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

特許第4775066号
(P4775066)

(45)発行日　平成23年9月21日(2011.9.21)　(24)登録日　平成23年7月8日(2011.7.8)

(51)Int.Cl.			F I		
H04N	5/232	(2006.01)	H04N	5/232	Z
H04N	5/91	(2006.01)	H04N	5/91	R
H04N	1/387	(2006.01)	H04N	1/387	
G06T	1/00	(2006.01)	G06T	1/00	340A
G10L	15/00	(2006.01)	G10L	15/00	200G

請求項の数 11　(全 17 頁)　最終頁に続く

(21)出願番号　特願2006-86787(P2006-86787)
(22)出願日　平成18年3月28日(2006.3.28)
(65)公開番号　特開2007-266793(P2007-266793A)
(43)公開日　平成19年10月11日(2007.10.11)
審査請求日　平成20年12月19日(2008.12.19)

(73)特許権者　000001443
カシオ計算機株式会社
東京都渋谷区本町１丁目６番２号
(74)代理人　100088100
弁理士　三好　千明
(72)発明者　栗山　祐司
東京都羽村市栄町３丁目２番１号　カシオ計算機株式会社羽村技術センター内

審査官　田村　誠治

最終頁に続く

(54)【発明の名称】画像加工装置

(57)【特許請求の範囲】
【請求項１】
画像を取得する画像取得手段と、
この画像取得手段により取得された画像から人の口を識別する画像識別手段と、
この画像取得手段に対応して音を取得する音取得手段と、
この音取得手段により取得された音を認識し、この認識した音を表示データに変換する音認識手段と、
前記画像識別手段により識別された前記人の口の数が複数ある時は、識別されたこれら複数の人の口から動きを検出し、前記音認識手段により認識された音に対応する動きのある人の口を判定する判定手段と、
前記判定手段による判定に基づいて、前記画像識別手段により識別された前記人の口に対応する位置に、前記音認識手段により変換された表示データを、合成する画像合成手段とを備え、
前記画像合成手段は、前記画像識別手段により識別された前記人の口が一つの場合、前記判定手段による判定に基づかずに、この一つの人の口に対応する位置に、前記音認識手段により変換された表示データを合成することを特徴とする画像加工装置。
【請求項２】
前記判定手段により検出された画像中における人の口の動きに基づき、前記表示データを訂正処理することを特徴とする請求項１記載の画像加工装置。
【請求項３】

前記画像合成手段は、前記表示データを吹き出しとともに合成することを特徴とする請求項１又は２に記載の画像加工装置。
【請求項４】
前記画像識別手段は、識別した前記人の口の人物が誰であるかを識別し、
前記音認識手段は、前記画像識別手段が識別した人物に応じて、変換する表示データの表示形態を変化させることを特徴とする請求項１から３にいずれか記載の画像加工装置。
【請求項５】
前記画像識別手段は、更に前記人の口の人物の種別を識別し、
前記音認識手段は、前記画像識別手段が識別した人物の種別に応じて、変換する表示データの表示形態を変化させることを特徴とする請求項１から４にいずれか記載の画像加工装置
【請求項６】
前記画像識別手段は、更に前記画像の内容を識別し、
この画像識別手段が識別した画像の内容に応じて表示データを生成する内容表示データ生成手段を更に備え、
前記画像合成手段は、前記内容表示データ生成手段により生成された表示データを前記画像中に合成することを特徴とする請求項１から５にいずれか記載の画像加工装置。
【請求項７】
前記画像取得手段は、前記画像とともに当該画像に付随する情報を取得し、
この画像取得手段が取得した前記情報に基づき、表示データを生成する情報表示データ生成手段を更に備え、
前記画像合成手段は、前記情報表示データ生成手段により生成された表示データを前記画像中に合成することを特徴とする請求項１から６にいずれか記載の画像加工装置。
【請求項８】
前記画像合成手段は、前記画像識別手段により前記人の口の識別ができなかった場合、前記表示データを前記画像中における背景部分に合成することを特徴とする請求項１から７にいずれ記載の画像加工装置。
【請求項９】
前記画像合成手段は、前記表示データを前記画像中における識別された人の口の人物と重ならない位置に合成することを特徴とする請求項１から８にいずれか記載の画像加工装置。
【請求項１０】
前記画像合成手段により前記表示データが合成された画像を記録する記録手段及び／又は前記画像合成手段により前記表示データが合成された画像を表示する表示手段を更に備えることを特徴とする請求項１から９にいずれか記載の画像加工装置。
【請求項１１】
画像加工装置が備えるコンピュータを、
画像を取得する画像取得手段と、
この画像取得手段により取得された画像から人の口を識別する画像識別手段と、
この画像取得手段に対応して音を取得する音取得手段と、
この音取得手段により取得された音を認識し、この認識した音を表示データに変換する音認識手段と、
前記画像識別手段により識別された前記人の口の数が複数ある時は、識別されたこれら複数の人の口から動きを検出し、前記音認識手段により認識された音に対応する動きのある人の口を判定する判定手段と、
前記判定手段による判定に基づいて、前記画像識別手段により識別された前記人の口に対応する位置に、前記音認識手段により変換された表示データを、合成する画像合成手段として機能させ、
前記画像合成手段は、前記画像識別手段により判別された前記人の口が一つの場合、前記判定手段による判定に基づかずに、この一つの人の口に対応する位置に、前記音認識手

段により変換された表示データを合成することを特徴とする画像加工プログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、音認識結果に基づく表示データを生成して画像に合成する画像加工装置に関する。
【背景技術】
【０００２】
従来の画像加工装置として特許文献１記載のものが知られている。この画像加工装置は、撮影の際に被写体から発せられた音声を検出して認識し、文字コードに変換する。また、音声の検出はステレオマイクで行い、このステレオマイクで検出した音声に基づき三角法を用いて音声の発生位置を推測する。そして、画像中の推測した音声発生位置に、前記変換された文字コードに基づく文字イメージと吹き出しとからなる表示データを合成する。
【特許文献１】特開平１１－５５６１４号公報
【発明の開示】
【発明が解決しようとする課題】
【０００３】
しかしながら、前述した従来の画像加工装置にあっては、画像中における文字イメージ等の合成位置を、当該画像から直接的に得ることなく、画像との関係においては間接的な要素である音声に基づき推測する。このため、推測された文字イメージ等の合成位置が実際に音声を発生している被写体と一致しない場合が生じ、画像中の音声発生被写体に対応する位置に精度よく文字イメージ等を合成することができない。
【０００４】
本発明は、かかる従来の課題に鑑みてなされたものであり、画像中の適正位置に音に対応する表示データを合成することのできる画像加工装置を提供することを目的とする。
【課題を解決するための手段】
【０００５】
前記課題を解決するため請求項１記載の発明に係る画像加工装置にあっては、画像を取得する画像取得手段と、この画像取得手段により取得された画像から人の口を識別する画像識別手段と、この画像取得手段に対応して音を取得する音取得手段と、この音取得手段により取得された音を認識し、この認識した音を表示データに変換する音認識手段と、前記画像識別手段により識別された前記人の口の数が複数ある時は、識別されたこれら複数の人の口から動きを検出し、前記音認識手段により認識された音に対応する動きのある人の口を判定する判定手段と、前記判定手段による判定に基づいて、前記画像識別手段により識別された前記人の口に対応する位置に、前記音認識手段により変換された表示データを、合成する画像合成手段とを備え、前記画像合成手段は、前記画像識別手段により識別された前記人の口が一つの場合、前記判定手段による判定に基づかずに、この一つの人の口に対応する位置に、前記音認識手段により変換された表示データを合成する。
【００１１】
また、請求項２記載の発明に係る画像加工装置にあっては、前記判定手段により検出された画像中における人の口の動きに基づき、前記表示データを訂正処理する。
【００１２】
また、請求項３記載の発明に係る画像加工装置にあっては、前記画像合成手段は、前記表示データを吹き出しとともに合成する。
【００１３】
また、請求項４記載の発明に係る画像加工装置にあっては、前記画像識別手段は、識別した前記人の口の人物が誰であるかを識別し、前記音認識手段は、前記画像識別手段が識別した人物に応じて、変換する表示データの表示形態を変化させる。
【００１４】

