

(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2017-211995

(P2017-211995A)

(43) 公開日 平成29年11月30日(2017.11.30)

(51) Int.Cl.		F I		テーマコード (参考)
G 0 6 F 17/30 (2006.01)		G 0 6 F	17/30	1 7 0 E
H 0 4 N 5/76 (2006.01)		H 0 4 N	5/76	
G 1 0 L 25/51 (2013.01)		G 1 0 L	25/51	
G 1 0 L 15/00 (2013.01)		G 1 0 L	15/00	2 0 0 G
		G 0 6 F	17/30	3 1 0 Z
審査請求 有 請求項の数 9 O L (全 30 頁) 最終頁に続く				

(21) 出願番号	特願2017-122583 (P2017-122583)	(71) 出願人	000000376
(22) 出願日	平成29年6月22日 (2017. 6. 22)		オリンパス株式会社
(62) 分割の表示	特願2013-54258 (P2013-54258)		東京都八王子市石川町2951番地
	の分割	(74) 代理人	100076233
原出願日	平成25年3月15日 (2013. 3. 15)		弁理士 伊藤 進
		(74) 代理人	100101661
			弁理士 長谷川 靖
		(74) 代理人	100135932
			弁理士 篠浦 治
		(72) 発明者	本間 伸祐
			東京都八王子市石川町2951番地 オリ
			ンパス株式会社内
		(72) 発明者	佐藤 和宏
			東京都八王子市石川町2951番地 オリ
			ンパス株式会社内
最終頁に続く			

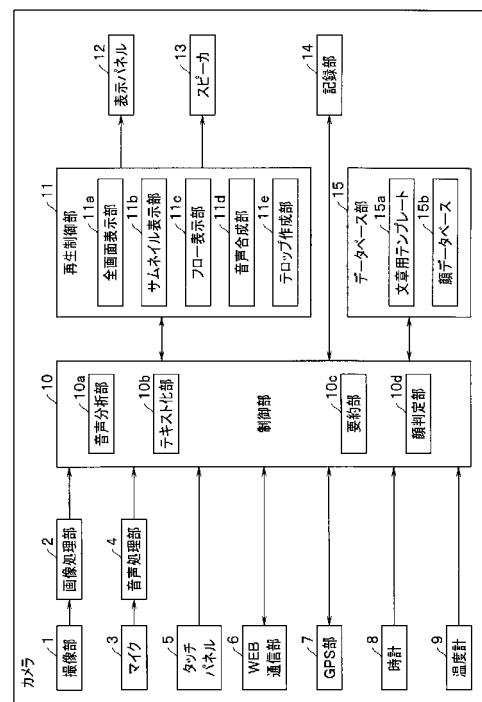
(54) 【発明の名称】 再生装置、再生方法、再生プログラム、音声要約装置、音声要約方法および音声要約プログラム

(57) 【要約】

【課題】情報の回想を効果的に行うことができる再生装置を提供する。

【解決手段】所定期間の音声データを取得する音声データ取得部と、上記音声データ取得部が取得した上記音声データから特徴部を抽出する特徴部抽出部と、上記特徴部に係るテキストデータを得るテキストデータ取得部と、上記テキストデータに係る情報を再生する再生部と、を備える

【選択図】 図 1



【特許請求の範囲】**【請求項 1】**

所定期間の音声データを取得する音声データ取得部と、
上記音声データ取得部が取得した上記音声データから特徴部を抽出する特徴部抽出部と

、
上記特徴部に係るテキストデータを得るテキストデータ取得部と、
上記テキストデータに係る情報を再生する再生部と、
を備えることを特徴とする再生装置。

【請求項 2】

上記特徴部抽出部は、上記音声データ取得部が取得した上記音声データから特徴的音声
を音声分析する音声分析部を備え、

上記テキストデータ取得部は、上記音声分析部において音声分析された音声をテキスト
データ化するテキストデータ化部を備え、

上記再生部は、上記テキストデータ化部においてテキストデータ化された上記音声に係
るテキスト情報を再生する

ことを特徴とする請求項 1 に記載の再生装置。

10

【請求項 3】

複数の文章テンプレートからタグ情報に応じて所定の文章テンプレートを取得する文章
テンプレート取得部

をさらに備え、

上記再生部は、上記文章テンプレート取得部において取得した上記文章テンプレートを
再生する

ことを特徴とする請求項 1 に記載の再生装置。

20

【請求項 4】

所定期間の音声データを取得する音声データ取得工程と、

上記音声データ取得工程において取得した上記音声データから特徴部を抽出する特徴部
抽出工程と、

上記特徴部に係るテキストデータを得るテキストデータ取得工程と、

上記テキストデータに係る情報を再生する再生工程と、

を有することを特徴とする再生方法。

30

【請求項 5】

所定期間の音声データを取得する音声データ取得工程と、

上記音声データ取得工程において取得した上記音声データから特徴部を抽出する特徴部
抽出工程と、

上記特徴部に係るテキストデータを得るテキストデータ取得工程と、

上記テキストデータに係る情報を再生する再生工程と、

をコンピュータに実行させるための再生プログラム。

【請求項 6】

取得した所定期間の音声データの音圧または音の周期性に基づいて、当該音声データが
人の声または環境音を含むか否かを分析し分類する音声分類部と、

上記音声データにおいて特徴的な環境音を第 1 の音声要約とする第 1 音声要約部と、

上記第 1 の音声要約から時間的に短縮した第 2 の音声要約を取得する第 2 音声要約部と

40

、
上記環境音であると分析された音声データを擬音テキストデータベースからテキストを
選択して擬音テキストとするテキスト化部と、

を備えることを特徴とする音声要約装置。

【請求項 7】

上記第 2 の音声要約は、人の声から選択した音声であることを特徴とする請求項 6 に記
載の音声要約装置。

【請求項 8】

50

取得した所定期間の音声データの音圧または音の周期性に基づいて、当該音声データが人の声または環境音を含むか否かを分析し分類する音声分類工程と、

上記音声データにおいて特徴的な環境音を第１の音声要約とする第１音声要約工程と、

上記第１の音声要約から時間的に短縮した第２の音声要約を取得する第２音声要約工程と、

上記環境音であると分析された音声データを擬音テキストデータベースからテキストを選択して擬音テキストとするテキスト化工程と、

を有することを特徴とする音声要約方法。

【請求項 9】

取得した所定期間の音声データの音圧または音の周期性に基づいて、当該音声データが人の声または環境音を含むか否かを分析し分類する音声分類工程と、

上記音声データにおいて特徴的な環境音を第１の音声要約とする第１音声要約工程と、

上記第１の音声要約から時間的に短縮した第２の音声要約を取得する第２音声要約工程と、

上記環境音であると分析された音声データを擬音テキストデータベースからテキストを選択して擬音テキストとするテキスト化工程と、

をコンピュータに実行させるための音声要約プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、画像再生モードとして複数のモードを設定可能な再生装置、再生方法、再生プログラムに関する。

【背景技術】

【0002】

撮影して得られた画像を鑑賞するための再生装置は、従来より、種々のものが提案されている。

【0003】

例えば、特開 2004 - 104426 号公報には、画像を撮影して撮影画像データとする画像生成手段と、画像生成手段が生成した撮影画像データを記憶する記憶手段と、画像生成手段が生成した撮影画像データと前記記憶手段が既に記憶している撮影画像データの少なくとも１つとを比較して、その差分を検出する差分検出手段と、差分検出手段による差分検出結果を表示する表示手段とを有し、過去に撮影した画像を保存して、現在の撮影画像と比較して差分を検出することができ、これにより良好な娯楽性の実現や利便性の向上を図った、付加価値の高い撮像装置が記載されている。また、該公報には、画像の表示と合わせて文字表示を行う技術も記載されている。

【先行技術文献】

【特許文献】

【0004】

【特許文献 1】特開 2004 - 104426 号公報

【発明の概要】

【発明が解決しようとする課題】

【0005】

しかし、思い出をたどって回想するときには、必ずしも一枚の画像のみに基づいて行うわけではなく、複数の画像を見ることによっても相乗効果が生まれて様々な感情が想起される場合もある。

【0006】

従って、一枚の画像のみを鑑賞する場合は、一枚の画像の貴重な一瞬の思い出が重要である。また、複数の画像を一度に鑑賞する場合は、イベントの総合的な思い出が簡単に思い出せることが重要であり、これによって、一枚の画像の検索を迅速に行うことができる。こうして、例えば撮影時の状況をよりリアルに回想することができるような、より効果

10

20

30

40

50

的な鑑賞法を提供し得る再生装置が望まれている。

【 0 0 0 7 】

本発明は上記事情に鑑みてなされたものであり、情報の回想を効果的に行うことができる再生装置、再生方法、再生プログラムを提供することを目的としている。

【課題を解決するための手段】

【 0 0 0 8 】

本発明の第 1 の態様による再生装置は、所定期間の音声データを取得する音声データ取得部と、上記音声データ取得部が取得した上記音声データから特徴部を抽出する特徴部抽出部と、上記特徴部に係るテキストデータを得るテキストデータ取得部と、上記テキストデータに係る情報を再生する再生部と、を備える。

10

【 0 0 0 9 】

本発明の第 2 の態様による再生方法は、所定期間の音声データを取得する音声データ取得工程と、上記音声データ取得工程において取得した上記音声データから特徴部を抽出する特徴部抽出工程と、上記特徴部に係るテキストデータを得るテキストデータ取得工程と、上記テキストデータに係る情報を再生する再生工程と、を有する。

【 0 0 1 0 】

本発明の第 3 の態様による再生プログラムは、所定期間の音声データを取得する音声データ取得工程と、上記音声データ取得工程において取得した上記音声データから特徴部を抽出する特徴部抽出工程と、上記特徴部に係るテキストデータを得るテキストデータ取得工程と、上記テキストデータに係る情報を再生する再生工程と、をコンピュータに実行させる。

20

【発明の効果】

【 0 0 1 1 】

本発明によれば、情報の回想を効果的に行うことができる再生装置、再生方法、再生プログラムを提供することができる。

【図面の簡単な説明】

【 0 0 1 2 】

【図 1】本発明の実施形態 1 の再生装置が適用された撮像装置としてのカメラの構成を示すブロック図。

【図 2】上記実施形態 1 において、静止画記録の前後に行われる音声録音の例を示すタイミングチャート。

30

【図 3】上記実施形態 1 において、画像再生時に再生される音声データの例を示す図表。

【図 4】上記実施形態 1 における画像再生モードの幾つかの例を説明するための図表。

【図 5】上記実施形態 1 における全画面表示モードの様子を示す図。

【図 6】上記実施形態 1 におけるサムネイル表示モードの様子を示す図。

【図 7】上記実施形態 1 におけるフロー表示モードの様子を示す図。

【図 8】上記実施形態 1 のカメラのメイン処理を示すフローチャート。

【図 9】上記実施形態 1 における撮影モードの処理を示すフローチャート。

【図 10】上記実施形態 1 における再生モードの処理を示すフローチャート。

【図 11A】上記実施形態 1 における 3W + 1H 要約テキスト化の処理の一部を示すフローチャート。

40

【図 11B】上記実施形態 1 における 3W + 1H 要約テキスト化の処理の他の一部を示すフローチャート。

【図 12】上記実施形態 1 における全画面表示の処理を示すフローチャート。

【図 13】上記実施形態 1 におけるフロー表示の処理を示すフローチャート。

【図 14】本発明の実施形態 2 において、タグ表示を伴う全画面表示の第 1 の例を示す図。

【図 15】上記実施形態 2 において、タグ表示を伴う全画面表示の第 2 の例を示す図。

【図 16】上記実施形態 2 において、タグ表示を伴う全画面表示の第 3 の例を示す図。

【図 17】上記実施形態 2 において、タグ表示を伴う組写真表示の様子を示す図。

50

【図 18】上記実施形態 2 における再生モードの処理を示すフローチャート。

【発明を実施するための形態】

【0013】

以下、図面を参照して本発明の実施の形態を説明する。

[実施形態 1]

【0014】

図 1 から図 13 は本発明の実施形態 1 を示したものであり、図 1 は再生装置が適用された撮像装置としてのカメラの構成を示すブロック図である。ここに、再生装置は、画像を表示する場合には画像再生装置（あるいは表示装置）、音声を再生する場合には音声再生装置、などとなる。

