

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7572654号
(P7572654)

(45)発行日 令和6年10月24日(2024.10.24)

(24)登録日 令和6年10月16日(2024.10.16)

(51)国際特許分類		F I			
H 0 4 N	7/18 (2006.01)	H 0 4 N	7/18		K
G 0 6 T	7/20 (2017.01)	G 0 6 T	7/20	3 0 0 Z	
G 0 6 T	7/00 (2017.01)	G 0 6 T	7/00	3 5 0 B	
G 0 6 V	10/774 (2022.01)	G 0 6 V	10/774		

請求項の数 8 (全21頁)

(21)出願番号	特願2023-522073(P2023-522073)	(73)特許権者	000004226 日本電信電話株式会社 東京都千代田区大手町一丁目5番1号
(86)(22)出願日	令和3年5月19日(2021.5.19)	(74)代理人	110001634 弁理士法人志賀国際特許事務所
(86)国際出願番号	PCT/JP2021/018964	(72)発明者	永井 隆昌 東京都千代田区大手町一丁目5番1号 日本電信電話株式会社内
(87)国際公開番号	WO2022/244135	(72)発明者	武田 翔一郎 東京都千代田区大手町一丁目5番1号 日本電信電話株式会社内
(87)国際公開日	令和4年11月24日(2022.11.24)	(72)発明者	松村 誠明 東京都千代田区大手町一丁目5番1号 日本電信電話株式会社内
審査請求日	令和5年8月28日(2023.8.28)	(72)発明者	志水 信哉

最終頁に続く

(54)【発明の名称】 学習装置、推定装置、学習モデルデータ生成方法、推定方法及びプログラム

(57)【特許請求の範囲】

【請求項1】

背景と競技者の動作とが記録された原映像データと、前記原映像データに含まれる複数の画像フレームの各々において前記競技者を囲む領域をマスクした競技者マスク映像データと、前記原映像データに含まれる複数の画像フレームの各々において前記競技者を囲む領域以外の領域をマスクした背景マスク映像データとを入力とし、前記原映像データを入力とした場合に、競技者の競技に対する評価値である真値競技スコアを出力とし、前記競技者マスク映像データを入力とした場合に、任意に定められる真値背景スコアを出力とし、前記背景マスク映像データを入力とした場合に、任意に定められる真値競技者スコアを出力とする学習モデルにおける学習モデルデータを生成する学習部

を備える学習装置。

【請求項2】

前記学習部は、

関数近似器を有しており、前記原映像データを前記関数近似器に与えた際に前記関数近似器の出力値として得られる推定競技スコアが、前記真値競技スコアに近づくように学習処理を行い、前記競技者マスク映像データを前記関数近似器に与えた際に前記関数近似器の出力値として得られる推定背景スコアが、前記真値背景スコアに近づくように学習処理を行い、前記背景マスク映像データを前記関数近似器に与えた際に前記関数近似器の出力値として得られる推定競技者スコアが、前記真値競技者スコアに近づくように学習処理を行うことにより前記関数近似器に適用する係数を更新して前記学習モデルデータを生成す

る、

請求項 1 に記載の学習装置。

【請求項 3】

前記学習部は、

前記学習処理の途中の任意のタイミングにおいて、前記競技者マスク映像データを前記関数近似器に与えた際に前記関数近似器の出力値として得られる推定背景スコアを、新たな前記真値背景スコアとし、前記背景マスク映像データを前記関数近似器に与えた際に前記関数近似器の出力値として得られる推定競技者スコアを、新たな前記真値競技者スコアとする、

請求項 2 に記載の学習装置。

10

【請求項 4】

前記真値競技スコアを、前記原映像データに記録される競技に対する採点結果の点数とし、前記真値背景スコアを、前記競技を評価しない場合の点数とし、前記真値競技者スコアを、前記真値競技スコアとする、

請求項 1 から請求項 3 のいずれか一項に記載の学習装置。

【請求項 5】

競技者の動作が記録された評価対象の映像データを取り込む入力部と、

背景と競技者の動作とが記録された原映像データと、前記原映像データに含まれる複数の画像フレームの各々において前記競技者を囲む領域をマスクした競技者マスク映像データと、前記原映像データに含まれる複数の画像フレームの各々において前記競技者を囲む領域以外の領域をマスクした背景マスク映像データとを入力とし、前記原映像データを入力とした場合に、競技者の競技に対する評価値である真値競技スコアを出力とし、前記競技者マスク映像データを入力とした場合に、任意に定められる真値背景スコアを出力とし、前記背景マスク映像データを入力とした場合に、任意に定められる真値競技者スコアを出力とする学習済みの学習モデルと、前記入力部が取り込む前記評価対象の映像データとに基づいて、前記評価対象の映像データに対する推定競技スコアを推定する推定部と、
を備える推定装置。

20

【請求項 6】

背景と競技者の動作とが記録された原映像データと、前記原映像データに含まれる複数の画像フレームの各々において前記競技者を囲む領域をマスクした競技者マスク映像データと、前記原映像データに含まれる複数の画像フレームの各々において前記競技者を囲む領域以外の領域をマスクした背景マスク映像データとを入力とし、前記原映像データを入力とした場合に、競技者の競技に対する評価値である真値競技スコアを出力とし、前記競技者マスク映像データを入力とした場合に、任意に定められる真値背景スコアを出力とし、前記背景マスク映像データを入力とした場合に、任意に定められる真値競技者スコアを出力とする学習モデルにおける学習モデルデータを生成する、
学習モデルデータ生成方法。

30

【請求項 7】

競技者の動作が記録された評価対象の映像データを取り込み、

背景と競技者の動作とが記録された原映像データと、前記原映像データに含まれる複数の画像フレームの各々において前記競技者を囲む領域をマスクした競技者マスク映像データと、前記原映像データに含まれる複数の画像フレームの各々において前記競技者を囲む領域以外の領域をマスクした背景マスク映像データとを入力とし、前記原映像データを入力とした場合に、競技者の競技に対する評価値である真値競技スコアを出力とし、前記競技者マスク映像データを入力とした場合に、任意に定められる真値背景スコアを出力とし、前記背景マスク映像データを入力とした場合に、任意に定められる真値競技者スコアを出力とする学習済みの学習モデルと、取り込んだ前記評価対象の映像データとに基づいて、前記評価対象の映像データに対する推定競技スコアを推定する、

40

推定方法。

【請求項 8】

50

請求項 1 から請求項 3 のいずれか一項に記載の学習装置又は請求項 4 に記載の推定装置としてコンピュータを実行させるためのプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、例えば、競技者の競技の採点の手法に関するノウハウを学習する学習装置、学習モデルデータ生成方法及び学習装置に対応するプログラム、並びに、学習結果に基づいて競技のスコアを推定する推定装置、推定方法及び推定装置に対応するプログラムに関する。

【背景技術】

【0002】

スポーツ競技において、高飛び込み、フィギュアスケート及び体操等の選手が行った競技に対して、オフィシャルな審判員がスコアを採点し、採点したスコアに基づいて個々の競技の順位を決める競技がある。このような競技には、採点における定量的な採点基準が存在している。

【0003】

近年、このような競技におけるスコアを自動的に推定するといったコンピュータビジョン分野での活動品質評価で使われる技術の検討が進められており、このような技術として AQA (Action Quality Assessment) という技術が知られている。

【0004】

例えば、非特許文献 1 に記載の技術では、競技者が競技した一連の動作を記録した映像データを入力データとし、深層学習により映像データから特徴を抽出してスコアを推定する手法が提案されている。

【0005】

図 8 は、非特許文献 1 に記載の技術における学習装置 100 と、推定装置 200 の概略構成を示すブロック図である。学習装置 100 の学習部 101 には、学習用データとして、競技者が競技した一連の動作を記録した映像データと、当該競技者の競技に対して審判員が採点した真値スコア t_{score} とが与えられる。学習部 101 は、DNN (Deep Neural Network) を備えており、DNN に学習モデルデータ記憶部 102 が記憶する重みやバイアス等の係数、すなわち学習モデルデータを適用する。

【0006】

学習部 101 は、DNN に映像データを与えることにより出力値として得られる推定スコア y_{score} と、当該映像データに対応する真値スコア t_{score} とを用いて損失 L_{SR} を算出する。学習部 101 は、算出した損失 L_{SR} を小さくするように誤差逆伝搬法により DNN に適用する新たな係数を算出する。学習部 101 は、算出した新たな係数を学習モデルデータ記憶部 102 に書き込むことにより係数を更新する。

【0007】

これらの係数を更新する処理を繰り返すことにより、係数が次第に収束し、最終的に収束した係数が、学習済みの学習モデルデータとして学習モデルデータ記憶部 102 に記憶されることになる。なお、非特許文献 1 では、損失 L_{SR} の算出に、 $L_{SR} = L1 \text{ 距離}(y_{score}, t_{score}) + L2 \text{ 距離}(y_{score}, t_{score})$ という損失関数を用いている。

