

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4347223号
(P4347223)

(45) 発行日 平成21年10月21日(2009.10.21)

(24) 登録日 平成21年7月24日(2009.7.24)

(51) Int. Cl.		F I	
HO4N	5/91	(2006.01)	HO4N 5/91
HO4N	5/781	(2006.01)	HO4N 5/781
G11B	27/00	(2006.01)	G11B 27/00

請求項の数 14 (全 8 頁)

(21) 出願番号	特願2004-563862 (P2004-563862)	(73) 特許権者	390009531
(86) (22) 出願日	平成15年12月19日 (2003.12.19)		インターナショナル・ビジネス・マシーンズ・コーポレーション
(65) 公表番号	特表2006-512007 (P2006-512007A)		INTERNATIONAL BUSINESS MACHINES CORPORATION
(43) 公表日	平成18年4月6日 (2006.4.6)		アメリカ合衆国10504 ニューヨーク州 アーモンク ニュー オーチャードロード
(86) 国際出願番号	PCT/US2003/040726		
(87) 国際公開番号	W02004/059536	(74) 代理人	100086243
(87) 国際公開日	平成16年7月15日 (2004.7.15)		弁理士 坂口 博
審査請求日	平成18年11月9日 (2006.11.9)	(74) 代理人	100091568
(31) 優先権主張番号	10/325,061		弁理士 市位 嘉宏
(32) 優先日	平成14年12月20日 (2002.12.20)	(74) 代理人	100108501
(33) 優先権主張国	米国 (US)		弁理士 上野 剛史

最終頁に続く

(54) 【発明の名称】 マルチメディア文書における多モード特性に注釈を付けるためのシステムおよび方法

(57) 【特許請求の範囲】

【請求項1】

マルチメディア・コンテンツに注釈付けをするための装置であって、
記憶媒体に格納されたマルチメディア・コンテンツを作業メモリに供給するための構成と、

前記供給されたマルチメディア・コンテンツに関連付けられた下記モード、すなわち、
ビデオを含む音声部分、ここで該音声部分はセグメントに区切られている、
音声を含むビデオ部分、ここで該ビデオ部分はショットにセグメント化されている、
ビデオ部分のみ、ここで該ビデオ部分はショットにセグメント化されている、
音声部分のみ、ここで該音声部分はセグメントに区切られている、

のうちの少なくとも1つの選択を可能にするための入力インターフェースと、

前記選択されたモードに注釈付けすることを可能にするための構成であって、該注釈付けが、前記セグメント又前記ショット毎に、注釈のためのリストボックス内にあり且つ前記選択されたモードによって表示される項目に対応するボックスをクリックすることにより行われる、前記構成と

を含む、前記装置。

【請求項2】

前記ビデオ部分の各セグメントへの注釈付けが、イベント、静止シーン又はキーオブジェクトのリストボックス内にある項目に対応するボックスをクリックすることにより行われる、請求項1に記載の装置。

10

20

【請求項 3】

前記ビデオ部分の各セグメントへの注釈付けが、キーワード・テキスト・ボックスに注記することによりさらに行われる、請求項 2 に記載の装置。

【請求項 4】

前記音声部分の各セグメントへの注釈付けが、前景サウンド又は背景サウンドのリストボックス内にある項目に対応するボックスをクリックすることにより行われる、請求項 1 に記載の装置。

【請求項 5】

前記音声部分の各セグメントへの注釈付けが、キーワード・テキスト・ボックスに注記することによりさらに行われる、請求項 4 に記載の装置。

10

【請求項 6】

前記選択モードの注釈が付けられた観察結果を保存するための作業メモリをさらに含む、請求項 1 に記載の装置

【請求項 7】

前記入力インターフェースが、
マルチメディア・コンテンツのビデオのみの選択を可能にするための構成と、
ビデオのみに音声を選択的に追加することを可能にするための構成と
を含む、請求項 1 に記載の装置。

