

(19) 日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2020-52999

(P2020-52999A)

(43) 公開日 令和2年4月2日 (2020. 4. 2)

(51) Int.Cl.		F I		テーマコード (参考)
G06T 7/00	(2017.01)	G06T 7/00	350B	5L096
G06N 20/00	(2019.01)	G06N 99/00	150	

審査請求 有 請求項の数 10 O L (全 18 頁)

(21) 出願番号 特願2019-12692 (P2019-12692)
 (22) 出願日 平成31年1月29日 (2019. 1. 29)
 (31) 優先権主張番号 10-2018-0114986
 (32) 優先日 平成30年9月27日 (2018. 9. 27)
 (33) 優先権主張国・地域又は機関
 韓国 (KR)

(71) 出願人 518414959
 クラウドワークス インコーポレイテッド
 CrowdWorks, Inc.
 大韓民国, 06621 ソウル, ソチョー
 グ, ソチョーデロ 78-ギル, 22, 5
 06ホ
 506Ho, 22, Seochod
 a ero 78-gil, Seoch
 -gu, Seoul, 06621 R
 epublic of Korea
 (74) 代理人 100121382
 弁理士 山下 託嗣

最終頁に続く

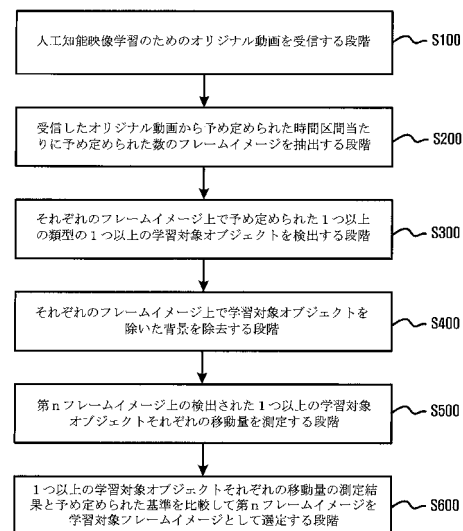
(54) 【発明の名称】 人工知能映像学習のための動画の学習対象フレームイメージサンプリング方法、装置、プログラム及びその映像学習方法

(57) 【要約】 (修正有)

【課題】人工知能映像学習のための動画の学習対象フレームイメージサンプリング方法、装置、プログラム及びその映像学習方法を提供する。

【解決手段】サンプリング方法は、オリジナル動画を受信する段階と、オリジナル動画から予め定められた時間区間当りに予め定められた数のフレームイメージを抽出する段階と、それぞれのフレームイメージ上で予め定められた1つ以上の種類の1つ以上の学習対象オブジェクトを検出する段階と、それぞれのフレームイメージ上で学習対象オブジェクトを除いた背景を除去する段階と、背景が除去された第n (nは2以上の自然数) フレームイメージ上の学習対象オブジェクトそれぞれの移動量を測定する段階と、第nフレームイメージ上の検出された1つ以上の学習対象オブジェクトそれぞれの移動量の測定結果と予め定められた基準を比較して第nフレームイメージを学習対象フレームイメージとして選定する段階と、を含む。

【選択図】図1



【特許請求の範囲】

【請求項 1】

コンピュータにより実行される方法であって、
人工知能映像学習のためのオリジナル動画を受信するオリジナル動画受信段階と、
受信した前記オリジナル動画から予め定められた時間区間あたりに予め定められた数のフレームイメージを抽出するフレームイメージ抽出段階と、
オブジェクト検出アルゴリズムを用いて、それぞれの前記フレームイメージ上で予め定められた 1 つ以上の種類の 1 つ以上の学習対象オブジェクトを検出する学習対象オブジェクト検出段階と、
それぞれの前記フレームイメージ上で前記学習対象オブジェクトを除いた背景を除去する背景除去段階と、
背景が除去された第 n (n は 2 以上の自然数) フレームイメージ上の検出された前記 1 つ以上の学習対象オブジェクトの位置と、前記第 n フレームイメージの直前の背景が除去された第 $n-1$ フレームイメージ上の検出された前記 1 つ以上の学習対象オブジェクトの位置と、を比較し、前記第 n フレームイメージ上の検出された前記 1 つ以上の学習対象オブジェクトそれぞれの移動量を測定するオブジェクトの移動量測定段階と、
前記第 n フレームイメージ上の検出された前記 1 つ以上の学習対象オブジェクトそれぞれの移動量の測定結果と予め定められた基準を比較して前記第 n フレームイメージを学習対象フレームイメージとして選定する学習対象フレームイメージ選定段階と、
を含む人工知能映像学習のための動画の学習対象フレームイメージサンプリング方法

10

20

【請求項 2】

前記学習対象フレームイメージ選定段階は、
検出された前記 1 つ以上の学習対象オブジェクトのうち予め定められた数以上の前記学習対象オブジェクトの移動量が前記予め定められた基準以上である場合には、前記第 n フレームイメージを前記学習対象フレームイメージとして選定し、そうでない場合には、前記第 n フレームイメージを前記学習対象フレームイメージとして選定しないことを特徴とする請求項 1 に記載の人工知能映像学習のための動画の学習対象フレームイメージサンプリング方法。

【請求項 3】

選定された前記学習対象フレームイメージを加工処理して学習用動画セットを生成する学習用動画セット生成段階を更に含むことを特徴とする請求項 1 に記載の人工知能映像学習のための動画の学習対象フレームイメージサンプリング方法。

30

【請求項 4】

前記学習用動画セット生成段階で前記学習対象フレームイメージは、
前記フレームイメージ抽出段階で前記 1 つ以上の学習対象オブジェクトが検出される前のフレームイメージであり、
前記学習用動画セット生成段階は、
前記学習対象フレームイメージ上で前記予め定められた 1 つ以上の種類の 1 つ以上の学習対象オブジェクトを検出するオブジェクト検出段階と、
前記学習対象フレームイメージ上の検出された前記 1 つ以上の学習対象オブジェクトに対して注釈化処理を行う注釈化処理段階と、
を更に含むことを特徴とする請求項 3 に記載の人工知能映像学習のための動画の学習対象フレームイメージサンプリング方法。

40

【請求項 5】

前記学習用動画セット生成段階で前記学習対象フレームイメージは、
前記学習対象オブジェクト検出段階で前記 1 つ以上の学習対象オブジェクトが検出されたフレームイメージであり、
前記学習用動画セット生成段階は、
前記学習対象フレームイメージ上の検出された前記 1 つ以上の学習対象オブジェクト

