

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6411274号
(P6411274)

(45) 発行日 平成30年10月24日(2018.10.24)

(24) 登録日 平成30年10月5日(2018.10.5)

(51) Int.Cl.		F I			
G09G	5/377	(2006.01)	G09G	5/36	520M
G09G	5/00	(2006.01)	G09G	5/00	510H
H04N	21/431	(2011.01)	G09G	5/00	550B
H04N	7/173	(2011.01)	H04N	21/431	
			H04N	7/173	610Z

請求項の数 11 (全 14 頁)

(21) 出願番号 特願2015-80590 (P2015-80590)
 (22) 出願日 平成27年4月10日(2015.4.10)
 (65) 公開番号 特開2016-200711 (P2016-200711A)
 (43) 公開日 平成28年12月1日(2016.12.1)
 審査請求日 平成29年4月25日(2017.4.25)

(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
 (74) 代理人 100121706
 弁理士 中尾 直樹
 (74) 代理人 100128705
 弁理士 中村 幸雄
 (74) 代理人 100147773
 弁理士 義村 宗洋
 (72) 発明者 鎌本 優
 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 白木 善史
 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

最終頁に続く

(54) 【発明の名称】 タイミング補正システム、その方法及びプログラム

(57) 【特許請求の範囲】

【請求項1】

所定の行為とその所定の行為を意味する視覚情報とが対応付けて記憶される記憶部と、
対象映像信号を見るものによって入力され、その対象映像信号に重畳して表示される視
覚情報が、前記記憶部に記憶される視覚情報と一致するか否かを判定し、一致する場合に
は、前記対象映像信号に重畳して表示される視覚情報、前記対象映像信号に重畳して表示
される視覚情報に対応する所定の行為、及び前記対象映像信号に重畳して表示される視覚
情報の入力時刻を含むメタデータを検出するキー入力検出部と、

(i)前記対象映像信号、(ii)前記対象映像信号に対応する対象音響信号、並びに、(iii)
 前記対象映像信号に重畳して表示されている既に入力済みの他の視覚情報及び前記既に入
 力済みの他の視覚情報のメタデータの少なくとも何れかから、前記キー入力検出部で検出
 された視覚情報の意味する行為の基準となるタイミングである基準タイミングを検出する
 タイミング検出部と、

前記キー入力検出部で検出された視覚情報を表示部に表示するためのタイミングを、前
 記基準タイミングに基づき補正する補正部とを含み、

前記記憶部に記憶される所定の行為を意味する視覚情報は前記補正部で補正された前記
基準タイミングで前記対象映像信号に重畳される、

タイミング補正システム。

【請求項2】

請求項1のタイミング補正システムであって、

前記タイミング検出部は、

(1)(i)前記対象映像信号に一定のテンポがある場合の当該テンポ、

(2)(ii)前記対象音響信号に一定のテンポがある場合の当該テンポ、

(3)(i)前記対象映像信号から抽出された特徴量と、予め所定の行為を撮影した映像信号から抽出した特徴量との類似度が閾値以上である場合の、所定の時間区間毎の代表値、

(4)(ii)前記対象音響信号から抽出された特徴量と、予め所定の行為を撮影した映像信号に対応する音響信号から抽出した特徴量との類似度が閾値以上である場合の、所定の時間区間毎の代表値、

の少なくとも何れかを前記基準タイミングとして検出する、

タイミング補正システム。

10

【請求項3】

請求項1のタイミング補正システムであって、

前記タイミング検出部は、少なくとも複数の(iii)前記対象映像信号に重畳して表示されている既に入力済みの他の視覚情報及び前記既に入力済みの他の視覚情報のメタデータの統計量に基づいて基準タイミングを求めるものであり、

既に入力済みの他の視覚情報のうち、前記キー入力検出部で検出したメタデータに含まれる所定の行為と一致する行為を意味する前記既に入力済みの他の視覚情報それぞれの表示時刻を抽出し、所定の時間区間毎に、抽出した表示時刻の前記所定の時間区間毎の最頻値を前記基準タイミングとして検出することを特徴とする、

タイミング補正システム。

20

【請求項4】

請求項1のタイミング補正システムであって、

前記タイミング検出部は、少なくとも複数の(iii)前記対象映像信号に重畳して表示されている既に入力済みの他の視覚情報及び前記既に入力済みの他の視覚情報のメタデータの統計量に基づいて基準タイミングを求めるものであり、

既に入力済みの他の視覚情報のうち、前記キー入力検出部で検出したメタデータに含まれる所定の行為と一致する行為を意味する前記既に入力済みの他の視覚情報それぞれの表示時刻を抽出し、抽出した表示時刻を用いて、所定の時間区間毎に、表示時刻の平均値と分散とを求め、前記求めた平均値と分散とを持つガウス分布に従う乱数を基準タイミングとする、

タイミング補正システム。

30

【請求項5】

請求項1から請求項3の何れかのタイミング補正システムであって、

前記タイミング検出部は、所定の行為の種類ごとに生じる分散の値を予め求めておき、前記基準タイミングを中心として、求めておいた分散のうちの、前記キー入力検出部で検出したメタデータに含まれる所定の行為に対応する分散のガウス分布に従う乱数を新たな基準タイミングとする、

タイミング補正システム。

【請求項6】

記憶部には、所定の行為とその所定の行為を意味する視覚情報とが対応付けて記憶されているものとし、

40

キー入力検出部が、対象映像信号を見るものによって入力され、その対象映像信号に重畳して表示される視覚情報が、前記記憶部に記憶される視覚情報と一致するか否かを判定し、一致する場合には、前記対象映像信号に重畳して表示される視覚情報、前記対象映像信号に重畳して表示される視覚情報に対応する所定の行為、及び前記対象映像信号に重畳して表示される視覚情報の入力時刻を含むメタデータを検出するキー入力検出ステップと

