

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4300938号
(P4300938)

(45) 発行日 平成21年7月22日(2009.7.22)

(24) 登録日 平成21年5月1日(2009.5.1)

(51) Int.Cl.	F I	
G06F 19/00 (2006.01)	G06F 19/00	140
G06F 13/00 (2006.01)	G06F 13/00	540E
G06F 17/30 (2006.01)	G06F 17/30	110C
G10L 15/00 (2006.01)	G06F 17/30	170G
G10L 15/28 (2006.01)	G10L 3/00	551A

請求項の数 4 (全 14 頁) 最終頁に続く

(21) 出願番号	特願2003-302302 (P2003-302302)	(73) 特許権者	000005108 株式会社日立製作所 東京都千代田区丸の内一丁目6番6号
(22) 出願日	平成15年8月27日(2003.8.27)	(74) 代理人	100100310 弁理士 井上 学
(65) 公開番号	特開2005-71195 (P2005-71195A)	(72) 発明者	大木 康幸 神奈川県川崎市幸区鹿島田890番地 株式会社日立製作所 ビジネスソリューション事業部内
(43) 公開日	平成17年3月17日(2005.3.17)	(72) 発明者	岩崎 一正 神奈川県川崎市幸区鹿島田890番地 株式会社日立製作所 ビジネスソリューション事業部内
審査請求日	平成17年9月9日(2005.9.9)	審査官	小原 正信

最終頁に続く

(54) 【発明の名称】 メディア認識サイト検索方法およびシステム

(57) 【特許請求の範囲】

【請求項1】

メディアデータを認識する複数のメディア認識サーバと、該メディア認識サーバとネットワークを介して接続されたユーザ端末を含むメディアデータ認識システムにおけるメディア認識サイト検索方法であって、

前記ユーザ端末は、予め当該ユーザ端末の記憶部に記憶されたサンプルデータおよび当該ユーザ端末の入力部を介した操作者からの前記サンプルデータに対する操作に従った入力配列で示されるメディア認識サイトを検索する基準となる第一のメディア特徴量を、前記記憶部に格納されたプログラムで構成される検索条件入力ツールに従って、ツール実行部において作成し、

前記ユーザ端末は、該予め記憶されたサンプルデータ、前記メディア認識サーバにおけるデータの入力データ型および出力データ型を含む検索条件データグラムを前記複数のメディア認識サーバそれぞれに、ネットワーク部を介して送信し、

前記複数のメディア認識サーバのそれぞれは、前記送信された検索条件データグラムを当該メディア認識サーバのネットワーク部を介して受信し、

前記複数のメディア認識サーバのそれぞれは、前記受信された検索条件データグラムに含まれる入力データ型および出力データ型が、当該メディア認識サーバ装置の仕様に合致するかを、検索条件照合部を用いて判断し、

前記複数のメディア認識サーバのそれぞれは、前記仕様が合致すると判断した場合、当該メディア認識サーバのメディア認識部を用いて、前記送信されたサンプルデータに対し

て、当該メディア認識サーバで予め規定されている認識処理し、

前記メディア認識サーバのそれぞれは、該認識結果である第二のメディア特徴量および当該メディア認識サーバを識別する識別情報を、当該メディア認識サーバのネットワーク部を介して前記ユーザ端末に送信し、

前記ユーザ端末は、該送信された第二のメディア特徴量と、前記作成された第一の特徴量とを、当該第一および第二それぞれの特徴量を構成する配列についてメディア特徴量比較部を用いて比較し、

前記ユーザ端末は、該比較の結果、前記第一の特徴量に最も類似する第二の特徴量を送信したメディア認識サーバを、該ユーザ端末が有するメディアデータの認識を依頼するメディア認識サイトとして選択することを特徴とするメディア認識サイト検索方法。

10

【請求項2】

前記ユーザ端末は、当該ユーザ端末のネットワーク部を介して、前記選択したメディア認識サイトに対し、該ユーザ端末が有するメディアデータの認識処理を依頼する依頼情報を送信し、

該ユーザ端末は、当該ユーザ端末のネットワーク部を介して、前記選択したメディア認識サイトから、前記依頼を受ける旨の情報を受信した場合は、前記ユーザ端末が有するメディアデータを、該メディア認識サイトに送信することを特徴とする請求項1記載のメディア認識サイト検索方法。

【請求項3】

メディアデータを認識する複数のメディア認識サーバと、該メディア認識サーバとネットワークを介して接続されたユーザ端末を含むメディアデータ認識システムであって、

20

前記ユーザ端末は、

予め当該ユーザ端末の記憶部に記憶されたサンプルデータおよび当該ユーザ端末の入力部を介した操作者からの前記サンプルデータに対する操作に従った入力データの配列で示されるメディア認識サイトを検索する基準となる第一のメディア特徴量を、前記記憶部に格納されたプログラムで構成される検索条件入力ツールに従って作成するツール実行部と、

該予め記憶されたサンプルデータ、前記メディア認識サーバにおけるデータの入力データ型および出力データ型を含む検索条件データグラムを前記複数のメディア認識サーバそれぞれへ送信するユーザ端末ネットワーク部とを有し、

前記複数のメディア認識サーバのそれぞれは、

30

前記送信された検索条件データグラムを受信するメディア認識サーバネットワーク部と

、前記受信された検索条件データグラムに含まれる入力データ型および出力データ型が、当該メディア認識サーバ装置の仕様に合致するかを判断する検索条件照合部とを有し、

前記仕様が合致すると判断した場合、前記送信されたサンプルデータに対して、当該メディア認識サーバで予め規定されている認識処理を行うメディア認識部と、

該認識結果である第二のメディア特徴量および当該メディア認識サーバを識別する識別情報を、前記ユーザ端末に送信するメディア認識サーバネットワーク部とを有し、

前記ユーザ端末は、

該送信された第二のメディア特徴量と、前記作成された第一の特徴量とを、当該第一および第二それぞれの特徴量を構成する配列について比較するメディア特徴量比較部をさらに有し、

