

Kinect を利用したギターの演奏動作認識システムの開発

柏木 喜貴^{*1}, 越智 洋司^{*2}

^{*1} 近畿大学大学院総合理工学研究科 ^{*2} 近畿大学理工学部

Development of Guitar Performance Recognition System Using Kinect Sensor

Yoshitaka Kashiwagi^{*1}, Youji Ochi^{*1}

^{*1} Graduate School of Science and Engineering, Kindai University

^{*2} Faculty of Science and Engineering, Kindai University

In guitar playing performance, there are fingering of the left hand and picking of the right hand. With the fingers of the left hand, the minimum movement and the selecting finger that holds the string are important. In picking with the right hand, the size of the stroke, the timing to play the strings, and the picking direction are important. In this way, correct guitar performance requires different skills of the left hand and the right hand, and it is possible to perform the correct performance by synchronizing the fingering of the left hand with the picking of the right hand. Therefore, we think that it is important to grasp how to play guitar objectively. In this paper, we describe a motion detection method for guitar players using Kinect Sensor and GK - 3 from Roland in combination.

キーワード: ギター演奏学習支援, 動作検出, Kinect

1. はじめに

近年, 教育システムの研究ではモーションセンサを使用した身体スキル学習支援システムの研究が注目されている. その理由として, モーションセンサを用いることで人体の運動動作を検出し定量的に評価することが出来る. つまり, 自身の運動動作の乱れや熟練者との差異を客観的に把握することで, 誤りを矯正し正しい運動動作を身に付けることで学習支援に繋がるためである. そして, モーションセンサを用いた学習支援システムの対象の一つとして, ギター演奏がある.

ギター演奏の練習では, 音楽教室に通い熟練者のレッスンを受ける方法(指南練習)と, 自宅で教則本やDVDなどを参考にして一人で練習する方法(自主練習)の2種類がある. 指南練習では, 定性的ではあるが熟練者から自身の演奏動作の改修や上達に必要なポイントを演奏中や演奏後に客観的に教わることが出来る. また, よく分からないことや不明な点についても質問

することで解決することも出来る. これにより, 正しいフォームや指の押さえ方などを学びことで, 演奏の上達に繋がる. しかしながら, レッスン時間内での練習時間と金銭的なコストを鑑みると, 指南練習は制限されてしまうと言える. そこで, ギター練習では多くの場合自主練習が中心になる. 自主練習では, 教則本に記された演奏動作を真似て練習する方法であるが, 自身の演奏動作が教則本通りに出来ているかを客観的に判断することは困難である. こういった, 自身の誤った演奏動作のままに反復練習を行うと, その演奏動作が身体に身に付いてしまい技術の向上を阻害するおそれがある. したがって, 自己練習においても誤った演奏動作を矯正し, 教則通りの演奏動作を身に付けることが重要であり, そのためにはギターをどのように弾いているのかを把握する必要があると考える.

そこで本稿では, 初心者のギター演奏の上達を最終目標とし, 演奏者の動作を検出し客観的に自身の演奏動作を判断することできるギター演奏認識システムの

開発を目的とする。

2. ギター演奏学習支援

2.1 ギターの演奏動作と演奏スキル

ギターは主に左手と右手を使って演奏する弦楽器であり、演奏方法には左手の指で1つの弦を押さえ、右手で1つの弦を弾いて演奏を行う「単音弾き」と左手の指で複数の弦を押さえ、右手で複数の弦を弾いて演奏を行う「コード弾き」がある。

ギターの演奏スキルで左手と右手のそれぞれの手で求められる演奏スキルが異なるので、演奏動作を練習する際にはそれぞれ異なった練習が必要となる上に、左手と右手の動きが同じタイミングで同期する必要がある、両方の演奏スキルが必要不可欠である。

練習で必要となる演奏スキルは左手においては、演奏する際の指の運びであり使い方でもある運指である。右手においては、弦を弾いて音を奏でる「ピッキング」である。それぞれについて、以下に詳細を説明する。

