(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第5866728号
(P5866728)

(45)発行日　平成28年2月17日(2016.2.17)　(24)登録日　平成28年1月15日(2016.1.15)

(51)Int.Cl.		F I		
G06F 13/00	(2006.01)	G06F	13/00	650B
G06F 17/30	(2006.01)	G06F	17/30	110C
G06Q 50/00	(2012.01)	G06F	17/30	310Z
G10L 15/28	(2013.01)	G06F	17/30	230Z
G10L 15/00	(2013.01)	G06F	17/60	

請求項の数 15　(全 73 頁)　最終頁に続く

(21)出願番号　特願2011-226792 (P2011-226792)
(22)出願日　平成23年10月14日 (2011.10.14)
(65)公開番号　特開2013-88906 (P2013-88906A)
(43)公開日　平成25年5月13日 (2013.5.13)
審査請求日　平成26年9月10日 (2014.9.10)

(73)特許権者　510003368
サイバーアイ・エンタテインメント株式会社
東京都世田谷区瀬田１－５－１
(74)代理人　100092093
弁理士　辻居　幸一
(74)代理人　100082005
弁理士　熊倉　禎男
(74)代理人　100067013
弁理士　大塚　文昭
(74)代理人　100086771
弁理士　西島　孝喜
(74)代理人　100109070
弁理士　須田　洋之

最終頁に続く

(54)【発明の名称】画像認識システムを備えた知識情報処理サーバシステム

(57)【特許請求の範囲】
【請求項１】
通信システムであって、
サーバ装置と、
第１の画像と、当該第１の画像に関連付けされている第１のメッセージと、当該第１の画像に関連付けされている場所の情報を少なくとも含む第１の情報とをネットワークを介して前記サーバ装置に送信する第１の装置と、
前記ネットワークを介して前記サーバ装置に接続されている第２の装置と、
を備え、
前記サーバ装置は、前記第１の画像に含まれている１以上の物体を特定し、前記第１のメッセージから１以上の単語を抽出し、当該１以上の単語を分析することによって前記第１の装置の第１のユーザが着目している物体を前記１以上の物体から特定し、前記第１のメッセージを当該着目物体に関連付けするよう構成され、
前記サーバ装置は、前記第１の画像と、前記第１のメッセージと、当該第１のメッセージが当該第１の画像における前記着目物体に関連付けされていることを示す情報とを前記第２の装置に前記ネットワークを介して送信するよう構成されている、通信システム。
【請求項２】
前記第２の装置は、当該第２の装置に関連付けされている場所の情報を少なくとも含む第２の情報を前記サーバ装置に送信するよう構成され、
前記サーバ装置は、前記第１の情報及び前記第２の情報に基づいて、前記第１の画像と

、前記第１のメッセージと、当該第１のメッセージが当該第１の画像における前記着目物体に関連付けされていることを示す情報とを前記第２の装置に前記ネットワークを介して送信することを決定するよう構成されている、請求項１に記載の通信システム。
【請求項３】
前記第１の装置は、前記第１の画像を送信した後に、前記第１のメッセージを前記サーバ装置に送信するよう構成されている、請求項１又は２に記載の通信システム。
【請求項４】
前記第２の装置は、第２のメッセージを前記ネットワークを介して前記サーバ装置に送信するよう構成されている、請求項１ないし３のいずれか一つに記載の通信システム。
【請求項５】
前記サーバ装置は、前記第１のメッセージ及び第２のメッセージを分析し、ユーザ間のインタレストグラフを取得するよう構成されている、請求項４に記載の通信システム。
【請求項６】
前記第１の情報は、第１の時間情報を更に含み、前記サーバ装置は、前記着目物体を当該第１の時間情報に関連付けるよう構成されている、請求項１ないし５のいずれか一つに記載の通信システム。
【請求項７】
前記サーバ装置は、前記第１の時間情報及び前記第１の画像を少なくとも用いて、アルバムを生成するよう構成されている、請求項６に記載の通信システム。
【請求項８】
前記第１の装置及び／又は前記第２の装置は、文字情報の書込み及び／又はユーザの声による語りかけによって、前記第１のメッセージ及び／又は第２のメッセージを入力するよう構成されている、請求項１ないし７のいずれか一つに記載の通信システム。
【請求項９】
前記第１の装置及び／又は前記第２の装置は、カメラ付き携帯電話を含む、請求項１ないし８のいずれか一つに記載の通信システム。
【請求項１０】
前記第１の装置及び／又は前記第２の装置は、一以上のマイクロフォン、一以上のイヤフォン、及び一以上の画像撮像素子（カメラ）を少なくとも有するヘッドセットと、当該ヘッドセットに接続されているネットワーク端末とを含み、当該ネットワーク端末は、前記ネットワークを介して前記サーバ装置に接続されている、請求項１ないし９のいずれか一つに記載の通信システム。
【請求項１１】
前記ヘッドセットは、２台以上の撮像視差を有するカメラ及び／又は対象物体までの深度（距離）を測定可能な三次元カメラを含む、請求項１０に記載の通信システム。
【請求項１２】
前記第１の装置及び／又は前記第２の装置は、さらに、生体認証（バイオメトリクス）センサを含み、これにより、ユーザ固有の生体識別情報を生体認証装置に問い合わせるよう構成されている、請求項１ないし１１のいずれか一つに記載の通信システム。
【請求項１３】
前記第１の装置、前記第２の装置及び／又は前記サーバ装置は、前記ヘッドセットの着脱を監視するよう構成されている、請求項１２に記載の通信システム。
【請求項１４】
前記第１の装置及び／又は前記第２の装置は、さらに、生体情報（バイタルサイン）センサを含み、これにより、当該生体情報を前記サーバ装置に送信するよう構成されている、請求項１ないし１３のいずれか一つに記載の通信システム。
【請求項１５】
サーバ装置であって、
第１の装置から、第１の画像と、当該第１の画像に関連付けされている第１のメッセージと、当該第１の画像に関連付けされている場所の情報を少なくとも含む第１の情報とを

ネットワークを介して受信し、
前記サーバ装置は、前記第１の画像に含まれている１以上の物体を特定し、前記第１のメッセージから１以上の単語を抽出し、当該１以上の単語を分析することによって前記第１の装置の第１のユーザが着目している物体を前記１以上の物体から特定し、前記第１のメッセージを当該着目物体に関連付けし、
前記第１の画像と、前記第１のメッセージと、当該第１のメッセージが当該第１の画像における前記着目物体に関連付けされていることを示す情報とを第２の装置に前記ネットワークを介して送信するよう構成されている、サーバ装置。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、ユーザの頭部に装着可能なヘッドセットシステムに組み込まれたカメラから得られる当該ユーザの主観的な視野を反映した画像信号を、当該ユーザのネットワーク端末経由でネットワークを介して画像認識システムを備えた知識情報処理サーバシステム側に適宜アップロードする事で、当該ユーザが関心を持って着目した特定物体、一般物体、人、写真、或いはシーン等の１以上の対象（以降「対象」と呼称）が、上記カメラ映像中のいずれに当るのかを、前記サーバシステムと当該ユーザ間の音声による双方向のコミュニケーションにより抽出可能にした上で、それら対象の抽出過程及び画像認識結果を、上記サーバシステム側が当該ユーザのネットワーク端末経由で、上記ヘッドセットシステムに組込まれたイヤフォンを通し、当該ユーザに対し音声情報により通知する事を特徴とする。
【０００２】
その上で当該ユーザが着目する様々な対象に対し、当該ユーザの音声によるメッセージやつぶやき、或いは質問等の音声タグを残す事を可能にする事で、異なる時空間内において自らを含む様々なユーザが当該対象に偶然遭遇する、或いはそれら対象を偶然目にした時に、前記サーバシステム側に蓄積された当該対象に係る様々なメッセージやつぶやき群を、当該対象への着目に同期して音声で受取る事を可能にし、それら個々のメッセージやつぶやきに対し、ユーザがさらなる音声応答を返す事を可能にする事で、様々なユーザの共通の着目対象に係る広範なソーシャル・コミュニケーションを喚起する事を特徴とする。
【０００３】
その上で、当該喚起された多数のユーザの視覚的関心に端を発する広範なソーシャル・コミュニケーションを、前記サーバシステム側で継続的に収集・解析・蓄積する事で、広範なユーザ、様々なキーワード、及び様々な対象を構成ノード群とする、動的なインタレストグラフとして獲得可能にし、それらを基に高度にカスタマイズされたサービスの提供、精度の高いリコメンデーションの提示、或いは動的な広告や告知等への効果的な情報提供サービスに繋げる事を可能にする、前記画像認識システムを備えた知識情報処理サーバシステムに関する。
【背景技術】
【０００４】
近年のインターネットの世界的な普及により、ネットワーク上の情報量が急激に増大しつつある事から、それら膨大な量の情報の海の中から目的とする情報を効果的且つ高速に探し出す手段としての検索技術が急速に進歩して来た。現在では、強力な検索エンジンを備えたポータルサイトがいくつも運営されている。また、閲覧者の検索キーワードやアクセス履歴等を解析し、閲覧者の嗜好にあったＷｅｂページや広告等を各々の検索結果に関連して配信する技術も開発され、閲覧者が多用するキーワードに基づく効果的なマーケティング活動等への応用も始まっている。
【０００５】
例えば、ユーザにとって有用な情報を精度良く且つ容易に提供する事が出来る情報提供装置がある（特許文献１）。この情報提供装置は、ユーザによる各コンテンツに対するア

クセスの頻度を表すアクセス頻度情報を、当該ユーザを識別するユーザ識別情報に対応付けて格納するアクセス履歴格納手段と、各ユーザ間における各コンテンツへのアクセス傾向の類似性を表すユーザ間類似度を、前記アクセス履歴格納手段に格納された前記アクセス頻度情報に基づいて算出するユーザ間類似度計算手段と、ユーザと各ユーザとの間の前記ユーザ間類似度により重み付けした、当該各ユーザの前記アクセス頻度情報から、当該ユーザにとってのコンテンツの有用度を表す情報であるコンテンツ・スコアを算出するコンテンツ・スコア計算手段と、前記コンテンツ・スコア計算手段によって算出された各コンテンツの前記コンテンツ・スコアを、前記ユーザ識別情報に対応付けて記憶するインデックス格納手段と、通信端末装置から送信されたユーザ識別情報を含むクエリの入力を受け付けるクエリ入力手段と、前記クエリ入力手段により受け付けられた前記クエリに適合するコンテンツのコンテンツ識別情報を取得し、当該クエリに含まれるユーザ識別情報に対応付けられて前記インデックス格納手段に記憶された前記コンテンツ・スコアを参照して、取得した前記コンテンツ識別情報から提供情報を生成する提供情報生成手段と、前記提供情報生成手段により生成された前記提供情報を、前記通信端末装置に出力する提供情報出力手段とを備える事を特徴とする、情報提供装置である。

【０００６】

これらのキーワード等の文字情報を検索クエリとする検索手段をさらに拡大する目的で、画像認識技術を備えた検索エンジンの開発が近年進み、文字に代わり画像そのものを入力クエリとする画像検索サービスが、広くインターネット上で提供される様になって来ている。画像認識技術の研究の始まりは、一般に４０年以上前に遡る事が出来る。以来、コンピュータの高速化と機械学習技術の進歩と共に、線画解釈（１９７０年代）、人手によるルールや幾何形状モデルによって構築された知識データベースに基づく認知モデル、３次元モデル表現（１９８０年代）といった研究が漸次行われる様になった。１９９０年代に入ると、特に顔画像の認識や学習による認識に関する研究が盛んになった。２０００年代になると、コンピュータの処理能力の一層の向上により、統計処理や機械学習の為に必要となる膨大な計算処理が比較的安価に実行可能になった為、一般物体認識に関する研究が進んだ。一般物体認識とは、実世界のシーンを撮影した画像に対して、コンピュータがその画像中に含まれる物体を一般的な名称で認識する技術である。１９８０年代には、全て人手によってルールやモデルの構築を試みていたが、大量のデータを手軽に扱える様になったこの時期には、コンピュータを活用した統計的機械学習によるアプローチが注目され、近年の一般物体認識ブームのきっかけとなった。一般物体認識技術によって、画像に対するキーワードを対象画像に自動的に付与する事が可能になり、画像をその意味内容によって分類及び検索する事も可能になる。近い将来には、コンピュータによって全ての人間の画像認識機能を実現する事が目標とされている（非特許文献１）。一般物体認識技術は、画像データベースからのアプローチと統計的確率手法の導入によって急速に進歩した。その中でも先駆的な研究として、画像に人手でキーワードを付与したデータから個々の画像との対応付けを学習し物体認識を行なう手法（非特許文献２）や、局所特徴量に基づく手法（非特許文献３）等がある。また、局所特徴量による特定物体認識に関する研究にＳＩＦＴ法（非特許文献４）、及びＶｉｄｅｏ　Ｇｏｏｇｌｅ（非特許文献５）等がある。その後、２００４年に入り、「Ｂａｇ－ｏｆ－Ｋｅｙｐｏｉｎｔｓ」あるいは「Ｂａｇ－ｏｆ－Ｆｅａｔｕｒｅｓ」と呼ばれる手法が発表された。この手法は、対象となる画像をビジュアル・ワード（ｖｉｓｕａｌ　ｗｏｒｄ）と呼ばれる代表的な局所パターン画像片の集合として扱い、その出現頻度を多次元のヒストグラムで表現する。具体的には、ＳＩＦＴ法に基づいた特徴点抽出を行い、予め求められた複数のビジュアル・ワードに基づいてＳＩＦＴ特徴ベクトルをベクトル量子化し、画像毎にヒストグラムを生成するものである。この様に生成されたヒストグラムの次元数は、通常、数百から数千次元のスパース（ｓｐａｒｓｅ）なベクトルになる。そして、これらのベクトルは、コンピュータ上の多次元ベクトルの分類問題として高速に処理される事により、一連の画像認識処理が行われる（非特許文献６）。

【０００７】

これらコンピュータによる画像認識技術の進展に伴い、カメラ付きネットワーク端末で撮影した画像を、ネットワーク経由でサーバ側に構築された画像認識システム側に問い合わせ、当該サーバ側に蓄積された膨大な画像データベースを基に、当該画像認識システム側がそれらの画像と、予め学習済みの物体毎の特徴を記述した画像特徴データベース群とを比較照合する事で、アップロードされた画像に含まれる主要な物体を画像認識し、その認識結果を前記ネットワーク端末側に速やかに提示するサービスが既に始まっている。画像認識技術の中でも特定の人間の顔の検出技術は、個々人を特定する手法の一つとして急速に応用開発が進んでいる。多数の顔画像の中から特定の人物の顔を精度良く抽出する為には、膨大な顔画像の事前学習が必要となる。その為に準備しなくてはならない知識データベースの量も極めて大きくなる事から、或る程度大規模な画像認識システムの導入が必要になる。一方、電子カメラにおけるオートフォーカスに用いられる様な一般的な「平均顔」の検出、或いは限られた人物の顔の特定であれば、電子カメラ等の小型の筐体内に十分収まる規模のシステムで今や容易に実現が可能である。また、近年供用が始まったインターネットを利用した地図提供サービスの中で、地図上の要所々々における路上写真（Ｓｔｒｅｅｔ　Ｖｉｅｗ）を居ながらにして俯瞰する事が出来る様になった。この様なアプリケーションでは、プライバシー保護の観点から偶然写り込んだ自動車のナンバープレートや歩行者の顔、或いは道路越しに垣間見えてしまう個人宅の様子等を、一定以上判別出来ない程度にフィルタ処理して再表示する必要性も出て来ている（非特許文献７）。

【０００８】

近年、現実空間を拡張して、コンピュータによる情報空間としてのサイバー空間とを相互に融合しようとする拡張現実感（Ａｕｇｍｅｎｔｅｄ　Ｒｅａｌｉｔｙ：略称ＡＲ）というコンセプトが提案され、既に一部のサービスが始まっている。一例として、ＧＰＳや無線基地局等から取得可能な位置情報を利用した三次元位置測位システム、カメラ、及び表示装置等を一体として備えたネットワーク携帯端末を用い、上記三次元位置測位システムから割り出した自身の位置情報を基に、カメラで撮影した現実世界の映像と、サーバ上にデジタル情報として蓄積されている注釈（アノテーション：Ａｎｎｏｔａｔｉｏｎ）とを重ね合わせ、サイバー空間に浮かぶエアタグ（Ａｉｒｔａｇ）として現実世界の映像に貼り付ける事が可能になっている（非特許文献８）。

【０００９】

１９９０年代後半になると、通信ネットワーク・インフラの整備拡張に伴い、インターネット上に構築されたユーザ相互の社会的関係を促進する目的で、ソーシャルネットワークに係るサイトが数多く開設され、数々のソーシャル・ネットワーキング・サービス（ＳＮＳ）が生まれた。ＳＮＳにおいては、ユーザ検索機能、メッセージ送受信機能、掲示板等のコミュニティ機能によって、ユーザ間のコミュニケーションが有機的に促進される。例えばＳＮＳのユーザは、趣味・嗜好を同じくするユーザが集う掲示板に積極的に参加して、文書や画像、音声等のパーソナル情報を交換し、また自分の友人を他の知人に紹介する事等により、人と人との相互の繋がりをさらに深め、ネットワーク上でコミュニケーションを有機的かつより広範に広げていく事が出来る。

【００１０】

ＳＮＳにおけるサービスの一形態として、ネットワーク上にアップロードされた動画を複数のユーザが選択共有し、当該動画シーン上の任意の位置にユーザが自由に当該動画内容に関連するコメントをアップロードする事を可能にし、それらコメント群を当該動画面上にスクロール表示する事で、複数のユーザ間で当該動画を媒介とした共有コミュニケーションを図る事が可能なコメント付き動画配信システムがある（特許文献２）。当該システムは、コメント情報をコメント配信サーバから受信し当該共有動画の再生を開始すると共に、当該コメント情報から再生する動画の、特定の動画再生時間に対応するコメントをコメント配信サーバから読み出し、読み出したコメント群に対応付けられた動画再生時間に、当該動画と共にそれらコメント群を表示可能にする。併せて、それらコメント情報をリストとしても個別に表示可能にし、表示されたコメント情報から特定のコメントデータが選択されると、選択されたコメントデータのコメント付与時間に対応する動画再生時間

から当該動画を再生し、読み出したコメントデータを表示部に再表示させる。また、ユーザによるコメントの入力操作を受け付けて、コメントが入力された時点の動画再生時間をコメント付与時間として、コメント内容と共に前記コメント配信サーバに送信する。

【００１１】

ＳＮＳの中でも、ネットワーク上で交換可能な情報パケットサイズを大幅に限定する事で、コミュニケーションのリアルタイム性をより重視しようという動きもある。これらマイクロブログとも呼ばれるユーザの短いつぶやきや、それらに関連するＵＲＬ等のアドレス情報を埋め込んだ１４０文字以内の文字データを、当該ユーザがインターネット上にリアルタイム且つ広範に発信する事で、当該ユーザのその時々の体験を当該ユーザの文字によるつぶやきのみならず、画像や音声データを加えた一体的な情報として広範なユーザ間で共有可能にし、さらにユーザがそれらつぶやきの中から特定の発信者や特定の話題を選択してフォローする機能も提供する事で、地球規模でのリアルタイム・コミュニケーションを喚起するサービスが既に始まっている（非特許文献９）。

【００１２】

ネットワークを介した情報サービスとは異なるものの、特定の対象に対峙した時に当該対象に関する詳細な音声説明を受取る事が出来るサービスとして、博物館や美術館の「音声ガイド」システムがある。これらは、対象となる絵画等の近傍に設置された音声信号送出部から送出される赤外線変調された音声信号を、それら対象物に近接したユーザの端末装置に組込まれた赤外線受信部で復調し、当該ユーザのイヤフォンに当該絵画等に係る詳細な説明を音声として提供するもので、この方式以外にも極めて指向性の高い音声トランスミッターを用いて、ユーザの耳元に直接当該音声情報を送り込める様な音声ガイドシステムも実用化されている。

【００１３】

コンピュータ・システムに対する音声による情報入力やコマンド入力方法として、ユーザの発話音声を音声言語として認識し、テキストデータや各種のコンピュータコマンドに変換して入力処理する技術がある。当該入力処理には高速の音声認識処理が必要となるが、これらを可能にする音声認識技術群として、音響処理技術、音響モデル作成・適応化技術、適合・尤度演算技術、言語モデル技術、対話処理技術等があり、これらの要素技術をコンピュータ上で組み合わせる事で、近年では十分実用に耐える高速の音声認識システムが構築可能となっている。近年では、大規模語彙連続音声認識エンジンの開発によって、ユーザにより発話される音声言語認識処理を、ネットワーク端末上でほぼ実時間で処理する事も可能となっている。

【００１４】

音声認識技術の研究の歴史は、１９５２年に米国のベル研究所でのゼロ交差回数を用いた数字認識の研究に始まり、１９７０年代に入ると発声時間の長さの変動を、動的計画法を用いて非線形に正規化する手法（Ｄｙｎａｍｉｃ　Ｔｉｍｅ　Ｗａｒｐｉｎｇ）が日本及びロシアの研究者によって提案され、米国においても統計確率的手法であるＨＭＭ（Ｈｉｄｄｅｎ　Ｍａｒｋｏｖ　Ｍｏｄｅｌ：隠れマルコフモデル）を用いた音声認識の基礎的な研究が進んだ。現在では、利用者の音声の特徴を適応的に学習させる事より、明瞭な発声で読み上げられた文章をほぼ完全に口述筆記する事が可能なレベルにまで到達している。この様な高度の音声認識技術を応用した従来技術として、会議による発言音声を入力とする話し言葉から、文語としての議事録を自動作成する技術も開発されている（特許文献３）。

【００１５】

すわなち、特許文献３に開示された技術は、音声を入力して文書情報を作成し出力する音声文書変換装置であり、文書情報出力を受信して画面に表示する表示装置を備え、この音声文書変換装置が、入力する音声を認識する音声認識部と、入力音声を漢字仮名混じりの文語に変換する変換テーブルと、前記音声認識部から認識した音声を受信して整列させ前記変換テーブルを検索して文語に変換し所定の書式で文書に編集する文書形成部と、この編集済み文書を記憶保存する文書メモリと、この保存された文書情報を送信すると共に

他の情報・信号を前記表示装置との間で授受する送受信部とを有し、かつ前記表示装置が前記音声文書変換装置の送受信部との間で情報・信号を送受信する送受信部と、受信した文書情報を表示情報として記憶する表示情報メモリと、この記憶する表示情報を画面表示する表示盤とを有する事を特徴としている。

【００１６】

また、コンピュータ上の文字情報からなる文章を、指定された言語で流暢に読み上げる音声合成システムは、近年最も進化の進んでいる領域の一つである。音声合成システムは、スピーチ・シンセサイザー（Ｓｐｅｅｃｈ　Ｓｙｎｔｈｅｓｉｚｅｒ）とも呼ばれ、テキストを音声に変換するテキスト読み上げシステムや、発音記号を音声に変換するシステム等を含む。歴史的には、１９６０年代末以降、コンピュータによる音声合成システムの開発が進んだものの、初期のスピーチ・シンセサイザーによる発声はいかにもコンピュータによる音声だと感じさせる人間味のない無機質なものが多かった。以降研究が進むにつれ、後述する様に、場面、状況、前後の文脈関係により声の抑揚や調子を自在に変化させる事が出来る様になり、人間の肉声と比べてほとんど遜色がない高品質の音声合成が可能になっている。特に、サーバ側に構築された音声合成システムは、膨大な辞書を活用可能なばかりではなく、その発声アルゴリズム自体も人間に近い複雑な発音が可能な様に多数のデジタルフィルタ類を組み込む事も可能になり、ネットワーク端末機器の急速な普及に伴い、近年その応用可能な範囲が一段と拡大している。

【００１７】

音声合成技術には、大きく分けてフォルマント合成と連結的合成とがある。フォルマント合成では、人間の音声を使用する事なく周波数や音色等のパラメータをコンピュータ上で調整して人工的な合成波形を生成する。これらは一般的に人工的な音声として聞こえる場合が多い。一方で連結的合成では、基本的に人間の音声を収録して、その音素断片等を滑らかに連結して肉声に近い音声を合成する方法である。具体的には、一定時間収録された音声を「音」「音節」「形態素」「単語」「成句」「文節」等に分割してインデックス化し、検索可能な音声ライブラリ群を作成する。こうした音声ライブラリは、テキスト読み上げシステム等により音声を合成する際に、適宜最適な音素や音節等が抽出され、適切なアクセントと共に最終的に人間の発話に近い流暢な一連の音声に変換される。

【００１８】

係る従来技術に加え、声調機能を備えたテキスト読み上げシステム等の開発により、バリエーションに富んだ音声を合成する技術も続々実用化されている。例えば、高度な音声編成システムによって、アクセント調整や音の高低・長さの調整を行う事によって、「うれしさを伴った声」「悲しみを伴った声」「怒りを伴った声」「冷たさを伴った声」等の感情の抑揚を調整する事が出来る他、音声編成システムが備えるデータベースに登録された特定の人のクセを反映した音声を、これらシステム上で自在に合成する事も出来る様になっている。

【００１９】

また、上述した音声合成についての先行技術に、合成音声区間と部分的に一致する肉声区間を検出して、その肉声区間の韻律（抑揚・リズム）情報を合成音声に付与し、肉声と合成音声を自然に結合させる技術も提案されている（特許文献４）。

【００２０】

即ち、特許文献４に開示された技術は、録音音声格納手段、入力テキスト解析手段、録音音声選択手段、接続境界算出手段、規則合成手段、接続合成手段に加えて、合成音声区間のうちで録音済みの肉声と部分的に一致する区間を決定する肉声韻律区間決定手段と、その一致部分の肉声韻律を抽出する肉声韻律抽出手段と、抽出された肉声韻律を使って合成音声区間全体の韻律情報を生成する、ハイブリッド韻律生成手段を備える事を特徴としている。

【先行技術文献】

【特許文献】

【００２１】

【特許文献１】特開２００９－２６５７５４号公報
【特許文献２】特開２００９－０７７４４３号公報
【特許文献３】特開１９９３－０１２２４６号公報
【特許文献４】特開２００９－０２０２６４号公報
【非特許文献】
【００２２】
【非特許文献１】柳井啓司, "一般物体認識の現状と今後", 情報処理学会論文誌, Vol.48, No.SIG16(CVIM19), pp.1-24, 2007
【非特許文献２】Pinar Duygulu, Kobus Barnard, Nando de Freitas, David Forsyth, "Object Recognition as Machine Translation: Learning a lexicon for a fixed image vocabulary," European Conference on Computer Vision (ECCV), pp.97-112, 2002.
【非特許文献３】R. Fergus, P. Perona, and A. Zisserman, "Object Class Recognition by Unsupervised Scale-invariant Learning," IEEE Conf. on Computer Vision and Pattern Recognition, pp.264-271, 2003.
【非特許文献４】David G.Lowe, "Object Recognition from Local Scale-Invariant Features," Proc. IEEE International Conference on Computer Vision, pp.1150-1157, 1999.
【非特許文献５】J. Sivic and A. Zisserman, "Video google: A text retrieval approach to object matching in videos", Proc. ICCV2003, Vol. 2, pp.1470-1477, 2003.
【非特許文献６】G. Csurka, C. Bray, C. Dance, and L. Fan, "Visual categorization with bags of keypoints," Proc. ECCV Workshop on Statistical Learning in Computer Vision, pp.1-22, 2004.
【非特許文献７】Ming Zhao, Jay Yagnik, Hartwig Adam, David Bau; Google Inc. "Large scale learning and recognition of faces in web videos" FG '08: 8th IEEE International Conference on Automatic Face & Gesture Recognition, 2008.
【非特許文献８】http://jp.techcrunch.com/archives/20091221sekai-camera/
【非特許文献９】Akshay Java, Xiaodan Song, Tim Finin, and Belle Tseng, “Why We Twitter: Understanding Microblogging Usage and Communities” Joint 9th WEBKDD and 1st SNA-KDD Workshop ’07.
【発明の概要】
【発明が解決しようとする課題】
【００２３】
しかしながら、従来の検索エンジンにおいては、検索対象に係るいくつかのキーワードを考え文字で入力する必要があった。それらの検索結果は、複数、時に夥しい数の候補群に係る文書タイトルと共に概略記述文章として提示される事から、目的とする検索結果に辿り着く為には、各候補群が示す情報の格納先をさらに個々に開いて読み進んでいく必要があった。近年は画像を直接入力クエリとする検索も可能になり、その検索出力として関連度の高い画像そのものを一覧的に閲覧可能な画像検索サービスも提供され始めている。しかし、ユーザが関心を持った対象や事象に対し、その好奇心をさらに喚起する様な関連情報を、快適且つ的確にユーザに提供出来る迄には至っていない。また従来の検索プロセスでは、ＰＣやネットワーク端末等に向かって一時的ではあるにせよ集中的な入力操作を行う必要がある事から、ユーザがハンズフリーで何か別の事をしながら日常の生活の中でふと誰かに語りかけ、身近な誰かが答えてくれる様な、普段我々が何気なく行っている自然なコミュニケーションが、従来のＩＴシステム上ではまだ実現出来ていない。
【００２４】
一例として、ユーザがふと調べたいと思った対象や事象に遭遇した場合、その名称等が判る場合には文字入力によるネットワーク検索を行うか、カメラ付き携帯電話やスマートフォン等を手に当該対象に近付き、当該ネットワーク端末に具備されているカメラで撮影した後、当該撮影画像を基に画像検索をかけるケースが多い。それでも思う様な検索結果が得られない場合は、ネットワーク上の他のユーザへ当該対象を問い合わせる事も可能で

はある。しかし、これら一連のプロセスはユーザにとって少々煩雑であるだけではなく、対象に直接携帯電話等をかざす等の行為が必要な事から時に対象から身構えられる、場合によっては失礼だと感じさせる、さらには携帯電話をかざす行為自体に対し周りから不審な目で見られる、といった嫌いがあった。また対象が動物や人物等の場合、対象と自分との間にカメラ付き携帯端末等が入る事により一種の視覚的な壁の様なものが出来てしまう点と、検索結果を先ずは当該携帯端末で確認しようとする事から、一時的にせよ当該対象や周囲の人々とのコミュニケーションが中断しがちであった。また、これら一連の検索プロセスには相応の時間がかかる事から、ユーザが外出中にふと目にした物体や人、動物、或いはシーン等に関心を持ったとしても、その場で上記一連の操作が完結出来ない場合も多く、一旦撮影した写真を自宅等に持ち帰って改めてＰＣ等で検索し直す必要もあった。

【００２５】

近年、実用化が始まった拡張現実と呼ばれるサービスにおいて、我々が存在する現実の空間と、コンピュータネットワーク網の中に構成されるサイバー空間とを紐付ける手法の一つとして、ＧＰＳ等から得られる測位情報に加え、カメラが向いている方位情報を併せて利用する手法がある。しかしこれら位置情報のみの利用では、対象物体自体の移動や、そもそも対象が観測時点で存在していない等、刻々と変化する現実の世界の状況に際し対応が困難なケースが多い。基本的に位置情報と固定的に紐付いている様々な建造物や都市のランドマーク等とは異なり、車などの移動・可搬可能な物体や、動き回る人や動物、或いは「夕焼け」等の概念的なシーンに対しては、当該システム内に画像認識機能を有していない場合には、本質的な意味での相互の対応付けが困難となる。

【００２６】

ＳＮＳにおけるサービスの一形態として、近年ユーザの間で人気のあるコメント付き動画共有サービスにおいては、共有視聴される動画が録画済みの動画である場合には、現実の世界で進行中の事象に対してリアルタイムの共有体験が得られないという問題がある。これに対し、ライブストリーム映像配信に対応したコメント付与サービスが既に始まっている。対象となるストリーム映像としては、記者会見、発表会、国会中継、イベント、スポーツ等に加えて、一般ユーザの投稿によるライブ映像配信がある。これらの動画共有サービスにおいては、ネットワークを介してリアルタイムで進行中の事象に係る「場」の共有が可能となる。しかし、延々と続くライブストリーム映像配信をフォローするには時間及び忍耐が必要である。そこからユーザ固有の或いは参加しているユーザ群に共通の関心の在り所等を効果的・効率的に抽出し、それらをインタレストグラフとして広範に体系付ける素材群として見ると、その収集可能な対象及び情報量には一定の限界があった。これは利用者数が急増しているネットワーク共有動画視聴サービスでも同じで、ユーザが様々な動画ファイルを連続視聴する為に消費する時間、及び配信サーバやネットワーク回線に係るコストに対し、ユーザが能動的に何か有用な情報をサーバ側に提供出来るチャンスはそれ程多くない。

【００２７】

これに対し、１４０文字以内という一定の制限は課されるものの、そのネットワーク上を流れるリアルタイムのトピックスの多彩さと参加者の急増も手伝って、これらマイクロブログと呼ばれるリアルタイム・メッセージ交換サービスから抽出可能なユーザ固有の、或いは特定のユーザ間で共通の、或いは広範なユーザ間において共通の、リアルタイムに収集可能なインタレストグラフの有用性に注目が集まっている。しかしながら、従来のマイクロブログにおいては、ユーザがその時点で自らが関心を持った対象や状況に係るつぶやきが中心で、当該ユーザの近傍或いは視野内に存在する他のユーザの関心の対象に対しては、有効な気付きを十分与える事が出来ているとは言えない。これらマイクロブログにおけるつぶやきの内容は極めて多岐に亘る為、特定のユーザ、特定の話題、或いは特定の場所等を指定して、テーマやトピックスを絞り込む方向の機能は提供されているものの、逆にその関心の対象をさらに拡げて行く方向として、個々のユーザ特有の潜在的な関心の反映や、当該ユーザの身近に存在する他のユーザによる顕在的な関心の在り処の通知等、さらに広範なＳＮＳを誘発する可能性については、まだ十分生かし切れているとは言えな

い。

【課題を解決するための手段】

【００２８】

上記課題を解決するために、本発明に係るネットワーク・コミュニケーションシステムは、一形態として、インターネットに接続可能なネットワーク端末に対し、有線或いは無線で接続可能な多機能入出力デバイスであって、少なくとも一以上のマイクロフォン、一以上のイヤフォン、一以上の画像撮像素子（カメラ）を一体として有する、ユーザの頭部に装着可能なヘッドセットシステムから得られる当該ユーザの主観的な視野、及び視点を反映した画像、及び音声信号を、前記ネットワーク端末経由でインターネット上の前記画像認識システムを備えた知識情報処理サーバシステム側にアップロード可能にし、当該画像に内包されている当該ユーザが着目した特定物体、一般物体、人、写真、或いはシーンに対し、音声認識システムとの協調動作により、当該ユーザ自身の音声による当該着目対象の指定、選択、及び抽出操作を、前記サーバシステム上で可能にした上で、当該ユーザによる上記一連の画像認識プロセス及び画像認識結果を、音声合成システムとの協調動作により、前記サーバシステム側がインターネットを介し、当該ユーザのネットワーク端末経由で、当該画像認識結果及びその認識プロセスを当該ユーザのヘッドセットシステムに組込まれたイヤフォンに対し音声情報として、及び／又は、当該ユーザのネットワーク端末に音声及び画像情報として通知する事を可能にし、当該画像認識可能になった対象に対し、当該ユーザが自らの声で語りかけたメッセージやつぶやきを、前記音声認識システムとの協調動作により、前記サーバシステム側がその内容を分析・分類・蓄積し、それらメッセージやつぶやきをネットワーク経由で、同様の対象を目にした自らを含む広範なユーザ間で共有可能にする事で、多数のユーザの視覚的な好奇心に端を発する広範なネットワーク・コミュニケーションを誘発させると共に、それら広範なユーザ間のコミュニケーションを、前記サーバシステム側で統計的に観察・蓄積・解析する事で、当該ユーザ特有の、或いは特定のユーザ群に特有の、或いはユーザ全体に共通の動的な関心や好奇心の在り所とその推移を、上記広範な「ユーザ」群、抽出可能な「キーワード」群、及び様々な着目「対象」に係るノード群との間を繋ぐ動的なインタレストグラフとして獲得可能にする事を特徴とする。

