(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7335274号
(P7335274)

(45)発行日　令和5年8月29日(2023.8.29)

(24)登録日　令和5年8月21日(2023.8.21)

(51)国際特許分類		FI		
G06T 7/00 (2017.01)		G06T	7/00	350B
G06N 3/08 (2023.01)		G06N	3/08	
G06N 20/00 (2019.01)		G06N	20/00	
G06V 10/82 (2022.01)		G06T	7/00	650A
G06V 20/40 (2022.01)		G06V	10/82	

請求項の数　14　（全30頁）　最終頁に続く

(21)出願番号　特願2020-566225(P2020-566225)
(86)(22)出願日　平成31年1月10日(2019.1.10)
(65)公表番号　特表2021-531541(P2021-531541A)
(43)公表日　令和3年11月18日(2021.11.18)
(86)国際出願番号　PCT/US2019/013024
(87)国際公開番号　WO2020/131140
(87)国際公開日　令和2年6月25日(2020.6.25)
審査請求日　令和3年1月21日(2021.1.21)
(31)優先権主張番号　62/781,276
(32)優先日　平成30年12月18日(2018.12.18)
(33)優先権主張国・地域又は機関
米国(US)

(73)特許権者　502208397
グーグル エルエルシー
Ｇｏｏｇｌｅ ＬＬＣ
アメリカ合衆国 カリフォルニア州 ９４０４３ マウンテン ビュー アンフィシアター パークウェイ １６００
１６００ Ａｍｐｈｉｔｈｅａｔｒｅ Ｐａｒｋｗａｙ ９４０４３ Ｍｏｕｎｔａｉｎ Ｖｉｅｗ， ＣＡ Ｕ．Ｓ．Ａ．
(74)代理人　100108453
弁理士　村山 靖彦
(74)代理人　100110364
弁理士　実広 信哉
(74)代理人　100133400
弁理士　阿部 達彦

最終頁に続く

(54)【発明の名称】　ジオロケーションの予測のためのシステムおよび方法

(57)【特許請求の範囲】

【請求項１】

画像から情報を抽出するためのコンピュータ実装方法であって、

1つまたは複数のプロセッサを含むコンピューティングシステムにおいて、画像のシーケンスを表すデータを取得するステップであって、画像の前記シーケンスのうちの少なくとも1つの画像が、物体を示す、ステップと、

前記コンピューティングシステムによって、画像の前記シーケンスから位置情報を抽出するように訓練される機械学習された情報抽出モデルに画像の前記シーケンスを入力するステップと、

画像の前記シーケンスを入力したことに応じて前記情報抽出モデルの出力として、前記コンピューティングシステムによって、画像の前記シーケンス内に示された前記物体に関連する実世界の位置を表すデータを取得するステップであって、

前記コンピューティングシステムによって、画像の前記シーケンス内に示された前記物体に関連する分類を表すデータを決定するステップと、

前記コンピューティングシステムによって、画像の前記シーケンスから抽出された画像の特徴のシーケンスを表すデータに少なくとも部分的に基づいて画像の前記シーケンスに関連する時間的アテンション値および空間的アテンション値を決定するステップと、

前記コンピューティングシステムによって、画像の特徴の前記シーケンス、前記時間的アテンション値、および前記空間的アテンション値に基づいて、前記物体に関連するカメラ座標空間内の座標および前記物体に関連するカメラ姿勢データを取得するステップと、

前記コンピューティングシステムによって、画像の特徴の前記シーケンス、前記時間的アテンション値、および前記空間的アテンション値のうちの少なくとも１つ、ならびに前記物体に関連するカメラ座標空間内の座標および前記物体に関連するカメラ姿勢データに基づいて前記物体に関連する前記実世界の位置を予測するステップと、

前記コンピューティングシステムによって、前記物体に関連する分類を表す前記決定されたデータに少なくとも部分的に基づいて、前記物体に関連する前記予測された実世界の位置を前記物体に対応する分類ラベルに関連付けるステップと、を含む、

ステップと、を含む、

コンピュータ実装方法。

【請求項２】

画像の前記シーケンスに関連する前記時間的アテンション値および前記空間的アテンション値を決定するステップが、

前記コンピューティングシステムによって、弱教師あり物体分類モデルに画像の特徴の前記シーケンスを入力するステップであって、前記物体分類モデルが、少なくとも1つの長期短期記憶ブロックを含む、ステップと、

画像の特徴の前記シーケンスを入力したことに応じて前記物体分類モデルの出力として、前記コンピューティングシステムによって、前記時間的アテンション値および前記空間的アテンション値を取得するステップとを含む、

請求項1に記載のコンピュータ実装方法。

【請求項３】

画像の前記シーケンス内に示された前記物体に関連する前記分類を決定するステップが、

前記コンピューティングシステムによって、弱教師あり物体分類モデルに画像の特徴の前記シーケンスを入力するステップと、

画像の特徴の前記シーケンスを入力したことに応じて前記物体分類モデルの出力として、前記コンピューティングシステムによって、前記物体に関連する前記分類を取得するステップとを含む、

請求項1または2に記載のコンピュータ実装方法。

【請求項４】

画像の前記シーケンスを表す前記データが、画像の前記シーケンスに関連する少なくとも1つの分類ラベルを含み、前記方法が、

前記コンピューティングシステムによって、画像の前記シーケンスに関連する前記少なくとも1つの分類ラベルに少なくとも部分的に基づいて前記物体分類モデルによって出力された前記分類に関連する損失を決定するステップと、

前記コンピューティングシステムによって、決定された損失に少なくとも部分的に基づいて前記物体分類モデルを訓練するステップとをさらに含む、

請求項3に記載のコンピュータ実装方法。

【請求項５】

前記物体に関連する前記実世界の位置を予測するステップが、

前記コンピューティングシステムによって、フレームレベル位置特徴抽出モデルに画像の特徴の前記シーケンス、前記時間的アテンション値、および前記空間的アテンション値を入力するステップと、

画像の特徴の前記シーケンス、前記時間的アテンション値、および前記空間的アテンション値を入力したことに応じて前記フレームレベル位置特徴抽出モデルの出力として、前記コンピューティングシステムによって、前記物体に関連する1つまたは複数の位置の特徴を含む位置の特徴のシーケンスを表すデータを取得するステップと、

前記コンピューティングシステムによって、フレームレベル位置予測モデルに位置の特徴の前記シーケンスを入力するステップと、

位置の特徴の前記シーケンスを入力したことに応じて前記フレームレベル位置予測モデルの出力として、前記コンピューティングシステムによって、前記物体に関連する前記カメラ座標空間内の前記座標を表すデータを取得するステップと、

前記コンピューティングシステムによって、前記カメラ座標空間内の前記座標および前記物体に関連する前記カメラ姿勢データに基づいて前記物体に関連する実世界の座標を決定するステップとを含む、
請求項1、2、3、または4に記載のコンピュータ実装方法。
【請求項６】
前記コンピューティングシステムによって、前記物体を示す画像の前記シーケンスの中の複数の画像にまたがる前記物体に関連する座標の間の分散に少なくとも部分的に基づいて位置の整合性の損失を決定するステップと、
前記コンピューティングシステムによって、前記位置の整合性の損失に少なくとも部分的に基づいて前記フレームレベル位置予測モデルを訓練するステップとをさらに含む、
請求項5に記載のコンピュータ実装方法。
【請求項７】
前記コンピューティングシステムによって、前記物体を示す画像の前記シーケンスの中の複数の画像にまたがって決定された外観の特徴の間の分散に少なくとも部分的に基づいて外観の整合性の損失を決定するステップと、
前記コンピューティングシステムによって、前記外観の整合性の損失に少なくとも部分的に基づいて前記フレームレベル位置予測モデルを訓練するステップとをさらに含む、
請求項5または6に記載のコンピュータ実装方法。
【請求項８】
前記コンピューティングシステムによって、前記物体を示す画像の前記シーケンスの中の複数の画像にまたがる前記物体に関連する前記カメラ座標空間内の前記座標および前記物体に関連する空間的アテンションに少なくとも部分的に基づいて照準の損失を決定するステップと、
前記コンピューティングシステムによって、前記照準の損失に少なくとも部分的に基づいて前記フレームレベル位置予測モデルを訓練するステップとをさらに含む、
請求項5、6、または7に記載のコンピュータ実装方法。
【請求項９】
前記コンピューティングシステムによって、前記物体に関連する前記実世界の座標と、前記物体を示す画像の前記シーケンスを撮影するために使用されたカメラに関連する視野とに少なくとも部分的に基づいて視野の損失を決定するステップと、
前記コンピューティングシステムによって、前記視野の損失に少なくとも部分的に基づいて前記フレームレベル位置予測モデルを訓練するステップとをさらに含む、
請求項5、6、7、または8に記載のコンピュータ実装方法。
【請求項１０】
画像の前記シーケンスが、画像の前記シーケンスの中の複数の画像にまたがって複数の物体を示し、前記情報抽出モデルの前記出力が、画像の前記シーケンス内に示された前記複数の物体に関連する実世界の位置を表すデータを含む、
請求項1から9のいずれか一項に記載のコンピュータ実装方法。
【請求項１１】
画像のシーケンス内に示された物体に関連する実世界の位置を表すデータを決定するように情報抽出モデルを訓練するためのコンピュータ実装方法であって、前記情報抽出モデルが、
画像特徴抽出モデル、
弱教師あり物体分類モデル、
フレームレベル位置予測モデル、および
ジオロケーション予測モデルを含み、
前記方法が、1つまたは複数のプロセッサを含むコンピューティングシステムにおいて、
ノイズのある分類付きの画像のシーケンスを表すデータを取得するステップであって、画像の前記シーケンスのうちの少なくとも1つの画像が、前記物体を示す、ステップと、
画像の前記シーケンスを入力したことに応じて前記画像特徴抽出モデルによって画像

の特徴のシーケンスを出力するステップと、
画像の特徴の前記シーケンスを入力したことに応じて前記物体分類モデルによって、画像の前記シーケンスに関連する1つまたは複数の分類ラベルを含む分類データを出力するステップであって、前記分類データが、画像の特徴の前記シーケンスに関連する1つまたは複数の時間的アテンション値および1つまたは複数の空間的アテンション値に少なくとも部分的に基づいて決定され、前記1つまたは複数の時間的アテンション値および前記1つまたは複数の空間的アテンション値が、前記物体分類モデルによって決定される、ステップと、
画像の前記シーケンスに関連する前記分類データおよび前記ノイズのある分類に少なくとも部分的に基づいて前記物体分類モデルを訓練するステップと、
画像の特徴の前記シーケンス、前記1つまたは複数の時間的アテンション値、および前記1つまたは複数の空間的アテンション値を入力したことに応じて前記フレームレベル位置予測モデルによって、前記物体に関連するカメラ座標空間内の座標および前記物体に関連するカメラ姿勢データを取得するステップと、
前記ジオロケーション予測モデルによって、画像の特徴の前記シーケンス、前記時間的アテンション値、および前記空間的アテンション値のうちの少なくとも１つ、ならびに前記物体に関連するカメラ座標空間内の座標および前記物体に関連するカメラ姿勢データに基づいて、画像の前記シーケンス内に示された前記物体に関連する実世界の位置を出力するステップと、
前記分類データに少なくとも部分的に基づいて、前記ジオロケーション予測モデルによって出力された前記物体に関連する実世界の位置を前記物体に対応する前記分類ラベルに関連付けるステップと、
少なくとも画像の特徴の前記シーケンス、前記時間的アテンション値、前記空間的アテンション値、および前記分類ラベルを使用して前記ジオロケーション予測モデルを訓練するステップとを含む、
コンピュータ実装方法。
【請求項１２】
画像から情報を抽出するためのコンピュータ実装方法であって、
1つまたは複数のプロセッサを含むコンピューティングシステムにおいて、1つまたは複数の画像を表すデータを取得するステップであって、前記1つまたは複数の画像のうちの少なくとも1つが、物体を示す、ステップと、
前記コンピューティングシステムによって、前記1つまたは複数の画像から位置情報を抽出するように訓練される機械学習された情報抽出モデルに前記1つまたは複数の画像を入力するステップと、
前記1つまたは複数の画像を入力したことに応じて前記情報抽出モデルの出力として、前記コンピューティングシステムによって、前記1つまたは複数の画像内に示された前記物体に関連する実世界の位置を表すデータを取得するステップであって、
前記コンピューティングシステムによって、前記1つまたは複数の画像に示された前記物体に関連する分類を表すデータを決定するステップと、
前記コンピューティングシステムによって、前記1つまたは複数の画像から抽出された1つまたは複数の画像の特徴を表すデータに少なくとも部分的に基づいて前記1つまたは複数の画像に関連する時間的アテンション値および空間的アテンション値を決定するステップと、
前記コンピューティングシステムによって、前記1つまたは複数の画像の特徴、前記時間的アテンション値、および前記空間的アテンション値に基づいて、前記物体に関連するカメラ座標空間内の座標および前記物体に関連するカメラ姿勢データを取得するステップと、
前記コンピューティングシステムによって、前記1つまたは複数の画像の特徴、前記時間的アテンション値、および前記空間的アテンション値のうちの少なくとも１つ、ならびに前記物体に関連するカメラ座標空間内の座標および前記物体に関連するカメラ姿勢データに基づいて前記物体に関連する前記実世界の位置を予測するステップと、