10

【0015】

カメラは、撮像部 1 と、画像処理部 2 と、マイク 3 と、音声処理部 4 と、タッチパネル 5 と、WEB 通信部 6 と、GPS 部 7 と、時計 8 と、温度計 9 と、制御部 10 と、再生制御部 11 と、表示パネル 12 と、スピーカ 13 と、記録部 14 と、データベース部 15 と、を備えている。

【0016】

撮像部 1 は、画像を撮像するものであり、被写体の光学像を光電変換して撮像信号を出力する。従って、本実施形態の再生装置は、撮像装置としての機能を備えている。

【0017】

画像処理部 2 は、撮像部 1 からの撮像信号に、増幅、デジタル化、ホワイトバランス、ノイズ除去、ガンマ補正等の各種の信号処理を施す。

20

【0018】

マイク 3 は、音声データを録音するための録音部の一部であり、動画撮像時の音声を音声信号に変換するだけでなく、静止画撮影時刻前後の音声も音声信号に変換している。

【0019】

音声処理部 4 は、録音部の一部であり、マイク 3 から出力される音声信号を増幅しデジタル化して、必要に応じて風切り音を低減するなどの音声処理を行う。

【0020】

タッチパネル 5 は、表示パネル 12 の表示面上に貼設されていて、ユーザがタッチ操作を行うことにより入力を行うためのデバイスである。なお、ここではタッチパネル 5 を設けたが、タッチパネル 5 に代えて、もしくはタッチパネル 5 に加えて、専用のスイッチ（例えば十字パッド等）をカメラに設けても勿論構わない。

30

【0021】

WEB 通信部 6 は、無線 LAN や電気通信事業のパケット通信等を介してネットワーク、例えばインターネット等へ接続するための装置である。この WEB 通信部 6 を介して、画像をインターネット等へアップロードしたり、インターネットから地図情報や天気情報などを取得することができる。ただし、天気情報については、WEB 通信部 6 を介して取得するに限るものではなく、例えば、撮像部 1 により撮影した画像や温度計 9 から取得した温度情報に基づき判定する（例えば、青空が撮影されているから「晴れ」である、取得した温度が 30 ° であるから「暑い」等）ようにしても構わない。

40

【0022】

GPS 部 7 は、GPS（グローバル・ポジショニング・システム：Global Positioning System）衛星からの信号を受信して、現在位置を測定する GPS 受信機である。この GPS 部 7 は、消費電力を削減する観点からオン/オフを切り換えることができるようになっており、GPS 位置情報を取得するモードに設定されている場合にはオンに切り換えられ、GPS 位置情報を取得するモードに設定されていない場合にはオフに切り換えられる。

【0023】

時計 8 は、現在時刻を出力する計時部である。

50

【 0 0 2 4 】

温度計 9 は、カメラの周囲の環境温度を取得する温度測定部である。ただし、温度情報を W E B 通信部 6 を介してインターネットから取得しても良く、この場合には、カメラに温度計 9 を設けなくても構わない。

【 0 0 2 5 】

制御部 1 0 は、カメラの各部を統括的に制御するものであり、画像を画像群に分類する画像分類部としての機能や、画像の特徴を検出する（例えば、画像における画面上方に青空があるか否かや、複数の画像同士が類似しているか否か等を検出する）特徴検出部としての機能も果たすようになっている。ここに、画像群とは、分類上における特定の性質を備える 1 つ以上の画像の集合であり、複数の画像が含まれる場合には、共通する特定の性質を備えるという点で互いに関連性のある複数の画像の集合となる。以下では簡単のために、画像群には複数の画像が含まれていると想定して説明を行うが、1 つの画像群に含まれる画像が 1 つであることを妨げるものではない。画像群の具体例としては、撮影時刻が近い複数の画像、同一人物が写っている複数の画像、被写体が類似している複数の画像、などが挙げられる。

【 0 0 2 6 】

再生制御部 1 1 は、カメラに設定されている画像再生モードに従って画像を表示パネル 1 2 に表示再生させると共に、画像再生モードが画面に画像を拡大して表示する第 1 の表示モード（第 1 の再生モード）（具体的な一例として、画面に画像を 1 つだけ表示する全画面表示モード）であるときには後述する第 1 の要約を、画像再生モードが画面に画像を縮小して複数表示する第 2 の表示モード（第 2 の再生モード）（後述する、サムネイル表示モード、フロー表示モード等）であるときには後述する第 2 の要約を再生させるものである。再生制御部 1 1 は、具体的に、要約が音声データであるときにはスピーカ 1 3 に音声再生させ、要約がテキスト（文字情報）である場合には表示パネル 1 2 に表示再生させる。第 1 の要約は、画像撮影時の状況を認識するのを補助する情報（補助情報）が表示可能な情報である場合には、そのまま表示再生しても良い。

【 0 0 2 7 】

表示パネル 1 2 は、再生制御部 1 1 から出力される画像（撮影された画像）や、カメラに関する各種情報などを画面に表示する表示部である。

【 0 0 2 8 】

スピーカ 1 3 は、再生制御部 1 1 から出力される音声データを音声として再生する音声再生部である。

【 0 0 2 9 】

記録部 1 4 は、撮像部 1 から取得された画像やマイク 3 から取得された音声などを記録する、コンピュータにより読み取り可能な一時的でない記録媒体である。従って、記録部 1 4 は録音部の一部であり、本実施形態の再生装置は録音装置としての機能を備えている。

【 0 0 3 0 】

データベース部 1 5 は、カメラにおいて用いる各種のデータを、不揮発かつ書き換え可能に記録するコンピュータにより読み取り可能な一時的でない記録媒体である。なお、ここではデータベース部 1 5 を記録部 1 4 とは別体に設けたが、記録部 1 4 内に設けても構わない。

【 0 0 3 1 】

次に、上述した制御部 1 0 は、音声分析部 1 0 a と、テキスト化部 1 0 b と、要約部 1 0 c と、顔判定部 1 0 d と、を備えている。

【 0 0 3 2 】

制御部 1 0 は、音声処理部 4 を介してマイク 3 から入力される音声から、撮影時点を含む所定時間（撮影時点の前の前録時間（例えば 5 秒）、および撮影時点の後の後録時間（例えば 5 秒）（つまり、例えば合計 1 0 秒））の音声を、図 2 に示すように第 1 の音声データ t s 1 として録音するように制御する。

10

20

30

40

50

【 0 0 3 3 】

ここに、図 2 は静止画記録の前後に行われる音声録音の例を示すタイミングチャートである。

【 0 0 3 4 】

音声分析部 1 0 a は、音声処理部 4 から入力された音声データを分析して、例えば波の音である、さざ波の音である、風音である、破裂音である、呼びかけ等の人の声である、などの音声の種類に分類する処理を行う。

【 0 0 3 5 】

テキスト化部 1 0 b は、上述した音声分析部 1 0 a の分析結果に基づき、人の声であると分類された場合には音声認識を行ってテキストデータに変換し、人の声以外の音であると分類された場合には擬音テキストデータベース（この擬音テキストデータベースは、後述するように、データベース部 1 5 の例えば文章用テンプレート 1 5 a に付帯して設けられている）の中から該当する擬音、波の音である場合には「ざざー」、さざ波の音である場合には「ちゃぶちゃぶ」、風音である場合には「そよそよ」、破裂音である場合には「バーン」等を選択する。

【 0 0 3 6 】

要約部 1 0 c は、画像撮影時の状況を認識するのを補助する情報である要約を作成するものである。

【 0 0 3 7 】

具体的に、要約部 1 0 c は、例えば第 1 の音声データ t s 1 自体を第 1 の要約とする。ここに、第 1 の要約は、いわゆる生録音声（マイク 3 から入力される音声をそのまま録音した音声）に限るものではなく、雑音などを除去して要約し上述した所定時間（例えば 1 0 秒）にした音声でも良いし、必要な部分のみを採用して所定時間（例えば 1 0 秒）にした音声であっても構わないし、その他であっても良い。つまり要約は、写真などの画像と共に再生することで撮影時を思い出すのに有効な情報であれば良い。具体例として、要約は、その画像にふさわしい特徴的な環境音や、撮影するときのかけ声や返事等の一連の会話などが相当する。

【 0 0 3 8 】

そして、要約部 1 0 c は、第 1 の音声データ t s 1 の中から特徴的な一部分、例えば図 2 に示すように音圧が最も高い部分を含む所定時間分（例えば 1 秒分）を抽出して、第 1 の要約よりも画像単位での認識に要する時間（具体的には、再生時間）が短い第 2 の要約である第 2 の音声データ t s 2 とする。なお、特徴的な部分の他の例としては、第 1 の音声データ t s 1 の中の、人の声であると分類された部分、周期的に発生している特徴音（例えば波の音）などが挙げられる。

【 0 0 3 9 】

さらに、要約部 1 0 c は、画像群に含まれる全画像の第 1 の音声データ t s 1 でなる第 1 の音声データ t s 1 群から、特徴部分である特徴音を所定時間分だけ抽出して第 1 の音声データ t s 1 よりも再生時間が短い第 3 の音声データ t s 3 を作成するとともに、作成した第 3 の音声データ t s 3 を他の第 2 の要約とする。この第 2 の要約において、大きな音の部分を選択したり、撮影者の声などは省いて、肝心の被写体の声だけを選択するような工夫も有効である。また、周期的に繰り返される音声を第 2 の要約とする場合には、2 周期分のみを採用する、などの短縮を行うことも有効である。つまり第 2 の要約においては、一枚の画像当たりの補助情報を時間的に短縮する。

【 0 0 4 0 】

ここに、第 3 の音声データ t s 3 として抽出される画像群の特徴音とは、画像群に含まれる各画像に係る全ての第 1 の音声データ t s 1 の中の特徴的な部分である。画像群の特徴音としては、例えば、画像群に含まれる各画像に係る複数の特徴的な第 2 の音声データ t s 2 の内の、最も特徴的な 1 つの第 2 の音声データ t s 2 が挙げられる。なお、「最も特徴的な 1 つの第 2 の音声データ t s 2」とは、例えば、複数の第 2 の音声データ t s 2 の音圧の各最高値の内の、最も高い値を示す第 2 の音声データ t s 2、あるいは音声認識

10

20

30

40

50

を最も高い精度で行うことができる第2の音声データ $t s 2$ 、などである。もしくは簡単に、画像群に含まれる複数の画像中の、撮影時刻が最も古い（あるいは撮影時刻が最も新しい、さらには最も古い撮影時刻と最も新しい撮影時刻との中間時刻に最も近い、等の）画像の第2の音声データ $t s 2$ を第3の音声データ $t s 3$ として設定しても構わない。

【0041】

第3の音声データ $t s 3$ として抽出される画像群の特徴音の他の例としては、画像群に含まれる各画像に係る複数の第1の音声データ $t s 1$ において、共通に含まれている音声成分（音波の性質（音圧、周波数など）がほぼ同一の音声部分）が挙げられる。ここに「共通に含まれている音声成分」とは、例えば、複数の第2の音声データ $t s 2$ に共通して波の音が含まれている場合に、何れか1つの第2の音声データ $t s 2$ から抽出した波の音の音声成分、あるいは複数の第2の音声データ $t s 2$ の全てから抽出した波の音を平均化した音声成分、などである。

【0042】

なお、第3の音声データ $t s 3$ の抽出方法は、上記に挙げた例に限定されるものではなく、その他の技術を適宜用いることができる。例えば、ステレオ式に複数の音声を複数のスピーカから同時に再生して、コラージュ風の表現をしても良い。また、1つのスピーカからモノラルで音声を再生しても良く、旅行などでは、その土地を一番良く表す音声を選んでも良い。このときに、GPS情報などを補助的に利用することが可能である。例えば、GPS情報に基づき海の近くであることが分かる場合には、録音した音声データに含まれる海の音の音声成分を、その他の音声成分よりも優先させて音量を上げる、等である。また、季節の情報なども有効に利用することが可能である。例えば、季節の情報が夏である場合には、録音した音声データに含まれる特徴的な音声としての蝉の鳴き声の音声成分を、その他の音声成分よりも優先させて音量を上げる、等である。こうした季節や土地に対して、どのような音声が特徴的となるかを示す優先度情報を表形式でもつようにして、この表を参照して優先させる音声成分を判断するようにしても良い。