【0008】

推定装置 200 は、学習部 101 と同一構成の DNN を備える推定部 201 と、学習装置 100 の学習モデルデータ記憶部 102 が記憶する学習済みの学習モデルデータを予め記憶する学習モデルデータ記憶部 202 とを備える。推定部 201 の DNN には、学習モデルデータ記憶部 202 が記憶する学習済みの学習モデルデータが適用される。推定部 201 は、任意の競技者が競技した一連の動作を記録した映像データを入力データとして DNN に与えることにより、DNN の出力値として当該競技に対する推定スコア y_{score} が得られる。

10

20

30

40

50

【先行技術文献】

【非特許文献】

【0009】

【文献】Paritosh Parmar and Brendan Tran Morris, “ Learning To Score Olympic Events ”, In CVPR Workshop. 2017

【発明の概要】

【発明が解決しようとする課題】

【0010】

非特許文献1に記載の技術に対して、以下のような実験を試みた。図9(a)に示す競技者が競技した一連の動作を記録した映像データ(以下「原映像データ」という。)と、
図9(b)に示す原映像データに含まれる複数の画像フレームの各々において競技者が表示されている領域を矩形形状の領域301, 302, 303で囲み、矩形形状の領域を画像フレームの平均色で塗りつぶした映像データ(以下「競技者マスク映像データ」という。)とを準備する。なお、領域301, 302, 303の範囲を点線の枠で示しているが、この点線の枠は、矩形形状の範囲を明確にするために示したものであり、実際の競技者マスク映像データには存在しない。

10

【0011】

図9(a)に示すように、推定部201に対して原映像データを与えた場合に得られる推定スコア y_{score} の正確度合いは「0.8890」であった。これに対して、図9(b)に示すように、推定部201に対して競技者マスク映像データを与えた場合に得られる推定スコア y_{score} の正確度合いは「0.8563」であった。この実験結果より、競技者マスク映像データを推定部201に与えた場合、競技者の動作が見えないにも関わらず、高精度でスコアを推定しており、競技者の動作が見える原映像データの場合と比較して、スコアの推定精度がほとんど下がっていないことが分かる。

20

【0012】

非特許文献1に記載の技術では、競技者の動作に関する特徴、例えば、関節座標などを明示的に与えずに、映像データのみを学習用のデータとして与えている。そのため、上記の実験結果より、非特許文献1に記載の技術は、競技者の動作に関係ない映像中の特徴、例えば、会場などの背景の特徴を抽出しており、学習モデルは、競技者の動作に汎化していないのではないかと推測される。会場などの背景の特徴を抽出していることから、非特許文献1に記載の技術は、未知の背景を含む映像データに対して精度が悪化するのではないかと推測される。

30

【0013】

人間の関節座標などの関節情報を明示的に与える手法も存在するが、関節は複雑な動作をするため推定が困難であり、不正確な関節情報は、逆に精度に悪影響を及ぼしてしまう。そのため、関節情報を明示的に与える手法は、回避したいという事情もある。

【0014】

上記事情に鑑み、本発明は、関節情報を明示的に与えることなく、競技者の動作を記録した映像データから、競技者の動作に汎化した学習モデルデータを生成して、競技における採点の精度を高めることを可能とする技術の提供を目的としている。

40

【課題を解決するための手段】

【0015】

本発明の一態様は、背景と競技者の動作とが記録された原映像データと、前記原映像データに含まれる複数の画像フレームの各々において前記競技者を囲む領域をマスクした競技者マスク映像データと、前記原映像データに含まれる複数の画像フレームの各々において前記競技者を囲む領域以外の領域をマスクした背景マスク映像データとを入力とし、前記原映像データを入力とした場合に、競技者の競技に対する評価値である真値競技スコアを出力とし、前記競技者マスク映像データを入力とした場合に、任意に定められる真値背景スコアを出力とし、前記背景マスク映像データを入力とした場合に、任意に定められる真値競技者スコアを出力とする学習モデルにおける学習モデルデータを生成する学習部を

50

備える学習装置である。

【0016】

本発明の一態様は、競技者の動作が記録された評価対象の映像データを取り込む入力部と、背景と競技者の動作とが記録された原映像データと、前記原映像データに含まれる複数の画像フレームの各々において前記競技者を囲む領域をマスクした競技者マスク映像データと、前記原映像データに含まれる複数の画像フレームの各々において前記競技者を囲む領域以外の領域をマスクした背景マスク映像データとを入力とし、前記原映像データを入力とした場合に、競技者の競技に対する評価値である真値競技スコアを出力とし、前記競技者マスク映像データを入力とした場合に、任意に定められる真値背景スコアを出力とし、前記背景マスク映像データを入力とした場合に、任意に定められる真値競技者スコアを出力とする学習済みの学習モデルと、前記入力部が取り込む前記評価対象の映像データとに基づいて、前記評価対象の映像データに対する推定競技スコアを推定する推定部と、を備える推定装置である。

10

【0017】

本発明の一態様は、背景と競技者の動作とが記録された原映像データと、前記原映像データに含まれる複数の画像フレームの各々において前記競技者を囲む領域をマスクした競技者マスク映像データと、前記原映像データに含まれる複数の画像フレームの各々において前記競技者を囲む領域以外の領域をマスクした背景マスク映像データとを入力とし、前記原映像データを入力とした場合に、競技者の競技に対する評価値である真値競技スコアを出力とし、前記競技者マスク映像データを入力とした場合に、任意に定められる真値背景スコアを出力とし、前記背景マスク映像データを入力とした場合に、任意に定められる真値競技者スコアを出力とする学習モデルにおける学習モデルデータを生成する、学習モデルデータ生成方法である。

20

【0018】

本発明の一態様は、競技者の動作が記録された評価対象の映像データを取り込み、背景と競技者の動作とが記録された原映像データと、前記原映像データに含まれる複数の画像フレームの各々において前記競技者を囲む領域をマスクした競技者マスク映像データと、前記原映像データに含まれる複数の画像フレームの各々において前記競技者を囲む領域以外の領域をマスクした背景マスク映像データとを入力とし、前記原映像データを入力とした場合に、競技者の競技に対する評価値である真値競技スコアを出力とし、前記競技者マスク映像データを入力とした場合に、任意に定められる真値背景スコアを出力とし、前記背景マスク映像データを入力とした場合に、任意に定められる真値競技者スコアを出力とする学習済みの学習モデルと、取り込んだ前記評価対象の映像データとに基づいて、前記評価対象の映像データに対する推定競技スコアを推定する、推定方法である。

30

【0019】

本発明の一態様は、上記の学習装置又は推定装置としてコンピュータを実行させるためのプログラムである。

【発明の効果】

【0020】

本発明により、関節情報を明示的に与えることなく、競技者の動作を記録した映像データから、競技者の動作に汎化した学習モデルデータを生成して、競技における採点の精度を高めることが可能になる。

40

【図面の簡単な説明】

【0021】

【図1】本発明の一実施形態による学習装置の構成を示すブロック図である。

【図2】本実施形態において用いられる原映像データに含まれる画像フレームの一例を示す図である。

【図3】本実施形態において用いられる競技者マスク映像データに含まれる画像フレームの一例を示す図である。

【図4】本実施形態において用いられる背景マスク映像データに含まれる画像フレームの

50

一例を示す図である。

【図 5】本実施形態の学習装置による処理の流れを示す図である。

【図 6】本実施形態による推定装置の構成を示すブロック図である。

【図 7】本実施形態の推定装置による処理の流れを示す図である。

【図 8】非特許文献 1 に記載の技術における学習装置と推定装置の構成を示すブロック図である。

【図 9】非特許文献 1 に記載の技術に対して行った実験の概要とその結果を示す図である。

【発明を実施するための形態】

【0022】

(学習装置の構成)

以下、本発明の実施形態について図面を参照して説明する。図 1 は、本発明の一実施形態による学習装置 1 の構成を示すブロック図である。学習装置 1 は、入力部 11、学習部 12 及び学習モデルデータ記憶部 15 を備える。

【0023】

入力部 11 は、競技者が行う動作の中で採点の評価の対象となる一連の動作が背景と共に記録された原映像データを取り込む。例えば、競技者が、水泳の高飛び込みの選手である場合、原映像データには、競技者が、飛び込み台に立ち、ジャンプして、ひねりや回転などの動作をして、プールへの入水が完了するまでの動作が背景と共に記録されることになる。図 2 (a), (b), (c) に示す画像フレームは、ある原映像データに含まれる複数の画像フレームの中から時系列順に任意に選択した画像フレームの一例である。

