

(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
2024年3月21日(21.03.2024)



(10) 国際公開番号

WO 2024/057469 A1

(51) 国際特許分類:

G06T 7/00 (2017.01)

(21) 国際出願番号: PCT/JP2022/034510

(22) 国際出願日: 2022年9月15日(15.09.2022)

(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: 日本電気株式会社 (NEC CORPORATION) [JP/JP]; 〒1088001 東京都港区芝五丁目7番1号 Tokyo (JP).

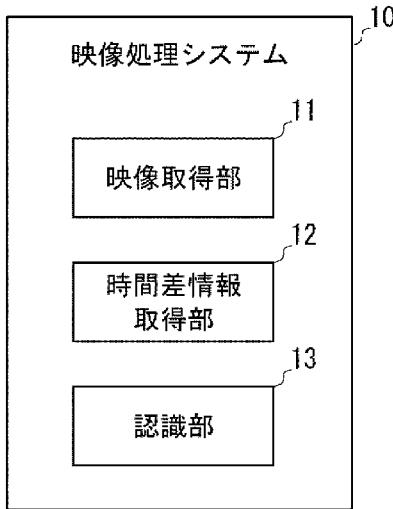
(72) 発明者: バイエ フロリアン (BEYE Florian); 〒1088001 東京都港区芝五丁目7番1号 日本電気株式会社内 Tokyo (JP). 岩井 孝法 (IWAI Takanori); 〒1088001 東京都港区芝五丁目7

番1号 日本電気株式会社内 Tokyo (JP). 二瓶 浩一 (NIHEI Koichi); 〒1088001 東京都港区芝五丁目7番1号 日本電気株式会社内 Tokyo (JP). 逸身 勇人 (ITSUMI Hayato); 〒1088001 東京都港区芝五丁目7番1号 日本電気株式会社内 Tokyo (JP). 高橋 勝彦 (TAKAHASHI Katsuhiko); 〒1088001 東京都港区芝五丁目7番1号 日本電気株式会社内 Tokyo (JP). 馬場崎 康敬 (BABAZAKI Yasunori); 〒1088001 東京都港区芝五丁目7番1号 日本電気株式会社内 Tokyo (JP). 安藤 隆平 (ANDO Ryuhei); 〒1088001 東京都港区芝五丁目7番1号 日本電気株式会社内 Tokyo (JP). 朴 君 (PIAO Jun); 〒1088001 東京都港区芝五丁目7番1号 日本電気株式会社内 Tokyo (JP).

(54) Title: VIDEO PROCESSING SYSTEM, VIDEO PROCESSING DEVICE, AND VIDEO PROCESSING METHOD

(54) 発明の名称: 映像処理システム、映像処理装置および映像処理方法

[図1]



10... VIDEO PROCESSING SYSTEM
11... VIDEO ACQUIRING UNIT
12... TIME DIFFERENCE INFORMATION ACQUIRING UNIT
13... RECOGNIZING UNIT

Fig. 1

(57) Abstract: The objective of the present invention is to provide a video processing system, a video processing device, and a video processing method that can be expected to improve the recognition accuracy of objects in a video. A video processing system (10) comprises a video acquiring means (11), a time difference information acquiring means (12), and a recognizing means (13). The video acquiring means (11) acquires an input video. The time difference information acquiring means (12) acquires first time difference information between frames of the input video. The recognizing means (13) recognizes an object in the input video by inputting the input video and the first time difference information between



(74) 代理人: 家入 健 (IEIRI Takeshi); 〒2210835 神奈川県横浜市神奈川区鶴屋町三丁目33番8 アーバンセンター横浜ウエスト5階 響国際特許事務所 Kanagawa (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, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, 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, SC, 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 frames of the input video into a trained recognition model that has been trained using a training video and second time difference information between frames of the training video.

(57) 要約: 映像内の物体の認識精度を向上することが期待できる映像処理システム、映像処理装置および映像処理方法を提供することを目的とする。映像処理システム(10)は、映像取得手段(11)、時間差情報取得手段(12)、認識手段(13)を備える。映像取得手段(11)は、入力映像を取得する。時間差情報取得手段(12)は、前記入力映像のフレーム間の第1の時間差情報を取得する。認識手段(13)は、学習映像と学習映像のフレーム間の第2の時間差情報とを用いて学習した学習済みの認識モデルに、入力映像と入力映像のフレーム間の第1の時間差情報とを入力し、入力映像内の物体を認識する。

明 細 書

発明の名称：

映像処理システム、映像処理装置および映像処理方法

技術分野

[0001] 本開示は、映像処理システム、映像処理装置および映像処理方法に関する。

背景技術

[0002] エッジ側の端末が撮影された映像をセンター側のサーバに送信し、センター側のサーバがAIエンジンによって当該映像内の物体を認識する技術が開発されている。例えば、サーバは、作業員が行う作業の種類を認識する。ここで、エッジ側の端末は、計算リソースの効率的な使用やネットワーク帯域の効率的な使用のためにフレームフィルタリングなどによって映像のフレームレート変更を行う。

[0003] 関連する技術として、特許文献1には、RNN (Recurrent Neural Network) などの深層学習アルゴリズムを用いて、映像から抽出された時系列のフレームから、映像シーン認識を行う技術が開示されている。

先行技術文献

特許文献

[0004] 特許文献1：特開2018-005638号公報

発明の概要

発明が解決しようとする課題

[0005] 特許文献1などの技術では、センター側のサーバは、映像のフレームレート変更に対応していないため、映像のフレームレート変更に対応した物体の認識ができなく、映像内の物体の認識精度に改善の余地があった。

[0006] 本開示は、そのような課題を鑑みることによって、映像内の物体の認識精度を向上することが期待できる映像処理システム、映像処理装置および映像処理方法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

- [0007] 本開示の映像処理システムは、
入力映像を取得する映像取得手段と、
前記入力映像のフレーム間の第1の時間差情報を取得する時間差情報取得手段と、
学習映像と前記学習映像のフレーム間の第2の時間差情報とを用いて学習した学習済みの認識モデルに、前記入力映像と前記入力映像のフレーム間の第1の時間差情報とを入力し、前記入力映像内の物体を認識する認識手段と、
を備える。
- [0008] 本開示の映像処理装置は、
入力映像を取得する映像取得手段と、
前記入力映像のフレーム間の第1の時間差情報を取得する時間差情報取得手段と、
学習映像と前記学習映像のフレーム間の第2の時間差情報とを用いて学習した学習済みの認識モデルに、前記入力映像と前記入力映像のフレーム間の第1の時間差情報とを入力し、前記入力映像内の物体を認識する認識手段と、
を備える。
- [0009] 本開示の映像処理方法は、
コンピュータが、
入力映像を取得し、
前記入力映像のフレーム間の第1の時間差情報を取得し、
学習映像と前記学習映像のフレーム間の第2の時間差情報とを用いて学習した学習済みの認識モデルに、前記入力映像と前記入力映像のフレーム間の第1の時間差情報とを入力し、前記入力映像内の物体を認識する。

発明の効果

- [0010] 本開示によって、映像内の物体の認識精度を向上することが期待できる映像処理システム、映像処理装置および映像処理方法を提供することができる。

図面の簡単な説明

[0011] [図1]実施形態の概要に係る映像処理システムの構成を示すブロック図である

。

[図2]実施形態の概要に係る映像処理装置の構成を示すブロック図である。

[図3]実施形態の概要に係る映像処理方法を示すフローチャートである。

[図4]第1の実施形態に係る映像処理システムの構成を示すブロック図である

。

[図5]第1の実施形態に係る端末の構成を示すブロック図である。

[図6]第1の実施形態に係るセンターサーバの構成を示すブロック図である。

[図7]第1の実施形態に係る映像処理システムの動作を示すフローチャートである。

[図8]第1の実施形態に係る学習済みの認識モデルの入力情報の一例を示す図である。

[図9]第1の実施形態に係る学習済みの認識モデルの構成および認識動作の一例を示す図である。

[図10]第2の実施形態に係るセンターサーバの構成を示すブロック図である

。

[図11]第2の実施形態に係る映像処理システムの動作の一例を示すフローチャートである。

[図12]第2の実施形態に係る学習済みの認識モデルの構成および認識動作の一例を示す図である。

[図13]第2の実施形態に係る認識モデルの第1の学習動作の一例を示す図である。

[図14]第2の実施形態に係る認識モデルの第2の学習動作の一例を示す図である。

[図15]第2の実施形態に係る学習済みの認識モデルの構成および認識動作の他の一例を示す図である

[図16]本実施形態に係るコンピュータの構成を示すブロック図である。

発明を実施するための形態

[0012] 以下では、本開示の実施形態について、図面を参照しながら詳細に説明する。各図面において、同一又は対応する要素には同一の符号が付されており、説明の明確化のため、必要に応じて重複説明は省略される。

[0013] (実施形態の概要)

まず、図1を用いて、実施形態の概要に係る映像処理システム10について説明する。図1は、実施形態の概要に係る映像処理システム10の構成を示すブロック図である。映像処理システム10は、例えば、ネットワークを介して映像を収集し、映像を認識する遠隔監視システムに適用可能である。

[0014] 図1に示すように、映像処理システム10は、映像取得部11、時間差情報取得部12および認識部13を備える。映像取得部11は、入力映像を取得する。時間差情報取得部12は、前記入力映像のフレーム間の第1の時間差情報を取得する。認識部13は、学習映像と学習映像のフレーム間の第2の時間差情報とを用いて学習した学習済みの認識モデルに、入力映像と入力映像のフレーム間の第1の時間差情報とを入力し、入力映像内の物体を認識する。

[0015] 続いて、図2を用いて、実施形態の概要に係る映像処理装置20の構成について説明する。図2は、実施形態の概要に係る映像処理装置20の構成を示すブロック図である。図2に示すように、映像処理装置20は、図1に示した映像取得部11、時間差情報取得部12および認識部13を備える。また、映像処理装置20は、エッジコンピューティングで実現される場合、映像処理装置20の一部または全部をエッジまたはクラウドに配置してもよい。例えば、エッジの端末に映像取得部11、時間差情報取得部12を配置し、クラウドのサーバに認識部13を配置してもよい。さらに、クラウドに各機能を分散配置してもよい。また、映像処理装置20は、仮想化サーバなど仮想化技術で実現されてもよい。また、映像処理装置20の一部または全部を現場側またはサーバ側に配置してもよい。端末が設置された現場、現場から近い場所に配置されるもの、またはネットワークの階層として端末に近い

装置を現場側に配置された装置とする。また、現場から離れたところにある装置をセンター側に配置された装置とする。センター側に配置された装置は、クラウド上に配置されることもあるため、センター側をクラウド側と称することもある。

[0016] 続いて、図3を用いて、実施形態の概要に係る映像処理方法について説明する。図3は、実施形態の概要に係る映像処理方法を示すフローチャートである。例えば、実施形態に係る映像処理方法は、図1の映像処理システム10や図2の映像処理装置20により実行される。

[0017] 図3に示すように、入力映像を取得する（ステップS11）。次に、入力映像のフレーム間の第1の時間差情報を取得する（ステップS12）。次に、学習映像と学習映像のフレーム間の第2の時間差情報とを用いて学習した学習済みの認識モデルに、入力映像と入力映像のフレーム間の第1の時間差情報とを入力し、入力映像内の物体を認識する（ステップS13）。

[0018] 上述したように、映像処理システム10は、入力映像のフレーム間の時間差情報を考慮することで、映像のフレームレート変更に対応した物体の認識ができる。映像処理システム10では、映像内の物体の認識精度の向上することが期待できる。

[0019] （映像処理システムの基本構成）

次に、図4を用いて、実施形態を適用するシステムの一例である映像処理システム1について説明する。図4は、第1の実施形態に係る映像処理システム1の構成を示すブロック図である。図4に示すように、映像処理システム1は、カメラが撮影した映像により、当該撮影されたエリアを監視するシステムである。本実施形態においては、以降現場における作業員の作業を遠隔で監視するシステムであるものとして説明する。例えば、現場は工事現場などの作業現場、人の集まる広場、学校など、人や機械が動作するエリアであってもよい。本実施形態においては、以降作業は建設作業や土木作業等として説明するが、これに限られない。なお、映像は、時系列の画像である時系列のフレームを含むため、映像と画像とは互いに言い換え可能である。す

なわち、映像処理システムは、映像を処理する映像処理システムであり、また、画像を処理する画像処理システムであるとも言える。

[0020] 映像処理システム１は、複数の端末１００、センターサーバ２００、基地局３００、ＭＥＣ４００を備えている。端末１００、基地局３００及びＭＥＣ４００は、現場側に配置され、センターサーバ２００は、センター側に配置されている。例えば、センターサーバ２００は、現場から離れた位置に配置されているデータセンタ等に配置されている。現場側はシステムのエッジ側であり、センター側はクラウド側でもある。

[0021] 端末１００と基地局３００との間は、ネットワークＮＷ１により通信可能に接続される。ネットワークＮＷ１は、例えば、４Ｇ、ローカル５Ｇ／５Ｇ、ＬＴＥ（Long Term Evolution）、無線ＬＡＮなどの無線ネットワークである。基地局３００とセンターサーバ２００との間は、ネットワークＮＷ２により通信可能に接続される。ネットワークＮＷ２は、例えば、５ＧＣ（5th Generation Core network）やＥＰＣ（Evolved Packet Core）などのコアネットワーク、インターネットなどを含む。端末１００とセンターサーバ２００との間は、基地局３００を介して、通信可能に接続されているとも言える。基地局３００とＭＥＣ４００の間は任意の通信方法により通信可能に接続されるが、基地局３００とＭＥＣ４００は、１つの装置でもよい。

