(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7638398号
(P7638398)

(45)発行日　令和7年3月3日(2025.3.3)

(24)登録日　令和7年2月20日(2025.2.20)

(51)国際特許分類		FI		
G06T 7/20 (2017.01)		G06T	7/20	300Z
G06V 10/82 (2022.01)		G06V	10/82	
H04N 7/18 (2006.01)		H04N	7/18	K
G06T 7/00 (2017.01)		G06T	7/00	350C

請求項の数　15　（全22頁）

(21)出願番号　特願2023-565608(P2023-565608)
(86)(22)出願日　令和4年4月24日(2022.4.24)
(65)公表番号　特表2024-516642(P2024-516642A)
(43)公表日　令和6年4月16日(2024.4.16)
(86)国際出願番号　PCT/CN2022/088692
(87)国際公開番号　WO2022/228325
(87)国際公開日　令和4年11月3日(2022.11.3)
審査請求日　令和5年10月25日(2023.10.25)
(31)優先権主張番号　202110459730.8
(32)優先日　令和3年4月27日(2021.4.27)
(33)優先権主張国・地域又は機関
中国(CN)

(73)特許権者　511151662
中興通訊股▲ふん▼有限公司
ＺＴＥ ＣＯＲＰＯＲＡＴＩＯＮ
中華人民共和国広東省深▲せん▼市南山区高新技術産業園科技南路中興通訊大厦
ＺＴＥ Ｐｌａｚａ，Ｋｅｊｉ Ｒｏａｄ Ｓｏｕｔｈ，Ｈｉ－Ｔｅｃｈ Ｉｎｄｕｓｔｒｉａｌ Ｐａｒｋ，Ｎａｎｓｈａｎ Ｓｈｅｎｚｈｅｎ，Ｇｕａｎｇｄｏｎｇ ５１８０５７ Ｃｈｉｎａ
(74)代理人　100112656
弁理士　宮田 英毅
(74)代理人　100089118
弁理士　酒井 宏明
(72)発明者　徐茜

最終頁に続く

(54)【発明の名称】　行動検出方法、電子機器およびコンピュータ読み取り可能な記憶媒体

(57)【特許請求の範囲】

【請求項１】

ビデオストリームから複数フレームのビデオ画像フレームデータを取得するステップと、

複数フレームの前記ビデオ画像フレームデータに基づき、前記ビデオストリームにおける通行人の行動を検出するステップと、を含み、

複数フレームの前記ビデオ画像フレームデータに基づき、前記ビデオストリームにおける通行人の行動を検出する前記ステップは、

複数フレームの前記ビデオ画像フレームデータを二次元畳み込みニューラルネットワークに入力し、複数フレームの前記ビデオ画像フレームデータ間の時系列関連関係と複数フレームの前記ビデオ画像フレームデータとに基づき、前記ビデオストリームにおける通行人の行動を認識するステップを少なくとも含み、

複数フレームの前記ビデオ画像フレームデータに基づき、前記ビデオストリームにおける通行人の行動を検出する前記ステップは、

複数フレームの前記ビデオ画像フレームデータを二次元畳み込みニューラルネットワークに入力し、複数フレームの前記ビデオ画像フレームデータ間の時系列関連関係と複数フレームの前記ビデオ画像フレームデータとに基づき、前記ビデオストリームにおける通行人の行動を認識した後に、前記二次元畳み込みニューラルネットワークの出力データに基づき、前記ビデオストリームにおける通行人の行動の空間位置を検出するステップをさらに含み、

前記二次元畳み込みニューラルネットワークは、少なくとも１つの畳み込み層を含み、

複数フレームの前記ビデオ画像フレームデータを二次元畳み込みニューラルネットワークに入力し、複数フレームの前記ビデオ画像フレームデータ間の時系列関連関係と複数フレームの前記ビデオ画像フレームデータとに基づき、前記ビデオストリームにおける通行人の行動を認識する前記ステップは、
前記少なくとも１つの畳み込み層により、複数フレームの前記ビデオ画像フレームデータについて特徴抽出を行い、複数フレームの前記ビデオ画像フレームデータに一対一に対応し、それぞれが複数の特徴チャネルを含む複数の特徴マップを取得するステップと、
複数の前記特徴マップの一部の特徴チャネルを交換して、複数フレームの前記ビデオ画像フレームデータの時系列情報を融合する特徴データを取得するステップ、とを含み、前記時系列情報は、複数フレームの前記ビデオ画像フレームデータ間の時系列関連関係を特徴付けるものであり、
前記特徴データに基づき前記ビデオストリームにおける通行人の行動を認識するステップと、を含む、
行動検出方法。
【請求項２】
前記二次元畳み込みニューラルネットワークはさらに、少なくとも１つの全結合層とを含み、
前記特徴データに基づき前記ビデオストリームにおける通行人の行動を認識する前記ステップは、
前記少なくとも１つの全結合層により、前記特徴データに基づき前記ビデオストリームにおける通行人の行動を認識するステップを含む、
請求項１に記載の行動検出方法。
【請求項３】
前記二次元畳み込みニューラルネットワークは複数の直列接続された前記畳み込み層を含み、
前記少なくとも１つの畳み込み層により、複数フレームの前記ビデオ画像フレームデータについて特徴抽出を行う前記ステップは、
各前記畳み込み層について、前記畳み込み層の入力データを前記畳み込み層に入力して特徴抽出を行うステップを含み、
複数の前記特徴マップの一部の特徴チャネルを交換して、複数フレームの前記ビデオ画像フレームデータの時系列情報を融合する特徴データを取得する前記ステップは、
複数の前記特徴マップの一部の特徴チャネルを交換して第１データを取得するステップを含み、
前記畳み込み層が最初の畳み込み層である時に、前記畳み込み層の入力データは複数フレームの前記ビデオ画像フレームデータであり、
前記畳み込み層が最後の畳み込み層ではなく、かつ最初の畳み込み層でもない時に、前記第１データを次の畳み込み層の入力データとし、
前記畳み込み層が最後の畳み込み層である時に、前記第１データを前記特徴データとする、
請求項１に記載の行動検出方法。
【請求項４】
複数フレームの前記ビデオ画像フレームデータは順に並んだビデオ画像フレームデータをＮフレーム含み、複数の前記特徴マップは順に並んだ特徴マップをＮ個含み、
複数の前記特徴マップの一部の特徴チャネルを交換して第１データを取得する前記ステップは、
各前記特徴マップにおける複数の特徴チャネルを、順に並んだＮ組の特徴チャネルに分けるステップと、
Ｎ個の順に並んだ特徴マップにおけるｉ番目の特徴マップについて、ｉ番目の特徴マップに対応するｊ番目の特徴マップを特定し、ｉ番目の特徴マップはＮ個の前記順に並んだ特徴マップにおけるいずれか１つであり、ｊ番目の特徴マップはＮ個の前記順に並んだ特