40

前記ユーザ端末は、該比較の結果、前記第一の特徴量に最も類似する第二の特徴量を送信したメディア認識サーバを、該ユーザ端末が有するメディアデータの認識を依頼するメディア認識サイトとして選択することを特徴とするメディアデータ認識システム。

【請求項4】

前記ユーザ端末は、当該ユーザ端末のネットワーク部を介して、前記選択したメディア認識サイトに対し、該ユーザ端末が有するメディアデータの認識処理を依頼する依頼情報を送信し、

該ユーザ端末は、当該ユーザ端末のネットワーク部を介して、前記選択したメディア認

50

識サイトから、前記依頼を受ける旨の情報を受信した場合は、前記ユーザ端末が有するメディアデータを、該メディア認識サイトに送信することを特徴とする請求項3記載のメディアデータ認識システム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、映像データなどのメディアを認識するメディア認識サイトを検索するシステムに関し、特に、ユーザの希望に合致する認識を行うメディア認識サイトを検索するシステムに関する。

【背景技術】

【0002】

現在、映像や音声などのメディアデータを認識するメディア認識ネットワークシステムが存在する。このシステムでは、メディアデータを所有する各エンドユーザが、ネットワーク上に配置されたメディアデータ認識処理用コンピュータ（以下、メディア認識サイトと呼ぶ）に接続し、エンドユーザ側からメディア認識サイトへメディアデータを送信する。メディア認識サイトからはその認識結果メタデータが返信される（特許文献1参照）。

【0003】

また、ネットワーク上に配置された各種処理サービスを検索する手法としては、Webサービスの検索ディレクトリUDDI (<http://www.uddi.org>)がある。UDDIでは、検索条件として、Webサービスの各種分類情報とともに、Webサービスの入力データと出力データの形式（データ型）を指定する。Webサービスを利用したいユーザは、Webサービスの種別情報と共に入力データの型と出力データの型を指定してWebサービスのサイトアドレスを入手し、その後サイトへの接続を行う。

【0004】

上記メディア認識ネットワークシステムにおいて、ユーザがメディア認識サイトの検索を行う場合、検索条件となる認識サイトの入力データの型として、認識対象となるメディアデータの種別（映像、音声、3D）や、そのフォーマット（画像の幅高さ、圧縮方式、色数、音声チャンネル数）を指定する。同様に認識サイトの出力データの型として、出力されるメタデータの型を指定する。

【0005】

【特許文献1】特開平10-282989号公報

【発明の開示】

【発明が解決しようとする課題】

【0006】

上記メディア認識ネットワークシステムにおいて、メディア認識サイトをその入出力データ型だけで検索する場合には、ユーザが本当に希望するメディア認識サイトを検索・選定できない可能性がある。その理由は、メディア認識方法が同じで、かつ認識精度が高いメディア認識サイトを選び出せたとしても、ユーザが認識したい認識対象が、メディア認識サイトの認識結果と合致しない場合があるからである。たとえば、映像中の物体動き追尾を用いてサッカー番組におけるサッカーボールの動きを追いかける場合、ある動き追尾認識サイトはサッカーボールを正しく追尾するのに対し、別の動き追尾認識サイトはサッカー選手を追尾してしまう場合がある。どちらの動き追尾認識サイトも入出力データ型が「映像、動き情報」と同じであり、どちらも独自のアルゴリズムにより高精度の動き追尾を行うのであるが、一方の動き追尾サイトはユーザが希望していないサッカー選手の動き追尾情報を返信してしまう。

【0007】

本発明の目的は、メディアデータの検索条件をもとに、ユーザの希望に添うメディア認識サイトを検索するメディア認識サイト検索システムを提供することにある。

【課題を解決するための手段】

【0008】

10

20

30

40

50

上記目的を達成するために本発明は、ユーザ端末は、予め記憶されたサンプル映像をもとにして、検索条件入力ツールを用いて、メディア認識サイトを検索する基準となる第一のメディア特徴量（正解特徴量）を作成する。メディア認識サーバは、サンプル画像を認識処理し、その認識結果である第二のメディア特徴量をユーザ端末に送信する。ユーザ端末は、作成した正解特徴量とメディア認識サーバ150の認識結果であるメディア特徴量とを比較し、ユーザが希望する認識処理を行うメディア認識サイトを選択する。

【発明の効果】

【0009】

本発明によれば、多数のメディア認識サイトの中から、ユーザの希望に沿ったメディア認識処理を行うサイトを選び出すことができる。

10

【発明を実施するための最良の形態】

【0010】

本発明における最良の実施形態を図面を用いて説明する。

【0011】

本実施形態においては、次のような動作シナリオを想定している。あるユーザが、サッカーの試合の分析をするために、自分が所有するあるサッカー映像を分析したいとする。映像の分析は、まずサッカーボールの動きの情報をまとめる処理を行い、次に各選手の詳細な動きを分析する処理を行う。このシナリオのうち、特にサッカーボールの動きを分析する処理について、本実施形態の認識サイト検索システムを用いて具体的に説明していく。

