

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2022年1月13日(13.01.2022)



(10) 国際公開番号

WO 2022/009331 A1

(51) 国際特許分類:

G06T 13/40 (2011.01) G06T 7/20 (2017.01)
G06T 7/00 (2017.01) G06N 3/08 (2006.01)

(21) 国際出願番号: PCT/JP2020/026683

(22) 国際出願日: 2020年7月8日(08.07.2020)

(25) 国際出願の言語: 日本語

(26) 国際公開の言語: 日本語

(71) 出願人: 日本電信電話株式会社 (NIPPON TELEGRAPH AND TELEPHONE CORPORATION) [JP/JP]; 〒1008116 東京都千代田区大手町一丁目5番1号 Tokyo (JP).

(72) 発明者: 謝 芳陸 (XIE Fanglu); 〒1808585 東京都武蔵野市緑町3丁目9-11 N T T 知的

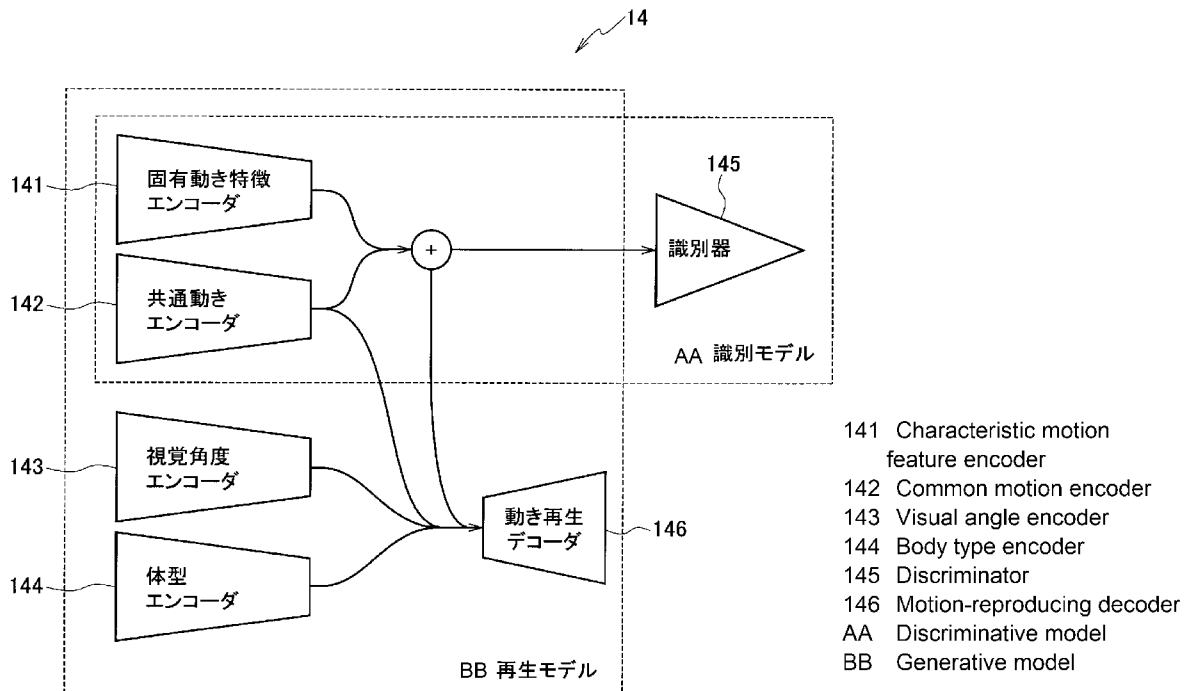
財産センタ内 Tokyo (JP). 入江 豪 (IRIE Go); 〒1808585 東京都武蔵野市緑町3丁目9-11 N T T 知的財産センタ内 Tokyo (JP). 松林 達史 (MATSUBAYASHI Tatsushi); 〒1808585 東京都武蔵野市緑町3丁目9-11 N T T 知的財産センタ内 Tokyo (JP). 日高 浩太 (HIDAKA Kota); 〒1808585 東京都武蔵野市緑町3丁目9-11 N T T 知的財産センタ内 Tokyo (JP).

(74) 代理人: 三好 秀和, 外 (MIYOSHI Hidekazu et al.); 〒1050001 東京都港区虎ノ門1丁目2番8号 虎ノ門琴平タワー Tokyo (JP).

(81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH,

(54) Title: TRAINING DEVICE, TRAINING METHOD, AND PROGRAM

(54) 発明の名称: 学習装置、学習方法、およびプログラム



(57) Abstract: The present invention provides a training device 1 comprising a data input unit 11 that inputs time-series data logging the motion of a skeleton and a training unit 15 that trains a model 14 including encoders 143 and 144 that use the time-series data to extract time-independent features of the skeleton, two or more encoders 141 and 142 that extract motion features that vary over time, a discriminator 145 that discriminates the skeleton on the basis of the motion features, and a motion-reproducing decoder 146 that outputs time-series data logging new motion of the skeleton. With regard to

CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

- (84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

添付公開書類:

- 一 国際調査報告(条約第21条(3))

the encoders 141 and 142 that extract the motion features, the training unit 15 uses adversarial training to train a unique motion feature encoder 141 that extracts motion features unique to the skeleton and a common motion encoder 142 that extracts common motion features.

(57) 要約: 本実施形態の学習装置 1 は、スケルトンの動きを記録した時系列データを入力するデータ入力部 11 と、時系列データを用いて、時間に依存しないスケルトンの特徴を抽出するエンコーダ 143, 144、時間変動する動きの特徴を抽出する 2 つ以上のエンコーダ 141, 142、動きの特徴からスケルトンを識別する識別器 145、およびスケルトンの新たな動きを記録した時系列データを出力する動き再生デコーダ 146 を含むモデル 14 を学習する学習部 15 を備える。学習部 15 は、動きの特徴を抽出するエンコーダ 141, 142 について、敵対的学習により、スケルトンの固有的な動きの特徴を抽出する固有動き特徴エンコーダ 141 と共通する動きの特徴を抽出する共通動きエンコーダ 142 を学習する。

明 細 書

発明の名称：学習装置、学習方法、およびプログラム

技術分野

[0001] 本発明は、学習装置、学習方法、およびプログラムに関する。

背景技術

[0002] 故人や有名人のフェイク映像を生成する方法として、口の部分だけ映像を生成し、生成した映像を本人の映像に部分的に合成する技術や、ターゲットのアバター（コンピュータグラフィックスで作成したキャラクタ）に動きを真似させる技術が知られている。

[0003] 非特許文献1には、深層学習を用いて、動的要素（モーション）と静的要素（スケルトンとビューアングル）を分離し、動的要素と静的要素を再構成する技術が開示されている。

先行技術文献

非特許文献

[0004] 非特許文献1：Kfir Aberman, et al., “Learning Character-Agnostic Motion for Motion Retargeting in 2D”, ACM Trans. Graph., Vol. 38, No. 4, Article 75.

非特許文献2：Yaroslav Ganin and Victor Lempitsky, “Unsupervised Domain Adaptation by Backpropagation”, Proceedings of the 32nd International Conference on Machine Learning

発明の概要

発明が解決しようとする課題

[0005] 従来の技術は、体型などの個性を反映することは考慮されているが、動きに含まれる個性を反映することは考慮されていない。有名人の動きを真似する人のモーションを反映する方法が考えられるが、多くは一部の癖を誇張する表現であり、利用できるシーンは限られている。

[0006] 本発明は、上記に鑑みてなされたものであり、ターゲットの固有の動きを

反映した動きを生成することを目的とする。

課題を解決するための手段

[0007] 本発明の一態様の学習装置は、動体の動きを記録した時系列データを入力する入力部と、前記時系列データを用いて、時間に依存しない前記動体の特徴を抽出するエンコーダ、時間変動する動きの特徴を抽出する2つ以上のエンコーダ、前記動きの特徴から前記動体を識別する識別器、および前記動体の新たな動きを記録した時系列データを出力するデコーダを含むモデルを学習する学習部を備え、前記学習部は、前記動きの特徴を抽出する2つ以上のエンコーダについて、敵対的学習により、前記動体の固有的な動きの特徴を抽出するエンコーダと共通する動きの特徴を抽出するエンコーダを学習する。

[0008] 本発明の一態様の学習方法は、コンピュータが実行する学習方法であって、動体の動きを記録した時系列データを入力し、前記時系列データを用いて、時間に依存しない前記動体の特徴を抽出するエンコーダ、時間変動する動きの特徴を抽出する2つ以上のエンコーダ、前記動きの特徴から前記動体を識別する識別器、および前記動体の新たな動きを記録した時系列データを出力するデコーダを含むモデルを学習し、前記モデルを学習する際、前記動きの特徴を抽出する2つ以上のエンコーダについて、敵対的学習により、前記動体の固有的な動きの特徴を抽出するエンコーダと共通する動きの特徴を抽出するエンコーダを学習する。

発明の効果

[0009] 本発明によれば、ターゲットの固有の動きを反映した動きを生成できる。

図面の簡単な説明

[0010] [図1]図1は、本実施形態の学習装置の構成の一例を示す機能ブロック図である。

[図2]図2は、学習装置が入力する時系列2次元データの一例を示す図である。

[図3]図3は、学習に用いるデータセットの分割の一例を示す図である。

[図4]図4は、データの分類の一例を示す図である。

[図5]図5は、学習に用いるデータセットの一例を示す図である。

[図6]図6は、本実施形態のモデルの構成の一例を示す図である。

[図7]図7は、識別モデルの学習の一例を説明する図である。

[図8]図8は、デコードに入力するコードを説明する図である。

[図9]図9は、共通動きを生成する再生モデルの学習の一例を示す図である。

[図10]図10は、共通動きを生成する再生モデルの学習の一例を示す図である。

[図11]図11は、固有動きを生成する再生モデルの学習の一例を示す図である。

[図12]図12は、固有動きを生成する再生モデルの学習の一例を示す図である。

[図13]図13は、固有動きを生成する再生モデルの学習の一例を示す図である。

[図14]図14は、固有動きを生成する再生モデルの学習の一例を示す図である。

[図15]図15は、ターゲットとなる有名人に新しい動きをさせる実施例を示す図である。

[図16]図16は、映像内の人物が本人であるか否かを判定する実施例を示す図である。

[図17]図17は、個性を無くした汎用的な動きをするスケルトンを生成する実施例を示す図である。

[図18]図18は、学習装置のハードウェア構成の一例を示す図である。

発明を実施するための形態

[0011] 以下、本発明の実施の形態について図面を用いて説明する。

[0012] 図1は、本実施形態の学習装置1の構成の一例を示す機能ブロック図である。同図に示す学習装置1は、データ入力部11、データ配置部12、データブロック13、モデル14、および学習部15を備えて、学習済みモデル

16 を出力する。

[0013] データ入力部11は、人（スケルトン）の動きを記録した時系列2次元データを入力し、トレーニングデータ、バリデーションデータ、およびテストデータの3つのタイプに分割する。時系列2次元データは、図2に示すように、15ポイントのスケルトンをTフレーム分（例えば $T=64$ ）有するスケルトンブロックである。スケルトンの1つのポイントはx座標値とy座標値を有する。つまり、1つのスケルトンで $15 \times 2 = 30$ の座標値を有する。なお、以下では、人のスケルトンを例に説明するが、これに限るものではない。例えば、動物、人形、あるいはロボットなど骨格を有する動体を撮影した画像から骨格（スケルトン）の時間変化を検出できるデータを利用できる。人の骨格に限らず、モーションマーカーを利用してもよいし、入力自体を人の映像にして、内部で骨格情報にして扱う事もできる。

[0014] データ入力部11は、多数の人物の多数の動きのスケルトンブロックを入力し、図3に示すように、入力した多数のスケルトンブロックをトレーニングデータ、バリデーションデータ、およびテストデータの3つのタイプに分割する。図3の例では、Mキャラクタ分のデータについて、5分の4をトレーニングデータ、5分の1をバリデーションデータとして設定し、残りのNキャラクタ分のデータをテストデータとした。テストデータの量はいくつでもよい。

[0015] データ配置部12は、データ入力部11の入力したスケルトンブロックを人物（Skeleton）、動き（Motion）、および視覚角度（View-Angle）で分類し、データブロック13に格納する。図4にデータ配置部12が分類する配置フォーマットの一例を示す。図4では、各スケルトンブロックについて、人物を大文字アルファベット「A, B, C, D, ..., N」、動きを数字「1, 2, 3, 4, ..., q」、視覚角度を記号「!, ", #, &, ..., (」、人の動きの個性を小文字アルファベット「_a, _b, _c, ..., _n」で表した。人物によって体型（スケルトン）が異なる。動きとは、スケルトンの動きのタイプ（種別）である。