[0022] 端末１００は、ネットワークＮＷ１に接続される端末装置であり、現場の映像を生成する映像生成装置でもある。端末１００は、現場に設置されたカメラ１０１が撮影した映像を取得し、取得した映像を、基地局３００を介して、センターサーバ２００へ送信する。なお、カメラ１０１は、端末１００の外部に配置されてもよいし、端末１００の内部に配置されてもよい。

[0023] 端末１００は、カメラ１０１の映像を所定のビットレートに圧縮し、圧縮した映像を送信する。端末１００は、圧縮効率を最適化する圧縮効率最適化機能１０２、映像配信機能１０３を有する。圧縮効率最適化機能１０２は、ＲＯＩ（Region of Interest；注視領域とも称する）の画質を制御するＲＯＩ制御を行う。圧縮効率最適化機能１０２は、人物や物体を含むＲＯＩの画

質を維持しながら、その周りの領域の画質を低画質にすることでビットレートを削減する。映像配信機能103は、画質が制御された映像をセンターサーバ200へ配信する。

[0024] 基地局300は、ネットワークNW1の基地局装置であり、端末100とセンターサーバ200の間の通信を中継する中継装置でもある。例えば、基地局300は、ローカル5Gの基地局、5GのgNB（next Generation Node B）、LTEのeNB（evolved Node B）、無線LANのアクセスポイント等であるが、その他の中継装置でもよい。

[0025] MEC（Multi-access Edge Computing）400は、システムのエッジ側に配置されたエッジ処理装置である。MEC400は、端末100を制御するエッジサーバであり、端末のビットレートを制御する圧縮ビットレート制御機能401、端末制御機能402を有する。圧縮ビットレート制御機能401は、適応映像配信制御やQoE（quality of experience）制御により端末100のビットレートを制御する。例えば、圧縮ビットレート制御機能401は、ネットワークNW1及びNW2の通信環境に応じてビットレートを抑えながら、得られる認識精度を予測し、認識精度が良くなるように各端末100のカメラ101にビットレートを割り当てる。端末制御機能402は、割り当てられたビットレートの映像を配信するように端末100を制御する。端末100は、割り当て得られたビットレートとなるように映像をエンコードし、エンコードした映像を配信する。

[0026] センターサーバ200は、システムのセンター側に設置されたサーバである。センターサーバ200は、1つまたは複数の物理的なサーバでもよいし、クラウド上に構築されたクラウドサーバやその他の仮想化サーバでもよい。センターサーバ200は、現場のカメラ映像から人物の作業を認識することで、現場の作業を監視する監視装置である。センターサーバ200は、端末100から送信された映像内の人物の行動等を認識する映像認識装置でもある。

[0027] センターサーバ200は、映像認識機能201、アラート生成機能202

、G U I 描画機能 2 0 3、画面表示機能 2 0 4 を有する。映像認識機能 2 0 1 は、端末 1 0 0 から送信された映像を A I エンジン（例えば学習済みの認識モデル）に入力することにより、作業員が行う作業、すなわち人物の行動の種類を認識する。アラート生成機能 2 0 2 は、認識された作業に応じてアラートを生成する。G U I 描画機能 2 0 3 は、表示装置の画面に G U I（Graphical User Interface）を表示する。画面表示機能 2 0 4 は、G U I に端末 1 0 0 の映像や認識結果、アラート等を表示する。

[0028]（第 1 の実施形態）

まず、図 4 を用いて、第 1 の実施形態に係る映像処理システム 1 の構成について説明する。図 4 に示すように、映像処理システム 1 は、複数の端末 1 0 0、センターサーバ 2 0 0、基地局 3 0 0、M E C 4 0 0 を備える。なお、各装置の構成は一例であり、後述の本実施形態に係る動作が可能であれば、その他の構成でもよい。例えば、端末 1 0 0 の一部の機能をセンターサーバ 2 0 0 や他の装置に配置してもよいし、センターサーバ 2 0 0 の一部の機能を端末 1 0 0 や他の装置に配置してもよい。

[0029] また、映像処理システム 1 は、実施形態の概要に係る映像処理システム 1 0 を具体化したものである。センターサーバ 2 0 0 は、実施形態の概要に係る映像処理装置 2 0 を具体化したものである。

[0030] 続いて、図 5 を用いて、第 1 の実施形態に係る映像処理システム 1 の端末 1 0 0 の構成について説明する。図 5 は、第 1 の実施形態に係る映像処理システム 1 の端末 1 0 0 の構成を示すブロック図である。図 5 に示すように、端末 1 0 0 は、映像取得部 1 1 0、フレームフィルタ部 1 2 0、エンコード部 1 3 0、端末通信部 1 4 0 を備えている。

[0031] 映像取得部 1 1 0 は、カメラ 1 0 1 が撮影した映像（入力映像とも称する）を取得する。入力映像は、例えば、現場で作業を行う作業員である人物や、人物が使用する作業物体等が撮影されたデータである。入力映像は、時系列のフレームを含む。

[0032] フレームフィルタ部 1 2 0 は、入力映像に含まれる時系列のフレームをフ

ィルタリング（選別）する。フレームフィルタ部１２０は、例えばセンターサーバ２００に送信する映像のビットレートを調整するためにフィルタリングを行う。ここで、入力映像に含まれるフレームの中でフィルタリングされなかったフレームはスキップされる。

[0033] エンコード部１３０は、フィルタリングされた入力映像をエンコードする。エンコード部１３０は、フレームのフィルタリングによって入力映像のフレームレートを変更する。

なお、エンコード部１３０は、フレームの注視領域が他の領域よりも高画質となるように、入力映像をエンコードしてもよい。具体的には、エンコード部１３０は、学習済みのニューラルネットワークのモデル（例えば畳み込みニューラルネットワークなどのモデル）を用いて、入力映像内の物体を検出し、検出された物体をボックスで囲う。なお、エンコード部１３０は、検出された物体をボックスに限られず、円形、楕円形、シルエットに合わせた不定形などで囲ってもよい。そして、エンコード部１３０は、ボックス内の物体を認識する。エンコード部１３０は、認識物体の中から、クラスが人物または作業物体である物体を抽出し、抽出された物体のボックス内を注視領域に決定する。エンコード部１３０は、注視領域が他の領域よりも高画質となるように、入力映像をエンコードする。

[0034] 端末通信部１４０は、エンコードされたエンコードデータをセンターサーバ２００に送信する。

[0035] 続いて、図６を用いて、第１の実施形態に係る映像処理システム１のセンターサーバ２００の構成について説明する。図６は、第１の実施形態に係る映像処理システム１のセンターサーバ２００の構成の一例を示すブロック図である。図６に示すように、センターサーバ２００は、センター通信部２１０、デコード部２２０、時間差情報取得部２３０、認識部２４０、記憶部２５０および学習部２６０を備える。

[0036] センター通信部２１０は、端末１００から送信されたエンコードデータを、基地局３００を介して受信する。センター通信部２１０は、インターネット

トやコアネットワークと通信可能なインタフェースであり、例えば、IP通信の有線インタフェースであるが、その他の任意の通信方式の有線または無線インタフェースでもよい。

[0037] デコード部220は、端末100から受信したエンコードデータをデコード（復号化）する。デコード部220は、端末100の符号化方式に対応し、例えば、H.264やH.265などの動画符号化方式によりデコードする。デコード部220は、フレームにおける各領域の圧縮率に応じてデコードし、デコードした入力映像を生成する。

[0038] 時間差情報取得部230は、映像圧縮コーデック等から取得されたタイムスタンプ情報に基づいて、時間差情報 ΔT （ ΔT は自然数）を取得する。時間差情報 ΔT は、入力映像に含まれる所定のフレームに対応し、当該所定のフレームにおける前のフレームとの時間差を表す情報である。つまり、時間差情報 ΔT は、当該所定のフレームと前のフレームとの間でフレームがスキップされていない場合、1となる。一方、時間差情報 ΔT は、当該所定のフレームと前のフレームとの間でフレームが n 枚スキップされている場合、 $1+n$ となる。また、タイムスタンプ情報は、入力映像に含まれる各フレームがカメラ101によって撮影されたタイミングを示す情報である。なお、タイムスタンプ情報は、各フレームが端末100のエンコード部130によってエンコードされたタイミングを示す情報であってもよい。

[0039] 認識部240は、入力映像に含まれる時系列のフレームと入力映像のフレーム間の時間差情報 ΔT とを入力情報として学習済みの認識モデルM1に入力し、入力映像内の物体を認識する。認識部240は、例えば、入力映像内の作業員が行う作業、すなわち人物の行動の種類を認識する。具体的には、学習済みの認識モデルM1は、入力映像に含まれる時系列のフレームを入力する再帰型ニューラルネットワーク（RNN: Recurrent Neural Network）のモデルであり、RNNの複数のセルを含む。RNNの複数のセルは、入力映像のフレーム間の時間差情報に対応するパラメータを入力する。より詳細には、RNNの複数のセルは、入力映像のフレーム間の時間差情報がデコードに

よってデコードされたパラメータを入力する。

[0040] 記憶部 250 は、学習済みの認識モデル M1 を記憶する。

学習部 260 は、学習映像と学習映像のフレーム間の時間差情報 ΔT と正解データとを用いた学習によって、学習済みの認識モデル M1 を生成する。

[0041] 続いて、図 7－図 9 を用いて、第 1 の実施形態に係る映像処理システム 1 の認識動作について説明する。

[0042] 図 7 は、第 1 の実施形態に係る映像処理システム 1 の動作を示すフローチャートである。図 7 に示すように、まず、映像処理システム 1 の端末 100 の映像取得部 110 は、カメラ 101 から現場を撮影した入力映像を取得する（ステップ S101）。入力映像は、時系列のフレームを含む。

[0043] フレームフィルタ部 120 は、入力映像に含まれる時系列のフレームをフィルタリングする（ステップ S102）。ここで、入力映像に含まれるフレームの中でフィルタリングされなかったフレームはスキップされる。

[0044] エンコード部 130 は、フィルタリングされた入力映像をエンコードする（ステップ S103）。次に、端末通信部 140 は、エンコードされたエンコードデータを、基地局 300 を介してセンターサーバ 200 に送信する（ステップ S104）。

[0045] 次に、センターサーバ 200 のセンター通信部 210 は、エンコードデータを端末 100 から受信する（ステップ S105）。次に、デコード部 220 は、エンコードデータをデコードすることによって、入力映像を取得する（ステップ S106）。

[0046] 次に、時間差情報取得部 230 は、入力映像のフレームに対応するフレーム間の時間差情報 ΔT を取得する（ステップ S107）。具体的には、時間差情報取得部 230 は、映像圧縮コーデック等から取得されたタイムスタンプ情報に基づいて、時間差情報 ΔT を取得する。タイムスタンプ情報は、例えばカメラ 101 において入力映像に含まれる各フレームが撮影されたタイミングの情報である。

[0047] 次に、認識部 240 は、入力映像に含まれる時系列のフレームと入力映像

のフレームに対応する時間差情報 ΔT とを入力情報として学習済みの認識モデルM1に入力する（ステップS108）。

[0048] 図8は、学習済みの認識モデルM1に入力される入力情報の一例を示す図である。図8に示すように、入力情報は、入力映像に含まれる時系列のフレームとフレームに対応する時間差情報 ΔT とを含む。時間差情報 ΔT は、対応する所定のフレームにおける前のフレームとの時間差を表す。例えば、時間差情報 ΔT は、対応する所定のフレームと前のフレームとの間でフレームがスキップされていない場合には、1となる。また、時間差情報 ΔT は、対応する所定のフレームと前のフレームとの間でフレームが n 枚スキップされている場合、 $1 + n$ となる。

[0049] 図7の説明に戻る。次に、認識部240は、学習済みの認識モデルM1によって入力映像内の物体を認識する（ステップS109）。認識部240は、例えば、入力映像内の作業員が行う作業、すなわち人物の行動の種類を認識する。

[0050] 図9は、学習済みの認識モデルM1の構成および認識動作の一例を示す図である。図9に示すように、学習済みの認識モデルM1は、再帰型ニューラルネットワーク（RNN）のモデルであり、RNNの時系列の複数のセルM11を含む。セルM11は、RNNの構造を入力層、中間層および出力層に分類した場合、RNNの中間層に該当する。さらに、学習済みの認識モデルM1は、各セルM11に対応するデコーダM12を備える。

[0051] 所定時刻では、デコーダM12は、時間差情報 ΔT の入力を受け付け、入力された時間差情報 ΔT をデコードしたパラメータをセルM11に出力する。次に、セルM11は、フレームと前の時刻でセルM11によって出力された状態ベクトルとデコーダM12によって出力されたパラメータ情報との入力を受け付け、後の時刻のセルM11に状態ベクトルを出力する。なお、セルM11に入力される状態ベクトルの初期状態は、例えばすべての要素が0であってもよい。

[0052] 例えば、時刻 t では、デコーダM12は、時間差情報 ΔT の1の入力を受

け付け、時間差情報 ΔT の1をデコードしたパラメータをセルM11に出力する。ここで、時刻 $t-1$ でセルM11に入力されるフレームと時刻 t でセルM11に入力されるフレームとの間でフレームスキップが起こらないため、時刻 t でデコーダM12に入力される時間差情報 ΔT は1となる。そして、時刻 t では、セルM11は、フレームと時刻 $t-1$ でセルM11によって出力された状態ベクトルとパラメータとの入力を受け付け、状態ベクトルを時刻 $t+1$ のセルM11に出力する。

