

(19)日本国特許庁(JP)

(12)公開特許公報(A)

(11)公開番号

特開2022-75661

(P2022-75661A)

(43)公開日 令和4年5月18日(2022.5.18)

(51)国際特許分類	F I	テーマコード(参考)
H 0 4 N 5/91 (2006.01)	H 0 4 N 5/91	5 C 0 5 3
G 0 6 T 7/00 (2017.01)	G 0 6 T 7/00 6 6 0	5 L 0 9 6
G 1 0 L 15/10 (2006.01)	G 1 0 L 15/10 2 0 0 W	

審査請求 有 請求項の数 18 O L (全24頁)

(21)出願番号	特願2022-13826(P2022-13826)	(71)出願人	520408744
(22)出願日	令和4年2月1日(2022.2.1)		株式会社 I ' m b e s i d e y o u
(62)分割の表示	特願2021-572314(P2021-572314)		東京都世田谷区成城2丁目34番2号
)の分割		村上ハイツ301
原出願日	令和2年10月27日(2020.10.27)	(74)代理人	110002790
			One i p特許業務法人
		(72)発明者	神谷 渉三
			東京都世田谷区成城2丁目34番2号村
			上ハイツ301 株式会社 I ' m b e s i
			d e y o u 内
		Fターム(参考)	5C053 FA14 LA01 LA11 LA14
			5L096 CA04 FA67 HA02 JA11

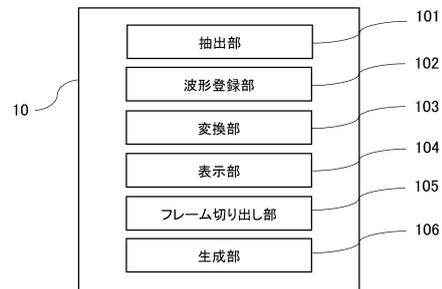
(54)【発明の名称】 情報抽出装置

(57)【要約】 (修正有)

【課題】演算処理に伴う負荷を軽減し得る情報抽出装置を提供すること。

【解決手段】在宅個別指導システムにおいて、情報抽出装置である教室映像配信装置10は、複数のフレームから構成される動画画像から、外部から指示される所定の特定条件に従って特定のフレーム群を抽出する抽出部と、所定の被写体を撮像して前記動画画像を取得する撮像部と、被写体を含む周囲の音を收音する收音部と、所定の波形データを予め登録する波形登録部と、を備え、收音部により收音された音の波形データと、波形登録部に登録されている波形データとが一致した場合に、当該一致した波形に対応するフレーム群を特定のフレーム群として動画画像から抽出する。

【選択図】図4



【特許請求の範囲】**【請求項 1】**

複数のフレームから構成される動画像を取得する取得部と、当該動画像内に含まれる所定のデータを特定するための特定条件を記憶する記憶部と、当該特定条件に従って、前記動画像から特定のフレーム群を複数抽出する抽出部と、抽出された前記特定のフレーム群同士を連結する連結部と、連結された複数のフレーム群を含むダイジェスト情報を入力する出力部と、を備える、情報抽出装置。

【請求項 2】

請求項 1 に記載の情報抽出装置であって、
所定の波形データを予め登録する波形登録部、を更に備え、
前記特定条件は、前記動画像内に含まれる音の波形データと前記登録されている波形データとが一致するか否かであって、
前記抽出部は、両波形データが一致した場合に、当該一致した波形に対応するフレーム群を前記特定のフレーム群として前記動画像から抽出する、
情報抽出装置。

10

【請求項 3】

請求項 2 に記載の情報抽出装置であって、
動画内に含まれる前記音を音声認識によりテキスト情報に変換する変換部を更に備え、
前記変換部は、前記特定のフレーム群とその前後所定フレーム数とを含む補助フレーム群
に対応する前記音を変換する、
情報抽出装置。

20

【請求項 4】

請求項 2 又は請求項 3 に記載の情報抽出装置であって、
前記被写体を含む周囲の音が示す情報には、会話情報と非会話情報とが混在する、
情報抽出装置。

【請求項 5】

請求項 4 に記載の情報抽出装置であって、
前記会話情報には、ポジティブな感情を示すワードと、ネガティブな感情を示すワードの
少なくとも何れかが含まれる、
情報抽出装置。

30

【請求項 6】

請求項 4 又は請求項 5 に記載の情報抽出装置であって、
前記非会話情報には、舌打ち、溜め息、相槌の少なくとも何れかを示す情報が含まれる、
情報抽出装置。

【請求項 7】

請求項 1 に記載の情報抽出装置であって、
顔の表情に関する所定の顔評価値を予め登録する顔情報登録部、を更に備え、
前記特定条件は、前記動画像内に含まれる顔の表情から算出される顔評価値と前記登録さ
れている顔評価値とが一致するか否かであって、
前記抽出部は、両顔評価値が一致した場合に、当該一致した顔評価値に対応するフレーム
群を前記特定のフレーム群として前記動画像から抽出する、
情報抽出装置。

40

【請求項 8】

請求項 7 に記載の情報抽出装置であって、
前記顔評価値には、前記人物の幸福感、退屈感又は緊張感の度合いを評価した評価値が含
まれる、
情報抽出装置。

【請求項 9】

請求項 7 又は請求項 8 に記載の情報抽出装置であって、

50

前記顔評価値には、前記人物の表情、前記人物の視線の向き、前記人物の顔の向きを評価した評価値が含まれる、

情報抽出装置

【請求項 10】

請求項 1 に記載の情報抽出装置であって、

人物の動作に関する所定の動作評価値を予め登録する動作情報登録部、を更に備え、

前記特定条件は、前記動画像内に含まれる人物から算出される動作評価値と前記登録されている動作評価値とが一致するか否かであって、

前記抽出部は、両動作評価値が一致した場合に、当該一致した動作評価値に対応するフレーム群を前記特定のフレーム群として前記動画像から抽出する、

10

情報抽出装置。

【請求項 11】

請求項 10 に記載の情報抽出装置であって、

前記動作評価値には、前記人物の身振り、手振り、ジェスチャ、ボディランゲージの少なくとも何れかの動作を評価した評価値が含まれる、

情報抽出装置。

【請求項 12】

請求項 1 に記載の情報抽出装置であって、

所定の生体情報に関する生体評価値を予め登録する生体情報登録部と、を備え、

前記特定条件は、前記動画像内に含まれる人物から算出可能な生体評価値と、前記登録されている生体評価値とが一致するか否かであって、

前記抽出部は、両生体評価値が一致した場合に、当該一致した生体評価値に対応するフレーム群を前記特定のフレーム群として前記動画像から抽出する、

20

情報抽出装置。

【請求項 13】

請求項 12 に記載の情報抽出装置であって、

前記生体評価値には、前記人物の血圧、脈拍、脈圧の少なくとも何れかが含まれる、

情報抽出装置。

【請求項 14】

請求項 1 乃至請求項 13 の何れか一項に記載の情報抽出装置であって、

前記特定のフレーム群に対して、当該特定のフレーム群と時系列的に前後に連続する追加フレームを追加するフレーム追加部を備えている、

30

情報抽出装置。

【請求項 15】

請求項 1 乃至請求項 14 の何れかに記載の情報抽出装置によって抽出されたダイジェスト情報に含まれる少なくとも顔画像又は音声を所定のフレーム単位ごとに識別する識別手段と、

識別した前記顔画像に関する評価値を算出する評価手段とを更に備える、

ビデオミーティング評価端末。

【請求項 16】

40

請求項 15 に記載のビデオミーティング評価端末であって、

ビデオミーティング評価端末は、前記評価値の時系列によるグラフ情報を提供する、

ビデオミーティング評価端末。

【請求項 17】

請求項 15 又は請求項 16 に記載のビデオミーティング評価端末であって、

前記ビデオミーティング評価端末は、前記顔画像を複数の異なる観点によって評価した複数の評価値を算出する、

ビデオミーティング評価端末。

【請求項 18】

請求項 15 乃至請求項 17 のいずれかに記載のビデオミーティング評価端末であって、

50

前記ビデオミーティング評価端末は、前記動画像に含まれる音声と共に前記評価値を算出する、

ビデオミーティング評価端末。

【請求項 19】

請求項 15 乃至請求項 18 のいずれかに記載のビデオミーティング評価端末であって、前記ビデオミーティング評価端末は、前記動画像内に含まれる前記顔画像以外の対象物と共に前記評価値を算出する、

ビデオミーティング評価端末。

【発明の詳細な説明】

【技術分野】

10

【0001】

本開示は、情報抽出装置に関する。

【背景技術】

【0002】

近年、各種コンテンツの配信を受ける配信サービスが普及しつつある。

【0003】

特許文献 1 には、ユーザが配信を希望する音楽コンテンツの曲名等がわからないときでも、検索対象である楽曲の鼻歌を入力することで、所望の音楽コンテンツを検出する処理を可能にした技術が開示されている。