例えば、動き1は歩く、動き2はダンス、動き3はサッカーの動作などである。視覚角度とは、スケルトンを構成する肢体を観察する方向の角度を示す。例えば、左90度、左60度、左30度、正面、右30度、右60度、右90度の7つの角度である。角度は、肢体が向く方向を示すものでもよい。一例として、スケルトンブロック「B1!__b」は、Bさんが動き1をしているスケルトンであって、角度「!」からみたスケルトンであり、Bさんの動きの個性を含むことを示す。

[0016] データブロック13は、図5に示すように、スケルトンブロックごとに人物を示すラベルを付与した学習用のデータセットを格納する。

[0017] モデル14は、図6に示すように、固有動き特徴エンコーダ (Unique Motion Encoder) 141、共通動きエンコーダ (Common Motion Encoder) 142、視覚角度エンコーダ (View-Angle Encoder) 143、体型エンコーダ (Skeleton Encoder) 144、識別器 (Classifier) 145、および動き再生デコーダ (Motion Reconstruction Decoder) 146を有する。モデル14は、ニューラルネットワークとも言う。

[0018] 固有動き特徴エンコーダ141と共通動きエンコーダ142は、時間変動する動きの特徴を抽出するエンコーダである。固有動き特徴エンコーダ141は、人物を識別できるように、識別ロス（誤差）を低くするようにスケルトンの動きを学習する。共通動きエンコーダ142は、人物を識別できないように、識別ロスを高くするようにスケルトンの動きを学習する。本実施形態では、時間変動する動きの特徴を抽出するエンコーダを2つ備えたが、3つ以上備えてもよい。

[0019] 視覚角度エンコーダ143と体型エンコーダ144は、時間に依存しないスケルトンの特徴を抽出するエンコーダである。視覚角度エンコーダ143は、視覚角度の特徴を抽出する。体型エンコーダ144は、スケルトンの体型の特徴を抽出する。本実施形態では、時間に依存しないスケルトンの特徴

を抽出するエンコーダを2つ備えたが、1つでもよいし、3つ以上備えてもよい。

[0020] 識別器145は、スケルトンの動きの特徴から人物を識別する。

[0021] 動き再生デコーダ146は、ソースとなるスケルトンブロックとターゲットとなるスケルトンブロックを入力して抽出された特徴から、ソースのスケルトンの動きをターゲットのスケルトンへ再構成したスケルトンブロックを出力する。

[0022] エンコーダ141, 142と識別器145で動きの特徴から人物を識別する識別モデルを構成し、エンコーダ141～144と動き再生デコーダ146で動きの特徴とスケルトンの特徴からスケルトンの動きを再構成する再生モデルを構成する。

[0023] 続いて、学習部15による識別モデルの学習について説明する。

[0024] 学習部15は、エンコーダ141, 142の抽出した動きの特徴で識別器145による敵対的学習を行い、識別モデルを学習する。より具体的には、固有動き特徴エンコーダ141の抽出した動きの固有特徴と共通動きエンコーダ142の抽出した共通動きを合わせた動き特徴の識別器145による識別結果は識別ロスを低くするように学習し、共通動きエンコーダ142の抽出した共通動きの識別器145による識別結果は識別ロスを高くするように学習する。

[0025] 図7に、識別モデルの学習の一例を示す。図7の例では、学習データとして「A1!__a」で示されるスケルトンブロックをエンコーダ141, 142に入力した。このスケルトンブロックは、Aさんの動き1を「!」で示される角度から見たスケルトンの時系列データである。このスケルトンブロックにはラベル「A」が付与されている。このスケルトンブロックを固有動き特徴エンコーダ141と共通動きエンコーダ142に入力し、エンコーダ141, 142から出力された動きの固有特徴と共通動きを合わせた動き特徴を識別器145に入力し、識別器145で識別した出力結果を得る。得られた出力結果とスケルトンブロックに付与されたラベル「A」を比較し、識別

ロスを低くするように、誤差関数の勾配を図中の破線矢印のように逆伝搬させて識別モデルに学習させる。誤差逆伝搬時に、Gradient Reversal Layer (GRL) 147を通し、前段の共通動きエンコーダ142へ伝わる勾配に負の定数をかける。固有動き特徴エンコーダ141へ伝わる勾配はそのまま入力する。これにより、固有動き特徴エンコーダ141は人を識別できるように学習させると同時に、共通動きエンコーダ142は人を識別できないように学習させることができる。なお、GRLについては非特許文献2に記載がある。

[0026] 続いて、学習部15による再生モデルの学習について説明する。再生モデルの学習の説明に際して、図8を参照し、動き再生デコーダ146に入力するコードの一例について説明する。

[0027] スケルトンブロックが入力されると、固有動き特徴エンコーダ141からはTフレーム分の固有特徴が抽出され、共通動きエンコーダ142からはTフレーム分の共通動きが抽出される。固有特徴と共通動きを合わせると共通動きに固有特徴を反映した固有動きとなる。

[0028] 視覚角度エンコーダ143からは視覚角度が抽出される。体型エンコーダ144からは体型が抽出される。視覚角度および体型は、時間に依存しない特徴である。

[0029] 図8に示すように、エンコーダ141、142で抽出された動きの特徴に、視覚角度と体型を時間軸に沿ってTフレーム分連結したコードを動き再生デコーダ146に入力する。動き再生デコーダ146からは、エンコーダ141、142で抽出された動きをする体型エンコーダ144から抽出された体型のスケルトンを視覚角度エンコーダ143から抽出された視覚角度から見たスケルトンブロックが出力される。出力されるスケルトンブロックは、入力されたスケルトンブロックと同じ形式である。なお、共通動きエンコーダ142で抽出された動きの特徴のみを用いると、人の固有の特徴を持たない共通動きのスケルトンブロックが得られる。共通動きエンコーダ142で抽出された共通動きに、固有動き特徴エンコーダ141で抽出された固有特

徴を加えることで、共通動きに人の固有の特徴を加えた固有動きのスケルトンブロックが得られる。

[0030] 図9および図10を参照し、共通動きを生成する再生モデルの学習例について説明する。

[0031] 図9の例では、学習部15は、Aさんのスケルトンブロックを共通動きエンコーダ142に入力し、Bさんのスケルトンブロックを視覚角度エンコーダ143と体型エンコーダ144に入力して、Aさんの動き（共通動き）をするBさんのスケルトンをBさんの視覚角度から見たスケルトンブロックを出力するようにモデル14を学習した。具体的には、「A1”__a」で示されるスケルトンブロックを共通動きエンコーダ142に入力し、「B2！__b」で示されるスケルトンブロックを視覚角度エンコーダ143と体型エンコーダ144に入力した。共通動きエンコーダ142で動き1の特徴を抽出し、視覚角度エンコーダ143でBさんのスケルトンを見る視覚角度を抽出し、体型エンコーダ144でBさんのスケルトンの体型を抽出した。エンコーダ142～144で抽出した特徴を再構成して、動き1をするBさんの体型のスケルトンを視覚角度「！」から見たスケルトンブロック「B1！’」の出力結果を動き再生デコーダ146から得る。得られた出力結果と正解に近いスケルトンブロック「B1！__b」を比較し、誤差を図中の破線矢印のように逆伝搬させて再生モデルに学習させる。

[0032] 図10の例では、学習部15は、Aさんのスケルトンブロックを共通動きエンコーダ142と視覚角度エンコーダ143に入力し、Bさんのスケルトンブロックを体型エンコーダ144に入力して、Aさんの動き（共通動き）をするBさんのスケルトンをAさんの視覚角度から見たスケルトンブロックを出力するようにモデル14を学習した。具体的には、「A1”__a」で示されるスケルトンブロックを共通動きエンコーダ142と視覚角度エンコーダ143に入力し、「B2！__b」で示されるスケルトンブロックを体型エンコーダ144に入力した。共通動きエンコーダ142で動き1の特徴を抽出し、視覚角度エンコーダ143でAさんのスケルトンを見る視覚角度を抽出し、

出し、体型エンコーダ144でBさんのスケルトンの体型を抽出した。エンコーダ142～144で抽出した特徴を再構成して、動き1をするBさんの体型のスケルトンを視覚角度「”」からみたスケルトンブロック「B1”'」の出力結果を動き再生デコーダ146から得る。得られた出力結果と正解に近いスケルトンブロック「B1”__b」を比較し、誤差を図中の破線矢印のように逆伝搬させて再生モデルに学習させる。

[0033] 図11～図14を参照し、固有動きを生成する再生モデルの学習例について説明する。

[0034] 図11の例では、学習部15は、Aさんのスケルトンブロックを共通動きエンコーダ142に入力し、Bさんのスケルトンブロックを固有動き特徴エンコーダ141、視覚角度エンコーダ143、および体型エンコーダ144に入力して、Bさんの個性を持ったAさんの動きをするBさんのスケルトンをBさんの視覚角度から見たスケルトンブロックを出力するようにモデル14を学習した。具体的には、「A1”__a」で示されるスケルトンブロックを共通動きエンコーダ142に入力し、「B2!__b」で示されるスケルトンブロックを固有動き特徴エンコーダ141、視覚角度エンコーダ143、および体型エンコーダ144に入力した。共通動きエンコーダ142で動き1の特徴を抽出し、固有動き特徴エンコーダ141でBさんの動きの固有特徴を抽出し、視覚角度エンコーダ143でBさんのスケルトンを見る視覚角度を抽出し、体型エンコーダ144でBさんのスケルトンの体型を抽出した。エンコーダ141～144で抽出した特徴を再構成して、Bさんの個性を持った動き1をするBさんの体型のスケルトンを視覚角度「！」からみたスケルトンブロック「B1!__b'」の出力結果を動き再生デコーダ146から得る。得られた出力結果と正解のスケルトンブロック「B1!__b」を比較し、誤差を図中の破線矢印のように逆伝搬させて再生モデルに学習させる。

[0035] 図12の例では、学習部15は、Aさんのスケルトンブロックを共通動きエンコーダ142に入力し、Aさんと同じ動きをするBさんの個性を持った

スケルトンブロックを固有動き特徴エンコーダ141に入力し、Bさんの別のスケルトンブロックを視覚角度エンコーダ143と体型エンコーダ144に入力して、Bさんの個性を持ったAさんと同じ動きをするBさんのスケルトンをBさんの視覚角度から見たスケルトンブロックを出力するようにモデル14を学習した。具体的には、「A1”__a」で示されるスケルトンブロックを共通動きエンコーダ142に入力し、「B1!__b」で示されるスケルトンブロックを固有動き特徴エンコーダ141に入力し、「B2!__b」で示されるスケルトンブロックを視覚角度エンコーダ143と体型エンコーダ144に入力した。共通動きエンコーダ142で動き1の特徴を抽出し、固有動き特徴エンコーダ141でBさんの動き1に依存する動きの固有特徴を抽出し、視覚角度エンコーダ143でBさんのスケルトンを見る視覚角度を抽出し、体型エンコーダ144でBさんのスケルトンの体型を抽出した。エンコーダ141～144で抽出した特徴を再構成して、Bさんの個性を持った動き1をするBさんの体型のスケルトンを視覚角度「!」から見たスケルトンブロック「B1!__b'」の出力結果を動き再生デコーダ146から得る。得られた出力結果と正解のスケルトンブロック「B1!__b」を比較し、誤差を図中の破線矢印のように逆伝搬させて再生モデルに学習させる。なお、固有動き特徴エンコーダ141へは、Bさんが動き1をする他のスケルトンブロック（例えば「B1”__b」で示されるスケルトンブロック）を入力しても学習可能である。

[0036] 図11の例と図12の例とでは、固有動き特徴エンコーダ141に入力するスケルトンブロックが異なっている。図12の例では、共通動きエンコーダ142に入力するスケルトンブロックの動き1と固有動き特徴エンコーダ141に入力するスケルトンブロックの動き1が同じである。その結果、図12の例は、図11の例よりも、より鮮明に動きの固有特徴を抽出できる。

[0037] 図13の例では、学習部15は、Aさんのスケルトンブロックを共通動きエンコーダ142と視覚角度エンコーダ143に入力し、Bさんのスケルトンブロックを固有動き特徴エンコーダ141と体型エンコーダ144に入力