また、請求項５記載の発明に係る画像加工装置にあっては、前記画像識別手段は、更に前記人の口の人物の種別を識別し、前記音認識手段は、前記画像識別手段が識別した人物の種別に応じて、変換する表示データの表示形態を変化させる。

【００１５】

また、請求項６記載の発明に係る画像加工装置にあっては、前記画像識別手段は、更に前記画像の内容を識別し、この画像識別手段が識別した画像の内容に応じて表示データを生成する内容表示データ生成手段を更に備え、前記画像合成手段は、前記内容表示データ生成手段により生成された表示データを前記画像中に合成する。

【００１６】

また、請求項７記載の発明に係る画像加工装置にあっては、前記画像取得手段は、前記画像とともに当該画像に付随する情報を取得し、この画像取得手段が取得した前記情報に基づき、表示データを生成する情報表示データ生成手段を更に備え、前記画像合成手段は、前記情報表示データ生成手段により生成された表示データを前記画像中に合成する。

【００１７】

また、請求項８記載の発明に係る画像加工装置にあっては、前記画像合成手段は、前記画像識別手段により前記人の口の識別ができなかった場合、前記表示データを前記画像中における背景部分に合成する。

【００１８】

また、請求項９記載の発明に係る画像加工装置にあっては、前記画像合成手段は、前記表示データを前記画像中における識別された人の口の人物と重ならない位置に合成する。

【００１９】

また、請求項１０記載の発明に係る画像加工装置にあっては、前記画像合成手段により前記表示データが合成された画像を記録する記録手段及び／又は前記画像合成手段により前記表示データが合成された画像を表示する表示手段を更に備える。

【００２０】

また、請求項１１記載の発明に係る画像加工プログラムにあっては、画像加工装置が備えるコンピュータを、画像を取得する画像取得手段と、この画像取得手段により取得された画像から人の口を識別する画像識別手段と、この画像取得手段に対応して音を取得する音取得手段と、この音取得手段により取得された音を認識し、この認識した音を表示データに変換する音認識手段と、前記画像識別手段により識別された前記人の口の数が複数ある時は、識別されたこれら複数の人の口から動きを検出し、前記音認識手段により認識された音に対応する動きのある人の口を判定する判定手段と、前記判定手段による判定に基づいて、前記画像識別手段により識別された前記人の口に対応する位置に、前記音認識手段により変換された表示データを、合成する画像合成手段として機能させ、前記画像合成手段は、前記画像識別手段により判別された前記人の口が一つの場合、前記判定手段による判定に基づかずに、この一つの人の口に対応する位置に、前記音認識手段により変換された表示データを合成する。

【発明の効果】

【００２１】

本発明によれば、画像中の音発生被写体を該画像に基づいて識別することから、画像から直接的に音発生被写体を識別することにより、精度よく音発生被写体を識別することができる。よって、精度よく画像中の音発生被写体に対応する位置に、音を変換した表示データを表示することが可能となる。

【発明を実施するための最良の形態】

【００２２】

以下、本発明の一実施の形態を図に従って説明する。図１は、本発明の一実施の形態を適用したデジタルカメラ１０の回路構成を示すブロック図であり、このデジタルカメラ１０は、後述するＡＦ機能とともにＡＥ、ＡＷＢ等の一般的な機能をも有するものである。すなわち、レンズブロック１１には、ズームレンズ、フォーカスレンズ等の光学系、及び光学系を駆動するための駆動機構が含まれており、前記光学系は、駆動機構に設けられて

いるモーター１２によって光軸方向に駆動される。なお、本実施の形態において、前記ＡＦは、フォーカスレンズを光軸方向に移動させながら、各位置で撮像した画像のＡＦ評価値（コントラスト値）を検出し、ＡＦ評価値のピーク位置を合焦位置とするコントラスト検出方式である。

【００２３】

デジタルカメラ１０全体を制御するＣＰＵ１３には、バス１４及びタイミング発生器（ＴＧ：Timing Generator）１５を介してモータードライバ１６が接続されており、モータードライバ１６は、ＣＰＵ１３の命令に従いタイミング発生器１５が発生するタイミング信号に基づき、モーター１２を駆動する。なお、ストロボ１７もタイミング発生器１５が発生するタイミング信号により駆動される。

【００２４】

また、このデジタルカメラ１０は撮像素子としてＣＣＤ１８を有している。ＣＣＤ１８は、レンズブロック１１の光軸上に配置されており、被写体は、レンズブロック１１によってＣＣＤ１８の受光面に結像される。ＣＣＤ１８は、ＣＰＵ１３の命令に従いタイミング発生器１５が生成するタイミング信号に基づき垂直及び水平ドライバ１９によって駆動され、被写体の光学像に応じたアナログの撮像信号をユニット回路２０に出力する。ユニット回路２０は、ＣＣＤ１８の出力信号に含まれるノイズを相関二重サンプリングによって除去するＣＤＳ回路や、ノイズが除去された撮像信号をデジタル信号に変換するＡ／Ｄ変換器等から構成され、デジタルに変換した撮像信号を画像処理部２１へ出力する。

【００２５】

画像処理部２１は、入力した撮像信号に対しペデスタルクランプ等の処理を施し、それを輝度（Ｙ）信号及び色差（ＵＶ）信号に変換するとともに、オートホワイトバランス、輪郭強調、画素補間などの画品質向上のためのデジタル信号処理を行う。画像処理部２１で変換されたＹＵＶデータは順次ＳＤＲＡＭ２２に格納されるとともに、ＲＥＣスルー・モードでは１フレーム分のデータ（画像データ）が蓄積される毎にビデオ信号に変換され、バックライト（ＢＬ）２４を備える液晶モニタ（ＬＣＤ）２３へ送られてスルー画像として画面表示される。

【００２６】

そして、スチル撮影モードにおいては、シャッターキー操作をトリガとして、ＣＰＵ１３は、ＣＣＤ１８、垂直及び水平ドライバ１９、ユニット回路２０、及び画像処理部２１に対してスルー画撮影モード（ＲＥＣスルー・モード）から静止画撮影モードへの切り替えを指示し、この静止画撮影モードによる撮影処理により得られ、ＳＤＲＡＭ２２に一時記憶された画像データは、ＣＰＵ１３により圧縮され、最終的には所定のフォーマットの静止画ファイルとして外部メモリ２５に記録される。また、ムービー録画モードにおいては、１回目のシャッターキーと２回目のシャッターキー操作との間に、ＳＤＲＡＭ２２に順次記憶される複数の画像データがＣＰＵ１３により順次圧縮されて、圧縮動画データが生成され動画ファイルとして外部メモリ２５に記録される。この外部メモリ２５に記録された静止画ファイル及び動画ファイルは、ＰＬＡＹ・モードにおいてユーザーの選択操作に応じてＣＰＵ１３に読み出されるとともに伸張され、ＹＵＶデータとしてＳＤＲＡＭ２２に展開された後、液晶モニタ（ＬＣＤ）２３に表示される。