、
タイミング検出部が、(i)前記対象映像信号、(ii)前記対象映像信号に対応する対象音響信号、並びに、(iii)前記対象映像信号に重畳して表示されている既に入力済みの他の視覚情報及び前記既に入力済みの他の視覚情報のメタデータの少なくとも何れかから、前

50

記キー入力検出ステップで検出された視覚情報の意味する行為の基準となるタイミングである基準タイミングを検出するタイミング検出ステップと、

補正部が、前記キー入力検出ステップで検出された視覚情報を表示部に表示するためのタイミングを、前記基準タイミングに基づき補正する補正ステップとを含み、

前記記憶部に記憶される所定の行為を意味する視覚情報は前記補正ステップで補正された前記基準タイミングで前記対象映像信号に重畳される、

タイミング補正方法。

【請求項 7】

請求項 6 のタイミング補正方法であって、

前記タイミング検出ステップにおいて、

(1)(i)前記対象映像信号に一定のテンポがある場合の当該テンポ、

(2)(ii)前記対象音響信号に一定のテンポがある場合の当該テンポ、

(3)(i)前記対象映像信号から抽出された特徴量と、予め所定の行為を撮影した映像信号から抽出した特徴量との類似度が閾値以上である場合の、所定の時間区間毎の代表値、

(4)(ii)前記対象音響信号から抽出された特徴量と、予め所定の行為を撮影した映像信号に対応する音響信号から抽出した特徴量との類似度が閾値以上である場合の、所定の時間区間毎の代表値、

の少なくとも何れかを前記基準タイミングとして検出する、

タイミング補正方法。

【請求項 8】

請求項 6 のタイミング補正方法であって、

前記タイミング検出ステップは、少なくとも複数の(iii)前記対象映像信号に重畳して表示されている既に入力済みの他の視覚情報及び前記既に入力済みの他の視覚情報のメタデータの統計量に基づいて基準タイミングを求めるものであり、

既に入力済みの他の視覚情報のうち、前記キー入力検出ステップで検出したメタデータに含まれる所定の行為と一致する行為を意味する前記既に入力済みの他の視覚情報それぞれの表示時刻を抽出し、所定の時間区間毎に、抽出した表示時刻の前記所定の時間区間毎の最頻値を前記基準タイミングとして検出することを特徴とする、

タイミング補正方法。

【請求項 9】

請求項 6 のタイミング補正方法であって、

前記タイミング検出ステップは、少なくとも複数の(iii)前記対象映像信号に重畳して表示されている既に入力済みの他の視覚情報及び前記既に入力済みの他の視覚情報のメタデータの統計量に基づいて基準タイミングを求めるものであり、

既に入力済みの他の視覚情報のうち、前記キー入力検出部で検出したメタデータに含まれる所定の行為と一致する行為を意味する前記既に入力済みの他の視覚情報それぞれの表示時刻を抽出し、抽出した表示時刻を用いて、所定の時間区間毎に、表示時刻の平均値と分散とを求め、前記求めた平均値と分散とを持つガウス分布に従う乱数を基準タイミングとする、

タイミング補正方法。

【請求項 10】

請求項 6 から請求項 8 の何れかのタイミング補正方法であって、

前記タイミング検出ステップは、所定の行為の種類ごとに生じる分散の値を予め求めておき、前記基準タイミングを中心として、求めておいた分散のうちの、前記キー入力検出ステップで検出したメタデータに含まれる所定の行為に対応する分散のガウス分布に従う乱数を新たな基準タイミングとする、

タイミング補正方法。

【請求項 11】

請求項 1 から請求項 5 の何れかのタイミング補正システムとしてコンピュータを機能させるためのプログラム。

10

20

30

40

50

【発明の詳細な説明】**【技術分野】****【0001】**

本発明は、映像を見るものによって入力されるテキスト情報を、その映像に重畳して表示する技術に関する。

【背景技術】**【0002】**

映像を見るものによって入力されるテキスト情報を、その映像に重畳して表示する技術の従来技術として非特許文献1が知られている。非特許文献1では、視聴者は、動画を視聴しながら、コメントを投稿することができる。

10

【先行技術文献】**【非特許文献】****【0003】**

【非特許文献1】「動画の視聴 コメントの投稿」、[online]、NIWANGO.INC、[平成27年2月2日検索]、インターネット<URL : <http://info.nicovideo.jp/help/player/howto/>>

【発明の概要】**【発明が解決しようとする課題】****【0004】**

しかしながら、従来技術では、動画に対してコメントしたいと思ってから、コメントを入力し、コメント投稿ボタンをクリックまたはエンターキーを押下する必要があるため、視聴者がコメントしたいと思ったタイミングから遅れてコメントが表示される場合がある。逆に動画の内容を予め知っている場合には、予めコメントを入力しておき、コメント投稿ボタンをクリックまたはエンターキーを押下するタイミングを視聴者が図ることもできるが、その場合であっても、視聴者がコメントしたいと思ったタイミングよりも早くなったり、または、遅くなったりする場合がある。例えば、ミュージックビデオやライブ映像の楽曲のテンポに合わせて、拍手を意味するテキスト情報「8」をコメントする場合、実際に拍手する場合よりも、ズレてしまう場合が多い、または、ズレ幅が大きくなりやすい。

20

【0005】

本発明は、コメントを映像に重畳して表示する際のタイミングを、映像に合わせて補正するタイミング補正システム、その方法及びプログラムを提供することを目的とする。

30

【課題を解決するための手段】**【0006】**

上記の課題を解決するために、本発明の一態様によれば、タイミング補正システムは、対象映像信号を見るものによって入力され、その対象映像信号に重畳して表示され、所定の行為を意味する視覚情報及びその視覚情報の入力時刻を含むメタデータを検出するキー入力検出部と、(i)対象映像信号、(ii)対象映像信号に対応する対象音響信号、並びに、(iii)対象映像信号に重畳して表示されている既に入力済みの他の視覚情報及びそのメタデータの少なくとも何れかから、キー入力検出部で検出された視覚情報の意味する行為の基準となるタイミングである基準タイミングを検出するタイミング検出部と、キー入力検出部で検出された視覚情報を表示部に表示するためのタイミングを、基準タイミングに基づき補正する補正部とを含む。