50

に対して注釈化処理を行う注釈化処理段階を更に含むことを特徴とする請求項 3 に記載の人工知能映像学習のための動画の学習対象フレームイメージサンプリング方法。

【請求項 6】

コンピュータにより実行される方法であって、

学習用動画セットを用いて人工知能映像学習を行う段階を含み、

前記学習用動画は、

人工知能映像学習のためのオリジナル動画を受信するオリジナル動画受信段階と、

受信した前記オリジナル動画から予め定められた時間区間あたりに予め定められた数のフレームイメージを抽出するフレームイメージ抽出段階と、

オブジェクト検出アルゴリズムを用いてそれぞれの前記フレームイメージ上で予め定められた 1 つ以上の種類の 1 つ以上の学習対象オブジェクトを検出する学習対象オブジェクト検出段階と、

それぞれの前記フレームイメージ上で前記学習対象オブジェクトを除いた背景を除去する背景除去段階と、

背景が除去された第 n (n は 2 以上の自然数) フレームイメージ上の検出された前記 1 つ以上の学習対象オブジェクトの位置と、前記第 n フレームイメージの直前の背景が除去された第 $n-1$ フレームイメージ上の検出された前記 1 つ以上の学習対象オブジェクトの位置と、を比較し、前記第 n フレームイメージ上の検出された前記 1 つ以上の学習対象オブジェクトそれぞれの移動量を測定するオブジェクトの移動量測定段階と、

前記第 n フレームイメージ上の検出された前記 1 つ以上の学習対象オブジェクトそれぞれの移動量の測定結果と予め定められた基準を比較して前記第 n フレームイメージを学習対象フレームイメージとして選定する学習対象フレームイメージ選定段階と、

選定された前記学習対象フレームイメージを加工処理して学習用動画セットを生成する学習用動画セット生成段階と、

を含む人工知能映像学習のための動画の学習対象フレームイメージサンプリング方法によって生成される人工知能映像学習方法。

【請求項 7】

人工知能映像学習のためのオリジナル動画を受信するオリジナル動画受信部と、

受信した前記オリジナル動画から予め定められた時間区間あたりに予め定められた数のフレームイメージを抽出するフレームイメージ抽出部と、

オブジェクト検出アルゴリズムを用いてそれぞれの前記フレームイメージ上で予め定められた 1 つ以上の種類の 1 つ以上の学習対象オブジェクトを検出する学習対象オブジェクト検出部と、

それぞれの前記フレームイメージ上で前記学習対象オブジェクトを除いた背景を除去する背景除去部と、

背景が除去された第 n (n は 2 以上の自然数) フレームイメージ上の検出された前記 1 つ以上の学習対象オブジェクトの位置と、前記第 n フレームイメージの直前の背景が除去された第 $n-1$ フレームイメージ上の検出された前記 1 つ以上の学習対象オブジェクトの位置と、を比較し、前記第 n フレームイメージ上の検出された前記 1 つ以上の学習対象オブジェクトそれぞれの移動量を測定するオブジェクトの移動量測定部と、

前記第 n フレームイメージ上の検出された前記 1 つ以上の学習対象オブジェクトそれぞれの移動量の測定結果と予め定められた基準を比較して前記第 n フレームイメージを学習対象フレームイメージとして選定する学習対象フレームイメージ選定部と、

を含む人工知能映像学習のための動画の学習対象フレームイメージサンプリング装置。

【請求項 8】

選定された前記学習対象フレームイメージを加工処理して学習用動画セットを生成する学習用動画セット生成部を更に含むことを特徴とする請求項 7 に記載の人工知能映像学習のための動画の学習対象フレームイメージサンプリング装置。

【請求項 9】

学習用動画セットを用いて人工知能映像学習を行う人工知能映像学習実行部を含み、
前記学習用動画は、
人工知能映像学習のためのオリジナル動画を受信するオリジナル動画受信部と、
受信した前記オリジナル動画から予め定められた時間区間あたりに予め定められた数のフレームイメージを抽出するフレームイメージ抽出部と、
オブジェクト検出アルゴリズムを用いてそれぞれの前記フレームイメージ上で予め定められた1つ以上の種類の1つ以上の学習対象オブジェクトを検出する学習対象オブジェクト検出部と、
それぞれの前記フレームイメージ上で前記学習対象オブジェクトを除いた背景を除去する背景除去部と、
背景が除去された第 n (n は2以上の自然数)フレームイメージ上の検出された前記1つ以上の学習対象オブジェクトの位置と、前記第 n フレームイメージの直前の背景が除去された第 $n-1$ フレームイメージ上の検出された前記1つ以上の学習対象オブジェクトの位置と、を比較し、前記第 n フレームイメージ上の検出された前記1つ以上の学習対象オブジェクトそれぞれの移動量を測定するオブジェクトの移動量測定部と、
前記第 n フレームイメージ上の検出された前記1つ以上の学習対象オブジェクトそれぞれの移動量の測定結果と予め定められた基準を比較して前記第 n フレームイメージを学習対象フレームイメージとして選定する学習対象フレームイメージ選定部と、
選定された前記学習対象フレームイメージを加工処理して学習用動画セットを生成する学習用動画セット生成部と、
を含む人工知能映像学習のための動画の学習対象フレームイメージサンプリング装置により生成される人工知能映像学習装置。

10

20

【請求項10】

ハードウェアであるコンピュータを用いて、請求項1～5の何れか一項に記載の方法を実行するために記録媒体に格納された人工知能映像学習のための動画の学習対象フレームイメージサンプリングプログラム。

【発明の詳細な説明】

【技術分野】

30

【0001】

本発明は、人工知能映像学習のための動画の学習対象フレームイメージサンプリング方法、その装置、プログラム及びその映像学習方法に関する。

【背景技術】

【0002】

人工知能 (Artificial Intelligence、AI) とは、機械から生成された知能を意味する。人間の知能により可能な思考、学習などをコンピュータが行えるようにする方法を研究するコンピュータ工学及び情報技術の一つの分野であって、コンピュータが人間の知能的な行動を模倣できるようにすることをいう。

40

【0003】

人工知能は、持続的に研究開発が行われており、イメージの知能化から音声及びテキストの知能化へ、現在はビデオ映像の知能化に関する研究開発が行われているため、急速に成長しており、ビデオ映像の知能化が産業に与える波及効果は非常に大きい。

【0004】

人工知能学習のための学習データを生成する過程において、獲得したデータを前処理する作業は、学習データを製作する時間のうち約70～80%を占める。

【0005】

また、ビデオ映像データの量は、従来のイメージ又は音声データの量と比較して、数十倍から数百倍まで差がある。

【先行技術文献】

50

【特許文献】

【0006】

【特許文献1】韓国登録特許公報第10-1888647号

【発明の概要】

【発明が解決しようとする課題】

【0007】

ビデオ映像データを用いて学習データの生成は、従来のイメージ又は音声データを用いた学習データの生成と比較して膨大なデータ容量のため、データの前処理作業における時間とコストが非常に増大するという問題がある。

【0008】

そこで、本発明は上記事情に鑑みてなされたものであって、その目的は、データの前処理作業における時間とコストを最小化できる人工知能映像学習のための動画の学習対象フレームイメージサンプリング方法を提供することにある。

