

(19)日本国特許庁(JP)

**(12)特許公報(B2)**

(11)特許番号  
**特許第7635779号**  
**(P7635779)**

(45)発行日 令和7年2月26日(2025.2.26)

(24)登録日 令和7年2月17日(2025.2.17)

(51)国際特許分類	F I
G 0 6 N 20/00 (2019.01)	G 0 6 N 20/00
G 0 6 N 5/045(2023.01)	G 0 6 N 5/045
G 0 6 T 7/00 (2017.01)	G 0 6 T 7/00 1 3 0
H 0 4 N 21/258(2011.01)	G 0 6 T 7/00 3 5 0 C
G 1 0 L 15/10 (2006.01)	H 0 4 N 21/258

請求項の数 19 (全42頁) 最終頁に続く

(21)出願番号	特願2022-512040(P2022-512040)	(73)特許権者	000002185 ソニーグループ株式会社 東京都港区港南1丁目7番1号
(86)(22)出願日	令和3年3月24日(2021.3.24)	(74)代理人	100093241 弁理士 宮田 正昭
(86)国際出願番号	PCT/JP2021/012368	(74)代理人	100101801 弁理士 山田 英治
(87)国際公開番号	WO2021/200503	(74)代理人	100095496 弁理士 佐々木 榮二
(87)国際公開日	令和3年10月7日(2021.10.7)	(74)代理人	100086531 弁理士 澤田 俊夫
審査請求日	令和6年2月2日(2024.2.2)	(74)代理人	110000763 弁理士法人大同特許事務所
(31)優先権主張番号	特願2020-65069(P2020-65069)	(72)発明者	シン アンドリュー 東京都港区港南1丁目7番1号 ソニー 最終頁に続く
(32)優先日	令和2年3月31日(2020.3.31)		
(33)優先権主張国・地域又は機関	日本国(JP)		
(31)優先権主張番号	特願2020-120049(P2020-120049)		
(32)優先日	令和2年7月13日(2020.7.13)		
(33)優先権主張国・地域又は機関	日本国(JP)		

(54)【発明の名称】 学習システム及びデータ収集装置

**(57)【特許請求の範囲】****【請求項1】**

感動するコンテンツを生成し又は感動するコンテンツを撮影するためのカメラ操作を推定する機械学習モデルの学習を行うための学習システムであって、

データを収集するデータ収集装置と、前記データ収集装置が収集したデータを用いて機械学習モデルの学習を行う学習装置を備え、

前記データ収集装置は、カメラで撮影した画像データ及び前記カメラの撮影時の操作情報を含む学習データを収集し、

前記学習装置は、前記機械学習モデルの学習に影響を与える学習データを分析した結果に基づいて収集した、前記機械学習モデルの学習に所定以上の影響を及ぼす学習データ、不足している学習データ、又はこれらに類似するデータを用いて、前記機械学習モデルの再学習を行う。

学習システム。

**【請求項2】**

X A I ( E x p l a i n a b l e A I ) 、信頼度スコア計算、影響関数、又はベイジアン D N N による前記分析を行う、

請求項1に記載の学習システム。

**【請求項3】**

前記学習装置は、前記機械学習モデルの学習に影響を与える学習データを分析し、その分析結果に基づいて、前記機械学習モデルの学習に所定以上の影響を及ぼす学習データ、不

足している学習データ、又はこれらに類似するデータの送信を要求する要求信号を前記データ収集装置に送信し、

前記データ収集装置は、受信した前記要求信号に基づいて収集したデータを前記学習装置に送信し、

前記学習装置は、前記要求信号に応じて前記データ収集装置から送信されたデータに基づいて前記機械学習モデルの再学習を行う、

請求項 1 に記載の学習システム。

【請求項 4】

前記データ収集装置は、画像を撮影するカメラ又はイメージャであり、前記機械学習モデルの学習への影響度に基づいて、解像度、フレームレート、輝度、色彩、画角、視点位置又は視線方向を変化させて撮影した画像データを前記学習装置に送信する、  
10 請求項 1 に記載の学習システム。

【請求項 5】

前記学習装置は、前記機械学習モデルの学習データの送信を要求する要求信号を前記データ収集装置に送信し、

前記データ収集装置は、収集したデータが前記機械学習モデルの学習に与える影響を分析した結果に基づいて、前記収集したデータのうち前記機械学習モデルの学習に所定以上の影響を及ぼす学習データ、不足している学習データ、又はこれらに類似するデータを前記学習装置に送信し、

前記学習装置は、前記要求信号に応じて前記データ収集装置から送信されたデータに基づいて前記機械学習モデルの再学習を行う、  
20 請求項 1 に記載の学習システム。

請求項 1 に記載の学習システム。

【請求項 6】

前記学習装置は、前記要求信号の送信時に、前記分析に必要な情報を前記データ収集装置に送信する、

請求項 5 に記載の学習システム。

【請求項 7】

請求項 1 に記載の学習システムにおける前記学習装置から、前記機械学習モデルの学習データの送信を要求する要求信号を受信する受信部と、

前記要求信号を受信したことに応じて、前記機械学習モデルの学習に所定以上の影響を及ぼす学習データ、不足している学習データ、又はこれらに類似するデータを収集するデータ収集部と、  
30 請求項 1 に記載の学習システム。

前記データ収集部が収集したデータを前記学習装置に送信する送信部と、  
を具備するデータ収集装置。

【請求項 8】

前記受信部は、前記機械学習モデルの学習に所定以上の影響を及ぼす学習データ、不足している学習データ、又はこれらに類似するデータを要求する前記要求信号を前記学習装置から受信し、

前記データ収集部は、受信した前記要求信号に基づいてデータを収集し、

前記送信部は、前記データ収集部が収集したデータを前記学習装置に送信する、  
40 請求項 7 に記載のデータ収集装置。

【請求項 9】

前記データ収集部は受信した前記要求信号に基づいて新たにデータを収集し、又は、前記送信部は前記データ収集部が既に収集したデータから前記要求信号に基づいて抽出したデータを送信する、

請求項 8 に記載のデータ収集装置。

【請求項 10】

前記データ収集部が収集したデータが前記機械学習モデルの学習に与える影響を分析する分析部をさらに備え、

前記送信部は、前記分析部による分析結果に基づいて、前記データ収集部が収集したデータを送信する、  
50 請求項 8 に記載のデータ収集装置。

タのうち前記機械学習モデルの学習に所定以上の影響を及ぼす学習データ、不足している学習データ、又はこれらに類似するデータを前記学習装置に送信する、  
請求項 7 に記載のデータ収集装置。

**【請求項 1 1】**

前記データ収集部は、前記機械学習モデルの学習への影響度に基づいて、画像を撮影するカメラ又はイメージャの解像度、フレームレート、輝度、色彩、画角、視点位置、又は視線方向を変化させて撮影した画像データを収集する、  
請求項 7 に記載のデータ収集装置。

**【請求項 1 2】**

コンテンツのコンテンツ評価情報とコンテンツを視聴する人間の生体情報に基づいて、コンテンツからセグメントを抽出するセグメント抽出部と、  
セグメントに写っている人物の顔を検出して顔の表情を識別する表情識別部と、  
セグメントに写っている人物間の関係性を推定する関係性推定部と、  
セグメントに写っている人物の顔の表情と、人物間の関係性に基づいてセグメントの感動ラベルを識別する感動識別部と、  
を具備し、

前記感動識別部は、請求項 1 に記載の学習システムにおける前記学習装置が学習を行った感動スコア推定モデルを用いて感動ラベルを識別する、  
情報処理装置。

**【請求項 1 3】**

前記生体情報は少なくとも脳波情報を含む、  
請求項 1 2 に記載の情報処理装置。

**【請求項 1 4】**

前記セグメント抽出部は、コンテンツのうち高評価とポジティブな感情を持つ生体情報が一致しているセグメントを抽出する、  
請求項 1 2 に記載の情報処理装置。

**【請求項 1 5】**

前記セグメント抽出部は、高評価とポジティブな感情を持つ生体情報が一致しないセグメントをさらに抽出する、  
請求項 1 4 に記載の情報処理装置。

**【請求項 1 6】**

前記関係性推定部は、前後のセグメントとのコンテキストと、人物の顔の表情に基づいて、セグメントに写っている人物間の関係性を推定する、  
請求項 1 2 に記載の情報処理装置。

**【請求項 1 7】**

前記関係性推定部は、前記コンテキストとして、過去のセグメントに写っている人物の顔検出及び検出顔の表情識別結果を用いて、現在のセグメントに写っている人物間の関係性を推定する、  
請求項 1 6 に記載の情報処理装置。

**【請求項 1 8】**

前記セグメントに含まれる音声を認識して得られたテキスト情報に基づいて感情を分析する第 1 の感情分析部をさらに備え、

前記感動識別部は、前記第 1 の感情分析部がテキスト情報から識別した感情をさらに考慮して、セグメントの感動ラベルを識別する、  
請求項 1 2 に記載の情報処理装置。

**【請求項 1 9】**

前記セグメントに含まれる音楽を検出して、その音楽が与える感情を分析する第 2 の感情分析部をさらに備え、

前記感動識別部は、前記第 2 の感情分析部が音楽から識別した感情をさらに考慮して、セグメントの感動ラベルを識別する、

10

20

30

40

50

請求項1\_2に記載の情報処理装置。

【発明の詳細な説明】

【技術分野】

【0001】

本明細書で開示する技術（以下、「本開示」とする）は、機械学習モデルの学習を行う学習システム、機械学習モデルを学習するための学習データの収集処理を行うデータ収集装置、並びにデータの分析処理を行う情報処理装置に関する。

【背景技術】

【0002】

現在、デジタルカメラは広範に普及している。例えば、各家庭では、日常生活や、子供の入学式や卒業式、結婚式といったさまざまな行事を記録するために、デジタルビデオカメラやデジタルスチルカメラが利用される。より感動的なシーンを撮影することが望ましいが、一般ユーザがファインダ越しに観察しているシーンが感動を与えるシーンかどうかを判別することは難しい。もちろん、プロのカメラマンにとっても、撮影しているシーンが感動的かどうかを判別することは難しい。

10

【0003】

例えば、学習済みニューラルネットワークモデルを搭載した人工知能技術を用いて、感動を与えるコンテンツを扱う方法が考えられる。特に、複数の畳み込みニューラルネットワークの層を備えたディープラーニングのニューラルネットワークモデル（以下、「DNN」とも呼ぶ）は、訓練すなわちディープラーニングを通じて開発者には想像できない特徴を多くのデータから抽出して、開発者がアルゴリズムを想定し得ないような複雑な問題解決を行うことのできる人工知能機能を開発することができる。

20

【0004】

人工知能技術を利用するには、人工知能を学習するための学習データが必要である。例えば、指定された能力の機械学習のための学習データを収集する学習データ収集部と、収集した学習データを用いて指定された能力を獲得させるように学習器の機械学習を行う学習処理部を備えた学習装置について提案がなされている（特許文献1を参照のこと）。人工知能技術を用いて感動を与えるコンテンツを扱うには、まず人間に感動を与えるコンテンツを生成し又は感動するコンテンツを撮影するためのカメラ操作を推定する機械学習モデルの学習に用いられる大量の学習データが必要である。

30

【0005】

ところが、感動の主観的な性質上、学習データの収集は大変に困難である。SNS（Social Network Service）などで現れる好みの情報を活用することによって、主観性を反映するデータ収集方法が考えられるが、プライバシーを脅かすそれがあるため実行が容易でない。また、コンテンツ中の個々のシーンの画像や背景音楽だけでなく前後のシーンのコンテキストも感動を生み出す要素と考えられるが、画像のみに依存してデータ収集を行う方法ではコンテキストを反映したデータを収集することは難しい。

【先行技術文献】

【特許文献】

【0006】

【文献】特開2018-149669号公報

40

【非特許文献】

【0007】

【文献】Using Brain Data for Sentiment Analysis. Gu et al. JLCL 2014.

【文献】Music Emotion Classification: A Fuzzy Approach. Yang et al. ACM MM 2006.

【文献】Visual Social Relationship Recognition. Li et al. arxiv 2018.

【文献】Understanding Black-box Predictions v

50

ia Influence Functions, Pang Wei Kho and Percy Liang <<https://arxiv.org/abs/1703.04730>>

【文献】Alex Kendall, Yarin Gal, "What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision", NIPS 2017

【発明の概要】

【発明が解決しようとする課題】

【0008】

本開示の目的は、感動するコンテンツを生成し又は感動するコンテンツを撮影するためのカメラ操作を推定する機械学習モデルの学習を行う学習システム、感動するコンテンツを生成し又は感動するコンテンツを撮影するためのカメラ操作を推定する機械学習モデルの学習データの収集処理を行うデータ収集装置、並びに収集したデータが与える感動を分析する情報処理装置を提供することにある。

10

【課題を解決するための手段】

【0009】

本開示は、上記課題を参照してなされたものであり、その第1の側面は、データを収集するデータ収集装置と、前記データ収集装置が収集したデータを用いて機械学習モデルの学習を行う学習装置を備え、

前記学習装置は、前記機械学習モデルの学習に影響を与える学習データを分析した結果に基づいて収集した、前記機械学習モデルの学習に所定以上の影響を及ぼす学習データ、不足している学習データ、又はこれらに類似するデータを用いて、前記機械学習モデルの再学習を行う、

20

学習システムである。

【0010】

但し、ここで言う「システム」とは、複数の装置（又は特定の機能を実現する機能モジュール）が論理的に集合した物のことを言い、各装置や機能モジュールが単一の筐体内にあるか否かは特に問わない。

【0011】

前記学習装置は、感動するコンテンツを生成し又は感動するコンテンツを撮影するためのカメラ操作を推定する機械学習モデルの学習を行う。また、前記分析は、XAI、信頼度スコア計算、影響関数、又はベイジアンDNNにより行うことができる。

30

【0012】

前記学習装置は、前記機械学習モデルの学習に影響を与える学習データを分析し、その分析結果に基づいて、前記機械学習モデルの学習に所定以上の影響を及ぼす学習データ、不足している学習データ、又はこれらに類似するデータの送信を要求する要求信号を前記データ収集装置に送信する。そして、前記データ収集装置は、受信した前記要求信号に基づいて収集したデータを前記学習装置に送信することで、前記学習装置は、前記要求信号に応じて前記データ収集装置から送信されたデータに基づいて前記機械学習モデルの再学習を行うことができる。

40

【0013】

あるいは、前記データ収集装置は、前記学習装置から前記要求信号を受信すると、収集したデータが前記機械学習モデルの学習に与える影響を分析した結果に基づいて、前記収集したデータのうち前記機械学習モデルの学習に所定以上の影響を及ぼす学習データ、不足している学習データ、又はこれらに類似するデータを前記学習装置に送信する。そして、前記学習装置は、前記要求信号に応じて前記データ収集装置から送信されたデータに基づいて前記機械学習モデルの再学習を行うことができる。前記学習装置は、前記要求信号の送信時に、前記分析に必要な情報を前記データ収集装置に送信するようにしてもよい。

【0014】

また、本開示の第2の側面は、

50

機械学習モデルの学習を行う学習装置から、前記機械学習モデルの学習データの送信を要求する要求信号を受信する受信部と、

前記要求信号を受信したことにより、前記機械学習モデルの学習に所定以上の影響を及ぼす学習データ、不足している学習データ、又はこれらに類似するデータを収集するデータ収集部と、

前記データ収集部が収集したデータを前記学習装置に送信する送信部と、  
を具備するデータ収集装置である。

#### 【0015】

前記受信部は、前記機械学習モデルの学習に所定以上の影響を及ぼす学習データ、不足している学習データ、又はこれらに類似するデータを要求する前記要求信号を前記学習装置から受信する。そして、前記データ収集部は、受信した前記要求信号に基づいてデータを収集し、前記送信部は、前記データ収集部が収集したデータを前記学習装置に送信する。10

#### 【0016】

あるいは、第2の側面に係るデータ収集装置は、前記データ収集部が収集したデータが前記機械学習モデルの学習に与える影響を分析する分析部をさらに備えている。そして、前記送信部は、前記分析部による分析結果に基づいて、前記データ収集部が収集したデータのうち前記機械学習モデルの学習に所定以上の影響を及ぼす学習データ、不足している学習データ、又はこれらに類似するデータを前記学習装置に送信する。

#### 【0017】

また、本開示の第3の側面は、20  
コンテンツのコンテンツ評価情報とコンテンツを視聴する人間の生体情報に基づいて、  
コンテンツからセグメントを抽出するセグメント抽出部と、  
セグメントに写っている人物の顔を検出して顔の表情を識別する表情識別部と、  
セグメントに写っている人物間の関係性を推定する関係性推定部と、  
セグメントに写っている人物の顔の表情と、人物間の関係性に基づいてセグメントの感動ラベルを識別する感動識別部と、  
を具備する情報処理装置である。

#### 【0018】

前記セグメント抽出部は、コンテンツのうち高視聴率とポジティブな感情を持つ生体情報が一致しているセグメントを抽出する。また、前記セグメント抽出部は、高視聴率とポジティブな感情を持つ生体情報が一致しないセグメントをさらに抽出する。30

#### 【0019】

前記関係性推定部は、前記コンテキストとして、過去のセグメントに写っている人物の顔検出及び検出顔の表情識別結果を用いて、現在のセグメントに写っている人物間の関係性を推定する。