して、Bさんの個性を持ったAさんの動きをするBさんのスケルトンをAさんの視覚角度から見たスケルトンブロックを出力するようにモデル14を学習した。具体的には、「A1!__a」で示されるスケルトンブロックを共通動きエンコーダ142と視覚角度エンコーダ143に入力し、「B2”__b」で示されるスケルトンブロックを固有動き特徴エンコーダ141と体型エンコーダ144に入力した。共通動きエンコーダ142で動き1の特徴を抽出し、視覚角度エンコーダ143でAさんのスケルトンを見る視覚角度を抽出し、固有動き特徴エンコーダ141でBさんの動きの固有特徴を抽出し、体型エンコーダ144でBさんのスケルトンの体型を抽出した。エンコーダ141～144で抽出した特徴を再構成して、Bさんの個性を持った動き1をするBさんの体型のスケルトンを視覚角度「!」から見たスケルトンブロック「B1!__b'」の出力結果を動き再生デコーダ146から得る。得られた出力結果と正解のスケルトンブロック「B1!__b」を比較し、誤差を図中の破線矢印のように逆伝搬させて再生モデルに学習させる。

- [0038] 図14の例では、学習部15は、Aさんのスケルトンブロックを共通動きエンコーダ142と視覚角度エンコーダ143に入力し、Aさんと同じ動きをするBさんの個性を持ったスケルトンブロックを固有動き特徴エンコーダ141に入力し、Bさんのスケルトンブロックを体型エンコーダ144に入力して、Bさんの個性を持ったAさんと同じ動きをするBさんのスケルトンをAさんの視覚角度から見たスケルトンブロックを出力するようにモデル14を学習した。具体的には、「A1!__a」で示されるスケルトンブロックを共通動きエンコーダ142と視覚角度エンコーダ143に入力し、「B1!__b」で示されるスケルトンブロックを固有動き特徴エンコーダ141に入力し、「B2”__b」で示されるスケルトンブロックを体型エンコーダ144に入力した。共通動きエンコーダ142で動き1の特徴を抽出し、視覚角度エンコーダ143でAさんのスケルトンを見る視覚角度を抽出し、固有動き特徴エンコーダ141でBさんの動き1に依存する動きの固有特徴を抽出し、体型エンコーダ144でBさんのスケルトンの体型を抽出した。エン

コード141～144で抽出した特徴を再構成して、Bさんの個性を持った動き1をするBさんの体型のスケルトンを視覚角度「！」からみたスケルトンブロック「B1!__b'」の出力結果を動き再生デコード146から得る。得られた出力結果と正解のスケルトンブロック「B1!__b」を比較し、誤差を図中の破線矢印のように逆伝搬させて再生モデルに学習させる。なお、固有動き特徴エンコード141へは、Bさんが動き1をする他のスケルトンブロック（例えば「B1”__b」で示されるスケルトンブロック）を入力しても学習可能である。

[0039] 図13の例と図14の例とでは、固有動き特徴エンコード141に入力するスケルトンブロックが異なっている。図14の例では、共通動きエンコード142に入力するスケルトンブロックの動き1と固有動き特徴エンコード141に入力するスケルトンブロックの動き1が同じである。その結果、図14の例は、図13の例よりも、より鮮明に動きの固有特徴を抽出できる。

[0040] 学習部15によるモデル14の学習を繰り返して、学習済みモデル16が得られる。

[0041] 次に、本実施形態の学習装置1で学習させたモデルを利用した実施例について説明する。

[0042] 図15は、再生モデルを利用して、ターゲットとなる有名人に新しい動きをさせる実施例である。

[0043] 有名人の動きを示すスケルトンの時系列2次元データを固有動き特徴エンコード141と体型エンコード144に入力し、有名人に再現させたい動きをしたソースの被写体のスケルトンの時系列2次元データを共通動きエンコード142と視覚角度エンコード143に入力する。これにより、ソースの被写体の動きに有名人の個性を加えた動きを有名人のスケルトンに行わせた時系列2次元データが得られる。例えば、有名人のアバターに得られたスケルトンの時系列2次元データを適用することで、ソースの被写体の動きに有名人の動きの個性が加えられたリアルなアバターを生成できる。

[0044] なお、有名人のデータの代わりに、ソースの被写体のデータを体型エンコ

ーダ 144 に入力すれば、ソースの被写体が有名人の動きをまねたデータを生成できる。

[0045] 図 16 は、識別モデルを利用して、映像内の人物が本物であるかなりすましであるかを識別する実施例である。

[0046] 判定対象の人物が写ったビデオを Open Pose など画像認識して動きデータを抽出する。抽出した動きデータを識別モデルのエンコーダ 141, 142 に入力すると、ビデオに写った人物が本人であるか否かを判定できる。なお、識別モデルは、判定対象の人物の動きを学習済みである。

[0047] 図 17 は、再生モデルを利用して、個性を無くした汎用的な動きをするスケルトンを生成する実施例である。

[0048] ソースの被写体のスケルトンの時系列 2 次元データを共通動きエンコーダ 142、視覚角度エンコーダ 143、および体型エンコーダ 144 に入力する。共通動きエンコーダ 142 は、個性をフィルタリングした共通動きを抽出するように学習されている。これにより、ソースの被写体の動きの個性を取り除いた共通動きをするスケルトンの時系列 2 次元データが得られる。個性を無くした汎用的な動きをするスケルトンの時系列 2 次元データを集めることで、動きのデータベースの作成に活用できる。また、様々な視覚角度の時系列 2 次元データを入力することで、様々な視覚角度の時系列 2 次元データを生成できる。

[0049] 以上説明したように、本実施形態の学習装置 1 は、スケルトンの動きを記録した時系列データを入力するデータ入力部 11 と、時系列データを用いて、時間に依存しないスケルトンの特徴を抽出するエンコーダ 143, 144、時間変動する動きの特徴を抽出する 2 つ以上のエンコーダ 141, 142、動きの特徴からスケルトンを識別する識別器 145、およびスケルトンの新たな動きを記録した時系列データを出力する動き再生デコーダ 146 を含むモデル 14 を学習する学習部 15 を備える。学習部 15 は、動きの特徴を抽出するエンコーダ 141, 142 について、敵対的学習により、スケルトンの固有的な動きの特徴を抽出する固有動き特徴エンコーダ 141 と共通す

る動きの特徴を抽出する共通動きエンコーダ142を学習する。本実施形態では、動きの特徴を抽出する2つ以上のエンコーダ141, 142を用意し、敵対的学習により、固有動き特徴エンコーダ141は個人識別を強くする学習を進めて固有の動きを抽出させるとともに、共通動きエンコーダ142は個人識別させない学習を進めて共通する動きを抽出させるので、ターゲット固有の動きを共通する動きに加えた新たな動きを記録した時系列2次元データを生成することができる。その結果、例えば、デジタルツインのサーバ空間内で、誰でも簡単にターゲットとする人物になりきることができる。

[0050] 上記説明した学習装置1には、例えば、図18に示すような、中央演算処理装置(CPU)901と、メモリ902と、ストレージ903と、通信装置904と、入力装置905と、出力装置906とを備える汎用的なコンピュータシステムを用いることができる。このコンピュータシステムにおいて、CPU901がメモリ902上にロードされた所定のプログラムを実行することにより、学習装置1が実現される。このプログラムは磁気ディスク、光ディスク、半導体メモリ等のコンピュータ読み取り可能な記録媒体に記録することも、ネットワークを介して配信することもできる。

符号の説明

- [0051] 1…学習装置
- 11…データ入力部
 - 12…データ配置部
 - 13…データブロック
 - 14…モデル
 - 15…学習部
 - 16…学習済みモデル
 - 141…固有動き特徴エンコーダ (Unique Motion Encoder)
 - 142…共通動きエンコーダ (Common Motion Encoder)

1 4 3…視覚角度エンコーダ視覚角度エンコーダ (V i e w - A n g l e
E n c o d e r)

1 4 4…体型エンコーダ (S k e l e t o n E n c o d e r)

1 4 5…識別器 (C l a s s i f i e r)

1 4 6…動き再生デコーダ (M o t i o n R e c o n s t r u c t i o
n D e c o d e r)

請求の範囲

- [請求項1] 動体の動きを記録した時系列データを入力する入力部と、
前記時系列データを用いて、時間に依存しない前記動体の特徴を抽出するエンコーダ、時間変動する動きの特徴を抽出する2つ以上のエンコーダ、前記動きの特徴から前記動体を識別する識別器、および前記動体の新たな動きを記録した時系列データを出力するデコーダを含むモデルを学習する学習部を備え、
前記学習部は、前記動きの特徴を抽出する2つ以上のエンコーダについて、敵対的学習により、前記動体の固有的な動きの特徴を抽出するエンコーダと共通する動きの特徴を抽出するエンコーダを学習する学習装置。
- [請求項2] 請求項1に記載の学習装置であって、
前記学習部は、前記識別器の誤差を低くするように誤差関数の勾配を逆伝搬させて学習し、誤差逆伝搬時に、前記共通する動きの特徴を抽出するエンコーダへは負の定数を乗算した前記勾配を逆伝搬する学習装置。
- [請求項3] 請求項1または2に記載の学習装置であって、
前記時間に依存しない前記動体の特徴は、前記動体の体型および前記動体の角度である
学習装置。
- [請求項4] コンピュータが実行する学習方法であって、
動体の動きを記録した時系列データを入力し、
前記時系列データを用いて、時間に依存しない前記動体の特徴を抽出するエンコーダ、時間変動する動きの特徴を抽出する2つ以上のエンコーダ、前記動きの特徴から前記動体を識別する識別器、および前記動体の新たな動きを記録した時系列データを出力するデコーダを含むモデルを学習し、
前記モデルを学習する際、前記動きの特徴を抽出する2つ以上のエ

ンコードについて、敵対的学習により、前記動体の固有的な動きの特徴を抽出するエンコードと共通する動きの特徴を抽出するエンコードを学習する

学習方法。

[請求項5]

請求項4に記載の学習方法であって、

前記モデルを学習する際、前記識別器の誤差を低くするように誤差関数の勾配を逆伝搬させて学習し、誤差逆伝搬時に、前記共通する動きの特徴を抽出するエンコードへは負の定数を乗算した前記勾配を逆伝搬する

学習方法。

[請求項6]