20

【0012】

まず図1を用いて、本発明の一実施形態であるメディア認識サイト検索システムのシステム構成を説明する。本システムは、ユーザ側に設置され、ユーザが操作するユーザ端末110と、映像や音声などのメディアデータを受け取り、その内容を分析・認識しその結果をメディア特徴量として返信するメディア認識サーバが複数個150、160、170と、メディア認識サイトの検索を支援する検索条件入力ツール取得サーバ140で構成され、上記サーバや端末がネットワーク130に接続されている。図1では、メディア認識サーバA150およびメディア認識サーバB160は、映像中の動いている物体を発見・追尾する動き追尾認識機能を持ち、メディア認識サーバC170は、音声を入力としてその発話内容を認識しテキストデータに変換する音声認識機能を持つものとする。

30

【0013】

ユーザ端末110は、プログラムコードである検索条件入力ツール111を実行する。ユーザの操作によって、ユーザ端末110は、この検索条件入力ツール111を用いてメディア認識サイトの検索・選定の処理を行う。このプログラムコードは、これを実行するツール実行装置113を備える。なおこのプログラムコードにはCPU依存型のネイティブコードなどを用いてもよい。検索条件入力ツール111は必要に応じてキーボードやマウスなどの入力装置118とユーザ操作結果を表示するディスプレイ装置117を備える。

【0014】

ユーザ端末110は、TCP/IPネットワーク接続等により外部装置と情報の送受信を行うネットワーク装置112、各種データを蓄積するハードディスク装置116、メディア特徴量比較装置114、およびこれらユーザ端末110内の各装置を制御するユーザ端末制御装置115から構成される。ユーザ端末制御装置115は、CPUとメモリを備える一般的なコンピュータ装置であって、図2に示すユーザ端末側の処理フローを示すプログラムを備える。なお本実施形態では、ハードディスク装置116に、認識サイトの検索に用いる一時的な映像であるサンプル映像119、実際に分析したい映像が録画されている実映像120、および、ユーザが希望するメタデータの正解値である正解特徴量121が記録されている。なお本実施形態では映像を用いたが、音声認識サイトの検索の場合は音声データを、顔認識サイトの検索の場合は写真データを記録することになる。

40

【0015】

50

検索条件入力ツール取得サーバ140は、複数の検索条件入力ツール143, 144等を記憶装置142内に保持することにより、ネット上にどのようなメディア認識サイトがあるのかといった認識方式のカテゴリを分類して管理するサーバである。検索条件入力ツール取得サーバ140は、主にユーザ端末110からアクセスされる。なお検索条件入力ツール取得サーバ140もネットワーク装置141を備える。

【0016】

メディア認識サーバ150、160、170は、ネットワーク経由でメディアデータを受信し、メディア認識装置153を用いて、受信したメディアデータを認識し、その認識結果であるメディア特徴量を返信する。このためメディア認識サーバ150、160、170は、ネットワーク接続を行うネットワーク装置151を有する。

10

【0017】

さらに、メディア認識サーバ150、160、170は、メディア認識サイトを検索する検索条件が、自身のメディア認識装置153と合致するか否かを調べる検索条件照合処理装置152と、メディア認識サーバ内の各装置を制御する認識サイト制御装置154を備える。認識サイト制御装置154は、ユーザ端末制御装置115と同様にコンピュータとプログラムから構成される。メディア認識サーバ160と170は、メディア認識サーバ150と同様の構成である。

【0018】

メディア認識装置153が行う認識処理は、映像中で動いている物体を自動追尾する認識処理、映像中のある色の部分を抽出して指し示す認識処理、および、音声を入力としてその発言内容を認識しテキストとして返す音声認識処理などが考えられる。これら処理を行うためには、公知のメディア認識製品（音声認識ソフトウェアや映像認識ソフトウェア）を用いることとし、ここでは詳細な説明は省略する。本実施形態においては、その認識処理がどのような形式のメディアデータを入力とし、どのような形式のメディア特徴量を出力するのかがポイントである。

20

【0019】

なお、本実施例では、サンプル映像119、実映像120、メディア特徴量比較装置114、およびツール実行装置113をユーザ端末110の中に配置したが、これらをネットワーク上の別のサイト（コンピュータやサーバ）上に配置してもよい。たとえば、映像（一般にはメディアデータ）そのものは別のサイトに保存されていて、ユーザ端末110上ではその置き場所を示すURLだけを記録し、ユーザ端末110やメディア認識サーバ150が必要な時にそのURLを元に映像の実データをダウンロードまたはストリーミングにより入手すれば、本実施例と同様の動作を実現することができる。同様に、検索条件入力ツール111とツール実行装置113も、ユーザ端末110ではなく別の検索条件入力ツール取得サーバ140上に配置されて、必要に応じて、検索条件入力ツール111とツール実行装置113が、ネットワーク経由でユーザ端末110内のディスプレイ装置117、入力装置118、ハードディスク装置116をアクセスするようにしてもよい。メディア特徴量比較装置114についても、ユーザ端末110上に配置したが、実際は様々なメディア特徴量に対する類似度比較処理を行わなければならないため、類似度比較サーバなどを別途設けて、このサーバが代理で認識処理する構成でもよい。

30

40

【0020】

ここで、メディア認識サイトの検索を行う場合の、入出力データ型の指定について説明する。入出力データ型の指定には、ISOのMPEG-7(ISO/IEC 15938)に定められるマルチメディアコンテンツのため情報記述方式を利用できる。MPEG-7ではW3C XML Schemaを元にした型定義言語を用いて、メディア情報記述のための様々な標準型を規定している。たとえば、映像の種別やフォーマットを記述する型として“mpeg7:MediaFormatType”（または<MediaFormat>タグ）というXML型を用意し、詳細なフォーマット情報を記述可能である。同様に、メタデータの型として、映像に関するもの（色、形、動き追尾情報）や音に関するもの（音声認識結果テキスト）など、様々な標準型を用意している。たとえば、動き追尾情報には“mpeg7:MovingRegionType”（または<MovingRegion>タグ）という型があり、物体