徴マップにおけるいずれか１つであるステップと、
ｉ番目の特徴マップにおける第ｉ組の特徴チャネルをｊ番目の特徴マップにおけるいずれか１組の特徴チャネルと交換し、前記第１データを取得するステップと、を含み、
Ｎ、ｉ、ｊは正の整数である、
請求項３に記載の行動検出方法。
【請求項５】
前記少なくとも１つの全結合層により、前記特徴データに基づき前記ビデオストリームにおける通行人の行動を認識する前記ステップは、
前記少なくとも１つの全結合層により、前記特徴データに基づき分類特徴ベクトルを取得し、前記分類特徴ベクトルの各要素が一種類の行動類型に対応するステップと、
前記分類特徴ベクトルに基づき、各種行動類型の分類確率を特定するステップと、
前記各種行動類型の分類確率に基づき前記ビデオストリームにおける通行人の行動を認識するステップと、を含む、
請求項２に記載の行動検出方法。
【請求項６】
各種行動類型の分類確率に基づき前記ビデオストリームにおける通行人の行動を認識する前記ステップは、
各種行動類型の分類確率がフィルタ閾値を上回るかどうか判断するステップと、
少なくとも一種類の行動類型の分類確率が前記フィルタ閾値を上回った時に、対象行動を認識したと判断するステップと、
分類確率が前記フィルタ閾値を上回る行動類型を前記対象行動の類型であると特定するステップと、
各種行動類型の分類確率がいずれも前記フィルタ閾値を上回らない時に、対象行動を認識していないと判断するステップと、を含む、
請求項５に記載の行動検出方法。
【請求項７】
前記二次元畳み込みニューラルネットワークはさらに、少なくとも１つの全結合層とを含み、前記二次元畳み込みニューラルネットワークの出力データは、前記少なくとも１つの全結合層により前記特徴データに基づき取得した分類特徴ベクトルと、対象畳み込み層が出力した前記複数の特徴マップとを含み、前記対象畳み込み層は前記少なくとも１つの畳み込み層における１つであり、前記分類特徴ベクトルの各要素が一種類の行動類型に対応し、
前記二次元畳み込みニューラルネットワークの出力データに基づき、前記ビデオストリームにおける通行人の行動の空間位置を検出する前記ステップは、
前記対象畳み込み層が出力する複数の特徴マップと前記分類特徴ベクトルとに基づき対象行動の空間位置を特定するステップを含む、
請求項１に記載の行動検出方法。
【請求項８】
前記対象畳み込み層が出力する複数の特徴マップと前記分類特徴ベクトルとに基づき対象行動の空間位置を特定する前記ステップは、
前記対象畳み込み層が出力する複数の特徴マップと前記分類特徴ベクトルとに基づき、前記対象行動のエッジ輪郭を特定するステップと、
前記対象行動のエッジ輪郭に基づき前記対象行動の空間位置を特定するステップと、を含む、
請求項７に記載の行動検出方法。
【請求項９】
複数フレームの前記ビデオ画像フレームデータは所定時間長さの複数フレームのビデオ画像フレームから収集し、
前記対象畳み込み層が出力する複数の特徴マップと前記分類特徴ベクトルとに基づき前記対象行動のエッジ輪郭を特定する前記ステップは、

前記対象畳み込み層が出力する複数の特徴マップに対する前記分類特徴ベクトルの微分係数を計算し、重みマップを取得することと、
前記重みマップと前記対象畳み込み層が出力する複数の特徴マップを乗算し、複数種類の行動類型に対応する第１空間予測マップを取得することと、
前記第１空間予測マップに基づき、分類信用度が最も高い行動類型に対応する第１空間予測マップを抽出し、第２空間予測マップとすることと、
前記第２空間予測マップに基づき第３空間予測マップを生成し、前記第３空間予測マップのサイズは前記ビデオ画像フレームのサイズと同一であることと、
前記第３空間予測マップのエッジを抽出し、前記対象行動のエッジ輪郭を特定することと、を含む、
請求項８に記載の行動検出方法。
【請求項１０】
前記対象行動のエッジ輪郭に基づき前記対象行動の空間位置を特定する前記ステップは、
前記対象行動のエッジ輪郭を複数フレームの前記ビデオ画像フレームにおいて描くことを含む、
請求項９に記載の行動検出方法。
【請求項１１】
前記対象行動のエッジ輪郭に基づき前記対象行動の空間位置を特定する前記ステップは、
前記対象行動のエッジ輪郭を複数フレームの前記ビデオ画像フレームにおいて描く前記ステップの後に、前記対象行動のエッジ輪郭が描かれた複数フレームの前記ビデオ画像フレームをビデオ生成キャッシュ領域に取り込むことをさらに含む、
請求項１０に記載の行動検出方法。
【請求項１２】
複数フレームの前記ビデオ画像フレームデータを二次元畳み込みニューラルネットワークに入力し、複数フレームの前記ビデオ画像フレームデータ間の時系列関連関係と複数フレームの前記ビデオ画像フレームデータとに基づき、前記ビデオストリームにおける通行人の行動を認識した結果は、対象行動を認識したということ、または対象行動を認識していないということを含み、対象行動を認識していない時に、
エッジ輪郭が描かれたビデオ画像フレームがビデオ生成キャッシュ領域に記憶されているかどうか判断するステップと、
エッジ輪郭が描かれたビデオ画像フレームが前記ビデオ生成キャッシュ領域に記憶されている時に、前記ビデオ生成キャッシュ領域に記憶されている、エッジ輪郭が描かれたビデオ画像フレームに基づきビデオセグメントを生成するステップと、
前記ビデオ生成キャッシュ領域から前記ビデオセグメントを取り出すステップと、を含む、
請求項１に記載の行動検出方法。
【請求項１３】
前記ビデオストリームから複数フレームのビデオ画像フレームデータを取得する前記ステップは、
前記ビデオストリーム内の目下のビデオ画像フレームにおける前景画像領域の面積を特定することと、
前記前景画像領域の面積が面積閾値を上回る時に、隣接する２つのビデオ画像フレームの運動量を特定することと、
隣接する２つのビデオ画像フレームの運動量が運動量閾値を上回る時に、目下のビデオ画像フレームをサンプリング開始点と決定することと、
所定時間長さの連続する複数フレームのビデオ画像フレームから、所定数の前記ビデオ画像フレームを均一サンプリングして前処理し、複数フレームの前記ビデオ画像フレームデータを取得することと、を含む、
請求項１に記載の行動検出方法。
【請求項１４】

少なくとも１つのプロセッサと、
少なくとも１つのコンピュータプログラムが記憶され、前記少なくとも１つのコンピュータプログラムが前記少なくとも１つのプロセッサにより実行される時に、前記少なくとも１つのプロセッサに請求項１から１３のいずれか一項に記載の行動検出方法を実現させるメモリと、
前記プロセッサと前記メモリとの間に接続され、前記プロセッサと前記メモリが情報のやり取りを実現するように配置された少なくとも１つのＩ／Ｏインターフェースと、を含む、
電子機器。
【請求項１５】
コンピュータプログラムが記憶され、前記コンピュータプログラムがプロセッサにより実行される時に請求項１から１３のいずれか一項に記載の行動検出方法を実現する、
コンピュータ読み取り可能な記憶媒体。
【発明の詳細な説明】
【技術分野】
【０００１】
本願は２０２１年４月２７日に提出された中国特許出願第２０２１１０４５９７３０．８号の優先権を主張し、当該中国特許出願の内容を参照によりここに援用する。
【０００２】
本願は画像認識分野に関するものであり、特に行動検出方法、電子機器およびコンピュータ読み取り可能な記憶媒体に関するものである。
【背景技術】
【０００３】
インテリジェントビデオ監視はコンピュータビジョン技術に基づき、ビデオデータをインテリジェント分析することができ、目下のところ、警備およびインテリジェント交通などの分野に広く適用されており、大幅に警備の反応速度向上、人的資源の節減を図っている。通行人はインテリジェントビデオ監視の重点フォロー対象であり、通行人の各種行動（例えば、異常行動など）の検出および認識は警備分野の重要なニーズの１つである。
【発明の概要】
【発明が解決しようとする課題】
【０００４】
いくつかの関連技術において、インテリジェントビデオ監視はインテリジェントビデオ分析技術を使用して、膨大な量の監視ビデオから通行人の各種行動を検出して認識することで、公共安全応急管理に重要な参考を与え、公共安全突発事件の危害の低減に利することができる。しかし、通行人の行動を検出および認識する関連の技術は、いくつかの実際の適用シーンにおける配置のニーズを満たすことができない。
【課題を解決するための手段】
【０００５】
第１態様において、本願実施例は、
ビデオストリームから複数フレームのビデオ画像フレームデータを取得するステップと、
複数フレームの前記ビデオ画像フレームデータに基づき、前記ビデオストリームにおける通行人の行動を検出するステップと、を含み、
複数フレームの前記ビデオ画像フレームデータに基づき、前記ビデオストリームにおける通行人の行動を検出するステップは、
複数フレームの前記ビデオ画像フレームデータを二次元畳み込みニューラルネットワークに入力し、複数フレームの前記ビデオ画像フレームデータ間の時系列関連関係と複数フレームの前記ビデオ画像フレームデータとに基づき、前記ビデオストリームにおける通行人の行動を認識するステップを少なくとも含む、行動検出方法を提供する。
【０００６】
第２態様において、本願実施例は、

