

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4557035号  
(P4557035)

(45) 発行日 平成22年10月6日(2010.10.6)

(24) 登録日 平成22年7月30日(2010.7.30)

(51) Int.Cl.		F I			
<b>HO4S</b>	<b>7/00</b>	<b>(2006.01)</b>	HO4S	7/00	F
<b>HO4N</b>	<b>5/225</b>	<b>(2006.01)</b>	HO4N	5/225	F

請求項の数 6 (全 22 頁)

(21) 出願番号	特願2008-97510 (P2008-97510)	(73) 特許権者	000002185
(22) 出願日	平成20年4月3日(2008.4.3)		ソニー株式会社
(65) 公開番号	特開2009-253555 (P2009-253555A)		東京都港区港南1丁目7番1号
(43) 公開日	平成21年10月29日(2009.10.29)	(74) 代理人	100122884
審査請求日	平成21年5月13日(2009.5.13)		弁理士 角田 芳末
		(74) 代理人	100133824
			弁理士 伊藤 仁恭
		(72) 発明者	近藤 哲二郎
			東京都港区港南1丁目7番1号 ソニー株式会社社内
		(72) 発明者	小久保 哲志
			東京都港区港南1丁目7番1号 ソニー株式会社社内

最終頁に続く

(54) 【発明の名称】 情報処理装置、情報処理方法、プログラム及び記録媒体

(57) 【特許請求の範囲】

【請求項1】

異なる撮影位置で撮像される第1及び第2の画像に同一の被写体が含まれ、前記第1の画像は前記被写体のパン画像であり、前記第2の画像は前記第1の画像に含まれる前記被写体のズーム画像である場合に、前記第1の画像を撮像する第1のカメラ及び前記第2の画像を撮像する第2のカメラの動き量をキャンセルした上で、入力された前記第1及び第2の画像に含まれる被写体が同一であることを特定する特徴量を求め、前記第1及び第2の画像を構成する画素毎に検出した所定の色座標系における色域の度数を各画像の色域毎に累積した値が所定の閾値より大きい色域が含まれる部分を、前記第1及び第2の画像における特徴部分として抽出する特徴量検出部と、

前記特徴量検出部によって求められた前記特徴量に基づいて前記被写体の動きを検出する画像処理部と、

前記第1及び第2の画像に共通する色域を特定し、前記第1の画像の特徴部分と前記第2の画像の特徴部分において共通する色域に該当する領域を、前記第2の画像から除去する除去領域検出部と、

前記共通する色域に該当する領域が除去された前記第2の画像を前記第1の画像に重ねて合成した画像を画像表示装置に出力する画像合成部と、

前記被写体の特徴量に基づいて、前記第1の画像に含まれる前記被写体の動きベクトルの変化量又は前記被写体の動き量を微分した動き量の変化量に応じて、音像を定位した音声<sub>10</sub>を出力する音声処理部と、を備える

情報処理装置。

【請求項 2】

請求項 1 に記載の情報処理装置において、

前記画像合成部によって合成される前記第 1 の画像に対する前記第 2 の画像は、前記第 1 の画像の時間変化によって異なる画像を重ねた

情報処理装置。

【請求項 3】

請求項 1 又は 2 に記載の情報処理装置において、

前記音声処理部によって音像が定位された前記第 1 の画像に含まれる前記被写体の音量を調整する操作部を備える

情報処理装置。

【請求項 4】

異なる撮影位置で撮像される第 1 及び第 2 の画像に同一の被写体が含まれ、前記第 1 の画像は前記被写体のパン画像であり、前記第 2 の画像は前記第 1 の画像に含まれる前記被写体のズーム画像である場合に、前記第 1 の画像を撮像する第 1 のカメラ及び前記第 2 の画像を撮像する第 2 のカメラの動き量をキャンセルした上で、入力された前記第 1 及び第 2 の画像に含まれる被写体が同一であることを特定する特徴量を求め、前記第 1 及び第 2 の画像を構成する画素毎に検出した所定の色座標系における色域の度数を各画像の色域毎に累積した値が所定の閾値より大きい色域が含まれる部分を、前記第 1 及び第 2 の画像における特徴部分として抽出するステップと、