前記コンピューティングシステムによって、前記物体に関連する分類を表す前記決定されたデータに少なくとも部分的に基づいて、前記物体に関連する前記予測された実世界の位置を前記物体に対応する分類ラベルに関連付けるステップとを含む、
ステップとを含む、
コンピュータ実装方法。
【請求項１３】
コンピューティングシステムであって、
1つまたは複数のプロセッサと、
1つまたは複数の機械学習された情報抽出モデルと、
前記1つまたは複数のプロセッサによって実行されるときに前記システムに請求項1から12のうちのいずれか一項に記載の方法を実行させる命令を記憶したコンピュータ可読記憶媒体とを含む、
システム。
【請求項１４】
1つまたは複数の機械学習された情報抽出モデルと、1つまたは複数のプロセッサによって実行されるときに前記1つまたは複数のプロセッサに請求項1から12のいずれか一項に記載の動作を実行させるコンピュータ可読命令とを記憶する1つまたは複数の有形の非一時的コンピュータ可読記憶媒体。
【発明の詳細な説明】
【技術分野】
【０００１】
本開示は、概して、1つまたは複数の物体に関する実世界の位置(たとえば、ジオロケーション)を予測することに関する。より詳細には、本開示は、物体に関する実世界の位置を予測するために教師なしで訓練され得る情報抽出モデルに関する。
【背景技術】
【０００２】
機械学習モデルを使用して画像からデータを抽出することの主なボトルネックのうちの1つは、そのようなモデルを訓練するための十分なデータを得る高いコストである。任意の物体に関する実世界の座標を含むグラウンドトゥルース(ground truth)データを得ることは、非常に時間のかかるプロセスである。ほとんどの場合、実世界の位置を予測するためにニューラルネットワークを訓練するための十分なグラウンドトゥルースデータを得ることは、非現実的である。別の問題は、画像から抽出される必要がある様々な種類のデータが急速に増えることである。
【発明の概要】
【課題を解決するための手段】
【０００３】
本開示の実施形態の態様および利点が、以下の説明に部分的に記載されるか、または説明から知られ得るか、または実施形態の実施を通じて知られ得る。
【０００４】
本開示の1つの例示的な態様は、画像から情報を抽出するためのコンピュータで実行される方法を対象とする。方法は、1つまたは複数のプロセッサを含むコンピューティングシステムにおいて、画像のシーケンスを表すデータを取得するステップであって、画像のシーケンスのうちの少なくとも1つの画像が、物体を示す、ステップを含む。方法は、コンピューティングシステムによって、画像のシーケンスから位置情報を抽出するように訓練される機械学習された情報抽出モデルに画像のシーケンスを入力するステップを含む。方法は、画像のシーケンスを入力したことに応じて情報抽出モデルの出力として、コンピューティングシステムによって、画像のシーケンス内に示された物体に関連する実世界の位置を表すデータを取得するステップを含む。
【０００５】
本開示の別の例示的な態様は、コンピューティングシステムを対象とする。コンピュー

ティングシステムは、1つまたは複数のプロセッサと、1つまたは複数の機械学習された情報抽出モデルと、1つまたは複数のプロセッサによって実行されるときにコンピューティングシステムに動作を実行させる命令を共同で記憶する1つまたは複数の有形の非一時的コンピュータ可読媒体とを含む。動作は、画像のシーケンスを表すデータを取得するステップであって、画像のシーケンスのうちの少なくとも1つの画像が、物体を示す、ステップを含む。動作は、画像のシーケンスから位置情報を抽出するように訓練される機械学習された情報抽出モデルに画像のシーケンスを入力するステップを含む。動作は、画像のシーケンスを入力したことに応じて情報抽出モデルの出力として、画像のシーケンス内に示された物体に関連する実世界の位置を表すデータを取得するステップを含む。

【０００６】

本開示の別の例示的な態様は、1つまたは複数の機械学習された情報抽出モデルと、1つまたは複数のプロセッサによって実行されるときに1つまたは複数のプロセッサに動作を実行させるコンピュータ可読命令とを記憶する1つまたは複数の有形の非一時的コンピュータ可読媒体を対象とする。動作は、画像のシーケンスを表すデータを取得するステップであって、画像のシーケンスのうちの少なくとも1つの画像が、物体を示す、ステップを含む。動作は、画像のシーケンスから位置情報を抽出するように訓練される機械学習された情報抽出モデルに画像のシーケンスを入力するステップを含む。動作は、画像のシーケンスを入力したことに応じて情報抽出モデルの出力として、画像のシーケンス内に示された物体に関連する実世界の位置を表すデータを取得するステップを含む。

【０００７】

本開示のその他の態様は、様々なシステム、装置、非一時的コンピュータ可読媒体、ユーザインターフェース、および電子デバイスを対象とする。

【０００８】

本開示の様々な実施形態のこれらのおよびその他の特徴、態様、および利点は、以下の説明および添付の請求項を参照するとより深く理解されるであろう。本明細書に組み込まれ、本明細書の一部をなす添付の図面は、本開示の例示的な実施形態を示し、説明とともに、関連する原理を説明する働きをする。

【０００９】

添付の図面を参照する当業者に向けた実施形態の詳細な検討が、本明細書に記載されている。

【図面の簡単な説明】

【００１０】

【図１Ａ】本開示の例示的な実施形態による画像のシーケンスから位置情報を抽出することができる例示的なコンピューティングシステム/デバイスのブロック図である。

【図１Ｂ】本開示の例示的な実施形態による画像のシーケンスから位置情報を抽出することができる例示的なコンピューティングシステム/デバイスのブロック図である。

【図１Ｃ】本開示の例示的な実施形態による画像のシーケンスから位置情報を抽出することができる例示的なコンピューティングシステム/デバイスのブロック図である。

【図２】本開示の例示的な実施形態による例示的な情報抽出モデルのブロック図である。

【図３】本開示の例示的な実施形態による例示的な情報抽出モデルのブロック図である。

【図４】本開示の例示的な実施形態による例示的なジオロケーション予測モデルのブロック図である。

【図５】本開示の例示的な実施形態による例示的な物体分類モデルのブロック図である。

【図６】本開示の例示的な実施形態による例示的な位置特徴抽出モデルのブロック図である。

【図７】本開示の例示的な実施形態による例示的な位置予測モデルのブロック図である。

【図８】本開示の例示的な実施形態による画像のシーケンスから位置情報を抽出するための例示的な方法の流れ図である。

【図９】本開示の例示的な実施形態による物体分類モデルを訓練するための例示的な方法の流れ図である。

【図１０】本開示の例示的な実施形態による位置予測モデルを訓練するための例示的な方法の流れ図である。
【発明を実施するための形態】
【００１１】
複数の図で繰り返される参照番号は、様々な実装の同じ特徴を特定するように意図される。
【００１２】
概要
本開示に合致するシステムおよび方法は、複数の画像に示された1つまたは複数の物体(たとえば、道路標識)に関する予測された実世界の位置(たとえば、緯度および経度)を決定するために使用されることが可能な情報抽出モデルを含み得る。情報抽出モデルは、ノイズのある(noisy)分類付きの画像データを使用して1つまたは複数の物体の実世界の位置を予測するように訓練され得る。大量のノイズのあるラベル付きデータおよび極めて大量のラベルなしデータを利用する能力を与えることによって、本開示は、より多くの応用のためにモデルをより速く、より安く開発することを可能にし得る。また、本開示は、十分なグラウンドトゥルース位置データが利用不可能な多くの新しい応用を可能にし得る。さらに、情報抽出モデルが(たとえば、回帰の目標値として)グラウンドトゥルース位置データを必要とせずに訓練され得るので、情報抽出モデルは、物体を分類する弱い教師(weak supervision)のみを用いて、物体の実世界の位置を予測するために教師なしでエンドツーエンドモデルとして訓練され得る。多数の実験が、本開示の情報抽出モデルがうまく機能し、同じ訓練データおよびテストデータセットを使用して、しかし、モデルの訓練にグラウンドトゥルースラベル、物体のバウンディングボックス(object bounding box)、および強い教師(strong supervision)を必要とすることなく、従来の完全な教師ありモデルに匹敵する精度に到達することができることを示した。
【００１３】
本開示の態様によれば、1つまたは複数のプロセッサを含むコンピューティングシステムが、情報抽出モデルを含む開示されるテクノロジーの態様を実装するのを助けるために使用され得る。一部の実装においては、コンピューティングシステムが、画像データを取得する。画像データは、たとえば、画像フレームのシーケンスなどの複数の画像を含み得る。画像フレームのシーケンスは、シーン内の1つまたは複数の物体(たとえば、道路標識)を示し得る。例として、画像フレームのシーケンスの中の1つまたは複数の画像フレームが、1つまたは複数の物体を示し得る。別の例として、画像フレームのシーケンスの中の複数の画像フレームが、同じ物体を示し得る。同じ物体を示す複数の画像フレームは、画像フレームのシーケンスの中の連続した画像フレームまたは連続しない画像フレームであり得る。一部の実装において、画像フレームのシーケンスは、街路を横断する乗り物の視点から街路近辺のシーンを示し得る。一部の実装において、画像フレームのシーケンスの中の1つまたは複数の画像フレームは、ビデオまたはその他の種類のモーションキャプチャ(motion capture)の1つまたは複数のフレームに対応し得る。
【００１４】
一部の実装において、画像データは、画像フレームのシーケンスに関連する分類を含み得る。たとえば、画像データは、画像フレームのシーケンスに関連する単一の分類ラベルを含み得る。代替的に、画像データは、画像フレームのシーケンスの中の1つまたは複数の画像フレームに関連する2つ以上の分類ラベルを含み得る。以下でさらに説明されるように、本開示の態様によれば、システムは、ノイズのある分類(たとえば、画像フレームのシーケンスに関連する単一の分類ラベル)付きの画像フレームのシーケンスに少なくとも部分的に基づいて物体の実世界の位置を表すデータを取得するために情報抽出モデルを使用することができる。
【００１５】
一部の実装において、画像データは、カメラ姿勢データを含み得る。カメラ姿勢データは、画像フレームのシーケンスの中の1つまたは複数の画像フレームを撮影するために使

用されたカメラの実世界の位置および/または向きを表し得る。たとえば、カメラ姿勢データは、4x4のカメラツーワールド(camera-to-world)射影行列を含み得る。

【００１６】

本開示の態様によれば、システムは、ジオロケーションデータを生成し得る。ジオロケーションデータは、画像フレームのシーケンス内に示された1つまたは複数の物体(たとえば、道路標識)に関する予測された実世界の位置(たとえば、緯度および経度)を含み得る。情報抽出モデルは、画像データを受け取り、画像データを受け取ったことに応じてジオロケーションデータを出力するように構成され得る。

【００１７】

システムは、情報抽出モデルに画像データを入力し、画像データを入力したことに応じて情報抽出モデルの出力としてジオロケーションデータを取得することができる。システムは、たとえば、道路標識に対応する道路のセグメント(たとえば、道路標識の緯度および経度にまたはその近くにある道路のセグメント)を特定するためにジオロケーションデータを使用することができる。たとえば、画像データは、速度制限の標識を示す画像フレームのシーケンスを含み得る。システムは、情報抽出モデルに画像データを入力し、情報抽出モデルの出力として速度制限の標識の予測された実世界の位置を含むジオロケーションデータを取得することができる。システムは、速度制限の標識に対応する道路のセグメント(たとえば、速度制限の標識の予測された実世界の座標にまたはその近くにある道路のセグメント)を特定するためにジオロケーションデータを使用することができる。

【００１８】

本開示の態様によれば、情報抽出モデルは、たとえば、画像特徴抽出モデル、物体分類モデル、およびジオロケーション予測モデルなどの複数の機械学習モデルを含み得る。一部の実装において、情報抽出モデルおよび/または情報抽出モデルに含まれる複数の機械学習モデル(たとえば、画像特徴抽出モデル、物体分類モデル、ジオロケーション予測モデルなど)は、ニューラルネットワーク(たとえば、深層ニューラルネットワーク)または非線形モデルおよび/もしくは線形モデルを含むその他の種類の機械学習モデルなどの様々な機械学習モデルであり得るかあるいはそうでなければそのような機械学習モデルを含み得る。ニューラルネットワークは、順伝播型ニューラルネットワーク、再帰型ニューラルネットワーク(たとえば、長期短期記憶再帰型ニューラルネットワーク)、畳み込みニューラルネットワーク、またはその他の形態のニューラルネットワークを含み得る。

