

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2010-224715

(P2010-224715A)

(43) 公開日 平成22年10月7日(2010.10.7)

(51) Int.Cl.	F I	テーマコード (参考)
G06F 3/048 (2006.01)	G06F 3/048 651B	5B050
G06F 17/30 (2006.01)	G06F 17/30 210A	5B075
G10L 15/00 (2006.01)	G10L 15/00 200T	5D015
G06T 1/00 (2006.01)	G06T 1/00 200D	5E501
G06F 3/16 (2006.01)	G06F 3/048 656C	

審査請求 未請求 請求項の数 25 O L (全 39 頁) 最終頁に続く

(21) 出願番号 特願2009-69398 (P2009-69398)
 (22) 出願日 平成21年3月23日 (2009. 3. 23)

(71) 出願人 000000376
 オリンパス株式会社
 東京都渋谷区幡ヶ谷2丁目4番2号
 (74) 代理人 100104710
 弁理士 竹腰 昇
 (74) 代理人 100124682
 弁理士 黒田 泰
 (74) 代理人 100090479
 弁理士 井上 一
 (72) 発明者 杉原 良平
 東京都渋谷区幡ヶ谷2丁目4番2号 オリンパス株式会社内
 (72) 発明者 龍田 成示
 東京都渋谷区幡ヶ谷2丁目4番2号 オリンパス株式会社内

最終頁に続く

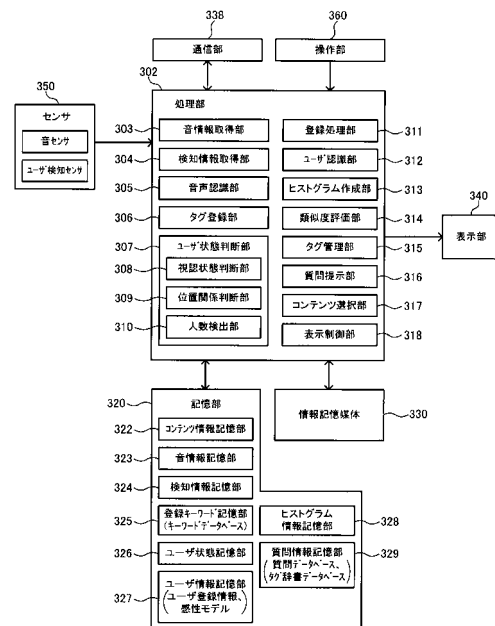
(54) 【発明の名称】 画像表示システム、デジタルフォトフレーム、情報処理システム、プログラム及び情報記憶媒体

(57) 【要約】

【課題】コンテンツに対する効率的なタグの付与を可能にする画像表示システム、デジタルフォトフレーム、情報処理システム、プログラム及び情報記憶媒体の提供。

【解決手段】画像表示システムは、画像を表示する表示部340の表示制御を行う表示制御部318と、コンテンツの表示中におけるユーザの発話の音声認識を行い、ユーザの発話キーワードを抽出する音声認識部305と、抽出された発話キーワードに基づいて、コンテンツのタグキーワードの登録処理を行うタグ登録部306を含む。

【選択図】図2



【特許請求の範囲】**【請求項 1】**

画像を表示する表示部の表示制御を行う表示制御部と、
コンテンツの表示中におけるユーザの発話の音声認識を行い、ユーザの発話キーワードを抽出する音声認識部と、
抽出された発話キーワードに基づいて、コンテンツのタグキーワードの登録処理を行うタグ登録部と、
を含むことを特徴とする画像表示システム。

【請求項 2】

請求項 1 において、
前記タグ登録部は、
抽出された前記発話キーワードの中に、コンテンツのタグキーワードに一致する一致発話キーワードが検出された場合に、検出された前記一致発話キーワード以外の発話キーワードを、コンテンツのタグキーワードとして登録することを特徴とする画像表示システム。

10

【請求項 3】

請求項 1 又は 2 において、
前記タグ登録部は、
抽出された前記発話キーワードの中に、コンテンツのタグキーワードに関連する関連発話キーワードが検出された場合に、検出された前記関連発話キーワードを、コンテンツのタグキーワードとして登録することを特徴とする画像表示システム。

20

【請求項 4】

請求項 1 乃至 3 のいずれかにおいて、
ユーザが前記表示部を注視している状態か否かを判断する視認状態判断部を含み、
前記タグ登録部は、
ユーザが注視状態であるときに抽出された発話キーワードに基づいて、コンテンツのタグキーワードの登録処理を行うことを特徴とする画像表示システム。

【請求項 5】

請求項 1 乃至 4 のいずれかにおいて、
ユーザと前記表示部との位置関係を判断する位置関係判断部を含み、
前記タグ登録部は、
ユーザと前記表示部との距離が所定距離以内であるときに抽出された発話キーワードに基づいて、コンテンツのタグキーワードの登録処理を行うことを特徴とする画像表示システム。

30

【請求項 6】

請求項 1 乃至 5 のいずれかにおいて、
ユーザの人数の検出処理を行う人数検出部を含み、
前記タグ登録部は、
ユーザの人数の検出結果に基づいて、コンテンツのタグキーワードの登録処理を行うことを特徴とする画像表示システム。

40

【請求項 7】

請求項 6 において、
前記音声認識部は、
ユーザの人数が一人であると検出された場合に、ユーザの発話キーワードを抽出し、
前記タグ登録部は、
抽出された一人のユーザの発話キーワードに基づいて、コンテンツのタグキーワードの登録処理を行うことを特徴とする画像表示システム。

【請求項 8】

請求項 6 又は 7 において、
前記タグ登録部は、

50

ユーザの人数が複数人であると検出され、複数人のユーザが同一の発話キーワードを発話したことが検出された場合に、検出された同一発話キーワードに基づいて、コンテンツのタグキーワードの登録処理を行うことを特徴とする画像表示システム。

【請求項 9】

請求項 1 乃至 8 のいずれかにおいて、
ユーザを認識するユーザ認識部を含み、
前記タグ登録部は、

認識されたユーザの発話キーワードに基づいて、コンテンツのタグキーワードが取得された場合に、取得されたタグキーワードを、認識されたユーザ及びコンテンツに関連づけて登録することを特徴とする画像表示システム。

10

【請求項 10】

請求項 1 乃至 9 のいずれかにおいて、
前記タグ登録部は、

ユーザの発話キーワードの出現回数情報の集計処理を行い、出現回数情報が所定しきい値を超えた発話キーワードを、コンテンツのタグキーワードとして登録することを特徴とする画像表示システム。

【請求項 11】

請求項 1 乃至 10 のいずれかにおいて、
前記音声認識部は、

ユーザの無発話状態を検知し、

前記表示制御部は、

ユーザの無発話状態が検知された場合に、表示されるコンテンツの切り替え処理を行うことを特徴とする画像表示システム。

20

【請求項 12】

請求項 1 乃至 11 のいずれかにおいて、

コンテンツの登録タグキーワードと、前記登録タグキーワードの登録回数情報とが関連づけられたヒストグラム情報を作成するヒストグラム作成部を含むことを特徴とする画像表示システム。

【請求項 13】

請求項 12 において、

ユーザに提示するコンテンツの選択処理を行うコンテンツ選択部を含み、

前記コンテンツ選択部は、

作成された前記ヒストグラム情報に基づいて、ユーザに提示するコンテンツの選択処理を行うことを特徴とする画像表示システム。

30

【請求項 14】

請求項 13 において、

複数のコンテンツ間の前記ヒストグラム情報の類似度を評価する類似度評価部を含み、
前記コンテンツ選択部は、

第 1 のコンテンツの次にユーザに提示する第 2 のコンテンツとして、前記第 1 のコンテンツに対する前記ヒストグラム情報の類似度が高いと評価されたコンテンツを選択することを特徴とする画像表示システム。

40

【請求項 15】

請求項 12 乃至 14 のいずれかにおいて、

前記ヒストグラム情報に基づいて、コンテンツに対して登録された登録タグキーワードの整理処理を行うタグ管理部を含むことを特徴とする画像表示システム。

【請求項 16】

請求項 15 において、

前記タグ管理部は、

コンテンツの登録タグキーワードの中から、前記ヒストグラム情報における前記登録回数情報が少ない登録タグキーワードを優先して削除する処理を行うことを特徴とする画像

50

表示システム。

【請求項 17】

請求項 15 又は 16 において、
前記タグ管理部は、
コンテンツの登録タグキーワードの中から、登録が古い登録タグキーワードを優先して
削除する処理を行うことを特徴とする画像表示システム。

【請求項 18】

請求項 1 乃至 17 のいずれかにおいて、
タグキーワードを取得するための質問の提示処理を行う質問提示部を含み、
前記タグ登録部は、
提示された質問に対するユーザの反応に基づいて、コンテンツのタグキーワードの登録
処理を行うことを特徴とする画像表示システム。

10

【請求項 19】

請求項 18 において、
前記質問提示部は、
コンテンツに関連づけられたタグキーワードに対応する質問を提示する処理を行うこと
を特徴とする画像表示システム。

【請求項 20】

請求項 18 又は 19 において、
前記質問提示部は、
コンテンツに関連づけられたタグキーワードの数が所定数以下である場合に、質問を提
示する処理を行うことを特徴とする画像表示システム。

20

【請求項 21】

請求項 18 乃至 20 のいずれかにおいて、
前記質問提示部は、
ユーザの発話から抽出された発話キーワードが、第 1 ~ 第 N の階層のタグキーワードの
うちの第 K の階層のタグキーワードである場合には、前記第 K の階層のタグキーワード又
は前記第 K の階層のタグキーワードよりも下位の第 K + 1 の階層のタグキーワードを用い
て質問を作成して提示する処理を行うことを特徴とする画像表示システム。

【請求項 22】

請求項 1 乃至 21 のいずれか記載の画像表示システムであることを特徴とするデジタル
フォトフレーム。

30

【請求項 23】

ユーザに提示するコンテンツの選択処理を行うコンテンツ選択部と、
選択されたコンテンツに基づいて、画像表示システムの表示部に表示される画像の表示
指示を行う表示指示部と、
コンテンツの表示中におけるユーザの発話の音声認識により抽出されたユーザの発話キ
ーワードに基づいて、コンテンツのタグキーワードの登録処理を行うタグ登録部と、
を含むことを特徴とする情報処理システム。

【請求項 24】

画像を表示する表示部の表示制御を行う表示制御部と、
コンテンツの表示中におけるユーザの発話の音声認識を行い、ユーザの発話キーワー
ドを抽出する音声認識部と、
抽出された発話キーワードに基づいて、コンテンツのタグキーワードの登録処理を行う
タグ登録部として、
コンピュータを機能させることを特徴とするプログラム。

40

【請求項 25】

コンピュータ読み取り可能な情報記憶媒体であって、請求項 24 に記載のプログラムを
記憶したことを特徴とする情報記憶媒体。

【発明の詳細な説明】

50

【技術分野】**【0001】**

本発明は、画像表示システム、デジタルフォトフレーム、情報処理システム、プログラム及び情報記憶媒体等に関する。

【背景技術】**【0002】**

近年、デジタルカメラ等の普及により、ユーザは撮影枚数を気にすることなく、手軽に写真を撮影できるようになったが、その反面、ユーザにより撮影された写真の多くは、閲覧されることなく、保存されるだけという状況になっている。例えば撮影しても半分以上の写真が長期間に亘って閲覧されることがなく、ハードディスク等のストレージに保存されたままになっている。

10

【0003】

このような問題を解消するために、写真等のコンテンツ画像にタグを付与し、付与されたタグに基づいて、ストレージ保存されたコンテンツ画像のスライドショー等を行う手法が考えられる。

【0004】

しかしながら、コンテンツ画像に対してタグを付与する作業はユーザにとって煩雑であり、手間がかかる作業である。このため、コンテンツ画像に対して効率的にタグを付与することが難しいという課題がある。

【0005】

また、近年、デジタルカメラで撮影した画像を手軽に再生できる装置として、デジタルフォトフレームが脚光を浴びている。このデジタルフォトフレームは、フォトスタンドの写真を入れる部分が液晶ディスプレイに置き換えられた形態の装置であり、メモリカードや通信装置を介して読み込まれたデジタルの画像データ（電子写真）の再生処理を行う。

20

【0006】

デジタルフォトフレームの従来技術としては例えば特許文献1に開示される技術がある。この従来技術では、電話回線接続装置を、デジタルフォトフレームであるデジタルフォトスタンドに具備させて、フォトスタンドと有線又は無線の電話回線との間の伝送路の形成を実現している。

【0007】

しかしながら、これまでのデジタルフォトフレームでは、デジタルカメラ等で撮影された画像を単に再生する機能しか有しておらず、再生画像が単調であり、ユーザに対して多様なコンテンツ画像を表示できないという課題がある。

30

【先行技術文献】**【特許文献】****【0008】**

【特許文献1】特開2000-324473号公報

【発明の概要】**【発明が解決しようとする課題】****【0009】**

本発明の幾つかの態様によれば、コンテンツに対する効率的なタグの付与を可能にする画像表示システム、デジタルフォトフレーム、情報処理システム、プログラム及び情報記憶媒体等を提供できる。

40

【課題を解決するための手段】**【0010】**

本発明の一態様は、画像を表示する表示部の表示制御を行う表示制御部と、コンテンツの表示中におけるユーザの発話の音声認識を行い、ユーザの発話キーワードを抽出する音声認識部と、抽出された発話キーワードに基づいて、コンテンツのタグキーワードの登録処理を行うタグ登録部とを含む画像表示システムに係る。また本発明は、上記各部としてコンピュータを機能させるプログラム、又は該プログラムを記憶したコンピュータ読

50

み取り可能な情報記憶媒体に係る。

【0011】

本発明の一態様によれば、コンテンツの表示中におけるユーザの発話の音声認識が行われて、発話キーワードが抽出される。そして抽出された発話キーワードに基づいて、コンテンツのタグキーワードの登録処理が行われる。例えば、抽出された発話キーワードがタグキーワードとして登録されたり、或いは発話キーワードにより取得されたキーワードがタグキーワードとして登録される。このように、コンテンツ表示中の発話を検知してそのコンテンツのタグキーワードの登録処理を行うようにすれば、コンテンツに対する効率的なタグの付与が可能になる。

【0012】

また本発明の一態様では、前記タグ登録部は、抽出された前記発話キーワードの中に、コンテンツのタグキーワードに一致する一致発話キーワードが検出された場合に、検出された前記一致発話キーワード以外の発話キーワードを、コンテンツのタグキーワードとして登録してもよい。

【0013】

このようにすれば、発話中のキーワードのうち、一致した発話キーワード以外のキーワードをタグキーワードとして登録できるようになり、効率的なタグの付与が可能になる。

【0014】

また本発明の一態様では、前記タグ登録部は、抽出された前記発話キーワードの中に、コンテンツのタグキーワードに関連する関連発話キーワードが検出された場合に、検出された前記関連発話キーワードを、コンテンツのタグキーワードとして登録してもよい。

【0015】

このようにすれば、発話中のキーワードのうち、関連発話キーワードをタグキーワードとして登録できるようになり、多様なタグの付与が可能になる。

【0016】

また本発明の一態様では、ユーザが前記表示部を注視している状態か否かを判断する視認状態判断部を含み、前記タグ登録部は、ユーザが注視状態であるときに抽出された発話キーワードに基づいて、コンテンツのタグキーワードの登録処理を行ってもよい。

【0017】

このようにすれば、ユーザの注視状態を条件にタグキーワードの登録処理が行われるようになるため、無用なタグキーワード等が登録されてしまう事態を防止できる。

【0018】

また本発明の一態様では、ユーザと前記表示部との位置関係を判断する位置関係判断部を含み、前記タグ登録部は、ユーザと前記表示部との距離が所定距離以内であるときに抽出された発話キーワードに基づいて、コンテンツのタグキーワードの登録処理を行ってもよい。

【0019】

このようにすれば、ユーザとの距離が所定距離以内であることを条件にタグキーワードの登録処理が行われるようになるため、無用なタグキーワード等が登録されてしまう事態を防止できる。

【0020】

また本発明の一態様では、ユーザの人数の検出処理を行う人数検出部を含み、前記タグ登録部は、ユーザの人数の検出結果に基づいて、コンテンツのタグキーワードの登録処理を行ってもよい。

【0021】

このようにすれば、ユーザの人数の検出結果を反映したタグキーワードの登録処理を実現できる。

【0022】

また本発明の一態様では、前記音声認識部は、ユーザの人数が一人であると検出された場合に、ユーザの発話キーワードを抽出し、前記タグ登録部は、抽出された一人のユーザ

10

20

30

40

50

の発話キーワードに基づいて、コンテンツのタグキーワードの登録処理を行ってもよい。

【0023】

このようにすれば、ユーザが1人であるときの発話キーワードに基づいて、タグキーワードの登録処理を実現できる。

【0024】

また本発明の一態様では、前記タグ登録部は、ユーザの人数が複数人であると検出され、複数人のユーザが同一の発話キーワードを発話したことが検出された場合に、検出された同一発話キーワードに基づいて、コンテンツのタグキーワードの登録処理を行ってもよい。