請求項4または5に記載の学習方法であって、

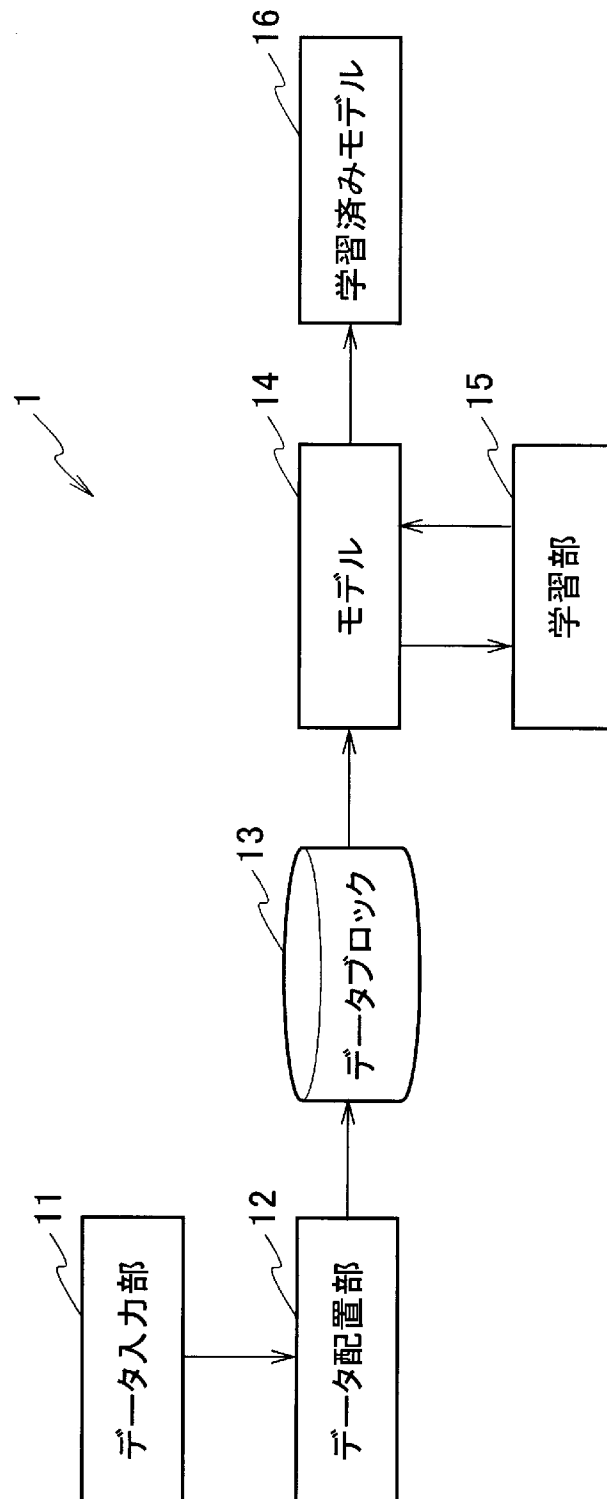
前記時間に依存しない前記動体の特徴は、前記動体の体型および前記動体の角度である

学習方法。

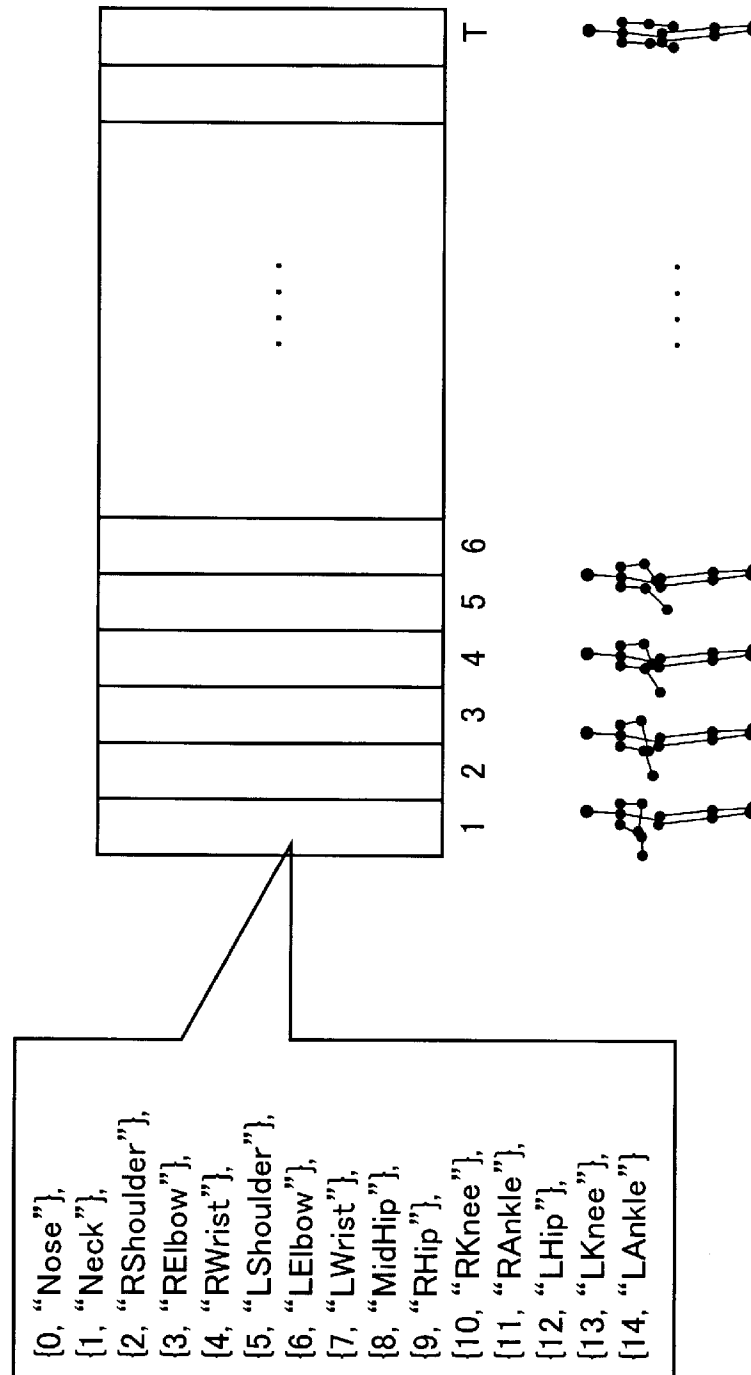
[請求項7]

請求項1ないし3のいずれかに記載の学習装置の各部としてコンピュータを動作させるプログラム。

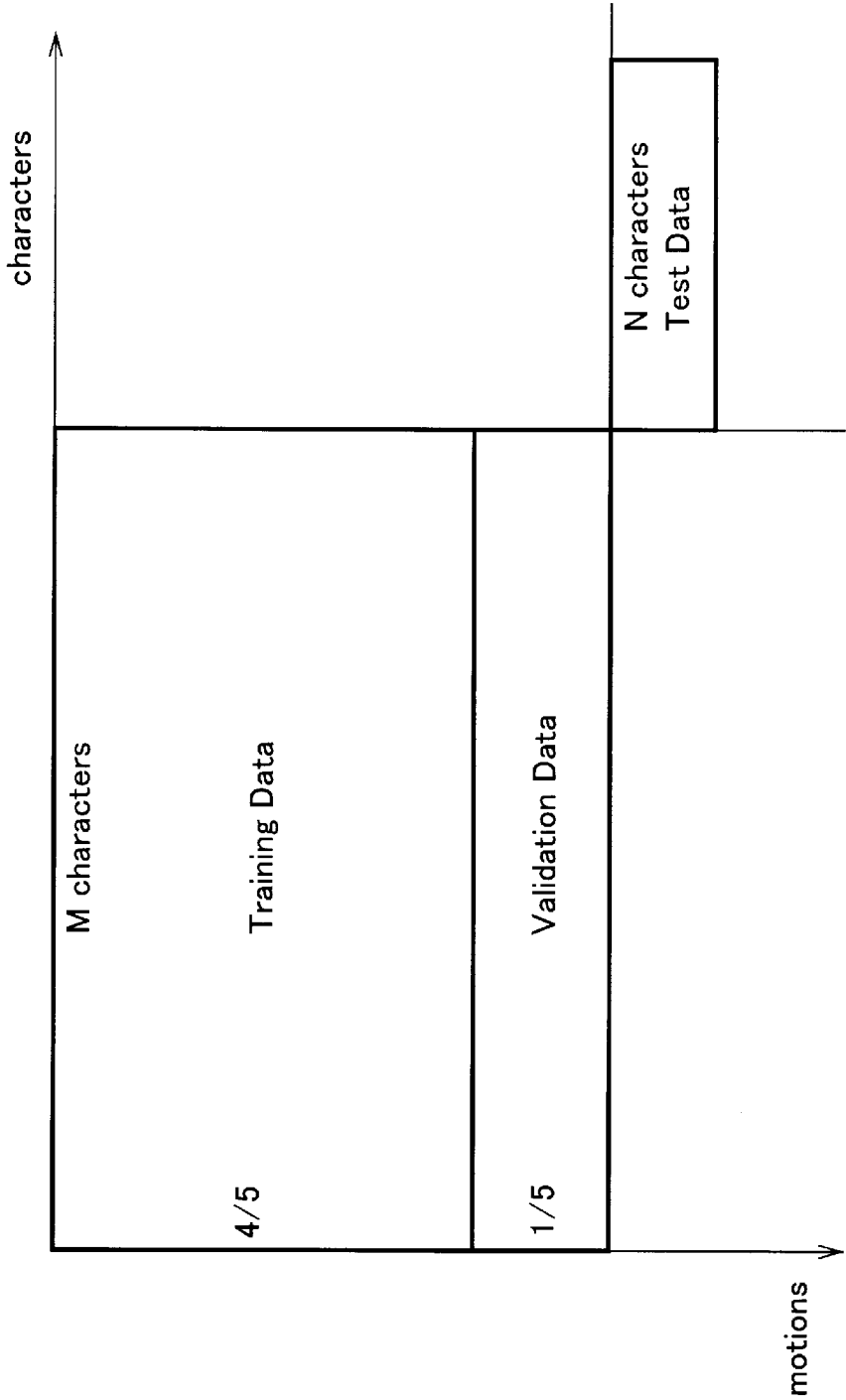
[図1]



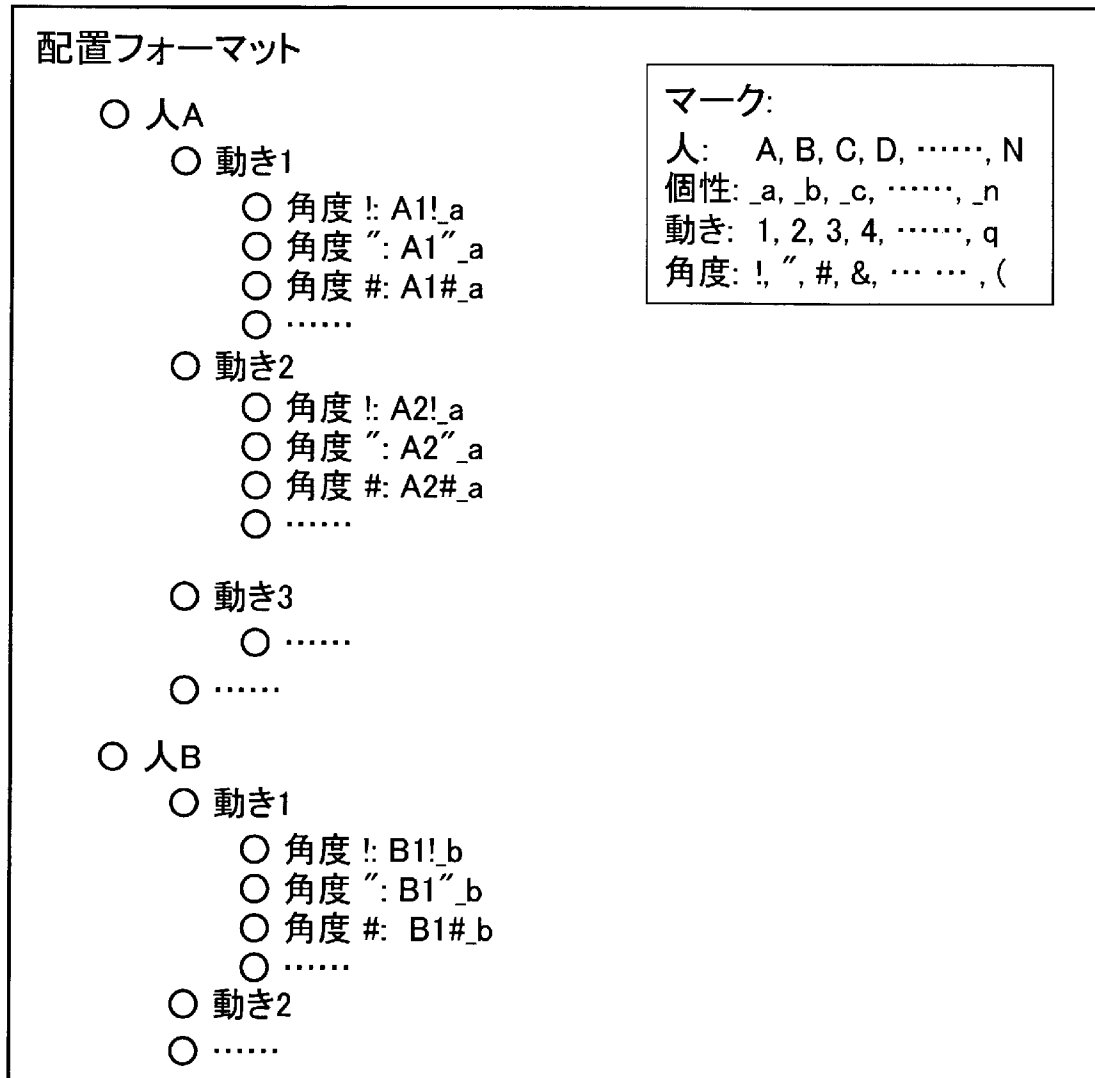
[図2]