【００１９】

一部の実装において、画像特徴抽出モデルは、画像フレームのシーケンスの中の1つまたは複数の画像フレームから抽出された1つまたは複数の画像の特徴を含む画像特徴データを生成し得る。画像特徴抽出モデルは、画像フレームのシーケンスを表すデータ(たとえば、画像データ)を受け取り、画像フレームのシーケンスを受け取ったことに応じて画像特徴データを出力するように構成され得る。以下でさらに説明されるように、本開示の態様によれば、画像特徴データ内の1つまたは複数の画像の特徴が、画像フレームのシーケンス内に示された1つまたは複数の物体を特定するおよび/または分類するために使用され得る。一部の実装において、画像特徴データは、画像の特徴のシーケンスを含み得る。たとえば、画像特徴データ内の1つまたは複数の画像の特徴は、画像特徴埋め込みのシーケンスへと編成され得る。画像特徴埋め込みのシーケンスの中の各画像特徴埋め込みは、画像フレームのシーケンスの中の画像フレームに対応することが可能であり、各画像特徴埋め込みは、対応する画像フレームから抽出された1つまたは複数の画像の特徴を表すことが可能である。一部の実装において、画像特徴抽出モデルは、たとえば、Inception v2、任意のSoTAの画像分類ネットワーク(またはその下部)などの畳み込みニューラルネットワーク(CNN)を含み得る。

【００２０】

システムは、画像特徴抽出モデルに画像フレームのシーケンスを表すデータ(たとえば、画像データ)を入力し、画像フレームのシーケンスを入力したことに応じて画像特徴抽出モデルの出力として画像の特徴のシーケンス(たとえば、画像特徴データ)を取得することが

できる。以下でさらに説明されるように、システムは、画像特徴データに少なくとも部分的に基づいてジオロケーションデータを決定するために情報抽出モデルを使用することができる。

【００２１】

一部の実装において、物体分類モデルは、分類データおよびアテンション(attention)値データを生成することができる。物体分類モデルは、画像の特徴のシーケンスを表すデータ(たとえば、画像特徴抽出モデルによって出力された画像特徴データ)を受け取り、画像の特徴のシーケンスおよび関連する画像特徴埋め込みを受け取ったことに応じて分類データおよびアテンション値データを出力するように構成され得る。一部の実装において、物体分類モデルは、弱教師あり(weakly supervised)再帰型ニューラルネットワーク(RNN)を含み得る。

【００２２】

分類データは、画像フレームのシーケンス内に示された1つまたは複数の物体に関連する分類を表し得る。たとえば、物体分類モデルは、画像の特徴のシーケンスに少なくとも部分的に基づいて1つまたは複数の物体を特定し、1つまたは複数の特定された物体に関連する1つまたは複数の分類ラベルを決定することができる。1つまたは複数の物体は、画像フレームのシーケンスの中の画像フレームの一部またはすべてに示され得る。別の例として、物体分類モデルは、速度制限の標識を示す画像フレームのシーケンスを含む画像データから抽出された画像の特徴のシーケンスを表すデータを受け取り得る。画像の特徴のシーケンスを受け取ったことに応じて、物体分類モデルは、速度制限の標識に対応する速度制限の値を示す分類ラベルを含む分類データを出力することができる。

【００２３】

アテンション値データは、たとえば、分類された物体が特定のフレーム内の特定のピクセルにある確率を表し得る。アテンション値データは、画像の特徴のシーケンスに関連する1つまたは複数の時間的アテンション値および1つまたは複数の空間的アテンション値を含み得る。例として、画像特徴データは、画像の特徴のシーケンスを表す画像特徴埋め込みのシーケンスを含み得る。物体分類モデルは、画像特徴埋め込みのシーケンスの中の各画像特徴埋め込みに関する時間的アテンション値および空間的アテンション値を決定し得る。各画像特徴埋め込みに関する時間的アテンション値および空間的アテンション値は、分類された物体が画像特徴埋め込みに対応する画像フレーム内の特定のピクセルにある確率を表し得る。追加的にまたは代替的に、物体分類モデルは、画像の特徴のシーケンス(たとえば、画像特徴埋め込みのシーケンス)に関して単一の時間的アテンション値および単一の空間的アテンション値を決定し得る。

【００２４】

システムは、物体分類モデルに画像の特徴のシーケンスを表すデータ(たとえば、画像特徴データ)を入力し、画像の特徴のシーケンスを入力したことに応じて物体分類モデルの出力として分類データおよびアテンション値データを取得することができる。以下でさらに説明されるように、システムは、分類データおよびアテンション値データに少なくとも部分的に基づいてジオロケーションデータを決定するために情報抽出モデルを使用することができる。

【００２５】

一部の実装において、物体分類モデルは、時空間アテンションメカニズムを伴う長期短期記憶(LSTM)を含み得る。時空間アテンションメカニズムは、アテンション値データを決定し、物体分類モデルが弱い教師のみを使用して効果的に訓練されることを可能にするために使用され得る。たとえば、物体分類モデルは、物体分類モデルに入力された画像の特徴のシーケンス(たとえば、画像特徴データ)に基づいてフレーム毎の埋め込みをそれぞれ出力する複数のLSTMブロックを含み得る。物体分類モデルは、物体分類モデルの出力を決定するためにLSTMブロックによって生成されたフレーム毎の埋め込みを重み付けするために時間的アテンションを使用することができる。このようにして、出力からの勾配(gradient)が、同じ時間ステップにおいて、一斉に時間的アテンションの対応する重みに

比例して各LSTMブロックの間に分散される。

【００２６】

一部の実装において、物体分類モデルは、物体分類モデルによって出力された分類データに関連する損失(loss)に少なくとも部分的に基づいて訓練され得る。たとえば、システムは、分類データ内の1つまたは複数の分類ラベルおよび画像データ内の画像フレームのシーケンスに関連する分類(たとえば、画像フレームのシーケンスに関連する単一の分類ラベル)に少なくとも部分的に基づいてソフトマックス交差エントロピー誤差(softmax cross entropy loss)を決定することができる。システムは、物体分類モデルを訓練するために決定されたソフトマックス交差エントロピーを使用することができる。

【００２７】

一部の実装において、情報抽出モデルは、ジオロケーションデータを生成することができるジオロケーション予測モデルを含み得る。ジオロケーション予測モデルは、画像の特徴のシーケンスを表すデータ(たとえば、画像特徴抽出モデルによって出力された画像特徴データ)、画像の特徴のシーケンスに対応する画像フレームのシーケンスを撮影するために使用された1つまたは複数のカメラに関連する位置および/または向きを表すデータ(たとえば、画像データ内のカメラ姿勢データ)、ならびに画像の特徴のシーケンスに関連するアテンション値を表すデータ(たとえば、物体分類モデルによって生成されたアテンション値データ)を受け取るように構成され得る。ジオロケーション予測モデルは、画像の特徴のシーケンスを受け取ったことに応じてジオロケーションデータ、カメラの位置および/または向きの情報、ならびに画像の特徴のシーケンスに関連するアテンション値を出力するように構成され得る。システムは、画像特徴データ、カメラ姿勢データ、およびアテンション値データをジオロケーション予測モデルに入力し、画像特徴データ、カメラ姿勢データ、およびアテンション値データを入力したことに応じてジオロケーション予測モデルの出力としてジオロケーションデータを取得することができる。

【００２８】

一部の実装において、ジオロケーション予測モデルは、1つまたは複数の分類された物体の各々に関連する単一の埋め込みベクトルを生成し得る。たとえば、単一の埋め込みベクトルは、画像の特徴のシーケンスからの関連する分類された物体に関するすべてのデータを符号化し得る。ジオロケーション予測モデルは、関連する分類された物体に関連する実世界の位置を予測するために単一の埋め込みベクトルを使用することができる。

【００２９】

一部の実装において、ジオロケーション予測モデルは、フレームレベル位置特徴抽出モデルおよびフレームレベル位置予測モデルを含み得る。フレームレベル位置特徴抽出モデルは、画像の特徴のシーケンスを表すデータ(たとえば、画像特徴抽出モデルによって出力された画像特徴データ)を受け取り、画像の特徴のシーケンスを受け取ったことに応じて1つまたは複数の分類された物体に関連する1つまたは複数の位置の特徴を含む位置特徴データを出力するように構成され得る。下でさらに説明されるように、本開示の態様によれば、位置特徴データ内の1つまたは複数の位置の特徴は、1つまたは複数の分類された物体に関する実世界の位置を予測するために使用され得る。一部の実装において、位置特徴データは、位置の特徴のシーケンスを含み得る。たとえば、位置特徴データ内の1つまたは複数の位置の特徴は、位置特徴埋め込みのシーケンスへと編成され得る。位置特徴埋め込みのシーケンスの中の各位置特徴埋め込みは、画像フレームのシーケンスの中の画像フレームに対応することが可能であり、各位置特徴埋め込みは、対応する画像フレーム内に示された1つまたは複数の分類された物体に関連する1つまたは複数の位置の特徴を表すことが可能である。

【００３０】

システムは、画像特徴データをフレームレベル位置特徴抽出モデルに入力し、フレームレベル位置特徴抽出モデルの出力として位置特徴データを取得することができる。たとえば、画像特徴データは、画像の特徴のシーケンスを表し、画像フレームのシーケンスに対応する画像特徴埋め込みのシーケンスを含み得る。システムは、画像フレームに関連する

画像の特徴を表すデータ(たとえば、画像特徴埋め込みのシーケンスの中の画像特徴埋め込み)をフレームレベル位置特徴抽出モデルに入力し、フレームレベル位置特徴抽出モデルの出力として画像フレームに関連する(たとえば、画像フレーム内に示された1つまたは複数の分類された物体に関連する) 1つまたは複数の位置の特徴を表す位置特徴埋め込みを含む位置特徴データを取得することができる。このようにして、システムは、画像特徴埋め込みのシーケンスの中の各画像特徴埋め込みをフレームレベル位置特徴抽出モデルに入力し、位置の特徴のシーケンスを表し、画像フレームのシーケンスに対応する位置特徴埋め込みのシーケンスを含む位置特徴データを取得することができる。

【００３１】

フレームレベル位置予測モデルは、位置の特徴のシーケンスを表すデータ(たとえば、フレームレベル位置特徴抽出モデルによって出力された位置特徴データ)を受け取り、位置の特徴のシーケンスを受け取ったことに応じて1つまたは複数の分類された物体に関連する座標を含む座標データを出力するように構成され得る。一部の実装において、座標データは、画像フレームのシーケンスに対応する座標埋め込みのシーケンスを含み得る。各座標埋め込みは、対応する画像のフレーム内に示された1つまたは複数の分類された物体に関連する座標を表し得る。画像フレーム内に示された分類された物体に関連する座標は、画像フレームに関連するカメラ座標空間内の分類された物体の三次元位置を示し得る。

【００３２】

システムは、位置特徴データをフレームレベル位置予測モデルに入力し、フレームレベル位置予測モデルの出力として座標データを取得することができる。たとえば、位置特徴データは、位置の特徴のシーケンスを表し、画像フレームのシーケンスに対応する位置特徴埋め込みのシーケンスを含み得る。システムは、画像フレームに関連する位置の特徴を表すデータ(たとえば、位置特徴埋め込みのシーケンスの中の位置特徴埋め込み)をフレームレベル位置予測モデルに入力し、フレームレベル位置予測モデルの出力として画像フレーム内に示された1つまたは複数の分類された物体に関連する座標を表す座標埋め込みを含む座標データを取得することができる。このようにして、システムは、位置特徴埋め込みのシーケンスの中の各位置特徴埋め込みをフレームレベル位置予測モデルに入力し、座標のシーケンスを表し、画像フレームのシーケンスに対応する座標埋め込みのシーケンスを含む座標データを取得することができる。

【００３３】

ジオロケーション予測モデルは、フレームレベル位置予測モデルによって出力された座標データおよびカメラ姿勢データに少なくとも部分的に基づいて1つまたは複数の分類された物体に関連する予測された実世界の位置を決定するように構成され得る。一部の実装において、ジオロケーション予測モデルは、カメラ座標空間からの座標値データ内の実世界内の分類された物体に関連する座標を実世界の座標(たとえば、緯度および経度)に変換することによって1つまたは複数の分類された物体に関する予測された実世界の位置を決定するように構成され得る。たとえば、ジオロケーション予測モデルは、分類された物体が画像フレームのシーケンスの中の複数の画像フレーム内に示されると判定し得る。ジオロケーション予測モデルは、座標データに少なくとも部分的に基づいて複数の画像フレームの各々に関して分類された物体に関連する座標を取得し、カメラ姿勢データに少なくとも部分的に基づいて分類された物体に関連する複数の座標を実世界の座標(たとえば、緯度および経度)に変換することができる。たとえば、ジオロケーション予測モデルは、画像フレームを撮影するために使用されたカメラの位置および/または向きに基づいて画像フレームに関連するカメラ座標空間内の分類された物体の三次元位置を分類された物体の実世界の座標に変換することができる。このようにして、システムは、分類された物体の予測された実世界の位置を決定するために、分類された物体に関連する複数の実世界の座標の時間的な加重平均を決定することができる。

