(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2009-141555
(P2009-141555A)
(43)公開日　平成21年6月25日(2009.6.25)

(51)Int.Cl.			F I			テーマコード（参考）
H04N	***5/225***	***(2006.01)***	H04N	5/225	F	5C053
H04N	***5/91***	***(2006.01)***	H04N	5/91	C	5C122

審査請求 未請求　請求項の数 15　ＯＬ　（全 13 頁）

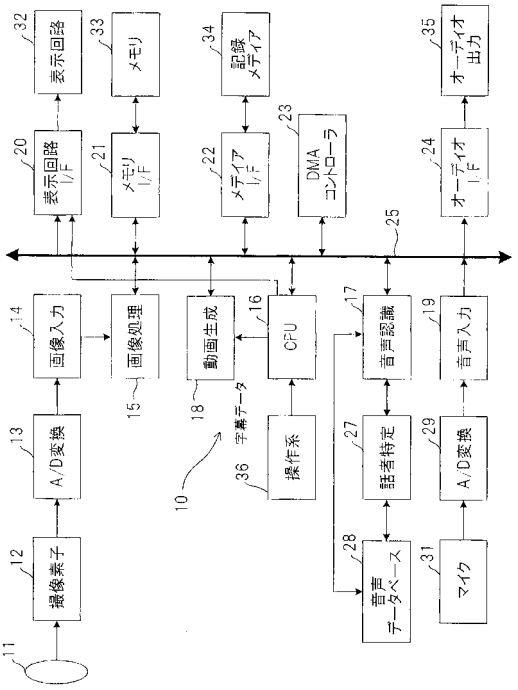
(21)出願番号　特願2007-314454 (P2007-314454)
(22)出願日　平成19年12月5日 (2007.12.5)

(71)出願人　306037311
富士フイルム株式会社
東京都港区西麻布２丁目２６番３０号
(74)代理人　100115107
弁理士　高松　猛
(74)代理人　100132986
弁理士　矢澤　清純
(72)発明者　穐山　俊文
埼玉県朝霞市泉水３丁目１１番４６号　富士フイルム株式会社内

Ｆターム(参考)　5C053 FA08 JA05 JA16 LA01
5C122 DA04 EA61 FJ01 FJ04 FJ15 FK37 FK42 GA20 GA24 GA31 HA04 HB01 HB05

(54)【発明の名称】音声入力機能付き撮像装置及びその音声記録方法

(57)【要約】
【課題】撮影後の編集によらなくても主要被写体の発する音声をテキストデータで被写体画像に合成する。
【解決手段】被写体画像を撮像する撮像手段１２と、撮像手段１２が被写体画像を撮像するとき同時に音声を取り込む音声入力手段３１と、該音声をテキストデータ化する音声認識手段１７と、該テキストデータを被写体画像に合成する画像合成手段１８とを備える。
【選択図】図１



【特許請求の範囲】
【請求項１】
被写体画像を撮像する撮像手段と、該撮像手段が前記被写体画像を撮像するとき同時に音声を取り込む音声入力手段と、該音声をテキストデータ化する音声認識手段と、該テキストデータを前記被写体画像に合成する画像合成手段とを備えることを特徴とする音声入力機能付き撮像装置。
【請求項２】
前記音声の話者を特定する話者特定手段を備え、前記画像合成手段は、特定された前記話者が予め登録された登録者である場合のみ前記合成を行うことを特徴とする請求項１に記載の音声入力機能付き撮像装置。
【請求項３】
前記話者が特定され該話者の音声データを前記テキストデータで合成するとき該話者を特定する名前を該テキストデータと一緒に合成することを特徴とする請求項２に記載の音声入力機能付き撮像装置。
【請求項４】
前記被写体画像中に前記登録者の顔が写っているか否かを識別する顔認識手段を備え、前記画像合成手段は、前記登録者の顔が写っており且つ該登録者が前記話者のとき前記テキストデータを前記被写体画像中の前記登録者の顔画像からの吹き出し形状で合成することを特徴とする請求項１に記載の音声入力機能付き撮像装置。
【請求項５】
前記画像合成手段は、前記被写体画像中に複数の登録者が写っており前記音声が複数人の音声のときは各音声データに対応したテキストデータを各人の画像の近くに吹き出し形状で合成することを特徴とする請求項４に記載の音声入力機能付き撮像装置。
【請求項６】
前記被写体画像中に前記登録者の顔が写っているか否かを識別する顔認識手段を備え、前記画像合成手段は、前記登録者の顔が写っておらず且つ該登録者が前記話者のとき前記テキストデータを字幕データとして合成することを特徴とする請求項１に記載の音声入力機能付き撮像装置。
【請求項７】
前記登録者が前記被写体画像中に複数人写っている場合には該複数人の中の一人の登録者を指定する手段を備え、前記画像合成手段は、指定された登録者の音声を前記テキストデータで該被写体画像中に合成することを特徴とする請求項１に記載の音声入力機能付き撮像装置。
【請求項８】
指向性を持つ前記音声入力手段の該指向性を前記被写体画像中の登録者の方向に制御する制御手段を備えることを特徴とする請求項１に記載の音声入力機能付き撮像装置。
【請求項９】
被写体画像を撮像する撮像手段と、該撮像手段が前記被写体画像を撮像するとき同時に音声を取り込む音声入力手段と、該音声をテキストデータ化する音声認識手段とを備える音声入力機能付き撮像装置の音声記録方法において、前記テキストデータを前記被写体画像に合成することで前記音声を記録すること特徴とする音声入力機能付き撮像装置の音声記録方法。
【請求項１０】
前記音声の話者を特定し、特定された前記話者が予め登録された登録者である場合のみ前記テキストデータの合成を行うことを特徴とする請求項９に記載の音声入力機能付き撮像装置の音声記録方法。
【請求項１１】
前記話者が特定され該話者の音声を前記テキストデータで合成するとき該話者を特定する名前を該テキストデータと一緒に合成することを特徴とする請求項１０に記載の音声入力機能付き撮像装置の音声記録方法。