【0009】

本発明が解決しようとする課題は、以上で言及された課題に限定されず、言及されていない他の課題は、以下の記載から通常の技術者が明確に理解できるはずである。

【課題を解決するための手段】

【0010】

上述した課題を解決するための本発明の一実施形態に係る人工知能映像学習のための動画の学習対象フレームイメージサンプリング方法は、コンピュータにより実行される方法であって、人工知能映像学習のためのオリジナル動画を受信するオリジナル動画受信段階と、受信した前記オリジナル動画から予め定められた時間区間当たりにより予め定められた数のフレームイメージを抽出するフレームイメージ抽出段階と、オブジェクト検出アルゴリズムを用いてそれぞれの前記フレームイメージ上で予め定められた1つ以上の種類の1つ以上の学習対象オブジェクトを検出する学習対象オブジェクト検出段階と、それぞれの前記フレームイメージ上で前記学習対象オブジェクトを除いた背景を除去する背景除去段階と、背景が除去された第 n (n は2以上の自然数)フレームイメージ上の検出された前記1つ以上の学習対象オブジェクトの位置と前記第 n フレームイメージの直前の背景が除去された第 $n-1$ フレームイメージ上の検出された前記1つ以上の学習対象オブジェクトの位置を比較し、前記第 n フレームイメージ上の検出された前記1つ以上の学習対象オブジェクトそれぞれの移動量を測定するオブジェクトの移動量測定段階と、前記第 n フレームイメージ上の検出された前記1つ以上の学習対象オブジェクトそれぞれの移動量の測定結果と予め定められた基準を比較して前記第 n フレームイメージを学習対象フレームイメージとして選定する学習対象フレームイメージ選定段階とを含む。

【0011】

前記学習対象フレームイメージ選定段階は、検出された前記1つ以上の学習対象オブジェクトのうち予め定められた数以上の前記学習対象オブジェクトの移動量が前記予め定められた基準以上である場合には、前記第 n フレームイメージを前記学習対象フレームイメージとして選定し、そうでない場合には、前記第 n フレームイメージを前記学習対象フレームイメージとして選定しない。

【0012】

上述した課題を解決するための本発明の一実施形態に係る人工知能映像学習のための動画の学習対象フレームイメージサンプリング方法は、選定された前記学習対象フレームイメージを加工処理して学習用動画セットを生成する学習用動画セット生成段階を更に含む。

【0013】

前記学習用動画セット生成段階で前記学習対象フレームイメージは、前記フレームイメージ抽出段階で前記1つ以上の学習対象オブジェクトが検出される前のフレームイメージであり、前記学習用動画セット生成段階は、前記学習対象フレームイメージ上で前記予め定められた1つ以上の種類の1つ以上の学習対象オブジェクトを検出するオブジェクト

10

20

30

40

50

検出段階と、前記学習対象フレームイメージ上の検出された前記１つ以上の学習対象オブジェクトに対して注釈化処理を行う注釈化処理段階とを更に含む。

【００１４】

前記学習用動画セット生成段階で前記学習対象フレームイメージは、前記学習対象オブジェクト検出段階で前記１つ以上の学習対象オブジェクトが検出されたフレームイメージであり、前記学習用動画セット生成段階は、前記学習対象フレームイメージ上の検出された前記１つ以上の学習対象オブジェクトに対して注釈化処理を行う注釈化処理段階を更に含む。

【００１５】

上述した課題を解決するための本発明の一実施形態に係る人工知能映像学習方法は、コンピュータにより実行される方法であって、学習用動画セットを用いて人工知能映像学習を行う段階を含み、前記学習用動画は、人工知能映像学習のためのオリジナル動画を受信するオリジナル動画受信段階と、受信した前記オリジナル動画から予め定められた時間区間あたりに予め定められた数のフレームイメージを抽出するフレームイメージ抽出段階と、オブジェクト検出アルゴリズムを用いてそれぞれの前記フレームイメージ上で予め定められた１つ以上の種類の１つ以上の学習対象オブジェクトを検出する学習対象オブジェクト検出段階と、それぞれの前記フレームイメージ上で前記学習対象オブジェクトを除いた背景を除去する背景除去段階と、背景が除去された第 n （ n は２以上の自然数）フレームイメージ上の検出された前記１つ以上の学習対象オブジェクトの位置と前記第 n フレームイメージの直前の背景が除去された第 $n-1$ フレームイメージ上の検出された前記１つ以上の学習対象オブジェクトの位置を比較し、前記第 n フレームイメージ上の検出された前記１つ以上の学習対象オブジェクトそれぞれの移動量を測定するオブジェクトの移動量測定段階と、前記第 n フレームイメージ上の検出された前記１つ以上の学習対象オブジェクトそれぞれの移動量の測定結果と予め定められた基準を比較して前記第 n フレームイメージを学習対象フレームイメージとして選定する学習対象フレームイメージ選定段階と、選定された前記学習対象フレームイメージを加工処理して学習用動画セットを生成する学習用動画セット生成段階と、を含む人工知能映像学習のための動画の学習対象フレームイメージサンプリング方法によって生成されることを特徴とする。

【００１６】

上述した課題を解決するための本発明の一実施形態に係る人工知能映像学習のための動画の学習対象フレームイメージサンプリング装置は、人工知能映像学習のためのオリジナル動画を受信するオリジナル動画受信部と、受信した前記オリジナル動画から予め定められた時間区間あたりに予め定められた数のフレームイメージを抽出するフレームイメージ抽出部と、オブジェクト検出アルゴリズムを用いてそれぞれの前記フレームイメージ上で予め定められた１つ以上の種類の１つ以上の学習対象オブジェクトを検出する学習対象オブジェクト検出部と、それぞれの前記フレームイメージ上で前記学習対象オブジェクトを除いた背景を除去する背景除去部と、背景が除去された第 n （ n は２以上の自然数）フレームイメージ上の検出された前記１つ以上の学習対象オブジェクトの位置と前記第 n フレームイメージの直前の背景が除去された第 $n-1$ フレームイメージ上の検出された前記１つ以上の学習対象オブジェクトの位置を比較し、前記第 n フレームイメージ上の検出された前記１つ以上の学習対象オブジェクトそれぞれの移動量を測定するオブジェクトの移動量測定部と、前記第 n フレームイメージ上の検出された前記１つ以上の学習対象オブジェクトそれぞれの移動量の測定結果と予め定められた基準を比較して前記第 n フレームイメージを学習対象フレームイメージとして選定する学習対象フレームイメージ選定部とを含む。

【００１７】

上述した課題を解決するための本発明の一実施形態に係る人工知能映像学習のための動画の学習対象フレームイメージサンプリング装置は、選定された前記学習対象フレームイメージを加工処理して学習用動画セットを生成する学習用動画セット生成部を更に含む。