50

の形状やその時間的な動き情報（映像内の座標位置 x,y と映像時刻 t のリスト）を一括して記述できる。なお、メタデータと呼ばれるメディアデータの関連情報のうち、特に二つのメタデータ間の類似度合いを数学的に計算できるものを、メディア特徴量（または単に特徴量）と呼ぶ。

【0021】

次に図2の全体処理フローと、図3，図4のユーザ端末のインタフェース画面を用いて、本発明の実施例の流れを説明する。

【0022】

図2は、メディア認識サーバの検索・選定の処理を示す。

【0023】

まずユーザ端末110は、検索条件入力ツール取得サーバ140に接続する（ステップ211）。ユーザ端末110のディスプレイ装置117は、図3に示す認識種別メニュー画面310を表示する（ステップ212）。認識種別メニュー画面310において、ユーザによってメディア認識の種別が選択されると、ユーザ端末110は、その情報が検索条件入力ツール取得サーバ140に送信し、検索条件入力ツール取得サーバ140の記憶装置142に格納された検索条件入力ツールであって、選択されたメディア認識種別に対応する検索条件入力ツールが、ユーザ端末110にダウンロードされる（ステップ213）。図3に示す例では「動き追尾」ボタン312が選択されたので、「動き追尾」の検索条件入力ツール144がユーザ端末110にダウンロードされる。

【0024】

次に、ユーザ端末110は、ダウンロードされた検索条件入力ツール144を実行し、ユーザの操作によってユーザ端末110内に正解特徴量121を作成する（ステップ221）。本実施の形態では、正解特徴量は、サンプル映像において「ボールを追いかける」というものである。

【0025】

ステップ221で正解特徴量121を作成すると、ユーザ端末110は、ネットワーク上の全メディア認識サイトに向けて検索条件データグラムを送信する（ステップ231）。検索条件データグラムには、メディア認識サイトの入力データ型、メディア認識サイトの出力データ型、サンプル用メディアデータ（ここではサンプル映像119）が含まれる。検索条件データグラムの詳細は後述する。

【0026】

ステップ231にて検索条件データグラムがネットワーク上に配信されると、それを受信した各メディア認識サーバ150、160、170は、検索条件データグラムのうち入力データ型と出力データ型が、自己のメディア認識装置の仕様に合致するか否かを照合する（ステップ241A，B，C）。この場合、メディア認識サーバC170は音声認識サーバのため、入力データ型が“映像”であるこのサンプルデータ（サンプル映像119）は、このサーバでは処理できない（241C）。このように検索条件の照合がNOの場合、メディア認識サーバC170では、以降の認識処理、返信処理は行わない。

【0027】

メディア認識サーバA150、B160は「動き追尾」の認識処理を行うサーバであり、検索条件の照合が成功する。この場合、検索条件データグラムに含まれるサンプル映像119を元に、各々のメディア認識装置153を用いて動き追尾処理を行う（ステップ242A，B）。メディア認識サーバA150、B160は、動き追尾処理の結果（ x,y,t ）の並びを、MPEG-7特徴量<MovingRegion>の形式で記述し、メディア認識サイトA150，B160を識別するURLと共に、ユーザ端末110に返信する（243A，B）。

【0028】

次にユーザ端末110は、各メディア認識サイトから返信されたMPEG-7<MovingRegion>特徴量と、ユーザ端末110が持つ正解特徴量121ととの類似度を比較する（ステップ251）。ユーザ端末110は、比較の結果、正解特徴量116と一番類似する認識結果

10

20

30

40

50

(特徴量)を出力する認識サイトを選ぶ。ステップ251の具体的な処理フローは図6に示す。ここでは、正解特徴量に一番類似している特徴量を返信したサイトとしてメディア認識サイトA150が選ばれたとする。

【0029】

ステップ221において説明したように、今回の正解特徴量121は「ボールを追いかける」特徴量である。各メディア認識サイトから返信された特徴量の中で、この正解特徴量121に一番類似しているものを選ぶことは、数ある「動き追尾」を行う認識サイトのうち、一番ユーザの意図どおりにボールを追いかけている認識サイトを選ぶことである。このように正解特徴量とメディア認識サイトからの返信特徴量を比較することにより、ユーザの希望する最適なメディア認識サイトの検索・選定が可能になる。

10

【0030】

次に、ユーザ端末110は、選定されたメディア認識サイトA150に選定通知を送信し、実映像120の配信接続依頼を発行する(ステップ261)。メディア認識サイトA150は、接続OKを示すACK信号をユーザ端末110に返信する(ステップ262)。ユーザ端末110は、ACK信号を受信すると、実映像120をメディア認識サイトA150にストリーミング配信し(ステップ263)、メディア認識サイトA150は、受信した実映像120に対して順次動き追尾処理を行い、その認識結果をユーザ端末110に返信する(ステップ264)。このストリーミング配信は、ユーザ端末110が配信を打ち切るまで継続される。

【0031】

20

なお、ステップ231において配信される検索条件データグラムに関して、入力データ型、出力データ型を検索条件データグラム上に表現する方法として、本実施形態では、MP EG-7記述形式を用いる。たとえば「352x240サイズ, 2Mbps映像, 音なし」を表す場合、次のように記述すればよい。

```
<MediaFormat xmlns=" http://www.mpeg7.org/2001/MPEG-7_Schema " >
```

```
<Format>
```

```
<VisualCoding>
```

```
<BitRate>2000000</BitRate>
```

```
<Frame width=" 352 " height=" 240 " />
```

```
</VisualCoding>
```

```
</Format>
```

```
</MediaFormat>
```