【００２７】

フラッシュメモリ２６には、ＣＰＵ１３に前記各部を制御させるための各種のプログラム、例えばＡＥ、ＡＦ、ＡＷＢ制御用のプログラムや、さらには、後述するフローチャートに示す処理を実行するためのプログラム等の各種のプログラムが格納されている。

【００２８】

また、デジタルカメラ１０は、電源スイッチ、モード選択キー、シャッターキー、ズームキー、後述するピント枠を手動選択するためのピント枠選択キー等の複数の操作キー及びスイッチを含むキー入力部（ＫＥＹ）２７、ニッケル水素電池等の充電可能なバッテリー２８、このバッテリー２８の電力を各部に供給するための電源制御回路２９、及びこれらを制御するマイコン３０を有している。マイコン３０は、キー入力部２７における前記

操作キーの操作の有無を定常的にスキャンしており、ユーザーによっていずれかの操作キーが操作されると、その操作内容に応じた操作信号をＣＰＵ１３へ送る。なお、シャッターキーは、半押しと全押しとが可能な所謂ハーフシャッター機能を有するものである。

【００２９】

また、このデジタルカメラ１０は、前記ムービー録画モードにおいて、周囲音を記録する録音機能を備えており、ＣＰＵ１３には、音声処理回路を有する音声チップ３２を介して、スピーカ（ＳＰ）３３と、マイクロホン（ＭＩＣ）３４とが接続されている。音声チップ３２は、ムービー録画モード時には、マイクロホン３４から入力された音声波形を処理して、音声波形データをＣＰＵ１３に入力する。そして、ＣＰＵ１３は、ムービー録画モードにおいて１回目と２回目のシャッターキー操作間に、音声チップ３２から入力された音声波形データを圧縮し、この圧縮周囲音データと前記圧縮動画データとを含む音声付き動画ファイルを生成して外部メモリ２５に記録する。この外部メモリ２５に記録された音声付き動画ファイルは、ＰＬＡＹ・モードにおいて動画データが再生される際に、周囲音データが音声チップ３２で音声波形に変換されてスピーカ３３により再生される。

【００３０】

さらに、バス１４にはＧＰＳ３５が接続されており、前記フラッシュメモリ２６には前記プログラム等とともに地図データが記憶されている。したがって、ＣＰＵ１３はＧＰＳ３５により検出された現在位置の緯度・経度と、フラッシュメモリ２６内の地図データとに現在位置の地名を取得することが可能である。フラッシュメモリ２６には、音声をテキストデータに変換するための音声－テキストデータ変換テーブルや、音声以外の音を擬音表示データ（例えば、クラッカーの破裂音を擬音表示データ「パン」、自動車の音を擬音表示データ「ブー」）に変換する音－表示データ変換テーブルが記憶されている。また、画像の動きを擬音表示するための「ビュー」や表情を擬音表示するための「プンプン」「ニコニコ」等の表示データ、あるいは画像の弧の動きを強調表示するための「（（」等、暑さや寒さや擬音表示するための「ジリジリ」「ヒュー」、汗マーク等の画像内容－表示データ変換テーブル、「もうかりまっか」「ぼちぼちでんな」等の複数の慣用語からいずれかをランダムに選択するためのランダムテキストデータ、口の動きをテキストデータに変換するための口の動き－テキストデータ変換テーブル、テキストデータを対応する方言テキストデータに変換するための方言変換テーブル等が記憶されている。

【００３１】

加えて、前記フラッシュメモリ２６には、図２に示す被写体種別判定テーブル２６１が格納されている。被写体種別判定テーブル２６１には、「人」、「人の口」、・、「自動車」・・・等の被写体となり得る被写体種別毎にその画像の特徴を示す特徴量データＤが記憶されている。さらに、被写体種別判定テーブル２６１には、顔に関しては、「怒っている顔」、「泣いている顔」等の顔の表情種別毎に特徴量データＤが記憶されているとともに、「個人名Ａ」、「個人名Ｂ」等の個人名に対応して画像の特徴を示す特徴量データＤも記憶されている。これら各画像の特徴量データＤは、色相＝ＨＨＨ、彩度＝ＳＳＳ、明度＝ＶＶＶ、輪郭形状＝ＦＦＦ、大きさ＝ＬＬＬ・・・等の複数種の特徴量で構成されている。

【００３２】

以上の構成に係る本実施の形態において、前述のようにムービー録画モードにおいては、１回目のシャッターキーと２回目のシャッターキー操作との間に、ＳＤＲＡＭ２２に順次記憶される複数の画像データがＣＰＵ１３により順次圧縮される。また、１回目と２回目のシャッターキー操作間に、音声チップ３２から入力された音声波形データが圧縮され、この圧縮周囲音データと圧縮動画データとを含む音声付き動画ファイルを生成されて外部メモリ２５に記録される。さらに、この音声付き動画ファイルの記録に際してＣＰＵ１３は、ＧＰＳ３５により検出された緯度・経度と前記地図データとに基づき、撮影地域を検出して、動画ファイルのヘッダーに記憶するとともに、撮影日時、撮影時の明るさ等の撮影条件データもヘッダーに記録する。したがって、音声付き動画ファイルには、圧縮周囲音データと圧縮動画データが記憶されているとともに、付加情報として撮影地域、撮影

日時、撮影条件等が記憶されている。

【００３３】

そして、ＰＬＡＹ・モードにおいて画像加工モードを設定し、外部メモリ２５からいずれかの音声付き動画ファイルを選択すると、ＣＰＵ１３は図３～図１０に示すフローチャートに従って処理を実行する。すなわち、図３に示すように、選択された音声付き動画ファイルからの画像データ及び周囲音データの読み出しを開始する（ステップＳ１０１）。この読み出した周囲音データに関しては、再生することなく後述する周囲音認識処理を実行する（ステップＳ１０２）。なお、周囲音に関しても、音声チップ３２で再生しスピーカ３３から放音するようにしてもよい。引き続き、後述する画像加工処理を実行し（ステップＳ１０３）、この画像加工処理された画像データを含む動画データをＳＤＲＡＭ２２に順次記憶するとともに、この画像加工処理された画像データを含む動画データを再生して、液晶モニタ２３に表示させる（ステップＳ１０４）。

【００３４】

しかる後に、前記動画ファイルから読み出している動画データの再生を終了したか、又は動作再生を停止させるキー操作がなされたか否かの終了判断を行い（ステップＳ１０５）、終了と判断したならば、ステップＳ１０４で順次記憶した複数の画像データからなる動画データを圧縮し、別動画ファイルとして外部メモリ２５に記録する（ステップＳ１０６）。したがって、後日これら動画ファイルに基づく加工動画を再生することもできるし、加工動画中の任意のフレームを選択して静止画としてプリントアウトすることもできる。

【００３５】

図４は、前記周囲音認識処理（ステップＳ１０２）の処理手順を示すフローチャートである。先ず、前記音声付き動画ファイルから動画データと同期して順次読み出される音声データに周囲音が含まれているか否かを判断する（ステップＳ２０１）。周囲音が含まれている場合には、その波形、スペクトル等の音声データの特徴と読み出された周囲音の特徴とを比較することにより、該読み出された周囲音が音声であるか否かを判断する（ステップＳ２０２）。つまり、周囲音を音声認識し、音声認識不可能であれば、音声ではないと判断する。この判断の結果、読み出された周囲音が音声以外の音であった場合には、当該音を擬音表示データに変換する（ステップＳ２０３）。例えば、周囲音がクラッカーの破裂音であれば、「パン」の文字からなる擬音データに変換し、周囲音が音楽であれば音符の画像からなる擬音データに変換する。

