

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4441879号  
(P4441879)

(45) 発行日 平成22年3月31日(2010.3.31)

(24) 登録日 平成22年1月22日(2010.1.22)

(51) Int.Cl.			F I		
<b>HO4N</b>	<b>5/225</b>	<b>(2006.01)</b>	HO4N	5/225	F
<b>GO6T</b>	<b>1/00</b>	<b>(2006.01)</b>	GO6T	1/00	340A
<b>GO6T</b>	<b>7/00</b>	<b>(2006.01)</b>	GO6T	7/00	100C
<b>HO4N</b>	<b>5/232</b>	<b>(2006.01)</b>	HO4N	5/232	Z

請求項の数 10 (全 23 頁)

(21) 出願番号	特願2005-187700 (P2005-187700)	(73) 特許権者	000002185
(22) 出願日	平成17年6月28日 (2005.6.28)		ソニー株式会社
(65) 公開番号	特開2007-13255 (P2007-13255A)		東京都港区港南1丁目7番1号
(43) 公開日	平成19年1月18日 (2007.1.18)	(74) 代理人	100082131
審査請求日	平成19年2月23日 (2007.2.23)		弁理士 稲本 義雄
		(72) 発明者	小沢 一彦
			東京都品川区北品川6丁目7番35号 ソニー株式会社内
		審査官	高野 美帆子

最終頁に続く

(54) 【発明の名称】 信号処理装置および方法、プログラム、並びに記録媒体

(57) 【特許請求の範囲】

【請求項1】

画像の信号と、この画像に同期した音声の信号を処理する信号処理装置であって、  
入力された画像の信号に基づいて、前記画像の中に特定の被写体が写っている可能性の大きさを表す情報を生成する生成手段と、

前記生成手段により生成された情報に基づいて、前記画像の中に特定の被写体が写っているか否かを判定する判定手段と、

前記判定手段により、前記画像の中に特定の被写体が写っていると判定された場合、前記被写体方向に向けて收音した音声の信号のレベルを上げ、および、または前記被写体方向以外の方向に向けて收音した音声の信号のレベルを下げるように、前記音声の信号の指向特性を変更する指向特性変更手段と、

前記判定手段により、前記画像の中に特定の被写体が写っていると判定された場合、前記被写体の発する音声が含まれる周波数帯域のレベルを上げ、および、または前記被写体の発する音声が含まれる周波数帯域以外の周波数帯域のレベルを下げるように、前記音声の信号の周波数特性を変更する周波数特性変更手段とを備え、

前記生成手段は、

入力された信号に対応する画像の中で、特定の色の画素により構成される領域を抽出する抽出手段と、

前記抽出手段により抽出された領域の画素の輝度または色差に基づいて前記領域の画素の特徴量のパターンを検出する検出手段と、

予め前記特定の被写体を含む画像から検出された特徴量のパターンを記憶する記憶手段と、

前記検出手段により検出された前記特徴量のパターンと、前記記憶手段に記憶されている前記特徴量のパターンとを比較し、それぞれのパターンの相関係数を演算する演算手段とを有し、

前記演算手段により演算された相関係数が前記画像の中に特定の被写体が写っている可能性の大きさを表す情報として生成される

ことを特徴とする信号処理装置。

【請求項 2】

前記画像の中の所定の領域を切り出す切出し手段をさらに備え、

前記抽出手段は、前記切出し手段により切出された画像の中で、特定の色の画素により構成される領域を抽出する

ことを特徴とする請求項 1 に記載の信号処理装置。

【請求項 3】

前記特定の被写体は、人物であり、

前記判定手段により、前記画像の中に人物が写っていると判定された場合、

前記指向特性変更部は、前方向に向けて收音した音声の信号のレベルを上げ、および、または前方向以外の方向に向けて收音した音声の信号のレベルを下げるように、前記音声の信号の指向特性を変更し、

前記周波数特性変更部は、人の声が含まれる周波数帯域のレベルを上げ、および、またはそれ以外の周波数帯域のレベルを下げるように、前記音声の信号の周波数特性を変更する

ことを特徴とする請求項 2 に記載の信号処理装置。

【請求項 4】

前記指向特性変更部による前記音声の信号の指向特性の変更の度合い、または前記周波数特性変更部による前記音声の信号の周波数特性の変更の度合いを制御する制御レベルを設定する設定手段をさらに備える

ことを特徴とする請求項 2 に記載の信号処理装置。

【請求項 5】

前記設定手段は、前記抽出手段により抽出された領域の大きさを表す情報を正規化し、前記制御レベルとして設定する

ことを特徴とする請求項 4 に記載の信号処理装置。

【請求項 6】

前記設定手段は、前記演算手段により演算された相関係数を正規化し、前記制御レベルとして設定する

ことを特徴とする請求項 4 に記載の信号処理装置。

【請求項 7】

前記画像の信号と、前記音声の信号は、記録媒体に記録されたデータが再生装置により再生されることにより出力される画像の信号と、音声の信号である

ことを特徴とする請求項 6 に記載の信号処理装置。

【請求項 8】

画像の信号と、この画像に同期した音声の信号を処理する信号処理装置の信号処理方法であって、

入力された画像の信号に基づいて、前記画像の中に特定の被写体が生じている可能性の大きさを表す情報を生成する生成ステップと、

前記生成ステップの処理により生成された情報に基づいて、前記画像の中に特定の被写体が生じているか否かを判定する判定ステップと、

前記判定ステップの処理により、前記画像の中に特定の被写体が生じていると判定された場合、前記被写体方向に向けて收音した音声信号のレベルを上げ、および、または前記被写体方向以外の方向に向けて收音した音声信号のレベルを下げるように、前記音声信号

10

20

30

40

50

の指向特性を変更する指向特性変更ステップと、

前記判定ステップの処理により、前記画像の中に特定の被写体が写っていると判定された場合、前記被写体の発する音声が含まれる周波数帯域のレベルを上げ、および、または前記被写体の発する音声が含まれる周波数帯域以外の周波数帯域のレベルを下げるように、前記音声の信号の周波数特性を変更する周波数特性変更ステップとを含み、

前記生成ステップの処理では、

入力された信号に対応する画像の中で、特定の色の画素により構成される領域が抽出され、

前記抽出された領域の画素の輝度または色差に基づいて前記領域の画素の特徴量のパターンが検出され、

前記検出された前記特徴量のパターンと、予め前記特定の被写体を含む画像から検出された特徴量のパターンとが比較され、それぞれのパターンの相関係数が演算され、

前記演算された相関係数が前記画像の中に特定の被写体が写っている可能性の大きさを表す情報として生成される

ことを特徴とする信号処理方法。

【請求項 9】

画像の信号と、この画像に同期した音声の信号を処理する信号処理装置に信号処理を実行させるプログラムであって、

入力された画像の信号に基づいて、前記画像の中に特定の被写体が写っている可能性の大きさを表す情報の生成を制御する生成制御ステップと、

前記生成制御ステップの処理により生成された情報に基づいて、前記画像の中に特定の被写体が写っているか否かの判定を制御する判定制御ステップと、

前記判定制御ステップの処理により、前記画像の中に特定の被写体が写っていると判定された場合、前記被写体方向に向けて収音した音声信号のレベルを上げ、および、または前記被写体方向以外の方向に向けて収音した音声信号のレベルを下げるように、前記音声信号の指向特性の変更を制御する指向特性変更制御ステップと

前記判定制御ステップの処理により、前記画像の中に特定の被写体が写っていると判定された場合、前記被写体の発する音声が含まれる周波数帯域のレベルを上げ、および、または前記被写体の発する音声が含まれる周波数帯域以外の周波数帯域のレベルを下げるように、前記音声の信号の周波数特性を変更する周波数特性変更制御ステップとをコンピュータに実行させ、

前記生成制御ステップの処理では、

入力された信号に対応する画像の中で、特定の色の画素により構成される領域が抽出され、

前記抽出された領域の画素の輝度または色差に基づいて前記領域の画素の特徴量のパターンが検出され、

前記検出された前記特徴量のパターンと、予め前記特定の被写体を含む画像から検出された特徴量のパターンとが比較され、それぞれのパターンの相関係数が演算され、

前記演算された相関係数が前記画像の中に特定の被写体が写っている可能性の大きさを表す情報として生成される

ことを特徴とするプログラム。

【請求項 10】

請求項 9 に記載のプログラムが記録されていることを特徴とする記録媒体。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、信号処理装置および方法、プログラム、並びに記録媒体に関し、特に、撮影された画像内において特定の被写体から発せられる音声を強調することができるようにする信号処理装置および方法、プログラム、並びに記録媒体に関する。

【背景技術】

10

20

30

40

50

## 【 0 0 0 2 】

近年、ビデオカメラなどの機器において、オートフォーカス機能などとともに、音声信号の抽出に関する技術も向上している。例えば、撮影方向の中央部分に位置する音源から発せられる音声信号を抽出し、強調する技術が提案されている（例えば、特許文献 1 参照。）。これは映像信号とともに音声信号を記録及び再生するビデオカメラなどの機器においては、撮影目的となる被写体がほとんどの場合に、画面中央部分に位置することを踏まえて、その被写体の発する音声をクリアに記録及び再生することを目的としたものであった。

## 【 0 0 0 3 】

また、音場空間の全周囲方向から入力する音声を、映像と共にマルチチャンネルで記録または再生することで臨場感のある映像と音声の試聴が可能なビデオカメラも提案されている（例えば、特許文献 2 参照。）。