30

同様に、動きの特徴量を出力型として表す場合、次のように記述すればよい。

```
<outputType xmlns:mpeg7=" http://www.mpeg7.org/2001/MPEG-7_Schema "
```

```
name=" mpeg7:MovingRegionType " />
```

この場合、<outputType>は本実施例が定義したタグであり、この記述は「MPEG-7のうち<MovingRegion>等として記述される特徴量である“MovingRegionType”型」を表している。ここで、MovingRegionTypeの内容は、xmlns:mpeg7に示される場所にあるスキーマで定義されている。

【0032】

40

ステップ231で送信されるサンプル映像119については、本実施形態では説明を分かりやすくするため、送信される検索条件データグラムに映像データ全てを追加している。別の構成としては、サンプル映像の置き場所のURL等を検索条件データグラムに記述するに留め、送信された検索条件データグラムを受け取ったメディア検索サイトが、必要に応じてそのURLを経由してサンプル映像にアクセスする構成も考えられる。この構成の場合は、通信トラフィックが軽減できるため望ましい。同様に、検索条件データグラムの配信も、本実施形態ではネットワーク全域にマルチキャスト配信しているが、マルチキャスト範囲を絞り込むような中間的なセンタサーバ(検索条件のキャッシュ&プロキシサーバ)を設け、そこに対して検索条件データグラムを送信する構成でもよい。この構成の方が、通信トラフィックが軽減できる(その代わりにセンタサーバの処理負荷がかかる)。

50

図3は、図2のステップ212において表示される認識種別メニュー画面310を示す。認識種別メニュー画面310はWebCGIなどで構築され、メディアの認識種別（音声認識、動き追尾、顔認識）に対応するダウンロードボタン311、312、313を含む。この認識種別はネットワーク上に多数存在するメディア認識サイトをその認識方法別に分類したものである。たとえば、同じ映像の物体の動き追尾認識機能でも、その実現方法としては、物体の特定の色を追いかけたり、映像の差分を元に物体の移動情報を抽出して追いかけたり、特定の物体の形をパターン化して追いかけたりするなど、様々である。本実施形態では、これら多様なメディア認識サイトをユーザに分かりやすく提示するため、これら全てを「動き追尾」に分類する。

【0033】

図3に示す認識種別メニュー画面310を構築する場合、検索条件入力ツール取得サーバ140は、記録装置142内に検索条件入力ツールを格納するにあたって、認識種別の分類を管理する必要がある。本実施形態では、カテゴリ情報をメディア認識処理の（入力データ型，出力データ型）の組として管理している。例えば、動き追尾検索条件入力ツール144の場合、図2のステップ231で述べたように、入力データ型と出力データ型を、MPEG-7を使って（入力データ型＝映像，出力データ型＝動き情報）のように記述できる。同様に音声認識の場合は、入力データ型と出力データ型を（入力データ型＝音声，出力データ型＝テキスト）と記述できる。検索条件入力ツール取得サーバ140は、これらの入出力データ型の組に、「動き追尾」や「音声認識」などの認識種別名称と、その認識処理種別に対応する検索条件入力ツールのプログラムを含めてデータベースで管理する。これにより、検索条件入力ツール取得サーバ140は、その認識種別名称の一覧を、WebCGI画面形式として認識種別メニュー画面310を構築することが可能である。なお、認識種別メニュー画面310において、認識種別の検索を行う構成でもよい。例えば、先のDBに認識種別ごとの概要説明文もあわせて格納してき、DBの全文検索機能を用いて検索を行えば、よりユーザにわかりやすい認識種別メニュー画面を提供することができる。

【0034】

図4は、図2のステップ221における検索条件入力ツール144（図2のステップ213で選択された図3の検索条件入力ツール144）を実行するときの画面を示す。図4の画面は、サンプル映像119であるサッカー番組の映像を開いて、映像の中のサッカーボールの動きを追いかけるような検索条件を設定する例を示す。このように、検索条件入力ツールは、ユーザ画面付きプログラムの形式をとっているため、様々なメディアの認識処理に特化したユーザ画面を提供することができる。このため、ユーザは認識技術についての深い知識を知ることなく、わかりやすい形で「動き追尾」を行う認識サイトの検索条件（つまり正解特徴量121）を入力できる。

【0035】

図4のディスプレイ画面117について説明する。この画面は、数ある動き追尾認識サイトの中からユーザの希望に添った認識サイトを検索・選定する検索条件を入力する際に用いられる。具体的には、検索条件入力ツール144は、検索・選定用に用いるサンプル映像119を入力し、ユーザの操作によって、正解特徴量121を設定して出力する。本実施形態では、サンプル映像119としてサッカーの短編映像411が指定されている。サンプル映像119は、サッカー番組の実映像120とは別の短編映像であるが、実映像120を直接用いても、またはネットワークに接続されたファイルサーバの映像リストから入手してもよい。本実施形態では、手軽に検索条件（つまり正解特徴量）を入力できるようにサンプル映像を短編映像とし、また、正解特徴量をユーザから隠蔽する点を分かりやすく説明するために、そのユーザしか知らない（つまりネットワーク上に公開されていない）独自の映像をサンプル映像として用いる。サンプル映像119の現在の再生時刻の画面411には、サッカー選手423とサッカーボール421が表示されている。またこの画面には、ユーザが入力したサッカーボールの軌跡ライン422と、それに用いたマウスカーソル415が表示されている。この画面は、「私が期待しているメディア認識サイ

10

20

30

40

50

トの認識結果とは、サッカー選手の追尾ではなく、サッカーボールの追尾である」ことがユーザから入力されたことを表している。本ツールにより、サッカー選手の追尾とサッカーボールの追尾の区別といったユーザが希望する検索条件を、メディア認識サイトの検索・選定の際に容易に指定できる。

