

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2020-46500
(P2020-46500A)

(43) 公開日 令和2年3月26日(2020.3.26)

(51) Int.Cl.	F I	テーマコード (参考)
G 1 O G 3/04 (2006.01)	G 1 O G 3/04	5 D 1 8 2
G 1 O H 1/00 (2006.01)	G 1 O H 1/00 Z	5 D 4 7 8

審査請求 未請求 請求項の数 20 O L (全 22 頁)

(21) 出願番号 特願2018-173676 (P2018-173676)
(22) 出願日 平成30年9月18日 (2018.9.18)

(特許庁注：以下のものは登録商標)

1. Z I G B E E

(71) 出願人 000002185
ソニー株式会社
東京都港区港南1丁目7番1号
(74) 代理人 100082762
弁理士 杉浦 正知
(74) 代理人 100123973
弁理士 杉浦 拓真
(72) 発明者 何 信瑩
東京都港区港南1丁目7番1号 ソニー株式会社内
Fターム(参考) 5D182 AC03 AD01 AD06
5D478 CC22

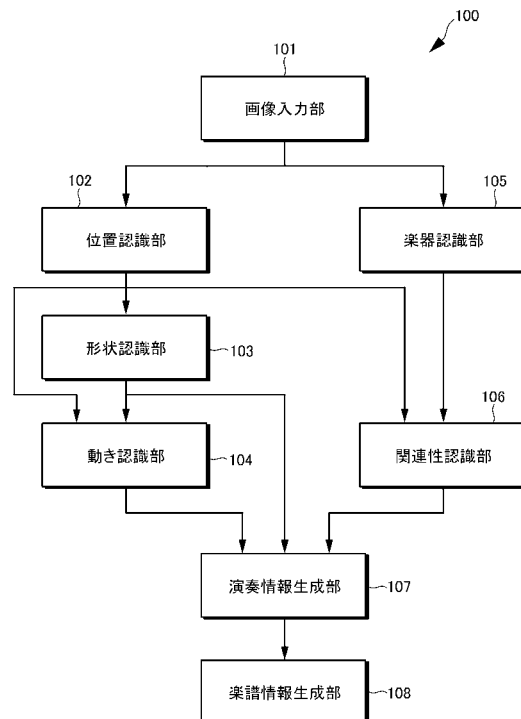
(54) 【発明の名称】 情報処理装置、情報処理方法および情報処理プログラム

(57) 【要約】 (修正有)

【課題】 画像から楽器の演奏を示す情報を取得することができる情報処理装置、情報処理方法および情報処理プログラムを提供する。

【解決手段】 情報処理装置100は、入力画像から演奏者の身体の部位の位置を認識する位置認識部と、入力画像から楽器を認識する楽器認識部と、部位の位置と楽器との関連性に基づき、演奏者による楽器の演奏を示す演奏情報を生成する演奏情報生成部とを備える。

【選択図】 図2



【特許請求の範囲】**【請求項 1】**

入力画像から演奏者の身体の部位の位置を認識する位置認識部と、
前記入力画像から楽器を認識する楽器認識部と、
前記部位の位置と前記楽器との関連性に基づき、前記演奏者による前記楽器の演奏を示す演奏情報を生成する演奏情報生成部と
を備える情報処理装置。

【請求項 2】

前記位置認識部により認識された前記部位の形状を認識する形状認識部を備え、
前記演奏情報生成部は、前記部位の形状と前記楽器の関連性に基づき前記演奏情報を生成する
請求項 1 に記載の情報処理装置。

10

【請求項 3】

前記位置認識部により認識された前記部位の動きを認識する動き認識部を備え、
前記演奏情報生成部は、前記部位の動きと前記楽器の関連性に基づき前記演奏情報を生成する
請求項 1 に記載の情報処理装置。

【請求項 4】

前記演奏情報は、前記演奏者が前記楽器を演奏している状態に対応した第 1 演奏要素を含む
請求項 1 に記載の情報処理装置。

20

【請求項 5】

前記第 1 演奏要素は、前記演奏者により演奏されている音階を含む
請求項 4 に記載の情報処理装置。

【請求項 6】

前記演奏情報は、前記演奏者が前記楽器を演奏してない状態に対応した第 2 演奏要素を含む
請求項 1 に記載の情報処理装置。

【請求項 7】

前記第 2 演奏要素は、前記演奏者により演奏されていない休みの長さを含む
請求項 6 に記載の情報処理装置。

30

【請求項 8】

前記演奏情報は、複数の前記入力画像間に跨る要素である第 3 演奏要素を含む
請求項 1 に記載の情報処理装置。

【請求項 9】

前記第 3 演奏要素は、前記演奏者により演奏されている曲のテンポを含む
請求項 8 に記載の情報処理装置。

【請求項 10】

前記演奏情報生成部は、一の前記入力画像に対応した前記演奏情報を生成する
請求項 1 に記載の情報処理装置。

40

【請求項 11】

前記演奏情報生成部は、複数の前記入力画像により構成される前記楽器の演奏の一部または全部に対応する演奏情報を生成する
請求項 1 に記載の情報処理装置。

【請求項 12】

前記関連性は、前記楽器に対する前記部位の接触位置である
請求項 1 に記載の情報処理装置。

【請求項 13】

前記関連性は、前記楽器に対する前記部位の動作の方向である
請求項 1 に記載の情報処理装置。

50

- 【請求項 14】
前記部位は前記演奏者の手である
請求項 1 に記載の情報処理装置。
- 【請求項 15】
前記演奏情報から楽譜情報を生成する楽譜情報生成部を備える
請求項 1 に記載の情報処理装置。
- 【請求項 16】
前記入力画像において前記部位の一部が映っていない場合、前記部位の一部の位置を推定し、前記演奏情報生成部は推定結果に基づき前記演奏情報を生成する
請求項 1 に記載の情報処理装置。 10
- 【請求項 17】
前記入力画像において前記楽器の一部が映っていない場合、前記演奏情報生成部は、前記部位と前記楽器の一部との前記関連性を推定し、推定結果に基づき前記演奏情報を生成する
請求項 1 に記載の情報処理装置。
- 【請求項 18】
前記演奏情報生成部は、補助情報として音情報を用いて前記演奏情報を生成する
請求項 1 に記載の情報処理装置。
- 【請求項 19】
入力画像から演奏者の身体の部位の位置を認識し、
前記入力画像から楽器を認識し、
前記部位の位置と前記楽器との関連性に基づき、前記演奏者による前記楽器の演奏を示す演奏情報を生成する
情報処理方法。 20
- 【請求項 20】
入力画像から演奏者の身体の部位の位置を認識し、
前記入力画像から楽器を認識し、
前記部位の位置と前記楽器との関連性に基づき、前記演奏者による前記楽器の演奏を示す演奏情報を生成する
情報処理方法をコンピュータに実行させる情報処理プログラム。 30
- 【発明の詳細な説明】
- 【技術分野】
- 【0001】
本技術は、情報処理装置、情報処理方法および情報処理プログラムに関する。
- 【背景技術】
- 【0002】
従来から、ダンスなどの人のパフォーマンスをデータ化するシステムが提案されている（特許文献 1）。
- 【先行技術文献】
- 【特許文献】 40
- 【0003】
【特許文献 1】特開 2016 - 24740 号公報
- 【発明の概要】
- 【発明が解決しようとする課題】
- 【0004】
特許文献 1 に記載のシステムは、3次元空間におけるパフォーマーのダンス動作が記録された譜面データを生成するものである。このような動作をデータ化する手法においては、動作の種別によってデータ化のために必要な情報や処理が異なるため、そのまま他の動作、例えば楽器演奏などに適用することは難しい。
- 【0005】 50

本技術はこのように鑑みなされたものであり、画像から楽器の演奏を示す情報を取得することができる情報処理装置、情報処理方法および情報処理プログラムを提供することを目的とする。

【課題を解決するための手段】

【0006】

上述した課題を解決するために、第1の技術は、入力画像から演奏者の身体の部位の位置を認識する位置認識部と、入力画像から楽器を認識する楽器認識部と、部位の位置と楽器との関連性に基づき、演奏者による楽器の演奏を示す演奏情報を生成する演奏情報生成部とを備える情報処理装置である。

【0007】

また、第2の技術は、入力画像から演奏者の身体の部位の位置を認識し、入力画像から楽器を認識し、部位の位置と楽器との関連性に基づき、演奏者による楽器の演奏を示す演奏情報を生成する情報処理方法である。

【0008】

さらに、第3の技術は、入力画像から演奏者の身体の部位の位置を認識し、入力画像から楽器を認識し、部位の位置と楽器との関連性に基づき、演奏者による前記楽器の演奏を示す演奏情報を生成する情報処理方法をコンピュータに実行させる情報処理プログラムである。

