

距離動画像を用いた学習データの最適選択に基づく マーカレスピアノ運指認識

岡 明也¹ 橋本 学¹

概要: ピアノ運指の認識は、ピアノ初心者への演奏スキル向上や、熟練アーティストの技の記録のために重要な技術である。本稿では、レンジセンサを用いて演奏者の手指の動きを連続的に撮影することにより、特別なマーカを用いることなく運指認識する手法を提案する。まず学習時には、あらかじめ辞書データベースとして正しい運指情報と手首位置が付与された大量の手指の距離画像を登録しておく。次に認識時には、入力された距離画像を辞書内のデータと照合することで運指を判定する。このとき、ピアノから得られた音名信号と手首位置を利用して探索範囲を限定し、速度向上と信頼性向上を図る。解の探索には ANN を用いる。実験の結果、初心者向けの Menuet (J.S.Bach) を自然に演奏した場合において、和音に対する認識性能、打鍵持続時間に関する認識性能を明らかにした。10 小節全 60 音における認識成功率 93.3 % を確認し本手法の有効性が示された。

キーワード: 距離画像, 手指認識, ピアノ運指, ANN

Range Image-based Marker-less Piano Fingering Recognition by Efficient Selection of Training Data

OKA AKIYA¹ HASHIMOTO MANABU¹

Keywords: depth image, finger recognition, piano fingering, ANN

1. はじめに

ピアノ演奏において、正しい運指で演奏することは初学者にとって重要である [1]。ここで、運指とはピアノ演奏における打鍵指の運び (指使い) であり、打鍵した指の種類と打鍵された音名の組み合わせ情報である。ピアノ練習時における音名の弾き間違いは、音を聴くことによって演奏者自身にとっても発見しやすいが、運指の間違ひは楽譜に記載された運指指定との比較が必要であり、初心者には判断が難しいことが多いことから、練習時に運指の状況を自動的に認識し、間違いを指摘してくれるシステムの開発が望まれる。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 [2][3][4] にも有用であることから、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を計測・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一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はできないため、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適していると考えられる。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多くの研究がある [5]。手指を用いたジェスチャ認識の例として、指先を検出し追跡する手法 [6] があり、机上の複数の指先を実時間で追跡することができる。しかし、認識できるジェスチャの数が少ないことから、ピアノ演奏の繊細で多様な動きに適応することが難しいと考えられる。ピアノ演奏の認識の例として、手指の肌色領域を抽出する手法 [7] があるが、手指と白鍵の色相差やコントラストが十分ではないため、安定した指先検

¹ 中京大学大学院 情報科学研究科
Graduate School of Information Science and Technology,
Chukyo University, Toyota, Aichi 470-0348, Japan

出には困難が予想される。また、多視点カメラで撮影することによって手指の3次元位置を計測するモーションキャプチャの応用が報告されており、リアリティの高いCG生成の分野で優れた成果を挙げているが [8], 手指になんらかのマーカの装着が必要のために、ピアノ練習時においては演奏者への負荷が懸念される。また、単眼カメラによる手法も提案され [9][10], 精度的には実用的な性能を達成しているが、爪先にカラーマーカの装着を必要とすることから、たとえ初心者といえども十分な自然な練習を妨げる要因になり得ると考えられる。ところで、手指の姿勢推定に関する研究としては、手の輪郭情報から特徴量を算出し階層構造学習を用いて姿勢推定する手法がある [11]. この手法は、手指の姿勢推定を高速に行える手法であるが、手に近接物がある場合では、正しく輪郭情報を抽出できないことが懸念される。よって、この手法を手に近接物があるピアノ運指認識にそのまま適用することは難しいと考えられる。

そこで本研究では、マーカを用いることなく、ピアノ演奏者の運指情報を獲得する手法を提案することを目的とする。運指情報のうち打鍵されたキーの音名に関しては、電子ピアノを使用した場合には鍵盤内のセンサの利用することによって取得できる。したがって、本研究では音名が取得可能であることを前提として、打鍵指の認識手法を提案する。