【 0 0 3 6 】

図 4 に示す画面の操作について説明する。画面 1 1 7 において、まず映像選択ボタン 4 1 2 を押してサンプル映像 1 1 9 を指定する。その後、映像操作パネル 4 1 3 を操作してサンプル映像 1 1 9 のサッカーボールが写っている先頭の時刻 t_1 を表示する。時刻 t_1 の表示画面 4 1 1 にて、サッカーボールの位置をマウスカーソルでクリックすると、その時刻 t_1 とマウスカーソル座標 x_1, y_1 が、正解特徴量の一要素 (x_1, y_1, t_1) として追加される。同様に時刻を少し進めてサッカーボール位置をクリックすることを繰り返すと、時刻 t_1 から現在時刻 t_n までのサッカーボールの軌跡 (x_1, y_1, t_1) (x_2, y_2, t_2) ... を正解特徴量 4 2 2 として登録することができる。正解特徴量の座標データ 4 2 2 がある程度まとまったら、最後に正解保存 & サイト検索ボタン 4 1 4 を押すことにより、正解特徴量のデータ 4 2 2 (この場合座標データ) をユーザ端末 1 1 0 のハードディスク装置 1 1 6 の正解特徴量格納領域 1 2 1 に保存する。

10

【 0 0 3 7 】

図 5 は、図 2 のステップ 2 2 1 における動き追尾検索条件入力ツール 1 4 4 (図 3) が行う処理フローを示す。まず、映像が未選択状態であるとして `null` に初期化される (ステップ 5 0 1)。同様に、正解の特徴量の配列がクリアされ、その個数を示す N も 0 に初期化される (ステップ 5 0 2)。その後、画面表示が行われ (ステップ 5 0 3)、ユーザ操作イベント待ちループ (ステップ 5 0 4) に入る。

20

【 0 0 3 8 】

画面上でどのような操作が行われたかを判断し (ステップ 5 1 0)、映像選択ボタン 4 1 2 (図 4) が押下されたと判断した場合、使用する映像がユーザの指定した映像ファイル (サンプル映像) に初期化される (ステップ 5 2 1)。ステップ 5 1 0 で映像操作パネル 4 1 3 が操作されたと判断した場合、ユーザが指定した操作に従って、映像の再生・停止・位置移動処理が行われる (ステップ 5 2 3)。ステップ 5 1 0 でマウスがクリックされた場合は、正解を示す配列に (マウスの x 座標、 y 座標、映像の現在時刻) という組を正解の配列に追加したのち、その配列を時刻順にソートする (ステップ 5 2 5)。ユーザがマウスをクリックするたびに、正解の (座標点、時刻) の組が追加される。本実施形態では簡単のため正解の配列の削除機能は説明していない。実際には、図面作成ソフトにおける折れ線の描画機能のように、折れ線の制御点の上にマウスカーソルを置き、`[Ctrl]+` クリックをすると、制御点を削除するような処理と同様の実装をすればよい。ステップ 5 1 0 で正解保存 & サイト検索ボタン 4 1 4 が押下されたと判断した場合は、まずユーザ端末 1 1 0 のハードディスク装置 1 1 6 に正解特徴量を保存する (ステップ 5 2 7)。そして上記したように、検索条件データグラムを (入力データ型 = 映像, 出力データ型 = 動き追尾特徴量 "mpeg7:MovingRegionType", サンプル用メディアデータ = サンプル映像 1 1 9) として作成する (ステップ 5 2 8)。その後、メディア認識サイトの検索処理を実行する (ステップ 5 2 9)。

30

40

【 0 0 3 9 】

なお、ユーザ操作を判断する処理 (ステップ 5 1 0) を終えた後は、正解の配列のデータを映像画面 4 1 1 上に動きの軌跡 4 2 2 として表示する。具体的には、まず正解の配列全てに対してループする (ステップ 5 1 1)。ここで、二点間の線分を引くためループ開始値を 2 としている。ループの中では、過去から映像の現在時刻までの時間区間のみの正解特徴量を画面に描画する必要があるため、正解 $[k]$ の時刻を確認する (ステップ 5 3 1)。もし正解の情報が映像の現在時刻より前の場合、その x, y 座標の組を用いて画面上に線分を表示する (ステップ 5 4 1)。

【 0 0 4 0 】

図 6 は、図 5 のステップ 5 2 9 の処理を詳細に示すフローである。つまり、ユーザ端末

50

110において、正解特徴量が指定された後に行われる処理フローである。検索処理529は、図2のフローのうち、ステップ231～ステップ264までの処理を具体的に示したものである。本処理529の入力は、正解特徴量と、検索条件データグラムである。

【0041】

まず、ユーザ端末110は、検索条件データグラムをネットワーク上にマルチキャストする(ステップ610)。次に返信データグラムを一定期間待ち、その一定時間の間にユーザ端末110に対して返信された返信データグラムを返信の配列に追加する(ステップ611)。次に、返信された特徴量の中から、一番正解特徴量に近い返信データグラムを探す。具体的には、まず最小類似度minを無限大に、最適な認識サイトURLをnullに初期化する(ステップ612)。次に全ての返信データに対してステップ620～630のループを行う(ステップ613)。ステップ613では、まず返信データグラム[k]中の特徴量と、正解特徴量121の類似度を計算する。類似度計算の詳細は省略するが、本実施例のような(x,y,t)の配列からなる動き追尾特徴量AとBがあった場合、例えば、簡単な類似度を計算するには次式を用いればよい。

10

【0042】

【数1】

数1

20

$$\text{類似度 Diff}(A,B) = 1/NT \sum |xy(A,t) - xy(B,t)|$$

(全てのt∈T)