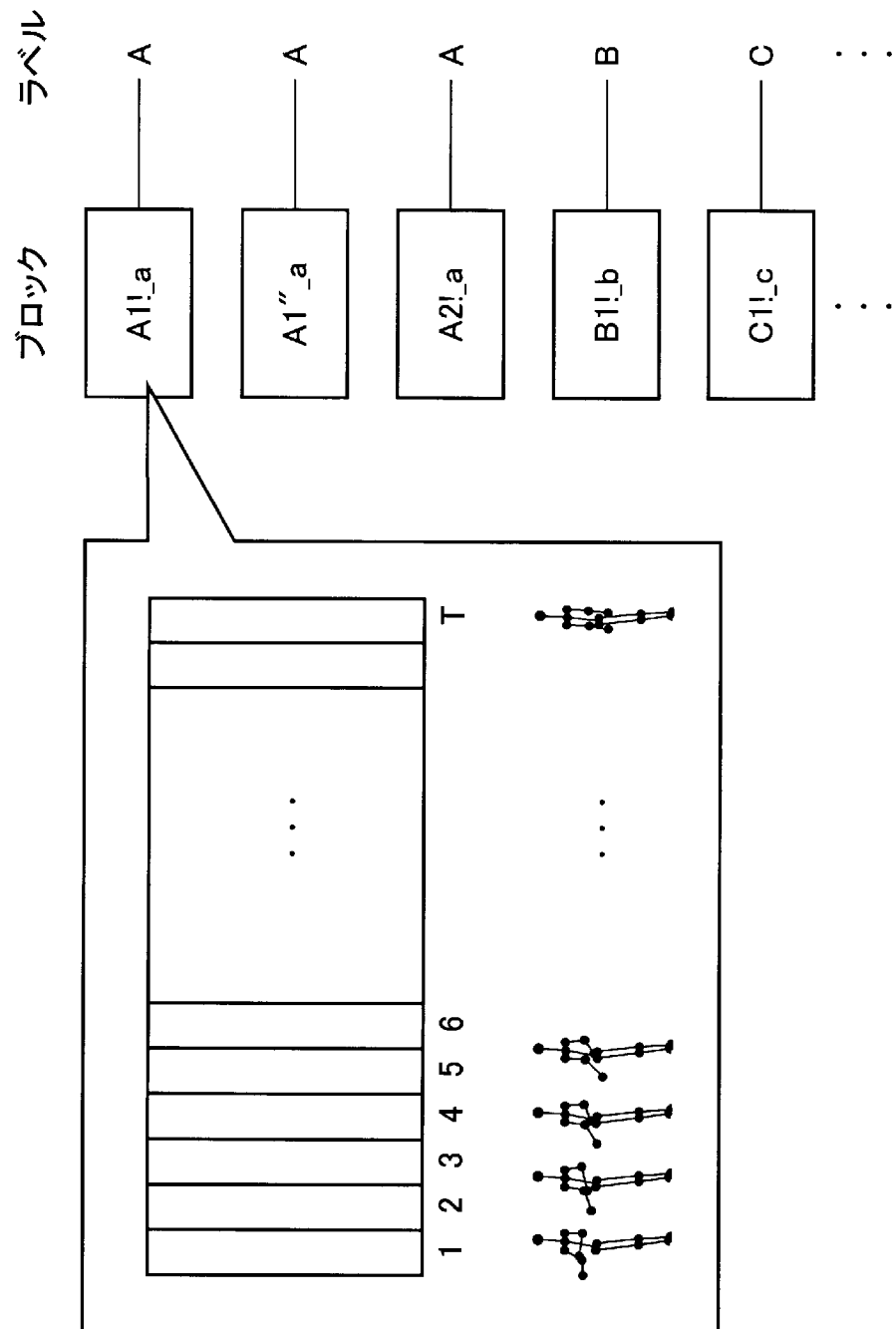
[図3]



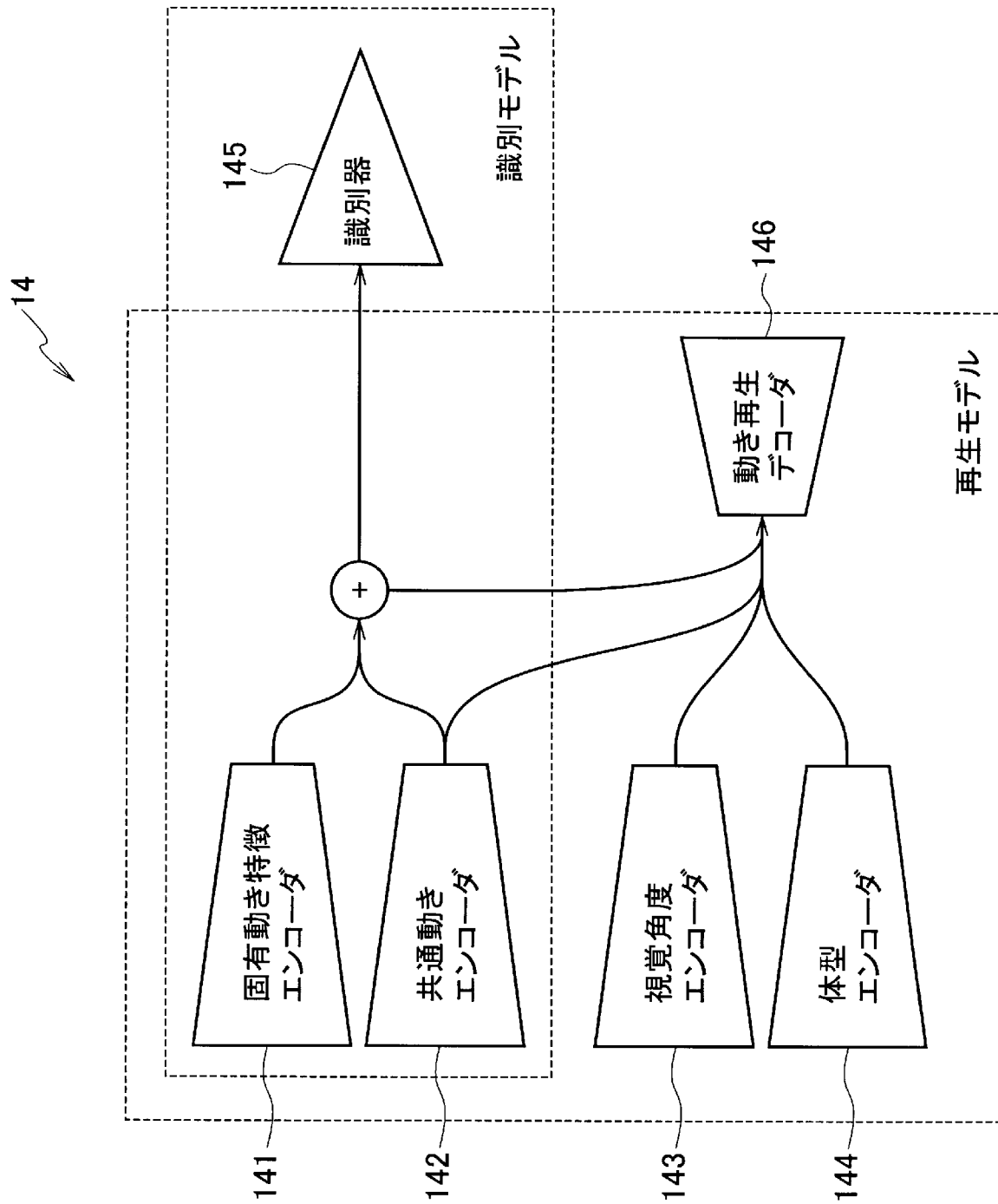
[図4]



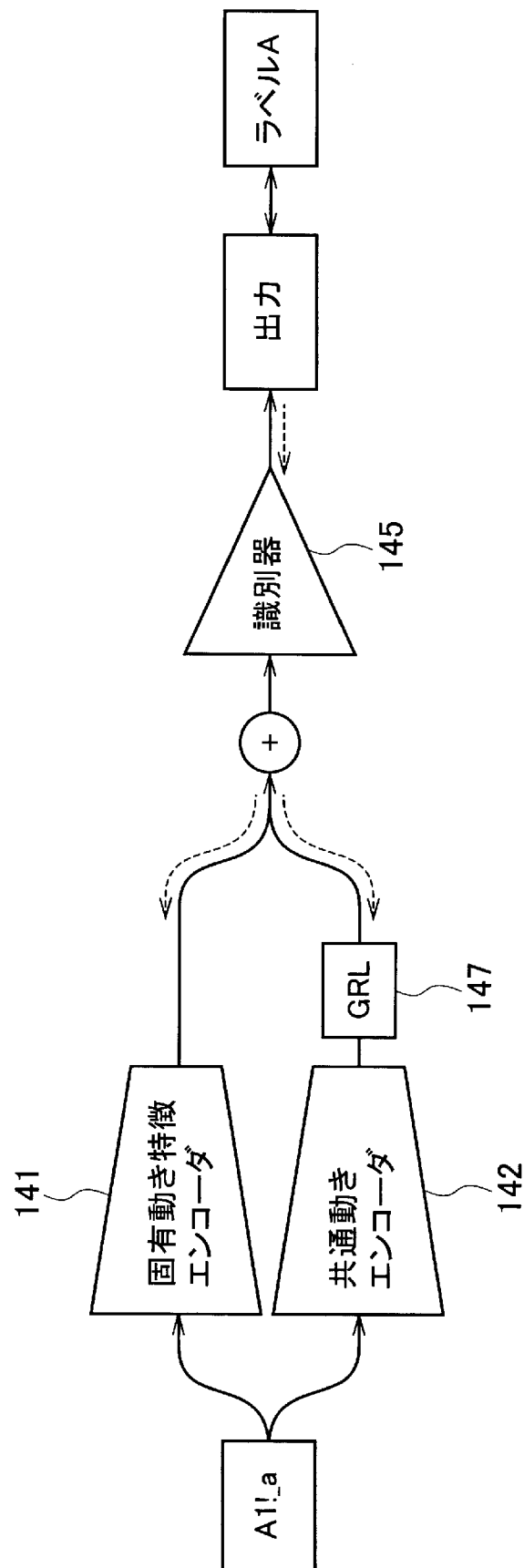
[図5]



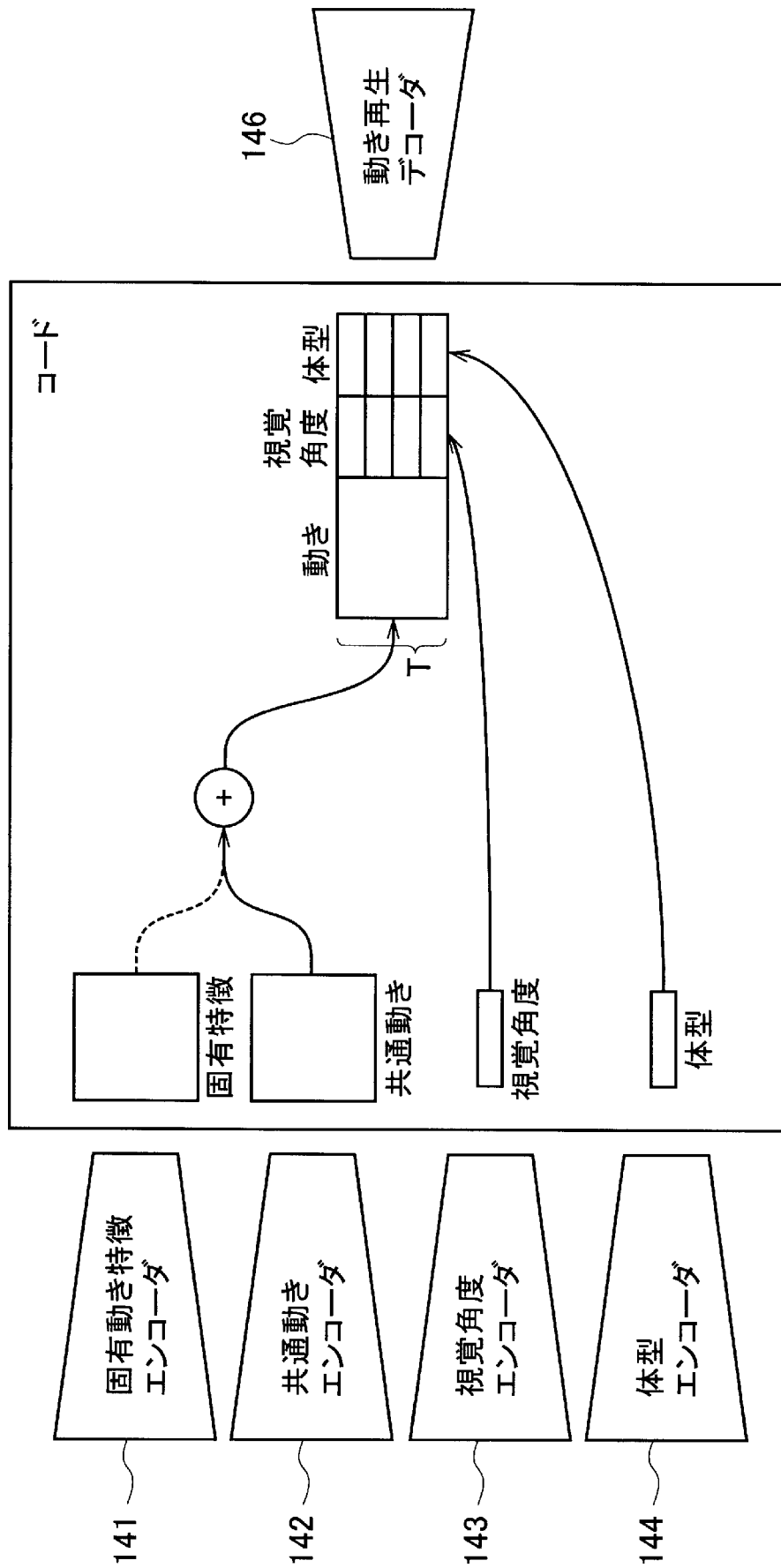
[図6]