【図面の簡単な説明】

【0009】

【図1】端末装置の構成を示すブロック図である。

【図2】第1の実施の形態における情報処理装置の構成を示すブロック図である。

【図3】第1の実施の形態における入力画像の一例を示す図である。

【図4】演奏者の手の認識についての説明図である。

【図5】演奏者の手の認識についての説明図である。

【図6】和音（コード）の認識についての説明図である。

【図7】第1の実施の形態における部分演奏情報の生成処理を示すフローチャートである。

【図8】第1の実施の形態における複合演奏情報の生成処理を示すフローチャートである。

【図9】第2の実施の形態における入力画像の一例を示す図である。

【図10】第2の実施の形態に係る情報処理装置の構成を示すブロック図である。

【図11】第2の実施の形態における部分演奏情報の生成処理を示すフローチャートである。

【図12】第2の実施の形態における複合演奏情報の生成処理を示すフローチャートである。

【図13】第3の実施の形態における入力画像の一例を示す図である。

【図14】第3の実施の形態に係る情報処理装置の構成を示すブロック図である。

【図15】第3の実施の形態における部分演奏情報の生成処理を示すフローチャートである。

【図16】第3の実施の形態における複合演奏情報の生成処理を示すフローチャートである。

【発明を実施するための形態】

【0010】

以下、本技術の実施の形態について図面を参照しながら説明する。なお、説明は以下の順序で行う。

< 1. 第1の実施の形態 >

[1 - 1. 端末装置の構成]

[1 - 2. 情報処理装置の構成]

[1 - 3. 情報処理装置による処理]

10

20

30

40

50

[1 - 3 - 1 . 部分演奏情報の生成]

[1 - 3 - 2 . 複合演奏情報の生成]

< 2 . 第 2 の実施の形態 >

[2 - 1 . 情報処理装置の構成]

[2 - 2 . 情報処理装置の処理]

< 3 . 第 3 の実施の形態 >

[3 - 1 . 情報処理装置の構成]

[3 - 2 . 情報処理装置の処理]

< 4 . 変形例 >

【 0 0 1 1 】

10

< 1 . 第 1 の実施の形態 >

[1 - 1 . 端末装置の構成]

まず図 1 を参照して端末装置 1 0 について説明する。端末装置 1 0 は、制御部 1 1 、記憶部 1 2 、通信部 1 3 、表示部 1 4 、入力部 1 5 、カメラ部 1 6 および情報処理装置 1 0 0 を備えている。

【 0 0 1 2 】

制御部 1 1 は、C P U (Central Processing Unit)、R A M (Random Access Memory) および R O M (Read Only Memory) などから構成されている。R O M には、C P U により読み込まれ動作されるプログラムなどが記憶されている。R A M は、C P U のワークメモリとして用いられる。C P U は、R O M に記憶されたプログラムに従い様々な処理を実行してコマンドの発行を行うことによって端末装置 1 0 全体の制御を行う。

20

【 0 0 1 3 】

記憶部 1 2 は、例えば、ハードディスク、半導体メモリなどを用いた大容量記憶媒体である。記憶部 1 2 は、カメラ部 1 6 により撮影された撮影画像、撮影映像や、情報処理装置 1 0 0 により生成された演奏情報や楽譜情報、さらにコンテンツやアプリケーションなどを保存することができる。

【 0 0 1 4 】

通信部 1 3 は、他の装置、インターネットなどと通信するための通信モジュール、通信用コネクタなどである。通信部 1 3 による通信は、U S B 通信などの有線通信、Wi-Fi などの無線 L A N、Bluetooth (登録商標)、ZigBee、4 G (第 4 世代移動通信システム)、ブロードバンドなどの無線通信などなんでもよい。

30

【 0 0 1 5 】

表示部 1 4 は、例えば、L C D (Liquid Crystal Display)、P D P (Plasma Display Panel)、有機 E L (Electro Luminescence) パネルなどにより構成された表示デバイスである。表示部 1 4 には、端末装置 1 0 のユーザインターフェース、情報処理装置 1 0 0 による処理のためにユーザに提示するインターフェースなどが表示される。

【 0 0 1 6 】

入力部 1 5 は、ユーザの端末装置 1 0 に対する操作入力を受け付けるものである。入力部 1 5 に対してユーザから入力となされると、その入力に応じた入力信号が生成されて制御部 1 1 へ出力される。そして、制御部 1 1 はその入力信号に対応した演算処理、端末装置 1 0 の制御を行う。入力部 1 5 としては、表示部 1 4 と一体に構成されたタッチパネル、トラックパッドやタッチパッドと称される表示部 1 4 とは一体となっていない平板状のセンサーを指でなぞって操作するポインティングデバイス、キーボード、マウスなどがある。

40

【 0 0 1 7 】

カメラ部 1 6 は撮像素子、画像処理用 L S I などを備え、静止画像および映像の撮影が可能なカメラ機能を備えるものである。カメラ部 1 6 により撮影された静止画像または映像は情報処理装置 1 0 0 における演奏情報生成処理に用いることができる。なお、カメラ部 1 6 は端末装置 1 0 の必須の構成要素ではない。

【 0 0 1 8 】

50

情報処理装置 100 は、本技術に係る演奏情報生成処理を行うものである。情報処理装置 100 の詳細は後述する。

【0019】

端末装置 10 は以上のようにして構成されている。端末装置 10 の具体例としてはパーソナルコンピュータ、ノートパソコン、タブレット端末、スマートフォン、電子キーボード、シンセサイザー、DAW (Digital Audio Workstation) などが挙げられる。

【0020】

[1-2. 情報処理装置の構成]

次に図2を参照して情報処理装置 100 の構成について説明する。情報処理装置 100 は、画像入力部 101、位置認識部 102、形状認識部 103、動き認識部 104、楽器認識部 105、関連性認識部 106、演奏情報生成部 107、楽譜情報生成部 108 とから構成されている。

10

【0021】

画像入力部 101 には処理対象となる入力画像としての複数枚の連続する静止画像、または動画を構成する連続する複数のフレーム画像が入力される。画像入力部 101 は、入力画像を位置認識部 102 と楽器認識部 105 に供給する。本技術における処理対象である入力画像は、複数枚の連続する静止画像のそれぞれ、動画を構成する複数のフレーム画像のそれぞれである。

【0022】

なお、入力画像は端末装置 10 が備えるカメラ部 16 で撮影したものでよいし、カメラ部 16 以外のカメラで撮影して端末装置 10 を介して情報処理装置 100 に取り込んだものでよい。また、外部の別の装置から端末装置 10 を介して情報処理装置 100 に供給したものでよい。また、現実に情報処理装置 100 の使用者の目の前で行われている演奏を撮影したものでよいし、テレビ、パーソナルコンピュータなどのディスプレイに表示されている映像を撮影したものでよい。また、市販のDVD、Blue ray (登録商標) に収録された映像、インターネット上で取得可能な静止画像や映像などでもよい。すなわち、入力画像は、演奏者が演奏している様子が映っている画像であればどのようなものでよい。また、入力画像はRGB (Red, Green, Blue) 画像の他、IR 画像などでもよい。

20

【0023】

第1の実施の形態における入力画像は図3に示すように、演奏者の両手、演奏者が演奏する楽器において演奏者の手が接触する演奏のための領域 (演奏領域) の全体が写っているものである。

30

【0024】

位置認識部 102 は、入力画像からHand Pose Detection、Hand Pose Estimation、Hand segmentationなどの人体の手認識技術や、HOG (Histogram of Oriented Gradient)、SIFT (Scale Invariant Feature Transform) などの特徴点抽出方法、Boosting、SVM (Support Vector Machine) などのパターン認識による被写体認識方法、Graph Cut などによる領域抽出方法、CNN (Convolutional Neural Network) などにより、入力画像中における演奏者の身体の一部である手の3次元位置を認識する。また位置認識部 102 は、手に加えて、演奏情報生成のために必要に応じて演奏者の身体の一部としての手の指の位置、腕の位置、肘の位置なども認識する。手の3次元位置情報は形状認識部 103、動き認識部 104 および関連性認識部 106 に供給される。

40

【0025】

手の3次元位置を認識するための手の特徴点としては指先、指の関節、手首などがある。位置情報は入力画像中における演奏者の手の3次元位置を示す情報であるため、例えば、入力画像の所定の位置を原点 (0, 0, 0) とした (x, y, z) の座標で表される。連続する入力画像の番号を t (t = 1, 2, 3, ...) とし、手の特徴点を P (P = 1, 2, 3, ...) とすると、位置情報は (x_{tP}, y_{tP}, z_{tP}) という形式で表される。

【0026】

50

例えば図4 Aに示すように、入力画像 ($t = 1$) では、手の特徴点が5つ認識された場合、それらは、

特徴点 P 1 : (x_{11}, y_{11}, z_{11})

特徴点 P 2 : (x_{12}, y_{12}, z_{12})

特徴点 P 3 : (x_{13}, y_{13}, z_{13})

