

音名信号を用いた学習データ選択に基づく 距離画像ベースのマーカレスピアノ運指認識手法の提案

岡 明 也^{†1} 橋 本 学^{†1}

ピアノ運指の認識は、ピアノ初心者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が付与された大量の手指の距離画像特徴ベクトルを登録しておく。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、音名を利用して打鍵指のみを認識した場合には 100%、音名と打鍵指両方を同時に認識させた場合でも約 85% の認識率、さらに実際の楽曲による実験では約 80% の認識率となり、本手法の有効性が示された。

A Proposal of Marker-less Piano Fingering Recognition based on Training Data Selection using the Note Name

AKIYA OKA^{†1} and MANABU HASHIMOTO^{†1}

Automatic recognition of the piano fingering is an important technique for skill up of beginners or recording and archiving of performance by skillful artists. In this paper, we propose a method for recognizing piano fingering without any markers by utilizing sequential range images. In the phase of learning, a lot of range feature vectors are enrolled with correct fingering information as the dictionary data. Next, in the phase of recognition, an acquired unknown range image is compared to all data in the dictionary, and fingering pattern will be decided. Here, the note name from an electronic piano keyboard is utilized for efficient search of the best-match data in high-dimensional feature space. The ANN (Approximated Nearest Neighbor) searching algorithm is used. As for experimental results, we have proved that the recognition success rate was 100% when the note name signal was used. In another experiment using actual easy and short piano music for beginners, 80% recognition success rate has been achieved.

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、運指の間違い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、ジェスチャー認識等の多くの研究があるが⁵⁾、ピアノ演奏における手指の動きは繊細で多様なために研究例は少ない。例えば、手指の肌色領域を抽出する手法⁶⁾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⁷⁾、手指になんらかのマーカの装着が必要なために、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⁸⁾⁹⁾、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。一方、手指認識に関する他の技術としては、手話認識を目的とした一連の研究事例がある。たとえば文献¹⁰⁾では、比較的動作が大きい手話認識の用途では高い性能を実現しているが、領域内の画像濃度の勾配を表現した HoG (Histograms

^{†1} 中京大学 情報理工学部

School of Information Science and Technology, Chukyo University

of Oriented Gradients) 特徴を利用しているために情報取得には一定面積の領域が必要となり、この技術を指先の正確な打鍵位置計測に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情報に関しては、電子ピアノを使用した場合には鍵盤内のセンサの利用や、あるいはマイクで集音された音響解析によってある程度の精度で取得が期待できる。したがって、本研究では音名が利用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。

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連続的な距離画像群を取得して学習クラスと定義し、これ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、入力された未知入力パターン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、未知入力パターンが属するクラス(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未知入力パターン両方に関して、前処理として時間的な画素変動の抑制処理や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くなっても高速な識別が実現する。提案手法により、マーカが不要なピアノ運指認識の実現が期待できる。

以下、2章では運指情報の表現方法を定義し、3章では本研究で提案する手法について述べる。4章では実験準備について説明し、5章では提案手法の評価に関する一連の実験を実施した結果を示すとともに内容を詳細に分析する。最後に6章にて本研究の成果をまとめる。

2. 運指情報の表現方法

本研究では、運指情報を押下されたキーの音名と打鍵した指の組み合わせとして定義する。したがって、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1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

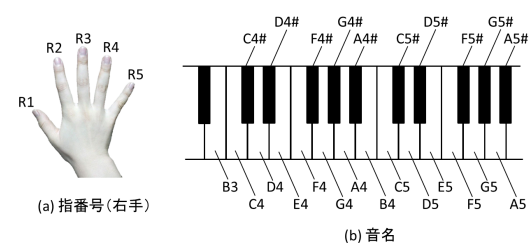


図1 各指とキーに設定した識別記号

表1 運指情報例

運指情報の表現	意味
R1C4	R1(右手の親指)でC4(ド)を押す。
R1C4 + R3E4	R1(右手親指)でC4を、R3(右手中指)でE4を押下する(和音)。

わせた識別記号をピアノ鍵盤の各キーに付与する。

例えば、識別記号が“R1C4”ならば、R1(右手親指)でC4(ド)のキーを押していることを表す。また“R1C4+R3E4”ならば、R1(右手親指)でC4(ド)のキーを押す、同時に右手のR3(右手中指)でE4(ミ)のキーを押していることを表す。

