

(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2016-208200

(P2016-208200A)

(43) 公開日 平成28年12月8日(2016.12.8)

(51) Int.Cl.	F I	テーマコード (参考)
HO4N 5/225 (2006.01)	HO4N 5/225	F 5C053
HO4N 5/232 (2006.01)	HO4N 5/232	Z 5C122
G1OL 21/0332 (2013.01)	HO4N 5/225	A
HO4N 5/91 (2006.01)	G1OL 21/0332	C
	HO4N 5/91	

審査請求 未請求 請求項の数 13 O L (全 15 頁)

(21) 出願番号 特願2015-86328 (P2015-86328)
 (22) 出願日 平成27年4月21日 (2015. 4. 21)

(71) 出願人 000001007
 キヤノン株式会社
 東京都大田区下丸子3丁目30番2号
 (74) 代理人 100086818
 弁理士 高梨 幸雄
 (72) 発明者 山内 裕子
 東京都大田区下丸子3丁目30番2号 キヤノン株式会社内
 Fターム(参考) 5C053 GB06 JA01 LA01
 5C122 DA03 EA42 FH11 FJ01 FJ03
 FJ04 FJ06 FJ15 FK24 FK41
 GA34 HA01 HB01 HB05 HB07

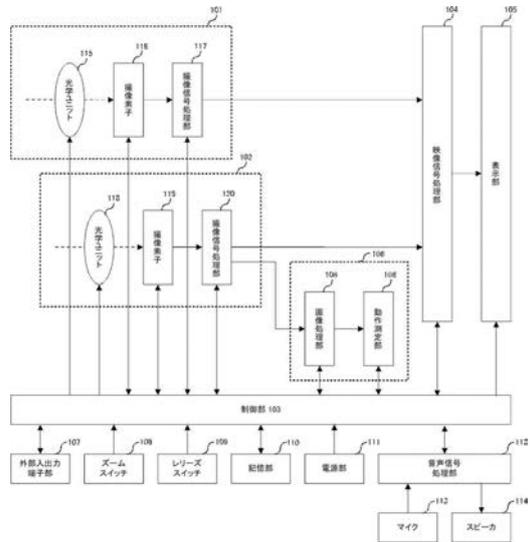
(54) 【発明の名称】 撮像装置

(57) 【要約】 (修正有)

【課題】 撮影者の声を認識し、その期間の音量を視聴者の意図に合わせて調節する撮像装置を提供する。

【解決手段】 音声信号を入力する音声入力手段113と、被写体を撮像して第一の動画ファイルを出力する第一の撮像光学系101と、第一の撮像光学系を操作する人物を撮像して第二の動画ファイルを出力する第二の撮像光学系102と、第二の撮像光学系から人物の口唇領域を検出し、第一の撮像光学系を操作する人物が声を発しているかを認識する撮影者発声認識手段を具備する。撮影者発声認識手段により撮影者が声を発していると認識した期間は、音声入力手段に入力される音声信号を調節する音声信号調節手段を備える。

【選択図】 図1(a)



【特許請求の範囲】**【請求項 1】**

第一の動画像を撮影する第一の撮影手段と、
第二の動画像を撮影する第二の撮影手段と、
前記第二の撮影手段は前記第一の撮影手段を操作する人物を撮影し、
前記第二の動画像から人物の口唇領域を検出し、撮影者が声を発しているかを認識する撮影者発声認識手段と、
音声信号を入力する音声入力手段と、
前記音声入力手段により入力された音声进行处理する音声処理手段と、
前記音声信号を前記第一の動画像に記録する音声記録手段を具備し、
前記撮影者発声認識手段により撮影者が声を発していると認識した期間は、前記音声入力手段に入力される音声信号を調節する音声信号調節手段を備えることを特徴とする撮像装置。

10

【請求項 2】

第一の動画像を撮影する第一の撮影手段と、
第二の動画像を撮影する第二の撮影手段と、
前記第二の撮影手段は前記第一の撮影手段を操作する人物を撮影し、
前記第二の動画像から人物の口唇領域を検出し、撮影者が声を発しているかを認識する撮影者発声認識手段と、
音声信号を入力する音声入力手段と、
前記音声入力手段により入力された音声进行处理する音声処理手段と、
前記音声信号を前記第一の動画像に記録する音声記録手段を具備し、
前記撮影者発声認識手段により撮影者が声を発していると認識した期間は、声を発したという情報を付与する撮影者発声情報付与手段を備えることを特徴とする撮像装置。

20

【請求項 3】

前記撮影者発声情報付与手段は前記撮影者発声認識手段により撮影者が声を発していると認識した画像フレームに付与することを特徴とする請求項 2 に記載の撮像装置。

【請求項 4】

前記第一の動画像に撮影時の詳細情報を記録できる管理ファイルを具備し、
前記撮影者発声情報付与手段は前記撮影者発声認識手段により撮影者が声を発していると認識した期間を特定し、前記管理ファイルに付与することを特徴とする請求項 2 に記載の撮像装置。

30

【請求項 5】

前記撮影手段により撮影された動画像を表示する表示手段と
前記第一の動画像を再生する際に前記撮影者発声情報を元に前記音声出力手段から出力される音声信号を調節する音声信号調節手段を持つこと特徴とする再生装置。

【請求項 6】

前記表示手段は前記撮影者発声認識手段により撮影者が声を発していると認識した期間前記第一の動画像に前記第二の動画像を組み込み表示することを特徴とする請求項 1 又は請求項 2 に記載の撮像装置。

40

【請求項 7】

前記撮影手段により撮影された動画像を表示する表示手段を具備し、
前記表示手段は前記第一の動画像に前記第二の動画像を組み込み第三の動画像を表示する動画像組込手段を有することを特徴とする請求項 1 又は請求項 2 に記載の撮像装置。

【請求項 8】

前記動画組込手段は前記音声入力手段に入力された音声の大きさに応じて第二の動画像の大きさを変える、動画組込変更手段を有することを特徴とする請求項 1 又は請求項 2 に記載の撮像装置。