【0024】

入力部 11 は、原映像データに記録される競技者の動作に対する評価値である真値競技スコアを取り込む。真値競技スコアは、例えば、原映像データが記録された際に、原映像データに記録される競技者の動作に対して、実際に審判員が当該競技で採用されている定量的な採点基準に基づいて採点した採点結果の点数である。入力部 11 は、取り込んだ原映像データと、当該原映像データに対応する真値競技スコアとを関連付けて原映像データの訓練データセットとする。

【0025】

入力部 11 は、原映像データに対応する競技者マスク映像データを取り込む。ここで、競技者マスク映像データとは、原映像データに含まれている複数の画像フレームの各々において競技者の領域を囲った矩形領域をマスクした映像データである。図 3 (a), (b), (c) に示す画像フレームは、それぞれ図 2 (a), (b), (c) に示す原映像データの画像フレームの各々に対応する競技者マスク映像データの画像フレームである。なお、図 3 (a), (b), (c) において、矩形領域 41, 42, 43 の範囲を点線の枠で示しているが、この点線の枠は、矩形領域 41, 42, 43 の範囲を明確にするために示したものであり、実際の競技者マスク映像データには存在しない。図 3 (a), (b), (c) において、矩形領域 41, 42, 43 がマスクされている状態をハッチングで示しているが、実際には、矩形領域 41, 42, 43 の各々は、例えば、矩形領域 41, 42, 43 の各々を含む画像フレームの平均色で塗りつぶしてマスクされる。

【0026】

入力部 11 は、競技者マスク映像データに対応する真値背景スコアを取り込む。真値背景スコアは、競技者マスク映像データに対する評価値である。競技者マスク映像データは、競技者が完全に見えない映像データである。そのため、審判員は、採点することができないことを考慮して、当該競技において評価しない場合の点数、例えば、当該競技における最低点が、真値背景スコアとして定められる。例えば、当該競技において評価しない場合の点数が「0」である場合、「0」という値が、真値背景スコアとして予め定められる。入力部 11 は、取り込んだ競技者マスク映像データと、当該競技者マスク映像データに対応する真値背景スコアとを関連付けて競技者マスク映像データの訓練データセットとする。

【0027】

10

20

30

40

50

入力部 11 は、原映像データに対応する背景マスク映像データを取り込む。ここで、背景マスク映像データとは、原映像データに含まれている複数の画像フレームの各々において競技者の領域を囲った矩形領域以外の領域をマスクした映像データである。図 4 (a) , (b) , (c) に示す画像フレームは、それぞれ図 2 (a) , (b) , (c) に示す原映像データの画像フレームに対応する背景マスク映像データの画像フレームである。なお、図 4 (a) , (b) , (c) において、矩形領域 41 , 42 , 43 の範囲を点線の枠で示しているが、この点線の枠は、矩形領域 41 , 42 , 43 の範囲を明確にするために示したものであり、実際の背景マスク映像データには存在しない。図 4 (a) , (b) , (c) において、矩形領域 41 , 42 , 43 以外の領域がマスクされている状態をハッチングで示しているが、実際には、矩形領域 41 , 42 , 43 以外の領域は、例えば、矩形領域 41 , 42 , 43 の各々を含む画像フレームの平均色で塗りつぶしてマスクされる。

10

【 0028 】

入力部 11 は、背景マスク映像データに対応する真値競技者スコアを取り込む。真値競技者スコアは、背景マスク映像データに対する評価値である。背景マスク映像データは、競技者が見えている映像データである。そのため、例えば、背景マスク映像データに対応する原映像データの真値競技スコアが、当該背景マスク映像データに対応する真値競技者スコアとして予め定められる。入力部 11 は、取り込んだ背景マスク映像データと、当該背景マスク映像データに対応して取り込んだ真値競技者スコアとを関連付けて背景マスク映像データの訓練データセットとする。

【 0029 】

20

入力部 11 は、複数の原映像データの訓練データセットを取り込んだ場合、複数の原映像データの訓練データセットの各々に対応する競技者マスク映像データの訓練データセット及び背景マスク映像データの訓練データセットを取り込むことになる。

【 0030 】

図 3 (a) , (b) , (c) 及び図 4 (a) , (b) , (c) に示した矩形領域 41 , 42 , 43 の範囲は、例えば、下記の参考文献に示す技術によって映像データに含まれる画像フレームの各々から自動的に検出するようにしてもよいし、映像データに含まれる全ての画像フレームを目視で確認しながら矩形領域 41 , 42 , 43 の範囲を手動で定めるようにしてもよい。

【 0031 】

30

[参考文献 : Kaiming He, Georgia Gkioxari, Piotr Dollar and Ross Girshick, " Mask R-CNN ", In ICCV, 2017]

【 0032 】

上記の参考文献に示す技術を採用する場合、例えば、入力部 11 が、原映像データを取り込み、取り込んだ原映像データから矩形領域の範囲を検出し、検出した矩形領域の範囲に基づいて、原映像データから競技者マスク映像データと、背景マスク映像データとを生成するようにしてもよい。この場合、真値背景スコアとして、例えば、上記した「 0 」を適用することが定められており、真値競技者スコアとして、真値競技スコアを適用することが定められているとする。この場合、入力部 11 は、原映像データ及び真値競技スコアのみを取り込むことにより、原映像データの訓練データセットと、競技者マスク映像データの訓練データセットと、背景マスク映像データの訓練データセットとを生成することができる。

40

【 0033 】

なお、真値競技スコア、真値背景スコア、真値競技者スコアの各々は、上記したような評価値に限られるものではなく、任意に定めてもよい。例えば、原映像データに記録される競技者の競技を、当該競技で採用されている定量的な採点基準以外の基準によって採点した採点結果の点数を真値競技スコアとしてもよい。真値競技者スコアとして、真値競技スコア以外の値を採用するようにしてもよい。真値背景スコア及び真値競技者スコアについては、処理の途中で変更するようにしてもよい。

【 0034 】

50

学習部 12 は、学習処理部 13 と、関数近似器 14 を備える。関数近似器 14 として、例えば、DNN が適用される。なお、DNN は、どのようなネットワーク構造を有していてもよい。関数近似器 14 は、学習処理部 13 によって学習モデルデータ記憶部 15 が記憶する係数が与えられる。ここで、係数とは、関数近似器 14 が DNN である場合、DNN に含まれる複数のニューロンの各々に適用される重みやバイアスである。

【0035】

学習処理部 13 は、原映像データの訓練データセットに含まれる原映像データを関数近似器 14 に与えることにより、関数近似器 14 の出力値として得られる推定競技スコアが、関数近似器 14 に与えた原映像データに対応する真値競技スコアに近づくように係数を更新する学習処理を行う。学習処理部 13 は、競技者マスク映像データの訓練データセットに含まれる競技者マスク映像データを関数近似器 14 に与えることにより、関数近似器 14 の出力値として得られる推定背景スコアが、関数近似器 14 に与えた競技者マスク映像データに対応する真値背景スコアに近づくように係数を更新する学習処理を行う。学習処理部 13 は、背景マスク映像データの訓練データセットに含まれる背景マスク映像データを関数近似器 14 に与えることにより、関数近似器 14 の出力値として得られる推定競技者スコアが、関数近似器 14 に与えた背景マスク映像データに対応する真値競技者スコアに近づくように係数を更新する学習処理を行う。

10

【0036】

学習モデルデータ記憶部 15 は、関数近似器 14 に適用する係数、すなわち学習モデルデータを記憶する。学習モデルデータ記憶部 15 は、初期状態において、係数の初期値を予め記憶する。学習モデルデータ記憶部 15 が記憶する係数は、学習処理部 13 が学習処理により新たな係数を算出するごとに、学習処理部 13 によって新たな係数に書き替えられる。

20

【0037】

すなわち、学習部 12 は、学習処理部 13 が行う学習処理によって、原映像データと、競技者マスク映像データと、背景マスク映像データとを入力とし、原映像データを入力とした場合に、真値競技スコアを出力とし、競技者マスク映像データを入力とした場合に、真値背景スコアを出力とし、背景マスク映像データを入力とした場合に、真値競技者スコアを出力とする学習モデルにおける学習モデルデータを生成することになる。ここで、学習モデルとは、学習モデルデータ記憶部 15 が記憶する係数、すなわち学習モデルデータが適用された関数近似器 14 のことである。

30

【0038】

(学習装置による処理)

次に、図 5 を参照しつつ、学習装置 1 による処理について説明する。図 5 は、学習装置 1 による処理の流れを示すフローチャートである。学習装置 1 が備える学習処理部 13 において学習ルールが予め定められており、以下、予め定められる学習ルールごとの処理について説明する。

【0039】

(学習ルール(その1))