3. 提案手法

3.1 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2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が取得され、次節で述べる一連の前処理が施される。これに教師信号として運指情報が付与され、辞書データベースに格納される。

ここで、教師信号として付与される運指情報のうち、押下したキーの音名についてはMIDI仕様のキーボードを利用することによって簡単に取得できる。MIDI (Musical Instrument Digital Interface) 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。また、各々の信号には打鍵されたキーの音名情報も含まれ、C, C#, D, D#, といった半階音に0~127のノートナンバーを割り当てて音名を表して

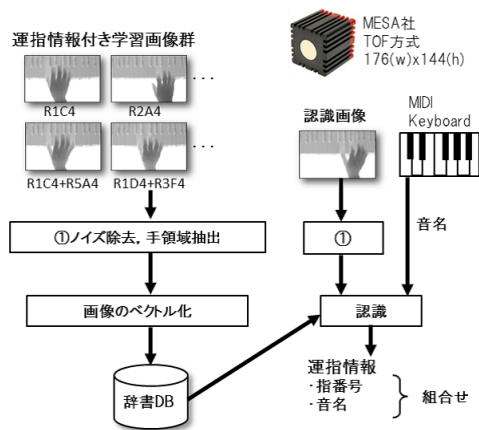


図 2 提案手法の学習・認識のブロック図

いる。例えば、88 鍵からなるピアノの全音域は中央 C=60 を中心に 21~108 の範囲からなるノートナンバーで表現される。

以上のように、学習時には演奏者があらかじめ指定された運指によってさまざまなキーを打鍵することによって、上述の MIDI 信号から得られる打鍵キー名と、打鍵した指の種類が結びつけられ、教師信号として学習画像とともに記録されることになる。

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を用いて学習データベースの選択を行う。特徴空間内で入力された認識画像と選択された学習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。

3.2 前処理

本研究では、レンジセンサを用いて距離画像を取得する。同一座標で計測される距離の値に時間方向に不安定な点が現れることがあるため、時間方向に連続した複数の距離画像の同一座標における濃度値のメディアンを計算することで不安定画素値を除外する。また、このようにして得られた画像に対してさらに空間的にメディアンフィルタを作用させることでノイズ低減を行う。

次に手指領域を抽出する。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背景の距離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着目すると、式 (1) のように各座標 について入力画像 から背景画像 の差分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。なお、 t_h は距離画像差分値のしきい値であり、実験的に求める。図 3 に手領域抽出例を示す。

$$S(i, j) = \begin{cases} I(i, j) & \text{when } |I(i, j) - B(i, j)| > t_h \\ 255 & \text{otherwise} \end{cases} \quad (1)$$



図 3 背景差分による手領域抽出例

3.3 MIDI 信号の利用による辞書作成システム

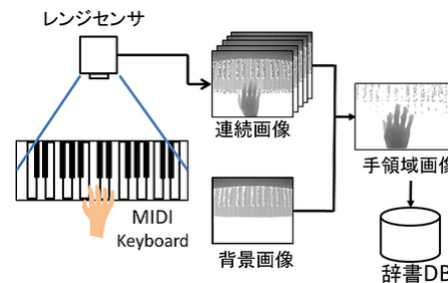


図 4 MIDI 信号を利用した辞書データベース作成システム

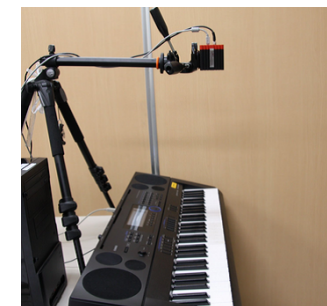


図 5 実験風景

3.1 節で説明したように、辞書データベース作成の際に MIDI キーボードから PC へ出力される MIDI 信号のノートオン、ノートオフイベントと、レンジセンサの距離計測トリガーを連動させる。図 4 に MIDI 信号を利用した辞書データベース作成システムの図を、図 5

に実験の様子を写真で示す。これにより、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。打鍵する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を統合し、距離画像の属性として付与し、辞書データとして画像ベクトルを格納する。なお距離情報については 3.2 節で述べた前処理を施しておく。