特徴点 P 4 : (x_{14}, y_{14}, z_{14})

特徴点 P 5 : (x_{15}, y_{15}, z_{15})

のように表される。

【0027】

また、図4 Bに示すように、入力画像 ($t = 2$) では、手の特徴点が5つ認識された場合、それらは、

特徴点 P 1 : (x_{21}, y_{21}, z_{21})

特徴点 P 2 : (x_{22}, y_{22}, z_{22})

特徴点 P 3 : (x_{23}, y_{23}, z_{23})

特徴点 P 4 : (x_{24}, y_{24}, z_{24})

特徴点 P 5 : (x_{25}, y_{25}, z_{25})

のように表される。

【0028】

なお、手の3次元位置情報はカメラ原点のグローバル座標系でもよいし、入力画像上のローカル座標系 + 奥行き情報でもよい。また、Hand Segmentationで求めた領域の重心およびDepth情報を用いて手の3次元位置を求めてもよい。

【0029】

なお、図4は説明の便宜上手の5本の各指の先端に特徴点が認識された図であるが、実際には図5の手に重畳して表された複数の黒点が示すように例えば各指の関節部分、水かき部分および手首など多数の特徴点が認識される。このように多数の特徴点を認識したほうがより正確に演奏情報を生成することができる。

【0030】

また、入力画像の一部領域を切り出した切り出し画像においては、(x, y, z)の座標系とは異なる座標系である、切り出し画像の所定の位置を原点とした(u_{tP}, v_{tP}, d_{tP})の座標で表してもよい。

【0031】

形状認識部103は、CNN、パターンマッチング、Boostingなどの技術を用いて、位置認識部102から供給された位置情報で示される手の形状を認識する。手の形状情報は動き認識部104と演奏情報生成部107に供給される。

【0032】

動き認識部104は、CNN、Hand Trackingなどの技術を用いて、位置および形状が認識された演奏者の手の動きを認識する。手の動き情報は演奏情報生成部107に供給される。手の動きは、複数の連続する入力画像のうちの一の入力画像(t)と、時系列でその入力画像(t)以降の入力画像($t+n$)との動きベクトルの変化から認識することができる。

【0033】

楽器認識部105は、CNN、パターンマッチングなどの技術を用いて、入力画像中における楽器およびその楽器において演奏者の手が接触する演奏のための領域(演奏領域)を認識するものである。演奏領域とは、例えば楽器がピアノであれば鍵盤、楽器がギターであればピックアップ部分(アコースティックギターであればサウンドホール)およびネックである。楽器認識情報は関連性認識部106に供給される。

【0034】

関連性認識部106は、CNN、パターンマッチングなどの技術を用いて演奏者の手の位置と楽器の演奏領域の関連性を認識する。関連性とは、楽器演奏のための演奏者と楽器の関連、すなわち、演奏者の手が楽器の演奏領域のどこに接触しているかを示す接触位

10

20

30

40

50

置である。また、関連性は、楽器の演奏領域に対する演奏者の手、腕、肘などの部位の動作の方向である。関連性情報は演奏情報生成部 107 に供給される。

【0035】

演奏情報生成部 107 は、CNN などの技術を用いて演奏者が演奏状態にあるか否かを認識する。そして、入力画像において演奏者が演奏している状態に基づく演奏要素（第 1 演奏要素）、演奏者が演奏していない状態に基づく演奏要素（第 2 演奏要素）、複数の入力画像に跨る演奏要素（第 3 演奏要素）とから入力画像の一枚に対応した演奏情報（部分演奏情報）を生成する。

【0036】

第 1 演奏要素は楽器によって異なるものではあるが、ピアノなどの鍵盤楽器では、音階、音の長さ、テンポ、強弱などがある。また、ギターなどの弦楽器でも同様に音階、音の長さ、音の強弱などがある。さらに、ドラムなどの打楽器では叩くドラムセットの種類、音の長さ、テンポ、強弱などがある。

【0037】

また、第 2 演奏要素としてはいずれの楽器においても、休みの長さ、などがある。第 3 演奏要素としては、テンポ、音の長さ、休みの長さ、調、音の強弱などがある。音の強弱、音の長さなどは第 1 の演奏要素でもあり、第 3 の演奏要素でもあるが、これは入力画像 1 枚で音の強弱や音の長さを推定することができる場合もあれば、推定に複数枚の入力画像を必要とする場合もあるからである。例えば、1 枚の入力画像において演奏者の指が楽器の演奏領域から大きく離れている位置にある場合はその 1 枚の入力画像から音が強いことを推定することができるが、指が楽器の演奏領域の近くで細かく動いているような場合は 1 枚の画像では強弱は推定できず、複数枚の入力画像を参照して演奏者の指の動きを認識して強弱を推定する必要がある。

【0038】

演奏者の手の位置と楽器の演奏領域との関連性情報から演奏要素を取得する方法としては、位置認識部 102 により認識された手の位置と、形状認識部 103 により認識された手の形状と、楽器認識部 105 により認識された楽器と演奏領域に基き、演奏者の手の指が楽器の演奏領域のどの鍵盤に接触しているかを認識する。それにより、その入力画像における演奏者の状態において演奏により音階のどの音を鳴らしているかを認識することができる。また、複数の音により構成されるどのような和音（コード）を鳴らしているかも認識することもできる。

【0039】

また、演奏者の指が演奏領域の同一箇所にとどのくらい接触し続けているかを認識することにより音の長さを認識することもできる。

【0040】

和音（コード）の認識は、例えば楽器がギターである場合は、図 6 A、図 6 B に示すように予め演奏情報生成部 107 に和音（コード）を演奏する場合の指の位置および形状を示すテンプレート画像を和音（コード）の種類ごとに複数保持させておく。そして、入力画像から抽出された指の位置情報、指の形状情報とテンプレート画像を比較（テンプレートマッチング）することにより指の位置および形状が最も近似する和音（コード）を決定する。

【0041】

また、和音（コード）の認識は、図 6 C に示すように予め演奏情報生成部 107 に和音（コード）を演奏する場合の指の位置を示す指の特徴点の座標情報を和音（コード）の種類ごとに複数保持させておき、その座標情報と入力画像から抽出された指の位置情報（座標情報）を比較することによっても可能である。

【0042】

また、動き認識部 104 により認識された手の動きに基づいて、一の入力画像（ t ）と、時系列でその入力画像（ t ）以降の入力画像（ $t+n$ ）とから認識することができる演奏者の手の略垂直方向の動きから演奏しているか否か、演奏の強弱、テンポなどを認識す

10

20

30

40

50

ることができる。

【0043】

この場合の略垂直方向とは、楽器がピアノの場合、鍵盤が並ぶ方向に対して略垂直の方向である。演奏者が演奏しているか否かは楽器の鍵盤に手が離れているか否かに基づいて判断することができる。また、演奏の強弱は略垂直方向における手の位置（手の高さ）から判断することができる。例えば、手が鍵盤から垂直方向に離れているほど音が強く、手が鍵盤に垂直方向に近づいているほど音が弱いと判断することができる。また、手の垂直方向における規則的な上下動作の時間間隔から曲のテンポを認識することができる。このように曲のテンポや音の長さなど時間に関連する演奏要素を認識するためには映像を構成するフレームレートと実時間に対応付けて、演奏者の規則的な動きの実時間での動作間隔と映像の再生時間とから求めることができる。

10

【0044】

また、同様に一の入力画像（ t ）と、時系列でその入力画像（ t ）以降の入力画像（ $t+n$ ）とから認識することができる演奏者の手の略水平方向の動きから音階を認識することができる。この場合の略水平方向とは、楽器がピアノの場合、鍵盤が並ぶ方向に対して略水平の方向である。具体的にはピアノに対する手の略水平方向の位置が変わることにより、ピアノの鍵盤のどの領域を演奏しているかがわかり、それにより音域、オクターブの変化など演奏されている音階を認識することができる。

【0045】

手の略垂直方向の動きと略水平方向の動きは、複数の連続する入力画像のうちの一の入力画像（ t ）と、時系列でその入力画像（ t ）以降の入力画像（ $t+n$ ）との動きベクトルの変化から認識することができる。

20

【0046】

第3演奏要素は複数の入力画像に跨った演奏者の指や腕の変化に基づいて生成することができる。例えば、手が複数の入力画像に跨って鍵盤から垂直方向に離れている時間が長いほど次に鳴らされる音が強いとして第3演奏要素とすることができる。また、複数の入力画像跨る手の垂直方向における上下動作の時間間隔から曲のテンポを認識して第3演奏要素とすることができる。

【0047】

さらに演奏情報生成部107は、複数の入力画像のそれぞれに対応した部分演奏情報を時系列に従って?いでいくことにより、それら複数の入力画像により構成されるフレーズ、曲の一部または全部の複合演奏情報を生成する。フレーズや曲の一部の複合演奏情報とは、1または複数の小節単位での演奏情報である。