【 0 0 1 8 】

上述した課題を解決するための本発明の一実施形態に係る人工知能映像学習装置は、学習用動画セットを用いて人工知能映像学習を行う人工知能映像学習実行部を含み、前記学習用動画は、人工知能映像学習のためのオリジナル動画を受信するオリジナル動画受信部と、受信した前記オリジナル動画から予め定められた時間区間あたりに予め定められた数のフレームイメージを抽出するフレームイメージ抽出部と、オブジェクト検出アルゴリズムを用いてそれぞれの前記フレームイメージ上で予め定められた1つ以上の種類の1つ以上の学習対象オブジェクトを検出する学習対象オブジェクト検出部と、それぞれの前記フレームイメージ上で前記学習対象オブジェクトを除いた背景を除去する背景除去部と、背景が除去された第 n (n は2以上の自然数)フレームイメージ上の検出された前記1つ以上の学習対象オブジェクトの位置と前記第 n フレームイメージの直前の背景が除去された第 $n-1$ フレームイメージ上の検出された前記1つ以上の学習対象オブジェクトの位置を比較し、前記第 n フレームイメージ上の検出された前記1つ以上の学習対象オブジェクトそれぞれの移動量を測定するオブジェクトの移動量測定部と、前記第 n フレームイメージ上の検出された前記1つ以上の学習対象オブジェクトそれぞれの移動量の測定結果と予め定められた基準を比較して前記第 n フレームイメージを学習対象フレームイメージとして選定する学習対象フレームイメージ選定部と、選定された前記学習対象フレームイメージを加工処理して学習用動画セットを生成する学習用動画セット生成部と、を含む人工知能映像学習のための動画の学習対象フレームイメージサンプリング装置により生成されることを特徴とする。

10

20

【 0 0 1 9 】

上述した課題を解決するための本発明の一実施形態に係る人工知能映像学習のための動画の学習対象フレームイメージサンプリングプログラムは、ハードウェアであるコンピュータを用いて、上述した方法のうち何れか1つの方法を実行するために記録媒体に格納される。

【 0 0 2 0 】

本発明のその他の具体的な事項は、詳細な説明及び図面に含まれている。

【 発明の効果 】

【 0 0 2 1 】

本発明によれば、動画の学習データの生成において動画データのうち一部のみを学習データとして選定することによって、データの前処理作業における時間及びコストを最小化できるという効果を奏する。

30

【 0 0 2 2 】

また、本発明によれば、学習データの選定において不要なデータは低減し、必要なデータのみを選定することができる。

【 0 0 2 3 】

更に、本発明によれば、学習対象オブジェクトと背景とを区分して学習対象オブジェクトの変化量を正確に測定できる。

【 0 0 2 4 】

本発明の効果は、以上で言及された効果に制限されず、言及されていない他の効果は、以下の記載から通常の技術者が明確に理解できる。

40

【 図面の簡単な説明 】

【 0 0 2 5 】

【 図 1 】 本発明の一実施形態に係る人工知能映像学習のための動画の学習対象フレームイメージサンプリング方法を説明する順序図である。

【 図 2 】 本発明の学習対象オブジェクトの移動量を測定するための方法を説明する図である。

【 図 3 】 本発明の学習対象フレームイメージの選定方法を説明する図である。

【 図 4 】 本発明の一実施形態に係る学習対象フレームイメージのサンプリング過程を説明する図である。

50

【図5】学習用動画セットを生成する段階を含む人工知能映像学習のための動画の学習対象フレームイメージサンプリング方法を説明する順序図である。

【図6】本発明の学習対象フレームイメージが、学習対象オブジェクトが検出される前のフレームイメージである場合、学習用動画セットを生成する段階を説明する順序図である。

【発明を実施するための形態】

【0026】

本発明の利点及び特徴、そしてそれらを達成する方法は、添付の図面と共に詳細に後述されている実施形態を参照すれば明確になる。しかし、本発明は、以下で開示される実施形態に限定されるものではなく、互いに異なる多様な形態で実現することができる。但し、本実施形態は、本発明の開示を完全なものにし、本発明が属する技術分野における通常の知識を有する者に本発明の範囲を完全に理解させるために提供されるものであり、本発明は特許請求の範囲の範囲により定義される。

10

【0027】

本明細書で用いられた用語は、実施形態を説明するためのものであり、本発明を制限しようとするものではない。本明細書において、単数型は特に言及しない限り複数型も含む。明細書で用いられる「含む (comprises)」及び/又は「含んでいる (comprising)」は、言及された構成要素以外に1つ以上の他の構成要素の存在又は追加を排除しない。明細書全体に亘って同一の図面符号は同一の構成要素を示し、「及び/又は」は言及された構成要素のそれぞれ及び1つ以上のあらゆる組み合わせを含む。たとえば、「第1」、「第2」などが多様な構成要素を示すために用いられていても、当然、これらの構成要素はこれらの用語により制限されない。これらの用語は、単に1つの構成要素を他の構成要素と区別するために用いる。従って、以下で言及される第1構成要素は、当然、本発明の技術思想内で第2構成要素であってもよい。

20

【0028】

他の定義がなければ、本明細書で用いられる全ての用語（技術及び科学的用語を含む）は、本発明が属する技術分野における通常の知識を有する者が共通して理解できる意味として使用できる。また、一般に用いられる辞典に定義されている用語は、特に明白に定義されていない限り、理想的に又は過度に解釈されない。

【0029】

空間的に相対的な用語である「の下 (below)」、「の真下 (beneath)」、「下部 (lower)」、「の上 (above)」、「上部 (upper)」などは、図示されているように、1つの構成要素と他の構成要素との相関関係を容易に記述するために使用される。空間的に相対的な用語は、図示されている方向に加えて、使用時又は動作時に構成要素の互いに異なる方向を含む用語として理解されるべきである。例えば、図示されている構成要素をひっくり返す場合、他の構成要素の「の下 (below)」又は「の真下 (beneath)」と記述されている構成要素は、他の構成要素の「の上 (above)」に置くことができる。従って、例示的な用語である「下」は下と上の方向を何れも含む。構成要素は他の方向にも向けることができ、これにより空間的に相対的な用語はその向きによって解釈できる。

30

40

【0030】

以下、添付の図面を参照し、本発明の実施形態を詳細に説明する。

【0031】

図1は、本発明の一実施形態に係る人工知能映像学習のための動画の学習対象フレームイメージサンプリング方法を説明する順序図である。

【0032】

図1を参照すれば、本発明の一実施形態に係る動画の学習対象フレームイメージサンプリング方法は、人工知能映像学習のためのオリジナル動画を受信するオリジナル動画受信段階 (S100)、受信したオリジナル動画から予め定められた時間区間あたりに予め定められた数のフレームイメージを抽出するフレームイメージ抽出段階 (S200)、そ