１つまたは複数のプロセッサと、
１つまたは複数のコンピュータプログラムが記憶され、前記１つまたは複数のコンピュータプログラムが前記１つまたは複数のプロセッサにより実行される時に、前記１つまたは複数のプロセッサに第１態様において本願実施例が提供する前記行動検出方法を実現させるメモリと、
前記プロセッサと前記メモリとの間に接続され、前記プロセッサと前記メモリが情報のやり取りを実現するように配置された１つまたは複数のＩ／Ｏインターフェースと、を含む、電子機器を提供する。
【０００７】
第３態様において、本願実施例は、
コンピュータプログラムが記憶され、前記コンピュータプログラムがプロセッサにより実行される時に、第１態様において本願実施例が提供する前記行動検出方法を実現する、コンピュータ読み取り可能な記憶媒体を提供する。
【図面の簡単な説明】
【０００８】
【図１】図１は本願実施例における行動検出方法のフローチャートである。
【図２】図２は本願実施例の行動検出方法における一部のステップのフローチャートである。
【図３】図３は本願実施例の行動検出方法における一部のステップのフローチャートである。
【図４】図４は本願実施例の行動検出方法における一部のステップのフローチャートである。
【図５】図５は本願実施例の行動検出方法における一部のステップのフローチャートである。
【図６】図６は本願実施例の行動検出方法における一部のステップのフローチャートである。
【図７】図７は本願実施例の行動検出方法における一部のステップのフローチャートである。
【図８】図８は本願実施例の行動検出方法における一部のステップのフローチャートである。
【図９】図９は本願実施例の行動検出方法における一部のステップのフローチャートである。
【図１０】図１０は本願実施例の行動検出方法における一部のステップのフローチャートである。
【図１１】図１１は本願実施例における電子機器の組成概略ブロック図である。
【図１２】図１２は本願実施例におけるコンピュータ読み取り可能な媒体の組成概略ブロック図である。
【図１３】図１３は本願実施例における実例の行動検出装置およびシステム構成の概略図である。
【発明を実施するための形態】
【０００９】
当業者が本願の技術案をよりよく理解できるように、以下に図面を組み合わせて本願が提供する行動検出方法、電子機器およびコンピュータ読み取り可能な記憶媒体ついて詳細に説明する。
【００１０】
以下では図面を参照して例示的な実施例についてより十分に説明するが、前記例示的な実施例は異なる形式で具体化されてよく、本明細書にて説明された実施例に限定されると解釈すべきではない。むしろ、これらの実施例を提供するのは、本願を詳らかで十全なものにするとともに、当業者に本願の範囲を十分に理解させることを目的とする。
【００１１】

矛盾することがなければ、本願の各実施例および実施例における各特徴は互いに組み合わせることができる。

【００１２】

本明細書にて使用する、「および／または」という用語は１つまたは複数の関連列挙アイテムの任意のおよび全ての組み合わせを含む。

【００１３】

本明細書にて使用する用語は特定の実施例を説明するためにのみ用いられ、本願を限定する意図はない。本明細書にて使用する単数形の「１つ」および「当該」という用語は、文脈が別途明らかに示さない限り複数形も含むことを意図する。さらに、本明細書にて「含む」および／または「……からなる」という用語を使用する場合は、特定の特徴、全体、ステップ、操作、素子および／またはコンポーネントが存在することを指すが、１つまたは複数のその他の特徴、全体、ステップ、操作、素子、コンポーネントおよび／またはそのグループが存在すること、または追加することを排除しない。

【００１４】

別途限定しない限り、本明細書にて使用する全ての用語（技術および科学用語を含む）の意味は当業者が一般的に理解する意味と同じである。さらに、一般的な辞書において限定されたこれらの用語は、関連技術および本願の背景での意味と一致する意味を有すると解釈されるべきであり、本明細書にて明確に限定しない限り、理想的された、または過度の形式上の意味を有すると解釈されない。

【００１５】

いくつかの関連技術では、主にフレーム間差分法を使用して通行人の行動を検出する。つまり、ビデオにおける連続する画像フレームのフレーム間差分の変化を分析することで異常行動領域を大まかに位置決めしてから、大まかな位置決め領域について行動認識を行い、異常行動が存在するかどうか、または異常行動の類型を特定する。フレーム間差分法に基づく通行人の異常行動検出および認識技術は光線の変化に敏感であり、かつ連続する画像フレームのフレーム間差分変化は全てが通行人の異常行動によって生じるものではなく、異常ではない多くの行動も画像フレームに大きな変化をもたらす可能性があり、ある種の異常行動は画像フレームの大きな変化をもたらすことはない。このほか、フレーム間差分法は連続した画像フレームを使用して異常行動について位置決めする必要があるため、大まかな位置決め領域について行動認識を行う際にも連続する画像フレームに基づいており、行動認識を行う際に使用する連続する画像フレームのフレーム数が少なければタイムドメイン情報の無駄を招き、行動認識を行う際に使用する連続する画像フレームのフレーム数が多ければ異常行動検出および認識の時間コストとリソース消費の増加を招くことになる。よって、当該フレーム間差分法は、通行人が少なく背景が単一のシーンにより適している。

【００１６】

別のいくつかの関連技術では、各通行人の行動を分析することで異常行動が存在するかどうか特定し、例えば、通行人の検出または骨格点の分析により、各通行人の空間位置を特定して追跡し、各通行人の時間緯度での軌跡を取得してから、通行人個々の運動画像シーケンスまたは骨格点運動シーケンスをまとめて、行動認識を行い、異常行動の類型を特定する。通行人の検出または骨格点の分析をする際には、監視装置の画角に厳しい要求があり、監視装置がハイアングルであれば通行人の骨格点の特定において問題があり、監視装置がアイアングルであれば通行人同士で互いに遮られることによって通行人の空間位置を特定できないため、誤検出および検出漏れが生じる可能性がある。このほか、通行人の検出と骨格点検出はともに大量の計算リソースの消費を必要とし、処理速度も遅いため、異常行動検出および認識のリアルタイム分析要件を満たすことができない。

【００１７】

いくつかの関連技術では三次元畳み込みまたはデュアルネットワークを使用して、異常行動の時系列情報を学習した上で異常行動の検出と認識を実現するが、三次元畳み込みおよびデュアルネットワークは短時間の時系列情報しか学習できず、例えば３×５×５の三

次元畳み込みの一回の演算は３フレームの画像しか関連付けることができず、デュアルネットワークにおけるオプティカルフロー計算は隣接するフレームの計算によって得られ、三次元畳み込みおよびデュアルネットワークはいずれも実行時に大量のリソースを消費する。いくつかの関連技術では二次元畳み込みを使用して特徴を重ねてから、三次元畳み込みを使用して時系列情報を融合して異常行動検出と認識を実現するが、三次元畳み込みを使用しているため、実行速度の向上には限りがある。いくつかの関連技術では、単一フレームのビデオ画像フレームを直接使用して、または複数フレームのビデオ画像フレームの特徴を重ねてから異常行動について分類を行っているが、フレームとフレームの間の関連関係を無視し、時系列情報を無駄にしており、検出と認識の精度は低下することになる。

【００１８】

以上を踏まえ、関連の異常行動検出と認識技術は、いくつかの実際のシーンにおける配置のニーズを満たすのが困難である。

【００１９】

この点に鑑み、第１態様において、図１を参照すれば分かるように、本願実施例は、ステップＳ１００およびステップＳ２００を含む行動検出方法を提供する。

【００２０】

ステップＳ１００では、ビデオストリームから複数フレームのビデオ画像フレームデータを取得する。

【００２１】

ステップＳ２００では、複数フレームの前記ビデオ画像フレームデータに基づき、前記ビデオストリームにおける通行人の行動を検出する。

【００２２】

上記ステップＳ２００は少なくともステップＳ３００を含む。

【００２３】

ステップＳ３００では、複数フレームの前記ビデオ画像フレームデータを二次元畳み込みニューラルネットワークに入力し、複数フレームの前記ビデオ画像フレームデータ間の時系列関連関係と複数フレームの前記ビデオ画像フレームデータとに基づき、前記ビデオストリームにおける通行人の行動を認識する。

【００２４】

いくつかの実施の形態において、ステップＳ１００におけるビデオストリームは監視装置により取得したものである。本願実施例において、ビデオストリームは監視装置によりリアルタイムで取得してよく、監視装置により取得してからコンピュータ読み取り可能な媒体に記憶してもよい。本願実施例ではこれについて特に限定しない。なお、本願実施例において、毎回行われる行動検出はいずれも一定時間長さのビデオストリームに対応し、つまり、複数フレームのビデオ画像フレームデータは一定時間長さのビデオストリームから取得する。