前記特徴量に基づいて前記被写体の動きを検出するステップと、

前記第 1 及び第 2 の画像に共通する色域を特定し、前記第 1 の画像の特徴部分と前記第 2 の画像の特徴部分において共通する色域に該当する領域を、前記第 2 の画像から除去するステップと、

前記共通する色域に該当する領域が除去された前記第 2 の画像を前記第 1 の画像に重ねて合成した画像を出力するステップと、

前記被写体の特徴量に基づいて、前記第 1 の画像に含まれる前記被写体の動きベクトルの変化量又は前記被写体の動き量を微分した動き量の変化量に応じて、音像を定位した音声<sub>1</sub>を出力するステップと、を含む

情報処理方法。

【請求項 5】

異なる撮影位置で撮像される第 1 及び第 2 の画像に同一の被写体が含まれ、前記第 1 の画像は前記被写体のパン画像であり、前記第 2 の画像は前記第 1 の画像に含まれる前記被写体のズーム画像である場合に、前記第 1 の画像を撮像する第 1 のカメラ及び前記第 2 の画像を撮像する第 2 のカメラの動き量をキャンセルした上で、入力された前記第 1 及び第 2 の画像に含まれる被写体が同一であることを特定する特徴量を求め、前記第 1 及び第 2 の画像を構成する画素毎に検出した所定の色座標系における色域の度数を各画像の色域毎に累積した値が所定の閾値より大きい色域が含まれる部分を、前記第 1 及び第 2 の画像における特徴部分として抽出するステップと、

前記特徴量に基づいて前記被写体の動きを検出するステップと、

前記第 1 及び第 2 の画像に共通する色域を特定し、前記第 1 の画像の特徴部分と前記第 2 の画像の特徴部分において共通する色域に該当する領域を、前記第 2 の画像から除去するステップと、

前記共通する色域に該当する領域が除去された前記第 2 の画像を前記第 1 の画像に重ねて合成した画像を出力するステップと、

前記被写体の特徴量に基づいて、前記第 1 の画像に含まれる前記被写体の動きベクトルの変化量又は前記被写体の動き量を微分した動き量の変化量に応じて、音像を定位した音声<sub>2</sub>を出力するステップと、を含む

コンピュータに実行させるためのプログラム。

【請求項 6】

10

20

30

40

50

請求項 5 に記載のプログラムを格納した記録媒体。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、一画面に表示された複数の画像を効率よく視聴する場合に適用して好適な情報処理装置、情報処理方法、プログラム及び記録媒体に関する。

【背景技術】

【0002】

従来、一画面に複数の画像を合成し、表示する場合がある。一般に複数の画像を同時に表示する場合、例えば、縮小した複数の画像を並べたり、主画像の一部にアングルが異なる副画像をはめ込んだりするピクチャ・イン・ピクチャ（PinP: Picture in Picture）方式が用いられていた。このとき、主画像、副画像のいずれか一方の音声を放音していた。

【0003】

例えば、特許文献 1 には、注目する画像の絶対位置を合わせて音声を定位する技術について開示されている。

【特許文献 1】特開 2005 - 29518 号公報

【発明の開示】

【発明が解決しようとする課題】

【0004】

ところで、複数の画像を一画面に提示する場合における音声を定位は考察されていなかった。このため、同じ被写体が含まれる主画像と副画像が一画面に表示された場合、一方の画像で注目している場面は、他方の画像のどの場面に対応しているかを直感的に理解することは難しい。例えば、主画像でサッカー選手のズーム画面が表示され、副画像でグラウンドのパン画面が表示された場合、ズーム画面に表示されるサッカー選手はパン画面のどこにいるかをすぐに知ることは難しかった。

【0005】

本発明はこのような状況に鑑みて成されたものであり、同一の被写体が含まれる複数の画像を一画面に表示して同時に視聴する場合に、効率のよい視聴を実現することを目的とする。

