

連続デプス画像解析に基づく仮説検証型ピアノ運指認識手法

有賀 治樹¹ 岡 明也¹ 橋本 学^{1,a)} 長田 典子^{2,b)}

概要：本稿では、演奏者によらず使用できる、マーカレス・ピアノ運指認識手法を提案する。距離画像中のわずかな手掛かりをもとに指先候補位置を多数検出し、ピアノキー位置との対応を確率的に表現したものを仮説群として生成する。各仮説から手全体の姿勢を推定して画像化し、入力距離画像との手全体の整合性を求めることによって最尤仮説を決定する。このとき、各指先（親指～小指）の存在確率マップを用いて指先候補位置を確率的に表現し、仮説の尤度に反映させる。汎用 3-D ハンドモデルを用いて仮説をオンラインで画像化することにより、個人ごとの事前の実データ学習が不要となり、演奏者固有の演奏時の手の姿勢に影響されにくい、汎用的な運指認識を実現した。計 128 音からなる初心者向け楽曲を用いた実験により、認識成功率 88 %を確認した。

1. はじめに

音楽・楽器産業は、子供の情操教育において大きな役割を担っている。なかでも、ピアノは最もポピュラーな楽器であり、演奏者人口も非常に多い。ピアノ演奏のスキル向上には、特に初心者にとってはミスタッチ（打鍵音名ミス）と運指（指使い）間違いの 2 つの低減が重要な要素である [1][2]。

ミスタッチについては、電子ピアノから取得される音名信号と楽譜との照合によって判定できる [3]。しかし、運指の間違いは目視での判断が必要であることから、画像を利用した自動化が望まれている。また、運指情報の正確な記録や分析は、難度の高い曲の演奏における運指法の研究 [4][5] にも有効であることから、熟練したピアノ演奏者や研究者にとっても価値が高いと考えられる。さらに、運指認識を人間の手指の繊細な動きの計測・認識する課題として捉えれば、各種製造業や伝統工芸分野における熟練作業者の「匠の技」を後世に残すための技術の 1 つと考えることもでき、発展性が期待される。

ピアノ演奏においては、通常手指にデータグローブなどの特殊な装置を装着することができないため、カメラ等を用いた非接触手法として画像処理ベースの手法が有効であると考えられる。画像処理による人間の手指の認識に関しては、ジェスチャ認識等の研究があるが、ピアノ演奏にお

ける手指の動きは繊細で多様であり、手指以外の物体との接触があるため適用が困難である。

従来のピアノ運指認識の研究として、鍵盤の直上にカメラを設置し、カメラ画像からカラーマーカを装着した指先を検出することによって、運指を認識する手法 [6] がある。しかし、カラーマーカの装着は演奏者の自然な演奏を妨げる恐れがあり、日常用途には不向きである。また、肌色領域の輪郭から求めた主曲率をもとに指先を検出することによって、運指を認識する手法 [7] がある。しかし、手指と白鍵のコントラストが十分でないため、安定した指先検出は困難である。一方、距離画像を用いることによって、肌の色によらず手領域を抽出し、学習ベースで運指を認識する手法 [8] がある。しかし、ピアノ演奏時における多種多様な手の動きを全て学習するには約 3 万枚の画像が必要となる。さらに、演奏時の手の姿勢は演奏者により異なるため個人ごとの学習画像が必要であり、汎用性が低いと考えられる。

以上のように従来手法では、マーカによる自然な演奏の妨げ、肌色情報を用いることによる不安定な手領域抽出という問題点があるが、距離画像を用いる従来手法 [8] により解決されている。しかし、従来手法 [8] には、大量の学習画像が必要、汎用性が低いという 2 つの問題点が残されている。

そこで本研究では、距離画像を用いた、事前の学習が不要で汎用性の高い運指認識手法を提案する。本研究では、仮説検証型アプローチを採り、従来手法の問題点を解決する 2 つの基本アイデアがある。1 つめは、1 枚の距離画像のみの情報から手の姿勢を推測することによって、事前の学