例えば、学習処理部 13 において、以下のような学習ルールが予め定められているとする。すなわち、原映像データの訓練データデータセット、競技者マスク映像データの訓練データセット及び背景マスク映像データの訓練データセットの各々の数が、例えば、N 個であり、ミニバッチサイズが M であり、1 エポック分の処理として、原映像データの訓練データデータセット、競技者マスク映像データの訓練データセット及び背景マスク映像データの訓練データセットの全てを用いるという学習ルールが予め定められているとする。学習ルールにおいて、原映像データの訓練データデータセット、競技者マスク映像データの訓練データセット、背景マスク映像データの訓練データセットの順に処理が行われることが予め定められているとする。ここで、N と M は、1 以上の整数であって、 $M < N$ であればどのような値であってもよい。以下では、一例として、N が、「300」であり、M が、「10」である場合について説明する。

40

50

【 0 0 4 0 】

学習装置 1 の入力部 1 1 は、3 0 0 個の原映像データと、3 0 0 個の原映像データの各々に対応する真値競技スコアとを取り込み、取り込んだ 3 0 0 個の原映像データと、取り込んだ原映像データの各々に対応する真値競技スコアとを関連付けて 3 0 0 個の原映像データの訓練データセットを生成する。

【 0 0 4 1 】

入力部 1 1 は、3 0 0 個の原映像データの各々に対応する 3 0 0 個の競技者マスク映像データと、競技者マスク映像データの各々に対応する真値背景スコアとを取り込み、取り込んだ 3 0 0 個の競技者マスク映像データと、取り込んだ競技者マスク映像データの各々に対応する真値背景スコアとを関連付けて 3 0 0 個の競技者マスク映像データの訓練データセットを生成する。

10

【 0 0 4 2 】

入力部 1 1 は、3 0 0 個の原映像データの各々に対応する 3 0 0 個の背景マスク映像データと、背景マスク映像データの各々に対応する真値競技者スコアとを取り込み、取り込んだ 3 0 0 個の背景マスク映像データと、取り込んだ背景マスク映像データの各々に対応する真値競技者スコアとを関連付けて 3 0 0 個の背景マスク映像データの訓練データセットを生成する。

【 0 0 4 3 】

入力部 1 1 は、それぞれ 3 0 0 個ずつの原映像データの訓練データセットと、競技者マスク映像データの訓練データセットと、背景マスク映像データの訓練データセットとを学習処理部 1 3 に出力する。学習処理部 1 3 は、入力部 1 1 が出力する、それぞれ 3 0 0 個ずつの原映像データの訓練データセットと、競技者マスク映像データの訓練データセットと、背景マスク映像データの訓練データセットとを取り込む。学習処理部 1 3 は、取り込んだ 3 0 0 個ずつの原映像データの訓練データセットと、競技者マスク映像データの訓練データセットと、背景マスク映像データの訓練データセットとを内部の記憶領域に書き込んで記憶させる。

20

【 0 0 4 4 】

学習処理部 1 3 は、内部の記憶領域にエポック数、すなわちエポックの回数の値を記憶する領域を設けて、エポック数を「0」に初期化する。学習処理部 1 3 は、内部の記憶領域にミニバッチ学習のパラメータ、すなわち、原映像データ、競技者マスク映像データ及び背景マスク映像データの各々を関数近似器 1 4 に対して与えた回数を示す処理回数を記憶する領域を設けて、原映像データ、競技者マスク映像データ及び背景マスク映像データの各々の処理回数を「0」に初期化する（ステップ S a 1）。

30

【 0 0 4 5 】

学習処理部 1 3 は、内部の記憶領域が記憶する原映像データ、競技者マスク映像データ及び背景マスク映像データの各々の処理回数と、予め定められる学習ルールとにしたがって、選択する訓練データセットを選択する（ステップ S a 2）。ここでは、原映像データ、競技者マスク映像データ及び背景マスク映像データの各々の処理回数は、いずれも「0」であり、それぞれ 3 0 0 個の原映像データ、競技者マスク映像データ及び背景マスク映像データの全てが処理に用いられていない状態である。上記したように、学習ルールにおいて、原映像データの訓練データデータセット、競技者マスク映像データの訓練データセット、背景マスク映像データの訓練データセットの順に処理が行われることが予め定められている。したがって、学習処理部 1 3 は、最初に、原映像データの訓練データセットを選択する（ステップ S a 2、原映像データ）。

40

【 0 0 4 6 】

学習処理部 1 3 は、学習モデルデータ記憶部 1 5 が記憶する係数を読み出し、読み出した係数を関数近似器 1 4 に適用する（ステップ S a 3 - 1）。学習処理部 1 3 は、ステップ S a 2 の処理において選択した原映像データの訓練データセットを対象として、先頭から順に、学習ルールにおいて定められているミニバッチサイズ M の数の原映像データの訓練データセットを内部の記憶領域から読み出す。

50

【 0 0 4 7 】

ここでは、ミニバッチサイズMは、「10」であるため、学習処理部13は、10個の原映像データの訓練データセットを内部の記憶領域から読み出す。学習処理部13は、読み出した10個の原映像データの訓練データセットから1つの原映像データを選択して関数近似器14に与える。学習処理部13は、原映像データを与えることにより関数近似器14が出力する推定競技スコアを取り込む。学習処理部13は、取り込んだ推定競技スコアと、関数近似器14に与えた原映像データに対応する真値競技スコアとを関連付けて内部の記憶領域に書き込んで記憶させる。学習処理部13は、関数近似器14に原映像データを与えることに、内部の記憶領域が記憶する原映像データの処理回数に1を加算する(ステップSa4-1)。

10

【 0 0 4 8 】

学習処理部13は、10個の原映像データの訓練データセットに含まれる10個の原映像データの各々に対してステップSa4-1の処理を繰り返し行い(ループL1s~L1e)、内部の記憶領域において、10個の推定競技スコアと真値競技スコアの組み合わせを生成する。

【 0 0 4 9 】

学習処理部13は、内部の記憶領域が記憶する10個の推定競技スコアと真値競技スコアの組み合わせに基づいて、予め定められる損失関数に基づいて損失を算出する。学習処理部13は、算出した損失に基づいて、例えば、誤差逆伝搬法により、関数近似器14に適用する新たな係数を算出する。学習処理部13は、学習モデルデータ記憶部15が記憶する係数を、算出した新たな係数に書き換えて更新する(ステップSa5-1)。

20

【 0 0 5 0 】

学習処理部13は、内部の記憶領域が記憶する原映像データ、競技者マスク映像データ及び背景マスク映像データの各々の処理回数を参照し、1エポック分の処理を終了しているか否かを判定する(ステップSa6)。上記したように、学習ルールにおいて、1エポックの処理として、原映像データの訓練データデータセット、競技者マスク映像データの訓練データセット及び背景マスク映像データの訓練データセットが全て用いられるということが定められている。そのため、1エポック分の処理が終了している状態とは、原映像データ、競技者マスク映像データ及び背景マスク映像データの各々の処理回数が「300」以上になっている状態である。ここでは、原映像データの処理回数が「10」であり、競技者マスク映像データ及び背景マスク映像データの各々の処理回数が「0」である。そのため、学習処理部13は、1エポック分の処理を終了してないと判定し(ステップSa6、No)、処理をステップSa2に進める。

30

【 0 0 5 1 】

学習処理部13は、再び行うステップSa2の処理において、原映像データの処理回数が「300」以上になっていない場合、ステップSa2の処理において、再び原映像データの訓練データセットを選択し(ステップSa2、原映像データ)、ステップSa3-1以降の処理を行う。

【 0 0 5 2 】

一方、学習処理部13は、再び行うステップSa2の処理において、原映像データの処理回数が「300」以上になっている場合、学習ルールにしたがって、次に、競技者マスク映像データの訓練データセットを選択する(ステップSa2、競技者マスク映像データ)。

40

【 0 0 5 3 】

学習処理部13は、学習モデルデータ記憶部15が記憶する係数を読み出し、読み出した係数を関数近似器14に適用する(ステップSa3-2)。

【 0 0 5 4 】

学習処理部13は、ステップSa2の処理において選択した競技者マスク映像データの訓練データセットを対象として、先頭から順に10個の競技者マスク映像データの訓練データセットを内部の記憶領域から読み出す。学習処理部13は、読み出した10個の競技

50

者マスク映像データの訓練データセットから1つの競技者マスク映像データを選択して関数近似器14に与える。学習処理部13は、競技者マスク映像データを与えることにより関数近似器14が出力する推定背景スコアを取り込む。学習処理部13は、取り込んだ推定背景スコアと、関数近似器14に与えた競技者マスク映像データに対応する真値背景スコアとを関連付けて内部の記憶領域に書き込んで記憶させる。学習処理部13は、関数近似器14に競技者マスク映像データを与えるごとに、内部の記憶領域が記憶する競技者マスク映像データの処理回数に1を加算する(ステップS a 4 - 2)。