【課題を解決するための手段】

【0006】

本発明は、異なる撮影位置で撮像される第 1 及び第 2 の画像に同一の被写体が含まれ、第 1 の画像は被写体のパン画像であり、第 2 の画像は第 1 の画像に含まれる被写体のズーム画像である場合に、第 1 の画像を撮像する第 1 のカメラ及び第 2 の画像を撮像する第 2 のカメラの動き量をキャンセルした上で、入力された第 1 及び第 2 の画像に含まれる被写体が同一であることを特定する特徴量を求め、第 1 及び第 2 の画像を構成する画素毎に検出した所定の色座標系における色域の度数を各画像の色域毎に累積した値が所定の閾値より大きい色域が含まれる部分を、第 1 及び第 2 の画像における特徴部分として抽出し、特徴量に基づいて被写体の動きを検出する。次に、第 1 及び第 2 の画像に共通する色域を特定し、第 1 の画像の特徴部分と第 2 の画像の特徴部分において共通する色域に該当する領域を、第 2 の画像から除去する。次に、共通する色域に該当する領域が除去された第 2 の画像を第 1 の画像に重ねて合成した画像を出力する。そして、被写体の特徴量に基づいて、第 1 の画像に含まれる被写体の動きベクトルの変化量又は被写体の動き量を微分した動き量の変化量に応じて、音声を定位した音声を出力するものである。

【0007】

このようにしたことで、画像に含まれる被写体の音声を定位すること（以下、「音声定位」とも言う。）によって被写体の動きが明確に示される。このため、ユーザは、1 画面

10

20

30

40

50

に複数の画像が表示される場合に、視線移動により各画像に含まれる被写体の動きを追わなくても、被写体の動きを容易に理解できる。

【発明の効果】

【0008】

本発明によれば、ユーザは、音像定位によって被写体の動きを容易に理解しやすくなるという効果がある。

【発明を実施するための最良の形態】

【0009】

以下、本発明の一実施の形態例について、図1～図18を参照して説明する。本実施の形態例では、アングル（撮影地点）やズーム（撮影倍率）が異なる複数のカメラから得られた画像を、本発明に係る情報処理装置によって一画面に表示できる情報処理システム1に適用した例について説明する。

10

【0010】

図1は、本実施の形態に係る画像処理システム1の配置例を示す。

情報処理システム1は、第1の情景を撮像し、第1の画像を生成する第1のカメラ3と、第1の情景より狭い範囲の第2の情景を撮像し、第2の画像を生成する第2のカメラ4を備える。そして、情報処理システム1は、第1のカメラ3から入力された第1の画像と第2のカメラ4から入力された第2の画像に基づいて、画像を合成する処理を施す情報処理装置10と、情報処理装置10で合成された画像を表示する出力装置6を備える。

【0011】

20

第1のカメラ3と第2のカメラ4は、互いにアングルやズームが異なる。出力装置6は、画像を表示し、音声を放音する装置である。第1の情景と第2の情景には、同一の被写体が含まれる。ただし、本実施の形態例において、第1のカメラ3は、パンして撮像し第1の画像を生成する。一方、第2のカメラ4は、ズームして撮像し第2の画像を生成する。このため、第1の情景には、第2の情景に比べて広い範囲の情景が含まれる。

【0012】

情報処理装置10は、マルチアングル放送の画像・音声や、ユーザ自身が撮影した同一の情景を異なるアングルから撮像した画像と収録した音声を保存部11（後述の図2を参照）に保存する。また、情報処理装置10は、第1のカメラ3と第2のカメラ4から入力される複数の画像のうち、一つの画像を主画像とし、主画像より画サイズが小さい他の画像を副画像として、主画像と副画像を重畳した合成画像を生成する。

30

【0013】

出力装置6としては、例えば、液晶ディスプレイ装置、有機EL（Electro-Luminescence）ディスプレイ装置、プロジェクタ装置等が含まれる。さらに、出力装置6は、音声を放音するスピーカを備える。画像の出力方法としては、複数の画像を重畳して提示することを基本とするが、ユーザが特に注目したい画像のみを出力してもよい。また、複数のディスプレイを使って複数の画像を提示することも可能であり、画像の出力形態は限定されない。