【0043】

加えて、要約部10cは、画像の撮影に伴って得られた撮影時情報、つまり、撮影で得られた画像自体（具体的には、画像に基づき顔判定部10dにより検出された顔、あるいは画像を解析して得られる画像の特徴など）、第1の音声データ $t s 1$ 自体（具体的には、第1の音声データ $t s 1$ から抽出される環境音やテキスト化された音声など）、GPS部7から取得したGPS位置情報、WEB通信部6から受信した地図情報や天気情報、時計8から取得した時間情報、温度計9から取得した温度情報等（列挙したこれらの例に限定されるものではなく、その他の情報を用いても勿論構わない）に基づいて、画像の特徴を表すタグ情報（このタグ情報は、画像の特徴を表し、下記に説明するように、例えば画像の説明文を作成するために用いられる）を、画像に関連付け可能な文字情報として生成する（タグ情報生成の処理の一例については、後で図11Aおよび図11Bを参照して説明する）。

【0044】

その後、要約部10cは、必要に応じて、データベース部15の文章用テンプレート15aに保存されている複数の文章テンプレート（空欄に言葉を詰め込むことにより文章を構成するためのテンプレート）の中からタグ情報に応じて適切な1つの文章テンプレートを読み込み、読み込んだ文章テンプレートの空欄個所にタグ情報を詰め込むことにより、画像の説明文を生成する。一例を挙げれば、次の文章テンプレート、「（X1）月（X2）日、（X3）と（X4）に行った。その日は（X5）日だった。」を用いて、時間情報に基づき「X1 = 5」および「X2 = 5」を、顔情報に基づき「X3 = 太郎さん」を、GPS位置情報および地図情報に基づき「X4 = 京都」を、天気情報（あるいは天気情報および温度情報）に基づき「X5 = 晴れていてとても暑い」を、それぞれ詰め込み、次の画像の説明文を作成する。「5月5日、太郎さんと京都に行った。その日は晴れていてとても暑い日だった。」このようにして要約部10cにより生成された画像の説明文は、要約

10

20

30

40

50

部 1 0 c の制御に基づき後述する音声合成部 1 1 d により音声生成されて第 1 の要約である第 4 の音声データ t s 4 に変換される。

【 0 0 4 5 】

なお、要約部 1 0 c による画像の説明文の作成は、画像が撮像された時点で行うに限るものではなく、画像を鑑賞する時点で行っても良い。画像を鑑賞する時点で画像の説明文を作成すれば、鑑賞する時点と撮像時点との時間差を考慮した文章を作成することが可能になる利点がある。一例としては、撮像時点が 1 年前であれば「昨年」という文言を選択することができ、2 年前であれば「一昨年」という文言を選択することができる、等である。

【 0 0 4 6 】

そして、これらの音声データ t s 1 ~ t s 4 の例をまとめたのが図 3 である。ここに、図 3 は画像再生時に再生される音声データの例を示す図表である。

【 0 0 4 7 】

顔判定部 1 0 d は、画像に基づき顔を検出するものである。すなわち、顔判定部 1 0 d は、画像処理部 2 を介して撮像部 1 から得られた画像データから、人物の顔部分を抽出し、抽出した顔の特徴点データから、データベース部 1 5 の後述する顔データベース 1 5 b に既に登録済みの顔であると判断される場合（つまり、抽出した顔の特徴点データと同一と判定される顔の特徴点データが、人物名等の人物情報に既に関連付けされている場合）には、関連付けられた人物情報を読み出して撮像して得られた画像データに関連付けする。また、顔判定部 1 0 d は、必要に応じて、顔の状態（例えば、笑顔であるか否か等）につ

10

20

【 0 0 4 8 】

なお、記録部 1 4 には、画像データが記録されると共に、画像データに関連付けて第 1 の音声データ t s 1 とタグ情報とが記録される。また、要約である第 2 の音声データ t s 2 も画像データに関連付けて記録部 1 4 に記録されても良いが、第 2 の音声データ t s 2 自体に代えて、第 1 の音声データ t s 1 中の特定範囲を示すポインタ等が記録されても構わない。さらに、画像群に係る第 3 の音声データ t s 3 が画像群に関連付けて記録部 1 4 に記録されても良いが、例えば旅行に係る画像群はその旅行が終了すれば確定するのに対して、特定の人物が写っている画像群はその特定の人物が含まれる画像が将来撮影されることもあり得るために、現時点の画像群が既に確定したものであるとはいえない。従って、記録部 1 4 に第 3 の音声データ t s 3 が記録されている場合であっても、画像群が更新された場合には第 3 の音声データ t s 3 も更新すると良い。あるいは、第 3 の音声データ t s 3 は、必要になる毎に画像群に含まれる各画像に係る複数の第 1 の音声データ t s 1 から作成しても良い。

30

【 0 0 4 9 】

続いて、再生制御部 1 1 は、全画面表示部 1 1 a と、サムネイル表示部 1 1 b と、フロー表示部 1 1 c と、音声合成部 1 1 d と、テロップ作成部 1 1 e と、を備えている。

【 0 0 5 0 】

全画面表示部 1 1 a は、画像の全画面表示を行うための表示用データを作成する。

【 0 0 5 1 】

サムネイル表示部 1 1 b は、画像のサムネイル表示を行うための表示用データを作成する。

40

【 0 0 5 2 】

フロー表示部 1 1 c は、画像のフロー表示を行うための表示用データを作成する。

【 0 0 5 3 】

音声合成部 1 1 d は、テキスト化部 1 0 b により作成されたテキスト、または要約部 1 0 c により作成されたタグ情報（このタグ情報もテキストである）に基づき音声合成を行って、読み上げ用の音声データを作成する。さらに、音声合成部 1 1 d は、要約部 1 0 c により作成された画像の説明文に基づき音声合成を行って、画像の説明文に係る第 4 の音声データ t s 4 を作成する。

50

【 0 0 5 4 】

テロップ作成部 1 1 e は、上述したテキスト化部 1 0 b により作成されたテキスト、要約部 1 0 c により作成されたタグ情報、または要約部 1 0 c により作成された画像の説明文に基づき、フォントデータを用いて表示用のテロップデータを作成する。

【 0 0 5 5 】

そして、データベース部 1 5 は、文章用テンプレート 1 5 a と、顔データベース 1 5 b と、を備えている。

【 0 0 5 6 】

文章用テンプレート 1 5 a は、上述したような、タグ情報を嵌め込む文章テンプレートを保持するものである。文章テンプレートは、撮影シーンに合わせて各種が予め用意されているが、WEB通信部 6 を介してインターネット等から新たな文章テンプレートや所望の文章テンプレートをダウンロードするようにしても構わない。このとき、インターネットへ接続するタイミング等は、所望のタイミングであっても良いし、画像をインターネットへアップロードするタイミングであっても構わない。なお、文章用テンプレート 1 5 a には、テキスト化部 1 0 b により用いられる擬音テキストデータベースや、画像の撮影時刻と再生時刻との時間差を表す言葉を収納する時差テンプレートなどが付帯して設けられている（ただし、擬音テキストデータベースや時差テンプレートを文章用テンプレート 1 5 a とは別体に設けても構わない）。

10

【 0 0 5 7 】

顔データベース 1 5 b は、顔（具体的には、顔の特徴点データ）と人物名等の人物情報とを関連付けて記憶するデータベースである。ここに、顔の特徴点データは画像から顔判定部 1 0 d により抽出され、人物情報は例えばユーザが入力する。

20

【 0 0 5 8 】

次に、図 8 は、カメラのメイン処理を示すフローチャートである。

【 0 0 5 9 】

カメラの電源スイッチがオンされる等によりこの処理が開始され、まず、カメラが撮影モードに設定されているか否かを判定する（ステップ S 1 ）。

【 0 0 6 0 】

ここで撮影モードに設定されている場合には、後で図 9 を参照して説明する撮影モードの処理を実行する（ステップ S 2 ）。

30

【 0 0 6 1 】

また、ステップ S 1 において撮影モードに設定されていないと判定された場合には、カメラが再生モードに設定されているか否かを判定する（ステップ S 3 ）。

【 0 0 6 2 】

ここで再生モードに設定されている場合には、後で図 1 0 を参照して説明する再生モードの処理を実行する（ステップ S 4 ）。

【 0 0 6 3 】

また、ステップ S 3 において再生モードに設定されていないと判定された場合には、カメラが通信モードに設定されていると判定して、画像通信の処理を実行する（ステップ S 5 ）。この画像通信の処理は、WEB通信部 6 等を介して画像をパーソナルコンピュータへ送信したり、あるいは画像をインターネットへアップロードする処理を含み、公知の技術を広く適用可能であるためにここでは詳細には説明しない。

40

【 0 0 6 4 】

上述したステップ S 2 、ステップ S 4 、またはステップ S 5 の処理を行ったら、このメイン処理を終了するか否かを判定する（ステップ S 6 ）。ここにメイン処理の終了は、例えば、電源スイッチがオフに操作された場合、あるいは何の操作もなされていない時間が自動電源オフ設定時間（あるいはスリープ設定時間）に達した場合などに実行されるようになっている。

【 0 0 6 5 】

ここで、メイン処理をまだ終了しない場合にはステップ S 1 へ戻って上述したような処

50

理を繰り返して行い、終了する場合にはメイン処理を終える。

【0066】

続いて、図9は、撮影モードの処理を示すフローチャートである。図8に示したステップS2に入ると、この撮影モード処理が開始される。

【0067】

まず、カメラがGPS位置情報を取得するモードに設定されていて、GPS部7がオンになっているか否かを判定する(ステップS11)。

【0068】

ここで、GPS部7がオンになっている場合には、GPS部7によりGPS位置情報を取得する(ステップS12)。

10

【0069】

次に、撮像部1により撮像を開始し、画像処理部2や再生制御部11等により処理を行って表示パネル12にスルー画を表示開始する(ステップS13)と共に、マイク3により録音を開始する(ステップS14)。ここにスルー画は、構図等を決定する際にユーザが観察することができるよう表示パネル12に表示されるリアルタイムの動画であり、例えば60fpsのフレームレートで撮影される。なお、例えば60fpsのフレームレートで撮像部1の全画素を読み出すことは困難であったり消費電力を要したりするために、例えば間引き読み出しや加算読み出し等が行われ、静止画よりも画素数が少なくなっている。従って、図2においては、記録画となる静止画よりも小さい画像として図示している。

20

【0070】

そして、スルー画における顔判定を行うモードに設定されているか否かを判定する(ステップS15)。

【0071】

ここで、スルー画における顔判定を行うモードに設定されている場合には、顔判定部10dにより人物の顔部分を抽出して、スルー画における人物の顔部分に例えば四角の枠を表示する等の顔判定表示を行う(ステップS16)。

【0072】

そして、認証可能であるか否か、つまり顔判定の対象となる人物のデータが顔データベース15bに既に登録されているか否かを判定する(ステップS17)。

30

【0073】

ここで認証可能である場合には、顔認証処理として、スルー画における人物の顔部分に人物のデータを関連付ける(ステップS18)。

【0074】

ステップS15においてスルー画の顔判定を行うモードに設定されていないと判定された場合、ステップS17において認証不可能であると判定された場合、またはステップS18における顔認証処理が終了した場合には、静止画撮影を指示するリリース操作(なお、カメラにおいては2段押圧式のリリーススイッチにより操作が行われることが多いために、この場合にはセカンド(2nd)リリース操作)が行われたか否かを判定する(ステップS19)。

40

【0075】

ここでリリース操作が行われていない場合には、ステップS14において録音を開始した音声データの内の、最新の前録時間(上記例では5秒)分の音声データ以前の部分をクリア(削除)して(ステップS20)、ステップS13に戻り、スルー画の表示と、前録時間分の音声データの記録と、必要に応じた顔判定と、を継続して行いながら、リリース操作が行われるのを待機する。

【0076】

そして、ステップS19においてリリース操作が行われたと判定された場合には、図2にも示すように、記録画としての静止画を撮影する(ステップS21)。

【0077】

50

静止画撮影後も、後録時間（上記例では5秒）が経過するまでは音声データの録音を行い、後録時間が経過した時点で録音を終了する（ステップS22）。こうして録音された、リリース時点を含む前録時間および後録時間の音声データが第1の音声データts1である。

【0078】

続いて、要約部10cが、第1の音声データts1から第2の音声データts2を上述したように作成する（ステップS23）。

【0079】

さらに、要約部10cが、後で図11Aおよび図11Bを参照して説明する3W+1H要約テキスト化の処理を実行して、各種のタグ情報を生成する（ステップS24）。ここに「3W+1H」とは、タグ情報の幾つかの例としての、いつ（WHEN）、どこで（WHERE）、誰が（WHO）、どのように（HOW）、を意味している。