【0025】

このようにすれば複数人が発話した同一の発話キーワードに基づいて、タグキーワードの登録処理を実現できる。

【0026】

また本発明の一態様では、ユーザを認識するユーザ認識部を含み、前記タグ登録部は、認識されたユーザの発話キーワードに基づいて、コンテンツのタグキーワードが取得された場合に、取得されたタグキーワードを、認識されたユーザ及びコンテンツに関連づけて登録してもよい。

【0027】

このようにすれば、各ユーザの発話に対応したタグキーワードをコンテンツに付与できるようになる。

【0028】

また本発明の一態様では、前記タグ登録部は、ユーザの発話キーワードの出現回数情報の集計処理を行い、出現回数情報が所定しきい値を超えた発話キーワードを、コンテンツのタグキーワードとして登録してもよい。

【0029】

このようにすれば、無用なタグキーワード等が登録されてしまう事態を防止できる。

【0030】

また本発明の一態様では、前記音声認識部は、ユーザの無発話状態を検知し、前記表示制御部は、ユーザの無発話状態が検知された場合に、表示されるコンテンツの切り替え処理を行ってもよい。

【0031】

このようにすれば、ユーザが表示コンテンツに興味等を持って発話している状態では、現在の表示コンテンツの表示が維持され、無発話状態が検知されると、次のコンテンツに切り替わるようになる。

【0032】

また本発明の一態様では、コンテンツの登録タグキーワードと、前記登録タグキーワードの登録回数情報とが関連づけられたヒストグラム情報を作成するヒストグラム作成部を含んでもよい。

【0033】

このようにすれば、登録タグキーワードのヒストグラム情報を利用した種々の処理を実現できる。

【0034】

また本発明の一態様では、ユーザに提示するコンテンツの選択処理を行うコンテンツ選択部を含み、前記コンテンツ選択部は、作成された前記ヒストグラム情報に基づいて、ユーザに提示するコンテンツの選択処理を行ってもよい。

【0035】

このようにすれば、登録タグキーワードのヒストグラム情報を有効活用してコンテンツを選択できるようになる。

【0036】

また本発明の一態様では、複数のコンテンツ間の前記ヒストグラム情報の類似度を評価

10

20

30

40

50

する類似度評価部を含み、前記コンテンツ選択部は、第1のコンテンツの次にユーザに提示する第2のコンテンツとして、前記第1のコンテンツに対する前記ヒストグラム情報の類似度が高いと評価されたコンテンツを選択してもよい。

【0037】

このようにすれば、ヒストグラム情報に基づき第1のコンテンツに類似すると評価された第2のコンテンツを、次のコンテンツとして表示できるようになる。

【0038】

また本発明の一態様では、前記ヒストグラム情報に基づいて、コンテンツに対して登録された登録タグキーワードの整理処理を行うタグ管理部を含んでもよい。

【0039】

このようにすれば、ヒストグラム情報を有効利用して、無駄な登録タグキーワード等を整理することが可能になる。

【0040】

また本発明の一態様では、前記タグ管理部は、コンテンツの登録タグキーワードの中から、前記ヒストグラム情報における前記登録回数情報が少ない登録タグキーワードを優先して削除する処理を行ってもよい。

【0041】

このようにすれば、間違って登録されたと考えられる登録タグキーワード等を削除することが可能になる。

【0042】

また本発明の一態様では、前記タグ管理部は、コンテンツの登録タグキーワードの中から、登録が古い登録タグキーワードを優先して削除する処理を行ってもよい。

【0043】

このようにすれば、あまり利用されていないと考えられる古い登録タグキーワード等を削除できる。

【0044】

また本発明の一態様では、タグキーワードを取得するための質問の提示処理を行う質問提示部を含み、前記タグ登録部は、提示された質問に対するユーザの反応に基づいて、コンテンツのタグキーワードの登録処理を行ってもよい。

【0045】

このようにすれば、質問によりユーザに働きかけてコンテンツのタグキーワードを登録できるようになる。

【0046】

また本発明の一態様では、前記質問提示部は、コンテンツに関連づけられたタグキーワードに対応する質問を提示する処理を行ってもよい。

【0047】

このようにすれば、コンテンツに関連した質問をユーザに提示して、タグキーワードを登録できるようになる。

【0048】

また本発明の一態様では、前記質問提示部は、コンテンツに関連づけられたタグキーワードの数が所定数以下である場合に、質問を提示する処理を行ってもよい。

【0049】

このようにすれば、タグキーワード数が足りないコンテンツに対してタグキーワードを付与できるようになる。

【0050】

また本発明の一態様では、前記質問提示部は、ユーザの発話から抽出された発話キーワードが、第1～第Nの階層のタグキーワードのうち第Kの階層のタグキーワードである場合には、前記第Kの階層のタグキーワード又は前記第Kの階層のタグキーワードよりも下位の第K+1の階層のタグキーワードを用いて質問を作成して提示する処理を行ってもよい。

10

20

30

40

50

【 0 0 5 1 】

このようにすれば、ユーザの発話を手がかりとして、より具体的なタグを付与するための質問を作成して、ユーザに提示できるようになる。

【 0 0 5 2 】

また本発明の他の態様は、上記のいずれか記載の画像表示システムであるデジタルフォトフレームに関する。

【 0 0 5 3 】

また本発明の他の態様は、ユーザに提示するコンテンツの選択処理を行うコンテンツ選択部と、選択されたコンテンツに基づいて、画像表示システムの表示部に表示される画像の表示指示を行う表示指示部と、コンテンツの表示中におけるユーザの発話の音声認識により抽出されたユーザの発話キーワードに基づいて、コンテンツのタグキーワードの登録処理を行うタグ登録部とを含む情報処理システムに関する。

【 図面の簡単な説明 】

【 0 0 5 4 】

【 図 1 】 図 1 (A)、図 1 (B) は、画像表示システムの 1 つであるデジタルフォトフレームの例を示す図。

【 図 2 】 本実施形態の画像表示システムの構成例。

【 図 3 】 ユーザの発話キーワードからタグキーワードを登録する手法の説明図。

【 図 4 】 ユーザの発話キーワードからタグキーワードを登録する手法の説明図。

【 図 5 】 ユーザの発話キーワードからタグキーワードを登録する処理のフローチャート。

【 図 6 】 図 6 (A)、図 6 (B) はユーザの注視を条件に発話キーワードからタグキーワードを登録する手法の説明図。

【 図 7 】 ユーザの注視を条件に発話キーワードからタグキーワードを登録する処理のフローチャート。

【 図 8 】 ユーザの注視を条件に発話キーワードからタグキーワードを登録する他の処理のフローチャート。

【 図 9 】 注視状態の判断処理を説明するためのフローチャート。

【 図 1 0 】 ユーザの接近を条件に発話キーワードからタグキーワードを登録する処理のフローチャート。

【 図 1 1 】 図 1 1 (A) ~ 図 1 1 (C) はユーザとの位置関係の検出手法の説明図。

【 図 1 2 】 ユーザの人数を検出して発話キーワードからタグキーワードを登録する処理のフローチャート。

【 図 1 3 】 ユーザの人数を検出して発話キーワードからタグキーワードを登録する他の処理のフローチャート。

【 図 1 4 】 図 1 4 (A)、図 1 4 (B) はユーザの登録情報についての説明図。

【 図 1 5 】 タグ登録処理及びコンテンツ切り替え処理の詳細例のフローチャート。

【 図 1 6 】 図 1 6 (A)、図 1 6 (B) はコンテンツテーブル、登録回数テーブルの例。

【 図 1 7 】 ヒストグラム情報に基づいて次に表示するコンテンツを選択する手法の説明図。

。

【 図 1 8 】 ヒストグラム情報の説明図。

【 図 1 9 】 ヒストグラム情報の説明図。

【 図 2 0 】 図 2 0 (A)、図 2 0 (B) はヒストグラム情報の類似度の評価処理の説明図。

。

【 図 2 1 】 ヒストグラム情報に基づくタグの整理処理のフローチャート。

【 図 2 2 】 タグキーワード付与のための質問提示処理のフローチャート。

【 図 2 3 】 タグ辞書データベースの例。

【 図 2 4 】 質問データベースの例。

【 図 2 5 】 本実施形態のシステム構成の変形例。

【 発明を実施するための形態 】

【 0 0 5 5 】

10

20

30

40

50

以下、本実施形態について説明する。なお、以下に説明する本実施形態は、特許請求の範囲に記載された本発明の内容を不当に限定するものではない。また本実施形態で説明される構成の全てが、本発明の必須構成要件であるとは限らない。

【0056】

1. 構成

図1(A)に、本実施形態の画像表示システムの一例であるデジタルフォトフレーム300(デジタルフォトプレーヤ、画像再生装置)の例を示す。

【0057】

図1(A)は、いわゆるフォトスタンドタイプのデジタルフォトフレームの例である。このデジタルフォトフレーム300は、家の中などの任意の場所にユーザにより設置される。そして、デジタルの画像データや音データなどのコンテンツ情報の再生処理(画像再生、音再生)を実行する。デジタルフォトフレーム300は、例えばユーザの明示的な再生指示がなくても、画像等のコンテンツ情報(メディア情報)を自動的に再生することができる。例えば写真のスライドショーを自動的に実行したり、映像の自動再生を行う。

【0058】

なお図1(A)はフォトスタンドタイプのデジタルフォトフレームの例であるが、例えば図1(B)に示すように、壁掛けタイプのものであってもよい。この壁掛けタイプのデジタルフォトフレームとしては、例えば電気泳動型ディスプレイ等により実現される電子ペーパーなどを用いることができる。また、デジタルフォトフレームに、コンテンツ情報の再生指示ボタンなどを設けたり、リモコンを用いて再生指示ができるようにしてもよい。

【0059】

デジタルフォトフレーム300は、例えばSDカード等のメモリカードのインターフェースを備えることができる。或いは、無線LAN、Bluetoothなどの無線通信のインターフェースや、USB等の有線の通信インターフェースを備えることができる。例えばユーザが、メモリカードにコンテンツ情報を保存して、デジタルフォトフレーム300のメモリカードインターフェースに装着すると、デジタルフォトフレーム300は、メモリカードに保存されたコンテンツ情報の自動再生(スライドショー等)を実行する。或いは、デジタルフォトフレーム300は、無線通信や有線通信により外部からコンテンツ情報を受信すると、このコンテンツ情報の再生処理(自動再生処理)を実行する。例えば、ユーザが所持するデジタルカメラや携帯電話機などの携帯型電子機器がBluetooth等の無線機能を有する場合には、この無線機能を利用して、携帯型電子機器からデジタルフォトフレーム300にコンテンツ情報を転送する。すると、デジタルフォトフレーム300は、転送されたコンテンツ情報の再生処理を実行する。

【0060】

図2に本実施形態の画像表示システムの構成例を示す。この画像表示システムは、処理部302、記憶部320、通信部338、表示部340、センサ350、操作部360を含む。なおこれらの一部の構成要素(例えば通信部、操作部、センサ)を省略したり、他の構成要素(例えばスピーカ)を追加するなどの種々の変形実施が可能である。

【0061】

処理部302は、各種の制御処理や演算処理を行う。例えば記憶部320や表示部340などの上述の各部の制御を行ったり全体的な制御を行う。この処理部302の機能は、各種プロセッサ(CPU等)、ASIC(ゲートアレイ等)などのハードウェアや、当該処理部302に接続された情報記憶媒体330に記憶されたプログラムなどにより実現できる。

【0062】

記憶部320は、処理部302、通信部338などのワーク領域となるものであり、その機能はRAMなどのメモリやHDD(ハードディスクドライブ)などにより実現できる。この記憶部320は、画像や音などのコンテンツ情報を記憶するコンテンツ情報記憶部322、取得された音情報を記憶する音情報記憶部323、取得された検知情報を記憶する検知情報記憶部324、登録キーワードを記憶する登録キーワード記憶部325、特定

10

20

30

40

50

されたユーザ状態を記憶するユーザ状態記憶部 3 2 6、ユーザの登録情報や感性モデル情報等のユーザ情報を記憶するユーザ情報記憶部 3 2 7、タグのヒストグラム情報を記憶するヒストグラム情報記憶部 3 2 8、ユーザに提示する質問情報を記憶する質問情報記憶部 3 2 9（質問データベース、タグ辞書データベース）を含む。

【0063】

情報記憶媒体 3 3 0（コンピュータにより読み取り可能な媒体）は、プログラムやデータなどを格納するものであり、その機能は、メモリカードや光ディスクなどにより実現できる。処理部 3 0 2 は、情報記憶媒体 3 3 0 に格納されるプログラム（データ）に基づいて本実施形態の種々の処理を行う。即ち情報記憶媒体 3 3 0 には、本実施形態の各部としてコンピュータ（操作部、処理部、記憶部、出力部を備える装置）を機能させるためのプログラム（各部の処理をコンピュータに実行させるためのプログラム）が記憶される。

10

【0064】

通信部 3 3 8（通信インターフェース）は、無線や有線の通信などにより外部デバイス（例えばサーバ、携帯型電子機器）との間で情報のやり取りを行うものであり、その機能は、通信用 A S I C 又は通信用プロセッサなどのハードウェアや、通信用ファームウェアにより実現できる。

【0065】

表示部 3 4 0 は、コンテンツ情報である画像を表示するためのものであり、例えば液晶ディスプレイや、有機 E L などの発光素子を用いたディスプレイや、電気泳動型ディスプレイなどにより実現できる。なお表示部 3 4 0 をタッチパネルディスプレイ（タッチスクリーン）により構成してもよい。

20

【0066】

センサ 3 5 0（音センサ、ユーザ検知センサ等）は、検知結果に基づいて検知情報を出力するデバイスである。センサ 3 5 0 としては、例えば音センサを用いることができる。音センサは音を電気信号等に変換するセンサであり、音の物理量である音圧を測定するマイクロフォンなどである。マイクロフォンとしては、ムービング・コイル型やリボン型のダイナミックマイクロフォンや、音声信号の振動による静電容量の変化を検知するコンデンサ型のマイクロフォンや、圧電効果を利用する圧電型のマイクロフォンや、カーボン型のマイクロフォンなどを用いることができる。

【0067】

またセンサ 3 5 0 としてはユーザ検知センサを用いることができる。ユーザ検知センサとしては、焦電センサなどの人感センサや、C C D、C M O S センサなどの撮像センサや、超音波センサなどの距離センサや、ユーザの動作状態（手や体の動き）を検知するモーションセンサなどを用いることができる。

30

【0068】

焦電センサは、人等が発生する赤外線を受光し、赤外線を熱に変換し、その熱を素子の焦電効果で電荷に変えるセンサである。この焦電センサを用いることで、検知範囲（検知エリア）にユーザ（人）が存在するか否かや、検知範囲に存在するユーザの動きや、検知範囲内に存在するユーザの人数などを検知できる。撮像センサ（イメージセンサ）は、1次元又は2次元の光学情報を、時系列の電気信号に変換する光センサである。この撮像センサを用いることで、検知範囲にユーザが存在するか否かや、検知範囲に存在するユーザの動きや、検知範囲内に存在するユーザの人数などを検知できる。また撮像センサを用いた顔画像認識により、ユーザの人物認証を実現できる。また撮像センサを用いた顔検出により、ユーザと表示部 3 4 0 との距離や表示部 3 4 0 に対するユーザの視線の角度などの位置関係を検出できる。或いは、ユーザの視野範囲内に表示部 3 4 0 が入っている状態か否かや、ユーザが表示部 3 4 0 を注視している状態か否かなどのユーザの視認状態を検出できる。或いはユーザが接近中なのか否かなども検出できる。

40

【0069】

なおセンサ 3 5 0 は、センサデバイス自体であってもよいし、センサデバイスの他に制御部や通信部等を含むセンサ機器であってもよい。また検知情報は、センサから直接得ら

50

れる 1 次情報であってもよいし、1 次情報を加工処理（情報処理）することで得られる 2 次情報であってもよい。

【0070】

またセンサ 350 は、画像表示システムに直接に取り付けてもよいし、ホームセンサなどをセンサ 350 として利用してもよい。センサ 350 を取り付ける場合には、図 1（A）に示すように、センサ 350 を例えば枠部分に取り付けることができる。或いは有線のケーブル等を用いてセンサ 350 を接続する形態にしてもよい。

【0071】

操作部 360 は、ユーザが各種情報を入力するためのものであり、例えば操作ボタンやリモコン装置などにより実現できる。ユーザは、この操作部 360 を用いて、ユーザ登録を行ったり、自身が所望する再生コンテンツ（お気に入り画像）の登録などを行うことができる。例えばユーザは、操作部 360 を用いて、ユーザ登録情報を入力することができる。なお表示部 340 がタッチパネルディスプレイにより構成される場合には、表示部 340 が操作部 360 の機能を兼ねることになる。