【0055】

学習処理部13は、10個の競技者マスク映像データの訓練データセットに含まれる10個の競技者マスク映像データの各々に対してステップS a 4 - 2の処理を繰り返し行い(ループL 2 s ~ L 2 e)、内部の記憶領域において、10個の推定背景スコアと真値背景スコアの組み合わせを生成する。

10

【0056】

学習処理部13は、内部の記憶領域が記憶する10個の推定背景スコアと真値背景スコアの組み合わせを用いて、予め定められる損失関数に基づいて損失を算出する。学習処理部13は、算出した損失に基づいて、例えば、誤差逆伝搬法により、関数近似器14に適用する新たな係数を算出する。学習処理部13は、学習モデルデータ記憶部15が記憶する係数を、算出した新たな係数に書き換えて更新する(ステップS a 5 - 2)。

【0057】

学習処理部13は、1エポック分の処理を終了しているか否かの判定を行う(ステップS a 6)。学習処理部13は、競技者マスク映像データの処理回数が「300」以上になっていない場合、1エポック分の処理を終了していないと判定し(ステップS a 6、No)、処理をステップS a 2に進める。

20

【0058】

学習処理部13は、再び行うステップS a 2の処理において、競技者マスク映像データの処理回数が「300」以上になっていない場合、再び競技者マスク映像データの訓練データセットを選択する(ステップS a 2、競技者マスク映像データ)。その後、学習処理部13は、ステップS a 3 - 2以降の処理を行う。

【0059】

一方、学習処理部13は、再び行うステップS a 2の処理において、競技者マスク映像データの処理回数が「300」以上になっている場合、学習ルールにしたがって、次に、背景マスク映像データの訓練データセットを選択する(ステップS a 2、背景マスク映像データ)。

30

【0060】

学習処理部13は、学習モデルデータ記憶部15が記憶する係数を読み出す。学習処理部13は、読み出した係数を関数近似器14に適用する(ステップS a 3 - 3)。

【0061】

学習処理部13は、ステップS a 2の処理において選択した背景マスク映像データの訓練データセットを対象として、先頭から順に10個の背景マスク映像データの訓練データセットを内部の記憶領域から読み出す。学習処理部13は、読み出した10個の背景マスク映像データの訓練データセットから1つの背景マスク映像データを選択して関数近似器14に与える。学習処理部13は、背景マスク映像データを与えることにより関数近似器14が出力する推定競技者スコアを取り込む。学習処理部13は、取り込んだ推定競技者スコアと、関数近似器14に与えた背景マスク映像データに対応する真値競技者スコアとを関連付けて内部の記憶領域に書き込んで記憶させる。学習処理部13は、関数近似器14に背景マスク映像データを与えるごとに、内部の記憶領域が記憶する背景マスク映像データの処理回数に1を加算する(ステップS a 4 - 3)。

40

【0062】

学習処理部13は、10個の背景マスク映像データの訓練データセットに含まれる10個の背景マスク映像データの各々に対してステップS a 4 - 3の処理を繰り返し行い(ル

50

ープL3s~L3e)、内部の記憶領域において、10個の推定競技者スコアと真値競技者スコアの組み合わせを生成する。

【0063】

学習処理部13は、内部の記憶領域が記憶する10個の推定競技者スコアと真値競技者スコアの組み合わせに基づいて、予め定められる損失関数に基づいて損失を算出する。学習処理部13は、算出した損失に基づいて、例えば、誤差逆伝搬法により、関数近似器14に適用する新たな係数を算出する。学習処理部13は、学習モデルデータ記憶部15が記憶する係数を、算出した新たな係数に書き換えて更新する(ステップSa5-3)。

【0064】

学習処理部13は、1エポック分の処理を終了しているか否かの判定を行う(ステップSa6)。学習処理部13は、背景マスク映像データの処理回数が「300」以上になっていない場合、1エポック分の処理を終了していないと判定する(ステップSa6、No)。この場合、学習処理部13は、処理をステップSa2に進める。

10

【0065】

学習処理部13は、再び行うステップSa2の処理において、背景マスク映像データの処理回数が「300」以上になっていない場合、ステップSa2の処理において、再び背景マスク映像データの訓練データセットを選択し(ステップSa2、背景マスク映像データ)する。その後、学習処理部13は、ステップSa3-3以降の処理を行う。

【0066】

一方、学習処理部13は、ステップSa6の処理において、1エポック分の処理を終了している、すなわち、原映像データ、競技者マスク映像データ及び背景マスク映像データの各々の処理回数が「300」以上になっている場合、1エポック分の処理が終了していると判定する(ステップSa6、Yes)。学習処理部13は、内部の記憶領域が記憶するエポック数の数に1を加算する。学習処理部13は、内部の記憶領域が記憶するミニバッチ学習のパラメータを「0」に初期化する(ステップSa7)。すなわち学習処理部13は、原映像データ、競技者マスク映像データ及び背景マスク映像データの各々の処理回数を「0」に初期化する。

20

【0067】

学習処理部13は、内部の記憶領域が記憶するエポック数が、終了条件を満たしているか否かを判定する(ステップSa8)。学習処理部13は、例えば、エポック数が予め定められる上限値に達している場合、終了条件を満たしていると判定する。一方、学習処理部13は、例えば、エポック数が予め定められる上限値に達していない場合、終了条件を満たしていないと判定する。

30

【0068】

学習処理部13は、エポック数が、終了条件を満たしていると判定した場合(ステップSa8、Yes)、処理を終了する。一方、学習処理部13は、エポック数が、終了条件を満たしていないと判定した場合(ステップSa8、No)、処理をステップSa2の処理に進める。ステップSa8の処理の後に再び行われるステップSa2の処理において、学習処理部13は、再び学習ルールにしたがって、原映像データの訓練データセット、競技者マスク映像データの訓練データセット及び背景マスク映像データセットの順に選択を行う。その後、学習処理部13は、選択したそれぞれに対して、ステップSa3-1以降の処理、ステップSa3-2以降の処理及びステップSa3-3以降の処理を行う。

40

【0069】

これにより、学習処理部13が、処理を終了した際に、学習モデルデータ記憶部15には、学習済みの係数、すなわち学習済みの学習モデルデータが生成されることになる。なお、学習処理部13が行う学習処理とは、図5におけるステップSa2~ステップSa8に示す繰り返しの処理により関数近似器14に適用する係数を更新する処理のことである。

【0070】

なお、上記した図5の処理において、学習処理部13は、2回目以降に行うステップSa4-1, Sa4-2, Sa4-3の各々の処理において、内部の記憶領域から次の10

50

個の訓練データセットを読み出す際、前回の同一ステップの処理において選択した10個の訓練データセットの後に続く10個の訓練データセットを読み出すものとする。

【0071】

上記した図5の処理において、学習処理部13が、ステップSa5-1, Sa5-2, Sa5-3の処理において用いる損失関数は、例えば、L1距離を算出する関数であってもよいし、L2距離を算出する関数であってもよいし、L1距離とL2距離の合計を算出する関数であってもよい。

【0072】

(学習ルール(その2))

学習ルールとして、例えば、エポック数の上限値が「100」に予め定められており、エポック数が「50」に到達するまでは、学習処理を安定させるため、すなわち、係数の収束を穏やかにするため、学習処理部13は、ステップSa2の処理において、原映像データの訓練データセット、競技者マスク映像データの訓練データセットの順に選択し、背景マスク映像データについては選択しないようにする。エポック数が「50」に到達して以降、次の50エポックについては、学習処理部13は、ステップSa2の処理において、原映像データの訓練データセット、競技者マスク映像データの訓練データセット、背景マスク映像データの訓練データセットの順に選択するという学習ルールを定めるようにしてもよい。これにより、上記の図5の処理において、エポック数が「50」に到達するまでは、ステップSa3-3~ステップSa5-3の処理が行われなくなり、エポック数が「50」に到達して以降、次の50エポックについては、上記の図5の処理が行われることになる。このように、エポック数に応じて、ステップSa2の処理において選択する訓練データセットを変えるという学習ルールを定めるようにしてもよい。

【0073】

なお、上記のエポック数が「50」というのは一例であり、別の値を定めるようにしてもよい。選択する訓練データセットの組み合わせを変えるエポック数を1つに限定するのではなく、選択する訓練データセットの組み合わせを変えるエポック数を複数定めておき、学習処理部13は、複数定めたエポック数に到達するごとに、選択する訓練データセットを変えるという学習ルールを定めるようにしてもよい。この場合に、学習処理部13がステップSa2の処理において選択する訓練データの組み合わせは、上記した訓練データの組み合わせの例に限られず、任意の組み合わせとしてもよい。エポック数が増加するごとに、ステップSa2の処理において学習処理部13が選択する訓練データセットをランダムに変えるという学習ルールにしてもよい。