【００３６】

また、読み出された周囲音が音声認識可能であれば、これを音声であると判断し、この音声を認識処理してテキストデータに変換する処理を開始する（ステップＳ２０４）。また、音声と同期して順次読み出される動画中における人間の口の動きを認識する（ステップＳ２０５）。このステップ２０５での処理に際しては、図１０において後述するように先ずフレーム画像中における人間の口の存在を検出する。そして、この検出したフレーム画像中における口の変化を時系列的に検出することにより、口の動きを認識する。この認識した口の動きに対応するテキストデータを前記口の動き－テキストデータ変換テーブルから読み出すことにより、口の動きに対応するテキストデータを得る。なお、言うまでもなく、動画中に人間の口が存在しない場合や人間の口が存在しても口が動いていない場合にはステップＳ２０５～Ｓ２０８の処理をスキップすることになる。

【００３７】

次に、この口の動きに対応するテキストデータと、ステップＳ２０４で音声からの変換を開始しているテキストデータとを照合し（ステップＳ２０６）、両者に不一致があるか否かを判断する（ステップＳ２０７）。両者に不一致がある場合には、音声から変換しているテキストデータの不一致部分を、口の動きに対応するテキストデータに訂正する（ステップＳ２０８）。なお、これとは逆に、口の動きに対応するテキストデータの不一致部分を、音声から変換しているテキストデータに訂正するようにしてもよい。

【００３８】

また、音声が終了したか否かを判断し（ステップＳ２０９）、音声が終了するまでステップＳ２０５からの処理を繰り返す。音声が終了したならば、音声が強く終わったか否かを判断し（ステップＳ２１０）、強く終わった場合にはテキストデータの末尾に感嘆符“！”を追加する（ステップＳ２１１）。さらに、音声が上がって終わったか否かを判断し（ステップＳ２１２）、上がって終わった場合にはテキストデータの末尾に疑問符“？”を追加する（ステップＳ２１３）。

【００３９】

図５～図７は、前記画像加工処理（ステップＳ１０３）の処理手順を示す一連のフローチャートである。先ず、図５に示すように、動きの早い被写体があるか否かを判断する（ステップＳ３０１）。この判断に際しては、予め動画における画像変化速度の基準値Ａｍｍ／ｓを定めておき、動画中にこの基準値Ａｍｍ／ｓよりも速い速度で動いた被写体があるか否かを判断する。そして、この判断した被写体の動画を構成するフレーム画像中における位置（位置座標）を検出する（ステップＳ３０２）。また、フラッシュメモリ２６から動きの早い被写体に対応する擬音を示す表示データ（本例では前記「ビュー」）を読み出し（ステップＳ３０３）、この読み出した擬音を示す表示データを前記ステップＳ３０２で検出した位置の近傍に合成する（ステップＳ３０４）。したがって、このステップＳ３０１～Ｓ３０４での処理により、例えば投げられたボールの近傍に擬音表示データ「ビュー」が合成される。

【００４０】

また弧の動きの被写体があるか否かを判断する（ステップＳ３０５）。この判断に際しては、動画を構成するフレームの前後の関係から、弧の動きの被写体の有無を判断する。そして、弧の動きの被写体があった場合には、フレーム画像中における位置を検出する（ステップＳ３０６）。また、フラッシュメモリ２６から弧の動きを線を示す表示データ（本例では前記「（（」を読み出し（ステップＳ３０７）、この読み出した擬音を示す表示データを前記ステップＳ３０６で検出した位置の近傍に合成する（ステップＳ３０８）。したがって、このステップＳ３０５～Ｓ３０８での処理により、例えば尻尾を振る犬の尻尾の近傍に「（（」を合成することができる。

【００４１】

引き続き、周囲音があるか否か（周囲音が読み出されたか否か）を判断し（図６ステップＳ３０９）、周囲音がない場合には、再生画像中に人の顔があるか否かを判断する（ステップＳ３１０）。

【００４２】

この判断に際しては、図１０のフローチャートに示すように、動画を構成するフレーム内の抽出領域を検出する（ステップＳ１）。この抽出領域の検出は、フレーム画像の画像データの輝度信号及び色差信号から、近い輝度又は色差信号別に、同系色の色相別等に領域を分割し、さらに、領域の境界線となる輪郭線を抽出し、この輪郭線で囲まれた部分を一つの抽出領域として検出する。引き続き、この検出した抽出領域を順次選択し（ステップＳ２）、この選択した抽出領域におけるフレーム画像の特徴抽出処理を実行する（ステップＳ３）。つまり、選択した抽出領域において、前記特徴量データＤが有する特徴種別の特徴量を抽出する。したがって、本例においては、特徴量データＤは、色相、彩度、明度、輪郭形状、大きさ・・・であったことから、抽出領域にこれら色相、彩度、明度、輪郭形状、大きさ・・・の特徴量を抽出する。

【００４３】

そして、このステップＳ３で抽出した特徴量と、被写体種別判定テーブル２６１に記憶されている比較対照となっている被写体種別（ステップＳ３１０の場合「人の顔」）の特徴量データＤの色相＝ＨＨＨ、彩度＝ＳＳＳ、明度＝ＶＶＶ、輪郭形状＝ｆｆｆ、大きさ＝ＬＬＬ・・・と各々比較し類似度を各々算出する（ステップＳ４）。つまり、被写体種別判定テーブル２６１に記憶されている判断対象の被写体種別の特徴量データＤの各値と抽出した特徴量の各値との比率を算出する。次に、この算出した比率である類似度が所定値以上である否かを判断し（ステップＳ５）、類似度が所定値以上である場合には、当該

被写体があると判断する（ステップＳ６）。そして、あると判断した被写体の画像上における位置を検出し、この検出した位置をその被写体種別と共にＳＤＲＡＭ２２に記憶する（ステップＳ７）。

【００４４】

また、類似度が所定値未満である場合には、最後の抽出領域まで以上のステップＳ２～ステップＳ５の処理を実行したか否かを判断し（ステップＳ８）、最後の抽出領域となるまでステップＳ２からの処理を繰り返す。したがって、後述するように画像中に複数の口が存在する場合には、各口に対応してステップＳ６とステップＳ７の処理が実行されて、複数の各口に対応してその位置がＳＤＲＡＭ２２に記憶されることとなる。よって、最後の抽出領域となるまで、ステップＳ５の判断がＮＯであって、類似度が所定値以上の抽出領域がない場合には、ＳＤＲＡＭ２２には被写体の画像上における位置、及び被写体種別が記憶されない。したがって、ＳＤＲＡＭ２２に被写体の画像上における位置、及び被写体種別が記憶されているか否かにより、当該被写体があるか否かを判断することができる。

【００４５】

そして、ステップＳ３１０の判断がＮＯであって、人の顔の被写体がない場合には、前記ヘッダーに記憶されている撮影条件データ等に基づきフラッシュメモリ２６から表示データを読み出し（ステップＳ３１１）、この読み出した表示データを画像の任意の位置に合成する（ステップＳ３１２）。したがって、このステップＳ３１１及びＳ３１２での処理により、周囲音がない場合であっても、ヘッダーに記憶されている明るさや撮影日時に応じて、「ジリジリ」や「ヒュー」の擬音表示データを、画像の適宜の位置に合成することができる。