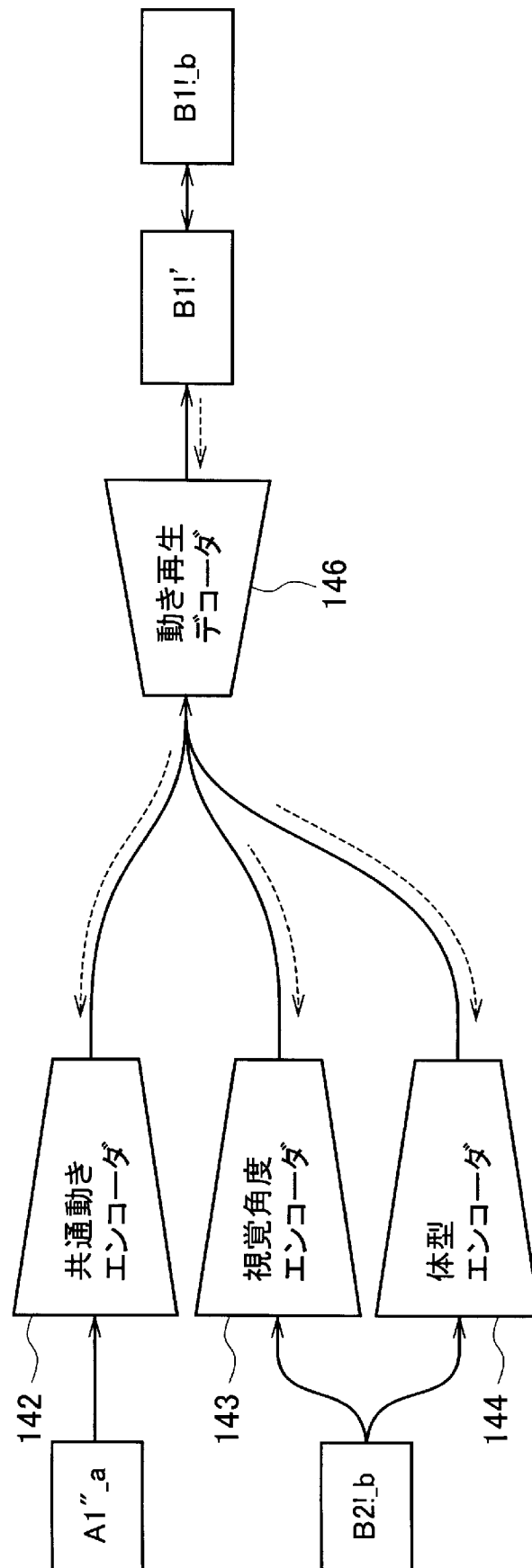
[図7]



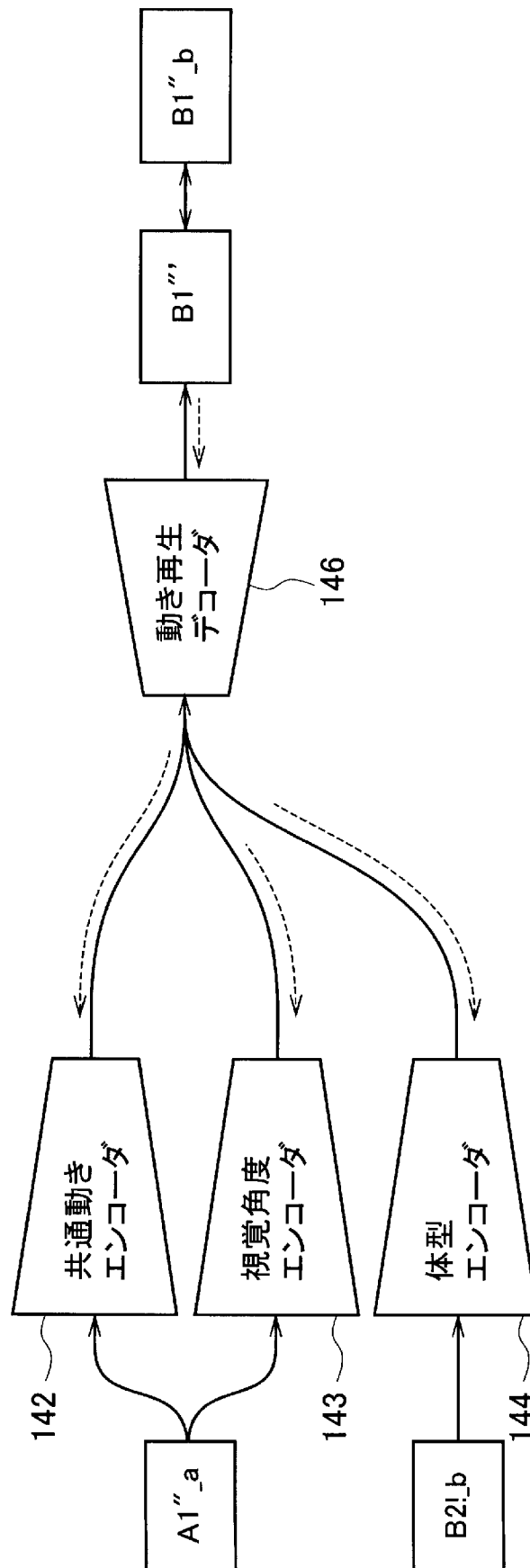
[図8]



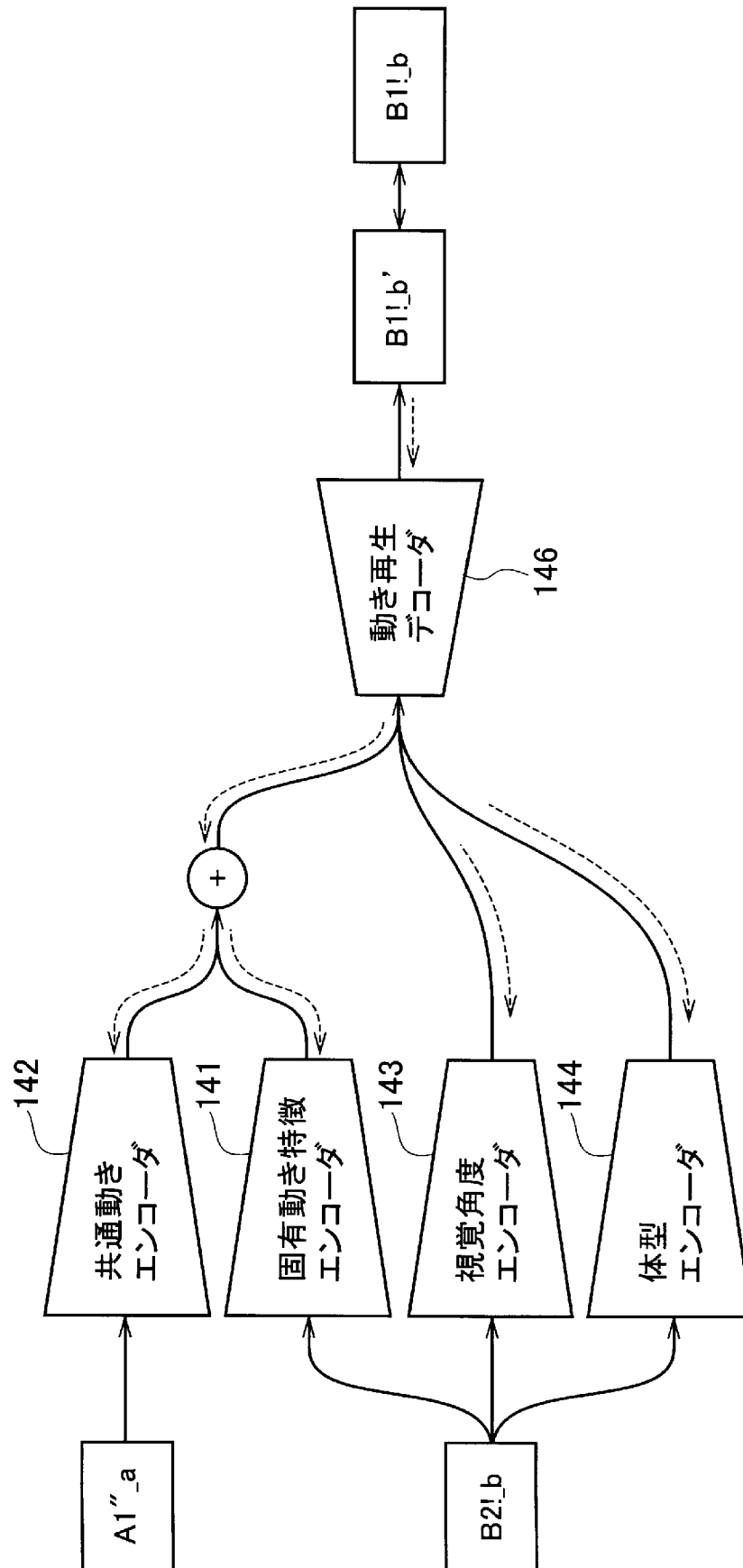
[図9]



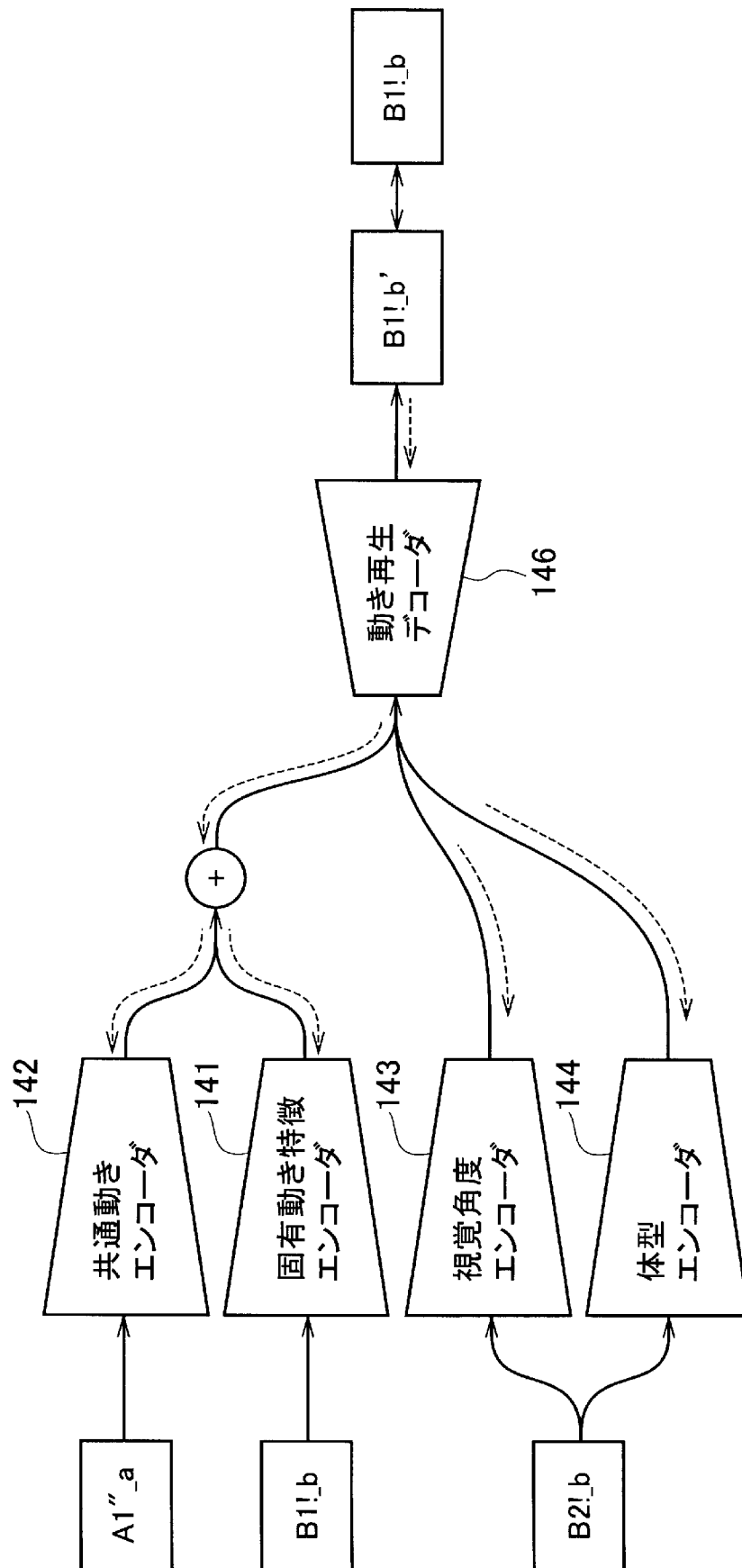
[図10]