#### 【0020】

第3の側面に係る情報処理装置は、前記セグメントに含まれる音声を認識して得られたテキスト情報に基づいて感情を分析する第1の感情分析部と、前記セグメントに含まれる音楽を検出して、その音楽が与える感情を分析する第2の感情分析部をさらに備える。そして、前記感動識別部は、前記第1の感情分析部がテキスト情報から識別した感情と、前記第2の感情分析部が音楽から識別した感情をさらに考慮して、セグメントの感動ラベルを識別する。40

#### 【0021】

前記データ収集装置は、信頼度スコア計算により前記分析を行う。この場合、前記学習装置は、前記分析に必要な情報として、現時刻までに学習済みの前記ニューラルネットワークモデルの情報を前記データ収集装置に送信する。

#### 【発明の効果】

#### 【0022】

本開示によれば、感動するコンテンツを生成し又は感動するコンテンツを撮影するためのカメラ操作を推定する機械学習モデルの学習を行う学習システム、感動するコンテンツ50

を生成し又は感動するコンテンツを撮影するためのカメラ操作を推定する機械学習モデルの学習データの収集処理を行うデータ収集装置、並びにコンテンツが与える感動を識別する情報処理装置を提供することができる。

#### 【0023】

なお、本明細書に記載された効果は、あくまでも例示であり、本開示によりもたらされる効果はこれに限定されるものではない。また、本開示が、上記の効果以外に、さらに付加的な効果を奏する場合もある。

#### 【0024】

本開示のさらに他の目的、特徴や利点は、後述する実施形態や添付する図面に基づくより詳細な説明によって明らかになるであろう。

10

#### 【図面の簡単な説明】

#### 【0025】

【図1】図1は、データ収集システム100の機能的構成を示した図である。

【図2】図2は、データ収集システム100の動作を示したフローチャートである。

【図3】図3は、セグメントのコンテキストと顔検出及び表情識別結果に基づいて人物間の関係性を推定する仕組みを示した図である。

【図4】図4は、収集した学習データを利用した感動識別機の学習プロセスの全体の流れを示した図である。

【図5】図5は、デジタルカメラ500の構成例を示した図である。

【図6】図6は、デジタルカメラ500で撮影したコンテンツの感動識別ラベリングを行うための機能的構成を示した図である。

20

【図7】図7は、撮影したコンテンツの感動識別結果に基づいてカメラワークを自動制御するためのデジタルカメラ500の機能的構成を示した図である。

【図8】図8は、撮影したコンテンツの感動識別結果に基づいてキャプションを自動生成するためのデジタルカメラ500の機能的構成を示した図である。

【図9】図9は、撮影したコンテンツの感動識別結果に基づいて背景音楽を自動付与するためのデジタルカメラ500の機能的構成を示した図である。

【図10】図10は、カメラの自動操作のためのニューラルネットワークモデルを効率的に学習するための学習システム1000の構成を示した図である。

【図11】図11は、観測予測モデル1100の構成を示した図である。

30

【図12】図12は、操作モデル1200の構成を示した図である。

【図13】図13は、操作推定モデル1300の構成を示した図である。

【図14】図14は、感動スコア推定モデル1400を示した図である。

【図15】図15は、制御対象と平均感動スコアの関係を示した図である。

【図16】図16は、出力の誤差を推定するように学習されたニューラルネットワークモデルを示した図である。

【図17】図17は、データ収集装置1010の内部構成例を示した図である。

【図18】図18は、データ収集装置1010の他の内部構成例を示した図である。

【図19】図19は、学習装置1030の内部構成例を示した図である。

【図20】図20は、学習装置1030の他の内部構成例を示した図である。

40

【図21】図21は、学習装置1030のさらに他の内部構成例を示した図である。

【図22】図22は、モデル利用装置1020の内部構成例を示した図である。

【図23】図23は、エッジAIとしてのモデル利用装置1020の内部構成例を示した図である。

#### 【発明を実施するための形態】

#### 【0026】

以下、図面を参照しながら本開示の実施形態について、以下の順に従って詳細に説明する。

#### 【0027】

A．感動を与えるコンテンツを生成する機械学習モデルに関するデータの収集処理を行う

50

## システムの構成

B . 感動を与えるコンテンツを生成する機械学習モデルに関するデータの収集処理を行うシステムの動作

C . コンテンツに含まれる人物間などの関係性の推定

D . 感動識別

E . 収集したデータを用いた学習プロセス

F . 推論

G . アプリケーション

H . 機械学習モデルの効率的学習

### 【 0 0 2 8 】

A . 感動を与えるコンテンツを生成する学習モデルに関するデータの収集処理を行うシステムの構成

図 1 には、本開示に係るデータ収集処理を行うデータ収集システム 1 0 0 の機能的構成を模式的に示している。

### 【 0 0 2 9 】

図示のデータ収集システム 1 0 0 は、膨大な映像コンテンツ 1 0 1 の中から人間に感動を与えるコンテンツを生成し又は感動するコンテンツを撮影するためのカメラ操作を推定する機械学習モデルの学習に用いられるデータの収集処理を行う。データ収集システム 1 0 0 で収集したデータは、基本的には、人間に感動を与えるコンテンツを扱うための（例えば、人間に感動を与えるシーンやコンテンツであるかどうかを推論するための）機械学習モデル（例えば、ニューラルネットワーク）を学習させるための学習データに用いられるが、もちろんそれ以外の用途に利用してもよい。

### 【 0 0 3 0 】

本実施形態では、データ収集システム 1 0 0 は、映像コンテンツ 1 0 1 として、テレビやインターネット放送局などで放送され、視聴率情報などのコンテンツ評価情報が付与されたコンテンツを扱うものとし、コンテンツは、主に映像データとオーディオデータなどで構成される。また、映像コンテンツ 1 0 1 として、例えば「いいね！（好き、楽しい、支持できる）」など視聴率情報に類似する（又は、視聴率情報に置き換えることが可能な）、コンテンツを評価するコンテンツ評価情報が付与されていれば、動画共有サイトなどさまざまなコンテンツを含んでいてもよい。また、データ収集システム 1 0 0 は、固定長又は可変長のセグメント単位で、映像コンテンツ 1 0 1 から取り出したコンテンツの処理を行うものとする。

### 【 0 0 3 1 】

データ収集システム 1 0 0 は、コンテンツ評価情報取得部 1 0 2 と、生体情報取得部 1 0 3 と、比較部 1 0 4 と、セグメント抽出部 1 0 5 と、コンテキスト抽出部 1 0 6 と、音声認識部 1 0 7 と、音楽検出部 1 0 8 と、顔検出部 1 0 9 と、第 1 の感情分析部 1 1 0 と、第 2 の感情分析部 1 1 1 と、表情識別部 1 1 2 と、関係性推定部 1 1 3 と、感動識別部 1 1 4 を備えている。データ収集システム 1 0 0 は、例えばクラウド、すなわちインターネットなどの広域的なコンピュータネットワーク上でコンピュータ資源が提供されるサービスとして構成されることを想定している。データ収集システム 1 0 0 の各構成要素は、1 台のコンピュータ内に集約される場合と複数のコンピュータに分散して配置される場合がある。以下、各部について説明する。

### 【 0 0 3 2 】

コンテンツ評価情報取得部 1 0 2 は、映像コンテンツ 1 0 1 から取り出されたコンテンツに付与されている視聴率などのコンテンツ評価情報を取得する。コンテンツ評価情報取得部 1 0 2 は、対象とするコンテンツの視聴率情報を、視聴率調査会社から取得するようにもよい。本実施形態では、データ収集システム 1 0 0 は、コンテンツの各シーンが人間に感動を与えるかどうかを判別するために、視聴率情報を利用する。視聴率情報を利用するメリットとして、プライバシー問題に抵触しないことや、既に多人数の反応が反映される情報であることなどが挙げられる。また、視聴率情報は、コンテンツのストーリー

10

20

30

40

50

上の起承転結などのコンテキストも反映されているので、視聴率が高いシーンの前後のコンテキスト情報も有効に使うことができる。さらに、視聴率情報が付与されているコンテンツの場合、高い頻度でそのシーンに付随する追加情報（例えば、セリフや背景音楽など）などが付いている。後述するように、感動的なシーンのセリフと背景の音楽は、人間に感動を与えることに関して大きな役割を持つとともに、アプリケーションの余地を広げるために、活用するべきである。なお、コンテンツ評価情報取得部102は、視聴率情報に加えて（又は、視聴率情報の代わりに）、SNSなどで付けられる「いいね！（好き、楽しい、支持できる）」などのコンテンツ評価情報を取得するようにしてもよい。

#### 【0033】

生体情報取得部103は、映像コンテンツ101から取り出されたコンテンツを視聴する人間の生体情報を取得する。例えば、家庭内に脳波や発汗、視線、筋電位などの生体情報を検出する生体情報センサーを設置しておき、テレビやインターネット放送などのコンテンツを視聴する際に検出された生体情報を、生体情報取得部103が収集するようにしてもよい。生体情報取得部103は、多数の家庭から膨大量の生体情報を収集した場合には、平均化などの統計処理を行った後に利用するようにしてもよい。

10

#### 【0034】

再生コンテンツを視聴する人間は、緊張感や悲しみ、怒りなどの感動を覚えたシーンでは、脳波や発汗、視線、筋電位などの生体情報に反応が現れる。他方、視聴率情報はコンテンツがどれくらいの世帯や人々に見られているのかを示す指標であるが、必ずしも人が感動したシーンが視聴率情報に反映されるとは限らない。そこで、データ収集システム100では、視聴率などのコンテンツ評価情報のみではコンテンツの各シーンが人間に感動を与えるかどうかを正確に判別できないというリスクを考慮して、コンテンツ評価情報を補間するために、生体情報取得部103が取得する脳波情報などの生体情報をを利用する。脳波情報から感情を識別する研究は、既に行われている（例えば、非特許文献1を参照のこと）。したがって、コンテンツの各シーンを視聴する人間の脳波情報に基づいて、高視聴率がどのような感情に起因しているのかを推定することができる。なお、脳波は、脳から生じる電気活動を頭皮から計測した電位のことである。脳波計は、一般に、頭皮上に設置した電極から脳電位を計測するように構成される。電極の配置位置として国際10-20法が知られているが、本開示は特にこれに限定されない。

20

#### 【0035】

比較部104は、コンテンツ評価情報取得部102が取得する視聴率などのコンテンツ評価情報と、生体情報取得部103が取得する人間の生体情報を比較する。そして、セグメント抽出部105は、映像コンテンツ101から取り出したコンテンツのうち、高視聴率とポジティブな感情（「感動した」など）を持つ生体情報が一致しているセグメントを、感動識別の学習に用いるデータとして抽出する。また、セグメント抽出部105は、高視聴率とポジティブな感情を持つ生体情報が一致しないセグメントも、ネガティブ（「感動しない」など）なサンプルとして抽出する。

30

#### 【0036】

コンテキスト抽出部106は、映像コンテンツ101から取り出したコンテンツのコンテキストを抽出する。より具体的には、コンテキスト抽出部106は、視聴率が高いシーンの前後のコンテキストや、セグメント抽出部105が抽出するセグメントの前後のコンテキストを抽出する。コンテキスト抽出部106は、映像及びオーディオデータからなるコンテンツからコンテキストを推定するように学習された学習済みニューラルネットワークモデルを用いて、コンテキストを抽出するようにしてもよい。

40

#### 【0037】

音声認識部107は、セグメント抽出部105が抽出したセグメントに含まれるオーディオデータのうち音声の成分に対して音声認識処理（ASR：Automatic Speech Recognition）を適用して、セリフなどの音声を認識してテキストに変換（speech transcription）する。

#### 【0038】

50

第1の感情分析部110は、音声認識部107から出力されるセリフなどのテキスト情報に対して感情分析を行う自然言語処理を適用して、そのセグメント（シーン）がどのような感情で構成されているのかを分析する。第1の感情分析部110は、前後のセグメント（シーン）のセリフを参考にして、より広いコンテキストの中で感情分析を行うようにしてもよい。セリフの感情分析結果に基づいて、セグメント（シーン）の感情をより正確に把握することができる。

#### 【0039】

音声認識部107や第1の感情分析部110はそれぞれ、畳み込みニューラルネットワーク（Convolutional Neural Network : CNN）などの学習済み機械学習モデルを用いて構成することができる。また、音声認識部107と第1の感情分析部110を併せて「音声識別機」として、1つのCNNなどの機械学習モデルで構成することもできる。10

#### 【0040】

音楽検出部108は、セグメント抽出部105が抽出したセグメントに含まれるオーディオデータに対して音楽検出処理を適用して、映像に付与された背景音楽を検出する。

#### 【0041】

第2の感情分析部111は、音楽検出部108が検出した背景音楽が持つムードや感動的であるかどうかを識別する。音楽が与える感情を識別する技術はいくつか提案されている（例えば、非特許文献2を参照のこと）。背景音楽の感情分析結果に基づいて、セグメント（シーン）の感情をより正確に把握することができる。20

#### 【0042】

音楽検出部108や第2の感情分析部111は、それぞれCNNなどの学習済み機械学習モデルを用いて構成することができる。また、音楽検出部108と第2の感情分析部111を併せて「音楽識別機」として、1つのCNNなどの機械学習モデルで構成することもできる。

#### 【0043】

顔検出部109は、セグメント抽出部105が抽出したセグメントに含まれる映像データに写っている人物の顔検出処理（face detection）を実行する。

#### 【0044】

表情識別部112は、顔検出部109が検出した顔の表情を識別する。表情識別部112は、例えば、検出顔から顔画像のパターンを解析し、その顔画像のパターンがどの感情に対応するかを推定する。顔画像のパターンと人間の感情との相関関係を事前学習した人工知能を用いて、人間の感情を推定することができる。なお、顔画像のパターンは、眉毛、目、鼻、口、頬といった顔パーツの組み合わせで構成することもできるが、顔パーツに分けて顔画像全体のイメージで構成することもできる。30

#### 【0045】

顔検出部109や表情識別部112は、それぞれCNNなどの学習済み機械学習モデルを用いて構成することができる。また、顔検出部109と表情識別部112を併せて「表情識別機」として、1つのCNNなどの機械学習モデルで構成することもできる。

#### 【0046】

関係性推定部113は、セグメントに写っている人物が二人以上の場合において、表情識別部112が識別した個々の人物の表情と、コンテキスト抽出部106が抽出した前後のセグメントとのコンテキストに基づいて、人物間の関係性や親密度を推定する。何故ならば、ある人物が同じ表情をしていても、一緒に写っている他の人物との関係性や親密度（例えば、家族や親しい友人と一緒に写っているか）に応じて、それを見た人間の感動に繋がり易さに違いがあり、後段の感動識別処理に影響を与えるからである。40

#### 【0047】

感動識別部114は、音声認識部107及び第1の感情分析部110によるセグメント内の音声（セリフなど）の感情識別結果と、音楽検出部108及び第2の感情分析部111によるセグメント内の背景音楽の感情やムードの識別結果と、顔検出部109及び表情識

別部 112 によるセグメントに写っている人物の表情の識別結果と、関係性推定部 113 によるセグメントに写っている人物間の関係性や親密度の推定結果に基づいて、セグメントの感動を識別する。

#### 【0048】

感動識別部 114 は、例えば学習済みニューラルネットワークモデルを用いて、上記の入力データからセグメントの感動を推定処理して、セグメントの感動のレベルを示す感動識別ラベルを出力する。感動識別ラベルは、ポジティブ（「感動した」など）とネガティブ（「感動しない」など）の 2 種類の簡単なバイナリラベルでもよいが、「怒り」、「嫌悪」、「恐れ」、「幸福」、「悲しみ」、「驚き」、…などさまざまな種類に分化した感動を表すラベルであってもよい。感動識別部 114 は、例えば W u n d t モデルや P l u t c h i k モデルなどの感情モデルに基づいて定義された感動識別ラベルを出力するようにもよい。10

#### 【0049】

B. 感動を与えるコンテンツを生成する機械学習モデルに関するデータの収集処理を行うシステムの動作

図 2 には、データ収集システム 100 の動作をフローチャートの形式で示している。

#### 【0050】

まず、テレビやインターネット放送局などで放送され、視聴率などのコンテンツ評価情報が付与されたコンテンツがデータ収集システム 100 に取り込まれる（ステップ S S 2 0 1）。コンテンツ評価情報取得部 102 は、このコンテンツに付与されているコンテンツ評価情報を取得する。コンテンツは複数のセグメントで構成され、セグメント毎にコンテンツ評価情報が付与されているものとする。20

#### 【0051】

生体情報取得部 103 は、ステップ S 2 0 1 で取り込んだコンテンツを視聴中の人間の反応を表す脳波などの生体情報を取得する（ステップ S 2 0 2）。コンテンツがテレビ放送の場合、その番組を視聴する多数の家庭から、少なくとも脳波を含む生体情報を収集することを想定している。

#### 【0052】

比較部 104 は、セグメント毎に、コンテンツ評価情報取得部 102 が取得したコンテンツ評価情報と、生体情報取得部 103 が取得する人間の生体情報を比較して、高視聴率とポジティブな感情を持つ生体情報が一致するかどうかをチェックする（ステップ S 2 0 3）。

#### 【0053】

ここで、セグメント抽出部 105 は、映像コンテンツ 101 から取り出したコンテンツのうち、高視聴率とポジティブな感情を持つ生体情報が一致しているセグメントを（ステップ S 2 0 3 の Y e s ）、感動識別の学習に用いるデータとして抽出する（ステップ S 2 0 4）。また、セグメント抽出部 105 は、高視聴率とポジティブな感情を持つ生体情報が一致しないセグメントを（ステップ S 2 0 3 の N o ）、ネガティブなサンプルとして抽出する（ステップ S 2 1 5）。