【００２９】

また、前記ネットワーク・コミュニケーションシステムにおいて、ユーザが関心を持った着目対象がどの様な特徴を有しているか、及び／又は、どの様な位置関係にあるか、及び／又は、どの様な運動状態にあるかを、前記画像認識システムを備えた知識情報処理サーバシステム側にユーザが明示的に指し示す手段として、当該ユーザの音声による対象の選択指定（ポインティング）操作を可能にし、これら一連の選択指定の過程で当該ユーザが発声する当該対象に係る様々な特徴群を基に、前記音声認識システムとの協調動作により前記サーバシステム側が当該対象を正確に抽出・認識し、その画像認識結果に係る前記サーバシステム側から当該ユーザに向けての再確認内容として、当該ユーザが前記サーバシステム側に対し明示的に音声で指し示した特徴群以外に、当該ユーザの主観的視野を反映したカメラ映像を基に、前記サーバシステム側が当該対象に共起する新たな物体や事象群を抽出し、当該対象をさらに正確に言い表す事が可能な共起事象として加え、それらを一連の文章に構成し、前記音声合成システムとの協調動作により、当該ユーザに対し音声により再確認を求める事を可能にする事を特徴とする。

【発明の効果】

【００３０】

本発明は、ユーザの頭部に装着可能なヘッドセットシステムに組み込まれたカメラから得られるユーザの主観的な視野を反映した画像信号を、当該ユーザのネットワーク端末経由でネットワークを介し前記画像認識システムを備えた知識情報処理サーバシステム側に適宜アップロードする事で、当該ユーザが関心を持って着目した特定物体、一般物体、人、写真、或いはシーン等の１以上の対象（以降「対象」と呼称）が、前記カメラ映像中のいずれに当るのかを、前記サーバシステムと当該ユーザ間の音声による双方向のコミュニ

ケーションにより抽出可能にする事で、従来の画像認識システムが不得意として来たユーザの「主観」を反映した対象の抽出及び認識処理を可能にし、画像認識率そのものを向上させる効果を与えると同時に、そこにユーザの音声による対象指定（ポインティング）操作と、それに対するサーバ側からの音声による再確認という双方向のプロセスを組み入れる事で、当該画像認識システムに対し継続的な機械学習が可能となる。

【００３１】

また、ユーザによる前記音声指示を前記サーバシステム側で適宜解析する事で、当該対象に係る有用なキーワード群の抽出、及び当該ユーザによる当該対象に対する関心の抽出を可能にし、そこから広範なユーザ、様々なキーワード、及び様々な対象を構成ノード群とする、動的なインタレストグラフが獲得可能になる。

【００３２】

その上で、当該インタレストグラフの対象となるノード群をネットワーク上でさらに広範なユーザ、様々な対象、及び様々なキーワードに対し拡大取得する事により、当該インタレストグラフの対象領域のさらなる拡大に加え、その収集頻度をさらに高める事が出来る。これにより、コンピュータ・システムによる継続的な学習プロセスに、人類の「知」をより効果的に組み入れて行く事が可能となる。

【００３３】

また本発明は、前記画像認識システムを備えた知識情報処理システムにより認識可能になったユーザの着目対象に対し、当該ユーザが残した音声によるメッセージやつぶやきをネットワーク経由で前記サーバシステム内にアップロードし分類・蓄積しておく事で、異なる時空間において同様或いは類似の対象に近付いた、或いは着目した他のユーザ、或いはユーザ群に対し、前記サーバシステム側がネットワークを介し、当該ユーザのネットワーク端末経由で、前記メッセージやつぶやきを、当該ユーザとの音声コミュニケーションにより、インタラクティブに送り込む事を可能にする。これにより、多数のユーザに及ぶ様々な視覚的好奇心に端を発する広範なユーザコミュニケーションを、ネットワーク上で継続的に喚起する事が可能になる。

【００３４】

また、ユーザが様々な対象に対して残した前記メッセージやつぶやきに係る内容の解析及び分類を前記サーバシステム側でリアルタイムに実行する事で、当該サーバシステム内に保持されている前記インタレストグラフの記述を基に、当該メッセージやつぶやきに含まれる主たる話題を抽出し、当該抽出された話題を中心ノードとするさらに関連性の高い他の話題群を抽出し、それらを抽出された話題に関心の高い他のユーザ及びユーザ群と、ネットワークを介して相互に共有可能にする事で、広範なユーザが目にする様々な対象や事象に端を発したネットワーク・コミュニケーションを継続的に誘発する事が可能となる。

【００３５】

また本発明においては、当該ユーザ側から発した前記メッセージやつぶやきのみならず、当該サーバシステム自身側から発する様々な関心、好奇心、或いは疑問を当該ユーザ、或いはユーザ群に対し提起する事が出来る。例えば前記インタレストグラフ内に記載の対象ノード間の関連性から想定可能な範囲を超えて、特定のユーザが特定の対象に対して一定以上の関心を示す場合や、或いは逆に一定以下の関心しか示さない場合や、当該サーバシステム側だけでは認識が困難な対象や事象が存在した場合、或いはそれらに遭遇した場合等に、当該サーバシステム側から関連する質問やコメントを、当該ユーザ、或いは特定のユーザ群、或いは広範なユーザ群に対し積極的に提起する事を可能にする。これにより、前記サーバシステム側が様々な事象を通じて人類の「知」を継続的に吸収し、学習の上で自らの知識データベース内に体系立てて取り込んで行くプロセスが構成可能となる。

【００３６】

近年では超高速光ファイバー網によるネットワークのさらなる高速化と相俟って、巨大なデータセンタの敷設が進み、超並列演算可能なスーパーコンピュータの開発も一段と加速している事から、コンピュータ・システム自身の自動学習プロセスにおいて、そこに人

類の「知」が効果的、有機的、かつ継続的に加わって行く事で、ネットワークを介してこれらの高性能コンピュータ・システム群による様々な事象の自動認識、及び機械学習が急速に発展して行く可能性がある。その為には、人類の「知」をいかにコンピュータ側が効果的に取得し、ネットワークを介して広範に共有可能な「知」の体系として再利用可能な状態に整理して行けるかが重要となる。言い換えると、いかにコンピュータの「好奇心」を刺激し、人とのコミュニケーションの中で継続的にコンピュータ・システムが進化して行ける効果的な方法を見つけられるかが重要となる。本発明においては、これらサーバ側に構築されたコンピュータ・システム自身による学習を、広範な対象に対する人々の視覚的関心と直接結び付ける具体的な方法を与える。

【図面の簡単な説明】

【００３７】

【図１】本発明の一実施形態におけるネットワーク・コミュニケーションシステムの構成に関しての説明図である。

【図２】本発明の一実施形態におけるヘッドセットシステム及びネットワーク端末の構成に関しての説明図である。

【図３Ａ】本発明の一実施形態における音声による対象画像抽出処理に関しての説明図である。

【図３Ｂ】本発明の一実施形態における音声による対象画像抽出処理に関しての説明図である。

【図４Ａ】本発明の一実施形態における音声によるポインティングに関しての説明図である。

【図４Ｂ】本発明の一実施形態における学習によるグラフ構造の成長に関しての説明図である。

【図４Ｃ】本発明の一実施形態における複数対象候補の選択優先度処理に関しての説明図である。

【図５】本発明の一実施形態における知識情報処理サーバシステムの構成に関しての説明図である。

【図６Ａ】本発明の一実施形態における画像認識システムの構成に関しての説明図である。

【図６Ｂ】本発明の一実施形態における一般物体認識部の構成及び処理フローに関しての説明図である。

【図６Ｃ】本発明の一実施形態における一般物体認識システムの構成及び処理フローに関しての説明図である。

【図６Ｄ】本発明の一実施形態におけるシーン認識システムの構成及び処理フローに関しての説明図である。

【図６Ｅ】本発明の一実施形態における特定物体認識システムの構成及び処理フローに関しての説明図である。

【図７】本発明の一実施形態における生体認証手順に関する説明図である。

【図８Ａ】本発明の一実施形態におけるインタレストグラフ部の構成及び処理フローに関する説明図である。

【図８Ｂ】本発明の一実施形態におけるグラフデータベースの基本要素及び構成に関する説明図である。

【図９】本発明の一実施形態における状況認識部の構成及び一グラフ構造例に関する説明図である。

【図１０】本発明の一実施形態におけるメッセージ保管部の構成及び処理フローに関する説明図である。

【図１１】本発明の一実施形態における再生処理部の構成及び処理フローに関する説明図である。

【図１２】本発明の一実施形態におけるＡＣＬ（アクセス制御リスト）に関する説明図である。

【図１３Ａ】本発明の一実施形態におけるユースケース・シナリオに関する説明図である。
【図１３Ｂ】本発明の一実施形態における共通の対象への視覚的な好奇心に誘起されるネットワーク・コミュニケーションに関する説明図である。
【図１４】本発明の一実施形態におけるインタレストグラフに関するグラフ構造の説明図である。
【図１５】本発明の一実施形態における画像認識プロセスからのグラフ抽出手順に関する説明図である。
【図１６】本発明の一実施形態におけるインタレストグラフの獲得に関する説明図である。
【図１７】本発明の一実施形態における獲得されたインタレストグラフのスナップショットの一部に関する説明図である。
【図１８Ａ】本発明の一実施形態における時空間及び対象を指定可能なメッセージやつぶやきの記録と再生手順に関する説明図である。
【図１８Ｂ】本発明の一実施形態における時間／時間帯の指定手順に関する説明図である。
【図１８Ｃ】本発明の一実施形態における場所／地域の指定手順に関する説明図である。
【図１９】本発明の一実施形態におけるユーザが指定した時空間でのメッセージやつぶやきの再生手順に関しての説明図である。
【図２０】本発明の一実施形態におけるユーザの手指による対象指示手順に関する説明図である。
【図２１】本発明の一実施形態における視野の固定による対象指示の手順に関する説明図である。
【図２２】本発明の一実施形態における写真の検出手法に関する説明図である。
【図２３Ａ】本発明の一実施形態における対象との対話手順に関する説明図である。
【図２３Ｂ】本発明の一実施形態における会話エンジンの構成と処理フローに関する説明図である。
【図２４】本発明の一実施形態における複数のヘッドセットからの共有ネットワーク端末の利用に関する説明図である。
【図２５】本発明の一実施形態における音声によるＷｉｋｉ利用に関する処理手順の説明図である。
【図２６】本発明の一実施形態における位置情報を利用した誤差補正に関する説明図である。
【図２７】本発明の一実施形態における視点マーカーのキャリブレーションに関する説明図である。
【図２８】本発明の一実施形態におけるサーバとのネットワーク接続が一時的に切断されている状況におけるネットワーク端末単体での処理に関する説明図である。
【図２９】本発明の一実施形態における同一の時空間内に撮影された画像から抽出された特定物体、及び一般物体の事例である。
【図３０】本発明の一実施形態におけるアップロードされた画像に含まれる特定の時空間情報の抽出及び特定の時間軸の選択指定表示に関する説明図である。
【図３１】本発明の一実施形態における特定の時空間への視点移動時に特定の対象に係る会話を促す仕組みに関する説明図である。
【発明を実施するための形態】
【００３８】
以下、本発明の一実施形態を図１から図３１を用いながら説明する。
【００３９】
図１を用いて、本発明の一実施形態におけるネットワーク・コミュニケーションシステム１００の構成に関し説明する。前記ネットワーク・コミュニケーションシステムは、ヘッドセットシステム２００、ネットワーク端末２２０、知識情報処理サーバシステム３０

０、生体認証システム３１０、音声認識システム３２０、音声合成システム３３０から構成される。前記ヘッドセットシステムは１以上存在し、１以上の前記ヘッドセットシステムが１個の前記ネットワーク端末にネットワーク２５１で接続される。前記ネットワーク端末は１以上存在し、インターネット２５０に接続される。前記知識情報処理サーバシステムは、生体認証システム３１０、音声認識システム３２０、及び音声合成システム３３０と、各々ネットワーク２５２、２５３、及び２５４で接続される。前記生体情報処理システムは、インターネット２５０と接続されていても良い。本実施例におけるネットワークは専用回線であっても良いし、インターネットを含む公衆回線であっても良いし、公衆回線上にＶＰＮ技術を用いて仮想的な専用回線を構築したものであっても良い。以下、特に断らない限りネットワークを前記の通り定義する。

【００４０】

図２（Ａ）に、本発明の一実施形態におけるヘッドセットシステム２００の構成例を示す。前記ヘッドセットシステムは、図２（Ｂ）に示す様な、ユーザが装着する事で当該ネットワーク・コミュニケーションシステム１００を利用可能なインターフェース装置である。図１において、ヘッドセットシステム２００ａから２００ｃは、接続２５１ａから２５１ｃでネットワーク端末２２０ａに対し接続され、ヘッドセットシステム２００ｄから２００ｅは、接続２５１ｄから２５１ｅでネットワーク端末２２０ｂに対し接続され、ヘッドセットシステム２００ｆは、接続２５１ｆでネットワーク端末２２０ｃに接続されている。つまり、ヘッドセット２００ａから２００ｆは、ネットワーク端末２２０ａから２２０ｃを介して、インターネット経由で知識情報処理サーバシステム３００に繋がっている様子を表わしている。以下、ヘッドセットシステム２００と記載した場合にはヘッドセットシステム２００ａから２００ｆのいずれか一台を指す。ヘッドセットシステム２００ａから２００ｆは、全て同一機種である必要はない。同等の機能、或いは実施可能な最低限の機能を備えた同様の装置であれば良い。

【００４１】

ヘッドセットシステム２００は以下の要素群で構成されるが、これらに限らず、そのいくつかを選択して搭載しても良い。マイクロフォン２０１は１以上存在し、当該ヘッドセットシステムを装着したユーザの音声や、当該ユーザの周辺の音を収集する。イヤフォン２０２は１以上存在し、モノラル或いはステレオで、他のユーザのメッセージやつぶやき、サーバシステムからの音声による応答等を含む様々な音声情報を、当該ユーザに通知する。カメラ（画像撮像素子）２０３は１以上存在し、当該ユーザの主観的な視野を反映した映像以外に、ユーザの背後や側面、或いは上部等の死角となっているエリアからの映像も含んでも良い。また、静止画であるか動画であるかを問わない。生体認証センサ２０４は１個以上存在し、一実施例としてユーザの有用な生体識別情報の一つである静脈情報（鼓膜や外耳部から）を取得し、前記生体認証システム３１０と連携して、当該ユーザ、当該ヘッドセットシステム、及び前記知識情報処理サーバシステム３００間を、認証し紐付ける。生体情報センサ２０５は１以上存在し、ユーザの体温、心拍、血圧、脳波、呼吸、眼球移動、発声、体の動き等の検出可能な各種生体情報（バイタルサイン）を取得する。深度センサ２０６は、前記ヘッドセットシステムを装着したユーザに近付く、人間を含む或る程度以上の大きさの生体の移動を検知する。画像出力装置２０７は、前記知識情報処理サーバシステム３００からの各種通知情報を表示する。位置情報センサ２０８は、前記ヘッドセットシステムを装着したユーザの位置（緯経度、高度、向き）を検知する。一例として、当該位置情報センサに６軸モーションセンサ等を装備する事で、移動方向、向き、回転等を前記に追加して検出する様に構成しても良い。環境センサ２０９は、前記ヘッドセットシステム周辺の明るさ、色温度、騒音、音圧レベル、温湿度等を検知する。視線検出センサ２１０は、一実施例として前記ヘッドセットシステムの一部からユーザの瞳、又は網膜に向けて安全な光線を照射し、その反射光を計測する事で、ユーザの視線方向を直接検知する。無線通信装置２１１は、ネットワーク端末２２０との通信、及び前記知識情報処理サーバシステム３００との通信を行う。電源部２１２は、前記ヘッドセットシステム全体に電力を供給する為の電池等を指すが、有線で前記ネットワーク端末に接続可能

な場合は、外部からの電力供給によっても良い。

【００４２】

図２（Ｃ）に、本発明の一実施形態におけるネットワーク端末２２０の構成例を示す。図１において、ネットワーク端末２２０ａから２２０ｆは広くユーザが利用するクライアント端末装置であり、ＰＣ、携帯情報端末（ＰＤＡ）、タブレット、インターネット接続可能な携帯電話、スマートフォン等が含まれ、これらがインターネットに接続されている様子を表している。以下、ネットワーク端末２２０と記載した場合には、インターネットに接続されたネットワーク端末２２０ａから２２０ｆのいずれか一台を指す。ネットワーク端末２２０ａから２２０ｆは同一機種である必要はない。同等の機能、或いは実施可能な最低限の機能を備えた端末装置であれば良い。

【００４３】

ネットワーク端末２２０は以下の要素群で構成されるが、これらに限らずそのいくつかを選択して搭載しても良い。操作部２２１は、表示部２２２と共にネットワーク端末２２０のユーザインターフェース部である。ネットワーク通信部２２３は、インターネットとの通信、及び１以上のヘッドセットシステムとの通信を担当する。前記ネットワーク通信部は、ＩＭＴ－２０００、ＩＥＥＥ８０２．１１、Ｂｌｕｅｔｏｏｔｈ、ＩＥＥＥ８０２．３、或いは独自の有線／無線規格、及びルータを経由したその混合形態であっても良い。認識エンジン２２４は、知識情報処理サーバシステム３００の主要な構成要素である画像認識システム３０１が有する画像認識処理機能から、限定された対象に関する画像認識処理に特化した前記ネットワーク端末に最適化した画像認識プログラムを前記知識情報処理サーバシステム側からダウンロードし実行する。これにより、前記ネットワーク端末側にも一定の範囲内で画像検出・認識機能の一部を持たせる事で、前記サーバ側の画像認識システム側に対する処理負担の軽減、及びネットワーク回線の負荷の軽減を図る事が出来ると共に、その後のサーバ側での認識プロセスに際し、後述の図３Ａにおけるステップ３０－２０から３０－３７に対応する予備的な前処理を実行する事が可能となる。同期管理部２２５は、ネットワークの不具合等により回線の一時的な切断が発生し、再び回線が復帰した際にサーバ側との同期処理を行う。ＣＰＵ２２６は中央処理装置であり、記憶部２２７は主メモリ装置であり、又フラッシュメモリ等を含む一次、及び二次記憶装置である。電源部２２８は、当該ネットワーク端末全体に電力を供給する為の電池等の電源である。これらネットワーク端末は、ネットワーク網に対し緩衝的な役割を果たす。例えば、ユーザにとって重要ではない情報をネットワーク側にアップロードしても、それは知識処理サーバシステム３００にとっては当該ユーザとの紐付けという意味ではノイズであり、ネットワーク回線に対しても不要なオーバーヘッドとなる。従って、可能な範囲で或る程度のスクリーニング処理をネットワーク端末側で行う事で、ユーザに対する有効なネットワークバンド幅の確保や、ローカリティが高い処理に関し応答速度の向上を図る事が可能になる。

【００４４】

図３Ａを用いて、本発明の一実施例としてユーザが関心を持った対象に着目する際のユーザの音声による対象画像抽出処理３０－０１のフローを説明する。前記で定義した様に本実施例では特定物体、一般物体、人、写真、或いはシーンを「対象」と総称する事にする。前記対象画像抽出処理は、ステップ３０－０２のユーザによる音声入力トリガで始まる。前記音声入力トリガには、特定の言葉や一連の自然言語を用いても良いし、音圧レベルの変化を検出する事によりユーザの発声を検出しても良いし、またネットワーク端末２２０上のＧＵＩ操作によっても良い。前記ユーザの音声入力トリガによりユーザのヘッドセットシステムに具備されているカメラの撮影が開始され、そこから取得可能になる動画像、連続した静止画、或いは静止画を、前記知識情報処理サーバシステム３００に対しアップロードを開始し（３０－０３）、その後ユーザからの音声コマンド入力待ち状態（３０－０４）に入る。

【００４５】

一連の対象画像抽出、及び画像認識処理フローは、音声認識処理、画像特徴抽出処理、

着目対象抽出処理、そして画像認識処理の順番で実行される。具体的には、音声入力コマンド待ち（３０－０４）からユーザの発話を認識し、当該音声認識処理によりユーザの発声した一連の言葉から単語列を抽出し、当該単語列に基づいて画像の特徴抽出処理を行い、抽出可能になった画像特徴群を基に画像認識処理を実行し、対象が複数に亘る場合や、対象自体からの特徴抽出が困難である場合等に、ユーザに対しさらなる画像特徴群の入力を求める事で、ユーザが着目した対象をサーバ側がより確実に認識するプロセスを構成する。上記ユーザの発話による「再確認」のプロセスを加える事で、画像認識システムの全ての処理プロセスをコンピュータ・システム側のみで対処しなくてはならないという従来の発想を転換して、従来画像認識システムが不得意として来た対象画像の正確な抽出、或いは従来の音声認識システムが不得意として来た同音異義語への対応問題等への効果的な対処が可能になる。実際の導入に当たっては、これらの一連の画像認識プロセスを、いかにユーザにとり煩わしい作業と思わせずに楽しいコミュニケーションと思わせられるかが重要となる。前記一連の画像特徴抽出処理では、図３Ａに示す事例よりもさらに多様な画像特徴群に対応する画像特徴抽出処理部群を多数並列に配置して一気に並列処理する事が可能で、それにより画像認識精度の一層の向上と併せて処理の大幅な高速化を図る事が可能となる。

【００４６】

ユーザの音声による対象のポインティング方法としては、当該ステップ３０－０６から３０－１５で例示した様な、各画像特徴群に対しユーザがそれらを各々単独に選択しながらポインティングして行く事例より、複数の画像特徴群を含んだ一連の言葉として一括してポインティングする事例の方が多いものと想定される。この場合は、複数の画像特徴群による対象の抽出処理が同時並列に行われ、そこから当該対象を表現する複数の画像特徴要素群が得られる可能性が高い。そこからより多くの特徴が抽出可能になれば、当該着目対象のポインティングの確度は一段と高まる。それら抽出可能になった画像特徴群を手掛かりに、前記画像認識システムによる画像認識処理３０－１６が開始される。画像認識は、一般物体認識システム１０６、特定物体認識システム１１０、及びシーン認識システム１０８により実行される。図３Ａでは、これらを連続したフローで表現しているが、当該画像認識処理は各々並列、或いは各一般物体認識、特定物体認識、及びシーン認識処理の中でさらに並列化する事が可能で、当該画像認識処理の認識速度に係る処理時間を大幅に短縮する事が出来る。上記の結果として、当該画像認識された対象に係る様々な認識結果を、音声で当該対象に係る画像認識結果として、ユーザに通知する事が可能になる。

【００４７】

この場合であっても、上記画像認識結果に加えて当該ユーザが指し示した特徴要素群のみを引用してユーザに再確認を求めたとしても、果たしてそれで本当にユーザが着目した対象をシステム側が正しく抽出したのか疑問が残る場合もある。例えば、ユーザの視野を反映したカメラ画像の中には、類似の物体が複数存在している可能性もある。本特許では、当該不確実性に対応する為、前記画像認識システムを備えた知識情報処理サーバシステム側が、当該対象の近傍状況を、当該カメラ映像を基に精査する事で当該対象と「共起」している新たな物体や事象を抽出（３０－３８）し、当該ユーザが明示的に指し示していないそれら新たな特徴要素群を上記再確認の要素に加え（３０－３９）、当該ユーザに対し音声による再確認（３０－４０）を求める事で、ユーザの着目対象と上記サーバシステム側が抽出した対象が同一である事を再確認する事を可能に構成することが出来る。

【００４８】

上記一連の処理は、基本的に同一の対象に関する処理であり、ユーザはその行動において常に他の対象に興味を移行し得るので、図３Ａにおける前記ステップ群を包含するさらに大きな外側の処理ループも存在する。なお、前記画像認識処理ループは、前記ヘッドセットシステムをユーザが装着した時点で開始しても良いし、ステップ３０－０２同様の音声トリガによっても開始しても良いし、前記ネットワーク端末を操作する事によって開始しても良いが、必ずしもそれらには限らない。前記処理ループの停止は、前記処理ループの開始における手段と同様に、前記ヘッドセットをユーザが外した時としても良いし、音

声トリガによっても良いし、前記ネットワーク端末を操作する事によって停止しても良いが、必ずしもそれらには限らない。さらに、ユーザの着目の結果認識された対象は、当該時空間情報を付して後述のグラフデータベース３６５に記録する事で、後日の問い合わせに回答出来る様に構成しても良い。前記図３Ａに記載の対象画像抽出処理は本発明における重要なプロセスであり、以下その各ステップを説明する。

【００４９】

最初に、ユーザによる音声入力トリガ（３０－０２）が発生し、カメラ画像のアップロード（３０－０３）開始後、音声認識処理３０－０５によりユーザの対象検出コマンドから単語列が抽出され、前記単語列が条件群３０－０７から３０－１５のいずれかの特徴に適合した場合には、係る画像特徴抽出処理に引き渡される。前記単語列が「対象の名称」である場合（３０－０６）、例えば、ユーザが当該対象に係る固有名詞を発話した場合、当該アノテーションはユーザの一定の認識判断を反映したものとして、係る特定物体認識の実行（１１０）処理を行う。その照合結果と、当該アノテーションに齟齬がある場合、或いは疑問がある場合は、当該ユーザによる誤認識の可能性もあるとして、当該ユーザに喚起を促す。或いはユーザが、当該対象に係る一般名詞を発話した場合、当該一般名詞に係る一般物体認識の実行（１０６）処理を行い、その画像特徴から対象を抽出する。或いはユーザが当該対象に係るシーンを発話した場合、当該シーンに係るシーン認識の実行（１０８）処理を行い、その画像特徴から対象領域を抽出する。またそれらの特徴を一つだけ指し示すのではなくて、複数の特徴を含む情景として指定しても良い。例えば、道路（一般物体）の左側（位置）を走る（状態）黄色い（色）タクシー（一般物体）、ナンバーは「１２３４（特定物体）」という様な指定の方法である。これらの対象指定を一連の言葉としても良いし、各々個別に指定を行っても良い。対象が複数個発見される場合には、前記画像認識システムによる再確認プロセスを経て、さらに新たな画像特徴を追加して対象を絞り込んで行く事が出来る。当該画像抽出結果は、一例としてユーザに対し音声による質問、例えば「それは～ですか？」を発行して再確認処理される（３０－４０）。当該再確認内容に対し、着目対象の抽出がユーザの意図通りである場合は、ユーザはその旨を示す言葉或いは単語を発話して、ステップ３０－５０「カメラ画像アップロード終了」を実行し、当該対象画像抽出処理を終了する（３０－５１）。一方、ユーザの意図とは違う場合には、再びステップ３０－０４「音声コマンド入力待ち」に戻り、さらなる画像特徴群を入力する。また、何度入力しても対象の特定に至らない場合や、そもそも対象自体が視野外に移動してしまった場合等には、処理を中断（ＱＵＩＴ）して当該対象画像抽出処理を終了する。

【００５０】

例えば音声認識処理３０－０５の結果が図３Ａで示す条件３０－０７に適合した場合、即ちユーザが対象の「色」に関する特徴を発話した場合には、色抽出処理３０－２０が行われる。当該色抽出処理には、ＲＧＢ３原色において色毎に範囲を設定して抽出する手法を用いても良いし、それらをＹＵＶ色空間上で抽出しても良い。またこれら特定の色空間表現には限定されない。当該色抽出処理後に対象を分離抽出し（３０－２９）、セグメンテーション（切り出し領域）情報を得る。次に当該セグメンテーション情報を手掛かりに対象の画像認識処理（３０－１６）を行う。その後は当該画像認識処理結果を利用して共起物体や共起事象を抽出（３０－３８）し、抽出可能になった全特徴群に関する記述を生成（３０－３９）し、当該記述をもってユーザに再確認を求める（３０－４０）。その結果がＹＥＳであれば、カメラ画像のアップロードを終了（３０－５０）し、音声による対象画像の抽出処理を終了（３０－５１）する。

【００５１】

例えば音声認識処理３０－０５の結果が図３Ａで示す条件３０－０８に適合した場合、即ちユーザが対象の「形状」に関する特徴を発話した場合には、形状特徴抽出３０－２１が行われる。当該形状特徴抽出処理では、対象に係るエッジ追跡を行いながら輪郭や主要な形状特徴を抽出後、形状のテンプレート・適合処理を行うが、それ以外の手法を用いても良い。当該形状抽出処理後に対象を分離し（３０－３０）、セグメンテーション情報を

得る。次に当該セグメンテーション情報を手掛かりに対象の画像認識処理（３０－１６）を行う。その後は当該画像認識処理結果を利用して共起物体や共起事象を抽出（３０－３８）し、抽出可能になった全特徴群に関する記述を生成（３０－３９）し、当該記述をもってユーザに再確認を求める（３０－４０）。その結果がＹＥＳであれば、カメラ画像のアップロードを終了（３０－５０）し、音声による対象画像の抽出処理を終了（３０－５１）する。

【００５２】

例えば音声認識処理３０－０５の結果が図３Ａで示す条件３０－０９に適合した場合、即ちユーザが対象の「大きさ」に関する特徴を発話した場合には、物体サイズ検出処理３０－２２が行われる。その一例として、当該物体サイズ検出処理ではサイズ以外の他の特徴抽出処理等により切り分けされた当該対象物体に対し、周囲にある他の物体との相対的なサイズ比較がユーザとのインタラクティブな音声コミュニケーションにより実行される。例えば「左隣の～よりも大きな～」という様な指示である。その理由としては、対象が単独で存在する場合、その大きさの比較になる様な具体的な指標がないと、単に画角から見た大きさのみでそのサイズを一意に判断出来ない事によるが、それ以外の手法を用いても良い。当該サイズ検出後に対象を分離し（３０－３１）、セグメンテーション情報を得る。次に当該セグメンテーション情報を手掛かりに対象の画像認識処理（３０－１６）を行う。その後は当該画像認識処理結果を利用して共起物体や共起事象を抽出（３０－３８）し、抽出可能になった全特徴群に関する記述を生成（３０－３９）し、当該記述をもってユーザに再確認を求める（３０－４０）。その結果がＹＥＳであれば、カメラ画像のアップロードを終了（３０－５０）し、音声による対象画像の抽出処理を終了（３０－５１）する。

【００５３】

例えば音声認識処理３０－０５の結果が図３Ａで示す条件３０－１０に適合した場合、即ちユーザが対象の「明るさ」に関する特徴を発話した場合には、輝度検出処理３０－２３が行われる。当該輝度検出処理では、ＲＧＢ３原色から、或いはＹＵＶ色空間から特定領域の輝度を求めるが、それら以外の手法を用いても良い。当該対象の輝度検出処理では、対象の周囲と比較した相対輝度の抽出が、ユーザとのインタラクティブな音声コミュニケーションにより実行される。例えば「周りより明るく輝いている～」という様な指示である。その理由としては、対象が単独で存在する場合、その明るさの比較になる様な具体的な指標がないと、単に画素が有する輝度値のみでユーザが感じた輝度を一意に判断出来ない理由によるが、それ以外の手法を用いても良い。当該輝度検出後に対象を分離し（３０－３２）、セグメンテーション情報を得る。次に当該セグメンテーション情報を手掛かりに対象の画像認識処理（３０－１６）を行う。その後は当該画像認識処理結果を利用して共起物体や共起事象を抽出（３０－３８）し、抽出可能になった全特徴群に関する記述を生成（３０－３９）し、当該記述をもってユーザに再確認を求める（３０－４０）。その結果がＹＥＳであれば、カメラ画像のアップロードを終了（３０－５０）し、音声による対象画像の抽出処理を終了（３０－５１）する。

【００５４】

例えば音声認識処理３０－０５の結果が図３Ａで示す条件３０－１１に適合した場合、即ちユーザが「対象との距離」に関する特徴を発話した場合には、奥行き検出処理３０－２４が行われる。当該奥行き検出処理では、ユーザのヘッドセットシステム２００に具備された深度センサ２０６を用いて奥行きを直接測定しても良いし、２台以上のカメラ映像から得られる視差情報から計算により算出しても良い。また、これら以外の手法を用いても良い。当該距離検出後に対象を分離し（３０－３３）、セグメンテーション情報を得る。次に当該セグメンテーション情報を手掛かりに対象の画像認識処理（３０－１６）を行う。その後は当該画像認識処理結果を利用して共起物体や共起事象を抽出（３０－３８）し、抽出可能になった全特徴群に関する記述を生成（３０－３９）し、当該記述をもってユーザに再確認を求める（３０－４０）。その結果がＹＥＳであれば、カメラ画像のアップロードを終了（３０－５０）し、音声による対象画像の抽出処理を終了（３０－５１）

する。
【００５５】
例えば音声認識処理３０－０５の結果が図３Ａで示す条件３０－１２に適合した場合、即ちユーザが「対象の存在する位置／領域」に関して発話した場合には、対象の領域検出３０－２５が行われる。当該領域検出処理では、一例としてユーザの主たる視野を反映したカメラ画像全体を予め等間隔のメッシュ状に領域分割し、ユーザからのインタラクティブな指示として「右上の～」という様な領域指定から対象を絞り込んでも良いし、「机の上の～」という様な、対象が存在する場所の指定で行っても良い。また、他の位置／領域に係る指定であっても良い。当該対象の存在する位置／領域検出後に対象を分離し（３０－３４）、セグメンテーション情報を得る。次に当該セグメンテーション情報を手掛かりに対象の画像認識処理（３０－１６）を行う。その後は当該画像認識処理結果を利用して他の共起物体や共起事象を抽出（３０－３８）し、抽出可能になった当該共起特徴群を含む記述を生成（３０－３９）し、当該記述をもってユーザに再確認を求める（３０－４０）。その結果がＹＥＳであれば、カメラ画像のアップロードを終了（３０－５０）し、音声による対象画像の抽出処理を終了（３０－５１）する。
【００５６】
例えば音声認識処理３０－０５の結果が図３Ａで示す条件３０－１３に適合した場合、即ちユーザが「対象と他物体との位置関係」に関して発話した場合には、当該対象に係る共起関係検出３０－２６が行われる。当該共起関係検出処理では、図３Ａに記載の各処理（１０６、１０８、１１０、３０－２０から３０－２８）により抽出された対応特徴に係るセグメンテーション情報を用いて、それらのセグメンテーション情報に対応する各特徴との共起関係を精査する事で、対象の抽出を行う。一例として「～と一緒に写っている～」という様な指示であるが、これ以外の手法を用いても良い。これにより、当該対象と他物体との位置関係を基に対象を分離し（３０－３５）、当該対象に係るセグメンテーション情報を得る。次に当該セグメンテーション情報を手掛かりに対象の画像認識処理（３０－１６）を行う。その後は当該認識結果を利用して他の共起物体や共起事象を抽出（３０－３８）し、抽出可能になった当該共起特徴群を含む記述を生成（３０－３９）し、当該記述をもってユーザに再確認を求める（３０－４０）。その結果がＹＥＳであれば、カメラ画像のアップロードを終了（３０－５０）し、音声による対象画像の抽出処理を終了（３０－５１）する。
【００５７】
例えば音声認識処理３０－０５の結果が図３Ａで示す条件３０－１４に適合した場合、即ちユーザが「対象の動き」に関して発話した場合には、動き検出処理３０－２７が行われる。当該動き検出処理では、時間軸上に連続的に展開された複数枚の画像を参照し、各画像を複数のメッシュ領域に分割し、当該領域を相互に比較する事によって、カメラ自体の移動による全体画像の平行移動以外に、相対的に個別移動している領域を見つけ出し、その領域の差分抽出（３０－３６）処理を行い、周囲に比べて相対的に移動している領域に係るセグメンテーション情報を得る。また、これら以外の手法を用いても良い。次に当該セグメンテーション情報を手掛かりに、対象の画像認識処理（３０－１６）を行う。その後は当該画像認識処理結果を利用して他の共起物体や共起事象を抽出（３０－３８）し、抽出可能になった当該共起特徴群を含む記述を生成（３０－３９）し、当該記述をもってユーザに再確認を求める（３０－４０）。その結果がＹＥＳであれば、カメラ画像のアップロードを終了（３０－５０）し、音声による対象画像の抽出処理を終了（３０－５１）する。
【００５８】
例えば音声認識処理３０－０５の結果が図３Ａで示す条件３０－１５に適合した場合、即ちユーザが「対象の様子」に関して発話した場合には、状態検出処理３０－２８が行われる。当該状態検出処理では、物体の状態、例えば、運動状態（静止、移動、振動、浮遊、上昇、下降、飛翔、回転、泳動、接近、離遠等）、動作状態（走っている、跳んでいる、しゃがんでいる、座っている、寝ている、横たわっている、眠っている、食べている、

飲んでいる、観察可能な喜怒哀楽等を含む）を、当該状態に係る特徴を記述した知識データベース（未図示）を参照しながら、連続する複数の画像群から推定・抽出（３０－３７）し、セグメンテーション情報を得る。次に当該セグメンテーション情報を手掛かりに、対象の画像認識処理（３０－１６）を行う。その後は当該画像認識処理結果を利用して、他の共起物体や共起事象を抽出（３０－３８）し、抽出可能になった当該共起特徴群を含む記述を生成（３０－３９）し、当該記述をもってユーザに再確認を求める（３０－４０）。その結果がＹＥＳであれば、カメラ画像のアップロードを終了（３０－５０）し、音声による対象画像の抽出処理を終了（３０－５１）する。