本研究では、レンジセンサを用いて運指認識する。レンジセンサを用いることで照明変動や肌の色に起因することなく安定的に手領域を抽出することができる利点がある。さらに、ピアノ鍵盤上部にレンジセンサを設置することで、二次元濃淡カメラに比べ打鍵時の指の変化を捉えることができるため、ピアノ運指の認識にはレンジセンサが適していると考えられる。提案手法は、学習モジュールと認識モジュールから構成される。学習モジュールは、さまざまな打鍵パターンに対応する距離画像群を取得して学習クラスと定義し、画像から手首位置を検出し、さらに運指情報を付加して辞書データベース内に格納する。認識モジュールは、運指未知の入力画像を辞書データベース内の距離画像群(クラス)と比較することによって、入力画像が属するクラス(打鍵指)を判定する。このとき、ピアノから得られた音名情報と、入力画像から得られる手首位置をもとに、当該音名に関係するクラスのみを効率よく事前抽出することによって、認識の高速化と高信頼化を図る。

本研究では、学習パターンおよび運指未知の入力画像両方に関して、前処理として空間的ノイズの低減、さらに手領域部分の抽出をおこなう。一般に打鍵キー名と打鍵指の組み合わせとして表現される識別クラス数が非常に多いことから、識別には、kd ツリーを利用して近似的な最近傍探索を実現する ANN(Approximate Nearest Neighbor) アルゴリズムを用いる。これによりクラス数が多い場合でも高速な識別が可能になる。以上のように提案手法を用いるこ

とにより、マーカが不要なピアノ運指認識の実現できる。

以下、2章では本研究で提案する手法、3章では楽曲を用いた実験の結果を述べ、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

提案手法は学習モジュールと認識モジュールの2つからなる。図1にブロック図を示す。学習時には、さまざまな運指により鍵盤を打鍵した場合の距離画像に対し、一連の前処理を施し、手首位置を検出する。教師信号として運指情報と手首位置が付与され、辞書データベースに格納される。また、認識時には学習時と同様の前処理後、ピアノから得られる音名信号と入力画像から手首位置を求めて、辞書データベースの絞り込みを行う。運指未知の入力画像と選択された辞書データベース間で最近傍探索を行うことによって運指情報が推定される。