【００３４】

一部の実装において、ジオロケーション予測モデルは、フレームレベル位置予測モデルによって出力された座標データを確認し、確認に基づいて予測された実世界の座標を決定

するように構成され得る。例として、ジオロケーション予測モデルは、特定された物体に関連する座標が正確であると確認し得る。別の例として、ジオロケーション予測モデルは、複数の画像フレームにまたがる特定された物体に関連する座標が同じ特定された物体に対応すると確認し得る。

【００３５】

一部の実装において、ジオロケーション予測モデルは、予測された実世界の位置が正確であり、関心のある分類された物体に対応することを確かめるために複数の損失値のうちの1つまたは複数に少なくとも部分的に基づいて訓練され得る。複数の損失値は、位置の整合性の損失(location consistency loss)、外観の整合性の損失(appearance consistency loss)、照準の損失(aiming loss)、および視野(FOV)の損失(field-of-view (FOV) loss)を含み得る。例として、システムは、複数の画像フレームにまたがる特定された物体に関連する座標の間の分散(variance)に少なくとも部分的に基づいて位置の整合性の損失を決定し得る。システムは、ジオロケーション予測モデルによって決定された座標が分類された物体に関する複数の画像フレームにまたがって整合性があるようにジオロケーション予測モデルを訓練するために決定された位置の整合性の損失を使用することができる。

【００３６】

別の例として、システムは、(たとえば、画像特徴抽出モデルによって出力された)画像特徴データおよび(たとえば、物体分類モデルによって出力された)アテンション値データに少なくとも部分的に基づいて外観の整合性の損失を決定し得る。特に、システムは、複数の画像フレームに関する外観の特徴を決定するためにアテンション値データに含まれる空間的アテンション値によって画像フレームに対応する画像の特徴を重み付けすることができ、システムは、複数の画像フレームにまたがる決定された外観の特徴の間の分散に少なくとも部分的に基づいて外観の整合性の損失を決定することができる。システムは、ジオロケーション予測モデルによって分類される1つまたは複数の物体が、物体が見える各画像フレームにおいて類似した視覚的外観を有するように、ジオロケーション予測モデルを訓練するために決定された外観の整合性の損失を使用することができる。

【００３７】

別の例として、システムは、(たとえば、フレームレベル位置予測モデルによって出力された)座標データおよび(たとえば、物体分類モデルによって出力された)アテンション値データに少なくとも部分的に基づいて照準の損失を決定し得る。システムは、画像フレーム内に示された分類された物体に関連する座標データ内の座標が、画像フレームに関連するカメラ座標空間内で、分類された物体に関連する空間的アテンションが最も高いエリア内に射影されるようにジオロケーション予測モデルを訓練するために照準の損失を使用することができる。

【００３８】

別の例として、システムは、予測された実世界の座標を、予測された実世界の座標がそれらに基づいて決定される画像フレームを撮影するために使用されたカメラの実際の可能なFOV内に制約するためにFOVの損失を決定することができる。システムは、予測された実世界の座標の範囲に対する意味のある制限(たとえば、妥当な空間)を含めるためにジオロケーション予測モデルを訓練するために決定されたFOVの損失を使用することができる。

【００３９】

本明細書において説明されるシステムおよび方法は、いくつかの技術的効果および利点を提供する可能性がある。たとえば、コンピューティングシステムは、ノイズのある分類付きの画像を表すデータから位置情報を抽出することが可能な1つまたは複数の情報抽出モデルを含み得る。情報抽出モデルは、ノイズのある分類付きの画像内に示された1つまたは複数の物体の実世界の位置を予測するためにエンドツーエンドで訓練され得る。たとえば、情報抽出モデルは、様々な種類の道路標識、家屋番号(house number)、右/左折禁止規制(turn restriction)、街路名などの実世界の位置を予測するために使用され得る。より詳細には、情報抽出モデルは、物体分類モデルおよび/またはジオロケーション予測モ

デルを含み得る。物体分類モデルは、教師信号(supervision signal)として弱い分類ラベル付きの画像データ(たとえば、画像フレームのシーケンスに関連する単一の分類ラベル)を使用して訓練されることが可能であり、ジオロケーション予測モデルは、(たとえば、決定された位置の整合性の損失、外観の整合性の損失、照準の損失、および/またはFOVの損失に基づいて)教師なしで訓練され得る。大量のノイズのあるラベル付きデータおよび極めて大量のラベルなしデータを利用する能力を与えることによって、本開示は、より多くの応用のためにモデルをより速く、より安く開発することを可能にし得る。さらに、情報抽出モデルによって抽出された位置情報が、十分なグラウンドトゥルース位置データが以前利用不可能であった多くの新しい応用の開発を可能にするために使用され得る。たとえば、情報抽出モデルは、乗り物に搭載されたコンピューティングシステムに、またはドライブレコーダ(dashcam)アプリケーションの一部として含まれることが可能であり、オフライン処理のためにデータを送信する必要なしに実世界の物体を検出するために使用され得る。さらに、情報抽出モデルの1つまたは複数の構成要素(たとえば、画像特徴抽出モデル、物体分類モデル、ジオロケーション予測モデルなど)は、1つまたは複数のその他の機械学習モデルに統合され得る。たとえば、時空間アテンションを用いる物体分類モデルは、(たとえば、動画共有プラットフォームにおいて暴力的なまたは攻撃的なコンテンツを分類するために)ノイズのある分類付きの様々なビデオコンテンツに関連する分類を決定するために使用され得る。

【００４０】

例示的なデバイスおよびシステム

図1Aは、本開示の例示的な実施形態による情報抽出を実行する例示的なジオロケーションシステム100のブロック図を示す。特に、ジオロケーションシステム100は、複数の画像内に示される1つまたは複数の物体に関する実世界の位置を予測し得る。ジオロケーションシステム100は、ネットワーク180を介して通信可能なように結合されるユーザコンピューティングデバイス102、サーバコンピューティングシステム130、および訓練コンピューティングシステム150を含むコンピューティングシステムに対応し得る。

【００４１】

ユーザコンピューティングデバイス102は、たとえば、パーソナルコンピューティングデバイス(たとえば、ラップトップもしくはデスクトップ)、モバイルコンピューティングデバイス(たとえば、スマートフォンもしくはタブレット)、ゲームコンソールもしくはコントローラ、ウェアラブルコンピューティングデバイス、組み込みコンピューティングデバイス、または任意のその他の種類のコンピューティングデバイスなどの任意の種類のコンピューティングデバイスであることが可能である。

【００４２】

ユーザコンピューティングデバイス102は、1つまたは複数のプロセッサ112およびメモリ114を含む。1つまたは複数のプロセッサ112は、任意の好適な処理デバイス(たとえば、プロセッサコア、マイクロプロセッサ、ASIC、FPGA、コントローラ、マイクロコントローラなど)であることが可能であり、1つのプロセッサまたは動作可能なように接続される複数のプロセッサであることが可能である。メモリ114は、RAM、ROM、EEPROM、EPROM、フラッシュメモリデバイス、磁気ディスクなど、およびこれらの組合せなどの1つまたは複数の非一時的コンピュータ可読ストレージ媒体を含み得る。メモリ114は、データ116と、ユーザコンピューティングデバイス102に動作を実行させるためにプロセッサ112によって実行される命令118とを記憶することができる。

【００４３】

一部の実装において、ユーザコンピューティングデバイス102は、1つまたは複数の情報抽出モデル120を記憶するかまたは含むことができる。たとえば、情報抽出モデル120は、ニューラルネットワーク(たとえば、深層ニューラルネットワーク)、または非線形のモデルおよび/もしくは線形モデルを含むその他の種類の機械学習モデルなどの様々な機械学習モデルであることが可能であるかまたはそうでなければそのような機械学習モデルを含むことが可能である。ニューラルネットワークは、順伝播型ニューラルネットワーク、

再帰型ニューラルネットワーク(たとえば、長期短期記憶再帰型ニューラルネットワーク)、畳み込みニューラルネットワーク、またはその他の形態のニューラルネットワークを含み得る。例示的な情報抽出モデル120が、図2～図7を参照して検討される。

【００４４】

一部の実装において、1つまたは複数の情報抽出モデル120は、ネットワーク180を介してサーバコンピューティングシステム130から受信され、ユーザコンピューティングデバイスのメモリ114に記憶され、それから、1つまたは複数のプロセッサ112によって使用されるかまたはそうでなければ実施されることが可能である。一部の実装において、ユーザコンピューティングデバイス102は、(たとえば、複数のインスタンスにまたがって並列的な情報抽出を実行するために)単一の情報抽出モデル120の複数の並列的なインスタンスを実施することができる。

【００４５】

より詳細には、情報抽出モデル120は、画像データを受け取り、画像データを受け取ったことに応じてジオロケーションデータを出力するように構成され得る。ジオロケーションシステム100は、情報抽出モデル120に画像データを入力し、画像データを入力したことに応じて情報抽出モデル120の出力としてジオロケーションデータを取得することができる。

【００４６】

追加的にまたは代替的に、1つまたは複数の情報抽出モデル140は、クライアント-サーバの関係によりユーザコンピューティングデバイス102と通信するサーバコンピューティングシステム130に含まれるかまたはそうでなければ記憶され、実施されることが可能である。たとえば、情報抽出モデル140は、ウェブサービス(たとえば、ジオロケーション情報抽出サービス)の一部としてサーバコンピューティングシステム130によって実施され得る。したがって、1つまたは複数のモデル120が、ユーザコンピューティングデバイス102に記憶され、実施されることが可能であり、および/または1つもしくは複数のモデル140が、サーバコンピューティングシステム130に記憶され、実施されることが可能である。

【００４７】

ユーザコンピューティングデバイス102は、ユーザ入力を受け取る1つまたは複数のユーザ入力構成要素122も含み得る。たとえば、ユーザ入力構成要素122は、ユーザ入力オブジェクト(たとえば、指またはスタイラス)のタッチを感知可能であるタッチ感知式構成要素(たとえば、タッチ式ディスプレイスクリーンまたはタッチパッド)であることが可能である。タッチ感知式構成要素は、仮想キーボードを実施するように働き得る。その他の例示的なユーザ入力構成要素は、マイクロフォン、通常のキーボード、またはユーザがユーザ入力を与えることができるその他の手段を含む。

【００４８】

サーバコンピューティングシステム130は、1つまたは複数のプロセッサ132およびメモリ134を含む。1つまたは複数のプロセッサ132は、任意の好適な処理デバイス(たとえば、プロセッサコア、マイクロプロセッサ、ASIC、FPGA、コントローラ、マイクロコントローラなど)であることが可能であり、1つのプロセッサまたは動作可能なように接続される複数のプロセッサであることが可能である。メモリ134は、RAM、ROM、EEPROM、EPROM、フラッシュメモリデバイス、磁気ディスクなど、およびこれらの組合せなどの1つまたは複数の非一時的コンピュータ可読ストレージ媒体を含み得る。メモリ134は、データ136と、サーバコンピューティングシステム130に動作を実行させるプロセッサ132によって実行される命令138とを記憶することができる。

【００４９】

一部の実装において、サーバコンピューティングシステム130は、1つもしくは複数のサーバコンピューティングデバイスを含むか、またはそうでなければ1つもしくは複数のサーバコンピューティングデバイスによって実装される。サーバコンピューティングシステム130が複数のサーバコンピューティングデバイスを含む場合、そのようなサーバコンピューティングデバイスは、逐次コンピューティングアーキテクチャ、並列コンピューテ