【請求項 9】

前記表示手段は、前記撮影者発声認識手段により撮影者が声を発していないと認識した期

50

間は前記第一の動画像を表示し、前記撮影者発声認識手段により撮影者が声を発していると認識した期間は前記第三の動画像を表示することを特徴とする請求項 8 に記載の撮像装置。

【請求項 10】

前記撮影した動画像を記録する記録手段を具備し、
前記記録手段は前記第一の動画像と前記第二の動画像をそれぞれ記録することを特徴とする請求項 1 又は請求項 2 に記載の撮像装置。

【請求項 11】

前記撮影した動画像を記録する記録手段を具備し、
前記記録手段は前記表示手段により表示された動画像と同様に前記第一の動画像に前記第二の動画像を組み込み記録することを特徴とする請求項 1 又は請求項 2 に記載の撮像装置。

10

【請求項 12】

前記撮影者発声認識手段により撮影者が声を発していると認識した期間、それを撮影者に知らせる撮影者発声認識通知手段を具備することを特徴とする請求項 1 又は請求項 2 に記載の撮像装置。

【請求項 13】

前記撮影者発声認識通知手段は前記表示手段により通知することを特徴とする請求項 12 に記載の撮像装置。

【発明の詳細な説明】

20

【技術分野】

【0001】

本発明は、通常の撮像装置にさらに撮影者も撮影可能な撮像系を有する撮像装置に関し、特に被写体から発せられる音声と撮影者から発せられる音声のバランスを自由に調節できる撮影モードを提供する技術に関する。

【背景技術】

【0002】

従来から、カムコーダ等で撮影された動画において、BGM (Background Music) や効果音等を付加する編集作業が行われている。またその際に、撮影時に録音された音声を残すべく、音声信号より人の声を認識し録音された音声やBGMの音量を調節する編集作業が行われている。

30

【0003】

特許文献 1 には、撮影時の画像と録音された音声信号から人の顔と声を認識し、それぞれが認識された区間の音量を調節する技術が開示されている。

【先行技術文献】

【特許文献】

【0004】

【特許文献 1】特開 2010 - 10780 号公報

【発明の概要】

【発明が解決しようとする課題】

40

【0005】

上述の特許文献に開示された従来技術では、画像に人の顔が映っているか、音声信号に人の声が入っているかという情報を用いているのみであるため、誰が声を発しているかまで判断することはできない。例えば、人物を撮影しながら撮影者がなにか声を発した場合でも、それが被写体の声なのか、撮影者の声なのかは判断できない。

【0006】

そのため音量の調節も一律で行われるが、被写体に対し撮影者の方がカメラ近くおり、その分大きな音量で録音されることになる。このとき、被写体と撮影者が会話をしていた場合にはバランスが低下するし、撮影者の声の方をはっきり聞きたいもしくは被写体の声の方をはっきり聞きたいと言った要望には応えることができない。

50

【 0 0 0 7 】

そこで、本発明の目的は、人の声でも特に撮影者の声を認識し、その期間の音量を視聴者の意図に合わせて調節することを可能にした撮像装置及びその制御方法を提供することにある。

【 課題を解決するための手段 】

【 0 0 0 8 】

上記の目的を達成するために、本発明の撮像装置は、音声信号を入力する音声入力手段と、被写体を撮像して第一の動画ファイルを出力する第一の撮像光学系と、前記第一の撮像光学系を操作する人物を撮像して第二の動画ファイルを出力する撮像光学系と、前記第二の撮像光学系から人物の口唇領域を検出し、前記第一の撮像光学系を操作する人物が声を発しているかを認識する撮影者発声認識手段を具備し、前記撮影者発声認識手段により撮影者が声を発していると認識した期間は、前記音声入力手段に入力される音声信号を調節する音声信号調節手段を備えることを特徴とする。

10

【 発明の効果 】

【 0 0 0 9 】

本発明によれば、撮影者の声が含まれる動画において、視聴者の用途に合わせて音量を調節する撮像装置を提供することができる。

【 図面の簡単な説明 】

【 0 0 1 0 】

【 図 1 (a) 】本発明の実施形態に係る撮像装置の全体構成を示すブロック図である。

20

【 図 1 (b) 】本発明の実施形態に係る撮像装置の全体構成を示すブロック図である。

【 図 2 】本発明に係る人物の口唇画像から、声を発しているかの判断を行う手段について説明する図である。

【 図 3 】本発明に係る撮影者の声と考えられる周波数帯の音量を調節する手段について説明する図である。

【 図 4 (a) 】本発明に係る撮影者の声の周波数を決定する手段について説明する図である。

【 図 4 (b) 】本発明に係る撮影者の声の周波数を決定する手段について説明する図である。

【 図 4 (c) 】本発明に係る撮影者の声の周波数を決定する手段について説明する図である。

30

【 図 4 (d) 】本発明に係る撮影者の声の周波数を決定する手段について説明する図である。

【 図 5 (a) 】本発明の第 1 の実施形態における動画撮影時に撮影者の画像から録音レベルを調節する処理の一例を示すフローチャートである。

【 図 5 (b) 】本発明の第 1 の実施形態における動画撮影時に撮影者の画像から録音レベルを調節する処理の一例を示すフローチャートである。

【 図 5 (c) 】本発明の第 1 の実施形態における動画撮影時に撮影者の画像から録音レベルを調節する処理の一例を示すフローチャートである。

【 図 6 (a) 】本発明の第 2 の実施形態における動画再生時に撮影者の画像から再生音レベルを調節する処理の一例を示すフローチャートである。

40

【 図 6 (b) 】本発明の第 2 の実施形態における動画再生時に撮影者の画像から再生音レベルを調節する処理の一例を示すフローチャートである。

【 図 7 】本発明の第 3 の実施形態における動画撮影時もしくは動画再生時での画面表示の例である。

【 発明を実施するための形態 】

【 0 0 1 1 】

以下に、本発明の好ましい実施の形態を、添付図面に基づいて詳細に説明する。

【 0 0 1 2 】

図 1 (a) 、 (b) は本発明の実施形態にかかわる撮像装置の構成例を示す図である。

50

【0013】

図1に示す撮像装置は、複数の撮像光学系である第1の撮像光学系101、第2の撮像光学系102と、制御部103と、映像信号処理部104と、表示部105と、撮影者発声認識部106と、外部入出力端子部107と、ズームスイッチ108と、シャッターリリーススイッチ109と、記憶部110と、電源部111と、音量信号処理部112と、マイク113とスピーカ114を備える。