## 【 0 0 0 4 】

【特許文献 1】特開 2 0 0 1 - 8 2 8 5 号公報

【特許文献 2】特開 2 0 0 0 - 2 9 9 8 4 2 号公報

【発明の開示】

【発明が解決しようとする課題】

## 【 0 0 0 5 】

しかしながら、特許文献 1 によれば、常に画面中央方向から入射した音声抽出されて強調されるために、人物に限らず、画面中央方向に音源がある場合には、環境音であっても抽出されてしまい、目的とする音声強調されない場合があった。

## 【 0 0 0 6 】

また、特許文献 2 によれば、臨場感を重視すれば、映像に見えないところの周囲からの音声は収録され、被写体の発する音声がクリアに収録されず、逆に被写体の音声を重視すれば、臨場感に乏しい収録になってしまう。

## 【 0 0 0 7 】

本発明はこのような状況に鑑みてなされたものであり、撮影された画像内において特定の被写体から発せられる音声を強調することができるようにするものである。

【課題を解決するための手段】

## 【 0 0 0 8 】

本発明の信号処理装置は、画像の信号と、この画像に同期した音声の信号を処理する信号処理装置であって、入力された画像の信号に基づいて、画像の中に特定の被写体が写っている可能性の大きさを表す情報を生成する生成手段と、生成手段により生成された情報に基づいて、画像の中に特定の被写体が写っているか否かを判定する判定手段と、判定手段により、画像の中に特定の被写体が写っていると判定された場合、被写体方向に向けて収録した音声の信号のレベルを上げ、および、または被写体方向以外の方向に向けて収録した音声の信号のレベルを下げるように、音声の信号の指向特性を変更する指向特性変更手段と、判定手段により、画像の中に特定の被写体が写っていると判定された場合、被写体の発する音声が含まれる周波数帯域のレベルを上げ、および、または被写体の発する音声が含まれる周波数帯域以外の周波数帯域のレベルを下げるように、音声の信号の周波数特性を変更する周波数特性変更手段とを備え、生成手段は、入力された信号に対応する画像の中で、特定の色の画素により構成される領域を抽出する抽出手段と、抽出手段により抽出された領域の画素の輝度または色差に基づいて領域の画素の特徴量のパターンを検出する検出手段と、予め特定の被写体を含む画像から検出された特徴量のパターンを記憶する記憶手段と、検出手段により検出された特徴量のパターンと、記憶手段に記憶されている特徴量のパターンとを比較し、それぞれのパターンの相関係数を演算する演算手段とを有し、演算手段により演算された相関係数が前記画像の中に特定の被写体写っている可能性の大きさを表す情報として生成されることを特徴とする。

## 【 0 0 1 1 】

前記画像の中の所定の領域を切り出す切出し手段をさらに備え、抽出手段は、切出し手

10

20

30

40

50

段により切り出された画像の中で、特定の色の画素により構成される領域を抽出することができる。

【0012】

前記特定の被写体は、人物であり、判定手段により、画像の中に人物が写っていると判定された場合、指向特性変更部は、前方向に向けて収音した音声の信号のレベルを上げ、および、または前方向以外の方向に向けて収音した音声の信号のレベルを下げるように、音声の信号の指向特性を変更し、周波数特性変更部は、人の声が含まれる周波数帯域のレベルを上げ、および、またはそれ以外の周波数帯域のレベルを下げるように、音声の信号の周波数特性を変更することができる。

【0013】

前記指向特性変更部による音声の信号の指向特性の変更の度合い、または周波数特性変更部による音声の信号の周波数特性の変更の度合いを制御する制御レベルを設定する設定手段をさらに備えるようにすることができる。

【0014】

前記設定手段は、抽出手段により抽出された領域の大きさを表す情報を正規化し、制御レベルとして設定するようすることができる。

【0015】

前記設定手段は、演算手段により演算された相関係数を正規化し、制御レベルとして設定するようすることができる。

【0017】

前記画像の信号と、音声の信号は、記録媒体に記録されたデータが再生装置により再生されることにより出力される画像の信号と、音声の信号であるようにすることができる。

【0018】

本発明の信号処理方法は、画像の信号と、この画像に同期した音声の信号を処理する信号処理装置の信号処理方法であって、入力された画像の信号に基づいて、画像の中に特定の被写体が写っている可能性の大きさを表す情報を生成する生成ステップと、生成ステップの処理により生成された情報に基づいて、画像の中に特定の被写体が写っているか否かを判定する判定ステップと、判定ステップの処理により、画像の中に特定の被写体が写っていると判定された場合、被写体方向に向けて収音した音声信号のレベルを上げ、および、または被写体方向以外の方向に向けて収音した音声信号のレベルを下げるように、音声信号の指向特性を変更する指向特性変更ステップと、判定ステップの処理により、画像の中に特定の被写体が写っていると判定された場合、被写体の発する音声が含まれる周波数帯域のレベルを上げ、および、または被写体の発する音声が含まれる周波数帯域以外の周波数帯域のレベルを下げるように、音声の信号の周波数特性を変更する周波数特性変更ステップとを含み、生成ステップの処理では、入力された信号に対応する画像の中で、特定の色の画素により構成される領域が抽出され、抽出された領域の画素の輝度または色差に基づいて領域の画素の特徴量のパターンが検出され、検出された特徴量のパターンと、予め特定の被写体を含む画像から検出された特徴量のパターンとが比較され、それぞれのパターンの相関係数が演算され、演算された相関係数が画像の中に特定の被写体が写っている可能性の大きさを表す情報として生成されることを特徴とする。

【0019】

本発明のプログラムは、画像の信号と、この画像に同期した音声の信号を処理する信号処理装置に信号処理を実行させるプログラムであって、入力された画像の信号に基づいて、画像の中に特定の被写体が写っている可能性の大きさを表す情報の生成を制御する生成制御ステップと、生成制御ステップの処理により生成された情報に基づいて、画像の中に特定の被写体が写っているか否かの判定を制御する判定制御ステップと、判定制御ステップの処理により、画像の中に特定の被写体が写っていると判定された場合、被写体方向に向けて収音した音声信号のレベルを上げ、および、または被写体方向以外の方向に向けて収音した音声信号のレベルを下げるように、音声信号の指向特性の変更を制御する指向特性変更制御ステップと、判定制御ステップの処理により、画像の中に特定の被写体が写っ

10

20

30

40

50

ていると判定された場合、被写体の発する音声が含まれる周波数帯域のレベルを上げ、および、または被写体の発する音声が含まれる周波数帯域以外の周波数帯域のレベルを下げるように、音声の信号の周波数特性を変更する周波数特性変更制御ステップとをコンピュータに実行させ、生成制御ステップの処理では、入力された信号に対応する画像の中で、特定の色の画素により構成される領域が抽出され、抽出された領域の画素の輝度または色差に基づいて領域の画素の特徴量のパターンが検出され、検出された特徴量のパターンと、予め特定の被写体を含む画像から検出された特徴量のパターンとが比較され、それぞれのパターンの相関係数が演算され、演算された相関係数が画像の中に特定の被写体が写っている可能性の大きさを表す情報として生成されることを特徴とする。

#### 【0020】

本発明の信号処理装置および方法、並びにプログラムにおいては、入力された画像の信号に基づいて、画像の中に特定の被写体が写っている可能性の大きさを表す情報が生成され、生成された情報に基づいて、画像の中に特定の被写体が写っているか否かが判定され、画像の中に特定の被写体が写っていると判定された場合、被写体方向に向けて収音した音声信号のレベルを上げ、および、または被写体方向以外の方向に向けて収音した音声信号のレベルを下げるように、音声信号の指向特性が変更され、画像の中に特定の被写体が写っていると判定された場合、被写体の発する音声が含まれる周波数帯域のレベルを上げ、および、または被写体の発する音声が含まれる周波数帯域以外の周波数帯域のレベルを下げるように、音声の信号の周波数特性が変更される。また、入力された信号に対応する画像の中で、特定の色の画素により構成される領域が抽出され、抽出された領域の画素の輝度または色差に基づいて領域の画素の特徴量のパターンが検出され、検出された特徴量のパターンと、予め特定の被写体を含む画像から検出された特徴量のパターンとが比較され、それぞれのパターンの相関係数が演算され、演算された相関係数が画像の中に特定の被写体が写っている可能性の大きさを表す情報として生成される。

#### 【発明の効果】

#### 【0021】

本発明によれば、撮影された画像内において特定の被写体から発せられる音声を強調することができる。

#### 【発明を実施するための最良の形態】

#### 【0022】

以下に本発明の実施の形態を説明するが、本明細書に記載した発明と、発明の実施の形態との対応関係を例示すると、次のようになる。この記載は、本明細書に記載されている発明をサポートする実施の形態が明細書に記載されていることを確認するためのものである。従って、明細書には記載されているが、ここには記載されていない実施の形態があったとしても、そのことは、その実施の形態が、その発明に対応するものではないことを意味するものではない。逆に、実施の形態が発明に対応するものとしてここに記載されていたとしても、そのことは、その実施の形態が、その発明以外の発明には対応しないものであることを意味するものでもない。