【００５９】

ユーザは前記ステップに係る音声による図３Ａで示す再確認（３０－４０）のステップにおいて、前記対象画像抽出処理をユーザの発話により中止する事が出来る。音声認識処理３０－０５において、前記中止コマンドが認識された場合には、ステップ３０－５０に移行しカメラ画像アップロードを終了し、音声による対象画像抽出処理を終了する（３０－５１）。前記記載の各々の対象の検出、抽出、或いは認識処理において、処理時間が一定以上長引く場合には、ユーザに対して興味を引き続ける目的で処理の経過を示す状況や、関連する情報を音声で伝える事が出来る。例えば、「今着目している～の認識処理を、引き続きサーバに問合わせ中です。現在～人の方が同様の対象に注目しています。もう少しお待ち下さい」「～までの処理が終わりました。途中経過は～です」の様な経過メッセージを当該ユーザに対し、音声で返す事が出来る。

【００６０】

ここで、図３Ｂを用いて、図３Ａをデータの流れから説明する。入力は画像３５－０１と発話３５－０２である。認識・抽出処理制御３５－０３では、発話３５－０２の入力による図３Ａにおけるステップ３０－０６から３０－１５を１以上実行し、画像３５－０１に対して図３Ａにおけるステップ３５－１６を実行する際には、一般物体認識処理システム１１０による一般部隊認識処理、特定物体認識システム１１０による特定物体認識処理、及びシーン認識システム１０８によるシーン認識処理のいずれか１以上を実行する。画像認識システム１０６、１０８、１１０の各々の機能ブロックは、実行ユニット毎にさらなる並列化が可能であり、画像認識処理ディスパッチ３５－０４により１以上の処理に振り分けられて並列に実行される。また、発話３５－０２の入力に対し、図３Ａにおけるステップ３０－０７から３０－１５を実行する場合には、特徴抽出処理３０－２０から３０－２８、及び分離抽出処理３０－２９から３０－３７を実行する。上記特徴抽出処理及び分離抽出処理は各々１以上存在し、特徴抽出ディスパッチ３５－０５により１以上の処理に分けられて並列に実行される。前記認識・抽出処理制御３５－０３では、ユーザの発話に処理順序に影響を与える単語が含まれている場合（例えば、「～の上」という場合には「～」を画像認識する必要があり、その次に「上」を処理する）には、順序制御を行う。

【００６１】

入力画像３５－０１に関して、認識・抽出処理制御３５－０３は、後述のグラフデータベース３６５にアクセスして、代表ノード３５－０６を抽出（当該データベースに当該ノードが存在しなければ新しい代表ノードを生成）する。前記一連の処理より、画像３５－０１が発話３５－０２に従って処理され、前記同時実行される各認識・抽出処理群に係る結果のグラフ構造３５－０７がグラフデータベース３６５に蓄積される。この様にして、入力画像３５－０１に対する認識・抽出処理制御３５－０３による一連のデータの流れは、発話３５－０２が当該入力画像に関して有効にある限り続く。

【００６２】

次に図４Ａを用いて、本発明の一実施例におけるユーザの音声による対象のポインティング操作を説明する。これは、図３Ａに記載の手順に対する応用例である。図４Ａ（Ａ）の場所は、ニューヨーク州マンハッタン島タイムズ・スクエア界隈である。この場所にいるユーザ、或いはこの写真を見たユーザが仮に発話４１「Ａ　ｙｅｌｌｏｗ　ｔａｘｉ　ｏｎ　ｔｈｅ　ｒｏａｄ　ｏｎ　ｔｈｅ　ｌｅｆｔ　ｓｉｄｅ」をつぶやいたとする。ここから音声認識システム３２０は、当該発話４１から複数の文字列或いは単語列を抽出す

る。当該発話から抽出可能な単語としては「一台」の「黄色」の「タクシー」が「左側」の「道路上」に見える、の５個である。ここから、前記図３Ａで示した対象画像抽出フローにおける「対象の名称」「対象の色情報」「対象の位置」「対象の存在する領域」及び着目している対象が複数ではなく、単一の対象である事が判る。これらの手掛かりから、当該画像特徴群を有する対象の検出・抽出処理が開始され、それが点線円（５０）のタクシーである可能性を前記画像認識システム側がユーザに音声により返答する事が可能となった場合、前述した様にその再確認内容として、上記ユーザが明示的に示した特徴要素群のみで再確認するだけでは、今一歩確実性に欠ける場合がある。これらの不確実性に対処する為に、ユーザがまだ指し示していない当該対象に係る他の共起特徴要素群を検出し、それらを再確認内容に加える必要がある。例えば「それは手前の横断歩道に差し掛かっているタクシーで、前に人が見えますね？」という様に、前記画像認識システムを備えた知識情報処理サーバシステム側が検出した当該対象に係る新たな共起事象を加えユーザに再確認を求める事が出来れば、よりユーザの意に沿った対象の検出・抽出・絞り込み処理が可能となる。本事例では、点線円（５０）を含む領域の拡大画像図４Ａ（Ｂ）から、「横断歩道」（５５）「人」（５６）が検出可能となっている様子を示している。

【００６３】

同様に、大きな看板があるビルを見上げているユーザが、発話４５「Ｉ’ｍ　ｓｔａｎｄｉｎｇ　ｏｎ　ｔｈｅ　Ｔｉｍｅｓ　Ｓｑｕａｒｅ　ｉｎ　ＮＹ　ｎｏｗ」とつぶやけば、カメラ画像を用いた適合処理により、そこが「ニューヨーク」州「タイムズ・スクウェア」で、ユーザが有名なランドマークとなっている建物を着目していると推測可能になる。

【００６４】

同様に、発話４２「Ａ　ｒｅｄ　ｂｕｓ　ｏｎ　ｔｈｅ　ｒｏａｄ　ｉｎ　ｆｒｏｎｔ」という表現から、「１台（対象の数）」の「赤（対象の色特徴）」い「バス（対象の名称）」が「正面（対象の存在する位置）」の「道路（一般物体）」「上（対象の位置関係）」が抽出可能になり、ユーザが点線円５１内のバスを着目していると推定可能になる。

【００６５】

同様に、発話４４「Ｔｈｅ　ｓｋｙ　ｉｓ　ｆａｉｒ　ｉｎ　ＮＹ　ｔｏｄａｙ」という表現から、「今日」の「ＮＹ」の天気は「晴れ」が抽出可能になり、ユーザが点線円（５２）の領域「空」を見上げていると推定可能になる。

【００６６】

少し複雑なつぶやき４３「Ａ　ｂｉｇ　ａｄ－ｂｏａｒｄ　ｏｆ　“ｔｈｅ　Ｐｈａｎｔｏｍ　ｏｆ　ｔｈｅ　Ｏｐｅｒａ”，　ｔｏｐ　ｏｎ　ｔｈｅ　ｂｕｉｌｄｉｎｇ　ｏｎ　ｔｈｅ　ｒｉｇｈｔ　ｓｉｄｅ」からは、「右端」に見える「ビル」の「屋上」にある、点線円（５３）で示した「オペラ座の怪人」の「広告ボード」をユーザが着目していると推定可能になる。

【００６７】

これら検出可能な単語列は、各々「固有の名称」「一般名詞」「シーン」「色」「位置」「領域」「場所」等を示しており、それらに対応した画像検出・画像抽出処理が実行される。その結果が当該時空間情報、及び画像情報と共に、前記知識情報処理サーバシステム３００上に引き渡される。なお、図４Ａに記載のイメージは本発明の一実施例を説明したもので、それに限定されない。

【００６８】

ここで、図４Ｂを用いて、本発明の一実施例における図３Ａに記載の手順を実行する過程の学習機能に関して、図４Ａのシーンを例に説明する。図４Ｂ（Ａ）は図４Ａに記載のユーザの主たる視野を反映した画像に関して獲得されたグラフ構造（後述）の一部のスナップショットである。まず画像認識プロセスとグラフ構造との関係を説明する。

【００６９】

ノード（６０）は図４Ａを代表するノードであり、図４Ａの画像データを記録しているノード（６１）とリンクしている。以下、ノードとノードのリンクを用いて情報を表現す

る。ノード（６０）はまた、場所を表わすノード（６２）と、時間を表わすノード（６３）に対してもリンクしている事で、撮影場所と時間の情報を保持している。さらにノード（６０）は、ノード（６４）とノード（６５）とリンクしている。ノード（６４）は、図４Ａ中の点線円（５０）の対象を代表するノードであり、前記発話４１により、特徴量Ｔ１（６５）、特徴量Ｔ２（６６）、色属性（６７）、切り抜き画像（６８）、及び画像内の位置座標（６９）の各情報を保持している。前記特徴量は、図３Ａの手順の過程における後述の一般物体認識システム１０６の処理結果として得られる。ノード（６５）は、図４Ａの点線円（５１）の対象を代表するノードであり、前記ノード（６４）と同様の情報を保持している。なお、ノード（６０）即ち図４Ａは、ユーザ１の主観視画像としてノード（７７）とリンクしている。

【００７０】

次に、ユーザ２を表すノード（８０）の主観視を代表するノード（８１）の保持する情報を、図４Ｂ（Ｂ）に示す。図では簡略化のため、図４Ｂ（Ａ）に記載のノードのうちいくつかは省略している。ノード（８２）は、ユーザ２の主観視における図４Ａの点線円（５１）に相当する対象の代表ノードである。同様に、特徴量Ｃ１（８４）とＣ２（８５）を情報として保持している。

【００７１】

前記ノード（６５）にリンクする特徴量であるＢ１（７０）及びＢ２（７１）と、前記ノード（８２）にリンクする特徴量であるＣ１（８４）及びＣ２（８５）は、一般物体認識システム１０６において比較され、同一対象であると判断された場合（即ち同じカテゴリに属した場合）、或いは統計的に新たな重心となり得る場合には、代表特徴量Ｄ（９１）が算出され学習に付される。本実施例では、当該学習結果をＶｉｓｕａｌ　Ｗｏｒｄ辞書１１０－１０に記録する。さらに、対象を代表するノード（９０）、及びそのサブノード群（９１から９３と７５から７６）をリンクした部分グラフが生成され、ノード（６０）は、ノード（６５）とのリンクをノード（９０）とのリンクに置き換える。同様にノード８１は、ノード８２とのリンクをノード９０とのリンクに置き換える。

【００７２】

次に、他のユーザが異なる時空間において、図４Ａで点線円（５０）に相当する対象に着目した場合には、前記同様のグラフ構造を構築するが、当該対象に対して一般物体認識システム１０６は、前記学習により当該対象の特徴量がノード（９０）に記録された特徴量と同じクラスにも属すると判断出来るので、ノード（９０）とリンクする様にグラフ構造を構築する事が出来る。

【００７３】

図３Ａに記載の、ステップ３０－２０から３０－２８に対応する特徴抽出処理において抽出された特徴群は、ユーザの発話と、セグメンテーション情報と、当該特徴とをノードに持つグラフ構造として表現出来る。例えば、図４Ａの点線円（５０）のセグメンテーション領域の場合で、特徴抽出処理がステップ３０－２０の場合には、色に関する特徴ノードを保持するグラフ構造となる。当該グラフ構造は、既に対象に関する代表ノードが存在する時には、その部分グラフと比較される。図４Ｂの例では、ノード（６７）の色特徴“ｙｅｌｌｏｗ”と近いと判断出来るので、当該グラフ構造は代表ノード（６４）の部分グラフになる。この様なグラフ構造の統合を記録しておいても良い。それにより、当該例では、ユーザの発話と色特徴との関係を記録する事が出来るので、“ｙｅｌｌｏｗ”に対応する色特徴の確からしさを高める事になる。

【００７４】

上記記載の手順により、後述の画像認識に係るデータベース群（１０７、１０９、１１１、１１０－１０）と、後述のグラフデータベース３６５は成長（新しいデータを獲得）する。上記記載では一般物体の場合を説明したが、特定物体、人、写真、或いはシーンであっても、同様に当該データベース群に対象に関する情報が蓄積される。

【００７５】

次に図４Ｃを用いて、本発明の一実施形態におけるグラフデータベース３６５から複数

の対象候補ノードが抽出された場合に、ユーザがどれに着目しているかを算出する手段に関して説明する。当該手順は、例えば、図３Ａにおける手順のステップ３０－３８及びステップ３０－３９において抽出可能になった複数の対象候補から、ユーザの着目対象を選び出す際に利用出来る。

【００７６】

ステップ（Ｓ１０）は、前記ステップ３０－３８の結果の共起物体・事象に対応する代表ノードをグラフデータベース３６５から抽出する（Ｓ１１）。当該ステップは、図３Ａに記載のステップ３０－１６、及びステップ３０－２０から３０－２８において、前記グラフデータベースをアクセスする事で、例えば色特徴抽出３０－２０では図４Ａに関係する色ノードから、対象ノード（６４）と（６５）を、図４Ａノード６０と、２つの色ノード（６７）と（７２）のリンクから抽出する事が出来る。

【００７７】

前記ステップ（Ｓ１１）では、１以上の代表ノードが抽出され得る。その全ての代表ノードに対して、次のステップを繰り返す（Ｓ１２）。ステップ（Ｓ１３）では、一つの代表ノードを変数ｉに格納する。そして、当該変数ｉの代表ノードを参照しているノード数を、変数ｎ＿ｒｅｆ［ｉ］に格納する（Ｓ１４）。例えば、図４Ｂ（Ｃ）ではノード（９０）を参照しているノードからのリンクは点線円（９４）のリンクであり、「３」となる。次にｎ＿ａｌｌ［ｉ］にノードｉの部分グラフの全ノード数を代入（Ｓ１５）する。図４Ｂ（Ｃ）のノード（９０）では「５」を代入する。次に、ｎ＿ｒｅｆ［ｉ］が規定値以上か？が判断される。ＹＥＳの場合にはｎ＿ｆｅａ［ｉ］に１を代入（Ｓ１７）し、ＮＯの場合には０を代入（Ｓ１８）する。ステップ（Ｓ１９）ではｎ＿ｆｅａ［ｉ］に、前記ノードｉの部分グラフ中で図３Ａに記載の手順で、ユーザの発話した特徴に対応するノードの数をｎ＿ａｌｌ［ｉ］で除した数値を加算する。例えば、図４Ｂ（Ｃ）の例で、ノード（９０）に関して、ユーザが“ｒｅｄ”のみを発話した場合には１／５を加算し、ユーザは“ｒｅｄ”と“ｏｎ”と“ｒｏａｄ”を含む発話をした場合には３／５を加算する。その結果、｛ｎ＿ａｌｌ［ｉ］，ｎ＿ｆｅａ［ｉ］｝の二項組を、ノードｉに対する選択優先度とする。

【００７８】

上記の構成により、前記画像認識プロセスによる学習結果を反映したグラフ構造を算出基準とする事になり、当該学習結果を選択優先度に反映する事が出来る。例えば、図３Ａの記載の、ステップ３０－２０から３０－２８を含む特徴とユーザの発話が一致する場合には、代表ノードに当該特徴に関するノードが追加されるので、前記ステップにより算出された選択優先度が変化する。なお、選択優先度の算出は当該手法には限らない。例えばリンクの重みを考慮しても良い。また、図４Ｂ（Ｃ）ではノード（７４）とノード（７５）を他のノードを同じ重みとしてノード数をカウントしたが、当該ノード（７４）とノード（７５）は強関係にあるとして、１つのノードとしてカウントしても良い。この様にノード間の関係を考慮しても良い。

【００７９】

ステップ３０－３９の抽出可能になった全特徴群の記述の生成では、前記選択優先度の第１項の値が大きな順に並べたノード群の中で、第２項が値「１」以上のノードを選び、後述の会話エンジン４３０を利用して、音声による再確認をユーザに対して行う事が出来る。当該第２項は、ステップ（Ｓ１６）にて規定値との関係から算出している。即ち、前記代表ノードの非参照数から算出している。例えばステップ（Ｓ１６）の規定値を「２」にした場合には、２以上の複数のユーザがリンクしている（即ち一度はユーザの着目対象になっている）代表ノードを選び出す。即ちユーザ対して再確認をする候補に加える事を意味している。以上記載の手順により、ステップ３０－３８の共起物体・事象の抽出による当該対象候補群の中から、よりユーザの意にかなった対象を選び出す事が可能になる。

【００８０】

なお、前記選択優先度に係る二項組の値は、前記組み合わせの利用手段以外を用いても良い。例えば、前記二項組で表現された選択優先度を２次元ベクトルとして正規化して比

較しても良い。また、例えば、代表ノードに係る部分グラフにある特徴量ノード、図４Ｂ（Ｃ）の例ではノード（９１）の対応クラス内での代表特徴量（例えば、Ｖｉｓｕａｌ Ｗｏｒｄ辞書１１０－１０における特徴量）との距離を考慮して、前記選択優先度を算出しても良い。

【００８１】

さらに、前記再確認において、ユーザが規定時間無言の場合には、ユーザの意にかなった対象を認識した可能性と見做して、カメラ画像のアップロードを終了（３０－５０）しても良い。

【００８２】

図５を用いて、本発明の一実施形態に係る知識情報処理サーバシステム３００における機能ブロックを説明する。本発明では画像認識システム３０１、生体認証部３０２、インタレストグラフ部３０３、音声処理部３０４、状況認識部３０５、メッセージ保管部３０６、再生処理部３０７、ユーザ管理部３０８から構成しているが、これらに限定されず、そのいくつかを選択して構成しても良い。

【００８３】

上記音声処理部３０４部は、ユーザが装着したヘットセットシステム２００が拾うユーザの発声を、音声認識システム３２０を利用して発話単語列に変換する。また、後述の再生処理部３０６からの出力を、音声合成システム３３０を利用して当該ユーザに前記ヘッドセットシステムを通して音声として通知する。

【００８４】

次に図６Ａから図６Ｅを用いて、本発明の一実施形態における画像認識システム３０１の機能ブロックを説明する。前記画像認識システムでは、ヘッドセットシステム２００からの画像に対して、一般物体認識、特定物体認識、シーン認識等の画像認識処理を行う。

【００８５】

最初に図６Ａを用いて、本発明の一実施形態における画像認識システム３０１の構成例を説明する。画像認識システム３０１は、一般物体認識システム１０６、シーン認識システム１０８、特定物体認識システム１１０、画像カテゴリデータベース１０７、シーン構成要素データベース１０９、及びマザーデータベース（以下ＭＤＢと略す）１１１で構成される。一般物体認識システム１０６は、一般物体認識部１０６－０１、カテゴリ検出部１０６－０２、カテゴリ学習部１０６－０３、及び新規カテゴリ登録部１０６－０４とで構成され、シーン認識システム１０８は、領域抽出部１０８－０１、特徴抽出部１０８－０２、重み学習部１０８－０３、及びシーン認識部１０８－０４とで構成され、特定物体認識システム１１０は、特定物体認識部１１０－０１、ＭＤＢ検索部１１０－０２、ＭＤＢ学習部１１０－０３、及び新規ＭＤＢ登録部１１０－０４とで構成され、画像カテゴリデータベース１０７は、カテゴリ分類データベース１０７－０１、及び不特定カテゴリデータ１０７－０２で構成され、シーン構成要素データベース１０９は、シーン要素データベース１０９－０１、及びメタデータ辞書１０９－０２とで構成され、ＭＤＢ１１１は、詳細設計データ１１１－０１、付帯情報データ１１１－０２、特徴量データ１１１－０３、及び不特定物体データ１１１－０４とで構成される。画像認識システム３０１の機能ブロックは必ずしもこれらに限定されるものではないが、これら代表的な機能について簡単に説明する。

【００８６】

一般物体認識システム１０６は、画像中に含まれる物体を一般的な名称、或いはカテゴリで認識する。ここでいうカテゴリは階層的であり、同じ一般物体として認識されているものでも、さらに細分化されたカテゴリ（同じ椅子でも４本足の「椅子」もあれば、全く足の無い「座椅子」の様なものまで含まれる）や、さらに大域的なカテゴリ（椅子も机もタンスも含めて、これらは全て「家具」のカテゴリとして大分類される）としても分類及び認識が可能である。カテゴリ認識は、この分類を意味する「Ｃｌａｓｓｉｆｉｃａｔｉｏｎ」、即ち既知のクラスに物体を分類するという命題であり、カテゴリはまたクラスとも呼ばれる。

【００８７】
一般物体認識プロセスにおいて、入力画像中の物体と参照物体画像との比較照合を行った結果、それらが同一形状であるか類似形状である場合、あるいは極めて類似した特徴を併せ持ち、他のカテゴリが有する主要な特徴において明らかに類似度が低いと認められる場合に、認識された物体に対し対応する既知のカテゴリ（クラス）を意味する一般名称を付与する。それらの各カテゴリを特徴付ける必須要素を詳細に記述したデータベースがカテゴリ分類データベース１０７－０１であり、それらのいずれにも分類する事が出来ない物体は、不特定カテゴリデータ１０７－０２として一時的に分類し、将来の新たなカテゴリ登録、あるいは既存カテゴリの定義範囲の拡大に備える。
【００８８】
一般物体認識部１０６－０１では、入力された画像中の物体の特徴点から局所特徴量を抽出し、それらの局所特徴量が予め学習によって得られた所定の特徴量の記述と似ているか似ていないかを相互に比較して、前記物体が既知の一般物体であるかどうかを判別するプロセスを実行する。
【００８９】
カテゴリ検出部１０６－０２では、一般物体認識可能となった物体がどのカテゴリ（クラス）に属するかを、カテゴリ分類データベース１０７－０１との照合において特定あるいは推定し、その結果、特定カテゴリにおいてデータベースに追加あるいは修正を加える様な追加の特徴量が見出された場合には、カテゴリ学習部１０６－０３において再学習した上で、カテゴリ分類データベース１０７－０１の前記一般物体に関する記述をアップデートする。また一旦、不特定カテゴリデータ１０７－０２とされた物体とその特徴量が別に検出された他の不特定物体の特徴量と極めて類似であると判定された場合には、それらは新たに発見された同一の未知のカテゴリ物体である可能性が高いとして、新規カテゴリ登録部１０６－０４において、カテゴリ分類データベース１０７－０１にそれらの特徴量が新規登録され、当該物体に対し新たな一般名称が付与される。
【００９０】
シーン認識システム１０８では、入力画像全体あるいは一部を支配している特徴的な画像構成要素を、性質の異なる複数の特徴抽出システムを用いて検出し、それらをシーン構成要素データベース１０９に記載されているシーン要素データベース１０９－０１と多次元空間上で相互に参照する事で、各々の入力要素群が当該特定シーン内に検出されるパターンを統計処理により求め、画像全体あるいは一部を支配している領域が当該特定のシーンであるかどうかを認識する。併せて、入力画像に付帯しているメタデータ群と、シーン構成要素データベース１０９に予め登録済みのメタデータ辞書１０９－０２に記載されている画像構成要素とを照合し、シーン検出の精度を一段と向上させる事が可能となる。領域抽出部１０８－０１では、画像全体を必要に応じて複数の領域に分割して、領域毎にシーン判別を可能にする。例えば、都市空間内のビルの壁面や屋上に設置した監視カメラからは、交差点や数多くの店舗のエントランス等の複数のシーンを見渡す事が出来る。特徴抽出部１０８－０２は、指定した画像領域内における検出された複数の特徴点の局所特徴量、色情報や物体の形状等、利用可能な様々な画像特徴量から得られる認識結果を後段の重み学習部１０８－０３に入力し、各々の要素が特定のシーンにおいて共起する確率を求め、シーン認識部１０８－０４に入力して最終的な入力画像に対するシーン判別を行う。
【００９１】
特定物体認識システム１１０は、入力された画像から検出された物体の特徴を、予めＭＤＢ１１１内に収納されている特定物体群の特徴と逐次照合し、最終的に物体を同定処理（Ｉｄｅｎｔｉｆｉｃａｔｉｏｎ）する。地球上に存在する特定物体の総数は膨大で、それら全ての特定物体との照合を行う事はおよそ現実的ではない。従って、後述する様に、特定物体認識システムの前段において、予め一定の範囲内に物体のカテゴリや探索範囲を絞り込んでおく必要がある。特定物体認識部１１０－０１では、検出された画像特徴点における局所特徴量と、学習によって得られたＭＤＢ１１１内の特徴パラメータ群とを相互に比較し、前記物体がどの特定物体に当て嵌まるかの判別を統計処理により判別する。Ｍ

ＤＢ１１１には、その時点で入手可能な当該特定物体に関する詳細なデータが保持されている。一例として、それら物体が工業製品であるならば、詳細設計データ１１１－０１として設計図やＣＡＤデータ等から抽出された物体の構造、形状、寸法、配置図、可動部、可動範囲、重量、剛性、仕上げ等、物体を再構成し製造する為に必要な基本情報等がＭＤＢ１１１内に保持される。付帯情報データ１１１－０２には、物体の名称、製造者、部品番号、日時、素材、組成、加工情報等、物体に関する様々な情報が保持される。特徴量データ１１１－０３には、設計情報に基づいて生成される個々の物体の特徴点や特徴量に係る情報が保持される。不特定物体データ１１１－０４は、その時点ではどの特定物体にも属していない不明な物体等のデータとして、将来の解析に備えＭＤＢ１１１内に暫定的に収納される。ＭＤＢ検索部１１０－０２は、当該特定物体に対応する詳細データを検索する機能を提供し、ＭＤＢ学習部１１０－０３は、適応的かつ動的な学習プロセスを通して、ＭＤＢ１１１内の当該物体に係る記載内容に対し追加・修正を行う。また一旦、不特定物体として不特定物体データ１１１－０４とされた物体も、その後に類似の特徴を有する物体が頻繁に検出された場合、新規ＭＤＢ登録部１１０－０４により、新たな特定物体として新規登録処理される。

【００９２】

図６Ｂに、本発明の一実施形態における一般物体認識部１０６－０１のシステム構成、及び機能ブロックの実施例を示す。一般物体認識部１０６－０１の機能ブロックは必ずしもこれらに限定されるものではないが、代表的な特徴抽出手法としてＢａｇ－ｏｆ－Ｆｅａｔｕｒｅｓ（以下、ＢｏＦと略す）を適用した場合の一般物体認識手法について、以下に簡単に説明する。一般物体認識部１０６－０１は、学習部１０６－１０、比較部１０６－１１、ベクトル量子化ヒストグラム部（学習）１１０－１１、ベクトル量子化ヒストグラム部（比較）１１０－１４、及びベクトル量子化ヒストグラム識別部１１０－１５で構成され、学習部１１０－１６は、局所特徴量抽出部（学習）１１０－０７、ベクトル量子化部（学習）１１０－０８、Ｖｉｓｕａｌ　Ｗｏｒｄ作成部１１０－０９、及びＶｉｓｕａｌ　Ｗｏｒｄ辞書（ＣｏｄｅＢｏｏｋ）１１０－１０とで構成される。

【００９３】

ＢｏＦは、画像中に現れる画像特徴点を抽出し、その相対位置関係を用いずに物体全体を複数の局所特徴量（Ｖｉｓｕａｌ　Ｗｏｒｄ）の集合体として表現し、それらを学習によって得られたＶｉｓｕａｌ　Ｗｏｒｄ辞書（ＣｏｄｅＢｏｏｋ）１１０－１０と比較照合して、それら局所特徴量の構成がどの物体に最も近いかを判別する。

【００９４】

図６Ｂを用いて、本発明の一実施形態における一般物体認識部１０６－０１における処理を説明する。学習部１０６－１０を構成する局所特徴量抽出部（学習）１１０－０７により得られた多次元の特徴ベクトルは、後段のベクトル量子化部（学習）１１０－０８によって一定次元数の特徴ベクトル群にクラスタ分割され、Ｖｉｓｕａｌ　Ｗｏｒｄ作成部１１０－０９で各々の重心ベクトルを元に、特徴ベクトル毎にＶｉｓｕａｌ　Ｗｏｒｄが生成される。クラスタリングの手法として、ｋ－ｍｅａｎｓ法やｍｅａｎ－ｓｈｉｆｔ法が知られている。生成されたＶｉｓｕａｌ　Ｗｏｒｄは、Ｖｉｓｕａｌ　Ｗｏｒｄ辞書（ＣｏｄｅＢｏｏｋ）１１０－１０に収納され、それを基に入力画像から抽出された局所特徴量を相互に照合し、ベクトル量子化部（比較）１１０－１３においてＶｉｓｕａｌ　Ｗｏｒｄ毎にベクトル量子化を行う。その後、ベクトル量子化ヒストグラム部（比較）１１０－１４において、全てのＶｉｓｕａｌ　Ｗｏｒｄに対するヒストグラムを生成する。

【００９５】

当該ヒストグラムの各ビンの総数（次元数）は通常数千から数万と多く、入力画像によっては特徴の一致が全くないヒストグラムのビンも数多く存在する一方、特徴の一致が顕著なビンもあり、それらを一括してヒストグラムの全ビンの値の総和が１になる様な正規化処理を行う。得られたベクトル量子化ヒストグラムは、後段のベクトル量子化ヒストグラム識別部１１０－１５へと入力され、一例として代表的な識別器であるＳｕｐｐｏｒｔ　Ｖｅｃｔｏｒ　Ｍａｃｈｉｎｅ（以下ＳＶＭと呼称）において、物体の属するクラス、

即ち当該対象が如何なる一般物体であるかを認識処理する。ここでの認識結果は、前記Ｖｉｓｕａｌ　Ｗｏｒｄ辞書に対する学習プロセスとしても利用可能である。また、他の手法（メタデータや集合知の利用）から得られた情報も、同様に前記Ｖｉｓｕａｌ　Ｗｏｒｄ辞書に対する学習フィードバックとして利用が可能で、同一クラスの特徴を最も適切に記述し、且つ他のクラスとの分離度を良好に保つ様に、適応的な学習を継続する事が重要となる。

【００９６】

図６Ｃに、本発明の一実施形態における前記一般物体認識部１０６－０１を含む一般物体認識システム１０６全体の概略構成ブロック図を示す。一般物体（クラス）は様々なカテゴリに属していて、それらは多重的な階層構造を成している。一例を挙げると、人間は「哺乳類」という上位カテゴリに属し、哺乳類は「動物」というさらに上位のカテゴリに属している。人間はまた、髪の色や目の色、大人か子供か？といった別のカテゴリでも認識が可能である。これらの認識判断を行うには、カテゴリ分類データベース１０７－０１の存在が欠かせない。これは人類の「知」の集積庫であり、将来の学習や発見によって、そこにさらに新たな「知」が加わり継続的な進化が図られるものでもある。一般物体認識部１０６－０１で同定されたクラス（およそ人類がこれまでに識別している全ての名詞の総数に及ぶ）は、様々な多次元的且つ階層的な構造体として、当該カテゴリ分類データベース１０７－０１内に記述されている。継続的な学習において認識された一般物体は、カテゴリ分類データベース１０７－０１と照合され、カテゴリ検出部１０６－０２で所属カテゴリが認識される。その後、カテゴリ学習部１０６－０３に当該認識結果が引き渡され、カテゴリ分類データベース１０７－０１内の記述との整合性がチェックされる。一般物体認識された物体は、時に複数の認識結果を内包する場合が多い。例えば「昆虫」であると認識した場合に、目の構造や手足の数、触角の有無、全体の骨格構造や羽の大きさ、胴体の色彩や表面のテクスチャ等でも新たな認識・分類が可能で、前記カテゴリ分類データベース１０７－０１内の詳細記述を基に照合される。カテゴリ学習部１０６－０３では、これらの照合結果を基に、カテゴリ分類データベース１０７－０１への追加・修正が必要に応じて適応的に行われる。その結果、既存カテゴリのいずれにも分類出来ない場合、「新種の昆虫」である可能性も高いとして、新規カテゴリ登録部１０６－０４がこれらの新規物体情報をカテゴリ分類データベース１０７－０１内に登録する。一方、その時点で不明な物体は、不特定カテゴリデータ１０７－０２として、将来の解析や照合に備え一時的にカテゴリ分類データベース１０７－０１内に収納される。

【００９７】

図６Ｄに、本発明の一実施形態における入力画像に含まれるシーンを認識判別する、シーン認識システム１０８の本発明における代表的な実施例をブロック図で示す。学習画像及び入力画像からは、一般に複数の物体が認識可能となるケースが多い。例えば、「空」「太陽」「地面」等を表す領域と同時に、「木」や「草」そして「動物」等の物体が同時に認識可能となる場合、それらが「動物園」なのか「アフリカの草原」なのかは、全体の景色やそれ以外に発見される物体との共起関係等から類推する事になる。例えば、檻や案内板等が同時に発見され多くの見物客で賑わっていれば、そこが「動物園」である可能性が高まるが、全体のスケールが大きく、遠くに「キリマンジャロ」の様な雄大な景色を臨み、様々な動物が混在して草原上にいる様な場合には、そこが「アフリカの草原」である可能性が一気に高まる。この様な場合、さらに認識可能な物体や状況、共起事象等を知識データベースであるシーン構成要素データベース１０９に照合し、より総合的な判断を下す必要も出てくる。例えば、全画面の９割が「アフリカの草原」を指し示していると推定されても、後述の図２２に記載の事例における手順と共に、それらが矩形の枠で切り取られ全体が平面状であれば、ポスターや写真である確率が極めて高くなる。

【００９８】

シーン認識システム１０８は、領域抽出部１０８－０１、特徴抽出部１０８－０２、強識別器（重み学習部）１０８－０３、シーン認識部１０８－０４、及びシーン構成要素データベース１０９から構成され、特徴抽出部１０８－０２は、局所特徴量抽出部１０８－

０５、色情報抽出部１０８－０６、物体形状抽出部１０８－０７、コンテキスト抽出部１０８－０８、及び弱識別器１０８－０９から１０８－１２とで構成され、シーン認識部１０８－０４は、シーン分類部１０８－１３、シーン学習部１０８－１４、及び新規シーン登録部１０８－１５で構成され、シーン構成要素データベース１０９は、シーン要素データベース１０９－０１、及びメタデータ辞書１０９－０２で構成される。

【００９９】

領域抽出部１０８－０１は、背景や他の物体の影響を受けずに目的とする物体の特徴を効果的に抽出する為に、対象画像に係る領域抽出を行う。領域抽出手法の例として、グラフベースの領域分割法（Ｅｆｆｉｃｉｅｎｔ Ｇｒａｐｈ－Ｂａｓｅｄ Ｉｍａｇｅ Ｓｅｇｍｅｎｔａｔｉｏｎ）等が知られている。抽出された物体画像は、局所特徴量抽出部１０８－０５、色情報抽出部１０８－０６、物体形状抽出部１０８－０７、及びコンテキスト抽出部１０８－０８に各々入力され、それらの各抽出部から得られた特徴量が弱識別器１０８－０９から１０８－１２において識別処理され、多次元の特徴量群として統合的にモデリングされる。それらモデリング化された特徴量群を、重み付け学習機能を有する強識別器１０８－０３に入力し、最終的な物体画像に対する認識判定結果を得る。前記の弱識別器の例としてＳＶＭ、強識別器の例としてはＡｄａＢｏｏｓｔ等が代表的である。

【０１００】

一般に入力画像には複数の物体や、それらの上位概念である複数のカテゴリが含まれている場合が多く、人間はそこから一目で特定のシーンや状況（コンテキスト）を思い浮かべる事が出来る。一方、単独の物体や単一のカテゴリのみを提示された場合、それだけで入力画像がどういうシーンを表わしているのかを判断するのは困難である。通常は、それらの物体が存在している周囲の状況や相互の位置関係、また各々の物体やカテゴリの共起関係（同時に出現する確率）が、当該シーンの判別に対して重要な意味を持ってくる。前項で画像認識可能となった物体群やカテゴリ群は、シーン要素データベース１０９－０１内に記述されているシーン毎の構成要素群の出現確率を基に照合処理され、後段のシーン認識部１０８－０４において、係る入力画像がいかなるシーンを表現しているのかを統計的手法を用いて決定する。