【請求項 8】

マルチメディア・コンテンツに注釈付けをするための方法であって、
記憶媒体に格納されたマルチメディア・コンテンツを作業メモリに供給するステップと

20

前記供給されたマルチメディア・コンテンツに関連付けられた下記モード、すなわち、
ビデオを含む音声部分、ここで該音声部分はセグメントに区切られている、
音声を含むビデオ部分、ここで該ビデオ部分はショットにセグメント化されている、
ビデオ部分のみ、ここで該ビデオ部分はショットにセグメント化されている、
音声部分のみ、ここで該音声部分はセグメントに区切られている、
のうちの少なくとも 1 つの選択を可能にするステップと、

前記選択されたモードに注釈付けを可能にするステップであって、該注釈付けが、前記セグメント又前記ショット毎に、注釈のためのリストボックス内にあり且つ前記選択されたモードによって表示される項目に対応するボックスをクリックすることにより行われる、前記注釈付けを可能にするステップと
を含む、前記方法。

30

【請求項 9】

前記注釈付けを可能にするステップが、前記ビデオ部分の各セグメントへの注釈付けである場合に、イベント、静止シーン又はキーオブジェクトのリストボックス内にある項目に対応するボックスをクリックすることを可能にするステップをさらに含む、請求項 8 に記載の方法。

【請求項 10】

前記ビデオ部分の各セグメントへの注釈付けを可能にするステップが、キーワード・テキスト・ボックスに注記することを可能にするステップをさらに含む、請求項 9 に記載の方法。

40

【請求項 11】

前記注釈付けを可能にするステップが、前記音声部分の各セグメントへの注釈付けである場合に、前景サウンド又は背景サウンドのリストボックス内にある項目に対応するボックスをクリックすることを可能にするステップをさらに含む、請求項 8 に記載の方法。

【請求項 12】

前記音声部分の各セグメントへの注釈付けを可能にするステップが、キーワード・テキスト・ボックスに注記することを可能にするステップをさらに含む、請求項 11 に記載の方法。

50

【請求項 13】

前記選択を可能にするステップが、
 マルチメディア・コンテンツのビデオ・モードの選択を可能にするステップと、
 ビデオ・モードに音声を選択的に追加することを可能にするステップと
 を含む、請求項 8 に記載の方法。

【請求項 14】

コンピュータに、請求項 8 ~ 13 のいずれか一項に記載の方法の各ステップを実行させる、コンピュータ・プログラム。

【発明の詳細な説明】

【技術分野】

10

【0001】

本発明は、マルチメディア・ファイルのコンピュータ処理に関する。より具体的に言えば、本発明は、マルチメディア・ファイルで発生する多モードのイベント、オブジェクト、シーン、および音声に手動で注釈を付けることに関する。

【背景技術】

【0002】

マルチメディア・コンテンツは、ワールド・ワイド・ウェブおよびローカル・コンピュータの両方において、ますます一般的になりつつある。マルチメディア・コンテンツのコーパスが増加するにつれて、コンテンツ内の機能に索引付けすることもますます重要になる。音声とビデオの両方を同時に観察すること、およびその観察結果 (o b s e r v a t i o n s) に注釈を付けることにより、結果として信頼レベルが高くなる。

20

【0003】

既存のマルチメディア・ツールは、全体としてではなく音声またはビデオのいずれかに別々に注釈を付ける機能を提供する。(ビデオ専用注釈付けツールの一例は、[http://www.alphaworks.ibm.com/tech/videoannexから入手可能な発明者] J. S m i t h 等による I B M M P E G 7 注釈付けツールである。他の従来の配置構成については、P a r k 等による 2 0 0 1 年 1 1 月北朝鮮での帰納的モデリングに関する国際会議 (I C I M) の議事録「i M E D I A - C A T : インテリジェント・メディア・コンテンツ注釈付けツール」、および M i n k a 等による T R # 3 4 9、1 9 9 7 年、パターン認識第 3 0 巻、5 6 5 ページ「モデル社会を使用した対話型学習」に記載されている。)