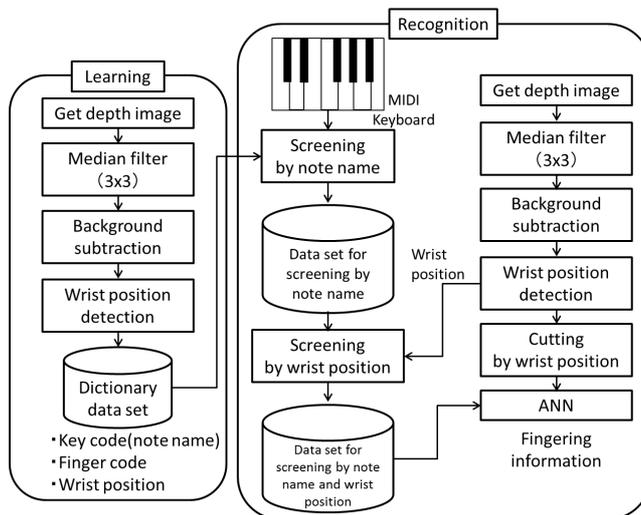


図1 提案手法の学習・認識のブロック図

2.2 運指情報の表現方法

本研究では、ピアノ鍵盤の各キーと、指について識別記号を設定し、その組み合わせ文字列として運指を表現する。

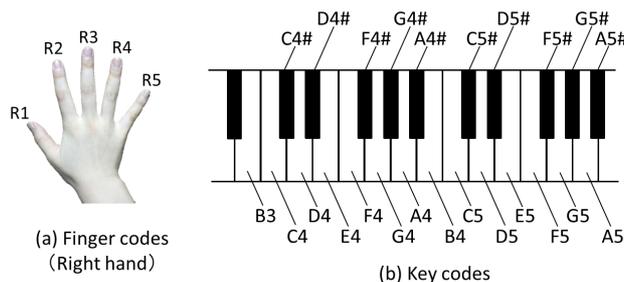


図2 各指とキーに設定した識別記号

図2に設定した各指とキーの識別記号を示す。指の記号としては同図(a)のように右手の親指から順にR1, R2・・・のように割り当てる。左手も同様である。またキーの音名については、図(b)のようにド(C)からシ(B)からなる音名に、音の高さを表す数字を組み合わせた識別記号をピアノ鍵盤の各キーに付与する。

表1に運指情報の表現例について示す。識別記号が“R1C4”ならば、R1(右手親指)でC4(ド)のキーを弾いていることを表す。また“R1C4+R3E4”ならば、R1(右手親指)でC4(ド)のキーを弾き、同時に右手のR3(右手中指)でE4(ミ)のキーを弾いていることを表す。

表1 運指情報の表現例

運指情報の表現	意味
R1C4	R1(右手の親指)でC4(ド)を弾く
R1C4+R3E4	R1(右手の親指)でC4を、 R3(右手の中指)でE4を弾く(和音)

2.3 MIDI信号を用いた辞書データベース作成

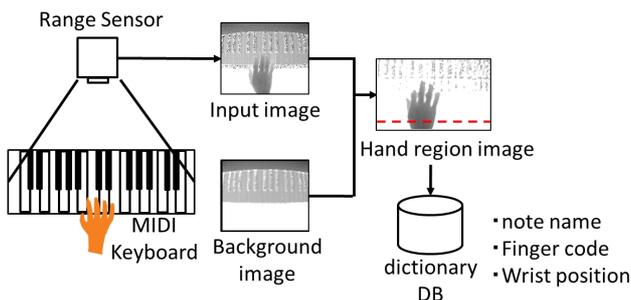


図3 MIDI信号を利用した学習画像生成の流れ

MIDI (Musical Instrument Digital Interface) [14]とは、電子楽器の演奏データを機器間でデジタル転送するための標準規格であり、定義されている代表的な信号として例えばノートオン(打鍵イベント)、ノートオフ(離鍵イベント)がある。そこで、辞書データベース作成の際にMIDIキーボードからPCに出力される打鍵イベントと離鍵イベントを用いて、レンジセンサの距離計測トリガーを連動させることで、キーを打鍵した瞬間の距離画像を自動的に取得できるだけでなく、打鍵された音名も自動取得できる。

図3にMIDI信号を利用した学習画像生成の流れを示す。まず、得られた距離画像に対し、空間的ノイズを低減させるためにメディアンフィルタを施す。そして、手領域を抽出するために、あらかじめ取得しておいた背景画像を用いて差分処理を施し、手領域画像を生成する。次に手領域画像から手首位置を求める。ピアノ演奏において、手首は常に鍵盤に対して平行に動くことから、あらかじめ手首位置検出ラインを設定する。同図の破線は手首位置検出ラ

インである。このライン上を走査し、手首の両端位置の中点を手首位置とする。

打鍵指の番号はユーザが手入力する必要があるが、これにMIDIキーボードから得られた音名と手首位置を統合し、距離画像の属性として付与し、辞書データベースに格納する。左右の手に対しそれぞれ辞書データベースを作成する。以上の処理により、運指情報と手首位置を伴った辞書データベースが作成される。

2.4 音名と手首位置によって絞り込まれた辞書データを用いた運指認識

本節では、認識時に打鍵された音名、手首位置を用いて、さまざまなクラスが混在した辞書データベースから辞書データを絞り込んで認識する方法について説明する。

図4に音名と手首位置による絞り込みの図を示す。まず、運指未知の入力画像に対し、打鍵時の音名を用いて、2.3で作成した辞書データベースから入力音名に対応した辞書データを絞り込む。そして、さらに辞書データを絞り込むために、学習時と同様の処理で入力画像から手首位置を求め、入力画像と辞書データベース内の手首位置の差を求めて、差がd[pixels]以上ある辞書データは距離計算候補から除外する。図5に、手首位置から求めた切り出し領域を示す。