¹ 中京大学
Chukyo University

² 関西学院大学
Kwansei University

a) mana@isl.sist.chukyo-u.ac.jp

b) nagata@kwansei.ac.jp

習を不要とすることである。2つめは、大きさ可変な3-Dハンドモデルを用いて手の姿勢を表現することによって、汎用性を向上させることである。

以下、2章では本研究で提案する手法について述べ、3章では提案手法の評価に関する実験結果を示す。最後に4章にて本研究の成果をまとめる。

2. 提案手法

2.1 概要

提案手法の流れを図1に示す。まず、ピアノ鍵盤の直上に設置されたレンジセンサから距離画像を取得し、手領域を抽出し、手指画像を生成する。この画像から指先の画像パターンを用いて指先候補位置を複数検出し、その位置に存在するキーと各指（親指～小指）の対応を仮説群として生成する。仮説は、図2(a)の指番号と(b)のキー番号を対応付けた(c)のようなマップで表現する。この際、各指先候補位置の各指らしさを求め、各仮説の確率に反映させる。各仮説から生成した仮説距離画像と入力距離画像との手全体の整合性を確率に反映させ、最尤仮説を決定する。これにより、電子ピアノから取得された音名信号（打鍵されたキー番号）を対応づけることにより、運指情報として、打鍵されたキー番号と指番号の組み合わせを出力する。

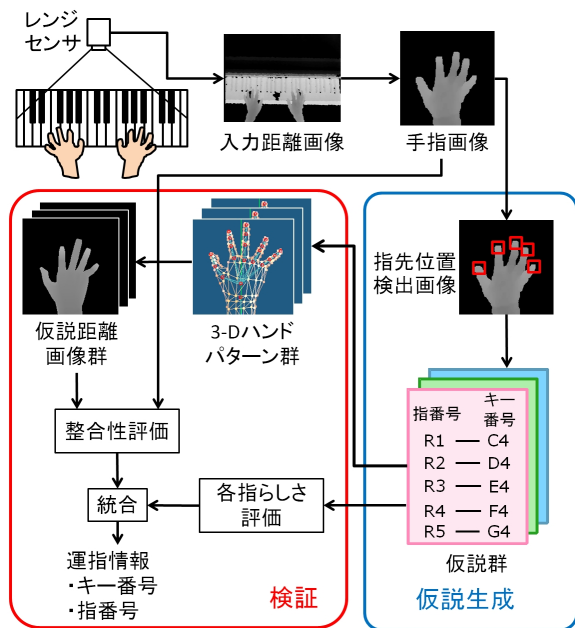


図1 提案手法の流れ

Fig. 1 Flow of proposed method.

2.2 前処理：手指画像生成

汎用MIDIインタフェースをもつ電子ピアノを用いて、打鍵した瞬間の画像を取得する。MIDI信号からは、音名のほか、打鍵や離鍵のタイミングを取得可能であることから、打鍵タイミングと同期した画像を取得できる。

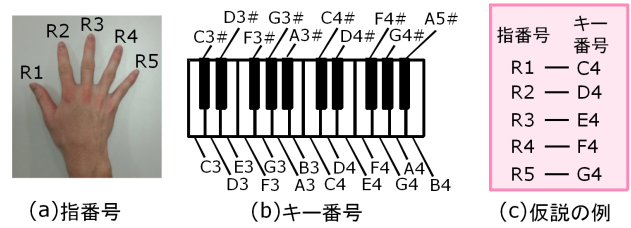


図2 仮説の構造

Fig. 2 Hypothetical structure.

通常の2次元濃淡カメラによる画像では背景が照明変動の影響を受けることに起因して背景差分が困難になることが少なくないが、距離画像の場合は背景の変動が比較的少なく、背景差分の効果が大きい。

提案手法においては、手指を含む距離画像とあらかじめ取得しておいた鍵盤のみの画像の間で差分を計算する。レンジセンサをピアノ鍵盤の上方に設置した場合には、センサと鍵盤までの距離が一定であり、抽出したい手指までの距離だけが変化することに注目すると、式(1)のように各座標の入力画像と背景画像の差分から手領域を抽出することができる。なお、 Th は距離画像差分値のしきい値であり、実験的に求める。しかし、ノイズにより、手領域以外が抽出されることがあるため、背景差分画像に対してラベリングを施し、面積の大きい領域を手領域として抽出する。これにより手領域を安定して求めることができ、手領域画像を生成することができる。図3(a)に入力画像、(b)に背景画像、(c)に背景差分画像、(d)に手領域画像を示す。(d)より、手領域だけを抽出できていることがわかる。さらに、オープニングを施すことによって、手指周りのノイズの低減を行う。また、手首位置をもとに手首位置より上部の領域を抽出し、手指画像を生成する。(e)に手指画像を示す。手首位置の詳細な検出方法は、2.4節で述べる。

$$S(i, j) = \begin{cases} I(i, j) & \text{when } |I(i, j) - B(i, j)| > Th \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