#### 【0054】

ステップ S 2 0 4 で抽出したセグメントが音声データを含む場合には（ステップ S 2 0 5 の Y e s ）、音声認識部 107 が音声認識してテキストに変換し、第 1 の感情分析部 10 がそのテキスト情報から感情を識別する（ステップ S 2 0 6）。

#### 【0055】

また、ステップ S 2 0 4 で抽出したセグメントが音楽データを含む場合には（ステップ S 2 0 7 の Y e s ）、音楽検出部 108 がその音楽を検出し、第 2 の感情分析部 111 がその音楽が与える感情を識別する（ステップ S 2 0 8）。

#### 【0056】

また、ステップ S 2 0 4 で抽出したセグメントに人物が写っている場合には（ステップ S 2 0 9 の Y e s ）、顔検出部 109 が人物の顔を検出し、表情識別部 112 が検出顔の40

表情を識別する（ステップ S 210）。

【0057】

また、ステップ S 204で抽出したセグメントに二人以上的人物が写っている場合には（ステップ S 211の Yes）、関係性推定部 113は、表情識別部 112が識別した個々の人物の表情と、コンテキスト抽出部 106が抽出した前後のセグメントとのコンテキストに基づいて、人物間の関係性を推定する（ステップ S 212）。

【0058】

感動識別部 114は、音声認識部 107及び第1の感情分析部 110によるセグメント内の音声（セリフなど）の感情識別結果と、音楽検出 108及び第2の感情分析部 111によるセグメント内の背景音楽の感情やムードの識別結果と、顔検出部 109及び表情識別部 112によるセグメントに写っている人物の表情の識別結果と、関係性推定部 113によるセグメントに写っている人物間の関係性の推定結果に基づいて、セグメントの感動を識別する（ステップ S 213）。

10

【0059】

感動識別部 114は、セグメントの感動を推定処理して、セグメントの感動のレベルを示す感動識別ラベルを出力する。感動識別ラベルは、ポジティブとネガティブの2種類の簡単なバイナリラベルでもよいが、「怒り」、「嫌悪」、「恐れ」、「幸福」、「悲しみ」、「驚き」、…などさまざまな種類の感動を表すラベルであってもよい。この結果、ポジティブ、又は、「怒り」、「嫌悪」、「恐れ」、「幸福」、「悲しみ」、「驚き」、…などさまざまな種類の感動を表すラベルを持つサンプルを獲得することができる（ステップ S 214）。また、人間にポジティブな感情を与えないネガティブなサンプルも獲得する（ステップ S 215）。

20

【0060】

データ収集システム 100は、図 2 に示した処理手順に従って、テレビやインターネット放送などのコンテンツから、人間に感動を与えるようなセグメント（シーン）を抽出し、さらにセグメントに含まれる音声や背景音楽の感情分析結果、セグメントに写っている人物の表情識別結果、さらにセグメントに二人以上の人物が映っている場合の人物間の関係性の推定結果に基づいて、各セグメントの感動ラベルを識別することができる。そして、データ収集システム 100によって収集された感情ラベル付きのセグメントは、人工知能（ニューラルネットワークモデル）を用いた感動識別機を訓練するための学習データに使用することができる。

30

【0061】

### C. コンテンツに含まれる人物間などの関係性の推定

この項では、関係性推定部 113で実施される、映像中的人物間の関係性を推定する処理について、詳細に説明する。

【0062】

映像に写っている人物が一人の場合、その人物の検出顔の表情から感情を識別することができる。これに対し、映像に二人の人物が写っている場合には、人物間の関係性が生じるので、人物から感動を識別する際に人物間の関係性も考慮する必要がある。

【0063】

例えば、同じ人物 A が写っている映像であっても、人物 A と一緒に写っている人物 B が家族や友人など深い関係性があり親密度が高いと、その映像は感動に繋がり易い。一方、人物 A と人物 B の親密度が低いと、どの感動も与えられない場合もある。このように、被写体間の関係性を把握することは、映像が感動を与えるかを推定する上で非常に重要である。

40

【0064】

SNS などに投稿された情報に基づいて被写体の情報を機械学習する技術は既に知られている。しかしながら、被写体間の関係性などの人間関係を SNS の投稿情報から把握しようとすると、被写体のプライバシーを脅かすおそれがあるため実行が容易でない。これに対し、本開示では、関係性推定部 113は、表情識別部 112が識別した個々の人物の

50

表情と、コンテキスト抽出部 106 が抽出した前後のセグメントとのコンテキストに基づいて、人物間の関係性を推定するので、プライバシーの問題を生じない。

#### 【0065】

各人物の表情識別に基づいて人物間の関係性を推定する技術は既に知られているが、この種の技術の多くは、単純にポジティブな表情をポジティブな感情に繋ぐという 1 次元的なレベルにとどまる。しかしながら、表情から表面的に識別された感情ラベルが必ずしもその人物の感情と一致するとは限らない。例えば、友達と撮った写真で、わざと怒っているような表情を作る場合がある。また、重たい雰囲気の公式イベントで撮った写真では、笑顔を作っていても、明らかに親密な関係ではないことが分かる場合もある。すなわち、人物の表情を識別して単純に人物間の関係性を推定することはできない。写っている人物の顔の表情から単純に感動を与えるかどうかを識別することだけではなく、写っている 2 人以上の人物間の親密度を把握する必要がある。親密度が高い人物が写っている映像であれば、それが感動を与える可能性は必然的に高くなる。

#### 【0066】

視覚データに基づいて人物間の関係性を認識する研究は既に活発に行われている（例えば、非特許文献 3 を参照のこと）。本開示では、関係性推定部 113 は、セグメント抽出部 105 が抽出した当該セグメントの情報（例えば、セグメント内の映像から識別される人物の表情）だけでなく、その前後のセグメントとのコンテキストを利用して人物間の関係性を推定するので、推定の正確性が向上することを期待できる。さらに、インターネットなどの情報源から簡単に得られるコンテンツ自体のメタ情報（例えば、放送番組の情報）を、人物間の関係性の推定に利用することも考えられる。

#### 【0067】

図 3 には、関係性推定部 113 がセグメントのコンテキストと顔検出及び表情識別結果に基づいて人物間の関係性を推定する仕組みを図解している。図 3 に示す例では、関係性推定部 113 は、セグメント抽出部 105 が抽出した現在のセグメントに写っている人物の顔検出及び検出顔の表情識別結果とともに、過去のセグメントに写っている人物の顔検出及び検出顔の表情識別結果をコンテキストとして利用する。

#### 【0068】

まず、過去のセグメント 301 が、顔検出部 109 と表情識別部 112 を構成する CNN 311 に入力され、この CNN 311 が過去のセグメント 301 に写っている各人物の顔検出及び検出顔の表情識別を行う。そして、関係性推定部 113 は、過去のセグメント 301 から顔検出された人物とその表情識別結果 303 を人物登録部 313 に登録しておく。

#### 【0069】

次いで、セグメント抽出部 105 が抽出した現在のセグメント 302 が、顔検出部 109 と表情識別部 112 を構成する CNN 312 に入力され、この CNN が現在のセグメント 302 に写っている各人物の顔検出及び検出顔の表情識別を行い、顔検出及び検出顔の表情識別結果 304 を出力する。そして、関係性推定部 113 は、現在のセグメント 302 から検出された人物が、人物登録部 313 に問い合わせて、過去のセグメント 301 から事前登録されている人物であるかどうかをチェックする。

#### 【0070】

ここで、現在のセグメント 302 から検出された人物が、過去のセグメント 301 から事前に登録されている人物である場合には、関係性推定部 113 は、同じ人物についての、現在のセグメント 302 から推定された表情の識別結果に加えて、過去のセグメント 301 から推定された表情識別の結果を用いて、人物間の関係性を推定する。関係性推定部 113 からセグメントのコンテキストを利用して人物間の関係性をより正確に推定する、後段の感動識別部 114 は、現在のセグメントの感動ラベルをより正確に識別することができるようになる。

#### 【0071】

また、現在のセグメント 302 から検出された人物が、過去のセグメント 301 から事

10

20

30

40

50

前に登録されている人物でない場合には、関係性推定部 113 は、現在のセグメント 302 から推定された表情の識別結果のみに基づいて、人物間の関係性を推定する。あるいは、後段の感動識別部 113 は、人物間の関係性を利用しないで、現在のセグメントから推定された表情の識別結果に基づいて、現在のセグメントの感動ラベルを識別する。

#### 【0072】

##### D. 感動識別

この項では、感動識別部 114 で実施される、セグメントが与える感動を識別する処理について、詳細に説明する。

#### 【0073】

感動識別部 114 は、以下の多元的な情報をまとめて、現在のセグメントの感動ラベルを識別する。10

#### 【0074】

(1) 音声認識部 107 及び第 1 の感情分析部 110 によるセグメント内の音声（セリフなど）の感情識別結果

(2) 音楽検出 108 及び第 2 の感情分析部 111 によるセグメント内の背景音樂の感情やムードの識別結果

(3) 顔検出部 109 及び表情識別部 112 によるセグメントに写っている人物の表情の識別結果

(4) 関係性推定部 113 によるセグメントに写っている人物間の関係性や親密度の推定結果20

#### 【0075】

上記の(1)～(4)がすべてポジティブな感情を示す場合には、感動識別部 114 は、現在のセグメントに感動的な場面としてラベルを付ける。

#### 【0076】

また、上記の(1)～(4)のうち一部の情報が抜けている場合には（例えば、現在のセグメントに背景音樂が不在である場合や、人物が写っておらず、表情識別結果が不在である場合）、存在しているすべての情報がポジティブであれば、感動識別部 114 は、現在のセグメントに感動的な場面としてラベルを付けるようにしてもよい。一部の情報が抜けていると、感動識別の信頼性が低下する可能性はあるが、セグメント自体が既にセグメント抽出部 105 においてコンテンツ評価情報と脳波情報の一一致というフィルターを通過しているので、信頼性が急激に低下することはないと想料される。30

#### 【0077】

また、上記(1)～(4)のいずれの情報からも現在のセグメントに対して感動的なラベルを付けられない場合には、感動識別部 114 は、そのセグメントにネガティブのラベルを付けるようにしてもよい。ネガティブのラベルが付けられたセグメントは、ネガティブサンプルとして感動識別機を訓練するための学習データに使用することができる。

#### 【0078】

上述したように、セグメントに含まれる音声、音樂、人物の顔の表情を識別する各識別機は、それぞれ CNN などの学習済み機械学習モデルで構成される。各識別機がどのように事前学習したかによって、感動識別部 114 は、ポジティブとネガティブの 2 種類の簡単なバイナリラベルでなく、「怒り」、「嫌悪」、「恐れ」、「幸福」、「悲しみ」、「驚き」、…などさまざまな種類に分化した感動を表すラベルを付けることができる。すなわち、データ収集システム 100 は、さまざまな感動識別ラベルを持つ学習データを収集することができ、さまざまな種類の感動を識別する感動識別機の訓練に使用することができる。40

#### 【0079】

##### E. 収集したデータを用いた学習プロセス

この項では、本開示に係るデータ収集システム 100 を用いて収集した学習データを利用して、感動識別機を訓練する学習プロセスについて、詳細に説明する。感動識別機は、コンテンツがどのような感動を与えるかを識別する装置である。ここで言う装置は、専用

のハードウェアで構成される装置と、ソフトウェアを実行して所定の機能を実現する装置の両方を意味する。

#### 【0080】

図4には、収集した学習データを利用した感動識別機の学習プロセスの全体の流れを模式的に示している。

#### 【0081】

データ収集システム100を用いて収集した学習データは、データ蓄積部410に格納されている。個々の学習データは、上記A～D項で「セグメント」と呼ばれるものに相当する。学習データとしてのセグメントは、テレビやインターネット放送などのコンテンツの要素であり、フレームシーケンスからなる映像データと、映像に同期したオーディオデータで構成される。また、学習データとしてのセグメントにはデータ収集システム100における処理過程で最終的に感動識別ラベルが教師データとして付与されている。感動識別ラベルは、さまざまな種類に分化される場合もあるが、この項では説明の簡素化のため、ポジティブとネガティブの2種類の簡単なバイナリラベルが付与されているものとする。

10

#### 【0082】

感動識別機420は、学習データとしての映像とオーディオデータを入力する複数のネットワークモデルからなるネットワーク部421と、ネットワーク部421で獲得した特徴量に基づいてコンテンツの感動ラベルを識別する識別部422で構成される。ネットワーク部421内の各ネットワークは、パラメータ（ノード間の結合重み係数など）を持つ。

20

#### 【0083】

映像の場合は、フレーム毎にネットワーク部421内の学習済みのニューラルネットワークモデル（CNN）を通して各々の特徴量を獲得して、それらをまとめることによって映像全体の特徴量（Video Feature）が得られる。フレームの特徴量のまとめ方としてはさまざまな方法があるが、単純なconcatenate、又はLSTM（Long Short Term Memory）などの時系列ネットワークに順番に入れる方法などが挙げられる。本実施形態では、フレームに写っている人物が二人以上の場合において、連続する複数のフレームのコンテキストに基づいて推定される人物間の関係性を映像特徴量として獲得する。

#### 【0084】

一方、オーディオデータの場合は、メル周波数ケプストラム係数（Mel-Frequency Cepstrum Coefficients: MFCC）やメルスペクトログラム（mel-spectrogram）などの特徴量に変換し、それをネットワーク部421内のLSTMなどの時系列ネットワークに入れることによってオーディオ特徴量（Audio Feature）を獲得することができる。また、音声の特徴量をテキストに書き換えて（transcription）、同じくネットワーク部421内のLSTMなどの時系列ネットワークに入れることによってテキスト特徴量（Text Feature）を獲得することができる。

30

#### 【0085】

識別部422は、上記のようにして得られた映像全体の特徴量（Video Feature）、オーディオ特徴量（Audio Feature）、及びテキスト特徴量（Text Feature）を共通空間に射影して、感動識別機420に入力された学習データに対して感動識別ラベルを付与する。ここでは、識別部422は、ポジティブ又はネガティブのいずれかのバイナリラベルを付与するものとする。

40

#### 【0086】

評価部423は、識別部422が学習データに付与した感動識別ラベルと、データ収集システム100によって学習データに付与された教師データとしての感動識別ラベルとの誤差に基づくソフトマックスなどの損失関数Lossを計算する。あるいは、損失関数Lossは、映像の特徴量に基づく感動識別ラベルと、音声の特徴量に基づく感動識別ラベルと、オーディオの特徴量に基づく感動識別ラベルを個別に識別して、各々の誤差L<sub>video</sub>、L<sub>text</sub>、L<sub>audio</sub>の合計に基づいて損失関数Lossを求めるようにしてもよい。そして

50

、誤差に基づいて求めた損失関数 Loss は、ネットワーク部 421 内のパラメータを持つ各ネットワークに逆伝播 (Backpropagate) され、損失関数 Loss が最小となるようにネットワーク部 421 内の各ネットワーク (CNN、LSTM) のパラメータが更新される。これによって、感動識別機 420 は、入力された学習データに対して教師データと等しくなるような感動識別ラベルを出力するように、学習が進む。

#### 【0087】

##### F. 推論

上記 A 項では、データ収集システム 100 が、テレビやインターネット放送などのコンテンツからコンテンツ評価情報と脳波などの生体情報をを利用してデータを収集する方法について説明した。また、上記 E 項では、データ収集システム 100 によって収集されたデータを用いて、感動識別機 420 の学習を行うプロセスについて説明した。

10

#### 【0088】

学習プロセスを経て学習された感動識別機 420 は、感動を与えるコンテンツ、あるいはコンテンツがどのような感動を与えるかを識別することができる。ここで言うコンテンツは、テレビやインターネット放送などのコンテンツを始め、動画共有サイトで共有される動画コンテンツ、ユーザがデジタルカメラで撮影した動画及び静止画のコンテンツなど、さまざまなコンテンツを含み、コンテンツ評価情報や脳波などの生体情報が付与されている必要はない。

#### 【0089】

##### G. アプリケーション

これまで説明してきたように、本開示に係るデータ収集システムが収集した学習データを使って、コンテンツが人間に与える感情を識別する感情識別機用の機械学習モデルを訓練することができる。また、このようにして開発した感情識別機にはさまざまなアプリケーションが期待される。

20

#### 【0090】

##### G-1. デジタルカメラの構成

本開示に基づいて開発された感情識別機は、例えばデジタルカメラを始めとする、コンテンツの記録、再生、編集などコンテンツに対して処理を行うさまざまなコンテンツ処理装置に搭載することができる。

#### 【0091】

30

図 5 には、デジタルカメラ 500 の構成例を示している。図示のデジタルカメラ 500 は、光学系 501 と、撮像部 502 と、AFE (Analog Front End : アナログ前処理) 部 503 と、カメラ信号処理部 504 と、コーデック 505 と、メイン処理部 506 と、マイク 514 と、A/D 変換部 515 と、表示部 516 と、音再生部 517 と、記録部 518 を備えている。このデジタルカメラ 500 には、本開示に基づいて開発された感情識別機が搭載されることを想定している。

#### 【0092】

光学系 501 は、被写体からの光を撮像部 502 の撮像面に集光するためのレンズ、レンズを移動させてフォーカス合わせやズーミングを行なうための駆動機構、開閉操作により被写体からの光を所定時間だけ撮像面に入射させるシャッター機構、被写体からの光線束の方向並びに範囲を限定するアイリス (絞り) 機構 (いずれも図示しない) を含んでいる。ドライバ (図示しない) は、後述するメイン処理部 506 からの制御信号に基づいて、光学系 501 内の各機構の駆動 (例えば、被写体のフォーカス合わせやアイリス、パン及びチルト、シャッター又はセルフ・タイマー設定のタイミング) などのカメラワークを制御する。