次に入力画像の手首位置を中心とし、入力画像と絞り込まれた辞書データの画像の切り出しを行う。両手による交差は発生しないと仮定し、左右それぞれの画像を切り出す。図5に左右の手の切り出し領域を示す。切り出しサイズは、手領域が含まれるようあらかじめ設定しておく。手領域の切り出し結果を図6に示す。左右の切り出された手領域画像を、画像ベクトル化し、それぞれ左右の辞書データベースと照合することによって運指を認識する。

入力データと学習データ群との照合アルゴリズムとしてはkdツリーを利用して近似的な最近傍探索を実現するANN (Approximate Nearest Neighbor) アルゴリズム [12]を用いる。

3. 実験結果と考察

3.1 実験に用いた装置

本研究では、ピアノ鍵盤として、88鍵デジタルピアノを用いてデータを取得した。このピアノにはMIDIインタフェースが搭載されており、PCとの接続により各種MIDIイベントを読み取ることができる。

また、レンジセンサとしては、Microsoft社製のKinectを用いた。図7に実験装置外観を示す。我々はこのKinectセンサを鍵盤の上方70cmの高さに設置した。これにより約4オクターブ程度の範囲の鍵盤が計測可能である。

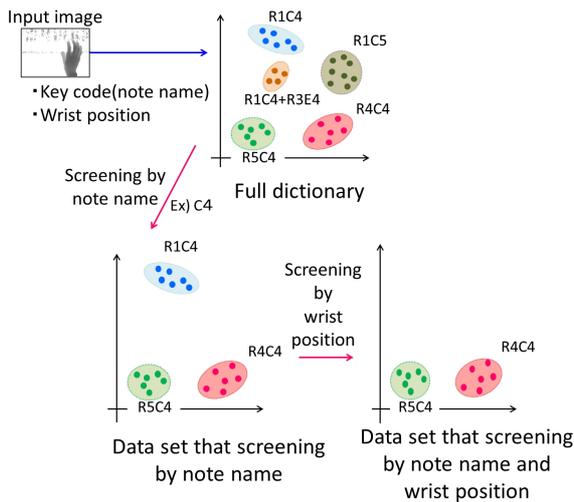


図 4 音名と手首位置による絞込み

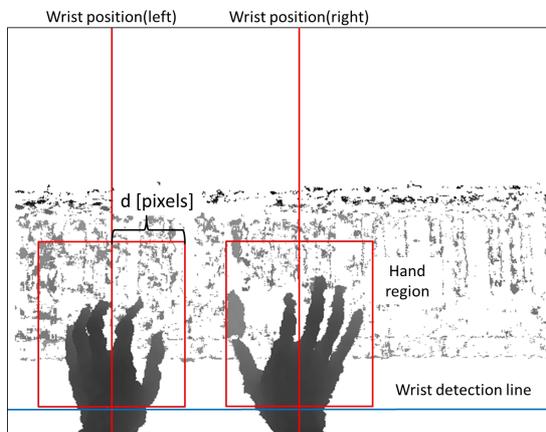


図 5 手首検出位置ラインと手首位置から求めた切り出し領域

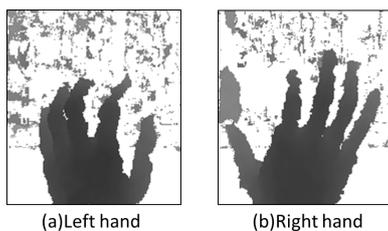


図 6 切り出された左右の手領域画像

3.2 実験方法

本節では、実験に使用した辞書データベース、評価用楽曲について説明する。

本手法の有効性を示すための評価用楽曲として Menuet (J.S.Bach) [13] を用いた。図 9 に Menuet の楽譜を示す。Menuet は、和音が含まれている点や、長さの異なる音符が含まれている点や、左右の手が異なる動きをする点などから、ピアノ初心者の練習に向いている楽曲であるといわれている。テストデータとしてこの楽曲を自然に演奏した距離動画像を実験に用いた。なお、修飾音は演奏していない。運指については、楽譜上で指定されている場合はそれ