30

【0004】

ビデオまたは音声機能に分離して注釈付けすると、結果としてその機能の識別に対する信頼性レベルが低くなることがわかっている。

【非特許文献 1】 P a r k 等による 2 0 0 1 年 1 1 月北朝鮮での帰納的モデリングに関する国際会議 (I C I M) の議事録「i M E D I A - C A T : インテリジェント・メディア・コンテンツ注釈付けツール」

【非特許文献 2】 M i n k a 等による T R # 3 4 9、1 9 9 7 年、パターン認識第 3 0 巻、5 6 5 ページ「モデル社会を使用した対話型学習」

【発明の開示】

【発明が解決しようとする課題】

40

【0005】

前述の内容を考えると、マルチメディア・ファイル内で生じる多モードのイベント、オブジェクト、シーン、および音声を観察および注釈付けするための改良されたシステムおよび方法を提供することが求められている。

【課題を解決するための手段】

【0006】

本発明の現時点で好ましい少なくとも 1 つの実施形態に従って、ユーザがビデオ単独、音声付きビデオ、音声単独、またはビデオ付き音声を観察すること、ならびに観察された内容に注釈を付けることを可能にする、マルチメディア注釈付けシステムおよび方法が広く企図されている。

50

【 0 0 0 7 】

一実施形態では、1つまたは複数のマルチメディア・ファイルが作業メモリに格納されたコンピュータ・システムが提供される。多モード注釈付けプロセスでは、ユーザが選択したマルチメディア・ファイルを表示し、ファイル・コンテンツを観察するために1つまたは複数モードの選択を可能にし、観察結果に注釈付けを行い、注釈付けした内容を作業メモリに保存する(MPEG-7 XMLファイルなど)。

【 0 0 0 8 】

要約すると、本発明の一態様はマルチメディア・コンテンツを管理するための装置を提供し、この装置は、マルチメディア・コンテンツを供給するための配置構成と、観察のためにマルチメディア・コンテンツに関連付けられたビデオを含む音声部分モードおよび音声を含むビデオ部分モードのうちの少なくとも1つの選択を可能にするための入力インターフェースと、選択されたモードの観察結果に注釈付けするための配置構成とを有する。

10

【 0 0 0 9 】

本発明の他の態様は、マルチメディア・コンテンツを管理する方法を提供し、この方法は、マルチメディア・コンテンツを供給するステップと、観察のためにマルチメディア・コンテンツに関連付けられたビデオを含む音声部分モードおよび音声を含むビデオ部分モードのうちの少なくとも1つの選択を可能にするステップと、選択されたモードの観察結果に注釈付けするステップとを有する。

【 0 0 1 0 】

さらに本発明の他の態様は、マルチメディア・コンテンツを管理するための方法をコンピュータに実行させるためのプログラムを記録したコンピュータによる読取りが可能なプログラム記憶デバイスを提供し、この方法は、マルチメディア・コンテンツを供給するステップと、観察のためにマルチメディア・コンテンツに関連付けられたビデオを含む音声部分モードおよび音声を含むビデオ部分モードのうちの少なくとも1つの選択を可能にするステップと、選択されたモードの観察結果に注釈付けするステップとを有する。

20

【 0 0 1 1 】

本発明ならびにその機能および利点をより良く理解するために、添付の図面と併せて以下の説明が参照され、本発明の範囲は添付の特許請求の範囲で指摘される。

【 発明を実施するための最良の形態 】

【 0 0 1 2 】

図1は、本発明に従った多モード注釈付けシステムの好ましい一実施形態を示すブロック図である。マルチメディア・コンテンツおよび以前の注釈は記憶媒体100に格納される。ユーザ130が注釈付けツールを介して記憶媒体100からマルチメディア・ファイルを選択すると、これは作業メモリ110にロードされ、その一部が注釈付けツール120に表示される。ユーザ130はいつでも、現在の多モード・ファイルに関連付けられた以前に保存された注釈を、記憶媒体100から作業メモリ110へロードするように要求することもできる。ユーザ130は、注釈付けツール120を介して要求することによってマルチメディア・データを表示する。その後ユーザ130は自分の観察結果に注釈を付け、注釈付けツール120はこれらの注釈内容を作業メモリ110に保存する。ユーザはいつでも、記憶媒体100上に注釈を保存するよう注釈付けツール120に要求することができる。