【請求項１２】
前記被写体画像中に前記登録者の顔が写っているか否かを識別し、前記登録者の顔が写っており且つ該登録者が前記話者のとき前記テキストデータを前記被写体画像中の前記登録者の顔画像からの吹き出し形状で合成することを特徴とする請求項９に記載の音声入力機能付き撮像装置の音声記録方法。
【請求項１３】
前記被写体画像中に複数の登録者が写っており前記音声が複数人の音声のときは各音声データに対応したテキストデータを各人の画像の近くに吹き出し形状で合成することを特徴とする請求項１２に記載の音声入力機能付き撮像装置の音声記録方法。
【請求項１４】
前記被写体画像中に前記登録者の顔が写っているか否かを識別し、前記登録者の顔が写っておらず且つ該登録者が前記話者のとき前記テキストデータを字幕データとして合成することを特徴とする請求項９に記載の音声入力機能付き撮像装置の音声記録方法。
【請求項１５】
前記登録者が前記被写体画像中に複数人写っている場合に、該複数人の中の一人の登録者の指定操作に従って、指定された登録者の音声を前記テキストデータで該被写体画像中に合成することを特徴とする請求項９に記載の音声入力機能付き撮像装置の音声記録方法。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、マイク等の音声入力機能を持った撮像装置及びその音声記録方法に関する。
【背景技術】
【０００２】
特許文献１には、会議などにおける音声をマイクで拾い、この音声をテキスト化して記録する技術が開示されている。また、特許文献２には、デジタルカメラ等で映像の他に音声をマイクで拾い、撮影された画像及び音声を解析して、音声をテキスト化したデータを画像ファイルに関連付けて記録する技術が開示されている。
【０００３】
また、特許文献３には、撮影された動画像データを解析し、顔検出機能により顔の特徴量と位置とを検出し、音声識別機能によって音声の特徴量を検出し、これらを基に動画像データ中の特定話者の位置を特定すると共に特定話者の音声をテキストデータ化し、特定話者が喋った内容を字幕として動画像データ中に合成する技術が開示されている。
【０００４】
また、特許文献４は、画像上の人物と音声との対応付けの精度を向上させ、音声をテキストデータ化して話者に対応させて表示させる技術を開示している。特許文献５は、外部から送信されてきた画像データおよび音声データを受信し、音声をテキストデータ化して画面上に表示する技術を開示する。
【０００５】
特許文献６は、入力された音声をテキストデータ化し、カメラ部で撮影された画像に対応させて記録させると共に、所望の画像を対応のテキストデータで検索できる様にする技術を開示する。
【特許文献１】特開２００６－１８９６２６号公報
【特許文献２】特開２００６－１３３４３３号公報
【特許文献３】特開２００７－２７９９０号公報
【特許文献４】特開２００４－５６２８６号公報
【特許文献５】特開平９－２３３４４２号公報
【特許文献６】特開２００５－２０４４０号公報
【発明の開示】
【発明が解決しようとする課題】

【０００６】
上述した各特許文献に記載されている様に、マイクで拾った音声をテキストデータ化する技術が一般的となり、撮像された画像に対応付けて音声を字幕として、あるいは吹き出しとして、表示することが行われる様になってきている。
【０００７】
しかし、従来技術は、いずれも、撮影済みの画像を後でパソコン等を用いて解析して、話者を特定すると共に音声をテキストデータ化し、両者を対応付ける解析を行い、字幕表示や吹き出し表示を行うという、撮影後の編集処理であるため、面倒であるという問題がある。また、話者と音声との対応付けの精度が低く、撮影中の主要人物の音声を高精度に抽出して字幕表示や吹き出し表示を精度良く対応付けて行うことができないという問題がある。
【０００８】
本発明の目的は、撮影後の編集によらなくても主要被写体の発する音声をテキストデータで被写体画像に合成することができ、また、撮影中の主要人物の音声を高精度に対応つけて字幕表示や吹き出し表示することができる音声入力機能付き撮像装置及びその音声記録方法を提供することにある。
【課題を解決するための手段】
【０００９】
本発明の音声入力機能付き撮像装置は、被写体画像を撮像する撮像手段と、該撮像手段が前記被写体画像を撮像するとき同時に音声を取り込む音声入力手段と、該音声をテキストデータ化する音声認識手段と、該テキストデータを前記被写体画像に合成する画像合成手段とを備えることを特徴とする。
【００１０】
本発明の音声入力機能付き撮像装置は、前記音声の話者を特定する話者特定手段を備え、前記画像合成手段は、特定された前記話者が予め登録された登録者である場合のみ前記合成を行うことを特徴とする。
【００１１】
本発明の音声入力機能付き撮像装置は、前記話者が特定され該話者の音声データを前記テキストデータで合成するとき該話者を特定する名前を該テキストデータと一緒に合成することを特徴とする。
【００１２】
本発明の音声入力機能付き撮像装置は、前記被写体画像中に前記登録者の顔が写っているか否かを識別する顔認識手段を備え、前記画像合成手段は、前記登録者の顔が写っており且つ該登録者が前記話者のとき前記テキストデータを前記被写体画像中の前記登録者の顔画像からの吹き出し形状で合成することを特徴とする。
【００１３】
本発明の音声入力機能付き撮像装置の前記画像合成手段は、前記被写体画像中に複数の登録者が写っており前記音声が複数人の音声のときは各音声データに対応したテキストデータを各人の画像の近くに吹き出し形状で合成することを特徴とする。
【００１４】
本発明の音声入力機能付き撮像装置は、前記被写体画像中に前記登録者の顔が写っているか否かを識別する顔認識手段を備え、前記画像合成手段は、前記登録者の顔が写っておらず且つ該登録者が前記話者のとき前記テキストデータを字幕データとして合成することを特徴とする。
【００１５】
本発明の音声入力機能付き撮像装置は、前記登録者が前記被写体画像中に複数人写っている場合には該複数人の中の一人の登録者を指定する手段を備え、前記画像合成手段は、指定された登録者の音声を前記テキストデータで該被写体画像中に合成することを特徴とする。
【００１６】

本発明の音声入力機能付き撮像装置は、指向性を持つ前記音声入力手段の該指向性を前記被写体画像中の登録者の方向に制御する制御手段を備えることを特徴とする。

【００１７】

本発明の音声入力機能付き撮像装置の音声記録方法は、被写体画像を撮像する撮像手段と、該撮像手段が前記被写体画像を撮像するとき同時に音声を取り込む音声入力手段と、該音声をテキストデータ化する音声認識手段とを備える音声入力機能付き撮像装置の音声記録方法において、前記テキストデータを前記被写体画像に合成することで前記音声を記録すること特徴とする。