A,B = 動き追尾特徴量 = (x,y,t)の集合

T = AとBに含まれる全てのtの集合

NT = Tの要素数

30

xy(C,t)

$$= (C[k].x, C[k].y) \cdots \text{if } C[k].t \leq t < C[k+1].t$$

$$(C[1].x, C[1].y) \cdots \text{if } t < C[1].t$$

$$(C[NC].x, C[NC].y) \cdots \text{if } C[NC].t \leq t$$

NC = Cの要素数

|xy| … ベクトル xy のノルム

ここで計算された類似度が、現在のminより小さいかどうかを判断し(ステップ621)、小さい場合は、ステップ620で計算された類似度をminに入力してminを更新し、また認識サイトURLを返信データグラムに記録されている認識サイトのURLに更新する(ステップ630)。最後に、認識サイトURLがnullでないかどうかを判断し(ステップ614)、nullでない場合、最適な認識サイトが検索・選定できたことを示す。そして、認識サイトURLに示されるメディア認識サイトに接続し(ステップ640)、実映像120の送信が終了するまでループし(ステップ641)、実映像120のデータをストリーミング送信してメディア認識サーバが認識処理を行い、その認識結果をユーザ端末110が受信する(ステップ642)という処理を繰り返す。

40

【0043】

図7は、メディア認識サーバ150が実行される検索条件照合処理(図2のステップ2

50

41)の処理フローを示す図である。図2におけるステップ241B, Cでも同様の処理を行う。図7において示す検索条件照合処理(ステップ701)の入力パラメータは、受信元情報(ユーザ端末110のIPアドレスやURLなど)と、検索条件データグラムである。

【0044】

まず、メディア認識サーバ150は、検索条件データグラムの中の入力データ型が「映像」であるか否かを判断する(ステップ702)。本実施形態のMPEG-7記述例の場合、<MediaFormat>のタグの中に<VideoCoding>タグが含まれていれば「映像」と判断する。もし「映像」でない場合(「音声」などの場合)は、メディア認識サーバ150が処理できないデータなので検索条件処理701を終了する(ステップ710)。次に、メディア認識サーバ150は、検索条件の出力データ型が“mpeg7:MovingRegionType”であるか否かを確認する(ステップ703)。もし“mpeg7:MovingRegionType”ではない場合(例えば色情報“mpeg7:DominantColorType”などの場合)は、本メディア認識サイトでは処理できないデータなので、検索条件処理を終了する(ステップ711)。もし入力データ型も出力データ型も本サイトで処理可能な場合、検索条件データグラムに含まれるサンプル用メディアデータ(サンプル映像119)をもとに、メディア認識サーバ150は動き追尾認識処理を実行する(ステップ704)。そして、その結果を認識結果特徴量として図示しない記憶装置に保存し、認識結果特徴量と自身のメディア認識サイトのURLとを組にして返信データグラムにまとめ、それをユーザ端末110に返信する(ステップ705)。

【0045】

以上が、本発明の一実施例における処理の全体フローである。このように、検索条件入力ツール取得サイト140、検索条件入力ツール143, 144, 145、正解特徴量121、サンプル映像119を活用することにより、多様の認識技術の中から分かりやすく認識技術を選びだし、かつ、ユーザの主観も含めた検索条件に合致したメディア認識サイトを検索・選定できるようになる。

【0046】

本実施の形態では、検索条件入力ツールを用いて、サッカー選手かボールのどちらをユーザが欲しているかをインタラクティブに設定できるため、ユーザの主観に沿った検索条件を入力できる。また、そのユーザの入力した検索条件を正解特徴量としてユーザ端末に保存し、これとは別にメディア認識サイトに同一のサンプルメディアデータを認識させ、これらを類似比較させることによって、よりユーザの主観に近いメディア認識サイトを選び出すことができる。

【図面の簡単な説明】

【0047】

【図1】本発明の一実施形態となるメディア認識サイト検索システムのシステム構成図を示す。

【図2】本発明の一実施形態の全体処理フローを示す。

【図3】認識種別メニュー画面例および検索条件入力ツール取得サーバ140が記憶する検索条件入力ツールを示す。

【図4】検索条件入力ツール111の実行画面の画面例を示す。

【図5】検索条件入力ツール111の検索条件入力処理フローを示す。

【図6】ユーザ端末110におけるメディア認識サイト検索処理を示す処理フローである。

【図7】メディア認識サーバにおける検索条件照合処理を示す処理フローである。

【符号の説明】

【0048】

110：ユーザ端末、111：検索条件入力ツール、113：ツール実行装置、114：メディア特徴量比較装置、115：ユーザ端末制御装置、116：ハードディスク装置、140：検索条件入力ツール取得サーバ、150：メディア認識サーバ、