[0053] 一方、時刻 $t+1$ では、デコーダM12は、時間差情報 ΔT の2の入力を受け付け、時間差情報 ΔT の2をデコードしたパラメータをセルM11に出力する。ここで、時刻 t でセルM11に入力されるフレームと時刻 $t+1$ でセルM11に入力されるフレームとの間で1枚のフレームスキップが発生するため、時刻 $t+1$ でデコーダM12に入力される時間差情報 ΔT は2となる。そして、時刻 $t+1$ では、セルM11は、フレームと時刻 t でセルM11によって出力された状態ベクトルとパラメータとの入力を受け付け、時刻 $t+2$ のセルM11に状態ベクトルを出力する。

[0054] 続いて、第1の実施形態に係る映像処理システム1における認識モデルM1の学習動作について説明する。

[0055] 学習部260は、学習映像に含まれる時系列のフレームとフレームに対応する時間差情報 ΔT とを認識モデルM1に入力する。学習映像は、例えば予め決められたパターンによってフレームスキップが起こされている時系列のフレームを含む。認識モデルM1の構成は上述した（図9参照）。学習部260は、認識モデルM1による出力結果を正解データと比較することによって認識モデルM1を学習し、学習済みの認識モデルM1を生成する。

[0056] 上述したように、映像処理システム1の学習済みの認識モデルM1は、時間差情報 ΔT をデコードしてセルM11に入力するパラメータを動的に決定する。つまり、学習済みの認識モデルM1は、映像のフレームレート変更などでフレームがスキップされた場合などを考慮して、フレームの時間差情報を物体の認識に反映することで、物体の認識精度を向上することができる。

[0057] (第2の実施形態)

続いて、第2の実施形態に係る映像処理システム2の構成について説明する。映像処理システム2は、第1の実施形態に係る映像処理システム1と同様に、複数の端末100、センターサーバ200、基地局300、MEC400を備える。ここで、映像処理システム2のセンターサーバ200は、映像処理システム1のセンターサーバ200と比較して次の構成が異なる。

[0058] 図10は、映像処理システム2のセンターサーバ200の構成を示す図である。図10に示すように、映像処理システム2のセンターサーバ200は、センター通信部210、デコード部220、時間差情報取得部230、認識部270、記憶部280および学習部290を備える。

[0059] 認識部270は、入力映像に含まれる時系列のフレームと入力映像のフレーム間の時間差情報 ΔT とを入力情報として学習済みの認識モデルM2に入力し、入力映像内の物体を認識する。学習済みの認識モデルM2は、入力映像に含まれる時系列のフレームを入力すると共に状態ベクトルを時系列に入出力する再帰型ニューラルネットワーク(RNN)の複数のセルを含む。学習済みの認識モデルM2は、フレームスキップが発生したセルの間など所定のセルの間に、フレーム間の時間差情報 ΔT に基づいて状態ベクトルを予測する状態予測器を挿入する。

記憶部280は、認識モデルM2を記憶する。

[0060] 学習部290は、学習映像に含まれる予め定められたパターンでフレームスキップを起こした時系列のフレームと学習映像のフレーム間の時間差情報と正解データとを用いて、状態予測器を挿入した学習済みの認識モデルM2を学習する。

[0061] また、学習部290は、学習映像に含まれるフレームスキップを起こしていない時系列のフレームと正解データとを用いて学習済みの認識モデルM2の複数のセルを学習する。そして、学習部290は、学習映像に含まれるフレームスキップを起こしていない時系列のフレームを学習された認識モデルM2の複数のセルに入力する。学習部290は、その場合に複数のセルによ

って時刻 t (t は自然数) に出力される状態ベクトルと時刻 $t + N$ (N は自然数) に出力される状態ベクトルとを用いて状態予測器を学習する。

[0062] 他方、認識部 270 は、学習映像と学習映像のフレーム間の時間差情報と入力映像のフレーム間の動きとを用いて学習した学習済みの認識モデルに、入力映像と入力映像のフレーム間の時間差情報と入力映像のフレーム間の動きとを入力し、入力映像内の物体を認識する。

[0063] 学習部 290 は、学習映像に含まれる予め定められたパターンでフレームスキップを起こした時系列のフレームと学習映像のフレーム間の時間差情報と、入力映像のフレーム間の動きと正解データとを用いて、状態予測器を挿入した学習済みの認識モデル $M2$ を学習する。

[0064] 続いて、図 11 を用いて、第 2 の実施形態に係る映像処理システム 2 の認識動作について説明する。

[0065] 図 11 は、第 2 の実施形態に係る映像処理システム 2 の動作の一例を示すフローチャートである。図 11 に示すように、まず、映像処理システム 2 は、上述したステップ $S101$ の処理～ステップ $S107$ の処理（図 7 を参照）を実行する。ステップ $S101$ の処理～ステップ $S107$ の処理に関しては説明を省略する。

[0066] 次に、センターサーバ 200 の認識部 270 は、入力映像に含まれる時系列のフレームとフレームに対応する時間差情報 ΔT とを入力情報として学習済みの認識モデル $M2$ に入力する（ステップ $S201$ ）。入力情報の一例は上述した（図 8 参照）。ただし、本実施形態では、認識部 240 は、 $\Delta T \neq 1$ の時間差情報 ΔT を学習済みの認識モデル $M2$ への入力情報とする。

[0067] 次に、認識部 270 は、学習済みの認識モデル $M2$ によって、入力映像内の物体を認識する（ステップ $S202$ ）。認識部 240 は、例えば、入力映像内の作業員が行う作業、すなわち人物の行動の種類を認識する。

[0068] 図 12 は、第 2 の実施形態に係る学習済みの認識モデル $M2$ の構成および認識動作の一例を示す図である。

図 12 に示すように、学習済みの認識モデル $M2$ は、再帰型ニューラルネ

ットワーク（RNN）であり、時系列のRNNの複数のセルM21を含む。セルM21は、所定時刻では、フレームと前の時刻でセルM21によって出力された状態ベクトルとの入力を受け付け、後の時刻のセルM21に状態ベクトルを出力する。なお、状態ベクトルの初期状態は、例えばすべての要素が0であってもよい。

[0069] さらに、学習済みの認識モデルM2は、所定時刻のセルM21に入力されるフレームと当該所定時刻のセルM21の前の時刻のセルM21に入力されるフレームとの間でフレームスキップが発生する場合、所定時刻のセルM21と前の時刻のセルM21との間に状態予測器M22を挿入する。フレームスキップの発生は、所定時刻のセルM21に入力されるフレームに対応する時間差情報 ΔT から判断できる。挿入された状態予測器M22は、前の時刻のセルM21によって出力された状態ベクトルと所定時刻のセルM21に入力されるフレームに対応する時間差情報 ΔT との入力を受け付ける。そして、状態予測器M22は、状態ベクトルを予測し、予測された状態ベクトルを所定時刻のセルM21に出力する。

[0070] 例えば、時刻 $t+1$ でセルM21に入力されるフレームと時刻 t でセルM21に入力されるフレームの間では、フレームスキップが発生する。その場合、学習済みの認識モデルM2は、時刻 $t+1$ のセルM21と時刻 t のセルM21との間に状態予測器M22を挿入する。状態予測器M22は、時刻 t のセルM21によって出力された状態ベクトルと時刻 $t+1$ でセルM21に入力されるフレームに対応する時間差情報 ΔT の2との入力を受け付ける。ここで、入力される時間差情報 ΔT は、時刻 $t+1$ でセルM21に入力されるフレームと時刻 t でセルM21に入力されるフレームの間では1枚のフレームスキップが発生しているため、2である。状態予測器M22は、状態ベクトルを予測し、予測された状態ベクトルを時刻 $t+1$ のセルM21に出力する。

[0071] 続いて、図13～図14を用いて、第2の実施形態に係る映像処理システム2の認識モデルM2の学習動作について説明する。

[0072] 図13は、認識モデルM2の第1の学習動作の一例を示す図である。

図13に示すように、学習部290は、学習映像に含まれる時系列のフレームとフレームに対応する時間差情報 ΔT ($\Delta T \neq 1$) とを認識モデルM2に入力する。具体的には、学習部290は、学習映像に含まれる時系列のフレームを認識モデルM2の複数のセルM21に入力する。入力される時系列のフレームは、予め決められたパターンによってフレームスキップが起こされている。また、学習部290は、認識モデルM2の所定時刻のセルM21に入力されるフレームと当該所定時刻の前の時刻のセルM21に入力されるフレームとの間でフレームスキップが発生する場合、所定時刻のセルM21と前時刻のセルM21との間に状態予測器M22を挿入する。学習部290は、所定時刻のセルM21に入力されるフレームに対応する時間差情報 ΔT を状態予測器M22に入力する。例えば、時刻 $t+1$ で所定のセルM21に入力されるフレームと時刻 t でセルM21に入力されるフレームとの間で1枚のフレームスキップが発生する。その場合、学習部290は、時刻 $t+1$ のセルM21と時刻 t のセルM21との間に状態予測器M22を挿入し、時刻 $t+1$ で所定のセルM21に入力されるフレームに対応する時間差情報 ΔT の2を入力する。

[0073] そして、学習部290は、状態予測器M22を挿入した認識モデルM2による出力結果を正解データと比較することによって状態予測器M22を挿入した認識モデルM2を学習する。なお、学習部290は、状態予測器M22の学習を認識モデルM2の学習とは別にしてもよい。

[0074] 図14は、認識モデルM2の第2の学習動作の一例を示す図である。

図14に示すように、学習部290は、学習映像に含まれる時系列のフレームを認識モデルM2の複数のセルM21に入力する。入力される時系列のフレームは、フレームスキップが起こされていない。そして、学習部290は、認識モデルM2による出力結果を正解データと比較することによって認識モデルM2を学習する。

[0075] 次に、学習部290は、学習映像に含まれる時系列のフレームを学習した

認識モデルM2の複数のセルM21に入力する。入力される時系列のフレームは、フレームスキップが起こされていない。

[0076] 次に、学習部290は、時刻 t でセルM21から出力された状態ベクトルと時刻 $t+N$ （ N は自然数）でセルM21から出力された状態ベクトルとからなるデータセットを取得し、取得されたデータセットを学習データとして状態予測器M22を学習させる。詳細には、学習部290は、時刻 t の状態ベクトルと N とを状態予測器M22に入力した際の出力結果を、時刻 $t+N$ の状態ベクトルに近づけるように回帰分析するなどして、状態予測器M22を学習させる。

[0077] 図15は、第2の実施形態に係る学習済みの認識モデルM2の認識動作の他の一例を示す図である。

[0078] 図15に示すように、学習済みの認識モデルM2は、再帰型ニューラルネットワーク（RNN）であり、時系列のRNNの複数のセルM21を含む。セルM21は、所定時刻では、フレームと前の時刻でセルM21によって出力された状態ベクトルとの入力を受け付け、後の時刻のセルM21に状態ベクトルを出力する。なお、状態ベクトルの初期状態は、例えばすべての要素が0であってもよい。

[0079] さらに、学習済みの認識モデルM2は、所定時刻のセルM21に入力されるフレームと当該所定時刻のセルM21の前の時刻のセルM21に入力されるフレームとの間でフレームスキップが発生する場合、所定のセルM21と前の時刻のセルM21との間に状態予測器M23を挿入する。状態予測器M23は、前の時刻のセルM21によって出力された状態ベクトルと、所定時刻のセルM21に入力されるフレームに対応する時間差情報 ΔT と、動きベクトルとの入力を受け付ける。動きベクトルは、所定時刻のフレームにおける前の時刻のフレームとの間の差異、つまり動きをベクトル化した情報である。状態予測器M23は、状態ベクトルを予測し、予測された状態ベクトルを所定時刻のセルM21に出力する。

[0080] 例えば、時刻 $t+1$ でセルM21に入力されるフレームと時刻 t でセルM

21に入力されるフレームとの間では、フレームスキップが発生する。その場合、学習済みの認識モデルM2は、時刻 $t+1$ のセルM21と時刻 t のセルM21との間に状態予測器M23を挿入する。状態予測器M23は、時刻 t のセルM21によって出力された状態ベクトルと時刻 $t+1$ でセルM21に入力されるフレームに対応する時間差情報 ΔT の2と共に、動きベクトルの入力を受け付ける。入力される動きベクトルは、時刻 t でセルM21に入力されるフレームと時刻 $t+1$ でセルM21に入力されるフレームとの差異、すなわち動きを表す。状態予測器M23は、状態ベクトルを予測し、予測された状態ベクトルを時刻 $t+1$ のセルM21に出力する。

[0081] 続いて、映像処理システム2における認識モデルM2の学習動作を説明する。

学習部290は、学習映像に含まれる時系列のフレームとフレームに対応する時間差情報 ΔT ($\Delta T \neq 1$) と動きベクトルとを認識モデルM2に入力する。学習部290は、状態予測器M22を挿入した認識モデルM2による出力結果を正解データと比較することによって状態予測器M22を挿入した認識モデルM2を学習する。なお、学習部290は、状態予測器M22の学習を認識モデルM2の学習とは別にしてもよい。

[0082] 上述したように、第2の実施形態に係る映像処理システム2の学習済みの認識モデルM2は、フレームスキップが発生した場合、セルM21の間に状態予測器M22又は状態予測器M23を挿入し、状態ベクトルを予測する。学習済みの認識モデルM1は、映像のフレームレート変更などでフレームがスキップされた場合などを考慮して、フレームの時間差情報を物体の認識に反映することで、物体の認識精度を向上することができる。