#### 【0023】

さらに、この記載は、明細書に記載されている発明が、全て請求されていることを意味するものではない。換言すれば、この記載は、明細書に記載されている発明であって、この出願では請求されていない発明の存在、すなわち、将来、分割出願されたり、補正により出願、または追加される発明の存在を否定するものではない。

#### 【0024】

請求項1に記載の信号処理装置は、画像の信号と、この画像に同期した音声の信号を処理する信号処理装置（例えば、図2の被写体強調処理部107）であって、入力された画像の信号に基づいて、前記画像の中に特定の被写体が写っている可能性の大きさを表す情報を生成する生成手段（例えば、図2の画像認識部131）と、前記生成手段により生成された情報に基づいて、前記画像の中に特定の被写体が写っているか否かを判定する判定手段（例えば、図2の判定部132）と、前記判定手段により、前記画像の中に特定の被

10

20

30

40

50

写体が写っていると判定された場合、前記被写体方向に向けて收音した音声の信号のレベルを上げ、および、または前記被写体方向以外の方向に向けて收音した音声の信号のレベルを下げるように、前記音声の信号の指向特性を変更する指向特性変更手段（例えば、図2の指向性操作部133）と、判定手段により、画像の中に特定の被写体が写っていると判定された場合、被写体の発する音声が含まれる周波数帯域のレベルを上げ、および、または被写体の発する音声が含まれる周波数帯域以外の周波数帯域のレベルを下げるように、音声の信号の周波数特性を変更する周波数特性変更手段とを備え、生成手段は、入力された信号に対応する画像の中で、特定の色の画素により構成される領域を抽出する抽出手段と、抽出手段により抽出された領域の画素の輝度または色差に基づいて領域の画素の特徴量のパターンを検出する検出手段と、予め特定の被写体を含む画像から検出された特徴量のパターンを記憶する記憶手段と、検出手段により検出された特徴量のパターンと、記憶手段に記憶されている特徴量のパターンとを比較し、それぞれのパターンの相関係数を演算する演算手段とを有し、演算手段により演算された相関係数が前記画像の中に特定の被写体が写っている可能性の大きさを表す情報として生成されることを特徴とする。

10

## 【0027】

請求項2に記載の信号処理装置は、前記画像の中の所定の領域（例えば、画面中央の領域）を切り出す切出し手段（例えば、図3のウィンドウ切出し部161）をさらに備え、前記抽出手段は、前記切出し手段により切り出された画像の中で、特定の色の画素により構成される領域を抽出するようにすることができる。

## 【0028】

請求項4に記載の信号処理装置は、前記指向特性変更部による前記音声の信号の指向特性の変更の度合い、または前記周波数特性変更部による前記音声の信号の周波数特性の変更の度合いを制御する制御レベルを設定する設定手段（例えば、レベル設定部142）をさらに備えるようにすることができる。

20

## 【0030】

請求項7に記載の信号処理装置は、前記画像の信号と、前記音声の信号は、記録媒体に記録されたデータが再生装置（例えば、図18の映像音声記録再生装置400）により再生されることにより出力される画像の信号と、音声の信号であるようにすることができる。

## 【0031】

請求項11に記載の信号処理方法は、画像の信号と、この画像に同期した音声の信号を処理する信号処理装置（例えば、図2の被写体強調処理部107）の信号処理方法であって、入力された画像の信号に基づいて、前記画像の中に特定の被写体が写っている可能性の大きさを表す情報を生成する生成ステップ（例えば、図11のステップS101）と、前記生成ステップの処理により生成された情報に基づいて、前記画像の中に特定の被写体が写っているか否かを判定する判定ステップ（例えば、図11のステップS102）と、前記判定ステップの処理により、前記画像の中に特定の被写体が写っていると判定された場合、前記被写体方向に向けて收音した音声信号のレベルを上げ、および、または前記被写体方向以外の方向に向けて收音した音声信号のレベルを下げるように、前記音声信号の指向特性を変更する指向特性変更ステップ（例えば、図11のステップS103）と、判定ステップの処理により、画像の中に特定の被写体が写っていると判定された場合、被写体の発する音声が含まれる周波数帯域のレベルを上げ、および、または被写体の発する音声が含まれる周波数帯域以外の周波数帯域のレベルを下げるように、音声の信号の周波数特性を変更する周波数特性変更ステップとを含み、生成ステップの処理では、入力された信号に対応する画像の中で、特定の色の画素により構成される領域が抽出され、抽出された領域の画素の輝度または色差に基づいて領域の画素の特徴量のパターンが検出され、検出された特徴量のパターンと、予め特定の被写体を含む画像から検出された特徴量のパターンとが比較され、それぞれのパターンの相関係数が演算され、演算された相関係数が画像の中に特定の被写体が写っている可能性の大きさを表す情報として生成される。

30

40

## 【0032】

50

請求項12に記載のプログラムは、画像の信号と、この画像に同期した音声の信号を処理する信号処理装置（例えば、図2の被写体強調処理部107）に信号処理を実行させるプログラムであって、入力された画像の信号に基づいて、前記画像の中に特定の被写体が写っている可能性の大きさを表す情報の生成を制御する生成制御ステップ（例えば、図11のステップS101）と、前記生成制御ステップの処理により生成された情報に基づいて、前記画像の中に特定の被写体が写っているか否かの判定を制御する判定制御ステップ（例えば、図11のステップS102）と、前記判定制御ステップの処理により、前記画像の中に特定の被写体が写っていると判定された場合、前記被写体方向に向けて收音した音声信号のレベルを上げ、および、または前記被写体方向以外の方向に向けて收音した音声信号のレベルを下げるように、前記音声信号の指向特性の変更を制御する指向特性変更制御ステップ（例えば、図11のステップS103）と、判定制御ステップの処理により、画像の中に特定の被写体が写っていると判定された場合、被写体の発する音声が含まれる周波数帯域のレベルを上げ、および、または被写体の発する音声が含まれる周波数帯域以外の周波数帯域のレベルを下げるように、音声の信号の周波数特性を変更する周波数特性変更制御ステップとをコンピュータに実行させ、生成制御ステップの処理では、入力された信号に対応する画像の中で、特定の色の画素により構成される領域が抽出され、抽出された領域の画素の輝度または色差に基づいて領域の画素の特徴量のパターンが検出され、検出された特徴量のパターンと、予め特定の被写体を含む画像から検出された特徴量のパターンとが比較され、それぞれのパターンの相関係数が演算され、演算された相関係数が画像の中に特定の被写体が写っている可能性の大きさを表す情報として生成される。

10

20

【0033】

以下、図面を参照して、本発明の実施の形態について説明する。

【0034】

図1は、本発明を適用した映像音声記録装置100の一実施形態に係る構成例を示す図である。

【0035】

図1において、映像音声記録装置100は、例えば、ビデオカメラとして構成され、レンズ101を介して集光される光に基づく画像（映像）と、マイクロホン104または105を介して集音される音声を記録する。

【0036】

レンズ101を介して集光された光は、撮像素子102において例えば、光の強度に応じた電気信号に変換（光電変換）される。撮像素子102は、例えば、CCD（Charge Coupled Device）センサまたはCMOS（Complementary Metal-Oxide Semiconductor）センサなどの光電変換素子により構成される。

30

【0037】

画像信号処理部103は、撮像素子102から出力される電気信号を、A/D変換して得られるデータに基づいて、RGB画像信号のデータを生成する。

【0038】

一方、マイクロホン104またはマイクロホン105を介して集音された音声の信号は、音声信号処理部106においてA/D変換される。例えば、マイクロホン104は、映像音声記録装置（ビデオカメラ）100の前方から発せられる音声を中心に集音するように設けられており、マイクロホン105は、ビデオカメラの左右方向から発せられる音声を中心に集音するように設けられている。音声信号処理部106は、例えば、マイクロホン104またはマイクロホン105から入力される音声信号に対応する音声データであって、それぞれビデオカメラの前方の音声または左右方向の音声であることを表すデータの付加された音声データを生成する。

40

【0039】

画像信号処理部103または音声信号処理部106から出力されるデータは、被写体強調処理部107に入力される。被写体強調処理部107は、ビデオカメラにより撮影される画像の中に、例えば、ユーザが所望する被写体（例えば、人物など）の画像が含まれて

50



いるか否かをチェックして、人物などの画像が含まれている場合、その人物から発せられたと推測される前方から発せられた音声を強調する処理を行う。なお、被写体強調処理部 107 の詳細な構成については後述する。

【0040】

被写体強調処理部 107 から出力される画像信号のデータおよび音声信号のデータは、エンコード処理部 108 に入力され、エンコード処理部 108 は、入力された画像信号のデータおよび音声信号のデータを、例えば、MPEG (Moving Picture Experts Group) などの方式に基づいて圧縮符号化などの処理を行い符号化 (エンコード) し、記録ストリームを生成する。

【0041】

記録部 109 は、エンコード処理部 108 から出力される記録ストリームを、例えば、DVD (Digital Versatile Disk)、HDD (Hard Disk Drive) などの記録媒体に記録する。

【0042】

図 2 は、被写体強調処理部 107 の詳細な構成例を示すブロック図である。同図において、画像認識部 131 は、画像信号処理部 103 から供給される画像信号のデータに基づいて、例えば、ビデオカメラにより撮影された画像の中に人物などの被写体が含まれている確からしさの度合いを表す評価値を算出して、判定部 132 に出力する。