2.1.1 左手の運指

左手の運指では、押さえるべき弦付近のフレットに左手全体をポジション移動した後に、各指で弦を押さえる必要がある。その際に、フレット移動の際に指のばたつきやムダな動きのない運指と綺麗な音を奏でるためにフレット付近を押さえることが重要となる。また、「コード弾き」の場合には、押さえるフレット位置とどの指で押さえるか(以下、押弦指)は決まっているが、「単音弾き」の場合には、押弦指が楽譜に示されていないことも多く、同一音であったとしても演奏の前後関係により押弦指は変化する。つまり、同じ楽譜を演奏したとしても演奏者により異なる押さえ方となり、初心者場合には自身の使い慣れている指を多く使ってしまうが故に非効率な押さえ方や、次のポジションへの移動に悪影響となる押さえ方をすることがある。

2.1.2 右手のピッキング

右手で必要とされるピッキングスキルには、ストロークの大きさ、ピッキングタイミング、そしてピッキング方向がある。ピッキング方向には、弦を上方向から下方向に弾くダウンピッキング、反対に下方向から上方向に弦を弾くアップピッキングがある。これらを交互に繰り返す演奏方法をオルタネイトピッキングと

呼び、ピッキング技術の基礎の1つとして重要視されている。また、ピッキング方向についても楽譜に示されていないことも多く、演奏者によりピッキング方向が異なる場合がある。

2.2 関連研究

ギター演奏学習支援のアプローチには、大きく分けて2つの手法がある。1つ目に、ウェアラブルデバイスをを用いた手法であり、ピッキングを行う利き手側にウェアラブルインターフェースを装着し、手首の角運動量の時間変化からギター演奏における運動様態の推定を試みる研究[1]や、左手の運指を計測する方法として、Light-glove[2]などのデータグローブを使用する手法がある。

以上のウェアラブルデバイスを用いた研究では、ウェアラブルデバイスを身体に直接取り付けることで詳細な動きの変化を検出することができる。一方で、装着機器が演奏の妨げとなってしまう演奏者本来の動きを制限してしまう問題がある。

2つ目に、非接触型のモーションカメラを用いた手法として、Microsoft Kinect Sensorを用いたビートトラッキングを行う研究[3]がある。これは、複数の弦を弾くピッキングを対象としたものであり、テンポの把握を目的にギター領域のマスキングによる右手領域ならびに動きを検出している。また、坂牛[4]らは低コストであるWebカメラを用いて演奏に支障を出さずに運指情報を取得し、その運指情報に対応した音を出力する研究を行っている。具体的には、画像処理を用いることでフレットの検出と弦推定を行った後に、利用者の爪に張られた青色のテープの位置を検出することで、検出された青色の領域と推定した弦の位置が一致する場合に、その弦を押さえていると判定している。

これらのアプローチでは、演奏者本来の動作センシングを行うことが出来るが、右手か左手のどちらか片方の動きのみを対象とした動作検出であり、練習を行う際の上達箇所が限定されてしまう問題がある。

2.3 検出アプローチ

本研究では、指を覆うセンサや手首に固定するタイプのウェアラブルデバイスはギター演奏の妨げとなると考えモーションカメラを使用し「単音弾き」時の左

手と右手の動作検出を行う。また、正確な音程、タイミング、強さで音を奏でているかを検出するために、MIDI ピックアップを用いてギターの変弦の検出を行う。そして、これら2つの機器を組み合わせることで変弦時の演奏者動作を認識し、可視化を行う。

3. 認識手法

モーションカメラには Microsoft 社の Kinect V2(以下, Kinect)を使用する。モーションカメラである Kinect は, RGB 情報だけでなく深度情報, 人体領域, ならびに関節位置の検出する機能を有する。しかしながら, 関節位置の検出において, ギターのような道具を持つとその物体を人体領域と認識してしまい, 手などの関節座標を正しく認識することが出来ない。本アプローチでは, Kinect の RGB, デプスカメラと画像認識技術を組み合わせて手領域を検出する。

また, MIDI ピックアップには Roland 社の GK-3 を使用する。そして, 取得した MIDI データを PC に伝送する中継機器として Boss 社の GP-10 を使用する。本システムモデルの流れを図 2 に示す。

3.1 ヘッド領域とネック領域の検出方法

Kinect から取得した RGB, 深度情報から, 以下の方法によりヘッド領域とネック領域の検出を行う。

(1) 演奏者の右肩の Z 座標(深度情報)を閾値とし, デプスフィルタを用いて背景マスクを実装する。この前処理により, 以降の処理で使用するカラーフィルタや弦の直線情報を検出する際に使用する Hough 変換において背景画像に対する誤検出を軽減する。