ィングアーキテクチャ、またはこれらの何らかの組合せによって動作し得る。
【００５０】
上述のように、サーバコンピューティングシステム130は1つまたは複数の機械学習された情報抽出モデル140を記憶するかまたはそうでなければ含むことが可能である。たとえば、モデル140は、様々な機械学習モデルであることが可能であり、またはそうでなければそのような機械学習モデルを含むことが可能である。例示的な機械学習モデルは、ニューラルネットワークまたはその他の多層非線形モデルを含む。例示的なニューラルネットワーク、順伝播型ニューラルネットワーク、深層ニューラルネットワーク、再帰型ニューラルネットワーク、および畳み込みニューラルネットワークを含む。例示的なモデル140が、図2～図7を参照して検討される。
【００５１】
ユーザコンピューティングデバイス102および/またはサーバコンピューティングシステム130は、ネットワーク180を介して通信可能なように結合される訓練コンピューティングシステム150とのインタラクションによってモデル120および/または140を訓練することができる。訓練コンピューティングシステム150は、サーバコンピューティングシステム130と別れていることが可能であり、またはサーバコンピューティングシステム130の一部であることが可能である。
【００５２】
訓練コンピューティングシステム150は、1つまたは複数のプロセッサ152およびメモリ154を含む。1つまたは複数のプロセッサ152は、任意の好適な処理デバイス(たとえば、プロセッサコア、マイクロプロセッサ、ASIC、FPGA、コントローラ、マイクロコントローラなど)であることが可能であり、1つのプロセッサまたは動作可能なように接続される複数のプロセッサであることが可能である。メモリ154は、RAM、ROM、EEPROM、EPROM、フラッシュメモリデバイス、磁気ディスクなど、およびこれらの組合せなどの1つまたは複数の非一時的コンピュータ可読ストレージ媒体を含み得る。メモリ154は、データ156と、訓練コンピューティングシステム150に動作を実行させるプロセッサ152によって実行される命令158とを記憶することができる。一部の実装において、訓練コンピューティングシステム150は、1つもしくは複数のサーバコンピューティングデバイスを含むか、またはそうでなければ1つもしくは複数のサーバコンピューティングデバイスによって実装される。
【００５３】
訓練コンピューティングシステム150は、たとえば、誤差逆伝播法などの様々な訓練または学習技術を使用してユーザコンピューティングデバイス102および/またはサーバコンピューティングシステム130に記憶された機械学習モデル120および/または140を訓練するモデルトレーナ160を含み得る。一部の実装において、誤差逆伝播法を実行することは、打ち切り型通時的逆伝播(truncated backpropagation through time)を実行することを含み得る。モデルトレーナ160は、訓練されているモデルの汎化能力を高めるためにいくつかの汎化技術(たとえば、重み減衰、ドロップアウトなど)を実行することができる。
【００５４】
特に、モデルトレーナ160は、訓練データ162のセットに基づいて情報抽出モデル120および/または140を訓練することができる。例として、訓練データ162は、画像フレームのシーケンスに関連する単一の分類ラベルを含む画像データなどの弱く分類された画像データを含み得る。モデルトレーナ160は、弱い分類付きの画像データを使用することによって情報抽出モデル120および/または140に含まれる物体分類モデルを訓練することができる。別の例として、訓練データ162は、画像フレームのシーケンスに関連する2つ以上の分類ラベルを含むことが可能であり、モデルトレーナ160は、2つ以上の分類ラベル付きの画像データを使用することによって情報抽出モデル120および/または140に含まれる物体分類モデルを訓練することができる。別の例として、訓練データ162は、情報抽出モデル120および/または140に入力として提供されるデータと、入力データに応じて情報抽出モデル120および/または140の出力として提供されるデータとを含み得る。モデルトレ

ーナ160は、入力データおよび出力データを使用することによって教師なしで情報抽出モデル120および/または140に含まれるジオロケーション予測モデルを訓練することができる。

【００５５】

一部の実装において、ユーザが同意を与えた場合、訓練例は、ユーザコンピューティングデバイス102によって提供され得る。したがって、そのような実装において、ユーザコンピューティングデバイス102に提供されるモデル120は、ユーザコンピューティングデバイス102から受信されたユーザに固有のデータに対して訓練コンピューティングシステム150によって訓練され得る。場合によっては、このプロセスは、モデルのパーソナライズと呼ばれ得る。

【００５６】

モデルトレーナ160は、所望の機能を提供するために利用されるコンピュータ論理を含む。モデルトレーナ160は、ハードウェア、ファームウェア、および/または汎用プロセッサを制御するソフトウェアに実装され得る。たとえば、一部の実装において、モデルトレーナ160は、ストレージデバイスに記憶され、メモリにロードされ、1つまたは複数のプロセッサによって実行されるプログラムファイルを含む。その他の実装において、モデルトレーナ160は、RAM、ハードディスク、または光学式もしくは磁気式媒体などの有形のコンピュータ可読ストレージ媒体に記憶されるコンピュータが実行可能な命令の1つまたは複数のセットを含む。

【００５７】

ネットワーク180は、ローカルエリアネットワーク(たとえば、イントラネット)、広域ネットワーク(たとえば、インターネット)、またはこれらの何らかの組合せなどの任意の種類の通信ネットワークであることが可能であり、任意の数の有線またはワイヤレスリンクを含むことが可能である。概して、ネットワーク180を介した通信は、多種多様な通信プロトコル(たとえば、TCP/IP、HTTP、SMTP、FTP)、符号化もしくはフォーマット(たとえば、HTML、XML)、および/または保護方式(たとえば、VPN、セキュアHTTP、SSL)を使用して任意の種類の有線および/またはワイヤレス接続を介して運ばれ得る。

【００５８】

図1Aは、本開示を実施するために使用され得る1つの例示的なコンピューティングシステムを示す。その他のコンピューティングシステムも、使用され得る。たとえば、一部の実装においては、ユーザコンピューティングデバイス102が、モデルトレーナ160および訓練データセット162を含み得る。そのような実装において、モデル120は、ユーザコンピューティングデバイス102のローカルで訓練されかつ使用されることが可能である。そのような実装の一部において、ユーザコンピューティングデバイス102は、ユーザに固有のデータに基づいてモデル120をパーソナライズするためにモデルトレーナ160を実装し得る。

【００５９】

図1Bは、本開示の例示的な実施形態による情報抽出を実行する例示的なコンピューティングデバイス10のブロック図を示す。コンピューティングデバイス10は、ユーザコンピューティングデバイスまたはサーバコンピューティングデバイスであることが可能である。

【００６０】

コンピューティングデバイス10は、いくつかのアプリケーション(たとえば、アプリケーション1からN)を含む。各アプリケーションは、独自の機械学習ライブラリおよび機械学習モデルを含む。たとえば、各アプリケーションは、機械学習モデルを含み得る。例示的なアプリケーションは、テキストメッセージングアプリケーション、電子メールアプリケーション、ディクテーションアプリケーション、仮想キーボードアプリケーション、ブラウザアプリケーションなどを含む。

【００６１】

図1Bに示されるように、各アプリケーションは、たとえば、1つもしくは複数のセンサ、コンテキストマネージャ(context manager)、デバイス状態構成要素、および/または

追加的な構成要素などのコンピューティングデバイスのいくつかのその他の構成要素と通信することができる。一部の実装において、各アプリケーションは、API (たとえば、パブリックAPI)を使用してそれぞれのデバイスの構成要素と通信することができる。一部の実装において、各アプリケーションによって使用されるAPIは、そのアプリケーションに固有である。

【００６２】

図1Cは、本開示の例示的な実施形態による情報抽出を実行する例示的なコンピューティングデバイス50のブロック図を示す。コンピューティングデバイス50は、ユーザコンピューティングデバイスまたはサーバコンピューティングデバイスであることが可能である。

【００６３】

コンピューティングデバイス50は、いくつかのアプリケーション(たとえば、アプリケーション1からN)を含む。各アプリケーションは、中央インテリジェンス層(central intelligence layer)と通信する。例示的なアプリケーションは、テキストメッセージングアプリケーション、電子メールアプリケーション、ディクテーションアプリケーション、仮想キーボードアプリケーション、ブラウザアプリケーションなどを含む。一部の実装において、各アプリケーションは、API (たとえば、すべてのアプリケーションにまたがる共通のAPI)を使用して中央インテリジェンス層(およびそこに記憶されたモデル)と通信し得る。

【００６４】

中央インテリジェンス層は、いくつかの機械学習モデルを含む。たとえば、図1Cに示されるように、それぞれの機械学習モデル(たとえば、モデル)が、各アプリケーションのために提供され、中央インテリジェンス層によって管理され得る。その他の実装においては、2つ以上のアプリケーションが、単一の機械学習モデルを共有し得る。たとえば、一部の実装において、中央インテリジェンス層は、アプリケーションのすべてのために単一のモデル(たとえば、単一モデル)を提供し得る。一部の実装において、中央インテリジェンス層は、コンピューティングデバイス50のオペレーティングシステムに含まれるかまたはそうでなければコンピューティングデバイス50のオペレーティングシステムによって実装される。

【００６５】

中央インテリジェンス層は、中央デバイスデータ層(central device data layer)と通信することができる。中央デバイスデータ層は、コンピューティングデバイス50のためのデータの集中化されたリポジトリであることが可能である。図1Cに示されるように、中央デバイスデータ層は、たとえば、1つもしくは複数のセンサ、コンテキストマネージャ、デバイス状態構成要素、および/または追加的な構成要素などのコンピューティングデバイスのいくつかのその他の構成要素と通信することができる。一部の実装において、中央デバイスデータ層は、API (たとえば、プライベートAPI)を使用してそれぞれのデバイスの構成要素と通信することができる。

【００６６】

図2は、本開示の例示的な実施形態による例示的な情報抽出モデル200のブロック図を示す。一部の実装において、情報抽出モデル200は、たとえば、画像フレームのシーケンスなどの複数の画像を示す入力データ204 (たとえば、画像データ)のセットを受け取り、入力データ204の受け取りの結果として、画像フレームのシーケンス内に示された1つまたは複数の物体に関する予測された実世界の位置を含む出力データ206 (たとえば、ジオロケーションデータ)を提供するように訓練される。一部の実装において、入力データ204は、画像フレームのシーケンスの中の1つまたは複数の画像フレームを撮影するために使用されたカメラの実世界の位置および/または向きを示すカメラ姿勢データを含み得る。

【００６７】

図3は、本開示の例示的な実施形態による例示的な情報抽出モデル300のブロック図を示す。情報抽出モデル300は、情報抽出モデル300が画像特徴抽出モデル302、物体分類モデル306、およびジオロケーション予測モデル310をさらに含むことを除いて図2の情報抽出モデル200と同様である。

【００６８】
一部の実装において、画像特徴抽出モデル302は、入力データ204またはその一部(たとえば、画像フレームのシーケンスを表すデータ)を受け取り、入力データ204の受け取りの結果として、入力データ204内の1つまたは複数の画像フレームから抽出された1つまたは複数の画像の特徴を含む画像特徴データ304を提供するように訓練される。一部の実装において、画像特徴データ304は、1つまたは複数の画像フレームから抽出された画像の特徴のシーケンスを表す画像特徴埋め込みのシーケンスを含み得る。
【００６９】
一部の実装において、物体分類モデル306は、画像特徴データ304 (たとえば、画像の特徴のシーケンスを表すデータ)を受け取り、画像特徴データ304の受け取りの結果として、アテンション値データ308および分類データ309を提供するように訓練される。アテンション値データ308は、画像特徴データ304内の画像の特徴のシーケンスに関連する1つまたは複数の時間的アテンション値および1つまたは複数の空間的アテンション値を含み得る。分類データ309は、入力データ204の1つまたは複数の画像フレーム内に示された1つまたは複数の分類された物体に関連する1つまたは複数の分類ラベル(たとえば、画像フレーム内に示された速度制限の標識に対応する速度制限値)を含み得る。一部の実装において、アテンション値データ308は、1つまたは複数の時間的アテンション値を表す時間的アテンションデータ(たとえば、図5に示される時間的アテンションデータ504)および1つまたは複数の空間的アテンション値を表す空間的アテンションデータ(たとえば、図5に示される空間的アテンションデータ506)を含み得る。
【００７０】
一部の実装において、ジオロケーション予測モデル310は、画像特徴データ304 (たとえば、画像の特徴のシーケンスを表すデータ)、入力データ204またはその一部(たとえば、カメラ姿勢データを表すデータ)、ならびにアテンション値データ308を受け取り、データの受け取りの結果として、入力データ204内に示された(たとえば、入力データ204内の1つまたは複数の画像フレーム内に示された) 1つまたは複数の物体に関する予測された実世界の位置を含む出力データ206 (たとえば、ジオロケーションデータ)を提供するように訓練される。情報抽出モデル300は、分類データ309に少なくとも部分的に基づいて物体に関する予測された実世界の位置を物体に対応する分類ラベルに関連付けることができる。
【００７１】
図4は、本開示の例示的な実施形態による例示的なジオロケーション予測モデル400のブロック図を示す。ジオロケーション予測モデル400は、ジオロケーション予測モデル400が位置特徴抽出モデル402、位置予測モデル406、および座標変換モデル410をさらに含むことを除いて図3のジオロケーション予測モデル310と同様である。
【００７２】
一部の実装において、位置特徴抽出モデル402は、画像特徴データ304およびアテンション値データ308を受け取り、データの受け取りの結果として、入力データ204内に示された1つまたは複数の分類された物体に関連する1つまたは複数の位置の特徴を含む位置特徴データ404を提供するように訓練される。一部の実装において、位置特徴データは、入力データ204内の画像フレームのシーケンスに対応する位置特徴埋め込みのシーケンスを含み得る。位置特徴埋め込みのシーケンスの中の各位置特徴埋め込みは、対応する画像フレーム内に示された1つまたは複数の分類された物体に関連する位置の特徴を表すことが可能である。
【００７３】
一部の実装において、位置予測モデル406は、位置特徴データ404を受け取り、位置特徴データ404の受け取りの結果として、入力データ204内に示された1つまたは複数の分類された物体に関連する座標を含む座標データ408を提供するように訓練される。一部の実装において、座標データ408は、入力データ204内の画像フレームのシーケンスに対応する座標埋め込みのシーケンスを含み得る。座標埋め込みのシーケンスの中の各座標埋め込みは、対応する画像のフレーム内に示された1つまたは複数の分類された物体に関連す

る座標を表し得る。一部の実装において、座標データ408は、分類された物体を示す画像フレームに関連するカメラ座標空間内の分類された物体の三次元位置を示す分類された物体に関連する座標を含み得る。