[0083] 上述の実施形態における各構成は、ハードウェア又はソフトウェア、もしくはその両方によって構成され、1つのハードウェア又はソフトウェアから構成してもよいし、複数のハードウェア又はソフトウェアから構成してもよい。各装置及び各機能（処理）を、図19に示すような、CPU（Central Processing Unit）等のプロセッサ1001及び記憶装置であるメモリ100

2を有するコンピュータ1000により実現してもよい。例えば、メモリ1002に実施形態における方法（映像処理方法）を行うためのプログラムを格納し、各機能を、メモリ1002に格納されたプログラムをプロセッサ1001で実行することにより実現してもよい。

[0084] これらのプログラムは、コンピュータに読み込まれた場合に、実施形態で説明された1又はそれ以上の機能をコンピュータに行わせるための命令群（又はソフトウェアコード）を含む。プログラムは、非一時的なコンピュータ可読媒体又は実体のある記憶媒体に格納されてもよい。限定ではなく例として、コンピュータ可読媒体又は実体のある記憶媒体は、random-access memory (RAM)、read-only memory (ROM)、フラッシュメモリ、solid-state drive (SSD) 又はその他のメモリ技術、CD-ROM、digital versatile disc (DVD)、Blu-ray（登録商標）ディスク又はその他の光ディスクストレージ、磁気カセット、磁気テープ、磁気ディスクストレージ又はその他の磁気ストレージデバイスを含む。プログラムは、一時的なコンピュータ可読媒体又は通信媒体上で送信されてもよい。限定ではなく例として、一時的なコンピュータ可読媒体又は通信媒体は、電氣的、光学的、音響的、またはその他の形式の伝搬信号を含む。

[0085] 上記の実施形態の一部又は全部は、以下の付記のようにも記載されうるが、以下には限られない。

（付記1）

入力映像を取得する映像取得手段と、

前記入力映像のフレーム間の第1の時間差情報を取得する時間差情報取得手段と、

学習映像と前記学習映像のフレーム間の第2の時間差情報とを用いて学習した学習済みの認識モデルに、前記入力映像と前記入力映像のフレーム間の第1の時間差情報とを入力し、前記入力映像内の物体を認識する認識手段と、を備える

映像処理システム。

(付記 2)

前記学習済みの認識モデルは、

前記入力映像に含まれる時系列のフレームを入力する再帰型ニューラルネットワーク (RNN: Recurrent Neural Network) の複数のセルを含むモデルであり、

前記複数のセルは、前記入力映像のフレーム間の第 1 の時間差情報に対応するパラメータを入力する

付記 1 に記載の映像処理システム。

(付記 3)

前記複数のセルは、

前記入力映像のフレーム間の第 1 の時間差情報がデコードされたパラメータを入力する

付記 2 に記載の映像処理システム。

(付記 4)

前記学習済みの認識モデルは、

前記入力映像に含まれる時系列のフレームを入力すると共に、状態ベクトルを時系列に入出力する再帰型ニューラルネットワークの複数のセルを含み、

所定のセルの間に、前記入力映像のフレーム間の第 1 の時間差情報に基づいて状態ベクトルを予測する状態予測器を挿入したモデルである

付記 1 に記載の映像処理システム。

(付記 5)

前記状態予測器を挿入した前記学習済みの認識モデルは、

前記学習映像に含まれる予め定められたパターンでフレームスキップを起こした時系列のフレームと前記学習映像のフレーム間の第 2 の時間差情報と正解データとを用いて、学習される

付記 4 に記載の映像処理システム。

(付記 6)

前記学習済みの認識モデルの複数のセルは、
前記学習映像に含まれるフレームスキップを起こしていない時系列のフレームと正解データとを用いて学習され、
前記学習済みの認識モデルに挿入される前記状態予測器は、
前記学習映像に含まれるフレームスキップを起こしていない時系列のフレームを前記学習された複数のセルに入力した場合に、前記複数のセルによって時刻 t (t は自然数) に出力される状態ベクトルと時刻 $t + N$ (N は自然数) に出力される状態ベクトルとを用いて学習される

付記 4 に記載の映像処理システム。

(付記 7)

前記認識手段は、

前記学習映像と前記学習映像のフレーム間の第 2 の時間差情報と前記学習映像のフレーム間の動きとを用いて学習した学習済みの認識モデルに、前記入力映像と前記入力映像のフレーム間の第 1 の時間差情報と前記入力映像のフレーム間の動きとを入力し、前記入力映像内の物体を認識する

付記 1 に記載の映像処理システム。

(付記 8)

入力映像を取得する映像取得手段と、

前記入力映像のフレーム間の第 1 の時間差情報を取得する時間差情報取得手段と、

学習映像と前記学習映像のフレーム間の第 2 の時間差情報とを用いて学習した学習済みの認識モデルに、前記入力映像と前記入力映像のフレーム間の第 1 の時間差情報とを入力し、前記入力映像内の物体を認識する認識手段と、
を備える

映像処理装置。

(付記 9)

前記学習済みの認識モデルは、

前記入力映像に含まれる時系列のフレームを入力する再帰型ニューラルネ

ットワーク（RNN）の複数のセルを含むモデルであり、

前記複数のセルは、前記入力映像のフレーム間の第1の時間差情報に対応するパラメータを入力する

付記8に記載の映像処理装置。

（付記10）

前記複数のセルは、

前記入力映像のフレーム間の第1の時間差情報がデコードされたパラメータを入力する

付記9に記載の映像処理装置。

（付記11）

前記学習済みの認識モデルは、

前記入力映像に含まれる時系列のフレームを入力すると共に、状態ベクトルを時系列に入出力する再帰型ニューラルネットワークの複数のセルを含み、

所定のセルの間に、前記入力映像のフレーム間の第1の時間差情報に基づいて状態ベクトルを予測する状態予測器を挿入したモデルである

付記8に記載の映像処理装置。

（付記12）

前記状態予測器を挿入した前記学習済みの認識モデルは、

前記学習映像に含まれる予め定められたパターンでフレームスキップを起こした時系列のフレームと前記学習映像のフレーム間の第2の時間差情報と正解データとを用いて、学習される

付記11に記載の映像処理装置。

（付記13）

前記学習済みの認識モデルの複数のセルは、

前記学習映像に含まれるフレームスキップを起こしていない時系列のフレームと正解データとを用いて学習され、

前記学習済みの認識モデルに挿入される前記状態予測器は、

前記学習映像に含まれるフレームスキップを起こしていない時系列のフレームを前記学習された複数のセルに入力した場合に、前記複数のセルによって時刻 t (t は自然数) に出力される状態ベクトルと時刻 $t + N$ (N は自然数) に出力される状態ベクトルとを用いて学習される

付記 11 に記載の映像処理装置。

(付記 14)

前記認識手段は、

前記学習映像と前記学習映像のフレーム間の第 2 の時間差情報と前記学習映像のフレーム間の動きとを用いて学習した学習済みの認識モデルに、前記入力映像と前記入力映像のフレーム間の第 1 の時間差情報と前記入力映像のフレーム間の動きとを入力し、前記入力映像内の物体を認識する

付記 8 に記載の映像処理装置。

(付記 15)

コンピュータが、

入力映像を取得し、

前記入力映像のフレーム間の第 1 の時間差情報を取得し、

学習映像と前記学習映像のフレーム間の第 2 の時間差情報とを用いて学習した学習済みの認識モデルに、入力映像と前記入力映像のフレーム間の第 1 の時間差情報とを入力し、前記入力映像内の物体を認識する

映像処理方法。

(付記 16)

前記学習済みの認識モデルは、

前記入力映像に含まれる時系列のフレームを入力する再帰型ニューラルネットワーク (RNN) の複数のセルを含むモデルであり、

前記複数のセルは、前記入力映像のフレーム間の第 1 の時間差情報に対応するパラメータを入力する

付記 15 に記載の映像処理方法。

(付記 17)

前記複数のセルは、

前記入力映像のフレーム間の第 1 の時間差情報がデコードされたパラメータを入力する

付記 16 に記載の映像処理方法。

(付記 18)

前記学習済みの認識モデルは、

前記入力映像に含まれる時系列のフレームを入力すると共に、状態ベクトルを時系列に入出力する再帰型ニューラルネットワークの複数のセルを含み、

所定のセルの間に、前記入力映像のフレーム間の第 1 の時間差情報に基づいて状態ベクトルを予測する状態予測器を挿入したモデルである

付記 15 に記載の映像処理方法。

(付記 19)

前記状態予測器を挿入した前記学習済みの認識モデルは、

前記学習映像に含まれる予め定められたパターンでフレームスキップを起こした時系列のフレームと前記学習映像のフレーム間の第 2 の時間差情報と正解データとを用いて、学習される

付記 18 に記載の映像処理方法。

(付記 20)

前記学習済みの認識モデルの複数のセルは、

前記学習映像に含まれるフレームスキップを起こしていない時系列のフレームと正解データとを用いて学習され、

前記学習済みの認識モデルに挿入される前記状態予測器は、

前記学習映像に含まれるフレームスキップを起こしていない時系列のフレームを前記学習された複数のセルに入力した場合に、前記複数のセルによって時刻 t (t は自然数) に出力される状態ベクトルと時刻 $t + N$ (N は自然数) に出力される状態ベクトルとを用いて学習される

付記 18 に記載の映像処理方法。

(付記 2 1)

コンピュータが、

前記学習映像と前記学習映像のフレーム間の第 2 の時間差情報と前記学習映像のフレーム間の動きとを用いて学習した学習済みの認識モデルに、前記入力映像と前記入力映像のフレーム間の第 1 の時間差情報と前記入力映像のフレーム間の動きとを入力し、前記入力映像内の物体を認識する

付記 1 5 に記載の映像処理方法。

符号の説明

[0086] 1、2、1 0 映像処理システム

1 1 映像取得部

1 2 時間差情報取得部

1 3 認識部

2 0 映像処理装置

1 0 0 端末

1 0 1 カメラ

1 0 2 圧縮効率最適化機能

1 1 0 映像取得部

1 2 0 フレームフィルタ部

1 3 0 エンコード部

1 4 0 端末通信部

2 0 0 センターサーバ

2 0 1 映像認識機能

2 0 2 アラート生成機能

2 0 3 G U I 描画機能

2 0 4 画面表示機能

2 1 0 センター通信部

2 2 0 デコード部

2 3 0 時間差情報取得部

240、270 認識部
250、280 記憶部
260、290 学習部
300 基地局
401 圧縮ビットレート制御機能
1000 コンピュータ
1001 プロセッサ
1002 メモリ
M1、M2 認識モデル
M11、M21 セル
M12 デコーダ
M22、M23 状態予測器

請求の範囲

- [請求項1] 入力映像を取得する映像取得手段と、
 前記入力映像のフレーム間の第1の時間差情報を取得する時間差情報取得手段と、
 学習映像と前記学習映像のフレーム間の第2の時間差情報とを用いて学習した学習済みの認識モデルに、前記入力映像と前記入力映像のフレーム間の第1の時間差情報とを入力し、前記入力映像内の物体を認識する認識手段と、を備える
 映像処理システム。
- [請求項2] 前記学習済みの認識モデルは、
 前記入力映像に含まれる時系列のフレームを入力する再帰型ニューラルネットワーク（RNN: Recurrent Neural Network）の複数のセルを含むモデルであり、
 前記複数のセルは、前記入力映像のフレーム間の第1の時間差情報に対応するパラメータを入力する
 請求項1に記載の映像処理システム。
- [請求項3] 前記学習済みの認識モデルは、
 前記入力映像に含まれる時系列のフレームを入力すると共に、状態ベクトルを時系列に入出力する再帰型ニューラルネットワークの複数のセルを含み、
 所定のセルの間に、前記入力映像のフレーム間の第1の時間差情報に基づいて状態ベクトルを予測する状態予測器を挿入したモデルである
 請求項1に記載の映像処理システム。
- [請求項4] 前記状態予測器を挿入した前記学習済みの認識モデルは、
 前記学習映像に含まれる予め定められたパターンでフレームスキップを起こした時系列のフレームと前記学習映像のフレーム間の第2の時間差情報と正解データとを用いて、学習される

請求項 3 に記載の映像処理システム。

[請求項5]

前記学習済みの認識モデルの複数のセルは、

前記学習映像に含まれるフレームスキップを起こしていない時系列のフレームと正解データとを用いて学習され、

前記学習済みの認識モデルに挿入される前記状態予測器は、

前記学習映像に含まれるフレームスキップを起こしていない時系列のフレームを前記学習された複数のセルに入力した場合に、前記複数のセルによって時刻 t (t は自然数) に出力される状態ベクトルと時刻 $t + N$ (N は自然数) に出力される状態ベクトルとを用いて学習される

請求項 3 に記載の映像処理システム。

[請求項6]

前記認識手段は、

前記学習映像と前記学習映像のフレーム間の第 2 の時間差情報と前記学習映像のフレーム間の動きとを用いて学習した学習済みの認識モデルに、前記入力映像と前記入力映像のフレーム間の第 1 の時間差情報と前記入力映像のフレーム間の動きとを入力し、前記入力映像内の物体を認識する

請求項 1 に記載の映像処理システム。

[請求項7]

入力映像を取得する映像取得手段と、

前記入力映像のフレーム間の第 1 の時間差情報を取得する時間差情報取得手段と、

学習映像と前記学習映像のフレーム間の第 2 の時間差情報とを用いて学習した学習済みの認識モデルに、前記入力映像と前記入力映像のフレーム間の第 1 の時間差情報とを入力し、前記入力映像内の物体を認識する認識手段と、を備える