30

40

【 0 0 1 3 】

図2は、ビデオ・シーン、オブジェクト、およびイベントに注釈付けするシステムを示す図である。(図1も同時に参照されたい。)マルチメディア・データが記憶媒体100から作業メモリ110へとロードされている。ビデオ・タブ290が選択されている。マルチメディア・ビデオが、変更シーン検出を使用してショットにセグメント化されている。ショット・リスト・ウィンドウ200は、マルチメディア内のショットの一部を表示する。ここでユーザ130は、ショット・リスト・ウィンドウ200でハイライト表示されているショット210を選択した。好ましくは、ショットのフレーム内で代表的ショットであるキー・フレーム220が表示される。加えて、再生コントロール240を使用して

50

ショットのフレームをビデオ・ウィンドウ 230 に表示することもできる。ビデオは、ミュート・ボタン 250 の選択に応じて、音声付きまたは音声なしで表示することができる。ユーザ 130 は、イベント 260、静止シーン 270、またはキー・オブジェクト 280 のリスト・ボックス内にあるボックスをクリックすることによって、このショットに関する注釈付けを選択することができる。チェック・ボックスに含まれていない重要な観察結果があれば、キーワード・テキスト・ボックス 300 に注記することができる。

【0014】

図 3 は、ビデオ付き音声に注釈付けするシステムを示す図である。(図 1 も同時に参照されたい。) マルチメディア・データが記憶媒体 100 から作業メモリ 110 へとロードされている。ビデオ付き音声タブ 370 が選択されている。マルチメディア・ビデオが、
10
変更シーン検出を使用してショットにセグメント化されている。ショット・リスト・ウィンドウ 200 は、マルチメディア内のショットの一部を表示する。現在の音声ポジションに関連付けられたショット 210 が、ショット・リスト・ウィンドウ 200 でハイライト表示される。音声データがウィンドウ 390 に表示される。音声のセグメント 340 が注釈付けのために区切られており、すなわちその後の注釈付けのために音声の限界または境界が固定されている。音声に関連付けられたビデオは 230 に表示される。ユーザ 130 が再生コントロール 360 を使用すると、音声データ表示 390 は現在の音声データを表示するように更新され、ビデオ・ウィンドウ 230 は現在のビデオ・フレームを反映するように変更される。このようにして、ユーザ 130 は音声に注釈付けをしながら、ビデオを観察すると同時に音声を聞くことができる。好ましくは、ユーザ 130 はボタン 350
20
を使用して音声セグメントを区切る。前景サウンド(320)(セグメント内で最も顕著なサウンド)および背景サウンド(330)(存在するが他のサウンドに対して二次的なサウンド)に応じてチェック・ボックスにチェックを入れて、音声セグメント 340 内で聞こえるサウンドを指示することができる。チェック・ボックスに含まれていない重要な観察結果があれば、キーワード・テキスト・ボックス 300 に注記することができる。

【0015】

図 4 は、ビデオなし音声に注釈付けするシステムを示す図である。(図 1 も同時に参照されたい。) マルチメディア・データが記憶媒体 100 から作業メモリ 110 へとロードされている。ビデオなし音声タブ 400 が選択されている。音声データがウィンドウ 390
30
に表示される。音声のセグメント 340 が注釈付けのために区切られている。ユーザ 130 が再生コントロール 360 を使用すると、音声データ表示 390 は現在の音声データを表示するように更新される。このようにしてユーザ 130 は、音声に注釈を付けながら音声を聞くだけにすることができる。ユーザ 130 はボタン 350 を使用して音声セグメントを区切る。前景サウンド用 320 および背景サウンド用 330 のチェック・ボックスにチェックを入れて、音声セグメント 340 内で聞こえるサウンドを指示することができる。チェック・ボックスに含まれていない重要な観察結果があれば、キーワード・テキスト・ボックス 300 に注記することができる。