【 図 1 】

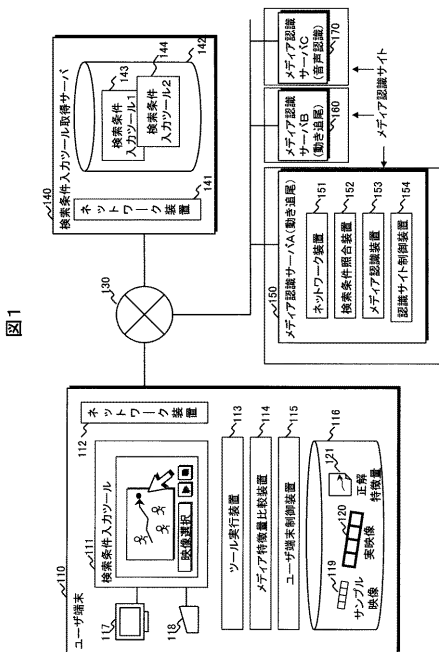


図1

【 図 2 】

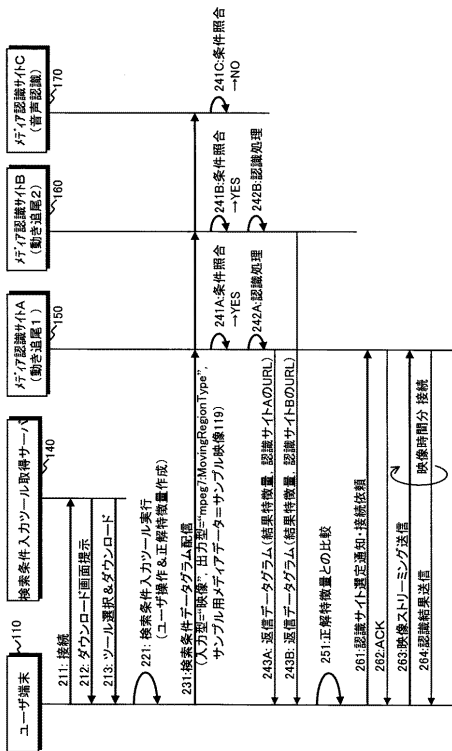


図2

【 図 3 】

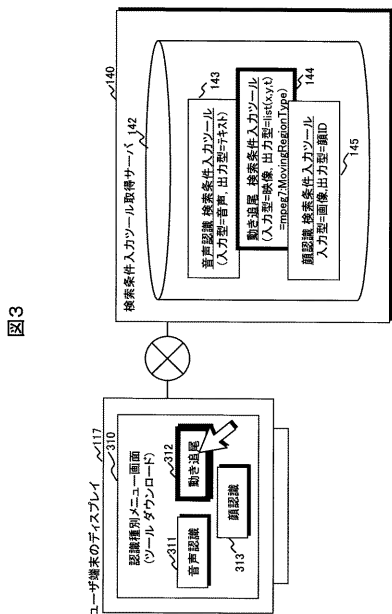


図3

【 図 4 】

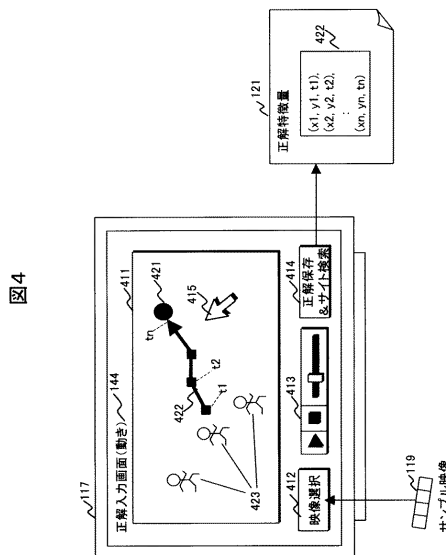
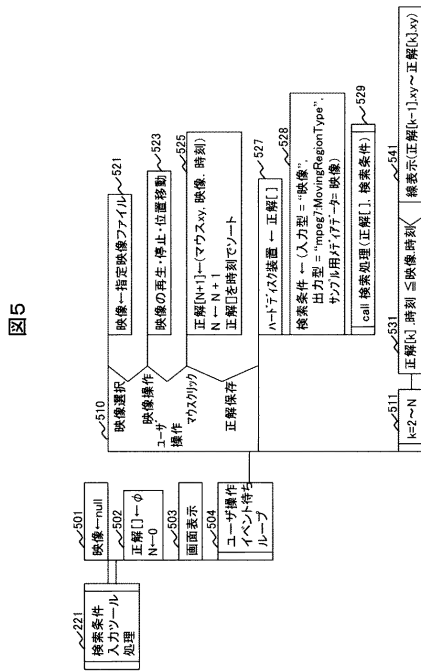
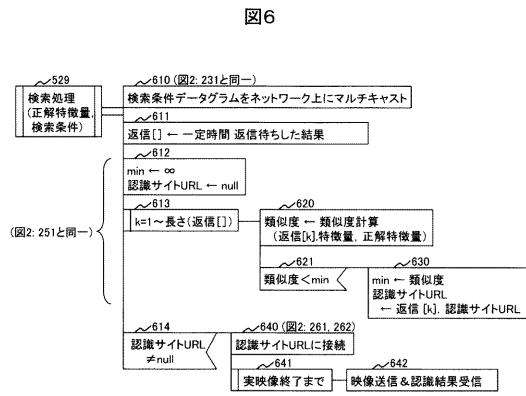


図4

【 図 5 】

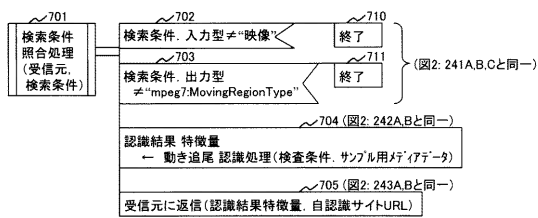


【 図 6 】



【 図 7 】

図7



フロントページの続き

(51)Int.Cl.

F I

G 1 0 L 3/00 5 7 1 A

(56)参考文献 特開平 1 1 - 1 7 5 5 5 3 (J P , A)
特開 2 0 0 0 - 1 9 4 8 6 1 (J P , A)
特開平 0 9 - 0 9 1 0 1 1 (J P , A)
特開 2 0 0 0 - 0 8 9 9 9 1 (J P , A)
特開 2 0 0 1 - 2 9 7 0 8 9 (J P , A)

(58)調査した分野(Int.Cl. , DB名)

G 0 6 F 1 9 / 0 0