50

それぞれのフレームイメージ上で予め定められた1つ以上の種類の1つ以上の学習対象オブジェクトを検出する学習対象オブジェクト検出段階(S300)、それぞれのフレームイメージ上で学習対象オブジェクトを除いた背景を除去する背景除去段階(S400)、第nフレームイメージ上の検出された1つ以上の学習対象オブジェクトそれぞれの移動量を測定する学習対象オブジェクト移動量測定段階(S500)、及び、1つ以上の学習対象オブジェクトそれぞれの移動量の測定結果と予め定められた基準を比較して第nフレームイメージを学習対象フレームイメージとして選定する学習対象フレームイメージ選定段階(S600)を含む。

【0033】

オリジナル動画受信段階(S100)で、オリジナル動画は、各種カメラなどで収集される動画を含む。

10

【0034】

一実施形態として、自動車の自律走行のための人工知能映像学習データを生成するためにオリジナル動画を受信する場合、オリジナル動画は、車両用ブラックボックスの映像、自動車が通行できる道路を撮影しているCCTV映像又は自律走行のための学習データの生成のためにカメラを取り付けた自動車から獲得した動画など、自律走行のための学習データの生成が可能な動画が全て含まれ、前記例に限定されない。

【0035】

他の実施形態として、病変又は疾患を診断するための人工知能映像学習データを生成するためにオリジナル動画を受信する場合、オリジナル動画は、各種医療用映像機器により獲得された動画を含み、例えば、コンピュータ断層撮影(Computer tomography、CT)画像、核磁気共鳴コンピュータ断層撮影(Nuclear Magnetic Resonance Computed Tomography、NMR-CT)画像、陽電子放射断層撮影(Positron Emission Tomography; PET)画像、CBCT(cone beam CT)、電子ビーム断層撮影(Electron beam tomography)画像、X線(X-Ray)、磁気共鳴画像(Magnetic Resonance Imaging)などが含まれるが、医療用映像機器により獲得された動画は全て含まれ、前記例に限定されない。

20

【0036】

更に他の実施形態として、犯罪現場を検知するための人工知能映像学習データを生成するためにオリジナル動画を受信する場合、オリジナル動画は、公共施設により設置されたCCTV及び個人で設置したCCTVなどにより獲得された動画を含む。

30

【0037】

フレームイメージ抽出段階(S200)は、受信したオリジナル動画からコンピュータ又はユーザの設定により定められた数のフレームイメージを抽出する。

【0038】

受信したオリジナル動画から定められた数のフレームイメージを抽出することは、予め定められた時間の区間で予め定められた数のフレームを抽出することを意味し、例えば、コンピュータは、1秒当たり30フレーム又は1秒当たり60フレームを抽出することができ、ユーザ又はコンピュータにより予め定められた基準に従ってフレームを抽出することは何れも含まれ、前記例に限定されない。

40

【0039】

学習対象オブジェクト検出段階(S300)は、抽出されたフレームイメージそれぞれに対して学習対象オブジェクトを検出する段階であって、抽出されたフレームイメージのそれぞれから1つ以上の学習対象オブジェクトを検出する。検出される学習対象オブジェクトは1つ以上の類型を含む。

【0040】

学習対象オブジェクトの種類としては、例えば、人、自動車、自転車、建物、電信柱、オートバイ、木、花、子犬、猫、道路、交通表示板、過速防止用段差、ロードコーン、車線などを含み、前記例に限定されず、オブジェクトとして識別が可能なものを何れも含む

50

。

【0041】

各学習対象オブジェクトの類型としては、例えば、前面、後面、右側面、左側面などを含み、各学習対象オブジェクトの類型は前記例に限定されず、前記例よりも細分化させて区分することができ、前記例とは全く異なる類型として区分することもできる。

【0042】

学習対象オブジェクトの検出として、1つ以上の類型の1つ以上のオブジェクトを検出することは、オブジェクト検出アルゴリズムを用いて実行される。オブジェクト検出アルゴリズムは、例えば、R-CNNモデルを含む。

【0043】

背景除去段階（S400）は、抽出したフレームイメージ上で検出された学習対象オブジェクトを除いたものなどを背景として処理し、背景部分を全て除去するものである。

【0044】

フレームイメージ上で背景を除去する方法としては、一実施形態として背景に該当する領域を0又は1に処理して除去する。

【0045】

学習対象オブジェクト移動量測定段階（S500）は、背景が除去された第 n （ n は2以上の自然数）フレームイメージ上の検出された1つ以上の学習対象オブジェクトの位置と、第 n フレームイメージの直前の背景が除去された第 $n-1$ フレームイメージ上の検出された1つ以上の学習対象オブジェクトの位置を比較し、第 n フレームイメージ上の検出された1つ以上の学習対象オブジェクトそれぞれの移動量を測定する段階である。

【0046】

学習対象オブジェクト移動量の測定に関する具体的な実施形態は、図2で後述する。

【0047】

学習対象フレームイメージ選定段階（S600）は、第 n フレームイメージ上の検出された1つ以上の学習対象オブジェクトそれぞれの移動量の測定結果と予め定められた基準を比較して、第 n フレームイメージを学習対象フレームイメージとして選定する。

【0048】

学習対象フレームイメージの選定に関する具体的な方法は、図3及び図4で後述する。

。

【0049】

図2は、本発明の学習対象オブジェクトの移動量を測定するための方法を説明する図である。

【0050】

図2を参照して、学習対象オブジェクト移動量測定段階（S500）での移動量の測定方法について説明する。

【0051】

図2の（a）は第 $n-1$ フレームイメージ11、図2の（b）は第 n フレームイメージ12を示している。

【0052】

学習対象オブジェクトの移動量の測定は、第 $n-1$ フレームイメージ11上での学習対象オブジェクト21と第 n フレームイメージ12上での学習対象オブジェクト22の位置を比較するものである。

【0053】

第 $n-1$ フレームイメージ11上での学習対象オブジェクト21と第 n フレームイメージ12上での学習対象オブジェクト22は同一の形態のオブジェクトであって、学習対象オブジェクトの同一の位置に該当する部分をまず選定する。

【0054】

学習対象オブジェクトの同一の位置に該当する部分の選定は、コンピュータが第 $n-1$ フレームイメージ11上の学習対象オブジェクト21で特定部分をAとして選定したと

10

20

30

40

50

するとき、第 n フレームイメージ 12 上の学習対象オブジェクト 22 上で A と同一の位置に該当する部分を A' として選定する。

【0055】

コンピュータは、学習対象オブジェクトの同一の位置に該当する部分を選定し、第 $n-1$ フレームイメージ 11 と第 n フレームイメージ 12 を同一の平面上に置いた後、 A 及び A' に対する座標を抽出する。