【００７４】

一部の実装において、座標変換モデル410は、座標データ408および入力データ204の少なくとも一部(たとえば、カメラ姿勢データ)を受け取り、データの受け取りの結果として、出力データ206 (たとえば、ジオロケーションデータ)を提供するように訓練される。特に、座標変換モデル410は、カメラ座標空間内の分類された物体に関連する座標を実世界の座標(たとえば、緯度値および経度値)に変換することができる。

【００７５】

一部の実装において、ジオロケーション予測モデル400は、予測された実世界の位置が正確であり、関心のある分類された物体に対応することを確かめるために複数の損失値のうちの1つまたは複数に少なくとも部分的に基づいて訓練され得る。例として、ジオロケーションシステム100は、複数の画像フレームにまたがる特定された物体に関連する座標の間の分散に少なくとも部分的に基づいて位置の整合性の損失を決定し得る。ジオロケーションシステム100は、ジオロケーション予測モデルによって決定された座標が分類された物体に関する複数の画像フレームにまたがって整合性があるようにジオロケーション予測モデル400を訓練するために決定された位置の整合性の損失を使用することができる。別の例として、ジオロケーションシステム100は、画像特徴データ304およびアテンション値データ308に少なくとも部分的に基づいて外観の整合性の損失を決定し得る。特に、ジオロケーションシステム100は、複数の画像フレームに関する外観の特徴を決定するためにアテンション値データ308に含まれる空間的アテンション値によって画像フレームに対応する画像の特徴を重み付けすることができ、ジオロケーションシステム100は、複数の画像フレームにまたがる決定された外観の特徴の間の分散に少なくとも部分的に基づいて外観の整合性の損失を決定することができる。ジオロケーションシステム100は、ジオロケーション予測モデルによって分類される1つまたは複数の物体が、物体が見える各画像フレームにおいて類似した視覚的外観を有するようにジオロケーション予測モデル400を訓練するために決定された外観の整合性の損失を使用することができる。別の例として、ジオロケーションシステム100は、座標データ408およびアテンション値データ308に少なくとも部分的に基づいて照準の損失を決定し得る。ジオロケーションシステム100は、画像フレーム内に示された分類された物体に関連する座標データ408内の座標が、画像フレームに関連するカメラ座標空間内で、分類された物体に関連する空間的アテンションが最も高いエリア内に射影されるようにジオロケーション予測モデル400を訓練するために照準の損失を使用することができる。別の例として、ジオロケーションシステム100は、予測された実世界の座標を、予測された実世
界の座標がそれらに基づいて決定される画像フレームを撮影するために使用されたカメラの実際の可能なFOV内に制約するために視野(FOV)の損失を決定することができる。ジオロケーションシステム100は、予測された実世界の座標の範囲に対する意味のある制限(たとえば、妥当な空間)を含めるためにジオロケーション予測モデル400を訓練するために決定されたFOVの損失を使用することができる。

【００７６】

図5は、本開示の例示的な実施形態による例示的な物体分類モデル500のブロック図を示す。物体分類モデル500は、物体分類モデル500がアテンション値データ308に加えて分類データ309を出力することを除いて図3の物体分類モデル306と同様である。

【００７７】

一部の実装において、物体分類モデル500は、画像の特徴のシーケンスを表すデータ(たとえば、画像特徴データ304)を受け取り、画像特徴データ304の受け取りの結果として、分類データ309、時間的アテンションデータ504、および空間的アテンションデータ506を提供するように訓練される。分類データ309は、画像フレームのシーケンス内に示された1つまたは複数の物体に関連する分類を含み得る。時間的アテンションデータ504は、

画像の特徴のシーケンスに関連する1つまたは複数の時間的アテンション値を含むことが可能であり、空間的アテンションデータ506は、画像の特徴のシーケンスに関連する1つまたは複数の空間的アテンション値を含むことが可能である。一部の実装において、物体分類モデル500は、分類データ309に少なくとも部分的に基づいて訓練され得る。たとえば、ジオロケーションシステム100は、分類データ内の1つまたは複数の分類ラベルおよび入力データ204内の画像フレームのシーケンスに関連する分類に少なくとも部分的に基づいてソフトマックス交差エントロピー誤差を決定することができる。ジオロケーションシステム100は、決定されたソフトマックス交差エントロピー誤差に少なくとも部分的に基づいて物体分類モデル500を訓練することができる。

【００７８】

一部の実装において、物体分類モデル500は、時空間アテンションメカニズム層510、複数のLSTMブロックを含む長期短期記憶(LSTM)層512、および全結合(FC)層514を含み得る。時空間アテンションメカニズム層510は、画像特徴データ304に少なくとも部分的に基づいて時間的アテンションデータ504および空間的アテンションデータ506を決定することができる。LSTM層512内の各LSTMブロックは、画像特徴データ304に少なくとも部分的に基づいてフレーム毎の埋め込みを決定し、複数の画像フレームにまたがって持続する1つまたは複数の物体を決定するためにFC層514にフレーム毎の埋め込みを提供することができる。物体分類モデル500は、分類データ309を決定するために時間的アテンションデータ504に基づいてフレーム毎の埋め込みを重み付けすることができる。

【００７９】

図6は、本開示の例示的な実施形態による例示的な位置特徴抽出モデル600のブロック図を示す。位置特徴抽出モデル600は、位置特徴抽出モデル600が単一の画像フレームに対応する1つまたは複数の画像の特徴を表すデータ(たとえば、画像特徴データ304)を受け取るように訓練されることを除いて図4の位置特徴抽出モデル402と同様である。1つまたは複数の画像の特徴およびアテンション値データ308の受け取りの結果として、位置特徴抽出モデル600は、単一の画像フレームに対応する1つまたは複数の位置の特徴を含む位置特徴データ404を提供する。一部の実装において、ジオロケーションシステム100は、入力データ204内の各画像フレームに関して、画像フレームに対応する1つまたは複数の画像の特徴を表すデータを順に入力し得る。一部の実装において、情報抽出モデル300は、複数の位置特徴抽出モデル600を含むことが可能であり、ジオロケーションシステム100は、1つまたは複数の画像の特徴を表すデータを並列に入力することが可能である。たとえば、情報抽出モデル300が第1のおよび第2の位置特徴抽出モデル600を含む場合、ジオロケーションシステム100は、第1の位置特徴抽出モデル600に第1の画像フレームに対応する1つまたは複数の画像の特徴を表すデータを、第2の位置特徴抽出モデル600に第2の画像フレームに対応する1つまたは複数の画像の特徴を表すデータを同時に入力し得る。このようにして、位置特徴抽出モデル600は、入力データ204内の画像フレームのシーケンスに対応する位置の特徴のシーケンス(たとえば、位置特徴データ404)を提供し得る。

【００８０】

図7は、本開示の例示的な実施形態による例示的な位置予測モデル700のブロック図を示す。位置予測モデル700は、位置予測モデル700が複数の長期短期記憶(LSTM)ブロックを含むLSTM層712および全結合(FC)層714を含むことを除いて図4の位置予測モデル406と同様である。位置予測モデル700は、入力データ204内の画像フレームのシーケンスに対応する位置の特徴のシーケンスを表すデータ(たとえば、位置特徴データ404)を受け取るように訓練され、位置の特徴のシーケンスの受け取りの結果として、位置予測モデル700は、画像フレームのシーケンス内に示された分類された物体に関する画像フレームのシーケンスに対応する座標のシーケンスを表すデータを提供する。座標のシーケンスは、たとえば、分類された物体を示す各画像フレーム内の分類された物体に関連する座標を含み得る。たとえば、位置予測モデル700は、位置特徴埋め込みのシーケンスを含む位置特徴データ404を受け取ることが可能であり、各位置特徴埋め込みは、画像フレームのシーケンスの中の画像フレームに対応する1つまたは複数の位置の特徴を表す。位置予測モデ

ル700は、各位置特徴埋め込みをLSTM層712内の対応するLSTMブロックに提供し得る。各LSTMブロックからの出力は、対応する画像フレーム内の物体の予測された位置を表し得る。このようにして、LSTM層712は、物体に関する予測された位置のシーケンスを出力することが可能であり、予測された位置のシーケンスは、物体を示す画像フレームのシーケンスの中の各画像フレーム内の物体の予測された位置に対応する。LSTM層712の出力は、物体に関する座標のシーケンスを含む座標データ408を決定するためにFC層714に提供され得る。

【００８１】

一部の実装において、ジオロケーションシステム100は、入力データ204内の画像フレームのシーケンス内に示された複数の分類された物体に関する座標のシーケンスを順に決定するために位置予測モデル700を使用することができる。たとえば、位置予測モデル700の各反復は、画像フレームのシーケンス内に示された異なる物体に関連する座標のシーケンスを出力し得る。一部の実装において、情報抽出モデル300は、複数の位置予測モデル700を含むことが可能であり、ジオロケーションシステム100は、位置特徴データ404を複数の位置予測モデル700の各々に並列に入力することが可能である。たとえば、情報抽出モデル300が第1のおよび第2の位置予測モデル700を含む場合、ジオロケーションシステム100は、位置特徴データ404を第1のおよび第2の位置予測モデル700に同時に入力し、第1の位置予測モデル700の出力として第1の分類された物体に関連する座標の第1のシーケンスを取得し、第2の位置予測モデル700の出力として第2の分類された物体に関連する座標の第2のシーケンスを取得することができる。

【００８２】

例示的な方法

図8は、本開示の例示的な実施形態による情報抽出を実行するための例示的な方法の流れ図を示す。図8は説明および検討を目的として特定の順序で実行されるステップを示すが、本開示の方法は、特に示される順序または配列に限定されない。方法800の様々なステップは、本開示の範囲を逸脱することなく様々な方法で省略され、再配列され、組み合わされ、および/または適応され得る。

【００８３】

802において、コンピューティングシステムが、画像のシーケンスを表すデータを取得し得る。たとえば、ジオロケーションシステム100が、画像のシーケンスを表すデータを含む入力データ204を取得し得る。ジオロケーションシステム100は、画像のシーケンスから位置情報を抽出するように訓練される機械学習された情報抽出モデル120/140に画像のシーケンスを入力し得る。一部の実装において、画像のシーケンスは、画像のシーケンスの中の複数の画像にまたがって複数の物体を示すことが可能であり、情報抽出モデル120/140の出力は、画像のシーケンス内に示された複数の物体に関連する実世界の位置を表すデータを含み得る。

【００８４】

804において、コンピューティングシステムが、画像のシーケンスから抽出された画像の特徴のシーケンスに少なくとも部分的に基づいて画像のシーケンスに関連する分類ラベルおよびアテンション値を決定し得る。たとえば、ジオロケーションシステム100は、(たとえば、画像特徴抽出モデル302によって)画像のシーケンスから抽出された画像の特徴のシーケンスを表すデータ(たとえば、画像特徴データ304)に少なくとも部分的に基づいて分類データ309、時間的アテンション値を含む時間的アテンションデータ504、および画像のシーケンスに関連する空間的アテンション値を含む空間的アテンションデータ506を決定し得る。特に、ジオロケーションシステム100は、弱教師あり物体分類モデル306に画像の特徴のシーケンスを入力し、画像の特徴のシーケンスを入力したことに応じて物体分類モデル306の出力として分類データ309、時間的アテンションデータ504、および空間的アテンションデータ506を取得し得る。ジオロケーションシステム100は、画像の特徴のシーケンス、時間的アテンションデータ504、および空間的アテンションデータ506に少なくとも部分的に基づいて物体に関連する実世界の位置を予測し得る。