【００４６】

また、ステップＳ３１０での判断の結果、人の顔があった場合には、フレーム画像中におけるその位置を前記図１０のステップＳ７においてＳＤＲＡＭ２２に人の顔と共に記憶された検出位置を取得する（ステップＳ３１３）。次に、この検出された位置の画像である顔に表情があるか否かを判断する（ステップＳ３１４）。この判断も図１０に示したフローチャートに従って行い、表情がある場合には、フラッシュメモリ２６から表情に応じた表示データ（本例では前記「ニコニコ」「プンプン」）を読み出し（ステップＳ３１５）、この読み出した擬音を示す表示データを前記ステップＳ３１３で取得した位置の近傍に合成する（ステップＳ３１６）。したがって、このステップＳ３１３～Ｓ３１６での処理により、周囲音がない場合であっても、被写体の顔の近傍に「ニコニコ」「プンプン」等を合成して表示ことができる。

【００４７】

また、ステップＳ３１４での判断の結果、表情がないと判断された場合には、フラッシュメモリ２６から前記ランダムテキストデータのいずれかをランダムに選択する（ステップＳ３１７）。引き続き、ステップＳ３１３で取得した検出位置に最も近い背景領域を検出する（ステップＳ３０８）。この背景領域の検出は、図１０に示したフローチャートのステップを利用して行うことができる。

【００４８】

すなわち、前述したように、図１０のステップＳ１においては、動画を構成するフレーム内の抽出領域を検出する。この抽出領域の検出は、フレーム画像の画像データの輝度信号及び色差信号から、近い輝度又は色差信号別に、例えば同系色の色相別等に領域を分割し、さらに、領域の境界線となる輪郭線を抽出し、この輪郭線で囲まれた部分を一つの抽出領域として検出する。したがって、このように、抽出領域と検出された領域以外の領域を背景領域であるとして検出することができる。

【００４９】

そして、検出位置に最も近い背景領域を検出したならば、この検出した検出領域内に吹き出しを合成し（ステップＳ３１９）、この吹き出し内に前記ステップＳ３１７で選択したテキストデータを合成する（ステップＳ３２０）。したがって、音声がない場合であっ

ても、画像内の人物が「もうかりまっか」等を発言しているかのような画像を合成して表示することができる。

【００５０】

他方、ステップＳ３０９での判断の結果、周囲音がある場合には、前記図１０に示したフローチャートに従って処理を実行することにより、フレーム画像中に人が存在するか否かを判断する（図７のステップＳ３２１）。人が存在する場合には、同様の処理により被写体種別判定テーブル２６１に個人名がある被写体であるか否かを判断する（ステップＳ３２２）。ある場合には、フレーム画像中における前記図１０のステップＳ７においてＳＤＲＡＭ２２に人と共に記憶された検出位置を取得し（ステップＳ３２３）、この取得した位置の被写体に個人名を合成する（ステップＳ３２４）。

【００５１】

また、音声があるか否かを判断し（ステップＳ３２５）、音声がない場合にはステップＳ３３３に進む。音声がある場合には、前記同様の処理により人の口が存在するか否かを判断し（ステップＳ３２６）、人の口が存在しない場合、つまり音声があり（ステップＳ３２５；ＹＥＳ）、人も写っているが（ステップＳ３２１；ＹＥＳ）、口は写っていない場合には（ステップＳ３２６；ＮＯ）、後述する第１の吹き出し合成処理を実行する（ステップＳ３２７）。

【００５２】

また、口が存在する場合には、複数の口が存在するか否かを判断する（ステップＳ３２８）。つまり、前述のように図１０のフローチャートに従った処理より、複数の口が存在する場合には、フレーム画像中における各口は特定されていることから、これに基づき複数の口の有無を判断する。この判断がＮＯであって単一の口のみが写っている場合には、次のステップＳ３２９の判断を行うことなく、後述する第２の吹き出し合成処理を実行する（ステップＳ３３２）。また、複数の口が写っている場合には、動いている口があるか否かを判断する（ステップＳ３２９）。つまり、前述のように図１０のフローチャートに従った処理より、複数の口が存在する場合には、フレーム画像中における各口は特定されていることから、このフレーム画像中における各口の変化の有無を時系列的に検出することにより、動いている口があるか否かを判断することができる。

【００５３】

そして、動いている口がない場合には、後述する第１の吹き出し合成処理（ステップＳ３２７）を実行する。また、動いている口がある場合には、該動いている口は１つであるか否かを判断し（ステップＳ３３０）、１つである場合には後述する第２の吹き出し合成処理を実行する（ステップＳ３３２）。しかし、動いている口が１つではなく、複数ある場合には、前記ステップＳ３２５でＹＥＳ（音声あり）と判断された音声に対応する口を検出する（ステップＳ３３１）。

【００５４】

すなわち、前述の図４のフローチャートにおいては、ステップＳ２０４で音声を認識処理してテキストデータに変換する処理を開始し、また、ステップＳ２０５では音声とともに順次読み出される動画中における人間の口の動きを認識する。したがって、動いている複数の口において、音声認識により順次変換されるテキストデータと前記ステップＳ２０５で認識される動きとが同期する口を検出することにより、音声に対応する口、つまりテキストデータに変換されている音声に対応して動いている口を検出することができる。したがって、このステップＳ３３１の処理は、図４のフローチャートに示した周囲音認識処理で実行されるテキストデータ変換処理と口の動き認識処理とを利用して、判断を行う。

【００５５】

なお、第２の吹き出し合成処理は、後述するようにテキストデータに基づき実行される処理、つまりは音声の存在を前提として実行される処理である。したがって、本実施の形態においては、単一の口が写っているか又は動いている口が写っている場合には、音声も録音されていることが前提となる。

【００５６】

そして、前記ステップＳ３２５で音声がないと判断された場合、第１の吹き出し合成処理（ステップＳ３３２）又は第２の吹き出し合成処理（ステップＳ３３２）を実行した後、同様に図１０のフローチャートに従った処理を実行することにより、人以外の他の音発生被写体があるか否かを判断する（ステップＳ３３３）。ある場合には、前記図１０のステップＳ７においてＳＤＲＡＭ２２に人以外の他の音発生被写体と共に記憶された検出位置を取得し（ステップＳ３３４）、この取得した位置の近傍に、前記ステップＳ２０３で変換された擬音表示データを合成する（ステップＳ３３５）。したがって、図１１（Ａ）に示すように、加工前の画像においてクラッカーＰ１が検出されると、同図（Ｂ）の加工後の画像に示すように、クラッカーＰ１の近傍に擬音表示データＰ２「パン」を合成することができる。

【００５７】

さらに、前記人又は音発生被写体以外の背景に前記ステップＳ２０３で変換された擬音表示データを合成する（ステップＳ３３６）。したがって、周囲音が例えば拍手であれば、図１１（Ｂ）に示すように、「パチパチ」なる表示データＰ３が合成される。また、音楽が流れていれば、音符からなる表示データＰ４を合成される。

【００５８】

他方、前記ステップＳ３２１で人が存在しないと判断された場合には、音声があるか否かを判断する（ステップＳ３３７）。そして、音声がある場合には第１の吹き出し処理を実行し（ステップＳ３３８）、音声がない場合にはステップＳ３３３に進む。

【００５９】

図８は、前記第１の吹き出し合成処理（ステップＳ３２７、ステップＳ３３８）の処理手順を示すフローチャートである。先ず、前記ステップＳ３１８での説明と同様の処理を行うことにより、フレーム画像中において背景領域を検出する（ステップＳ４０１）。なお、ステップＳ３２７でこの第１の吹き出し合成処理を実行する場合には、ステップＳ３２１で人が存在すると判断されているので、このステップＳ３２１で存在すると判断された人の近傍に背景領域を検出する。