【0014】

第1の撮像光学系101と第2の撮像光学系102とは、同じ構成を有する。具体的には、第1の撮像光学系101は、光学ユニット115、撮像素子116、撮像信号処理部117を備える。また第2の撮像光学系102は、光学ユニット118、撮像素子119、撮像信号処理部120を備える。

10

【0015】

第1の撮像系101は主に被写体の撮影を行うため、画像データを撮像する画像撮像手段としての機能を有する。第1の撮像系101が備える光学系は3群構成である。第1の撮像系101が有する光学ユニット115は、対物レンズ121、ズームレンズ122、シャッター・絞りユニット123、シフトレンズ124、フォーカスレンズ125、ズーム駆動制御部126、シャッター・絞り駆動制御部127、シフトレンズ駆動制御部128、フォーカス駆動制御部129からなる。

【0016】

図1(b)を用い説明する。対物レンズ121は、被写体に焦点を合わせるためのものである。ズームレンズ122は、ズーム制御に関与する1群レンズである。シフトレンズ123は振れ補正を実行する2群レンズである。フォーカスレンズ124は、焦点調節処理を実行する3群レンズである。ズームレンズ122は、光軸方向に沿って位置を変更可能に構成されており、倍率変更を行う。ズーム駆動制御部126は、ズームレンズ122を駆動する。ズームレンズ122の後段に配置されているシャッター・絞りユニット123は、露光量を調節する。

20

【0017】

シャッター・絞り駆動制御部127は、シャッター・絞りユニット123を駆動制御して、露光量の調節すなわち露出調節処理を行う露出調節手段である。シフトレンズ124は、光軸に対して略垂直な平面内での位置を変更することが可能に配置されており、振れ補正光学系を構成する。シフトレンズ駆動制御部128は、シフトレンズ124を駆動制御する。フォーカスレンズ125は、ピント調節用レンズである。フォーカスレンズ125は光軸方向に沿って位置を変更可能に構成されている。フォーカス駆動制御部229は、フォーカスレンズ125を駆動制御して、焦点調節処理を実行する焦点調節手段としての機能を有する。

30

【0018】

第2の撮像系102は、第1の撮像系101と同一の光軸上を逆向きに配置することで主に撮影者の撮影を行い、撮影者が声を発しているかを認識するための処理機能を有する。第2の撮像系102が備える光学ユニット118は、第1の撮像系101と同様に3群構成としても良いが、ここでは対物レンズのみとする。

40

【0019】

撮像素子116および119は、光学ユニット115および118を通ってきた光像を受光して、受光した光像を電気信号に変換する。ここで、第1の撮像光学系101で用いられている撮像素子119は本発明の撮像手段であり、映像信号生成手段に相当する。また第2の撮像光学系102で用いられている撮像素子118は撮影者が発声しているかどうかを認識するための画像を撮像する撮像素子である。撮像信号処理部117および120は、撮像素子116および119がそれぞれ出力した電気信号を映像信号に変換し、変換した映像信号を画像データとして出力する。

【0020】

撮像信号処理部117および120が出力した画像データに対して、映像信号処理部1

50

04は所定の処理を施して表示部105に表示可能な画像信号とし、該画像信号を出力する。撮影者発声認識部106は、撮像信号処理部120から得られた発話者の口唇部分の映像から、当該発話者が発話した発話内容を識別し、その情報を制御部103に伝えるもので、画像処理部130と、動作測定部131を備えている。

【0021】

画像処理部130は、撮像信号処理部120で撮影された口唇部分の映像を構成する画像の画素値の差に基づいて、当該口唇部分における予め設定した特徴点の位置を抽出するものである。ここで、図2を参照して、抽出する特徴点について説明する。図2に示すように、口唇部分の特徴点は、口唇部分の上端A点、下端B点、左端C点、右端D点の4点である。なお、各点は、口唇部分を示す赤色の画素値を有する画素と、顔面部分を示す肌色の画素値を有する画素との画素値の差が最大となる画素（境界となる画素）、つまり、口唇上部（上唇）及び口唇下部（下唇）の最端座標に位置する画素が対象となっている。

10

【0022】

また、この実施形態では、画像処理部130は、図2に示したように4点を特徴点として抽出しているが、これに限定されるものではなく、任意個数（4点未満でもよいし、5点以上でもよい）の点を特徴点として抽出してもよい。例えば、口唇部分の映像として、発話者の顔面下部が撮影されている場合、下顎の頂点を特徴点として抽出してもよい。この場合、顎部の肌色の画素値を有する画素と首部の肌色の画素値を有する画素との画素値の差（輝度の差）が最大となる画素で、且つ、最下端に位置する画素を抽出すればよい。

【0023】

そして、この画像処理部130は、撮像系102から映像が継続して入力される限り、各画像について、特徴点を抽出し続ける。つまり、画像処理部130は、1つの画像について、4点を抽出し、映像を構成する複数の画像ごとに抽出していき、抽出した特徴点を、口唇部分の映像を撮影していた時刻に沿った時系列データとして、出力することとなる。

20

【0024】

動作測定部131は、画像処理部130で抽出された特徴点の位置について、当該特徴点の位置の変化を、口唇動作の履歴である動作履歴として測定するものである。そして、この動作測定部131では、画像処理部130で抽出された4点（又は5点）の特徴点それぞれについて、特徴点それぞれの変化について、動作履歴を測定する。なお、画像処理部130に口唇部分の映像が継続して入力され、特徴点が抽出された際に、一定時間、特徴点の変化が無い場合、つまり、口唇部分の映像には変化がない（動きがない）場合、この動作測定部5によって、発話者は発話していないと判定される。