40

#### 【0093】

撮像部 502 は、CCD (Charge Coupled Device) や CMOS (Complementary Metal Oxide Semiconductor) などの撮像素子からなり、光電変換効果を持つ各画素が 2 次元に配列された撮像面を有し、被写体からの入射光を電気信号に変換する。受光側には、例えば G 市松 R/B 色コーディン

50

グ单板が配設されている。各色フィルターを通した入射光量に対応する信号電荷が各画素に蓄積され、各画素から読み出される3色の各信号電荷量からその画素位置における入射光の色を再現することができる。なお、撮像部502から出力されるアナログ画像信号は、R G B各色の原色信号であるが、補色系の色信号であってもよい。

#### 【0094】

A F E 部 5 0 3 は、撮像信号の低雑音を高精度に抑圧した後 (Correlated Double Sampling (相関二重サンプリング) : C D S ) 、サンプル・ホールドを行ない、さらに A G C ( A u t o m a t i c G a i n C o n t r o l : 自動利得制御回路 ) により適正なゲイン・コントロールをかけて、A D 変換を施してデジタル画像信号を出力する。また、A F E 部 5 0 3 は、撮像素子を駆動するためのタイミングパルス信号と、このタイミングパルス信号に従って撮像素子の各画素の電荷を垂直方向にライン単位で出力するための駆動信号を撮像部 5 0 2 に出力する。10

#### 【0095】

カメラ信号処理部 5 0 4 は、A F E 部 5 0 3 から送られてくる画像信号に対して、欠陥画素補正やデジタルクランプ、デジタルゲイン制御などの前処理を施した後、A W B によりホワイトバランスゲインを掛けるとともに、シャープネス・再度コントラスト調整などの画質補正処理を施して適正な色状態を再現し、さらにデモザイク処理により R G B 画面信号を作成する。また、カメラ信号処理部 5 0 4 は、撮影画像を表示部 5 1 6 にスルー画像として表示出力するか又は記録部 5 1 8 に保存するかに応じて解像度変換を行ったり、M P E G ( M o v i n g P i c t u r e E x p e r t s G r o u p ) などのコーデック処理を行ったりする。20

#### 【0096】

メイン処理部 5 0 6 は、プロセッサと R A M ( R a n d o m A c c e s s M e m o r y ) 及び R O M ( R e a d O n l y M e m o r y ) で構成され、デジタルカメラ 5 0 0 全体の動作を統括的に制御する。プロセッサは、C P U ( C e n t r a l P r o c e s s i n g U n i t ) やマルチコアを持つ G P U ( G r a p h i c P r o c e s s i n g U n i t ) などである。メイン処理部 5 0 6 は、記録時には、撮影部 5 0 2 で撮像した映像データとマイク 5 1 4 で収音したオーディオデータを記録部 5 1 8 に保存する。また、メイン処理部 5 0 6 は、再生時には、記録部 5 1 8 から映像及びオーディオデータを読み出して、表示部 5 1 6 及び音再生部 5 1 7 で出力する。また、本実施形態では、メイン処理部 5 0 6 は、本開示に基づいて開発された感情識別機が搭載されることを想定している。30

#### 【0097】

表示部 5 1 6 は、デジタルカメラ 5 0 0 に搭載された液晶表示パネル、外部のテレビやプロジェクタなど、撮影中の映像又は記録した映像を表示する装置である。

#### 【0098】

音再生部 5 1 7 は、デジタルカメラ 5 1 0 0 に搭載されたスピーカ、外部のスピーカなど、記録した音声を再生する装置である。

#### 【0099】

記録部 5 1 8 は、H D D ( h a r d D i s c D r i v e ) や S S D ( S o l i d S t a t e D r i v e ) などの大容量の記録装置である。記録部 5 1 8 には、撮像部 5 0 2 で撮影した映像及び映像と同期してマイク 5 1 4 で収音したオーディオデータからなるコンテンツが記録される。また、感情識別機用の機械学習モデルのパラメータ（例えば、ニューラルネットワークモデルにおけるニューロン間の結合重み係数）が記録部 5 1 8 に記録されている。40

#### 【0100】

### G - 2 . 撮影したコンテンツのラベリング

続いて、本開示に基づいて開発された感情識別機をデジタルカメラ 5 0 0 に搭載して実現される機能について説明する。

#### 【0101】

デジタルカメラ 5 0 0 で撮影したコンテンツを感動識別機にかけることによって、コン50

テンツの感動識別ラベルを付与したり、コンテンツ内で感動を与えるシーン（セグメント）を抽出したりすることができる。

#### 【0102】

図6には、デジタルカメラ500で撮影したコンテンツの感動識別ラベリングを行うための機能的構成を示している。同図中、感動識別機420の構成は図4に示したものと同じとする。感動識別機420は、デジタルカメラ500内のメイン処理部506で動作することを想定している。

#### 【0103】

デジタルカメラ500で撮影したコンテンツの映像は、フレーム毎にネットワーク部421内の学習済みのニューラルネットワークモデル（CNN）を通して各々の特徴量を獲得して、それらをまとめることによって映像全体の特徴量が得られる。また、映像と同期して記録されたオーディオデータは、MFCCなどの特徴量に変換し、ネットワーク部421内のLSTMなどの時系列ネットワークに入れることによって、オーディオ特徴量が得られる。また、音声の特徴量をテキストに書き換えて、同じくネットワーク部421内のLSTMなどの時系列ネットワークに入れることによってテキスト特徴量が得られる。

10

#### 【0104】

識別部422は、上記のようにして得られた映像全体の特徴量、オーディオ特徴量、及びテキスト特徴量を共通空間に射影して、感動識別機420に入力された学習データに対して感動識別ラベルを付与する。識別部422は、ポジティブ又はネガティブのいずれかのバイナリラベルを付与する。あるいは、識別部422は、さまざまな種類に分化した感動を表すラベルを付与してもよい。

20

#### 【0105】

いずれにせよ、識別部422は、撮影したコンテンツに対して、シーン（又はセグメント）毎に感動識別ラベルを付与する。付与された感動識別ラベルは、例えばコンテンツのメタ情報として記録される。録画したコンテンツを再生するときには、感動識別ラベルを検索キーに用いて、感動的なシーンを選択して視聴することができる。また、録画したコンテンツを編集するときには、感動識別ラベルを検索キーにして、「うれしい」シーンや「悲しい」シーンなど、特定の感情を抱くシーンだけを取捨選択することができる。

#### 【0106】

このように、本開示に係るデータ収集システム100は、入力された各コンテンツ又はコンテンツ中のセグメント毎に、人間に与えることができる感動の種類を識別する感動識別ラベルを付与することができる。したがって、データ収集システム100を用いて、膨大な映像コンテンツ101の中から、特定の感動識別ラベル（例えば、「うれしい」、「悲しい」など）が付与されたコンテンツ又はセグメントを収集して、その特定の感動を与えるコンテンツを生成し又は感動するコンテンツを撮影するためのカメラ操作を推定する機械学習モデルの学習に用いることができる。

30

#### 【0107】

##### G - 3 . カメラワークの支援又は制御

デジタルカメラ500で撮影したコンテンツを感動識別機にかけて得られた結果に基づいて、カメラワークの支援や自動制御を行うことができる。例えば、感動の度合いが増すように、あるいは特定の種類の感動識別ラベルを獲得できるように、被写体を撮影する視線方向やズームを自動制御したり、又は撮影画像の輝度、色彩、画角、構図、フォーカスなどを自動制御したり、表示部516のガイダンス表示や音再生部517からの音声ガイダンスを使っておすすめのアングルを教えたりするようにしてもよい。

40

#### 【0108】

図7には、撮影したコンテンツの感動識別結果に基づいてカメラワークを自動制御するためのデジタルカメラ500の機能的構成を示している。同図中、感動識別機420の構成は図4に示したものと同じとする。感動識別機420は、デジタルカメラ500内のメイン処理部506で動作することを想定している。

#### 【0109】

50

デジタルカメラ 500 で撮影したコンテンツの映像は、フレーム毎にネットワーク部 421 内の学習済みのニューラルネットワークモデル（CNN）を通して各々の特徴量を獲得して、それらをまとめることによって映像全体の特徴量が得られる。また、映像と同期して記録されたオーディオデータは、MFCCなどの特徴量に変換し、ネットワーク部 421 内の LSTM などの時系列ネットワークに入れることによって、オーディオ特徴量が得られる。また、音声の特徴量をテキストに書き換えて、同じくネットワーク部 421 内の LSTM などの時系列ネットワークに入れることによってテキスト特徴量が得られる。

#### 【0110】

識別部 422 は、上記のようにして得られた映像全体の特徴量、オーディオ特徴量、及びテキスト特徴量を共通空間に射影して、感動識別機 420 に入力された学習データに対して感動識別ラベルを付与する。そして、評価部 423 は、感動の度合いが増すように、あるいは特定の種類の感動識別ラベルを獲得できるように、光学系 501 に対してカメラワークの制御信号を出力する。

#### 【0111】

##### G - 4 . キャプションの追加

デジタルカメラ 500 で撮影したコンテンツを感動識別機にかけて得られた結果に基づいて、映像のシーンに対して適切で、且つ感動の度合いが増すようなキャプションを自動付与することができる。

#### 【0112】

図 8 には、撮影したコンテンツの感動識別結果に基づいてキャプションを自動生成するためのデジタルカメラ 500 の機能的構成を示している。同図中、感動識別機 420 の構成は図 4 に示したものと同じとする。感動識別機 420 は、デジタルカメラ 500 内のメイン処理部 506 で動作することを想定している。

#### 【0113】

デジタルカメラ 500 で撮影したコンテンツの映像は、フレーム毎にネットワーク部 421 内の学習済みのニューラルネットワークモデル（CNN）を通して各々の特徴量を獲得して、それらをまとめることによって映像全体の特徴量が得られる。また、映像と同期して記録されたオーディオデータは、MFCCなどの特徴量に変換し、ネットワーク部 421 内の LSTM などの時系列ネットワークに入れることによって、オーディオ特徴量が得られる。また、音声の特徴量をテキストに書き換えて、同じくネットワーク部 421 内の LSTM などの時系列ネットワークに入れることによってテキスト特徴量が得られる。また、キャプション生成部 801 は、例えば学習済みの機械学習済みモデルを用いて、シーン毎のキャプションを生成する。キャプション生成部 801 が生成したキャプションのテキスト情報は、コンテンツの音声認識から得られたテキスト情報に重畠されて、ネットワーク部 421 内の LSTM などの時系列ネットワークに入れられる。

#### 【0114】

識別部 422 は、上記のようにして得られた映像全体の特徴量、オーディオ特徴量、及びテキスト特徴量を共通空間に射影して、感動識別機 420 に入力された学習データに対して感動識別ラベルを付与する。評価部 423 は、識別部 422 から出力される感動識別ラベルについての損失関数を計算する。そして、キャプション生成部 801 は、感動の度合いが増すように、あるいは特定の種類の感動識別ラベルを獲得できるように再学習する。キャプション生成部 801 が生成したキャプションは、コンテンツの音声データに重畠して記録される。

#### 【0115】

##### G - 5 . 背景音楽の付与

デジタルカメラ 500 で撮影したコンテンツを感動識別機にかけて得られた結果に基づいて、映像のシーンに対して適切で、且つ感動の度合いが増すような背景音楽を自動付与することができる。

#### 【0116】

図 9 には、撮影したコンテンツの感動識別結果に基づいて背景音楽を自動付与するため

10

20

30

40

50

のデジタルカメラ 500 の機能的構成を示している。同図中、感動識別機 420 の構成は図 4 に示したものと同じとする。感動識別機 420 は、デジタルカメラ 500 内のメイン処理部 506 で動作することを想定している。

#### 【0117】

デジタルカメラ 500 で撮影したコンテンツの映像は、フレーム毎にネットワーク部 421 内の学習済みのニューラルネットワークモデル（CNN）を通して各々の特徴量を獲得して、それらをまとめることによって映像全体の特徴量が得られる。また、映像と同期して記録されたオーディオデータは、MFCCなどの特徴量に変換し、ネットワーク部 421 内の LSTM などの時系列ネットワークに入れることによって、オーディオ特徴量が得られる。また、音声の特徴量をテキストに書き換えて、同じくネットワーク部 421 内の LSTM などの時系列ネットワークに入れることによってテキスト特徴量が得られる。また、音楽検索部 901 は、例えば学習済みの機械学習済みモデルを用いて、シーンの背景音楽となる楽曲を検索する。音楽検索部 901 は、例えば膨大量の音楽データを蓄積している音楽データベース（図示しない）上で、シーンの背景音楽となる楽曲を検索するようにもよい。音楽検索部 901 が見つけ出した楽曲の音楽データは、コンテンツのオーディオデータに重畠されて、ネットワーク部 421 内の LSTM などの時系列ネットワークに入れられる。

10

#### 【0118】

識別部 422 は、上記のようにして得られた映像全体の特徴量、オーディオ特徴量、及びテキスト特徴量を共通空間に射影して、感動識別機 420 に入力された学習データに対して感動識別ラベルを付与する。評価部 423 は、識別部 422 から出力される感動識別ラベルについての損失関数を計算する。そして、音楽検索部 901 は、感動の度合いが増す、あるいは特定の種類の感動識別ラベルを獲得できる背景音楽を見つけ出すように再学習する。音楽検索部 901 が見つけた楽曲の音楽データは、コンテンツのオーディオデータに重畠して記録される。

20

#### 【0119】

##### H . モデルの効率的学習

上記 G 項で説明したように、本開示によれば、デジタルカメラ 500 にニューラルネットワークモデルを搭載することにより、コンテンツ（又は、コンテンツを構成するセグメント毎）に感動のレベルを表す感動識別ラベルを付与したり、高レベルの感動が得られるコンテンツを撮影するためのカメラワークの支援又は制御（カメラの自動操作）を提供したり、コンテンツの感動スコアを向上させるためのキャプションや背景音楽を付与したりすることができる。この H 項では、デジタルカメラ 500 に搭載されたニューラルネットワークモデルを効率的に学習するための方法について説明する。

30

#### 【0120】

なお、便宜上、カメラの自動操作のためのニューラルネットワークモデルの学習方法に限定して説明するが、キャプションの追加や背景音楽の付与を行うニューラルネットワークモデルに関しても同様の方法により効率的に学習を行うことができるこを理解されたい。

40

#### 【0121】

##### H - 1 . 学習システムの構成

図 10 には、カメラの操作支援や自動操作のためのニューラルネットワークモデルを効率的に学習するための学習システム 1000 の構成を模式的に示している。本実施形態では、学習の対象とするニューラルネットワークモデルとして、主に、感動するコンテンツを生成し又は感動するコンテンツを撮影するためのカメラ操作を推定するニューラルネットワークモデルを想定している。もちろん、学習システム 1000 を使って、その他のタイプのニューラルネットワークモデルの学習にも活用することができる。

#### 【0122】

図 10 に示す学習システム 1000 は、学習データを収集するデータ収集装置 1010 と、データ収集装置 1010 が収集した学習データに基づいてニューラルネットワークモ

50

デルの学習を行う学習装置 1030 と、学習データに関する分析を行うデータ分析装置 1040 と、学習装置 1030 によって学習されたニューラルネットワークモデル 1050 を使用するモデル利用装置 1020 で構成される。

#### 【0123】

データ収集装置 1010 は、例えばカメラで撮影した画像データと、カメラ撮影時の操作情報などからなるデータを収集する。データ収集装置 1010 は、スチルカメラやビデオカメラ、カメラで使用されるイメージセンサー、スマートフォンなどの多機能情報端末、T V、ヘッドホン又はイヤホン、ゲーム機、冷蔵庫や洗濯機などの I o T デバイス、ドローンやロボットなどの移動体装置など、多種類の膨大な装置を含み、これら膨大数のデータ収集装置 1010 から膨大量のデータを収集することができる。例えば、データ収集装置 1010 が、プロのカメラマンなどのエキスパートが使用するカメラであれば、感動を与える撮影画像と、そのような画像を撮影するためのカメラ操作情報を収集することができる。また、データ収集装置 1010 は、上記で例示したようなリアルタイムでデータを収集する装置だけでなく、コンテンツデータベースのように既に大量のデータを蓄積している装置を含んでいてもよい。

10

#### 【0124】

学習装置 1030 は、各データ収集装置 1010 に対して収集したデータの送信を要求する要求信号を送信する。もちろん、データ収集装置 1010 は、要求信号に応じてではなく、自発的にデータを送信するようにしてもよい。そして、学習装置 1030 は、多数のデータ収集装置 1010 によって収集された膨大なデータを使って、「観測予測モデル」、「操作モデル」、「操作推定モデル」、「感動スコア推定モデル」などのさまざまなニューラルネットワークモデルの学習及び再学習を行う。学習システム 1000 において利用されるニューラルネットワークモデルの詳細については、後述に譲る。

20

#### 【0125】

膨大量のデータを使用することでニューラルネットワークモデルの学習精度は向上するが、学習にあまり寄与しないデータを使って学習や再学習を行うのは非効率的である。そこで、本開示に係る学習システム 1000 では、データ分析装置 1040 は、学習の対象とするニューラルネットワークモデルの学習に影響を与える学習データの分析を行い、その分析結果に基づいて、ニューラルネットワークモデルの学習に所定以上の影響を及ぼす学習データや不足している学習データ、これらに類似するデータなど、有意義な学習データを抽出し、学習装置 1030 は、データ分析部 1040 によって抽出された有意義な学習データを使って、ニューラルネットワークモデルの学習及び再学習を効率的に行うよう正在実施