40

【0007】

上記の課題を解決するために、本発明の他の態様によれば、タイミング補正方法は、キー入力検出部が、対象映像信号を見るものによって入力され、その対象映像信号に重畳して表示され、所定の行為を意味する視覚情報及びその視覚情報の入力時刻を含むメタデータを検出するキー入力検出ステップと、タイミング検出部が、(i)対象映像信号、(ii)対象映像信号に対応する対象音響信号、並びに、(iii)対象映像信号に重畳して表示されている既に入力済みの他の視覚情報及びそのメタデータの少なくとも何れかから、キー入力検出ステップで検出された視覚情報の意味する行為の基準となるタイミングである基準タ

50

イミングを検出するタイミング検出ステップと、補正部が、キー入力検出ステップで検出された視覚情報を表示部に表示するためのタイミングを、基準タイミングに基づき補正する補正ステップとを含む。

【発明の効果】

【0008】

本発明によれば、コメントを映像に重畳して表示する際のタイミングを、映像に合わせて補正することができるという効果を奏する。

【図面の簡単な説明】

【0009】

【図1】第一実施形態に係るタイミング補正システムの機能ブロック図。

10

【図2】第一実施形態に係るタイミング補正システムの処理フローの例を示す図。

【図3】キー入力検出部の備える記憶部に記憶されるデータの例を示す図。

【図4】補正部の補正例を示す図。

【図5】テキスト情報 $txt_{in}(r)$ が入力時刻 $t_{in}(r)$ に表示部の右端から表示され、左端に向かって、移動し、左端から消えていく場合の例を示す図。

【図6】テキスト情報 $txt_{in}(q)$ がタイミング $t_{out}(q)$ に表示部の右端から表示され、左端に向かって、移動し、左端から消えていく場合の例を示す図。

【発明を実施するための形態】

【0010】

以下、本発明の実施形態について、説明する。なお、以下の説明に用いる図面では、同じ機能を持つ構成部や同じ処理を行うステップには同一の符号を記し、重複説明を省略する。

20

【0011】

<第一実施形態に係る動画配信システム1>

図1は第一実施形態に係るタイミング補正システム100の機能ブロック図を、図2はその処理フローを示す。

【0012】

動画配信システム1は、1台以上の視聴者端末91と、タイミング補正システム100と、対象映像信号を視聴者端末91に配信する動画配信サーバ92とを含む。各視聴者端末91、タイミング補正システム100及び動画配信サーバ92は、通信回線を介して通信可能とされている。なお、動画配信システム1に含まれる各視聴者端末91、タイミング補正システム100、動画配信サーバ92は、NTP(Network Time Protocol)等により、時刻を同期しておくことが望ましい。

30

【0013】

<視聴者端末91>

視聴者端末91は、対象映像信号(例えば、動画)を見るもの(例えば、動画の視聴者)によって操作され、入力部(キーボード、マウス、タッチパネル等)と、表示部(ディスプレイ、タッチパネル等)とを含み、例えば、パーソナルコンピュータ、スマートホン、タブレット等からなる。視聴者は、視聴者端末91の入力部を介して、動画配信サーバ92に対して対象映像の再生を要求することができる。また、視聴者端末91の表示部を介して、対象映像信号を視聴することができる。さらに、視聴者は、入力部を介して、対象映像信号に重畳して表示されるテキスト情報(例えば、コメント)を入力することができる。視聴者端末91は、テキスト情報が入力されると、例えばシステム時刻に基づき入力時刻を取得して、メタデータとしてテキスト情報と一緒に送信する。

40

【0014】

<動画配信サーバ92>

動画配信サーバ92は、動画データベース及びビデオカメラから動画を受け取り、視聴者端末91の要求に応じて、動画データベース内に格納されている動画、または、ビデオカメラで収録した動画をリアルタイムで配信する。また、ビデオカメラで収録された動画に限らず、リアルタイムで合成・編集されたCGやモーションキャプチャ等から合成され

50

たCGをリアルタイム配信することもある。なお、本実施形態において、動画とは、時間軸に同期させた音響信号と共に提供される映像信号を意味する。動画データベースには、動画と共に動画に付加されたテキスト情報が記憶される。さらに、テキスト情報にはメタデータが付加されている。メタデータとしては、テキスト情報の入力時刻、テキスト情報の大きさ、その色、その出現方法、その移動速度や、移動位置等がある。例えば、大きさ、色、出現方法、移動速度、移動位置等は、テキスト情報の入力者が選択できるものとしてもよく、視聴者端末91がメタデータとしてテキスト情報と一緒に送信し、動画データベースに動画と共に記憶される。

【0015】

< タイミング補正システム100 >

タイミング補正システム100は、テキスト情報 $txt_{in}(p)$ とそのそのテキスト情報の入力時刻 $t_{in}(p)$ を示すメタデータ及びテキスト情報付の対象映像信号が入力され、テキスト情報 $txt_{in}(p)$ とそのテキスト情報を表示部(例えばディスプレイ)に表示するためのタイミング $t_{in}(p)$ または $t_{out}(p)$ を含むメタデータとを出力する。なお、 p は入力されるテキスト情報全てに付与されるインデックスを示す。なお、テキスト情報 $txt_{in}(p)$ は、対象映像信号を見るものによって入力され、その対象映像信号に重畳して表示されるものある。