【００２５】

いくつかの実施の形態において、ステップＳ１００における複数フレームのビデオ画像フレームデータは、本願実施例における二次元畳み込みニューラルネットワークで処理できるデータである。本願実施例では、ビデオストリームのデコード後に複数フレームのビデオ画像を取得することができ、各フレームのビデオ画像フレームデータはいずれも１フレームのビデオ画像フレームに対応する。

【００２６】

通行人の行動には時間的な連続性があり、これに応じて、ビデオストリームのデコードで得られる複数フレームのビデオ画像フレームの間にも時系列上関連関係があり、複数フレームのビデオ画像フレームにそれぞれ対応する複数フレームのビデオ画像フレームデータの間にも時系列上関連関係がある。本願実施例にて使用する二次元畳み込みニューラルネットワークは、各フレームのビデオ画像フレームデータの特徴を学習できるだけでなく、複数フレームのビデオ画像フレームデータ間の時系列関連関係を学習することもできることから、複数フレームのビデオ画像フレームデータ間の時系列関連関係と複数フレーム

のビデオ画像フレームデータとに基づき通行人の行動検出を行うことができる。
【００２７】
いくつかの実施の形態において、複数フレームのビデオ画像フレームに基づき前記ビデオストリームにおける通行人の行動を検出する前記ステップでは、ビデオストリームにおける通行人の行動についてのみ認識してよく、前記の通行人の行動についての認識は、通行人の行動が存在するかどうか特定すること、通行人の行動の類型を特定することなどを含むがこれらに限定されない。いくつかの実施の形態において、複数フレームのビデオ画像フレームに基づき前記ビデオストリームにおける通行人の行動を検出する前記ステップでは、ビデオストリームにおける通行人の行動を認識してから、通行人の行動の空間位置を検出してよい。本願実施例では通行人の行動の類型について特に限定しない。いくつかの実施の形態において、通行人の行動は例えば、転倒、殴り合いなどの異常行動を含んでよく、駆け足、飛び跳ねなどの正常行動を含んでもよい。
【００２８】
本願実施例が提供する行動検出方法では、二次元畳み込みニューラルネットワークに基づき通行人の行動を検出し、使用する二次元畳み込みニューラルネットワークは、各フレームのビデオ画像フレームデータの特徴を学習できるだけでなく、複数フレームのビデオ画像フレームデータ間の時系列関連関係を学習することもできることから、複数フレームのビデオ画像フレームデータ間の時系列関連関係と複数フレームのビデオ画像フレームデータとに基づき、ビデオストリームにおける通行人の行動を認識することができる。三次元畳み込みまたはデュアルネットワークを使用して行動検出を行うものと比べて、本願実施例における、二次元畳み込みニューラルネットワークを使用した通行人の行動検出は、計算量が少なく、実行速度が速く、リソース消費が少なく、実際の配置におけるリアルタイム性要件を満たすことができる。単一フレームのビデオ画像フレームを直接使用するもの、または複数フレームのビデオ画像フレームの特徴を重ねてから行動検出を行うものと比べて、本願実施例における二次元畳み込みニューラルネットワークは通行人の行動の時系列特徴を学習していることから、誤検出、検出漏れを効果的に回避することができるとともに、検出精度を向上できる。このほか、本願実施例では、通行人の行動の類型を直接認識すること、または通行人の行動の類型を認識してから、通行人の行動の空間位置を特定することができ、通行人の行動の領域を大まかに位置決めしてから、行動の類型を特定することによる適用シーンの制限を回避することができ、行動検出のシーン適応性を大幅に向上させている。
【００２９】
本願実施例では二次元畳み込みニューラルネットワークの構造について特に限定しない。いくつかの実施の形態において、二次元畳み込みニューラルネットワークは、少なくとも１つの畳み込み層と少なくとも１つの全結合層とを含む。いくつかの実施の形態において、複数フレームのビデオ画像フレームデータはバッチ処理方式で二次元畳み込みニューラルネットワークに入力され、二次元畳み込みニューラルネットワークにおける畳み込み層は入力データについて特徴抽出を行うことができ、全結合層は畳み込み層により取得した特徴データに基づき今回の通行人の行動検出に対応するビデオストリームにおける通行人の行動を特定する。
【００３０】
関連して、いくつかの実施の形態において、前記二次元畳み込みニューラルネットワークは、少なくとも１つの畳み込み層と少なくとも１つの全結合層とを含み、図２を参照すれば分かるように、ステップＳ３００はステップＳ３１０およびＳ３２０を含む。
【００３１】
ステップＳ３１０では、前記少なくとも１つの畳み込み層により、複数フレームの前記ビデオ画像フレームデータについて特徴抽出を行い、特徴データを取得し、前記特徴データは複数フレームの前記ビデオ画像フレームデータの時系列情報を融合し、前記時系列情報は複数フレームの前記ビデオ画像フレームデータ間の時系列関連関係を特徴付けるものである。

【００３２】
ステップＳ３２０では、前記少なくとも１つの全結合層により、前記特徴データに基づき前記ビデオストリームにおける通行人の行動を認識する。
【００３３】
なお、いくつかの実施の形態において、前記特徴データは複数フレームのビデオ画像フレームデータの時系列情報を融合するということは、特徴データは各フレームのビデオ画像フレームデータの特徴を特徴付けることができ、複数フレームのビデオ画像フレームデータ間の時系列関連関係を特徴づけることもできることをいう。
【００３４】
いくつかの実施の形態において、前記二次元畳み込みニューラルネットワークはプーリング層をさらに含む。
【００３５】
いくつかの実施の形態において、前記二次元畳み込みニューラルネットワークにおける複数の畳み込み層は直列接続されており、各畳み込み層が入力データについて特徴抽出を行い、入力データに対応する特徴マップを取得する。複数フレームのビデオ画像フレームデータがバッチ処理方式で二次元畳み込みニューラルネットワークに入力される時に、各畳み込み層が入力データについて特徴抽出を行えば複数の特徴マップを取得できる。本願実施例において、バッチ処理方式で二次元畳み込みニューラルネットワークに入力される複数フレームのビデオ画像フレームデータは時系列順に並べられ、各畳み込み層が取得する複数の特徴マップも時系列順に並べられる。いくつかの実施の形態では、各畳み込み層後に、複数の特徴マップの一部の特徴チャネルを交換することで、異なる時系列情報の相互融合を実現し、最終的に特徴データにおいて複数フレームの前記ビデオ画像フレームデータの時系列データの融合を実現する。
【００３６】
関連して、いくつかの実施の形態において、前記二次元畳み込みニューラルネットワークは複数の直列接続された前記畳み込み層を含み、図３を参照すれば分かるように、ステップＳ３１０はステップＳ３１１およびステップＳ３１２を含む。
【００３７】
ステップＳ３１１では、各前記畳み込み層について、前記畳み込み層の入力データを前記畳み込み層に入力して特徴抽出を行い、複数の特徴マップを取得し、複数の前記特徴マップは複数フレームの前記ビデオ画像フレームデータに一対一で対応し、各前記特徴マップは複数の特徴チャネルを含む。
【００３８】
ステップＳ３１２では、複数の前記特徴マップの一部の特徴チャネルを交換して第１データを取得する。
【００３９】
いくつかの実施の形態では、前記畳み込み層が最初の畳み込み層である時に、前記畳み込み層の入力データは複数フレームの前記ビデオ画像フレームデータである。
【００４０】
いくつかの実施の形態では、前記畳み込み層が最後の畳み込み層ではなく、かつ最初の畳み込み層でもない時に、前記第１データを次の畳み込み層の入力データとする。
【００４１】
いくつかの実施の形態では、前記畳み込み層が最後の畳み込み層である時に、前記第１データを前記特徴データとする。
【００４２】
なお、いくつかの実施の形態において、複数の特徴マップの一部の特徴チャネルの交換はデータの移動を行うだけで、加算乗算操作はないため、時系列情報のやり取りを行う際に計算量は増加せず、データ移動の速度が速く、行動検出の実行効率に影響しない。
【００４３】
関連して、いくつかの実施の形態において、複数フレームの前記ビデオ画像フレームデ

ータは順に並んだビデオ画像フレームデータをＮフレーム含み、複数の前記特徴マップは順に並んだ特徴マップをＮ個含み、図４を参照すれば分かるように、ステップＳ３１２はステップＳ３１２１からステップＳ３１２３を含む。