【０１０１】

これとは別の判断材料として、画像に付帯しているメタデータも有用な情報源となり得る。しかし、時には人間が付したメタデータ自体が、思い込みや明らかな誤り、或いは比喩として画像を間接的に捉えている場合等もあり、必ずしも当該画像中に存在する物体やカテゴリを正しく表わしているとは限らない場合がある。この様な場合にも、前記画像認識システムを備えた知識情報処理サーバシステムから抽出可能な当該対象に係る共起事象等を参考に総合的に判断し、最終的な物体やカテゴリの認識処理が行われる事が望ましい。また、一つの画像からは複数のシーンが得られる場合も多い。例えば、「夏の海」であると同時に「海水浴場」であったりもする。その場合は、複数のシーン名が当該画像に付される。さらに画像に付すべきシーン名として、例えば「夏の海」或いは「海水浴場」のいずれがより適当であるかは、当該画像のみからでは判断が難しく、前後の状況や全体との関係、各々の要素群の共起関係等を参考に、それらの要素間の関連性を記述した知識データベース（未図示）を基に最終的に判断が必要な場合もある。

【０１０２】

図６Ｅに、本発明の一実施形態における特定物体認識システム１１０のシステム全体の構成例、及び機能ブロックを示す。特定物体認識システム１１０は、一般物体認識システム１０６、シーン認識システム１０８、ＭＤＢ１１１、特定物体認識部１１０－０１、ＭＤＢ検索部１１０－０２、ＭＤＢ学習部１１０－０３、及び新規ＭＤＢ登録部１１０－０４とで構成され、特定物体認識部１１０－０１は、二次元写像部１１０－０５、個別画像切り出し部１１０－０６、局所特徴量抽出部（学習）１１０－０７、ベクトル量子化部（学習）１１０－０８、Ｖｉｓｕａｌ Ｗｏｒｄ作成部１１０－０９、Ｖｉｓｕａｌ Ｗｏｒｄ辞書（ＣｏｄｅＢｏｏｋ）１１０－１０、ベクトル量子化ヒストグラム部（学習）１１０－１１、局所特徴量抽出部（比較）１１０－１２、ベクトル量子化部（比較）１１０

－１３、ベクトル量子化ヒストグラム部（比較）１１０－１４、ベクトル量子化ヒストグラム識別部１１０－１５、形状特徴量抽出部１１０－１６、形状比較部１１０－１７、色情報抽出部１１０－１８、及び色彩比較部１１０－１９とで構成される。

【０１０３】

一般物体認識システム１０６により、対象物体の属するクラス（カテゴリ）が認識可能になった時点で、物体がさらに特定物体としても認識可能か？という絞り込みのプロセスに移る事が出来る。クラスが或る程度特定されないと、無数の特定物体群からの検索を余儀なくされ、時間的にもコスト的にも実用的とは言えない。これらの絞り込みプロセスには、一般物体認識システム１０６によるクラスの絞り込み以外にも、シーン認識システム１０８の認識結果から当該対象の絞り込みを行う事も有用となる。また特定物体認識システム１１０から得られる特徴量を用いて、さらなる絞り込みが可能になるだけではなく、物体の一部にユニークな識別情報（商品名とか、特定の商標やロゴ等）が認識可能な場合、或いは有用なメタデータ等が予め付されているケースでは、さらなるピンポイントの絞り込みも可能となる。

【０１０４】

それら絞り込まれたいくつかの可能性の中から、複数の物体候補群に係る詳細データや設計データをＭＤＢ検索部１１０－０２がＭＤＢ１１１内から順次引き出し、それらを基に入力画像との適合プロセスが実行される。物体が工業製品でない場合や、詳細な設計データ自体が存在していない場合においても、写真等があれば各々検出可能な画像特徴及び画像特徴量を詳細に突き合わせる事で、或る程度の特定物体認識が可能となる。しかし、入力画像と比較画像の見え方が全く同じというケースは稀で、例え同じであっても各々を違う物体として認識してしまう事例もある。反面、物体が工業製品であり、ＣＡＤ等の詳細なデータベースが利用可能な場合には、一例として二次元写像部１１０－０５が入力画像の見え方に応じＭＤＢ１１１内の三次元データを二次元画像に可視化（レンダリング）する事により、精度の高い特徴量の適合処理を行う事が可能になる。この場合、二次元写像部１１０－０５における二次元画像へのレンダリング処理を全視点方向からくまなく写像して実行する事は、計算時間と計算コストの不要な増大を招く事から、入力画像の見え方に応じた絞り込み処理が必要となる。一方、ＭＤＢ１１１を用いた高精度のデータから得られる各種特徴量は、学習プロセスにおいて予め求めておく事が可能である。

【０１０５】

特定物体認識部１１０－０１では、物体の局所特徴量を局所特徴量抽出部１１０－０７で検出し、ベクトル量子化部（学習）１１０－０８で各々の局所特徴量を複数の類似特徴群に分離した後、Ｖｉｓｕａｌ　Ｗｏｒｄ作成部１１０－０９で多次元の特徴量セットに変換し、それらをＶｉｓｕａｌ　Ｗｏｒｄ辞書１１０－１０に登録する。これらは多数の学習画像に対し、十分な認識精度が得られるまで継続して行われる。学習画像が例えば写真等である場合は、画像の解像度不足やノイズの影響、オクルージョンの影響、対象以外の物体から受ける影響等が避けられないが、ＭＤＢ１１１を基にしている場合は、ノイズのない高精度のデータを基に理想的な状態で対象画像の特徴抽出を行う事が可能な事から、従来の手法に比べて大幅に抽出・分離精度を高めた認識システムを構成する事が可能となる。入力画像は、個別画像切り出し部１１０－０６で目的とする特定物体に係る領域が切り出された後に、局所特徴量抽出部（比較）１１０－１２において局所特徴点及び特徴量が算出され、予め学習により用意されたＶｉｓｕａｌ　Ｗｏｒｄ辞書１１０－１０を用い個々の特徴量毎にベクトル量子化部（比較）１１０－１３にてベクトル量子化された後に、ベクトル量子化ヒストグラム部（比較）１１０－１４にて多次元の特徴量に展開され、ベクトル量子化ヒストグラム識別部１１０－１５にて、物体が当該学習済み物体と同一か、似ているか、それとも否かが識別判断される。識別器の例として、ＳＶＭ（Ｓｕｐｐｏｒｔ　Ｖｅｃｔｏｒ　Ｍａｃｈｉｎｅ）が広く知られているが、他にも識別判断の重み付けを学習の上で可能にするＡｄａＢｏｏｓｔ等も有効な識別器として広く活用されている。これらの識別結果は、ＭＤＢ学習部１１０－０３を通じてＭＤＢ自体への追加修正、或いは新たな項目の追加というフィードバックループにも利用可能となる。対象が依然と

して未確認となる場合には、新規ＭＤＢ登録部１１０－０４に保留され、次なる解析再開に備える。

【０１０６】

また、局所特徴量のみならず、検出精度をさらに向上させる目的で、物体の形状特徴を利用する事も有用となる。入力画像から切り出された物体は、形状特徴量抽出部１１０－１６を経由して形状比較部１１０－１７に入力され、物体の各部の形状的な特徴を用いた識別が行われる。その識別結果はＭＤＢ検索部１１０－０２にフィードバックされ、それによりＭＤＢ１１１に対する絞り込み処理が可能となる。形状特徴量抽出手段の例として、ＨｏＧ（Ｈｉｓｔｏｇｒａｍｓ ｏｆ Ｏｒｉｅｎｔｅｄ Ｇｒａｄｉｅｎｔｓ）等が知られている。形状特徴は、またＭＤＢ１１１を用いた二次元写像を得る為の多視点方向からのレンダリング処理を大幅に減らす目的でも有用となる。

【０１０７】

また、物体の色彩的な特徴やテクスチャ（表面処理）も、画像認識精度を上げる目的で有用である。切り出された入力画像は、色情報抽出部１１０－１８に入力され、色彩比較部１１０－１９で物体の色情報、あるいは当該テクスチャ等の抽出が行われ、その結果をＭＤＢ検索部１１０－０２にフィードバックする事で、ＭＤＢ１１１においてさらなる絞り込み処理を行う事が可能となる。これら、一連のプロセスを通じて、特定物体認識処理をより効果的に行う事が可能となる。

【０１０８】

次に、図７を用いて、本発明の一実施形態における生体認証部３０２の処理手順３４０を説明する。ユーザが前記ヘッドセットシステム２００を装着する事で（３４１）、以下の生体認証処理が始まる。ユーザと前記知識情報処理サーバシステムとの間の通信において、個々のユーザに対応する生体認証情報や、個々のユーザのプロファイル等の個人情報をやり取りする場合には、通信途中でのデータの抜き取りや改竄等の不正な行為からの強力な保護が必須になる。そこで、まず上記生体認証システムとの間で、強力な暗号化通信路を確立する（３４２）。ここではＳＳＬ（Ｓｅｃｕｒｅ Ｓｏｃｋｅｔｓ Ｌａｙｅｒ）や、ＴＬＳ（Ｔｒａｎｓｐｏｒｔ Ｌａｙｅｒ Ｓｅｃｕｒｉｔｙ）等の技術（例えば、ｈｔｔｐ：／／ｗｗｗ．ｏｐｅｎｓｓｌ．ｏｒｇ／）を用いる事が可能になるが、他の同様の暗号化手法を導入しても良い。次に、前記ヘッドセットシステムに具備された生体認証センサ２０４から、生体認証情報３４５を取得する。生体認証情報には、前記ヘッドセットシステムを装着するユーザの外耳部や鼓膜における静脈パターン情報等を用いる事が出来るが、これらを選択して組み合わせても良いし、これらには限らない。前記生体認証情報はテンプレートとして、前記生体認証システムに送付される。図７のステップ３５５は、前記生体認証システム側での処理を説明している。ステップ３５６にて、当該テンプレートを知識情報処理サーバシステム３００にユーザ登録する。ステップ３５７にて、当該テンプレートから署名＋暗号化関数ｆ（ｘ，ｙ）を生成し、ステップ３５８にて前記関数を当該ヘッドセットシステムに返す。ここで、ｆ（ｘ，ｙ）における“ｘ”は署名暗号化されるデータであり、“ｙ”は署名暗号化の際に用いる生体認証情報である。判断３４５では、前記関数を入手出来たかどうかを確認し、ＹＥＳの場合には当該ヘッドセットシステムと前記知識情報処理サーバシステム間の通信に前記関数を利用する（３４６）。判断３４５がＮＯの場合には、規定回数、前記判断３４５がＮＯであるかを判断（３４９）し、ＹＥＳの場合には認証エラーをユーザに通知する（３５０）。当該判断３４９がＮＯの場合には、ステップ３４４から処理を繰り返す。その後、ステップ（３４７）で規定時間待ってから、ループ（３４３）を繰り返す。ユーザが当該ヘッドセットシステムを取り外した場合、或いは前記認証エラーの場合には、前記生体認証システムとの間の暗号化通信路を切断する（３４８）。

【０１０９】

図８Ａに、本発明の一実施形態におけるインタレストグラフ部３０３の構成例を示す。本実施例においては、グラフデータベース３６５へのアクセスを、グラフデータベース３６５、及びユーザデータベース３６６への直接アクセスとして記述しているが、具体的な

実装においては、システムを利用中のユーザに係るインタレストグラフ適用処理の高速化を図る目的で、グラフ記憶部３６０はグラフデータベース３６５内に収納されているグラフ構造データの中から必要な部分のみ、及びユーザデータベース３６６内に記載の当該ユーザに係る必要な部分情報を自らの高速メモリ上に選択的に読み出し、内部にキャッシュする事が可能である。

【０１１０】

グラフ演算部３６１は、前記グラフ記憶部３６０から部分グラフの抽出、又は前記ユーザに係るインタレストグラフの演算を行う。関連性演算部３６２は、ノード間の関連性に関して、ｎ（＞１）次繋がりのノードの抽出、フィルタリング処理、及びノード間のリンクの生成・破壊等を行う。統計情報処理部３６３は、前記グラフデータベース内のノードとリンクデータを統計情報として処理し、新規の関連性を発見する。例えば、或る部分グラフが別の部分グラフと情報距離が近く、同じ様な部分グラフが同一クラスタ内に分類出来る時は、新しい部分グラフは前記クラスタに含まれる確率が高いと判断可能になる。

【０１１１】

ユーザデータベース３６６は、当該ユーザに関する情報を保持しているデータベースであり、前記生体認証部３０２にて利用される。本発明では、前記ユーザデータベース内部の当該ユーザに対応したノードを中心としたグラフ構造を、当該ユーザのインタレストグラフとして扱う。

【０１１２】

図８Ｂを用いて、本発明の一実施形態におけるグラフデータベース（３６５）に関して説明する。図８Ｂ（Ａ）に、前記グラフデータベース（３６５）に対する基本アクセス手法を示す。ｖａｌｕｅ（３７１）は、ｋｅｙ（３７０）からｌｏｃａｔｅ演算（３７２）により得られる。前記ｋｅｙ（３７０）は、ｖａｌｕｅ（３７３）をハッシュ（ｈａｓｈ）関数で計算して導出する。例えば、ハッシュ関数にＳＨＡ－１アルゴリズムを用いた場合には、ｋｅｙ（３７０）は１６０ビット長になる。Ｌｏｃａｔｅ演算（３７２）には、分散ハッシュテーブル（Ｄｉｓｔｒｉｂｕｔｅｄ　Ｈａｓｈ　Ｔａｂｌｅ）法を利用出来る。図８Ｂ（Ｂ）に示す様に、本発明では、前記ｋｅｙとｖａｌｕｅの関係を（ｋｅｙ，｛ｖａｌｕｅ｝）で表現し、前記グラフデータベースへの格納単位とする。

【０１１３】

例えば、図８Ｂ（Ｃ）の様に、２つのノードがリンクされている場合、ノードｎ１（３７５）は、（ｎ１，｛ノードｎ１｝）で、ノードｎ２（３７６）は、（ｎ２，｛ノードｎ２｝）で表現する。ｎ１やｎ２は、各々ノードｎ１（３７５）、ノードｎ２（３７６）のｋｅｙであり、ノード実体ｎ１（３７５）、ノード実体ｎ２（３７６）を各々ｈａｓｈ演算し、各々のｋｅｙを得る。また、リンクｌ１（３７７）は、ノードと同様に（ｌ１，｛ｎ１，ｎ２｝）で表現し、｛ｎ１，ｎ２｝をｈａｓｈ演算する事で、そのｋｅｙ（ｌ１）３７７を得る。

【０１１４】

図８Ｂ（Ｄ）は、前記グラフデータベースの構成要素の一例である。ノード管理部３８０）は前記ノードを、リンク管理部３８１は前記リンクを管理し、各々をノード・リンク格納部３８５に記録する。データ管理部３８２は、ノードに関連したデータをデータ格納部３８６に記録すべく管理する。

【０１１５】

図９を用いて、本発明の一実施例における状況認識部３０５の構成例を説明する。図９（Ａ）における履歴管理部４１０は、ユーザ毎にネットワーク・コミュニケーションシステム１００内での利用履歴を管理する。例えば、対象に対する着目を足跡（フットプリント）として残す事を可能にする。或いは、同じメッセージやつぶやきを繰り返して再生しない様に、前回どこまで再生したか？を記録する。或いは、メッセージやつぶやきの再生を途中で中止した時には、以降の継続再生の為に当該再生を中止した箇所を記録する。例えば、図９（Ｂ）では、その一実施例として、グラフデータベース３６５に記録されたグラフ構造の一部を示す。ユーザ（４１７）ノード、対象（４１５）ノード、及びメッセー

ジやつぶやき（４１６）ノードは、各々リンクで繋がっている。ノード（４１６）に再生位置を記録したノード（４１８）をリンクする事で、ユーザ（４１７）の着目した対象（４１５）に関するメッセージやつぶやきの再生を、ノード（４１８）として記録した再生位置から再開する。なお、本実施例における前記利用履歴はこれらの手法には限定されず、同様の効果が期待出来る他の手法を用いても良い。

【０１１６】

メッセージ選択部４１１はユーザ毎に管理され、ユーザが着目した対象に複数のメッセージやつぶやきが記録されていた場合に、適切なメッセージやつぶやきを選択する。例えば、記録された時刻順で再生しても良い。当該ユーザに係るインタレストグラフから、当該ユーザの関心の高い話題を選択的に選び出し再生しても良い。また、当該ユーザを明示的に指定したメッセージやつぶやきを優先的に再生しても良い。なお、本実施例におけるメッセージやつぶやきの選択手順は、これらに限定されない。

【０１１７】

カレント・インタレスト４１２は、インタレストグラフ部３０３中の当該ユーザに係る現在の関心を表すノード群として、ユーザ毎に管理され収納されている。前記メッセージ選択部では、前記カレント・インタレストにおける当該ユーザの現在の関心に対応したノード群から上記グラフ構造を探索する事で、当該ユーザが当該時点において関心度の高いノード群を選び出し、後述の会話エンジン４３０の入力要素とし、それらを一連の文章に変換し再生する。

【０１１８】

当該ユーザの関心の対象や度合いは、例えば後述の図１７におけるグラフ構造から求める。図１７において、ユーザ（１００１）ノードは，ノード（１００５）とノード（１００２）へのリンクを有している。即ち、このリンクから，「ワイン」と「車」に関心があるとする。前記ユーザが「ワイン」と「車」のどちらに関心が高いかは、ノード「ワイン」から繋がるグラフ構造と、ノード「車」から繋がるグラフ構造とを比較し、ノード数が多い方をより関心が高いとしても良いし、ノードに関連した着目履歴から、着目回数の多い方により関心が高いとしても良いし、前記ユーザが自らの関心の強さを指定しても良いし、これらには限定されない。

【０１１９】

図１０を用いて、本発明の一実施形態におけるメッセージ保管部３０６に関して説明する。ユーザが発話したメッセージやつぶやき３９１、及び／又は、ヘッドセットシステム２００で撮影した画像４２１は、当該メッセージ保管部によりメッセージデータベース４２０に記録される。メッセージノード生成部４２２は、インタレストグラフ部３０３から前記メッセージやつぶやきの対象となる情報を取得し、メッセージノードを生成する。メッセージ管理部４２３は、当該メッセージノードに前記メッセージやつぶやきを関連付けて、前記メッセージやつぶやきを前記グラフデータベース３６５に記録する。なお、前記ヘッドセットシステムで撮影した画像４２１を、同様に前記グラフデータベース３６５に記録しても良い。なお、前記メッセージやつぶやきの記録には、ネットワークを経由してネットワーク上の同様のサービスを利用しても良い。

【０１２０】

図１１を用いて、本発明の一実施形態における再生処理部３０７に関して説明する。ユーザのメッセージやつぶやき３９１を含むユーザの発話は、音声認識システム３２０で認識処理され、単数の或いは複数の単語列に変換される。前記単語列は、状況認識部３０４において「ユーザが現在何かの対象に着目している？」「時空間情報を指示している？」「或いは何かの対象に向かい話しかけている？」という状況識別子を付与され、再生処理部３０６の構成要素である会話エンジン４３０に送付される。なお、前記状況認識部３０４の出力としての識別子は、前記の各々の状況には限定されないし、当該識別子を用いない手法で構成しても良い。

【０１２１】

前記再生処理部３０７は、前記会話エンジン４３０、着目処理部４３１、コマンド処理

部４３２、ユーザメッセージ再生部４３３から構成されるが、これらを選択して構成しても良いし、新たな機能を追加して構成しても良く、当該構成には限定されない。前記着目処理部は、前記状況認識部から対象を着目中であるとの識別子が付された場合に実行され、図３Ａに記載の一連の処理を担う。前記ユーザメッセージ再生部は、対象に残されたメッセージやつぶやき、及び／又は、関連付けられた画像の再生を行う。

【０１２２】

図１２を用いて、本発明の一実施形態に係るユーザ管理部３０８に関し説明する。前記ユーザ管理部は、許可されたユーザのＡＣＬ（アクセス制御リスト）をグラフ構造で管理する。例えば、図１２（Ａ）は、一人のユーザ（４５１）ノードが、許可（４５０）ノードとリンクを有している状態を示す。これにより、当該ユーザに対し、当該許可ノードとリンクしたノードに対する許可が与えられる。当該ノードがメッセージやつぶやきであれば、それらを再生する事が出来る。

【０１２３】

図１２（Ｂ）は、特定のユーザ群に許可を与えている例である。許可（４５２）ノードは、ユーザグループ（４５３）ノードにリンクする、ユーザ１（４５４）ノード、ユーザ２（４５５）ノード、及びユーザ３（４５６）ノードに対し、一括して許可を与えている様子を示している。また、図１２（Ｃ）は、全員（４５８）ノードに対し、一括して許可（４５７）ノードが与えられている例である。

【０１２４】

さらに、図１２（Ｄ）は、特定のユーザ（４６０）ノードに対し、特定の時間或いは時間帯（４６１）ノード、特定の場所／地域（４６２）ノードに限り許可（４５９）ノードを与えている様子を示している。

【０１２５】

なお、本実施例におけるＡＣＬは、図１２以外の構成をとっても良い。例えば、不許可ノードを導入して、許可を与えないユーザを明示する様に構成しても良い。また、前記許可ノードをさらに詳細化して、再生許可ノードと記録許可ノードを導入する事で、メッセージやつぶやきを再生する場合と記録する場合で、許可の形態を変える様に構成しても良い。

【０１２６】

図１３Ａを用いて、本発明の一実施形態に係るネットワーク・コミュニケーションシステム１００を利用するユーザを中心とした、ユースケース・シナリオの一事例を説明する。

【０１２７】

本発明では、ユーザが装着しているヘッドセットシステム２００に具備されたカメラの撮影可能範囲を視野５０３と呼び、ユーザが主に見ている方向を当該ユーザの主観的な視野：主観視５０２と呼ぶ。ユーザは、ネットワーク端末２２０を装着しており、ユーザの発話（５０６又は５０７）を前記ヘッドセットシステムに組み込まれたマイクロフォン２０１で拾い、ユーザの主観視を反映した前記ヘッドセットシステムに組み込まれたカメラ２０３が撮影する映像と共に、前記知識情報処理サーバシステム３００側にアップロードされている。前記知識情報処理サーバシステム側からは、前記ヘッドセットシステムに組み込まれたイヤフォン２０２、或いはネットワーク端末２２０に対し、音声情報、及び映像／文字情報等を返す事が可能になっている。

【０１２８】

図１３Ａにおいて、ユーザ５００は物体群５０５を見ているとし、ユーザ５０１はシーン５０４を見ているとする。例えば、ユーザ５００に関して、図３Ａに記載の手順に従って当該ユーザのカメラの視野５０３には、物体群５０５が撮影され、その画像が前記知識情報処理サーバシステム３００側にアップロードされる。前記画像認識システム３０１は、そこから認識可能な特定物体、及び／又は一般物体を抽出する。この時点で当該画像認識システムとしては、ユーザ５００がどの対象に着目しているかまでは判らないので、ユーザ５００は音声によって、例えば「右上」とか「ワイン」といった様な当該ユーザの音

声による着目対象のポインティング操作を行い、前記画像認識システムに当該ユーザが現在物体５０８に着目している事を通知する。この際、前記知識情報処理サーバシステム側は「アイスペールに入っているワインですね？」という様な当該ユーザが明示的に示していない共起事象を加えた再確認の問い合わせを、当該ユーザ５００のヘッドセットシステム２００に対し音声で通知する事を可能とする。その再確認通知内容がユーザの意とは違っていた場合には、一例として「違う」と発話して、ユーザの追加的な対象選択指示を前記サーバシステム側に音声で発行し、改めて着目対象の再検出を求めるプロセスを可能にしても良い。或いは、当該ユーザは、前記ネットワーク端末上のＧＵＩにて着目中の対象を直接指定、又は修正しても良い。

【０１２９】

一例として、ユーザ５０１はシーン５０４を見ているが、ユーザの主観的視野５０３を反映したカメラ画像を、前記画像認識エンジンを備えた知識情報処理サーバシステム側にアップロードする事で、前記サーバシステム側に組み込まれた前記画像認識システムは、対象シーン５０４はおそらく「山の風景」であろうと推測する。ユーザ５０１は、前記シーンに対して自らのメッセージやつぶやき、例えば「懐かしい里山だ」を音声で発話する事で、当該ユーザのヘッドセットシステム２００経由で、当該メッセージやつぶやきが前記サーバシステム側に当該カメラ映像と共に記録される。その後、他のユーザが異なる時空間内において同様、或いは類似のシーンに遭遇した場合に、当該ユーザに対して、前記ユーザ５０１のつぶやき「懐かしい里山だ」を前記サーバシステム側からネットワークを介して、当該ユーザに対し音声情報で送り込む事が可能となる。この事例の様に、実際目にした景色自体やその場所等は異なっても、誰でも思い浮かべる共通の印象的なシーン、例えば「夕焼け」等に対して、共有体験に係るユーザコミュニケーションを喚起する事が可能になる。

【０１３０】

また、ユーザの音声による指示、或いはネットワーク端末２２０上での直接操作により、上記ユーザが予め設定した条件に従い、上記ユーザ５００やユーザ５０１が特定の対象に対して残したメッセージやつぶやきを、特定のユーザのみ、或いは特定のユーザグループのみ、或いはユーザ全員に対し、選択的に残す事を可能にする。

【０１３１】

また、ユーザの音声による指示、或いはネットワーク端末２２０上での直接操作により、当該ユーザが予め設定した条件に従い、当該ユーザ５００やユーザ５０１が特定の対象に対して残したメッセージやつぶやきを、特定の時間、或いは時間帯、及び／又は、特定の場所、特定の地域、及び／又は、特定のユーザ、特定のユーザグループ、或いはユーザ全員に対し、選択的に残す事を可能にする。

【０１３２】

図１３Ｂを用いて、前記ユースケース・シナリオから導出される、共通の対象への視覚的な好奇心により誘起されるネットワーク・コミュニケーションの事例を説明する。当該視覚的な好奇心により誘起されるネットワーク・コミュニケーションとして、異なる時空間内において、複数のユーザが各々に異なる状況で「桜」を眺めている様子で説明する。偶然桜の花（５６０）を目にしたユーザ１（５５０）が、「綺麗な桜だ」とつぶやき、別の時空間でユーザ２（５５１）が、「桜が満開だ」（５６１）とつぶやいている。一方で、離れた場所で水面を流れる花びらを見たユーザ４（５５３）が、「桜の花びらかな？」とつぶやくシーンである。この時、ユーザ３（５５２）が川面に桜の花びらが舞い落ちる様子を見て（５６２）、「花筏（はないかだ）だ」とつぶやいたとすると、このつぶやきは、同じ「花筏」を眺めている前記ユーザ４に、前記ユーザ３のつぶやきとして届ける事が可能になる。そして、偶然別の場所で桜の花を眺めているユーザ５（５５４）に対し、同じ時期に別の場所で「桜」を鑑賞している前記ユーザ１からユーザ４のつぶやきとして送り込む事が可能となり、その結果前記ユーザ５は「そうか、今週はちょうど桜の見頃を迎えているのだな」と、眼前の桜を前に各所の春の到来を感じる事が可能になる。この事例で示す様に、同様の対象やシーンに対し、それらを偶然目にする可能性のある異なる時

空間内に存在する複数のユーザ間で、共通する視覚的な関心に端を発した、広範な共有ネットワーク・コミュニケーションを誘起する事が可能となる。

【０１３３】

図１４で、リンク構造を用いて、本発明の一実施形態におけるユーザ、対象、キーワード、時間、時間帯、場所、地域、メッセージやつぶやき、及び／又は着目した対象が含まれる映像、及び特定のユーザ、特定のユーザ群、或いはユーザ全体をノードとした各要素間の許可の関係を説明する。本実施例では、これらの関係を全てグラフ構造で表現し、グラフデータベース３６５に記録する。全の関係をノード群とそれら相互のリンクからなるグラフ構造で表現する事で、例えば、リレーショナル・データベース（表構造）等を採用した場合に、事前に全てのノードの存在やノード間の関係や関連性を組み込んでおかなければならない、という実現不可能な要件から本質的に逃れる事が出来る。これらのノード群の中には、時間の経過と共に刻々と変化、及び成長する構造である性質を持っているノード群もある為、事前に全ても構造を予想し、設計しておく事は凡そ困難である。

【０１３４】

図１４に示す基本形では、対象６０１は、ユーザ（６００）ノード、キーワード（６０２）ノード、対象画像特徴（６０３）ノード、時間／時間帯（６０４）ノード、場所／地域（６０５）ノード、メッセージやつぶやき６０７の各々のノードとリンクしている。対象６０１には、ＡＣＬ（６０６）がリンクしている。メッセージやつぶやき（６０７）ノードには、ＡＣＬ（６０８）ノード、時間／時間帯（６０９）ノード、場所／地域（６１０）ノードがリンクしている。即ち、図１４は、ユーザの着目した対象と、その時間／時間帯、場所／地域、図３Ａに記載の手順３０－０１の過程で抽出された、及び／又は統計情報処理部３６３にて抽出された、及び／又は後述の会話エンジン４３０で抽出された、関連するキーワード及び着目対象に残されたユーザのメッセージやつぶやきが、ＡＣＬにて許可されている様子を表しているデータ構造である。なお、図１４に記載のグラフ構造は、ノードを追加、或いは削除する事で、前記記載の時間／時間帯、場所／地域、ＡＣＬには限定されない情報を記録する事が出来る様に構成しても良い。

【０１３５】

図１５を用いて、本発明の一実施例における一般物体認識システム１０６、特定物体認識システム１１０、及びシーン認識システム１０８に係るグラフ構造の抽出プロセスを説明する。まず一般物体認識システム１０６において当該対象が属するカテゴリを検出（９０１）する。次に、グラフデータベース３６５からカテゴリノードを検索し（９０２）、当該カテゴリがグラフデータベース３６５上に存在しているかの確認を行う（９０３）。存在していなければ新規カテゴリノードが追加されグラフデータベースに記録される（９０４）。次に特定物体認識システム１１０にて特定物体の検出を行い（９０５）、前記グラフデータベース上に既に存在しているかの確認を行う（９０７）。存在していなければ新規当該特定物体ノードを追加し（９０８）、それらをグラフデータベース上に記録する（９０９）。もう一方のパスにおいては、シーン認識システム１０８においてシーンの検出（９１０）を行い、グラフデータベース３６５からシーンノードを検索して（９１１）、当該シーンがグラフデータベースに存在しているかの確認を行う（９１２）。存在していなければ当該シーンに係るノードを生成し、前記グラフデータベースに追加する（９１３）。これら一連の処理が終了した時点で、当該カテゴリノード、特定物体ノード、或いはシーンノードに、上記処理を行ったタイムスタンプ情報をグラフデータベース上に追加記録し（９１４）、当該処理を終了する。

【０１３６】

前記図１５に記載のグラフデータベース３６５への登録の為の新規ノード群生成は、図３Ａに記載のユーザによる再確認処理の際に行っても良い。前記再確認処理では、前記音声認識システムにより抽出された単語列と、前記画像認識システムを備えた知識情報処理サーバシステム側で抽出された各種特徴とを対応付ける事が可能である。一例として、図４Ａに記載のタクシー５０に関し、前記サーバシステム側が、対象５１に対する画像認識結果として「それは赤いバスですか？」とユーザに音声による確認を求めてきた場合、ユ

ーザが「いいえ、黄色いタクシーです」と答えたとすると、前記サーバシステム側が再追加的な画像特徴抽出処理を行う事で最終的にタクシー５０を認識し、当該ユーザに対して「左側の黄色いタクシーを検出しました」と音声による再確認を発行し、それに対し当該ユーザは「そうです」と答えたとする。その結果、前記タクシー５０に係る検出された全ての特徴群を当該ビュー（シーン）に係る関連ノード群として、当該ユーザが確認した単語「タクシー」「黄色」に係るノード群と共に、前記グラフデータベース３６５内に登録可能になる。

【０１３７】

また、前記図１５に記載のカテゴリノード、特定物体ノード、或いはシーンノードにリンクされた上記タイムスタンプと、当該ユーザとの関係付けを行う事が出来る。この場合、当該ユーザの上記着目履歴を、上記獲得したインタレストグラフの部分グラフとして構成する事が出来る。これにより、当該対象に着目した特定の時空間における当該ユーザの着目対象、及びそれらに関連付けられた他のノード群に係る状況を、当該ユーザの音声或いはネットワーク端末２２０上のＧＵＩ経由で、前記画像認識システムを備えた知識情報処理サーバシステム３００側に問い合わせる事が可能になる。その結果として、前記サーバシステム側から、上記獲得したインタレストグラフの部分グラフにより導く事が可能な特定の時空間における当該着目対象に係る様々な状態を、当該ユーザに音声、或いは文字、写真、図形情報等で通知する事が可能となる。

【０１３８】

さらに、前記着目履歴は、画像認識システム３０１との協調動作により認識可能になった、特定物体、一般物体、人、写真、或いはシーンの名称に加え、当該操作を行った時空間情報、ユーザ情報、及び対象となる画像情報と共に、グラフデータベース３６５内にグラフ構造として蓄積される。従って前記着目履歴を、前記グラフ構造を直接参照・解析する事が可能な様に構成する事も可能となる。

【０１３９】

図１６を用いて、本発明の一実施例における画像認識システムを備えた知識情報処理サーバシステム３００において実行されるインタレストグラフの獲得に関して説明する。グラフ構造（１０００）は、或る時点でのユーザ（１００１）ノードのインタレストグラフである。当該ユーザは特定物体としての車種Ａ（１００３）ノードと車種Ｂ（１００４）ノードに興味があり、それらはカテゴリ「車」（１００２）ノードに属している。当該ユーザは、また、３つの対象（特定物体１００６から１００８）ノードに興味があり、それらはワイン（１００５）ノードに属している。次に、ユーザが対象車種Ｘ（１０１１）ノードに着目したとする。前記対象車種Ｘ（１０１１）ノードには、画像（１０１２）ノードと、他のユーザのメッセージやつぶやき（１０１３）ノードがリンクしているとする。前記サーバシステムは、前記対象車種Ｘ（１０１１）ノードを含むグラフ構造（１０１０）を車（１００２）ノードに繋ぐリンク（１０４０）を生成する。一方、前記統計情報処理部３６３により、例えば共起確率を計算する事で、ワイン（１００５）ノードに図中の３本のワイン（１００６から１００８）ノードがリンクされている時には、囲み１０２０にある２本のワイン（１０２１から１０２２）ノードも同様にリンクされている可能性が高まる。これにより前記サーバシステムは、当該ユーザに囲み（１０２０）を提案する事が出来る。その結果、当該ユーザが当該囲み（１０２０）に興味を示した場合には、それら囲み１０２０にある２本のワイン（１０２１から１０２２）ノードをワイン（１００５）ノードに直接繋ぐリンク（１０４１）を生成する事により、当該ユーザ（１００１）に係るインタレストグラフを継続的に成長させる事が可能になる。

【０１４０】

前記図１６に記載のインタレストグラフの成長がさらに進んだ状態における、ユーザ（１００１）ノードを中心とするグラフ構造のスナップショット例を図１７に示す。図は次の状態を表現している。ユーザ（１００１）ノードは、車（１００２）ノードとワイン（１００５）ノード以外に、特定のシーン（１０３０）ノードに関心がある。車（１００２）ノードでは、特に特定物体として車種Ａ（１００３）、車種Ｂ（１００４）、及び車種

Ｘ（１０１１）の各ノードに関心があり、ワイン（１００５）ノードでは５種のワイン（１００６、１００７、１００８、１０２１、及び１０２２）ノードに関心がある。特定のシーン（１０３０）ノードは、画像（１０３１）ノードで代表されるシーンであり、特定の時間（１０３３）ノードにおいて、特定の場所（１０３４）ノードで撮影され、ＡＣＬ（１０３２）ノードにリストされたユーザに対してのみ再生が許されている。車種Ｘ（１０１１）ノードは画像（１０１２）ノードで表現されており、そこに様々なユーザのメッセージやつぶやき（１０１３）ノードが残されていて、ＡＣＬ（１０３６）ノードにリストされたユーザ群に対してのみ、それらの再生が許可されている。車種Ａには、エンジンの仕様と色がノードとして記載されている。以下、５種のワイン（１００６、１００７、１００８、１０２１、及び１０２２）ノードに関しても同様の属性が記載されている。なお、これらのノードの一部は、他のユーザ２（１０３６）から直接リンクされても良い。