30

【0048】

部分演奏情報および複合演奏情報は、五線譜で記された楽譜に限らず、その情報に基づいて演奏者、コンピュータ、音楽演奏用ソフトウェア、音楽作成用ソフトウェアなどが楽曲を再現することができればどのような形式の情報でもよい。例えば、MIDI（Musical Instrument Digital Interface）形式の情報やプログラミング形式の情報、音楽演奏/制作用ソフトウェア独自のフォーマットの情報などでもよい。

【0049】

楽譜情報生成部108は、演奏情報生成部107から部分演奏情報が供給された場合には入力画像一枚に対応する部分楽譜情報を生成する。また、演奏情報生成部107から複合演奏情報が供給された場合には複数の入力画像により構成されるフレーズ、曲の一部または全部の楽譜情報である複合楽譜情報を生成する。ここでいう楽譜とは五線譜で記された楽譜であり、楽譜情報を構成する情報としては、音符、休符、拍子記号、テンポ、臨時記号、調号、強弱などがある。臨時記号情報は演奏者が演奏している状態に基づく第1演奏要素である、演奏されている音階と、複数の入力画像に跨る第3演奏要素である調とから導き出すことができる。

40

【0050】

情報処理装置100は以上のようにして構成されている。情報処理装置100はプログ

50

ラムで構成され、そのプログラムは予め端末装置 10 にインストールされていてもよいし、ダウンロード、記憶媒体などで配布されて、ユーザが自ら端末装置 10 にインストールするようにしてもよい。また、情報処理装置 100 は、プログラムによって実現されるのみでなく、その機能を有するハードウェアによる専用の装置、回路などを組み合わせて実現されてもよい。

【0051】

[1 - 3 . 情報処理装置による処理]

[1 - 3 - 1 . 部分演奏情報の生成]

次に図 7 のフローチャートを参照して情報処理装置 100 における処理の流れについて説明する。図 7 のフローチャートの処理は、入力画像一枚に対応した部分演奏情報を生成するものである。なお、上述したように入力画像の一枚とは、複数枚の連続する静止画像のうち一枚または、動画を構成する複数のフレーム画像のうち一枚である。

10

【0052】

まずステップ S 101 で、画像入力部 101 に対して入力画像が入力される。この入力画像は一枚の静止画像またはフレーム画像でもよいし、連続する複数の静止画像でもよいし、動画を構成する連続する複数のフレーム画像でもよい。複数の入力画像が入力されると、以下のステップ S 102 以降の処理は、まず ($t = 1$) の一番目の入力画像に対して行われる。また、連続する複数の静止画像または動画を構成する連続する複数のフレーム画像が入力された場合、どの入力画像の部分演奏情報を生成するかをユーザが選択できるようにしてもよい。

20

【0053】

次にステップ S 102 で位置認識部 102 により入力画像中における演奏者の手の 3 次元位置が認識され、手の位置情報が形状認識部 103、動き認識部 104 および関連性認識部 106 に供給される。

【0054】

次にステップ S 103 で、形状認識部 103 により入力画像中において位置が認識された手の形状が認識される。手の形状情報は動き認識部 104 と演奏情報生成部 107 に供給される。さらにステップ S 104 で、動き認識部 104 により、位置および形状が認識された手の動きが認識される。手の動き情報は演奏情報生成部 107 に供給される。

30

【0055】

次にステップ S 105 で楽器認識部 105 により入力画像中における楽器および演奏領域が認識される。楽器情報および演奏領域情報は関連性認識部 106 に供給される。なお、ステップ S 102 乃至ステップ S 104 における演奏者の手の位置、形状、動きの認識処理とステップ S 105 における楽器および演奏領域の認識は並行して行うようにしてもよいし、楽器および演奏領域の認識を先に行ってもよい。

【0056】

次にステップ S 106 で関連性認識部 106 により手の各指とそれに対応する楽器の演奏領域の位置の関連性が認識される。関連性とは演奏者の手が楽器の演奏領域のどこに位置しているかを示すものであり、関連性情報は演奏情報生成部 107 に供給される。

【0057】

次にステップ S 107 で演奏情報生成部 107 は手の動き情報および関連性情報から演奏者が入力画像において楽器を演奏している状態であるか否かを判定する。

40

【0058】

判定の結果、演奏者が演奏している場合、処理はステップ S 108 からステップ S 109 に進む (ステップ S 108 の Yes)。そしてステップ S 109 で演奏情報生成部 107 は手の 3 次元位置情報、手の形状情報、手の動き情報、関連性情報とから第 1 演奏要素を生成する。

【0059】

一方、ステップ S 107 での判定の結果、演奏者が演奏していない場合、処理はステップ S 108 からステップ S 110 に進む (ステップ S 108 の No)。そしてステップ S

50

110で演奏情報生成部107は第2演奏要素を生成する。

【0060】

次にステップS111で演奏情報生成部107は、第1演奏要素または第2演奏要素から入力画像に対応した部分演奏情報を生成する。そしてステップS112でその部分演奏情報を出力する。

【0061】

出力された部分演奏情報は端末装置10の表示部14において表示したり、端末装置10が備える音楽演奏用ソフトウェア、音楽制作用ソフトウェアなどにおいて使用可能である。また、ユーザ、演奏者などからの要求に応じて楽譜情報生成部108によって演奏情報に基づいて楽譜情報を生成してもよい。また、部分演奏情報を端末装置10の記憶部12に保存しておき、必要に応じて記憶部12から読み出して使用することも可能である。

10

【0062】

以上のようにして入力画像に対する演奏情報生成処理が行われる。

【0063】

[1-3-2. 複合演奏情報の生成]

次に図8のフローチャートについて説明する。図8のフローチャートの処理は、複数の入力画像により構成されるフレーズ、曲の一部または全部の演奏情報である複合演奏情報を生成する処理である。

【0064】

まずステップS101で、画像入力部101に対して入力画像として連続する複数の静止画像または動画を構成する連続する複数のフレーム画像が入力される。複数の入力画像が入力されると以下のステップS102以降の処理はまず入力画像($t=1$)の一番目の入力画像に対して行われる。また、連続する複数の静止画像または動画を構成する連続する複数のフレーム画像が入力された場合、どの入力画像から処理を開始するかをユーザが選択できるようにしてもよい。

20

【0065】

ステップS101からステップS111までの処理は図7のフローチャートと同様であるため、説明を省略する。

【0066】

ステップS111の後、次にステップS121で、演奏情報生成部107は複数の入力画像間に跨る演奏要素である第3演奏要素があるか否かが判定する。複数の入力画像間に跨る第3演奏要素があるか否かは、以下のように判断できる。例えば音の強弱(大きさ)の場合、現在処理中の入力画像(t)において認識された演奏の強弱が一つ前の入力画像である入力画像($t-1$)で認識された強弱よりも強くなる場合、入力画像($t-1$)から入力画像(t)まで、「だんだん強く」という演奏要素が導き出せる。また、同様に、例えば、入力画像($t+1$)の音程において認識された演奏の強弱が入力画像(t)の強弱より大きい場合は、入力画像($t-1$)、入力画像(t)、入力画像($t+1$)とも「だんだん強く」という演奏要素が導き出せる。このように、処理対象である複数の入力画像それぞれの状態により、現在の入力画像における演奏要素に基づいて過去の入力画像における演奏要素が認識される場合「フレーム間に跨る演奏要素である第3演奏要素がある」と判断することができる。

30

40

【0067】

複数の入力画像間に跨る第3演奏要素がある場合、処理はステップS122に進み(ステップS121のYes)、ステップS111で生成した部分演奏情報に第3演奏要素を付加することにより部分演奏情報を更新する。そして処理はステップS122からステップS123に進む。なお、第3演奏要素は部分演奏情報において第1演奏要素、第2演奏要素と同様に部分演奏情報の構成要素としてもよいし、部分演奏情報とは別情報としたまま紐付けにより対応付けてもよい。

【0068】

一方、ステップS121で複数の画像間に跨る第3演奏要素がない場合処理はステップ

50

S 1 2 3 に進む (ステップ S 1 2 1 の N o)。

【 0 0 6 9 】

次にステップ S 1 2 3 で処理対象である次の入力画像があるか否かが判定される。ステップ S 1 0 1 で画像入力部 1 0 1 に対して入力された、連続する複数の静止画像または動画を構成する連続する複数のフレーム画像にまだ未処理の画像がある場合には次の入力画像があるとして処理はステップ S 1 0 2 に戻る (ステップ S 1 2 3 の Y e s)。そして、時系列で次の順の入力画像 (フレーム画像である場合には次のフレーム番号の画像) に対してステップ S 1 0 2 乃至ステップ S 1 2 3 の処理が行われる。そして、入力された全ての入力画像のそれぞれに対して処理が行われるまでステップ S 1 0 2 乃至ステップ S 1 2 3 が繰り返される。