【0080】

そして、撮影された画像を記録部14に記録すると共に、画像に関連付けて第1の音声データts1とタグ情報、あるいはさらに第2の音声データts2等が記録される（ステップS25）。

【0081】

このステップS25の処理を終えたら、この撮影モードの処理から図8に示すメイン処理に復帰する。

【0082】

次に、図10は、再生モードの処理を示すフローチャートである。図8に示したステップS4に入ると、この再生モード処理が開始される。

【0083】

まず、記録部14に記録されている全画像を、上述したように画像群に分類する（ステップS31）。記録部14に記録されている全画像が画像群に全く分類されていない場合には、全画像の分類を行うが、再生モード処理を既に1回以上行っている場合には、前回再生モード処理を行って以降に撮影された画像のみを画像群に分類（つまり、既存の画像群への追加、または新規の画像群の生成を）すれば良い。これにより、記録部14に記録されている全画像の、現時点での画像群への分類が行われたことになるために、ここでは更新が必要な第3の音声データts3を要約部10cが作成して、記録部14へ記録する処理も行う。

【0084】

続いて、再生モードが、第2の表示モードであるサムネイル表示モードに設定されているか否かを判定する（ステップS32）。

【0085】

ここでサムネイル表示モードに設定されていない場合には、全画面表示モードに設定されていると判定して、後で図12を参照して説明する全画面表示の処理を行う（ステップS33）。

【0086】

また、ステップS32においてサムネイル表示モードに設定されていると判定された場合には、サムネイル表示部11bがサムネイル表示用の画像データを作成し、サムネイル表示の処理を行う（ステップS34）。このサムネイル表示は、デジタルカメラ等において広範に利用されているためにここでは詳細な説明を省略するが、図6に示すように、表示パネル12の画面12a全体に複数のサムネイル画像psを同じ大きさに配列して表示する（従って、複数のサムネイル画像psが同時に表示される）ものである。ここに、図6はサムネイル表示モードの様子を示す図である。

【0087】

なお、ここではサムネイル表示を、例えば画像群毎に区切って（つまり、画面に表示される複数のサムネイル画像が同一の画像群に属するように）行うものとする。具体的に、第1の画像群に属する画像が20枚、第2の画像群に属する画像が10枚で、12枚のサ

10

20

30

40

50

ムネイル画像を配列してサムネイル表示する場合を考えると、まず、第1の画像群のサムネイル画像12枚を表示し、次に第1の画像群の残りのサムネイル画像8枚を表示し、その後第2の画像群のサムネイル画像10枚を表示する、等である。ただし、このような表示例に限定されるものではなく、第1の画像群の残りのサムネイル画像8枚と第2の画像群のサムネイル画像4枚とを1つの画面にサムネイル表示しても良いし、その他の種々の表示方法を適宜利用しても構わない。

【0088】

サムネイル表示を行っているときに、フロー表示に移行する操作がなされたか否かを監視している(ステップS35)。

【0089】

ここで、フロー表示に移行する操作がなされた場合には、フロー表示を行う際の基準画像として、撮影時刻が最新となる画像を設定する(ステップS36)。

【0090】

そして、後で図13を参照して説明するフロー表示の処理を行う(ステップS37)。

【0091】

ステップS37のフロー表示が終了したら、フロー表示から全画面表示へ移行する操作がなされたか否かを判定し(ステップS38)、操作がなされた場合にはステップS33の全画面表示の処理へ移行する。

【0092】

一方、ステップS35において、フロー表示に移行する操作がなされていないと判定された場合には、配列して表示されている複数のサムネイル画像psの内の1つが選択されて着目画像となったか否かを判定する(ステップS39)。

【0093】

ここで、何れのサムネイル画像psも選択されていない場合には、サムネイル画像psが配列されていない周辺の余白部分に、タッチパネル5を介したタッチ操作がなされたか否かを判定する(ステップS40)。

【0094】

ここで、周辺の余白部分にタッチ操作がなされた場合(従って、サムネイル表示において着目画像が存在しない場合)には、再生制御部11の制御により、サムネイル表示されている画像群の第3の音声データts3をスピーカ13から音声再生する(ステップS41)。

【0095】

このステップS41の処理を開始した後、または、ステップS40において周辺の余白部分にタッチ操作がなされていないと判定された場合には、ステップS35へ戻ってフロー表示への移行を再び判定する。

【0096】

また、ステップS39において、サムネイル画像psの内の1つ(ひいては、サムネイル画像psにより表される1つの画像)がシングルタッチ(あるいはシングルタップ)により選択されたと判定された場合には、選択された着目画像の第2の音声データts2をスピーカ13から音声再生する(ステップS42)。また、サムネイル画像psの内の1つがダブルタッチ(あるいはシングルタップ)により選択された場合には、カメラが全画面表示モードに設定されるために、ステップS42の処理は実質的にスキップされ、ステップS43、ステップS45の分岐を経て、ステップS32の分岐を「NO」へ移行し、ステップS33の全画面表示の処理を行うことになる。

【0097】

このステップS42の処理を開始した後に、次のサムネイル表示候補を選択する操作(つまり、サムネイル表示を次頁へ進める操作)がなされたか否かを判定する(ステップS43)。

【0098】

ここで、操作がなされた場合には、現在表示中の画像群の中に未表示の画像があるとき

10

20

30

40

50

にはその画像の選択を行い、現在表示中の画像群の全てがサムネイル表示済みのときには次の画像群から画像の選択を行う（ステップS44）。このステップS44の処理を行ったら、ステップS34へ戻って選択した画像をサムネイル表示する。

【0099】

一方、ステップS43において、次のサムネイル表示候補を選択する操作がなされていないと判定された場合には、表示を終了するか否かを判定する（ステップS45）。

【0100】

ここで、表示を終了しない場合にはステップS32へ戻る。また、表示を終了する場合、あるいはステップS33の処理を終了した場合には、この再生モードの処理から図8に示すメイン処理に復帰する。

【0101】

なお、上述では、サムネイル表示を行っただけでは音声は再生されず、サムネイル表示における周辺の余白部分にタッチ操作がなされたときに画像群に係る第3の音声データts3を再生するようにしているが、これに代えて、サムネイル表示を行っただけで第3の音声データts3が自動的に再生されるように構成しても構わない。

【0102】

ここで、サムネイル表示モードは、複数の画像を縮小して並べて表示するモードであり、少ない枚数（1～2枚）ごとに画像を鑑賞するモード（第1の表示モードであり、例えば全画面表示モード）とは異なる第2の表示モードの1種である。そして、サムネイル画像は、このサムネイル表示モードにおいて並べて表示される縮小された画像のことである。第2の表示モードとしては、グループで（少なくない枚数の画像が）表示されれば良く、縮小して並べて表示するに限らず、縮小することなく重ねて表示する表示方法を取っても良い。そして、複数の中から特定画像を選択するのにふさわしい表示方法であるという。

【0103】

次に、図11Aおよび図11Bを参照して、図9のステップS24の処理の詳細を説明する。ここに、図11Aは3W+1H要約テキスト化の処理の一部を示すフローチャート、図11Bは3W+1H要約テキスト化の処理の他の一部を示すフローチャートである。

【0104】

この3W+1H要約テキスト化の処理を開始すると、まず、画像に関連して録音された第1の音声データts1の中から、音声分析部10aにより環境音が検出されたか否かを判定する（ステップS51）。

【0105】

ここで環境音を検出された場合には、要約部10cが、文章用テンプレート15aに付帯して設けられた擬音テキストデータベースから、検出された環境音に対応する擬音テキストを選択して、HOWのタグ情報として設定する（ステップS52）。タグ情報の幾つかの具体例を挙げれば、波→「ザザー」、さざ波→「ちゃぶちゃぶ」、風音→「そよそよ」、破裂音→「バーン」、呼びかける声→テキスト化部10bによるテキスト化、等である。なお、これらのタグ情報は、テキストとして再生するに限るものではなく、例えばアイコン化して（つまり図として）表示再生しても良い。これにより、聴覚に自信のないユーザや、聞き取りが困難な騒音環境下にいるユーザでも、タグ情報をより容易に認識することが可能となる。

【0106】

ステップS51において環境音を検出されないと判定された場合、またはステップS52の処理を行った場合には、次に、顔判定部10dにより記録画像である静止画像中に顔部分が検出されたか否かを判定する（ステップS53）。

【0107】

ここで顔部分が検出された場合には、検出された顔部分の数が所定数以上であるか否かを判定する（ステップS54）。そして所定数以上である場合には、要約部10cは、W H Oのタグ情報として例えば「みんな一緒」を設定する（ステップS55）。

10

20

30

40

50

【 0 1 0 8 】

また、検出された顔部分の数が所定数未満である場合、またはステップ S 5 4 の処理を行った場合には、検出された顔部分の認証が可能であるか否かを判定する（ステップ S 5 6）。

【 0 1 0 9 】

ここで認証可能である場合には、要約部 1 0 c は、顔判定部 1 0 d により検出された顔に基づきデータベース部 1 5 の顔データベース 1 5 b から人物情報を取得して、顔が検出された画像の W H O のタグ情報として認証された人物情報、例えば「○○さん」を設定する（ステップ S 5 7）。なお、ステップ S 5 5 において既に W H O のタグ情報を設定している場合であって、このステップ S 5 7 の処理をさらに行った場合には、W H O のタグ情報が追記され、つまり 1 つの種類のタグに複数のタグ情報が保存されることになる。このように、タグ情報は 1 種類に対して 1 つ設定するに限るものではなく、複数を列記しても構わない。

10

【 0 1 1 0 】

また、ステップ S 5 6 において認証不可能であると判定された場合には、図 9 のステップ S 1 8 における静止画像撮影前後の動画像における認証結果を利用可能であるか否かを判定する（ステップ S 5 8）。

【 0 1 1 1 】

ここで利用可能である場合には、ステップ S 5 7 へ行って上述したように W H O のタグ情報を設定する。

20

【 0 1 1 2 】

ステップ S 5 3 において顔部分が検出されないと判定された場合、ステップ S 5 8 においてステップ S 1 8 の認証結果を利用不可能であると判定された場合、またはステップ S 5 7 の処理を行った場合には、記録画像である静止画像の画面上方に特徴が検出されたか否かを判定する（ステップ S 5 9）。ここに、カメラには図示しない重力センサ等が設けられていて、静止画像には重力方向上側の情報が付随して記録されていることを想定している。

【 0 1 1 3 】

そして、画面上方に特徴が検出された場合には、検出された特徴に対応するキーワードを W H E R E のタグ情報として設定する（ステップ S 6 0）。画面上方の特徴に応じたタグ情報の幾つかの具体例を挙げれば、青い→「青空の下」、暗い→「夜空の下」、人工光→「室内で」、等である。なお、これらのキーワードは、データベース部 1 5 にキーワードテンプレートとして予め用意しておいても構わないし、ユーザが入力しても良いし、W E B 通信部 6 を介してインターネット等からダウンロードしても構わない。

30

【 0 1 1 4 】

ステップ S 5 9 において画面上方に特徴が検出されないと判定された場合、またはステップ S 6 0 の処理を行った場合には、G P S 位置情報および地図情報を取得可能であるか否かを判定する（ステップ S 6 1）。

【 0 1 1 5 】

ここで取得可能である場合には、取得された G P S 位置情報および地図情報に基づき、W H E R E のタグ情報として、例えば「東京」などの地名やその他の地理情報を設定する（ステップ S 6 2）。なお、上述と同様に、ステップ S 6 0 において設定された W H E R E のタグ情報が存在する場合には、ステップ S 6 2 において設定した W H E R E のタグ情報が列記されることになる。

40

【 0 1 1 6 】

ステップ S 6 1 において G P S 位置情報または地図情報が取得できないと判定された場合、またはステップ S 6 2 の処理を行った場合には、静止画像の撮影時刻の情報を取得可能であるか否かを判定する（ステップ S 6 3）。

【 0 1 1 7 】

一般的なカメラ等であれば静止画像には撮影時刻の情報が付随しているために取得可能

50

であり、取得した撮影時刻に基づいて、W H E Nのタグ情報を設定する（ステップS 6 4）。タグ情報の幾つかの具体例を挙げれば、撮影時刻→「年月日時分」、月日→「春」「夏」「秋」「冬」、特別な日→「誕生日」「クリスマス」、時分→「朝」「夜」、等である。なお、例示したこれらのタグ情報は、上述したように、複数を列記しても構わない。