【０１４１】

図１８Ａを用いて、本発明の一実施形態におけるユーザのメッセージやつぶやきを音声として記録する手段、或いは再生する手段を説明する。まず、ユーザは図３Ａに記載の手順で対象を特定（１１０１）して変数Ｏにバインドする。次に当該メッセージやつぶやきを記録した時間、或いは再生を可能にする時間／時間帯（１１０２）を指定して変数Ｔにバインドし、当該メッセージやつぶやきを記録した場所、或いは再生を可能にする場所／地域（１１０３）を指定して変数Ｐにバインドする。次に、それらメッセージやつぶやきを受取る事が可能な受領者を指定（ＡＣＬ）して変数Ａにバインドする。そして、記録するか再生するかを選択（１１０５）し、記録処理の場合には当該メッセージやつぶやきの記録手順を実行する（１１０６）。その後、前記４つの変数（Ｏ、Ｔ、Ｐ、Ａ）から必要なノード群を生成し、グラフデータベース３６５に記録する（１１０７）。前記選択（１１０５）が再生処理の場合には、前記４つの変数（Ｏ、Ｔ、Ｐ、Ａ）から該当するノード群をグラフデータベース３６５から抽出（１１０８）し、前記ノードに残されたメッセージやつぶやきを再生する（１１０９）手順を実行して、一連の処理を終了する。

【０１４２】

図１８Ｂに、図１８Ａにおける再生時のステップ１１０２を詳細化して説明する。ユーザは音声によって時間／時間帯を指定するか、或いはネットワーク端末２２０上のＧＵＩによって直接時間／時間帯を指定するかを選択する（１１１１）。発話による場合には、ユーザは時間／時間帯を発話（１１１２）し、前記音声認識システム３２０で認識処理（１１１３）される。その結果が時間／時間帯であるか確認（１１１４）し、その結果が正しい場合は、指定時間／時間帯データを変数Ｔに格納する（１１１６）。違っている場合は、時間／時間帯を発話（１１１２）に戻る。処理を中断（ＱＵＩＴ）する場合は発話により終了する。一方、前記ネットワーク端末のＧＵＩにより時間／時間帯を指定する場合（１１１５）には、入力された時間／時間帯を直接前記変数Ｔに格納（１１１６）して、一連の終了処理をする。

【０１４３】

図１８Ｃに、図１８Ａにおける再生時のステップ１１０３を詳細化して説明する。ステップ１１２１で、ユーザは音声によって場所／地域を指定するか、ネットワーク端末２２０上のＧＵＩによって直接場所／地域を指定するかを選択する。発話による場合には、ユーザは場所／地域を発話（１１２２）し、前記音声認識システム３２０で音声認識処理（１１２３）される。その結果が発話された場所／地域であるか確認（１１２４）し、その結果が正しい場合は、緯度・経度データに変換（１１２７）してから変数Ｐに格納する（１１２８）。違っている場合は、場所／地域を発話（１１２２）するに戻る。処理を中断（ＱＵＩＴ）する場合は発話により終了する。一方、前記ネットワーク端末のＧＵＩにて地図を表示（１１２５）し、当該ネットワーク端末の画面上で直接場所／地域を指定する場合（１１２６）し、当該緯度・経度データを変数Ｐに格納して、一連の処理を終了する（１１２８）。

【０１４４】

図１９を用いて、本発明の一実施例として、特定の対象に残された複数のメッセージや

つぶやきの中から、受領対象者がそれらメッセージやつぶやきが残された時間或いは時間帯、及び／又は、残された場所或いは地域、及び／又は、残したユーザ名を指定可能にする事で、絞り込み再生する手順を説明する。説明の為の前提条件として、上記受領対象となるユーザは、図３Ａに記載した手順に従って当該対象に着目し、予め対応する対象となる各ノード群が選択されているとする（１１４０）。

【０１４５】

まず、当該対象に関して再生したい時間／時間帯、及び場所／地域を、図１８Ｂ、及び図１８Ｃに記載の手順で指定する（１２０１）。次に、誰の残したメッセージやつぶやきを再生するかを指定する（１２０２）。次にＡＣＬを確認し（１２０３）、当該指定条件に合致したメッセージやつぶやきに対応するノード、及び／又は、当該映像に対応したノードからデータを取り出す（１２０４）。この段階では、複数のノードが取り出される可能性があるので、その場合には、当該全ノードに関して次の処理を繰り返し適用する（１２０５）。

【０１４６】

次に当該メッセージやつぶやきを残したユーザに係る情報を、受領対象であるユーザに通知するか否かを選択する（１２０６）。通知する場合は、前記ノードに関連した当該メッセージやつぶやきを残したユーザ情報をグラフデータベース３６５から入手し、図１１に記載の再生処理部３０６を利用して上記受領対象ユーザが装着しているヘッドセットシステム２００、或いは／又は、上記受領対象ユーザに紐付けられているネットワーク端末２２０に音声、及び／又は、文字で通知する（１２０８）。通知内容が音声の場合には、ヘッドセットシステムに組み込まれたイヤフォンから再生され、文字、写真、及び／又は図形の場合には、前記ネットワーク端末上にそれら音声以外の情報が当該メッセージやつぶやきに同期して表示される（１２０９）。上記ユーザ情報を通知しない場合には、当該音声ノードから上記メッセージやつぶやき、及び／又は、当該映像ノードから対応する画像データを取り出し、前記再生処理部３０６を利用して、上記受領対象ユーザが装着しているヘッドセットシステム２００、及び／又は、上記受領対象ユーザに紐付けられているネットワーク端末２２０に、当該メッセージやつぶやきを残したユーザ情報を含まない音声、及び／又は、画像情報として送出し（１２０７）、それらの一連の処理を、前記取り出された全ノードに関して繰り返し終了する。

【０１４７】

前記実施例では、ループ（１２０５）で取り出された全ノードに関して繰り返し処理しているが、他の手段を用いても良い。例えば、状況認識部３０５を利用して受領対象ユーザに適切なメッセージやつぶやきを選び出し、上記メッセージやつぶやきのみ、及び／又は、付帯している映像情報と共に再生しても良い。前記、時間／時間帯と場所／地域の指定（１２０１）に係る説明では、過去に記録されたメッセージやつぶやき、及びそれらの基になる画像情報に関して時空間を過去に遡って受領する目的で、特定の時間／時間帯、及び場所／地域を指定する事例を示したが、逆に未来の時間／時間帯及び場所／地域を指定しても良い。その場合には、当該指定された未来の時空間に、当該メッセージやつぶやき、及びそれらの基となる映像情報を“タイムカプセル”に乗せて届ける事が可能になる。

【０１４８】

また、当該メッセージやつぶやきの再生に同期して、当該着目対象に関する詳細情報を前記ネットワーク端末上に表示しても良い。さらに、受領対象ユーザの主観的視野外となっている対象に向け、前記画像認識システムを備えた知識情報処理サーバシステム側が音声情報により、当該受領対象ユーザに対し、当該メッセージやつぶやきが残された対象に向け頭を動かす、或いは当該対象の存在する方向に向かって移動する等の指示を与え、その結果、受領対象ユーザが当該対象をその主観的視野内に捉えた時に、当該対象に残されたメッセージやつぶやきを再生する様に構成しても良い。また、類似の効果が得られる別の手段を用いても良い。

【０１４９】

上記、メッセージやつぶやきの再生においては、前記状況認識部の一構成要素である履歴管理部４１０によって、その時々の再生位置が該当するノード内に記録されるので、受領対象ユーザが同一対象に再び着目した場合、以前と同一のメッセージやつぶやきを再び繰り返す事なく、前回の続きから、或いはそれ以降に更新されたメッセージやつぶやきを加え、受領する事を可能とする。

【０１５０】

次に、図２０を用いて、ユーザが眼前のとある対象に着目している事を、前記画像認識システムを活用して前記知識情報処理サーバシステム側に明示的に指し示す一つの方法として、当該ユーザの音声による指示によらず、当該着目対象に向けユーザが直接手指でポインティングする、或いは当該対象に手指で直接触れる事により、当該ユーザのヘッドセットシステムに組み込まれたカメラ映像から得られる画像情報を基に、前記画像認識システム側がリアルタイムに画像解析して当該着目対象を特定する実施例について説明する。

【０１５１】

図２０（Ａ）は、ユーザの主観視（１３００）事例である。ここでは、ワイン（１３０１）、アイスペール（１３０４）、及びそれ以外の２つの物体（１３０２、１３０３）が検出されている。ここでユーザは左側のワイン（１３０１）に着目している事を前記サーバシステム側に明示的に通知する為に、当該ユーザの手指（１３１０）でワインを直接指し示している状態を表している。ユーザはまた着目対象であるワイン（１３０１）に直接触れる事も出来る。また、指で指し示す代わりに、身近にある棒状の道具を使って指し示し、或いはレーザーポインター等の光線を対象に直接照射しても良い。

【０１５２】

図２０（Ｂ）に、手指（１３１０）による対象のポインティング手順を説明する。前提条件として、図２０（Ａ）の画面はユーザの主観的な視野を反映したカメラからの映像であるとする。まず、画面中から、手指（１３１０）を含むユーザの手（１３１１）を検出する。当該カメラ映像を前記画像認識システムにより画像解析し、そこから検出された手指（１３１０）及び手（１３１１）の形状特徴から主要なオリエンテーション（１３１２）を求め、手指（１３１０）が指し示す方向を抽出する。上記オリエンテーション（１３１２）の検出は、ネットワーク端末２２０側に組み込まれた画像認識エンジン２２４によりローカルに実行しても良い。

【０１５３】

前記オリエンテーションが検出されれば（１３２２）、そのベクトル線上にユーザが指し示す対象が存在する可能性が高い。次に、図２０（Ａ）の画像から、前記画像認識システム３０１との協調動作により当該ベクトル線上に存在する物体を検出し（１３２３）、当該対象物体の画像認識処理を実行する（１３２４）。当該画像検出及び認識処理は、ユーザのネットワーク端末２２０側の一構成要素である認識エンジン２２４上で行う事も可能で、ネットワーク側の負荷を大幅に軽減する事が出来る。また、ユーザによる素早いポインティング操作に対しても、レイテンシ（時間遅れ）の少ない高速なトラッキング処理が可能になる。最終的な画像認識結果は、ネットワークを介して前記画像認識システムを備えた知識情報処理サーバシステム３００側に問い合わせする事で確定され、ユーザに当該認識対象の名称等が通知される（１３２５）。当該ポインティング対象の画像認識結果がユーザの意にかなえば当該ポインティング処理を終了し（１３２５）、結果がユーザの意と異なる場合は、追加の指示要求を発行（１３２７）してステップ（１３２２）に戻り、引き続きポインティング操作を続ける。同様に、当該ユーザが着目対象のポインティングを明示的に確認しなかった場合に、当該検出結果がユーザの意図通りではなかったと推定して上記の処理を繰り返す、或いは無言の同意と見做して当該検出処理を終了するかを予め設定しておく、或いは前後の流れから、或いは個々のユーザの癖を学習する事により、適応的に当該判断内容を振り分ける事が出来る様に構成しておく事が可能である。これらのユーザによる確認にはユーザの音声による指示を用いるが、それに代わる同様の効果をもたらす手段を用いても良い。

【０１５４】

また、当該ユーザにおける前記一連のポインティング操作の過程で、前記画像認識システムを備えた知識情報処理サーバシステム３００と当該ユーザの間で、インタラクティブなコミュニケーションを行う事が可能である。例えば図２０（Ａ）の画像において、前記オリエンテーション１３１２が指し示す方向が前記１３０２上に向かった時に、「対象は１３０２ですか？」と前記知サーバシステムが当該ユーザに対し確認する事で、当該ユーザが「そう。けれども、これは一体何かな？」と改めて質問し直す事も可能となる。

【０１５５】

次に、本発明の一実施例において、前記ヘッドセットシステム２００に具備された位置情報センサ２０８を用い、当該ヘッドセットシステムの移動状態を都度検出する事で、当該ヘッドセットシステムを装着したユーザが、或る対象へ着目し始めた可能性を検出する手順を説明する。

【０１５６】

図２１は、当該ヘッドセットシステム２００の動作に関しての状態遷移を表している。動作開始（１４００）状態は、当該ヘッドセットシステムが一定の静止状態から動き出す状態である。当該ヘッドセットシステムの動きには、当該ヘッドセットシステム自体の並行移動（上下、左右、前後）に加えて、当該ヘッドセットシステム自体の位置はそのままで、ユーザの首振り動作によりその向きを変える（左右を見る、上下を見る）動きを含む。停止（１４０３）は、当該ヘッドセットシステムが静止している状態である。短時間静止（１４０４）状態は、一時的に当該ヘッドセットシステムが静止している状態である。長時間静止（１４０５）状態は、当該ヘッドセットシステムがしばらくの間静止している状態である。当該ヘッドセットシステムが一定の動作状態から静止した場合、停止（１４０３）状態に遷移（１４１０）する。停止（１４０３）状態が一定時間以上続いた場合、短時間静止（１４０４）状態に遷移（１４１１）する。短時間制状態（１４０４）がその後一定時間以上継続し、さらに長時間静止している場合には、長時間静止状態（１４０５）に遷移（１４１３）する。短時間静止状態（１４０４）、或いは長時間静止状態（１４０５）から当該ヘッドセットシステムが再び動き出すと、再び動作開始（１４００）状態に遷移（１４１２、或いは１４１４）する。

【０１５７】

これにより、例えば短時間静止（１４０４）状態に前記ヘッドセットがある時には、ユーザが何か眼前の対象を着目し始めている可能性があると判断して、前記画像認識システムを備えた知識情報処理サーバシステム３００側に対し着目開始を予告すると同時に、前記ヘッドセットシステムに組込まれたカメラを自動的に撮影開始状態に投入し、引き続く一連の処理に備えるきっかけとする事が出来る。また、前記ヘッドセットシステムを装着したユーザの言外の反応、例えば首を傾げる（疑問）、首を左右に振る（否定）、首を上下に振る（同意）等の動作を、当該ヘッドセットシステムに具備された位置情報センサ２０８から検出可能なデータから検出する事も可能になる。これらのユーザが多用する首振りのジェスチャーは、地域の風習やユーザ毎の癖によって異なる可能性がある。従って、前記サーバシステム側で、それらユーザ個々の、或いは地域特有のジェスチャーを学習の上で取得して、当該属性を保持し反映する必要がある。

【０１５８】

図２２に、本発明の一実施例における写真抽出の事例を示す。写真画像は、視点位置に従いアフィン変換された矩形領域に囲われている閉領域と想定し、当該領域内から検出される物体のサイズがその領域外にある物体のサイズと大幅に異なるスケールで存在している場合、或いは特定の領域に含まれる本来立体であるべき一般物体、或いは特定物体から抽出される各特徴点が、ユーザの視点移動に伴う相対位置変移を起こさず、当該特定の閉領域内で平行移動する場合、或いは画像の奥行き情報を直接検出可能なカメラから獲得可能な対象との距離情報、或いは複数のカメラ画像による両眼視差から獲得可能な物体の奥行き情報等が取得可能な場合において、本来立体であるべき物体やシーンに係る特徴点が同一平面上に存在する場合に、当該閉領域が平面的な印刷物や写真である可能性が高いと推定する事が可能となる。似た様な状況として、窓外の景色も同様の条件を満たし得るが

、それが窓であるか平面画像であるかは周囲の状況から或る程度推定可能になる場合もある。また、それらが写真である可能性が高いと推定された場合、それらの写真自体を一つの特定物体と見なして、前記画像認識システムを備えた知識情報処理サーバシステム３００側に問い合わせる事で、類似写真の検索が可能になる。その結果、同様或いは類似の写真画像が発見されれば、以降異なる時空間内において同様或いは類似の写真画像を眺めている、或いは眺めた、或いは眺める可能性のある、他のユーザ群を繋ぐことが可能になる。

【０１５９】

図２３Ａ及び図２３Ｂを用いて、本発明の一実施形態における着目対象との会話に関して説明する。前提としてユーザの着目画像をカメラが捉えているとする（１６００）。ユーザの主観的視野を反映したカメラ画像から、ネットワーク上の画像認識システム３０１との協調動作により、図３Ａに記載の着目対象の抽出プロセスにより、当該対象となる画像を認識する（１６０２）。次に、グラフデータベース３６５から着目対象に関するグラフ構造を抽出し、当該着目対象に残されたメッセージやつぶやきに係るノード群を抽出する（１６０３）。次に、それらメッセージやつぶやきの受領対象者を指定したＡＣＬを確認し（１６０４）、その結果として上記対象ノード群に関連付けられたメッセージやつぶやきを、当該ユーザのヘッドセットシステム２００、或いはネットワーク端末２２０に、音声、画像、図形、イラスト、或いは文字情報で通知する（１６０５）事が出来る。

【０１６０】

本発明では、上記メッセージやつぶやきに対して、当該ユーザが発話（１６０６）によってさらに着目対象に向かい会話的に話しかける仕組みを提供する。前記発話内容は、前記音声認識システム３２０との協調動作により認識され（１６０７）、発話文字列に変換される。当該文字列は会話エンジン４３０に送られ、当該ユーザに係るインタレストグラフを基に、前記知識情報処理サーバシステム３００側の前記会話エンジン４３０によって、時々の最適な話題が選択され（１６０８）、前記音声合成システム３３０経由で当該ユーザのヘッドセットシステム２０１に、音声情報として届ける事が可能になる。これにより当該ユーザは、継続的な音声コミュニケーションを前記サーバシステムとの間で続ける事が可能になる。

【０１６１】

前記会話内容が、ユーザによる当該着目対象そのものに係る質問等の場合は、前記知識情報処理サーバシステム３００が、当該質問に対する応答を前記ＭＤＢ１１１内に記載の詳細情報、或いは当該着目対象に係る関連ノード群から引き出し、当該ユーザに音声情報により通知する。

【０１６２】

逆に、前記サーバシステム側から当該ユーザに対し、当該ユーザのインタレストグラフ基にその時々の話題に係る関連ノード群を辿って継続的な話題を抽出し、タイムリーに提供する事が出来る。その場合には、同じ話題が不必要に繰り返し提供されない様に、当該会話の流れの中で以前触れた事のある話題に係るノード群それぞれに対し、上記会話の履歴情報を記録しておく事で回避が可能になる。また、当該ユーザにとり関心がない話題に不必要に向かう事により、当該ユーザの好奇心が殺がれない様にする事も大事となる事から、当該ユーザに係るインタレストグラフを基に、抽出される話題を選択する事が出来る。上記継続的な会話は、当該ユーザによる発話が続く限り、ステップ１６０６に戻り繰り返され、当該ユーザの発話がなくなるまで続き（１６０９）、その後終了する。

【０１６３】

上記における広範なユーザと前記知識情報処理サーバシステム３００間の双方向の会話は、前記インタレストグラフ部３０３自体の学習パスとしても重要な役割を果たす事が出来る。特に、ユーザが特定の対象、或いは話題に対して頻繁に会話を促す場合には、当該ユーザが当該対象、或いは話題に対し極めて強い関心があるとして、それら関心に係るノードと当該ユーザに係るノードの直接或いは間接のリンクに対し、重み付けを加える事が可能となる。逆に、ユーザが特定の対象、或いは話題に対して継続的な会話を拒む場合に

は、当該ユーザが当該対象、或いは話題に対し興味を失った可能性があるとして、それら対象や話題に係るノードと当該ユーザに係るノードの直接或いは間接のリンクに対し、重み付けを減じる事も可能となる。

【０１６４】

前記実施例では、ユーザが着目対象をその視野内に捉えてからのステップを、順を追って説明したが、他の実施形態をとっても良い。例えば、図３Ａに記載の手順において、途中のステップから当該ユーザと前記知識情報処理サーバシステム３００間の双方向の会話を始める様に本実施形態を構成しても良い。

【０１６５】

図２３Ｂに、本発明の一実施形態における会話エンジン４３０の一構成例を示す。前記会話エンジンへの入力は、対象ノードを中心とするグラフ構造１６４０と、音声認識システム３２０からの発話文字列１６４１である。前者は関連ノード抽出１６５１により前記対象に関連する情報を取り出し、キーワード抽出１６５０に送る。ここでは、前記発話文字列と前記情報を基に、オントロジー辞書１６５２を参照して複数のキーワード群を抽出する。次に、話題抽出１６５３にて前記複数のキーワード群から１つを選択する。ここでは、同じ会話を繰り返さない為の話題の履歴管理を行う。また、上記キーワード抽出に当たっては、新しい、他のユーザにより参照頻度の高い、或いは当該ユーザの関心の高いキーワード群を優先して抽出するように構成する事も出来る。適切な話題が抽出された後は、反応文生成１６５４にて会話パターン辞書１６５５を参照しながら、自然な口語体に変換された反応文が作成１６４２され、後段の音声合成システム３３０に引き渡される。

【０１６６】

本実施例における前記会話パターン辞書１６５５は、前記キーワード群から想起される文章のルールを記述している。例えば、「Ｈｅｌｌｏ！」とのユーザ発話に対しては「Ｉ’ｍ　ｆｉｎｅ　ｔｈａｎｋ　ｙｏｕ．　Ａｎｄ　ｙｏｕ？」と返答するとか、「Ｉ」とのユーザ発話に際しては「ｙｏｕ」と返答するか、「Ｉ　ｌｉｋｅ　ｉｔ．」とのユーザ発話に対しては「Ｗｏｕｌｄ　ｙｏｕ　ｌｉｋｅ　ｔｏ　ｔａｌｋ　ａｂｏｕｔ　ｉｔ？」と返答するといった代表的な会話のルールを記述している。返答のルールには変数を含めて良い。その場合、当該変数はユーザの発話から充当される。

【０１６７】

前記構成により、前記知識情報処理サーバシステム３００側が、当該サーバシステム内に収納された前記インタレストグラフ部３０３内に記載の内容から、当該ユーザの関心に沿ったキーワード群を選び出し、前記インタレストグラフを基に適切な反応文を生成する事で当該ユーザにとって引き続き会話を続ける強い動機になると同時に、対象と会話しているような感覚を抱くように構成する事も可能になる。

【０１６８】

また、グラフデータベース３６５には、自らを含む特定のユーザ、或いは特定のユーザ群、或いはユーザ全体に対応するノード群が記録され、それらは、特定物体、一般物体、人、写真、或いはシーンに関するノード群、及びそれらに対して残されたメッセージやつぶやきを記録したノード群が相互にリンクされ、グラフ構造を構成している。統計情報処理部３６３により、前記メッセージやつぶやきに関連するキーワード群を抽出し、状況認識部３０５により選択的に当該ユーザのヘッドセットシステム２００、或いはネットワーク端末２２０に、関連する音声や画像、図形、イラスト、或いは文字情報で通知する様に本実施例を構成しても良い。

【０１６９】

図２４を用いて、本発明の一実施例として２以上のヘッドシステム２００が一台のネットワーク端末２２０に接続された際の、前記ヘッドセットシステム間の協調動作に関して説明する。図２４では、４人のユーザが各々ヘッドセットシステム２００を装着しており、各々のユーザが見ている方向が図示されている。この際に、共有する前記ネットワーク端末上に位置のキャリブレーションを行うマーカー等を表示し（１７０１から１７０４）、それを各ユーザのヘッドセットシステムに組込まれたカメラで常時モニタリングする事

で、各々のユーザの相互の位置関係、及びその動きを把握する事が出来る。或いは、時間軸変調された画像パターンを当該共有ネットワーク端末の表示デバイス上に表示して、それらを各ユーザのヘッドセットシステムに具備されたカメラ映像で捉えた後に復調して、同様の位置関係を求めても良い。これらにより、各々のカメラの視野と視線のキャリブレーション、及び各ユーザのヘッドセットシステムと当該共有ネットワーク端末とのキャリブレーション、及びトラッキング処理を自動的に行う事で、前記ネットワーク端末は各々のユーザの位置を常に知る事が出来る。それにより、当該共有ネットワーク端末上のＧＵＩ操作に関して、どのユーザからの入力操作であるかを当該ネットワーク端末側が認識する事が可能になる。それにより、当該共有ネットワーク端末の共有表示デバイス上で、各々のユーザの位置を考慮した、各ユーザに向けたアライメントを有するサブ画面群の表示が可能になる。

【０１７０】

図２５を用いて、本発明の一実施例として、前記画像認識システムを備えた知識情報処理サーバシステム３００では認識出来なかった不明な着目対象に対し、当該ユーザが当該対象に係る質問をネットワーク上に残す事を可能にし、他のユーザがネットワーク経由でそれらの不明な対象に対する新たな情報や回答を寄せる事で、当該不明となった着目対象を、前記サーバシステム側が、それらユーザ間のやりとりの中から必要な情報を選択抽出し学習する手順を説明する。

【０１７１】

前記手順１８００は、ユーザによる音声入力トリガ１８０１から始まる。前記音声入力トリガは、ユーザによる特定の単語の発話、マイクが拾う音圧レベルの急変、或いは前記ネットワーク端末部２２０のＧＵＩによっても良い。また、それらの方法に制限されない。それによりカメラ画像のアップロードが開始され（１８０２）、音声コマンド待ち（１８０３）となる。次に、ユーザが着目対象抽出の為のコマンド群を音声により発話する事で、それらが音声認識処理され（１８０４）、例えば図３Ａに記載の手段を使って音声による着目対象のポインティング処理が正しく完了したかが判断される（１８０５）。上記ポインティング処理が困難で認識対象をうまく指定出来ない場合には（１８０６）、新たな特徴追加による再試行が可能か判断される（１８０７）。再試行が可能な場合にはユーザからの音声コマンド入力待ち（１８０３）に戻り、再試行する。一方、特徴の追加が困難な場合には、ネットワーク上のＷｉｋｉへの問い合わせを開始する（１８０８）。

【０１７２】

前記問い合わせ処理では、当該問い合わせ対象に係るカメラ画像、及びユーザの音声による質問やコメントをセットにして、ネットワーク上に発行する（１８０９）。それに対しＷｉｋｉから新たな情報提供や回答があれば回収し（１８１０）、その内容を当該ユーザ、或いは多数のユーザ群、及び／又は、前記知識情報処理サーバシステム３００側が検証する（１８１１）。当該検証処理では、寄せられた回答の正当性を判断する。検証に合格すれば、対象を新規登録する（１８１２）。当該新規登録に当たっては、前記質問、コメント、情報、回答に対応する各ノード群を生成し、当該対象に係るノード群として関連付け、グラフデータベース３６５に記録する。前記検証に不合格の場合には、保留処理１８２２を行う。当該保留処理では、ステップ１８０８或いはステップ１８１８におけるＷｉｋｉへの問い合わせ処理が未完了である旨を記録し、前記検証に合格する回答が収集されるまでステップ１８１０のＷｉｋｉからの情報・回答収集処理をバックグラウンドで続行する。

【０１７３】

前項ステップ１８０５にて、対象の音声によるポインティング処理が可能だった場合、当該対象の画像認識プロセスに移行する（１８１３）。当該画像認識処理は、本実施例では特定物体認識システム１１０にて特定物体認識を行い、認識出来なかった場合には一般物体認識システム１０６にて一般物体認識を行い、さらに認識出来なかった場合にはシーン認識システム１０８にてシーン認識を行う様を図示しているが、これらの各画像認識処理自体は、本事例のように必ずしも直列的に実行せず、各々を個別に並列、或いは各々の

認識ユニットの中をさらに並列化して実行しても良い。或いは、その各々を最適化した上で組み合わせても良い。

【０１７４】

前記画像認識処理が成功し、対象が認識可能となった場合、ユーザに対する音声による再確認のメッセージが発行され（１８２０）、それをユーザが正しく確認出来た場合には、カメラ画像のアップロードを終了（１８２１）して前記一連の対象画像認識処理を終了する（１８２３）。一方、ユーザが正しく確認出来なかった場合には、当該対象は未確認のままであるとして（１８１７）、ネットワーク上のＷｉｋｉへの問い合わせが開始される（１８１８）。Ｗｉｋｉへの問い合わせに際しては、当該問い合わせ対象画像も一緒に発行する（１８１９）必要がある。ステップ１８１０では、Ｗｉｋｉから寄せられた新たな情報や回答群に対し、その内容及び正当性を検証する（１８１１）。検証に合格すれば、対象を登録する（１８１２）。当該登録に当たっては、前記質問・コメント及び情報・回答に対応するノード群を生成し、当該対象に係るノード群に関連付けてグラフデータベース３６５に記録する。

【０１７５】

図２６を用い、前記ヘッドセットシステム２００に具備された位置情報センサ２０８を利用する一実施例を説明する。前記位置情報センサには、ＧＰＳ（Ｇｌｏｂａｌ　Ｐｏｓｉｔｉｏｎｉｎｇ　Ｓｙｓｔｅｍ：全地球測位システム）を利用しても良いが、それには限定されない。前記位置情報センサで検出された位置情報及び絶対時間を、前記ヘッドセットシステムに具備されたカメラ２０３が撮影した画像に付加し、前記画像認識システムを備えた知識情報処理サーバシステム３００側にアップロードする事で、グラフデータベース３６５が記録している情報を較正する事が出来る。図２６（Ａ）は、当該アップロード前の、前記グラフデータベースの画像５０４（図１３Ａ）に関係するグラフ構造の一実施例である。「太陽」が「真上」であるので、時間帯は昼頃であると推定可能になる。図２６（Ｂ）は、前記画像アップロード後の、グラフ構造の一例である。「絶対時間」ノードの追加により、当該画像に対応した時刻が正確に確定可能になる。また、上記位置情報センサ２０８により検出された位置情報自体に内在する誤差を、カメラの撮像画像から前記サーバシステムによる認識結果により較正する事が可能になる。

【０１７６】

さらに、前記画像５０４が、前記グラフデータベース３６５内に存在しなかった場合、前記図２５における一実施例と同様の手順を用いて、前記画像５０４に関係する情報をグラフ構造として前記グラフデータベース３６５に記録する。その際に、前記位置情報と絶対時間を利用して、近傍にいる他のユーザ群に対して、前記画像５０４に関する質問を発行する事で、ユーザ間の新たなネットワーク・コミュニケーションを誘発する事が可能になり、そこから得られる有用な情報群を、前記画像５０４に係るグラフ構造に追加する様に前記サーバシステムを構成する事が可能になる。

【０１７７】

さらに、前記画像認識システムを備えた知識情報処理サーバシステム３００においてアップロードされた画像中の物体が不審物体として判断された場合には、当該不審物体を画像解析して入手可能になった情報を前記グラフデータベース３６５に、かかる不審物体に係る情報群として記録する事が出来る。当該不審物体の存在或いは発見を、事前に設定可能な特定のユーザ、或いは機関に速やかに自動通知しても良い。前記不審物体か否かの判断には、予め登録済みの不審物体、或いは平常状態における物体との照合を前記グラフデータベース３６５との協調動作により行う事が出来る。その他、不審な状況、或いは不審なシーンが検出された場合にも、係る不審な状況、或いはシーンが検出可能になる様に本システムを構成しても良い。

【０１７８】

また、ユーザが予め指定可能な発見対象とした特定物体、一般物体、人、写真、或いはシーンを、ユーザのヘッドセットシステム２００に装着したカメラが偶然捉えた場合、当該ヘッドセットシステムに有線或いは無線で接続されるユーザのネットワーク端末２２０

上に、前記画像認識システムを備えた知識情報処理サーバシステム３００側からネットワーク経由で予めダウンロードされ常駐可能となっている特定画像検出フィルタ群が、当該特定物体、一般物体、人、写真、或いはシーンの初期的な抽出及び対象の暫定的な認識を行い、その結果としてさらに詳細な画像認識処理が必要となった場合には、ネットワーク経由で前記サーバシステム側にそれらを詳細に問い合わせる事で、探し物や忘れ物等、或いは発見したい対象をユーザが前記サーバシステム側に登録しておく事で、効果的に見つけ出す事が可能になる。

【０１７９】

なお、当該発見対象の指定には、ユーザのネットワーク端末２２０上でのＧＵＩを用いても良い。或いは、前記画像認識システムを備えた知識情報処理サーバシステム３００側が、特定の発見対象画像に係るデータ、及び必要な検出フィルタ群を前記ユーザのネットワーク端末上にプッシュして、当該サーバシステム側が指定した発見対象を、広範なユーザ間で共同して探索する事が可能になる様に構成しても良い。

【０１８０】

前記特定画像検出フィルタ群を、前記画像認識システムを備えた知識情報処理サーバシステム３００側から抽出する一実施事例として、前記サーバシステム内の前記グラフデータベース３６５内から前記指定された発見対象に係るノード群を部分グラフとして取り出し、当該指定された発見対象に係る画像特徴群を、それら部分グラフを基に抽出する事で、当該対象を検出する為に最適化された前記特定画像検出フィルタ群を獲得する事が可能になる様に構成しても良い。

【０１８１】

また、本発明に係る一実施例として、ユーザが装着しているヘッドセットシステム２００とネットワーク端末２２０を一体として構成しても良い。また、前記ヘッドセットシステムにネットワークに直接接続可能な無線通信システム、及びユーザの視野の一部を覆う形で半透明の表示ディスプレイを組み込み、前記ヘッドセットシステム自体に前記ネットワーク端末の一部、或いは全体の機能を組み込んで一体として構成しても良い。これらの構成により前記ネットワーク端末を利用しなくとも、前記画像認識システムを備えた知識情報処理サーバシステム３００側と直接通信する事が可能になる。その際には、前記ネットワーク端末に組込まれたいくつかの構成要素は、一部統合・修正する必要がある。例えば、電源部２２７は当該ヘッドセットの電源部２１３と統合可能になる。また、表示部２２２も画像出力装置２０７に統合する事が可能になる。当該ヘッドセットシステムにおける無線通信装置２１１は、前記ネットワーク端末間の通信を担っていたが、それらもネットワーク通信部２２３に統合可能になる。その他の画像特徴検出部２２４、ＣＰＵ２２５、及び記憶部２２６は、当該ヘッドセットに組み込む事が可能になる。

【０１８２】

図２８に、サーバとのネットワーク接続が一時的に切断されている状況下における、ネットワーク端末２２０単体での処理の一実施例を示す。ネットワーク接続の一時的な中断は、トンネル内やコンクリートで覆われた建物内への移動、航空機での移動中等で頻繁に発生する可能性がある。また、様々な理由で電波状況が悪化する場合や、無線基地局毎に設定されているセル最大接続数を超えてしまった場合等に、ネットワーク接続速度が大幅に低下する傾向がある。この様な状況下でも、前記画像認識を行う対象の種類と数を必要最小限度に絞り込み、音声コミュニケーション機能を特定の会話内容に限定する事で、予めネットワーク接続が確立している時に、前記ネットワーク端末側の一時記憶メモリ容量内、或いはフラッシュメモリ等の二次記憶メモリ容量内にユーザが指定可能な限定された数の特定物体、一般物体、人、写真、或いはシーンの検出、判別、及び認識に必要な学習済みの特徴データ群、及び当該限定された数の対象群の検出・認識する為に最適な画像検出・認識プログラムのサブセットを、上記各特徴データ群と共に一体として前記サーバシステム側から前記ネットワーク端末側に予めダウンロードしておく事で、ネットワーク接続が一時的に中断した場合でも一定の基本動作が可能になる様に構成する事が出来る。

【０１８３】

上記の機能を実現する為の一実施例を以下に示す。図２８（Ａ）及び（Ｆ）にユーザが装着するヘッドセットシステム２００、及びユーザのネットワーク端末２２０の主要機能ブロック構成を示す。一般的なネットワーク端末は、内蔵するＣＰＵ２２６により様々なアプリケーションがネットワーク・ダウンロード可能なソフトウェアの形で常駐可能となっている。それらの実行可能なプログラム規模や参照可能な情報量或いはデータ量自体は、サーバ上における構成に比べて大幅な制約は課されるものの、前記画像認識システムを備えた知識情報処理サーバシステム３００側に構築される各種プログラムやデータの実行サブセットを一時的にユーザの前記ネットワーク端末に常駐させる事で、前記の様に最小限度の実行環境の構築が可能となる。

【０１８４】

図２８（Ｄ）に、サーバ側に構築された画像認識システム３０１の主要機能ユニット構成を示す。この中で、特定物体認識システム１１０、一般物体認識システム１０６、シーン認識システム１０８においては、本来その要求される画像認識対象として、過去も含め現在に至るまで存在する、或いは存在していた全ての固有名詞／一般名詞を付す事が可能な、物体、人、写真、或いはシーン全体に及ぶ。これら無限とも言える種類及び対象に本来は備えなくてはならない事と、今後の継続的な物体や事象の発見や認識対象アイテムの増加に伴う追加学習も必要となり、その全体の実行環境自体は極めて限られた情報処理能力やメモリ容量しか持ち合わせないネットワーク端末の手に到底及ぶものではなく、それらの包括的な機能はネットワークを介しサーバ側の強力なコンピュータ・リソース、及び巨大なデータベースシステム上に置かれる事になる。その上で、その時々で都度必要な機能部分について、非力なクライアント機器でも実行可能な画像認識機能のサブセットや、予め学習済みの知識データ等の必要な部分を、ネットワーク経由で当該ネットワーク端末上に選択的にダウンロードする事で、ネットワーク接続の切断に或る程度備える事が出来る。これには、不測のネットワーク切断に備えると言う目的以外に、サーバ・リソースへのアクセス集中による負荷軽減や、ネットワーク回線の不要なトラフィックを抑制するという実用的な側面もある。