30

#### 【0126】

データ分析装置 1040 は、例えば、 X A I ( e x p l a i n a b l e A I ) 、学習データの信頼度 ( C o n f i d e n c e ) スコア計算、影響関数 ( I n f l u e n c e F u n c t i o n ) 計算、ベイジアン D N N ( D e e p N e u r a l N e t w o r k ) によるデータ不足推定などの手法に基づいて、データ収集装置 1010 が収集したデータの分析を行うことができる。

40

#### 【0127】

モデル利用装置 1020 は、学習装置 1030 によって学習されたニューラルネットワークモデル 1050 を使用して、感動するコンテンツの撮影を行う装置である。モデル利用装置 1020 は、例えば、カメラ操作に慣れていない一般的のユーザなどが使用するカメラである。モデル利用装置 1020 は、学習装置 1030 による学習済みのニューラルネットワークモデルを使用して、プロのカメラマンなど撮影のエキスパートに匹敵するカメラの自動操作を行ったり、感動スコアの高い映像の自動撮影を行ったりすることができる。

#### 【0128】

ここで、学習システム 1000 の変形例について説明しておく。

#### 【0129】

学習装置 1030 は、さまざまなニューラルネットワークモデルの学習を行うために膨

50

大な計算リソースが必要である。したがって、図10では、学習装置1030は、例えばクラウド上に構築されること（すなわち、クラウドAI（Artificial Intelligence））を想定している。また、学習装置1030は、複数の計算ノードを用いて分散学習を行うようにしてもよい。但し、学習装置1030は、学習済みニューラルネットワークモデルを利用するモデル利用装置と一体で構成されること（すなわち、エッジAI）も想定される。あるいは、学習装置1030は、学習データを提供するデータ収集装置1010と一体で構成されていてもよい。

#### 【0130】

また、データ分析装置1040は、クラウド又はエッジのいずれに構築されていてもよい。例えばデータ分析装置1040は、学習装置1030と一体の装置として構成されていてもよい。この場合、学習装置1030は、内部でニューラルネットワークモデルの学習に影響を与える学習データを分析して、ニューラルネットワークモデルの学習に所定以上の影響を及ぼす学習データ、不足している学習データ、又はこれらに類似するデータの送信を、データ収集装置1010に対して要求するようにしてもよい。10

#### 【0131】

あるいは、データ分析装置1040は、データ収集装置1010と一体の装置として構成されていてもよい。この場合、学習装置1030は、学習データの送信要求時などに、データ分析に必要な情報（例えば、その時点で学習済みのニューラルネットワークモデルの情報）を、データ収集装置1010に提供する。そして、データ収集装置1010は、収集したデータがニューラルネットワークモデルに与える影響を分析して、収集したデータのうちニューラルネットワークモデルの学習に所定以上の影響を及ぼす学習データ、不足している学習データ、又はこれらに類似するデータを学習装置1030に送信するようにしてもよい。20

#### 【0132】

また、図10では、学習データを収集するデータ収集装置1010と、ニューラルネットワークモデルを利用するモデル利用装置1020を別々の装置として描いているが、1つの装置がデータ収集装置1010及びモデル利用装置1020の双方として動作することも想定される。例えば1台のカメラが、マニュアル操作モードではデータ収集装置1010として動作して、撮影データとカメラ操作情報などのデータを収集して学習装置1030に送信する一方、自動操作モードに切り替えるとモデル利用装置1020として動作して、学習装置1030で学習済みのニューラルネットワークモデルを用いて自動撮影を行うようにしてもよい。30

#### 【0133】

##### H-2. ニューラルネットワークモデルの構成

図10に示した学習システム1000において利用されるニューラルネットワークモデルについて説明しておく。学習システム1000では、学習装置1030は、データ収集装置1040によって収集されたデータを使って、「観測予測モデル」、「操作モデル」、「操作推定モデル」、「感動スコア推定モデル」などのニューラルネットワークの学習及び再学習を行う。また、モデル利用装置1020は、学習装置1030によって学習が行われたこれらのニューラルネットワークモデルのうち少なくとも一部を利用する。40

#### 【0134】

##### 観測予測モデル：

図11には、観測予測モデル1100の構成を模式的に示している。観測予測モデル1100は、カメラで撮影した現時刻までの映像情報1101と、カメラに対する現時刻までの操作情報1102から、次の時刻にカメラで撮影される画像（すなわち、「次の時刻の画像」）1111を予測するニューラルネットワークモデルである。

#### 【0135】

ここで言う操作情報1102は、例えば、フレームレートや絞り、露出値、倍率、焦点などの撮像条件を決めるためにカメラに対して行われた操作に関する情報である（以下、同様）。また、ロボットやドローンなどの移動体装置にカメラを搭載して行う場合には、50

移動体装置に対して行われたりモコン操作（ロール、ピッチ、ヨーで示されるカメラワークなど）も操作情報に含んでもよい（以下、同様）。

#### 【0136】

また、観測予測モデル1100は、予測した次の時刻の画像1111に対する信頼度スコア1112も併せて出力する。信頼度スコア1112は、次の時刻の画像1111がどの程度正しく予測できているかを示す値である。本開示では、学習に不足しているデータ、又は学習への影響度の高いデータを特定するために、信頼度スコアが用いられる。あるいは、データ分析装置1040が、XAIによる観測予測モデル1100の予測の根拠の説明、信頼度スコア計算、影響関数計算、ベイジアンDNNによるデータ不足推定などの手法に基づいて、観測予測モデル1100の学習に所定以上の影響を及ぼす学習データや不足している学習データ、これらに類似するデータなど、有意義な学習データを抽出する。観測予測モデル1100による信頼度スコア1112の計算機能は、データ分析装置1040の一部として実装されていてもよい。

10

#### 【0137】

、  
学習システム1000では、学習装置1030が、データ収集装置1010から送信された映像情報と操作情報からなるデータセットに基づいて、観測予測モデル1100の学習を行うことができる。学習装置1030は、強化学習により、より感動を与えることができる画像を予測できるように観測予測モデル1100の学習を行うことができる。また、データ分析装置1040は、データ収集装置1010から送信されたデータセットの中から有意義な学習データを抽出し、学習装置1030はこの有意義な学習データを用いて観測予測モデル1100の再学習を効率的に行うことができる。

20

#### 【0138】

##### 操作モデル：

図12には、操作モデル1200の構成を模式的に示している。操作モデル1200は、カメラで撮影した現時刻までの映像情報1201と、カメラに対する現時刻までの操作情報1202から、次の時刻にカメラに対して行われる操作1211を予測するニューラルネットワークモデルである。また、操作モデル1200は、予測した次の時刻の操作1211に対する信頼度スコア1212も併せて出力する。信頼度スコア1212は、次の時刻の画像1111がどの程度正しく予測できているかを示す値である。あるいは、データ分析装置1040が、XAIによる操作モデル1200の予測の根拠の説明、信頼度スコア計算、影響関数計算、ベイジアンDNNによるデータ不足推定などの手法に基づいて、操作モデル1200の学習に所定以上の影響を及ぼす学習データや不足している学習データ、これらに類似するデータなど、有意義な学習データを抽出する。操作モデル1200による信頼度スコア1212の計算機能は、データ分析装置1040の一部として実装されていてもよい。

30

#### 【0139】

学習システム1000では、学習装置1030が、データ収集装置1010から送信された映像情報と操作情報からなるデータセットに基づいて、操作モデル1200の学習を行うことができる。学習装置1030は、強化学習により、より感動を与えることができる画像を撮影できるカメラ操作情報を予測できるように操作モデル1200の学習を行うことができる。また、データ分析装置1040は、データ収集装置1010から送信されたデータセットの中から有意義な学習データを抽出し、学習装置1030はこの有意義な学習データを用いて操作モデル1200の再学習を効率的に行うことができる。

40

#### 【0140】

##### 操作推定モデル：

図13には、操作推定モデル1300の構成を模式的に示している。操作推定モデル1300は、現時刻までの映像時系列情報1301をカメラで撮影するための時系列操作情報1311を推定するニューラルネットワークモデルである。例えば、操作推定モデル1300を使って、プロのカメラマンなどカメラ操作に精通したエキスパートが撮影した高

50

品質な映像時系列情報から、エキスパートが行うカメラの時系列操作情報を推定することができる。

#### 【0141】

操作推定モデル1300は、推定した時系列操作情報1311に対する信頼度スコアも併せて出力するようにしてもよい。あるいは、データ分析装置1040が、XAIによる操作推定モデル1300の予測の根拠の説明、信頼度スコア計算、影響関数計算、ベイジアンDNNによるデータ不足推定などの手法に基づいて、操作推定モデル1300の学習に所定以上の影響を及ぼす学習データや不足している学習データ、これらに類似するデータなど、有意義な学習データを抽出する。操作推定モデル1300による信頼度スコアの計算機能は、データ分析装置1040の一部として実装されていてもよい。

10

#### 【0142】

学習システム1000では、学習装置1030が、データ収集装置1010から送信された映像情報と操作情報からなるデータセットに基づいて、操作推定モデル1300の学習を行うことができる。学習装置1030は、強化学習により、より感動を与えることができる画像を撮影できる時系列操作情報を予測できるように操作推定モデル1300の学習を行うことができる。また、データ分析装置1040は、データ収集装置1010から送信されたデータセットの中から有意義な学習データを抽出し、学習装置1030はこの有意義な学習データを用いて操作推定モデル1300の再学習を効率的に行うことができる。

20

#### 【0143】

##### 感動スコア推定モデル：

図14には、感動スコア推定モデル1400の構成を模式的に示している。感動スコア推定モデル1400は、映像情報1401の感動スコア1411を推定するニューラルネットワークモデルであり、上記の感情識別機に相当する。例えば図4に示した学習プロセスに従って、感動スコア推定モデル1400の学習を行うことができる。

#### 【0144】

図15には、制御対象（フレームレート、解像度など）と平均感動スコアの関係を示している。但し、ニューラルネットワークモデルから出力される予測値を実線で示し、分散を点で示している。図15は、あるフレームレートで取得した映像データに基づいて感動スコアが高くなるように操作モデルを強化学習により学習し、学習した操作モデルを用いてカメラを操作したときに得られた平均感動スコアを示している。同図中、黒丸で示すデータ点は、既にデータのある点である。データのある点は分散が小さくなる。データのない点は分散が大きい。分散が大きく、且つ高いスコアが期待できる点ほど観測する価値が大きい。例えば楽観的に分散を考慮した上で最も高いスコアが期待できる点のデータを取得するという実装が考えられる。

30

#### 【0145】

感動スコア推定モデル1400は、推定した感動スコア1411に対する信頼度スコアも併せて出力するようにしてもよい。あるいは、データ分析装置1040が、XAIによる感動スコア推定モデル1400の予測の根拠の説明、信頼度スコア計算、影響関数計算、ベイジアンDNNによるデータ不足推定などの手法に基づいて、感動スコア推定モデル1400の学習に所定以上の影響を及ぼす学習データや不足している学習データ、これらに類似するデータなど、有意義な学習データを抽出する。感動スコア推定モデル1400による信頼度スコアの計算機能は、データ分析装置1040の一部として実装されていてもよい。

40

#### 【0146】

学習システム1000では、学習装置1030が、データ収集装置1010から送信された映像情報と操作情報からなるデータセットに基づいて、感動スコア推定モデル1400の学習を行うことができる。学習装置1030は、強化学習により、感動を与えることができる映像情報に対してより高い感動スコアを推定できるように感動スコア推定モデル1400の学習を行うことができる。また、データ分析装置1040は、データ収集装置

50

1010から送信されたデータセットの中から有意義な学習データを抽出し、学習装置1030はこの有意義な学習データを用いて感動スコア推定モデル1400の再学習を効率的に行うことができる。

#### 【0147】

##### H-3. 学習データの分析について

学習システム1000では、データ分析装置1040が、学習対象となるニューラルネットワークモデルの学習データを分析して、学習に所定以上の影響を及ぼす学習データや不足している学習データ、これらに類似するデータなど、有意義な学習データを特定して、学習装置1020はそのような有意義な学習データを用いてニューラルネットワークモデルの効率的な学習や再学習を行う。

10

#### 【0148】

データ分析装置1040は、クラウド又はエッジのいずれに構築されていてもよい。例えばデータ分析装置1040は、学習装置1030と一体の装置として構成されていてもよい。この場合、学習装置1030は、内部でニューラルネットワークモデルの学習に影響を与える学習データを分析して、ニューラルネットワークモデルの学習に所定以上の影響を及ぼす学習データ、不足している学習データ、又はこれらに類似するデータの送信を、データ収集装置1010に対して要求する。

#### 【0149】

あるいは、データ分析装置1040は、データ収集装置1010と一体の装置として構成されていてもよい。この場合、学習装置1030は、学習データの送信要求時などに、データ分析に必要な情報（例えば、その時点で学習済みのニューラルネットワークモデルの情報）を、データ収集装置1010に提供する。そして、データ収集装置1010は、収集したデータがニューラルネットワークモデルに与える影響を分析して、収集したデータのうちニューラルネットワークモデルの学習に所定以上の影響を及ぼす学習データ、不足している学習データ、又はこれらに類似するデータを学習装置1030に送信する。

20

#### 【0150】

データ分析装置1040がデータを分析する方法として、XAI、学習データの信頼度スコア計算、影響関数計算、ベイジアンDNNによるデータ不足推定などが挙げられる。

#### 【0151】

信頼度スコアは、ニューラルネットワークモデルによる予測値の正しさの程度を数値化したものである。上記で説明したニューラルネットワークモデルのうち、観測予測モデル1100及び操作モデル1200は、予測値とともに信頼度スコアを出力するように構成されている。

30

#### 【0152】

学習システム1000では、データ収集装置1010が、信頼度スコアを用いて学習装置1030に送信する学習データのフィルタリングを行うが、この点の詳細については後述に譲る。

#### 【0153】

ニューラルネットワークモデルの出力値の信頼度スコアを計算する方法はいくつかあり、本開示では特定の計算方法に限定されない。ここでは、3種類の信頼度スコアの計算方法(1)～(3)について説明しておく。

40

#### 【0154】

##### (1) 出力の誤差を推定するように学習されたニューラルネットワークモデル

図16に示すように、ニューラルネットワークモデル1500において、本来の出力とともに、その出力の誤差を信頼度スコアとして出力するように学習する。

#### 【0155】

##### (2) ベイズ推定を用いる方法

Image Augmentation(画像拡張)及びDropoutなど乱数を用い、入力データや中間ニューロンの構成に摂動を加え、摂動の下でも正しい予測が得られるような学習を行う。推論時は摂動を加えながら何度も推定を行う。結果の分散が大きい

50

ほど、信頼度スコアが低いことを表す。

**【0156】**

(3) 予測確率を用いる方法(分類問題の場合)

0.0 ~ 1.0 の確率で予測が得られる分類問題の場合、0.0、1.0などの結果が得られた場合は信頼度スコアが高い、2値分類の場合は0.5(50%に近い)、他クラス分類の場合は最も確率の高いクラスの確率が低い場合は信頼度スコアが低いと判断できる。

**【0157】**

また、影響関数は、個々の学習データの有無や振動がニューラルネットワークモデルの予測結果に与える影響を定式化したものである(例えば、非特許文献4を参照のこと)。また、ベイジアン DNN は、ベイズ推定とディープラーニングを組み合わせて構成されるが、ベイズ推定を使うことにより、ニューラルネットワークモデルが予測結果を出力する際のデータ不足による不確実性を評価することができる。

10

**【0158】**

H-4. データ収集装置の構成

データ収集装置 1010 は、スマートフォンなどの多機能情報端末、TV、ヘッドホン又はイヤホン、ゲーム機、冷蔵庫や洗濯機などの IoT デバイス、ドローンやロボットなどの移動体装置など、多種類の装置である。

**【0159】**

図 17 には、データ収集装置 1010 の内部構成例を示している。図 17 に示すデータ収集装置 1010 は、センサー部 1011 と、操作入力部 1012 と、制御部 1013 と、ログ送信部 1014 と、データ分析部 1015 を備えている。但し、図 17 は、多種類のデータ収集装置 1010 のうち、本開示の実現に関わる代表的な機能的構成を抽象化して描いたものであり、個々のデータ収集装置 1010 は図示しないさまざまな構成要素を備えていることが想定される。

20

**【0160】**

センサー部 1011 は、CMOS などで構成されるイメージセンサーや、データ収集装置 1010 に装備されるその他のセンサーからなり、画像や映像の撮影を始めとする観測を行う。また、データ収集装置 1010 がロボットやドローンなどの移動体装置に搭載されている場合には、IMU (Inertial Measurement Unit) などの移動体装置に装備される各種センサーもセンサー部 1011 に含まれるものとする。

30

**【0161】**

操作入力部 1012 は、データ収集装置 1010 における撮影条件などの操作情報を調整するための入力操作を行う機能モジュールである。操作入力部 1012 は、ボタンやつまみなどの操作子やタッチパネル画面などからなる。また、データ収集装置 1010 がロボットやドローンなどの移動体装置に搭載されている場合には、移動体装置を遠隔操作する際に使用するリモコンも操作入力部 1012 に含まれる。

**【0162】**

制御部 1013 は、データ収集装置 1010 全体の動作を統括的に制御する。また、制御部 1013 は、操作入力部 1012 を介して入力された操作情報に基づいて、センサー部 1011 における画像や映像の撮影とする観測の制御を行う。