30

【0025】

制御部103はシステム全体を制御する。具体的には、制御部103は、各々の撮像光学系が備えるズーム駆動制御部126、シャッタ・絞り駆動制御部127、シフトレンズ駆動制御部128、フォーカス駆動制御部129、撮像素子116および119、撮像信号処理部117および120、映像信号処理部104、撮影者発声認識部106の処理を制御する。また、制御部103は、表示部105、外部入出力端子部10、記憶部118、音声信号処理部120を制御する。制御部103は、図示を省略するCPU（Central Processing Unit）等により解釈及び実行されるプログラムに従って処理を実行する。

40

【0026】

本実施形態に特有の動作として、制御部103は、前述した撮影者発声認識部106の認識結果に基づいて、音声信号処理部112に指令を出す。また、制御部103は、表示部105に指示して、映像信号処理部104が出力した画像信号を表示画面上に画面表示させる。表示部105は、撮像光学系毎に画像信号を画面表示する。表示部105は、制御部103の指示に従った画面表示処理を実行する。

【0027】

ズームスイッチ108は、ズームレンズ103を操作する操作手段である。ズームスイ

50

タッチ08は、ユーザの操作入力に応じて、操作信号を入力し、入力した操作信号を制御部103に送信する。

【0028】

外部入出力端子部117は、図示を省略する外部装置との間の通信を媒介する。具体的には、外部入出力端子部117には、映像信号及び音声信号が入力される。また、映像信号及び音声信号が外部入出力端子部117から出力される。

【0029】

シャッターリリーススイッチ109は、押し込み量に応じて、第1スイッチ(以下「SW1」と記述)及び第2スイッチ(以下「SW2」と記述)が順にオン状態となるように構成されている。具体的には、ユーザがシャッターリリースボタン109を約半分押し込んだ場合に、SW1がオン状態となる。ユーザが、更に、シャッターリリースボタン109を最後まで深く押し込んだ場合に、SW2がオン状態となる。そして、SW1、SW2がオン状態となったことを示す信号が制御部103に送信される。

10

【0030】

記憶部110には、撮像信号処理部117および120が出力した画像データ、映像信号処理部104が出力した映像信号等が記憶される。なお、記憶部110には、制御部103によって解釈されて実行される、本実施形態の撮像装置の制御プログラムを記憶するためのメモリ装置も含まれる。

【0031】

電源部111は、図1に示す撮像装置の各構成部に対して、必要に応じて電源電圧を供給する。

20

【0032】

マイク113は音声信号入力部であり、外部の音を録音する際に用いられる。スピーカ114は音声信号出力部であり、動画再生時など外部に音声を出力する際に用いられる。マイク113およびスピーカ114はそれぞれ、音声処理部112に接続される。音声処理部112は、マイク113など外部から入力された音声信号に、デジタル変換処理やエンコード処理等の種々の信号処理を施し、デジタル音声信号として制御部103へ入力する。また、制御部103より音声出力のために音声信号が送られてきた場合には、アナログ変換等の種々の信号処理を施し、当該音声信号をスピーカ114などに出力する。

【0033】

また音声出力の際、音声処理部112は制御部103からの指示に従い音量を調節しながら出力する。特に撮影者が声を発したとされる期間に関しては、動画撮影もしくは再生を行いながら随時調節を行う。単純にその期間の音量を上げ下げして調節しても良いし、撮影者の声と考えられる周波数帯のみを調節しても良い。後者の方法をとれば、仮に被写体と撮影者が同時に声を発した場合でも、撮影者の声のみを調節することができる。

30

【0034】

前述撮影者声と考えられる周波数帯のみを調節する方法について、図3を用いて説明する。301は撮像光学系102により撮影された撮影者の顔画像である。顔画像301はa、b、c、d、eとフレームが進むとする。ここで、撮影者の顔画像301の中から、301cから301dの期間で撮影者が発声していると判断されたとする。このとき、発声したとされるフレーム301cとその直前のフレーム301bの音を周波数であらわしたものが302bと302cである。この302bと302cを比較し、差異のある周波数帯を調節する。ここでは差異が見られる周波数帯Aを調節する。また、このとき一般的に人の声の周波数帯であるとされる100Hzから700Hzのみの差異を見ると限定しても良い。直前のフレームも撮影者が声を発していたと認識されていた場合には、直前のフレームにて決定した周波数帯を継続して調節対象とする。

40

【0035】

ただし、この方法では撮影者と被写体が同時に発声した場合に、被写体の声までも調節対象となってしまう。被写体の声でなくとも、周りの音声等も撮影者の声と認識されてしまうことが考えられる。

50

【0036】

図4(a)は撮影者と被写体が同時に発声したときの音を周波数で表した一例である。300Hzと600Hzにそれぞれピークがあるとする。そこで、撮影者が声を発していないと認識されたフレーム、声を発したと認識されたフレームそれぞれから平均値を算出し、その差分から撮影者の声であると考えられる周波数を決定する。図4(b)は撮影者が声を発していないと認識されたフレームの平均、図4(c)は声を発したと認識されたフレームの平均である。また図4(d)は、図4(b)と図4(c)の差分をとったものである。これより、最も値の大きい周波数Cが撮影者の声であると判断し、図4(a)の例でも周波数Cに近いと考えられる600Hzの方のピークを調節する。またここで説明した差分から撮影者の声の周波数を求める方法は撮影時に随時行っても良いし、動画記録終了後に声を発したと認識されたフレームを見直し、撮影者の声の周波数と比較して逸脱するものがあつたら声を発していないと情報を修正する処理にも利用できる。

10

【0037】

本実施形態の撮像装置の制御方法は、図1に示す撮像装置が備える各処理部によって実現される。また、図1に示す撮像装置の機能は、CPUとその上で実行されるコンピュータプログラムにより実現される。このコンピュータプログラムは、コンピュータが読み取り可能な記録媒体に格納することができ、この記録媒体に記録して提供され、又は、通信インタフェースを介してネットワークを利用した送受信により提供される。