以上の処理により、運指情報を持った辞書データベースが作成される。

3.4 ANN を用いた運指認識

提案手法では音名と打鍵指の組み合わせで 1 つのクラスを形成するが、クラス数が非常に多い。入力距離画像は辞書データベース中の各クラスとの距離を計算することによって最近傍探索されるが、クラス数が多いことから処理時間がかかる懸念がある。そこで、本研究では ANN を利用して高速化を図る。

ANN は近似最近傍探索アルゴリズム¹¹⁾であり、各学習サンプル間の関係を木構造で表現しておくことによって、すべての学習データと直接的な距離計算をおこなうことなく、高速に、かつ効率よく最近傍データを検出することができる。木構造には多次元ベクトルを二分木で保持する kd-tree(k-dimensional tree) を用いる。kd-tree 構築の方法としては、再帰的な処理により特徴空間を分割していくことで、セルと呼ばれる同じ分割ルールによって分けられた特徴が集まる領域をまとめていく。このようにして生成されたセルを葉ノード、分割ノード、分割ルールを内部ノードとすることで木構造(kd-tree)を構築する。図 6 に kd-tree による特徴空間分割の図を示す。ノード A は特徴空間を左右に、ノード B,C は上下に分割することを表している。各ノードは特徴空間を分割することで末端ノード 1 つにつき 1 つの特徴ベクトルが対応する。

次に、作成した kd-tree を用いて認識対象画像の特徴に対して最近傍探索を行う。図 7 に ANN による近似最近傍探索の概念図を示す。入力画像として特徴ベクトル q が与えられると、まず入力画像の特徴ベクトルが特徴空間内のどのセルに属するかを kd-tree の末端ノードから求める。そして、そのセル内の特徴ベクトル p_1 との距離を計算し、その距離を半径 r とする超球を探索することで、真の最近傍点を求める。この場合、最近傍点は p_3 となる。その際、計算量を削減するパラメータ ϵ を用いることで探索する範囲を限定し、処理の高速化を図ることができる。同図の場合、近似探索を行うと p_2 と p_4 との距離計算を省くことができるが、 ϵ の値を大きくして探索する範囲を限定し過ぎると高速化とともに探索の精度が低くなるというトレードオフの関係が存在するため、 ϵ の設定には注意を要する。

以上の処理によって認識画像から抽出した特徴ベクトルが辞書データベース内のどの特徴

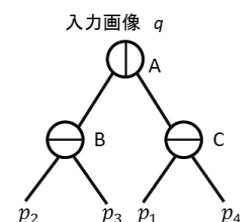


図 6 kd-tree による特徴空間分割

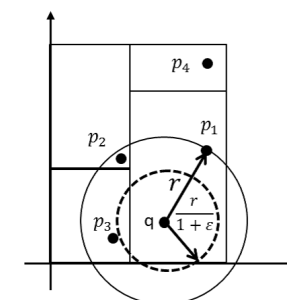


図 7 ANN による近似最近傍探索の概念図

ベクトルに近いかを調べ、多数決判定を行う。なお、ベクトル同士の距離計算にはユークリッド距離を用いる。

4. 実験準備

本研究では、ピアノ鍵盤として、カシオ社製の 61 鍵キーボード CTK-6000 を用いてデータを取得した。同製品には 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、TOF (Time of Flight) 方式の MESA 社製の SR-4000 を用いた。TOF センサとは、センサの周囲に配置された LED より照射される近赤外光が対象物体の表面で反射し、センサに帰還して観測されるまでの時間を計測することにより、物体までの距離を計測するセンサである。SR-4000 は、視野角 43.6×34.6 [deg] であり、画素数は 176×144 である。距離計測精度としては、動作範囲 0.3-5.0m において ± 1 cm、フレームレートは最大 54fps の仕様である。我々はこの TOF センサを鍵盤の上方 50cm の高さに設置した。これにより約 1 オクターブ程度の範囲の鍵盤を計測可能である。