【0 1 1 8】

ステップS 6 3において何らかの理由により撮影時刻が取得できないと判定された場合、またはステップS 6 4の処理を行った場合には、W H E N、W H E R E、W H O、H O Wの各タグ情報を画像と関連付けて（例えば画像ファイルのヘッダ情報などとして）記録部1 4に記録し（ステップS 6 5）、この3 W + 1 H要約テキスト化の処理から図9に示す撮影モードの処理に復帰する。

10

【0 1 1 9】

なお、上述では、静止画像の撮影時の状況を表すタグ情報として、3 W + 1 Hのタグ情報を用いたが、その他のタグ情報を適宜取捨選択しても構わないことは勿論である。

【0 1 2 0】

続いて、図1 2を参照して、図1 0のステップS 3 3の処理の詳細を説明する。ここに、図1 2は全画面表示の処理を示すフローチャートである。

【0 1 2 1】

この全画面表示は、必ずしも文字通りに画像を全画面に拡大して表示する必要はなく、余白付きの画像表示であっても良く、さらにこの余白に情報表示があっても構わず、加えて余白の情報表示が次の画像の予告であっても良い。

20

【0 1 2 2】

この全画面表示の処理を開始すると、まず、全画面表示部1 1 aが選択画像を全画面表示するための表示用データを作成し、図5に示すように全画面表示を行う（ステップS 7 1）。ここに図5は、全画面表示モードの様子を示す図である。この全画面表示モードにおいては、表示パネル1 2の画面1 2 aのほぼ全体に（つまり画像を主体として）1枚の選択画像pのみを表示する。ただし、画面1 2 aの全面積を用いて1つの選択画像pを表示するに限らず、画面1 2 aの中央に1つの選択画像pを比較的大きな面積で表示して、その周辺に各種の情報を表示する等の表示形態であっても構わない。

【0 1 2 3】

次に、画像の説明文に係る第4の音声データt s 4を音声再生する設定がなされているか否かを判定する（ステップS 7 2）。

30

【0 1 2 4】

ここで、第4の音声データt s 4を音声再生する場合には、画像の撮像時点の前後に録音された第1の音声データt s 1と、第4の音声データt s 4と、を例えば同時に（あるいは順次に）音声再生する（ステップS 7 3）。

【0 1 2 5】

例えば同時に音声再生する場合には、撮影時の第1の音声データt s 1を背景音として説明文に係る第4の音声データt s 4をナレーションとする感覚であり、このようなバランスをより明瞭にするために、第4の音声データt s 4の再生音量を第1の音声データt s 1の再生音量よりも大きくするようにしても良い。画像の説明文に係る第4の音声データt s 4が音声合成部1 1 dにより作成されるのは、上述した通りである。

40

【0 1 2 6】

そして、同時音声再生の場合には、第4の音声データt s 4の再生時間が、第1の音声データt s 1の再生時間（ひいては第1の音声データt s 1の録音時間：上述した例では1 0秒）以下となるようにすると良い。このためには、要約部1 0 cが画像の説明文を作成する際に、読み上げ時間長さが第1の音声データt s 1の再生時間以下となるような文章テンプレートを文章用テンプレート1 5 aから選択して読み込むようにすると良い。従って、各文章テンプレートに、標準的な読み上げ時間等の情報を予め関連付けておくようにすると良い。

【0 1 2 7】

50

また、ステップ S 7 2 において第 4 の音声データ t s 4 を音声再生しないと判定された場合には、第 1 の音声データ t s 1 を音声再生すると共に、第 4 の音声データ t s 4 を作成する基となった、要約部 1 0 c により作成された画像の説明文を、選択画像 p の表示に重畳して例えばテロップ表示再生する（ステップ S 7 4）。このときに用いる表示用のテロップデータがテロップ作成部 1 1 e により作成されるのは、上述した通りである。なお、ここでは画像の説明文のテロップ再生を行ったが、テロップ再生は行わずに第 1 の音声データ t s 1 のみの音声再生を行うようにしても構わない。

【 0 1 2 8 】

また、上述では第 1 の音声データ t s 1 を必ず再生しているが、第 1 の音声データ t s 1 のみの音声再生に代えて、第 4 の音声データ t s 4 のみの音声再生を行っても構わない。

10

【 0 1 2 9 】

こうして、ステップ S 7 3 またはステップ S 7 4 の処理が行われたら、音声再生を終了するか否かを判定する（ステップ S 7 5）。この判定は、全画面表示における音声再生が、繰り返し再生として設定されているか、あるいは 1 回再生として設定されているかに基づき行われる。そして、ここで音声再生を終了しない場合には、ステップ S 7 2 へ戻って、音声再生を繰り返して行う。

【 0 1 3 0 】

一方、ステップ S 7 5 において音声再生を終了すると判定された場合には、全画面表示する画像を次の画像に進める操作が行われたか否かを判定する（ステップ S 7 6）。

20

【 0 1 3 1 】

そして、次の画像に進める操作が行われた場合には、ステップ S 7 1 へ戻って、次の画像について上述したような音声再生を伴う全画面表示を行う。

【 0 1 3 2 】

また、ステップ S 7 6 において次の画像に進める操作が行われていないと判定された場合には、この全画面表示の処理から、図 1 0 に示す再生モードの処理に復帰する。

【 0 1 3 3 】

次に、図 1 3 を参照して、図 1 0 のステップ S 3 7 の処理の詳細を説明する。ここに、図 1 3 はフロー表示の処理を示すフローチャートである。

【 0 1 3 4 】

この処理に入ると、ステップ S 3 6 において設定した最新画像を規準としてフロー表示部 1 1 c がフロー表示用データ作成し、例えば図 7 に示すようなフロー表示を開始する（ステップ S 8 1）。ここに図 7 は、フロー表示モードの様子を示す図である。

30

【 0 1 3 5 】

フロー表示モードは、表示パネル 1 2 の画面 1 2 a 内に複数の縮小画像 p r を撮影時刻順に配置して表示し、時間軸方向（図 7 に示す例では、時間軸方向が画面 1 2 a の左右方向であることを想定している）の操作入力に応じて表示を移動させる第 2 の表示モードである。

【 0 1 3 6 】

本実施形態のフロー表示モードは、さらに、ある画像がどの画像群 P G に含まれるのかが見ただけで容易に判別されるように、画像群 P G に含まれる画像同士を近接させて（例えば、画像群 P G に含まれる任意の画像は、同一の画像群 P G に含まれる他の少なくとも 1 つの画像と一部が重複するように）縮小画像 p r として表示し、さらに、一の画像群 P G と他の画像群 P G とは所定の距離を離すように（重複部分がなく、時間軸方向において離隔するように）表示するようにしている。従って、このフロー表示モードにおいて利用される画像分類は、撮影時刻に沿った画像群への画像の分類である。

40

【 0 1 3 7 】

図 7 に示す例においては、画面 1 2 a の左側により古い画像が表示され、画面 1 2 a の右側により新しい画像が表示されるものとする。また、この図 7 の例では、縮小画像 p r の大きさも、例えば大小のサイズがランダムに混在するように表示されている（ただし、

50

大きさをランダムにしなくても勿論構わない)。

【0138】

そして、フロー表示モードにおいては、フリック入力 of 操作 (タッチパネル 5 を指先等でスライドしたり、パッと払ったりする入力操作) が行われると、その操作が時間軸方向における未来方向か過去方向かに応じて、表示パネル 12 に表示する画像をスライドさせる。

【0139】

すなわち、まずフリック入力により、フロー表示の流れを停止させる操作が行われたか否かを判定する (ステップ S 82)。

【0140】

ここでフロー表示の流れを停止させる操作が行われていない場合には、画面を左側へ移動させるような左向きのフリック入力の操作が行われたか否かを判定する (ステップ S 83)。

【0141】

左向きのフリック入力の操作が行われた場合には、画面 12 a 内に新しく表示される画像をリサイズして縮小画像 p r として、時間進行方向へフロー移動を行う (ステップ S 84)。これにより、画面 12 a 内の右側に、撮影時刻がより後の (より新しい) 画像が新たに表示される。

【0142】

また、ステップ S 83 において画面を右側へ移動させるような右向きのフリック入力の操作が行われたと判定された場合には、画面 12 a 内に新しく表示される画像をリサイズして縮小画像 p r として、時間後退方向へフロー移動を行う (ステップ S 85)。これにより、画面 12 a 内の左側に、撮影時刻がより前の (より古い) 画像が新たに表示される。

【0143】

ステップ S 84 またはステップ S 85 の処理を行ったら、次に、時間軸方向の操作入力によるフロー表示の移動速度が所定値未満であるか否かを判定する (ステップ S 86)。なお、フロー表示の移動速度は、フリック入力の操作の仕方によって変化するようにになっている。

【0144】

ここで、移動速度が所定値未満である場合には、各画像毎の音声再生が可能であると判断して、時間軸方向 (図示の例では画面 12 a の左右方向) における画面 12 a の中央 C を通過している縮小画像 p r に係る画像の第 2 の音声データ t s 2 を音声再生する (ステップ S 87)。

【0145】

一方、ステップ S 86 において移動速度が所定値以上であると判定された場合には、各画像毎の音声再生が不可能 (あるいは困難) であると判断して、時間軸方向における画面 12 a の中央 C を通過している画像群 P G に係る第 3 の音声データ t s 3 を音声再生する (ステップ S 88)。このときにはもちろん、第 2 の音声データ t s 2 は音声再生されない。

【0146】

このように、フロー表示の移動速度が所定値よりも遅い場合には画像に係る第 2 の音声データ t s 2 が再生され、フロー表示の移動速度が所定値以上に速い場合には画像群 P G に係る第 3 の音声データ t s 3 が再生されるようになっている。このときの移動速度の区分は、例えば、画像群 P G に含まれる各縮小画像 p r が画面 12 a の中央 C を通過する縮小画像 p r 毎の時間の内の最短時間が、第 2 の音声データ t s 2 を再生するに要すると見込まれる典型時間以上である場合に移動速度が所定値未満であるとし、典型時間未満である場合に移動速度が所定値以上であるとするなどが考えられる。

【0147】

ステップ S 87 またはステップ S 88 の処理を行ったら、ステップ S 82 へ戻ってフロ

10

20

30

40

50

ー表示の流れを停止させる操作が行われたか否かを判定する。こうして、ステップS82において、フロー表示の流れを停止させる操作が行われたと判定された場合には、このフロー表示の処理から、図10に示す再生モードの処理に復帰する。

【0148】

上述したように、各表示モードが設定されたときに再生される音声データは、例えば図4に示すようになっていいる。ここに図4は、画像再生モードの幾つかの例を説明するための図表である。

【0149】

すなわち、全画面表示モードが設定されると、画像に係る音声データの内の、比較的長い再生時間を要すると考えられる、撮影時刻前後の第1の音声データts1と、画像の説明文を読み上げる第4の音声データts4と、の少なくとも一方が音声再生される。

10

【0150】

この全画面表示モードは、必ずしも全画面で表示する必要はなく、1枚あるいは2枚の少ない枚数の画像を比較的大きめに表示して鑑賞するものを想定している。このときさらに、画像表示部分の外に様々なアイコンや要約を表示することができるスペースがあっても良い。

【0151】

また、サムネイル表示モードが設定されると、特定の画像が選択されていないときには音声再生されないか、または画像群の特徴音声である第3の音声データts3が音声再生され、特定の画像が着目画像として選択されたときに、第1の音声データts1の中の特徴部分の第2の音声データts2が音声再生される。

20

【0152】

上述したように、サムネイル表示モードは、複数の画像を縮小して並べて表示するモードであり、少ない枚数(1~2枚)ごとに画像を鑑賞する第1の表示モード(例えば、叙述したような全画面表示モード)とは異なる第2の表示モードである。そして、第2の表示モードとしては、縮小して並べて表示するに限らないことも上述した通りである。

【0153】

さらに、フロー表示モードが設定されると、フロー表示の移動速度が遅いときには画像に係る第2の音声データts2が再生され、フロー表示の移動速度が速いときには画像群に係る第3の音声データts3が再生される。

30

【0154】

ここに、音声データts1, ts2, ts3, ts4の各音声再生時間Tts1, Tts2, Tts3, Tts4は、概略、次のような大小関係にあると考えられる。ここに、下記の不等式において用いた記号「~」は、近似した時間長さであることを示している。

$$Tts1 \sim Tts4 > Tts2 \sim Tts3$$

【0155】

ただし、画像群として考えたときに、第2の音声データts2は画像群に含まれる各画像毎に音声再生されるのに対して、第3の音声データts3は画像群全体で1つ音声再生されるのみであるために、1画像当たりの再生時間を考えたときには、次のような大小関係になると捉えることもできる。