40

**【0163】**

ログ送信部 1014 は、センサー部 1011 で観測された観測ログ及び操作入力部 1012 に入力された操作ログからなるデータセットを、学習装置 1030 に送信する。基本的には、ログ送信部 1014 は、学習装置 1030 からデータの送信を要求する要求信号を受信したことに応じて学習装置 1030 にデータセットを送信する。データ収集装置 1010 は、学習装置 1030 から要求信号を受信したことに応じて新たにデータを収集して学習装置 1030 に送信するが、既に収集したデータから要求信号に基づいて抽出したデータを学習装置 1030 に送信するようにしてもよい。もちろん、データ収集装置 1010

50

0は、要求信号に応じてではなく自発的にデータを送信するようにしてもよい。

**【0164】**

データ分析部1015は、センサー部1011で観測された観測ログ及び操作入力部1012に入力された操作ログからなるデータセットが、学習対象となる各ニューラルネットワークモデルの学習に与える影響を分析して、学習に所定以上の影響を及ぼす学習データ、不足している学習データ、又はこれらに類似するデータなどの、学習に有意義なデータであるかどうかを特定する。

**【0165】**

学習装置1030から、学習に有意義なデータを指定して要求信号が送信される場合と、収集したデータの分析を指示して要求信号が送信される場合がある。前者の場合、データ分析部1015は、センサー部1011で観測された観測ログ及び操作入力部1012に入力された操作ログからなるデータセットが要求信号で指定されたデータに該当するかどうかをチェックして、ログ送信部1014は、要求に適合するデータのみを学習装置1030に送信する。

10

**【0166】**

また、後者の場合、データ分析部1015は、図10中のデータ分析装置1040に相当する。この場合のデータ分析部1015は、学習装置1030において学習対象とするニューラルネットワークモデルの学習に影響を与える学習データを分析して、センサー部1011で観測された観測ログ及び操作入力部1012に入力された操作ログからなるデータセットが、ニューラルネットワークモデルの学習に所定以上の影響を及ぼす学習データや不足している学習データ、これらに類似するデータなど、有意義な学習データに該当するかどうかをチェックする。そして、ログ送信部1014は、有意義な学習データとなるデータセットのみを学習装置1030に送信する。データ分析部1015は、例えばXAI、信頼度スコア、影響関数、ベイジアンDNNのうち少なくとも1つ、又は2以上の組み合わせを用いて、ニューラルネットワークモデルの分析を行う。例えば、データ分析部1015が信頼度スコアを用いてデータの分析を行う場合には、学習装置1030から現時刻までに学習済みのニューラルネットワークモデルの情報を受け取り、データセットを入力したときのニューラルネットワークモデルによる推論の信頼度スコアを計算する。信頼度スコアを計算する方法については、上記H-3項を参照されたい。そして、信頼度スコアが高い観測ログ及び操作ログは不足していない学習データであり価値は低いが、信頼度スコアが低い観測ログ及び操作ログは不足している学習データであり価値が高いということができる。信頼度スコアが低く不足している学習データを収集して学習装置1030に提供することにより、学習装置103側ではニューラルネットワークモデルの学習や再学習を効率的に行うことができるようになる。

20

**【0167】**

また、データ分析部1015は、分析結果をセンサー部1011及び操作入力部1012にフィードバックするようにしてもよい。そして、センサー部1011及び操作入力部1012は、学習対象のニューラルネットワークモデルの学習への影響度に基づいて、解像度、フレームレート、輝度、色彩、画角、視点位置又は視線方向を変化させて撮影を行うようにしてもよい。

30

**【0168】**

図18には、データ収集装置1010の他の内部構成例を示している。図18に示すデータ収集装置1010は、センサー部1011と、操作入力部1012と、制御部1013と、ログ送信部1014を備えている。図17に示した構成例との主な相違は、データ分析部1015を備えていない点である。図18に示すデータ収集装置1010は、収集したデータの分析を行わず、言い換えれば、収集したデータがニューラルネットワークモデルの学習や再学習に有意義な学習データであるかどうかに関わらず、すべて学習装置1030に送信する。

40

**【0169】**

図19には、学習装置1030の内部構成例を示している。図19に示す学習装置1030は、モデル学習部1031と、観測・操作ログ蓄積部1032と、観測予測モデル・操作モデル蓄積部1033と、操作推定モデル蓄積部1034と、操作推定部1035と、データ分析部1036を備えている。

#### 【0170】

モデル学習部1031は、学習データを使って各種のニューラルネットワークモデルの学習を行う。具体的には、モデル学習部1031は、観測予測モデル（図11を参照のこと）、操作モデル（図12を参照のこと）、操作推定モデル（図13を参照のこと）、感動スコア推定モデル（図14を参照のこと）などの学習を行う。学習装置1030がクラウド上に設置されるクラウドAIの場合、例えば複数の計算ノードを用いてモデル学習部1031を構成して、ニューラルネットワークモデルの分散学習を行うようにしてもよい。

10

#### 【0171】

観測・操作ログ蓄積部1032は、データ収集装置1010から送信された観測ログと操作ログを蓄積する。データ収集装置1010からは、有意義な学習データとなる観測ログ及び操作ログが送られてくるものとする。

#### 【0172】

モデル学習部1031は、観測・操作ログ蓄積部1032に蓄積されている観測ログと操作ログからなるデータセットに基づいて、観測予測モデルの学習を行う。そして、学習済みの観測予測モデルは観測予測モデル・操作モデル蓄積部1033に蓄積される。

20

#### 【0173】

また、モデル学習部1031は、観測・操作ログ蓄積部1032に蓄積されている観測ログと操作ログからなるデータセットに基づいて、操作モデルの学習を行う。そして、学習済みの操作モデルは観測予測モデル・操作モデル蓄積部1033に蓄積される。

#### 【0174】

また、モデル学習部1031は、観測・操作ログ蓄積部1032に蓄積されている観測ログと操作ログからなるデータセットに基づいて、操作推定モデルの学習を行う。そして、学習済みの操作推定モデルは操作推定モデル蓄積部1034に蓄積される。

30

#### 【0175】

データ分析部1036は、図10中のデータ分析装置1040に相当する。データ分析部1036は、観測予測モデル・操作モデル蓄積部1033及び操作推定モデル蓄積部1034に蓄積されている学習済みの各ニューラルネットワークの学習に影響を与える学習データを分析する。そして、この分析結果に基づいて、所定以上の影響を及ぼす学習データ、不足している学習データ、又はこれらに類似するデータの送信を要求する要求信号が、学習装置1030からデータ収集装置1010へ送信される。したがって、データ収集装置1010からは、有意義な学習データとなる観測ログ及び操作ログが送られてくる。データ分析部1036は、例えば、XAI、信頼度スコア計算、影響関数、又はベイジアンDNNなどの手法に基づいて、学習データの分析を行う。

#### 【0176】

操作推定部1035は、学習済みの操作推定モデルを用いて、プロのカメラマンなどのカメラ操作に精通したエキスパートが撮影した映像（以下、「プロの映像」とも呼ぶ）から、その映像を撮影するための時系列操作情報を推定する。そして、操作推定部1035に入力されたプロの映像（観測ログ）と、操作推定部1035で推定された操作情報（推定された操作ログ）からなるデータセットが、プロ並みの操作モデルを学習するための高品質な学習データとして、観測・操作ログ蓄積部1032に蓄積される。したがって、モデル学習部1031は、観測・操作ログ蓄積部1032に蓄積されている高品質な学習データを用いて、プロ並みのカメラ操作を予測することが可能な操作モデルを学習することができる。

40

#### 【0177】

なお、プロの映像コンテンツは任意であり、ネットワーク経由でクラウドに提供されることを想定している。プロの映像コンテンツを使って大量の高品質な学習データを収集し

50

ていくことで、プロ並みのカメラ自動操作を実現するための操作モデルの学習を行うことが可能になる。

#### 【0178】

図20には、学習装置1030の他の内部構成例を示している。図20に示す学習装置1030は、モデル学習部1031と、観測・操作ログ蓄積部1032と、観測予測モデル・操作モデル蓄積部1033と、操作推定モデル蓄積部1034と、操作推定部1035を備えている。図19に示した構成例との主な相違点は、データ分析部1036を備えていない点である。

#### 【0179】

図20に示す学習装置1030は、自らは学習データの分析を行わず、その代わりに学習対象となるニューラルネットワークモデルの情報をデータ収集装置1010に送信する。この場合、データ収集装置1010側では、データ分析部1015が、学習装置1030において学習対象とするニューラルネットワークモデルの学習に影響を与える学習データを分析して、センサー部1011で観測された観測ログ及び操作入力部1012に入力された操作ログからなるデータセットが、ニューラルネットワークモデルの学習に所定以上の影響を及ぼす学習データや不足している学習データ、これらに類似するデータなど、有意義な学習データに該当するかどうかをチェックする(前述)。そして、データ収集装置1010から学習装置1030へ、有意義な学習データとなるデータセットのみが送信される。

10

#### 【0180】

図21には、学習装置1030のさらに他の内部構成例を示している。図21に示す学習装置1030は、モデル学習部1031と、観測・操作ログ蓄積部1032と、観測予測モデル・操作モデル蓄積部1033と、操作推定モデル蓄積部1034と、操作推定部1035と、データ分析部1037を備えている。図19及び図20に示した構成例との主な相違点は、学習装置1030は、図18に示した構成例からなるデータ収集装置1010から、すべての収集データを受信して、データ分析部1037において各受信データが有意義な学習データかどうかをチェックする点である。

20

#### 【0181】

データ分析部1037は、図10中のデータ分析装置1040に相当する。データ分析部1037は、観測・操作ログ蓄積部1032に蓄積されている各受信データが、観測予測モデル・操作モデル蓄積部1033及び操作推定モデル蓄積部1034に蓄積されている学習済みの各ニューラルネットワークの学習に与える影響を分析する。そして、データ分析部1037は、各ニューラルネットワークモデルの学習に所定以上の影響を及ぼす学習データや不足している学習データ、これらに類似するデータなど、有意義な学習データのみを、観測・操作ログ蓄積部1032から抽出して、モデル学習部1031に出力する。したがって、モデル学習部1031は、有意義な学習データを使って各種のニューラルネットワークモデルの学習や再学習を効率的に行うことができる。

30

#### 【0182】

### H - 6 . モデル利用装置の構成

図22には、モデル利用装置1020の内部構成例を示している。モデル利用装置1020は、センサー部1021と、自動操作部1022と、制御部1023と、提示部1024を備えている。

40

#### 【0183】

センサー部1021は、CMOSなどで構成されるイメージセンサーや、モデル利用装置1020に装備されるその他のセンサーからなり、画像や映像の撮影を始めとする観測を行う。また、モデル利用装置1020がロボットやドローンなどの移動体装置に搭載されている場合には、IMUなどの移動体装置に装備される各種センサーもセンサー部1021に含まれるものとする。

#### 【0184】

自動操作部1022は、学習装置1030から提供された操作モデルを用いて、センサ

50

一部 1021 の観測情報（現時刻までの映像情報及び現時刻までの操作情報）から次の時刻の操作情報を予測するとともに、予測した操作情報の信頼度スコアなどのデータ分析結果を出力する。

#### 【0185】

制御部 1023 は、自動操作部 1022 によって予測された次の時刻の操作情報に基づいて、センサー部 1021 における画像や映像の撮影とする観測の制御を行う。

#### 【0186】

学習装置 1030 において十分な学習が行われた操作モデルが提供される場合、モデル利用装置 1020 のユーザ自身はカメラ操作に精通していないなくても、自動操作部 1022 によって予測された操作情報に基づいて、プロのカメラマンなどカメラ操作に精通したエキスパートと同等の撮影を行うことができる。10

#### 【0187】

提示部 1024 は、自動操作部 1022 において予測した操作情報の信頼度スコアなどのデータ分析結果を提示する。モデル利用装置 1020 のユーザは、提示された信頼度スコアから、自動操作によりプロ並みの腕前で映像の撮影が行われるかどうかを判断することができる。

#### 【0188】

### H-7. エッジAIの構成

学習済みのニューラルネットワークモデルを利用するモデル利用装置 1020 が、ニューラルネットワークモデルの学習を行う学習装置の機能と一体となって構成されること（すなわち、エッジAI）も想定される。図 23 には、エッジAIとしてのモデル利用装置 1020 の内部構成例を示している。図示のモデル利用装置 1020 は、センサー部 1021 と、自動操作部 1022 と、制御部 1023 と、提示部 1024 に加えて、モデル学習部 1031 と、観測・操作ログ蓄積部 1032 と、観測予測モデル・操作モデル蓄積部 1033 と、操作推定モデル蓄積部 1034 と、操作推定部 1035 を備えている。20

#### 【0189】

観測・操作ログ蓄積部 1032 は、データ収集装置 1010 から送信された観測ログと操作ログを蓄積する。モデル学習部 1031 は、観測・操作ログ蓄積部 1032 に蓄積されている観測ログと操作ログからなるデータセットに基づいて、観測予測モデルの学習を行う。また、モデル学習部 1031 は、観測・操作ログ蓄積部 1032 に蓄積されている観測ログと操作ログからなるデータセットに基づいて、操作モデルの学習を行う。そして、学習済みの観測予測モデル及び操作モデルは観測予測モデル・操作モデル蓄積部 1033 に蓄積される。30

#### 【0190】

また、モデル学習部 1031 は、観測・操作ログ蓄積部 1032 に蓄積されている観測ログと操作ログからなるデータセットに基づいて、操作推定モデルの学習を行う。そして、学習済みの操作推定モデルは操作推定モデル蓄積部 1034 に蓄積される。操作推定部 1035 は、学習済みの操作推定モデルを用いて、プロのカメラマンなどのカメラ操作に精通したエキスパートが撮影した映像から、その映像を撮影するための時系列操作情報を推定して、プロ並みの操作モデルを学習するための高品質な学習データとして、観測・操作ログ蓄積部 1032 に蓄積される。したがって、モデル学習部 1031 は、観測・操作ログ蓄積部 1032 に蓄積されている高品質な学習データを用いて、プロ並みのカメラ操作を予測することが可能な操作モデルを学習することができる。40

#### 【0191】

センサー部 1021 は、CMOS などで構成されるイメージセンサーや、モデル利用装置 1020 に装備されるその他のセンサーからなり、画像や映像の撮影を始めとする観測を行う。

#### 【0192】

自動操作部 1022 は、観測予測モデル・操作モデル蓄積部 1033 から読み出した操作モデルを用いて、センサー部 1021 の観測情報（現時刻までの映像情報及び現時刻ま50

での操作情報)から次の時刻の操作情報を予測するとともに、予測した操作情報の信頼度スコアなどのデータ分析結果を出力する。

#### 【0193】

制御部1023は、自動操作部1022によって予測された次の時刻の操作情報に基づいて、センサー部1011における画像や映像の撮影とする観測の制御を行う。学習装置1030において十分な学習が行われた操作モデルが提供される場合、モデル利用装置1020のユーザ自身はカメラ操作に精通していなくても、自動操作部1022によって予測された操作情報に基づいて、プロのカメラマンなどカメラ操作に精通したエキスパートと同等の撮影を行うことができる。

#### 【0194】

提示部1024は、自動操作部1022において予測した操作情報の信頼度スコアなどのデータ分析結果を提示する。モデル利用装置1020は、提示された信頼度スコアから、自動操作によりプロ並みの腕前で映像の撮影が行われるかどうかを判断することができる。

#### 【産業上の利用可能性】

#### 【0195】

以上、特定の実施形態を参照しながら、本開示について詳細に説明してきた。しかしながら、本開示の要旨を逸脱しない範囲で当業者が該実施形態の修正や代用を成し得ることは自明である。

#### 【0196】

本開示は、人間に感動を与えるコンテンツを生成する機械学習モデルの学習に用いられるデータの収集処理に適用することができる。本開示に基づいて収集したデータは、人間に感動を与えるコンテンツを扱うための(具体的には、感動するコンテンツを生成し又は感動するコンテンツを撮影するためのカメラ操作を推定するための)機械学習モデル(例えば、ニューラルネットワークモデル)を学習させるための学習データに用いられるが、もちろんそれ以外の用途の機械学習モデルの学習に利用することも可能である。

#### 【0197】

また、本開示に基づいて収集したデータを使って、コンテンツが人間に与える感情を識別する機械学習モデルを訓練することができ、このようにして開発した機械学習モデルを備えた感情識別機を、デジタルカメラを始めとするコンテンツ処理装置に搭載することができる。

#### 【0198】

要するに、例示という形態により本開示について説明してきたのであり、本明細書の記載内容を限定的に解釈するべきではない。本開示の要旨を判断するためには、特許請求の範囲を参照すべきである。

#### 【0199】

なお、本開示は、以下のような構成をとることも可能である。

#### 【0200】

(1) データを収集するデータ収集装置と、前記データ収集装置が収集したデータを用いて機械学習モデルの学習を行う学習装置を備え、

前記学習装置は、前記機械学習モデルの学習に影響を与える学習データを分析した結果に基づいて収集した、前記機械学習モデルの学習に所定以上の影響を及ぼす学習データ、不足している学習データ、又はこれらに類似するデータを用いて、前記機械学習モデルの再学習を行う、

学習システム。

#### 【0201】

(2) 前記学習装置は、感動するコンテンツを生成し又は感動するコンテンツを撮影するためのカメラ操作を推定する機械学習モデルの学習を行う、  
上記(1)に記載の学習システム。