10

【 0 0 7 0 】

ステップ S 1 2 3 で処理対象の画像がない場合、処理はステップ S 1 2 4 に進む (ステップ S 1 2 3 の N o)。

【 0 0 7 1 】

次にステップ S 1 2 4 で演奏情報生成部 1 0 7 は、複数の入力画像のそれぞれに対応した部分演奏情報を時系列に従ってつないでいくことにより、それら複数の入力画像により構成されるフレーズ、曲の一部または全部の複合演奏情報を生成する。

【 0 0 7 2 】

次にステップ S 1 2 5 で、演奏情報生成部 1 0 7 は複合演奏情報を出力する。出力された複合演奏情報は端末装置 1 0 の表示部 1 4 において表示したり、端末装置 1 0 が備える音楽演奏用ソフトウェア、音楽制作用ソフトウェアなどにおいて使用可能である。また、ユーザ、演奏者などからの要求に応じて楽譜情報生成部 1 0 8 が複合演奏情報に基づいて複合楽譜情報を生成してもよい。また、複合演奏情報を出力する際に部分演奏情報も出力してもよい。

20

【 0 0 7 3 】

以上のようにして第 1 の実施の形態における処理が行われる。本技術の第 1 の実施の形態によれば、複数枚の連続する静止画像または動画を構成する複数のフレーム画像に基づいて演奏情報と楽譜情報を生成することができる。

これにより、専門的な知識のない人でも手軽に演奏情報、楽譜情報を得ることができる。また、例えば、音声がない映像データ、音声が悪化 / 破損している映像データなどに基づいても演奏情報と楽譜情報を生成することができる。また、音声を出力することができない環境においても映像データのみに基づいて演奏情報を生成することができる。

30

【 0 0 7 4 】

なお、第 1 の実施の形態において演奏情報を生成するための入力画像は、例えば楽器がピアノの場合には、ピアノの演奏領域である鍵盤と演奏者の両手を認識することができる上方から撮影したものが好ましい。楽器がギターの場合にはギターの演奏領域であるピックアップ部分 (アコースティックギターであればサウンドホール) およびネックと演奏領域の両手を認識することができる正面から撮影したものが好ましい。

【 0 0 7 5 】

本技術は、自分または自分以外の他の演奏者の即興演奏の楽譜化、楽器練習の楽譜化、好きなアーティスト曲を演奏するための楽譜作成、作曲、編曲などの用途に用いることができる。また、作曲、編曲の際には、楽器でいろいろな演奏、フレーズなどを試し、必要な演奏パターンまたは全ての演奏パターンを用意に演奏情報、楽譜情報として得ることができる。また、「楽譜を書いて、楽器で演奏してみる」、または「楽器で演奏してみて、良かったら楽譜を書く」の繰り返し作業が必要なくなる。

40

【 0 0 7 6 】

< 2 . 第 2 の実施の形態 >

[2 - 1 . 情報処理装置の構成]

次に本技術の第 2 の実施の形態について説明する。第 2 の実施の形態は図 9 に示すように、入力画像において演奏者の身体の一部が遮蔽されて隠れているまたは

50

写っていない場合において演奏情報の生成を行うものである。図9においては演奏者の左手の一部が隠れている。なお、情報処理装置100が動作する端末装置10の構成は第1の実施の形態と同様であるためその説明を省略する。

【0077】

図10に示すように情報処理装置200は、画像入力部101、センサ情報取得部201、第1位置認識部202、第2位置認識部203、形状認識部103、動き認識部104、楽器認識部105、関連性認識部106、演奏情報生成部107、楽譜情報生成部108とから構成されている。画像入力部101、形状認識部103、動き認識部104、楽器認識部105、関連性認識部106、演奏情報生成部107、楽譜情報生成部108は第1の実施の形態と同様のものである。

10

【0078】

センサ情報取得部201は端末装置10が備える、または端末装置10に接続された外部のセンサで取得されたセンサ情報を取得して第2位置認識部203に供給するものである。センサとしては、マイクロホン、圧力センサ、動きセンサなどがある。

【0079】

第1位置認識部202は入力画像中において隠れていない演奏者の手の位置を認識するものであり、第1の実施の形態における位置認識部102と同様のものである。

【0080】

第1位置認識部202は第1の実施の形態における位置認識部102と同様に、入力画像からHand Pose Detection、Hand Pose Estimationなどと称される人体の手認識技術やHOG、SIFTなどの特徴点抽出方法、Boosting、SVMなどのパターン認識による被写体認識方法、Graph Cutなどによる領域抽出方法、CNNなどにより、入力画像中における演奏者の身体の一部である手の3次元位置を認識する。

20

【0081】

第2位置認識部203は、入力画像中において遮蔽されることによって一部が隠れている演奏者の手の3次元位置を補助情報を用いて認識するものである。補助情報としては、センサ情報取得部201から供給されるセンサ情報などがある。センサ情報としては、マイクロホンで集音される演奏の音、手または指が楽器を押圧する力を示す圧力センサ情報、演奏者の腕/手/指の動きを示す動きセンサ情報などがある。さらに補助情報としては、第1位置認識部202と同様の手法を用いて認識した演奏者の腕および/または肘の位置/形状/動き情報などもある。

30

【0082】

例えば、演奏者の腕および肘の位置、形状の情報から演奏者の肘から先の腕の先端にある手(隠れている手)が楽器の演奏領域のどこに位置しているかを推定して認識することができる。

【0083】

第1位置認識部202および第2位置認識部203により取得された位置情報は3次元位置を示す情報であるため、例えば、入力画像の所定の位置を原点とした(x, y, z)の座標で表される。また、入力画像の一部領域を切り出した切り出し画像においては、切り出し画像の所定の位置を原点とした(u, v, d)の座標で表される。この点は第1の実施の形態と同様である。位置情報は形状認識部103および関連性認識部106に供給される。

40

【0084】

第2の実施の形態における情報処理装置200は以上のように構成されている。

【0085】

[2-2. 情報処理装置の処理]

次に第2の実施の形態における情報処理装置200の処理の流れについて説明する。図11のフローチャートは第1の実施の形態で説明した、一つの入力画像に対応する部分演奏情報を生成するための処理に対応したものである。

【0086】

50

まずステップ S 1 0 1 で、画像入力部 1 0 1 に対して入力画像が入力されると、次にステップ S 2 0 1 で入力画像において手の一部が隠れているか否かが判定される。これは、例えば、第 1 位置認識部 2 0 2 において 2 つの手の全体が認識された否かに基づいて判定することができる。

【 0 0 8 7 】

手の一部が隠れている場合、処理はステップ S 2 0 2 に進み（ステップ S 2 0 1 の Yes）、第 2 位置認識部 2 0 3 により補助情報を用いて一部が隠れている演奏者の手が認識される。

【 0 0 8 8 】

一方、手の一部が隠れていない場合処理はステップ S 1 0 3 に進み、第 1 位置認識部 2 0 2 により演奏者の手が認識される。

【 0 0 8 9 】

これ以降の処理は第 1 の実施の形態におけるものと同様である。

【 0 0 9 0 】

また、図 1 2 のフローチャートに示すように、複数の入力画像により構成されるフレーズ、曲の一部または全部の複合演奏情報を生成する処理においても図 1 1 のフローチャートにおけるステップ S 2 0 1 とステップ S 2 0 2 と同様の処理が行われる。

【 0 0 9 1 】

この第 2 の実施の形態によれば、入力画像において演奏者の手の一部が隠れていても第 1 の実施の形態と同様に演奏情報、楽譜情報の生成を行うことができる。

【 0 0 9 2 】

< 3 . 第 3 の実施の形態 >

[3 - 1 . 情報処理装置の構成]

次に本技術の第 3 の実施の形態について説明する。第 3 の実施の形態は図 1 3 に示すように、入力画像において楽器の一部が隠れているまたは映っていない場合において演奏情報の生成を行うものである。図 1 3 においては、楽器であるピアノの鍵盤の一部のみが映っており、鍵盤の一部が入力画像の画角外に存在している。なお、情報処理装置 3 0 0 が動作する端末装置 1 0 の構成は第 1 の実施の形態と同様であるためその説明を省略する。

【 0 0 9 3 】

図 1 4 に示すように、情報処理装置 3 0 0 は、画像入力部 1 0 1、センサ情報取得部 3 0 1、位置認識部 1 0 2、形状認識部 1 0 3、動き認識部 1 0 4、楽器認識部 1 0 5、関連性認識部 1 0 6、演奏情報生成部 1 0 7、楽譜情報生成部 1 0 8 とから構成されている。画像入力部 1 0 1、位置認識部 1 0 2、形状認識部 1 0 3、動き認識部 1 0 4、演奏情報生成部 1 0 7、楽譜情報生成部 1 0 8 は第 1 の実施の形態と同様のものである。