【０１８５】

これらを実現する一実施形態として、図２８（Ｄ）に示す特定物体認識システム１１０、一般物体認識システム１０６、シーン認識システム１０８から選択した画像認識プログラムの必要なプログラム群を、ネットワークを介し図２８（Ａ）に示すネットワーク端末２２０上で実行可能な画像認識プログラム２２９として、認識エンジン２２４上にサーバ側からダウンロードの上で常駐させ、併せて各認識対象に即し必要な学習済みの特徴データ群を画像カテゴリデータベース１０７、シーン構成要素データベース１０９、及びＭＤＢ１１１から抽出し、同様にユーザのネットワーク端末２２０上の記憶部２２７上に選択的に常駐させる。これら対象となる認識対象候補群と、他のユーザによる当該対象候補群に対するメッセージやつぶやきを関連付ける為に、サーバ側の前記画像認識システムを備えた知識情報処理サーバシステム３００側から、必要な当該対象との関連性を前記グラフデータベース３６５から抽出すると共に、前記メッセージデータベース４２０から必要な会話候補群を抽出し、ネットワークを介し予めユーザのネットワーク端末２２０上のメッセージ管理プログラム２３２上にダウンロードしておく。これらユーザのメッセージやつぶやきの候補群は、限られた容量のメモリを効果的に使用する目的で、圧縮して当該ネットワーク端末２２０上の記憶部２２７内に格納する事が出来る。

【０１８６】

一方、前記画像認識システムを備えた知識情報処理サーバシステム３００側との双方向の音声による会話機能については、ネットワーク端末２２０上の音声認識プログラム２３０、及び音声合成プログラム２３１により一定の制限下で実行可能になる。その為には前記一実施例において、前記サーバシステム側とのネットワーク接続が確立しているタイミングで、前記サーバシステムを構成する会話エンジン４３０内の音声認識システム３２０、音声合成システム３３０、及びそれらに対応する知識データベースである音声認識辞書データベース３２１、会話パターン辞書１６５５から、必要最小限の実行プログラム群、

及びデータセットをユーザのネットワーク端末２２０上の記憶部２２７内に予めダウンロードしておく必要がある。

【０１８７】

上記において、ユーザのネットワーク端末２２０の処理能力、或いは記憶部２２７の記憶容量に十分な余裕がない場合には、予め会話の候補群をネットワーク上の音声合成システム３３０で音声化した後に、圧縮音声データとしてユーザのネットワーク端末２２０上の記憶部２２７上にダウンロードしておいても良い。これにより、ネットワーク接続に一時的に障害が生じても、主要な音声コミュニケーション機能は限定的ではあるが保持する事が可能になる。

【０１８８】

次に、ネットワークへの再接続時のプロセスについて説明する。ユーザが着目した様々な対象に係るカメラ画像、及び当該対象に対してユーザが残したメッセージやつぶやき等が、関連する様々な情報と共にユーザのネットワーク端末２２０上の記憶部２２７内に一時的に保持されているとする。そこで再びネットワーク接続が復帰した時点で、ネットワーク上の生体認証システム３１０内の生体認証処理サーバシステム３１１、及び個々のユーザ毎の詳細な生体認証情報を保持している生体認証情報データベース３１２に対し、当該ユーザのヘッドセットシステム２００に紐付けられたユーザのネットワーク端末２２０から得られる生体認証データを問い合わせる。その結果、紐付けされた当該ユーザのネットワーク端末２２０と、サーバ側の前記画像認識システムを備えた知識情報処理サーバシステム３００内にそれまで蓄積されている情報及びデータとの同期処理を行う事で、関連するデータベース群を最新の状態に更新すると共に、ネットワークのオフライン時に先に進んだ会話ポインタ等の更新も併せて行う事で、オフラインからオンライン、或いはオンラインからオフラインの状態への移行がシームレスに可能になる。

【０１８９】

また本発明により、ＰＣやカメラ付きスマートフォン等に代表されるネットワーク端末、或いは前記ヘッドセットシステムから、インターネット経由で前記画像認識システムを備えた知識情報処理サーバシステム３００側に様々な画像（カメラ画像、写真、動画等）をアップロードする事により、前記サーバシステム側が当該画像、或いは当該画像に内包されている、特定物体、一般物体、人、或いはシーン中から、認識可能になった様々な画像構成要素群に対応するノード群、及び／又は当該画像に付帯するメタデータ、及び／又は当該画像に係るユーザのメッセージやつぶやき、及び／又は当該画像に係るユーザ間のコミュニケーションから抽出可能なキーワード群を、ノード群として抽出する事が可能となる。

【０１９０】

これら抽出された各ノードを中心とする部分グラフから、前記グラフデータベース３６５に記載の関連ノード群を参照する事で、ユーザが指定可能な特定の対象やシーン、或いは特定の場所や地域に係る画像の選択・抽出を可能にし、それらを基に同様或いは類似の対象やシーンを集めたアルバムの作成、或いは一定の場所や地域に係る画像群の抽出処理を行う事が出来る。その上で、前記サーバシステム側が当該抽出された画像群に係る画像特徴群、或いはメタデータ群を基に、それらが特定の物体を撮影したものである場合には複数の視点方向からの映像、或いは異なる環境下で撮影した映像として集約、或いはそれらが特定の場所や地域に係る画像群であるなら、連続的、及び／又は離散的なパノラマ画像に繋ぎ合わせる事で、様々な視点の移動が可能とする。

【０１９１】

前記場所や地域を特定可能なパノラマ画像の構成要素群となっている、インターネット経由でアップロードされるそれぞれの画像に付帯しているメタデータ、或いは前記画像認識システムを備えた知識情報処理サーバシステム３００により認識可能になった当該画像中の特定物体に関し、当該物体が存在していた時点或いは期間をインターネット上の各種知識データベース、或いはインターネットを介して広範なユーザに問い合わせる事で推定或いは獲得し、それら時間軸情報を基に当該画像群を時間軸に沿って振り分け、それら振

り分けられた画像群を基に、ユーザが指定可能な任意の時点或いは期間における前記パノラマ画像を再構成する事が可能となる。これにより、ユーザは任意の場所や地域を含む、任意の「時空間」を指定して、当該「時空間」上に存在していた現実世界の映像を、前記パノラマ画像として視点移動可能な状態で楽しむ事が出来る様になる。

【０１９２】

その上で、特定の対象、或いは特定の場所や地域毎に編成された前記画像群を基に、当該対象に関心が高い、或いは特定の場所や地域に関わりの深いユーザ群を、前記グラフデータベース３６５を基に抽出し、それら多数のユーザ群による当該対象、或いは特定の場所や地域毎に編成されたネットワーク・コミュニケーションを誘発し、そこから特定の対象、或いは特定の場所や地域に係る様々なコメント、メッセージやつぶやきの共有、或いは参加ユーザによる新規情報の提供、或いは特定の不明・不足・欠落情報の探索要求等を可能にするネットワーク・コミュニケーションシステムが構築可能になる。

【０１９３】

図２９を用いて、本発明に係る一実施例における前記サーバシステム上にアップロードされた画像群の中から、特定の「時空間」を指定する事によって抽出した３枚の写真、写真（Ａ）、写真（Ｂ）、写真（Ｃ）を事例として示す。ここでは、１９００年前半における東京日本橋界隈の様子を示す。

【０１９４】

写真（Ａ）では、手前の「日本橋」に加えて、画面左側中央のランドマーク的な建物として知られている「野村証券」本社ビルが特定物体認識可能になり、また画面左側奥には「倉庫」らしき建物、橋の上には「路面電車」２両が一般物体認識可能になっている様子を示す。

【０１９５】

写真（Ｂ）では、別の方向から俯瞰した「日本橋」であり、画面右側に同じく「野村証券」本社ビル、画面左手には「帝国製麻ビル」、また「日本橋」の橋上の装飾的な「外灯」が新たに特定物体認識可能になっている様子を示す。

【０１９６】

写真（Ｃ）では、画面左側に、同じ「帝国製麻ビル」と思われる建物がある事から、「野村証券」本社ビル屋上と思われる場所から「日本橋」方面を撮影したシーンである事が判り、画面上部に文字で『日本橋上ヨリ三越呉服店及ビ神田方面盛観』と読み取れる事からも、「日本橋」「三越呉服店」「神田」の３つのキーワード群が抽出可能となり、そこから画面奥の白い大きな建物は「三越呉服店」の可能性が高いと推定可能になっている様子を示す。

【０１９７】

また、「日本橋」橋梁上に「路面電車」の形状がはっきり写っている事で前記画像認識システムによる精査が可能となり、この「路面電車」が写真（Ｄ）と同じ「１０００型」車両であると特定物体認識可能になっている様子を示す。

【０１９８】

上記一連の画像認識処理は、前記画像認識システム３０１内に備わった特定物体認識システム１１０、一般物体認識１０６、シーン認識システム１０８との協調動作により実行される。

【０１９９】

図３０を用いて、アップロードされた画像群の中から、ユーザが任意の時空間情報を指定する事によって当該時空間内に撮影された画像群のみを抽出し、それらを基に当該時空間を連続的、或いは離散的なパノラマ画像に再構築して、ユーザが自由に当該空間内で視点の移動を行う、或いは自由に当該空間内で時間の移動が可能な、時空間移動表示システムについて、概略的な実施事例を用いて説明する。

【０２００】

最初に、インターネットを介し前記画像認識システムを備えた知識情報処理サーバシステム３００側に、ユーザのネットワーク端末２２０経由で画像のアップロード（２２００

）が開始される。アップロードされた画像は前記画像認識システム３０１にて画像認識処理が開始される（２２０１）。当該画像ファイルに予めメタデータが付与されている場合は、メタデータ抽出処理（２２０４）が実行される。また、当該画像中に文字情報が発見された場合には、ＯＣＲ（Ｏｐｔｉｃａｌ　Ｃｈａｒａｃｔｅｒ　Ｒｅｃｏｇｎｉｔｉｏｎ）等を用いて、文字情報抽出処理（２２０３）が行われ、そこからメタデータ抽出処理（２２０４）を経て、有用なメタデータ群を得る。

【０２０１】

一方、アップロードされた一枚の画像の中から、ユーザのネットワーク端末２２０上のＧＵＩ，或いは図３Ａに記載の前記音声による着目対象のポインティング処理により、当該画像中の個々の物体に係る画像の切り抜き（２２０２）処理を行い、当該対象に対して一般物体認識システム１０６、及びシーン認識システム１０８にて画像認識したクラス情報に従いＭＤＢ検索部１１０－０２で物体の絞り込み処理を行い、当該画像に関する詳細情報を記述したＭＤＢ１１１を参照して、特定物体認識システム１１０により当該物体との比較照合処理を行い、最終的に同定された特定物体に関し、前記メタデータ群を参照して、当該画像に時間軸情報が存在するか？否かを判別（２２０５）する。

【０２０２】

当該画像に時間軸情報が存在する場合、画像中の物体群が存在した時間情報をＭＤＢ１１１内の記述から抽出し、参照の上で物体が当該時間内に存在するか否かを判別（２２０６）する。前記存在が確認された場合は、当該物体以外に画像認識可能になった他の物体について、同様に当該時間内に存在し得ない物体がないかどうか（２２０７）前記同様にＭＤＢ１１１内の記述から判別し、当該全ての整合性が確認された時点で、当該画像に関する撮影時間の推定（２２０８）処理が行われる。それ以外の場合は、時間情報が不明（２２０９）として、当該ノード情報が更新される。

【０２０３】

次に、当該画像に場所に係る情報が存在する場合（２２１０）、画像中の物体群が存在した場所に係る情報をＭＤＢ１１１内の記述から抽出し、参照の上で物体が当該場所において存在するか否かを判別（２２１０）する。前記存在が確認された場合は、当該物体以外に画像認識可能になった他の物体について、同様に当該場所において存在し得ない物体がないかどうか（２２１１）前記同様にＭＤＢ１１１内の記述から判別し、当該全ての整合性が確認された時点で、当該画像に関する撮影された場所の推定（２２１２）処理が行われる。それ以外の場合は、場所情報が不明（２２１３）として、当該ノード情報が更新される。

【０２０４】

前記一連の処理に加えて、前記獲得可能になった当該画像自体から抽出可能な、或いは当該画像自体に付帯するメタデータ群と、前記推定可能になった時空間情報とを再度照合し、その整合性が確認された時点で、当該画像全体に係る時空間情報の獲得（２２１４）が完了し、当該時空間情報を当該画像に係るノードにリンク（２２１５）する。また上記整合性に齟齬のある場合には、メタデータ自体の誤り、画像認識システムの認識誤り、或いはＭＤＢ１１１内に記載の内容に誤りや不備があるとして、以降の再検証処理に備える。

【０２０５】

これらの時空間情報の付与が行われた画像群に対し、ユーザは任意の時空間を指定して当該条件に合致した画像群を抽出する事が可能になる（２２１６）。まず、多数の画像群の中から任意の場所（２２１７）、任意の時間（２２１８）に撮影された画像群を、当該指定した時空間に係るノードを辿って抽出する（２２１９）。これら抽出された複数の画像群を基に、画像中の共通の特定特徴点を探索する事で、検出された特定特徴点同士を連続的に繋いでパノラマ画像を再構成（２２２０）する事が可能になる。この場合、パノラマ画像中に欠落或いは欠損画像がある場合は、ＭＤＢ１１１記載の地図、図面、或いは設計図等の利用可能な情報から広範に推定処理する事で、離散的なパノラマ画像として再構成が可能になる。

【０２０６】
前記一連の時空間情報獲得の為の学習プロセスを、アップロードされる多数の写真（動画を含む）画像に対して、前記画像認識システムを備えた知識情報処理サーバシステム３００が継続的に行う事により、時空間情報を有する連続的なパノラマ画像が取得可能になる。これにより、ユーザは任意の時間／空間を指定して、任意の視点移動、或いは同一空間における任意の時間に係る画像体験（２２２１）を楽しむ事が可能になる。
【０２０７】
図３１を用いて、本発明に係る一実施例における、ユーザが前記画像認識システムを備えた知識情報処理サーバシステムに対してアップロードした画像に対して、当該ユーザのネットワーク端末上のＧＵＩ操作、或いは前記音声処理によるポインティング操作による当該ユーザが着目した特定物体、一般物体、人、或いはシーンに係る選択抽出処理により、前記サーバシステムが認識した結果を、当該入力画像と共に当該ユーザを含むあらかじめ指定可能な広範なユーザ間で共有可能にすることによるネットワーク・コミュニケーションシステムの構成を説明する。
【０２０８】
当該時空間を指定したユーザの視点の移動により発見可能になった特定物体、一般物体、人、或いはシーンに対しても、これまで述べて来た様な特定の着目対象に係る一連のメッセージやつぶやきの記録、及び再生体験が可能になる。
【０２０９】
当該ユーザによるアップロードされた画像２１０１は、前記サーバシステムにおいて選択・抽出処理２１０３が行われる。この際に、ユーザは図３Ａに記載の手順での選択・抽出処理を実行しても良いし、図３０に示した選択・抽出コマンドを、ＧＵＩ２１０４を操作することによって選択・抽出処理を実行しても良い。当該選択・抽出処理により切り出された画像は、画像認識システム３０１において認識処理される。その結果は、インタレストグラフ部３０３おいて分析・分類・蓄積され、キーワード群や時空間情報と共にグラフデータベース３６５に記録される。当該ユーザは、画像のアップロードに際して、メッセージやつぶやき２１０６、或いは文字情報２１０５による書き込みを行っても良い。これら当該ユーザの発したメッセージやつぶやき、或いは文字情報もインタレストグラフ部にて分析・分類・蓄積される。当該ユーザ、或いは当該ユーザを含むユーザ群、或いはユーザ全体は、前記対象に係るキーワード群、及び／又は時空間情報（２１０６）を基に、インタレストグラフ部から記録された画像を選択する事が可能であり、当該画像に係る広範なネットワーク・コミュニケーションを誘発させることが出来る。さらに、前記広範なユーザ間のコミュニケーションを、前記サーバシステム側で観察・蓄積し、インタレストグラフ部３０３の１構成要素である統計情報処理部３６３において分析することで、当該ユーザ特有の、或いは特定のユーザ群に特有の、或いはユーザ全体に共通の動的な関心や好奇心の在り所とその推移を、上記広範なユーザ群、抽出可能なキーワード群、及び様々な着目対象に係るノード間を繋ぐ動的なインタレストグラフとして獲得する事が可能となる。
【０２１０】
［周辺技術］
本発明に係るシステムは、既存の様々な技術と組み合わせる事によって、さらに利便性の高いシステムとして構成する事が可能となる。以下に、例示する。
【０２１１】
本発明に係る一実施例として、ユーザの発話をヘッドセットシステム２００に組み込まれたマイクロフォンが拾い、前記音声認識システム３２０により発話中に含まれる単語列及び構文を抽出した後、ネットワーク上の自動翻訳システムを活用する事で異なる言語に翻訳し、当該翻訳された単語列を前記音声合成システム３３０により音声変換した上で、他のユーザに当該ユーザのメッセージやつぶやきとして伝える事が可能になる。或いは、前記画像認識システムを備えた知識情報処理サーバシステム３００側からの音声情報を、当該ユーザが指定可能な言語で受け取る事が出来る様に構成する事が出来る。

【０２１２】
本発明に係る一実施例として、ユーザのヘッドセットシステムに組込まれたカメラがその視野内に捉えた映像の中から、規定の認識マーカーと共に特定の画像変調パターンを抽出した場合、当該信号源の存在をユーザに喚起し、当該信号源が表示装置或いはその近傍にある場合、当該変調された画像パターンを前記認識エンジン２２４との協調動作により復調する事によって、そこから得られるＵＲＬ等のアドレス情報をインターネット経由で参照し、当該表示装置上に表示されている画像に係る音声情報を当該ユーザのヘッドセットシステム経由で送り込む事を可能にする。これにより、ユーザが偶然目にした様々な表示装置から、当該表示画像に係る音声情報を当該ユーザに効果的に送り込む事が可能になる。これにより、電子広告媒体としてのデジタル・サイネージの有効性を一段と高める事が出来る。反面、ユーザが目にする事が出来る全てのデジタル・サイネージから音声情報が一斉に送り届けられると、場合によってはそれらを不要なノイズと感じてしまう可能性もある事から、それぞれのユーザに係る前記インタレストグラフを基に、ユーザ毎に異なる嗜好を反映した広告等のみを選択して、個々のユーザ毎に異なる音声情報として送り届ける事が出来る様に構成しても良い。
【０２１３】
本発明に係る一実施例として、様々な生体情報（バイタルサイン）をセンシング可能な複数の生体センサ群をユーザのヘッドセットシステムに組み込む事で、当該ユーザが関心を持って着目した対象と、当該生体情報との相関を、前記画像認識システムを備えた知識情報処理サーバシステム３００側で統計処理した上で当該ユーザに係る特殊なインタレストグラフとして登録しておく事によって、当該ユーザが当該特定の対象或いは事象に遭遇した場合、或いは遭遇の可能性が高まった場合に、当該ユーザの生体情報値が急変する事態に備える事が出来る様に、前記サーバシステム側を構成する事が可能である。取得可能になる生体情報としては、ユーザの体温、心拍、血圧、発汗、皮膚表面の状態、筋電位、脳波、眼球運動、発声、頭の動き、体の動き等が含まれる。
【０２１４】
この為の学習パスとして、カメラが捉えたユーザの主観視内に特定の特定物体、一般物体、人、写真、或いはシーンが現れた時に、測定可能な前記生体情報値が一定以上変化する場合、当該ユーザに関わる特異的な反応として係る事態を、前記画像認識システムを備えた知識情報処理サーバシステム３００側に通知する事で、当該サーバシステム側は関連する生体情報の蓄積・分析を開始すると同時に、当該カメラ映像の解析を開始し、そこから抽出可能な画像構成要素群を係る事態に関連する可能性のある原因要素群として前記グラフデータベース３６５、及びユーザデータベース３６６内に登録する事を可能にする。
【０２１５】
以降、様々な事例で前記学習を繰り返す事で、前記各種生体情報値の変化に係る要因の分析・推定を統計処理から求める事が可能になる。
【０２１６】
上記の一連の学習プロセスから、個々のユーザ毎に異なる当該生体情報値の異常な変化の要因となっていると予測可能な特定物体、一般物体、人、写真、或いはシーンに、当該ユーザが再び遭遇する、或いは遭遇する可能性が高いと予測可能な場合、前記サーバシステム側から当該ユーザに対し、ネットワークを介して音声、及び／又は、文字、画像、バイブレーション等で、係る可能性を速やかに通知する様に当該サーバシステムを構成する事が可能となる。
【０２１７】
さらに、観測可能な前記生体情報値が急変し、ユーザの容体に一定以上の危機の可能性があると推定可能な場合、速やかに当該ユーザに係る事態の確認を求め、当該ユーザから一定の反応が得られない場合、当該ユーザに一定以上の緊急事態が発生した可能性が高いと判断し、予め設定可能な緊急連絡網、或いは特定の機関等に通知する事が可能な様に、前記画像認識システムを備えた知識情報処理サーバシステム３００側を構成する事が出来る。

【０２１８】
本発明に係る生体認証システムにおいて、ユーザが頭部に装着可能な前記ヘッドセットシステムから、ユーザ固有の声紋、静脈パターン、或いは網膜パターン等を取得して生体認証が可能な場合、ユーザと前記画像認識システムを備えた知識情報処理サーバシステム３００側とを一意にバインドする様に本システムを構成する事が出来る。当該生体認証デバイスはユーザの前記ヘッドセットシステムに組み込み可能な事から、当該ヘッドセットシステムの着脱に合わせて自動的にログイン、ログアウト可能にする様に構成する事も可能になる。これら生体情報を活用した紐付けを常時上記サーバシステム側で監視する事により、異なるユーザによる不正なログイン、不正な利用が排除可能になる。当該ユーザ認証が正常に行われた場合、以下の情報群が当該ユーザにバインドされる。
（１）ユーザが設定可能なユーザプロファイル
（２）ユーザの音声
（３）カメラ画像
（４）時空間情報
（５）生体情報
（６）その他のセンサ情報
【０２１９】
本発明に係る一実施例として、複数のユーザ間で共有される画像に関し、プライバシー保護の観点から、ユーザが予め指定可能なルールに従い、当該ユーザ毎の顔部分、及び/又は、当該ユーザを特定可能な画像の特定部分を、前記画像認識システムを備えた知識情報処理サーバシステム３００側に組込まれた画像認識システム３０１により抽出及び検出し、それらの特定画像領域に対し、判別不能なレベルにまで自動的にフィルタ処理を施す様に構成する事が出来る。これにより、プライバシー保護を含む一定の閲覧制限を設ける事が可能となる。
【０２２０】
本発明に係る一実施例として、ユーザが頭部に装着可能なヘッドセットシステムに複数のカメラを設置する事が出来る。この場合、一実施例として複数のカメラに撮像視差を設ける事が出来る。或いは、性質の異なる複数の撮像素子を使って、対象物体までの深度（距離）を直接測定可能な三次元カメラを組み込む様に構成する事も出来る。
その上で、前記画像認識システムを備えた知識情報処理サーバシステム３００側からの音声による指示により、当該サーバシステムにより指定された特定のユーザに対し、当該サーバシステムが指定した特定の対象、或いは周囲の様子等を、当該サーバシステムが当該ユーザに対して様々な視点から撮影する様に依頼する事で、前記サーバシステム側が当該対象の立体的な把握、或いは周囲の状況等の立体的な把握が容易になると共に、当該画像認識結果により前、記サーバシステム内のＭＤＢ１１１を含む関連データベース群の更新が可能となる様に、当該サーバシステムを構成する事が出来る。
【０２２１】
本発明に係る一実施例として、ユーザが頭部に装着可能なヘッドセットシステムに、指向性を有する深度センサを組み込む事が出来る。これにより、当該ヘッドセットシステムを装着したユーザに近付く人間を含む生体や物体の動きを検知し、前記ユーザに音声で係る事態を通知する事が可能となる。同時に、当該ユーザのヘッドセットシステムに組み込まれたカメラ及び画像認識エンジンを自動的に起動し、不測の物体の急接近に即時に対応可能な様にリアルタイム処理が要求される部分をユーザのネットワーク端末側で、高度の情報処理を必要とする部分に関して前記画像認識システムを備えた知識情報処理サーバシステム３００側で分担して実行可能にする様にシステムを構成する事で、ユーザに近付く特定の物体、特定の人間、特定の動物等を高速に識別／解析し、その結果を音声情報、或いはバイブレーション等により当該ユーザに速やかに喚起する事が可能となる。
【０２２２】
本発明に係る一実施例として、ユーザが頭部に装着可能なヘッドセットシステムに、当該ユーザを中心とした周囲、或いはその上部や下部も含めた全方位を撮影する事が可能な

撮像システムを組み込む事が出来る。或いは、ユーザの主観的視野外となる後方や側面からの視野を撮影する事が可能な複数のカメラを、当該ユーザのヘッドセットシステムに追加する事が可能となる。この様な構成を採る事により、当該ユーザの主観視野外にあるものの、当該ユーザが特に関心や注意を払わなければならない対象が近傍に存在する場合に、当該ユーザに対して速やかに音声、或いはそれに代わる手段を用いて係る状況の喚起を促す事が可能になるように、当該画像認識システムを備えた知識情報処理サーバシステム３００を構成する事が出来る。

【０２２３】

本発明に係る一実施例として、ユーザが頭部に装着可能なヘッドセットシステムに、以下の様な環境値を測定可能な環境センサ群を任意に組み込む事が可能である。

（１）周囲の明るさ（光度）
（２）照明や外光の色温度
（３）周囲の環境騒音
（４）周囲の音圧レベル

これにより周囲の環境雑音の低減、最適なカメラ露光状態への対応が可能になり、前記画像認識システムの認識精度、及び前記音声認識システムの認識精度を向上させる事が可能になる。

【０２２４】

本発明に係る一実施例として、ユーザが頭部に装着可能なヘッドセットシステムに、当該ユーザの視野の一部を覆う形で半透明のディスプレイ装置を組み込む事が出来る。或いは、当該ヘッドヘッドシステムをヘッドマウントディスプレイ（ＨＭＤ）、或いはスカウター（Ｓｃｏｕｔｅｒ）として表示ディスプレイと一体的に構成する事も出来る。この様な表示システムを可能とする装置には、ユーザの網膜に直接画像情報を走査投影するレチナール・センシングと呼ばれる画像投影システム、或いは眼前に配置した半透明の反射板に画像を投影するデバイス等が知られている。上記の様な表示システムを採用する事により、ユーザのネットワーク端末の表示画面に表示される画像の一部、或いは全部を、当該表示デバイス上に映し出す事が可能になり、前記ネットワーク端末をユーザの眼前に取り出す事なく、インターネット経由で直接前記画像認識システムを備えた知識情報処理サーバシステム３００側とのコミュニケーションが可能となる。

【０２２５】

本発明の一実施形態としてユーザが頭部に装着可能な前記ＨＭＤ、前記スカウター、或いはそれらに併設する形態で視線検出センサを具備しても良い。当該視線検出センサには光センサアレイを用いても良く、そこから照射される光線の反射光を計測する事で、当該ユーザの瞳の位置を検出し、当該ユーザの視線位置を高速に抽出する事が出来る。例えば、図２７において、点線枠２００１はユーザの装着する前記スカウター２００２の視野画像であるとする。この時、当該ユーザの視線方向にある対象に対して、視点マーカー２００３を重ねて表示しても良い。その場合、前記視点マーカーの位置が当該対象と同位置に表示される様に、ユーザの音声による指示でキャリブレーション可能にする事が出来る。

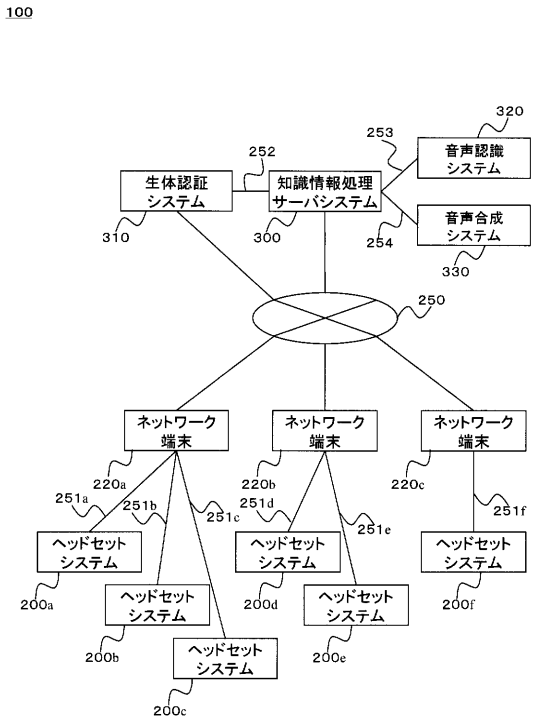
【符号の説明】

【０２２６】

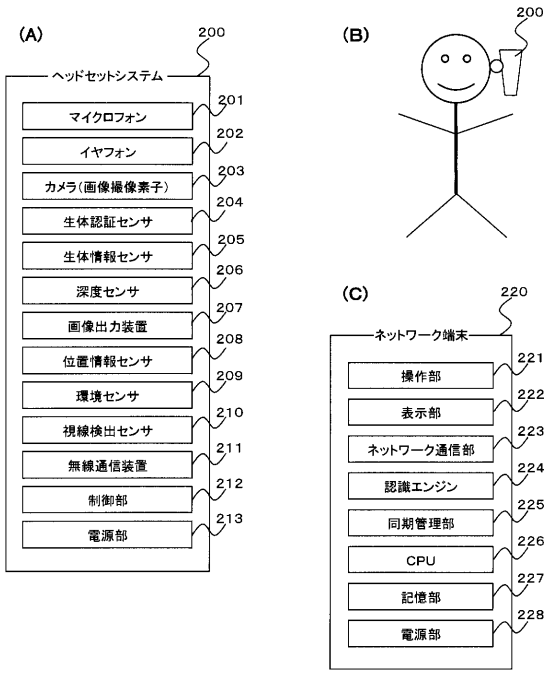
１００　ネットワーク・コミュニケションシステム
１０６　一般物体認識システム
１０７　画像カテゴリデータベース
１０８　シーン認識システム
１０９　シーン構成要素データベース
１１０　特定物体認識システム
１１１　マザーデータベース
２００　ヘッドセットシステム
２２０　ネットワーク端末
３００　知識情報処理サーバシステム

３０１　画像認識システム
３０３　インタレストグラフ部
３０４　状況認識部
３０６　再生処理部
３１０　生体認証システム
３２０　音声認識システム
３３０　音声合成システム
３６５　グラフデータベース
４３０　会話エンジン