【００４４】

ステップＳ３１２１では、各前記特徴マップにおける複数の特徴チャネルを、順に並んだＮ組の特徴チャネルに分ける。

【００４５】

ステップＳ３１２２では、Ｎ個の順に並んだ特徴マップにおけるｉ番目の特徴マップについて、ｉ番目の特徴マップに対応するｊ番目の特徴マップを特定し、ｉ番目の特徴マップはＮ個の前記順に並んだ特徴マップにおけるいずれか１つであり、ｊ番目の特徴マップはＮ個の前記順に並んだ特徴マップにおけるいずれか１つである。

【００４６】

ステップＳ３１２３では、ｉ番目の特徴マップにおける第ｉ組の特徴チャネルをｊ番目の特徴マップにおけるいずれか１組の特徴チャネルと交換し、前記第１データを取得し、Ｎ、ｉ、ｊは正の整数である。

【００４７】

本願実施例ではステップＳ３１２２においてｉ番目の特徴マップに対応するｊ番目の特徴マップを特定することを如何に実行するかについて特に限定しない。いくつかの実施の形態では、ｉとｊの代数関係に基づき、ｉ番目の特徴マップに対応するｊ番目の特徴マップを特定し、いくつかの実施の形態では、ｉとｊの隣接関係に基づき、ｉ番目の特徴マップに対応するｊ番目の特徴マップを特定し、いくつかの実施の形態では、Ｎ個の前記順に並んだ特徴マップから特徴マップを１つランダムに指定し、ｉ番目の特徴マップに対応するｊ番目の特徴マップとする。

【００４８】

いくつかの実施の形態では、二次元畳み込みニューラルネットワークの全結合層により分類特徴ベクトルを取得し、分類特徴ベクトルの各要素が一種類の行動類型の分類確率を特徴付ける。各種行動類型の分類確率に基づき、今回の通行人の行動検出に対応するビデオストリームから通行人の行動を特定することができる。今回の通行人の行動検出に対応するビデオストリームから通行人の行動を特定することは、今回の通行人の行動検出に対応するビデオストリームに、検出しようとする対象行動があるかどうか判断することと、存在する対象行動の類型を判断することと、を含むが、これらに限定されない。

【００４９】

関連して、いくつかの実施の形態では、図５を参照すれば分かるように、ステップＳ３２０はステップＳ３２１からステップＳ３２３を含む。

【００５０】

ステップＳ３２１では、前記少なくとも１つの全結合層により、前記特徴データに基づき分類特徴ベクトルを取得し、前記分類特徴ベクトルの各要素が一種類の行動類型に対応する。

【００５１】

ステップＳ３２２では、前記分類特徴ベクトルに基づき、各種行動類型の分類確率を特定する。

【００５２】

ステップＳ３２３では、前記各種行動類型の分類確率に基づき前記ビデオストリームにおける通行人の行動を認識する。

【００５３】

いくつかの実施の形態では、ステップＳ３２２において、ステップＳ３２１で取得した分類特徴ベクトルを分類器に入力し、各種行動類型の分類確率を取得する。

【００５４】

関連して、いくつかの実施の形態では、図６を参照すれば分かるように、ステップＳ３２３はステップＳ３２３１からステップＳ３２３４を含む。

【００５５】
ステップＳ３２３１では、各種行動類型の分類確率がフィルタ閾値を上回るかどうか判断する。
【００５６】
ステップＳ３２３２では、少なくとも一種類の行動類型の分類確率が前記フィルタ閾値を上回った時に、対象行動を認識したと判断する。
【００５７】
ステップＳ３２３３では、分類確率が前記フィルタ閾値を上回る行動類型を前記対象行動の類型であると特定する。
【００５８】
ステップＳ３２３４では、各種行動類型の分類確率がいずれも前記フィルタ閾値を上回らない時に、対象行動を認識していないと判断する。
【００５９】
二次元畳み込みニューラルネットワークの畳み込み層は空間不変性を有し、つまり、畳み込み層を介して取得した特徴マップと初期画像との間には空間対応関係があり、トレーニングを経て取得した二次元畳み込みニューラルネットワークの畳み込み層は、特徴マップにおける分類に関連する領域特徴値を大きくし、分類に関係のない領域特徴値を小さくすることができる。いくつかの実施の形態では、対象行動を認識した時に、二次元畳み込みニューラルネットワークの畳み込み層が出力する特徴マップについて輪郭分析を行うことで、対象行動に関連する領域のエッジ輪郭を特定して、認識した対象行動の空間位置を特定することができる。いくつかの実施の形態では、二次元畳み込みニューラルネットワークの少なくとも１つの畳み込み層の中から畳み込み層を１つ指定して対象畳み込み層とし、対象畳み込み層が出力する特徴マップと、二次元畳み込みニューラルネットワークの全結合層が取得した分類特徴ベクトルとに基づき輪郭分析を行う。
【００６０】
いくつかの実施の形態では、図７を参照すれば分かるように、ステップＳ３００の後に、ステップＳ２００はステップＳ４００をさらに含む。
【００６１】
ステップＳ４００では、前記二次元畳み込みニューラルネットワークの出力データに基づき、前記ビデオストリームにおける通行人の行動の空間位置を検出する。
【００６２】
いくつかの実施の形態において、ステップＳ４００は、ステップＳ３００により認識した通行人の行動の状況下で実行してよい。いくつかの実施の形態では、毎回ステップＳ３００を実行した後にいずれもステップＳ４００を実行し、つまり、通行人の行動を認識したかどうかに関わらずステップＳ４００を実行する。いくつかの実施の形態では、異なる適用シーンに対して、ステップＳ４００の実行またはスキップを選択してよい。
【００６３】
関連して、いくつかの実施の形態では、前記二次元畳み込みニューラルネットワークは、少なくとも１つの畳み込み層と少なくとも１つの全結合層とを含み、前記二次元畳み込みニューラルネットワークの出力データは、前記少なくとも１つの全結合層により取得した分類特徴ベクトルと、対象畳み込み層が出力した複数の特徴マップとを含み、前記対象畳み込み層は前記少なくとも１つの畳み込み層における１つであり、前記分類特徴ベクトルの各要素が一種類の行動類型に対応し、図８を参照すれば分かるように、ステップＳ４００はステップＳ４１０を含む。
【００６４】
ステップＳ４１０では、前記対象畳み込み層が出力する複数の特徴マップと前記分類特徴ベクトルとに基づき対象行動の空間位置を特定する。
【００６５】
関連して、いくつかの実施の形態では、図９を参照すれば分かるように、ステップＳ４１０はステップＳ４１１およびステップＳ４１２を含む。

ステップＳ４１１では、前記対象畳み込み層が出力する複数の特徴マップと前記分類特徴ベクトルとに基づき、前記対象行動のエッジ輪郭を特定する。

【００６６】

ステップＳ４１２では、前記対象行動のエッジ輪郭に基づき前記対象行動の空間位置を特定する。

【００６７】

関連して、いくつかの実施の形態において、上記ステップＳ４１２は、前記対象畳み込み層が出力する複数の特徴マップに対する前記分類特徴ベクトルの微分係数を計算し、重みマップを取得することと、前記重みマップと前記対象畳み込み層が出力する複数の特徴マップを乗算し、複数種類の行動類型に対応する第１空間予測マップを取得することと、前記第１空間予測マップに基づき、分類信用度が最も高い行動類型に対応する第１空間予測マップを抽出し、第２空間予測マップとすることと、前記第２空間予測マップに基づき第３空間予測マップを生成し、前記第３空間予測マップのサイズは前記ビデオ画像フレームのサイズと同一であることと、前記第３空間予測マップのエッジを抽出し、前記対象行動のエッジ輪郭を特定することと、を含んでよい。

【００６８】

関連して、いくつかの実施の形態において、複数フレームの前記ビデオ画像フレームデータは所定時間長さの複数フレームのビデオ画像フレームから収集し、上記ステップＳ４１２は、前記対象畳み込み層が出力する複数の特徴マップに対する前記分類特徴ベクトルの微分係数を計算して、ゼロ未満の微分係数値をゼロとし、重みマップを取得することと、前記重みマップと前記対象畳み込み層が出力する複数の特徴マップを乗算して、ゼロ未満の積をゼロとし、複数種類の行動類型に対応する第１空間予測マップを取得することと、前記第１空間予測マップに基づき、分類信用度が最も高い行動類型に対応する第１空間予測マップを抽出し、第２空間予測マップとすることと、前記第２空間予測マップを正規化処理することと、正規化処理後の第２空間予測マップのサイズを前記ビデオ画像フレームのサイズにスケーリングして二値化処理を行い、第３空間予測マップを取得することと、前記第３空間予測マップについてエッジ抽出を行い、前記対象行動のエッジ輪郭を特定することと、を含んでよい。