【0016】

タイミング補正システム100は、キー入力検出部110と、タイミング検出部120と、補正部130とを含む。

【0017】

< キー入力検出部110 >

キー入力検出部110は、テキスト情報 $txt_{in}(p)$ とそのテキスト情報の入力時刻 $t_{in}(p)$ とを受け取り、所定の行為を意味するテキスト情報 $txt_{in}(q)$ 及びそのテキスト情報の入力時刻 $t_{in}(q)$ を含むメタデータを検出し(S110)、出力する。なお、 q は所定の行為を意味するテキスト情報全てに付与されるインデックスを示す。入力される全てのテキスト情報の個数を P とし、入力されるテキスト情報の内、所定の行為を意味するテキスト情報の個数を Q とすると、 $P \geq Q$ であり、 $p=1,2,\dots,P$ 、 $q=1,2,\dots,Q$ である。

【0018】

例えば、キー入力検出部110は図示しない記憶部を備え、記憶部には所定の行為とその所定の行為を意味するテキスト情報とが対応付けて記憶される(図3参照)。キー入力検出部110は、受け取ったテキスト情報 $txt_{in}(p)$ と記憶部に記憶されるテキスト情報とが一致するか否かを判定し、一致する場合には、対応するテキスト情報 $txt_{in}(q)$ と所定の行為(または、所定の行為を示すインデックス)と入力時刻 $t_{in}(q)$ とを、補正部130に出力する。一方、一致しない場合には、対応するテキスト情報 $txt_{in}(r)$ と入力時刻 $t_{in}(r)$ とを、動画配信サーバ92に出力する。なお、 r は所定の行為を意味するテキスト情報以外の全てのテキスト情報に付与されるインデックスを示す。入力されるテキスト情報の内、所定の行為を意味するテキスト情報以外のテキストの個数を R とすると、 $P=Q+R$ であり、 $r=1,2,\dots,R$ である。

【0019】

< タイミング検出部120 >

タイミング検出部120は、動画配信サーバ92から(i)対象映像信号、(ii)対象映像信号に対応する対象音響信号、並びに、(iii)対象映像信号に重畳して表示される既に入力済みの他のテキスト情報及びそのメタデータの少なくとも何れかを受け取る。例えば、テキスト情報付きの動画には(i)~(iii)の全てが含まれる。

【0020】

タイミング検出部120は、(i)~(iii)の少なくとも何れかから、キー入力検出部110で検出されたテキスト情報の意味する行為の基準となるタイミングである基準タイミング $t_{out}(s)$ を検出し(S120)、出力する。なお、 s は(i)~(iii)の少なくとも何れかに含まれる所定の行為に出現番号を表すインデックスを示し、 $s=1,2,\dots,S$ であり、 S は(i)~(iii)の少なくとも何れかに含まれる所定の行為の出現回数である。

10

20

30

40

50

【 0 0 2 1 】

基準タイミング $t_{out}(s)$ を検出する方法としては、テキスト情報の意味する行為に応じて様々な方法が考えられる。

【 0 0 2 2 】

例えば、テキスト情報の意味する行為が「拍手」の場合に基準タイミングを検出する方法を、用いるデータに応じて例示する。テキスト情報の意味する行為が「拍手」の場合、さらに、テキスト情報の意味する行為として、一定のテンポに合わせて行う「手拍子(clap)」と、一定のテンポを持たず行う「拍手(applause)」とが考えられる。なお、以下、単に「拍手」といった場合、一定のテンポを持たず行う「拍手(applause)」を意味するものとする。「手拍子」と「拍手」とは、手を叩く時間的間隔や音量的差異が異なるため(参考文

10

献 1)、例えば、(ii)対象映像信号に対応する対象音響信号等に基づいて、テキスト情報が何れの行為を意味するのかを判別することができる。
(参考文献 1) 鎌本優, 河原一彦, 尾本章, 守谷健弘, 「音楽鑑賞時に励起される拍手音・手拍子音の低遅延伝送に向けた基礎的検討」、日本音響学会 2014年秋季研究発表会, 1-Q-17, 2014年。

この実施形態では、テキスト情報の意味する行為が「手拍子」の場合の、基準タイミングの検出方法について説明する。

【 0 0 2 3 】

(i) 行為が「手拍手」であり、データが対象映像信号の場合

対象映像信号に一定のテンポがある場合には、そのテンポに合わせて拍手すると仮定し、テンポを基準タイミングとして検出する(参考文献 2 参照)。

20

(参考文献 2) 三上弾、松本鮎美、門田浩二、川村春美、小島明、「動作学習のための遅延同期ビデオフィードバックシステム」、情報処理学会、情報処理学会論文誌 コンシューマ・デバイス & システム、2014年、vol.4、No.1、pp22-31。

(ii) 行為が「手拍手」であり、データが対象音響信号の場合

対象音響信号に一定のテンポがある場合には、そのテンポに合わせて拍手すると仮定し、テンポを基準タイミングとして検出する(参考文献 2 ~ 参考文献 5 参照)。

(参考文献 3) 角尾衣未留, 宮本賢一, 小野順貴, 嵯峨山茂樹, 「調波音・打楽器音分離手法を用いた音楽音響信号からのリズム特徴量の抽出」, 日本音響学会春季研究発表会講演集, Mar. 2008, pp. 905-906。

30

(参考文献 4) 角尾衣未留, 小野順貴, 嵯峨山茂樹, 「リズムマップ: 音楽音響信号からの単位リズムパターンの抽出と楽曲構造の解析」, 情報処理学会研究報告, Aug. 2008, vol. 2008-MUS-76, no. 25, pp. 149-154。

(参考文献 5) 角尾衣未留, 小野順貴, 嵯峨山茂樹, 「和声境界を考慮した単位リズムパターンの抽出に基づく音楽音響信号の小節境界推定」, 日本音響学会秋季研究発表会講演集, Sep. 2009, no. 3-5-10, pp. 897-898。