映像処理装置。

[請求項8]

前記学習済みの認識モデルは、

前記入力映像に含まれる時系列のフレームを入力する再帰型ニュー

ラルネットワーク（RNN）の複数のセルを含むモデルであり、

前記複数のセルは、前記入力映像のフレーム間の第1の時間差情報に対応するパラメータを入力する

請求項7に記載の映像処理装置。

[請求項9]

前記学習済みの認識モデルは、

前記入力映像に含まれる時系列のフレームを入力すると共に、状態ベクトルを時系列に入出力する再帰型ニューラルネットワークの複数のセルを含み、

所定のセルの間に、前記入力映像のフレーム間の第1の時間差情報に基づいて状態ベクトルを予測する状態予測器を挿入したモデルである

請求項7に記載の映像処理装置。

[請求項10]

前記状態予測器を挿入した前記学習済みの認識モデルは、

前記学習映像に含まれる予め定められたパターンでフレームスキップを起こした時系列のフレームと前記学習映像のフレーム間の第2の時間差情報と正解データとを用いて、学習される

請求項9に記載の映像処理装置。

[請求項11]

前記学習済みの認識モデルの複数のセルは、

前記学習映像に含まれるフレームスキップを起こしていない時系列のフレームと正解データとを用いて学習され、

前記学習済みの認識モデルに挿入される前記状態予測器は、

前記学習映像に含まれるフレームスキップを起こしていない時系列のフレームを前記学習された複数のセルに入力した場合に、前記複数のセルによって時刻 t （ t は自然数）に出力される状態ベクトルと時刻 $t + N$ （ N は自然数）に出力される状態ベクトルとを用いて学習される

請求項9に記載の映像処理装置。

[請求項12]

前記認識手段は、

前記学習映像と前記学習映像のフレーム間の第2の時間差情報と前記学習映像のフレーム間の動きとを用いて学習した学習済みの認識モデルに、前記入力映像と前記入力映像のフレーム間の第1の時間差情報と前記入力映像のフレーム間の動きとを入力し、前記入力映像内の物体を認識する

請求項7に記載の映像処理装置。

[請求項13]

コンピュータが、

入力映像を取得し、

前記入力映像のフレーム間の第1の時間差情報を取得し、

学習映像と前記学習映像のフレーム間の第2の時間差情報とを用いて学習した学習済みの認識モデルに、前記入力映像と前記入力映像のフレーム間の第1の時間差情報とを入力し、前記入力映像内の物体を認識する

映像処理方法。

[請求項14]

前記学習済みの認識モデルは、

前記入力映像に含まれる時系列のフレームを入力する再帰型ニューラルネットワーク（RNN）の複数のセルを含むモデルであり、

前記複数のセルは、前記入力映像のフレーム間の第1の時間差情報に対応するパラメータを入力する

請求項13に記載の映像処理方法。

[請求項15]

前記学習済みの認識モデルは、

前記入力映像に含まれる時系列のフレームを入力すると共に、状態ベクトルを時系列に入出力する再帰型ニューラルネットワークの複数のセルを含み、

所定のセルの間に、前記入力映像のフレーム間の第1の時間差情報に基づいて状態ベクトルを予測する状態予測器を挿入したモデルである

請求項13に記載の映像処理方法。

- [請求項16] 前記状態予測器を挿入した前記学習済みの認識モデルは、
前記学習映像に含まれる予め定められたパターンでフレームスキップを起こした時系列のフレームと前記学習映像のフレーム間の第2の時間差情報と正解データとを用いて、学習される
請求項15に記載の映像処理方法。
- [請求項17] 前記学習済みの認識モデルの複数のセルは、
前記学習映像に含まれるフレームスキップを起こしていない時系列のフレームと正解データとを用いて学習され、
前記学習済みの認識モデルに挿入される前記状態予測器は、
前記学習映像に含まれるフレームスキップを起こしていない時系列のフレームを前記学習された複数のセルに入力した場合に、前記複数のセルによって時刻 t (t は自然数) に出力される状態ベクトルと時刻 $t + N$ (N は自然数) に出力される状態ベクトルとを用いて学習される
請求項15に記載の映像処理方法。
- [請求項18] コンピュータが、
前記学習映像と前記学習映像のフレーム間の第2の時間差情報と前記学習映像のフレーム間の動きとを用いて学習した学習済みの認識モデルに、前記入力映像と前記入力映像のフレーム間の第1の時間差情報と前記入力映像のフレーム間の動きとを入力し、前記入力映像内の物体を認識する
請求項13に記載の映像処理方法。

[図1]

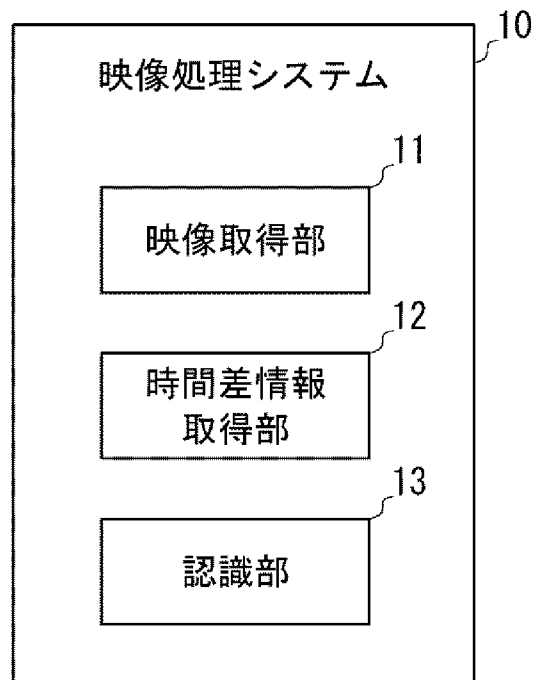


Fig. 1

[図2]

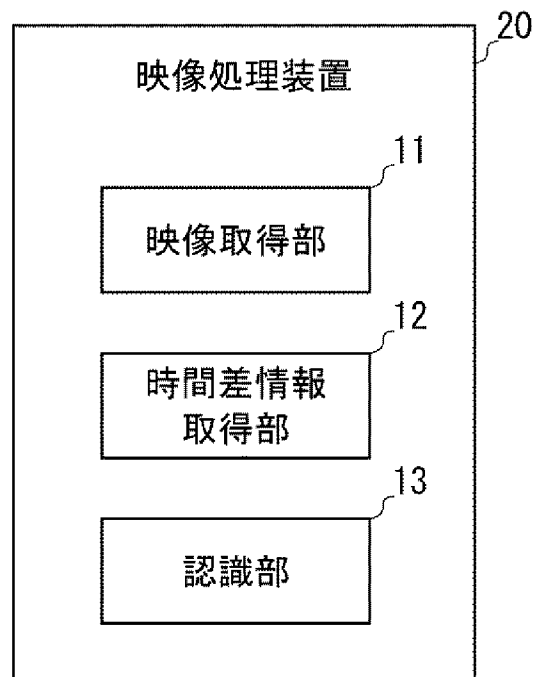


Fig. 2

[図3]

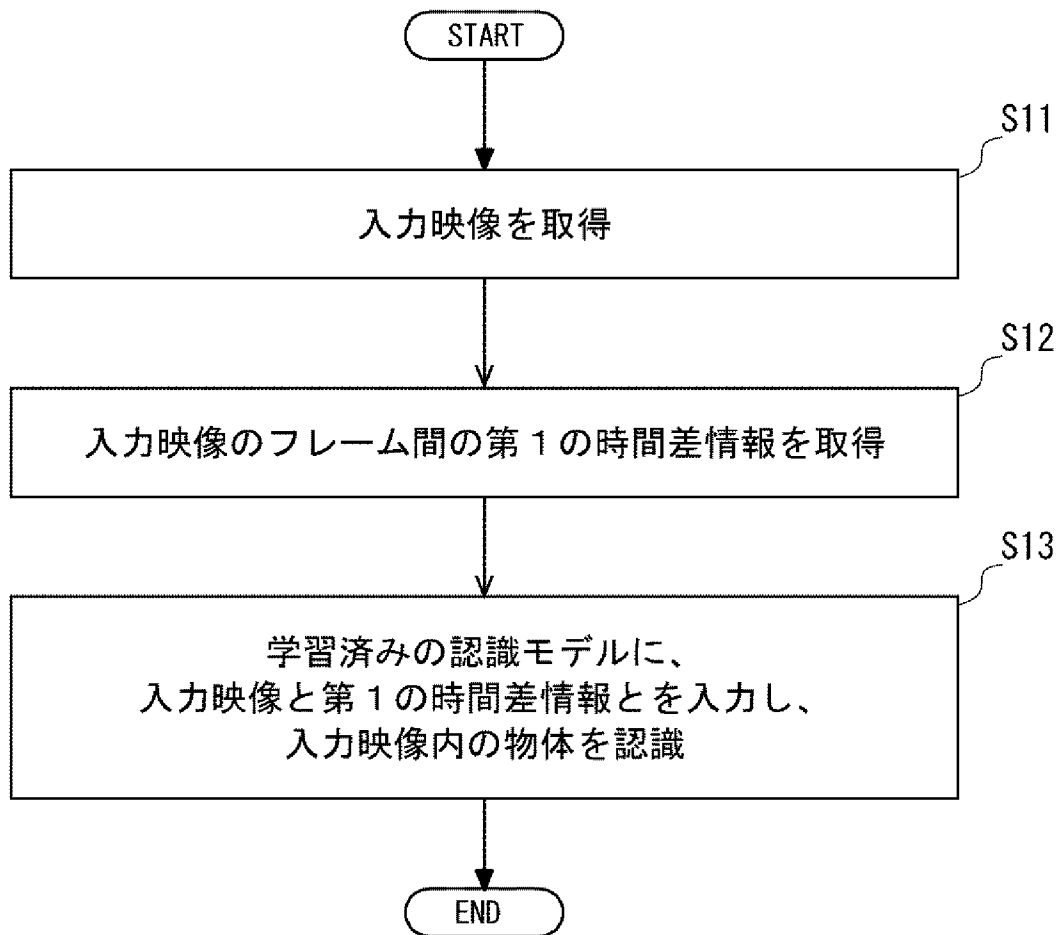


Fig. 3

[図4]

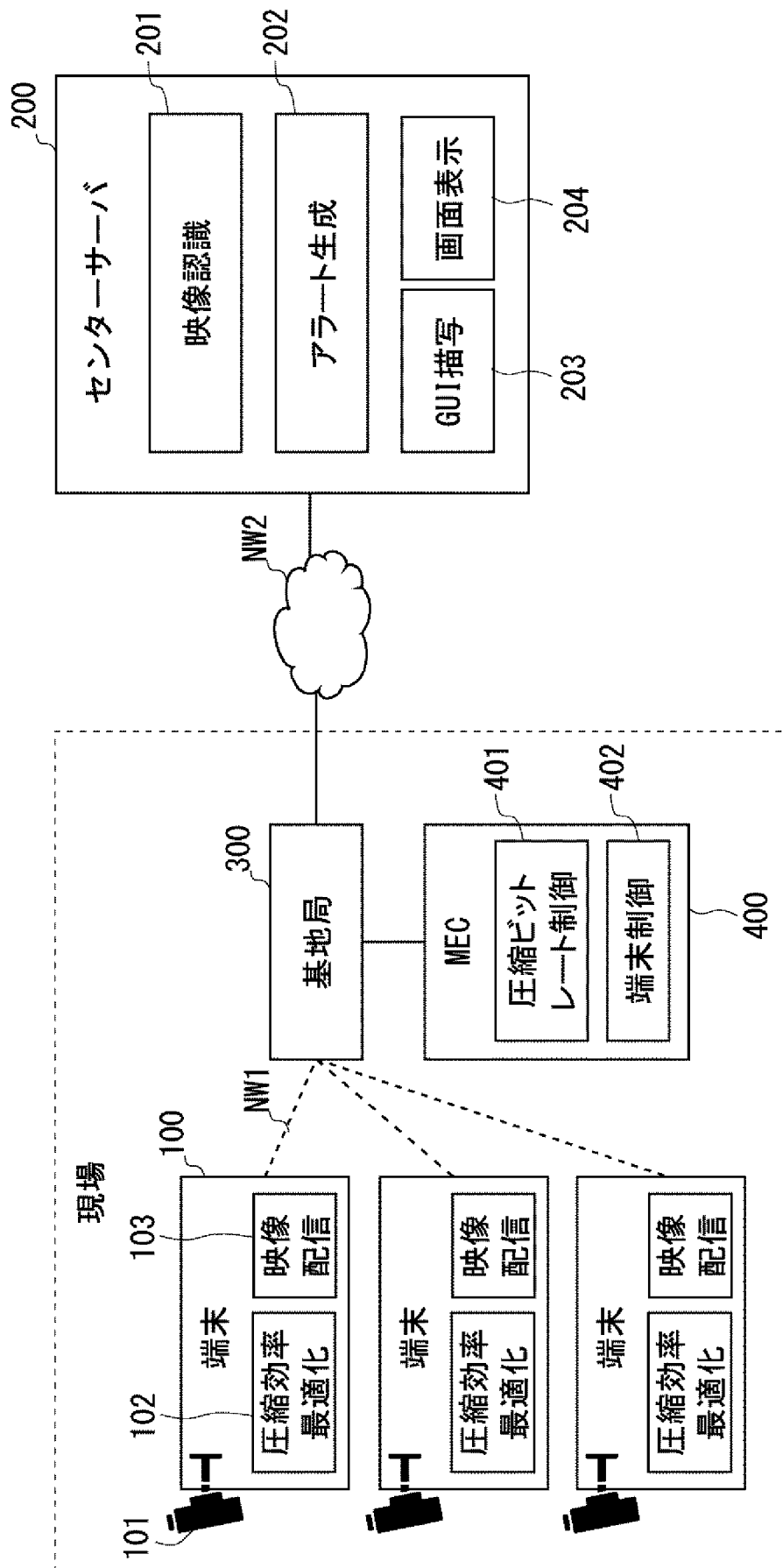


Fig. 4

[図5]