【先行技術文献】

20

【特許文献】

【0004】

【特許文献 1】特開 2002 - 55994 号公報

【発明の概要】

【発明が解決しようとする課題】

【0005】

ところで、特許文献 1 に記載の技術は、配信されるコンテンツが音楽コンテンツに限られるため、それ以外のあらゆる動画コンテンツに対して検索対象を検出する処理を行うには、コンピュータによる膨大な演算処理が必要となる。

【0006】

30

そこで、本開示は、このような状況に鑑みてなされたものであり、演算処理に伴う負荷を軽減し得る情報抽出装置を提供することを一つの目的とする。

【課題を解決するための手段】

【0007】

上記課題を解決するための本発明の主たる発明は、複数のフレームから構成される動画像から、外部から指示される所定の特定条件に従って特定のフレーム群を抽出する抽出部を備えることを特徴とする。

【発明の効果】

【0008】

本開示によれば、演算処理に伴う負荷を軽減し得る。

40

【図面の簡単な説明】

【0009】

【図 1】本開示の第 1 の実施形態に係る在宅個別指導システム 1 の構成例を示す概念図である。

【図 2】本開示の第 1 の実施形態に係る教室映像配信装置 10 を実現するコンピュータのハードウェア構成例を示す図である。

【図 3】本開示の第 1 の実施形態に係る受講生端末 20 を実現するコンピュータのハードウェア構成例を示す図である。

【図 4】本開示の第 1 の実施形態に係る教室映像配信装置 10 のソフトウェア構成例を示す図である。

50

【図 5】本開示の第 1 の実施の形態による教室映像配信方法の処理の流れを説明するフローチャートである。

【図 6】本開示の第 2 の実施形態に係る教室映像配信装置 10 のソフトウェア構成例を示す図である。

【図 7】本開示の第 2 の実施の形態による教室映像配信方法の処理の流れを説明するフローチャートである。

【図 8】本開示の第 3 の実施形態に係る教室映像配信装置 10 のソフトウェア構成例を示す図である。

【図 9】本開示の第 3 の実施の形態による教室映像配信方法の処理の流れを説明するフローチャートである。

10

【図 10】本開示の第 4 の実施形態に係る教室映像配信装置 10 のソフトウェア構成例を示す図である。

【図 11】本開示の第 4 の実施の形態による教室映像配信方法の処理の流れを説明するフローチャートである。

【発明を実施するための形態】

【0010】

本開示の実施形態の内容を列記して説明する。本開示は、以下のような構成を備える。

[項目 1]

複数のフレームから構成される動画像を取得する取得部と、
当該動画像内に含まれる所定のデータを特定するための特定条件を記憶する記憶部と、
当該特定条件に従って、前記動画像から特定のフレーム群を複数抽出する抽出部と、
抽出された前記特定のフレーム群同士を連結する連結部と、
連結された複数のフレーム群を含むダイジェスト情報を出力する出力部と、を備える、
情報抽出装置。

20

[項目 2]

項目 1 に記載の情報抽出装置であって、
所定の波形データを予め登録する波形登録部、を更に備え、
前記特定条件は、前記動画像内に含まれる音の波形データと前記登録されている波形データとが一致するか否かであって、
前記抽出部は、両波形データが一致した場合に、当該一致した波形に対応するフレーム群
を前記特定のフレーム群として前記動画像から抽出する、
情報抽出装置。

30

[項目 3]

項目 2 に記載の情報抽出装置であって、
動画内に含まれる前記音を音声認識によりテキスト情報に変換する変換部を更に備え、
前記変換部は、前記特定のフレーム群とその前後所定フレーム数とを含む補助フレーム群
に対応する前記音を変換する、
情報抽出装置。

[項目 4]

項目 2 又は項目 3 に記載の情報抽出装置であって、
前記被写体を含む周囲の音が示す情報には、会話情報と非会話情報とが混在する、
情報抽出装置。

40

[項目 5]

項目 4 に記載の情報抽出装置であって、
前記会話情報には、ポジティブな感情を示すワードと、ネガティブな感情を示すワードの
少なくとも何れかが含まれる、
情報抽出装置。

[項目 6]

項目 4 又は項目 5 に記載の情報抽出装置であって、
前記非会話情報には、舌打ち、溜め息、相槌の少なくとも何れかを示す情報が含まれる、

50

情報抽出装置。

[項目 7]

項目 1 に記載の情報抽出装置であって、
顔の表情に関する所定の顔評価値を予め登録する顔情報登録部、を更に備え、
前記特定条件は、前記動画像内に含まれる顔の表情から算出される顔評価値と前記登録されている顔評価値とが一致するか否かであって、
前記抽出部は、両顔評価値が一致した場合に、当該一致した顔評価値に対応するフレーム群を前記特定のフレーム群として前記動画像から抽出する、
情報抽出装置。

10

[項目 8]

項目 7 に記載の情報抽出装置であって、
前記顔評価値には、前記人物の幸福感、退屈感又は緊張感の度合いを評価した評価値が含まれる、
情報抽出装置。

[項目 9]

項目 7 又は項目 8 に記載の情報抽出装置であって、
前記顔評価値には、前記人物の表情、前記人物の視線の向き、前記人物の顔の向きを評価した評価値が含まれる、
情報抽出装置

20

[項目 10]

項目 1 に記載の情報抽出装置であって、
人物の動作に関する所定の動作評価値を予め登録する動作情報登録部、を更に備え、
前記特定条件は、前記動画像内に含まれる人物から算出される動作評価値と前記登録されている動作評価値とが一致するか否かであって、
前記抽出部は、両動作評価値が一致した場合に、当該一致した動作評価値に対応するフレーム群を前記特定のフレーム群として前記動画像から抽出する、
情報抽出装置。

[項目 11]

項目 10 に記載の情報抽出装置であって、
前記動作評価値には、前記人物の身振り、手振り、ジェスチャ、ボディランゲージの少なくとも何れかの動作を評価した評価値が含まれる、
情報抽出装置。

30

[項目 12]

項目 1 に記載の情報抽出装置であって、
所定の生体情報に関する生体評価値を予め登録する生体情報登録部と、を備え、
前記特定条件は、前記動画像内に含まれる人物から算出可能な生体評価値と、前記登録されている生体評価値とが一致するか否かであって、
前記抽出部は、両生体評価値が一致した場合に、当該一致した生体評価値に対応するフレーム群を前記特定のフレーム群として前記動画像から抽出する、
情報抽出装置。

40

[項目 13]

項目 12 に記載の情報抽出装置であって、
前記生体評価値には、前記人物の血圧、脈拍、脈圧の少なくとも何れかが含まれる、
情報抽出装置。

[項目 14]

項目 1 乃至項目 13 の何れか一項に記載の情報抽出装置であって、
前記特定のフレーム群に対して、当該特定のフレーム群と時系列的に前後に連続する追加フレームを追加するフレーム追加部を備えている、
情報抽出装置。

[項目 15]

50

項目 1 乃至項目 1 4 の何れかに記載の情報抽出装置によって抽出されたダイジェスト情報に含まれる少なくとも顔画像又は音声をもとに所定のフレーム単位ごとに識別する識別手段と、識別した前記顔画像に関する評価値を算出する評価手段とを更に備える、ビデオミーティング評価端末。

[項目 1 6]

項目 1 5 に記載のビデオミーティング評価端末であって、ビデオミーティング評価端末は、前記評価値の時系列によるグラフ情報を提供する、ビデオミーティング評価端末。

[項目 1 7]

項目 1 5 又は項目 1 6 に記載のビデオミーティング評価端末であって、前記ビデオミーティング評価端末は、前記顔画像を複数の異なる観点によって評価した複数の評価値を算出する、ビデオミーティング評価端末。

10

[項目 1 8]

項目 1 5 乃至項目 1 7 のいずれかに記載のビデオミーティング評価端末であって、前記ビデオミーティング評価端末は、前記動画画像に含まれる音声と共に前記評価値を算出する、ビデオミーティング評価端末。

[項目 1 9]

項目 1 5 乃至項目 1 8 のいずれかに記載のビデオミーティング評価端末であって、前記ビデオミーティング評価端末は、前記動画画像内に含まれる前記顔画像以外の対象物と共に前記評価値を算出する、ビデオミーティング評価端末。

20

【 0 0 1 1 】

以下に添付図面を参照しながら、本開示の好適な実施の形態について詳細に説明する。なお、本明細書及び図面において、実質的に同一の機能構成を有する構成要素については、同一の符号を付することにより重複説明を省略する。