【００６９】

いくつかの実施の形態において、上記ステップＳ４１２は、前記対象行動のエッジ輪郭を複数フレームの前記ビデオ画像フレームにおいて描くことを含む。

【００７０】

いくつかの実施の形態では、対象行動を認識して対象行動の空間位置を特定した後に、対象行動の輪郭が描かれたビデオ画像フレームをビデオ生成キャッシュ領域に書き込み、さらにビデオファイルを生成してファイルシステムに記憶する。

【００７１】

関連して、いくつかの実施の形態において、上記ステップＳ４１２は、前記対象行動のエッジ輪郭を複数フレームの前記ビデオ画像フレームにおいて描いた後に、前記対象行動のエッジ輪郭が描かれた複数フレームの前記ビデオ画像フレームをビデオ生成キャッシュ領域に取り込むことをさらに含む。

【００７２】

いくつかの実施の形態において、毎回の通行人の行動検出は一定時間長さのビデオストリームに対応し、ステップＳ１００により一定時間長さのビデオストリームから複数フレームのビデオ画像フレームデータを取得する。本願実施例では当該一定時間長さについて特に限定しない。いくつかの実施の形態では、検出を要する通行人の行動が一般的に発生する時間長さに基づき当該一定時間長さを決定する。例えば、殴り合いは２秒、転倒は１秒とするなどである。

【００７３】

いくつかの実施の形態において、通行人の行動が持続する時間長さは上述の一定時間長さよりも長いことがあり得る。いくつかの実施の形態では、毎回の検出で対象行動を認識

すると、対象行動の輪郭が描かれたビデオ画像フレームをビデオ生成キャッシュ領域に書き込み、ある回の検出で対象行動が認識されなくなり、対象行動が終了したことを表すまで続け、このような場合、ビデオ生成キャッシュ領域におけるビデオ画像フレームはビデオセグメントに変換され、対象行動の開始から終了までの全過程が記録されたビデオセグメントを取得することができ、当該ビデオセグメントに基づき、対象行動の開始時間、終了時間、持続時間などの情報を特定することもできる。関連して、図１０を参照すれば分かるように、いくつかの実施の形態では、複数フレームの前記ビデオ画像フレームデータを二次元畳み込みニューラルネットワークに入力し、複数フレームの前記ビデオ画像フレームデータ間の時系列関連関係と複数フレームの前記ビデオ画像フレームデータとに基づき、前記ビデオストリームにおける通行人の行動を認識した結果は、対象行動を認識したということ、または対象行動を認識していないということを含み、対象行動を認識していない時に、前記行動検出方法はステップＳ５０１からステップＳ５０３をさらに含む。

【００７４】

ステップＳ５０１では、エッジ輪郭が描かれたビデオ画像フレームがビデオ生成キャッシュ領域に記憶されているかどうか判断する。

【００７５】

ステップＳ５０２では、エッジ輪郭が描かれたビデオ画像フレームが前記ビデオキャッシュ領域に記憶されている時に、前記ビデオキャッシュ領域に記憶されている、エッジ輪郭が描かれたビデオ画像フレームに基づきビデオセグメントを生成する。

【００７６】

ステップＳ５０３では、前記ビデオ生成キャッシュ領域から前記ビデオセグメントを取り出す。

【００７７】

いくつかの実施の形態において、ビデオ生成キャッシュ領域から前記ビデオセグメントを取り出す前記ステップは、ビデオ生成キャッシュ領域におけるビデオセグメントをファイルシステムに記憶することと、ビデオ生成キャッシュ領域を空にすることと、を含む。

【００７８】

いくつかの実施の形態において、上記ステップＳ１００は、前記ビデオストリームを取得することと、前記ビデオストリームをデコードして、連続する複数のビデオ画像フレームを取得することと、連続する複数のビデオ画像フレームについてサンプリングを行い、複数の検出対象ビデオ画像フレームを取得することと、複数の前記検出対象ビデオ画像フレームについて前処理を行い、複数フレームの前記ビデオ画像フレームデータを取得することと、を含む。

【００７９】

なお、本願実施例ではビデオストリームを如何にデコードするかについて特に限定しない。いくつかの実施の形態では、グラフィックスプロセッサ（ＧＰＵ、Ｇｒａｐｈｉｃｓ Ｐｒｏｃｅｓｓｉｎｇ Ｕｎｉｔ）を使用してビデオストリームをデコードする。

【００８０】

なお、本願実施例では連続する複数のビデオ画像フレームを如何にサンプリングするかについて特に限定しない。いくつかの実施の形態では、複数のビデオ画像フレームにおいてランダムにサンプリングを行う。いくつかの実施の形態では、所定の間隔で複数のビデオ画像フレームにおいてサンプリングを行う。いくつかの実施の形態では、複数のビデオ画像フレームにおいて連続してサンプリングを行う。

【００８１】

なお、前記の所定の間隔で複数のビデオ画像フレームにおいてサンプリングを行うことは、前記の複数のビデオ画像フレームにおいて連続してサンプリングを行うことと比べて、より多くの時系列情報を取得することができるため、検出精度が向上する。

【００８２】

本願実施例では複数の検出対象のビデオ画像フレームを如何に前処理するかについて特に限定しない。いくつかの実施の形態において、複数の検出対象のビデオ画像フレームに

対して前処理を行うことは、各検出対象のビデオ画像フレームのサイズを所定のサイズに調整することと、所定のサイズに調整された検出対象ビデオ画像フレームについて色空間変換処理、画素値正規化処理、平均値を引いて標準偏差で除す処理を行い、複数フレームのビデオ画像フレームデータを取得することと、を含む。

【００８３】

いくつかの実施の形態において、毎回の通行人の行動検出は一定時間長さのビデオストリームに対応し、均一サンプリグ方式で一定時間長さのビデオ画像フレームにおいて所定のフレーム数のビデオ画像フレームデータを取得する。

【００８４】

関連して、いくつかの実施の形態において、上記ステップＳ１００は、前記ビデオストリーム内の目下のビデオ画像フレームにおける前景画像領域の面積を特定することと、前記前景画像領域の面積が面積閾値を上回る時に、隣接する２つのビデオ画像フレームの運動量を特定することと、隣接する２つのビデオ画像フレームの運動量が運動量閾値を上回る時に、目下のビデオ画像フレームをサンプリング開始点と決定することと、所定時間長さの連続する複数フレームのビデオ画像フレームから、所定数の前記ビデオ画像フレームを均一サンプリングして前処理し、複数フレームの前記ビデオ画像フレームデータを取得することと、を含む。

【００８５】

本願実施例では目下のビデオ画像フレームにおける前景画像領域の面積を如何に特定するかについて特に限定しない。いくつかの実施の形態では、フレーム間差分法を使用して目下のビデオ画像フレームの前景画像を取得する。

【００８６】

本願実施例では隣接する２つのビデオ画像フレームの運動量を如何に特定するかについて特に限定しない。いくつかの実施の形態では、まばらなオプティカルフローを使用して隣接する２つのビデオ画像フレームの運動量を計算する。

【００８７】

いくつかの実施の形態では、ステップＳ１００からステップＳ２００により行動検出を行う前に、二次元畳み込みニューラルネットワークをトレーニングするステップをさらに含み、ビデオストリームを取得し、ビデオストリームをデコードしてビデオ画像フレームを生成し、データクレンジングを行ってサンプルビデオセグメントを取得し、サンプルビデオセグメントにおける通行人の行動の類型をマークし、検出を要する通行人の行動がないサンプルビデオセグメントを背景としてマークし、マークされたサンプルビデオセグメントを使用して二次元畳み込みニューラルネットワークをトレーニングし、トレーニングされた二次元畳み込みニューラルネットワークについて定量化操作を行い、フォーマット変換する。

【００８８】

第２態様において、図１１を参照すれば分かるように、本願実施例は、

１つまたは複数のプロセッサ１０１と、

１つまたは複数のコンピュータプログラムが記憶され、前記１つまたは複数のコンピュータプログラムが１つまたは複数のプロセッサ１０１により実行される時に、１つまたは複数のプロセッサ１０１に第１態様において本願実施例が提供する前記行動検出方法を実現させるメモリ１０２と、