【００６０】

そして、背景領域を検出したならば、この検出した検出領域内に収まるような吹き出しを生成する（ステップＳ４０２）。しかる後に、前記図４のフローチャートに従った処理により得られているテキストデータを方言に変換する（ステップＳ４０３）。つまり前述のように、この音声付き動画ファイルの記録に際しては、ＧＰＳ３５により検出された緯度・経度と地図データとに基づき検出された撮影地域が、当該動画ファイルのヘッダーに記憶されている。したがって、この撮影地域を読み出し、前記テキストデータを、フラッシュメモリ２６内の方言変換テーブルを用いて、前記撮影地域に対応する方言のテキストデータに変換する。

【００６１】

さらに、この変換したテキストデータをステップＳ４０２で生成した吹き出し内に合成して、この吹き出しとテキストデータとからなる吹き出しテキストデータを生成する（ステップＳ４０４）。引き続き、表示色変更処理を実行して、この吹き出しテキストデータの表示色を、ステップＳ３０９で検出された周囲音（音声）の高さに応じて変更する（ステップＳ４０５）。また、表示サイズ変更処理を実行して、この吹き出しテキストデータの表示サイズを、ステップＳ３０９で検出された周囲音（音声）の音量に応じて変更する（ステップＳ４０６）。また、前記ステップＳ３２１で存在が検出された人に対応する個人名が被写体種別判定テーブル２６１にあるか否かを判断する（ステップＳ４０７）。ある場合には、フォント変更処理を実行して、この吹き出しテキストデータにおけるテキストデータのフォントを、前記ステップＳ４０７で個人名ありと判断された個人名（あるいは性別）に応じて変更する（ステップＳ４０８）。そして、以上の処理により確定した吹き出しテキストデータを前記ステップＳ４０１で検出した検出領域内に、合成する（ステップＳ４０９）。

【００６２】

したがって、この図８に示した第１の吹き出し合成処理により、人の口が写っていない場合であって、音声が検出された場合には、吹き出し内に音声に対応するテキストデータが合成された吹き出しテキストデータが、背景に合成されることとなる。

【００６３】

図９は、前記第２の吹き出し合成処理（ステップＳ３３２）の処理手順を示すフローチャートである。先ず、前記ステップＳ３２８で単一の口であると判断された口、又は前記ステップＳ３３０で動いている口は１つであると判断された当該口、又はステップＳ３３１で取得された口の位置をＳＤＲＡＭ２２から取得する（ステップＳ５０１）。引き続き、前記ステップＳ３１８での説明と同様の処理を行うことにより、検出位置に最も近い背景領域を検出する（ステップＳ５０２）。以下は、前記ステップＳ４０２～Ｓ４０９と同様の処理であり、この検出した検出領域内に収まるような吹き出しを生成し（ステップＳ５０３）、テキストデータを方言に変換する（ステップＳ５０４）。この変換したテキストデータをステップＳ５０３で生成した吹き出し内に合成して、この吹き出しとテキストデータとからなる吹き出しテキストデータを生成する（ステップＳ５０５）。引き続き、表示色変更処理を実行して、この吹き出しテキストデータの表示色を周囲音（音声）の高さに応じて変更する（ステップＳ５０６）。また、表示サイズ変更処理を実行して、この吹き出しテキストデータの表示サイズを、ステップＳ３０９で検出された周囲音（音声）の音量に応じて変更する（ステップＳ５０７）。

【００６４】

また、前記ステップＳ３２１で存在が検出された人に対応する個人名が被写体種別判定テーブル２６１にあるか否かを判断する（ステップＳ５０８）。ある場合には、フォント変更処理を実行して、この吹き出しテキストデータにおけるテキストデータのフォントを、個人名（あるいは性別）に応じて変更する（ステップＳ５０９）。そして、以上の処理により確定した吹き出しテキストデータを前記ステップＳ５０１で検出した検出領域内に、合成する（ステップＳ５１０）。

【００６５】

したがって、この図９に示した第２の吹き出し合成処理により、人の口が写っている場合であってその動きも検出され、音声も検出された場合には、吹き出し内に音声に対応するテキストデータが合成された吹き出しテキストデータが、動きのある人の口の近傍であって背景に表示されることとなる。これにより、図１１（Ｂ）に示すように、加工後の画像には、口Ｐ５を動かしている人Ｐ６の、該口５の近傍であって、他の被写体とは重ならない背景に、吹き出し内に音声に対応するテキストデータ（「おめでとう」）を有する吹き出しテキスト表示データＰが合成される。

【００６６】

なお、本実施の形態においては、予め撮影して記録した音声付き動画ファイルを再生する際に本発明を適用する場合を示したが、音声付き静止画を再生する際、音声付き静止画又は音声付き動画を撮影する際のスルー画像表示時、音声付き静止画又は音声付き動画を撮影記録する際に本発明を適用するようにしてもよい。また、テキストデータを方言に変換するようにしたが、方言に変換することなく合成するようにしてもよい。また、実施の形態においては、本発明をデジタルカメラに適用した場合について示したが、これに限ることなく、撮影機能のみを有するビデオカメラ等の撮影装置、再生機能のみを有するビデオデッキ等の映像機器、画像加工機能のみを有する画像加工機器、撮影機能と再生機能とを併有する各種映像機器に本発明を適用するようにしてもよい。

【図面の簡単な説明】

【００６７】

【図１】本発明の一実施の形態に係るデジタルカメラの電気的構成を示すブロック図である。

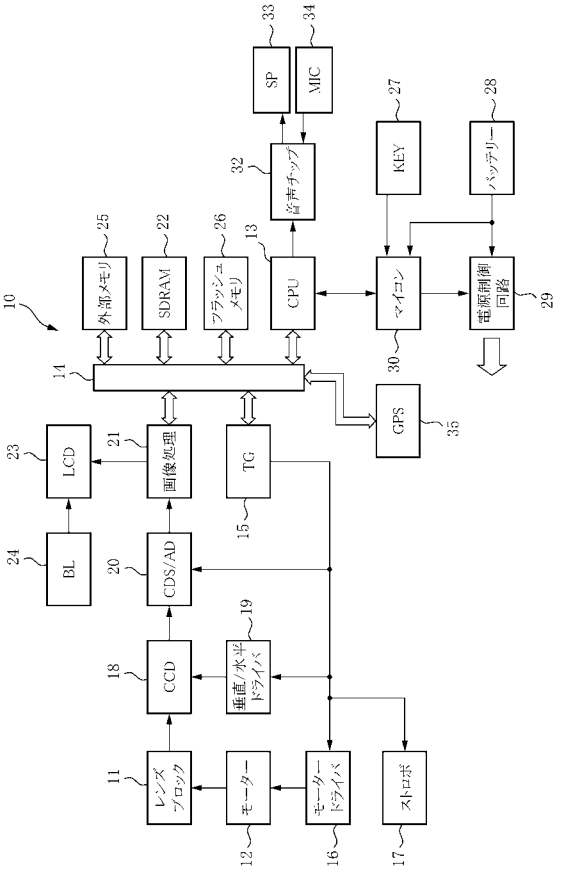
【図２】被写体種別判定テーブルを示す概念図である。

【図３】画像加工モードの処理手順を示すフローチャートである。

【図４】周囲音認識処理の処理手順を示すフローチャートである。

【図５】画像加工処理の処理手順を示すフローチャートである。
【図６】図５に続くフローチャートである。
【図７】図６に続くフローチャートである。
【図８】第１の吹き出し合成処理の処理手順を示すフローチャートである。
【図９】第２の吹き出し合成処理の処理手順を示すフローチャートである。
【図１０】被写体存在判定処理の処理手順を示すフローチャートである。
【図１１】（Ａ）は加工前、（Ｂ）は加工後の画像を示す図である。
【符号の説明】
【００６８】
１０　デジタルカメラ
１１　レンズブロック
１４　バス
１５　タイミング発生器
１８　ＣＣＤ
１９　水平ドライバ
２０　ユニット回路
２１　画像処理部
２２　ＳＤＲＡＭ
２３　液晶モニタ
２５　外部メモリ
２６　フラッシュメモリ
２７　キー入力部
３２　音声チップ
３３　スピーカ
３４　マイクロホン
３５　ＧＰＳ
２６１　被写体種別判定テーブル