【００８５】
806において、コンピューティングシステムが、画像の特徴のシーケンスおよびアテンション値に少なくとも部分的に基づいて位置の特徴のシーケンスを決定し得る。たとえば、ジオロケーションシステム100は、画像の特徴のシーケンスを表すデータ、時間的アテンションデータ504、および空間的アテンションデータ506をフレームレベル位置特徴抽出モデル600に入力し、画像の特徴のシーケンス、時間的アテンションデータ504、および空間的アテンションデータ506を入力したことに応じてフレームレベル位置特徴抽出モデル600の出力として物体に関連する1つまたは複数の位置の特徴を含む位置の特徴のシーケンスを表す位置特徴データ404を取得し得る。
【００８６】
808において、コンピューティングシステムが、位置の特徴のシーケンスおよびアテンション値に少なくとも部分的に基づいて画像のシーケンス内に示された1つまたは複数の物体に関連する座標を決定し得る。たとえば、ジオロケーションシステム100は、フレームレベル位置予測モデル406に位置特徴データ404を入力し、位置特徴データ404を入力したことに応じてフレームレベル位置予測モデル406の出力として物体に関連するカメラ座標空間内の座標を表す座標データ408を取得し得る。ジオロケーションシステム100は、座標データ408および入力データ204内の物体に関連するカメラ姿勢データに少なくとも部分的に基づいて物体に関連する実世界の座標を決定し得る。
【００８７】
810において、コンピューティングシステムが、決定された座標に少なくとも部分的に基づいて1つまたは複数の物体に関連する実世界の位置を予測し得る。たとえば、ジオロケーションシステム100は、入力データ204を入力したことに応じて情報抽出モデル120/140の出力として画像のシーケンス内に示された物体に関連する実世界の位置を表す出力データ206を取得し得る。ジオロケーションシステム100は、分類データ309に少なくとも部分的に基づいて物体に関する予測された実世界の位置を物体に対応する分類ラベルに関連付けることができる。
【００８８】
図9は、本開示の例示的な実施形態による情報抽出モデルを訓練するための例示的な方法の流れ図を示す。図9は説明および検討を目的として特定の順序で実行されるステップを示すが、本開示の方法は、特に示される順序または配列に限定されない。方法900の様々なステップは、本開示の範囲を逸脱することなく様々な方法で省略され、再配列され、組み合わされ、および/または適応され得る。
【００８９】
902において、コンピューティングシステム(たとえば、訓練コンピューティングシステム150、またはジオロケーションシステム100のその他の部分)が、ノイズのある分類付きの画像のシーケンスから抽出された画像の特徴のシーケンスを表すデータを取得し得る。たとえば、ジオロケーションシステム100は、画像のシーケンスに関連する単一の分類ラベル付きの画像のシーケンスから抽出された画像の特徴のシーケンスを表す画像データ(たとえば、画像特徴データ304)を取得し得る。ジオロケーションシステム100は、画像特徴抽出モデル302に画像データを入力し、画像データを入力したことに応じて画像特徴抽出モデル302の出力として画像特徴データ304を取得し得る。
【００９０】
904において、コンピューティングシステムが、画像の特徴のシーケンスに少なくとも部分的に基づいて画像のシーケンス内に示された1つまたは複数の物体に関連する分類を決定し得る。たとえば、ジオロケーションシステム100は、弱教師あり物体分類モデル306に画像特徴データ304を入力し、画像の特徴のシーケンスを入力したことに応じて物体分類モデル306の出力として画像フレームのシーケンス内に示された物体に関連する分類を表すデータ(たとえば、分類データ309)を取得し得る。
【００９１】
906において、コンピューティングシステムが、決定された分類に関連する損失を決定

し得る。たとえば、ジオロケーションシステム100は、画像のシーケンスに関連するノイズのある分類に少なくとも部分的に基づいて物体分類モデル306によって出力された分類データ309に関連する損失を決定し得る。

【００９２】

908において、コンピューティングシステムが、決定された分類に関連する損失に少なくとも部分的に基づいて物体分類モデルを訓練し得る。たとえば、ジオロケーションシステム100は、決定された損失に少なくとも部分的に基づいて物体分類モデル306を訓練することができる。

【００９３】

図10は、本開示の例示的な実施形態による情報抽出モデルを訓練するための例示的な方法の流れ図を示す。図10は説明および検討を目的として特定の順序で実行されるステップを示すが、本開示の方法は、特に示される順序または配列に限定されない。方法1000の様々なステップは、本開示の範囲を逸脱することなく様々な方法で省略され、再配列され、組み合わされ、および/または適応され得る。

【００９４】

1002において、コンピューティングシステム(たとえば、訓練コンピューティングシステム150、またはジオロケーションシステム100のその他の部分)が、ノイズのある分類付きの画像のシーケンスから抽出された画像の特徴のシーケンスを表すデータおよび画像のシーケンスに関連するアテンション値を表すデータを取得し得る。たとえば、ジオロケーションシステム100は、画像のシーケンス(たとえば、入力データ204)から抽出された画像の特徴のシーケンスを表すデータ(たとえば、画像特徴データ304)を取得し得る。ジオロケーションシステム100は、画像特徴抽出モデル302に入力データ204を入力し、入力データ204を入力したことに応じて画像特徴抽出モデル302の出力として画像特徴データ304を取得し得る。

【００９５】

1004において、コンピューティングシステムが、位置の特徴のシーケンスおよびアテンション値に少なくとも部分的に基づいて画像のシーケンス内に示された1つまたは複数の物体に関連する実世界の位置を予測し得る。たとえば、ジオロケーションシステム100は、入力データ204を入力したことに応じて情報抽出モデル120/140の出力として出力データ206を取得し得る。出力データ206は、画像のシーケンス内に示された物体に関連する実世界の位置を表し得る。特に、ジオロケーションシステム100は、位置特徴抽出モデル402に画像特徴データ304およびアテンション値データ308を入力し、結果として、入力データ204内に示された1つまたは複数の分類された物体に関連する1つまたは複数の位置の特徴を含む位置特徴データ404を取得し得る。ジオロケーションシステム100は、位置予測モデル406に位置特徴データ404を入力し、結果として、入力データ204内に示された1つまたは複数の分類された物体に関連する座標を含む座標データ408を取得し得る。ジオロケーションシステム100は、座標変換モデル410に座標データ408および入力データ204の少なくとも一部(たとえば、カメラ姿勢データ)を入力し、結果として、入力データ204内に示された1つまたは複数の物体(たとえば、道路標識)に関する予測された実世界の位置(たとえば、緯度および経度)を含む出力データ206 (たとえば、ジオロケーションデータ)を取得し得る。

【００９６】

1006において、コンピューティングシステムが、1つまたは複数の物体に関連する予測された実世界の位置に少なくとも部分的に基づいて位置の整合性の損失を決定し得る。たとえば、ジオロケーションシステム100は、物体を示す画像のシーケンスの中の複数の画像にまたがる物体に関連する座標の間の分散に少なくとも部分的に基づいて位置の整合性の損失を決定し得る。

【００９７】

1008において、コンピューティングシステムが、1つまたは複数の物体に関連する予測された実世界の位置に少なくとも部分的に基づいて外観の整合性の損失を決定し得る。た

とえば、ジオロケーションシステム100は、物体を示す画像のシーケンスの中の複数の画像にまたがって決定された外観の特徴の間の分散に少なくとも部分的に基づいて外観の整合性の損失を決定し得る。

【００９８】

1010において、コンピューティングシステムが、1つまたは複数の物体に関連する予測された実世界の位置に少なくとも部分的に基づいて照準の損失を決定し得る。たとえば、ジオロケーションシステム100は、物体を示す画像のシーケンスの中の複数の画像にまたがる物体に関連するカメラ座標空間内の座標および物体に関連する空間的アテンションに少なくとも部分的に基づいて照準の損失を決定し得る。

【００９９】

1012において、コンピューティングシステムが、1つまたは複数の物体に関連する予測された実世界の位置に少なくとも部分的に基づいて視野の損失を決定し得る。たとえば、ジオロケーションシステム100は、物体に関連する実世界の座標と、物体を示す画像のシーケンスを撮影するために使用されたカメラに関連する視野とに少なくとも部分的に基づいて視野の損失を決定し得る。

【０１００】

1014において、コンピューティングシステムが、決定された損失に少なくとも部分的に基づいて位置予測モデルを訓練し得る。たとえば、ジオロケーションシステム100は、位置の整合性の損失、外観の整合性の損失、照準の損失、および/または視野の損失に少なくとも部分的に基づいて位置予測モデル406を訓練し得る。

【０１０１】

追加的な開示

本明細書において検討されたテクノロジーは、サーバ、データベース、ソフトウェアアプリケーション、およびその他のコンピュータに基づくシステム、ならびに行われるアクション、およびそのようなシステムに送信され、そのようなシステムから送信される情報に言及する。コンピュータに基づくシステムの固有の柔軟性が、構成要素の間のタスクおよび機能の非常に多様な可能な構成、組合せ、および分割を可能にする。たとえば、本明細書において検討されたプロセスは、単一のデバイスもしくは構成要素または組合せで働く複数のデバイスもしくは構成要素を使用して実装され得る。データベースおよびアプリケーションは、単一のシステム上に実装され得るかまたは複数のシステムに分散され得る。分散された構成要素は、逐次的にまたは並列的に動作し得る。

【０１０２】

本対象がその様々な特定の例示的な実施形態に関連して詳細に説明されたが、各例は、本開示の限定ではなく説明の目的で提供されている。当業者は、以上のことを理解すると、そのような実施形態に対する改変、そのような実施形態の変更、およびそのような実施形態の均等物を容易に生み出し得る。したがって、対象の開示は、当業者に容易に分かるように、本対象に対するそのような修正、変更、および/または追加を包含することを除外しない。たとえば、一実施形態の一部として示されるかまたは説明される特徴は、さらなる実施形態を生み出すために別の実施形態によって使用され得る。したがって、本開示は、そのような改変、変更、および均等物を包含するように意図される。

【０１０３】

特に、図8から図10は説明および検討を目的として特定の順序で実行されるステップをそれぞれ示すが、本開示の方法は、特に示される順序または配列に限定されない。方法800、900、および1000の様々なステップは、本開示の範囲を逸脱することなく様々な方法で省略され、再配列され、組み合わされ、および/または適応され得る。

【符号の説明】

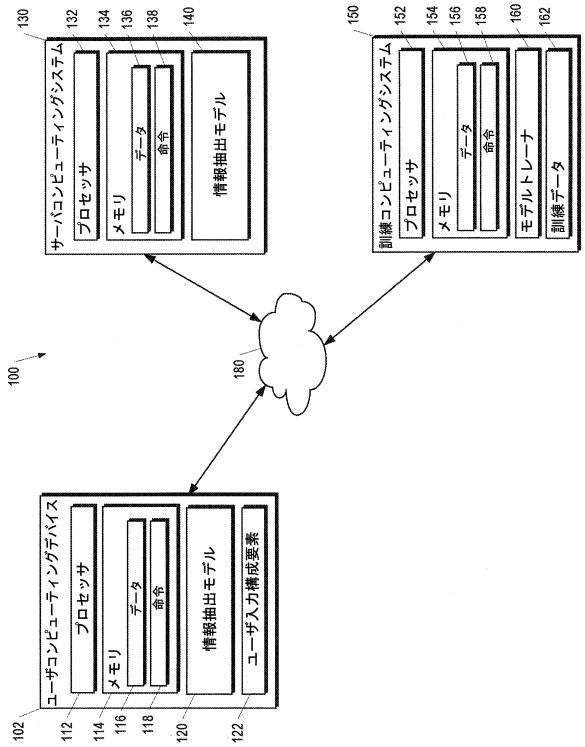
【０１０４】

10　コンピューティングデバイス
50　コンピューティングデバイス
100　ジオロケーションシステム

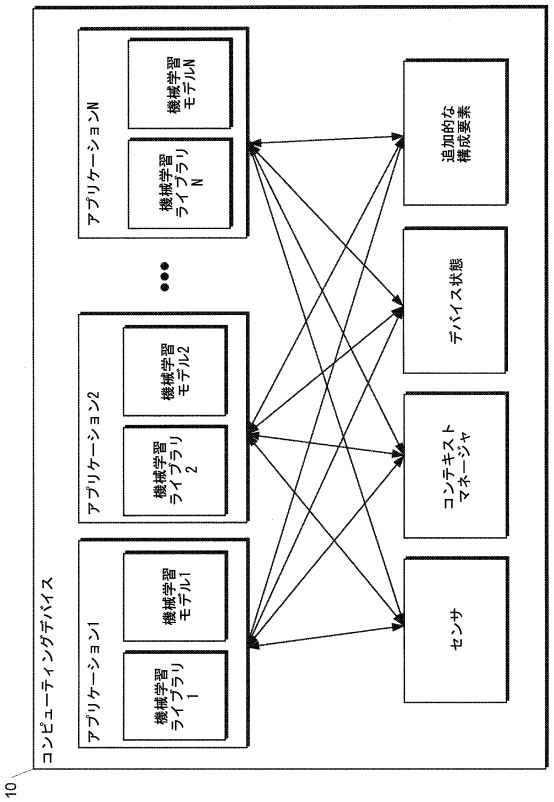
102 ユーザコンピューティングデバイス
112 プロセッサ
114 メモリ
116 データ
118 命令
120 情報抽出モデル
122 ユーザ入力構成要素
130 サーバコンピューティングシステム
132 プロセッサ
134 メモリ
136 データ
138 命令
140 情報抽出モデル
150 訓練コンピューティングシステム
152 プロセッサ
154 メモリ
156 データ
158 命令
160 モデルトレーナ
162 訓練データ
180 ネットワーク
200 情報抽出モデル
204 入力データ
206 出力データ
300 情報抽出モデル
302 画像特徴抽出モデル
304 画像特徴データ
306 物体分類モデル
308 アテンション値データ
309 分類データ
310 ジオロケーション予測モデル
400 ジオロケーション予測モデル
402 位置特徴抽出モデル
404 位置特徴データ
406 位置予測モデル
408 座標データ
410 座標変換モデル
500 物体分類モデル
504 時間的アテンションデータ
506 空間的アテンションデータ
510 時空間アテンションメカニズム層
512 長期短期記憶(LSTM)層
514 全結合(FC)層
600 位置特徴抽出モデル
700 位置予測モデル
712 長期短期記憶(LSTM)層
714 全結合(FC)層
800 方法