【0056】

コンピュータは、 A 及び A' に対する座標を抽出した後、 A 座標及び A' 座標の差を用いて移動量を測定する。

【0057】

10

図 3 は、本発明の学習対象フレームイメージの選定方法を説明する図である。

【0058】

図 3 を参照すれば、学習対象フレームイメージの選定方法は、検出された 1 つ以上の学習対象オブジェクトのうち予め定められた数以上の学習対象オブジェクトの移動量が予め定められた基準以上であるかを判断し (S610)、検出された 1 つ以上の学習対象オブジェクトのうち予め定められた数以上の学習対象オブジェクトの移動量が予め定められた基準以上である場合には、第 n フレームイメージを学習対象フレームイメージとして選定し (S611)、検出された 1 つ以上の学習対象オブジェクトのうち予め定められた数以上の学習対象オブジェクトの移動量が予め定められた基準以上に該当しない場合には、第 n フレームイメージを学習対象フレームイメージとして選定しない (S612)。

20

【0059】

学習対象オブジェクトの移動量の基準を設定し、予め定められた基準以上のものに該当するフレームイメージのみ、学習対象フレームイメージとして選定することによって、不要なデータを低減し、必要なデータのみを選定できる。

【0060】

従って、学習対象フレームイメージは、抽出されたフレームイメージが何れも学習対象フレームイメージとして選定されず、オブジェクトの移動量が少ないため、学習において影響が大きいデータを除き、一部のみ学習対象フレームイメージとして選定されることによって、学習データセットを、データの量が膨大であるにも拘らず、迅速、且つ、正確に生成できる。

30

【0061】

更に、学習データセットが不要なデータを除去して生成されることによって、学習にかかる時間も短縮できる。

【0062】

図 4 は、本発明の一実施形態に係る学習対象フレームイメージのサンプリング過程を説明する図である。

【0063】

図 4 を参照すれば、図 4 の (a) はフレームイメージ抽出段階 (S200) により抽出されたフレームイメージ 10 を示しており、フレームイメージ 10 は学習対象オブジェクト 20 及び背景 30 を含む。

40

【0064】

図 4 の (b) は、図 4 の (a) であるフレームイメージ 10 において、背景除去段階 (S400) により背景 30 が除去されたことを示す図であって、フレームイメージ 10 は学習対象オブジェクト 20 のみを含む。

【0065】

図 4 の (c) は、第 n (n は 2 以上の自然数) フレームイメージ上の検出された 1 つ以上の学習対象オブジェクト 21 の位置と、第 n フレームイメージの直前の背景が除去された第 $n-1$ フレームイメージ上の検出された 1 つ以上の学習対象オブジェクト 22 の位置を比較することを示す図である。

【0066】

50

コンピュータは、それぞれの学習対象オブジェクト 2 1、2 2 間の比較により学習対象オブジェクトの移動量を測定できる。

【0067】

学習対象オブジェクト 2 1、2 2 の位置比較により、移動量を測定した後、コンピュータは、検出された 1 つ以上の学習対象オブジェクト 2 0 のうち予め定められた数以上の学習対象オブジェクトの移動量が予め定められた基準以上である場合には、第 n フレームイメージを学習対象フレームイメージとして選定し、そうでない場合には、第 n フレームイメージを学習対象フレームイメージとして選定しない。

【0068】

一実施形態として、コンピュータは、第 1 フレームイメージ上で検出された学習対象オブジェクトと第 2 フレームイメージ上で検出された学習対象オブジェクトの移動量を比較し、移動量が予め定められた基準以上である場合には、第 2 フレームイメージを学習対象フレームイメージとして選定する。移動量が予め定められた基準以上に該当しない場合には、第 2 フレームイメージを学習対象フレームイメージとして選定しない。

【0069】

学習対象フレームイメージを選定する段階では、選定され抽出されたフレームイメージの全てを比較する。

【0070】

従って、コンピュータは、第 2 フレームイメージを学習対象フレームイメージとして選定又は選定しないことに止まらず、再び第 2 フレームイメージ上で検出された学習対象オブジェクトと第 3 フレームイメージ上で検出された学習対象オブジェクトの移動量を比較し、移動量が予め定められた基準以上である場合には、第 3 フレームイメージを学習対象フレームイメージとして選定する。

【0071】

フレームイメージ上のオブジェクトの移動量を測定し、学習対象フレームイメージとして選定又は選定しない段階は、第 n (n は 2 以上の自然数) フレームイメージ上の検出された 1 つ以上の学習対象オブジェクトと第 n - 1 フレームイメージ上の検出された 1 つ以上の学習対象オブジェクトの位置を比較するものであって、抽出された全てのフレームイメージ上の検出された 1 つ以上の学習対象オブジェクトに対して比較を実行して、学習対象フレームとして選定又は選定しないことを完了するまで繰り返される。

【0072】

また、学習対象オブジェクト 2 0 は 1 つ以上であってもよく、コンピュータは、1 つのフレームイメージ 1 0 上に複数の学習対象オブジェクト 2 0 がある場合にも、それぞれの学習対象オブジェクト 2 0 を比較して移動量を測定する。

【0073】

1 つのフレームイメージ 1 0 上に複数の学習対象オブジェクト 2 0 が含まれる場合には、一実施形態として、コンピュータは、複数の学習対象オブジェクト 2 0 全部の移動量を測定し、予め定められた数の学習対象オブジェクト 2 0 の移動量が予め定められた基準以上である場合、該当フレームイメージ 1 0 を学習対象フレームイメージとして選定する。

【0074】

他の実施形態として、コンピュータは、複数の学習対象オブジェクト 2 0 全部の移動量を測定し、複数の学習対象オブジェクト 2 0 全部の移動量が予め定められた基準以上である場合、該当フレームイメージ 1 0 を学習対象フレームイメージとして選定する。

【0075】

更に他の実施形態として、コンピュータは、複数の学習対象オブジェクト 2 0 のうち予め定められた数の学習対象オブジェクト 2 0 の移動量のみを測定し、測定した学習対象オブジェクト 2 0 の移動量のうち予め定められた数の学習対象オブジェクト 2 0 の移動量が予め定められた基準以上である場合、該当フレームイメージ 1 0 を学習対象フレームイメージとして選定する。

10

20

30

40

50

【0076】

更に別の実施形態として、コンピュータは、複数の学習対象オブジェクト20のうち予め定められた数の学習対象オブジェクト20の移動量のみを測定し、測定した学習対象オブジェクト20全部の移動量が予め定められた基準以上である場合、該当フレームイメージ10を学習対象フレームイメージとして選定する。

【0077】

図4の(c)の移動量の測定により選定された学習対象フレームイメージは、図4の(d)のように、抽出されたフレームイメージ11、12、13、14、15の中から学習対象フレームイメージ12、14として選定される。

【0078】

図5は、学習用動画セットを生成する段階を含む人工知能映像学習のための動画の学習対象フレームイメージサンプリング方法を説明する順序図である。

【0079】

図6は、本発明の学習対象フレームイメージが、学習対象オブジェクトが検出される前のフレームイメージである場合、学習用動画セットを生成する段階を説明する順序図である。

【0080】

図5を参照すれば、本発明の動画の学習対象フレームイメージサンプリング方法は、選定された学習対象フレームイメージを加工処理して学習用動画セットを生成する段階(S700)を更に含む。