(2) 背景マスクを実装した画像に対して, Hough 変換を用いて弦の直線検出を行い, ギターの傾きを推定する。

(3) ギターの傾き分アフィン変換を行い, ネック領域が水平になるように回転させる。

(4) アフィン変換を行った画像(RGB 表色系)を HSV 表色系に変換した後に, カラーフィルタを用いて, ヘッド領域のみを検出し, その領域に対してラベリングを行う。この際, ヘッド領域の HSV と類似する物体が画面上にある場合は画像の左端にある領域をヘッド領域と推定する。

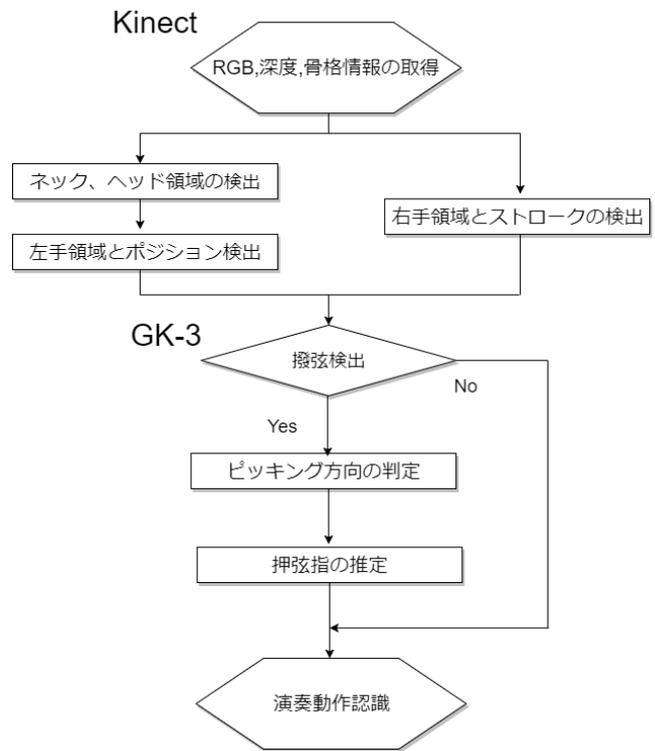
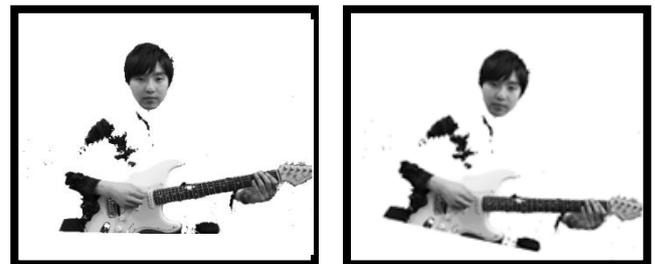
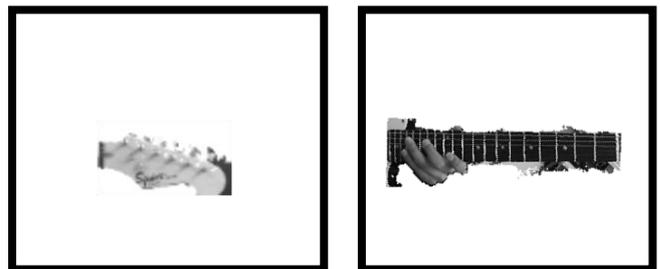


図 1 認識手法の流れ



(a) 背景マスク

(b) アフィン変換



(c) ヘッド領域の検出

(d) ネック領域の検出

図 2 ネック領域の検出結果例

(5) 検出したヘッド領域の最右端をネック領域との境界とし, 右側領域に対してトリミングを行い, ネック領域の矩形を検出する。ヘッド領域とネック領域の検出例を図 3 に示す。