【 0 0 9 4 】

センサ情報取得部 2 0 1 は、端末装置 1 0 が備える、または端末装置 1 0 に接続された外部のセンサで取得されたセンサ情報を取得して演奏情報生成部 1 0 7 に供給するものである。センサとしては、マイクロホン、圧力センサ、動きセンサなどがある。

【 0 0 9 5 】

楽器認識部 1 0 5 は、CNN、パターンマッチング、テンプレートマッチングなどの技術を用いて、入力画像中における楽器およびその楽器において演奏者の手が接触する演奏のための領域（演奏領域）を認識するものである。そこで、例えば、テンプレートマッチングで楽器の一部分のみがテンプレートと一致するような場合、認識された楽器は一部分が隠れているまたは映っていないと判断する。入力画像に楽器の一部分しか映っていないことを示す情報と共に楽器認識情報は関連性認識部 1 0 6 に供給される。

【 0 0 9 6 】

関連性認識部 1 0 6 は、CNN、パターンマッチングなどの技術を用いて演奏者の手の位置と楽器の演奏領域の関連性を認識する。関連性とは、演奏者の手が楽器の演奏領域のどこに接触しているかを示す接触位置である。また、関連性は、楽器の演奏領域に対する演奏者の手の動作の方向である。関連性認識部 1 0 6 は、入力画像中において楽器の一部

10

20

30

40

50

しか映っていない場合、手の位置情報、手の形状情報、楽器（例えばピアノ）の演奏領域である鍵盤が並ぶ方向に対する略水平方向における腕／肘の開き具合の角度、腕の動きから指が接触している演奏領域を推定ことにより演奏者の指と楽器の演奏領域の関連性を認識する。関連性情報は演奏情報生成部 107 に供給される。

【0097】

演奏情報生成部 107 は補助情報としてセンサ情報を用いて指が接触している鍵盤を推定する。センサ情報としては、マイクロホンで集音される演奏の音、手または指が楽器を押圧する力を示す圧力センサ情報、演奏者の腕／手／指の動きを示す動きセンサ情報などがある。さらに関連性認識部 106 は、複数の入力画像において楽器全体が写っている入力画像がある場合、その入力画像と腕、手の動き情報から指が接触している鍵盤を推定する

10

【0098】

このように指が接触している鍵盤を推定することによりその推定結果から第 1 の実施の形態と同様に第 1 演奏要素、第 2 演奏要素、第 3 演奏要素を生成することができる。

【0099】

第 3 の実施の形態に係る情報処理装置 300 は以上のように構成されている。

【0100】

[3 - 2 . 情報処理装置の処理]

次に第 3 の実施の形態における情報処理装置 300 の処理の流れについて説明する。図 15 のフローチャートは第 1 の実施の形態で説明した、一つの入力画像に対応する演奏情報および楽譜情報を生成するための処理に対応したものである。

20

【0101】

ステップ S 101 乃至ステップ S 105 は第 1 の実施の形態における処理と同様である。

【0102】

ステップ S 301 で、関連性認識部 106 は入力画像において楽器の演奏領域全体が映っているかを判定し、楽器の演奏領域全体が映っている場合処理はステップ S 106 に進む（ステップ S 301 の Yes）。そして、ステップ S 106 乃至ステップ S 112 の処理が第 1 の実施の形態と同様に行われる。

【0103】

一方、入力画像に楽器の演奏領域全体が映ってはいない場合、処理はステップ S 302 に進む（ステップ S 301 の No）。そしてステップ S 302 で関連性認識部 106 により、手の位置情報、センサ情報などを用いて関連性を推定する。

30

【0104】

その後はステップ S 106 乃至ステップ S 112 の処理が第 1 の実施の形態と同様に行われて、部分演奏情報が生成されて出力される。

【0105】

また、図 16 のフローチャートに示すように、複数の入力画像により構成されるフレーズ、曲の一部または全部の複合演奏情報を生成する処理においても図 15 のフローチャートにおけるステップ S 301 とステップ S 302 と同様の処理が行われる。

40

【0106】

この第 3 の実施の形態によれば、入力画像において楽器の一部が映っていなくても第 1 の実施の形態と同様に演奏情報、楽譜情報の生成を行うことができる。

【0107】

< 4 . 変形例 >

以上、本技術の実施の形態について具体的に説明したが、本技術は上述の実施の形態に限定されるものではなく、本技術の技術的思想に基づく各種の変形が可能である。

【0108】

実施の形態では演奏の音がなくても複数枚の連続する静止画像または動画を構成する複数のフレーム画像から演奏情報および楽譜情報を生成できると説明したが、本技術は音の

50

使用を除外するものではない。演奏情報および楽譜情報を生成の際の補助情報として音情報を用いてもよいし、生成した演奏情報および楽譜情報の精度を確認する際に音情報を用いてもよい。例えば、入力映像の音声に対して音声認識処理を施し、音の周波数から音階を認識する、音量から強弱や演奏しているか否かを認識するなどである。

【0109】

第2の実施の形態と第3の実施の形態を組み合わせることにより、入力画像において演奏者の手の一部および楽器の演奏領域の一部が映っていない場合でも演奏情報の生成を行うことができる。

【0110】

本技術は実施の形態で挙げたピアノ、ギター、ドラムに限られず、木琴、鉄琴、パーカッションなどの楽器の演奏に対しても使用可能である。

10

【0111】

実施の形態では主に押す、叩くなどのピアノの演奏方法、手をストロークさせる、爪弾くなどのギターの演奏方法を例にして説明を行ったが、それら以外の演奏方法、例えば、引っ張る、弾くなどの演奏動作を認識して演奏情報を生成してもよい。入力画像から認識できる楽器の演奏の動作であればどのような動作に基づいて演奏情報を生成してもよい。

【0112】

第3の実施の形態においては、入力画像に写っていない楽器の一部を推定し、その推定結果に基づいて演奏情報生成部107が演奏情報を生成するようにしてもよい。

【0113】

20

本技術は以下のような構成も取ることができる。

(1)

入力画像から演奏者の身体の一部の位置を認識する位置認識部と、
前記入力画像から楽器を認識する楽器認識部と、
前記部位の位置と前記楽器との関連性に基づき、前記演奏者による前記楽器の演奏を示す演奏情報を生成する演奏情報生成部と
を備える情報処理装置。

(2)

前記位置認識部により認識された前記部位の形状を認識する形状認識部を備え、
前記演奏情報生成部は、前記部位の形状と前記楽器の関連性に基づき前記演奏情報を生成する(1)に記載の情報処理装置。

30

(3)

前記位置認識部により認識された前記部位の動きを認識する動き認識部を備え、
前記演奏情報生成部は、前記部位の動きと前記楽器の関連性に基づき前記演奏情報を生成する(1)または(2)に記載の情報処理装置。

(4)

前記演奏情報は、前記演奏者が前記楽器を演奏している状態に対応した第1演奏要素を含む(1)から(3)のいずれかに記載の情報処理装置。

(5)

前記第1演奏要素は、前記演奏者により演奏されている音階を含む(4)に記載の情報処理装置。

40

(6)

前記演奏情報は、前記演奏者が前記楽器を演奏していない状態に対応した第2演奏要素を含む(1)から(5)のいずれかに請求項1に記載の情報処理装置。

(7)

前記第2演奏要素は、前記演奏者により演奏されていない休みの長さを含む(6)に記載の情報処理装置。

(8)

前記演奏情報は、複数の前記入力画像間に跨る要素である第3演奏要素を含む(1)から(7)のいずれかに記載の情報処理装置。

50

(9)

前記第3演奏要素は、前記演奏者により演奏されている曲のテンポを含む(8)に記載の情報処理装置。

(10)

前記演奏情報生成部は、一の前記入力画像に対応した前記演奏情報を生成する(1)から(9)のいずれかに記載の情報処理装置。

(11)

前記演奏情報生成部は、複数の前記入力画像により構成される前記楽器の演奏の一部または全部に対応する演奏情報を生成する(1)から(9)のいずれかに記載の情報処理装置。

10

(12)

前記関連性は、前記楽器に対する前記部位の接触位置である(1)から(11)のいずれかに記載の情報処理装置。

(13)

前記関連性は、前記楽器に対する前記部位の動作の方向である(1)から(12)のいずれかに記載の情報処理装置。

(14)

前記部位は前記演奏者の手である(1)から(13)のいずれかに記載の情報処理装置。

(15)

前記演奏情報から楽譜情報を生成する楽譜情報生成部を備える(1)から(14)のいずれかに記載の情報処理装置。

20

(16)

前記入力画像において前記部位の一部が映っていない場合、前記部位の一部の位置を推定し、前記演奏情報生成部は推定結果に基づき前記演奏情報を生成する(1)から(15)のいずれかに記載の情報処理装置。

(17)

前記入力画像において前記楽器の一部が映っていない場合、前記演奏情報生成部は、前記部位と前記楽器の一部との前記関連性を推定し、推定結果に基づき前記演奏情報を生成する(1)から(16)のいずれかに記載の情報処理装置。