10

【0072】

処理部 302 は、音情報取得部 303、検知情報取得部 304、音声認識部 305、タグ登録部 306、ユーザ状態判断部 307、登録処理部 311、ユーザ認識部 312、ヒストグラム作成部 313、類似度評価部 314、タグ管理部 315、質問提示部 316、コンテンツ選択部 317、表示制御部 318 を含む。なお、これらの一部の構成要素（例えば検知情報取得部、ユーザ状態判断部、登録処理部、ユーザ認識部、ヒストグラム作成部、類似度評価部、タグ管理部、質問提示部、コンテンツ選択部等）を省略したり、他の構成要素を追加するなどの種々の変形実施が可能である。

20

【0073】

音情報取得部 303 は、センサ 350 の 1 つである音センサにより検知された音情報を取得する処理を行う。例えば音センサにより音声や音楽等の音（周囲音）が検知されて、検知結果である音情報が出力されると、音情報取得部 303 は、その音情報を取り込む。そして取り込まれた音情報は記憶部 320 の音情報記憶部 323 に記憶される。同様に、検知情報取得部 304 は、センサ 350 の 1 つであるユーザ検知センサによりユーザ状態等が検知されて、検知結果である検知情報（撮像情報等）が出力されると、その検知情報を取り込む。そして、取り込まれた検知情報は検知情報記憶部 324 に記憶される。

30

【0074】

なおセンサ 350 として、ホームセンサ等の外部のセンサを用いる場合には、通信部 338 が音情報や検知情報を受信し、音情報取得部 303、検知情報取得部 304 は、受信した音情報や検知情報を取得することになる。

【0075】

音声認識部 305 は、音センサにより取得された音情報に基づいて、ユーザの発話（会話）の音声認識処理を行う。そして音声認識処理によりユーザの発話キーワード（単語）を抽出する。或いはユーザの発話音声からユーザの話者認識を行ったり、ユーザの感情状態を認識するようにしてもよい。また音声認識部 305 はユーザの発話検知を行ってもよい。例えば人間の音声に特有の周波数成分や振幅（パワー）を有する音を抽出して、ユーザの発話を雑音から区別して検知する。具体的には一定の振幅レベル以上の音に対して特定周波数帯域を通過させるフィルタ処理等を行ってユーザの発話を検知する。

40

【0076】

ここで音声認識は、人間（出演者、ユーザ）の話す音声言語をコンピュータにより解析し、テキストデータとして抽出する処理である。音声認識は、音響分析、認識デコーダ、音響モデル、辞書、言語モデルにより実現される。音響分析では、人間の音声をフーリエ解析等の信号処理により特徴量情報に変換する。認識デコーダでは、特徴量情報に基づいてテキストデータを出力する。具体的には音響情報と言語情報を総合的に判断して音声をテキストデータに変換する。この認識デコーダでの判断処理は、隠れマルコフモデルや動的時間伸縮法などの統計的手法により実現される。辞書は、認識対象の単語（ワード）を

50

データ化したものであり、音素列と単語を関連づけるものである。言語モデルは、辞書の単語についての確率をデータ化したものである。具体的には各単語の出現確率や接続確率をデータ化する。このような音響分析、認識デコーダ、音響モデル、辞書、言語モデルを用いた音声認識により、音センサにより取得された音情報からユーザの発話キーワードを適正に抽出できるようになる。

【0077】

タグ登録部306は、コンテンツのタグキーワードの登録処理（付与処理）を行う。例えばユーザの発話から抽出された発話キーワードに基づいて、コンテンツのタグキーワードの登録処理を行う。具体的には、発話キーワード自体や発話キーワードにより得られたキーワードを、コンテンツのタグキーワードとして、例えばコンテンツ情報記憶部322のコンテンツ情報に関連づけて登録する。

10

【0078】

ユーザ状態判断部307は、検知情報取得部304により取得された検知情報に基づいてユーザ状態を判断する。例えば、取得された検知情報に基づいて、表示部340に対するユーザの視認状態や、ユーザ（人物）と表示部340との位置関係等を判断する。或いはユーザの人数を検出する。そして、ユーザの視認状態や位置関係や人数を表すユーザ状態の情報は、ユーザ状態記憶部326に記憶される。

【0079】

ここで視認状態は、ユーザの視野範囲の状態や注視状態などであり、具体的には、ユーザの視野範囲（ビューボリューム）に表示部340が入っているか否かや、ユーザが表示部340を注視しているか否かなどである。このユーザの視認状態は視認状態判断部308が判断する。例えばユーザの視認状態として、ユーザが表示部340を注視している状態か否かを判断する。

20

【0080】

また位置関係は、ユーザと表示部340との距離や、表示部340に対するユーザの視線方向などであり、ユーザと表示部340との位置関係は、位置関係判断部309が判断する。例えば、ユーザと表示部340との位置関係として、ユーザと表示部340との間の距離（距離情報、距離パラメータ）を判断する。

【0081】

なおセンサ350として、ユーザを撮像する撮像センサが設けられたとする。この場合には、ユーザ状態判断部307（位置関係判断部）は、撮像センサからの撮像情報に基づいて、ユーザの顔領域（矩形の枠領域）を検出する。そして検出された顔領域のサイズに基づいて、ユーザと表示部340との間の距離を判断（推定）する。またユーザ状態判断部307は、検出された顔領域を内包し顔領域よりもサイズが大きな計測領域を設定する。即ち顔領域にオーバーラップする計測領域を設定する。そして計測領域内に顔領域が存在する時間を計測し、計測された時間に基づいて、ユーザが表示部340を注視しているか否かを判断する。例えば計測領域内への存在時間が所定時間以上であった場合に、ユーザが注視していたと判断する。

30

【0082】

或いは、ユーザ状態判断部307は、ユーザに対する自動焦点合わせ処理（オートフォーカス機能）により、ユーザと表示部340との間の距離を判断してもよい。例えばアクティブ方式を採用した場合には、画像表示システムに赤外線や超音波を射出するデバイスを設けると共に、センサ350として赤外線や超音波の受光センサを設ける。そしてユーザからの反射光を受光センサにより検知することで、ユーザとの距離等を検出すればよい。またパッシブ方式を採用した場合には、センサ350として撮像センサを設け、撮像画像に対して、位相差検出方式やコントラスト検出方式による画像処理を行うことで、ユーザとの距離等を検出すればよい。

40

【0083】

また、ユーザの人数は、例えばセンサ350の1つである焦電センサ（人感センサ）を用いて認識（特定）したり、撮像センサを用いて認識できる。例えば撮像センサからの画

50

像により、複数のユーザの顔領域が検出された場合には、検出された顔領域の個数に基づいてユーザの人数を検出できる。このユーザの人数の認識は人数検出部 310 により行われる。

【0084】

登録処理部 311 はユーザの登録処理を行う。例えばユーザ登録情報の設定処理を行う。具体的には、ユーザ登録画面等においてユーザが、操作部 360 等を用いてユーザ登録情報を入力した場合に、入力されたユーザ登録情報をユーザに関連づけてユーザ情報記憶部 327 に記憶する。ここでユーザ登録情報は、例えばユーザの ID、パスワードや、表示部 340 に表示される画像のカスタマイズ情報などを含むことができる。そして表示部 340 にはユーザ登録情報を反映させた画像が表示されることになる。

10

【0085】

なお、ユーザの登録処理の際に、ユーザの音声ユーザ登録情報として登録してもよい。例えばユーザの音声の特徴量情報をユーザ登録情報としてユーザ情報記憶部 327 に記憶して登録する。そして音声認識部 305 により話者認識を行う場合には、音センサにより取得された音情報と登録されたユーザの音声の特徴量情報とを比較して、発話しているユーザが登録ユーザであるか否かを判断する。また、センサ 350 として撮像センサを設け、この撮像センサによってユーザを撮像し、ユーザの顔画像の特徴量情報をユーザ登録情報として登録してもよい。この場合には、撮像センサにより取得された画像情報とユーザの顔画像の特徴量情報を比較して、撮像されたユーザが登録ユーザであるか否かを判断する人物認証を行う。

20

【0086】

ユーザ認識部 312 はユーザの認識処理を行う。例えば、撮像センサを用いた顔認識処理によりユーザの人物認証を行う。なお、目の網膜にある毛細血管の模様である網膜を用いたり、黒目の中の放射状の紋様である虹彩を用いて人物認証を行ってもよい。

【0087】

ヒストグラム作成部 313 は、タグのヒストグラムの作成処理（更新処理）を行う。例えば、コンテンツの登録タグキーワードと、登録タグキーワードの登録回数情報とが関連づけられたヒストグラム情報を作成する。ここで登録回数情報は、登録回数自体であってもよいし、登録回数と同等な情報（例えば登録頻度）であってもよい。そしてコンテンツ選択部 317 は、作成されたヒストグラム情報に基づいて、ユーザに提示するコンテンツの選択処理を行う。

30

【0088】

類似度評価部 314 は、ヒストグラム情報の類似度を評価する処理を行う。例えば複数のコンテンツ間のヒストグラム情報の類似度を評価する。類似度の評価処理としては、例えば後述するようなユークリッド距離、相関係数、Tanimoto係数を利用した評価処理がある。そしてコンテンツ選択部 317 は、第1のコンテンツの次にユーザに提示する第2のコンテンツとして、第1のコンテンツに対するヒストグラム情報の類似度が高いと評価されたコンテンツ（所定しきい値よりも高いと評価されたコンテンツ）を選択する。

【0089】

タグ管理部 315 はコンテンツの登録タグキーワードの整理処理（削除処理、更新処理）を行う。例えばヒストグラム作成部 313 により作成されたヒストグラム情報に基づいて、コンテンツに対して登録された登録タグキーワードの整理処理を行う。具体的には、コンテンツの登録タグキーワードの中から、ヒストグラム情報における登録回数情報（登録回数、登録頻度）が少ない登録タグキーワードを優先して削除する処理を行う。或いは、コンテンツの登録タグキーワードの中から、登録が古い登録タグキーワード（登録・更新・年月日・時間が古い登録タグキーワード）を優先して削除する処理を行う。

40

【0090】

質問提示部 316 は、ユーザへの質問の提示処理（質問画像や質問音声の出力処理）を行う。例えばタグキーワードを取得するための質問の提示処理を行う。具体的には質問提示部 316 は、コンテンツに関連づけられたタグキーワードに対応する質問を提示する処

50

理を行う。またコンテンツに関連づけられたタグキーワードの数が所定数以下である場合に、質問を提示する処理を行う。そしてタグ登録部306は、質問提示部316により提示された質問に対するユーザの反応（例えば発話キーワード、肯定・否定のキーワード、ユーザの動作等）に基づいて、コンテンツのタグキーワードの登録処理を行う。例えば質問に対してユーザが発話した場合に、その発話キーワードをタグキーワードとして登録する。

【0091】

また、例えばタグキーワードが階層構造を有しており、ユーザの発話から抽出された発話キーワードが、第1～第N（Nは2以上の整数）の階層のタグキーワードのうちの第K（ $1 < K < N$ ）の階層のタグキーワードであったとする。この場合には、質問提示部316は、第Kの階層のタグキーワード又は第Kの階層のタグキーワードよりも下位の第K+1の階層のタグキーワード（階層構造において下位のタグキーワード）を用いて質問を作成して提示してもよい。

10

【0092】

コンテンツ選択部317は、ユーザに提示するコンテンツの選択処理を行う。例えば、音声認識結果やヒストグラム情報の類似度の評価結果に基づいて、コンテンツ情報記憶部322から対応するコンテンツの情報を読み出して、ユーザに提示するコンテンツを選択する。或いは、通信部338を介して、ホームサーバ等の外部サーバからコンテンツの情報を受信することで、コンテンツを選択する。

20

【0093】

表示制御部318は、表示部340の表示制御を行う。例えば、コンテンツ選択部317が、コンテンツ情報記憶部322からコンテンツ情報を読み出したたり、通信部338によりコンテンツ情報を受信することで、コンテンツを選択すると、表示制御部318は、選択されたコンテンツの画像を表示部340に表示するための制御を行う。

【0094】

そして本実施形態では音声認識部305は、例えばコンテンツ（コンテンツ画像）の表示中におけるユーザの発話の音声認識を行い、ユーザの発話キーワードを抽出する。即ち音声認識によりユーザの会話中から抽出可能なキーワードである発話キーワードを抽出する。そしてタグ登録部306は、抽出された発話キーワードに基づいて、コンテンツのタグキーワードの登録処理を行う。例えば抽出された発話キーワード自体を、コンテンツのタグキーワードとしてコンテンツに関連づけて登録する。或いは、抽出された発話キーワードの中から、コンテンツのタグキーワードに一致する一致発話キーワードを検出する。そして、検出された一致発話キーワード以外の発話キーワードを、コンテンツのタグキーワードとして登録する。或いは、抽出された発話キーワードの中から、コンテンツのタグキーワードに関連する関連発話キーワードを検出する。そして検出された関連発話キーワードを、コンテンツのタグキーワードとして登録する。

30

【0095】

また、視認状態判断部308は、ユーザが表示部340を注視している状態か否かを判断し、タグ登録部306は、ユーザが注視状態であるときに抽出された発話キーワードに基づいて、コンテンツのタグキーワードの登録処理を行う。例えばユーザがコンテンツ（表示部）を注視していないときの発話キーワードについては、コンテンツのタグキーワードとして登録せずに、ユーザがコンテンツを注視しているときに抽出された発話キーワードを、コンテンツのタグキーワードとして登録する。

40

【0096】

また位置関係判断部309は、ユーザと表示部340との位置関係を判断し、タグ登録部306は、ユーザと表示部340との距離が所定距離以内であるときに抽出された発話キーワードに基づいて、コンテンツのタグキーワードの登録処理を行う。例えばユーザとの距離が所定距離よりも遠い場合の発話キーワードについては、コンテンツのタグキーワードとして登録せずに、ユーザとの距離が所定距離よりも近い場合に抽出された発話キーワードを、コンテンツのタグキーワードとして登録する。なお、ユーザと表示部340と

50

の距離は、距離そのもののみならず、距離と等価なパラメータであってもよい。

【0097】

また人数検出部310は、ユーザの人数の検出処理を行い、タグ登録部306は、ユーザの人数の検出結果に基づいて、コンテンツのタグキーワードの登録処理を行う。例えば音声認識部305は、ユーザが一人であると検出された場合に、その一人のユーザの発話キーワードを抽出する。そしてタグ登録部306は、抽出された一人のユーザの発話キーワードに基づいて、コンテンツのタグキーワードの登録処理を行う。例えばその発話キーワード自体や、発話中におけるその発話キーワード以外の発話キーワードや、関連発話キーワードを、コンテンツのタグキーワードとして登録する。

【0098】

一方、ユーザが複数人であると検出され、複数人のユーザが同一の発話キーワードを発話したことが検出されたとする。この場合にはタグ登録部306は、検出された同一発話キーワードに基づいて、コンテンツのタグキーワードの登録処理を行う。例えば第1のユーザの発話から第1の発話キーワードが抽出され、第2のユーザの発話からも第1の発話キーワードが抽出されたとする。この場合には、例えばその第1の発話キーワード自体や、発話中における第1の発話キーワード以外の発話キーワードや、関連発話キーワードを、コンテンツのタグキーワードとして登録する。

【0099】

またユーザ認識部312がユーザを認識（人物認証）したとする。そして、認識されたユーザの発話キーワードに基づいて、コンテンツのタグキーワードが取得された場合には、タグ登録部306は、取得されたタグキーワードを、認識されたユーザ及びコンテンツに関連づけて登録する。即ち、そのタグキーワードを、そのユーザ用のコンテンツのタグキーワードとして登録する。

【0100】

またタグ登録部306は、ユーザの発話キーワードの出現回数情報の集計処理を行う。そして抽出された発話キーワードのうち出現回数情報が所定しきい値を超えた発話キーワードを、コンテンツのタグキーワードとして登録する。ここで出現回数情報は、出現回数自体であってもよいし、出現回数と等価な情報（例えば出現頻度）であってもよい。

【0101】

また音声認識部305は、ユーザの無発話状態を検知する。例えば、取得された音情報の中に、人間の音声に特有の周波数成分や振幅を有する音情報が抽出されなかった場合に、無発話状態であると判断する。そして表示制御部318は、ユーザの無発話状態が検知された場合に、表示されるコンテンツの切り替え処理を行う。例えば無発話状態が検知されると、コンテンツのスライド表示において次のコンテンツ（スライド）への切り替え処理を行う。具体的には、ユーザが発話をしていると判断される場合には、現在のコンテンツ（現在のスライド）の表示を続け、無発話期間が所定時間を超えると、次のコンテンツ（次のスライド）に切り替える。