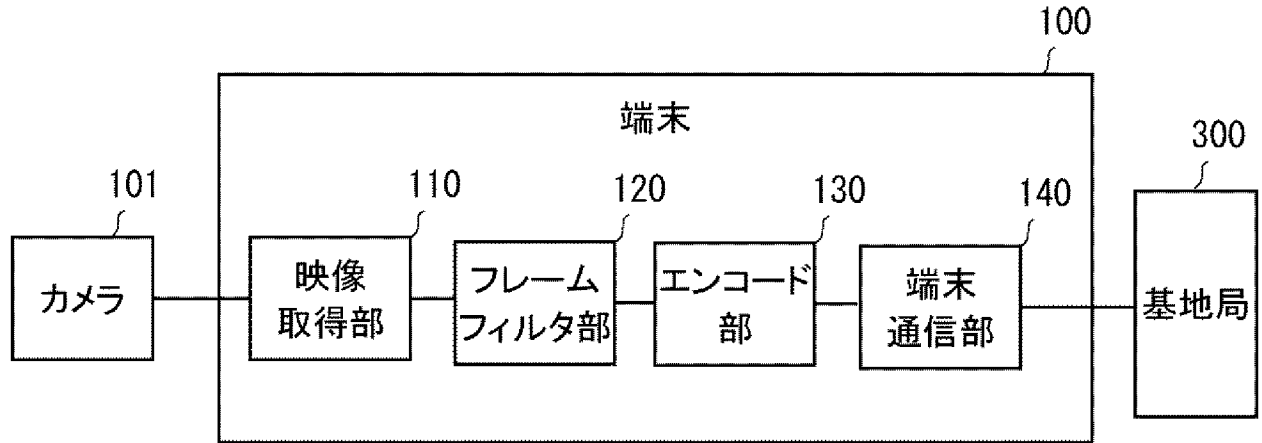


Fig. 5

[図6]

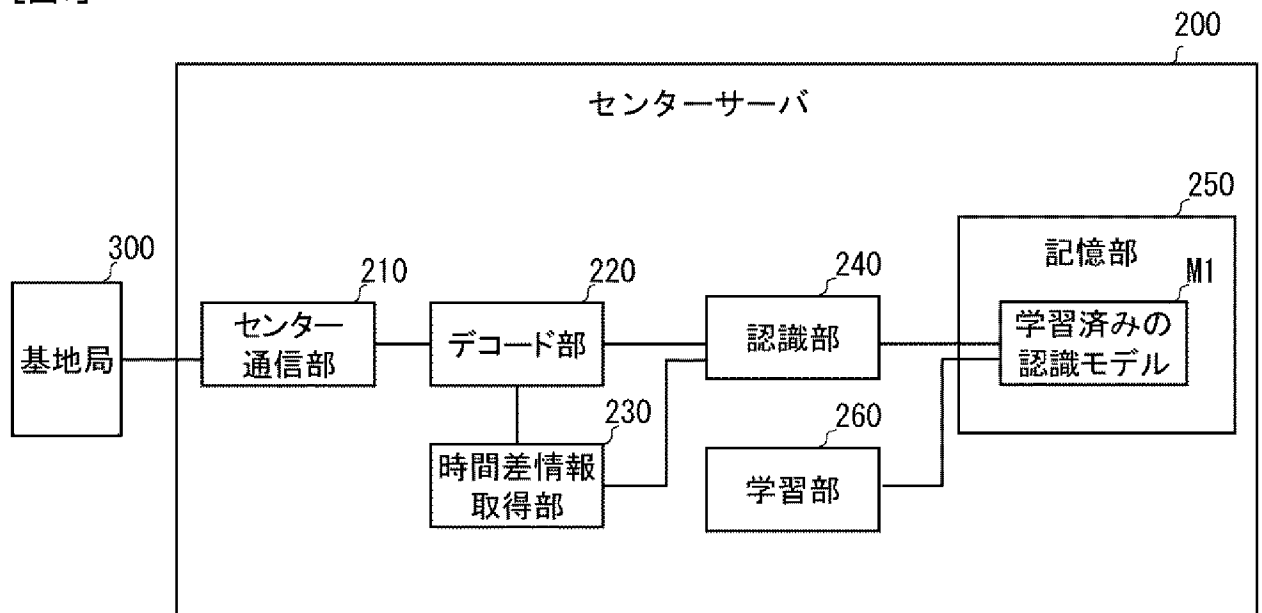


Fig. 6

[図7]

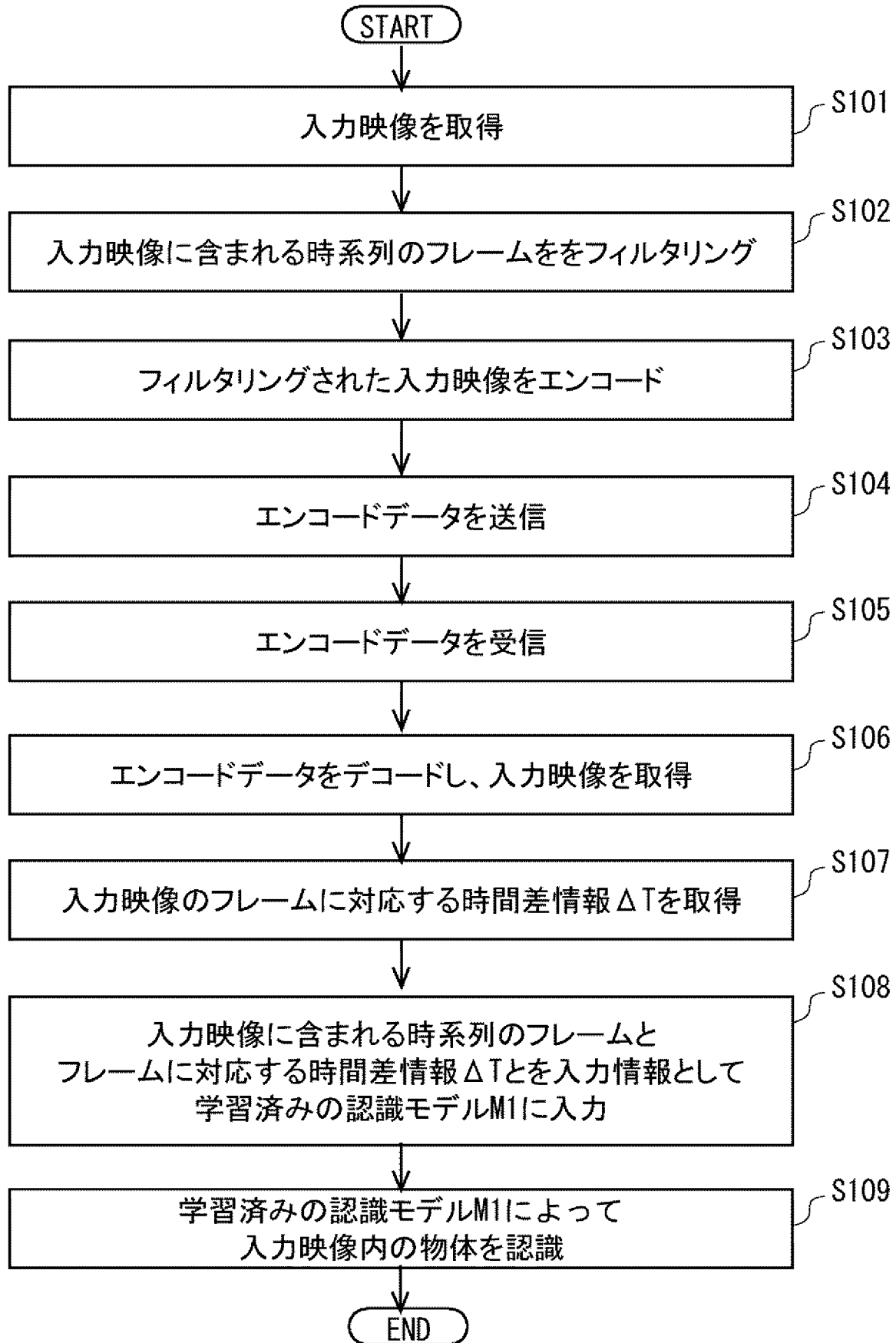


Fig. 7

[図8]

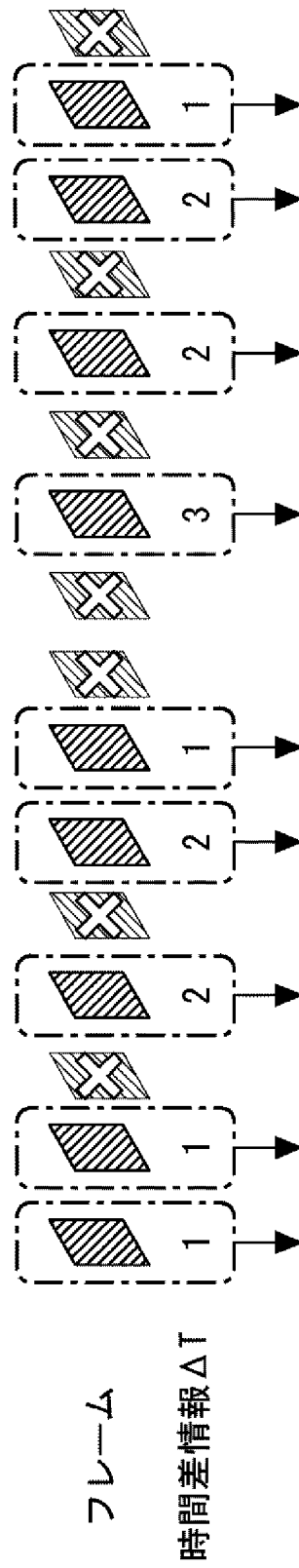


Fig. 8

[図9]

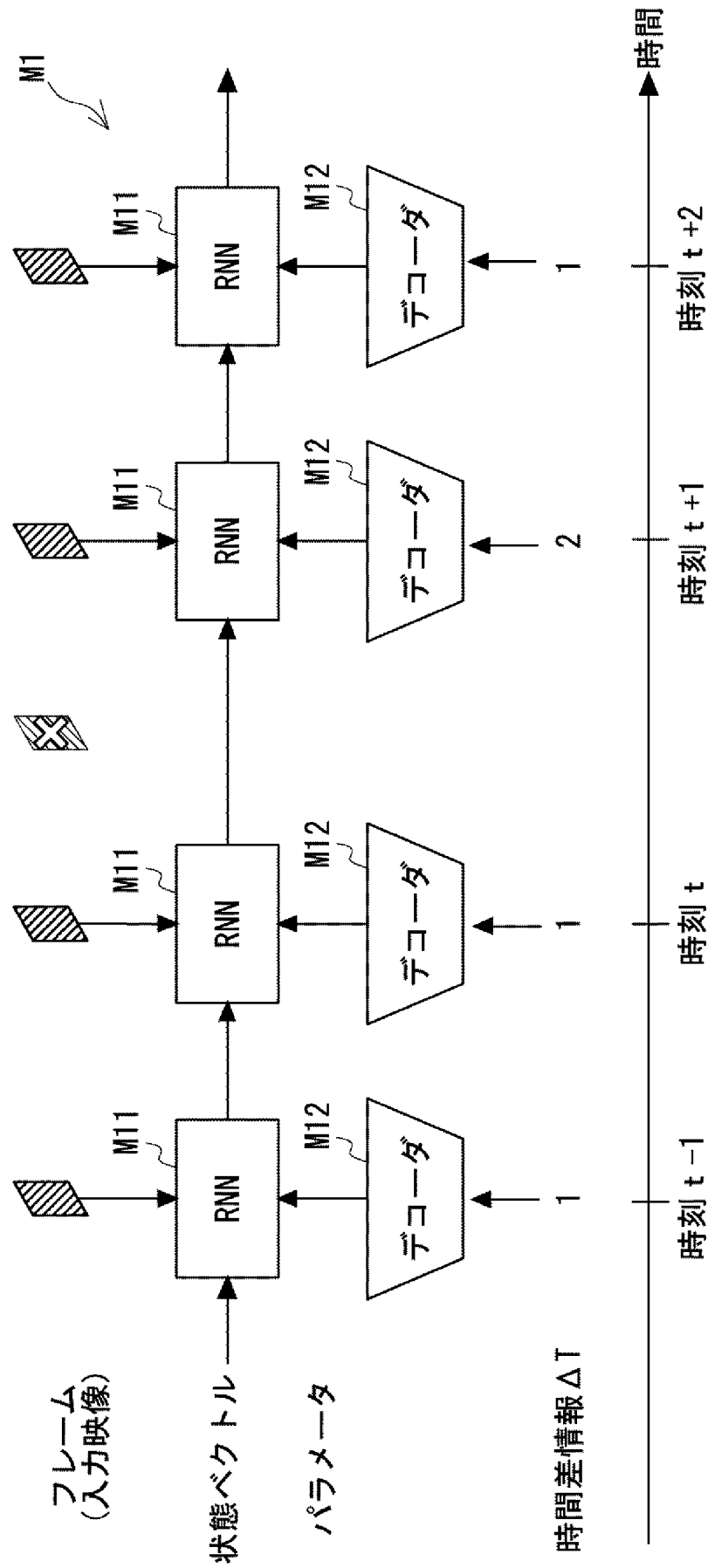


Fig. 9

[図10]