【0043】

図 3 は、画像認識部 131 の詳細な構成例を示すブロック図である。同図において、ウィンドウ切出し部 161 は、入力される画像の中の所定の部分 (例えば、画面中央の部分) をウィンドウとして切り出す。このとき、例えば、図 4 に示されるように、撮影された画像 190 の中で、画像 190 のサイズと比較して縦方向 (上下方向) および横方向 (左右方向) とともに小さいサイズとされる画面中央のウィンドウ 191 が切り出される。この例では、撮影された画像 190 の中央に人物が写っており、人物の顔の部分 (図中ハッチングされていない部分) がウィンドウ 191 として切り出されている。

【0044】

あるいはまた、ウィンドウ 191 が図 5 に示されるように切り出されるようにしてもよい。図 5 においては、撮影された画像 190 の中で、画像 190 のサイズと比較して縦方向のサイズは変わらず、横方向 (左右方向) が小さいサイズとされる画面中央のウィンドウ 191 が切り出されている。この例では、撮影された画像 190 の中央に 2 人の人物が写っており、画像 190 の中で 2 人の人物全体を含む部分がウィンドウ 191 として切り出されている。

【0045】

ウィンドウ 191 は、例えば、縦方向と横方向のサイズが予め定められた値に基づいて、設定されて切り出されるようにしてもよいし、例えば、図示せぬ操作部などを介して入力されるユーザの操作などに基づいて、縦方向と横方向のサイズが設定されて切り出されるようにしてもよい。

【0046】

図 3 に戻って、ウィンドウ切出し部 161 により切り出されたウィンドウ 191 の画素のデータは、色領域抽出部 162 に供給される。色領域抽出部 162 は、例えば、RGB 信号に対応するデータとして構成されるウィンドウ 191 の画素のデータを、次式により輝度と色差信号のデータに変換する。なお、次式において、「R」、「G」、「B」は、RGB 信号のそれぞれ「R」、「G」、「B」に対応する値を表しており、「Y」は輝度信号の値、「Cr」、「Cb」はそれぞれ色差信号の値を表している。

【0047】

$$Y = 0.299R + 0.587G + 0.114B$$

$$Cr = 0.701R - 0.587G - 0.114B$$

$$Cb = -0.299R - 0.587G + 0.886B$$

【0048】

色領域抽出部 162 は、輝度と色差信号のデータに基づいて、特定の被写体の色に対応

10

20

30

40

50

する画素と、それ以外の画素を区別し、ウィンドウ 191 の画素のデータを 2 値化する。例えば、特定の被写体として人物を抽出したい場合、個々の画素の輝度と色差信号のデータに基づいて、その画素の色が推定され、その画素が人物の色である肌色に対応する画素である場合、その画素のデータに、抽出すべき領域であることを表す所定のデータが付加され、その画素が肌色に対応する画素ではない場合、その画素のデータに、抽出すべき領域ではないことを表す所定のデータが付加されて 2 値化される。

【 0 0 4 9 】

特徴抽出部 163 は、色領域抽出部 162 により抽出された (2 値化) された、ウィンドウ 191 の画素のデータのうち抽出すべき領域の画素のデータについて、特徴量を演算する。特徴量は、例えば、画素の色の濃度または明るさなどの値に基づくヒストグラムとして演算される。

10

【 0 0 5 0 】

図 6 は、例えば、人物の顔の部分がウィンドウ 191 として切り出された場合、ウィンドウ 191 の中の肌色の画素を、所定の水平走査期間内において画素の濃度 (明るさ) によりヒストグラム化したときの例を示すグラフである。同図においては、縦軸が画素の明るさとされ、横軸が画面水平方向の時間とされ、画素のヒストグラムが線 211 で表されている。また、特徴抽出部 163 は、例えば、同様に所定の垂直走査期間内において画素の明るさと画面垂直方向の時間とに対応する画素のヒストグラムも生成する。

【 0 0 5 1 】

図 3 に戻って、特徴抽出部 163 により生成されたヒストグラムと、モデル記憶部 164 から供給されるヒストグラムに基づいて、パターンマッチング部 165 は、それぞれのヒストグラムの相類似度を表す相関係数を演算する。

20

【 0 0 5 2 】

モデル記憶部 164 には、人間の顔などに対応するヒストグラムのパターンが予め記憶されている。例えば、予め撮影された人間の顔の画像などに基づいて、図 7 のグラフにおいて線 212 で示されるようなヒストグラムが生成されて、モデル記憶部 164 に記憶されている。

【 0 0 5 3 】

図 6 の線 211 で示され、特徴抽出部 163 により生成されたヒストグラムを関数  $A[t]$  とし、図 7 の線 212 で示され、モデル記憶部 164 から供給されるヒストグラムを関数  $B[t]$  とすると、双方のヒストグラムの相関係数は、式 (1) により求めることができる。

30

【 0 0 5 4 】

【 数 1 】

$$\rho = \frac{\sum_{t=X}^Y (A[t] \cdot B[t])}{\sqrt{\sum_{t=X}^Y (A[t])^2 \cdot \sum_{t=X}^Y (B[t])^2}} \quad \dots (1)$$

40

【 0 0 5 5 】

図 3 に戻って、パターンマッチング部 165 は、演算した相関係数の値を、評価値として出力する。ここで、評価値は、特徴抽出部 163 により生成されたヒストグラムと、モデル記憶部 164 から供給されるヒストグラムとの相類似度を表しており、双方のヒストグラムの相類似度が高いほど、の値は 1 に近くなり、また、双方のヒストグラムの相類似度が低いほど、の値は 0 に近くなる。

【 0 0 5 6 】

図 2 に戻って、判定部 132 は、画像認識部 131 から出力される評価値に基づいて、ビデオカメラにより撮影された画像の中に特定の被写体 (いまの場合、人物の顔) が写っているか否かを判定する。例えば、画像認識部 131 から出力される評価値が、予め設定

50

された閾値を超えている場合、判定部 132 は、撮影された画像の中に特定の被写体が写っていると判定し、後述するように、指向性操作部 133 と音声帯域操作部 134 とを制御する制御信号を出力する。

【0057】

指向性操作部 133 は、ビデオカメラにより撮影される画像とともに入力される音声の指向特性を制御する機能ブロックであり、判定部 132 から出力される制御信号に基づいて、マイクロホン 104 または 105 を介して入力される音声の指向特性を変更する。音声帯域操作部 134 は、ビデオカメラにより撮影される画像とともに入力される音声の中で、周波数帯の強調（例えば、特定の周波数帯のレベルを増加させるなど）を制御する機能ブロックであり、判定部 132 から出力される制御信号に基づいて、集音された音声のうち所定の周波数帯のレベルを変更する。

10

【0058】

図 8 は、通常時、すなわち判定部 132 から制御信号が出力されていない場合、指向性操作部 133 により設定される、マイクロホン 104 または 105 を介して入力される音声の指向特性（ポーラパターン）の例を示す図である。同図においては、図中上側がビデオカメラの前方向とされ、図中左右がそれぞれビデオカメラの左方向または右方向とされ、実線 253 および破線 254 によりポーラパターンが示されている。なお、実線 253 および破線 254 は、それぞれその方向から発せられる音声の集音レベルを表している。ここでは例えば、臨場感を重視したポーラパターンであって、ビデオカメラの左前方に対応する矢印 251 を指向主軸とする実線 253 と、ビデオカメラの右前方に対応する矢印 252 を指向主軸とする破線 254 とがポーラパターンとして設定されている。

20

【0059】

すなわち、指向性操作部 133 は、通常時は、ビデオカメラの左前方および右前方から発生される音声が強いつレベルで集音されるように、矢印 251 および 252 を中心としたポーラパターンを設定する。

【0060】

図 9 は、ビデオカメラにより撮影された画像の中に特定の被写体が写っている場合、すなわち判定部 132 から制御信号が出力されている場合、指向性操作部 133 により設定される、マイクロホン 104 または 105 を介して入力される音声の指向特性（ポーラパターン）の例を示す図である。同図においては、図中上側がビデオカメラの前方向とされ、図中左右がそれぞれビデオカメラの左方向または右方向とされ、実線 272 によりポーラパターンが示されている。なお、実線 272 は、その方向から発せられる音声の集音レベルを表している。ここでは例えば、撮影される画像の中の中央の被写体を重視したポーラパターンであって、ビデオカメラの前方に対応する矢印 271 を指向主軸とする実線 272 がポーラパターンとして設定されている。

30

【0061】

すなわち、指向性操作部 133 は、画像の中に特定の被写体が写っているときは、ビデオカメラの前方から発生される音声が強いつレベルで集音されるように、矢印 271 を中心としたポーラパターンを設定する。

【0062】

このようにすることで、例えば、通常時（風景などを撮影しているとき）は、画面の左右方向からの音声がはっきりと聞こえる臨場感のある画像と音声を提供することができるとともに、人物などを撮影しているときは、画面の前方にいる人物が発する音声がはっきりと聞こえる画像と音声を提供することができる。

40

【0063】

図 10 は、音声帯域操作部 134 を介して出力される音声信号の周波数特性の例を示す図である。図 10 において縦軸は、レベルを表し、横軸は周波数を表す。

【0064】

音声帯域操作部 134 は、通常時、すなわち判定部 132 から制御信号が出力されていない場合、点線 291 で示されるような周波数特性を有する音声信号を出力する。点線 2