【0016】

本発明は、現時点で好ましい少なくとも 1 つの実施形態に従って、マルチメディア・コンテンツを供給するための配置構成と、観察のためにマルチメディア・コンテンツに関連
40
付けられたモードの選択を可能にするための入力インターフェースと、選択されたモードの観察結果に注釈付けするための配置構成とを含むことが理解されよう。これらの要素は、好適なソフトウェア・プログラムを実行している少なくとも 1 つの汎用コンピュータでまとめて実施することができる。これらは、少なくとも 1 つの集積回路または少なくとも 1 つの集積回路の一部でも実施可能である。したがって本発明は、ハードウェア、ソフトウェア、またはその両方の組合せで実施可能であることを理解されよう。

【0017】

本明細書で特に記載のない限り、本明細書で言及および記載したすべての特許、特許明細書、特許出版物、および他の出版物(ウェブ・ベースの出版物を含む)は、それらの全文が本明細書に記載されているかのように、参照により本明細書に完全に組み込まれてい
50

ると想定される。

【0018】

以上、本明細書の例示的な実施形態について添付の図面を参照しながら説明してきたが、本発明はそれらの精密な実施形態に限定されるものではないこと、ならびに、本発明の趣旨または精神を逸脱することなく当業者によって様々な他の変更および修正が実行可能であることを理解されよう。