【0074】

(学習ルール(その3))

例えば、真値背景スコアを「0」としている場合に、学習処理が一定程度行われた後であっても、関数近似器14に競技者マスク映像データを与えたときに、関数近似器14が出力する推定背景スコアは、完全に「0」にはならず、「1」や「2」を出力することがシミュレーションの結果として分かっている。このようになるのは、審判が背景に対してもわずかに点数をつけている状態になっている可能性があると考えられることができる。真値競技者スコアを、真値競技スコアとしている場合に、学習処理が一定程度行われた後であっても、関数近似器14に背景マスク映像データを与えたときに、関数近似器14は、真値競技スコアに完全に一致する値を出力するようにはならないことが分かっている。

【0075】

このように、審判の採点に背景が影響していることを想定し、学習ルールとして、エポック数が、予め定められる上限値未満の所定数になった際に、学習処理部13は、競技者マスク映像データの訓練データセットに含まれる全ての真値背景スコアを、その時点で、競技者マスク映像データを与えた際に関数近似器14が出力する推定背景スコアに置き換え、背景マスク映像データの訓練データセットに含まれる全ての真値競技者スコアを、その時点で、背景マスク映像データを与えた際に関数近似器14が出力する推定競技者スコアに置き換えるという学習ルールを定めるようにしてもよい。

【 0 0 7 6 】

この学習ルールが適用される場合、学習処理部 1 3 は、エポック数が上記の所定数になるまでは、上記した図 5 の処理を行い、エポック数が所定数になった際に、原映像データの訓練データセットと、学習ルールにしたがって真値背景スコアの置き換えを行った競技者マスク映像データの訓練データセットと、学習ルールにしたがって真値競技者スコアの置き換えを行った背景マスク映像データの訓練データセットとに基づいて、残りのエポック数について、ステップ S a 2 以降の処理を行うことになる。なお、学習処理部 1 3 は、学習ルールにしたがった置き換えを行った後、最初から処理をやり直すようにしてもよい。すなわち、学習処理部 1 3 は、エポック数を「0」に初期化し、ミニバッチ学習のパラメータを初期化して、ステップ S a 2 以降の処理を行うようにしてもよい。なお、最初から処理をやり直す場合に、学習モデルデータ記憶部 1 5 が記憶する係数については、そのまま継続して利用するようにしてもよいし、学習モデルデータ記憶部 1 5 が記憶する係数を初期化するようにしてもよい。

10

【 0 0 7 7 】

上記ではエポック数が所定数になった際に、真値背景スコア及び真値競技者スコアの置き換えを行うようにしているが、エポック数が所定数になったタイミング以外の予め定められる学習処理の途中の任意のタイミングで、真値背景スコア及び真値競技者スコアの置き換えを行ってもよい。例えば、関数近似器 1 4 が出力する推定背景スコアと、直前の推定背景スコアとの差が予め定められる回数において連続して一定値以下になり、かつ関数近似器 1 4 が出力する推定競技者スコアと、直前の推定競技者スコアとの差が予め定められる回数において連続して一定値以下になったことを学習処理部 1 3 が検出したタイミングであってもよい。

20

【 0 0 7 8 】

(その他の学習ルール)

上記した図 5 の処理では、ミニバッチサイズ M を、原映像データ、競技者マスク映像データ及び背景マスク映像データの各々の訓練データセットの数である N よりも小さい値としたミニバッチ学習による学習処理を示している。これに対して、ミニバッチサイズ M = N としたバッチ学習による学習処理を行うようにしてもよいし、ミニバッチサイズ M = 1 としたオンライン学習による学習処理を行うようにしてもよい。

【 0 0 7 9 】

上記した図 5 の処理では、学習処理部 1 3 が、繰り返し行うステップ S a 4 - 1 , 4 a - 2 , 4 a - 3 の処理において、内部の記憶領域が記憶する原映像データ、競技者マスク映像データ、背景マスク映像データの各々から、ミニバッチサイズ M の数のデータを選択する際、内部の記憶領域に記憶されている順に、ミニバッチサイズ M の数ずつ選択するようにしている。これに対して、学習処理部 1 3 は、内部の記憶領域からミニバッチサイズ M の数の訓練データをランダムに選択するようにしてもよいし、例えば、エポック数が、予め定められる上限値未満の所定数に到達するまでは、内部の記憶領域に記憶されている順に、ミニバッチサイズ M の数ずつ訓練データを選択するようにし、エポック数が、予め定められる上限値未満の所定数に到達した後は、ミニバッチサイズ M の数の訓練データをランダムに選択するようにしてもよい。

30

40

【 0 0 8 0 】

上記した図 5 の処理では、ステップ S a 5 - 1 の処理において、推定競技スコアと真値競技スコアの組み合わせに基づいて、損失を算出し、ステップ S a 5 - 2 の処理において、推定背景スコアと真値背景スコアの組み合わせに基づいて、損失を算出し、ステップ S a 5 - 3 の処理において、推定競技者スコアと真値競技者スコアの組み合わせに基づいて、損失を算出し、それぞれの損失に基づいて新たな係数を算出するようにしている。

【 0 0 8 1 】

これに対して、例えば、学習処理部 1 3 は、上記した図 5 の処理において、ループ L 1 s ~ L 1 e の処理が終了した後に、ステップ S a 5 - 1 を行わずに、処理を、ステップ S a 6 に進める。その後、ループ L 2 s ~ L 2 e の処理が終了した後にも、学習処理部 1 3

50

は、ステップ S a 5 - 2 の処理を行わずに、処理をステップ S a 6 に進める。その後、ループ L 3 s ~ L 3 e の処理が終了した後、学習処理部 1 3 は、ステップ S a 5 - 3 の処理において、内部の記憶領域において生成されている全ての推定競技スコアと真値競技スコアの組み合わせと、全ての推定背景スコアと真値背景スコアの組み合わせと、全ての推定競技者スコアと真値競技者スコアの組み合わせとに基づいて損失を算出し、算出した損失に基づいて新たな係数を算出するようにしてもよい。

【 0 0 8 2 】

例えば、学習処理部 1 3 は、上記した図 5 の処理において、ループ L 1 s ~ L 1 e の処理が終了した後に、ステップ S a 5 - 1 を行わずに、処理をステップ S a 6 に進める。その後、ループ L 2 s ~ L 2 e の処理が終了した後に、学習処理部 1 3 は、ステップ S a 5 - 2 の処理において、内部の記憶領域において生成されている全ての推定競技スコアと真値競技スコアの組み合わせと、全ての推定背景スコアと真値背景スコアの組み合わせとに基づいて損失を算出し、算出した損失に基づいて新たな係数を算出するようにしてもよい。

10

【 0 0 8 3 】

例えば、学習処理部 1 3 は、上記した図 5 の処理において、ループ L 2 s ~ L 2 e の処理が終了した後に、ステップ S a 5 - 2 を行わずに、処理をステップ S a 6 に進める。その後、ループ L 3 s ~ L 3 e の処理が終了した後に、学習処理部 1 3 は、ステップ S a 5 - 3 の処理において、内部の記憶領域において生成されている全ての推定背景スコアと真値背景スコアの組み合わせと、全ての推定競技者スコアと真値競技者スコアの組み合わせとに基づいて損失を算出し、算出した損失に基づいて新たな係数を算出するようにしてもよい。

20

【 0 0 8 4 】

上記した図 5 の処理では、ステップ S a 2 の処理において、学習処理部 1 3 は、原映像データの訓練データセット、競技者マスク映像データの訓練データセット、背景マスク映像データの訓練データセットの順に選択するようにしているが、この順に限られるものではなく、選択する順番を任意に変更するようにしてもよい。この場合、例えば、学習処理部 1 3 は、原映像データの訓練データセット、競技者マスク映像データの訓練データセット、背景マスク映像データの訓練データセットの順に選択するときには、例えば、ループ L 1 s ~ L 1 e の処理が終了した後に、ステップ S a 5 - 1 を行わずに、処理をステップ S a 6 に進める。その後、ループ L 3 s ~ L 3 e の処理が終了した後、学習処理部 1 3 は、ステップ S a 5 - 3 の処理において、内部の記憶領域において生成されている全ての推定競技スコアと真値競技スコアの組み合わせと、全ての推定競技者スコアと真値競技者スコアの組み合わせとに基づいて損失を算出し、算出した損失に基づいて新たな係数を算出するようにしてもよい。

30

【 0 0 8 5 】

このように、ステップ S a 2 の処理における原映像データの訓練データセット、競技者マスク映像データの訓練データセット、背景マスク映像データの訓練データセットの選択順は、任意に定めるようにしてもよい。学習処理部 1 3 は、推定競技スコアと真値競技スコアの組み合わせ、推定背景スコアと真値背景スコアの組み合わせ、推定競技者スコアと真値競技者スコアの組み合わせを任意に選択して、損失を算出し、算出した損失に基づいて新たな係数を算出するようにしてもよい。