50

91は、全ての周波数帯にわたってレベルがほぼ平坦であり、このような周波数特性であれば、集音された音声は特に強調されることなく出力されることになる。

【0065】

すなわち、音声帯域操作部134は、画像の中に特定の被写体（例えば、人物）が写っていないときは、集音された音声を特に強調せずに出力する。

【0066】

一方、ビデオカメラにより撮影された画像の中に特定の被写体が写っている場合、すなわち判定部132から制御信号が出力されている場合、音声帯域操作部134は、実線292で示されるような周波数特性を有する音声信号を出力する。実線292は、点線291と比較して、ボイス帯域とされる周波数帯のレベルが増大されており、ボイス帯域とされる周波数帯以外の周波数帯のレベルが減少されている。このような周波数特性であれば、集音された音声のうち、ボイス帯域に属する周波数帯の音声が強調されて出力されることになる。ここで、ボイス帯域は、人の声の周波数帯に対応する帯域であり、例えば、400Hzから4kHzまでの間の周波数帯域がボイス帯域とされる。

10

【0067】

すなわち、音声帯域操作部134は、画像の中に特定の被写体（例えば、人物）が写っているときは、人物が発する声が強いレベルで集音されるように、ボイス帯域を強調させる。

【0068】

このようにすることで、例えば、通常時（風景などを撮影しているとき）は、画面の中のすべての物体（被写体）からの音声がはっきりと聞こえる臨場感のある画像と音声を提供することができるとともに、人物などを撮影しているときは、人物が発する音声がはっきりと聞こえる画像と音声を提供することができる。

20

【0069】

次に、図11のフローチャートを参照して、被写体強調処理部107による音声信号出力処理について説明する。この処理は、例えば、図1の映像音声記録装置100により映像と音声記録される時実行される。

【0070】

ステップS101において、画像認識部131は、図12を参照して後述する評価値演算処理を実行する。これにより、上述した評価値が演算される。

30

【0071】

ここで、図12のフローチャートを参照して、図11のステップS101の評価値演算処理の詳細について説明する。

【0072】

ステップS121において、ウィンドウ切出し部161は、入力される画像の中でウィンドウを切り出す。このとき、例えば、図4または図5を参照して上述したように、画像190の中でウィンドウ191が切り出される。

【0073】

ステップS122において、色領域抽出部162は、例えば、RGB信号に対応するデータとして構成されるウィンドウ191の画素のデータを、輝度と色差信号のデータに変換することにより切り出されたウィンドウの画像の色空間を変換する。

40

【0074】

ステップS123において、色領域抽出部162は、輝度と色差信号のデータに基づいて、特定の被写体の色（例えば、肌色）に対応する画素と、それ以外の画素を区別し、ウィンドウ191の画素のデータを2値化する。

【0075】

ステップS124において、色領域抽出部162は、ステップS123の処理で2値化されたデータに基づいて、抽出すべき領域（いまの場合、肌色の領域）を設定する。

【0076】

ステップS125において、特徴抽出部163は、ステップS124の処理で設定され

50

た抽出すべき領域の画素のデータについて、画素の色の濃度または明るさなどの値に基づくヒストグラムを生成する。このとき、例えば、図6を参照して上述したようなヒストグラムが生成される。

【0077】

ステップS126において、パターンマッチング部165は、ステップS125の処理により生成されたヒストグラムと、モデル記憶部164に記憶されているヒストグラムとのパターンマッチングを行う。このとき、上述したように、双方のヒストグラムの相似度を表す相関係数が演算される。

【0078】

ステップS127において、パターンマッチング部165は、ステップS126の処理結果に基づいて評価値を出力する。

10

【0079】

このようにして、評価値が演算されて出力される。

【0080】

図11に戻って、ステップS101の処理の後、ステップS102において、判定部132は、ステップS101の処理の結果出力された評価値は、閾値を超えているか否かを判定する。閾値は、例えば、予め設定された閾値であり、ステップS101の処理で演算された閾値が評価値を超えている場合、撮影された画像のウィンドウの中に、例えば、人物の顔などの特定の被写体が写っている可能性が高いと推定される。

【0081】

20

ステップS101において、評価値は、閾値を超えていると判定された場合、処理は、ステップS103に進み、判定部132は制御信号を出力し、指向性操作部133は、マイクロホン104または105を介して集音される音声の指向特性を変更する。このとき例えば、図9に示されるように、撮影される画像の中の中央の被写体を重視したポーラパターンであって、ビデオカメラの前方を指向主軸とするポーラパターンが設定される。

【0082】

すなわち、画像の中に特定の被写体（人物）が写っているので、ビデオカメラの前方から発生される音声が強いレベルで集音されるように設定される。

【0083】

ステップS104において、判定部132は制御信号を出力し、音声帯域操作部134は、例えば図10を参照して上述したように、ボイス帯域とされる周波数帯のレベルを増大させ、ボイス帯域とされる周波数帯以外の周波数帯のレベルが減少させ、ボイス帯域の周波数帯を強調する。

30

【0084】

すなわち、画像の中に特定の被写体（人物）が写っているので、人物が発する声が強いレベルで集音されるように設定される。

【0085】

ステップS105において、ステップS103およびステップS104において、それぞれ指向性操作部133および音声帯域操作部134の処理を経た音声信号が出力される。

【0086】

40

一方、ステップS102において、ステップS101の処理の結果出力された評価値は、閾値を超えていないと判定された場合、ステップS103とS104の処理は、スキップされる。

【0087】

すなわち、画像の中に特定の被写体（人物）が写っていないので、ステップS105では、ビデオカメラの左右から発生される音声が強いレベルで集音され、また、特定の周波数帯が強調されることなく音声信号が出力されることになる。

【0088】

このようにして、被写体強調処理部107から音声信号が出力される。このようにすることで、風景などを撮影しているときは、画面の中のすべての物体（被写体）からの音声

50

がはっきりと聞こえる臨場感のある画像と音声を提供することができるとともに、人物などを撮影しているときは、人物が発する音声がはっきりと聞こえる画像と音声を提供することができる。

【 0 0 8 9 】

以上においては、被写体強調処理部 1 0 7 により、画像の中に特定の被写体（人物）が写っているときと、写っていないときに対応して音声信号が変更される例について説明したが、被写体強調処理部 1 0 7 により、段階的に音声信号が変更されるようにすることも可能である。例えば、図 3 のパターンマッチング部 1 6 5 により演算される相関関数の値に応じて段階的に音声信号が変更されるようにしてもよい。

【 0 0 9 0 】

図 1 3 は、この場合の被写体強調処理部 1 0 7 の詳細な構成例を示すブロック図である。同図において、図 2 と対応する部分には、それぞれ同一の符号が付されている。

【 0 0 9 1 】

図 1 3 の例では、図 2 の場合における判定部 1 3 2 に代わり、レベル設定部 1 4 2 が設けられている。レベル設定部 1 4 2 は、画像認識部 1 3 1 から出力される評価値に基づいて、指向性操作部 1 3 3 と音声帯域操作部 1 3 4 とを制御する制御レベルを設定する。レベル設定部 1 4 2 は、例えば、画像認識部 1 3 1 から出力される評価値を、10 段階の制御レベルに対応させて正規化して得られた制御レベルを、指向性操作部 1 3 3 および音声帯域操作部 1 3 4 に供給する。

【 0 0 9 2 】

被写体強調処理部 1 0 7 が図 1 3 のように構成される場合、指向性操作部 1 3 3 は、レベル設定部 1 4 2 により設定された制御レベルに応じて、段階的にマイクロホン 1 0 4 または 1 0 5 を介して入力される音声の指向特性（ポーラパターン）を変化させる。

【 0 0 9 3 】

図 1 4 と図 1 5 は、レベル設定部 1 4 2 の制御レベルに応じて設定される指向特性の例を示す図である。なお、同図においては、図 8 の場合と同様に、図中上側がビデオカメラの前方向とされ、図中左右がそれぞれビデオカメラの左方向または右方向とされ、矢印 2 5 1 と 2 5 2 は指向主軸を表し、実線 2 5 3 および破線 2 5 4 によりポーラパターンが示されている。

【 0 0 9 4 】

図 1 4 は、例えば制御レベルが 1 である場合の指向特性（ポーラパターン）の例を示す図である。この例では、図 8 に示される通常時のポーラパターンと比較して、指向主軸である矢印 2 5 1 と 2 5 2 が、より前方向に近づいている。すなわち、図 8 の場合と比較して、矢印 2 5 1 と 2 5 2 とが、それぞれ前方向の軸となす角度が小さくなっており、実線 2 5 3 および破線 2 5 4 により示されるポーラパターンは、より前方向が重視されたものになっている。

【 0 0 9 5 】

図 1 5 は、例えば、制御レベルが 2 である場合の指向特性（ポーラパターン）の例を示す図である。この例では、図 1 4 の場合と比較してさらに、指向主軸である矢印 2 5 1 と 2 5 2 が、より前方向に近づいている。すなわち、図 1 4 の場合と比較して、矢印 2 5 1 と 2 5 2 とが、それぞれ前方向の軸となす角度がさらに小さくなっており、実線 2 5 3 および破線 2 5 4 により示されるポーラパターンは、さらに前方向が重視されたものになっている。

【 0 0 9 6 】

このように、指向性操作部 1 3 3 は、レベル設定部 1 4 2 から供給される制御レベルが高くなるほど、前方向を重視した指向特性を設定する。このようにすることで、例えば、撮影される画像の中に人物が写っている確度が高い（評価値が高い）ほど、画面の前方にいる人物が発する音声がはっきりと聞こえるように、画像と音声を提供することができる。