【００１８】

本発明の音声入力機能付き撮像装置の音声記録方法は、前記音声の話者を特定し、特定された前記話者が予め登録された登録者である場合のみ前記テキストデータの合成を行うことを特徴とする。

【００１９】

本発明の音声入力機能付き撮像装置の音声記録方法は、前記話者が特定され該話者の音声を前記テキストデータで合成するとき該話者を特定する名前を該テキストデータと一緒に合成することを特徴とする。

【００２０】

本発明の音声入力機能付き撮像装置の音声記録方法は、前記被写体画像中に前記登録者の顔が写っているか否かを識別し、前記登録者の顔が写っており且つ該登録者が前記話者のとき前記テキストデータを前記被写体画像中の前記登録者の顔画像からの吹き出し形状で合成することを特徴とする。

【００２１】

本発明の音声入力機能付き撮像装置の音声記録方法は、前記被写体画像中に複数の登録者が写っており前記音声が複数人の音声のときは各音声データに対応したテキストデータを各人の画像の近くに吹き出し形状で合成することを特徴とする。

【００２２】

本発明の音声入力機能付き撮像装置の音声記録方法は、前記被写体画像中に前記登録者の顔が写っているか否かを識別し、前記登録者の顔が写っておらず且つ該登録者が前記話者のとき前記テキストデータを字幕データとして合成することを特徴とする。

【００２３】

本発明の音声入力機能付き撮像装置の音声記録方法は、前記登録者が前記被写体画像中に複数人写っている場合に、該複数人の中の一人の登録者の指定操作に従って、指定された登録者の音声を前記テキストデータで該被写体画像中に合成することを特徴とする。

【発明の効果】

【００２４】

本発明によれば、撮像装置で被写体画像を撮影するときに撮影と同時に主要被写体の発する音声をテキストデータ化して被写体画像中に合成するため、使い勝手の優れた撮像装置を提供することが可能となる。また、登録者の場合にのみ音声をテキストデータで画像に合成するため、主要被写体と音声との対応付けの精度が向上する。

【発明を実施するための最良の形態】

【００２５】

以下、本発明の一実施形態について、図面を参照して説明する。

【００２６】

図１は、本発明の第１実施形態に係る撮像装置の機能ブロック図である。この撮像装置１０は、撮影レンズ１１と、この撮影レンズ１１を通して結像された光像に応じた電気信号を出力する撮像素子１２と、撮像素子１２のアナログの撮像画像信号をデジタル信号に変換するアナログデジタル変換回路１３と、デジタルの撮像画像信号を取り込む画像入力インタフェース１４と、画像入力インタフェース１４から取り込まれた撮像画像信号を処理する画像処理手段１５とを備える。

【００２７】

この撮像装置１０は、更に、撮像装置１０の全体を統括制御するシステム制御部（ＣＰＵ）１６と、音声認識手段１７と、動画像生成手段１８と、音声入力インタフェース１９と、表示回路インタフェース２０と、メモリインタフェース２１と、メディアインタフェース２２と、ＤＭＡ（ダイレクトメモリアクセス）コントローラ２３と、オーディオインタフェース２４と、これらを相互接続するバス２５とを備える。

【００２８】

音声認識手段１７には、話者特定手段２７と、この特定手段２７に接続された音声データベース２８とが接続され、音声入力インタフェース１９には、アナログデジタル変換回路２９を介して音声入力手段であるマイク３１が接続される。

【００２９】

表示回路インタフェース２０には撮像装置１０の背面等に取り付けられる液晶等の表示回路３２が接続され、メモリインタフェース２１には撮像装置１０のメインメモリとなるフレームメモリ３３が接続され、メディアインタフェース２２には、着脱自在の外部メモリ（記録メディア）３４が接続され、オーディオインタフェース２４にはスピーカ等のオーディオ出力手段３５が接続される。

【００３０】

ＣＰＵ１６は、音声認識手段１７が音声認識しテキストデータ化した字幕データや吹き出しデータを、動画生成手段１８や表示回路インタフェース２０に出力する様になっている。また、ＣＰＵ１６には、シャッタボタンなどの操作系３６の信号が入力される。ＣＰＵ１６が音声認識手段１７の機能を実行する構成としても良い。

【００３１】

図１に示す音声データベース２８には、例えば、撮像装置１０のユーザが良く撮る家族等の各人の音声が、各人の名前と対応付けて登録されている。

【００３２】

図２は、図１に示す撮像装置１０が実行する処理手順を示すフローチャートである。動画像の記録が開始される（ステップＳ１）と、撮像素子１２から被写体画像が取り込まれると共に、マイク３１から音声が取り込まれる（ステップＳ２）。音声認識手段１７はこの音声を解析すると共にデータベース２８を参照し（ステップＳ３）、この音声がデータベース２８に登録されている音声であるか否かを判定する（ステップＳ４）。

【００３３】

マイク３１から取り込んだ音声が登録者の音声でないと判定した場合には、そのままステップＳ８に進む。登録者の音声であると判定した場合には、ステップＳ４からステップＳ５に進み、話者特定手段２７は、データベース２８から話者が誰であるかを特定し、次のステップＳ６で、音声認識手段１７は、その音声をテキストデータに変換する。

【００３４】

そして、次のステップＳ７で、ＣＰＵ１６は、テキストデータ化された音声データを動画生成手段１８と表示回路インタフェース２０とに渡し、動画生成手段１８は、図３に示す様に、動画像上に、特定された話者の名前“Ａ”と、その音声データの字幕データとを合成して、記録メディア３４に書き込み、ステップＳ８に進む。

【００３５】

尚、音声の字幕データを合成した画像データを記録メディア３４に書き込むのではなく、無線で外部記憶手段に伝送する構成としても良い。これは以下の実施形態でも同様である。

【００３６】

ステップＳ８では、動画記録が終了したか否かを判定し、終了した場合にはこの図２の処理を終了し、動画記録が終了しない場合にはステップＳ２に戻り、ステップＳ１～ステップＳ７の処理を繰り返す。