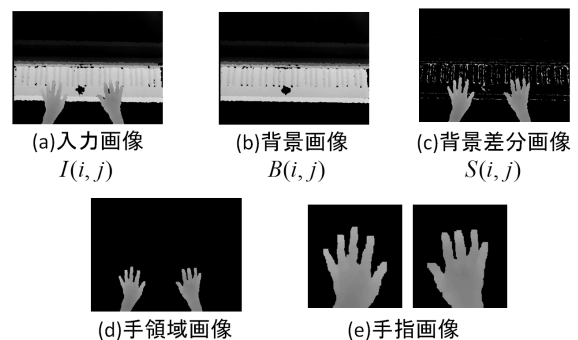


図3 手領域抽出と手指画像生成

Fig. 3 Extraction of hand region and generation of hand image.

2.3 仮説生成

仮説は、各指とキーの組み合わせとする。手指画像から検出した指先候補位置をもとに指先が存在する可能性のあるキーを推定する。それらのキーと各指との対応を複数生成し、仮説とする。

指先候補位置検出のために手指画像に対して、指先の画像パターンを用いたマッチングを適用し、類似度マップを作成する。類似度マップからしきい値以上の点を抽出する。図4(a)に指先の画像パターン、(b)に類似度マップ、(c)に類似度の高い点を抽出した結果を示す。その後、類似度の高い点を抽出した画像に対してラベリングを施し、複数のラベル付けされた領域を検出する。検出された領域ごとに重心を求め、指先候補位置とする。さらに、あらかじめキャリブレーションによって求めた各キーの画像上の座標を用いて、指先候補位置に存在するキーを指先候補位置に付随させる。このとき、指先検出の誤差を考慮し、指先候補位置を中心とした半径 r の円領域内に含まれるキーも同時に付随させる。(d)に指先候補位置と r の例を示す。その後、各指先候補位置と各指がとりうるすべての組み合わせを生成する。このとき、指が交差する組み合わせは除外しておくことで、仮説を削減する。この組み合わせをもとに指先候補位置に付随するキーと各指の対応マップを生成し、仮説とする。

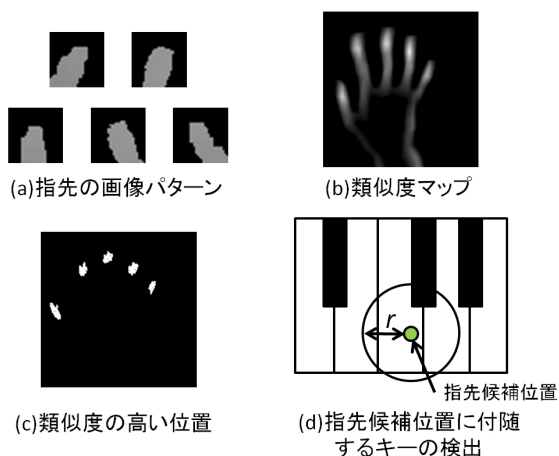


図4 指先候補位置の検出
Fig. 4 Detection of candidate for fingertip position.

2.4 検証

検証には、指先候補位置の各指らしき評価と手全体の整合性評価の2つの方法を用いる。指先の局所的な情報に対する評価と手全体の大局的な情報に対する評価を統合することにより、検証の信頼性を向上させる。これらを用いて各仮説の尤度を計算し、最尤仮説を決定する。この最尤仮説によって各指とキーの位置関係が決定され、MIDIの音名信号を用いることにより、運指情報（打鍵されたキー

番号と指番号の組み合わせ）が完成する。2つの検証方法の詳細を以下に説明する。

2.4.1 各指らしき評価

事前に作成した各指先の存在確率マップを用いて、各指先候補位置が各指である確率を求め、仮説の尤度に反映させる。図5に指先候補位置（N1～N5）を示す。表1にN1～N5の指先候補位置の各指らしき確率の例を示す。存在確率マップの作成方法と仮説の尤度の算出方法について説明する。

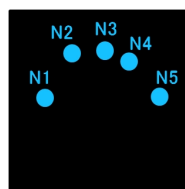


図5 指先候補位置
Fig. 5 Candidate for fingertip position.