【 0 0 1 2 】

本開示では、デジタル通信回線を介して学習塾・教育機関と生徒・受講生宅を結び、各生徒・受講生は自宅に居ながら学習塾・教育機関で講義されている講義映像を視聴して、学習塾・教育機関の授業を受けられる在宅個別指導システムに情報抽出装置を適用する例を説明する。

30

【 0 0 1 3 】

< 第 1 の実施形態 >

図 1 は、本開示の第 1 の実施形態に係る在宅個別指導システム 1 の構成例を示す概念図である。図示するように、この在宅個別指導システム 1 では、遠隔授業を行う講師 T の教室側に設けられた教室映像配信装置 1 0 と、それぞれの在宅で指導を受ける受講生群（受講生 A、B、C）に夫々関連する受講生端末 2 0 A、2 0 B、2 0 C と、がネットワーク N W を介して通信可能に接続されている。なお以下では、受講生端末 2 0 A、2 0 B、2 0 C を特に区別して説明する必要がない場合には、単に受講生端末 2 0 と略記する。同様に、受講生 A、B、C を特に区別して説明する必要がない場合には、単に受講生と略記する。

40

【 0 0 1 4 】

教室映像配信装置 1 0 は、請求の範囲に記載された情報抽出装置の一例となる。なお、本構成は一例であり、ある構成が他の構成を兼ね備えていたり、他の構成が含まれていたりしてもよい。なお、ここでは受講生 A、B、C の 3 名の場合を示しているが、講師が同時に指導できる人数又はネットワーク N W の接続回線数等に応じて、さらに多人数としてもよい。

【 0 0 1 5 】

本実施形態において、「講師」とは、教授、教諭、教師を含む概念である。「教室」とは

50

、学習塾、カルチャーセンター、教育機関（例えば、初等・中等・高等教育機関、高等学校、高等専門学校、専門学校、短期大学、四年制大学、大学院など、文部科学省に登録されている学校）を含む概念である。「受講生」とは、生徒、学生、聴講生を含む概念である。

【0016】

本実施形態においてネットワークNWはインターネットを想定している。ネットワークNWは、例えば、公衆電話回線網、携帯電話回線網、無線通信網、イーサネット（登録商標）などにより構築される。

【0017】

<ハードウェア構成>

図2は、本実施形態に係る教室映像配信装置10を実現するコンピュータのハードウェア構成例を示す図である。コンピュータは、少なくとも、通信部11と、撮像部12と、收音部13と、モニタ14と、メモリ15と、ストレージ16と、入出力部17と、制御部18等を備える。これらはバス19を通じて相互に電氣的に接続される。

【0018】

通信部11は、教室映像配信装置10をネットワークNWに接続する。通信部11は、例えば、有線LAN（Local Area Network）、無線LAN、Wi-Fi（Wireless Fidelity、登録商標）、赤外線通信、Bluetooth（登録商標）、近距離または非接触通信等の方式で、外部機器と直接またはネットワークアクセスポイントを介して通信する。

【0019】

撮像部12は、CMOS又はCCDなどの撮像素子を用いて電子撮影する機能を有する。撮像部12は、受講生に対する講義を行う講師Tを被写体として撮像して、講師映像を取得する。撮像部12は、講師Tが講義を進行する際に使用する黒板又はホワイトボードに記載した画像も撮像できる構成とするとよいが、黒板又はホワイトボードの為に独立したカメラを設けてもよい。

【0020】

收音部13は、講師Tを含む周囲の音を收音する。收音部13は、講師Tの音声を含む周囲の音を取得するためのマイクロフォン等を備える。さらに、收音部13は、取得した音を電気信号に変換する等の適宜処理を行い得る。

【0021】

モニタ14は、受講生端末20から送信される受講生映像と、撮像部12で取得される講師映像とを一覧可能な状態で表示し得る。もちろん、モニタ14は、受講生映像のみを単独で表示してもよく、講師映像のみを単独で表示してもよい。

【0022】

メモリ15は、DRAM（Dynamic Random Access Memory）等の揮発性記憶装置で構成される主記憶と、フラッシュメモリ又はHDD（Hard Disc Drive）等の不揮発性記憶装置で構成される補助記憶と、を含む。メモリ15は、制御部18のワークエリア等として使用され、また、教室映像配信装置10の起動時に実行されるBIOS（Basic Input/Output System）、及び各種設定情報等を格納する。

【0023】

ストレージ16は、アプリケーション・プログラム等の各種プログラムを格納する。各処理に用いられるデータを格納したデータベースがストレージ16に構築されていてもよい。

【0024】

入出力部17は、例えば、キーボード、マウス、タッチパネル等の情報入力機器である。

【0025】

制御部18は、教室映像配信装置10全体の動作を制御し、各要素間におけるデータの送受信の制御、及びアプリケーションの実行及び認証処理に必要な情報処理等を行う演算装

10

20

30

40

50

置である。例えば制御部 18 は、CPU (Central Processing Unit) 等のプロセッサであり、ストレージ 16 に格納されメモリ 15 に展開されたプログラム等を実行して各情報処理を実施する。

【0026】

バス 19 は、上記各要素に共通に接続され、例えば、アドレス信号、データ信号及び各種制御信号を伝達する。

【0027】

図 3 は、本実施形態に係る受講生端末 20 を実現するコンピュータのハードウェア構成例を示す図である。コンピュータは、少なくとも、通信部 21 と、撮像部 22 と、收音部 23 と、モニタ 24 と、メモリ 25 と、ストレージ 26 と、入出力部 27 と、制御部 28 等を備える。これらはバス 29 を通じて相互に電氣的に接続される。本実施形態に係る受講生端末 20 を実現するコンピュータ等のハードウェア構成は、図 2 に示す教室映像配信装置 10 のハードウェア構成例と同様であるため、相違点のみ説明する。

10

【0028】

通信部 21 は、受講生端末 20 をネットワーク NW に接続する。

【0029】

撮像部 22 は、講義を受講する受講生を被写体として撮像して、受講生映像を取得する。

【0030】

收音部 23 は、受講生を含む周囲の音を收音する。音声データを送受するために、受講生端末 20 においては、マイク付きヘッドフォンを設けてもよいが、当該端末に内蔵されたマイク並びにスピーカを用いてもよい。

20

【0031】

モニタ 24 は、教室映像配信装置 10 から送信される講師映像と、撮像部 22 で取得される受講生映像とを一覧可能な状態で表示し得る。もちろん、モニタ 24 は、講師映像のみを単独で表示してもよく、受講生映像のみを単独で表示してもよい。

【0032】

制御部 28 は、受講生端末 20 全体の動作を制御し、各要素間におけるデータの送受信の制御、及びアプリケーションの実行及び認証処理に必要な情報処理等を行う演算装置である。

【0033】

30

<ソフトウェア構成>

図 4 は、本実施形態に係る教室映像配信装置 10 のソフトウェア構成例を示す図である。教室映像配信装置 10 は、抽出部 101 と、波形登録部 102 と、変換部 103 と、表示部 104 と、フレーム切り出し部 105 と、生成部 106 と、を備える。

【0034】

抽出部 101 と、波形登録部 102 と、変換部 103 と、表示部 104 と、フレーム切り出し部 105 と、生成部 106 とは、制御部 18 がストレージ 16 に記憶されているプログラムをメモリ 15 に読み出して実行することにより実現され得る。

【0035】

抽出部 101 は、撮像部 12 で取得される講師映像と、撮像部 22 で取得される受講生映像とを適宜取捨選択して合成すると共に、收音部 13 又は收音部 23 で收音された音を組み合わせる教室映像を生成する。ここでの講師映像又は受講生映像は、請求の範囲に記載された複数のフレームの一例となる。また、教室映像は、請求の範囲に記載された動画の一例となる。教室映像は、テキストデータ、数値データ、図形データ、画像データ、動画データ、音声データ等、又はこれらの組み合わせであり、記憶、編集及び検索等の対象となり、システム又は利用者間で個別の単位として交換できるものをいい、これらに類似するものを含む。

40

【0036】

抽出部 101 は、かかる教室映像から、外部から指示される所定の特定条件に従って特定のフレーム群を抽出する機能を有する。例えば、外部からの指示は、合成する教室映像の

50

項目、画像の配置・画像の占有面積等を指示するものであり得る。外部からの指示は、例えば、講師 T 自身が講義の途中で映像構成を編集可能な簡便な操作であることが好ましい。

【 0 0 3 7 】

波形登録部 1 0 2 は、所定の波形データを予め登録する機能を有する。所定の特定条件とは、例えば、収録部 1 3 又は収録部 2 3 により収録された音の波形データと、波形登録部 1 0 2 に登録されている波形データとが一致するか否かであってよく、もちろん、他の条件であってもよい。本実施形態において波形登録部 1 0 2 に登録されている波形データは、講義におけるその場全体の雰囲気の評価するために用いられ得る。