40

$$Tts1 \sim Tts4 > Tts2 > Tts3$$

【0156】

そして、全画面表示モードのときに最も再生時間の長い音声データts1, ts4が音声再生されるのは、全画面表示モードが特定の1つの画像をじっくりと観察する際に利用される表示モードであると考えられるためである。

【0157】

また、1つの画面に複数の画像が表示されるサムネイル表示モードまたはフロー表示モードにおいては、1つの画像のみに対する音声再生を行うことがふさわしい場合には第2の音声データts2が音声再生され、そうでない場合には必要に応じて第3の音声データts3が音声再生される。

50

【 0 1 5 8 】

このような実施形態 1 によれば、画面に画像を拡大して表示する第 1 の表示モード（第 1 の再生モード）（例えば、画面に画像を 1 つだけ表示する全画面表示モード）が設定されているときには第 1 の要約を、画面に画像を縮小して複数表示する第 2 の表示モード（第 2 の再生モード）が設定されているときには第 1 の要約よりも画像単位での認識に要する時間が短い第 2 の要約を、作成して再生するようにしたために、画像撮影時の回想を、一枚毎だけでなく、画像群としても効果的に行うことが可能となる。

【 0 1 5 9 】

また、第 1 の要約を第 1 の音声データ $t s 1$ 、第 2 の要約を第 2 の音声データ $t s 2$ とした場合には、画像を観察すると共に音声を聴取することにより（つまり、視覚だけでなく聴覚を併用して）、画像撮影時の状況をよりリアルに認識することが可能となる。

10

【 0 1 6 0 】

さらに、サムネイル表示モードにおいては、着目画像となっているサムネイル画像に係る第 2 の音声データ $t s 2$ を音声再生するようにしたために、着目画像の撮影時の状況を簡潔に認識することが可能となる。

【 0 1 6 1 】

そして、サムネイル表示において前記着目画像が存在しないとき（例えば、周辺の余白部分にタッチ操作がなされたとき）に、第 3 の音声データ $t s 3$ を音声再生するようにしたために、サムネイル表示されている画像群の撮影時の状況を簡潔に認識することが可能となる。

20

【 0 1 6 2 】

一方、フロー表示モードにおいては、画面中央を通過する縮小画像に係る第 2 の音声データ $t s 2$ を音声再生するようにしたために、画面中央の縮小画像の撮影時の状況を簡潔に認識することが可能となる。

【 0 1 6 3 】

また、フロー表示の移動速度が所定値以上であるときには画像群に係る第 3 の音声データ $t s 3$ を音声再生するようにしたために、個々の画像に係る第 2 の音声データ $t s 2$ を音声再生するのが難しい場合でも、画面中央の画像群の撮影時の状況を簡潔に認識することが可能となる。

【 0 1 6 4 】

さらに、要約部 10 c が撮影時情報に基づきタグ情報を生成するようにしたために、タグ情報を用いて画像の特徴を明瞭に表すことが可能となる。このとき、文章テンプレートの空欄にタグ情報を嵌め込んで画像の説明文を生成する場合には、撮影時の状況認識を文章に基づき行うことが可能となる。

30

【 0 1 6 5 】

そして、画像の説明文を第 4 の音声データ $t s 4$ として音声再生する場合には、撮影時の状況を表す言葉を聴覚を用いて聞き取ることにより、撮影時の状況を明瞭に認識することが可能となる。ここに、第 4 の音声データ $t s 4$ の再生時間が第 1 の音声データ $t s 1$ の再生時間以下となるようにする場合には、第 1 の音声データ $t s 1$ の再生時間内において、第 1 の音声データ $t s 1$ を背景音とし第 4 の音声データ $t s 4$ をナレーションとする聴取が可能となる。

40

【 0 1 6 6 】

また、上述した再生装置は、撮像装置や録音装置として構成することも可能であるために、撮像装置や録音装置においても同様の効果を奏することができる。

[実施形態 2]

【 0 1 6 7 】

図 1 4 から図 1 8 は本発明の実施形態 2 を示したものであり、図 1 4 はタグ表示を伴う全画面表示の第 1 の例を示す図、図 1 5 はタグ表示を伴う全画面表示の第 2 の例を示す図、図 1 6 はタグ表示を伴う全画面表示の第 3 の例を示す図、図 1 7 はタグ表示を伴う組写真表示の様子を示す図、図 1 8 は再生モードの処理を示すフローチャートである。

50

【0168】

この実施形態2において、上述の実施形態1と同様である部分については同一の符号を付すなどして説明を適宜省略し、主として異なる点についてのみ説明する。

【0169】

本実施形態においては、全画面表示以外に、組写真表示を行う例について説明する。ただし、全画面表示および組写真表示に加えて、上述したサムネイル表示やフロー表示を行っても勿論構わない。また、本実施形態においては、音声再生を行う必要はなく、音声再生に代えて、あるいは音声再生と共に、要約部10cにより作成されたタグ情報を表示するようになっている。もちろん上述した実施形態1と同様に、音声でタグ情報を読み上げたり、タグ情報とテンプレートとを組み合わせで文章にして読み上げたりしても良い。

10

【0170】

また、上述した実施形態1においては、サムネイル表示として複数の画像を縮小して並べて表示するモードでの表示方法を詳しく説明したが、少ない枚数(1~2枚)ごとに画像を鑑賞する第1の表示モード(第1の再生モード)(例えば全画面表示モード)とは異なるモードであるという点で、本実施形態の組写真表示もサムネイル表示と同じ分類となり、上位概念としての第2の表示モード(第2の再生モード)であると考えることができる。なお、この組写真表示においても、グループで(少なくない枚数の画像が)表示されれば良く、縮小して並べて表示するに限らず、縮小することなく一部を重ねて表示する表示方法を取っても良い。また、複数の画像の中から一つの画像を選択することができるようにすれば、この組写真表示は検索用にも使えるモードである。そして組写真表示は、アルバム風のレイアウトをとる点が特徴となっている。

20

【0171】

まず、図14~図16は、タグ表示を伴う全画面表示の幾つかの例を示している。要約部10cは、全画面表示モードが設定されているときには、画面12aに表示される1つの画像pのタグ情報を第1の要約とする。つまり全画面表示においては、例えば、画像に関連付けられているタグ情報が全て(あるいは、1つの種類のタグ毎に代表的なタグ情報が1つ)表示再生されるような表示方法が採用されている。なお、タグ情報の列挙表示に代えて、要約部10cにおいて作成された画像の説明文をテロップ作成部11eにより表示用のテロップデータとして画像化し、第1の要約として表示再生しても構わない。ここでも全画面表示は、表示パネル12の画面12aの全表示面積を使う表示である必要はなく、少ない枚数の画像を一枚ずつ鑑賞することができる表示であれば良い。従って、単純化のためにここでは全画面表示モードと呼んで説明しているが、全画面表示をより広い概念化した第1の表示モードであれば良い。

30

【0172】

具体的に、図14においては、画像pに加えて、WHENのタグ情報として「去年」および「夏」が、WHEREのタグ情報として「伊豆」が、WHOのタグ情報として「Aちゃん」が、HOWのタグ情報として「ざざー」が、タグ表示12tとして表示されている。ただし、「去年」のタグ情報は、後で図18を参照して説明するように、画像の撮影時刻と再生時刻との時間差に基づき設定されたものである。

【0173】

また、図15においては、画像pに加えて、WHENのタグ情報として「去年」および「夏」が、WHEREのタグ情報として「伊豆」および「青空」が、HOWのタグ情報として「そよそよ」が、タグ表示12tとして表示されているが、人物が画像内にいないためにWHOのタグ情報は表示されていない。

40

【0174】

さらに、図16においては、画像pに加えて、WHENのタグ情報として「去年」および「夏」が、WHEREのタグ情報として「伊豆」および「山」が、タグ表示12tとして表示されているが、人物が画像内にいないためにWHOのタグ情報は表示されておらず、さらにHOWのタグ情報も設定されておらず表示されていない。

【0175】

50

次に、図 17 は、タグ表示を伴う組写真表示の例を示している。組写真表示モードは、表示パネル 12 の画面 12 a に、例えばユーザが選択した（あるいはカメラが自動選択した）複数枚（例えば 3 枚以上（ただし、全画面表示が 1 枚のみの表示である場合には、組写真表示は 2 枚以上であっても良い））の画像に係る縮小画像 p r が同時に表示される第 2 の表示モードである。この図 17 に示す例では、図 14 ~ 図 16 に示した画像 p の縮小画像 p r が適宜の大きさと同時に表示されている。

【0176】

要約部 10 c は、第 2 の表示モードである組写真表示モードが設定されているときには画面 12 a に表示される複数の縮小画像 p r のタグ情報における共通部分を重複させることなく第 2 の要約とする。すなわち、組写真においては、表示される複数枚の画像に共通するタグが表示される。なお、全画面表示において表示されたタグ情報に共通するタグ情報がない場合でも、画像に関連して記録されたタグ情報（つまり、全画面表示では未表示となっていたタグ情報も含む全タグ情報）中に共通するタグ情報が存在する場合には、そのタグ情報が表示される。このタグ情報には、被写体の姿勢情報や表情情報などを含んでも良く、タグ情報の変化を見れば、行動などを要約することができる。例えば、「座っている」タグが付いた画像と「立っている」タグが付いた画像が組写真中にあるときに、これらの画像の撮影時刻タグを含めて判定すれば、組写真に対する第 2 の要約として「座っている」タグと「立っている」タグとを並記するよりも、時系列順に、「座っている」→「立っている」となった場合にはタグ「立ち上がった」を、「立っている」→「座っている」となった場合にはタグ「座り込んだ」を、記載した方が分かり易く、かつ情報量も少なく要約されることになり好ましい。同じ赤ちゃんが立っている画像と座っている画像において、こうしたタグ情報が付されていると、愛らしい赤ちゃんの行動がリアルに回想される。

【0177】

また、要約部 10 c は、第 2 の表示モードである組写真表示モードが設定されているときには、画面 12 a に表示される複数の縮小画像 p r の各タグ情報の何れかに人物情報が含まれているときには、その人物情報を第 2 の要約にさらに含める。つまり、組写真に係る WHO のタグ情報は、（もちろん重複させることなく）全て表示するようになっている。これは、組写真の何れかに顔データベース 15 b に登録されている人物が写っている場合には、組写真表示された画像の全てに共通して写っていなくても観察者が容易に認識することができるためである。

【0178】

従って、図 17 に示す例においては、図 14 ~ 図 16 における共通するタグ情報として、WHEN のタグ情報「去年」および「夏」と、WHERE のタグ情報「伊豆」がタグ表示 12 t として表示され、さらに、図 14 ~ 図 16 における WHO のタグ情報「A ちゃん」がタグ表示 12 t として表示される。

【0179】

なお、図 14 ~ 図 17 に示したようなタグ情報は、文字として表示するだけでも構わないが、音声としての読み上げ再生を併用するようにしても良い。

【0180】

次に、図 18 を参照して、本実施形態における再生モードの処理について説明する。

【0181】

この再生モードの処理が開始されると、再生を行う現在時刻と、画像が撮影された時刻と、の時間差に基づいて、文章用テンプレート 15 a に付帯して設けられた時差テンプレートの中から、適切な言葉を選択する（ステップ S 91）。時差テンプレートには、例えば、「今年」、「去年」、「一昨年」、...等の言葉が予め記憶されているものとする。

【0182】

続いて、再生モードにおいて、組写真表示モードが設定されているか否かを判定する（ステップ S 92）。

【0183】

10

20

30

40

50

ここで、組写真表示モードが設定されている場合には、組写真表示部として機能する再生制御部 11 が組写真として表示する画像群を生成して再生し、要約部 10 c および再生制御部 11 が画像テキストくり表示、つまり共通するタグ情報をくくる表示を図 17 のタグ表示 12 t に示したように行う（ステップ S 93）。

【0184】

そして、組写真として表示する画像群を変更する操作が行われたか否かを判定し（ステップ S 94）、操作が行われた場合には画像群の変更処理を行う（ステップ S 95）。

【0185】

一方、ステップ S 92 において、組写真表示モードが設定されていない場合（つまり、全画面表示モードが設定されている場合）には、選択されている画像を再生して、選択画像に係るタグ情報を図 14 ~ 図 16 のタグ表示 12 t に示したようにテキスト表示する（ステップ S 96）。