【００３７】

表示回路３２には、動画像データが表示されるが、このとき、字幕データと話者の名前とが重ねて表示される。

【００３８】
尚、音声を字幕データとして記録する部分についてのみ説明したが、マイク３１から取り込んだ音声のままのデータでも記録することはいうまでもない。音声のままのデータを記録する場合には、音声データベース上に登録されている話者であるか否かに関係なく全て記録する。これは以下の実施形態でも同様である。
【００３９】
図４は、本発明の第２実施形態に係る撮像装置の機能ブロック図である。この撮像装置４０は、図１に示す撮像装置１０と殆どの機能が重複し、静止画像撮像用である点のみ異なる。このため、重複する部分には同一符号を付してその説明は省略し、異なる部分についてのみ説明する。
【００４０】
本実施形態の撮像装置４０は、静止画像撮像専用であり、図１に示した動画生成手段１８が設けられておらず、ＣＰＵ１６は、字幕データを画像処理部１５に渡す様になっている。
【００４１】
図５は、本実施形態に係る撮像装置が実行する処理手順を示すフローチャートである。先ず、２段シャッタボタンが半押し状態になっているか否かを判定する（ステップＳ１１）。半押し状態になっていなければ繰り返しステップＳ１１を実行し、半押し状態になっている場合には、音声をマイク３１から取り込む（ステップＳ１２）。次に、シャッタボタンが全押しになかったか否かを判定し（ステップＳ１３）、全押しになっていない場合にはステップＳ１２を繰り返し実行する。
【００４２】
シャッタボタンが全押しになった場合にはステップＳ１３からステップＳ１４に進み、マイク３１から取り込んだ音声データを記録し、次に、音声データベース２８を参照する（ステップＳ１５）。また、シャッタボタンが全押しされた場合には、ステップＳ１４以下の処理と並行して、被写体画像の撮像処理が行われる。
【００４３】
そして、音声認識手段１７は、この音声がデータベース２８に登録されている音声であるか否かを判定し（ステップＳ１６）、マイク３１から取り込んだ音声が登録者の音声でないと判定した場合には、この図５の処理を終了する。登録者の音声であると判定した場合には、ステップＳ１６からステップＳ１７に進み、話者特定手段２７は、データベース２８から話者が誰であるかを特定し、次のステップＳ１８で、音声認識手段１７は、その音声をテキストデータに変換する。
【００４４】
そして、次のステップＳ１９で、ＣＰＵ１６は、テキストデータ化された音声データを画像処理部１５と表示回路インタフェース２０とに渡し、画像処理部１５は、図３に示す様に、静止画像上に、特定された話者の名前“Ａ”と、その音声データの字幕データとを合成して、記録メディア３４に書き込み、図５の処理を終了する。表示回路３２には、静止画像データが表示されるが、このとき、字幕データと話者の名前とが重ねて表示される。
【００４５】
図６は、本発明の第３実施形態に係る撮像装置５０の機能ブロック図である。本実施形態の撮像装置５０は、図１に示す撮像装置１０と殆どの機能が重複するため、同一機能ブロックには同一符号を付してその説明は省略し、異なる部分についてのみ説明する。
【００４６】
本実施形態に係る撮像装置５０は、画像処理部１５の処理結果を受け取り撮像画像中に人間の「顔」が存在するか否かを検出すると共に、検出した「顔」が登録されている特定者の「顔」であるか否かを検出する顔検出／顔認識処理手段４１と、この顔検出／顔認識処理手段４１が人間の「顔」として検出するとき使用する顔検出用データ及び登録者（例えば家族等）の顔画像を登録者の名前と対応付けて格納した顔データベース４２とを備え

る。顔検出／顔認識処理手段４１はバス２５に接続されている。

【００４７】

図７は、本実施形態に係る撮像装置が実行する処理手順を示すフローチャートである。動画像の記録が開始される（ステップＳ２１）と、撮像素子１２から被写体画像が取り込まれると共に、マイク３１から音声が取り込まれる（ステップＳ２２）。音声認識手段１７はこの音声を解析すると共にデータベース２８を参照し（ステップＳ２３）、この音声がデータベース２８に登録されている音声であるか否かを判定する（ステップＳ２４）。

【００４８】

マイク３１から取り込んだ音声が登録者の音声でないと判定した場合には、ステップＳ２２に戻って次の音声入力を待機し、登録者の音声であると判定した場合には、ステップＳ２４からステップＳ２５に進み、話者特定手段２７は、データベース２８から話者が誰であるかを特定すると共に、音声認識手段１７はその音声をテキストデータに変換する。

【００４９】

次に、顔検出／顔認識処理手段４１は、撮像された画像の中に人間の「顔」が存在するか否かを解析し（ステップＳ２６）、「顔」が検出されなかった場合には、ステップＳ２５でテキストデータ化された音声データを図３に示す様に字幕として合成し（ステップＳ３０）、ステップＳ３１に進む。

【００５０】

人間の「顔」が検出された場合にはステップＳ２６からステップＳ２７に進んで顔データベース４２を参照する。そして、次にステップＳ２８では、検出された「顔」が登録者の「顔」であるか否か、及び、登録者である場合にはその登録者が、ステップＳ２５でテキストデータ化した音声の特定話者であるか否かを判定する（ステップＳ２８）。

【００５１】

登録者の顔で無い場合、あるいは特定話者と一致しない登録者の顔である場合には、上記のステップＳ３０に進み、テキストデータ化された音声データを字幕として合成する。

【００５２】

ステップＳ２８の判定の結果、登録者の顔であり、且つ特定話者と一致すると判定した場合には、次にステップＳ２９に進み、テキストデータ化した音声データを、特定話者と一致する撮像画像中の登録者の顔画像の近くに、図８に示す様に、吹き出し形状で合成し、ステップＳ３１に進む。

【００５３】

ステップＳ３１では、動画記録が終了したか否かを判定し、終了した場合にはこの図７の処理を終了し、動画記録が終了しない場合にはステップＳ２２に戻り、ステップＳ２２～ステップＳ３０の処理を繰り返す。

【００５４】

吹き出し形状で音声のテキストデータを画像に合成し表示する場合、音声データが画像データと重なることになる。このため、下の画像が見づらくならないように、吹き出し位置，字の大きさ、透明度等を設定可能にするのが好ましい。