30

(18)

前記演奏情報生成部は、補助情報として音情報を用いて前記演奏情報を生成する(1)から(17)のいずれかに記載の情報処理装置。

(19)

入力画像から演奏者の身体の一部の位置を認識し、
前記入力画像から楽器を認識し、
前記部位の位置と前記楽器との関連性に基づき、前記演奏者による前記楽器の演奏を示す演奏情報を生成する
情報処理方法。

40

(20)

入力画像から演奏者の身体の一部の位置を認識し、
前記入力画像から楽器を認識し、
前記部位の位置と前記楽器との関連性に基づき、前記演奏者による前記楽器の演奏を示す演奏情報を生成する
情報処理方法をコンピュータに実行させる情報処理プログラム。

【符号の説明】

【 0 1 1 4 】

1 0 0、2 0 0、3 0 0・・・情報処理装置

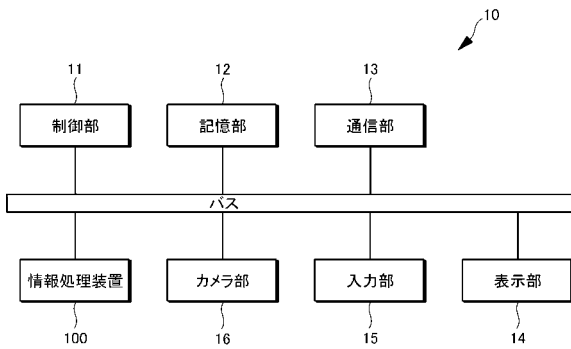
1 0 2・・・位置認識部

1 0 3・・・形状認識部

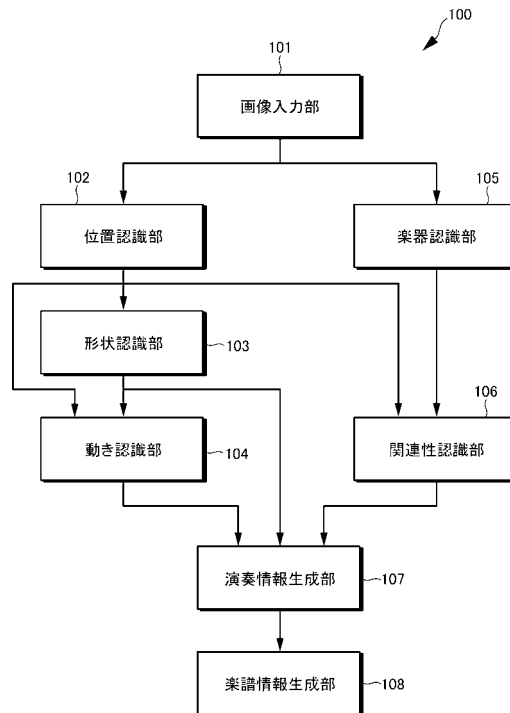
50

- 104 . . . 動き認識部
- 105 . . . 楽器認識部
- 107 . . . 演奏情報生成部
- 108 . . . 楽譜情報生成部
- 202 . . . 第1位置認識部
- 203 . . . 第2位置認識部

【図1】



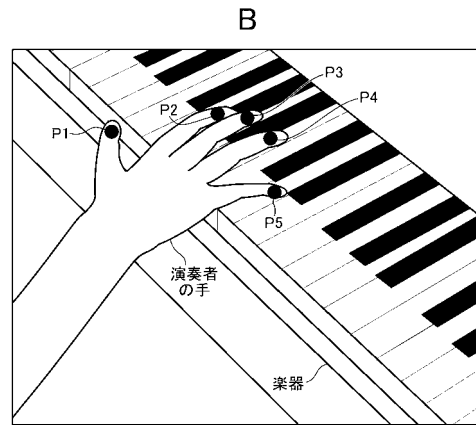
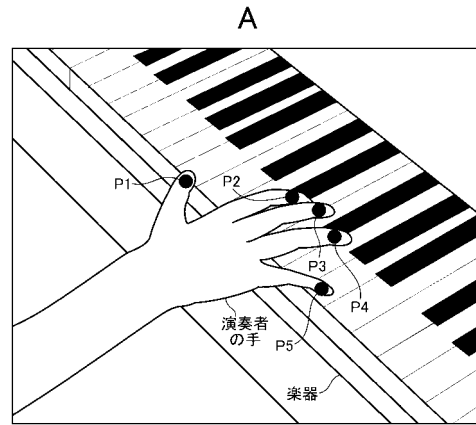
【図2】



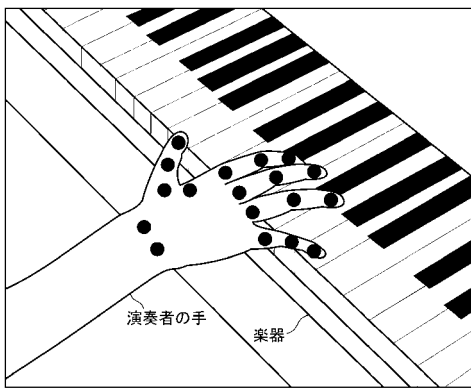
【 図 3 】



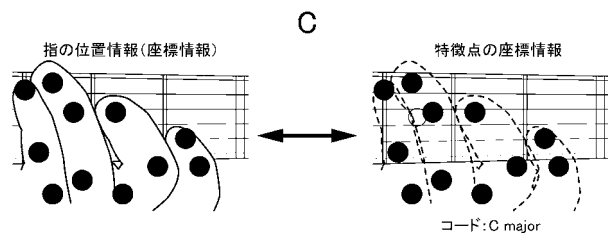
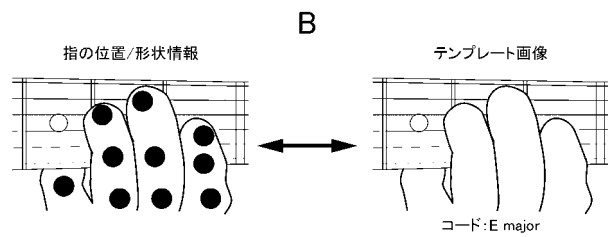
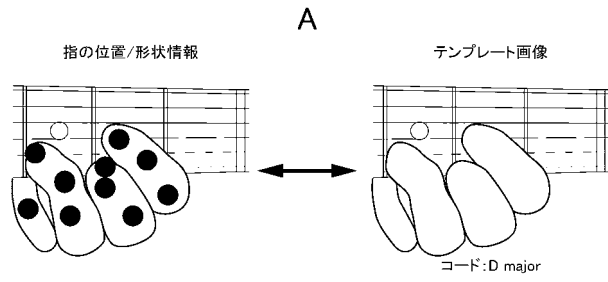
【 図 4 】