10

【0186】

さらに、全画面表示する画像を変更する操作が行われたか否かを判定し（ステップ S 97）、操作が行われた場合には画像の変更処理を行う（ステップ S 98）。

【0187】

ステップ S 94 において画像群を変更する操作が行われていないと判定された場合、ステップ S 97 において画像を変更する操作が行われていないと判定された場合、ステップ S 95 またはステップ S 98 の処理が行われた場合には、この再生モードの処理から図 8 に示すメイン処理に復帰する。従って、ステップ S 95 において変更された画像群、またはステップ S 98 において変更された画像の再生は、図 8 のメイン処理における次のループ処理において実行されることになる。

20

【0188】

なお、表示される文字情報をユーザが読むのに要すると考えられる時間 T は、概略、次のような大小関係にあると考えられる。ここに記号「~」は、上述したように、近似した時間長さであることを示している。

T（画像の説明文）~ T（画像のタグ情報） T（画像群のタグ情報）

【0189】

ただし、画像群として考えたときに、画像のタグ情報は画像群に含まれる各画像毎に表示再生されるのに対して、画像群のタグ情報は画像群全体に対して表示再生されるものであるために、文字情報をユーザが読むのに要する 1 画像当たりの時間を考えたときには、次のような大小関係になると捉えることもできる。

30

T（画像の説明文）~ T（画像のタグ情報）> T（画像群のタグ情報）

【0190】

そして、全画面表示モードのときに読むのに最も長い時間を要する「画像の説明文」または「画像のタグ情報」が表示再生されるのは、全画面表示モードが特定の 1 つの画像をじっくりと観察する際に利用される表示モードであると考えられるためである。

【0191】

また、1 つの画面に複数の画像が表示される組写真表示モードにおいては、読むのに要する 1 画像当たりの時間が短い「画像群のタグ情報」が表示再生される。

40

【0192】

また、本実施形態において、実施形態 1 で説明したサムネイル表示やフロー表示を行う場合には、音声再生に代えて、あるいは音声再生と共に、タグ情報を表示するようにすれば良い。

【0193】

具体的に、サムネイル表示においては、サムネイル表示が行われたときに表示されている画像群に関連するタグ情報（画像群タグ情報）を第 2 の要約として表示する。ここに画像群タグ情報は、画像群に含まれている全画像に係る（重複を除いた）全タグ情報であっても良いが、上述した組写真の例と同様に、画像群の全画像に共通するタグ情報、および画像群の何れかの画像に係る WHO のタグ情報であっても良い。そして、サムネイル表示

50

において、特定の画像がシングルタッチされて着目状態になると、シングルタッチされた画像に係るタグ情報を例えば簡易的に（つまり例えば、全種類のタグ情報の中の、重要と考えられる特定種類のタグ情報のみを）第２の要約として表示する（簡易タグ情報の表示）。

【０１９４】

その後、サムネイル表示において特定の画像がダブルタッチされると、図１４～図１６に示したような全画面表示に移行して画像のタグ情報または画像の説明文を第１の要約として表示する。

【０１９５】

また、フロー表示においては、上述したようなフロー表示の移動速度に応じて、時間軸方向における画面中央を通過している画像に関連する簡易タグ情報、もしくは画面中央を通過している画像群に関連するタグ情報（画像群タグ情報）を第２の要約として表示する。このときには、タグ情報の表示を、文字が流れて行くテロップ表示として行っても良い。

【０１９６】

このような実施形態２によれば、上述した実施形態１とほぼ同様の効果を奏するとともに、音声再生を要することなく表示再生のみによっても、画像撮影時の状況をよりリアルに認識することが可能となる。

【０１９７】

また、第２の表示モードが設定されているときには、画面に表示される複数の画像のタグ情報における共通部分を重複させることなく第２の要約として表示再生するようにしたために、複数の画像のタグ情報をそれぞれ認識する場合に比べて、認識に要する時間を有効に短縮することができる。また、時間的な特徴変化を要約する辞書を参照可能に記録しておき、これによって個々の姿勢を動作に変換して表示するなどの工夫も可能となる。例えば上述したように、「立っている」、「座っている」という二つの情報の時間変化を、「座っている」→「立っている」に時間変化した場合には「立ち上がった」、「立っている」→「座っている」に時間変化した場合には「座り込んだ」、というシンプルな動作情報に変換することも可能である。

【０１９８】

さらに、顔判定により検出された顔の人物情報を、第２の表示モードにおいて表示再生する第２の要約に含めるようにしたために、重要度が高いと考えられる、画像に写っている人物の名前等を、画像撮影時の状況に沿って明瞭に認識することが可能となる。

【０１９９】

なお、上述では主として再生装置について説明したが、再生を上述したように行うための再生方法、再生を上述したように制御するための再生制御方法であっても良いし、コンピュータに再生装置を上述したように制御させるための、あるいは再生方法や再生制御方法を上述したように実行するためのプログラム、該プログラムを記録するコンピュータにより読み取り可能な一時的でない記録媒体、等であっても構わない。

【０２００】

具体的に、上記で説明した技術の内の、主にフローチャートを参照して説明した制御に関しては、プログラムの処理より実行可能であることが多く、このプログラムは記録媒体や記録部に収められる場合もある。このプログラムの記録は、製品を出荷するときに出荷製品に対して行っても良いし、製品の出荷と共にあるいは製品の出荷とは異なる時点で配布する記録媒体に対して行っても良い。あるいは、プログラムを、インターネット等の通信回線を介してダウンロードすることができるようにしても良い。

【０２０１】

また、本発明は民生用のカメラ、ビデオカメラ、撮影機能付きの携帯機器、録音用装置、ＰＣのみならず、産業用、医療用の表示機器においても適用可能である。例えば、カプセル内視鏡で一枚の画像を拡大して診る場合と、複数の臓器の画像を通して診る場合とで、補助情報を変更することにより、一つ一つの病変とトータルな健康状態を要約を切り替

10

20

30

40

50

えて認識させても良い。顕微鏡や工業用内視鏡においても同様である。監視カメラの場合は、不審者の外見特徴として一枚の画像から性別や年齢、服装などの見た目から分かる要約を表示し、複数の画像から行動や動作、癖などその他の分かることを要約しても良い。例えば、一枚表示では、「黒服の40代の男で髪が長い」といった要約になり、複数枚表示では、位置情報タグや姿勢情報のタグの変化を解析して「男が走っていた」という要約にしても良い。

【0202】

さらに、上述の実施形態では、画像の表示を前提としていたが、要約機能だけを用いて、画面表示せずに楽しむことも考えられる。例えば、耳で聞く回想装置として使用する場合は、画面を消して省エネ効果を持たせてもよい。必要なのは、グループとしての画像を扱うか、特定の画像を扱うかの差異で、要約が切り替わる点である。思い出に浸る方法としては、画像群を選んだ鑑賞と、特定の画像（1枚でなくともよい）を選んだ鑑賞があるということで、この選択に従って再生される補助情報や要約が変わるといった特徴をもつ機器を提供することができる。このような構成によれば、視覚にこだわることなく、聴覚のみでの鑑賞、回想が可能となり、車を運転しながらの回想や、視覚に自信がない人の回想や、多人数での鑑賞などの一つの画面が見えない（あるいは見難い）状況での鑑賞に対応することが可能となる。

10

【0203】

そして、本発明は上述した実施形態そのままに限定されるものではなく、実施段階ではその要旨を逸脱しない範囲で構成要素を変形して具体化することができる。また、上記実施形態に開示されている複数の構成要素の適宜な組み合わせにより、種々の発明の態様を形成することができる。例えば、実施形態に示される全構成要素から幾つかの構成要素を削除しても良い。さらに、異なる実施形態にわたる構成要素を適宜組み合わせても良い。このように、発明の主旨を逸脱しない範囲内において種々の変形や応用が可能であることは勿論である。

20

【符号の説明】

【0204】

- 1 ... 撮像部
- 2 ... 画像処理部
- 3 ... マイク
- 4 ... 音声処理部
- 5 ... タッチパネル
- 6 ... WEB通信部
- 7 ... GPS部
- 8 ... 時計
- 9 ... 温度計
- 10 ... 制御部
- 10a ... 音声分析部
- 10b ... テキスト化部
- 10c ... 要約部
- 10d ... 顔判定部
- 11 ... 再生制御部
- 11a ... 全画面表示部
- 11b ... サムネイル表示部
- 11c ... フロー表示部
- 11d ... 音声合成部
- 11e ... テロップ作成部
- 12 ... 表示パネル
- 12a ... 画面
- 12t ... タグ表示

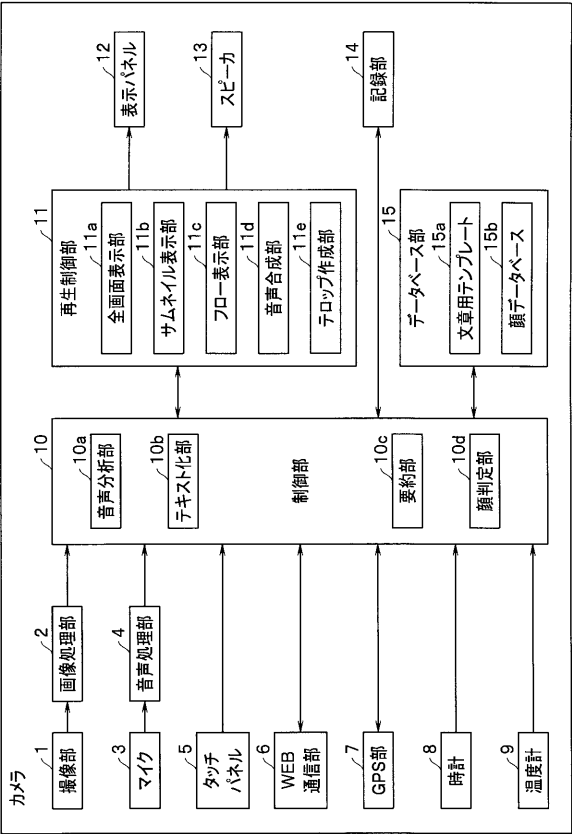
30

40

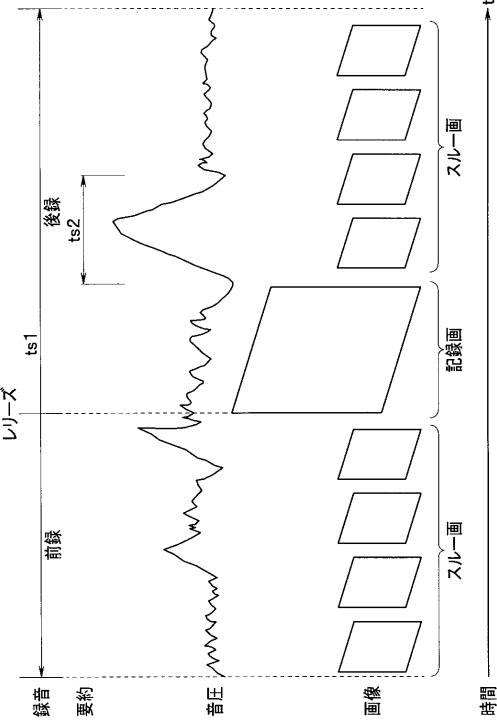
50

- 1 3 ... スピーカ
- 1 4 ... 記録部
- 1 5 ... データベース部
- 1 5 a ... 文章用テンプレート
- 1 5 b ... 顔データベース

【 図 1 】



【 図 2 】



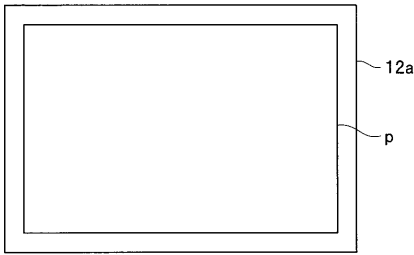
【 図 3 】

音声データ名	内容
ts1	2ndリリース前後5秒の音声データ
ts2	ts1中の特に音圧の高い音声データ
ts3	画像群の特徴音声データ
ts4	画像の説明文の音声データ