3.2 右手に関する検出手法

3.2.1 右手領域の検出

右手領域の検出方法について述べる。

(1) 3.1(1)で, 実装した背景マスクに対して, 肌色の

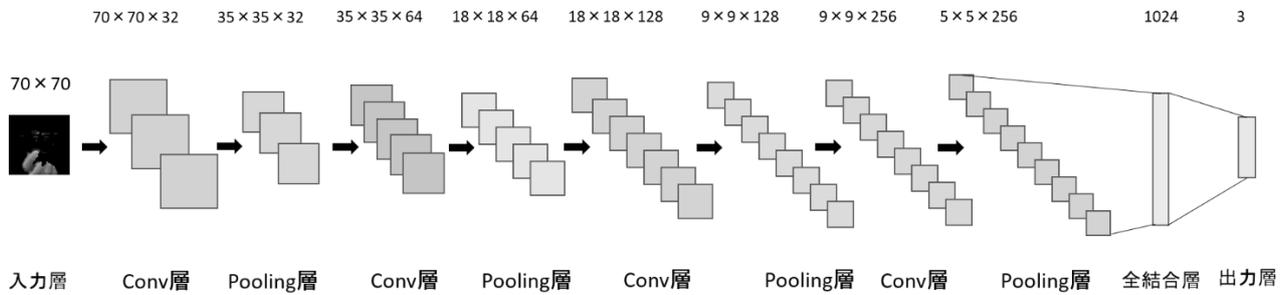


図 3 作成した CNN モデル

カラーフィルタを利用して肌色領域の抽出を行う。

- (2) 演奏者の右肩の下方領域にある肌色領域に対してラベリング処理を行い、検出した肌色領域を右手領域と推断する。

3.2.2 ストロークの検出

ラベリング処理を行った右手領域のモーメント情報から角度を計算し、その角度変化から右手の動きであるストロークを検出する。角度情報には、Kinect から取得した情報を使用しているためノイズや誤差情報も含まれるので、次式のローパスフィルタを実装した。

$$y[t] = 0.9 * y[t - 1] + 0.1 * x[t]$$

本式のxの値は、Kinect から得られる測定値、yはローパスフィルタファイタ適用後の値、tは時間とし、現在の計測値とひとつ前の値の重み付けによる加算平均をとる。本フィルタ処理を実装することで、ピッキング方向の判定での誤判定を軽減することができる。

3.2.3 ピッキング方向の判定

MIDI ピックアップである Roland 社の GK-3 を利用し、ギターの撥弦時刻を取得する。その際に、右手のモーメント情報の角度の値が1フレーム前の値より、大きければアップピッキングと判定する。反対に、1フレーム前の値より小さければダウンピッキングと判定する。ピッキング方向の判定式を以下に示す。

<アップピッキング>

$$y[t] - y[t - 1] > 0$$

<ダウンピッキング>

$$y[t] - y[t - 1] < 0$$

3.3 左手に関する検出手法

3.3.1 左手領域の検出

左手領域の検出方法について述べる。

- (1) 3.1(5)で検出したネック領域を HSV 表色系

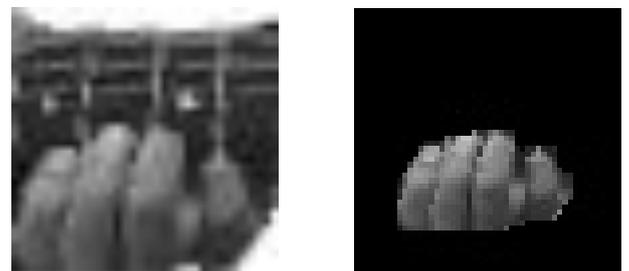
に変換する。

- (2) 肌色のカラーフィルタを用いてネック領域の矩形上の肌色領域を抽出し、その領域に対してラベリングとトリミングを行う。そして、このトリミングされた領域を左手領域と推断する。トリミングされた左手領域の検出結果を図 4(a)に示す。

3.3.2 画像サイズの正規化

トリミングした左手領域の矩形サイズは演奏者の手の大きさや異なる弦を押さえる際の指の形状、または Kinect と演奏者の距離によりばらつきが生じるので、以下の方法で正規化を行う。

- (1) 左手領域の矩形画像を黒色単色画像(70*70px)の背景画像の縦横中央揃えで重ね合わせる。
- (2) 重ね合わせた画像に対して HSV 表色系に変換した後に肌色のカラーフィルタを用いて左手の指の形状のみを検出する。左手の指の形状の検出結果例を図 4(b)に示す



(a) トリミング画像

(b) 重ねあわせ画像

図 4 左手の指の形状の取得結果

3.3.1 押弦指の推定

正規化を行った左手の指の形状画像(図 4(b))を畳み込みニューラルネットワーク(以下、CNN)の入力画像とし、押弦指の推定を行う。使用した CNN の構成を図 5 に示す。