【図面】

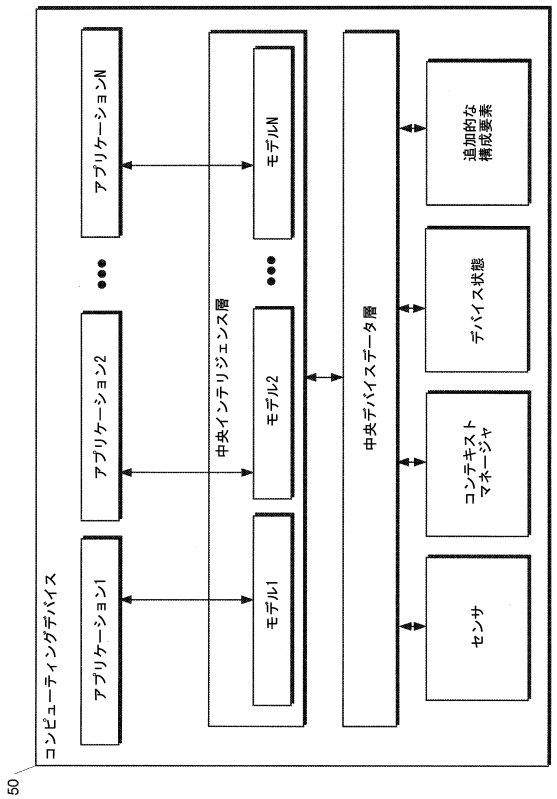
【図１Ａ】


100
102
112
114
116
118
120
122
130
132
134
136
138
140
150
152
154
156
158
160
162
180
ユーザコンピューティングデバイス
プロセッサ
メモリ
データ
命令
情報抽出モデル
ユーザ入力構成要素
サーバコンピューティングシステム
プロセッサ
メモリ
データ
命令
情報抽出モデル
訓練コンピューティングシステム
プロセッサ
メモリ
データ
命令
モデルトレーナ
訓練データ

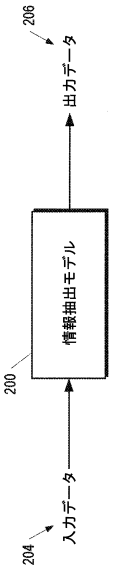
【図１Ｂ】


10
コンピューティングデバイス
アプリケーション1
機械学習ライブラリ1
機械学習モデル1
アプリケーション2
機械学習ライブラリ2
機械学習モデル2
アプリケーションN
機械学習ライブラリN
機械学習モデルN
センサ
コンテキストマネージャ
デバイス状態
追加的な構成要素

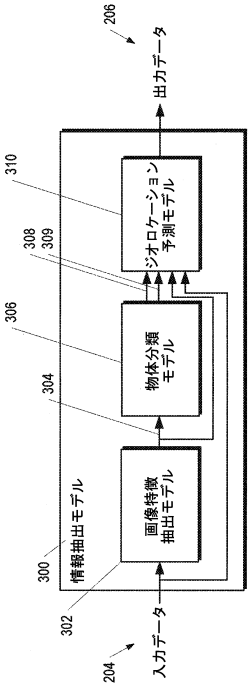
【図１Ｃ】


コンピューティングデバイス
50
アプリケーション1
アプリケーション2
アプリケーションN
中央インテリジェンス層
モデル1
モデル2
モデルN
中央デバイスデータ層
センサ
コンテキスト
マネージャ
デバイス状態
追加的な
構成要素

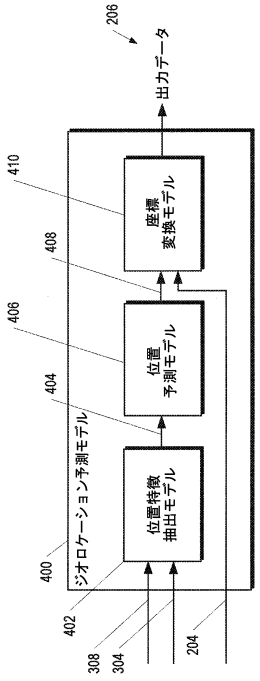
【図２】


206
出力データ
200
情報抽出モデル
204
入力データ

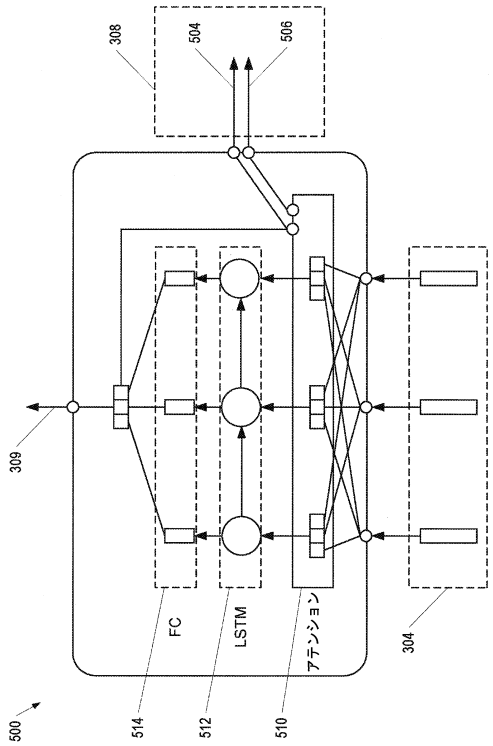
【図３】


206
出力データ
310
ジオロケーション
予測モデル
308
309
306
物体分類
モデル
304
画像特徴
抽出モデル
情報抽出モデル
300
302
204
入力データ

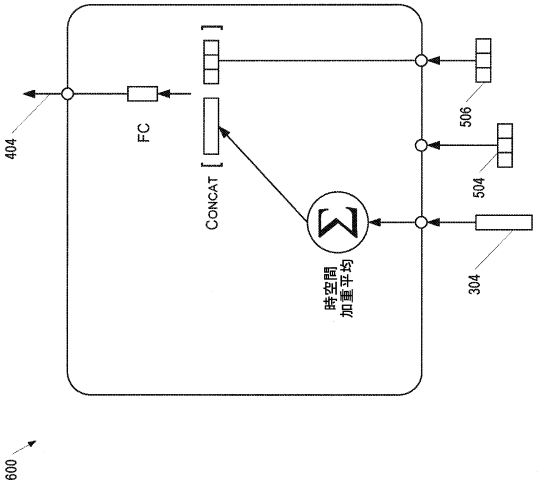
【図４】


206
出力データ
410
座標
変換モデル
408
406
位置
予測モデル
404
位置特徴
抽出モデル
ジオロケーション予測モデル
400
402
308
304
204

【図５】



【図６】



【図７】

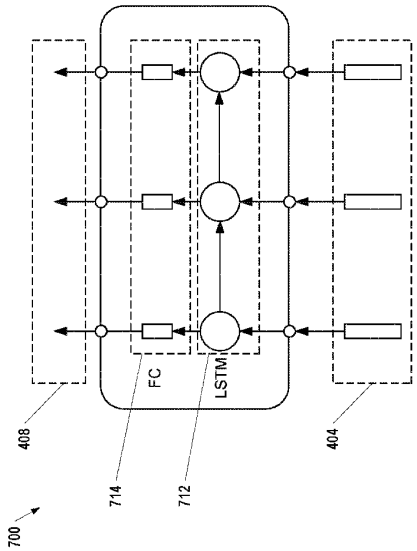
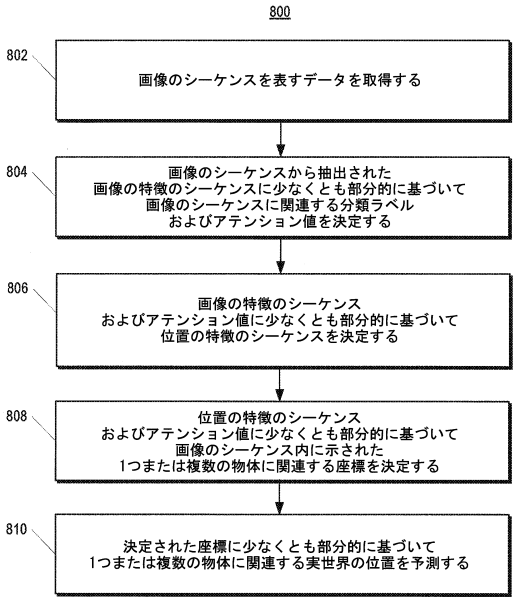
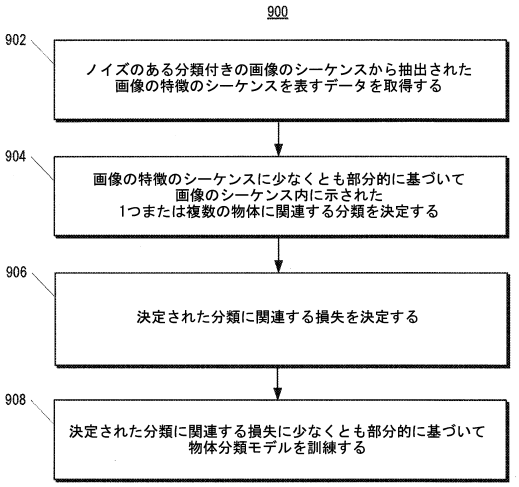


FIG. 7

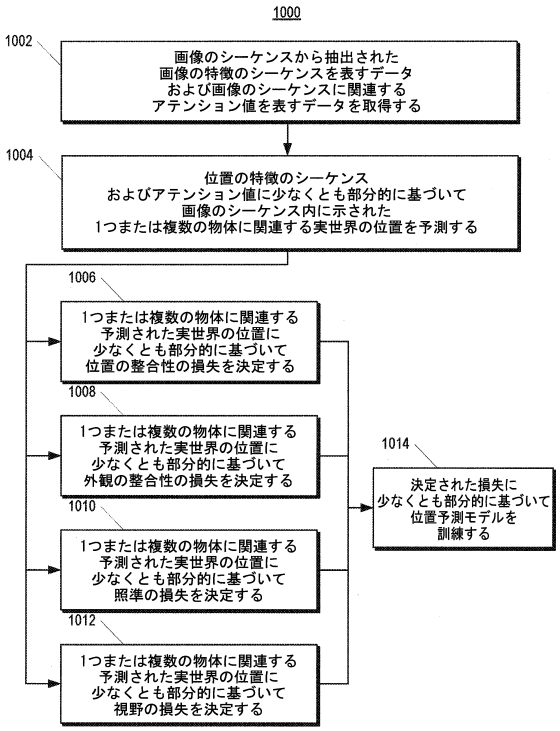
【図８】



【図９】


900
902
ノイズのある分類付きの画像のシーケンスから抽出された
画像の特徴のシーケンスを表すデータを取得する
904
画像の特徴のシーケンスに少なくとも部分的に基づいて
画像のシーケンス内に示された
1つまたは複数の物体に関連する分類を決定する
906
決定された分類に関連する損失を決定する
908
決定された分類に関連する損失に少なくとも部分的に基づいて
物体分類モデルを訓練する

【図１０】


1000
1002
画像のシーケンスから抽出された
画像の特徴のシーケンスを表すデータ
および画像のシーケンスに関連する
アテンション値を表すデータを取得する
1004
位置の特徴のシーケンス
およびアテンション値に少なくとも部分的に基づいて
画像のシーケンス内に示された
1つまたは複数の物体に関連する実世界の位置を予測する
1006
1つまたは複数の物体に関連する
予測された実世界の位置に
少なくとも部分的に基づいて
位置の整合性の損失を決定する
1008
1つまたは複数の物体に関連する
予測された実世界の位置に
少なくとも部分的に基づいて
外観の整合性の損失を決定する
1010
1つまたは複数の物体に関連する
予測された実世界の位置に
少なくとも部分的に基づいて
照準の損失を決定する
1012
1つまたは複数の物体に関連する
予測された実世界の位置に
少なくとも部分的に基づいて
視野の損失を決定する
1014
決定された損失に
少なくとも部分的に基づいて
位置予測モデルを
訓練する

フロントページの続き

(51)国際特許分類
G06V 20/70 (2022.01)

FI
G06V 20/40
G06V 20/70

(72)発明者 アレクサンダー・ゴルバン
アメリカ合衆国・カリフォルニア・９４０４３・マウンテン・ビュー・アンフィシアター・パークウェイ・１６００・グーグル・エルエルシー内
(72)発明者 ヤンシャン・ウ
アメリカ合衆国・カリフォルニア・９４０４３・マウンテン・ビュー・アンフィシアター・パークウェイ・１６００・グーグル・エルエルシー内

審査官 笠田 和宏

(56)参考文献 特開２００９－２５１７９３（ＪＰ，Ａ）
Tobias Weyand, Ilya Kostrikov, James Philbin，"PlaNet - Photo Geolocation with Convolutional Neural Networks"，Computer Vision - ECCV 2016，European Conference on Computer Vision，2016年10月11日，PP.37-55，https://link.springer.com/content/pdf/10.1007/978-3-319-46484-8_3.pdf

(58)調査した分野 (Int.Cl.，ＤＢ名)
Ｇ０６Ｔ ７／００
Ｇ０６Ｎ ３／０８
Ｇ０６Ｎ ２０／００
Ｇ０６Ｖ １０／８２
Ｇ０６Ｖ ２０／４０
Ｇ０６Ｖ ２０／７０