#### 【0202】

10

20

30

40

50

(3) X A I、信頼度スコア計算、影響関数、又はベイジアン D N N による前記分析を行う、

上記(1)又は(2)のいずれかに記載の学習システム。

**【0203】**

(4) 前記学習装置は、前記機械学習モデルの学習に影響を与える学習データを分析し、その分析結果に基づいて、前記機械学習モデルの学習に所定以上の影響を及ぼす学習データ、不足している学習データ、又はこれらに類似するデータの送信を要求する要求信号を前記データ収集装置に送信し、

前記データ収集装置は、受信した前記要求信号に基づいて収集したデータを前記学習装置に送信し、

10

前記学習装置は、前記要求信号に応じて前記データ収集装置から送信されたデータに基づいて前記機械学習モデルの再学習を行う、

上記(1)乃至(3)のいずれかに記載の学習システム。

**【0204】**

(5) 前記データ収集装置は、画像を撮影するカメラ又はイメージヤであり、前記機械学習モデルの学習への影響度に基づいて、解像度、フレームレート、輝度、色彩、画角、視点位置、又は視線方向を変化させて撮影した画像データを前記学習装置に送信する、上記(1)乃至(4)のいずれかに記載の学習システム。

**【0205】**

(6) 前記学習装置は、前記機械学習モデルの学習データの送信を要求する要求信号を前記データ収集装置に送信し、

20

前記データ収集装置は、収集したデータが前記機械学習モデルの学習に与える影響を分析した結果に基づいて、前記収集したデータのうち前記機械学習モデルの学習に所定以上の影響を及ぼす学習データ、不足している学習データ、又はこれらに類似するデータを前記学習装置に送信し、

前記学習装置は、前記要求信号に応じて前記データ収集装置から送信されたデータに基づいて前記機械学習モデルの再学習を行う、

上記(1)乃至(3)のいずれかに記載の学習システム。

**【0206】**

(7) 前記学習装置は、前記要求信号の送信時に、前記分析に必要な情報を前記データ収集装置に送信する、

30

上記(6)に記載の学習システム。

**【0207】**

(8) 機械学習モデルの学習を行う学習装置から、前記機械学習モデルの学習データの送信を要求する要求信号を受信する受信部と、

前記要求信号を受信したことにより、前記機械学習モデルの学習に所定以上の影響を及ぼす学習データ、不足している学習データ、又はこれらに類似するデータを収集するデータ収集部と、

前記データ収集部が収集したデータを前記学習装置に送信する送信部と、  
を具備するデータ収集装置。

40

**【0208】**

(9) 前記受信部は、前記機械学習モデルの学習に所定以上の影響を及ぼす学習データ、不足している学習データ、又はこれらに類似するデータを要求する前記要求信号を前記学習装置から受信し、

前記データ収集部は、受信した前記要求信号に基づいてデータを収集し、  
前記送信部は、前記データ収集部が収集したデータを前記学習装置に送信する、  
上記(8)に記載のデータ収集装置。

**【0209】**

(10) 前記データ収集部は受信した前記要求信号に基づいて新たにデータを収集し、又は、前記送信部は前記データ収集部が既に収集したデータから前記要求信号に基づいて抽

50

出したデータを送信する、  
上記(9)に記載のデータ収集装置。

**【0210】**

(11) 前記データ収集部が収集したデータが前記機械学習モデルの学習に与える影響を分析する分析部をさらに備え、

前記送信部は、前記分析部による分析結果に基づいて、前記データ収集部が収集したデータのうち前記機械学習モデルの学習に所定以上の影響を及ぼす学習データ、不足している学習データ、又はこれらに類似するデータを前記学習装置に送信する、

上記(8)に記載のデータ収集装置。

**【0211】**

(12) 前記データ収集部は、前記機械学習モデルの学習への影響度に基づいて、画像を撮影するカメラ又はイメージヤの解像度、フレームレート、輝度、色彩、画角、視点位置、又は視線方向を変化させて撮影した画像データを収集する、

上記(8)乃至(11)のいずれかに記載のデータ収集装置。

**【0212】**

(13) コンテンツのコンテンツ評価情報とコンテンツを視聴する人間の生体情報に基づいて、コンテンツからセグメントを抽出するセグメント抽出部と、

セグメントに写っている人物の顔を検出して顔の表情を識別する表情識別部と、

セグメントに写っている人物間の関係性を推定する関係性推定部と、

セグメントに写っている人物の顔の表情と、人物間の関係性に基づいてセグメントの感動ラベルを識別する感動識別部と、

を具備する情報処理装置。

20

**【0213】**

(14) 前記生体情報は少なくとも脳波情報を含む、

上記(13)に記載の情報処理装置。

**【0214】**

(15) 前記セグメント抽出部は、コンテンツのうち高評価とポジティブな感情を持つ生体情報が一致しているセグメントを抽出する、

上記(13)又は(14)のいずれかに記載の情報処理装置。

**【0215】**

(16) 前記セグメント抽出部は、高評価とポジティブな感情を持つ生体情報が一致しないセグメントをさらに抽出する、

上記(15)に記載の情報処理装置。

30

**【0216】**

(17) 前記関係性推定部は、前後のセグメントとのコンテキストと、人物の顔の表情に基づいて、セグメントに写っている人物間の関係性を推定する、

上記(13)乃至(16)のいずれかに記載の情報処理装置。

**【0217】**

(18) 前記関係性推定部は、前記コンテキストとして、過去のセグメントに写っている人物の顔検出及び検出顔の表情識別結果を用いて、現在のセグメントに写っている人物間の関係性を推定する、

上記(17)に記載の情報処理装置。

40

**【0218】**

(19) 前記セグメントに含まれる音声を認識して得られたテキスト情報に基づいて感情を分析する第1の感情分析部をさらに備え、

前記感動識別部は、前記第1の感情分析部がテキスト情報から識別した感情をさらに考慮して、セグメントの感動ラベルを識別する、

上記(13)乃至(18)のいずれかに記載の情報処理装置。

**【0219】**

(20) 前記セグメントに含まれる音楽を検出して、その音楽が与える感情を分析する第

50

2の感情分析部をさらに備え、

前記感動識別部は、前記第2の感情分析部が音楽から識別した感情をさらに考慮して、セグメントの感動ラベルを識別する、

上記(13)乃至(19)のいずれかに記載の情報処理装置。

#### 【0220】

(21)コンテンツのコンテンツ評価情報とコンテンツを視聴する人間の生体情報に基づいて、コンテンツからセグメントを抽出するセグメント抽出ステップと、

セグメントに写っている人物の顔を検出して顔の表情を識別する表情識別ステップと、

セグメントに写っている人物間の関係性を推定する関係性推定ステップと、

セグメントに写っている人物の顔の表情と、人物間の関係性に基づいてセグメントの感動ラベルを識別する感動識別ステップと、

を有する情報処理方法。

#### 【0221】

(22)コンテンツのコンテンツ評価情報とコンテンツを視聴する人間の生体情報に基づいて、コンテンツからセグメントを抽出するセグメント抽出部、

セグメントに写っている人物の顔を検出して顔の表情を識別する表情識別部、

セグメントに写っている人物間の関係性を推定する関係性推定部、

セグメントに写っている人物の顔の表情と、人物間の関係性に基づいてセグメントの感動ラベルを識別する感動識別部、

としてコンピュータを機能させるようにコンピュータ可読形式で記述されたコンピュータプログラム。

#### 【0222】

(23)コンテンツが与える感動を識別する学習済み機械学習モデルを生成する生成方法であって、

機械学習モデルにコンテンツを入力するステップと、

前記機械学習モデルが前記コンテンツから推定した映像特徴量を獲得するステップと、

前記機械学習モデルが前記コンテンツから推定したオーディオ特徴量を獲得するステップと、

前記機械学習モデルが前記コンテンツから推定した音声のテキスト特徴量を獲得するステップと、

前記機械学習モデルが推定した映像特徴量、オーディオ特徴量、及びテキスト特徴量に基づいてコンテンツの感動ラベルを識別するステップと、

前記識別した感動ラベルと前記コンテンツに付けられた感動ラベルとの誤差に基づく損失関数を計算するステップと、

前記損失関数に基づいて前記機械学習モデルのパラメータを更新するステップと、  
を有する学習済み機械学習モデルの生成方法。

#### 【0223】

(24)前記映像特徴量は、映像のフレームに写っている人物が2人以上の場合において、連続する複数のフレームのコンテキストに基づいて推定される人物間の関係性を含む、上記(23)に記載の学習済み機械学習モデルの生成方法。

#### 【符号の説明】

#### 【0224】

100...データ収集システム、101...映像コンテンツ

102...コンテンツ評価情報取得部、103...生体情報取得部

104...比較部、105...セグメント抽出部

106...コンテキスト抽出部、107...音声認識部

108...音楽検出部、109...顔検出部

110...第1の感情分析部、111...第2の感情分析部

112...表情識別部、113...関係性推定部

114...感動識別部

10

20

30

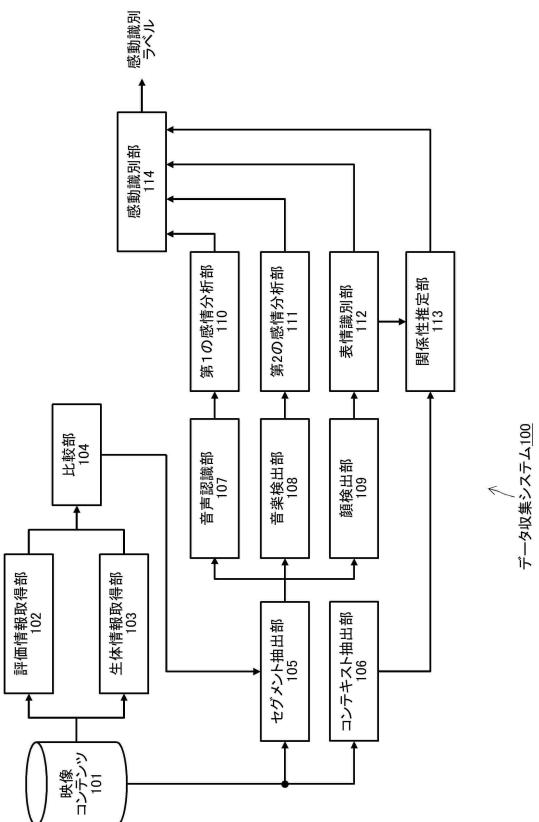
40

50

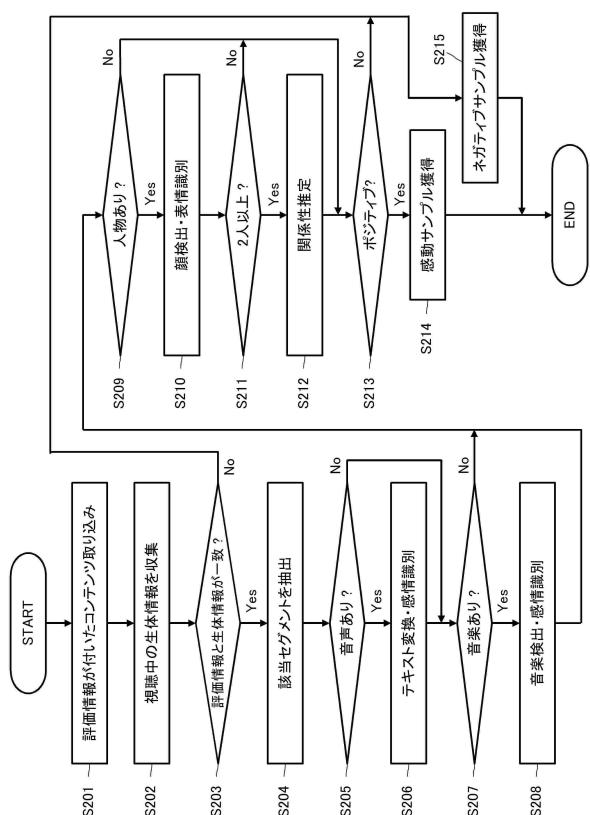
410...データ蓄積部、420...感動識別機  
 421...ネットワーク部、422...識別部、423...評価部  
 500...デジタルカメラ、501...光学系、502...撮像部  
 503...A F E部、504...カメラ信号処理部、506...メイン処理部  
 514...マイク、515...A / D変換部、516...表示部  
 517...音再生部、518...記録部  
 1000...学習システム、1010...データ収集装置  
 1011...センサー部、1012...操作入力部、1013...制御部  
 1014...ログ送信部、1015...データ分析部  
 1020...モデル利用装置、1021...センサー部  
 1022...自動操作部、1023...制御部、1024...提示部  
 1030...学習装置、1031...モデル学習部  
 1032...観測・操作ログ蓄積部  
 1033...観測予測モデル・操作モデル蓄積部  
 1034...操作推定モデル蓄積部、1035...操作推定部  
 1036...データ分析部、1037...データ分析部  
 1040...データ分析装置  
 1100...観測予測モデル、1200...操作モデル  
 1300...操作推定モデル、1400...感動スコア推定モデル