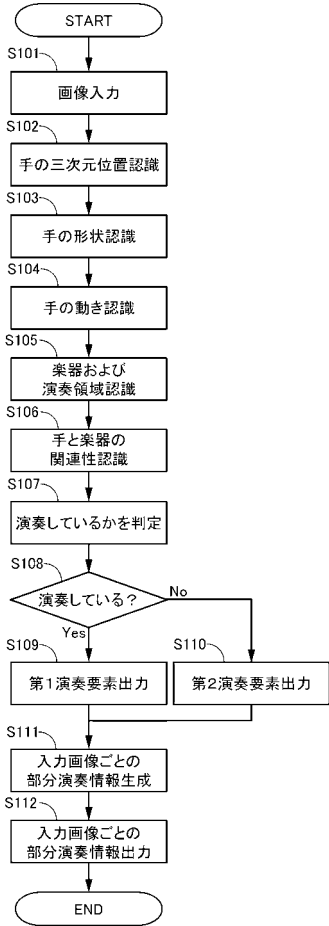
【 図 5 】



【 図 6 】



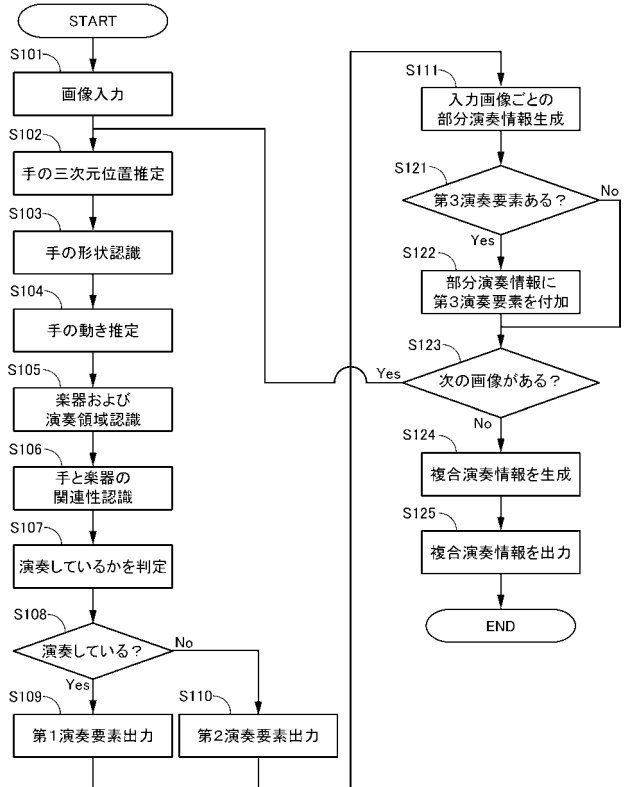
【 図 7 】



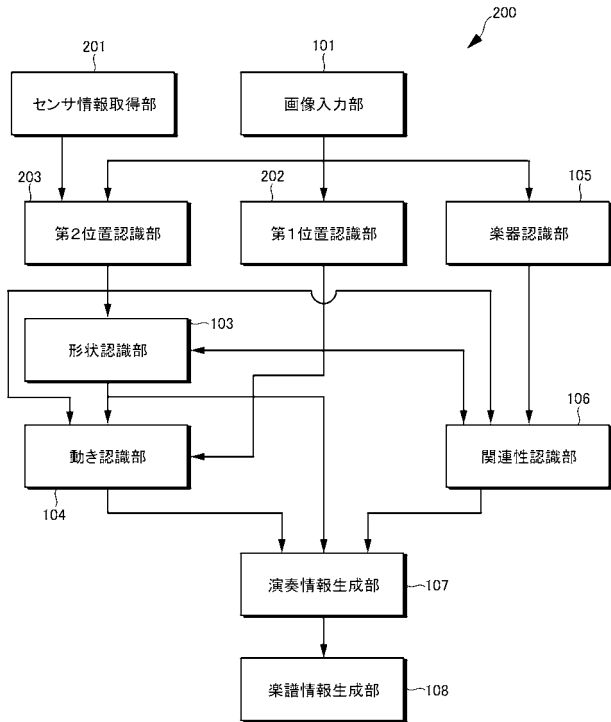
【 図 9 】



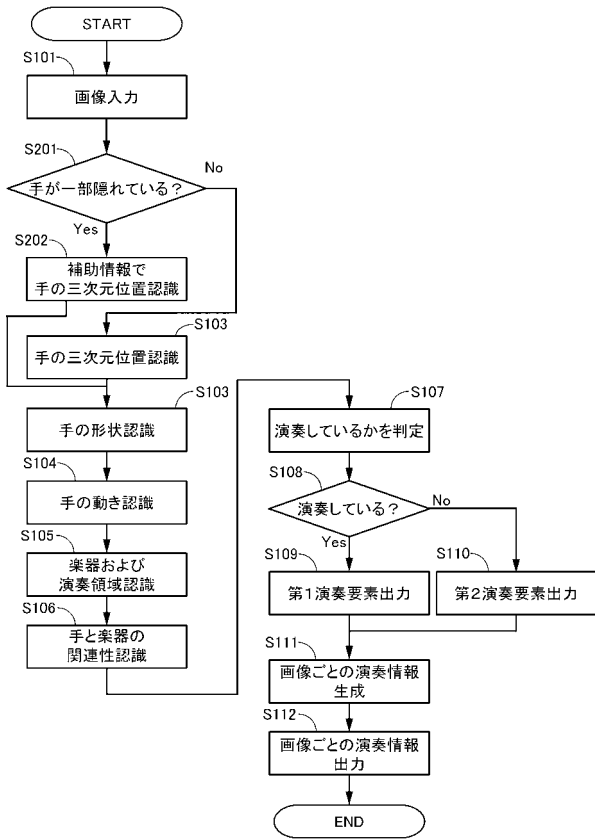
【 図 8 】



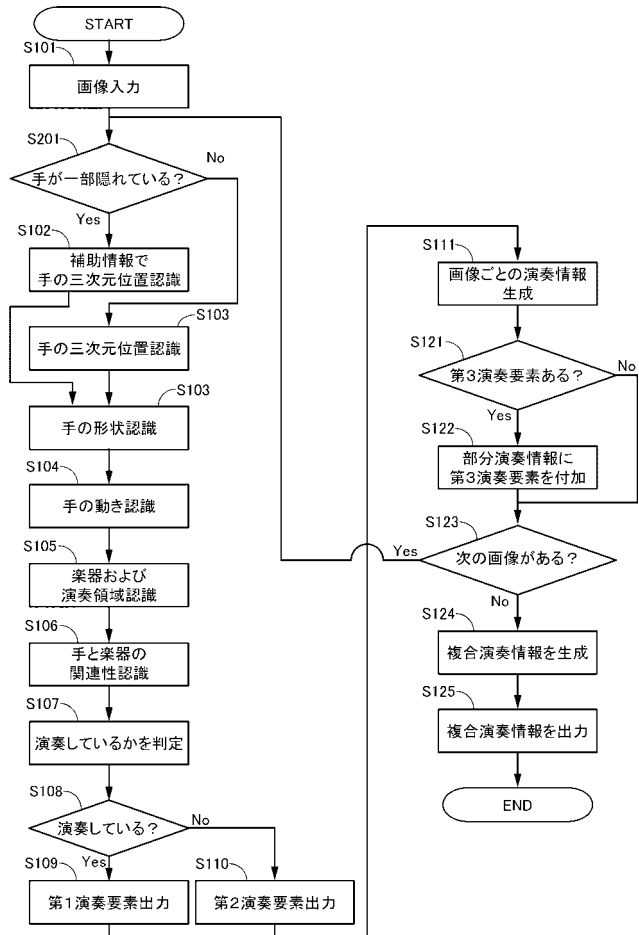
【 図 10 】



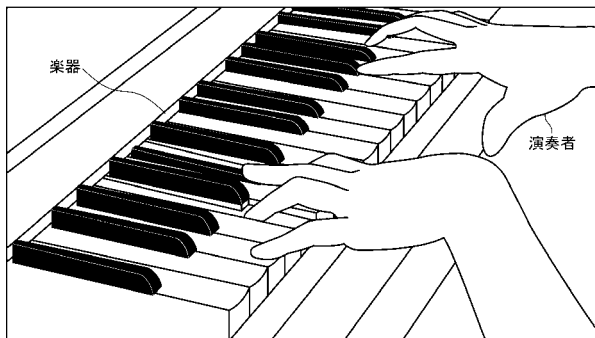
【図11】



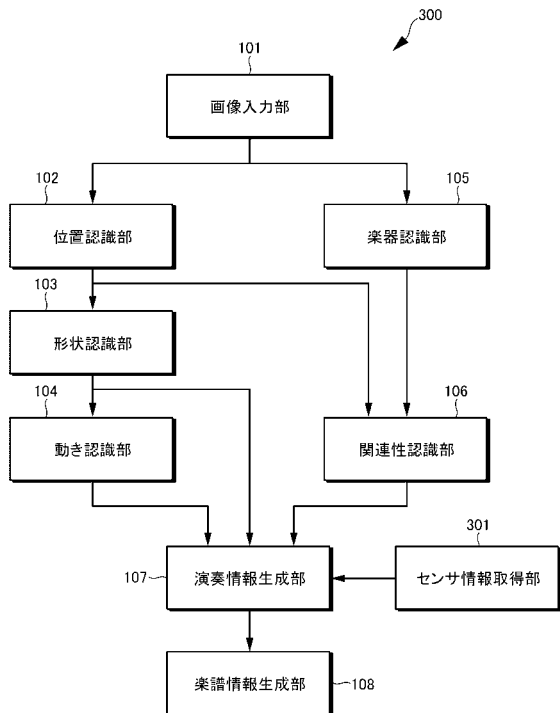
【図12】



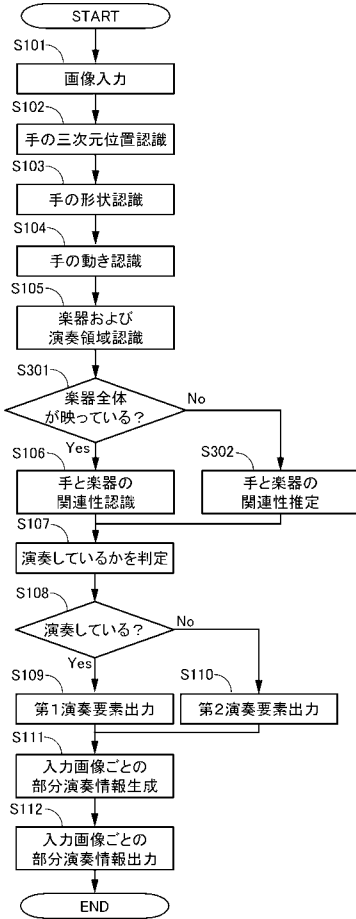
【図13】



【図14】



【図15】



【図16】