【0038】

次に、図1に示す撮像装置による、自動焦点調節(AF)処理、自動露出(AE)処理、及び、ズーム駆動制御処理について説明する。上述したように、シャッターリリーススイッチ109の押し込み量に応じて、SW1及びSW2が順にオン状態となる。シャッターリリーススイッチ109のSW1がオン状態となったことを示す信号が制御部103に入力されると、制御部103が、フォーカス駆動制御部129に指示して、制御部103が決定した指標を対象としてAF処理を実行させる。制御部103から指示を受けたフォーカス駆動制御部129は、フォーカスレンズを駆動することによって、上記指標を対象としたAF処理を実行する。また、制御部203が、シャッター・絞り駆動制御部127に指示して、制御部103が決定した指標を対象として、AE処理を実行させる。具体的には、シャッター・絞り駆動制御部127が、シャッター・絞りユニット123を駆動して、露光量を適正な値に設定する。

20

30

【0039】

SW2がオン状態となり、SW2がオン状態となったことを示す信号が制御部103に入力されると、制御部103が、撮像素子116および119に指示して、撮像素子116および119が受光した光像を電気信号に変換させる。撮像信号処理部117および120が、上記電気信号に基づいて画像データを出力する。映像信号処理部104が、撮像信号処理部117および120から得られる画像データに基づいて、表示部105に表示する画像信号を出力する。制御部103は、撮像信号処理部117および120が出力した画像データ、映像信号処理部104が出力した画像信号を記憶部110に記憶する。

【0040】

ユーザがズームスイッチ108を操作すると、その操作信号が制御部103に入力され、制御部103がズーム駆動制御部126に指示を与える。ズーム駆動制御部126は、制御部103によって指示されたズーム位置へズームレンズ122を移動させる。これによって、ズーム駆動制御処理が実行される。

40

【実施例1】

【0041】

以下、図5を参照して、本発明の第1の実施例による、動画撮影時に録音レベルの調節を行う方法について説明する。

【0042】

図5は撮像装置の動作処理について説明するフローチャートである。なお図中の「YES」は肯定的な判断結果を示し、「NO」は否定的な判断結果を示す。

50

【0043】

図5(a)は本発明の実施形態にかかわる撮像装置における動画撮影時に録音レベルの調節を行う処理例を示すフローチャートであり、撮影者発声認識処理と、撮影者発声Flag補正処理を含む。当該撮影者発声認識の処理例については図5(b)に示す。つまり図5(b)は得られた撮影者の画像から撮影者が声を発しているかを認識する処理を例示したフローチャートである。当該撮影者発声Flag補正処理については図5(c)に示す。つまり図5(c)は、撮影者発声認識処理に撮影者発声Flagが1とされたフレームに対し、認識に誤りがないかを判断し、誤りがある場合には補正する処理を示したフローチャートである。

【0044】

まず、図1に示す撮像装置の制御部103が、ユーザの操作入力に従って撮影モードに設定されたことを確認し、撮影処理を開始する。

【0045】

先ず、ステップS501では、撮影者の声を大きく録音するか、小さく録音するか、またどの程度調節するか等の音量調節の仕方を選択する。

【0046】

ステップS502では、本発明の撮像手段である撮像光学系101の撮像素子116の駆動を開始する。

【0047】

続くステップS503では、シャッターリリーススイッチ109のSW1がオン状態になったか否かを制御部103が判断する。その結果、SW1のオン状態が検知された場合(S503でYESの場合)には、ステップS504に進み、またSW1のオフ状態が検知された場合(S503でNOの場合)には、S503での判定が繰り返される。

【0048】

ステップS504において、焦点調節やシャッタースピードの決定等の撮影準備をAE及びAFにて行い、ステップS505において、撮影者発声Flagを0に初期化し、ステップS506において、撮影者発声認識処理を行う撮像光学系102の撮像素子119の駆動を開始する。

【0049】

ステップS507において、制御部103がシャッターリリーススイッチ109のSW2がオン状態になったかを判断する。制御部103が、SW2がオン状態になったと判断した場合(S507でYESの場合)には、ステップS508に進み動画の記録を開始する。制御部103が、SW2がオン状態でない、すなわち、オフ状態であると判断した場合(S507でNOの場合)には、上記ステップS507に戻り、SW2がON状態であるかを判断する。

【0050】

ステップS508において、撮影者発声認識部106が撮影者が声を発したかの認識処理を行い、続くステップS509において、撮影者発声Flagの補正処理を行い、ステップS510において、音声信号処理部112がマイク113から入力される音の録音レベルを調節する。撮影者発声認識処理の具体的内容については、図5(b)を用いて後で詳細に説明する。撮影者発声Flag補正処理の具体的内容については、図5(c)を用いて後で詳細に説明する。

【0051】

ステップS511において、撮影者発声Flagの値を見る。撮影者発声Flagが0である、すなわち画像フレームにて撮影者が声を発していないと判断された場合(S511でNOの場合)には、動画撮影を継続し、ステップS509に戻り撮影者発声認識処理を行う。撮影者発声Flagが1である、すなわち画像フレームにて撮影者が声を発していると判断された場合(S511でYESの場合)には、ステップS514に進み、動画撮影を終了し、ステップS515において、記憶部110に画像を保存し、撮影を終了する。

10

20

30

40

50

【 0 0 5 2 】

図 5 (b) は図 5 (a) のステップ S 5 0 9 における撮影者発声認識処理の一例を示す。

【 0 0 5 3 】

先ず、ステップ S 5 1 6 において、撮影者発声認識部 1 0 6 は画像処理部 1 3 0 によって、撮像信号処理部 1 2 0 から得られた口唇部分の映像から特徴点を抽出する。続くステップ S 5 1 7 において、撮影者発声認識部 1 0 6 は、動作測定部 1 3 1 によって、画像処理部 1 3 0 で抽出された特徴点の変化を、動作履歴 (動作履歴グラフ) として測定する。

【 0 0 5 4 】

ステップ S 5 1 8 では、動作履歴から一定時間内に変化があるかどうかを判断する。一定時間内に変化があった場合 (S 1 0 2 で Y E S の場合) 、制御部 1 0 3 は撮影者が声を発したと判断し、続くステップ S 5 1 9 において撮影者発声 F l a g を 1 にして、処理を終了する。逆に一定時間内に特徴点の変化が無い場合、つまり、口唇部分の映像には変化がない (動きがない) 場合、制御部 1 0 3 は撮影者が声を発していないと判断し、続くステップ S 5 2 0 において、撮影者発声 F l a g を 0 にして、処理を終了する。