表1 各指らしき確率
Table 1 Probability of each finger-likeness.

	親指	人差し指	中指	薬指	小指
N1	0.51	0.21	0.07	0.01	0.00
N2	0.01	0.32	0.08	0.01	0.00
N3	0.00	0.21	0.30	0.12	0.00
N4	0.00	0.00	0.07	0.45	0.05
N5	0.00	0.00	0.00	0.13	0.52

存在確率マップ作成

ピアノキーを打鍵した際の距離画像を学習画像とする。複数の人物の学習画像から手動で指先位置を求めることで各指先の存在確率マップを作成する。しかし、学習画像の手の位置と姿勢はさまざまであるため、各指先位置の分布は、大きく広がってしまう。そこで、手首位置を基準とした指先位置を求めることで、各指先位置の分布を収束させる。そのために学習画像から手首位置をもとに手指領域を切り出し、手指の傾きを補正した画像から指先位置を求める。図6(a)に学習画像、(b)に手指領域の画像、(c)に手指の傾きを補正した画像を示す。手首位置検出と手指の傾き補正について以下に述べる。

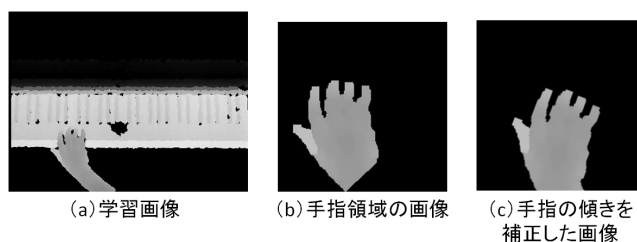


図6 各指先位置の手動検出
Fig. 6 Manual detection of candidate for fingertip position.

手首位置検出には、[10]で用いられている方法を用いた。一般に、人の腕の幅はほぼ一定で、手首から手のひらにかけて幅が広がっていく。このことを利用して、手領域画像の下部から腕の太さを求めていき、腕の太さが一定以上太くなった位置を手首位置として検出する。また、手指領域の画像の重心位置と手首位置を結んだ線と手首位置から画像上部へのびる直線のなす角度を手指の傾きとする。図

7 (a) の α に手指の傾きを示す. α だけ画像を回転させ, 手指の傾きを補正する. この画像から指先位置を手動で求める.

手首位置から手動で求めた指先位置までの距離 l , 手首位置と指先位置を結んだ線と手首位置から画像上部へのびる直線のなす角 θ を求める. (b) に l と θ を示す. このときの l と θ の値を用いて各指ごとに 2 次元の頻度分布を作成する. ただし, 人物によって手の大きさが異なるため, 手の大きさに応じて l の値を正規化し, 頻度分布を作成する. 図 8 に各指先位置の頻度分布を示す. 各指先の分布の重なりが少ないことから, 各指の識別に有効であると考えられる. この頻度分布を正規化し, 2 次元のガウス分布で近似する. このガウス分布を指先の存在確率マップとする.

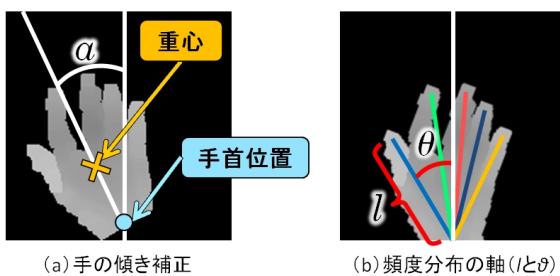


図 7 学習画像を用いた各指先位置の頻度分布の作成

Fig. 7 Creation of frequency distribution of each fingertip position using learning image.