認識のための辞書データベースとしては、ハ長調とニ長調の 2 種類を作成した。使用した音域は、ハ長調で C4~C5、ニ長調で D4~D5 の約 1 オクターブとした。その範囲内で右手に限定して、白鍵、黒鍵を含め、各種の単音、2 和音、3 和音について距離画像を取得し、前処理を施した後に辞書として登録した。ハ長調の辞書においては 1 クラスにつき 5 回、ニ長調については 3 回ずつデータを取得した。図 8 に実験に用いた画像例を示す。また、指の鉛直軸周りの回転ずれの影響を見るために、ハ長調辞書データベースの作成時には手の向きが一定になるように注意を払い、一方、ニ長調辞書データベースに関しては意識的に手の

向きを考慮せずデータを登録した。なお、ANN による識別では上位 k 位までについてクラスの多数決を行い、最も票が多くなったクラスを識別結果とするが、 k の値として、本研究では 1 および 3 とした。



図 8 学習に用いた画像の例

5. 実験結果と考察

5.1 入力音数の違いによる認識率評価結果

単音、2 和音、3 和音の入力音数と認識率の関係性を調べるために、音数ごとに分けて認識率を求めた。ハ長調データ 116 枚、ニ長調データ 88 枚を入力パターンとして、打鍵音数別に認識率を算出した結果を表 2 に示す。認識率を求めるため、近似探索における許容誤差 ϵ を 0 とした。打鍵音数などの情報が一切未知の場合は入力が 2 和音や 3 和音の方が高い認識率を示している。これは単音より 2 和音、3 和音の方が指の形に強い拘束があり、入力画像の変動が小さいことが理由であると考えられる。辞書データベースの作成の際に動きに注意を払いながら撮影したハ長調データの方が認識率が高くなっていることがわかる。さらに、MIDI 信号の利用等により音名情報をもとに当該音名に関するクラスのみを効率よく事前抽出した場合には、認識率が 100% に向上することが確認できた。

表 2 ハ長調・ニ長調テストデータによる認識成功率 [%] (辞書データ：ハ長調・ニ長調)

入力	ハ長調				ニ長調			
	音名利用しない		音名利用する		音名利用しない		音名利用する	
	k=1	k=3	k=1	k=3	k=1	k=3	k=1	k=3
単音	97.5	92.5	100.0	100.0	81.5	65.7	100.0	100.0
2 和音	98.1	96.2	100.0	100.0	89.1	86.4	100.0	100.0
3 和音	100.0	100.0	100.0	100.0	92.3	92.3	100.0	100.0

図 9 に音名情報を利用しない場合の失敗例を示す。正解運指は R4G4 であるがについては正しく認識できたが、誤って R2 で E4 を打鍵していると認識された。R2G4 の状態では R

は E4 のキー上に存在していることが原因であると推察される。このことを確認するために認識画像である R2G4 と学習画像 R2E4 との距離を計算したところ、誤認識結果のほうが特徴空間における距離が近いことが確認できた。

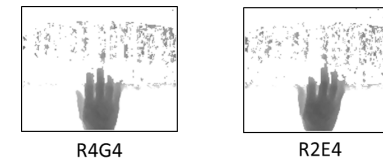


図 9 認識失敗例

5.2 実際の楽曲を用いた認識成功率評価実験

右手のみで演奏可能な 10 種類の簡単な曲を用いて運指の認識成功率を評価した。各曲中に現れるそれぞれの音（単音および 2 和音）に対して表 2 で得られた認識成功率結果を適用して、楽曲全体の認識率を計算により求めた。実験に用いた曲には、手の動きが激しくなく、親指の指くぐりなどのセルフオクルージョンの起こらないものを選んだ。なお辞書データベースには単音、2 和音、3 和音の計 88 クラス (264 データ) を用いた。

表 3 実際の楽曲における運指認識成功率 [%]

入力	ハ長調				ニ長調			
	音名利用しない		音名利用する		音名利用しない		音名利用する	
	k=1	k=3	k=1	k=3	k=1	k=3	k=1	k=3
1. ちょうちょう	100.0	100.0	100.0	100.0	72.2	72.2	100.0	100.0
2. かえるの合唱	100.0	100.0	100.0	100.0	86.2	79.3	100.0	100.0
3. メリーさんのひつじ	100.0	100.0	100.0	100.0	92.0	92.0	100.0	100.0
4. ロンドン橋	100.0	100.0	100.0	100.0	91.6	58.3	100.0	100.0
5. アルザスの踊り	100.0	98.0	100.0	100.0	80.7	80.7	100.0	100.0
6. マーチ	100.0	100.0	100.0	100.0	83.6	65.3	100.0	100.0
7. ゆうべの星	100.0	100.0	100.0	100.0	58.3	54.1	100.0	100.0
8. かっこう	100.0	100.0	100.0	100.0	73.0	73.0	100.0	100.0
9. ジングルベル	100.0	100.0	100.0	100.0	100.0	69.3	100.0	100.0
10. お人形の子もり歌	100.0	100.0	100.0	100.0	69.0	99.6	100.0	100.0