【００５５】

この様にすることで、動画記録時に画面上に複数人が撮影されている場合でも自動的に登録者，話者を特定して各人に対応した吹き出し合成を行うため、画像を見れば誰が喋った内容か直ぐ分かるようになる。

【００５６】

尚、図７では、音声による話者特定とテキストデータ化を先に行い、その後に顔検出，顔認識を行ったが、これを逆に行っても、また同時並行的に行っても良いことは言うまでもない。

【００５７】

図９は、本発明の第４実施形態に係る撮像装置６０の機能ブロック図である。本実施形態の撮像装置６０は、図６に示す撮像装置５０と殆どの機能が重複するため、同一機能ブロックには同一符号を付してその説明は省略し、異なる部分についてのみ説明する。

【００５８】
本実施形態の撮像装置６０は、マイク３１が指向性を持ち、且つその指向性の制御（集音範囲，集音方向，集音距離（感度）等の制御）が可能なマイクであり、マイク制御手段４３がマイク３１をＣＰＵ１６からの指示により制御する構成になっている。また、図６に示す話者特定手段２７と音声データベース２８とがこの撮像装置６０には設けられていない。
【００５９】
図１０は、本実施形態に係る撮像装置が実行する処理手順を示すフローチャートである。動画像の記録が開始される（ステップＳ４１）と、撮像素子１２から被写体画像が取り込まれる。次のステップＳ４２では、撮像画像中に登録者の「顔」が存在するか否かを判定し、登録者の顔が検出されるまでステップＳ４２を繰り返し実行する。
【００６０】
画面中に登録者の顔が検出された場合には、ステップＳ４２からステップＳ４３に進み、今度は１画面の中に複数の登録者の顔が存在するか否かを判定する。複数の登録者の顔が検出された場合には、ステップＳ４３からステップＳ４４に進み、ユーザの顔選択処理を行った後、ステップＳ４５に進む。１画面の中に一人の登録者の顔しか検出されない場合にはステップＳ４４を飛び越してステップＳ４５に進む。
【００６１】
図１１は、１画面の中に３人の人間が撮像されている状態を示しており、そのうちの二人の顔が名前“Ａ”“Ｂ”の登録者であり、もう一人が登録者でない人の場合を示している。図９の顔検出／顔認識処理手段４１は、検出した顔部分を矩形枠で示すため、図１０のステップＳ４４では、ユーザは、登録者Ａ，Ｂのいずれか一方を操作系３６のボタン操作により指定することになる。図１１に示す例では、登録者Ａが指定されたため矩形枠を二重枠で表示したところを示している。
【００６２】
ステップＳ４５では、画面中の一人の登録者あるいはステップＳ４４で指定された登録者に対して、マイク制御を行う。この登録者の撮像画像から撮像装置と登録者との間の距離，登録者が居る方向がズーム倍率等で判別できるため、この登録者が喋る音声を精度良く集音できるように、ＣＰＵ１６はマイク３１の指向性制御を行う。
【００６３】
次のステップＳ４６では、マイク３１から集音した音声データを取り込み、ステップＳ４７でこれをテキストデータ化し、ステップＳ４８でテキストデータした字幕（勿論、吹き出しでも良い。）を図１１に示す様に画像上に合成し、ステップＳ４９に進む。
【００６４】
ステップＳ４９では、動画記録が終了したか否かを判定し、終了した場合にはこの図１０の処理を終了し、動画記録が終了していない場合にはステップＳ４２に戻り、ステップＳ４３～ステップＳ４８を繰り返し実行する。この繰り返し時には、ステップＳ４４のユーザ指定は、何らかのユーザによるボタン操作が無い限り実行しない構成とすることで、指定した登録者をずっと追って集音することが可能となる。
【００６５】
尚、図１０の実施形態では、ユーザによる顔選択処理（ステップＳ４４）を動画記録時に行ったが、動画記録前に予め選択する構成としても良い。このステップＳ４４における選択時に、図１１に示したように、選択対象者の顔を二重枠で区別したり、人物名を画像中に表示することで、選択が容易且つ確実に行える様にするのが良い。
【００６６】
この様に、本実施形態によれば、ユーザが記録したい被写体の音声のみを取り込んで字幕化することが可能となる。
【００６７】
以上述べた様に、本発明の各実施形態によれば、画像の撮像中に、音声をテキストデータ化し、該当する被写体画像に対応付けて合成するため、話者と音声データ（テキストデ

ータ）との対応付けの精度が向上すると共に、後で編集する手間が省け、デジタルカメラ等の撮像装置の使い勝手が向上する。

【産業上の利用可能性】

【００６８】

本発明に係る撮像装置は、話者と音声データとの対応付けの精度が向上し字幕や吹き出しとして音声のテキストデータを撮像中の画像に合成できるため、デジタルスチルカメラやビデオカメラ等に適用すると有用である。