(iii) 行為が「手拍手」であり、データが対象映像信号に重畳して表示される既に入力済みの他のテキスト情報及びそのメタデータの場合

例えば、タイミング検出部 120 は、入力済みの他のテキスト情報の中から、キー入力検出部 110 で検出されたテキスト情報の意味する行為と同じ行為を意味する他のテキスト情報を抽出する。例えば、図示しない記憶部(図 3 参照)を参照して、タイミング検出部 120 は、受け取った入力済みの他のテキスト情報と、記憶部に記憶され、キー入力検出部 110 で検出されたテキスト情報の意味する行為に対応するテキスト情報とが一致するか否かを判定し、一致する場合には、その表示時刻とを、抽出する。

40

【 0 0 2 4 】

タイミング検出部 120 は、抽出した表示時刻の統計量に基づいて基準タイミングを求める。例えば、抽出した表示時刻を用いて、所定の時間区間(例えば、動画がミュージックビデオであり、曲のテンポが148BPM(Beats per Minutes)の場合、一拍の間隔は405ms程度なので、所定の時間区間を405msとする)毎に、時間区間毎の代表値(平均値、最頻値、最小値及び最大値等の複数の表示時刻を代表する何らかの値)を求め、基準タイミングと

50

して検出する。例えば、抽出した表示時刻を用いて、ヒストグラムを作成し、多数決により基準タイミングを求める。つまり、最頻値を基準タイミングとする。

【0025】

(タイミングに揺らぎを与える方法)

拍手のタイミングにゆらぎを与えてもよい(参考文献1参照)。

【0026】

例えば、(iii)の場合、抽出した表示時刻を用いて、所定の時間区間毎に、表示時刻の平均値と分散とを求め、その平均値と分散とを持つガウス分布に従う乱数を基準タイミングとしてもよい。この方法により、拍手のタイミングにゆらぎを与えることができ、より自然なタイミングで所定の行動に対応するテキスト情報を表示することができる。

10

【0027】

なお、同様の方法により、分散を求め、(i),(ii)の方法と組合せてもよい。例えば、(i),(ii)の方法で基準タイミングを求め、その基準タイミングを中心として、求めた分散を持つガウス分布に従う乱数を新たな(最終的に用いる)基準タイミングとする。このような方法により、より自然なタイミングで所定の行動に対応するテキスト情報を表示することができる。

【0028】

また、予め手拍子を行う際に一般的に生じる分散の値を求めておき、その分散に基づき、基準タイミングを求めてもよい。例えば、(i),(ii),(iii)の方法で基準タイミングを求め、その基準タイミングを中心として、手拍子を行う際に一般的に生じる分散を持つガウス分布に従う乱数を新たな(最終的に用いる)基準タイミングとする。

20

【0029】

(行為が「手拍子」以外の場合について)

なお、データが(iii)対象映像信号に重畳して表示される既に入力済みの他のテキスト情報及びそのメタデータの場合には、行為が「手拍子」以外の行為であっても容易に適用することができる。例えば、行為が「拍手」または「笑い」の場合、所定の時間区間を一連の行為「拍手」または「笑い」が、継続しうる最大の時間に設定する。例えば、何らかの事象に対して、「拍手」を送るのは、長くとも30秒程度であろうと想定される場合、最初に、「拍手」を意味する他のテキスト情報が表示されてから1分以内に表示される「拍手」を意味する他のテキスト情報から代表値を求め、基準タイミングとして検出する。前述の方法により「拍手」や「笑い」のタイミングにゆらぎを与えてもよい。

30

【0030】

ただし、初めて対象映像信号を配信する場合(対象映像信号にテキスト情報が付加されていない場合)、または、ビデオカメラで収録した動画をリアルタイムで配信する場合は、対象映像信号に既に入力済みの他のテキスト情報がないので、(iii)を用いることはできない。

【0031】

(i)対象映像信号、または、(ii)対象音響信号から「手拍子」以外の行為の基準タイミングを検出する方法としては、例えば、予め所定の行為(例えば「拍手」「笑い」)を撮影した映像信号または音響信号から特徴量を抽出し、図示しない記憶部に記憶しておく。さらに、動画配信サーバ92から受け取った(i)対象映像信号、または、(ii)対象音響信号から特徴量を取り出し、記憶部に記憶されている特徴量との類似度を求め、類似度が閾値以上となる場合に、所定の行為が行われていると判断し、代表値を求め、基準タイミングとして検出する。例えば、既存の顔認証技術を応用し、(i)対象映像信号から笑顔を検出し、検出した時刻から「笑い」の基準タイミングを求めてもよい。また、例えば、動画圧縮用の特徴量である動き補償ベクトル(参考文献6参照)を利用してもよい。

40

(参考文献6)村上篤道、浅井光太郎、関口俊一、「高効率映像符号化技術 HEVC/H.265とその応用」、オーム社、2013年、p.20-28,125-132

予め「拍手」を撮影した映像信号を圧縮符号化して動き補償ベクトルの時系列を取得し、図示しない記憶部に記憶しておく。さらに、動画配信サーバ92から受け取った(i)対象

50

映像信号を圧縮符号化して動き補償ベクトルの時系列を取得し(そもそも(i)対象映像信号が圧縮符号化されているものであれば動き補償ベクトルが含まれるため、それをそのまま利用すればよい)、記憶部に記憶されている動き補償ベクトルの時系列との類似度を求めればよい。