10

【 0 0 5 5 】

図 5 (c) は図 5 (a) のステップ S 5 1 0 における撮影者発声 F l a g 補正処理の一例を示す。

【 0 0 5 6 】

先ず、ステップ S 5 2 1 において、制御部 1 0 3 が撮影者発声 F l a g を確認する。制御部 1 0 3 が、撮影者発声 F l a g が 1 ではない、すなわち撮影者が声を発していないと判断した場合 (S 5 2 1 で N O の場合) には、そのまま処理を終了する。制御部 1 0 3 が、撮影者発声 F l a g が 1 である、すなわち撮影者が声を発したと判定した場合 (S 5 2 1 で Y E S の場合) には、続くステップ S 5 2 2 にて、制御部 1 0 3 が調節対象の周波数帯を決定する。決定の方法は前述のとおりである。

20

【 0 0 5 7 】

ステップ S 5 2 3 において、撮影した動画ファイルの先頭から直前までの撮影者発声 F l a g が 0 のフレームの音声信号の平均を、ステップ S 5 2 4 にて撮影者発声 F l a g が 1 のフレームの音声信号の平均を算出し、続くステップ S 5 2 5 において、その差分を算出する。その結果に基づき、ステップ S 5 2 6 において、撮影者の声と考えられる周波数を決定する。

30

【 0 0 5 8 】

ステップ S 5 2 7 において、ステップ S 5 2 2 で求めた調節対象の周波数帯と、ステップ S 5 2 6 において求めた撮影者の声の周波数に乖離があるかを判断する。この判断は、調節対象の周波数帯の中のピーク値と、撮影者の声と考えられる周波数がどれくらいずれているかを見て行う方法があるし、それ以外の方法を用いても良い。調節対象の周波数帯と撮影者の声の周波数に乖離がないと判断された場合 (S 5 2 7 で N O の場合) には、そのまま処理を終了するが、乖離があると判断された場合 (S 5 2 7 で Y E S の場合) には、ステップ S 5 2 8 にて撮影者発声 F l a g を 0 にし、処理を終了する。

40

【 実施例 2 】

【 0 0 5 9 】

以下、図 6 を参照して、本発明の第 2 の実施例による、動画撮影時に録音レベルの調節を行う方法について説明する。なお図中の「 Y E S 」は肯定的な判断結果を示し、「 N O 」は否定的な判断結果を示す。動画再生の際には、撮影者発声 F l a g を見て、再生音レベルを調節する。また動画撮影の際、撮影者発声 F l a g は撮影者が声を発していると認識したフレームに都度付与しても良いし、声を発したタイミングもしくはフレームを記憶部 1 1 0 に記憶しておき、動画記録の際に管理ファイルとして一緒に記録しても良い。

【 0 0 6 0 】

図 6 (a) は本発明の実施形態にかかわる撮像装置における動画再生時に再生音レベルの調節を行う処理例を示すフローチャートであり、撮影処理と、撮影者発声認識処理と、

50

撮影者発声 F l a g 補正処理を含む。当該撮影処理については図 6 (b) に示す。つまり図 6 (b) は撮像装置が撮影モードに設定されたときの処理動作を例示したフローチャートである。当該撮影者発声認識処理および当該撮影者発声 F l a g 補正処理については、実施例 1 で説明した通りである。

【 0 0 6 1 】

先ず、ステップ S 6 0 1 において、図 1 に示す撮像装置の制御部 1 0 3 がユーザの操作入力に従って再生モードに設定されているかを確認する。制御部 1 0 3 が再生モードであると判断した場合 (S 6 0 1 で Y E S の場合) には、ステップ S 6 0 2 に続く。制御部 1 0 3 が再生モードでない、すなわち撮影モードであると判断した場合 (S 6 0 1 で N O の場合) には、ステップ S 6 0 3 において撮影処理を行う。なお具体的内容については図 6 (b) を用いて後で詳細に説明する。

10

【 0 0 6 2 】

ステップ S 6 0 2 では、撮影者の声を大きく再生するか、小さく再生するか、またどの程度調節するか等の音量調節の仕方を選択する。このステップでなくとも、動画再生しながら随時変更しても良い。

【 0 0 6 3 】

ステップ S 6 0 4 では、記憶部 1 1 0 に保存されている動画の再生を開始する。続くステップ S 6 0 5 では、制御部 1 0 3 が再生中の動画が音調節モードで撮影されたものであるかを判断する。制御部 1 0 3 が音調節モードではないと判断した場合 (S 6 0 5 で N O の場合) には、ステップ S 6 0 8 に進む。制御部 1 0 3 が音調節モードであると判断した場合 (ステップ S 6 0 5 において Y E S の場合) には、続くステップ S 6 0 6 において制御部 1 0 3 は、再生中のフレームの撮影者発声 F l a g を確認する。撮影者発声 F l a g が 1 である、すなわち撮影者が声を発していると判断した場合 (S 6 0 6 で Y E S の場合) は、そのフレームでは撮影者の発した声が録音されているため、続くステップ S 6 0 7 において、再生者の用途に応じて再生音レベルを調節する処理を行う。撮影者発声 F l a g が 0 である、すなわち撮影者が声を発していないと判断した場合 (S 6 0 6 で N O の場合) は、ステップ S 6 0 8 に進む。調節の仕方に関してはステップ S 6 0 2 の処理に従う。

20

【 0 0 6 4 】

続くステップ S 6 0 8 では、制御部 1 0 3 が再生中の動画が終了したかどうかを判断する。制御部 1 0 3 が動画が終了したと判断した場合 (S 6 0 8 で Y E S の場合) には、そこで処理を終了する。制御部 1 0 3 が動画がまだ続くと判断した場合 (S 6 0 8 で N O の場合) には、ステップ S 6 0 5 の処理に戻る。