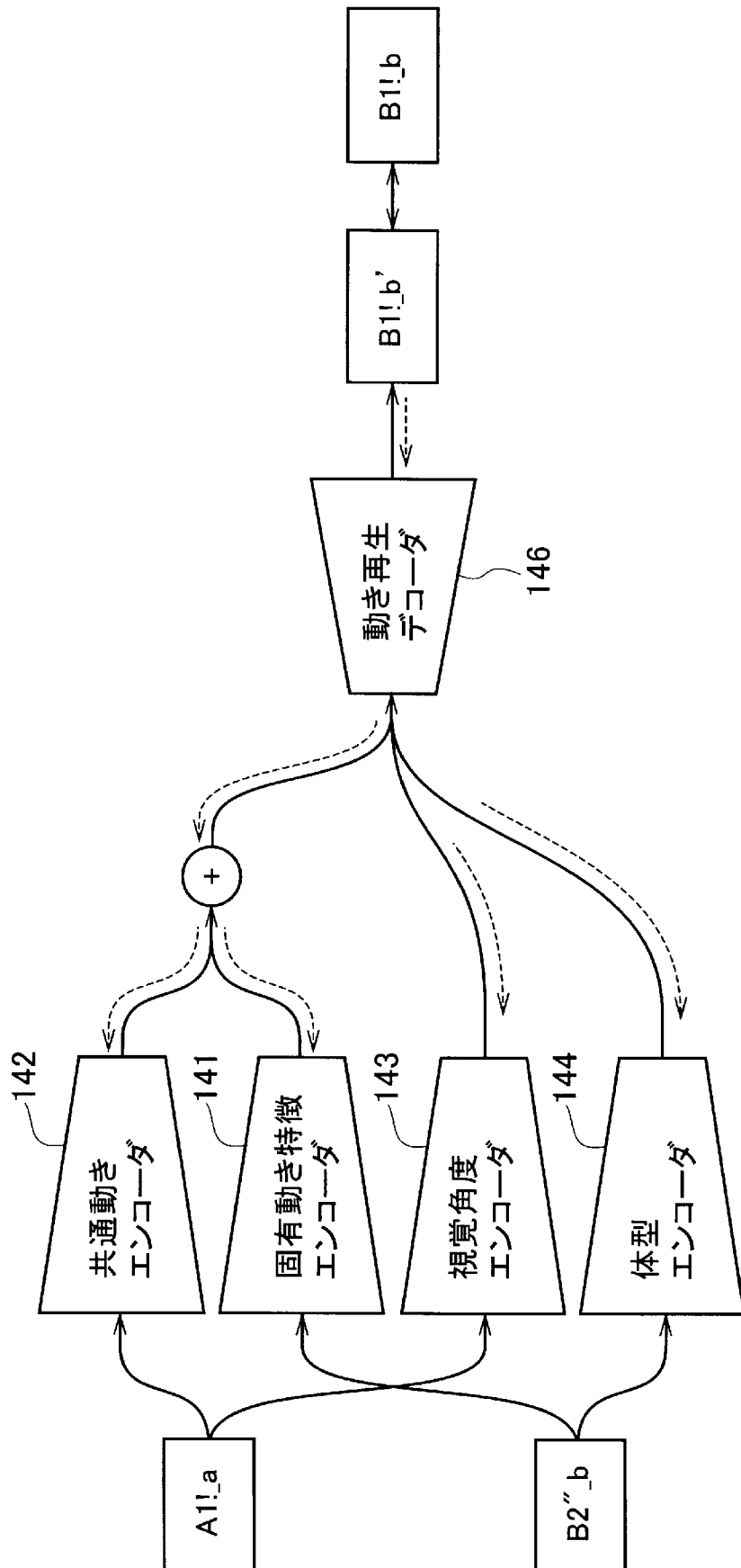
[図11]



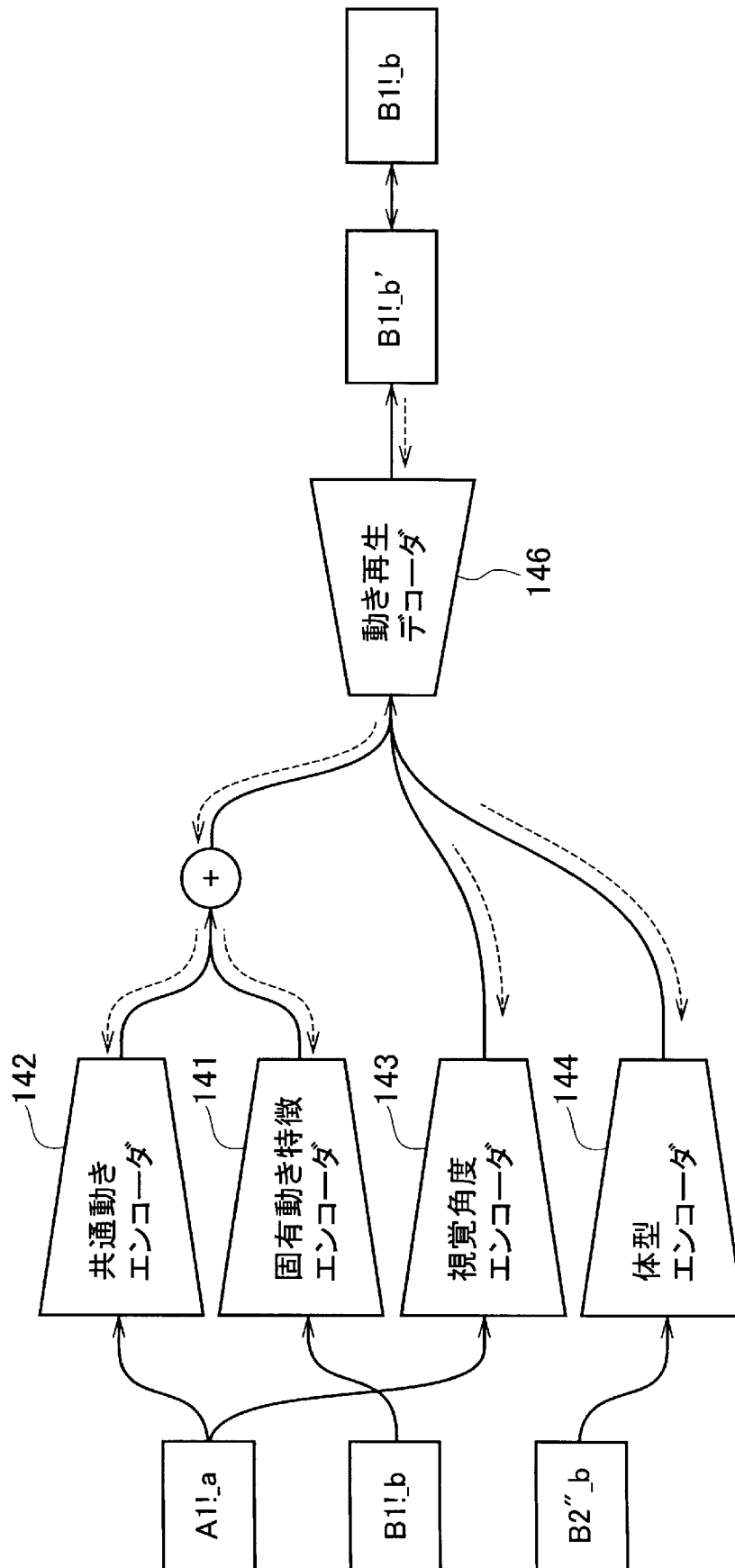
[図12]



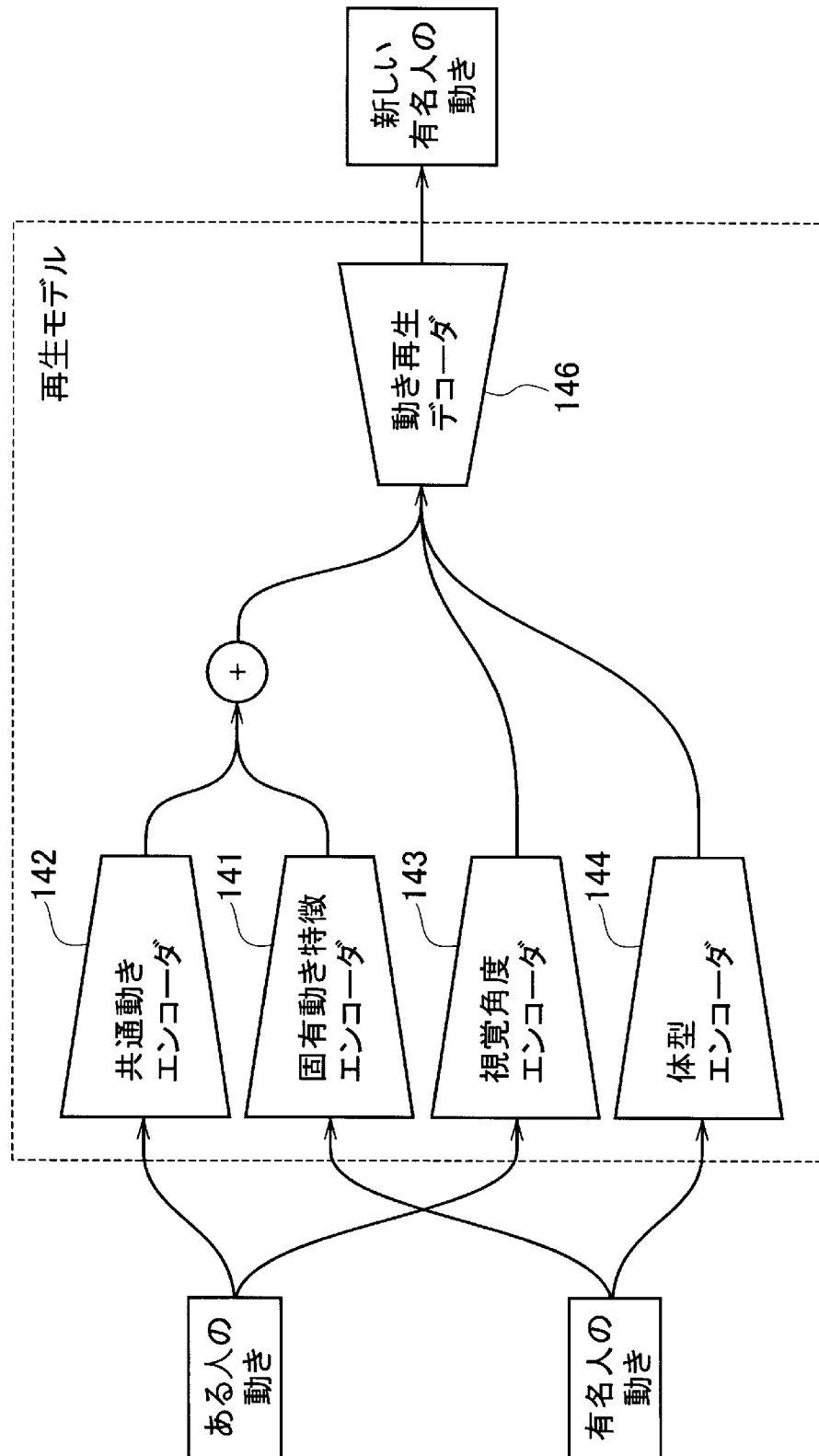
[図13]