【0102】

以下、本実施形態の種々の手法について説明する。なお、以下では画像表示システムがデジタルフォトフレームである場合を主に例に取り説明するが、本実施形態の手法はデジタルフォトフレーム以外の種々の画像表示システムに適用できる。

【0103】

2. 発話に基づく自動タグ付け

本実施形態ではコンテンツが表示されているときの人の状況をセンシングしてコンテンツにタグ（タグキーワード）を付与する。具体的にはユーザの発話内容から自動的にタグを付与する。例えば発話（会話）のキーワード自体を、コンテンツ（表示画像、写真）のタグとして登録（付与）したり、その発話キーワードにより取得されるキーワード（単語）をタグとして登録する。

【0104】

なおコンテンツのタグとしては、コンテンツのジャンルに関するタグ（天気、スポーツ

10

20

30

40

50

、グルメ、旅行、文化、経済、政治、健康、自然、動物)、コンテンツの内容に関するタグ(猫、りんご)、ユーザが付加したタグ(白い、かわいい)などを想定できる。また個人モードでユーザが付加した個人タグ、共有モードで付加した共有タグ、他の多くのユーザが付加したソーシャルタグなどでもよい。また注視回数等から算出された注目度を表すタグ(何パーセント等)、お勧め対象を表すタグ(女性、男性、年齢、地域、職業)、お勧め環境を表すタグ(1人で、みんなで、遠くで、近くで)、感情タグ(喜怒哀楽)なども想定できる。

【0105】

次に発話に基づくタグの付与処理の具体例について図3、図4を用いて説明する。図3では、現在表示されているコンテンツ(画像、写真)のタグキーワードと同じタグキーワードがユーザの発話から検出された場合に、その発話中の他のキーワードを新たなタグとして登録(付与)する。これにより、表示コンテンツを見ながら話しているユーザの発話のみからキーワードを抽出できるため、自動タグ付けの精度を向上できる。

10

【0106】

例えば図3では、表示されているコンテンツに対して「猫」のタグキーワードが既に登録されている。またユーザの発話を音声認識することで、ユーザの発話キーワードとして「白い」、「猫」、「かわいい」という発話キーワードが抽出されている。この場合に図3では、抽出された「白い」、「猫」、「かわいい」の発話キーワードの中に、コンテンツのタグキーワードである「猫」に一致する発話キーワードが存在するか否かを検出する。そして「猫」の発話キーワードが検出されると、「猫」の発話キーワード以外の発話キーワードである「白い」、「かわいい」を、コンテンツのタグキーワードとして登録する。こうすることで、「猫」だけであったコンテンツのタグキーワードを、「猫」、「白い」、「かわいい」に増やすことが可能になる。

20

【0107】

即ちユーザが発話した「猫」のキーワードは既にタグキーワードとして登録されているため、これを重複して登録することは無駄であるため省く。一方、「猫」のタグキーワードが付与されたコンテンツを見ることで、ユーザが「この白い猫、かわいいね～」と発話した場合には、その発話内容のうちの「猫」以外の「白い」、「かわいい」のキーワードは、そのコンテンツのタグとして適切なものであると類推できる。そこで、既に登録されている「猫」に加えて、「白い」、「かわいい」についてもそのコンテンツのタグキーワードとして登録する。

30

【0108】

図4では、現在表示されているコンテンツのタグキーワードに関連する発話キーワードがユーザの発話から検出された場合に、検出されたキーワードを新たなタグとして登録(付与)する。これにより、コンテンツを見たユーザの発話を利用して、関連するタグを新たに付与できるようになる。

【0109】

例えば図4では、表示されているコンテンツに対して「果物」のタグキーワードが既に登録されている。またユーザの発話を音声認識することで、ユーザの発話キーワードとして「りんご」、「美味しそう」という発話キーワードが抽出されている。この場合に図4では、抽出された「りんご」、「美味しそう」の発話キーワードが、コンテンツのタグキーワードである「果物」に関連するキーワードであることが検出されている。例えば、各キーワードの関連キーワードが登録される関連キーワードデータベースを用意する。そして関連キーワードデータベースを参照して、「りんご」、「美味しそう」の発話キーワードが、「果物」の関連キーワードであるか否かを判断する。そして「りんご」、「美味しそう」が、[果物]の関連キーワードであると判断された場合には、既に登録されている「果物」に加えて、「りんご」、「美味しそう」をコンテンツのタグキーワードとして登録する。

40

【0110】

即ち、「果物」のタグキーワードが付与されたコンテンツを見ることで、ユーザが「こ

50

の「りんご」美味しそうだね～」と発話した場合には、その発話内容を構成する「りんご」、「美味しそう」のキーワードは、そのコンテンツのタグとして適切で具体的なものであると類推できる。そこで、既に登録されている「果物」に加えて、「りんご」、「美味しそう」についてもそのコンテンツのタグキーワードとして登録する。このようにすれば、抽象的な「果物」というタグしか付与されていなかったコンテンツに対して、ユーザの発話を利用して、より具体的な「りんご」等のタグを付与できるようになる。

【0111】

なお、ユーザの発話から発話キーワードを抽出する処理は例えば以下のようにして実現できる。即ち、まず音センサにより周囲音を検知して音情報を取得する。次に、取得された音情報を音声認識によりテキスト情報に変換し、得られたテキスト情報の中から単語を抽出する。そして、抽出された単語と、登録キーワード記憶部325（キーワードデータベース）に記憶された登録キーワードとの照合処理を行い、抽出された単語が登録キーワードであるか否かを判定する。そして抽出された単語が登録キーワードである場合には、その単語を発話キーワードであると判定する。これにより、意味のない情報や不快な情報がユーザに提示されてしまう事態を防止できる。

【0112】

図5に、ユーザの発話キーワードからタグキーワードを登録する処理のフローチャートを示す。

【0113】

まずコンテンツを表示する（ステップS1）。例えばランダムに又はユーザ情報に基づいてコンテンツを選択して、表示部340にコンテンツの画像を表示する。次に、ユーザの発話が検知されたか否かを判断し、検知された場合には、ユーザの発話の音声認識処理を実行する（ステップS2、S3）。そして、ユーザの発話から発話キーワードを抽出し、抽出された発話キーワードと、表示コンテンツのタグキーワードの照合処理を行う（ステップS4、S5）。図3を例に取れば「白い」、「猫」、「かわいい」の発話キーワードと、「猫」のタグキーワードの照合処理が行われる。図4を例に取れば「りんご」、「美味しそう」の発話キーワードと、「果物」のタグキーワードの照合処理が行われる。

【0114】

次に、コンテンツのタグキーワードに一致する発話キーワードが検出されたか否かを判断する（ステップS6）。図3を例に取れば、「猫」が一致発話キーワードとして検出される。そして、一致発話キーワードが検出された場合には、検出された一致発話キーワード以外の発話キーワードを、表示コンテンツのタグキーワードとして登録する（ステップS7）。図3を例に取れば、「猫」以外の「白い」、「かわいい」がタグキーワードとして登録される。

【0115】

一方、ステップS6で一致発話キーワードが検出されなかった場合には、コンテンツのタグキーワードに関連する発話キーワードが検出された否かを判断する（ステップS8）。図4を例に取れば、「りんご」、「美味しい」が「果物」の関連発話キーワードとして検出される。そして関連発話キーワードが検出された場合には、検出された関連発話キーワードを、表示コンテンツのタグキーワードとして登録する（ステップS9）。図4を例にとれば、「りんご」、「美味しい」がタグキーワードとして登録される。ステップS8でコンテンツのタグキーワードに関連する発話キーワードが検出されなかった場合は、新規なタグキーワードは登録されずに終了となる。

【0116】

以上の本実施形態の手法によれば、ユーザの発話キーワードがコンテンツのタグキーワードとして自動的に登録されるため、ユーザの手間をかけずに効率的にコンテンツに対してタグを付与することが可能になる。即ち、ユーザがキーボード等の操作部を操作しなくても、表示コンテンツを見て発話するだけで、タグが自動的に付与されるため、ユーザの利便性を向上できる。また、表示コンテンツを見ているユーザの発話からタグを取得するようにすれば、表示コンテンツとは無関係な会話をしているユーザの発話からタグが取得

10

20

30

40

50

されてしまう事態を防止できるため、自動タグ付けの精度を向上できる。

【 0 1 1 7 】

3. 注視しているときの発話からの自動タグ付け

本実施形態では、例えばユーザがコンテンツを注視しているときの発話だけを用いて、音声認識を行い、その発話に含まれるキーワードをタグとしてコンテンツに付与して、自動タグ付けを行ってもよい。即ち、ユーザが表示コンテンツを注視して発話している場合には、その発話内容はコンテンツに関するものである可能性が高い。従って、ユーザの注視時の発話からキーワードを抽出すれば、表示コンテンツの発話のみを精度良く抽出することができ、自動タグ付けの精度を向上できる。

【 0 1 1 8 】

例えば図6(A)では、ユーザは、カニ等の北海道の旬の食材が表示されているコンテンツを注視しており、このユーザの注視状態がデジタルフォトフレーム300(視認状態判断部)により検出される。そして、ユーザが表示コンテンツを見て、「この「カニ」美味しそうだな」と発話すると、この発話がセンサ350により検出されて、音声認識が行われ、「カニ」という発話キーワードが抽出される。そして、この「カニ」の発話キーワードや、「カニ」の発話キーワードから図3、図4等の手法により取得されたキーワードが、表示コンテンツのタグキーワードとして登録される。このようにすることで、注視時の発話のみからキーワードを抽出して、タグキーワードを登録できる。

【 0 1 1 9 】

ここで、ユーザの注視状態の検知は、センサ350である撮像センサを用いて実現できる。即ち図6(B)に示すように、撮像センサからの撮像情報に基づいて、ユーザの顔領域FARを検出する。次に検出された矩形の顔領域FARに対応する計測領域SARを設定する。この計測領域SARは、顔領域FARを内包し、顔領域FARよりもサイズが大きな領域である。この計測領域SARは、例えば顔領域FARをオーバーサイジングすることで設定できる。そして、この計測領域SAR内に顔領域FARが存在する時間を計測し、計測された時間に基づいて、ユーザが表示部340を注視しているか否かを判断する。例えば顔領域FARが計測領域SAR内に一定時間以上位置していた場合には、ユーザが表示部340を注視していると判断する。

【 0 1 2 0 】

図7に、ユーザの注視を条件に発話キーワードからタグキーワードを登録する処理のフローチャートを示す。

【 0 1 2 1 】

まず、コンテンツを表示し、ユーザがコンテンツ(表示部)を注視しているか否かを判断する(ステップS11、S12)。そしてユーザが注視していると判断された場合には、ユーザの発話が検知されたか否かを判断する(ステップS13)。そして発話が検知された場合には、ユーザの発話の音声認識処理を行って、発話キーワードを抽出し、抽出された発話キーワードに基づき、表示コンテンツのタグキーワードの登録処理を行う(ステップS14、S15、S16)。例えば発話キーワード自体や発話キーワードに基づき取得されるキーワード(図3、図4参照)をタグキーワードとして登録する。なおステップS12においてユーザの注視状態を判断するか否かについては、後述するユーザの登録情報において、注視モードのオン・オフをユーザが入力することで設定できる。

【 0 1 2 2 】

図8は、ユーザの注視を条件に発話キーワードからタグキーワードを登録する他の処理例のフローチャートである。

【 0 1 2 3 】

図8のステップS21~S25は図7のステップS11~S15と同様である。そしてステップS25でユーザの発話キーワードが抽出されると、抽出された発話キーワードの中に指示代名詞が検出されたか否かを判断する(ステップS26)。そして、指示代名詞が検出された場合には、検出された指示代名詞以外の発話キーワードを、表示コンテンツのタグキーワードとして登録する(ステップS27)。

10

20

30

40

50

【0124】

即ち図8では、ユーザが表示コンテンツを注視しているときに使用される指示代名詞（こそあど言葉）を、ユーザの発話から検出する。そしてユーザの発話から指示代名詞が検出された場合には、その発話中の他のキーワードを、新たなタグキーワードとして登録する。

【0125】

例えば表示コンテンツとしてラーメンの写真が表示されているときに、ユーザが「これ、美味しそう」と発話したとする。この場合には「これ」という指示代名詞が検出されたため、他のキーワードである「美味しそう」というキーワードを、ラーメンの写真のタグキーワードとして登録する。

10

【0126】

また表示コンテンツとして寺の写真が表示されているときに、ユーザが「ここ、京都の寺だね」と発話したとする。この場合には「ここ」という指示代名詞が検出されたため、他のキーワードである「京都の寺」というキーワードを、寺の写真のタグキーワードとして登録する。このようにすれば、ユーザが発話した指示代名詞を手がかりとして、そのコンテンツに適切なタグを付与できるようになる。

【0127】

次に、注視状態の検出処理について図9のフローチャートを用いて説明する。まず、撮像センサ（カメラ）を用いた顔検出により、顔領域（枠領域）を検出する（ステップS31）。次に、検出された顔領域を内包し、顔領域よりもサイズが大きな計測領域を設定する（ステップS32）。即ち図6（B）に示すように、顔領域をオーバーサイジングした計測領域を設定する。そして計測領域内に顔領域が存在する時間をタイマを用いて計測する（ステップS33）。即ち計測領域の設定後、タイマの計測を開始し、顔領域が計測領域内に位置する時間を計測する。そして所定時間以上、経過したか否かを判断し、経過した場合には注視状態であると判断する（ステップS34、S35）。

20

【0128】

なお注視状態の検出手法は図9の手法に限定されない。例えばユーザの赤目を検出することで注視状態を検出してよい。或いは、2台のカメラ（ステレオカメラ）で撮影されたユーザの顔画像の目周辺の画像領域の明暗から、瞳孔の位置を検出し、検出された瞳孔の中心位置と眼球の中心位置から、ユーザの視線方向を検出し、ユーザが注視状態か否かを判断してもよい。

30

【0129】

4. 表示部からの距離が近いときの発話からの自動タグ付け

本実施形態では、表示部340（表示画面）とユーザとの距離が近いときの発話だけを用いて、音声認識を行い、その発話に含まれるキーワードをタグとしてコンテンツに付与して、自動タグ付けを行ってもよい。即ち、ユーザが表示部340に接近して発話している場合には、その発話内容はコンテンツに関するものである可能性が高い。従って、ユーザの接近時の発話からキーワードを抽出すれば、通常時の発話と、表示コンテンツに関する発話とを精度良く切り分けることができ、より効果的で精度の高い自動タグ付けを実現できる。

40

【0130】

図10に、ユーザの接近を条件に発話キーワードからタグキーワードを登録する処理のフローチャートを示す。

【0131】

まずコンテンツを表示し、コンテンツ表示中におけるユーザと表示部340との距離を検出する（ステップS41、S42）。そしてユーザとの距離が所定距離以内か否かを判断する（ステップS43）。そして所定距離以内である場合には、ユーザの発話の音声認識処理を行い、ユーザの発話から発話キーワードを抽出する（ステップS44、S45）。そして抽出された発話キーワードに基づき、表示コンテンツのタグキーワードの登録処理を行う（ステップS46）。例えば発話キーワード自体や発話キーワードに基づき取得

50

されるキーワードをタグキーワードとして登録する。

【0132】

ここでユーザとの位置関係の検出手法としては種々の手法が考えられる。例えば図11(A)では、センサ350として、CCD、CMOSセンサなどの撮像センサ(カメラ)を用いる。そして撮像センサからの撮像情報に基づいて、矩形の枠領域であるユーザの顔領域FARを検出する。また顔領域FARに映ったユーザの画像に対する画像認識処理を行い、ユーザの顔画像の特徴点データを抽出する。この特徴点データは、例えば顔画像の認識結果として、ユーザ登録情報と関連づけて登録される。

【0133】

ユーザと表示部340との間の位置関係を検出する場合には、撮像センサからの撮像情報に基づいて、顔領域FARのサイズを求める。そして求められたサイズに基づいて、ユーザと表示部340との間の距離を判断する。

【0134】

例えば図11(B)では、顔領域FARのサイズが小さいため(所定サイズ以下であるため)、ユーザとの距離は遠いと判断される。この場合には、ユーザの発話キーワードの抽出処理は行わないようにする。

【0135】

一方、図11(C)では、顔領域FARのサイズが大きいため(所定サイズよりも大きい)、ユーザとの距離は近いと判断される。そして、このようにユーザとの距離が近くなって、ユーザがデジタルフォトフレーム300に近づいたと判断された場合には、ユーザの発話キーワードの抽出処理を行い、抽出された発話キーワードに基づいてタグキーワードの登録処理を行う。

【0136】

ここで顔領域の検出手法としては種々の手法が考えられる。例えば、顔検出を行うためには、撮像センサで撮影された撮像画像において、顔がある場所と他の物体とを区別して、顔領域を切り出す必要がある。顔は、目、鼻、口等から構成され、これらの形状・位置関係は個人差はあるものの、ほぼ共通した特徴を有する。そこで、この共通な特徴を用いて、顔を他の物体から識別して画面の中から切り出す。このための手がかりとしては、肌の色、顔の動き、形、大きさ等がある。肌の色を用いる場合には、RGBデータを色相・輝度・彩度からなるHSVデータに変換し、人の肌の色相を抽出する手法を採用する。