【0014】

音像の定位位置は、上下・左右・前後の3次元空間方向に移動する。全体を撮影したパン画像と、パン画像の一部の部分を詳細に撮影したズーム画像が情報処理装置10に入力されると、音像の定位位置はパン画像に含まれる被写体に合わせて決定される。これは、パン画像の中でズーム画像に含まれる被写体がどの位置にあるかを、ユーザが複数の画像から直感的に理解するためである。

40

【0015】

また、入力信号の組み合わせは、画像・音声などに限定されない。ここで、オブジェクトの位置情報や、音源の移動情報などがメタデータとして画像又は音声データに含まれている場合は、それらのデータも使って音像を定位する処理の正確性を高めることを、「入力信号の組み合わせ」と称している。

そして、複数の画像信号と音声信号を混ぜることによって音像が定位されるため、ユー

50

ザが複数の画像を觀賞する際に自然に視線を移動できる。また、定位された音像が移動することによって視線が音の移動に追従するため、複数の画面間でオブジェクトの同定が容易となる。

#### 【 0 0 1 6 】

図 2 は、情報処理装置 1 0 の内部構成例を示す。

情報処理装置 1 0 は、第 1 のカメラ 3 と第 2 のカメラ 4 並びにチューナから入力される画像を保存する保存部 1 1 を備える。また、情報処理装置 1 0 は、入力された画像に含まれる被写体の特徴量を検出する特徴量検出部 2 1 と、保存部 1 1 から読み出した画像に所定の処理を加える画像処理部 2 2 と、保存部 1 1 から読み出した音声に所定の処理を加える音声処理部 2 3 を備える。画像処理部 2 2 で処理された合成画像は、出力装置 6 によって表示される。一方、音声処理部 2 3 で処理された音像定位された音声は、出力装置 6 によって放音される。

10

#### 【 0 0 1 7 】

特徴量検出部 2 1 は、入力される画像が異なる撮影位置で撮像される第 1 及び第 2 の画像であり、第 1 及び第 2 の画像に同一の被写体が含まれる場合に、第 1 及び第 2 の画像から被写体の特徴量を求める。

#### 【 0 0 1 8 】

画像処理部 2 2 は、特徴量検出部 2 1 によって求められた特徴量に基づいて被写体の動きを検出する。音声処理部 2 3 は、画像処理部 2 2 によって検出された被写体の動きに応じて、被写体の音像を定位する。特徴量検出部 2 1、画像処理部 2 2、音声処理部 2 3 は、例えば、中央処理回路 (CPU: Central Processing Unit) 等が適用できる。

20

#### 【 0 0 1 9 】

また、情報処理装置 1 0 は、センサ情報として、ユーザ操作の情報、ユーザ情報並びにユーザの室内環境等の情報が入力される外部信号入力部 1 2 を備える。ユーザ情報には、例えば提示された画像を視聴するユーザの人数、どのような方式で画面を提示するかを選択するための情報が含まれる。また、ユーザの室内環境には、例えば画像が提示される部屋の大きさや形状、スピーカやディスプレイ画面の数や配置等の情報が含まれる。外部信号入力部 1 2 を介して入力された情報は、特徴量検出部 2 1、画像処理部 2 2、音声処理部 2 3 に供給される。

#### 【 0 0 2 0 】

保存部 1 1 は、例えば、ハードディスクドライブ (HDD: Hard Disk Drive) 等で構成される大容量の記録装置又は記録媒体である。保存部 1 1 に保存される画像は、例えば、MPEG-2 (Moving Picture Expert Group-2) 形式で圧縮されたデジタルデータである。この画像には、カメラ、チューナ、ストレージ等から供給される動画像、静止画像が含まれる。以下、これらの情報をコンテンツとも称する。また、保存部 1 1 には、カメラ情報 (例えば、カメラの位置、アングル、倍率の情報) 等も保存される。保存部 1 1 に保存された情報は、ユーザが選択したり、情報処理装置 1 0 が自動判定したりすることによって、複数の画像と音声と同時に提示されるよう加工される。