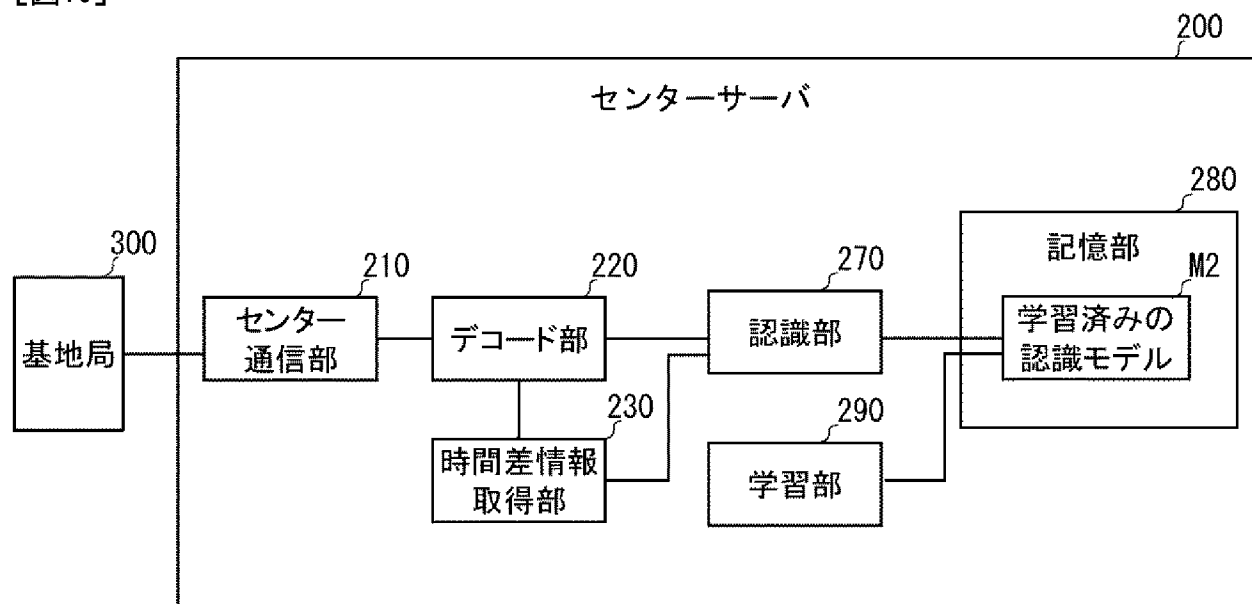


Fig. 10

[図11]

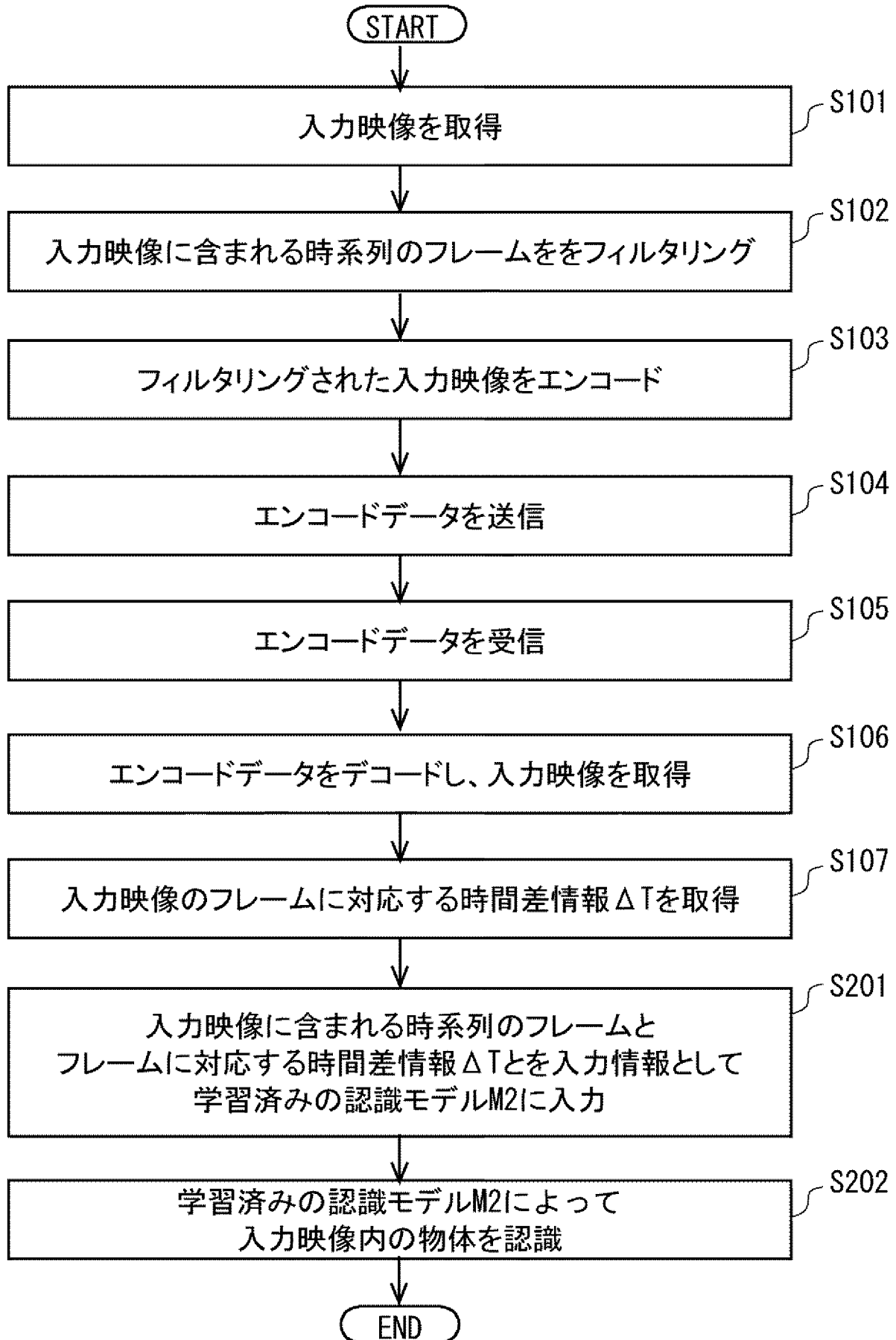


Fig. 11

[図12]

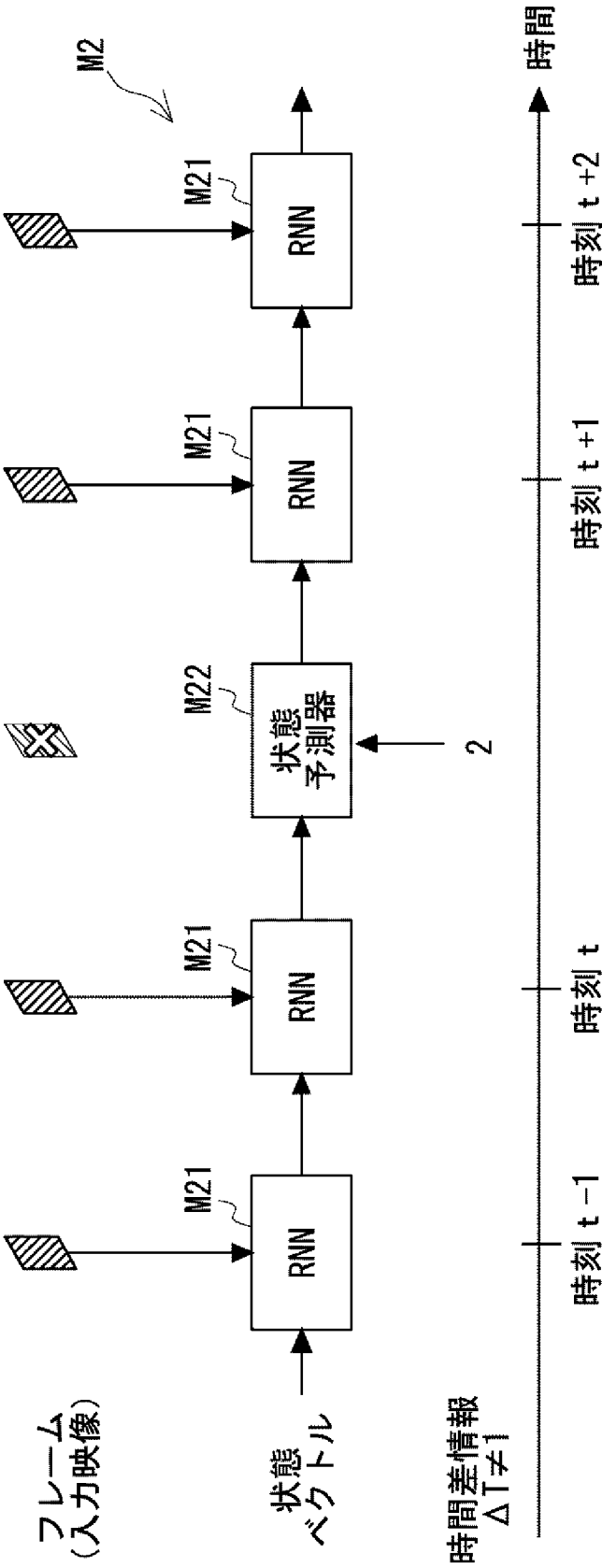


Fig. 12

[図13]

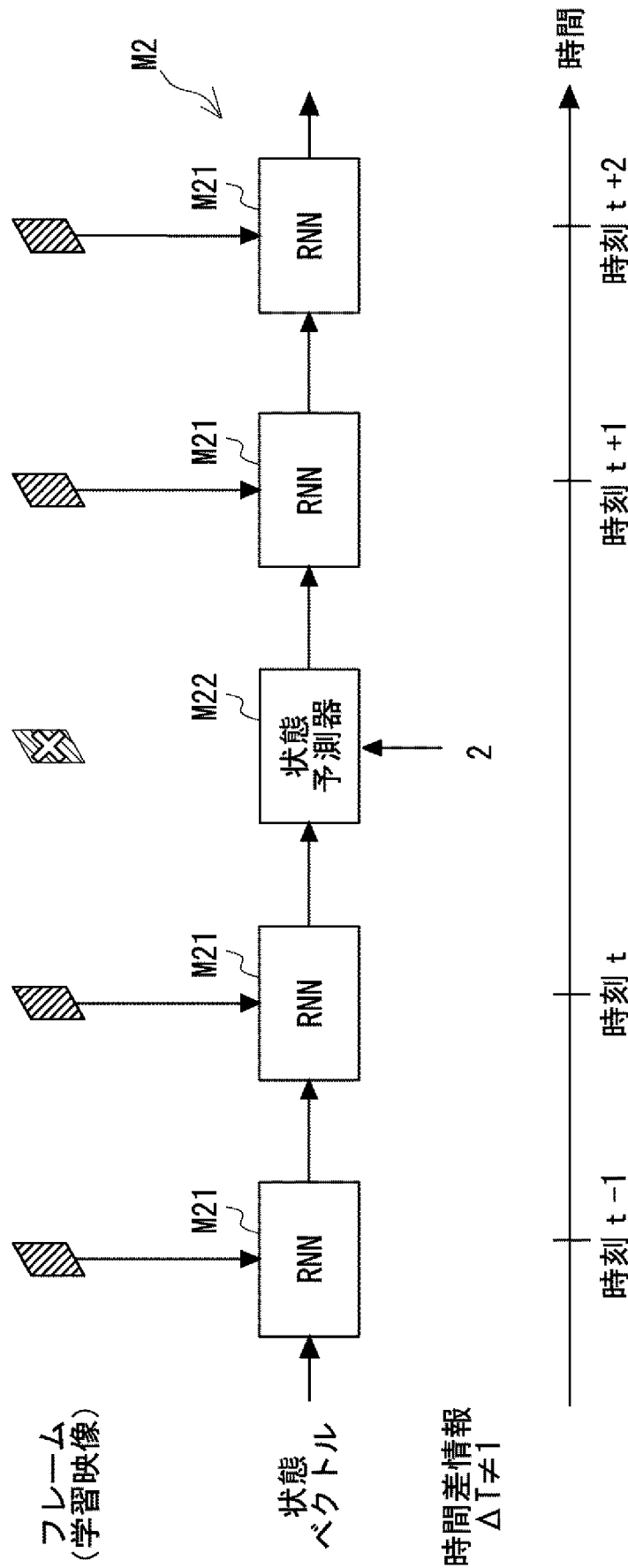


Fig. 13

[図14]