40

【 0 0 8 6 】

上記した図 5 の処理では、学習処理部 1 3 は、例えば、原映像データの訓練データセットを選択した場合、原映像データの処理回数が、N 以上となるまで、再び行われるステップ S a 2 の処理において、繰り返し原映像データの訓練データセットを選択するようにしている。しかしながら、学習処理部 1 3 は、前回のステップ S a 2 において選択した訓練データセットとは異なる他の訓練データセットを選択するようにしてもよい。

【 0 0 8 7 】

上記したその他の学習ルール（その 1）と、学習ルール（その 2）と、学習ルール（その 3）とを任意に組み合わせた学習ルールを予め定めるようにして

50

もよい。

【0088】

(推定装置の構成)

図6は、本発明の実施形態による推定装置2の構成を示すブロック図である。推定装置2は、入力部21、推定部22及び学習モデルデータ記憶部23を備える。学習モデルデータ記憶部23は、学習装置1が図5に示す処理を終了した際に学習モデルデータ記憶部15が記憶する学習済みの係数、すなわち学習済みの学習モデルデータを予め記憶する。入力部21は、任意の映像データ、すなわち任意の競技者が行う一連の動作を背景と共に記録した評価対象の映像データ(以下、評価対象映像データという)を取り込む。

【0089】

推定部22は、学習処理部13が備える関数近似器14と同一の構成である関数近似器を内部に備える。推定部22は、入力部21が取り込んだ評価対象映像データと、学習モデルデータ記憶部23が記憶する学習済みの係数を適用した関数近似器、すなわち学習済みの学習モデルとに基づいて、映像データに対応する推定スコアを算出する。

【0090】

(推定装置による推定処理)

図7は、推定装置2による処理の流れを示すフローチャートである。入力部21は、評価対象映像データを取り込み、取り込んだ評価対象映像データを推定部22に出力する(ステップSb1)。推定部22は、入力部21が出力する評価対象映像データを取り込む。推定部22は、学習モデルデータ記憶部23から学習済みの係数を読み出す。推定部22は、読み出した学習済みの係数を内部に備える関数近似器に適用する(ステップSb2)。

【0091】

推定部22は、取り込んだ評価対象映像データを関数近似器に与える(ステップSb3)。推定部22は、関数近似器の出力値を、評価対象映像データに対する推定スコアとして出力する(ステップSb4)。

【0092】

上記の実施形態の学習装置1は、原映像データと、競技者マスク映像データと、背景マスク映像データとを入力とし、原映像データを入力とした場合に、真値競技スコアを出力とし、競技者マスク映像データを入力とした場合に、真値背景スコアを出力とし、背景マスク映像データを入力とした場合に、真値競技者スコアを出力とする学習モデルにおける学習モデルデータを生成する。学習装置1は、原映像データと、競技者マスク映像データと、背景マスク映像データとを用いて学習処理を行うことにより、映像データの中の競技者の動作に関する特徴を抽出するように促進される。それにより、学習装置1は、関節情報を明示的に与えることなく、競技者の動作を記録した映像データから、競技者の動作に汎化した学習モデルデータを生成することが可能になる。このようにして学習装置1が生成した学習済みの学習モデルデータを関数近似器に適用して生成される学習済みの学習モデルを用いて行う推定装置2による推定処理において、競技における採点の精度を高めることが可能になる。

【0093】

なお、上記の実施形態では、原映像データに1人の競技者が含まれる例を示しているが、原映像データに記録される競技は、複数の競技者によって行われる競技であってもよく、この場合の矩形領域は、複数の競技者を囲む領域になる。

【0094】

上記の実施形態では、競技者の領域を囲む形状を矩形形状にしているが、矩形形状に限られるものではなく、矩形形状以外の形状であってもよい。

【0095】

上記の実施形態では、競技者マスク映像データ及び背景マスク映像データの映像データにおいて、マスクする際の色を、マスクを行う画像フレームにおける平均色としている。これに対して、競技者マスク映像データ及び背景マスク映像データの映像データの各々に

10

20

30

40

50

対応する原映像データに含まれる全ての画像フレームの平均色を、マスクする際の色として選択してもよい。映像データごとに、任意に定める色をマスクする際の色としてもよい。なお、マスクする際の色は、目立たないようにした方がよいため、画像フレームごとの全体の色合いに応じて、目立たない色が選択する必要があり、その点では、背景に溶け込んで目立たない色合いになる画像フレームごとの平均色をマスクする際の色として選択するのが最も効果的であると考えられる。

【0096】

上記の実施形態の学習装置1の学習部12が備える関数近似器14と、推定装置2の推定部22が内部に備える関数近似器は、例えば、DNNであるとしているが、DNN以外のニューラルネットワークや、機械学習による手段や、関数近似器において近似する関数の係数を算出する任意の手段を適用してもよい。

10

【0097】

学習装置1と推定装置2とは、一体化されて構成されてもよい。このように構成される場合、学習装置1と推定装置2とが一体化された装置は、学習モードと推定モードとを有する。学習モードは、学習装置1による学習処理を行って学習モデルデータを生成するモードである。すなわち、学習モードでは、学習装置1と推定装置2とが一体化された装置は、図5に示す処理を実行する。推定モードは、学習済みの学習モデル、すなわち学習済みの学習モデルデータが適用された関数近似器を用いて推定スコアを出力するモードである。すなわち、推定モードでは、学習装置1と推定装置2とが一体化された装置は、図7に示す処理を実行する。

20

【0098】

上述した実施形態における学習装置1及び推定装置2をコンピュータで実現するようにしてもよい。その場合、この機能を実現するためのプログラムをコンピュータ読み取り可能な記録媒体に記録して、この記録媒体に記録されたプログラムをコンピュータシステムに読み込ませ、実行することによって実現してもよい。なお、ここでいう「コンピュータシステム」とは、OSや周辺機器等のハードウェアを含むものとする。また、「コンピュータ読み取り可能な記録媒体」とは、フレキシブルディスク、光磁気ディスク、ROM、CD-ROM等の可搬媒体、コンピュータシステムに内蔵されるハードディスク等の記憶装置のことをいう。さらに「コンピュータ読み取り可能な記録媒体」とは、インターネット等のネットワークや電話回線等の通信回線を介してプログラムを送信する場合の通信線のように、短時間の間、動的にプログラムを保持するもの、その場合のサーバやクライアントとなるコンピュータシステム内部の揮発性メモリのように、一定時間プログラムを保持しているものも含んでもよい。また上記プログラムは、前述した機能の一部を実現するためのものであってもよく、さらに前述した機能をコンピュータシステムにすでに記録されているプログラムとの組み合わせで実現できるものであってもよく、FPGA(Field Programmable Gate Array)等のプログラマブルロジックデバイスを用いて実現されるものであってもよい。

30

【0099】

以上、この発明の実施形態について図面を参照して詳述してきたが、具体的な構成はこの実施形態に限られるものではなく、この発明の要旨を逸脱しない範囲の設計等も含まれる。

40

【産業上の利用可能性】

【0100】

スポーツ競技における競技の採点に利用することができる。

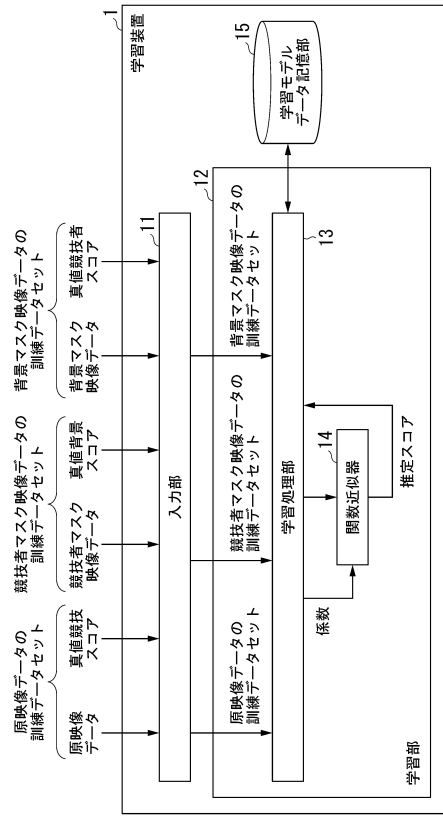
【符号の説明】

【0101】

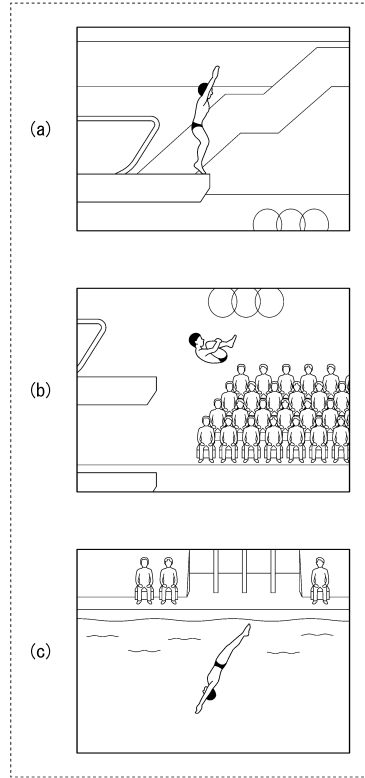
1...学習装置、11...入力部、12...学習部、13...学習処理部、14...関数近似器、15...学習モデルデータ記憶部、2...推定装置、21...入力部、22...推定部、23...学習モデルデータ記憶部