30

【 0 0 6 5 】

図 6 (b) は図 6 (a) のステップ S 6 0 3 における撮影処理の一例を示す。

【 0 0 6 6 】

先ず、ステップ S 6 0 9 では、本発明の撮像手段である撮像光学系 1 0 1 の撮像素子 1 1 6 の駆動を開始する。

【 0 0 6 7 】

続くステップ S 6 1 0 では、シャッターリリーススイッチ 1 0 9 の S W 1 がオン状態になったか否かを制御部 1 0 3 が判断する。その結果、S W 1 のオン状態が検知された場合 (S 6 1 0 で Y E S の場合) には、ステップ S 6 1 1 に進み、また S W 1 のオフ状態が検知された場合 (S 6 1 0 で N O の場合) には、S 6 1 0 での判定が繰り返される。

40

【 0 0 6 8 】

ステップ S 6 1 1 において、焦点調節やシャッタースピードの決定等の撮影準備を A E 及び A F にて行い、ステップ S 6 1 2 において、撮影者発声 F l a g を 0 に初期化し、ステップ S 6 1 3 において、撮影者発声認識を行う撮像光学系 1 0 2 の撮像素子 1 1 9 の駆動を開始する。

【 0 0 6 9 】

ステップ S 6 1 4 において、制御部 1 0 3 がシャッターリリーススイッチ 1 0 9 の S W 2

50

がオン状態になったかを判断する。制御部103が、SW2がオン状態になったと判断した場合(S614でYESの場合)には、ステップS615に進み動画の記録を開始する。制御部103が、SW2がオン状態でない、すなわち、オフ状態であると判断した場合(S614でNOの場合)には、上記ステップS614に戻り、SW2がON状態であるかを判断する。

【0070】

ステップS616において、撮影者発声認識部106が撮影者が声を発したかの認識処理を行い、続くステップS617において、撮影者発声Flagの補正処理を行う。撮影者発声認識処理および撮影者発声Flag補正処理の具体的内容については、実施例1で説明した通りである。

10

【0071】

ステップ618において、再び、制御部103がシャッターリリーススイッチ109のSW2がオン状態になったかを判断する。制御部103が、SW2がオン状態になったと判断した場合(S618でYESの場合)には、ステップ619に進み、動画撮影を終了し、ステップ620において、記憶部110に画像を保存し、撮影処理を終了する。制御部103が、SW2がオン状態でない、すなわち、オフ状態であると判断した場合(S618でNOの場合)には、動画記録を続け、撮影者発声認識処理を行う。

【実施例3】

【0072】

以下、図7を参照して、本発明の第3の実施例による、画面表示の方法について説明する。

20

【0073】

これまでも説明した通り、撮影者発声認識処理は撮影者の顔画像を撮影したそのデータを元に行う。このとき動画データには撮像光学系102で撮影した撮影者の顔データを付与しても良いし、付与しなくとも良い。また、撮影者の顔データを付与する場合には、撮影者発声認識部106が撮影者が声を発したと認識したタイミングで前記撮影者の顔データに効果をつけても良いし、もしくは撮影者が見て分かるようなアイコンを表示させても良い。

【0074】

撮影者の顔データを付与する場合の画面表示例について、図7を用いて説明する。図7は撮像装置における再生画面表示の例である。図に示す701は本発明における撮像装置である。702は、撮影者を撮影するための撮像光学系102である。703は表示部105であり、通常撮像光学系101により撮影された画像を表示する。704は撮像光学系102により撮影された画像である。705は撮影者が発声したと認識したタイミングで表示するアイコンである。

30

【0075】

撮影者が声を発していないと認識された場合(撮影者発声Flagが0の場合)の画面表示例が701aである。撮影者が声を発したと認識された場合(撮影者発声Flagが1の場合)の画面表示例が701bである。この例では、撮影者が声を発したと認識すると、撮影者の画像704を大きく表示し、かつアイコン705を表示させている。他にも、撮影者の画像704を点滅させるなどの効果を付与しても良い。また、効果をつける際には、撮影者の声のトーンや、大きさから効果を変化させても良い。

40

【0076】

実施例3により、再生時にどのタイミングで撮影者が声を発したと認識されたかが視覚的にも分かるようになる。そこで、もし視聴者が意図しないタイミングで撮影者が声を発したと認識されていた場合には、撮像装置701の操作からも撮影者発声Flagの修正や音調節の設定ができるようにしても良い。例えば、動画再生中にシャッターリリーススイッチ109が押された場合、そのシーンの撮影者発声Flagを1にする。ズームスイッチ108が操作された場合、撮影者発声Flagが1のフレームの音量を調節できる。

【0077】

50

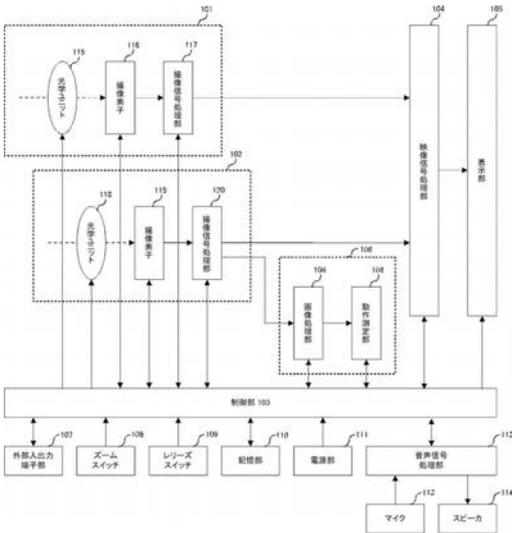
以上、本発明の好ましい実施形態について説明したが、本発明はこれらの実施形態に限定されず、その要旨の範囲内で種々の変形及び変更が可能である。