【0137】

或いは、多数の人の顔パターンから生成した平均顔パターンを顔テンプレートとして作成してもよい。そして、この顔テンプレートを撮像画像の画面上で走査して、撮像画像との相関を求め、最も相関値が高い領域を顔領域として検出する。

【0138】

なお検出精度を高めるため、複数の顔テンプレートを辞書データとして用意し、これらの複数の顔テンプレートを用いて顔領域を検出してもよい。或いは目、鼻、口の特徴や、これらの位置関係や、顔の中のコントラストなどの情報も考慮して、顔領域を検出してもよい。或いは、ニューラルネットワークモデルを用いた統計的なパターン認識により顔領域を検出することも可能である。

【0139】

図11(A)~図11(C)の検出手法によれば、顔領域FARのサイズによりユーザと表示部340の距離を検出できるのみならず、ユーザが表示部340を見ているか否かも同時に検出できるという利点がある。即ち、ユーザの視線が表示部340の方に向いていなかった場合には、顔テンプレートとの相関値が低くなるため、顔領域FARは非検出になる。従って、顔領域FARが検出されたということは、ユーザの視線が表示部340の方に向いており、ユーザの視野範囲内に表示部340が入っていることと等価になる。そして、この状態で、顔領域FARのサイズを検出し、ユーザの発話キーワードを抽出して、コンテンツのタグとして付与すれば、表示部340を見ているユーザの発話内容に応じたタグを、コンテンツに付与できるようになる。

10

20

30

40

50

【 0 1 4 0 】

5 . ユーザの人数の検出結果に応じた自動タグ付け

本実施形態では、ユーザの人数を検出し、検出結果に応じた自動タグ付けを行ってもよい。例えば検出されたユーザの人数が1人である場合に、個人モードに設定し、そのユーザの発話から自動タグ付けを行う。即ち、ユーザが1人でコンテンツを見ている時にだけ、発話からタグを抽出して、コンテンツのタグとして付与する。これにより、複数人による通常の会話と、1人のユーザによる表示コンテンツに関する発話(コメント)とを精度良く切り分けることが可能になり、より効果的で精度の高い自動タグ付けを実現できる。或いは、複数人のユーザが検出された場合には、複数人のユーザが同じキーワード(単語)を発話した場合に、そのキーワードをコンテンツのタグとして登録したり、登録処理におけるそのキーワードの重み付けを高くするようにしてもよい。

10

【 0 1 4 1 】

図12に、ユーザの人数を検出して発話キーワードからタグキーワードを登録する処理のフローチャートを示す。

【 0 1 4 2 】

まず、コンテンツを表示し、ユーザの人数を検出し、検出されたユーザの人数が1人であるか否かを判断する(ステップS51、S52、S53)。そして、ユーザの人数が1人である場合には、デジタルフォトフレーム300の動作モードを個人モードに設定する(ステップS54)。次に、ユーザが発話しているか否かを検知し、ユーザの発話が検知された場合には、ユーザの発話の音声認識処理を行って、発話キーワードを抽出する(ステップS55、S56、S57)。そして抽出された発話キーワードに基づき、表示コンテンツのタグキーワードの登録処理を行う(ステップS58)。

20

【 0 1 4 3 】

このようにすれば、ユーザが1人でデジタルフォトフレーム300を見ているときにだけ、そのユーザの発話キーワードを検出して、表示コンテンツのタグキーワードとして付与できる。

【 0 1 4 4 】

図13に、ユーザの人数を検出して発話キーワードからタグキーワードを登録する他の処理例のフローチャートを示す。

【 0 1 4 5 】

まず、コンテンツを表示し、ユーザの人数を検出し、検出されたユーザの人数が1人であるか否かを判断する(ステップS61、S62、S63)。そして、ユーザの人数が1人である場合には、個人モードに設定して音声認識を行い、音声認識により得られた発話キーワードを表示コンテンツのタグキーワードとして登録する(ステップS64、S65、S66)。一方、ユーザの人数が複数人である場合には、グループモード(複数人モード)に設定する(ステップS67)。そして、複数人のユーザの発話の音声認識処理を行い、複数人のユーザが同一の発話キーワードを発話したか否かを判断する(ステップS68、S69)。例えば第1のユーザが発話した1又は複数のキーワードと、第2のユーザが発話した1又は複数のキーワードの中に同一(共通)のキーワードがあるか否かを検出する。そして、検出された同一の発話キーワードを、表示コンテンツのタグキーワードとして登録する(ステップS70)。この場合に、例えば、検出された同一の発話キーワードについては、タグ登録の際の重み付け(例えば出現回数や登録回数に乘算する重み付け係数)を高くするようにしてもよい。

30

40

【 0 1 4 6 】

例えば複数人のユーザがグループでコンテンツを鑑賞している場合には、それらのユーザが共通に発したキーワードは、そのコンテンツに関連するキーワードである可能性が高い。従って、このキーワードを、そのコンテンツのタグキーワードとして登録することで、自動タグ付けの精度を向上できる。

【 0 1 4 7 】

なお図12、図13において、図7、図8、図10に示すような注視検出や距離検出を

50

行ってもよい。また図12のステップS54や図13のステップS64のように個人モードに設定された場合に、後述するユーザ登録情報にしたがった個人設定でコンテンツを表示するようにしてもよい。また図13のステップS67でグループモードに設定された場合に、グループ（例えば家族・友人グループ）についての登録情報にしたがったグループ設定でコンテンツを表示するようにしてもよい。またグループモードの場合には、各ユーザの音声情報（特徴量情報）をユーザ登録情報として予め登録しておき、この登録音声情報に基づいて、どのユーザが発話しているのかを判別してもよい。

【0148】

またユーザの人数の検出は例えば以下のような手法により実現できる。例えば焦電センサ等のユーザ検知センサで、デジタルフォトフレーム300の周囲にユーザが存在するかどうかを検出する。そしてユーザの存在が検出された場合には、図11(A)～図11(C)で説明したように撮像センサを用いてユーザの顔領域を検出する。そして検出された顔領域の個数等に基づいてユーザの人数を検出する。この際に、ユーザの顔画像（特徴量情報）がユーザ登録情報として予め登録されている場合には、この登録された顔画像に基づいて、各ユーザの人物認証を行い、個人モードやグループモードでのコンテンツの表示制御を行ってもよい。

10

【0149】

図14(A)にユーザ登録情報の入力画面の一例を示す。ユーザは、図14(A)のようにデジタルフォトフレーム300の表示部340に表示されたユーザ登録画面を見ながら、操作部等を用いてユーザ登録情報を入力する。これによりユーザ登録処理が実現される。なお、ユーザ登録情報をPC（パーソナルコンピュータ）等を用いて設定し、設定されたユーザ登録情報を、直接に或いはホームサーバ等を介してデジタルフォトフレーム300に転送するようにしてもよい。

20

【0150】

図14(A)では、ユーザ登録情報として、ユーザのID、パスワード、ユーザの趣味、お気に入りの情報、距離しきい値、注視制御のオン・オフ、表示時間（表示時間間隔、表示速度）、スライド表示方法（広義には表示方法）等が入力される。例えば図14(A)では、ユーザの趣味が釣りであり、お気に入り情報が株価の情報であり、距離しきい値が中ぐらいであり、注視制御がオンであり、表示時間が中ぐらいであり、スライド表示方法がフェードであることが入力されている。

30

【0151】

なお、距離しきい値は、図10のようにユーザが接近したときの発話からタグを付与する場合の距離のしきい値を決めるパラメータである。例えば距離しきい値が「短い」に設定されると、図10のステップS43の「所定距離」が短い距離に設定され、距離しきい値が「長い」に設定されると、「所定距離」が長い距離に設定される。また、注視制御は、図7、図8のようにユーザの注視を条件としたタグ付与のオン・オフを制御するものである。例えば、注視制御が「オフ」に設定されると、図7、図8のステップS12、S22の注視判定がスキップされ、「オン」に設定されると、ステップS12、S22の注視判定が行われるようになる。

40

【0152】

図14(B)はユーザ登録処理のフローチャートである。まず図14(A)に示すようなユーザ登録画面を表示する（ステップS71）。これによりユーザはユーザ登録情報の入力が可能になる。そしてユーザ登録情報の入力完了したか否かを判断し（ステップS72）、完了した場合には、ユーザ検知センサである撮像センサ（カメラ）を用いてユーザの顔領域の検出を行う（ステップS73）。そして撮像センサを用いてユーザの顔画像認識を行い（ステップS74）、認識された顔画像をユーザ登録情報に関連づけて記憶する（ステップS75）。

【0153】

このように、ユーザの顔画像をユーザ登録情報として登録すれば、ユーザ認識部312は、デジタルフォトフレーム300を見ているユーザを、撮像センサを用いて認識できる

50

ようになる。そしてタグ登録部306は、認識されたユーザの発話キーワードに基づいて、コンテンツのタグキーワードが取得されると、取得されたタグキーワードを、認識されたユーザ及びコンテンツに関連づけて登録する。例えば、コンテンツの表示中に第1のユーザが認識され、その第1のユーザの発話から抽出された発話キーワードに基づきタグキーワードが取得されたとする。すると、そのタグキーワードは、第1のユーザ及び表示コンテンツに関連づけて、コンテンツ情報記憶部322に登録される。またコンテンツの表示中に第2のユーザが認識され、その第2のユーザの発話から抽出された発話キーワードに基づきタグキーワードが取得されたとする。すると、そのタグキーワードは、第2のユーザ及び表示コンテンツに関連づけて、コンテンツ情報記憶部322に登録される。

【0154】

こうすることで、各ユーザの発話に対応したタグがコンテンツに付与されるようになる。従って、付与されたタグに基づいてコンテンツの表示制御を行った場合に、ユーザ毎に異なるコンテンツが選択されて表示されるようになるため、コンテンツ表示のバラエティ度を増すことができる。

【0155】

6. タグ登録処理及びコンテンツ切り替え処理の詳細例

次に、タグの登録処理やコンテンツの切り替え処理（表示制御）の詳細例について説明する。

【0156】

この詳細例では、コンテンツ表示中の発話キーワードの出現回数を記録し、出現回数にしきい値を超えた場合に、その発話キーワードをタグとして付与する。即ち、発話から単純にタグを付与すると、表示コンテンツとは無関係なタグが付与されてしまうので、出現回数を記録して、無関係なタグが付与されないようにする。また発話（会話）の途切れを検知して、そのタイミングで表示コンテンツを切り替える。

【0157】

また表示コンテンツの登録タグキーワードの登録回数についてのヒストグラム情報を作成し、前回に表示したコンテンツのヒストグラム情報に基づいて次に表示するコンテンツを決定する。タグキーワード、登録回数、出現回数等はユーザ毎に作成する。この詳細例によれば、コンテンツにタグキーワードが予め登録されていない状態から開始して、発話に応じたタグキーワードをコンテンツに付与できる。そして付与されたタグキーワードに基づいてコンテンツの表示を制御する。これにより多様なコンテンツ表示が可能になる。

【0158】

図15に、タグ登録処理及びコンテンツ切り替え処理の詳細例のフローチャートを示す。まず撮像センサ（カメラ）を用いてユーザを認識する（ステップS81）。例えば図14（A）、（B）で登録された顔画像と、撮像センサで撮像されたユーザの顔画像を照合してユーザを認識する。そして、 $i=1$ に設定して、コンテンツ $CT_i = CT_1$ を表示する（ステップS82、S83）。

【0159】

次に、無発話期間（発話を検知されない期間）が所定時間を超えたか否かを判断する（ステップS84）。そして無発話期間が所定時間を超えていない場合には、ユーザの発話を音声認識して、発話キーワードを抽出する（ステップS85）。

【0160】

次に抽出された発話キーワードの出現回数の集計処理を行う（ステップS86）。例えば抽出された発話キーワードの出現回数をカウントする。そして、抽出された発話キーワードの出現回数にしきい値を超えたか否かを判断し、超えた場合には、その発話キーワードを、コンテンツ CT_i のタグキーワードとして登録する（ステップS87、S88）。そしてコンテンツ CT_i の登録タグキーワードのヒストグラム情報を更新する（ステップS89）。即ち登録されたタグキーワードの登録回数を1だけインクリメントする。

【0161】

一方、ステップS84で、無発話期間が所定時間を超えたと判断した場合には、ユーザ

10

20

30

40

50

の無発話状態が検知されたと判断して、表示中のコンテンツ CT_i のヒストグラム情報を読み出す（ステップ S_{90} ）。即ちステップ S_{89} で更新されたヒストグラム情報をヒストグラム情報記憶部 328 から読み出す。そして $i = i + 1$ に設定し、読み出されたヒストグラム情報に基づいて、次に表示するコンテンツ $CT_i = CT_2$ を選択する（ステップ S_{91} 、 S_{92} ）。そして選択されたコンテンツ $CT_i = CT_2$ を表示する（ステップ S_{83} ）。

【0162】

図16(A)にコンテンツテーブル（コンテンツ情報）の例を示す。このコンテンツテーブルは、各コンテンツIDに対して、コンテンツURLや、登録タグキーワードが関連づけられている。

【0163】

図16(B)に登録回数テーブルの例を示す。この登録回数テーブルでは、各コンテンツの登録タグキーワードに対してその登録回数に関連づけられている。この登録回数テーブルは、ユーザIDとコンテンツIDの各組み合わせに対応して設けることができる。

【0164】

例えばユーザAとユーザBが以下のような会話をしたとする。

ユーザA：「この前のカニは美味しかったねー」

ユーザB：「そうそう、北海道でカニを食べたね」

ユーザA：「カニの身をとるのが難しくて」

ユーザB：「今度は飛行機じゃなくて、フェリーでのんびり北海道に行きたいね」

この会話から名詞のキーワードだけを抽出すると、カニ = 3回、北海道 = 2回、身 = 1回、飛行機 = 1回、フェリー = 1回になる。そして図15のステップ S_{87} の出現回数のしきい値が3回に設定されていた場合には、現在表示中のコンテンツに対して「カニ」というタグキーワードが登録されることになる。具体的には図16(A)のコンテンツテーブルにおいて、現在表示されているコンテンツのIDに対して、「カニ」のタグキーワードが関連づけて登録される。また登録回数テーブルに「カニ」のタグキーワードが追加されると共に、その登録回数が1だけインクリメントされる。

【0165】

また本実施形態では図15のステップ S_{89} 、 S_{90} に示すように、コンテンツのタグについてのヒストグラム情報を作成し、作成されたヒストグラム情報に基づいて、次に表示するコンテンツを選択する。

【0166】

例えば図17では、コンテンツである画像 I_1 が表示されているときに、ユーザが「この”うどん”美味しかったよねー」と発話している。すると「うどん」という発話キーワードが、画像 I_1 のタグとして登録される。そして、この画像 I_1 のタグのヒストグラム情報に基づいて、次に表示する画像 (I_2 、 I_3 等) を選択する。

【0167】

例えば図18に、画像 I_1 、 I_2 、 I_3 の各画像に付与されている登録タグキーワードとその登録回数の例を示す。例えば画像 I_1 に対しては「うどん」、「そば」、「熱い」、「めん」、「旅行」、「レストラン」のタグキーワードが登録されており、その登録回数は、各々、10、12、5、10、30、10回になっている。また画像 I_2 に対しては「うどん」、「そば」、「熱い」、「めん」、「レストラン」のタグキーワードが登録されており、その登録回数は、各々、5、10、5、5、8回になっている。また画像 I_3 に対しては「うどん」、「そば」、「コーヒー」、「熱い」、「めん」、「旅行」、「レストラン」、「カフェ」のタグキーワードが登録されており、その登録回数は、各々、1、1、10、5、2、10、1、10回になっている。

【0168】

本実施形態では、図18から例えば図19に示すようなヒストグラム情報を作成する。図19では横軸がタグキーワードになっており、縦軸が各タグキーワードの登録回数になっている。そして、作成されたヒストグラム情報に基づいて、現在表示されているコンテ

10

20

30

40

50

ンツに関連するコンテンツを選択し、選択されたコンテンツを次の表示コンテンツ（次のスライド）として表示する。

【0169】

この場合に、ヒストグラム情報に基づくコンテンツの選択手法としては種々の手法を想定でき、例えばヒストグラム情報のキーワードランキングによりコンテンツを選択する手法が考えられる。例えば、現在表示されているコンテンツのヒストグラム情報において、最も登録回数が大きいタグキーワードが「うどん」であったとする。或いは、今回のコンテンツ表示中に、「うどん」というタグが登録されたとする。この場合には、次に表示するコンテンツとして、現在表示されているコンテンツを除いて、「うどん」のタグキーワードが付与されたランキング上位のコンテンツを選択する。例えば「うどん」の登録回数が多い上位の所定数のコンテンツの中から選択されたコンテンツを表示する。このようにすることで、現在表示されているコンテンツに関連するコンテンツを次のスライドとして表示でき、これまでにないコンテンツ表示を実現できる。