【0081】

選定された学習対象フレームイメージを加工処理して学習用動画セットを生成する段階(S700)で、学習対象フレームイメージは、学習対象オブジェクトが検出される前のフレームイメージ又は学習対象オブジェクトが検出されたフレームイメージを含む。

【0082】

図6を参照すれば、学習対象フレームイメージが、学習対象オブジェクトが検出される前のフレームイメージである場合、学習用動画セットを生成する段階(S700)は、学習対象フレームイメージ上で予め定められた1つ以上の種類の1つ以上の学習対象オブジェクトを検出する段階(S710)及び学習対象フレームイメージ上の検出された1つ以上の学習対象オブジェクトに対して注釈化処理を行う段階(S720)を含む。

【0083】

学習対象フレームイメージ上で予め定められた1つ以上の種類の1つ以上の学習対象オブジェクトを検出する段階(S710)は、上述した図1の学習対象オブジェクトを検出する段階(S300)と同一に適用される。

【0084】

学習対象オブジェクトが検出される前のフレームイメージである学習対象フレームイメージが、学習対象オブジェクトを検出する段階を経るようになるのは、その後に行われる学習対象フレームイメージ上の学習対象オブジェクトに対して注釈化処理を行い、これを用いて学習用動画セットを生成するためである。

【0085】

学習対象フレームイメージ上の検出された1つ以上の学習対象オブジェクトに対して注釈化処理を行う段階(S720)で、注釈化処理は、ラベリング(labeling)、カラーリング(coloring)又はレイヤリング(layering)を含み、学習対象オブジェクトが何かを表示することは何れも注釈化処理に含めることができる。

【0086】

学習対象フレームイメージ上の検出された1つ以上の学習対象オブジェクトは、例えば、ボックス(box)などの形態であって、領域が区分されるように表示されている。

【0087】

従って、学習対象オブジェクトとして区分されて表示された領域に対して、各オブジェクトが何かについての説明を作成でき、ラベリング(labeling)は、簡単には

10

20

30

40

50

1つの単語で作成でき、1つの単語ではない文章で詳細に作成することもできる。

【0088】

1つ以上のユーザが直接ラベリングを行うことができ、コンピュータがユーザからラベリング命令を受けて入力することができる。

【0089】

ユーザがラベリングを行う場合、ラベリングの方法としては、直接説明を作成するか、複数のタブの中から選択することもできる。

【0090】

また、コンピュータが学習によって以前にラベリングされたオブジェクトと現在のラベリング対象オブジェクトが同一であると判断される場合には、コンピュータは、現在のラベリング対象オブジェクトに対して以前のオブジェクトと同様にラベリングすることができる。

10

【0091】

カラーリング (coloring) 又はレイヤリング (layering) によって注釈化処理を行う場合にも、ラベリングと同一に1つ以上のユーザが直接カラーリング又はレイヤリングを行うことができ、コンピュータがユーザからカラーリング又はレイヤリング命令を受けて入力することができる。

【0092】

また、コンピュータが学習によって以前にカラーリング又はレイヤリングされたオブジェクトと現在のカラーリング又はレイヤリング対象オブジェクトが同一であると判断される場合には、コンピュータは、現在のカラーリング又はレイヤリング対象オブジェクトに対して以前のオブジェクトと同様にカラーリング又はレイヤリングすることができる。

20

【0093】

一方、学習対象フレームイメージが、1つ以上の学習対象オブジェクトが検出されたフレームイメージである場合、学習用動画セットを生成する段階 (S700) は、学習対象フレームイメージ上の検出された1つ以上の学習対象オブジェクトに対して注釈化処理を行う段階 (S720) を含む。

【0094】

検出された1つ以上の学習対象オブジェクトに対して注釈化処理を行う段階 (S720) は、上述した内容と同一である。

30

【0095】

本発明の他の実施形態に係る人工知能映像学習方法は、学習用動画セットを用いて人工知能映像学習を行う段階を含み、学習用動画は、上述した人工知能映像学習のための動画の学習対象フレームイメージサンプリング方法によって生成された学習用動画である。

【0096】

従って、学習用動画の生成のための方法は、上述した人工知能映像学習のための動画の学習対象フレームイメージサンプリング方法と同一に適用され、本発明の他の実施形態に係る人工知能映像学習方法は、生成された学習用動画セットを用いて人工知能映像学習を行う。

【0097】

40

本発明の更に他の実施形態に係る人工知能映像学習のための動画の学習対象フレームイメージサンプリング装置は、オリジナル動画受信部、フレームイメージ抽出部、学習対象オブジェクト検出部、背景除去部、オブジェクトの移動量測定部及び学習対象フレームイメージ選定部を含む。

【0098】

オリジナル動画受信部は、人工知能映像学習のためにオリジナル動画を受信する。

【0099】

フレームイメージ抽出部は、受信したオリジナル動画から予め定められた時間区間ごとに予め定められた数のフレームイメージを抽出する。

【0100】

50

学習対象オブジェクト検出部は、オブジェクト検出アルゴリズムを用いてそれぞれのフレームイメージ上で予め定められた1つ以上の種類の1つ以上の学習対象オブジェクトを検出する。

【0101】

背景除去部は、それぞれの前記フレームイメージ上で学習対象オブジェクトを除いた背景を除去する。

【0102】

オブジェクトの移動量測定部は、背景が除去された第 n (n は2以上の自然数)フレームイメージ上の検出された1つ以上の学習対象オブジェクトの位置と第 n フレームイメージの直前の背景が除去された第 $n-1$ フレームイメージ上の検出された1つ以上の学習対象オブジェクトの位置を比較し、第 n フレームイメージ上の検出された1つ以上の学習対象オブジェクトそれぞれの移動量を測定する。

【0103】

学習対象フレームイメージ選定部は、第 n フレームイメージ上の検出された1つ以上の学習対象オブジェクトそれぞれの移動量の測定結果と予め定められた基準を比較して第 n フレームイメージを学習対象フレームイメージとして選定する。

【0104】

本発明の更に別の実施形態に係る人工知能映像学習のための動画の学習対象フレームイメージサンプリング装置は、選定された前記学習対象フレームイメージを加工処理して学習用動画セットを生成する学習用動画セット生成部を更に含む。

【0105】

人工知能映像学習のための動画の学習対象フレームイメージサンプリング装置の各構成は、図1～図6の人工知能映像学習のための動画の学習対象フレームイメージサンプリング方法で説明した内容と同一に適用される。

【0106】

本発明の更に他の実施形態に係る人工知能映像学習装置は、学習用動画セットを用いて人工知能映像学習を行う人工知能映像学習実行部を含み、学習用動画は、上述した人工知能映像学習のための動画の学習対象フレームイメージサンプリング装置により生成される学習用動画である。