【符号の説明】

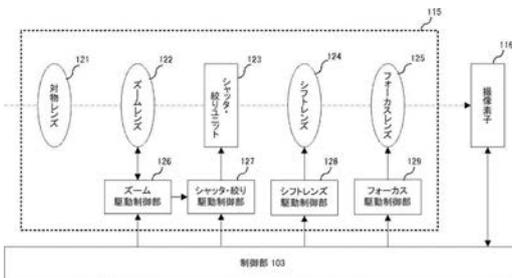
【0078】

- 101 撮像光学系 1、102 撮像光学系 2、103 制御部、
- 104 映像信号処理部、105 表示部、106 撮影者発声認識部、
- 107 外部入出力端子部、108 ズームスイッチ、
- 109 シャッターリリーススイッチ、110 記憶部、111 電源部、
- 112 音声信号処理部、113 マイク、114 スピーカ、
- 115 撮像光学系 1に含まれる光学ユニット 1、
- 116 撮像光学系 1に含まれる撮像素子 1、
- 117 撮像光学系 1に含まれる撮像信号処理部 1、
- 118 撮像光学系 2に含まれる光学ユニット 2、
- 119 撮像光学系 1に含まれる撮像素子 2、
- 120 撮像光学系 1に含まれる撮像信号処理部 2、
- 130 画像処理部、131 動作測定部

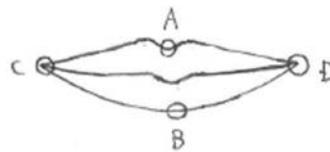
【図 1 (a)】



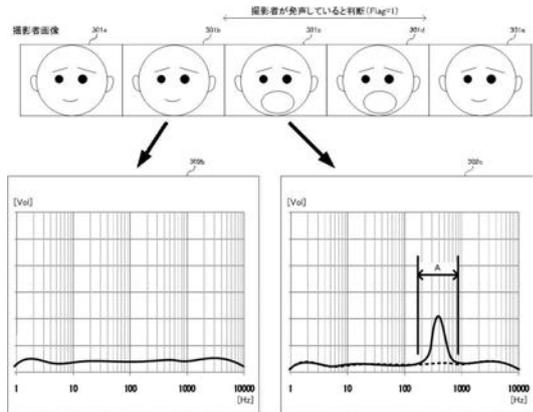
【図 1 (b)】



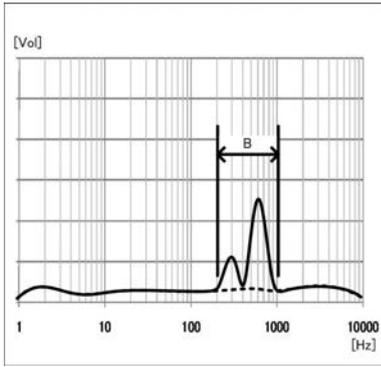
【図 2】



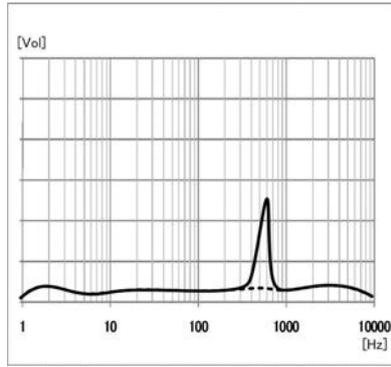
【図 3】



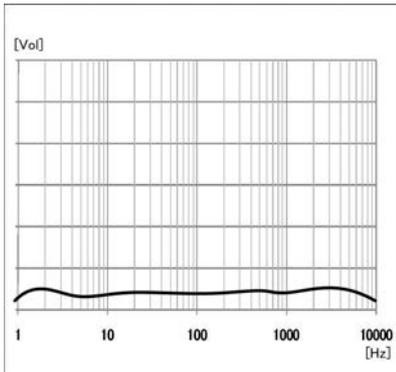
【図4(a)】



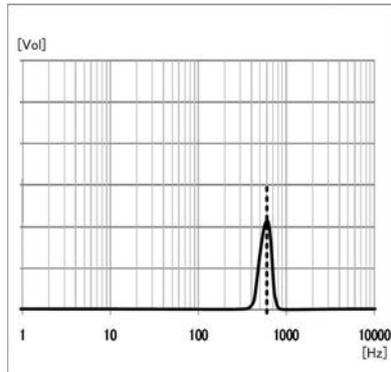
【図4(c)】



【図4(b)】



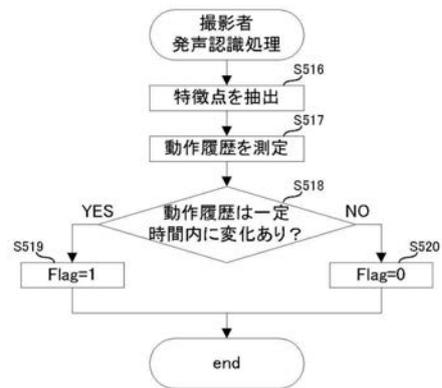
【図4(d)】



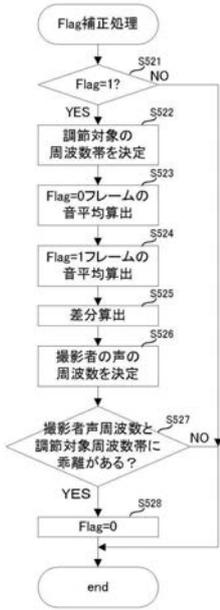
【図5(a)】



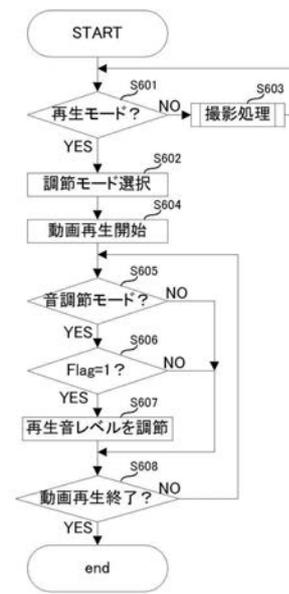
【図5(b)】



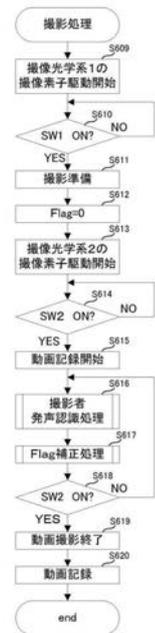
【図5(c)】



【図6(a)】



【図6(b)】



【図7】