【図面の簡単な説明】

【００６９】

【図１】本発明の第１実施形態に係る撮像装置の機能ブロック図である。

【図２】図１に示す撮像装置の処理手順を示すフローチャートである。

【図３】図２の処理手順により音声データが字幕表示された画像を示す図である。

【図４】本発明の第２実施形態に係る撮像装置の機能ブロック図である。

【図５】図４に示す撮像装置の処理手順を示すフローチャートである。

【図６】本発明の第３実施形態に係る撮像装置の機能ブロック図である。

【図７】図６に示す撮像装置の処理手順を示すフローチャートである。

【図８】図７に示す処理手順により音声データが吹き出し表示された画像を示す図である。

【図９】本発明の第４実施形態に係る撮像装置の機能ブロック図である。

【図１０】図９に示す撮像装置の処理手順を示すフローチャートである。

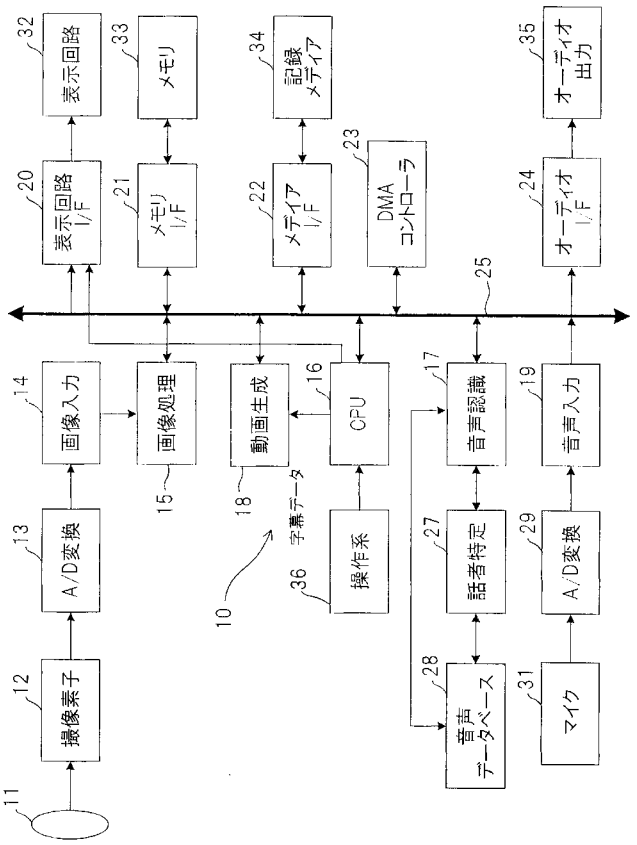
【図１１】図１０に示す処理手順により音声データが字幕表示された画像を示す図である。

【符号の説明】

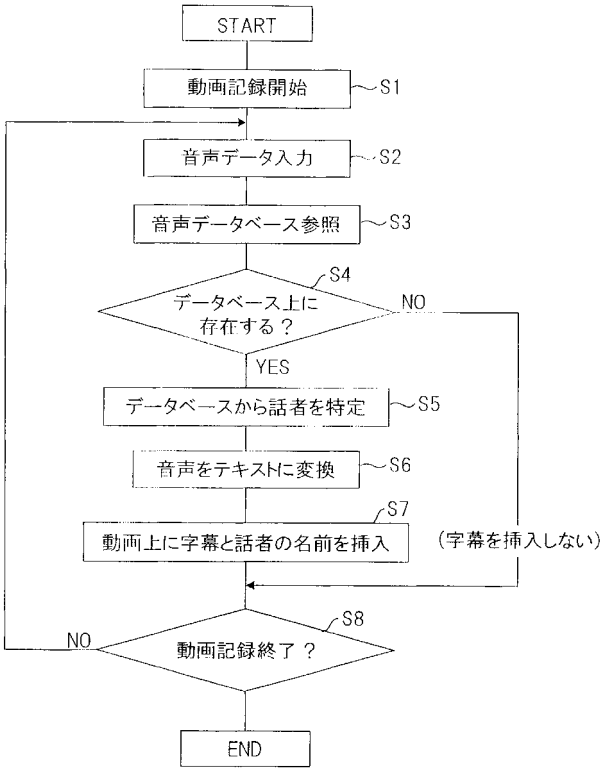
【００７０】

１０，４０，５０，６０　撮像装置
１２　撮像素子
１５　画像処理手段（画像合成手段）
１６　ＣＰＵ
１７　音声認識手段
１８　動画生成手段（画像合成手段）
２７　話者特定手段
２８　音声データベース
３１　マイク（音声入力手段）
４１　顔検出／顔認識処理手段
４２　顔データベース
４３　マイク制御手段

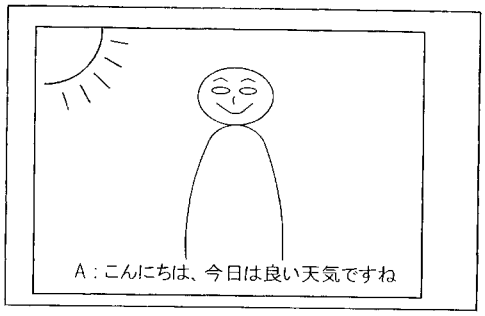
【図１】


表示回路
メモリ
記録メディア
オーディオ出力
表示回路I/F
メモリI/F
メディアI/F
DMAコントローラ
オーディオI/F
画像入力
画像処理
動画生成
CPU
音声認識
音声入力
A/D変換
操作系
話者特定
A/D変換
撮像素子
音声データベース
マイク
字幕データ

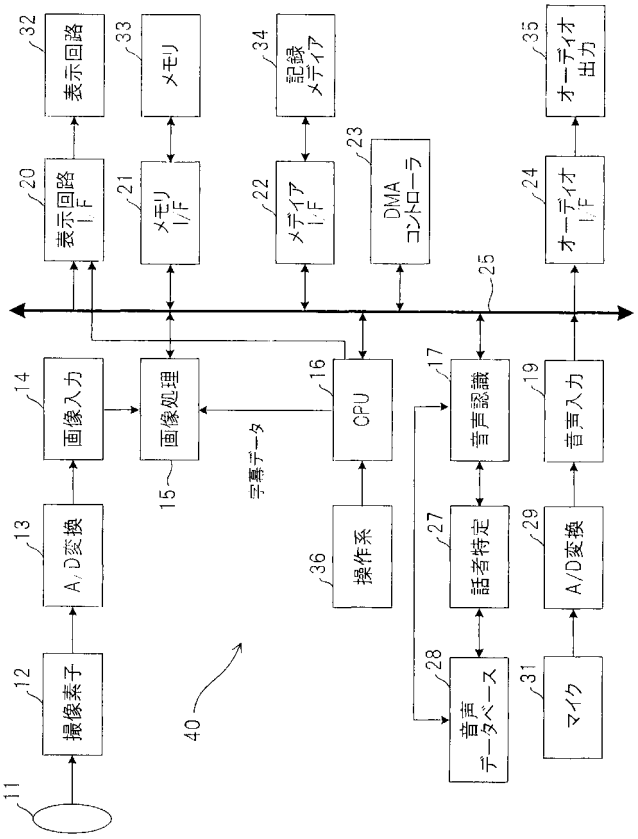
【図２】


START
動画記録開始
S1
音声データ入力
S2
音声データベース参照
S3
データベース上に存在する？
S4
NO
YES
データベースから話者を特定
S5
音声をテキストに変換
S6
動画上に字幕と話者の名前を挿入
S7
（字幕を挿入しない）
動画記録終了？
S8
NO
END