【 0 0 9 7 】

10

20

30

40

50

被写体強調処理部 107 が図 13 のように構成される場合、音声帯域操作部 134 は、例えば、レベル設定部 142 により設定された制御レベルに応じて、段階的にボイス帯域を強調させる。

【0098】

図 16 は、レベル設定部 142 の制御レベルに応じて変化する、音声帯域操作部 134 を介して出力される音声信号の周波数特性の例を示す図である。図 16 において縦軸は、レベルを表し、横軸は周波数を表す。

【0099】

同図において、線 301 で示される周波数特性は、例えば、制御レベルが 1 である場合の周波数特性とされ、通常時の周波数特性である点線 291 と比較して、ボイス帯域の周波数が強調された周波数特性とされている。

10

【0100】

また、線 302 で示される周波数特性は、例えば、制御レベルが 2 である場合の周波数特性とされ、線 301 の場合と比較して、ボイス帯域の周波数がさらに強調された周波数特性とされている。

【0101】

このように、音声帯域操作部 134 は、レベル設定部 142 から供給される制御レベルが高くなるほど、ボイス帯域の周波数を強調する。このようにすることで、例えば、撮影される画像の中に人物が写っている確度が高い（評価値が高い）ほど、人物が発する音声がはっきりと聞こえるように、画像と音声を提供することができる。

20

【0102】

ここまで、レベル設定部 142 が評価値に応じて制御レベルを設定する例について説明したが、例えば、レベル設定部 142 に、パターンマッチング部 165 により演算された評価値とともに、色領域抽出部 162 により抽出された領域の大きさを表す情報を供給し、レベル設定部 142 は、例えば、評価値が閾値を超える場合、色領域抽出部 162 により抽出された領域の大きさを表す情報を、10段階の制御レベルに対応させて正規化して得られた制御レベルを、指向性操作部 133 および音声帯域操作部 134 に供給するようにしてもよい。

【0103】

このようにすることで、例えば、撮影される画像の中で人物がズームアップされるなどした場合、ズームアップされた度合いが高い（抽出された領域が大きい）ほど、人物が発する音声がはっきりと聞こえるように、画像と音声を提供することができる。

30

【0104】

図 17 は、被写体強調処理部 107 が図 13 のように構成される場合の、音声信号出力処理の例を説明するフローチャートである。

【0105】

ステップ S201 の処理は、図 11 のステップ S101 の処理と同様の処理なので詳細な説明は省略する。

【0106】

ステップ S202 において、レベル設定部 142 は、ステップ S101 の処理により演算された評価値を、例えば、10段階の制御レベルに正規化し、制御レベルを設定する。なお、ここで上述したように、評価値が閾値を超える場合、色領域抽出部 162 により抽出された領域の大きさを表す情報を、10段階の制御レベルに対応させて正規化し、制御レベルを設定してもよい。

40

【0107】

ステップ S203 において、指向性操作部 133 は、ステップ S202 の処理で設定された制御レベルに応じて指向特性を変化させる。このとき、例えば、図 14 と図 15 を参照して上述したように、指向特性が段階的に変化させられる。

【0108】

ステップ S204 において、音声帯域操作部 134 は、ステップ S202 の処理で設定さ

50

れた制御レベルに応じてボイス帯域の周波数を強調する。このとき、例えば、図16を参照して上述したように、ボイス帯域の周波数が段階的に強調される。

【0109】

ステップS205において、ステップS203およびステップS204において、それぞれ指向性操作部133および音声帯域操作部134の処理を経た音声信号が出力される。

【0110】

このようにして、被写体強調処理部107から音声信号が出力される。このようにすることで、風景などを撮影しているときは、画面の中のすべての物体（被写体）からの音声がはっきりと聞こえる臨場感のある画像と音声を提供できるとともに、人物などを撮影しているときは、その画像における人物の写り方に応じて段階的に人物が発する音声がはっきりと聞こえる画像と音声を提供することができる。

10

【0111】

ここまで、例えば、ビデオカメラなどとして構成される映像音声記録装置100を例として特定の被写体の音声を強調する例について説明したが、例えば、DVDプレーヤなどで構成される映像音声再生装置により特定の被写体の音声が強調されるようにすることも可能である。

【0112】

この場合、映像音声再生装置に上述した図2または図13に示されるような被写体強調処理部107を組み込み、例えば、複数の指向特性（例えば、前方を重視した指向特性と左右方向を重視した指向特性）に対応する音声をそれぞれ異なるチャンネルに記録したマルチチャンネルのDVDなどを、映像音声再生装置により再生する。そして、映像音声再生装置に内蔵された被写体強調処理部107が再生される映像（画像）を、図12を参照して上述したように解析して、画像の中に特定の被写体が写っている場合、図11または図17を参照して上述したように、その被写体から発せられる音声が強調されるようにすればよい。

20

【0113】

ただし、この場合、指向性操作部133は、マイクロホンを通じて入力される音声の指向特性を変更させるのではなく、所定のチャンネルの出力レベルを上げるなどして、その被写体から発せられる音声が強調されるように指向特性を変更することになる。

【0114】

また、映像音声記録装置と、映像音声再生装置をあわせて1つの装置（映像音声記録再生装置）として構成することも可能である。図18は、映像音声記録再生装置400の構成例を示すブロック図である。

30

【0115】

同図において、レンズ401乃至エンコード処理部408は、図1のレンズ101乃至エンコード処理部108と同様なので詳細な説明は省略する。

【0116】

図18の例では、映像音声記録再生装置400にスイッチ409が設けられており、映像音声を記録する場合、スイッチ401は、エンコード処理部408と、記録再生部410を接続するように設定され、エンコード処理部408から出力される映像音声のストリームデータが、記録再生部の記録媒体などに記録される。

40

【0117】

また、映像音声を再生する場合、スイッチ401は、デコード処理部411と、記録再生部410を接続するように設定され、記録再生部410から出力される圧縮符号化された映像音声のストリームデータが、デコード処理部411により復号（デコード）され、デコードされた画像信号と音声信号に対応するデータが被写体強調処理部412に供給される。

【0118】

被写体強調処理部412は、図2または図13に示されるような被写体強調処理部107と同様の構成を有しており、デコード処理部411から供給される画像のデータを、図

50



12を参照して上述したように解析して、画像の中に特定の被写体が写っている場合、図11または図17を参照して上述したように、その被写体から発せられる音声を強調させる処理を行う。ただし記録時と異なり再生時は、例えば、複数の指向特性（例えば、前方を重視した指向特性と左右方向を重視した指向特性）に対応する音声それぞれ異なるチャンネルの音声のデータとして被写体強調処理部412に供給され、所定のチャンネルの出力レベルを上げるなどして、その被写体から発せられる音声が強調されるように指向特性が変更される。

【0119】

そして、被写体強調処理部412から出力された音声信号がアンプ413により増幅されて図示せぬスピーカなどから出力される。

10

【0120】

このようにすることで、記録時も再生時も、人物などが写っているときは、人物が発する音声がはっきりと聞こえる画像と音声を提供することができる。

【0121】

なお、上述した一連の処理は、ハードウェアにより実行させることもできるし、ソフトウェアにより実行させることもできる。上述した一連の処理をソフトウェアにより実行させる場合には、そのソフトウェアを構成するプログラムが、専用のハードウェアに組み込まれているコンピュータ、または、各種のプログラムをインストールすることで、各種の機能を実行することが可能な、例えば図19に示されるような汎用のパーソナルコンピュータ500などに、ネットワークや記録媒体からインストールされる。

20

【0122】

図19において、CPU(Central Processing Unit)501は、ROM(Read Only Memory)502に記憶されているプログラム、または記憶部508からRAM(Random Access Memory)503にロードされたプログラムに従って各種の処理を実行する。RAM503にはまた、CPU501が各種の処理を実行する上において必要なデータなども適宜記憶される。

【0123】

CPU501、ROM502、およびRAM503は、バス504を介して相互に接続されている。このバス504にはまた、入出力インタフェース505も接続されている。

【0124】

入出力インタフェース505には、キーボード、マウスなどよりなる入力部506、CRT(Cathode Ray Tube)、LCD(Liquid Crystal display)などよりなるディスプレイ、並びにスピーカなどよりなる出力部507、ハードディスクなどより構成される記憶部508、モデム、LANカードなどのネットワークインタフェースカードなどより構成される通信部509が接続されている。通信部509は、インターネットを含むネットワークを介しての通信処理を行う。

30

【0125】

入出力インタフェース505にはまた、必要に応じてドライブ510が接続され、磁気ディスク、光ディスク、光磁気ディスク、或いは半導体メモリなどのリムーバブルメディア511が適宜装着され、それらから読み出されたコンピュータプログラムが、必要に応じて記憶部508にインストールされる。

40

【0126】

上述した一連の処理をソフトウェアにより実行させる場合には、そのソフトウェアを構成するプログラムが、インターネットなどのネットワークや、リムーバブルメディア511などからなる記録媒体からインストールされる。