【0107】

本発明の実施形態と関連して説明された方法又はアルゴリズムの段階は、ハードウェアで直接実現するか、ハードウェアにより実行されるソフトウェアモジュールで実現するか、又はこれらの結合により実現することができる。ソフトウェアモジュールは、RAM (Random Access Memory)、ROM (Read Only Memory)、EPROM (Erasable Programmable ROM)、EEPROM (Electrically Erasable Programmable ROM)、フラッシュメモリ (Flash Memory)、ハードディスク、脱着型ディスク、CD-ROM又は本発明が属する技術分野において周知されている任意の形態のコンピュータ読み取り可能な記録媒体に存在することもできる。

【0108】

以上、添付の図面を参照して本発明の実施形態を説明したが、本発明が属する技術分野における通常の技術者は、本発明がその技術的思想や必須の特徴を変更することなく、他の具体的な形態で実施され得ることが理解できるはずである。従って、以上で記述した実施形態はあらゆる面で例示的なものであり、制限的ではないものとして理解すべきである。

【符号の説明】

【0109】

- 10 フレームイメージ
- 20 学習対象オブジェクト
- 30 背景

10

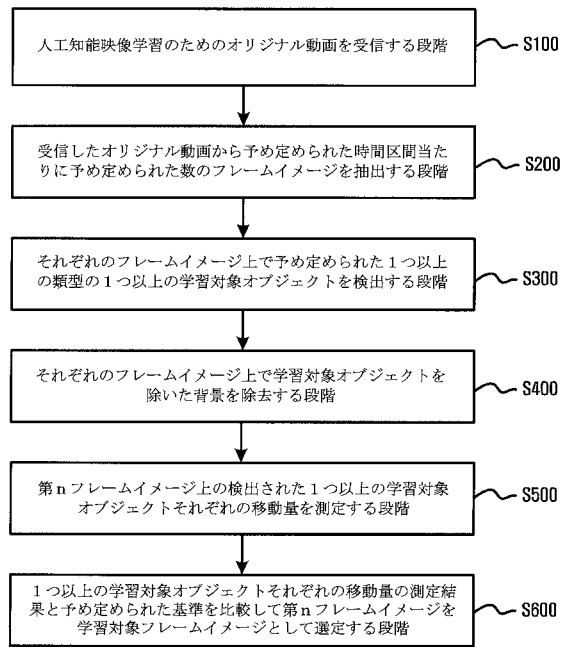
20

30

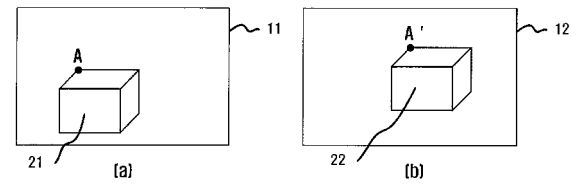
40

50

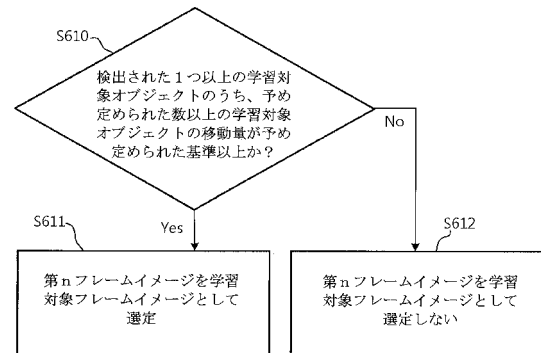
【図 1】



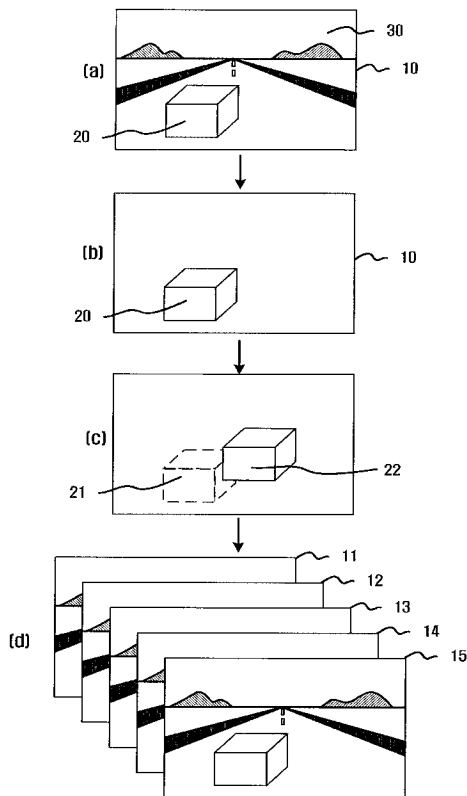
【図 2】



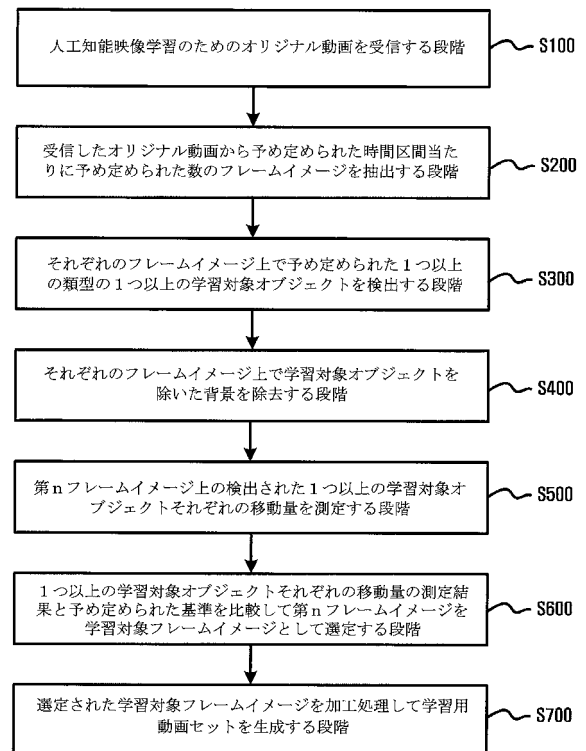
【図 3】



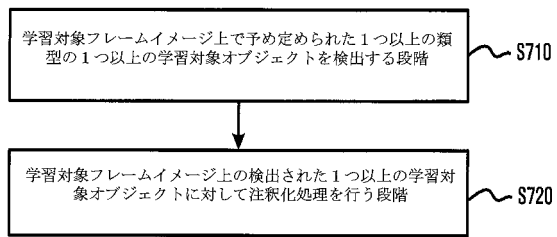
【図 4】



【図 5】



【図 6】



フロントページの続き

(72)発明者 バク, ミン ウ

大韓民国, 1 3 6 2 6 キョング - ド, ソンナム - シ, ブンダン - グ, グミ - ロ, 1 0 0 , 1 0 0
1 - 6 0 4

F ターム(参考) 5L096 AA06 BA02 BA04 BA06 BA08 BA13 CA04 DA01 GA34 HA04
KA04