【 0 0 3 8 】

抽出部 1 0 1 は、所定の特定条件が満たされた場合、例えば、両波形データが一致した場合に、当該一致した波形に対応するフレーム群を特定のフレーム群として前記教室映像から抽出する。

【 0 0 3 9 】

変換部 1 0 3 は、収録部 1 3 が収録した音を音声認識によりテキストに変換する機能を有する。このテキストとは、任意の文字列のことである。変換部 1 0 3 は、音声認識に成功した場合は、生成したテキストを含む音声認識結果を出力する。音声認識結果に、音声認識が成功したことを示す成功情報を含めてもよい。

【 0 0 4 0 】

表示部 1 0 4 は、変換部 1 0 3 により変換されたテキストをモニタ 1 4 又はモニタ 2 4 に表示する機能を有する。講師 T 又は受講生 A、B、C を含む周囲の音を示す情報には、会話情報と非会話情報とが混在する。会話情報には、例えば、ポジティブな感情を示すワードと、ネガティブな感情を示すワードの少なくとも何れかが含まれる。

【 0 0 4 1 】

ポジティブな感情を示すワードの一例としては、講師が受講生を褒めたり、応援したり、励ましたりする内容として、「よく頑張ったね」「努力したね」「すごいね」「素晴らしいね」「立派だね」「偉いね」等を挙げることができる。

【 0 0 4 2 】

一方、ネガティブな感情を示すワードの一例としては、講師が受講生を貶したり、非難したり、誹謗したりする内容として、「そんなんじゃないだ」「お前はダメだ」「なにやってんのよ」「落ちるぞ」「バカ」等を挙げることができる。

【 0 0 4 3 】

非会話情報は、会話情報以外のテキスト情報である。非会話情報には、舌打ち、溜め息、相槌の少なくとも何れかを示す情報が含まれる。これらの非会話情報は、講義を受講する受講生の感情を判断するための判断基準となり得る。受講生の感情は、例えば、「幸福感」、「退屈感」、「緊張感」の 3 つに分類され得る。

【 0 0 4 4 】

フレーム切り出し部 1 0 5 は、抽出部 1 0 1 により抽出された特定のフレーム群に対して、少なくとも時系列的に前後に連続するフレーム群を切り出す機能を有する。フレーム切り出し部 1 0 5 は、例えば、各ワードがどのような文脈で使用されたかを示す文脈情報を取得するために適用可能な任意のフレームレートを用いて、前後に連続するフレーム群を切り出すことができる。ここでの文脈情報とは、例えば、単語前後の任意範囲の文字列、単語間の共起関係等を示す情報である。

【 0 0 4 5 】

生成部 1 0 6 は、抽出部 1 0 1 により抽出された特定のフレーム群に対して、その前後に連続する先行フレーム群と後続フレームを連結して、ダイジェスト動画を生成する機能を有する。

【 0 0 4 6 】

次に、このように構成された在宅個別指導システム 1 の動作について説明する。図 5 は、本開示の第 1 の実施の形態による教室映像配信方法の処理の流れを説明するフローチャー

10

20

30

40

50

トである。

【0047】

ここでは、教室を運営する運営者等が、講師が受講生に対して不適切な発言・問題発言をしていないかどうかをチェックする場面を例に挙げて説明する。具体的に、予め登録された講師の声でネガティブな感情を示すワードとして「バカ」を表す波形データを用いて、ダイジェスト動画を生成する場面を例に説明する。

【0048】

まず、講義が開始される時刻になると、各受講生 A、B、C は受講生端末 20A、20B、20C をネットワーク NW 経由で在宅個別指導システム 1 に接続して、講師 T の講義開始を待つ。講師 T は、教室映像配信装置 10 が備えるモニタ 14 を見て各受講生 A、B、C が受講態勢にあるか否かを判断し、受講態勢が整っていれば、講義を開始する。すなわち、撮像部 12 並びに撮像部 22 による撮像動作が開始されるとともに、收音部 13 並びに收音部 23 による收音動作が開始される（ステップ S100）。

10

【0049】

そして、抽出部 101 は、撮像部 12 で取得される講師映像と、撮像部 22 で取得される受講生映像とを適宜取捨選択して合成すると共に、收音部 13 又は收音部 23 で收音された音を組み合わせる教室映像を生成すると共に、生成した教室映像をネットワーク NW 経由で受講生端末 20A、20B、20C に配信する（ステップ S102）。

【0050】

次に、講師が「バカ」と発声すると、その音声は收音部 13 で收音されて、その音声データを含む教室映像が特定のフレーム群として抽出部 101 によって抽出される（ステップ S104）。

20

【0051】

フレーム切り出し部 105 は、その教室映像の前後において例えば 10 秒～20 秒程度の時間間隔で連なるフレーム群を切り出す（ステップ S106）。

【0052】

そして、生成部 106 は、これらのフレーム群を連結したダイジェスト動画を生成すると共に、当該ダイジェスト動画に基づいて、講師の音声データに対して所定の音響分析を施す（ステップ S108）。

【0053】

この音響分析によれば、「バカ」というワードがどういう文脈で使われたかを把握することができる。

30

【0054】

例えば、（受講生以外の第三者に対して）「こういうバカなことを言ってる人はダメだね」という文脈で講師において「バカ」というワードが使われた場合には、講師は受講生のことをバカと言ったわけではないことが把握できる。

【0055】

また例えば、「俺はバカだから」と文脈で講師において「バカ」というワードが使われた場合には、講師は自身のことをバカと言っていることが把握できる。

【0056】

かくして、教室を運営する運営者等は、「バカ」というワードとその前後の文脈をテキスト文章として例えばレポート形式で取得し得るので、講師が受講生に対して不適切な発言・問題発言をしていないかどうかを容易にチェックできる。

40

【0057】

すなわち、予め登録済みの波形データを用いた照合により、演算処理に伴う負荷を軽減しながらも、教室映像内において講師が不適切な発言・問題発言を引き起こす可能性の高い状況を含む特徴的なシーンをピンポイントで引き出すことが可能となる。例えば、講義時間（例えば 90 分）に対して、講師によるネガティブな感情を示すワードの発声回数が比較的大きい所定の回数（例えば 10 回）に至るような場合には、講師の人間性を判断することも可能になる。もちろん、講師によるポジティブな感情を示すワードの発声回数も講

50

師の人間性を判断する材料になり得る。さらには、講師の側に限らず、受講生の側においても、講師からの発話に対する舌打ちの回数、溜め息の回数、相槌の回数は、受講生がどのような感情を抱いているかを判断する材料となり得る。

【0058】

<第2の実施形態>

以下、図6及び図7に基づいて、第2の実施形態に係る在宅個別指導システムについて説明する。この実施形態では、上述した第1実施形態で説明した要素と同一の要素について同じ符号を付し、詳細な説明を省略する。

【0059】

上記の第1実施形態に係る在宅個別指導システムでは、既に述べたように、予め登録済みの波形データを用いた照合により、教室映像の中から特定のフレーム群を抽出していたが、第2の実施形態に係る在宅個別指導システムでは、人物の顔の表情に関する評価値に基づいて、教室映像の中から特定のフレーム群を抽出する処理が行われる。

10

【0060】

<ソフトウェア構成>

図6は、本実施形態に係る教室映像配信装置10のソフトウェア構成例を示す図である。教室映像配信装置10は、抽出部101と、フレーム切り出し部105と、生成部106と、顔検出部107と、算出部108と、顔情報登録部109と、を備える。

【0061】

顔検出部107は、例えば、メモリ15に格納されている教室映像を構成する複数のフレームの夫々について、人物の顔検出を行う機能を有する。例えば、顔検出部107は、教室映像の中から特徴点を抽出して、講師T又は各受講生A、B、Cの顔領域、顔領域の大きさ・顔面積等を検出する。

20

【0062】

特徴点としては、例えば、眉、目、鼻、唇の各端点、顔の輪郭点、頭頂点、顎の下端点等が挙げられる。そして、顔検出部107は、顔領域の位置情報を特定する。例えば、顔検出部107は、画像の横方向をX軸とし、縦方向をY軸として、顔領域に含まれる画素のX座標及びY座標を算出する。さらに、顔検出部107は、上述した特徴点を用いた演算処理によって、検出した講師又は各受講生の表情・年齢などを判別し得る。

【0063】

算出部108は、教室映像を構成する複数のフレームの夫々について、講師T又は各受講生A、B、Cの顔に関する評価値を算出する機能を有する。算出部108において算出される各評価値は、以下に示す所定の評価値(1)~(6)が含まれる。これらの評価値(1)~(6)は、顔情報登録部109に予め登録される。