10

【0170】

但し、キーワードランキングを利用する手法は、処理も単純で実装も容易であるが、システムの使用時間が長くなると、登録回数（登録頻度）が高いコンテンツしか表示されなくなり、表示されるコンテンツが単調になってしまうという問題がある。

【0171】

このような問題を解決するために、例えばヒストグラム情報の類似度を評価し、類似度の評価結果に基づいて、次に表示するコンテンツを決定する。例えば第1のコンテンツの次に第2のコンテンツを選択する場合に、第1のコンテンツのヒストグラム情報と第2のヒストグラム情報の類似度を評価し、類似度の評価結果に基づいてコンテンツを選択する。例えば、第1のコンテンツに対するヒストグラム情報の類似度が高いと評価されたコンテンツを、次に表示する第2のコンテンツとして選択する。

20

【0172】

このようにすれば、キーワードランキングに基づいてコンテンツを選択する手法に比べて、多様なコンテンツ表示を実現できる。

【0173】

ここで、ヒストグラムの類似度の評価処理では、ヒストグラムの各値をベクトルの変量とみなして、類似性を比較する。このような類似度の評価処理の例として、ユークリッド距離を利用した手法、相関係数（ピアソン相関係数）を利用した手法、Tanimoto係数を利用した手法などがある。

30

【0174】

ユークリッド距離を利用した手法では下式（1）に示すように、各キーワードを軸とするN次元のユークリッド空間での距離 d_{ij} を求める。この距離 d_{ij} は、任意の正の値を取り、値が小さいほど類似であると評価される。

【0175】

【数1】

$$d_{ij} = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 \cdots (x_{in1} - x_{jn1})^2} \quad \cdots \cdots (1)$$

40

相関係数（ピアソン相関係数）を利用し手法では下式（2）に示すように、線形類似性を計算する。r は -1 から 1 までの値を取り、0 に近ければ近いほど相関がないと評価される。また r が負の値である場合には、逆相関である評価され、正の値である場合には、正の相関であると評価され、r が 1 に近いほど類似であると評価される。

【0176】

【数 2】

$$r = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{(\sum X^2 - \frac{(\sum X)^2}{N}) + (\sum Y^2 - \frac{(\sum Y)^2}{N})}} \quad \dots\dots (2)$$

Tanimoto係数を利用した手法では下式(3)に示すように、比較するコンテンツに共通したキーワードの数と、全体のキーワードの数の比 $T(A, B)$ を計算する。 $T(A, B)$ は、0 から 1 までの値を取り、1 に近いほど類似であると評価される。

【0177】

10

【数 3】

$$T(A, B) = \frac{A \cdot B}{\|A\|^2 + \|B\|^2 - A \cdot B} \quad \dots\dots (3)$$

例えば図18、図19において、画像I1に対する画像I2のユークリッド距離を計算すると、 $d_{ij} = 30.95$ になり、画像I1に対する画像I3のユークリッド距離を計算すると、 $d_{ij} = 30.77$ になる。従って、画像I3の方が画像I2よりも画像I1に類似するという結果になり、画像I1の次に表示するコンテンツとして、ユークリッド距離が小さい画像I3が選択されることになる。

【0178】

20

また画像I1に対する画像I2の相関係数を計算すると、 $r = 0.029$ になり、画像I1に対する画像I3の相関係数を計算すると、 $r = 0.027$ になる。従って、画像I2の方が画像I3よりも画像I1に類似するという結果になり、画像I1の次に表示するコンテンツとして、相関係数が1に近い画像I2が選択されることになる。

【0179】

また画像I1に対する画像I2のTanimoto係数を計算すると、 $T(A, B) = 0.833$ になり、画像I1に対する画像I3の相関係数を計算すると、 $T(A, B) = 0.75$ になる。従って、画像I2の方が画像I3よりも画像I1に類似するという結果になり、画像I1の次に表示するコンテンツとして、Tanimoto係数が1に近い画像I2が選択されることになる。

30

【0180】

ユークリッド距離(ユークリッド空間)は、純粹にタグの登録回数が結果に影響する。このため、コンテンツ間の類似度に関係無く、タグの登録回数が多いコンテンツのヒストグラムと、タグの登録回数が少ないコンテンツのヒストグラムの間の距離が必ず大きくなってしまふ。このため、前述したキーワードランキングを用いる手法と類似した問題が生じる。例えば図20(A)において、P1のヒストグラムを有するコンテンツの次に表示するコンテンツとして、P1の付近のP2、P3、P4に示すヒストグラムを有するコンテンツが選択されてしまふ。従って、コンテンツ表示が単調になってしまうおそれがある。

【0181】

一方、相関係数を用いる手法では、例えば図20(A)のP5に示すヒストグラムを有するコンテンツについても選択されるようになる。従って、意外性のあるコンテンツが選択されるようになり、コンテンツ表示の多様性を高めることができる。

40

【0182】

またTanimoto係数を用いる手法は、図20(B)において、T集合AとBに共通の要数 $A \cdot B$ を求めて、類似度を評価する。このため、タグが登録された回数ではなくタグの有無しか考慮されない。従って、ノイズデータの影響タグ登録回数が1回だけのものが最終結果に寄与するため、精度が低くなってしまうという問題がある。即ち、「うどん」のコンテンツに対して、間違っ「寺」というタグが1回でも登録されると、「うどん」とは類似しない「寺」のコンテンツが選択されるようになってしまふ。但し、Tanimoto係数を用いる手法においても、高性能の音声認識を利用することでノイズの影響を少なくする

50

ことも可能である。またタグの数が膨大になった場合の処理スピードは、タグの有無 (True/False) を判断する演算だけで済み、単純であるため、この点において他の手法に比べて有利である。

【0183】

このように、精度の観点からは相関係数を用いる手法が有効であるが、評価対象、パフォーマンス、精度の兼ね合いを考慮して、キーワードランキング、ユークリッド距離、相関係数、Tanimoto係数を用いる評価関数の中から、適切な評価関数を選択することが望ましい。具体的には、状況に応じて使用する評価関数を選択する。例えばタグの数次元数が多く、処理速度が求められる場合には、キーワードランキングやTanimoto係数を用いればよい。また高い精度を追求したい場合には、相関係数を用いればよい。また、ノイズデータが少なく、タグの登録回数を平準化又は正規化できる場合には、ユークリッド距離やTanimoto係数を用いればよい。

10

【0184】

なお、以上ではヒストグラム情報を用いて次に表示するコンテンツを選択する手法について説明したが、このヒストグラム情報を用いて、登録タグキーワードの整理処理を行うようにしてもよい。例えばヒストグラム情報を用いて、登録回数が少ない登録タグキーワードや古い登録タグキーワードを削除する処理を行う。

【0185】

図21に、ヒストグラム情報に基づくタグの整理処理のフローチャートを示す。まず、整理処理の対象となるコンテンツのヒストグラム情報をヒストグラム情報記憶部328から読み出す(ステップS101)。例えば、定期的に整理処理を行う場合には、ヒストグラム情報記憶部328の中から定期的に1又は複数のコンテンツのヒストグラム情報を読み出す。そして、読み出されたヒストグラム情報に基づいて、登録回数が少ない下位の所定数の登録タグキーワードを抽出する(ステップS102)。或いは登録回数が所定数以下である登録タグキーワードを抽出する。図18の画像I3を例にとれば、「うどん」、「そば」、「めん」、「レストラン」などの登録タグキーワードが抽出される。

20

【0186】

次に、ヒストグラム情報に基づいて、登録が古い下位の所定数の登録タグキーワードを抽出する(ステップS103)。例えば図18において、各登録タグキーワードに対して登録日時情報(時間、日、月、年等)を関連づけておく。そして所定の登録日時よりも古い登録タグキーワードを抽出する。そして、ステップS102やS103で抽出された登録タグキーワードを削除する(ステップS104)。即ち、これらの登録タグキーワードをヒストグラム情報の中から削除する。なお、既に登録されているタグキーワードが再度登録された場合には、登録日時情報は、例えば再度登録された日時(時間、日、月年等)になる。

30

【0187】

本実施形態のようにユーザの発話に基づきコンテンツにタグを自動的に付与する手法によれば、ユーザの手間を省くことができ、コンテンツに対して効率的にタグを付与することが可能になる。

【0188】

しかしながら、コンテンツを見たユーザが、そのコンテンツに関連しないキーワードを発話する場合もある。また、実際にはユーザはコンテンツを見ておらず、無関係な会話をしている場合もある。従って、このようにユーザの発話に基づきタグを付与すると、無関係な多数のタグが付与されてしまう可能性がある。

40

【0189】

この点、図21の手法によれば、登録回数が少ないタグや古いタグは削除されるため、このような無関係な多数のタグが付与されてしまった場合にも、これに対処できる。また、コンテンツの選択に使用されるヒストグラム情報を有効活用して、タグを整理できるという利点もある。

【0190】

50

7. タグ取得のための質問提示

以上では、ユーザに対してコンテンツを表示して、ユーザの発話に基づいてタグを取得する手法について説明した。しかしながら、ユーザに対してコンテンツを表示しても、ユーザがタグ取得のための適切な発話を行わない場合がある。また、ユーザが、そもそも表示コンテンツに興味を持たず、発話しない場合もある。

【0191】

このため、コンテンツに対して有効なタグを付与するためには、システム側からユーザに対して積極的に働きかける仕組みが望まれる。例えば表示コンテンツに対する質問（問いかけ）をユーザに対して提示することで、ユーザとの間のインタラクションが発生し、より多くのタグ情報を引き出すことが可能になる。

10

【0192】

具体的には、コンテンツのタグキーワードに関連する質問を提示し、その回答の発話キーワードを用いて、タグキーワードを登録する。例えば画像に対応する質問を音声や画像で提示する。具体的には食べ物の画像の場合には、「これは美味しそうかな？」という質問を、音声や画像でユーザに提示する。この場合に、食べ物の画像が否かは、画像に関連づけられたタグに基づき特定してもよいし、画像認識により食べ物の画像であることを認識することにより特定してもよい。

【0193】

またコンテンツのタグキーワードが所定数以下の場合に、タグ付与のための質問を提示するようにしてもよい。即ち有効なタグが不足しているときに、「これは何だろう？」というような質問を行う。また、質問後にユーザが発話した肯定、否定の言葉を音声認識して、タグを付与してよい。或いは、質問後にユーザのうなずき、首振りなどの反応（動作）を画像認識して、タグを付与してもよい。また質問の提示は、ユーザが画像を注視していることを条件に行うようにしてもよい。

20

【0194】

図22に、タグキーワード付与のための質問提示処理のフローチャートを示す。まず、コンテンツを表示し、ユーザがコンテンツを注視しているか否かを判断する（ステップS111、S112）。そしてユーザが注視していると判断した場合には、ユーザの発話が検知されたか否かを判断する（ステップS113）。

【0195】

ユーザの発話が検知された場合には、表示コンテンツの登録タグキーワード数が所定数以下か否かを判断する（ステップS114）。そして登録タグキーワード数が所定数以下である場合には、ユーザの発話を音声認識して、発話キーワードを抽出する（ステップS115）。

30

【0196】

次に、抽出された発話キーワード数が0個又は不明な発話キーワードか否かを判断する（ステップS116）。そして、抽出発話キーワード数が0個の場合（発話キーワードを抽出できなかった場合）や、抽出された発話キーワードが不明な発話キーワード（辞書に登録されていないキーワードや質問の作成が不能なキーワード）である場合には、コンテンツ内容をユーザに問うための質問を音声や画像で提示する（ステップS117）。例えば「これは何だろう？」というような質問を提示する。

40

【0197】

一方、ステップS116で、抽出発話キーワード数が0個ではなく、不明な発話キーワードではない場合には、抽出された発話キーワードに基づいて、質問データベースを参照する（ステップS118）。そして質問データベースへの参照結果に基づいて質問を提示する（ステップS119）。即ち、発話キーワードをキーにして質問データベースから読み出された質問を提示する。

【0198】

次に、質問に対するユーザの回答を音声認識し、回答から抽出された発話キーワードを表示コンテンツのタグキーワードとして登録する（ステップS120）。また、回答から

50

肯定、否定のキーワードを抽出し、それに応じたタグキーワードを登録する（ステップ S 1 2 1）。また、顔画像認識により、ユーザのうなずき、首振り等の反応を検出し、反応に応じたタグキーワードを登録する（ステップ S 1 2 2）。なおステップ S 1 2 0、S 1 2 1、S 1 2 2 の全ての処理を実行してもよいし、その一部の処理のみを実行するようにしてもよい。

【 0 1 9 9 】

図 2 3、図 2 4 にタグ辞書データベースや質問データベースの一例を示す。図 2 3 のタグ辞書データベースでは、タグが、階層構造に分類されている。例えば第 1 階層のタグは、より抽象的なタグであり、例えば動物、植物、食べ物等のタグである。第 2 階層のタグは、第 1 階層のタグの下位概念のタグ（抽象度が低い具体的なタグ）であり、例えば第 1 階層のタグの動物のタグに対応して、哺乳類、爬虫類、イヌ等のタグが用意され、植物のタグに対応して、花、木、果物、サクラ等のタグが用意される。第 3 階層のタグは、第 2 階層のタグの下位概念のタグであり、例えば第 2 階層のサクラのタグに対応して、ソメイヨシノ、八重桜などの具体的な名称のタグが用意される。主観タグは、美味しい、楽しい、かわいいなどの主観的なものを表すタグである。

10

【 0 2 0 0 】

図 2 4 の質問データベースでは、図 2 3 のように階層構造化されたタグ辞書データベースを利用して質問情報を取得する。例えば図 2 2 のステップ S 1 1 7 のように、発話キーワードが抽出されなかった場合等には、図 2 4 の質問データベースから「これは何だろう？」という質問を読み出して提示する。また、例えば動物などの第 1 階層のタグを用いて、「これは動物かな？」という質問を行ったり、主観タグを用いて、「これは美味しい？」という質問を行う。

20

【 0 2 0 1 】

また図 2 2 のステップ S 1 1 8、S 1 1 9 のように、抽出された発話キーワードを用いて質問を提示する場合には、タグ辞書データベースの階層構造等を利用して質問を提示する。例えば、発話キーワードから第 1 階層のタグである動物が抽出された場合には、抽出された第 1 階層のタグである動物を用いて、「この動物は何だろう？」という質問を行う。或いは、第 1 階層のタグの動物に対応する第 2 階層のタグのイヌを用いて、「これはイヌかな？」という質問を行う。或いは主観タグを用いて、「これはかわいい？」という質問を行う。

30

【 0 2 0 2 】

以上のように図 2 3、図 2 4 では、タグが第 1 ~ 第 N の階層のタグキーワードに分類されている。そして、発話から抽出された発話キーワードが、第 1 ~ 第 N の階層のタグキーワードのうちの第 K の階層のタグキーワードである場合には、第 K の階層のタグキーワードや第 K の階層のタグキーワードよりも下位の第 K + 1 の階層のタグキーワードを用いて質問を作成して提示する。こうすることで、ユーザの発話を手がかりとして、より具体的なタグを付与するための質問を作成して、ユーザに提示できるようになる。

【 0 2 0 3 】

8 . システム構成の変形例

図 2 5 に本実施形態のシステム構成の変形例について示す。この変形例のシステムでは、サーバ 2 0 0（広義には情報処理システム、狭義にはホームサーバ）が設けられている。このサーバ 2 0 0 は、処理部 2 0 2、記憶部 2 2 0、通信部 2 3 8 を含む。なおこれらの一部の構成要素を省略したり、他の構成要素を追加するなどの種々の変形実施が可能である。なお、図 2 と同様の構成要素については、同様の符号又は名称を付してその説明を省略する。

40

【 0 2 0 4 】

処理部 2 0 2 は、サーバ管理処理などの各種の処理を行うものであり、CPU 等のプロセッサや ASIC などにより実現できる。記憶部 2 2 0 は、処理部 2 0 2 や通信部 2 3 8 のワーク領域となるものであり、例えば RAM や HDD 等により実現できる。通信部 2 3 8 は、デジタルフォトフレーム 3 0 0 や、外部サーバ 6 0 0 との間で、有線又は無線で通

50

信を行うためのものであり、通信用 A S I C 又は通信用プロセッサなどにより実現できる。例えばデジタルフォトフレーム 3 0 0 とサーバ 2 0 0 は、例えば無線 L A N 等のネットワークで通信接続される。