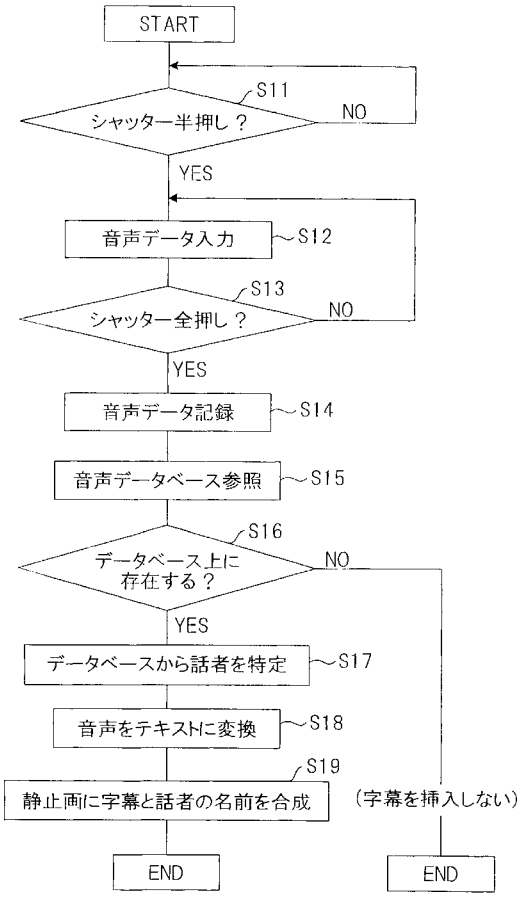
【図３】


A：こんにちは、今日は良い天気ですね

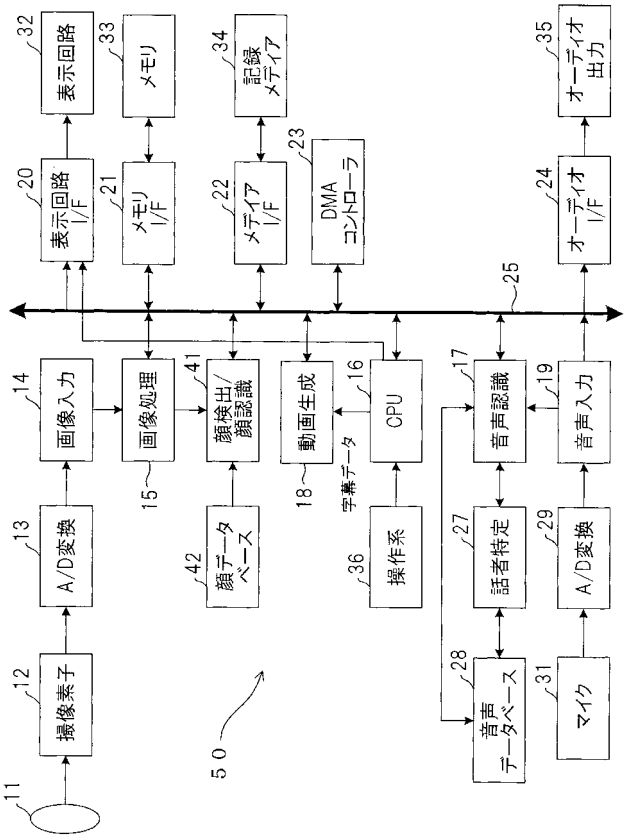
【図４】


表示回路
メモリ
記録メディア
オーディオ出力
表示回路I/F
メモリI/F
メディアI/F
DMAコントローラ
オーディオI/F
画像入力
画像処理
CPU
音声認識
音声入力
A/D変換
操作系
話者特定
A/D変換
撮像素子
音声データベース
マイク
字幕データ

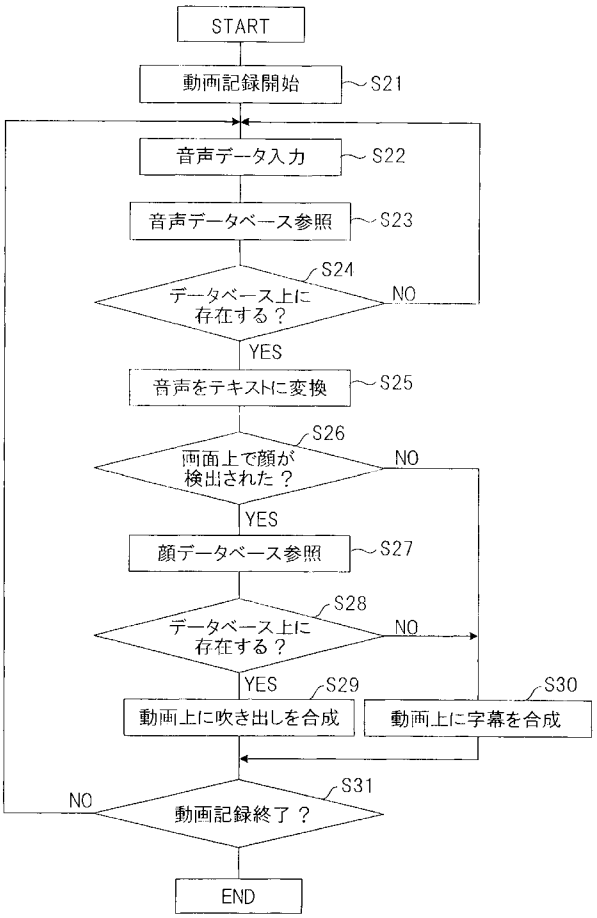
【図５】


START
S11
シャッター半押し？
NO
YES
音声データ入力
S12
S13
シャッター全押し？
NO
YES
音声データ記録
S14
音声データベース参照
S15
S16
データベース上に存在する？
NO
YES
データベースから話者を特定
S17
音声をテキストに変換
S18
S19
静止画に字幕と話者の名前を合成
（字幕を挿入しない）
END
END