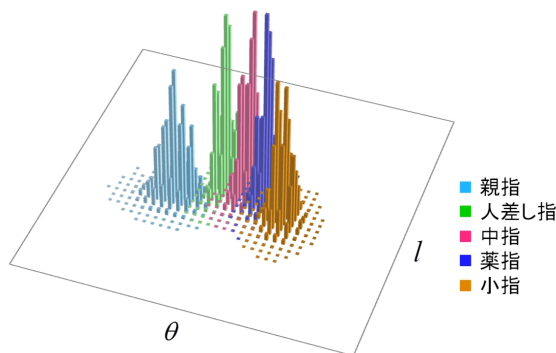


図 8 学習画像を用いた各指先位置の頻度分布

Fig. 8 Frequency distribution of each fingertip position using learning image.

仮説の尤度算出

入力距離画像から検出した指先候補位置を手指の傾きに応じて補正する. この指先候補位置の l と θ の値を求め, 各指先の存在確率マップを用いて各指先候補位置の各指である確率を求める. 図 9 に各仮説の尤度の求め方の例を示す. (a) に仮説, (b) に仮説を生成するための対応関係を示す. 仮説の親指 (R1) に対応するキーが C4 であり, その C4 が付随している指先候補位置が N1 である. N1 の親指 (R1) である確率がその仮説の親指らしさ P_k となる.

これを各指で求め, それらの平均値 P を仮説の尤度に反映させる.

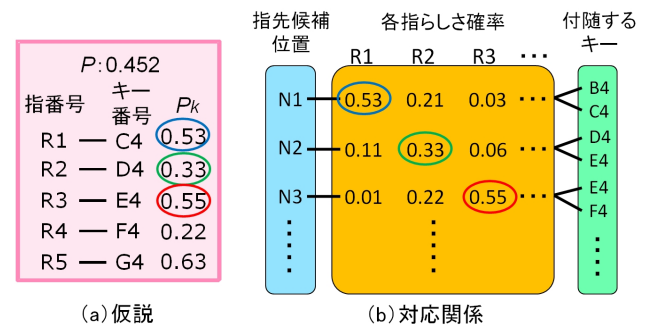


図 9 仮説の尤度の算出

Fig. 9 Calculation of hypothetical likelihood.

2.4.2 手全体の整合性評価

各仮説から推定される手の姿勢をコンピュータグラフィックスで 3-D ハンドモデルを用いて再現し, 3-D ハンドパターンを生成する. その後, 仮想的に距離画像を生成し, 入力距離画像との整合性を評価する. この値を各仮説の尤度に反映させる. 3-D ハンドモデル, 3-D ハンドパターンの生成, 仮説の尤度の算出方法について説明する.

3-D ハンドモデル

手は本来非剛体の物体であるが, 本研究では剛体の関節物体として近似する. 図 10(a) は, 形状モデル, (b) は, 3-D ハンドモデルに内包する関節モデルであり, 16 個の関節を 26 自由度で制御可能である.

本研究では, (b) に示すそれぞれの関節について, DIP 関節および PIP 関節に屈伸運動を行う回転の 1 自由度, MP 関節には屈伸および内転・外転を行うため回転の 2 自由度を考える. 手首の関節には, 左右の首振りの 1 自由度, 肘の関節には平行移動と左右の首振りの 3 自由度を考える. また, これらのモデルは肘を最上位として, それ以降接続する関節を子とする階層構造を持ち, 親の関節のパラメータの変更は子の関節に影響を与える. また, 手のひらの大きさや指の長さなどの静的なパラメータは, あらかじめ測定し, 手動で設定されるものとする.

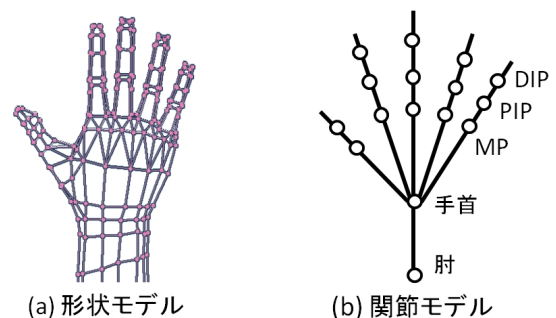


図 10 3-D ハンドモデルの構造

Fig. 10 Structure of 3-D hand model.