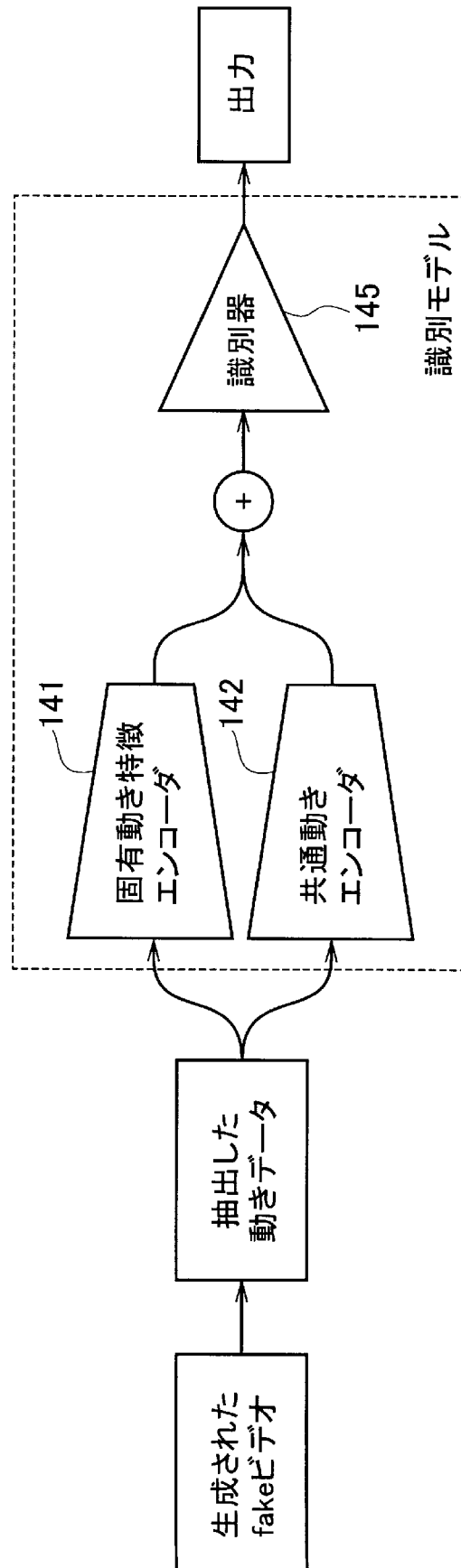
[図14]



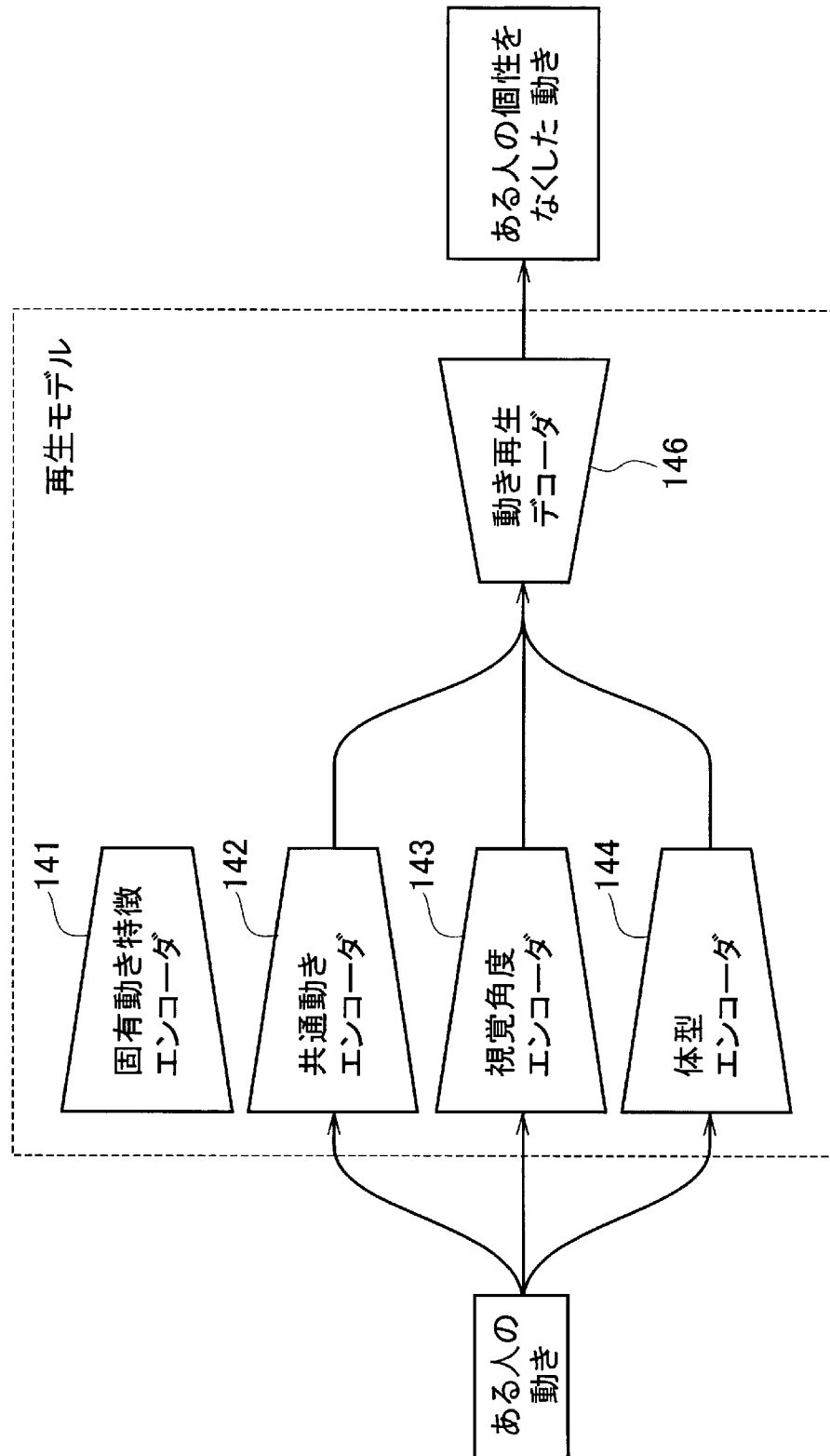
[図15]



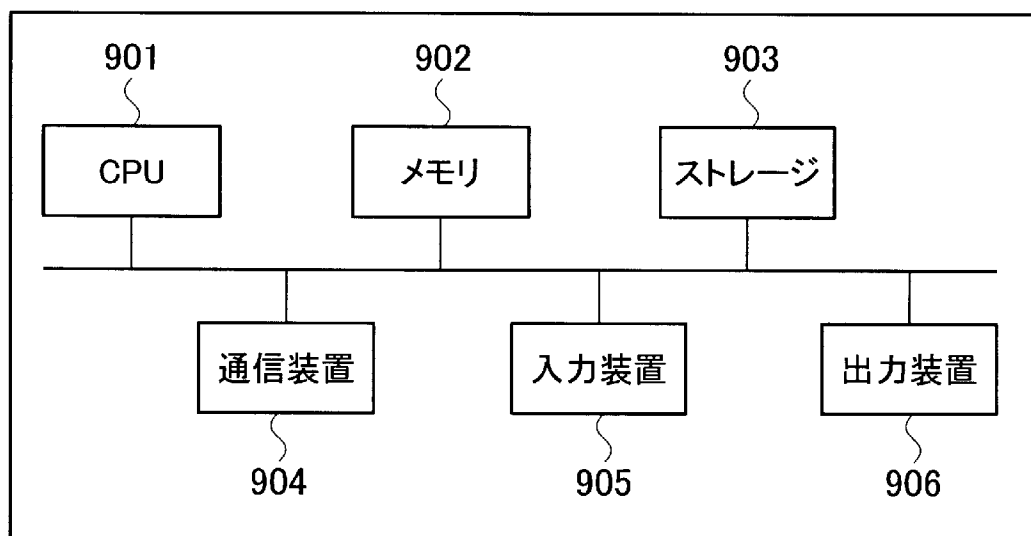
[図16]



[図17]



[図18]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2020/026683

A. CLASSIFICATION OF SUBJECT MATTER

Int. Cl. G06T13/40 (2011.01) i, G06T7/00 (2017.01) i, G06T7/20 (2017.01) i, G06N3/08 (2006.01) i

FI: G06N3/08, G06T7/00 350C, G06T7/20 300Z, G06T13/40

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Int. Cl. G06N3/00-99/00, G06T7/00-7/90, G06T13/40

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Published examined utility model applications of Japan 1922-1996

Published unexamined utility model applications of Japan 1971-2020

Registered utility model specifications of Japan 1996-2020

Published registered utility model applications of Japan 1994-2020

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	YANG, Zhuogian et al. TransMoMo: Invariance-Driven Unsupervised Video Motion Retargeting, arXiv.org [online], arXiv:2003.14401v2, Cornell University, April 2020, [retrieved on 02 November 2020], Internet: <URL: https://arxiv.org/pdf/2003.14401v2 > chapter 3	1-7

☐ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search
02.11.2020

Date of mailing of the international search report
17.11.2020

Name and mailing address of the ISA/
Japan Patent Office
3-4-3, Kasumigaseki, Chiyoda-ku,
Tokyo 100-8915, Japan

Authorized officer

Telephone No.

A. 発明の属する分野の分類（国際特許分類（IPC）） G06T 13/40(2011.01)i; G06T 7/00(2017.01)i; G06T 7/20(2017.01)i; G06N 3/08(2006.01)i FI: G06N3/08; G06T7/00 350C; G06T7/20 300Z; G06T13/40										
B. 調査を行った分野										
調査を行った最小限資料（国際特許分類（IPC）） G06N3/00-99/00; G06T7/00-7/90; G06T13/40										
最小限資料以外の資料で調査を行った分野に含まれるもの										
<table border="0"> <tr> <td>日本国実用新案公報</td> <td>1922 - 1996年</td> </tr> <tr> <td>日本国公開実用新案公報</td> <td>1971 - 2020年</td> </tr> <tr> <td>日本国実用新案登録公報</td> <td>1996 - 2020年</td> </tr> <tr> <td>日本国登録実用新案公報</td> <td>1994 - 2020年</td> </tr> </table>			日本国実用新案公報	1922 - 1996年	日本国公開実用新案公報	1971 - 2020年	日本国実用新案登録公報	1996 - 2020年	日本国登録実用新案公報	1994 - 2020年
日本国実用新案公報	1922 - 1996年									
日本国公開実用新案公報	1971 - 2020年									
日本国実用新案登録公報	1996 - 2020年									
日本国登録実用新案公報	1994 - 2020年									
国際調査でを使用した電子データベース（データベースの名称、調査に使用した用語）										
C. 関連すると認められる文献										
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号								
A	YANG, Zhuoqian et al., TransMoMo: Invariance-Driven Unsupervised Video Motion Retargeting, arXiv.org [online], arXiv:2003.14401v2, Cornell University, 2020.04, [検索日 2020.11.02], インターネット:<URL:https://arxiv.org/pdf/2003.14401v2> 第3章	1-7								
☐ C欄の続きにも文献が列挙されている。 ☐ パテントファミリーに関する別紙を参照。										
<table border="0"> <tr> <td style="vertical-align: top;"> * 引用文献のカテゴリー “A” 特に関連のある文献ではなく、一般的な技術水準を示すもの “E” 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの “L” 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す） “O” 口頭による開示、使用、展示等に言及する文献 “P” 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献 </td> <td style="vertical-align: top;"> “T” 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの “X” 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの “Y” 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの “&” 同一パテントファミリー文献 </td> </tr> </table>			* 引用文献のカテゴリー “A” 特に関連のある文献ではなく、一般的な技術水準を示すもの “E” 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの “L” 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す） “O” 口頭による開示、使用、展示等に言及する文献 “P” 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献	“T” 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの “X” 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの “Y” 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの “&” 同一パテントファミリー文献						
* 引用文献のカテゴリー “A” 特に関連のある文献ではなく、一般的な技術水準を示すもの “E” 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの “L” 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す） “O” 口頭による開示、使用、展示等に言及する文献 “P” 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献	“T” 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの “X” 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの “Y” 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの “&” 同一パテントファミリー文献									
国際調査を完了した日 02.11.2020	国際調査報告の発送日 17.11.2020									
名称及びあて先 日本国特許庁(ISA/JP) 〒100-8915 日本国 東京都千代田区霞が関三丁目4番3号	権限のある職員（特許庁審査官） 多賀 実 5B 9367 電話番号 03-3581-1101 内線 3545									