【0127】

なお、この記録媒体は、図19に示される、装置本体とは別に、ユーザにプログラムを配信するために配布される、プログラムが記録されている磁気ディスク（フロッピディスク（登録商標）を含む）、光ディスク（CD-ROM(Compact Disk-Read Only Memory)、DVD(Digital Versatile Disk)を含む）、光磁気ディスク（MD(Mini-Disk)（登録商標）を含む）、もしくは半導体メモリなどよりなるリムーバブルメディア511により構成されるも

50

のだけでなく、装置本体に予め組み込まれた状態でユーザに配信される、プログラムが記録されているROM 502や、記憶部508に含まれるハードディスクなどで構成されるものも含む。

【0128】

本明細書において上述した一連の処理を実行するステップは、記載された順序に沿って時系列的に行われる処理はもちろん、必ずしも時系列的に処理されなくとも、並列的あるいは個別に実行される処理をも含むものである。

【図面の簡単な説明】

【0129】

【図1】本発明を適用した映像音声記録装置の一実施形態に係る構成例を示す図である。 10

【図2】図1の被写体強調処理部の詳細な構成例を示すブロック図である。

【図3】図2の画像認識部の詳細な構成例を示すブロック図である。

【図4】ウィンドウの例を示す図である。

【図5】ウィンドウの別の例を示す図である。

【図6】図3の特徴抽出部により生成されるヒストグラムの例を示す図である。

【図7】図3のモデル記憶部に記憶されているヒストグラムの例を示す図である。

【図8】通常時の指向特性（ポーラパターン）の例を示す図である。

【図9】特定の被写体が写っているときの指向特性（ポーラパターン）の例を示す図である。

【図10】図2の音声帯域操作部を介して出力される音声信号の周波数特性の例を示す図である。 20

【図11】音声信号出力処理を説明するフローチャートである。

【図12】評価値演算処理を説明するフローチャートである。

【図13】図1の被写体強調処理部の別の構成例を示すブロック図である。

【図14】図13の指向性操作部により設定される指向特性（ポーラパターン）の例を示す図である。

【図15】図13の指向性操作部により設定される指向特性（ポーラパターン）の例を示す図である。

【図16】図13の音声帯域操作部を介して出力される音声信号の周波数特性の例を示す図である。 30

【図17】音声信号出力処理の別の例を説明するフローチャートである。

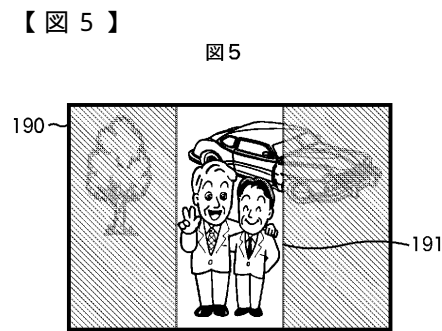
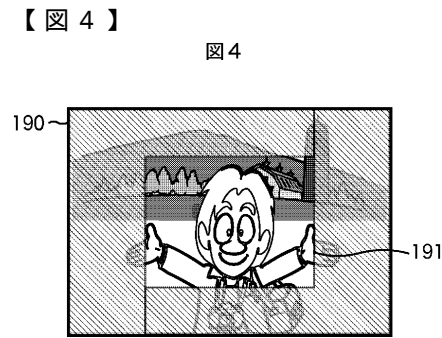
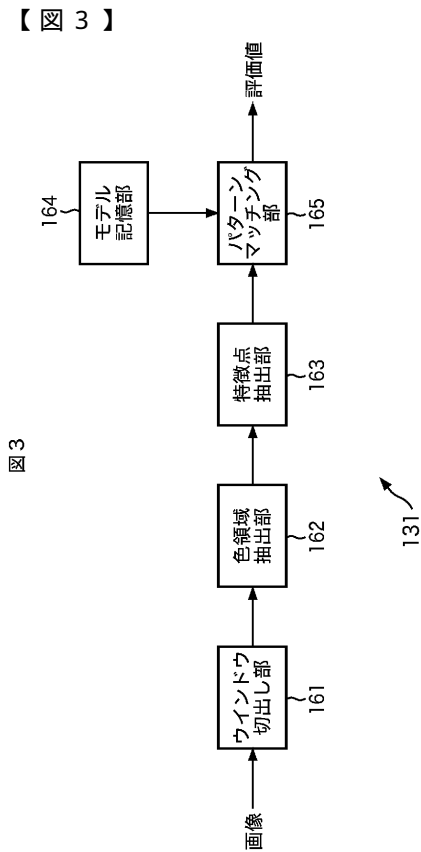
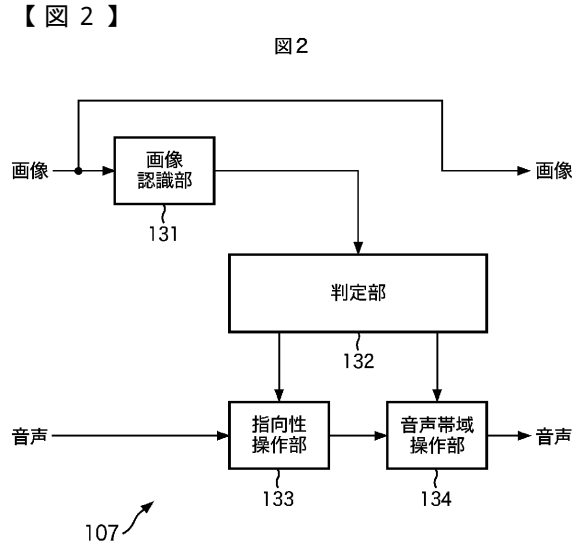
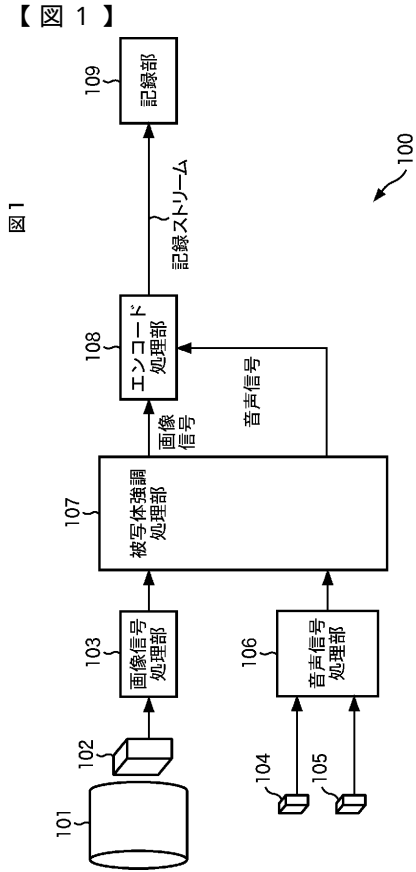
【図18】本発明を適用した映像音声記録再生装置の一実施形態に係る構成例を示す図である。

【図19】パーソナルコンピュータの構成例を示すブロック図である。

【符号の説明】

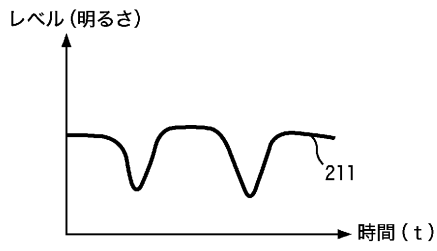
【0130】

100 映像音声記録装置, 107 被写体強調処理部, 131 画像認識部,  
132 判定部, 133 指向性操作部, 134 音声帯域操作部, 142 レベル判定部,  
161 ウィンドウ切出し部, 162 色領域抽出部, 163 特徴抽出部,  
164 モデル記憶部, 165 パターンマッチング部, 400 映像音声記録再生装置 40