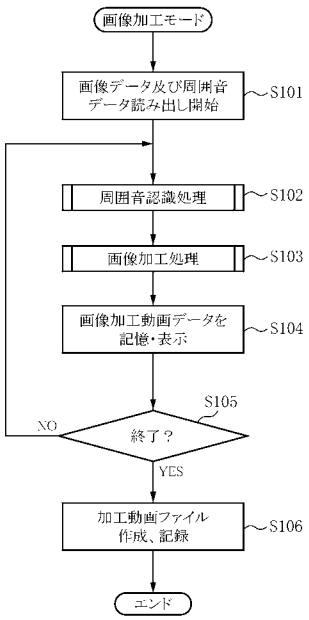
【図１】



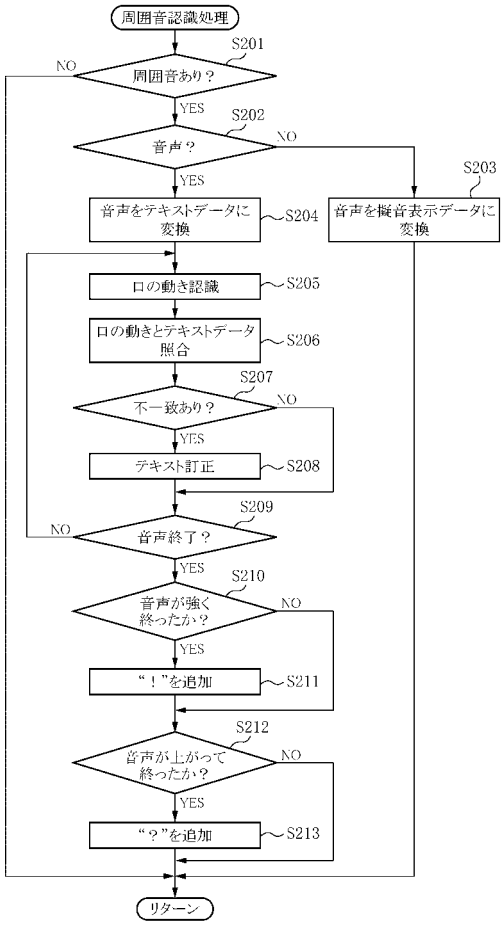
【図２】

被写体種別	特徴量データD
人	－ －
人の口	－ －
人の顔	－ －
自動車	－ －
電話	－ －
クラッカー	－ －
起っている顔	－ －
泣いている顔	－ －
個人名A	－ －
個人名B	－ －

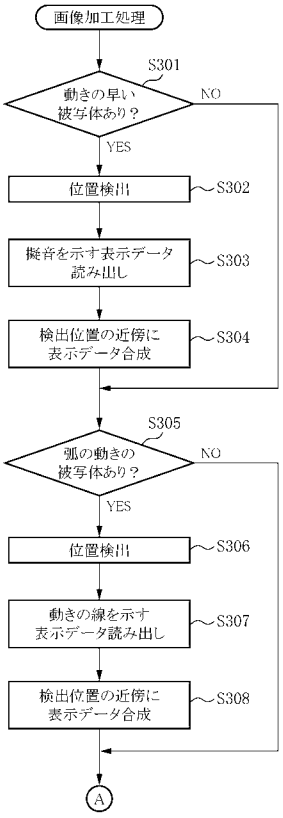
【図３】



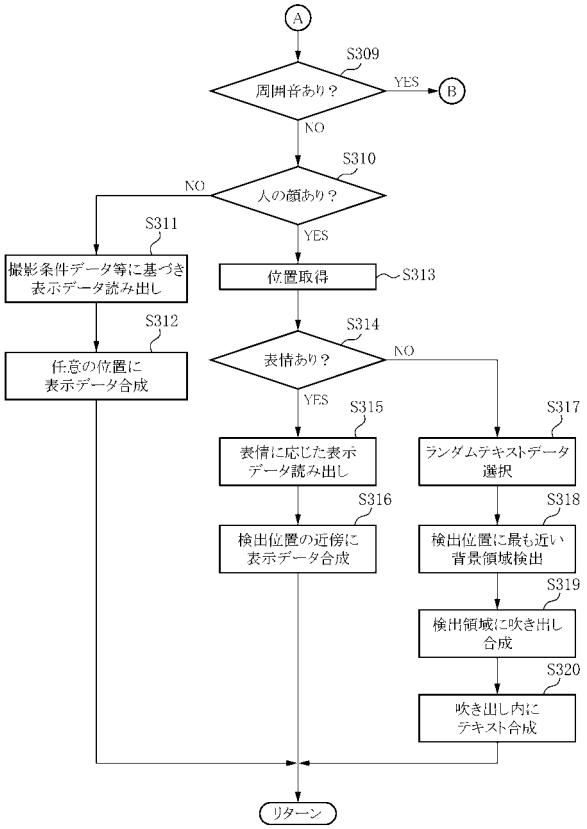
【図４】



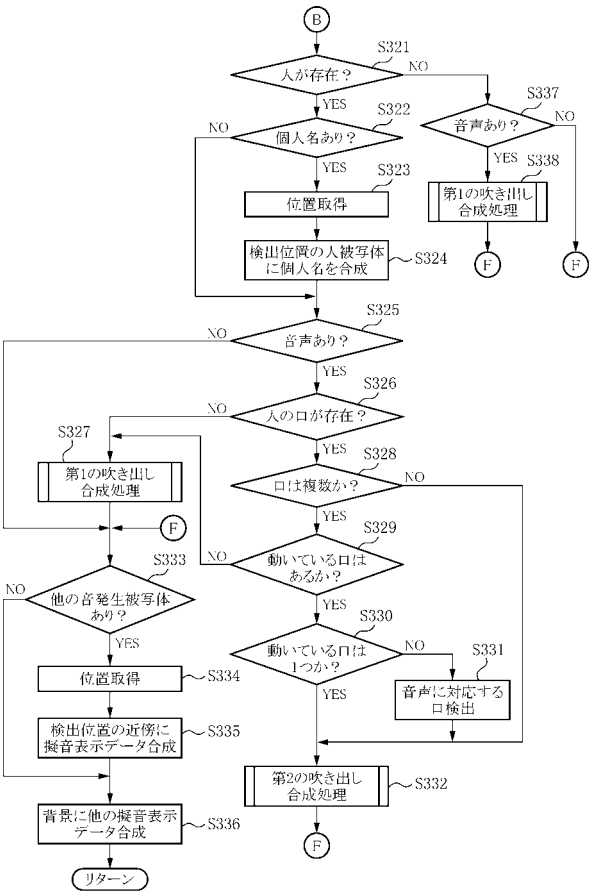
【図５】


画像加工処理
S301
動きの早い被写体あり？
NO
YES
位置検出
S302
擬音を示す表示データ読み出し
S303
検出位置の近傍に表示データ合成
S304
S305
弧の動きの被写体あり？
NO
YES
位置検出
S306
動きの線を示す表示データ読み出し
S307
検出位置の近傍に表示データ合成
S308
A