プロセッサ１０１とメモリ１０２との間に接続され、プロセッサ１０１とメモリ１０２が情報のやり取りを実現するように配置された１つまたは複数のＩ／Ｏインターフェース１０３と、を含む、電子機器を提供する。

【００８９】

プロセッサ１０１はデータ処理機能を有するデバイスであり、中央プロセッサ（ＣＰＵ）などを含むがこれらに限定されず、メモリ１０２はデータ記憶機能を有するデバイスであり、ランダムアクセスメモリ（ＲＡＭ。より具体的にはＳＤＲＡＭ、ＤＲＲなど）、読み取り専用メモリ（ＲＯＭ）、電気的に消去、プログラムが可能な読み取り専用メモリ（

ＥＥＰＲＯＭ）、フラッシュメモリ（ＦＬＡＳＨ）を含むがこれらに限定されず、Ｉ／Ｏインターフェース（読み書きインターフェース）１０３はプロセッサ１０１とメモリ１０２との間に接続され、プロセッサ１０１とメモリ１０２の情報のやり取りを実現することができ、データバス（Ｂｕｓ）などを含むがこれらに限定されない。

【００９０】

いくつかの実施の形態において、プロセッサ１０１、メモリ１０２、Ｉ／Ｏインターフェース１０３はバス１０４により互いに接続されて、さらにはコンピュータ装置のその他のコンポーネントに接続される。

【００９１】

第３態様において、図１２を参照すれば分かるように、本願実施例は、コンピュータプログラムが記憶され、前記コンピュータプログラムがプロセッサにより実行される時に、第１態様において本願実施例が提供する前記行動検出方法を実現する、コンピュータ読み取り可能な記憶媒体を提供する。

【００９２】

本願実施例が提供する技術案を当業者がより明確に理解できるように、以下では具体的な実例を通じて本願実施例が提供する技術案について詳細に説明する。

【００９３】

実例１

図１３は本願実施例における実例の行動検出装置およびシステム構成の概略図である。

【００９４】

図１３に示すように、前記行動検出装置は行動認識モジュールと、行動位置検出モジュールと、ビデオ自動記憶モジュールと、を含む。本実例１において、前記行動検出装置はサーバに配置され、サーバはＧＰＵと、ＣＰＵと、ネットワークインターフェースと、ビデオメモリと、内部メモリと、をさらに含み、内部バスにより互いに接続される。

【００９５】

実例２

本実例２において行動検出を行う際のフローは以下の通りである。

【００９６】

二次元畳み込みニューラルネットワークをビデオメモリまたは内部メモリにロードして初期化し、入力画像サイズ制限、バッチ処理の大きさ、フィルタ閾値、面積閾値、行動の一般的な発生時間長さ、フレームレートなどのパラメータを配置する。

【００９７】

カメラからビデオストリームを取得して、取得したビデオストリームをシステムサーバのＧＰＵに送りハードウェアデコードし、複数フレームのビデオ画像フレームを生成する。

【００９８】

複数フレームのビデオ画像フレームから行動の一般的な発生時間長さに対応するビデオ画像フレームセットを抽出し、そこからＮフレームを均一サンプリングし、Ｎは二次元畳み込みニューラルネットワークのトレーニング時に決定される。

【００９９】

均一サンプリングで取得したＮフレームのビデオ画像フレームを前処理し、その長辺を二次元畳み込みニューラルネットワーク指定のサイズに固定し、等比例スケーリングを行い、短辺塗りつぶし画素値を０としてからＲＧＢ色空間に変換し、最後に画素値を０から１の間に正規化して、平均値を引いて標準偏差で除し、前処理後のＮフレームのビデオ画像フレームデータを取得する。

【０１００】

前処理を経て取得したＮフレームのビデオ画像フレームを、Ｎの二次元データセットとしてバッチ処理され二次元畳み込みニューラルネットワークに送られるものとし、二次元畳み込みニューラルネットワークの各畳み込み層は全てサイズがＮＣＨＷであるとの特徴を得られ、ＮフレームのサイズをＣＨＷの特徴マップとして特徴付け、Ｃは特徴チャネル数を表し、ＨおよびＷは特徴マップの幅と高さをそれぞれ表す。

【０１０１】
各畳み込み層が出力するＮフレームの特徴マップの特徴チャネルをＮ組に分け、ｉ番目の特徴マップについて、ｉ番目の特徴マップ以外のその他の特徴マップから特徴マップを１つランダムに抽出し、ｊ番目の特徴マップと記し、ｊ番目の特徴マップのＮ組の特徴チャネルにおいて１組の特徴チャネルをランダムに選択し、これをｉ番目の特徴マップのi組の特徴チャネルと交換することで、計算量を別途増やさずに時系列情報のやり取りを行う。
【０１０２】
特徴チャネルの交換を経た特徴マップを次の畳み込み層に送り計算し、最後の畳み込み層後に特徴データを取得し、全結合層により分類特徴ベクトルを生成する。
分類特徴ベクトルを分類器に送り、各種行動類型の分類確率を取得する。
【０１０３】
少なくとも一種類の行動類型の分類確率がフィルタ閾値を上回った時に、対象行動を認識したと判断する。
【０１０４】
分類確率がフィルタ閾値を上回る行動類型を対象行動の類型であると特定する。
【０１０５】
各種行動類型の分類確率がいずれもフィルタ閾値を上回らない時に、対象行動を認識していないと判断する。
【０１０６】
対象行動を認識した時に、対象行動の空間位置を特定するステップは、
二次元畳み込みニューラルネットワークにおける対象畳み込み層の特徴マップと、全結合層が出力する分類特徴ベクトルを抽出し、対象畳み込み層の特徴マップに対する分類特徴ベクトルの微分係数を計算し、微分係数の値が０未満の微分係数の値を０とし、対象畳み込み特徴マップと空間の大きさが一致する重みマップを取得することと、
重みマップと対象畳み込み層の特徴マップを乗算して、積における０未満の値を０として第１空間予測マップを取得することと、
第１空間予測マップをＮ*Ｃｌａｓｓ*Ｈ*Ｗとの記憶形式に変換し、Ｃｌａｓｓは行動類型数を表し、その後、類型緯度において第１空間予測マップについてｓｏｆｔｍａｘ操作を行い、分類結果の信用度が最も高い類型に対応する緯度を抽出し、Ｎ*Ｈ*Ｗの空間予測マップを取得し、特徴マップにおける全ての要素の、バッチ処理数Ｎの所在緯度における最大値を計算し、Ｈ*Ｗ第２空間予測マップを取得することと、
第２空間予測マップにおける全ての要素から全ての要素の最小値を引き、さらに全ての要素の最大値で除して、第２空間予測マップを０から１の間に正規化することと、
正規化後の第２空間予測マップをビデオ画像フレームのサイズにスケーリングして二値化処理を行い（要素値が０．５を上回れば１、そうでなければ０とする）、第３空間予測マップを取得することと、
第３空間予測マップについてエッジ輪郭抽出を行い、取得したエッジ輪郭は対象行動が存在する空間位置であることと、
輪郭境界を検出結果としてビデオ画像フレームにおいて描いて、ビデオ画像フレームをビデオ生成キャッシュ領域に取り込み、次の検出を開始することと、を含む。
ビデオセグメントを記憶するステップは、
対象行動を認識していない時に、ビデオ生成キャッシュ領域にビデオ画像フレームがあるかどうか判断する（つまり、前回の検出が対象行動を認識したかどうか判断する）ことと、
ビデオキャッシュ領域にビデオ画像フレームがある（つまり、前回の検出が対象行動を認識した）時に、ビデオキャッシュ領域におけるビデオ画像フレームに基づきビデオセグメントを生成することと、
ビデオセグメントを記憶することと、
ビデオ生成キャッシュ領域を空にして次の検出を開始すること、または、
ビデオキャッシュ領域にビデオ画像フレームがない（つまり、前回の検出で対象行動を