50

【図面】
【図 1】



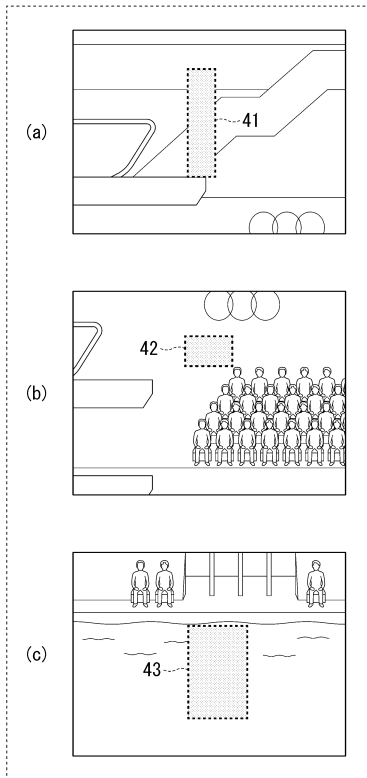
【図 2】



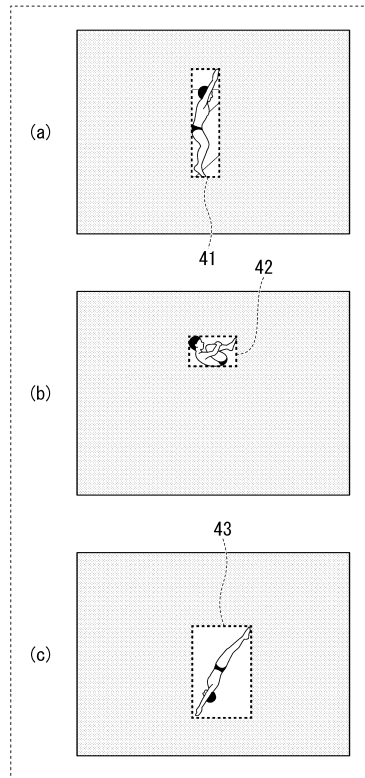
10

20

【図 3】



【図 4】

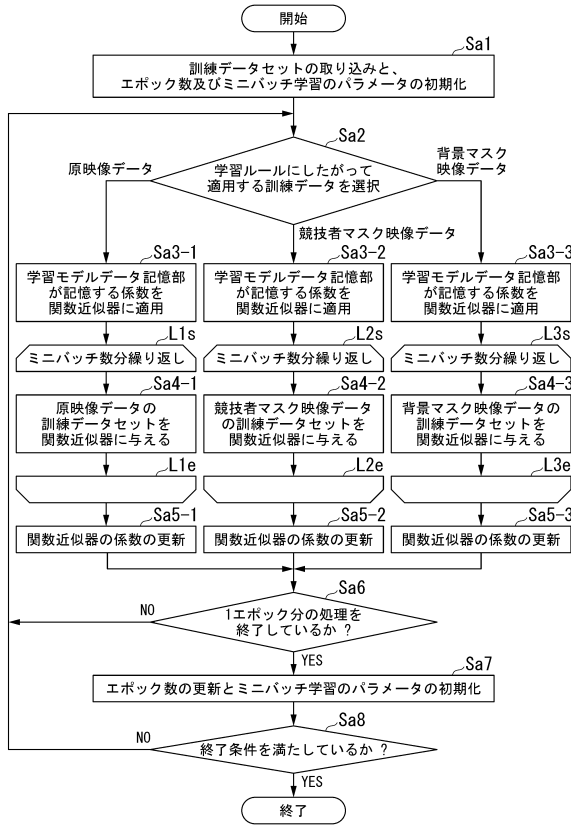


30

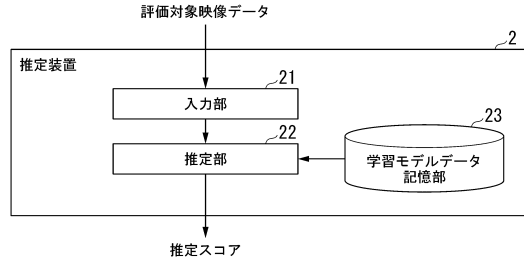
40

50

【図5】



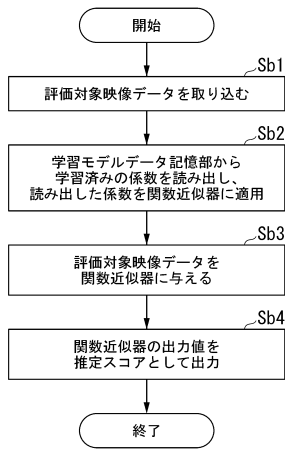
【図6】



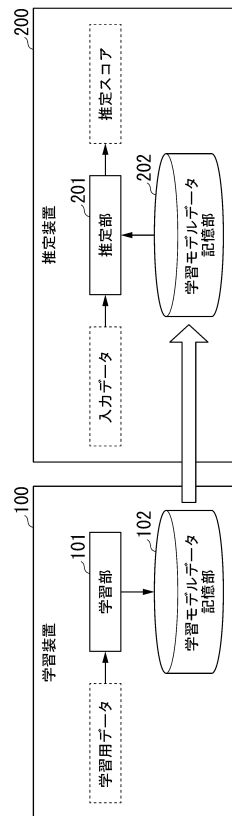
10

20

【図7】



【図8】

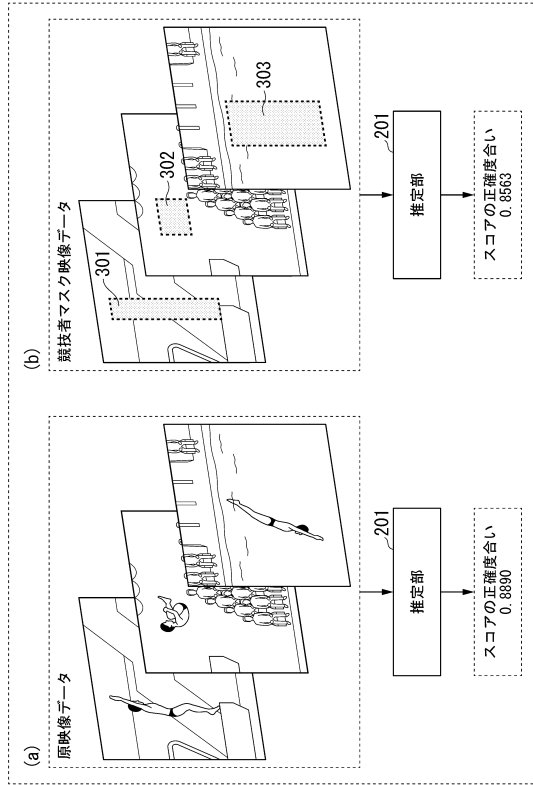


30

40

50

【図9】



10

20

30

40

50

フロントページの続き

東京都千代田区大手町一丁目5番1号 日本電信電話株式会社内

(72)発明者 山本 奏

東京都千代田区大手町一丁目5番1号 日本電信電話株式会社内

審査官 秦野 孝一郎

(56)参考文献

特開2021-71953(JP,A)

国際公開第2021/064963(WO,A1)

国際公開第2021/064960(WO,A1)

国際公開第2021/064830(WO,A1)

特開2021-47164(JP,A)

国際公開第2021/002025(WO,A1)

国際公開第2020/084667(WO,A1)

国際公開第2020/050111(WO,A1)

国際公開第2019/225692(WO,A1)

岩田 あきほ、川島 寛乃、大越 匡、中澤 仁、深層学習モデルを用いたフィギュアスケートにおけるステップシーケンスの元素認識，情報処理学会 研究報告 モバイルコンピューティングとパーベシブシステム(MBL) 2020-MBL-096 [online]，日本，情報処理学会，2020年09月22日，p.1-8

(58)調査した分野 (Int.Cl., DB名)

H04N 7/18

G06T 7/00 - 7/90

G06V 10/00 - 20/90