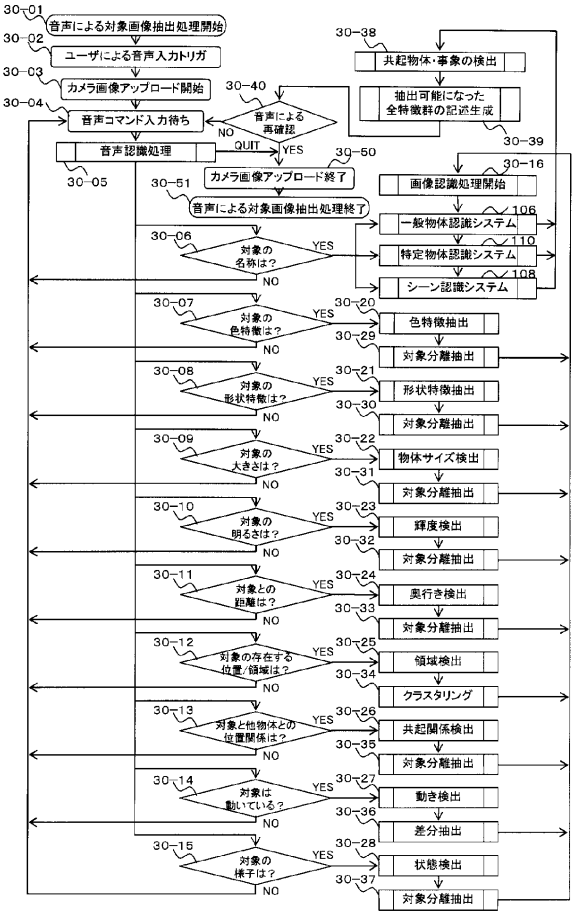
【図１】



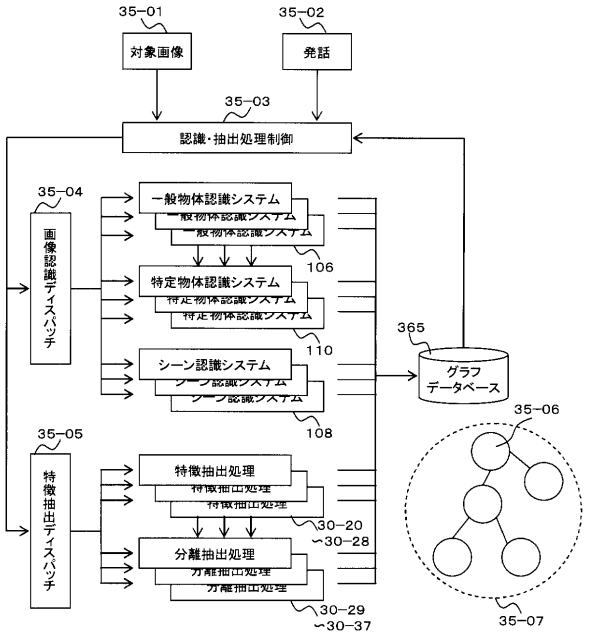
【図２】



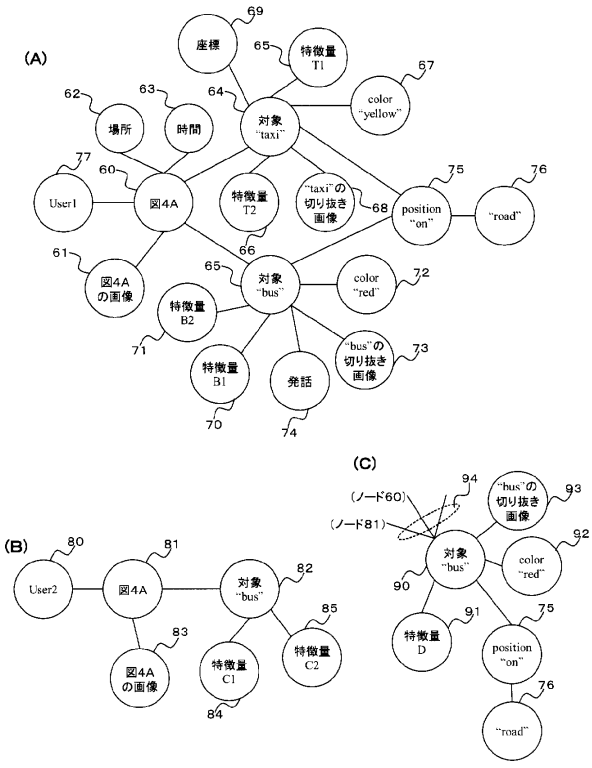
【図３Ａ】



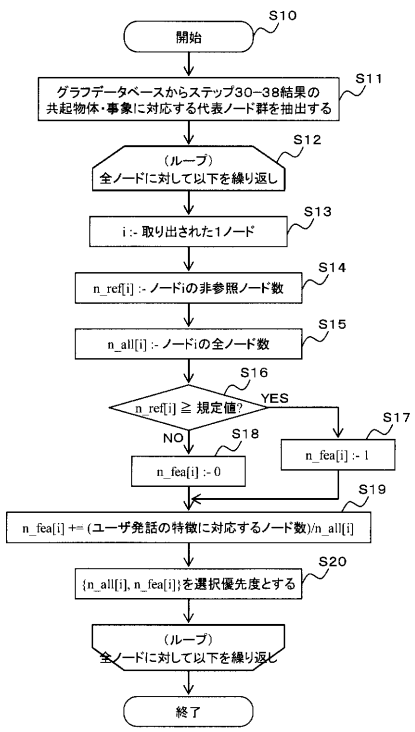
【図３Ｂ】



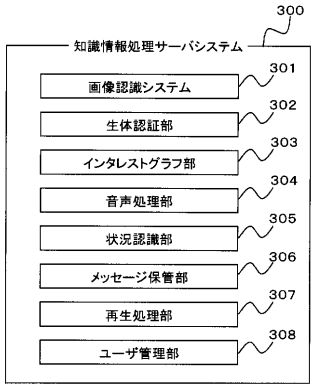
【図４Ｂ】



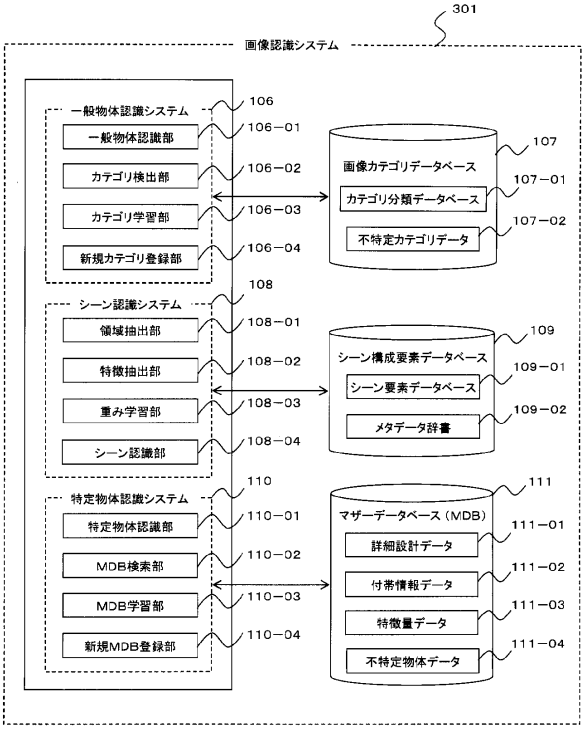
【図４Ｃ】



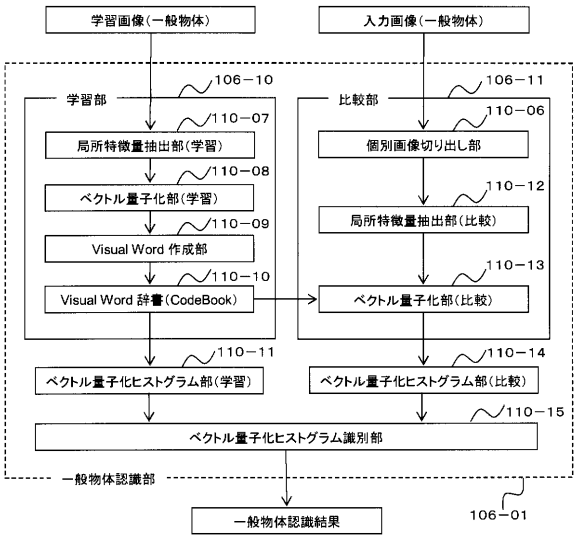
【図５】



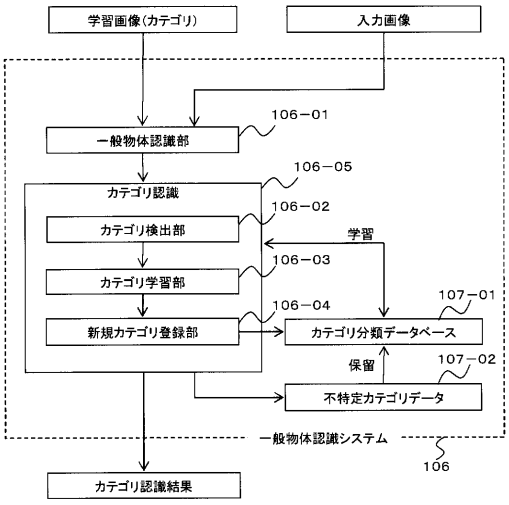
【図６Ａ】



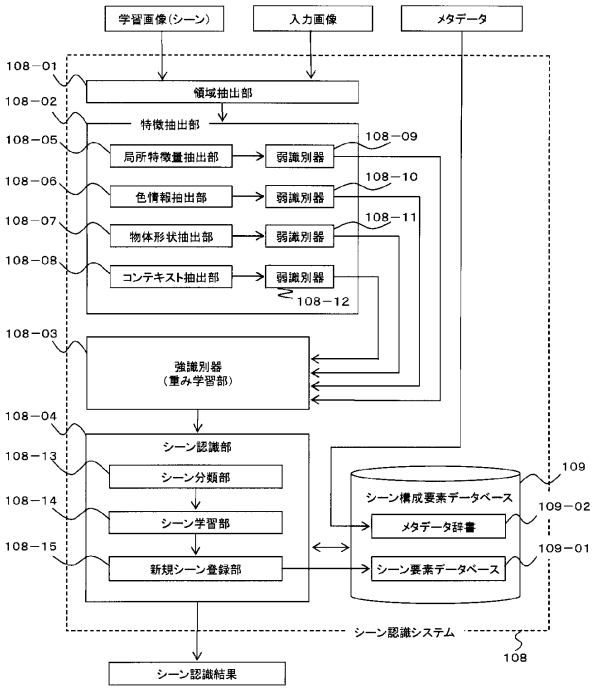
【図６Ｂ】



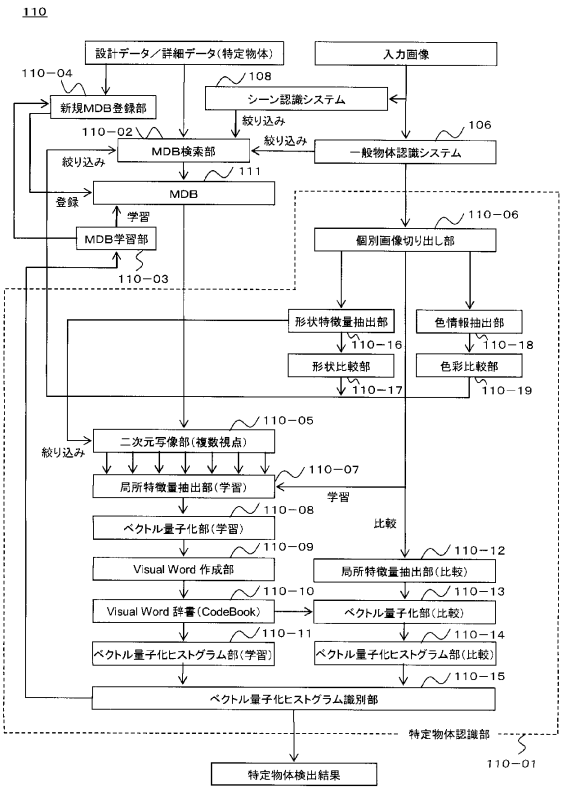
【図６Ｃ】



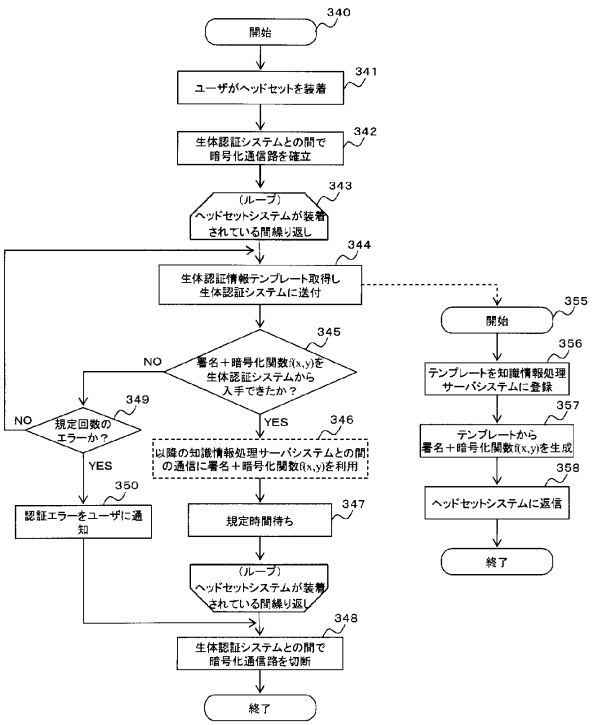
【図６Ｄ】



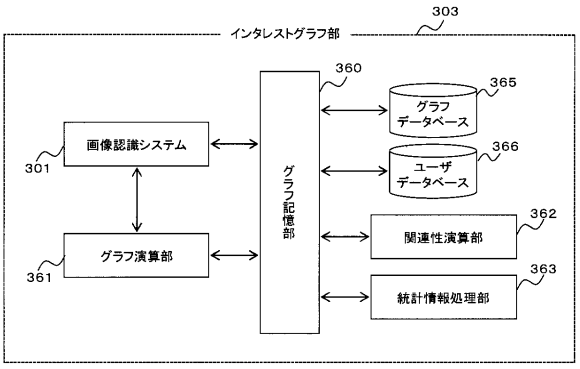
【図６Ｅ】



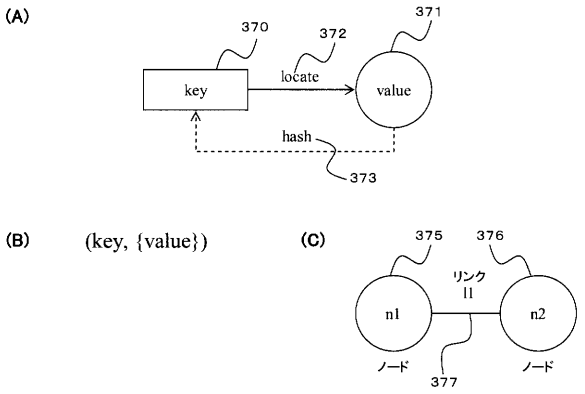
【図７】



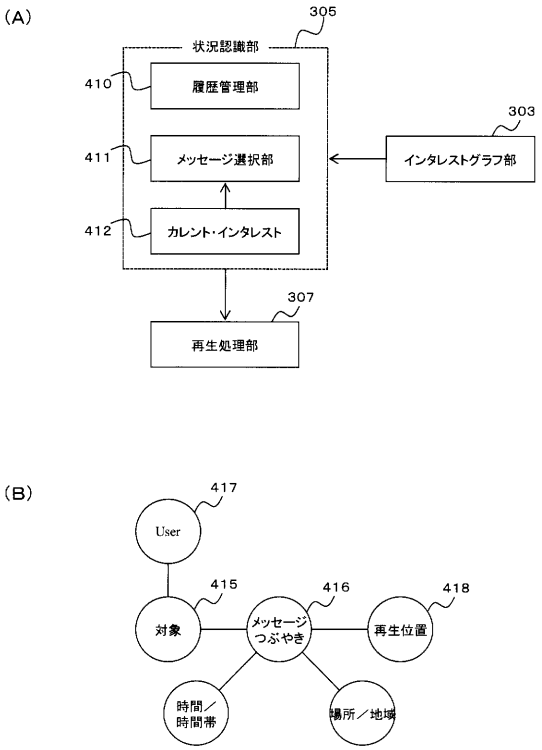
【図８Ａ】



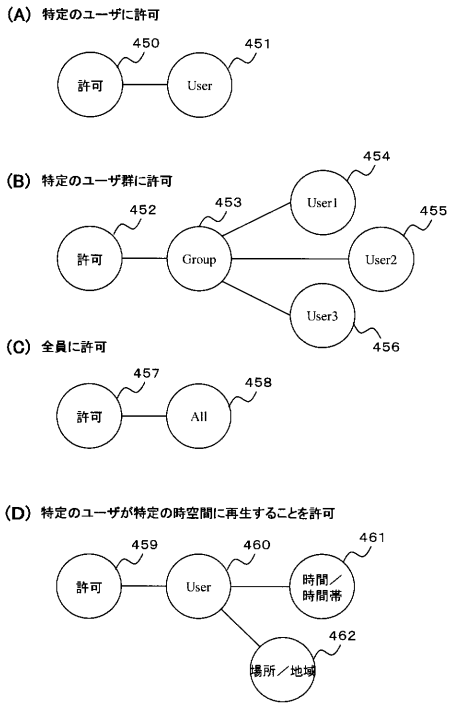
【図８Ｂ】



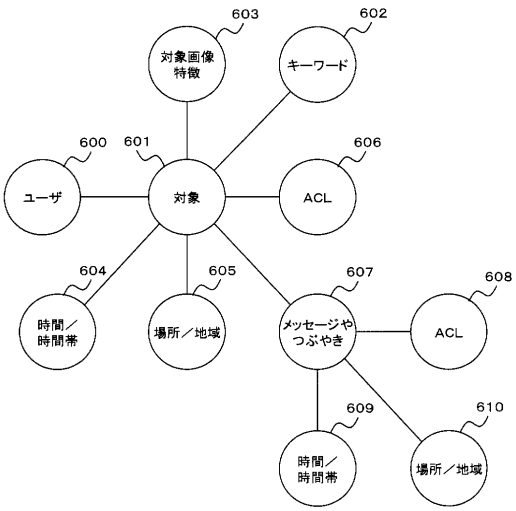
【図９】



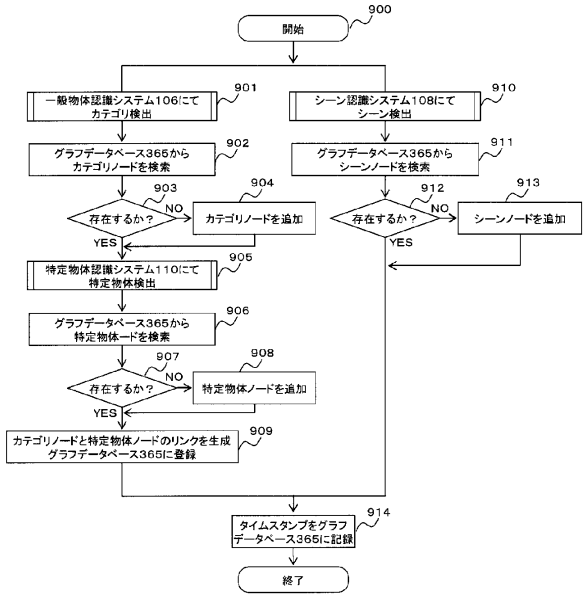
【図１２】



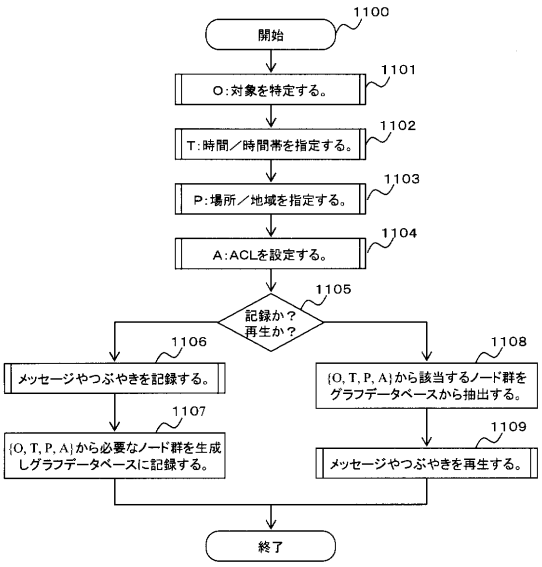
【図１４】



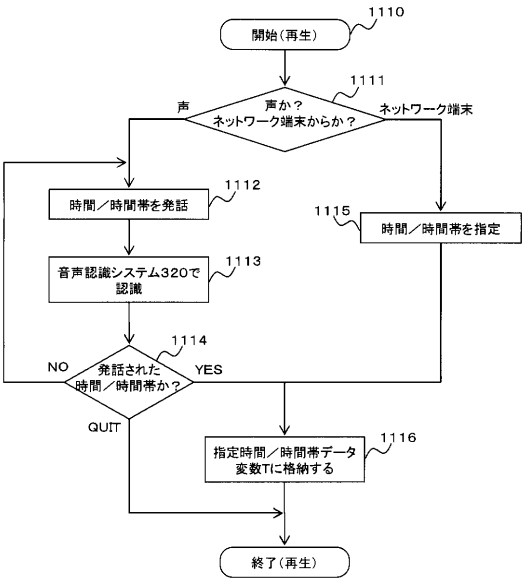
【図１５】



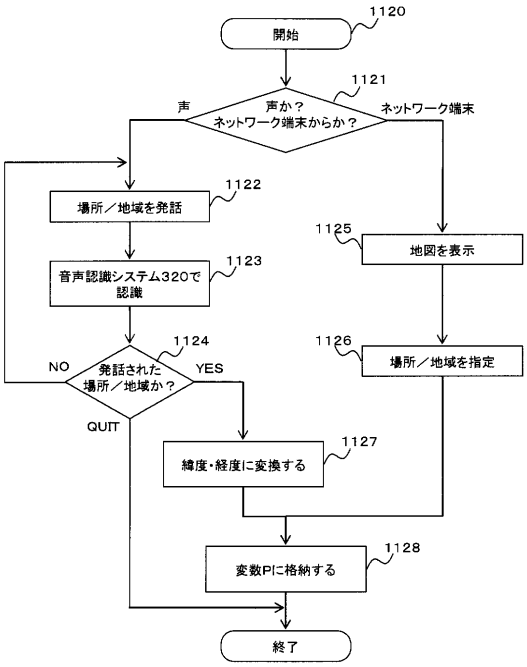
【図１８Ａ】



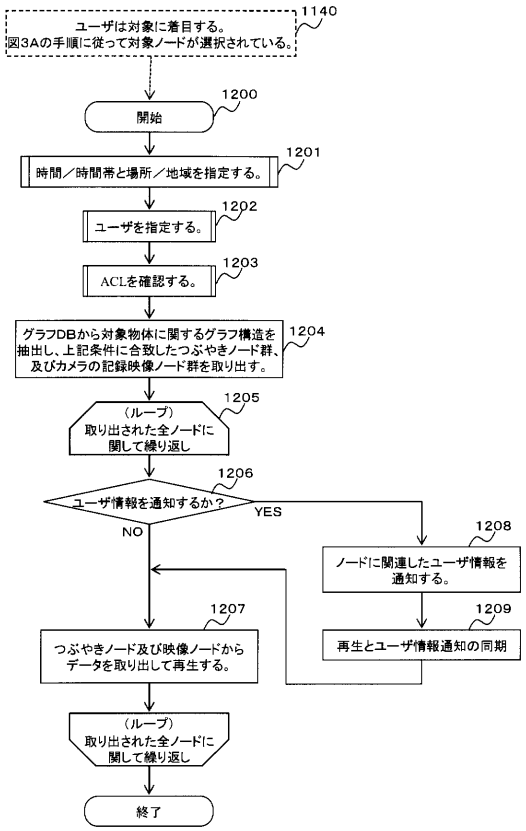
【図１８Ｂ】



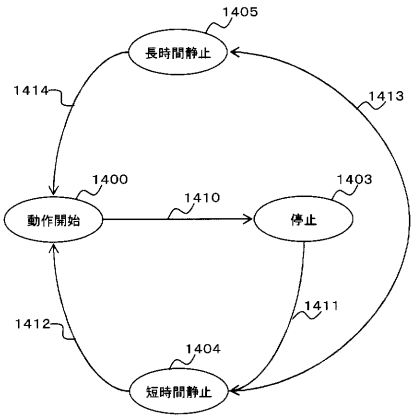
【図１８Ｃ】



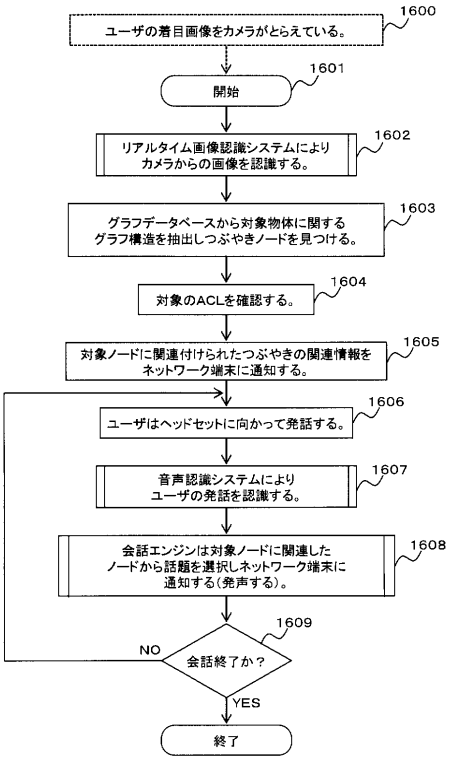
【図１９】



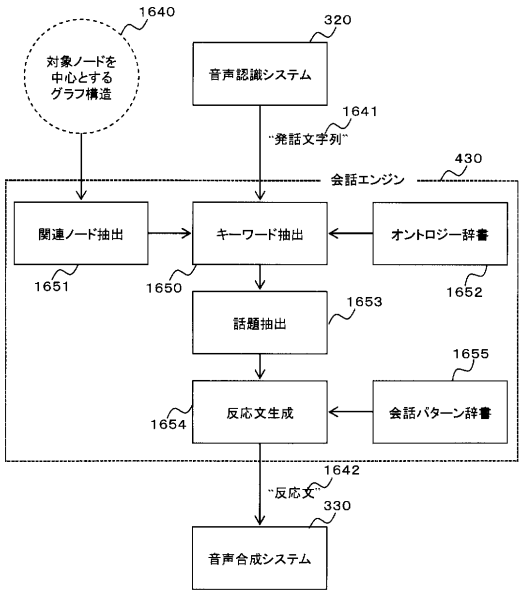
【図２１】



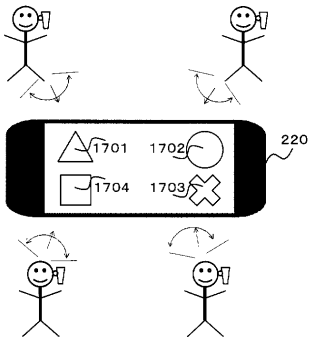
【図２３Ａ】



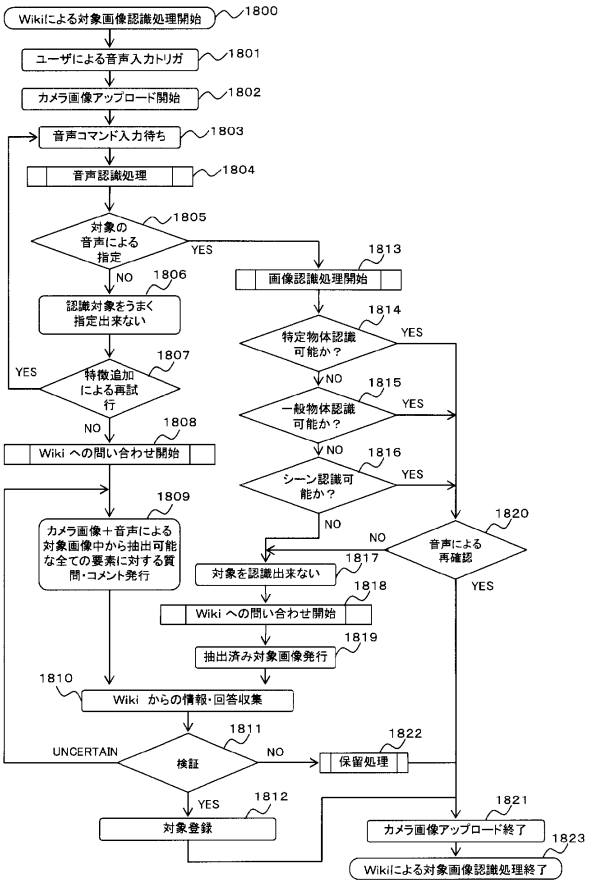
【図２３Ｂ】



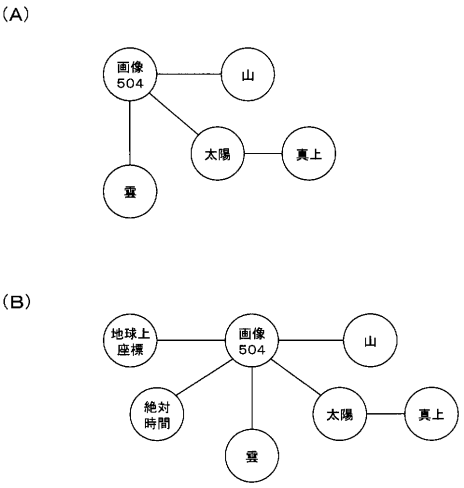
【図２４】



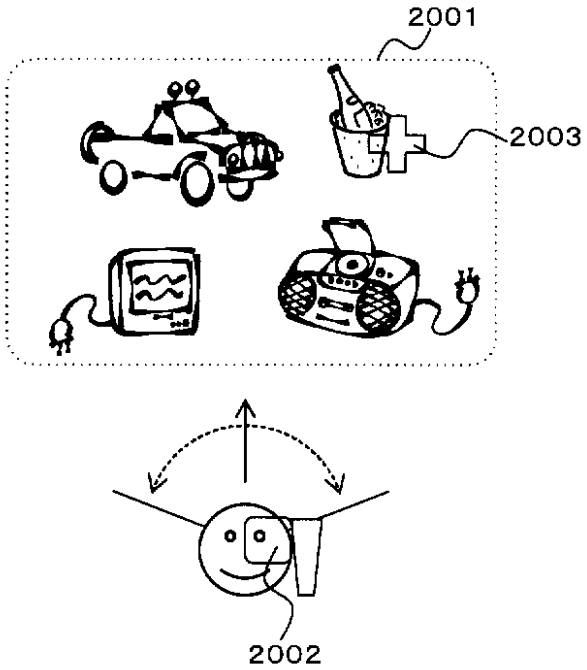
【図２５】



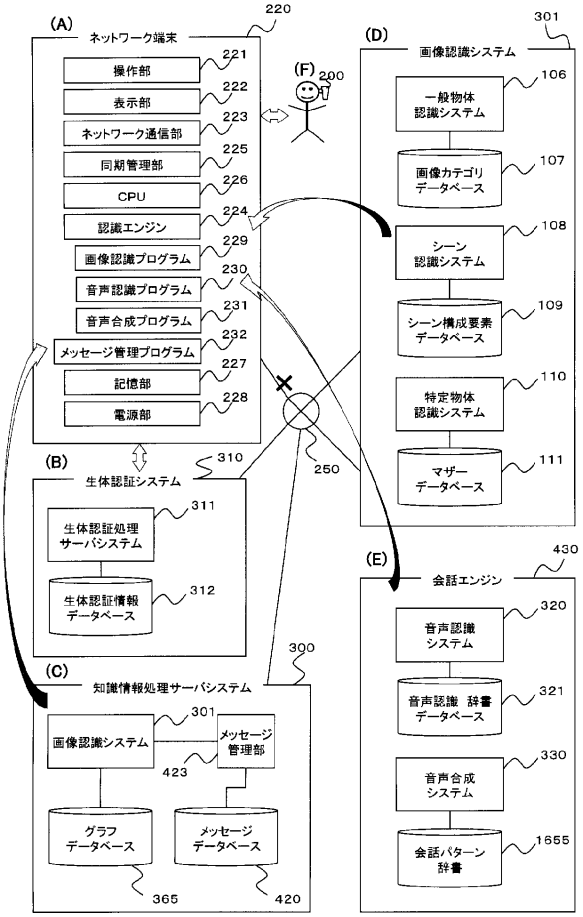
【図２６】



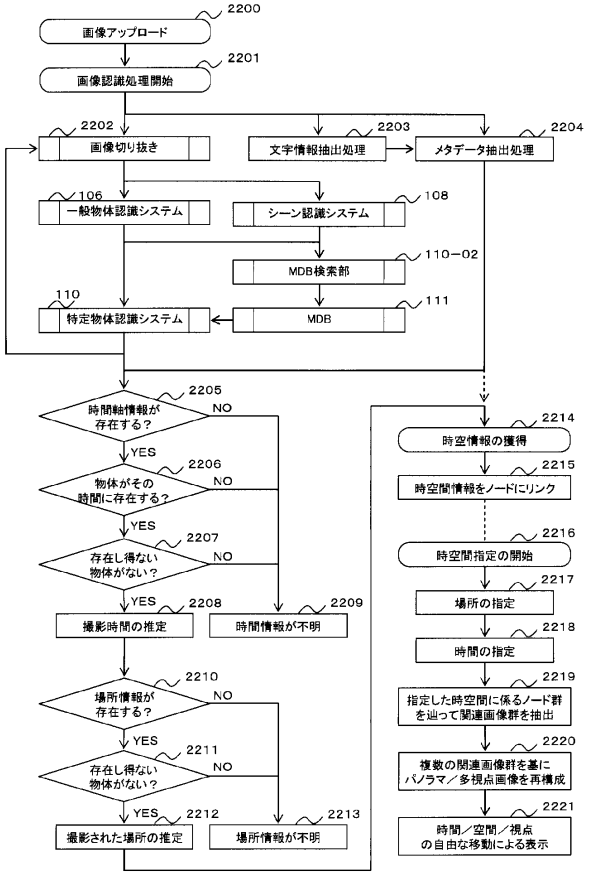
【図２７】



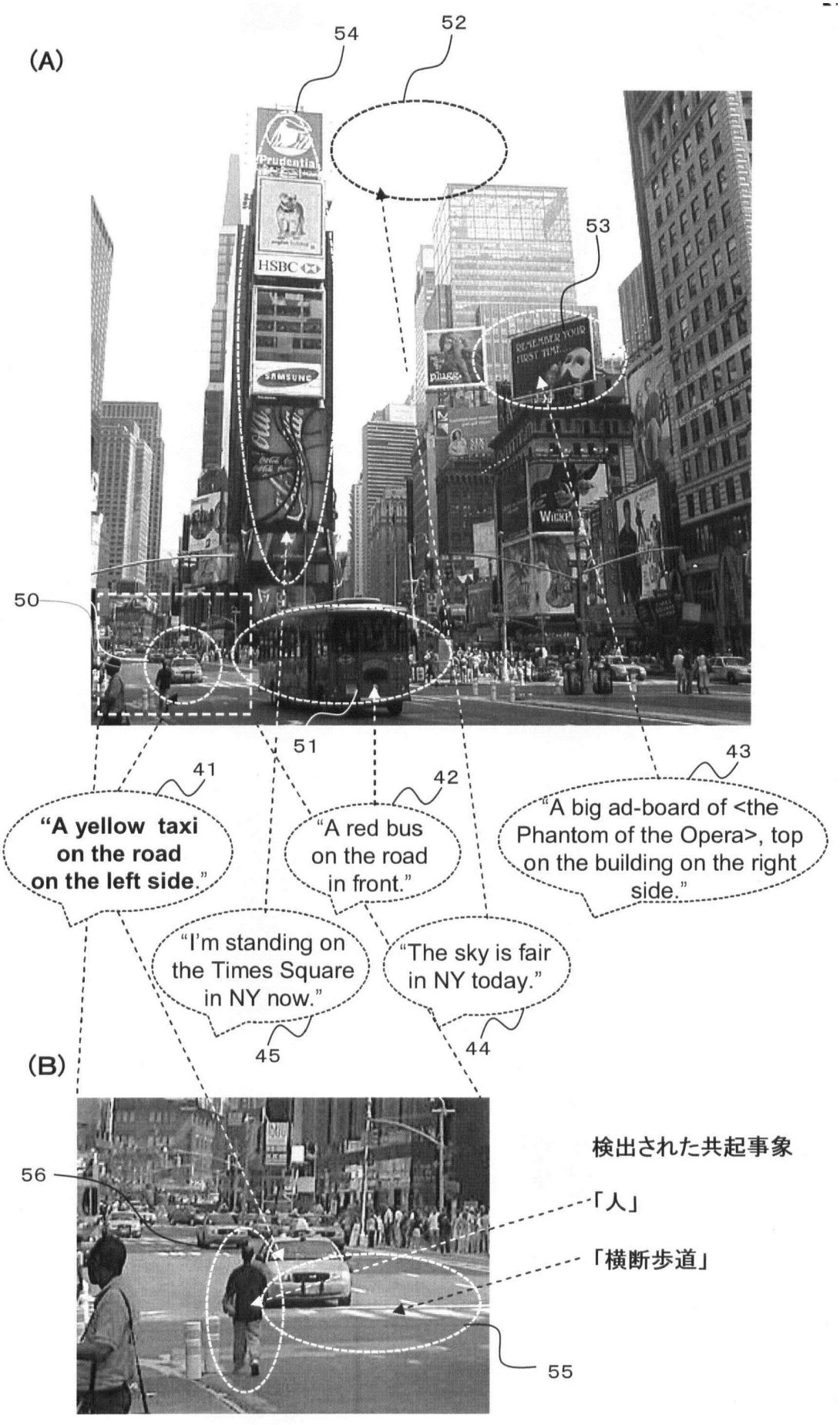
【図２８】


(A)
ネットワーク端末
220
操作部
221
表示部
222
ネットワーク通信部
223
同期管理部
225
CPU
226
認識エンジン
224
画像認識プログラム
229
音声認識プログラム
230
音声合成プログラム
231
メッセージ管理プログラム
232
記憶部
227
電源部
228
(F)
200
(D)
画像認識システム
301
一般物体認識システム
106
画像カテゴリデータベース
107
シーン認識システム
108
シーン構成要素データベース
109
特定物体認識システム
110
マザーデータベース
111
(B)
生体認証システム
310
生体認証処理サーバシステム
311
生体認証情報データベース
312
250
(C)
知識情報処理サーバシステム
300
画像認識システム
301
メッセージ管理部
423
グラフデータベース
365
メッセージデータベース
420
(E)
会話エンジン
430
音声認識システム
320
音声認識 辞書データベース
321
音声合成システム
330
会話パターン辞書
1655

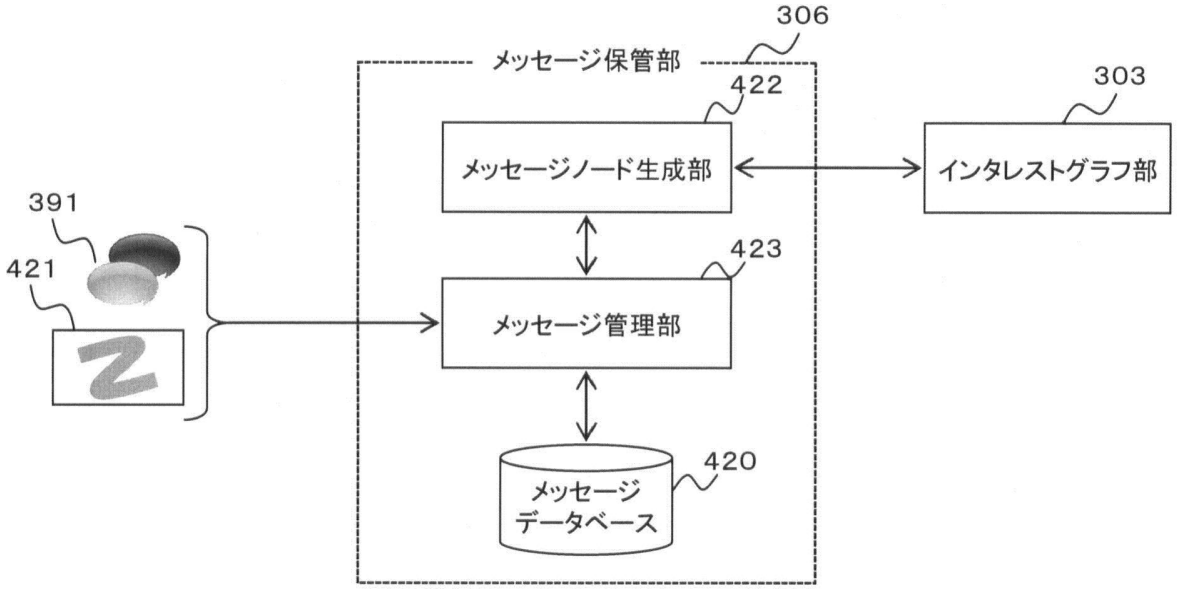
【図３０】


2200
画像アップロード
2201
画像認識処理開始
2202
画像切り抜き
2203
文字情報抽出処理
2204
メタデータ抽出処理
106
一般物体認識システム
108
シーン認識システム
110−02
MDB検索部
110
特定物体認識システム
111
MDB
2205
時間軸情報が存在する？
NO
YES
2206
物体がその時間に存在する？
NO
YES
2207
存在し得ない物体がない？
NO
YES
2208
撮影時間の推定
2209
時間情報が不明
2210
場所情報が存在する？
NO
YES
2211
存在し得ない物体がない？
NO
YES
2212
撮影された場所の推定
2213
場所情報が不明
2214
時空情報の獲得
2215
時空間情報をノードにリンク
2216
時空間指定の開始
2217
場所の指定
2218
時間の指定
2219
指定した時空間に係るノード群を辿って関連画像群を抽出
2220
複数の関連画像群を基にパノラマ／多視点画像を再構成
2221
時間／空間／視点の自由な移動による表示