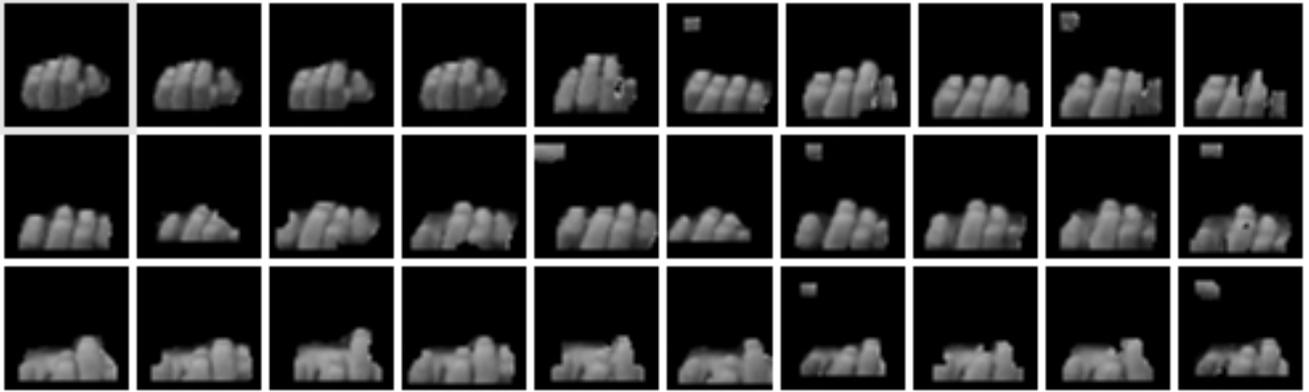


図5 作成したトレーニング画像の例

4. 評価実験

4.1 ピッキング方向の判定精度評価

4.1.1 概要

演奏者のピッキングの方向検出を検証するために、演奏者によるピッキングをシステムによって判定し、その判定精度を記録した。測定には5名の被験者に協力してもらい、1. ダウンピッキングのみ、2. アップピッキングのみを各100回演奏してもらいその判定精度を測定した。対象は、1弦の単ピッキングのみとした。

4.1.2 結果

ピッキング方向の判定精度を表1に示す。精度は小数点第4位までを求め、四捨五入した後にパーセント表示したものである。表1の結果より、平均90%を超える結果が得られた。

表1 ピッキングの精度評価の測定結果

	ダウン [%]	アップ [%]
A	95.0	98.3
B	90.9	96.5
C	81.2	88.1
D	91.7	95.6
E	96.3	98.5
平均	91.0	95.4

4.2 押弦指の推定精度評価

4.2.1 CNNの学習

単ピッキング時の押弦指の推定精度を評価するために用いたCNNの学習方法について述べる。

学習に用いたトレーニング画像は、一人当たりギタ

ーの第一弦の7フレットを人差し指で押さえた画像を200枚、8フレットを中指で押さえた画像を200枚、9フレットを薬指で押さえた画像を200枚の計600とし、これを8人の被験者に協力してもらった。つまり、一人あたり600枚、全体で4800枚の画像をトレーニング画像として使用した。トレーニングデータに使用した一例を図6に示す。また、学習に使用したハイパーパラメータを以下に示す。

- epoch : 20000
- batch size : 50

4.2.2 評価実験

4.2.1述べた学習済みのCNNを用いて、単ピッキング時の押弦指の推定精度評価を行った。テスト画像には、同一の8人の被験者に、一人当たり人差し指、中指、薬指を各50枚の計150枚、全体で1200枚の画像を使用した。また、テスト画像に使用した画像はトレーニング画像とすべてにおいて同一ではない。各押弦指の推定精度の結果を表2に示す。表2の結果よりすべての押弦指の推定において94%を超える結果が得られた。

表2 押弦指の測定結果

押弦指	認識精度 (Loss 値)
人差し指	0.9475(0.3905)
中指	0.94(0.5830)
薬指	0.955(0.30257)