30

【0064】

(1)笑顔の度合い

算出部108は、顔検出部107が検出した顔のそれぞれについて、例えば、パターンマッチングなどの公知技術を用いて、笑顔の度合いを評価値として算出する。本実施形態では、度合いの一例として、「0：笑顔なし」、「1：微笑」、「2：普通笑い」、「3：大笑い」までの4段階で笑顔の度合いを示す。

40

【0065】

(2)視線の向き

算出部108は、顔検出部107が検出した顔のそれぞれについて、公知の技術を用いて、視線の向きを評価値として算出する。本実施形態では、一例として、「0：視線正面」、「1：視線左右方向」、「2：視線右方向」、「3：視線検出不可」の4種類で視線の向きを示す。

【0066】

(3)顔の向き

算出部108は、顔検出部107が検出した顔のそれぞれについて、公知の技術を用いて、顔の向きを評価値として算出する。本実施形態では、一例として、「0：顔向き正面」

50

、「 1 : 顔向き左方向」、「 2 : 顔向き右方向」、「 3 : 検出不可」の 4 種類で顔の向きを示す。

【 0 0 6 7 】

(4) 顔面積

算出部 1 0 8 は、顔検出部 1 0 7 が検出した顔のそれぞれについて、顔部分の面積を評価値として算出する。

【 0 0 6 8 】

(5) 年齢

算出部 1 0 8 は、顔検出部 1 0 7 が検出した顔のそれぞれについて、公知の技術を用いて、その人物の年齢を評価値として算出する。

【 0 0 6 9 】

(6) 目つぶり度合い

算出部 1 0 8 は、顔検出部 1 0 7 が検出した顔のそれぞれについて、公知の技術を用いて、目つぶり度合いを評価値として算出する。本実施形態では、一例として、「 0 : 目つぶりなし」、「 1 : 一部目つぶりあり」、「 2 : 両目目つぶり」、「 3 : 目つぶり検出不可」の 4 種類で目つぶり度合いを示す。

【 0 0 7 0 】

これらの評価値 (1) ~ (6) は、講義を受講する受講生の感情を判断するための判断基準となり得る。受講生の感情は、例えば、「幸福感」、「退屈感」、「緊張感」の 3 つに分類され得る。

【 0 0 7 1 】

抽出部 1 0 1 は、所定の特定条件が満たされた場合、例えば、算出部 1 0 8 により算出された評価値と、顔情報登録部 1 0 9 に予め登録されている評価値とが一致した場合に、当該一致した評価値に対応するフレーム群を特定のフレーム群として前記教室映像から抽出する。

【 0 0 7 2 】

次に、このように構成された在宅個別指導システムの動作について説明する。図 7 は、本開示の第 2 の実施の形態による教室映像配信方法の処理の流れを説明するフローチャートである。

【 0 0 7 3 】

ここでは、教室を運営する運営者等が、受講生 A である子供を塾などに預ける保護者等からの要望であって、受講生 A の学習態度・学習状況を把握したいという要望に応える場面を例に挙げて説明する。

【 0 0 7 4 】

まず、講義が開始される時刻になると、各受講生 A、B、C は受講生端末 2 0 A、2 0 B、2 0 C をネットワーク NW 経由で在宅個別指導システムに接続して、講師 T の講義開始を待つ。講師 T は、教室映像配信装置 1 0 が備えるモニタ 1 4 を見て各受講生 A、B、C が受講態勢にあるか否かを判断し、受講態勢が整っていれば、講義を開始する。すなわち、撮像部 1 2 並びに撮像部 2 2 による撮像動作が開始されるとともに、收音部 1 3 並びに收音部 2 3 による收音動作が開始される (ステップ S 2 0 0) 。

【 0 0 7 5 】

そして、抽出部 1 0 1 は、撮像部 1 2 で取得される講師映像と、撮像部 2 2 で取得される受講生映像とを適宜取捨選択して合成すると共に、收音部 1 3 又は收音部 2 3 で收音された音を組み合わせる教室映像を生成すると共に、生成した教室映像をネットワーク NW 経由で受講生端末 2 0 A、2 0 B、2 0 C に配信する (ステップ S 2 0 2) 。

【 0 0 7 6 】

次に、顔検出部 1 0 7 は、教室映像について、受講生 A の顔検出を行う (ステップ S 2 0 4) 。顔検出の具体的な手法については、公知技術と同様であるため説明を省略する。

【 0 0 7 7 】

そして、算出部 1 0 8 は、顔検出部 1 0 7 が検出した受講生 A の顔について、公知の技術

10

20

30

40

50

を用いて、顔の向きを評価値として算出する（ステップ S 2 0 6）。続いて、抽出部 1 0 1 は、算出部 1 0 8 により算出された評価値が、顔情報登録部 1 0 9 に予め登録されている評価値「1：顔向き左方向」、「2：顔向き右方向」の何れかと一致した場合に、当該一致した評価値に対応するフレーム群を特定のフレーム群として教室映像から抽出する（ステップ S 2 0 8）。

【0078】

すなわち、受講生 A の顔の向きが正面ではなく左方向又は右方向を向いているような場合には、受講生 A の講義に対する集中度が低下していることが推認され得る。

【0079】

フレーム切り出し部 1 0 5 は、受講生 A の顔が正面を向いていない教室映像の前後において例えば 1 0 秒～2 0 秒程度の時間間隔で連なるフレーム群を切り出す（ステップ S 2 1 0）。

10

【0080】

そして、生成部 1 0 6 は、これらのフレーム群を連結したダイジェスト動画を生成すると共に、当該ダイジェスト動画に基づいて、講師の音声データに対して所定の音響分析を施す（ステップ S 2 1 2）。かかる音響分析によれば、受講生 A の顔が正面を向いていないときに、講師 T が発話しているワードが何であるかのみをフラグで管理できる。これにより、受講生 A の講義に対する集中度が低下した要因となり得るワードを、教室を運営する運営者等が収集できるとともに、受講生 A の学習態度・学習状況を把握したいという保護者等の要望に対しても効率的に応えることができる。

20

【0081】

もちろん、受講生 A の講義に対する集中度の判定は、受講生 A の顔の向きに限らず、受講生 A の視線の向きによっても行い得る。すなわち、受講生 A の視線の向きが正面ではなく左方向又は右方向を向いているような場合には、受講生 A の講義に対する集中度が低下していることが推認され得る。

【0082】

本実施形態に係る在宅個別指導システムには、さらに以下のような使用例が考えられる。

【0083】

具体的に、受講生 A が満足感・幸福感・充実感といったポジティブな感情を抱いたシーンのみを集めて編集した動画を受講生 A の保護者等に向けたダイジェスト動画として生成し得る。

30

【0084】

かかるダイジェスト動画を生成し得る具体的な処理の一例としては、まず、算出部 1 0 8 は、顔検出部 1 0 7 が検出した受講生 A の顔について、公知の技術を用いて、笑顔の度合いを評価値として算出することができる。続いて、抽出部 1 0 1 は、算出部 1 0 8 により算出された度合いが、顔情報登録部 1 0 9 に予め登録されている度合い「2：普通笑い」、「3：大笑い」の何れかと一致した場合に、当該一致した評価値に対応するフレーム群を特定のフレーム群として教室映像から抽出することができる。

【0085】

さらに、フレーム切り出し部 1 0 5 は、受講生 A の笑顔の度合いが「2：普通笑い」、「3：大笑い」の何れかであるときの教室映像の前後において例えば 1 0 秒～2 0 秒程度の時間間隔で連なるフレーム群を切り出すことができる。そして、最後に、生成部 1 0 6 は、これらのフレーム群を連結したダイジェスト動画を、受講生 A の保護者等に向けたダイジェスト動画として生成する。

40

【0086】

< 第 3 の実施形態 >

以下、図 8 及び図 9 に基づいて、第 3 の実施形態に係る在宅個別指導システムについて説明する。この実施形態では、上述した第 1 実施形態で説明した要素と同一の要素について同じ符号を付し、詳細な説明を省略する。

【0087】

50

上記の第1実施形態に係る在宅個別指導システムでは、既に述べたように、予め登録済みの波形データを用いた照合により、教室映像の中から特定のフレーム群を抽出していたが、第3の実施形態に係る在宅個別指導システムでは、人物の動作に関する動作情報に基づいて、教室映像の中から特定のフレーム群を抽出する処理が行われる。

【0088】

<ソフトウェア構成>

図8は、本実施形態に係る教室映像配信装置10のソフトウェア構成例を示す図である。教室映像配信装置10は、抽出部101と、フレーム切り出し部105と、生成部106と、特定部110と、動作情報登録部111と、を備える。

【0089】

特定部110は、例えば、メモリ15に格納されている教室映像を構成する複数のフレームの夫々について、人物の動作に関する動作情報を特定する機能を有する。この動作情報は、例えば、人物の動作を複数の姿勢の連続として捉えた情報であって、様々な姿勢に対応する人体の骨格を形成する各関節の情報を含み得る。動作情報には、例えば、人物の身振り、手振り、ジェスチャ、ボディランゲージの少なくとも何れかが含まれる。