【0032】

<補正部130>

補正部130は、所定の行為を意味するテキスト情報 $txt_{in}(q)$ 及びそのテキスト情報の入力時刻 $t_{in}(q)$ を含むメタデータと、基準タイミング $t_{out}(s)$ を受け取り、キー入力検出部110で検出されたテキスト情報を表示部に表示するためのタイミングを、基準タイミング $t_{out}(s)$ に基づき補正し(S130)、テキスト情報 $txt_{in}(q)$ とそのテキスト情報
10
を表示するためのタイミング $t_{out}(q)$ を含むメタデータとを動画配信サーバ92に出力する。例えば、S個の $t_{out}(s)$ の中から最も $t_{in}(q)$ に近い $t_{out}(s)$ を、テキスト情報 $txt_{in}(q)$ を表示するためのタイミング $t_{out}(q)$ とする($t_{out}(q) = t_{out}(s)$)。例えば、テキスト情報 $txt_{in}(q)$ が「8」の場合の例を図4に示す。

【0033】

なお、動画配信サーバ92は、対象映像信号にテキスト情報を重畳して出力する。動画配信サーバ92は、テキスト情報 $txt_{in}(q)$ または $txt_{in}(r)$ を、それぞれタイミング $t_{out}(q)$ または入力時刻 $t_{in}(r)$ で視聴者端末91の表示部に表示されるように、対象映像信号とともに配信する。例えば、テキスト情報 $txt_{in}(p)$ がタイミング $t_{out}(q)$ または入力時刻 $t_{in}(r)$ に表示部の右端から表示され、左端に向かって、移動し、左端から消えていく場合、
20
補正されていない場合(入力時刻 $t_{in}(r)$ に表示)の例を図5に示し、補正されている場合(タイミング $t_{out}(q)$ に表示)の例を図6に示す。

【0034】

<タイミング $t_{out}(q)$ の適用時期について>

なお、動画配信サーバ92は、テキスト情報 $txt_{in}(q)$ をタイミング $t_{out}(q)$ で視聴者端末91の表示部に表示されるように、対象映像信号とともに配信するタイミングとして二つのタイミング考えられる。

【0035】

(1)視聴者がテキスト情報 $txt_{in}(q)$ を入力した際、その再生時においては、テキスト情報 $txt_{in}(q)$ を入力時刻 $t_{in}(q)$ で視聴者端末91の表示部に表示されるように、対象映像信号とともに配信する。そして、動画配信サーバ92は、動画データベース内に、対象映像信号とともにテキスト情報 $txt_{in}(q)$ と(入力時刻 $t_{in}(q)$ ではなく)タイミング $t_{out}(q)$ とを格納しておき、次の再生時には、テキスト情報 $txt_{in}(q)$ をタイミング $t_{out}(q)$ で視聴者端末91の表示部に表示されるように、対象映像信号とともに配信する。
30

【0036】

(2)視聴者がテキスト情報 $txt_{in}(q)$ を入力した際、その再生時において、テキスト情報 $txt_{in}(q)$ をタイミング $t_{out}(q)$ で視聴者端末91の表示部に表示されるように、対象映像信号とともに配信する。この場合、タイミング $t_{out}(q)$ と入力時刻 $t_{in}(q)$ との大小関係に応じて配信方法が異なる。

【0037】

(2-A)タイミング $t_{out}(q)$ が入力時刻 $t_{in}(q)$ よりも遅い場合には、テキスト情報 $txt_{in}(q)$ をタイミング $t_{out}(q)$ で表示部に表示されるように、表示のタイミングを遅らせて、対象映像信号とともに配信すればよい。
40

【0038】

(2-B)タイミング $t_{out}(q)$ が入力時刻 $t_{in}(q)$ よりも早い場合には、以下の方法により、配信する。

【0039】

(2-B-1)例えば、図5及び図6のように、テキスト情報 $txt_{in}(q)$ の表示方法が時間の経過を表す場合には、その時間の経過を利用し、テキスト情報 $txt_{in}(q)$ がまるでタイミング $t_{out}(q)$ で表示されたように、対象映像信号とともに配信する。例えば、図6の場合、タ
50

イミング $t_{out}(q)$ が入力時刻 $t_{in}(q)$ よりも早いので、 $t=t_{out}(q)$ のとき（図6の一番の上の図）、動画配信サーバ92は、テキスト情報 $txt_{in}(q)$ を受け取っていないため、テキスト情報 $txt_{in}(q)$ を表示されるように、配信することができない。テキスト情報 $txt_{in}(q)$ を受け取ったタイミング（入力時刻 $t=t_{in}(q)$ ）で、仮に、タイミング $t_{out}(q)$ で視聴者端末91の表示部に表示されていた場合、 $t=t_{in}(q)$ においてテキスト情報 $txt_{in}(q)$ が表示される位置にテキスト情報 $txt_{in}(q)$ を表示する。つまり、図6の一番の上の図から図6の真ん中の図までのテキスト情報 $txt_{in}(q)$ の移動（遷移）を表示せずに、 $t=t_{in}(q)$ において、突然、テキスト情報 $txt_{in}(q)$ を図6の真ん中の図のように表示し、図6の一番の下の図までテキスト情報 $txt_{in}(q)$ の移動させる。なお、この場合、動画配信サーバ92は、動画データベース内に、対象映像信号とともにテキスト情報 $txt_{in}(q)$ とタイミング $t_{out}(q)$ とを格納し

10

【0040】

(2-B-2)また、テキスト情報が示す所定の行為が、手拍子などの周期的な行為の場合には、取得したテキスト情報 $txt_{in}(q)$ を次の周期で表示する。つまり、テキスト情報 $txt_{in}(q)$ をタイミング $t_{out}(q+1)$ で視聴者端末91の表示部に表示されるように、対象映像信号とともに配信する。なお、この場合、動画配信サーバ92は、動画データベース内に、対象映像信号とともにテキスト情報 $txt_{in}(q)$ とタイミング $t_{out}(q)$ とを格納してもよいし、テキスト情報 $txt_{in}(q)$ とタイミング $t_{out}(q+1)$ とを格納してもよい。次の再生時には、テキスト情報 $txt_{in}(q)$ をタイミング $t_{out}(q)$ またはタイミング $t_{out}(q+1)$ で視聴者端末91の