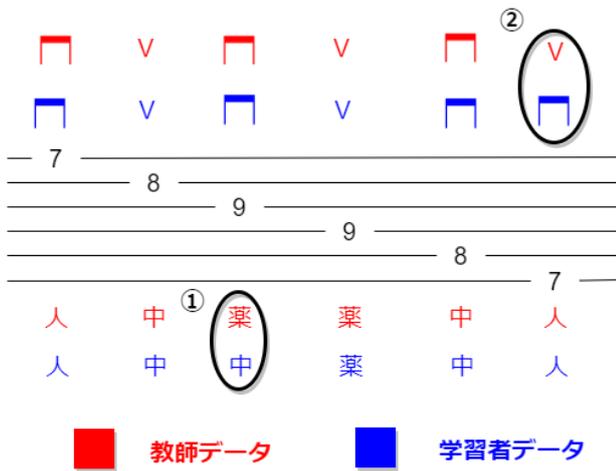


図 6 演奏情報の可視化

5. 本システムの学習支援への有用性の考察

本システムの学習支援への有用性として同じ楽譜を演奏した学習者と教師データを比較することにより、ピッキングタイミングのずれ、ピッキング方向、そして押弦指の違いのなど差異判定を行えると考える。図 7 は学習者と教師データを同譜面上に表示した結果となる。両者の演奏情報を可視化することにより、学習者は図 7 上の黒丸で示したような①押弦指の違いや②ピッキング方向の違いなどを確認し、教師データを真似るように自身の演奏動作を矯正することで学習支援に利用できると考える。

6. おわりに

本稿では、Kinect を利用したギター演奏者の演奏動作を客観的に判断することができるギター演奏認識システムを開発した。評価実験では、ピッキング方向の判定、CNN を用いた押弦指の推定ともに 90%を超える結果が得られた。

また、本システムは従来研究と比べて左手と右手の動きの両方を対象としているので、片方の動きに限定したものではなく実際のギター演奏に則った学習支援システムとしての応用性可能性を持つ。

今後の課題としては、左手の押弦指の推定において、第一弦の 7,8,9 フレットのみに限定するのではなく、他の弦、フレットに対しても同様に学習を行うことで、推定を行えるようにすることである。また、CNN にお

いて隠れ層などのハイパーパラメータを変化させ、推定精度の向上を図る。

謝辞

本研究は JSPS 科研費 JP17K01098 の助成を受けた。

参考文献

- (1) 浅野忠史, 松下宗一郎, “ギター演奏運動を評価するウェアラブルインターフェースの研究”, 情報処理学会第 78 回全国大会, no.2Y-06, pp.4-257-4-258, Sept.2012.
- (2) B.Howard,S.Howard, “Lightglove: wrist-worn virtual typing and pointing”, in Proc. of the 5th IEEE International Symposium on Wearable Computers(ISWC2001),pp.172-173,Oct.2001.
- (3) 糸原達彦, 大塚琢磨, 他, “Kinect による楽器マスキングを用いた視聴覚統合ビートトラッキング”, 情報処理学会第 74 回全国大会, no.5ZD-1, pp4-355-4-356,Mar.2012.
- (4) 坂牛和里, 植村あい子, 村岡眞伍, 甲藤二郎, “Web カメラを利用したギターの練習支援システムの検討”, 第 93 回オーディオビジュアル複合情報処理研究発表会, no.16, pp1-5, Aug.2016.
- (5) 元川洋一, 斉藤英雄, “ギター演奏支援のための構造特徴追跡を利用したマーカレス AR 表示”, 情報処理学会研究報告コンピュータビジョンとイメージメディア, Vol.3, pp.9-14, Jan.2008.
- (6) 青木直史, 棚橋誠, 岸本英一, 安田星季, 岩越睦郎, “画像処理によるギター運指動作のキャプチャリング”, 電子情報通信学会総合大会講演論文, no.D-11-110, p.110, Mar.2005.
- (7) 藤井創太, 浜中雅俊, 長谷川真晶一, “Fingering Simulator”, 情報処理学会研究報告音楽情報学科 (MUS), pp.167-172, Jul.2008.
- (8) N.Dalal, B.Tiggs, “Histograms of oriented gradients for human detection”, CVPR05,2005
- (9) 武藤亮介, 嶋田和考, 遠藤勤, “SVM と逐次学習を併用した HOG 特徴による手形状認識手法”, 電子情報通信学会, Mar.2010
- (10) Yoshitaka Kashiwagi, Youji Ochi, “Development of guitar performance recognition system Using Kinect Sensor”, ICEE 2016 Okinawa, 90309, Jul.2016