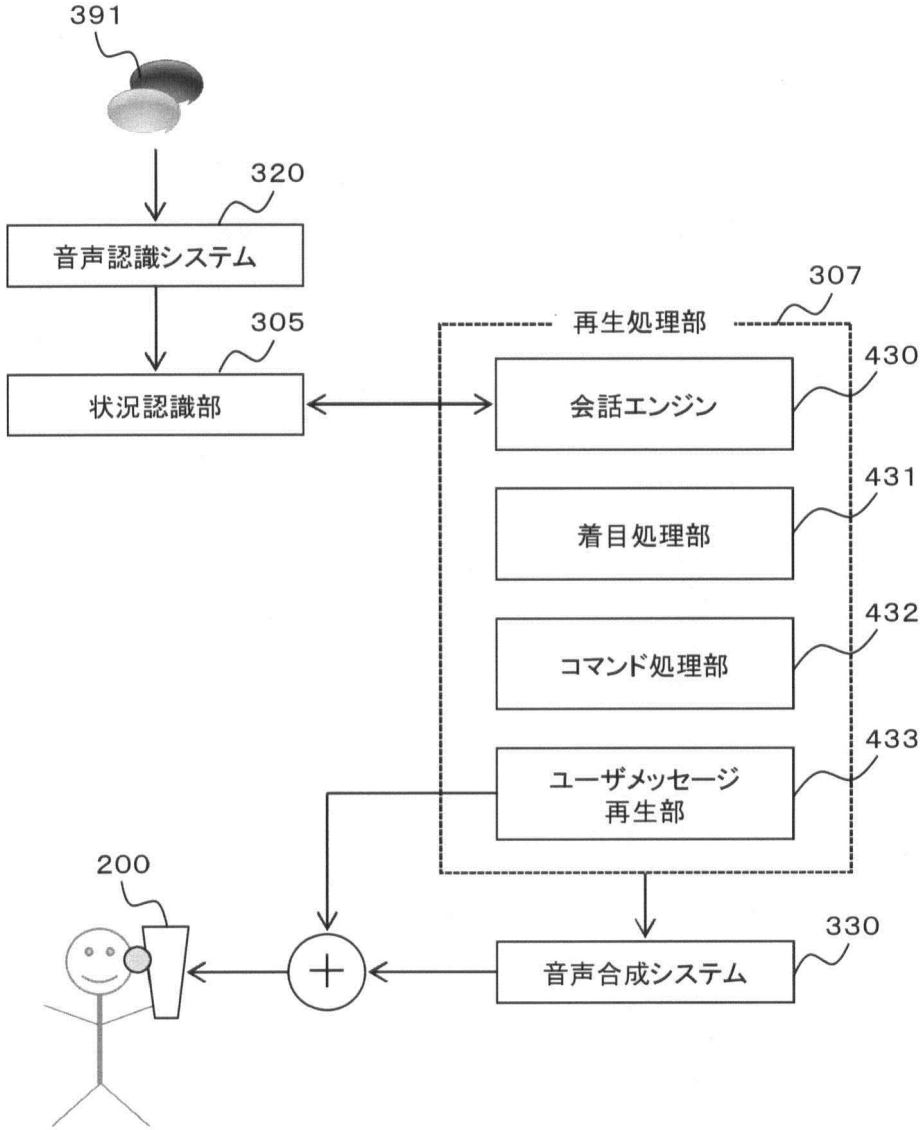
【図４Ａ】


(A)
54
52
53
50
51
41
42
43
"A yellow taxi on the road on the left side."
"A red bus on the road in front."
"A big ad-board of <the Phantom of the Opera>, top on the building on the right side."
"I'm standing on the Times Square in NY now."
"The sky is fair in NY today."
45
44
(B)
56
55
検出された共起事象
「人」
「横断歩道」

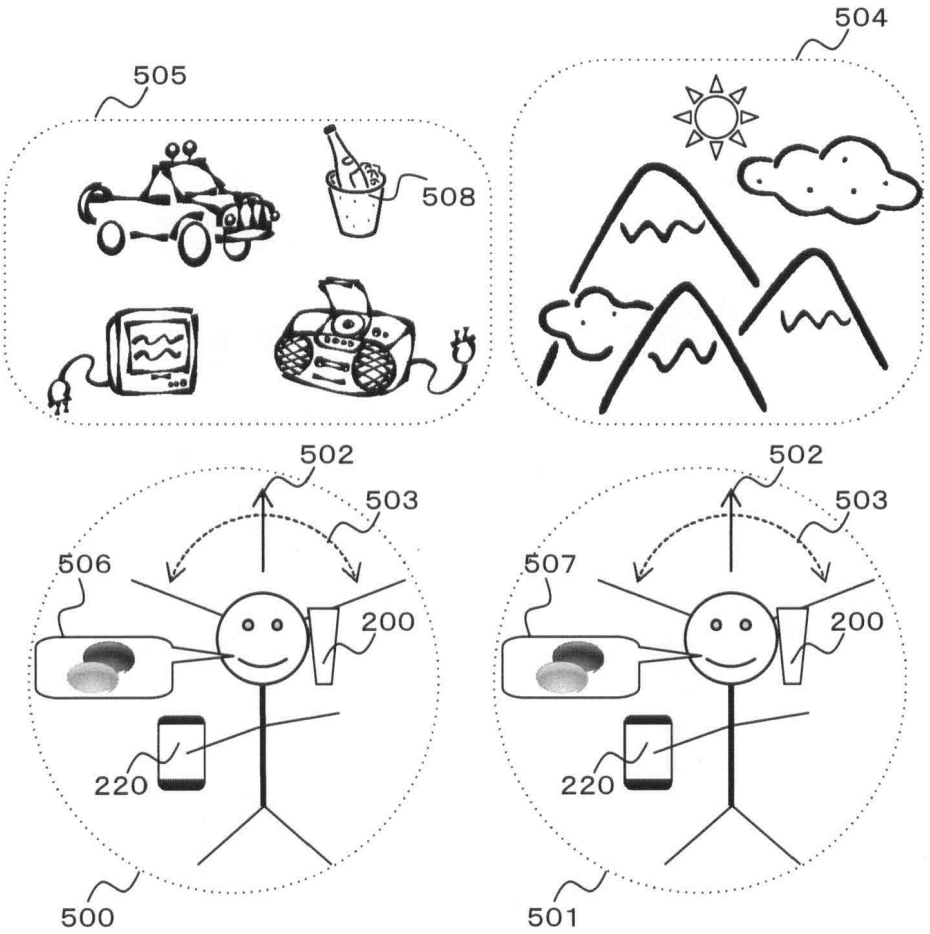
【図１０】


306
メッセージ保管部
422
303
メッセージノード生成部
インタレストグラフ部
391
421
423
メッセージ管理部
420
メッセージ
データベース

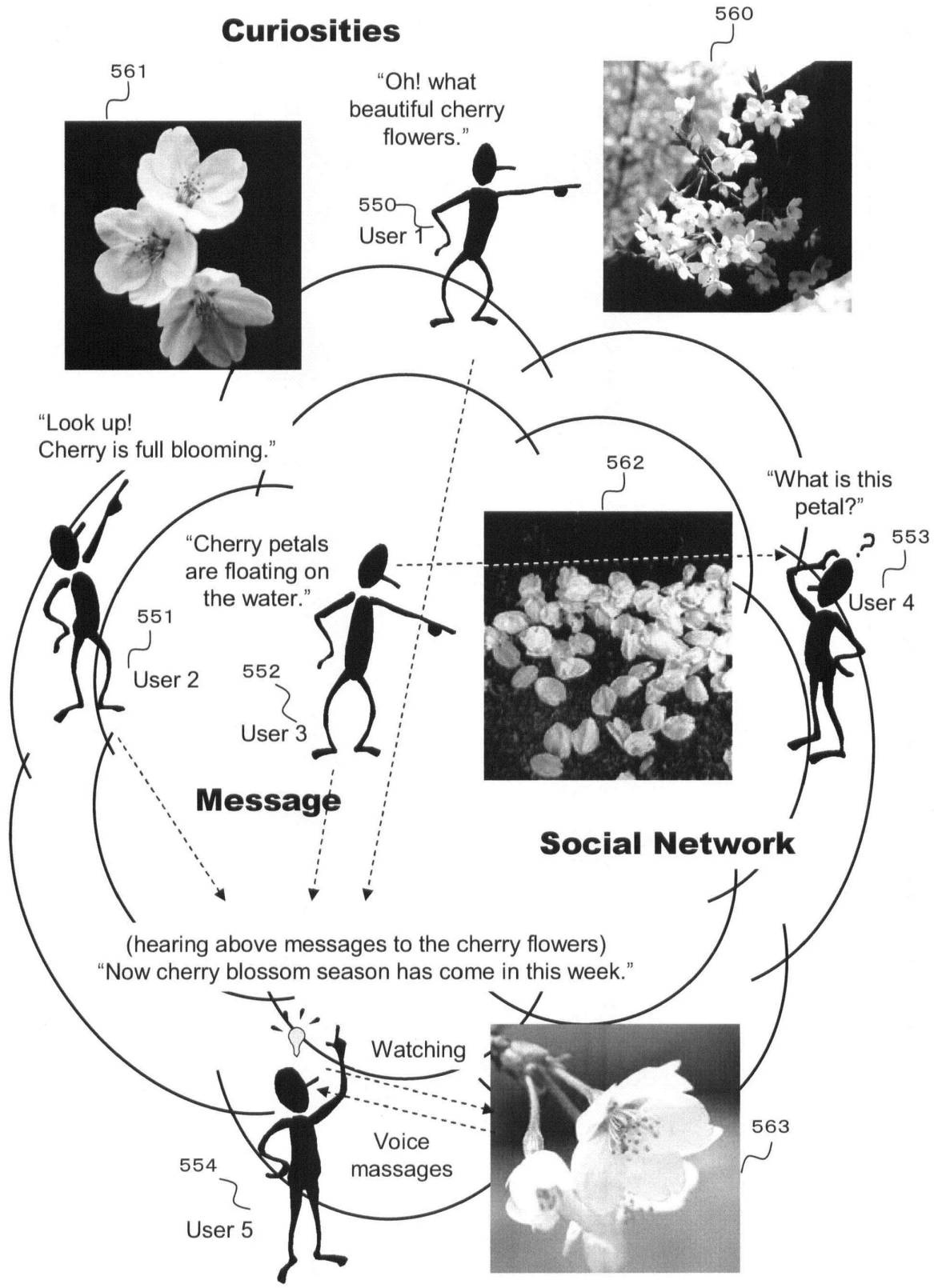
【図１１】


391
320
音声認識システム
307
再生処理部
305
状況認識部
430
会話エンジン
431
着目処理部
432
コマンド処理部
433
ユーザメッセージ
再生部
200
330
音声合成システム

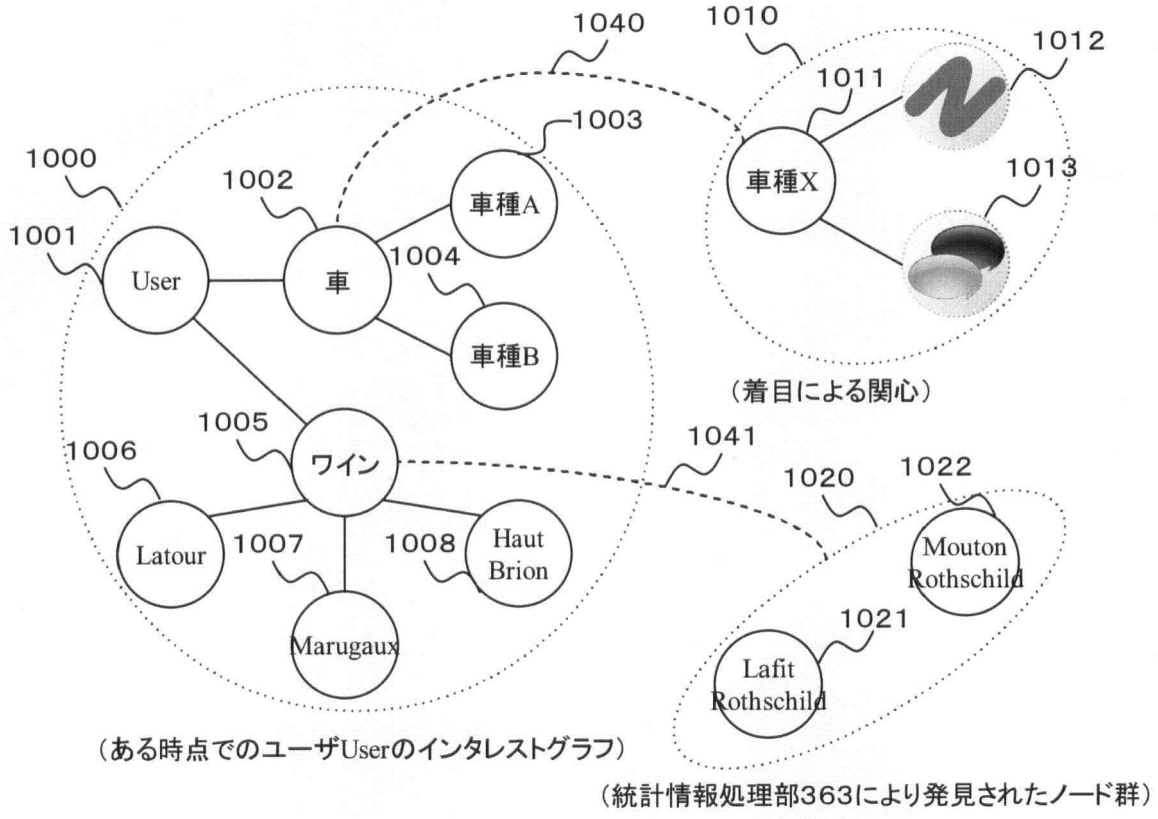
【図１３Ａ】



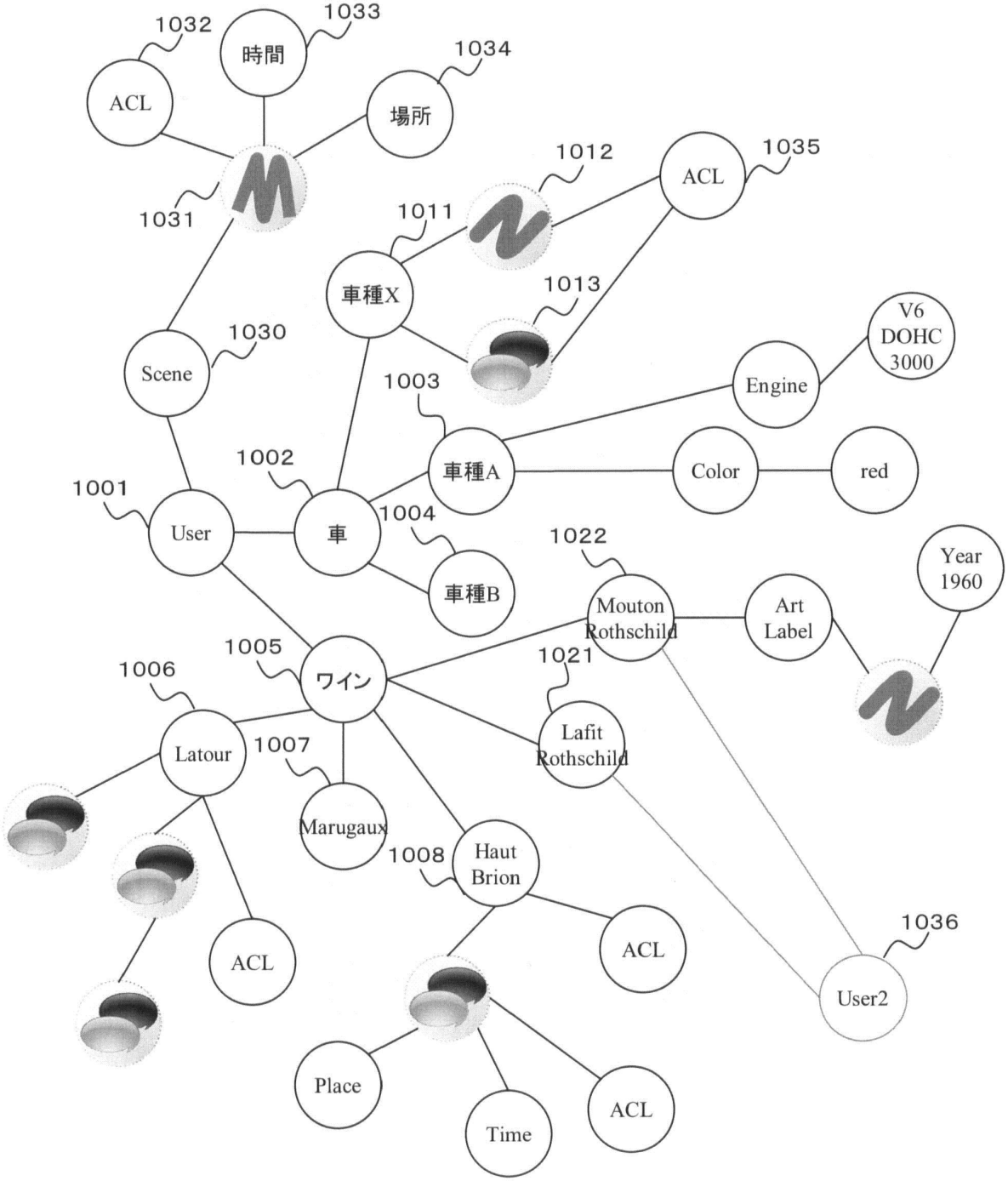
【図１３Ｂ】


Curiosities
560
561
"Oh! what beautiful cherry flowers."
550
User 1
"Look up! Cherry is full blooming."
562
"What is this petal?"
553
User 4
"Cherry petals are floating on the water."
551
User 2
552
User 3
Message
Social Network
(hearing above messages to the cherry flowers)
"Now cherry blossom season has come in this week."
Watching
Voice massages
563
554
User 5

【図１６】

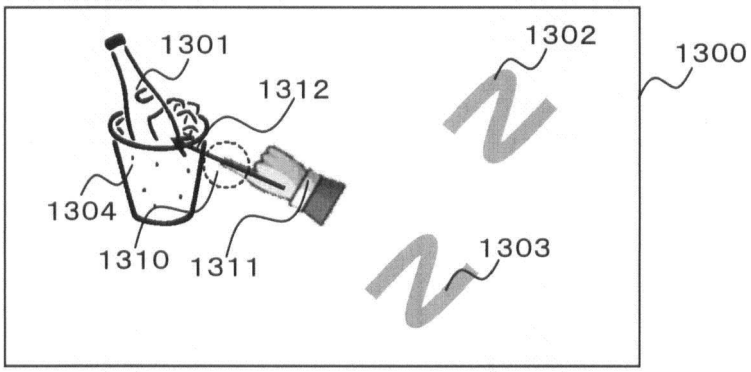

1040
1010
1012
1011
1003
1000
1002
車種A
車種X
1013
1001
User
車
1004
車種B
（着目による関心）
1005
1041
1006
ワイン
1020
1022
Latour
1007
1008
Haut
Brion
Mouton
Rothschild
1021
Marugaux
Lafit
Rothschild
（ある時点でのユーザUserのインタレストグラフ）
（統計情報処理部363により発見されたノード群）

【図１７】

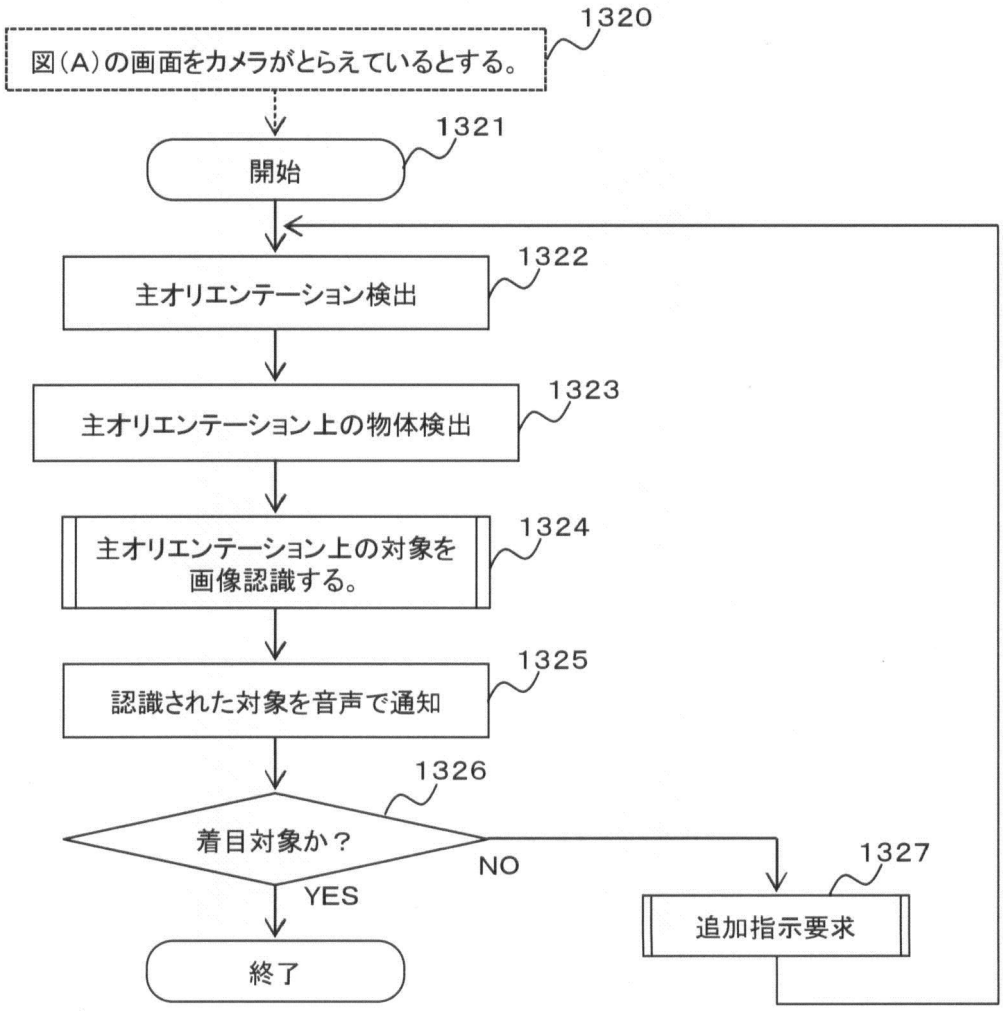

1032
時間
1033
ACL
1034
場所
1031
1012
ACL
1035
1011
車種X
1013
V6
DOHC
3000
Scene
1030
1003
Engine
1002
車種A
Color
red
1001
User
車
1004
車種B
1022
Year
1960
Mouton
Rothschild
Art
Label
1005
ワイン
1021
1006
Latour
1007
Lafit
Rothschild
Marugaux
1008
Haut
Brion
ACL
ACL
1036
User2
Place
Time
ACL

【図２０】

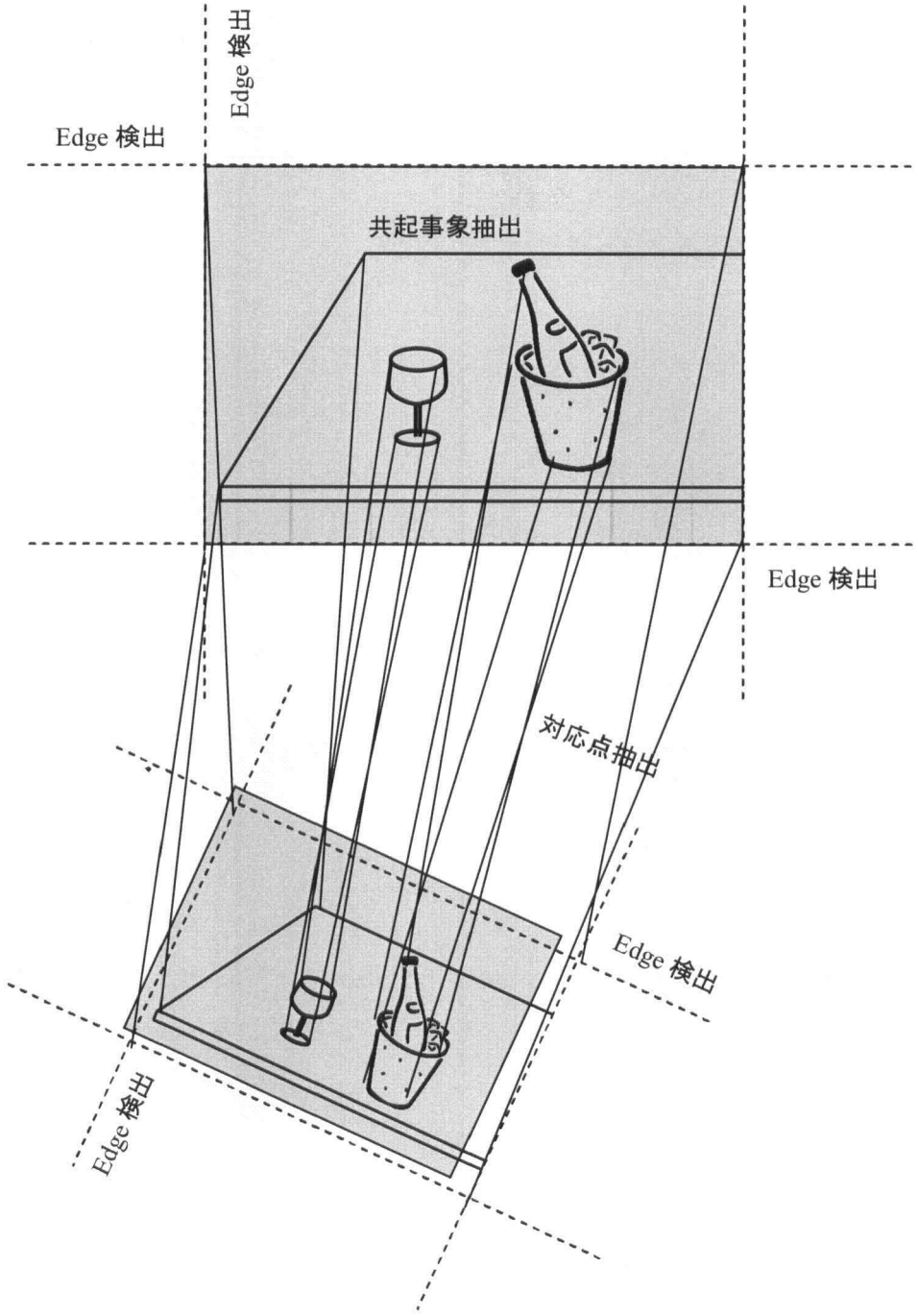
(A)


1301
1302
1300
1312
1304
1310
1311
1303

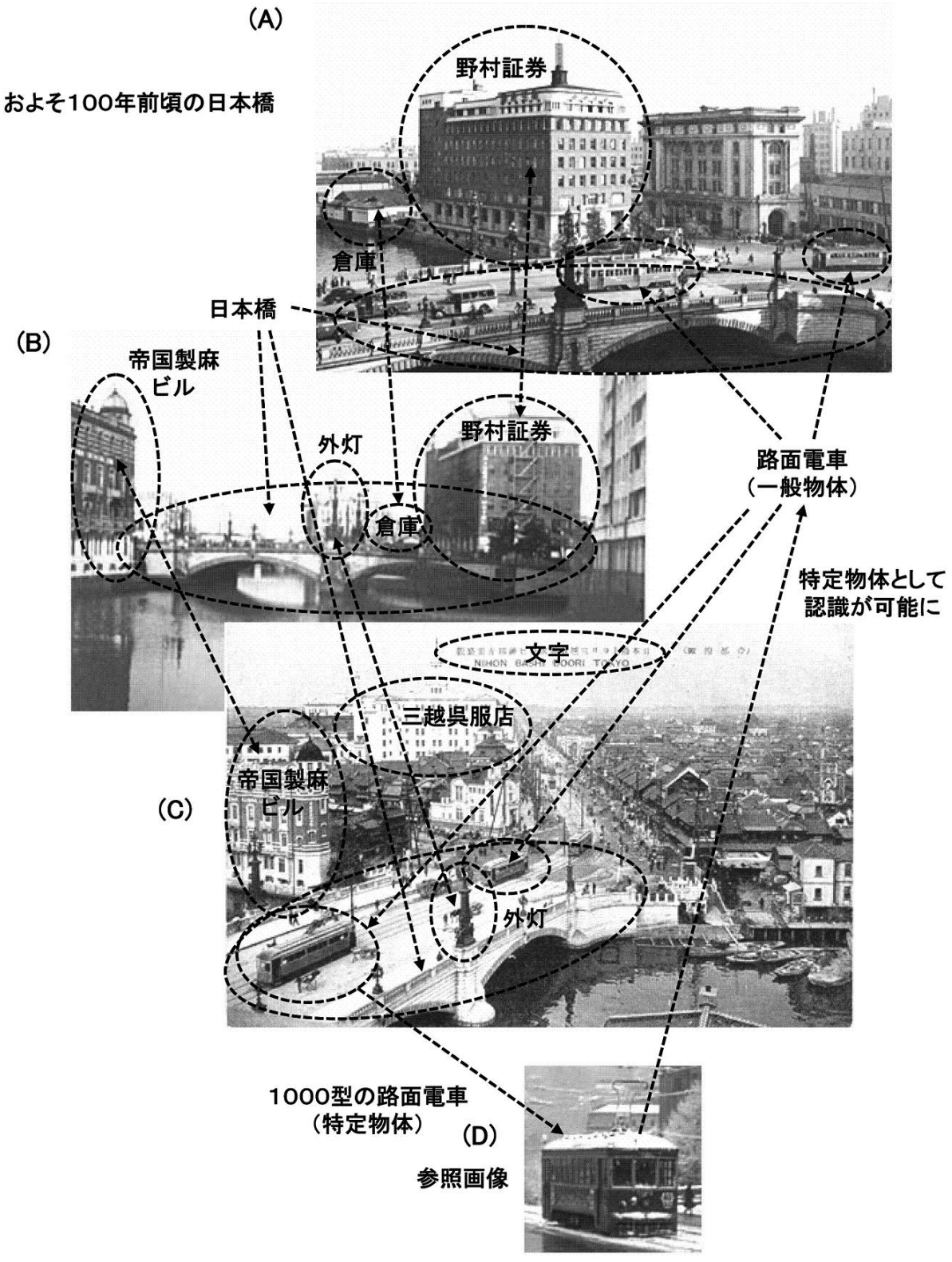
(B)


図(A)の画面をカメラがとらえているとする。
1320
開始
1321
主オリエンテーション検出
1322
主オリエンテーション上の物体検出
1323
主オリエンテーション上の対象を
画像認識する。
1324
認識された対象を音声で通知
1325
着目対象か？
1326
NO
YES
追加指示要求
1327
終了

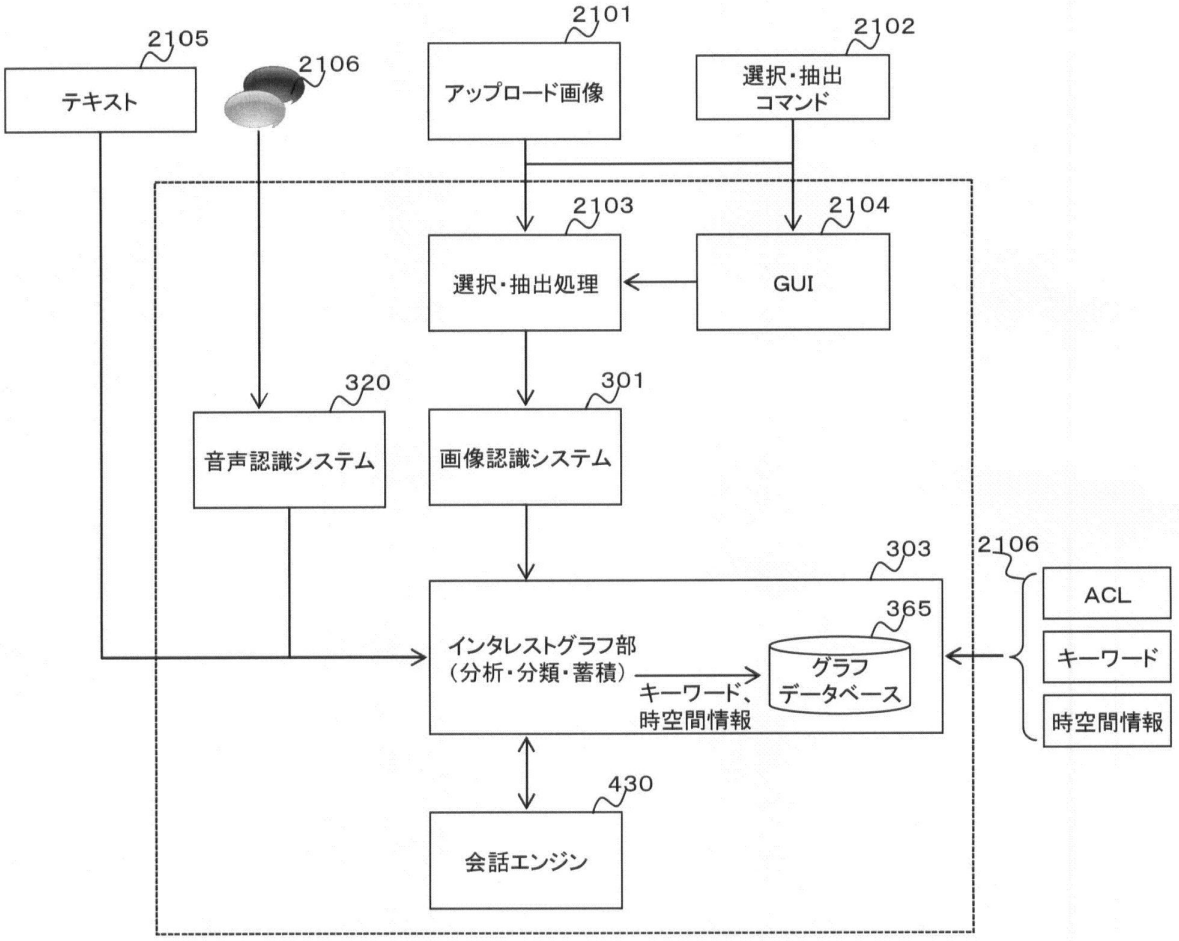
【図２２】



【図２９】


(A)
およそ100年前頃の日本橋
野村証券
倉庫
日本橋
(B)
帝国製麻
ビル
外灯
野村証券
倉庫
路面電車
（一般物体）
特定物体として
認識が可能に
文字
三越呉服店
帝国製麻
ビル
(C)
外灯
1000型の路面電車
（特定物体）
(D)
参照画像

【図３１】


2105
テキスト
2106
2101
アップロード画像
2102
選択・抽出
コマンド
2103
選択・抽出処理
2104
GUI
320
音声認識システム
301
画像認識システム
303
2106
ACL
365
インタレストグラフ部
（分析・分類・蓄積）
キーワード、
時空間情報
グラフ
データベース
キーワード
時空間情報
430
会話エンジン

フロントページの続き

(51)Int.Cl.		FI		
G10L 13/00	*(2006.01)*	G10L	15/28	500
		G10L	15/00	200A
		G10L	13/00	

(74)代理人 100109335
弁理士 上杉 浩
(74)代理人 100136744
弁理士 中村 佳正
(72)発明者 久夛良木 健
東京都世田谷区瀬田１－５－１ サイバーアイ・エンタテインメント株式会社内
(72)発明者 薄 隆
東京都世田谷区瀬田１－５－１ サイバーアイ・エンタテインメント株式会社内
(72)発明者 横手 靖彦
東京都世田谷区瀬田１－５－１ サイバーアイ・エンタテインメント株式会社内

審査官 木村 雅也

(56)参考文献 国際公開第２０１１／０８１１９４（WO，Ａ１）
特開２０１１－１３７６３８（ＪＰ，Ａ）
特開２００５－１９６４８１（ＪＰ，Ａ）
特開２００９－０７７４４３（ＪＰ，Ａ）
特開２００８－２７８０８８（ＪＰ，Ａ）
国際公開第２０１１／００４６０８（WO，Ａ１）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ １３／００
Ｇ０６Ｆ １７／３０
Ｇ０６Ｑ ５０／００
Ｇ１０Ｌ １３／００
Ｇ１０Ｌ １５／００
Ｇ１０Ｌ １５／２８