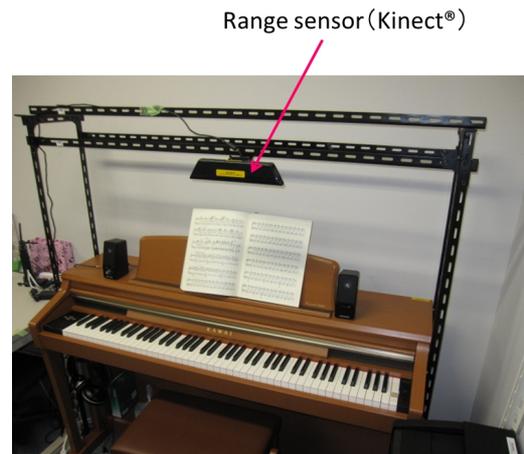


図 7 実験装置外観

を用い、指定されていない場合は自然な演奏になるように運指指定を追加し演奏した。

認識のための辞書データベースとしては、Menuet における右手パートに現れる音名は右手で、左手パートに現れる音名は左手でそれぞれ打鍵した距離画像を用いて作成した。使用した音域は、右手用データベースは D4~B5、左手用データベースは G2~E4 とした。それぞれの範囲内で、白鍵、黒鍵について距離画像を取得し、前処理を施した後に辞書データとして登録した。作成した辞書においては 1 クラスにつき 6 回ずつデータを取得した。右手、左手 DB はどちらも計 65 クラス (390 データ) を用いて実験を行った。図 8 に学習に用いた画像の例を示す。

手首位置による切り出しサイズは、予備実験によって、x 方向には手首位置から ± 100 [pixels]、y 方向に関しては手首検出位置ラインから 200 [pixels] とした。

なお、ANN による認識では、入力画像と辞書データの画像ベクトルのユークリッドノルムを計算し、ノルムが最小である辞書データを認識結果とする。

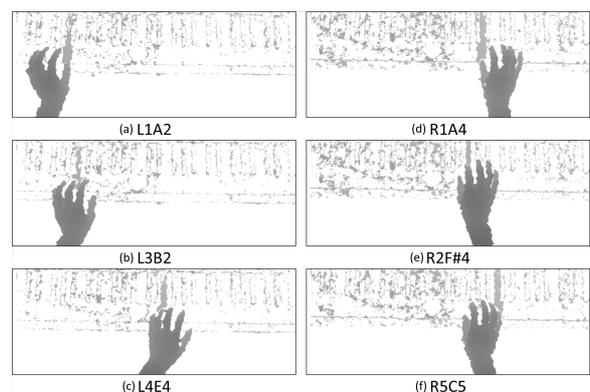


図 8 学習に用いた画像の例 (左: 左手 DB 右: 右手 DB)

次節以降の実験では、和音に関する認識性能の評価実験、打鍵持続時間に関する認識性能の評価実験、指くぐりに関する実験について説明する。



図 9 Menuet(J.S.Bach) の楽譜 (一部運指指定追加)

3.3 和音の認識性能に関する認識性能の評価実験

図 9 の A-1, A-2 を用いて、和音の認識性能に関する実験を行った。図 10 は A-1 に対応しており、右手で単音、左手で単音を弾き、合計 2 和音である。図 11 は A-2 に対応しており、右手で単音、左手で 3 和音で弾き、合計 4 和音である。それぞれの上段には入力画像と正解タグ、下段に最近傍画像と認識結果のタグを示す。A-1, A-2 のどちらにおいても、正解タグと認識結果のタグが一致していることから、和音数の違いによらず、正しく認識できることが確認できた。