結果を表 3 に示す。曲番号 1~7 は文献¹²⁾ から選んだ単音のみからなる楽曲である。曲番

号 8~10 は単音と 2 和音からなる楽曲であり、曲番号 8,9 は文献¹³⁾、曲番号 10 は文献¹⁴⁾から選んだ。両手のための楽曲については、右手パートのみを使用して実験した。これらの曲はすべてハ長調であるため、二長調の辞書データベースを利用する認識実験の際には、これらの曲をすべて二長調に移調して利用した。運指については楽譜上で指定されている場合はそれを用い、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た。

表より、認識時に音名信号を利用することで大幅に認識成功率が向上したことを確認した。

5.3 各モジュールの処理時間

提案手法において、学習時の kd-tree の作成、認識時の前処理、探索処理の 3 つのモジュールにおける処理時間を計測した。ハ長調データベース (116 クラス 580 データ) を用いて 1 枚の入力画像を認識するまでの各モジュールの処理時間を計測した。結果を表 4 に示す。k-d tree の作成に多少の時間がかかっているが、学習時の処理のためオフライン処理も可能であることから実用上はあまり問題にならないと推察される。前処理と探索時間の合計としては 1 画像あたり 36.8 ミリ秒であり、今回実験に用いた程度の初心者向けのテンポの遅い曲については十分適用できると考えられる。

表 4 各モジュールに対する処理時間

各モジュール		処理時間 [ms]
学習	k-dtree の作成	150.0
認識	前処理	31.0
	探索処理	5.8

6. 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をもとに、当該音名に関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った。提案手法では、ノイズ低減のために時間的なメディアンフィルタ、空間的なメディアンフィルタを施す。また、認識に ANN を用いることにより、クラス数 88 個における識別でも処理時間 36.8 ミリ秒であることを確認できた。認識率としては事前に打鍵キーを手掛かりとする認識では 100%、事前情報を手掛かりとしない認識では約 85%であり、さらに実際の楽曲を用いた認識では約 80%の認識率となった。

今後の課題としては、認識率を向上させる必要がある。提案手法では距離画像 1 枚ごとに

運指を推定したが、楽曲演奏においては直前の数音に対する運指情報を利用して不自然な認識結果を除外したり、修正することによって認識性能を改善する予定である。

参 考 文 献

- 1) 下山望: ピアノ運指法—譜例分類による—, ムジカノーヴァ (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) 長田典子, 大城英裕, 加藤邦人, 興水大和, 佐川立昌, 藤原孝幸, 山下淳, 橋本学: メディア分野における多次元センシング技術の現状と動向, 電気学会論文誌 D 部門, Vol.131, No.4, pp.433-440(2011).
- 6) 木村慎二, 子安大士, 前川仁: ピアノ演奏の運指解析のための指先追跡, 情報科学技術フォーラム, pp.169-170 (1990).
- 7) 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 表現と音楽演奏インタフェースへの応用, 情報処理学会研究報告, 2007-MUS-72, pp.79-84 (2007).
- 8) 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).
- 9) 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).
- 10) 山田寛, 松尾直志, 島田伸敬, 白井良: 手話認識のための見えの学習による手領域検出と形状識別, 画像の認識・理解シンポジウム, pp.635-642 (2009).
- 11) S.Arya, D.M.Mount, R.Silverman and A.Y. Wu. An optimal algorithm for approximate nearest neighbor searching, *Journal of the ACM*, Vol.45, No.6, pp.891-923 (1998).
- 12) 田丸信明: ピアノの森, 学習研究社, (1989).
- 13) 私はピアニスト 1, 全音楽譜出版社 (1974).
- 14) みんなのオルガン・ピアノの本 3, ヤマハミュージックメディア (2007).