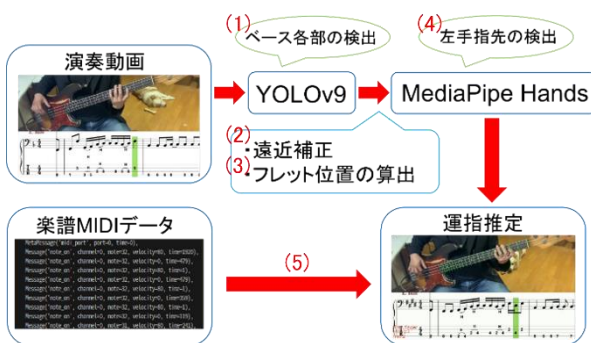


図 1 提案手法の概要

2.2 エレキベース本体の自動追跡と遠近補正

本手法では、12 番フレットが必ず弦長の中点に位置することを利用して、エレキベース本体の自動追跡と遠近補正を以下の手順で行う。

- (1) 図 2 のように YOLO が検出したブリッジ B から 12 番フレット M までのベクトル $V_{BM} = M - B$ と、12 番フレット M からナット N までのベクトル $V_{MN} = N - M$ を求める。
- (2) 12 番フレット M から画面の上下方向に垂直に伸ばした直線と、画面上下の両端との交点を A, C とする。
- (3) ベクトル V_{BM}, V_{MN} の長さを比較して、 $|V_{BM}| > |V_{MN}|$ なら B 点が $B \rightarrow B' = 2M - N$ に移動するよう、射影変換 $[A B' C N] = P \cdot [A B C N]$ を行う。
- (4) $|V_{BM}| < |V_{MN}|$ なら N 点が $N \rightarrow N' = 2M - B$ に移動するよう、射影変換 $[A B C N'] = P \cdot [A B C N]$ を行う。

これにより、指板が常にカメラに対して正対するよう補正されるため、動画内のエレキベース本体の動きに影響されることなく、各フレットの位置を求めることができる。

2.3 フレット位置の算出

ブリッジから各フレットまでの距離は、弦長を特定の比率で分割していくことで決定される。図 3 において、ブリッジ B から n 番フレットの中心 F_n までの距離は弦長の 0.9438^n 倍となる。よって、 n 番フレットの中心 F_n は、

$$F_n = B + 0.9438^n \times (N - B) \quad (2)$$

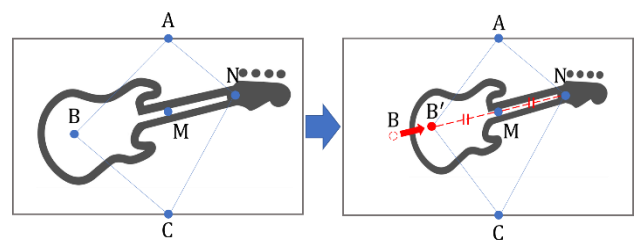


図 2 エレキベースの特徴点を利用した自動追跡と遠近補正の原理

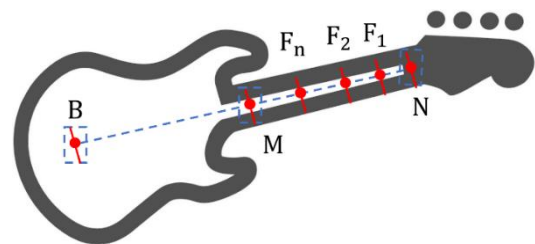


図 3 ブリッジ、ナットおよび各フレットの位置関係

† 神戸大学大学院工学研究科

Graduate School of Engineering, Kobe University

表 1 提案手法による運指推定の精度

演奏動画	演奏音	正解数	推定精度[%]
1	56	48	85.7
2	49	39	79.6
3	48	38	79.2
4	50	42	84.0
5	59	53	89.8
6	40	31	77.5
7	54	53	98.1
8	50	43	86.0
9	57	55	96.5
10	40	28	70.0
平均	50.3	43.0	84.6

と求めることができる。弦長方向のベクトル V_{NM} に垂直な単位ベクトルを $I = \frac{V_{NM}^\perp}{|V_{NM}^\perp|}$, 指板の幅を w とするとき, n 番フレット両端の座標 L_n, H_n は,

$$L_n = F_n + I \times \frac{w}{2} \quad (3)$$

$$H_n = F_n - I \times \frac{w}{2} \quad (4)$$

と求めることができる。線分 $L_n H_n$ は指とフレット間の距離の算出に利用する。

2.4 運指推定

運指推定は以下の手順で行う。

- (1) MediaPipe Hands を用いて, 左手の親指を除く 4 指先端の座標 T_m ($m = 1, 2, 3, 4$) を検出する。
- (2) T_m とそれに最も近いフレット F_n の組み合わせ (T_m, F_n) を運指候補とする。
- (3) 演奏音をもとに正しい組み合わせを選択する。

3. 評価実験

3.1 実験内容

10 本の演奏動画に対して運指推定を行い, その正誤判定を行った。YOLOv9 の学習には, 別の 10 本の学習用動画から取り出した 700 枚の画像を使用した。評価用動画には, 演奏者, エレキベースの機種, 背景がすべて異なるものを使用した。運指推定に用いる MIDI データは, 動画内で演奏と同期して表示される楽譜をもとに作成した。

正誤判定では, 1 つの演奏音に対する推定結果(押さえている弦, フレット, 運指に用いる指)が, すべて合致している場合に正, いずれか 1 つでも異なる場合は誤とした。評価項目は運指推定の精度とする。推定精度は,

$$\text{推定精度} = \frac{\text{推定結果が正であった音の数(正解数)}}{\text{動画内で演奏された音の数(演奏音)}} \times 100 \quad (4)$$

で求めるものとする。

3.2 実験結果と考察

運指推定の精度を表 1 に示す。また, 推定結果の一例を図 4 に示す。運指推定の精度は, 平均 84.6% であった。提案手法によって, 演奏者, エレキベースの機種, 背景に影響されず, 概ね正確に運指推定を行うことができたと考える。

推定結果が誤となった演奏音には, 隣の指との距離が近い, フレットから最も近い指以外で弦を押さえている, 開放弦を弾いている, などの特徴がみられた。

動画ごとに運指推定の精度を比較すると, 低いものは 70.0%, 高いものは 98.1% とばらつきがあった。特に, 左手が複雑な姿勢で映っている動画 6 や, エレキベース本体が画面の奥方向に大きく傾いている動画 10 において, フレットや指先の検出に誤差が生じ, 精度が低下したと考えられる。

4. まとめ

本研究では, 一般的な演奏動画にも適用可能な汎用性の高いエレキベースの運指推定手法の構築を目的として, 画像認識 AI を活用して, 演奏映像から自動で運指推定を行う手法の提案を行った。

提案手法では, YOLOv9 によるエレキベース各部の物体検出と, MediaPipe Hands による手の姿勢推定を活用して, 演奏動画から自動で運指推定を行った。評価実験の結果, 提案手法は, 演奏者, エレキベースの機種, 背景の影響を抑えつつ, 平均 84.6% の精度で運指推定を行えることを確認した。

今後は, 右手のピッキング動作の推定や, 演奏練度の定量評価に取り組む予定である。

5. 参考文献

- [1] 樋口拓也ら, “画像処理を用いたギター演奏支援プログラムの開発”, 情報処理学会第 79 回全国大会講演論文集, pp.107-108, 2017 年。
- [2] 坂牛和里ら, “Web カメラを利用したギターの練習支援システムの検討”, 映像情報メディア学会技術報告, 2016 年。



図 4 運指推定の結果の一例