【0090】

特定部110は、例えば、人体パターンを用いたパターンマッチングにより、教室映像を構成する複数のフレームから、人体の骨格を形成する各関節の座標を得る。座標取得の具体的な手法については、公知技術と同様であるため説明を省略する。そして、この座標系で表される各関節の座標が、例えば、1フレーム分の骨格情報となり得る。さらに、複数フレーム分の骨格情報が所定の動作情報となり得る。

【0091】

かかる所定の動作情報は、動作情報登録部111に予め登録されている。すなわち、動作情報登録部111は、様々な姿勢に対応する動作情報を、例えば、公知の人工知能技術を用いた機械学習により予め記憶している。例えば、本実施形態において、受講生が手を振る動きに対応するジェスチャは、講義の内容に納得ができなかったり、引っかかるところがあったりする受講生が講師に対して補充説明を求めるジェスチャパターンとして機械学習済みであるとする。

【0092】

抽出部101は、所定の特定条件が満たされた場合、例えば、特定部110により特定された動作情報と、動作情報登録部111に予め登録されている動作情報とが一致した場合に、当該一致した動作情報に対応するフレーム群を特定のフレーム群として教室映像から抽出する。

【0093】

次に、このように構成された在宅個別指導システムの動作について説明する。図9は、本開示の第3の実施の形態による教室映像配信方法の処理の流れを説明するフローチャートである。

【0094】

ここでは、講義の内容に納得ができない受講生Aが講師に対して補充説明を求める状況を含むシーンを教室映像からピックアップする場面を例に挙げて説明する。

【0095】

まず、講義が開始される時刻になると、各受講生A、B、Cは受講生端末20A、20B、20CをネットワークNW経由で在宅個別指導システムに接続して、講師Tの講義開始を待つ。講師Tは、教室映像配信装置10が備えるモニタ14を見て各受講生A、B、Cが受講態勢にあるか否かを判断し、受講態勢が整っていれば、講義を開始する。すなわち、撮像部12並びに撮像部22による撮像動作が開始されるとともに、收音部13並びに收音部23による收音動作が開始される(ステップS300)。

【0096】

そして、抽出部101は、撮像部12で取得される講師映像と、撮像部22で取得される受講生映像とを適宜取捨選択して合成すると共に、收音部13又は收音部23で收音され

10

20

30

40

50

た音を組み合わせる教室映像を生成すると共に、生成した教室映像をネットワークNW経由で受講生端末20A、20B、20Cに配信する(ステップS302)。

【0097】

次に、特定部110は、人体パターンを用いたパターンマッチングにより、教室映像を構成する複数のフレームから、受講生Aの骨格を形成する各関節の座標を得る。さらに、特定部110は、各関節の座標に基づいて、複数フレーム分の骨格情報を受講生Aのジェスチャとして特定する(ステップS304)。

【0098】

続いて、抽出部101は、特定部110により特定されたジェスチャが、動作情報登録部111において機械学習済みのジェスチャパターン(受講生Aが講師に対して補充説明を 10
求めるジェスチャパターン)と一致した場合に、当該一致したジェスチャに対応するフレーム群を特定のフレーム群として教室映像から抽出する(ステップS306)。

【0099】

フレーム切り出し部105は、受講生A講師に対して補充説明を求める教室映像の前後において例えば10秒~20秒程度の時間間隔で連なるフレーム群を切り出す(ステップS308)。

【0100】

そして、生成部106は、これらのフレーム群を連結したダイジェスト動画を生成する(ステップS310)。

【0101】

かくして、受講生Aが講師に対して補充説明を求めるジェスチャパターンが既に登録済みの状態であるので、今後上記フローと同様の状況があれば、動作情報登録部111に保持されているジェスチャパターンに従った照合により、教室映像の中から特定のフレーム群をピックアップしてくれば、演算処理に伴う負荷を増やさなくとも同様のダイジェスト動画を生成することが可能となる。 20

【0102】

<第4の実施形態>

以下、図10及び図11に基づいて、第4の実施形態に係る在宅個別指導システムについて説明する。この実施形態では、上述した第1実施形態で説明した要素と同一の要素について同じ符号を付し、詳細な説明を省略する。 30

【0103】

上記の第1実施形態に係る在宅個別指導システムでは、既に述べたように、予め登録済みの波形データを用いた照合により、教室映像の中から特定のフレーム群を抽出していたが、第4の実施形態に係る在宅個別指導システムでは、人物の生体情報に基づいて、教室映像の中から特定のフレーム群を抽出する処理が行われる。

【0104】

<ソフトウェア構成>

図10は、本実施形態に係る教室映像配信装置10のソフトウェア構成例を示す図である。教室映像配信装置10は、抽出部101と、フレーム切り出し部105と、生成部106と、生体情報検出部112と、生体情報登録部113と、を備える。 40

【0105】

生体情報検出部112は、例えば、メモリ15に格納されている教室映像を構成する複数のフレームの夫々について、人物の生体情報を検出する機能を有する。人物の生体情報には、人物の血圧、脈拍、脈圧の少なくとも何れかが含まれる。これら所定の生体情報は、各フレームに映り込んだ講師又は受講生の顔領域を一般的な顔検知技術等によって抽出したのちに、血流方向に沿って複数の領域に分割し、各領域における血流を示す色画像の時系列変化に基づいて取得することができる。

【0106】

かかる所定の生体情報は、生体情報登録部113に予め登録されている。すなわち、生体情報登録部113は、例えば緊張の有無等の精神状態、体調の良否等の身体状態の検知に 50

用いる生体情報を、例えば、公知の人工知能技術を用いた機械学習により予め記憶している。例えば、本実施形態において、受講生 A においてミリ秒単位での表情の変化、瞳孔の開き、脈拍の速さ（脈拍数）、顔面の紅潮、発汗具合等、受講生 A が無意識に支配されている情動を読み取り得る生体情報が学習済みであるとする。

【0107】

抽出部 101 は、所定の特定条件が満たされた場合、例えば、生体情報検出部 112 により検出された生体情報と、生体情報登録部 113 に予め登録されている生体情報とが一致した場合に、当該一致した生体情報に対応するフレーム群を特定のフレーム群として教室映像から抽出する。

【0108】

次に、このように構成された在宅個別指導システムの動作について説明する。図 11 は、本開示の第 4 の実施の形態による教室映像配信方法の処理の流れを説明するフローチャートである。

【0109】

ここでは、人間には肉体的安全を保つために遺伝的に備わっているバイアスがあり、見慣れないもの、理解しにくいものに対しては瞬間的に異常を感じるという知見のもとで、受講生 A が緊張状態に陥ったシーンを教室映像からピックアップする場面を例に挙げて説明する。

【0110】

まず、講義が開始される時刻になると、各受講生 A、B、C は受講生端末 20A、20B、20C をネットワーク NW 経由で在宅個別指導システムに接続して、講師 T の講義開始を待つ。講師 T は、教室映像配信装置 10 が備えるモニタ 14 を見て各受講生 A、B、C が受講態勢にあるか否かを判断し、受講態勢が整っていれば、講義を開始する。すなわち、撮像部 12 並びに撮像部 22 による撮像動作が開始されるとともに、收音部 13 並びに收音部 23 による收音動作が開始される（ステップ S400）。

【0111】

そして、抽出部 101 は、撮像部 12 で取得される講師映像と、撮像部 22 で取得される受講生映像とを適宜取捨選択して合成すると共に、收音部 13 又は收音部 23 で收音された音を組み合わせる教室映像を生成すると共に、生成した教室映像をネットワーク NW 経由で受講生端末 20A、20B、20C に配信する（ステップ S402）。

【0112】

次に、生体情報検出部 112 は、公知の技術を用いて、教室映像を構成する複数のフレームの夫々について、受講生 A の脈拍数を検出する（ステップ S404）。

【0113】

続いて、抽出部 101 は、生体情報検出部 112 により検出された脈拍数が、生体情報登録部 113 において機械学習済みの脈拍数（受講生 A が緊張状態にある脈拍数）と一致した場合に、当該一致した脈拍数に対応するフレーム群を特定のフレーム群として教室映像から抽出する（ステップ S406）。

【0114】

フレーム切り出し部 105 は、受講生 A が緊張状態にある教室映像の前後において例えば 10 秒～20 秒程度の時間間隔で連なるフレーム群を切り出す（ステップ S408）。

【0115】

そして、生成部 106 は、これらのフレーム群を連結したダイジェスト動画を生成する（ステップ S410）。

【0116】

かくして、受講生 A が緊張状態にある脈拍数が既に登録済みの状態であるので、今後上記フローと同様の状況があれば、生体情報登録部 113 に保持されている脈拍数に従った照合により、教室映像の中から特定のフレーム群をピックアップしてくれば、演算処理に伴う負荷を増やさなくとも同様のダイジェスト動画を生成することが可能となる。