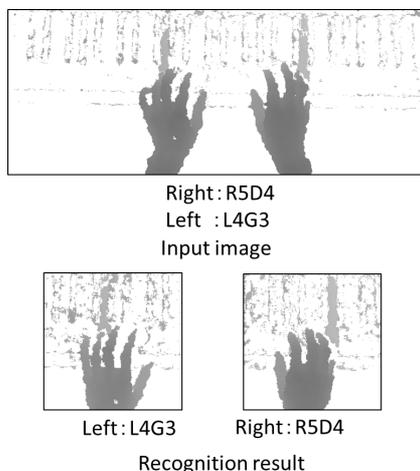
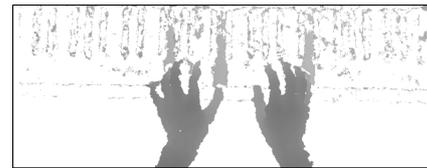
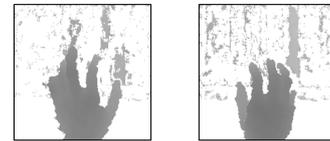


図 10 同時 2 和音



Right: R5D5
 Left : L1D4+L2B3+L4G3
 Input image



Left: L1D4+L2B3+L4G3
 Right: R5D5
 Recognition result

図 11 同時 4 和音

3.4 打鍵持続時間に関する認識性能の評価実験

1 枚当たりの認識にかかる処理時間を計測することにより、Menuet 演奏における打鍵持続時間内に認識が可能であるか実験した。入力画像 1 枚当たりの平均処理時間 46msec であった。そこで、Menuet における打鍵持続時間が最短である図 9 の B の部分 (8 分音符) に着目した。一般的な Menuet のテンポが Allegretto (108BPM) であるため、8 分音符の打鍵持続時間は 280msec である。提案手法の処理時間は 46msec であるため、十分適用可能な処理時間を達成したことを確認した。なお、16 分音符にも適用可能である。

3.5 指くぐりに関する実験

指くぐりとは、広い音域を弾くために、親指を他の指の下をくぐらせて弾くテクニックのことである。図 9 において、C では右手で F#4, G 4, A4 の 3 音が連続して弾かれる。図 9 における C の部分では、指くぐりが発生する。C の部分による指くぐりの様子を、距離画像と濃淡画像を用いて図 12 に示す。このように、同図 (b) からわかるように、濃淡画像であっても打鍵指の推定が困難であることが分かる。

次に、指くぐりが発生する場合の認識結果を図 13 の (a),(b),(c) に示す。

図 12 では、F#4 を弾く際に、曲全体の弾きやすさを考慮することによって、R1 よりも R2 で弾くほうが自然な運指になるため、指くぐりが発生している。

以上のことから、指くぐりが発生している場合のみに誤認識が起こっていることがわかる。対策案としては、指先を検出し、見えていない指を推定することで誤認識を低減することが考えられる。

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

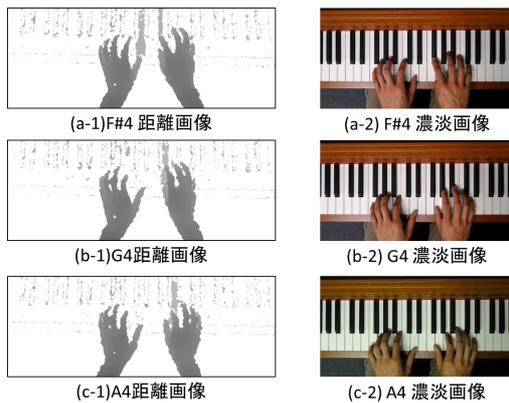
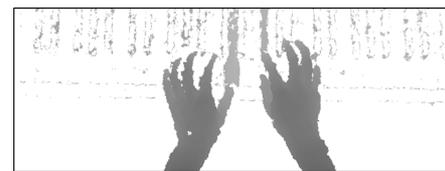