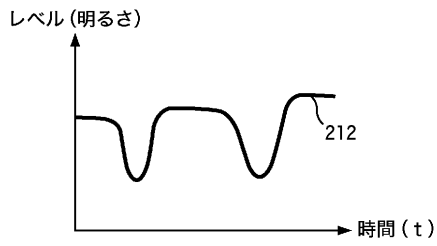
【図6】

図6



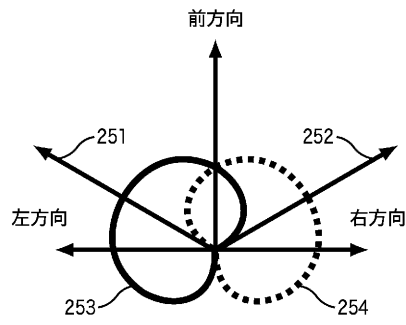
【図7】

図7



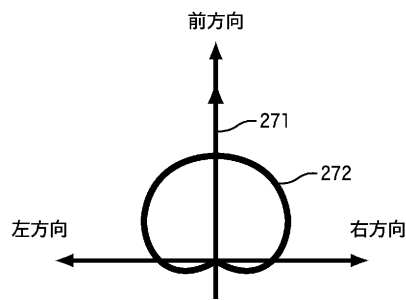
【図8】

図8

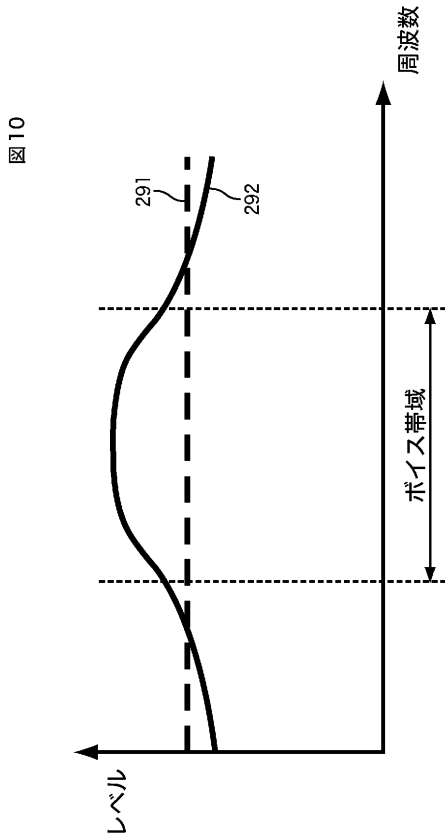


【図9】

図9

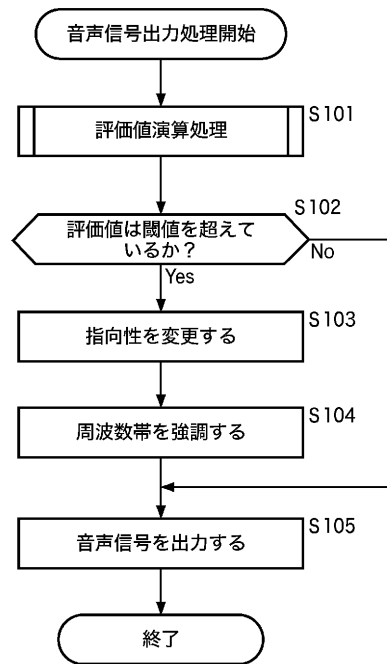


【図10】



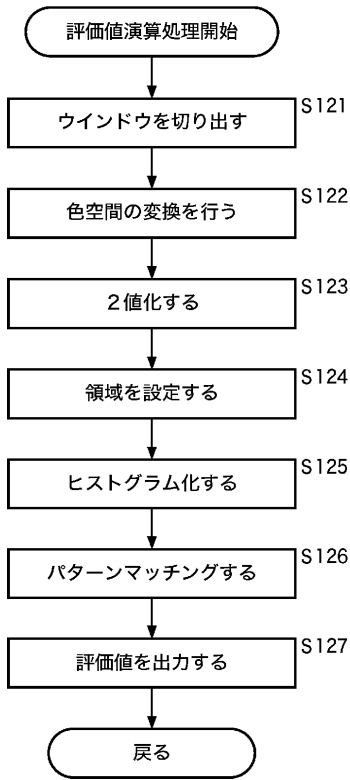
【図11】

図11



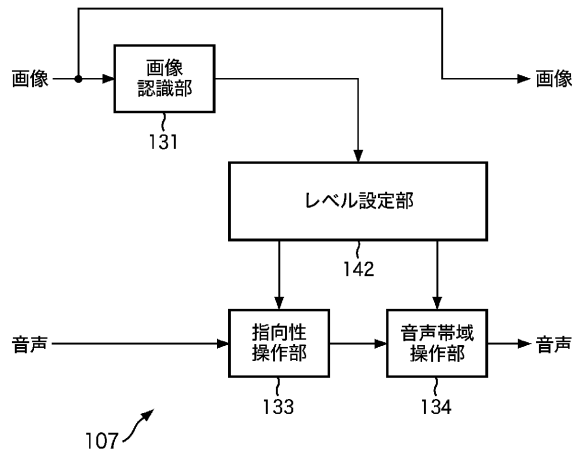
【図12】

図12



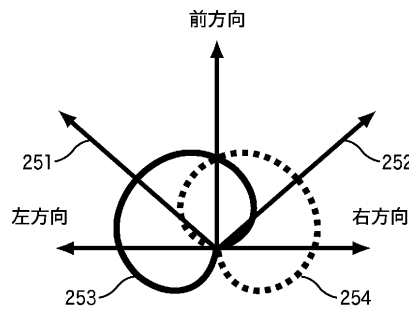
【図13】

図13



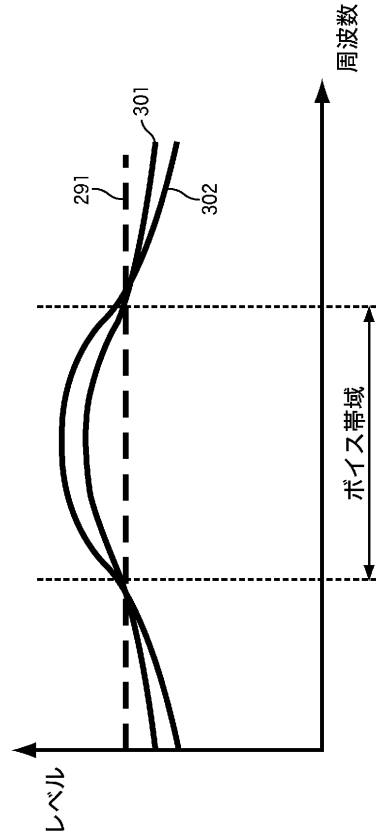
【図14】

図14



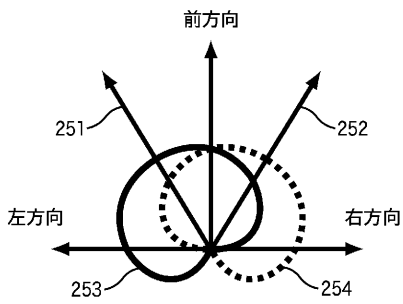
【図16】

図16

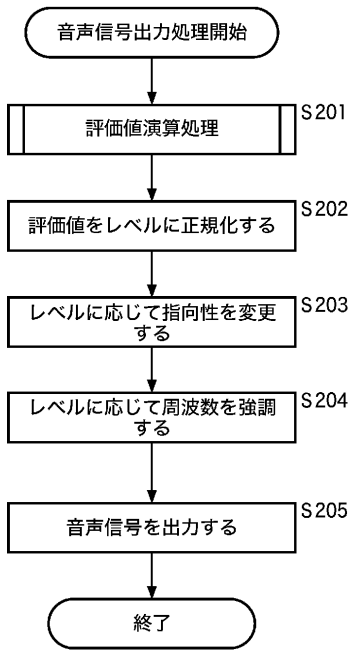


【図15】

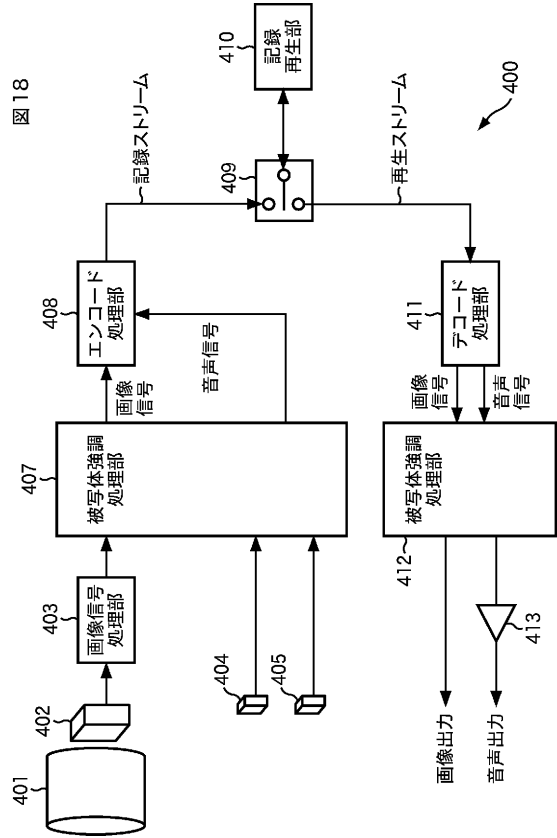
図15



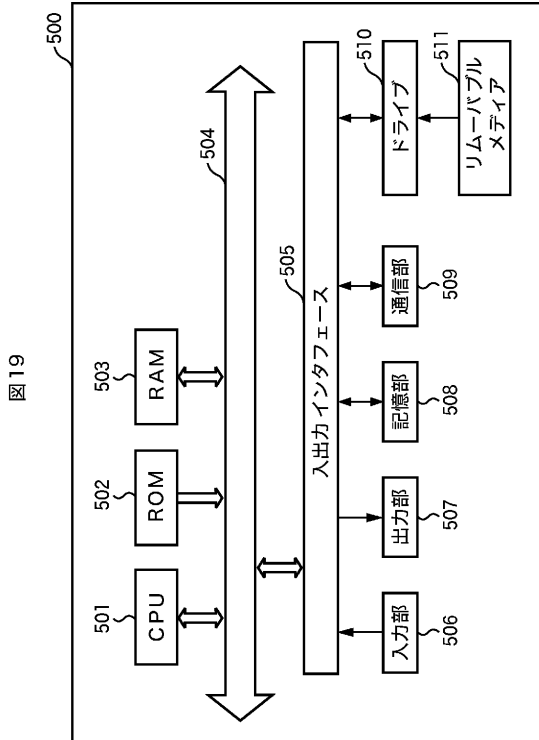
【図17】  
図17



【図18】  
図18



【図19】



---

フロントページの続き

- (56)参考文献 特開平09 - 200601 (JP, A)  
特開平11 - 098383 (JP, A)  
特開2001 - 352530 (JP, A)  
特開2005 - 151042 (JP, A)

(58)調査した分野(Int.Cl., DB名)

H04N	5 / 2 2 5
G06T	1 / 0 0
G06T	7 / 0 0
H04N	5 / 2 3 2