【0117】

10

20

30

40

50

以上、添付図面を参照しながら本開示の好適な実施形態について詳細に説明したが、本開示の技術的範囲はかかる例に限定されない。本開示の技術分野における通常の知識を有する者であれば、請求の範囲に記載された技術的思想の範疇内において、各種の変更例または修正例に想到し得ることは明らかであり、これらについても、当然に本開示の技術的範囲に属するものと了解される。

【0118】

上述した各実施形態では、遠隔授業を支援する在宅個別指導システムに情報抽出装置を適用する例について述べた。しかし、これに限らない。例えば、会議、講演会等のように、開始時刻および終了時刻が事前に決められており、主として教室、会議室などの特定の空間で行われる各種イベントを支援するシステムに情報抽出装置を適用してもよい。

10

【0119】

本明細書において説明した装置は、単独の装置として実現されてもよく、一部または全部がネットワークで接続された複数の装置（例えばクラウドサーバ）等により実現されてもよい。例えば、教室映像配信装置10のストレージ16又は制御部18は、互いにネットワークで接続された異なるサーバにより実現されてもよい。

【0120】

本明細書において説明した装置による一連の処理は、ソフトウェア、ハードウェア、及びソフトウェアとハードウェアとの組合せのいずれを用いて実現されてもよい。本実施形態に係る教室映像配信装置10の各機能を実現するためのコンピュータプログラムを作製し、PC等を実装することが可能である。また、このようなコンピュータプログラムが格納された、コンピュータで読み取り可能な記録媒体も提供することができる。記録媒体は、例えば、磁気ディスク、光ディスク、光磁気ディスク、フラッシュメモリ等である。また、上記のコンピュータプログラムは、記録媒体を用いずに、例えばネットワークを介して配信されてもよい。

20

【0121】

また、本明細書においてフローチャート図を用いて説明した処理は、必ずしも図示された順序で実行されなくてもよい。いくつかの処理ステップは、並列的に実行されてもよい。また、追加的な処理ステップが採用されてもよく、一部の処理ステップが省略されてもよい。

【0122】

また、本明細書に記載された効果は、あくまで説明的または例示的なものであって限定的ではない。つまり、本開示に係る技術は、上記の効果とともに、または上記の効果に代えて、本明細書の記載から当業者には明らかな他の効果を奏しうる。

30

【符号の説明】

【0123】

- 1 在宅個別指導システム
- 10 教室映像配信装置（情報抽出装置）
- 101 抽出部
- 102 波形登録部
- 103 変換部
- 104 表示部
- 105 フレーム切り出し部
- 106 生成部
- 107 顔検出部
- 108 算出部
- 109 顔情報登録部
- 110 特定部
- 111 動作情報登録部
- 112 生体情報検出部
- 113 生体情報登録部