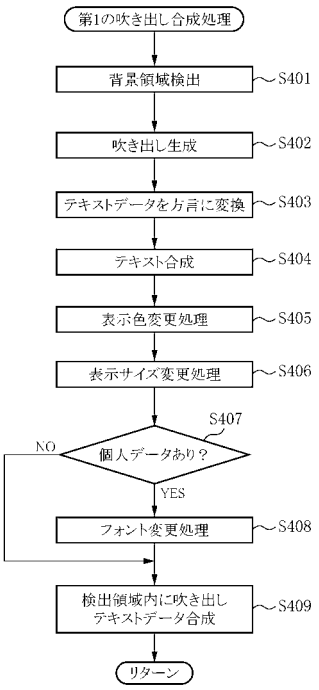
【図６】


A
S309
周囲音あり？
YES
B
NO
S310
人の顔あり？
NO
YES
S311
撮影条件データ等に基づき表示データ読み出し
S312
任意の位置に表示データ合成
位置取得
S313
S314
表情あり？
NO
YES
S315
表情に応じた表示データ読み出し
S316
検出位置の近傍に表示データ合成
S317
ランダムテキストデータ選択
S318
検出位置に最も近い背景領域検出
S319
検出領域に吹き出し合成
S320
吹き出し内にテキスト合成
リターン

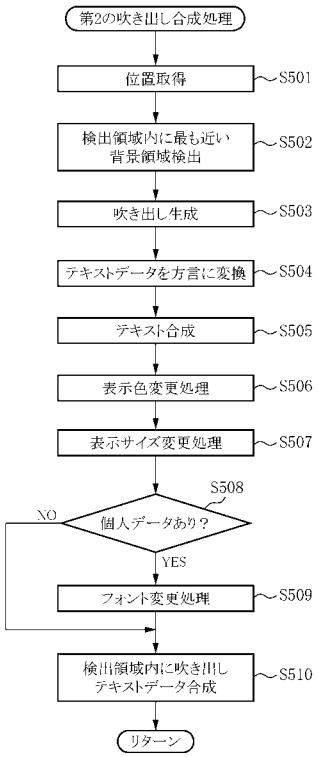
【図７】


B
S321
人が存在？
NO
YES
S322
個人名あり？
NO
YES
S323
位置取得
検出位置の人被写体に個人名を合成
S324
S337
音声あり？
NO
YES
S338
第1の吹き出し合成処理
F
F
S325
音声あり？
NO
YES
S326
人の口が存在？
NO
YES
S327
第1の吹き出し合成処理
F
S328
口は複数か？
NO
YES
S329
動いている口はあるか？
NO
YES
S330
動いている口は1つか？
NO
YES
S331
音声に対応する口検出
S332
第2の吹き出し合成処理
F
S333
他の音発生被写体あり？
NO
YES
位置取得
S334
検出位置の近傍に擬音表示データ合成
S335
背景に他の擬音表示データ合成
S336
リターン

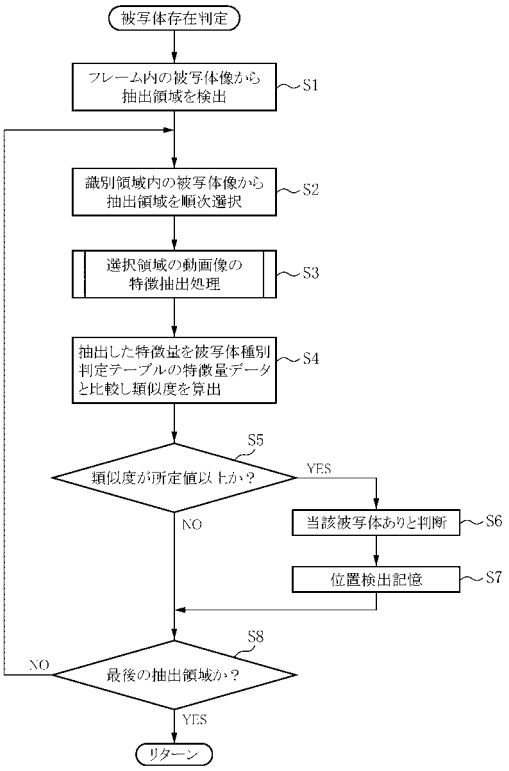
【図８】


第1の吹き出し合成処理
背景領域検出
S401
吹き出し生成
S402
テキストデータを方言に変換
S403
テキスト合成
S404
表示色変更処理
S405
表示サイズ変更処理
S406
S407
個人データあり？
NO
YES
フォント変更処理
S408
検出領域内に吹き出しテキストデータ合成
S409
リターン

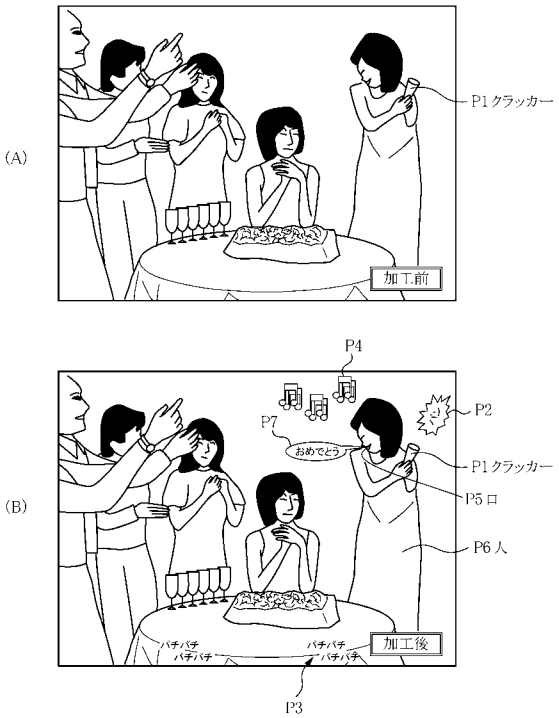
【図９】


第2の吹き出し合成処理
位置取得
S501
検出領域内に最も近い背景領域検出
S502
吹き出し生成
S503
テキストデータを方言に変換
S504
テキスト合成
S505
表示色変更処理
S506
表示サイズ変更処理
S507
S508
個人データあり？
NO
YES
フォント変更処理
S509
検出領域内に吹き出しテキストデータ合成
S510
リターン

【図１０】


被写体存在判定
フレーム内の被写体像から抽出領域を検出
S1
識別領域内の被写体像から抽出領域を順次選択
S2
選択領域の動画像の特徴抽出処理
S3
抽出した特徴量を被写体種別判定テーブルの特徴量データと比較し類似度を算出
S4
S5
類似度が所定値以上か？
YES
NO
当該被写体ありと判断
S6
位置検出記憶
S7
S8
最後の抽出領域か？
NO
YES
リターン

【図１１】


(A)
P1 クラッカー
加工前
(B)
P4
P2
P7
おめでとう
P1 クラッカー
P5 口
P6 人
パチパチ
パチパチ
パチパチ
パチパチ
加工後
P3

フロントページの続き

(51)Int.Cl. F I
H04N 101/00 (2006.01) H04N 101:00

(56)参考文献 特開２００４－０５６２８６（ＪＰ，Ａ）
特開平０８－３１７３６３（ＪＰ，Ａ）
特開２０００－１９６９３５（ＪＰ，Ａ）
特開２００５－１２４１６９（ＪＰ，Ａ）
特開２０００－３３３０８０（ＪＰ，Ａ）
特開２００１－０５１３３８（ＪＰ，Ａ）
特開２００５－０５１２７８（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｈ０４Ｎ ５／２３２
Ｇ０６Ｔ １／００
Ｇ１０Ｌ １５／００
Ｈ０４Ｎ １／３８７
Ｈ０４Ｎ ５／９１
Ｈ０４Ｎ １０１／００