【 0 2 0 5 】

図 2 5 では、サーバ側の処理部 2 0 2 が、タグ登録部 2 0 6、登録処理部 2 1 1、ヒストグラム作成部 2 1 3、類似度評価部 2 1 4、タグ管理部 2 1 5、質問提示部 2 1 6、コンテンツ選択部 2 1 7、表示指示部 2 1 8 を含む。またサーバ側の記憶部 2 2 0 が、コンテンツ情報記憶部 2 2 2、ユーザ情報記憶部 2 2 7、ヒストグラム情報記憶部 2 2 8、質問情報記憶部 2 2 9 を含む。なお、これらの一部をデジタルフォトフレーム 3 0 0 の処理部 3 0 2 や記憶部 3 2 0 に設けるようにしてもよい。

10

【 0 2 0 6 】

また図 2 5 では、デジタルフォトフレーム 3 0 0 側の処理部 3 0 2 が、音情報取得部 3 0 3、検知情報取得部 3 0 4、音声認識部 3 0 5、ユーザ状態判断部 3 0 7、ユーザ認識部 3 1 2、表示制御部 3 1 8 を含む。なおこれらの一部をサーバ 2 0 0 側の処理部 2 0 2 に設けるようにしてもよい。

【 0 2 0 7 】

図 2 5 では、コンテンツ選択部 2 1 7 は、ユーザに提示するコンテンツの選択処理を行う。例えば記憶部 2 2 0 のコンテンツ情報記憶部 2 2 2 からコンテンツ情報を読み出したり、外部サーバ 6 0 0 にアクセスしてコンテンツ情報を受信することで、ユーザに提示するコンテンツを選択する。

20

【 0 2 0 8 】

表示指示部 2 1 8 は、コンテンツ選択部 2 1 7 により選択されたコンテンツに基づいて、デジタルフォトフレーム 3 0 0 (広義には画像表示部) の表示部 3 4 0 に表示される画像の表示指示を行う。具体的にはコンテンツ選択部 2 1 7 により選択されたコンテンツの画像を表示部 3 4 0 に表示するための指示を行う。そしてデジタルフォトフレーム 3 0 0 の表示制御部 3 1 8 は、サーバ 2 0 0 の表示指示部 2 1 8 からの指示にしたがって、表示部 3 4 0 の表示制御を行う。これにより、コンテンツ選択部 2 1 7 により選択されたコンテンツの画像等が表示部 3 4 0 に表示されるようになる。

【 0 2 0 9 】

そしてタグ登録部 2 0 6 は、コンテンツの表示中におけるユーザの発話の音声認識により抽出されたユーザの発話キーワードに基づいて、コンテンツのタグキーワードの登録処理を行う。例えばデジタルフォトフレーム 3 0 0 側の音声認識部 3 0 5 が、センサ 3 5 0 である音センサからの音情報に基づいて、ユーザの発話の音声認識処理を行い、ユーザの発話キーワードを抽出する。すると、この抽出された発話キーワードの情報は、通信部 3 3 8、2 3 8 を介してサーバ 2 0 0 側に送られる。そして、タグ登録部 2 0 6 は、受信したユーザの発話キーワードの情報に基づいて、コンテンツのタグキーワードの登録処理を行う。なお音声認識部 3 0 5 をサーバ 2 0 0 側に設け、サーバ 2 0 0 側で音声認識や発話キーワードの抽出処理を行うようにしてもよい。また図 2 5 では、本実施形態で説明したユーザ情報の登録処理、ヒストグラムの作成処理、類似度の評価処理、タグの管理処理、質問の提示処理は、サーバ 3 0 0 側が行うが、これらの一部又は全部をデジタルフォトフ

30

40

【 0 2 1 0 】

図 2 5 の変形例によれば、コンテンツの選択処理やタグ登録処理等をサーバ 2 0 0 が行うため、デジタルフォトフレーム 3 0 0 の処理負荷を軽減できる。従って、デジタルフォトフレーム 3 0 0 の処理部 3 0 2 (C P U) の処理能力が低い場合も、本実施形態の処理を実現できるようになる。なお、これらの処理を、サーバ 2 0 0 とデジタルフォトフレーム 3 0 0 の分散処理により実現してもよい。

【 0 2 1 1 】

なお、上記のように本実施形態について詳細に説明したが、本発明の新規事項および効果から実体的に逸脱しない多くの変形が可能であることは当業者には容易に理解できるで

50

あろう。従って、このような変形例はすべて本発明の範囲に含まれるものとする。例えば、明細書又は図面において、少なくとも一度、より広義または同義な異なる用語と共に記載された用語は、明細書又は図面のいかなる箇所においても、その異なる用語に置き換えることができる。また画像表示システム、デジタルフォトフレーム、情報処理システムの構成、動作や、タグ登録手法、コンテンツの表示手法、音声認識手法、視認状態判定手法等も本実施形態で説明したものに限定されず、種々の変形実施が可能である。

【符号の説明】

【0212】

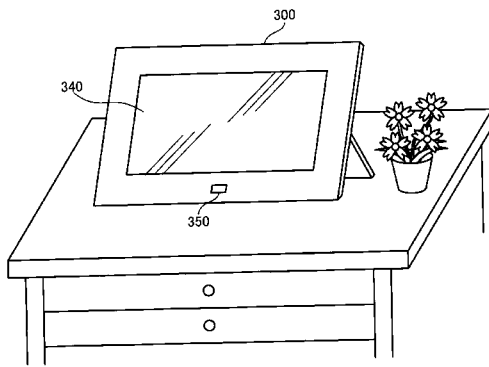
- 200 サーバ、202 処理部、206 タグ登録部、211 登録処理部、
- 213 ヒストグラム作成部、214 類似度評価部、215 タグ管理部、
- 217 コンテンツ選択部、218 表示指示部、220 記憶部、
- 222 コンテンツ情報記憶部、227 ユーザ情報記憶部、
- 228 ヒストグラム情報記憶部、229 質問情報記憶部、238 通信部、
- 300 デジタルフォトフレーム、302 処理部、303 音情報取得部、
- 304 検知情報取得部、305 音声認識部、306 タグ登録部、
- 307 ユーザ状態判断部、308 視認状態判断部、309 位置関係判断部、
- 310 人数検出部、311 登録処理部、312 ユーザ認識部、
- 313 ヒストグラム作成部、314 類似度評価部、315 タグ管理部、
- 316 質問提示部、317 コンテンツ選択部、318 表示制御部、
- 320 記憶部、322 コンテンツ情報記憶部、323 音情報記憶部、
- 324 検知情報記憶部、325 登録キーワード記憶部、326 ユーザ状態記憶部、
- 327 ユーザ情報記憶部、328 ヒストグラム情報記憶部、
- 329 質問情報記憶部、330 情報記憶媒体、338 通信部、340 表示部、
- 350 センサ、360 操作部、600 外部サーバ

10

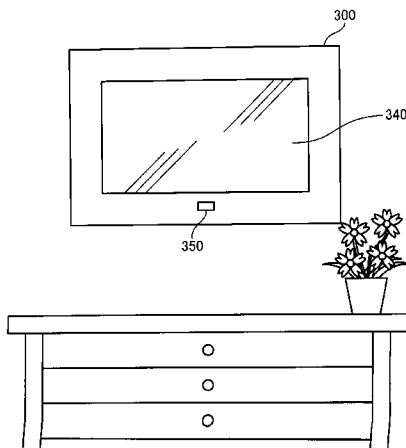
20

【図1】

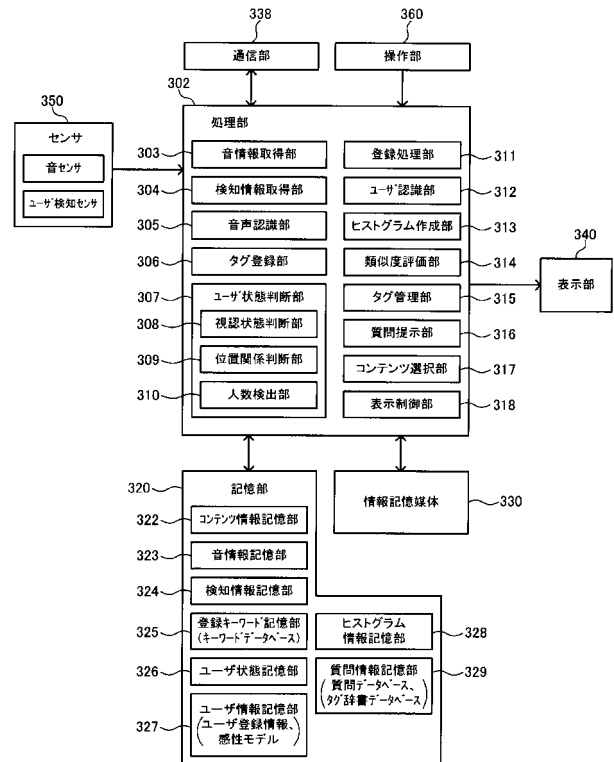
(A)



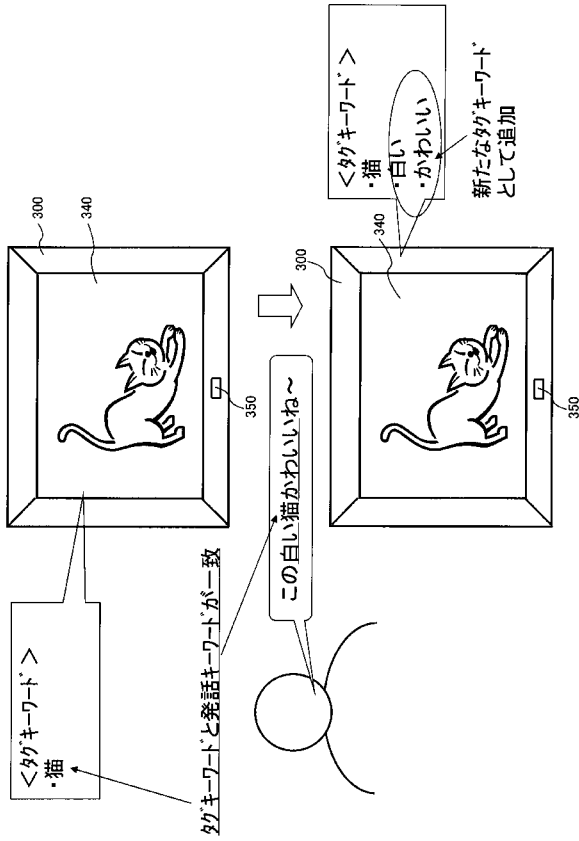
(B)



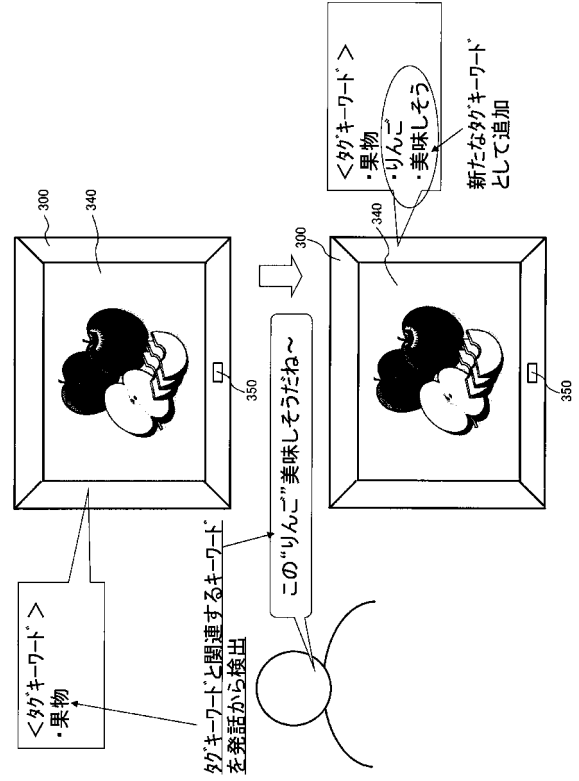
【図2】



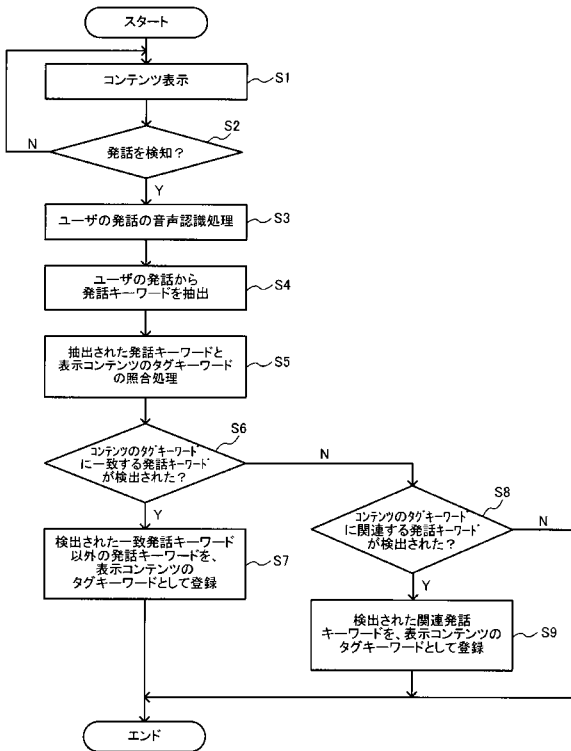
【 図 3 】



【 図 4 】

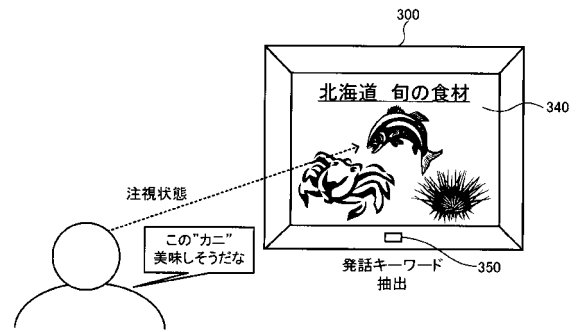


【 図 5 】

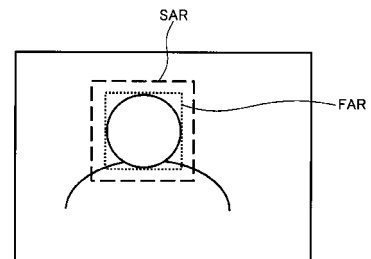


【 図 6 】

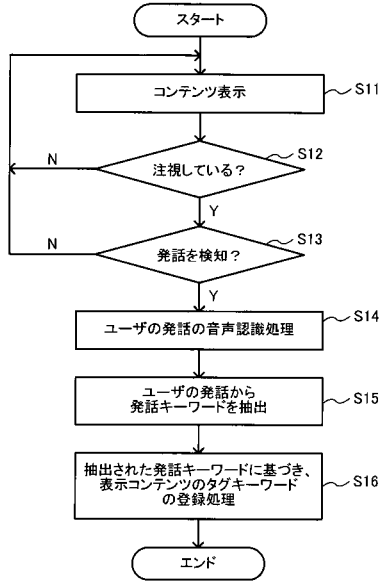
(A)



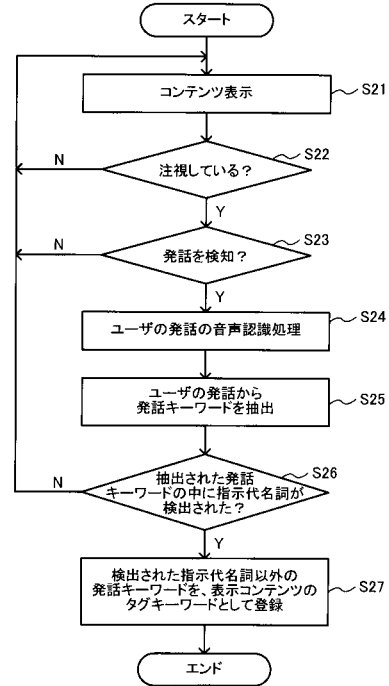
(B)