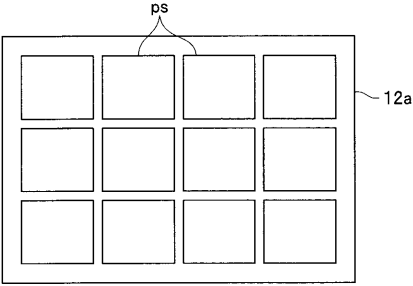
【 図 4 】

画像再生モード	表示例	対応音声	表示説明
全画面表示	図5	ts1もしくはts4,またはts1とts4の同時再生	画面全体に一枚の画像を表示
サムネイル表示	図6	ts2	サムネイル画像を画面全体に並べて表示
フロー表示	図7	流れが遅い時:ts2 流れが速い時:ts3	縮小画像をタッチの方向に応じて,撮影時刻順にフロー表示 その際,画像群単位でまとめ,画像群同士の間はスペースを空けて表示

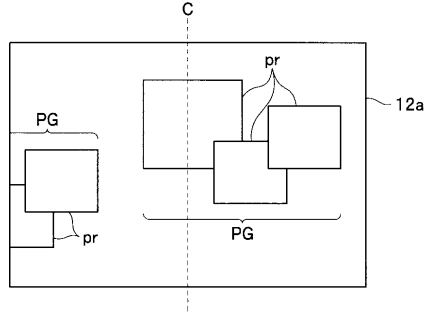
【 図 5 】



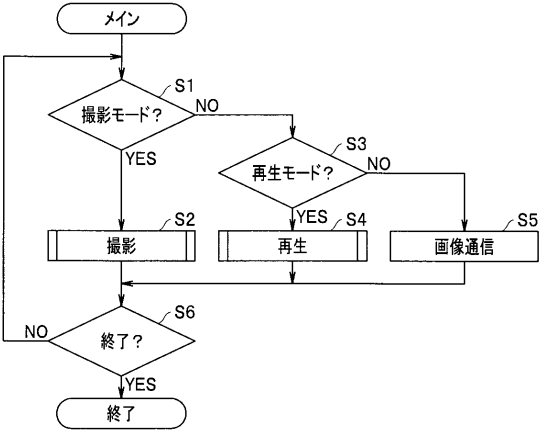
【 図 6 】



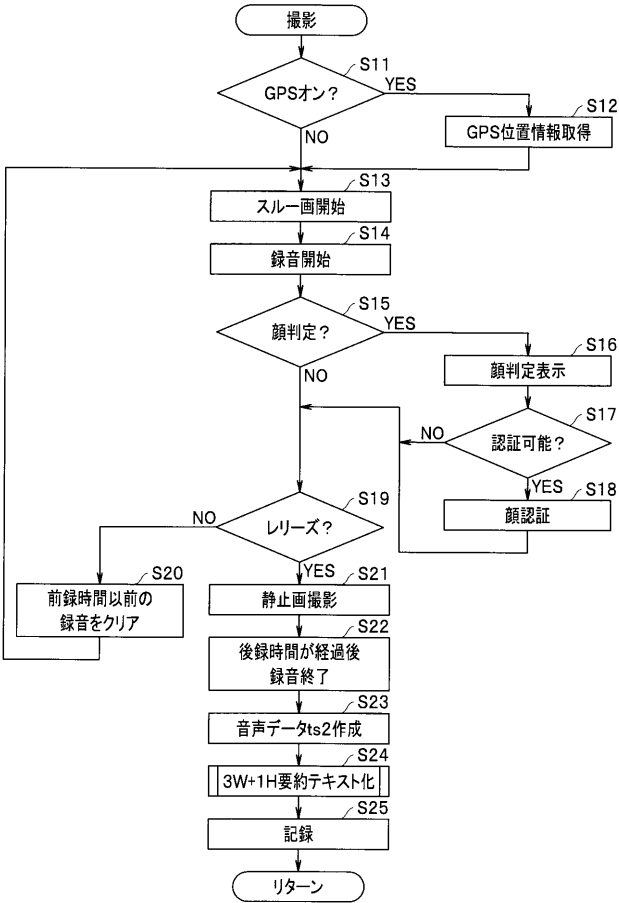
【 図 7 】



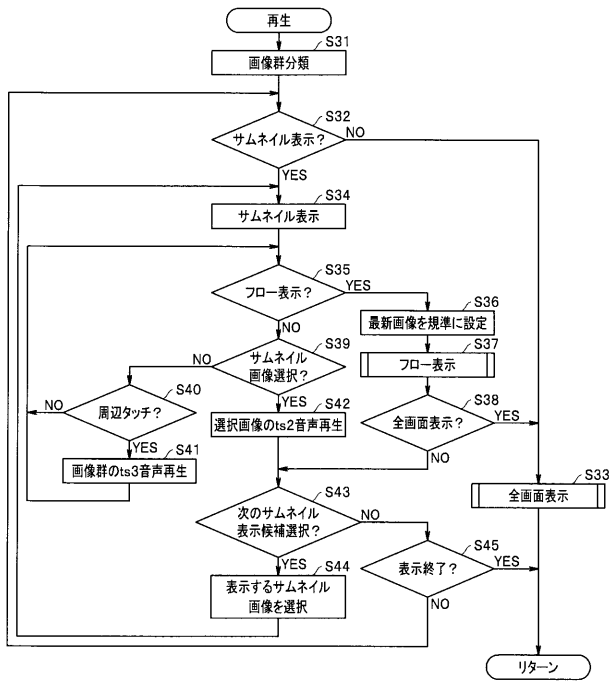
【 図 8 】



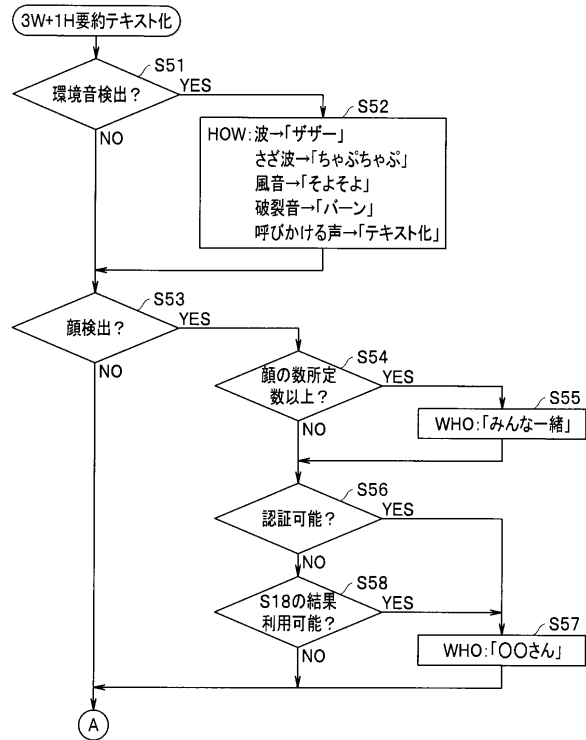
【 図 9 】



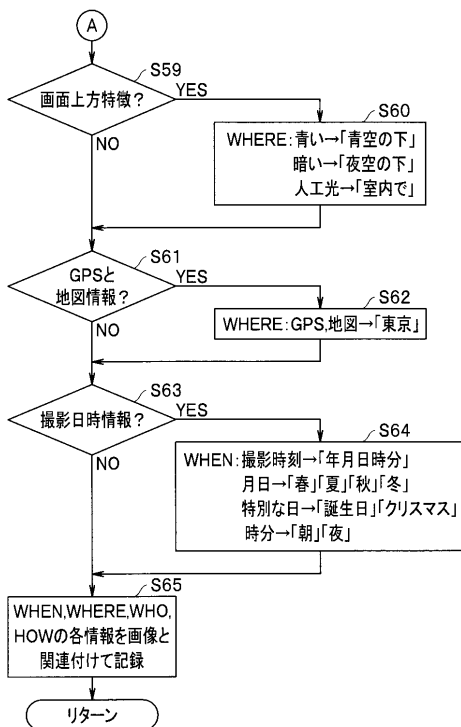
【図 10】



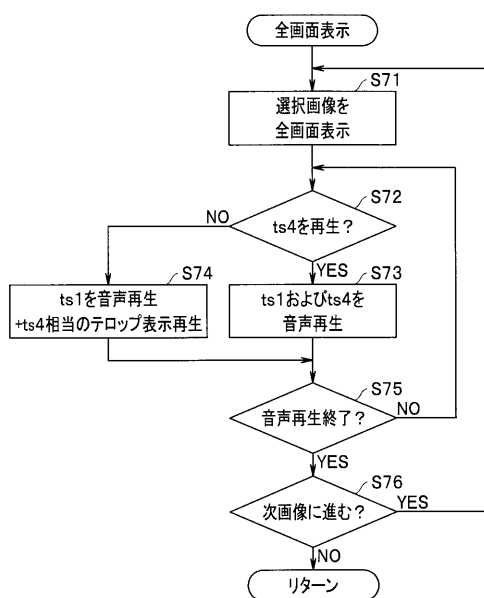
【図 11 A】



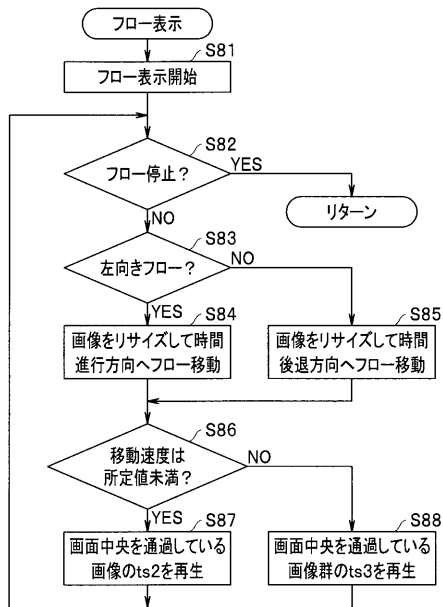
【図 11 B】



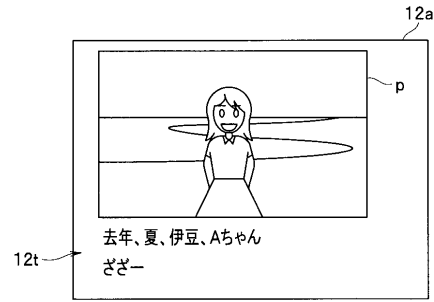
【図 12】



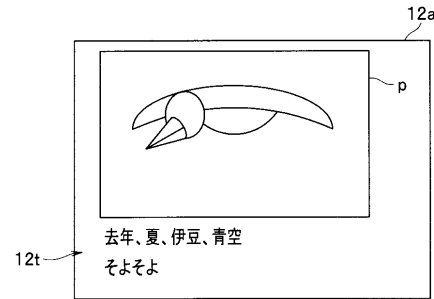
【図 13】



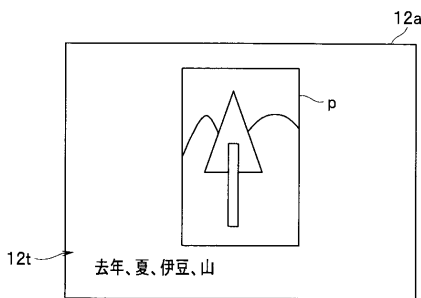
【図 14】



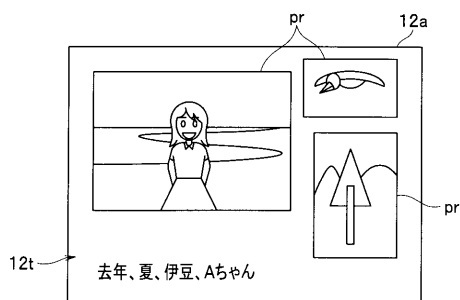
【図 15】



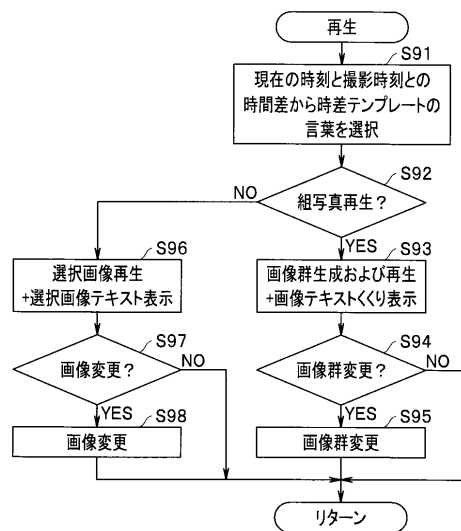
【図 16】



【図 17】



【図 18】



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 17/30 2 1 0 A	

(72) 発明者 野中 修
東京都八王子市石川町 2 9 5 1 番地 オリンパス株式会社内