3-D ハンドパターンの生成

3-D ハンドパターンとは、仮説から推定される手の姿勢を 3-D ハンドモデルを用いて表現したものである。図 11 に 3-D ハンドパターンの例を示す。3-D ハンドパターンの生成のために、まず、入力距離画像から手首位置を検出する。この手首位置に 3-D ハンドモデルの手首位置を合わせ、逆運動学を用いて各仮説における各指の対応キーの位置にモデルの各指先を対応させることにより 3-D ハンドパターンを生成する。このとき、手指関節の可動範囲 [9] [10] に基づき、手形状として自然でない仮説を除外する。それぞれの関節に対する可動範囲を表 2 に示す。

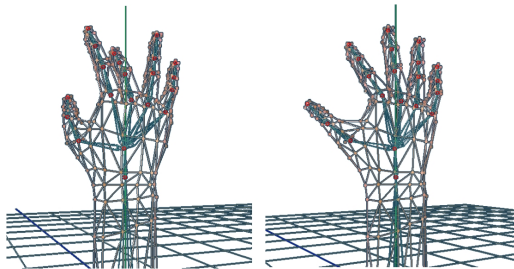


図 11 3-D ハンドパターンの例

Fig. 11 Example of 3-D hand pattern.

表 2 関節角の可動範囲 [°]

Table 2 Working volume of joint angle.

	DIP	PIP	MP	内転
親指		-20~70	-20~50	-40~40
人差し指	0~80	0~100	-40~90	-20~20
中指	0~80	0~100	-20~90	-20~20
薬指	0~80	0~100	-20~90	-20~20
子指	0~80	0~100	-30~90	-20~20

仮説の尤度算出

3-D ハンドパターンから各仮説を画像化し、仮説距離画像とする。図 12 に仮説距離画像の例を示す。これを入力距離画像と照合することにより整合性を判定する。照合には手領域全体の画素相関値を用い、相関値を各仮説の確率に反映させる。



図 12 仮説距離画像の例

Fig. 12 Example of hypothetical depth image.

3. 実験結果と考察

3.1 実験準備

本研究では、MIDI インターフェイスが搭載された電子ピアノを用いることで PC との接続により各種 MIDI イベントを読み取る。また、レンジセンサとして、Microsoft® の Kinect を用いた。Kinect を用いることにより約 4 オクターブ程度の範囲の鍵盤を計測可能である。実験環境を図 13 に示す。



図 13 実験環境

Fig. 13 Experiment environment.

3.2 演奏者に対する汎用性に関する実験

被験者 5 人を対象に提案手法を用いて運指の認識成功率を評価した。実験には F3-F5 の 14 鍵を各指で打鍵した距離画像を 70 枚用いた。結果を表 3 に示す。被験者によって、認識成功率にばらつきがあるが、平均して 90 % 以上の認識成功率であった。また、被験者全員が少なくとも 85 % 以上の認識率であり、演奏者に依存しにくい手法であることを確認した。図 14(a)(b)(c) に入力画像を示し、それぞれの画像に対応して最も整合性が高かった仮説距離画像を (d)(e)(f) に示す。

表 3 複数の演奏者に対する認識成功率 [%]

Table 3 Recognition rates for the performance of the multiple player.

	提案手法
被験者 A	86(60/70)
被験者 B	96(67/70)
被験者 C	89(62/70)
被験者 D	87(61/70)
被験者 E	96(67/70)
平均	91

3.3 初心者向けの楽曲を用いた実験

被験者 3 名が演奏した片手演奏の Menuet(J.S.Bach)[11] の演奏データを用いて提案手法の運指の認識成功率を評価した。結果を表 4 に示す。3 名の被験者の認識成功率が 85

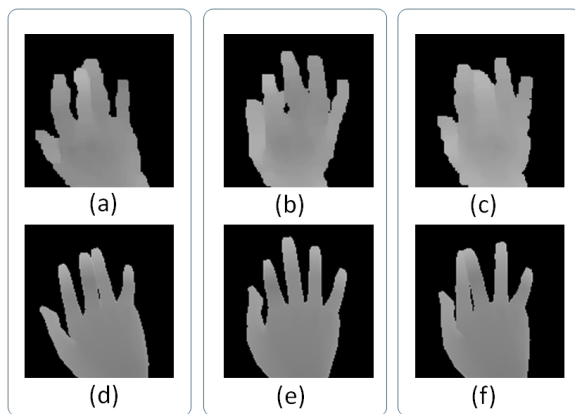


図 14 入力画像と最も整合性の高い仮説距離画像

Fig. 14 Input images and hypothetical depth images which have high consistency.