認識していない）時に、次の検出を開始することと、を含む。

【０１０７】

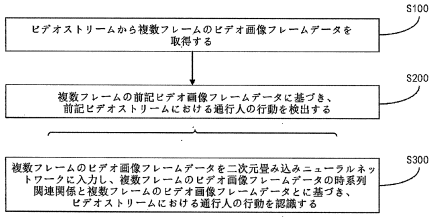
本明細書にて開示した方法のうちの全てまたはいくつかのステップ、装置における機能モジュール／ユニットはソフトウェア、ファームウェア、ハードウェアおよびその適切な組み合わせとして実施されてよいと当業者は理解できる。ハードウェアの実施の形態において、以上の説明で言及した機能モジュール／ユニットの間の区分は必ずしも物理的なコンポーネントの区分に対応せず、例えば、１つの物理的コンポーネントは複数の機能を有してよく、あるいは１つの機能またはステップは複数の物理的コンポーネントが連携して実行することができる。ある物理的コンポーネントまたは全ての物理的コンポーネントは、（中央処理装置、デジタル信号プロセッサまたはマイクロプロセッサのような）プロセッサにより実行されるソフトウェアとして実施されてよく、またはハードウェアとして、あるいは専用集積回路のような集積回路として実施されてもよい。このようなソフトウェアはコンピュータ読み取り可能な媒体に配置することができ、コンピュータ読み取り可能な媒体はコンピュータ記憶媒体（または非一時的な媒体）および通信媒体（または一時的な媒体）を含んでよい。当業者に知られているように、コンピュータ記憶媒体という技術用語は（コンピュータ読み取り可能な命令、データ構造、プログラムモジュールまたはその他のデータのような）情報を記憶するための任意の方法または技術において実施される揮発性および不揮発性、リムーバブルなおよび非リムーバブルな媒体を含む。コンピュータ記憶媒体はＲＡＭ、ＲＯＭ、ＥＥＰＲＯＭ、フラッシュメモリまたはその他のメモリ技術、ＣＤ－ＲＯＭ、デジタル多機能ディスク（ＤＶＤ）またはその他の光ディスクメモリ、磁気ボックス、磁気テープ、磁気ディスクメモリまたはその他の磁気記憶装置、または所望の情報を記憶しかつコンピュータによりアクセス可能な任意のその他の媒体を含むがこれらに限定されない。このほか、当業者であれば、通信媒体は一般的にコンピュータ読み取り可能な命令、データ構造、プログラムモジュールまたは搬送波あるいはその他の伝送機構のような変調データ信号におけるその他のデータを含み、また任意の情報伝送媒体を含むことができるということは公知の事項である。

【０１０８】

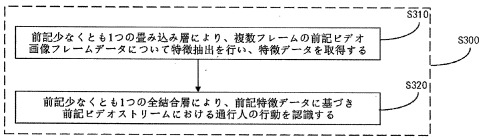
本明細書では例示的な実施例を開示し、かつ具体的な用語を使用しているが、これらは単に一般的な例示的な意味としてのみ使用され、またそのように解釈されるべきであり、制限の目的に用いられない。いくつかの実例において、別途明示しない限り、特定の実施例と組み合わせて説明された特徴、特性および／または要素を単独で使用することができ、またはその他の実施例と組み合わせて説明された特徴、特性および／または要素と組み合わせて使用することができるということは当業者にとって自明である。したがって、添付の請求項により説明された本願の範囲から逸脱しなければ、様々な形式および詳細において変更を加えることができると当業者は理解できる。

【図面】

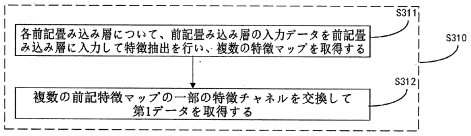
【図１】



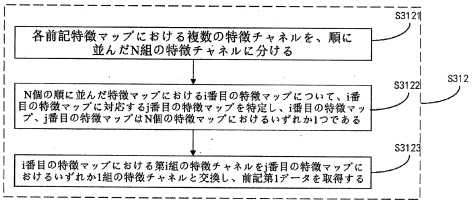
【図２】



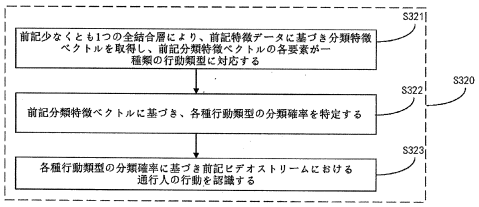
【図３】



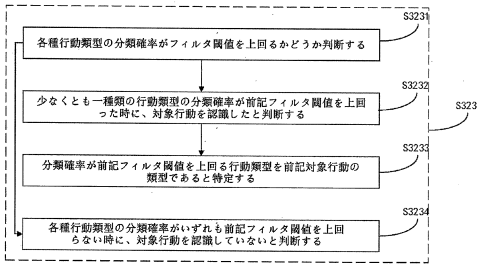
【図４】



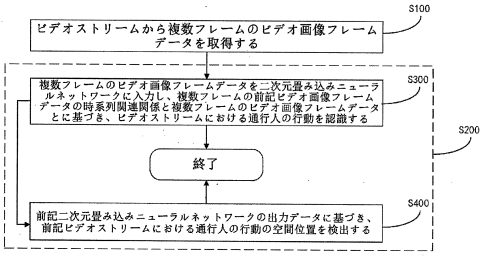
【図５】



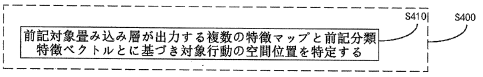
【図６】



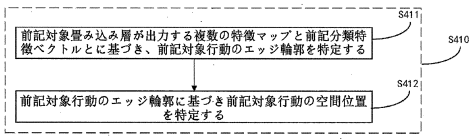
【図７】



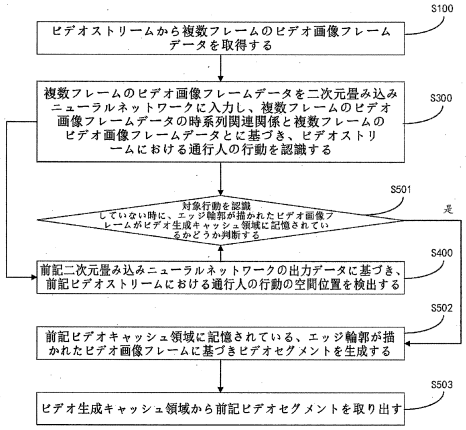
【図８】



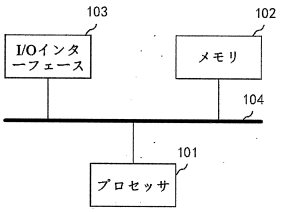
【図９】



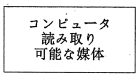
【図１０】



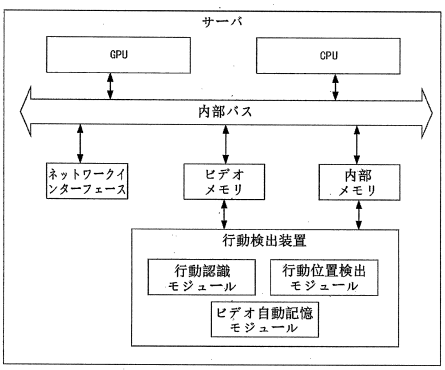
【図１１】



【図１２】



【図１３】


サーバ
GPU
CPU
内部バス
ネットワークインターフェース
ビデオメモリ
内部メモリ
行動検出装置
行動認識モジュール
行動位置検出モジュール
ビデオ自動記憶モジュール

フロントページの続き

中国広東省深▲せん▼市南山区高新技術産業園科技南路中興通訊大厦
(72)発明者　賈霞
中国広東省深▲せん▼市南山区高新技術産業園科技南路中興通訊大厦
(72)発明者　劉明
中国広東省深▲せん▼市南山区高新技術産業園科技南路中興通訊大厦
(72)発明者　張羽豐
中国広東省深▲せん▼市南山区高新技術産業園科技南路中興通訊大厦
(72)発明者　林巍▲ヨ▼
中国広東省深▲せん▼市南山区高新技術産業園科技南路中興通訊大厦

審査官　村山 絢子

(56)参考文献　Ji Lin, Chuang Gan, Song Han，TSM: Temporal Shift Module for Efficient Video Understanding，2019 IEEE/CVF International Conference on Computer Vision (ICCV)，米国，IEEE，2019年10月27日，pp. 7082-7092，https://ieeexplore.ieee.org/document/9008827/

(58)調査した分野　(Int.Cl.，ＤＢ名)
Ｇ０６Ｔ　７／００－７／９０
Ｇ０６Ｖ　１０／００－２０／９０
Ｈ０４Ｎ　７／１８