図 12 指くぐりの距離画像と濃淡画像

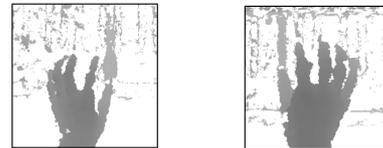
く運指情報を推定する手法を提案した。その際、ピアノから得られた音名情報と手首位置座標をもとに、効率よくデータベースを事前抽出することによって、認識の高速化と高信頼化を図った。楽曲 Menuet(J.S.Bach) を用いた実験により、和音に対する認識性能、打鍵持続時間に関する認識性能を明らかにした。今後の課題としては、指くぐり発生時の認識性能向上に取り組む予定である。

参考文献

- [1] 下山望：ピアノ運指法-譜例分類による-, ムジカノヴァ (1998).
- [2] E.Clarke, R.Parncutt, M.Raekallio, J.Sloboda: Talking fingers: an interview study of pianists' views on fingering, *Musicae Scientiae*, Vol.1, No.1, pp.87-107,(1997).
- [3] Melanie Hart, Robert Bosch, Elbert Tsai: Finding Optimal Piano Fingerings, *The. UMAP Journal*, Vol.21, No.2, pp.167-177(2000).
- [4] 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).
- [5] Xia Liu, Kikuo Fujimura: Hand Gesture Recognition using Depth Data, *Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, pp.529-534,(2004).
- [6] K. Oka, Y. Sato, and H. Koike: Real time fingertip tracking and gesture recognition, *Computer Graphics and Applications*, IEEE, Volume 22, Issue 6, pp.64-71,(2002).
- [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, *情報科学技術フォーラム*, pp.171-172, (2011).
- [8] 釘本望美, 山本和樹, 武田晴登, 片寄晴弘, 長田典子, 巳波弘佳: モーションキャプチャを用いたピアノ演奏動作のCG表現と音楽インタフェースへの応用, *情報処理学会研究報告*, 2007-MUS-72, pp79-84, (2007).
- [9] 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, *日本ソフトウェア科学会論文誌*, Vol.23, No.4, pp.51-59, (2006).
- [10] 竹川佳成, 寺田努, 塚本昌彦: 運指認識技術を活用したピアノ学習支援システム, *情報処理学会研究報告*, 2009-MUS-81, pp.1-6, (2009).
- [11] 藤本光一, 松尾直志, 島田伸敬, 白井良明: 輪郭部分特徴の階層構造学習による三次元手指姿勢推定の高速化, *画像の認識・理解シンポジウム*, IS3-64, (2010).
- [12] S.Arya, D.M.Mount, R.Silverman and A.Y. Wu: An optimal algorithm for approximate nearest neighbor



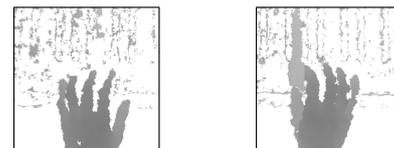
Right : R2FS4
 Left : L1D4
 Input image



Left: L1D4 Right: R1F#4
 (a) Recognition result(F#4 of Fig.9 part C)



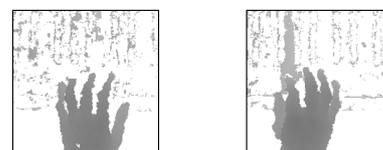
Right : R1G4
 Left : L3B3
 Input image



Left: L3B3 Right: R1G4
 (b) Recognition result(G4 of Fig.9 part C)



Right : R2A4
 Left : L3B3
 Input image



Left: L3B3 Right: R2A4
 (C) Recognition result(A4 of Fig.9 part C)

図 13 指くぐりが発生している一連の認識結果

searching, *Journal of the ACM*, Vol.45, No.6, pp.891-923,(1998).

- [13] J. S. Bach: *Notenbuchein für Anna Magdalena Bach*, G. Henle Verlag.
- [14] MIDI Manufacturers Association (<http://www.midi.org>)