%以上であり本手法が実際の曲を演奏した際にも演奏者に依存しにくい手法であることを確認した。図 15 に認識成功例を示す。(a) は楽譜に指定された正しい運指、(b) は入力画像、(c) は最も整合性の高かった仮説距離画像を示す。

表 4 初心者向けの楽曲を用いた認識成功率 [%]

Table 4 Recognition rates using music for beginners.

	提案手法
被験者 A	86(110/128)
被験者 B	88(112/128)
被験者 C	88(113/128)



図 15 認識成功例 (第 5 小節 2 音目)

Fig. 15 Example of recognition success.

認識失敗の原因としては、親指が他の指によって隠される指くぐり状態において指先検出の失敗があったと考えられる。図 16(a) に指くぐりが起こった際の入力画像、(b) に指先検出の類似度の高い位置を示す。親指が他の指に隠されており、指先位置の認識が困難である。Menuet では合計 9 回の指くぐりが発生しており、認識成功率が低下したと考えられる。

4. おわりに

距離画像を用いることによって、マーカを用いることなく、演奏者によらない運指認識手法を提案した。3-D ハンドモデルを用いることにより、仮説の画像化をオンラインで行うことを可能とし、事前の学習を不要とした。指先の

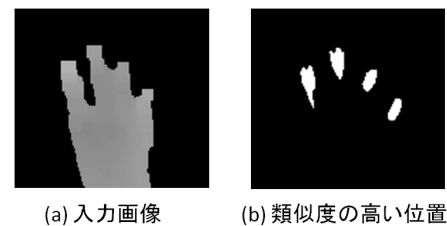


図 16 認識失敗例

Fig. 16 Example of recognition failure.

局所的な情報に対する評価と手全体の大局的な情報に対する評価を統合することにより、検証の信頼性を向上させた。また、計 128 音からなる初心者向け楽曲を用いた実験により、認識成功率 88 % を確認した。

今後の課題として、認識成功率の向上と指くぐりへの対応が必要であると考えられる。今後は、ピアノキーを打鍵した時だけでなく、その前後の動きを認識することで運指認識成功率の向上と指くぐりへの対応を目指す。

謝辞

本研究の一部は科研費 B (24300088) の助成を受けて遂行されたものである。

参考文献

- [1] 古屋晋一：ピアニストの脳を科学する超絶技巧のメカニズム春秋社,(2010).
- [2] 下山望：ピアノ運指法一譜例分類による一、ムジカノーヴァ,(1998).
- [3] Takayuki Hoshishiba and Susumu Horiguchi : Improved DP matching between a musical score and its performance using interpolation, Proc. Acoustical Science and Technology, pp.13-19, (2001).
- [4] Melanie Hart,Rovert Bosch,Elebert Tsai : Finding Optimal Piano Fingerings, The.UMAP Journal,vol.21,No.2,pp.167-177, (2000).
- [5] 米林裕一郎, 亀岡弘和, 嵯峨山茂樹：隠れマルコフモデルに基づくピアノ運指の自動決定, 情報処理学会研究報告, 2006-MUS-65, pp.7-12, (2006).
- [6] 竹川佳成, 寺田努, 西尾章治郎：鍵盤楽器のための実時間運指取得システムの構築, コンピュータソフトウェア(日本ソフトウェア科学会論文誌), vol.23, No.4, pp.51-59, (2006).
- [7] 子安大士, 木村慎二, 前川仁：ピアノ演奏動作解析のための 3 次元手指追跡, 情報科学フォーラム 2011, pp.171-172, (2011).
- [8] 岡明也, 橋本学：連続デプス画像解析に基づくマーカレス運指認識, 情報処理学会コンピュータビジョンとイメージメディア研究会, Vol.2012-CVIM-182, No.17, pp1-8, (2012).
- [9] 鎌倉矩子：手のかたち 手のうごき, 医歯薬出版株式会社, (1985).
- [10] 藤木隆司, 陳維英, 吉本廣雅, 有田大作, 谷口倫一郎：手の制約知識を用いた逆運動学解法による実時間三次元手形状推定, 画像の認識・理解シンポジウム, pp1380-1387, (2005).
- [11] "Notenbuechlein fuer Anna Magdalena Bach", G. Henle Verlag, (1983).