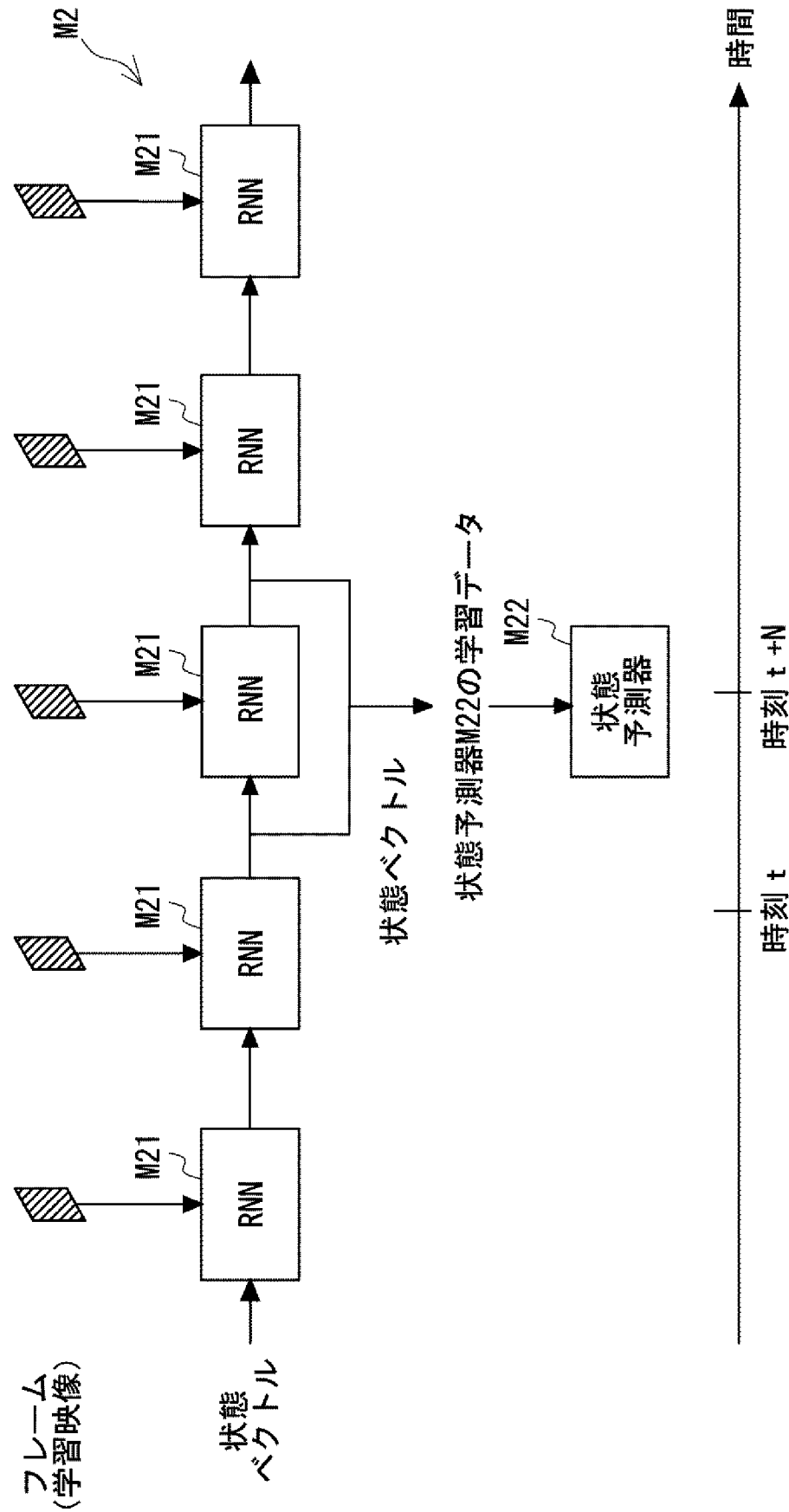


Fig. 14

[図15]

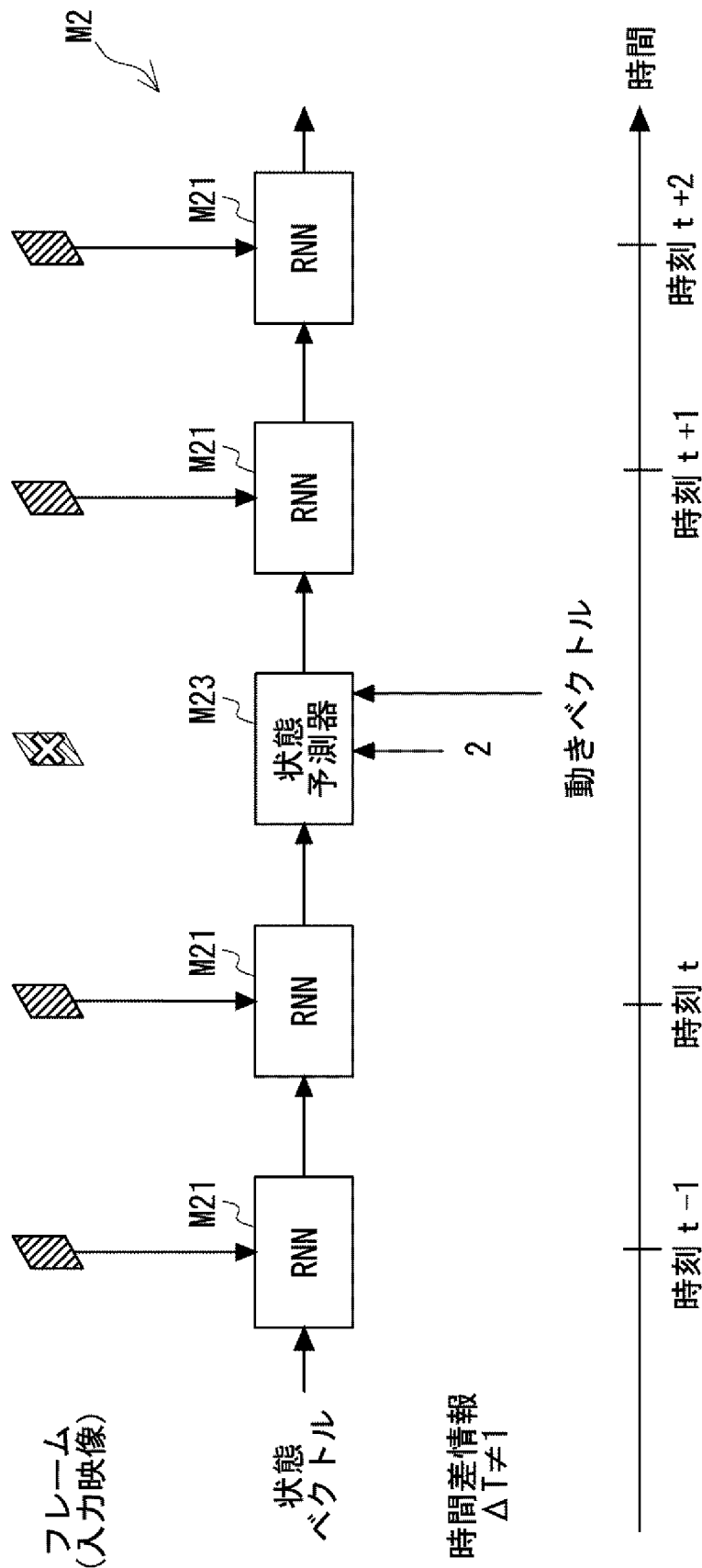


Fig. 15

[図16]

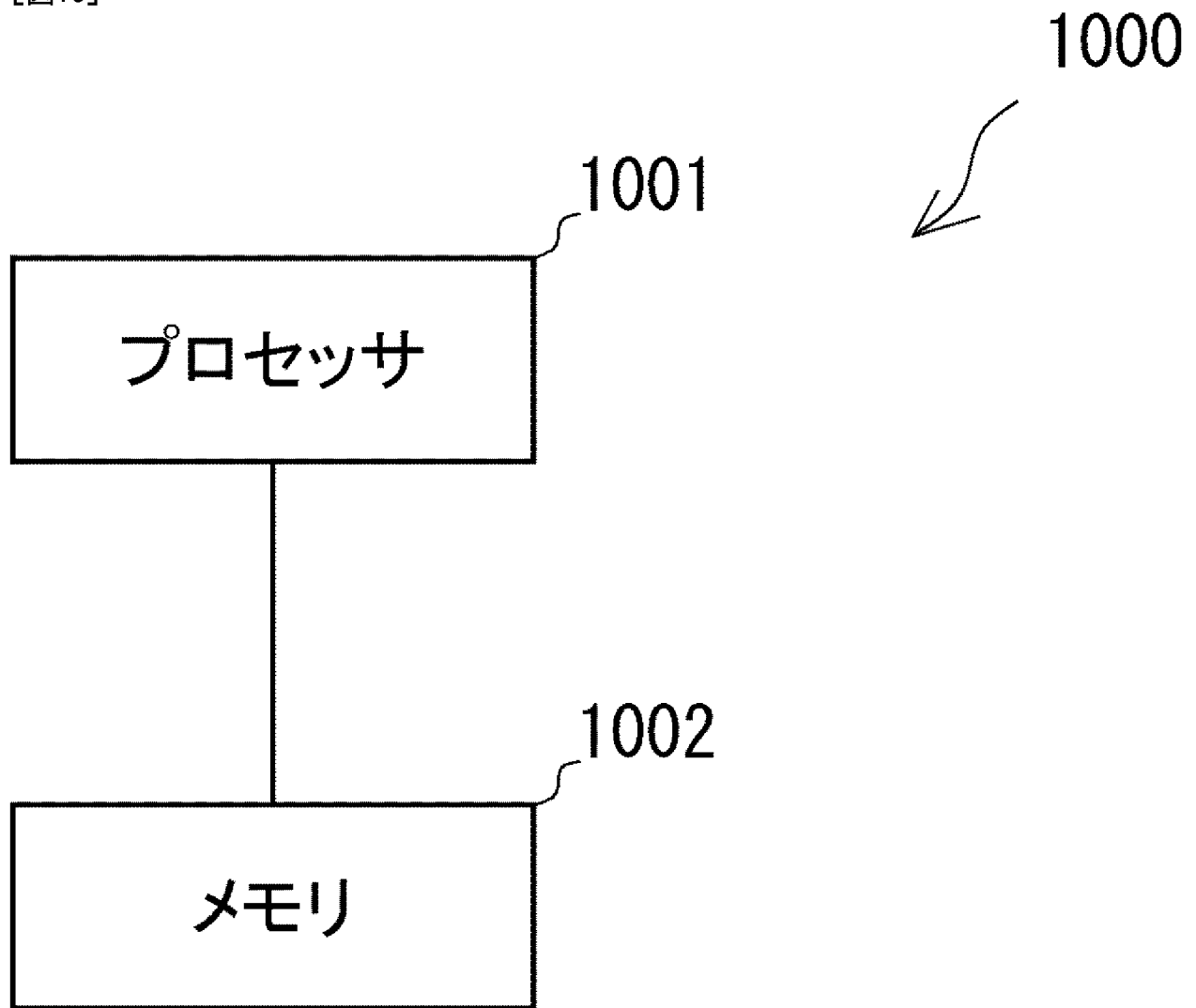


Fig. 16

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2022/034510

A. CLASSIFICATION OF SUBJECT MATTER**G06T 7/00**(2017.01)i

FI: G06T7/00 350C

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)

G06T7/00

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 1971-2022

Registered utility model specifications of Japan 1996-2022

Published registered utility model applications of Japan 1994-2022

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	JP 2012-73971 A (FUJIFILM CORP) 12 April 2012 (2012-04-12) entire text, all drawings	1-18
A	US 2018/0308253 A1 (SAMSUNG ELECTRONICS CO., LTD.) 25 October 2018 (2018-10-25) entire text, all drawings	1-18
A	JP 2018-5638 A (NIPPON TELEGRAPH & TELEPHONE) 11 January 2018 (2018-01-11) entire text, all drawings	1-18



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

22 November 2022

Date of mailing of the international search report

06 December 2022

Name and mailing address of the ISA/JP

Japan Patent Office (ISA/JP)

3-4-3 Kasumigaseki, Chiyoda-ku, Tokyo 100-8915

Japan

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/JP2022/034510

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
JP	2012-73971	A	12 April 2012	(Family: none)	
US	2018/0308253	A1	25 October 2018	CN 108734739 A entire text, all drawings	
JP	2018-5638	A	11 January 2018	(Family: none)	

A. 発明の属する分野の分類（国際特許分類（IPC）） G06T 7/00(2017.01)i FI: G06T7/00 350C										
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）） G06T7/00										
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>1971-2022年</td> </tr> <tr> <td>日本国実用新案登録公報</td> <td>1996-2022年</td> </tr> <tr> <td>日本国登録実用新案公報</td> <td>1994-2022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	1971-2022年	日本国実用新案登録公報	1996-2022年	日本国登録実用新案公報	1994-2022年
日本国実用新案公報	1922-1996年									
日本国公開実用新案公報	1971-2022年									
日本国実用新案登録公報	1996-2022年									
日本国登録実用新案公報	1994-2022年									
国際調査で使用了電子データベース（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	JP 2012-73971 A（富士フイルム株式会社）12.04.2012（2012-04-12） 全文、全図	1-18								
A	US 2018/0308253 A1（SAMSUNG ELECTRONICS CO., LTD.）25.10.2018（2018-10-25） 全文、全図	1-18								
A	JP 2018-5638 A（日本電信電話株式会社）11.01.2018（2018-01-11） 全文、全図	1-18								
☐ 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> * 引用文献のカテゴリー “A” 特に関連のある文献ではなく、一般の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> “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、一般の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、一般の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 22.11.2022	国際調査報告の発送日 06.12.2022									
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） 山田 辰美 5H 8383 電話番号 03-3581-1101 内線 3531									

国際調査報告
パテントファミリーに関する情報

国際出願番号

PCT/JP2022/034510

引用文献			公表日	パテントファミリー文献	公表日
JP	2012-73971	A	12.04.2012	(ファミリーなし)	
US	2018/0308253	A1	25.10.2018	CN 108734739 A	
				全文, 全図	
JP	2018-5638	A	11.01.2018	(ファミリーなし)	