【図面の簡単な説明】

【0019】

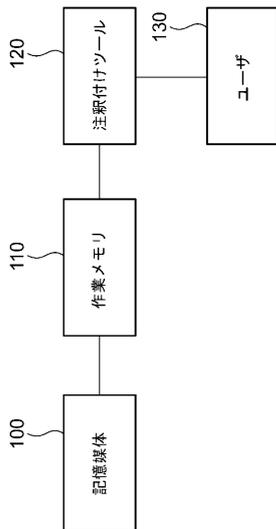
【図1】多モード注釈付けシステムを示すブロック図である。

【図2】ビデオ・シーン、オブジェクト、およびイベントに注釈付けするシステムを示す図である。

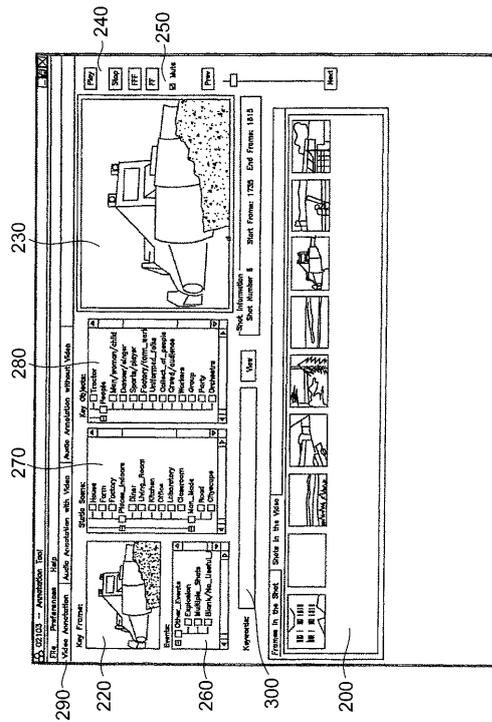
【図3】ビデオ付き音声に注釈付けするシステムを示す図である。

【図4】ビデオなし音声に注釈付けするシステムを示す図である。

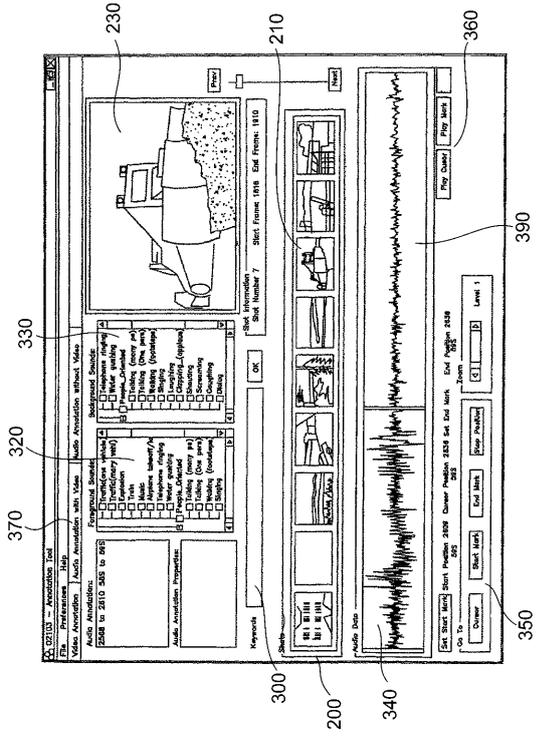
【図1】



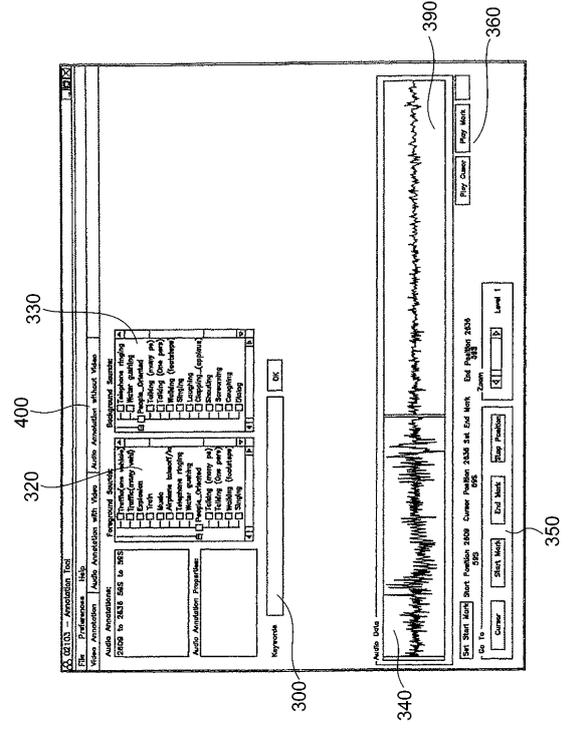
【図2】



【 3 】



【 4 】



フロントページの続き

- (72)発明者 アダムス、ヒュー、ダブリュ、ジュニア
アメリカ合衆国12590 ニューヨーク州ワッピングス・フォールズ ローリング・グリーン
・レーン17
- (72)発明者 イェンガー、ジリッドヘイレン
アメリカ合衆国10541 ニューヨーク州マホバック キャリー・ストリート80
- (72)発明者 リン、チン-ユン
アメリカ合衆国11375 ニューヨーク州フォレスト・ヒルズ ナンセン・ストリート68-4
3
- (72)発明者 ネティ、チャラパシー、ヴィー
アメリカ合衆国10598 ニューヨーク州ヨークタウン・ハイツ ハイ・リッジ・コート235
- (72)発明者 スミス、ジョン、アール
アメリカ合衆国11040 ニューヨーク州ニュー・ハイド・パーク ファレル・ストリート40
- (72)発明者 ツェン、ベル、エル
アメリカ合衆国11375 ニューヨーク州フォレスト・ヒルズ ナンス・ストリート68-49

審査官 堀 洋介

- (56)参考文献 特開平06-043839(JP,A)
特開2002-057981(JP,A)
特開平10-177511(JP,A)
特開2001-218160(JP,A)
特開2001-119653(JP,A)
特開2000-152144(JP,A)
特開平10-254886(JP,A)

(58)調査した分野(Int.Cl., DB名)

H04N 5/781
H04N 5/91
G11B 27/00