【図面】

【図1】



【図2】



10

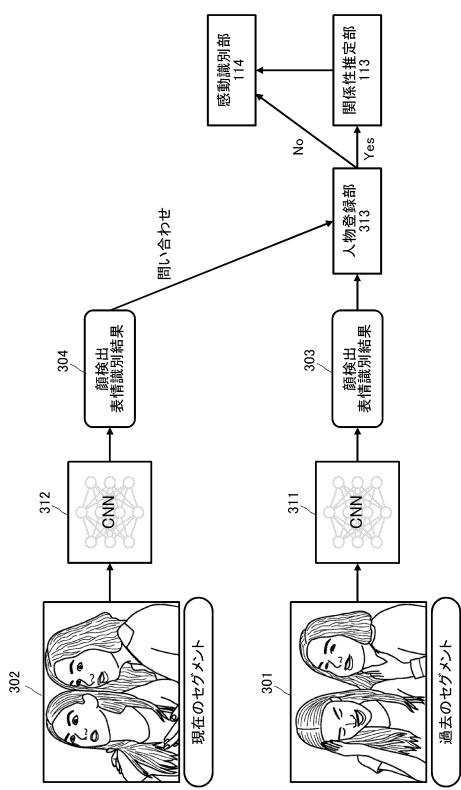
20

30

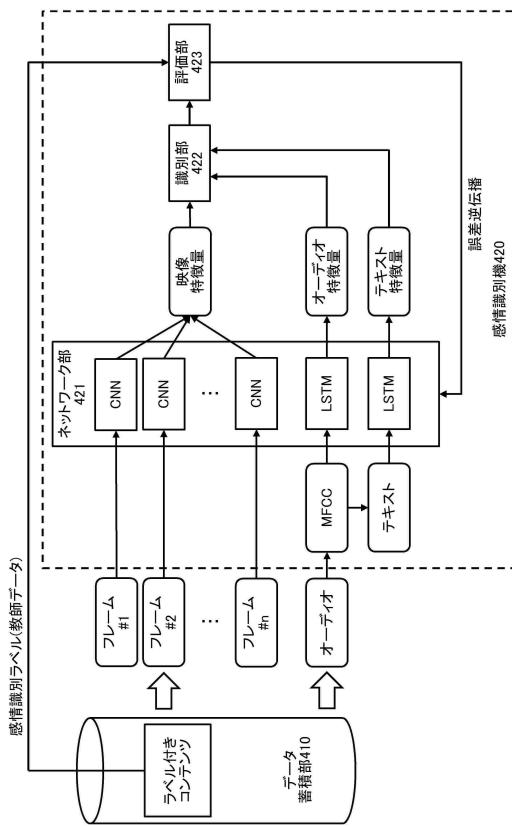
40

50

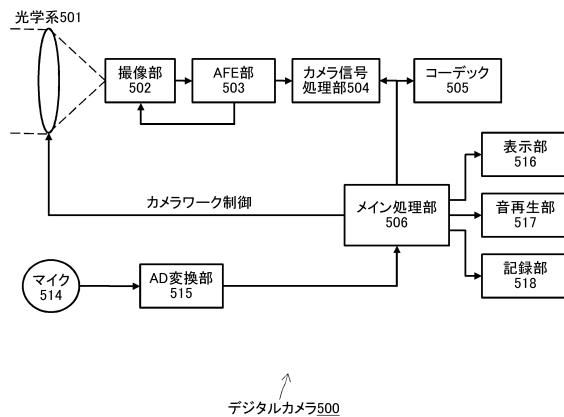
【図3】



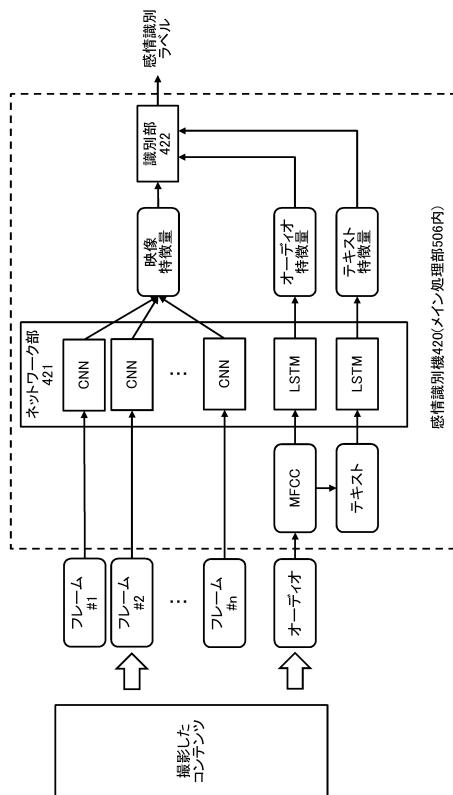
【図4】



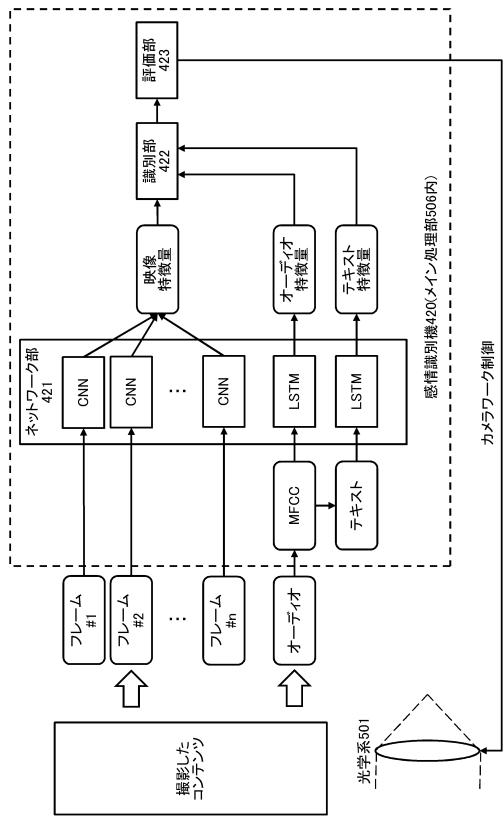
【図5】



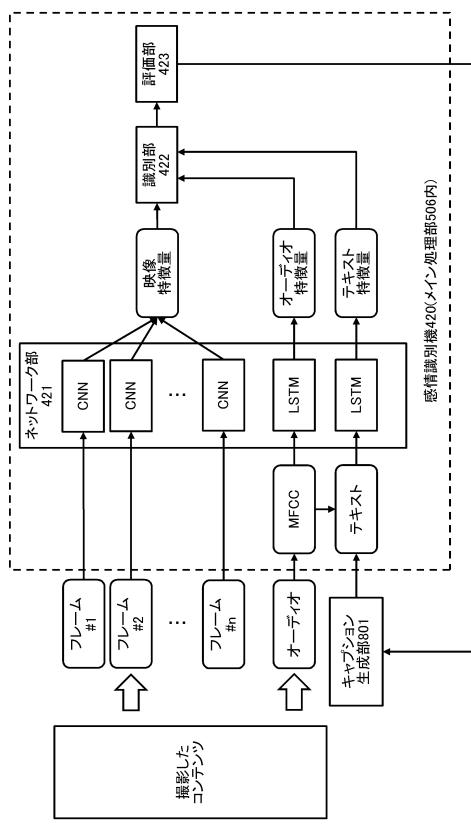
【図6】



【図 7】



【図 8】



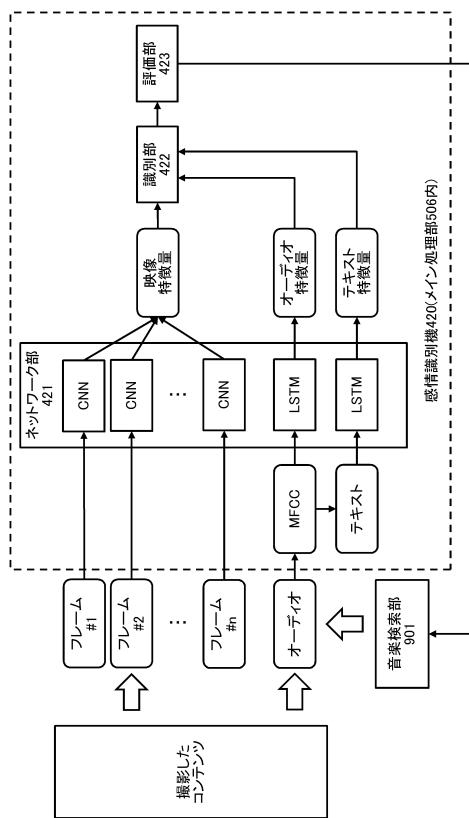
10

20

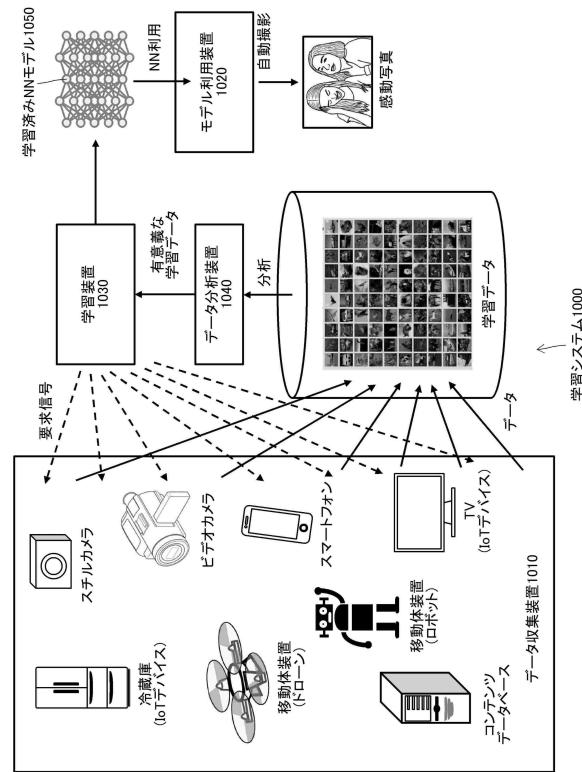
30

40

【図 9】

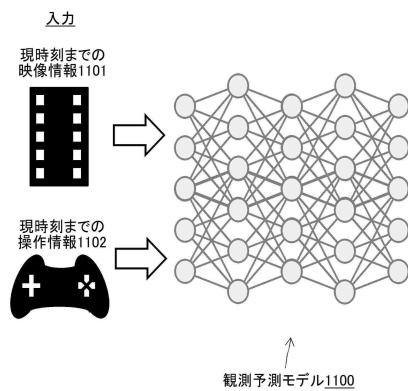


【図 10】

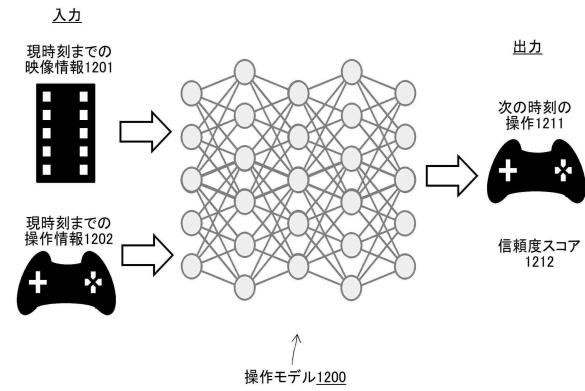


50

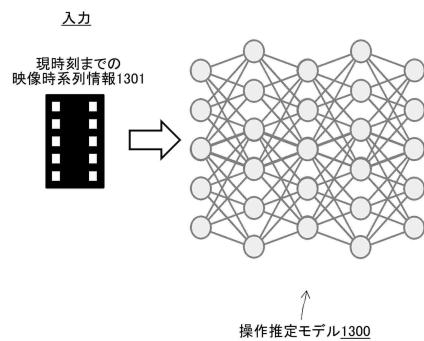
【図 1 1】



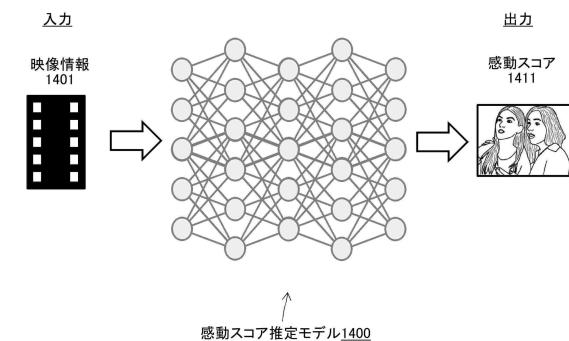
【図 1 2】



【図 1 3】



【図 1 4】

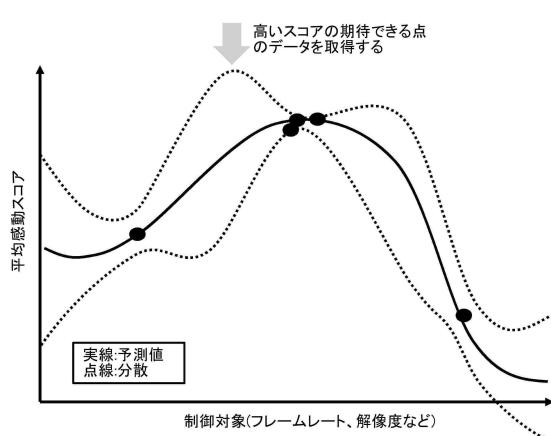


30

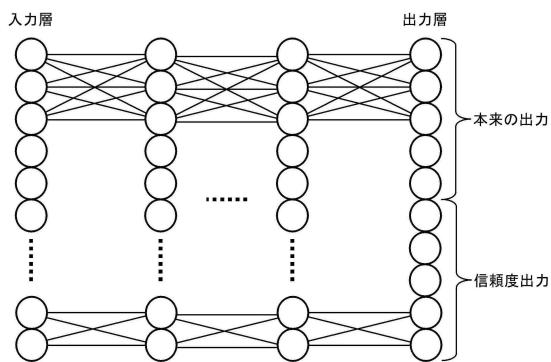
40

50

【図 1 5】

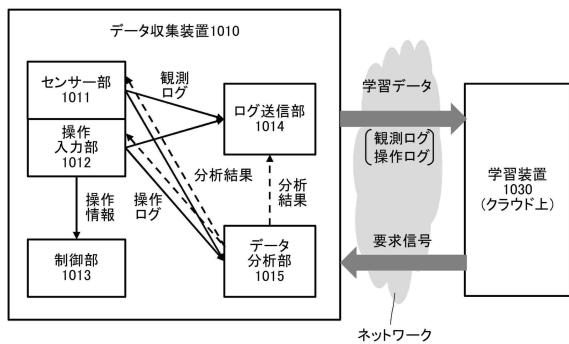


【図 1 6】

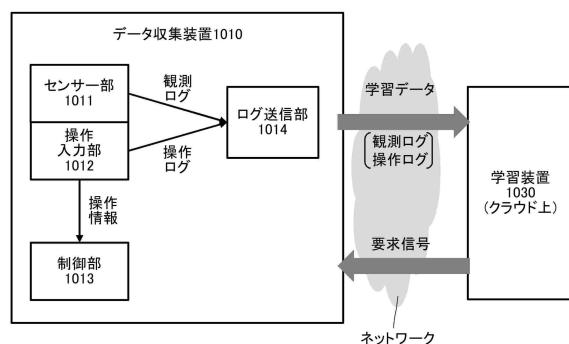


10

【図 1 7】



【図 1 8】



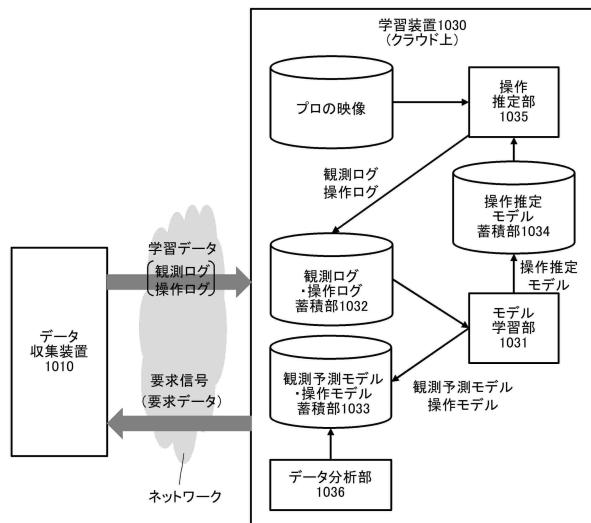
20

30

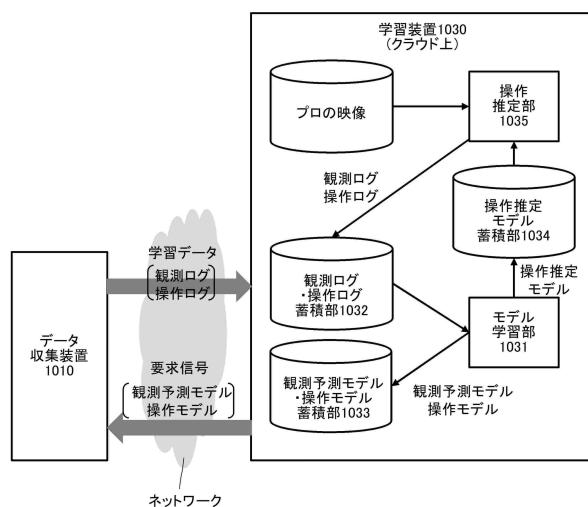
40

50

【図19】

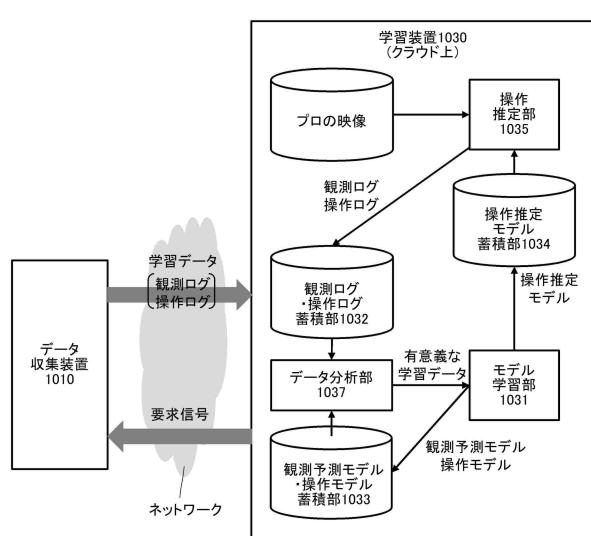


【図20】

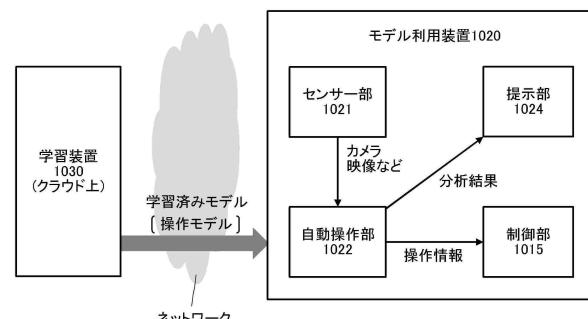


10

【図21】



【図22】



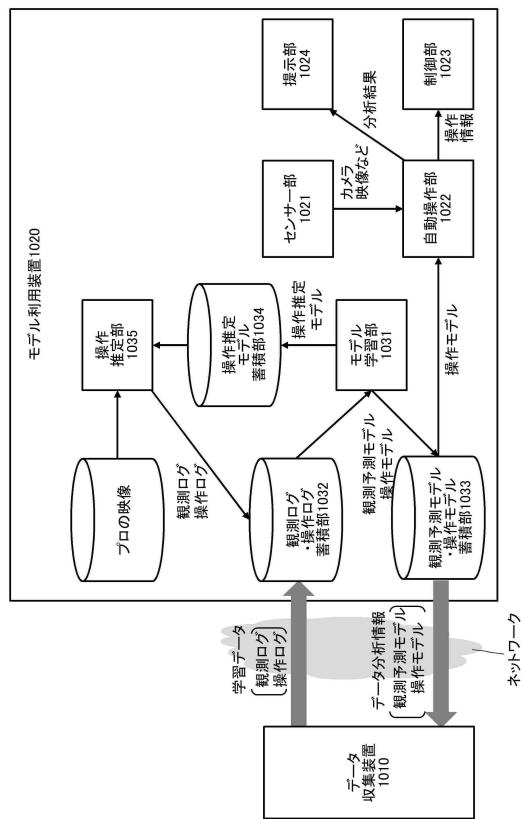
20

30

40

50

【図 2 3】



10

20

30

40

50

---

フロントページの続き

## (51)国際特許分類

G 10 L	25/63 (2013.01)	F I	G 10 L	15/10	500 N
A 61 B	5/16 (2006.01)		G 10 L	25/63	
			A 61 B	5/16	120

グループ株式会社内

## (72)発明者 小林 由幸

東京都港区港南1丁目7番1号 ソニーグループ株式会社内

## (72)発明者 鈴木 健二

東京都港区港南1丁目7番1号 ソニーグループ株式会社内

## 審査官 大倉 嶺吾

## (56)参考文献 特開平11-085719 (JP, A)

国際公開第2018/030206 (WO, A1)

国際公開第2019/215778 (WO, A1)

特開2010-093584 (JP, A)

特開2008-269065 (JP, A)

特開2012-160082 (JP, A)

特開2009-111938 (JP, A)

猪貝光祥, "ディープラーニング技術を用いた高速な画像認識ソリューション", 月刊自動認識, 2020年03月10日, Vol. 33, No. 3, p. 33-38, ISSN: 0915-1060

横井直明 ほか, "AIの予測結果に対する納得度を高める予測根拠解釈支援技術の提案", 電子情報通信学会技術研究報告, 一般社団法人電子情報通信学会, 2019年03月10日, Vol. 118, No. 513, p. 61-66

## (58)調査した分野 (Int.Cl., DB名)

G 06 N 3 / 00 - 99 / 00

G 06 T 7 / 00

H 04 N 21 / 258

G 10 L 15 / 10

G 10 L 25 / 63

A 61 B 5 / 16