20

【0041】

<効果>

このような構成により、テキスト情報を映像に重畳して表示する際のタイミングを、映像に合わせて補正することができる。例えば、楽曲のテンポに合わせる場合には、テキスト情報の表示のタイミングを合わせることで、視聴者の一体感を向上させることができる。

【0042】

<変形例>

本実施形態では、タイミング補正システム100を視聴者端末91または動画配信サーバ92とは別装置として構成したが、視聴者端末91または動画配信サーバ92に組み込まれる構成としてもよい。また、タイミング補正システム100は、キー入力検出部110と、タイミング検出部120と、補正部130とを含み、独立した装置として説明したが、各部が視聴者端末91または動画配信サーバ92に組み込まれる構成としてもよい。例えば、キー入力検出部110を視聴者端末91に組み込み、タイミング検出部120と補正部130とを動画配信サーバ92に組み込む構成としてもよい。

30

【0043】

また、動画配信サーバ92は、テキスト情報を付加せずに動画のみを配信するサーバであってもよい。その場合には、タイミング補正システム100においてテキスト情報用のデータベースを備え、テキスト情報を付加すればよい。

40

【0044】

本実施形態では、視聴者によって入力され、対象映像信号に重畳して表示される情報としてテキスト情報の例を示したが、他の視覚情報であってもよい。ここで、「視覚情報」とは、表示部を介して視覚的に認識可能な情報であって、例えば、文字、図形若しくは記号若しくはこれらの結合又はこれらと色彩との結合である。また、静止画に限らず、動く画像であってもよい。例えば、(1)本実施形態のように、「笑い」や「拍手」等の所定の行為を意味するテキスト情報（例えば「w」や「8」等）、(2)テキスト情報以外の「笑い」や「拍手」等の所定の行為を意味し、識別するためのコンピュータ上のビット情報、(3)顔文字、絵文字等、通常のテキスト情報で無いもの。例えば、キャリアの異なる携帯電話間で共通絵文字（参考文献7参照）、(4)アスキーアート等、全体としてはテキスト情

50

報とテキスト情報の配置情報を用いた絵のようになっているもの(参考文献8参照)、(5)上述の(1)~(4)に対応するネットスラング。例えば、「笑い」を意味するテキスト情報「w w w w ...」に対して「草生えた」等のネットスラングがある。

(参考文献7)「docomo/au共通絵文字」、株式会社NTTドコモ、[online]、[平成27年2月9日検索]、インターネット<URL: https://www.nttdocomo.co.jp/service/developer/smart_phone/make_contents/pictograph/>

(参考文献8)「アスキーアート」、[online]、2015年2月2日、ウィキペディア、[平成27年2月9日検索]、インターネット<URL: <http://ja.wikipedia.org/wiki/%E3%82%A2%E3%82%B9%E3%82%AD%E3%83%BC%E3%82%A2%E3%83%BC%E3%83%88>>

【0045】

<その他の変形例>

本発明は上記の実施形態及び変形例に限定されるものではない。例えば、上述の各種の処理は、記載に従って時系列に実行されるのみならず、処理を実行する装置の処理能力あるいは必要に応じて並列的あるいは個別に実行されてもよい。その他、本発明の趣旨を逸脱しない範囲で適宜変更が可能である。

【0046】

<プログラム及び記録媒体>

また、上記の実施形態及び変形例で説明した各装置における各種の処理機能をコンピュータによって実現してもよい。その場合、各装置が有すべき機能の処理内容はプログラムによって記述される。そして、このプログラムをコンピュータで実行することにより、上記各装置における各種の処理機能がコンピュータ上で実現される。

【0047】

この処理内容を記述したプログラムは、コンピュータで読み取り可能な記録媒体に記録しておくことができる。コンピュータで読み取り可能な記録媒体としては、例えば、磁気記録装置、光ディスク、光磁気記録媒体、半導体メモリ等のようなものでもよい。

【0048】

また、このプログラムの流通は、例えば、そのプログラムを記録したDVD、CD-ROM等の可搬型記録媒体を販売、譲渡、貸与等することによって行う。さらに、このプログラムをサーバコンピュータの記憶装置に格納しておき、ネットワークを介して、サーバコンピュータから他のコンピュータにそのプログラムを転送することにより、このプログラムを流通させてもよい。

【0049】

このようなプログラムを実行するコンピュータは、例えば、まず、可搬型記録媒体に記録されたプログラムもしくはサーバコンピュータから転送されたプログラムを、一旦、自己の記憶部に格納する。そして、処理の実行時、このコンピュータは、自己の記憶部に格納されたプログラムを読み取り、読み取ったプログラムに従った処理を実行する。また、このプログラムの別の実施形態として、コンピュータが可搬型記録媒体から直接プログラムを読み取り、そのプログラムに従った処理を実行することとしてもよい。さらに、このコンピュータにサーバコンピュータからプログラムが転送されるたびに、逐次、受け取ったプログラムに従った処理を実行することとしてもよい。また、サーバコンピュータから、このコンピュータへのプログラムの転送は行わず、その実行指示と結果取得のみによって処理機能を実現する、いわゆるASP(Application Service Provider)型のサービスによって、上述の処理を実行する構成としてもよい。なお、プログラムには、電子計算機による処理の用に供する情報であってプログラムに準ずるもの(コンピュータに対する直接の指令ではないがコンピュータの処理を規定する性質を有するデータ等)を含むものとする。

【0050】

また、コンピュータ上で所定のプログラムを実行させることにより、各装置を構成することとしたが、これらの処理内容の少なくとも一部をハードウェア的に実現することとしてもよい。