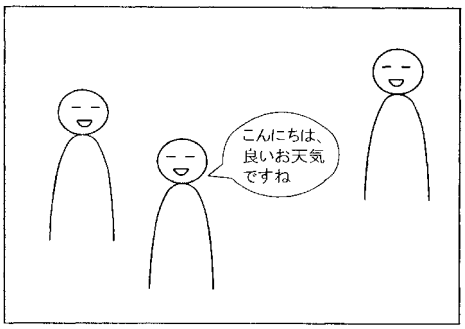
【図６】


32 表示回路
33 メモリ
34 記録メディア
35 オーディオ出力
20 表示回路I/F
21 メモリI/F
22 メディアI/F
23 DMAコントローラ
24 オーディオI/F
25
14 画像入力
15 画像処理
41 顔検出/顔認識
18 動画生成
16 CPU
17 音声認識
19 音声入力
字幕データ
13 A/D変換
42 顔データベース
36 操作系
27 話者特定
29 A/D変換
12 撮像素子
28 音声データベース
31 マイク
11
50

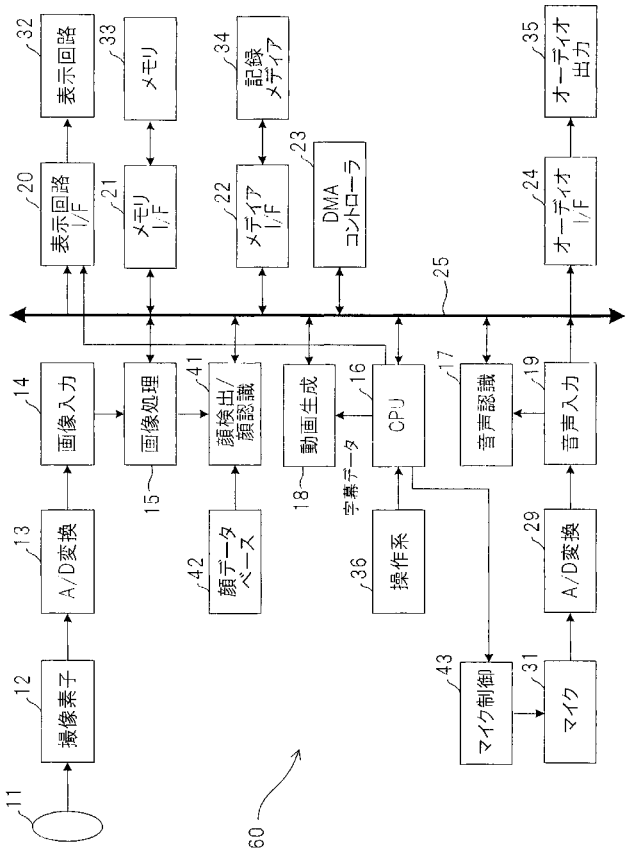
【図７】


START
動画記録開始
S21
音声データ入力
S22
音声データベース参照
S23
S24
データベース上に存在する？
NO
YES
音声をテキストに変換
S25
S26
画面上で顔が検出された？
NO
YES
顔データベース参照
S27
S28
データベース上に存在する？
NO
YES
S29
動画上に吹き出しを合成
S30
動画上に字幕を合成
S31
NO
動画記録終了？
END

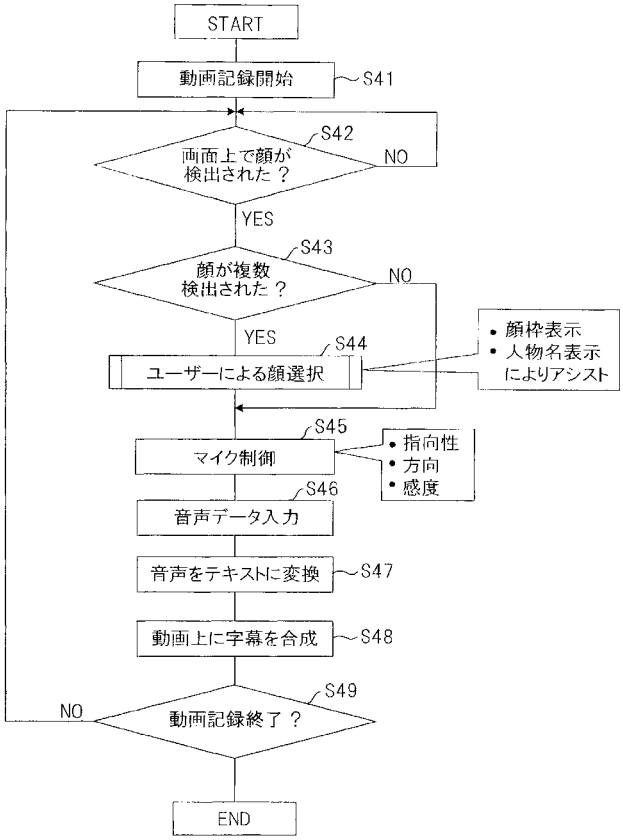
【図８】


こんにちは、良いお天気ですね

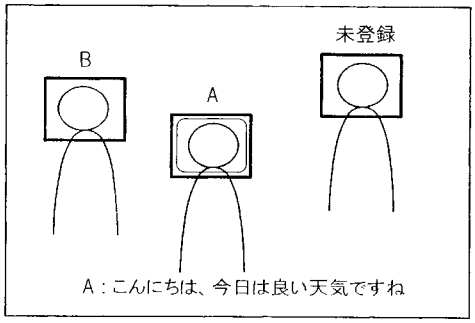
【図９】


表示回路
メモリ
記録メディア
オーディオ出力
表示回路I/F
メモリI/F
メディアI/F
DMAコントローラ
オーディオI/F
25
画像入力
画像処理
顔検出/顔認識
動画生成
CPU
音声認識
音声入力
A/D変換
顔データベース
操作系
A/D変換
撮像素子
マイク制御
マイク
字幕データ
11
60

【図１０】


START
動画記録開始
S41
画面上で顔が検出された？
S42
NO
YES
顔が複数検出された？
S43
NO
YES
ユーザーによる顔選択
S44
• 顔枠表示
• 人物名表示
によりアシスト
マイク制御
S45
• 指向性
• 方向
• 感度
音声データ入力
S46
音声をテキストに変換
S47
動画上に字幕を合成
S48
動画記録終了？
S49
NO
END

【図１１】


B
A
未登録
A：こんにちは、今日は良い天気ですね