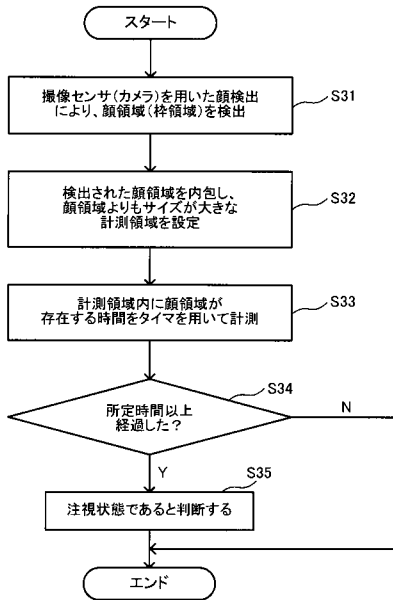
【 図 7 】



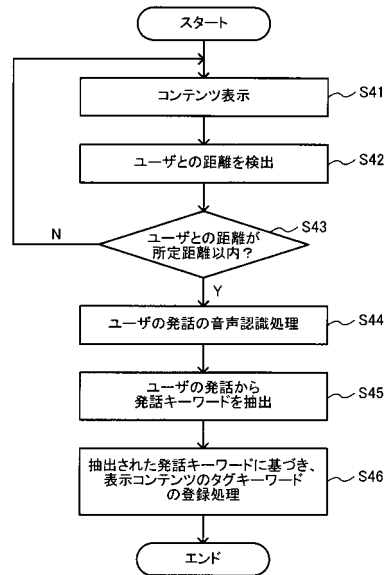
【 図 8 】



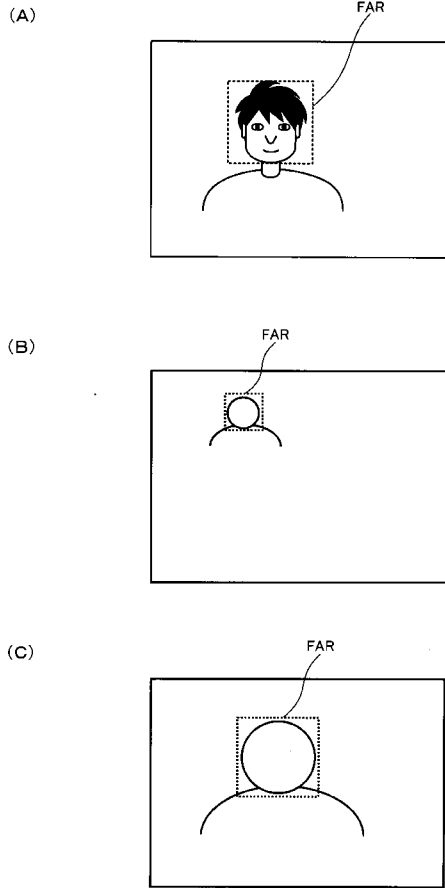
【 図 9 】



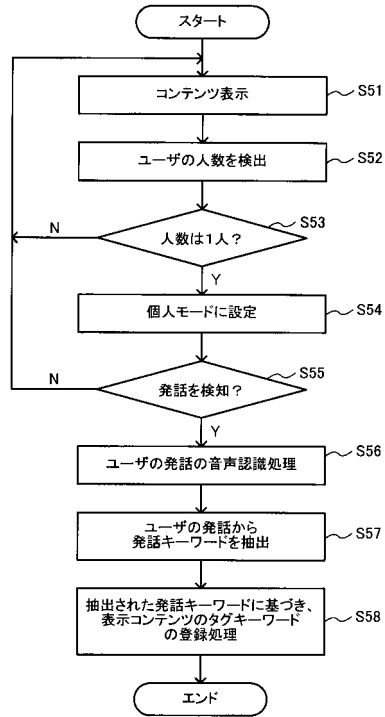
【 図 10 】



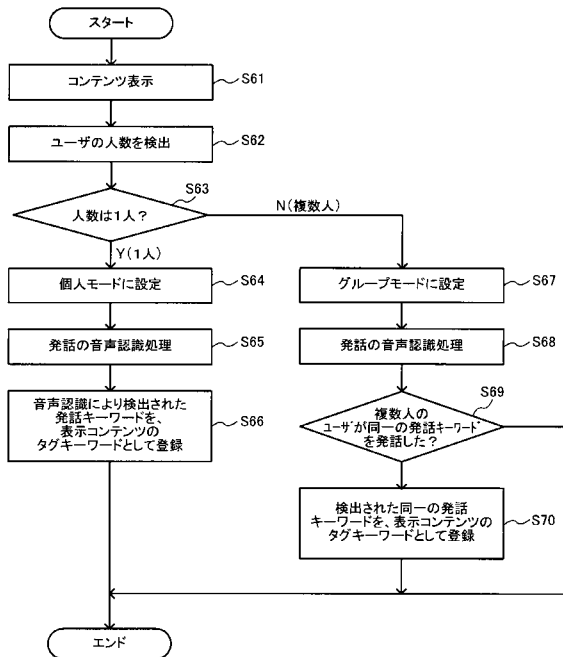
【図 1 1】



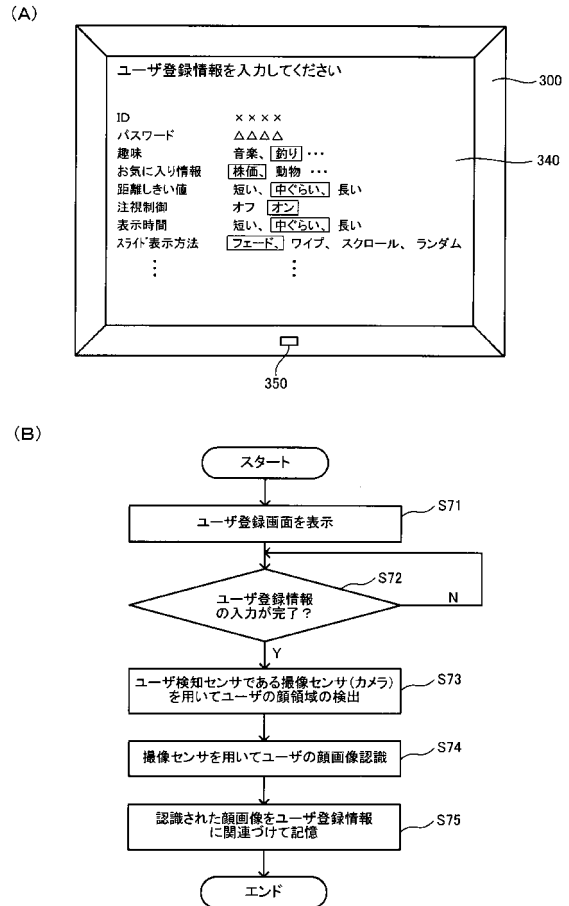
【図 1 2】



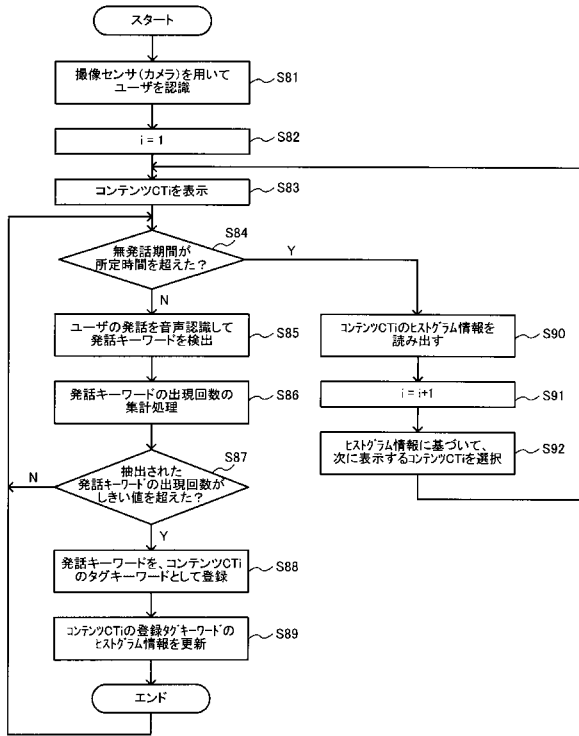
【図 1 3】



【図 1 4】



【図15】



【図16】

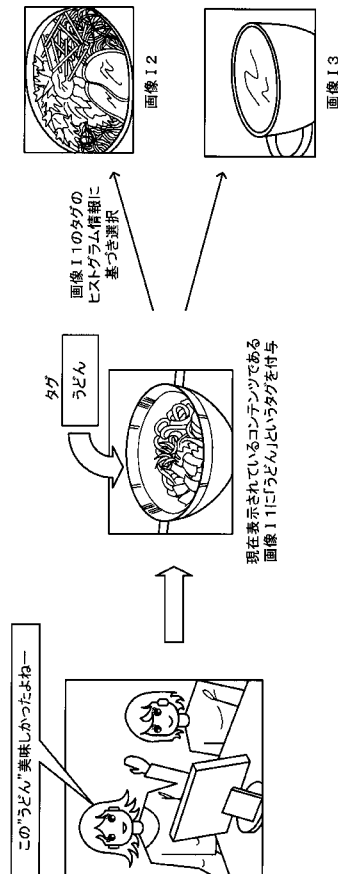
(A) コンテンツテーブル

コンテンツID	コンテンツURL	登録タグキーワード
CID1	URL1	TAG11, TAG12, TAG13...
CID2	URL2	TAG21, TAG22, TAG23...
⋮	⋮	⋮

(B) 登録回数テーブル(ユーザID_コンテンツID)

登録タグキーワード	登録回数
TAG11	5
TAG12	7
TAG13	2
⋮	⋮

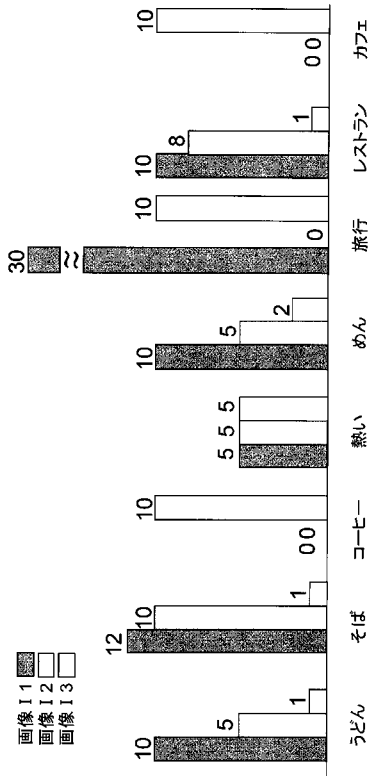
【図17】



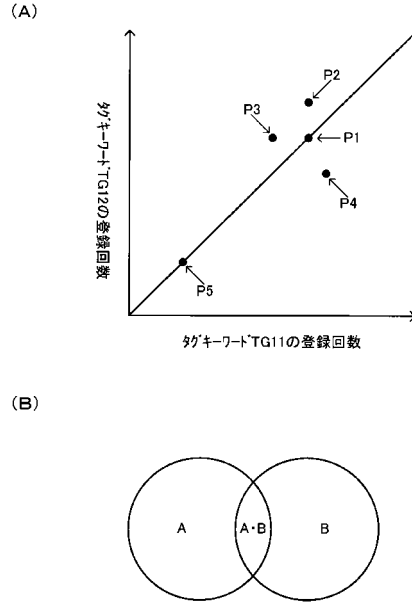
【図18】

登録タグキーワード	画像I1	画像I2	画像I3
うどん	10	5	1
そば	12	10	1
コーヒー	0	0	10
熱い	5	5	5
めん	10	5	2
旅行	30	0	10
レストラン	10	8	1
カフェ	0	0	10

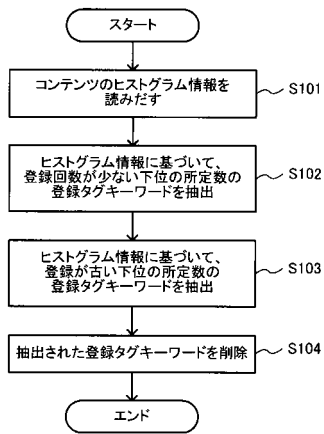
【図19】



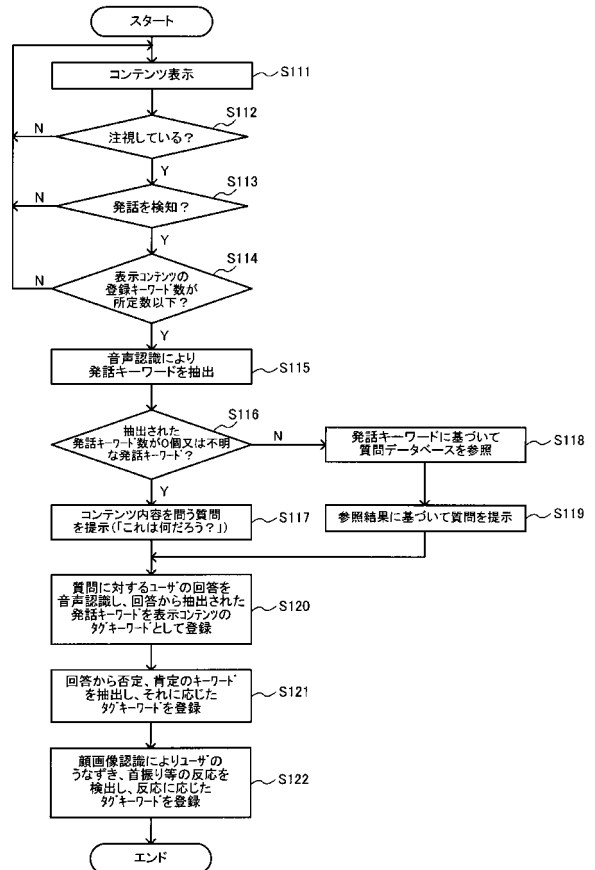
【図20】



【図21】



【図22】



【図 2 3】

タグ辞書データベース

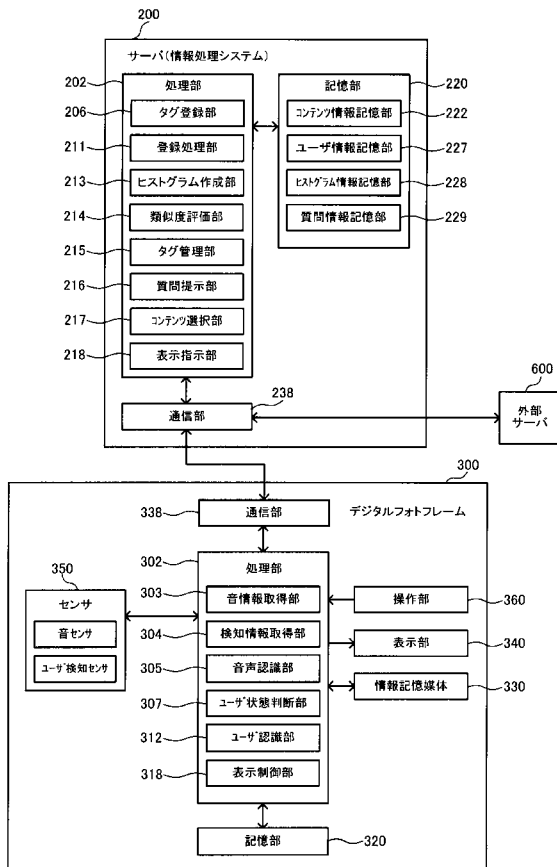
第1階層タグ	動物・植物・食べ物・物質・道具・電気製品・地名・人
第2階層タグ	<ul style="list-style-type: none"> 動物:哺乳類、爬虫類、昆虫、イヌ、ネコ、ゾウ、ライオン... 植物:花、木、葉物、サクラ、バラ... 食べ物:ラーメン、カレー、リンゴ... ...
第3階層タグ	<ul style="list-style-type: none"> イヌ:トイプードル、チワワ... サクラ:ソメイヨシノ、八重桜... ラーメン:醤油ラーメン、塩ラーメン... ...
主観タグ	<ul style="list-style-type: none"> 美味しい、楽しい、かわいい、面白い、綺麗、好き、嫌い...

【図 2 4】

質問データベース

第1階層	第2階層	第3階層	主観	質問
-	-	-	-	<ul style="list-style-type: none"> これは何だろう? これは〇〇(第1階層タグを入れて)かな? これは〇〇(主観タグを入れて)?
動物	-	-	-	<ul style="list-style-type: none"> この動物(対象のタグ)は何だろう? これはイヌ(第2階層タグを入れて)かな? これかわいい(主観タグを入れて)?
植物	サクラ	-	-	<ul style="list-style-type: none"> このサクラ(対象のタグ)は何だろう? これはソメイヨシノ(第3階層タグを入れて)かな? これ綺麗(主観タグを入れて)?
食べ物	ラーメン	塩ラーメン	-	<ul style="list-style-type: none"> 〇〇(第3階層タグ)は〇〇(主観タグを入れて)? 塩ラーメンは好き? これ美味しいかな? かわいいね。これは何というの?
-	-	-	かわいい	-

【図 2 5】



フロントページの続き

(51)Int.Cl. F I テーマコード(参考)
G 0 6 F 3/16 3 2 0 H
G 0 6 F 3/16 3 3 0 C

(72)発明者 井場 陽一
東京都渋谷区幡ヶ谷 2 丁目 4 3 番 2 号 オリンパス株式会社内

(72)発明者 亀山 未帆
東京都渋谷区幡ヶ谷 2 丁目 4 3 番 2 号 オリンパス株式会社内

(72)発明者 藤垣 勇人
東京都渋谷区幡ヶ谷 2 丁目 4 3 番 2 号 オリンパス株式会社内

F ターム(参考) 5B050 BA10 GA08
5B075 ND06 NK39 PP07 PP28
5D015 KK02
5E501 AA30 AC15 AC34 BA17 CA02 CB15 DA16 EA21 FA47