10

20

30

40

50

【図1】

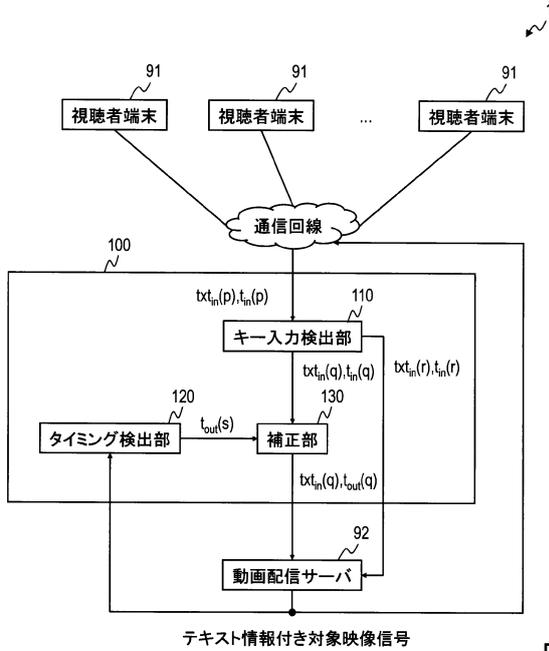


図1

【図2】

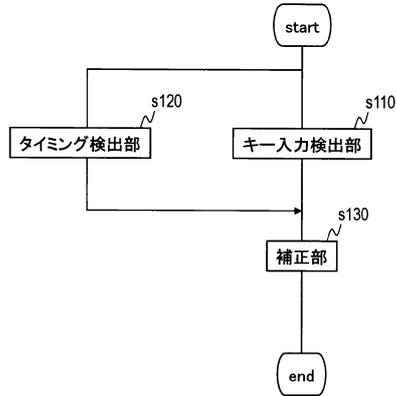


図2

【図3】

	行為	テキスト情報
1	笑う	「w」、「笑」
2	拍手	「8」
...

図3

【図5】

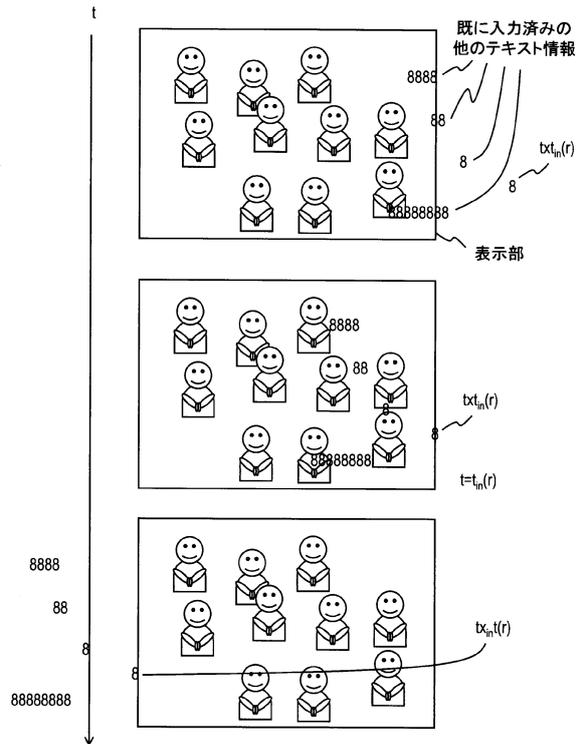


図5

【図4】

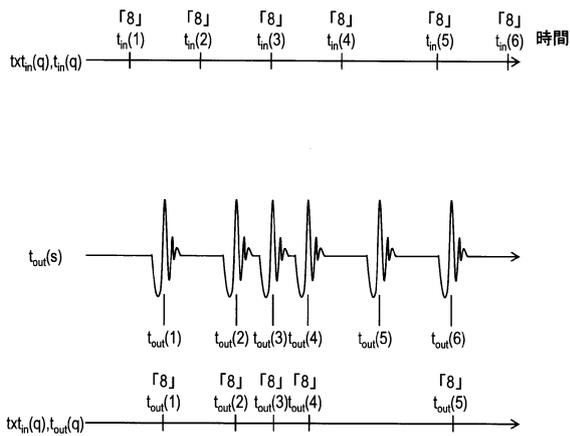


図4

【 図 6 】

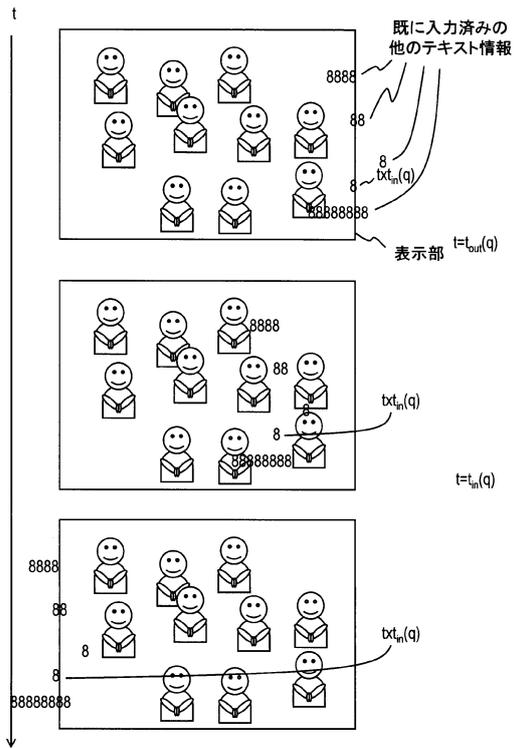


図6

フロントページの続き

- (72)発明者 佐藤 尚
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 ガブリエル パブロ ナバ
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 守谷 健弘
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 小野 健二

- (56)参考文献 特開2013-089052(JP,A)
特開2004-193871(JP,A)
特開2006-157689(JP,A)
特開2006-157687(JP,A)
特開2007-274090(JP,A)
特開2008-148071(JP,A)

(58)調査した分野(Int.Cl., DB名)

G09G 5/00 - 5/42
H04N 7/173
G06F 3/048