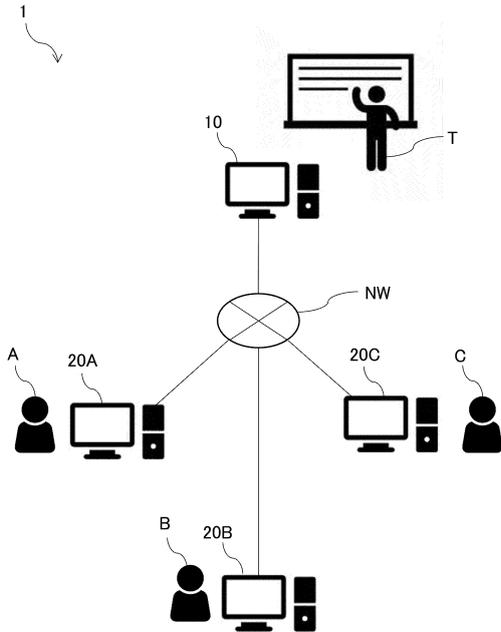
40

50

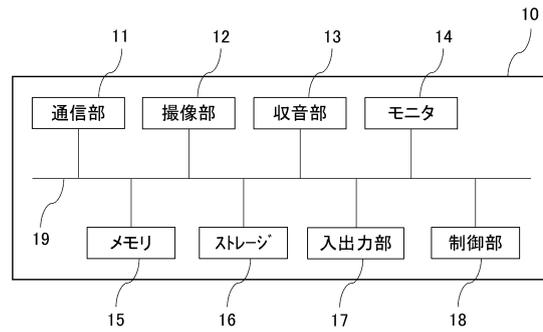
N W ネットワーク

【図面】

【図 1】



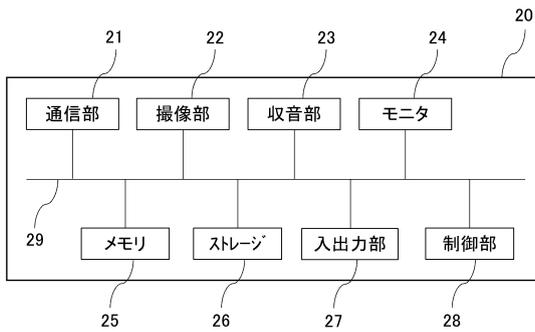
【図 2】



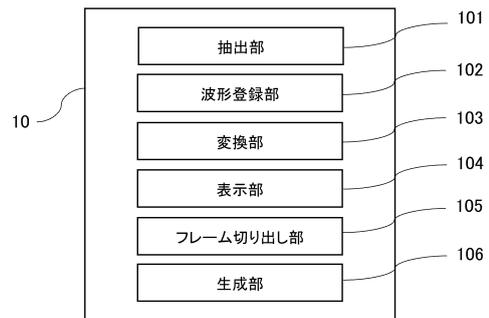
10

20

【図 3】



【図 4】

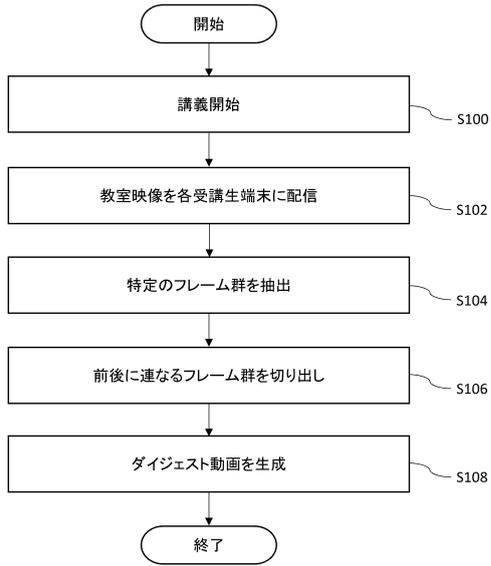


30

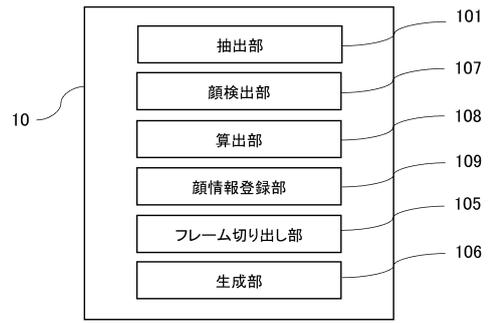
40

50

【図5】

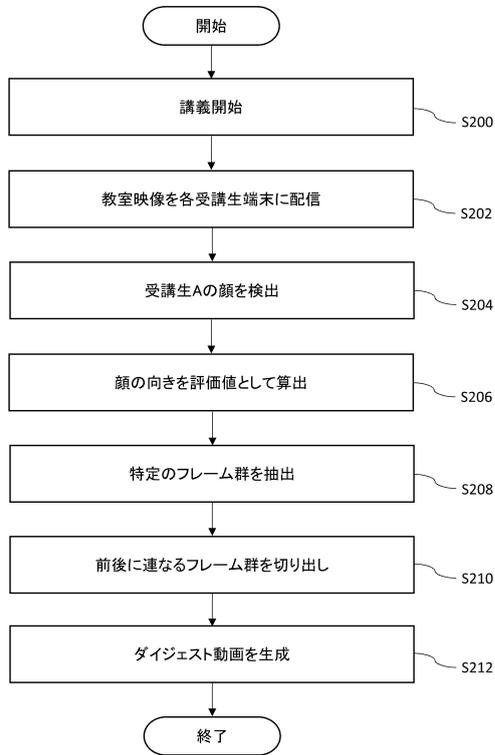


【図6】

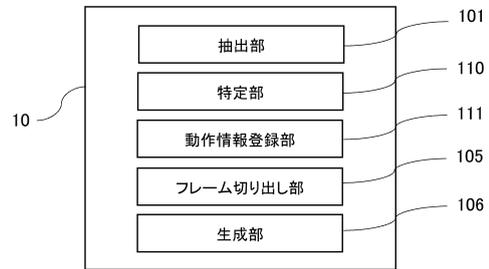


10

【図7】



【図8】



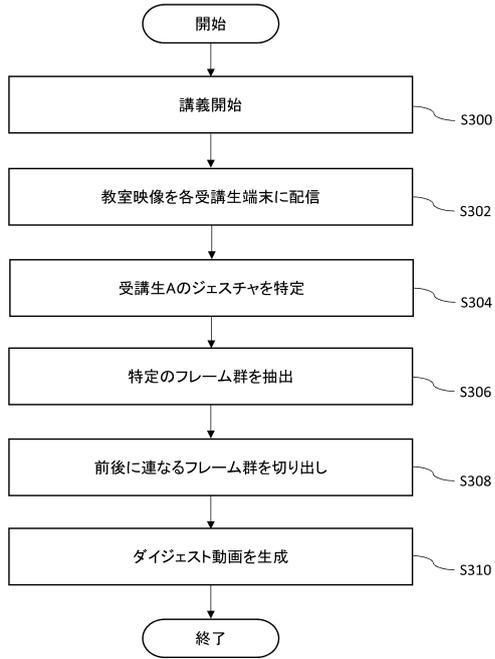
20

30

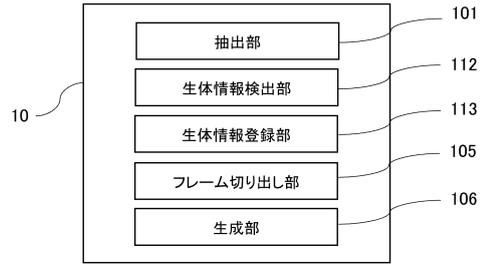
40

50

【 図 9 】



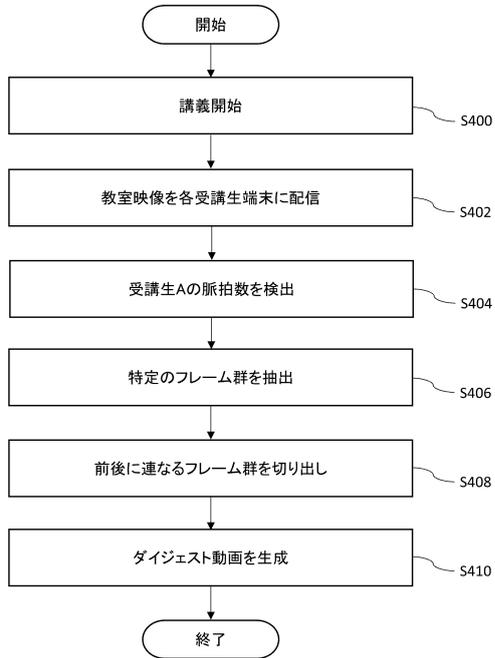
【 図 10 】



10

20

【 図 11 】



30

40

50

【手続補正書】

【提出日】令和4年4月26日(2022.4.26)

【手続補正1】

【補正対象書類名】特許請求の範囲

【補正対象項目名】全文

【補正方法】変更

【補正の内容】

【特許請求の範囲】

【請求項1】

複数のフレームから構成される動画像を取得する取得部と、
 当該動画像内に含まれる所定のデータを特定するための特定条件を記憶する記憶部と、
 当該特定条件に従って、前記動画像から特定のフレーム群を複数抽出する抽出部と、
抽出された前記特定のフレーム群に含まれる顔画像又は音声を所定のフレーム単位ごとに
識別する識別手段と、
識別した前記顔画像又は前記音声から評価値を算出する評価手段とを備える、
 情報抽出装置。

【請求項2】

請求項1に記載の情報抽出装置であって、
 所定の波形データを予め登録する波形登録部、を更に備え、
 前記特定条件は、前記動画像内に含まれる音の波形データと前記登録されている波形デー
 タとが一致するか否かであって、
 前記抽出部は、両波形データが一致した場合に、当該一致した波形に対応するフレーム群
 を前記特定のフレーム群として前記動画像から抽出する、
 情報抽出装置。

【請求項3】

請求項2に記載の情報抽出装置であって、
 動画内に含まれる前記音を音声認識によりテキスト情報に変換する変換部を更に備え、
 前記変換部は、前記特定のフレーム群とその前後所定フレーム数とを含む補助フレーム群
 に対応する前記音を変換する、
 情報抽出装置。

【請求項4】

請求項2又は請求項3に記載の情報抽出装置であって、
 前記被写体を含む周囲の音が示す情報には、会話情報と非会話情報とが混在する、
 情報抽出装置。

【請求項5】

請求項4に記載の情報抽出装置であって、
 前記会話情報には、ポジティブな感情を示すワードと、ネガティブな感情を示すワードの
 少なくとも何れかが含まれる、
 情報抽出装置。

【請求項6】

請求項4又は請求項5に記載の情報抽出装置であって、
 前記非会話情報には、舌打ち、溜め息、相槌の少なくとも何れかを示す情報が含まれる、
 情報抽出装置。

【請求項7】

請求項1に記載の情報抽出装置であって、
 顔の表情に関する所定の顔評価値を予め登録する顔情報登録部、を更に備え、
 前記特定条件は、前記動画像内に含まれる顔の表情から算出される顔評価値と前記登録さ
 れている顔評価値とが一致するか否かであって、
 前記抽出部は、両顔評価値が一致した場合に、当該一致した顔評価値に対応するフレーム
 群を前記特定のフレーム群として前記動画像から抽出する、

情報抽出装置。

【請求項 8】

請求項 7 に記載の情報抽出装置であって、

前記顔評価値には、前記人物の幸福感、退屈感又は緊張感の度合いを評価した評価値が含まれる、

情報抽出装置。

【請求項 9】

請求項 7 又は請求項 8 に記載の情報抽出装置であって、

前記顔評価値には、前記人物の表情、前記人物の視線の向き、前記人物の顔の向きを評価した評価値が含まれる、

10

情報抽出装置

【請求項 10】

請求項 1 に記載の情報抽出装置であって、

人物の動作に関する所定の動作評価値を予め登録する動作情報登録部、を更に備え、

前記特定条件は、前記動画像内に含まれる人物から算出される動作評価値と前記登録されている動作評価値とが一致するか否かであって、

前記抽出部は、両動作評価値が一致した場合に、当該一致した動作評価値に対応するフレーム群を前記特定のフレーム群として前記動画像から抽出する、

情報抽出装置。

【請求項 11】

20

請求項 10 に記載の情報抽出装置であって、

前記動作評価値には、前記人物の身振り、手振り、ジェスチャ、ボディランゲージの少なくとも何れかの動作を評価した評価値が含まれる、

情報抽出装置。

【請求項 12】

請求項 1 に記載の情報抽出装置であって、

所定の生体情報に関する生体評価値を予め登録する生体情報登録部と、を備え、

前記特定条件は、前記動画像内に含まれる人物から算出可能な生体評価値と、前記登録されている生体評価値とが一致するか否かであって、

前記抽出部は、両生体評価値が一致した場合に、当該一致した生体評価値に対応するフレーム群を前記特定のフレーム群として前記動画像から抽出する、

30

情報抽出装置。

【請求項 13】

請求項 12 に記載の情報抽出装置であって、

前記生体評価値には、前記人物の血圧、脈拍、脈圧の少なくとも何れかが含まれる、

情報抽出装置。

【請求項 14】

請求項 1 乃至請求項 13 の何れか一項に記載の情報抽出装置であって、

前記特定のフレーム群に対して、当該特定のフレーム群と時系列的に前後に連続する追加フレームを前記評価値の算出対象に追加するフレーム追加部を備えている、

40

情報抽出装置。

【請求項 15】

請求項 1 に記載の情報抽出装置であって、

前記評価値の時系列によるグラフ情報を提供する、

情報抽出装置。

【請求項 16】

請求項 1 に記載の情報抽出装置であって、

前記顔画像を複数の異なる観点によって評価した複数の評価値を算出する、

情報抽出装置。

【請求項 17】

50

請求項 1 に記載の情報抽出装置であって、
前記動画像に含まれる音声と共に前記評価値を算出する、
情報抽出装置。

【請求項 18】

請求項 1 乃至請求項 17 の何れかに記載の情報抽出装置であって、
抽出された前記特定のフレーム群同士を連結してダイジェスト情報を生成する手段を更に
備え、
前記識別手段は、前記ダイジェスト情報に含まれる少なくとも顔画像又は音声を所定のフ
レーム単位ごとに識別し、
前記評価手段は、識別した前記顔画像に関する前記評価値を算出する、
情報抽出装置。

10

20

30

40

50