30

#### 【 0 0 2 1 】

情報処理装置 1 0 は、例えば、色空間として YUV 表色系を採用して画像を処理する。YUV 表色系で用いられる YUV 座標系では、輝度信号 (Y) と、輝度信号と青色信号の差信号 (U)、輝度信号と赤色信号の差信号 (V) で定められる色空間によって色が表現される。そして、この色空間が表現できる範囲を色域という。情報処理装置 1 0 に入力された画像を構成する画素は、必ず YUV 表色系によって表現される。本実施の形態例では、画素ごとに定まる UV 座標の位置を、「色域」ともいう。

40

#### 【 0 0 2 2 】

特徴量検出部 2 1 は、同一の被写体を撮影した複数の画像に含まれる色域を特徴量として求める。このとき、主画像と副画像が決定される。色域は、画像を構成する画素毎に検出される。色域は 3 次元の度数分布図で表せる。画像の中で同じ色が多く含まれる場合、その色に対応する色域の度数が累積される。この累積された度数が所定の閾値より大きい

50

場合に、最も出現頻度が高い色域として、その色域が含まれる画像の一部（以下、「最頻部分」ともいう。）を、その画像の特徴部分として抽出できる。

【0023】

図3は、画像処理部22の内部構成例を示す。

画像処理部22は、副画像から除去領域を検出する除去領域検出部24と、検出された除去領域が除かれた副画像を主画像に合成する画像合成部25と、を備える。

【0024】

除去領域検出部24は、特徴量検出部21で求められた複数の画像の色域から共通の色域を特定し、副画像から共通の色域に該当する領域を除去する。このとき、主画像と副画像の特徴部分に共通する色域が含まれる領域を、主画像又は副画像から除去する。例えば、主画像と副画像の両方に、グラウンドの画像が含まれる場合、このグラウンドの色域が最も画面に多く現れる色域（以下、「最頻色域」ともいう。）として検出される。そして、最頻色域が含まれるグラウンドの画像が最頻部分として除去される。

10

【0025】

画像合成部25は、除去領域検出部24で共通の色域に該当する領域が除去された副画像を主画像に重ねて合成する。ユーザの操作は、図示しないリモートコントロール装置、ボタン、タッチパネル等の操作部（不図示）を用いて行われる。操作部は、外部信号入力部12に接続される。操作部は、主画像と副画像を入れ替えたり、音声処理部23によって画像が定位された第1の画像に含まれる被写体の音量を調整したりすることができる。操作部が発する操作信号は、外部信号入力部12を介して入力される。そして、特徴量検出部21によって求められる色域と、所定の閾値と、除去領域検出部24によって除去される領域が決定される。同様に、画像合成部25によって合成される主画像に対する副画像の位置は、操作部からの入力に基づいて決定される。この位置は、例えば、タッチパネルで指示された場所で決定される。この結果、表示される合成画像は、ユーザにとって違和感が生じない。

20

【0026】

図4は、情報処理装置10が行う主処理の例を示す。

初めに、音声アシスト処理が有効であるか否かを判断する（ステップS1）。音声アシスト処理とは、パン画像とズーム画像を合成する場合に、パン画像に含まれる被写体を基準として画像を定位させる処理である。音声アシスト処理が無効である場合、主処理を終了する。

30

【0027】

一方、音声アシスト処理が有効である場合、特徴量検出部21は保存部11に保存されたコンテンツを読み出す（ステップS2）。そして、特徴量検出部21は、読み出したコンテンツに含まれる複数のアングルから撮像された第1の画像と第2の画像に共通する特徴量を検出する（ステップS3）。

【0028】

読み出したコンテンツに含まれる複数の画像が同一の被写体を撮影したものである場合、これらの画像に含まれる同一の被写体（以下、「オブジェクト」ともいう。）の有無、形状を判定する。この判定は、各画像に含まれる被写体の特徴量をマッチングすることで行われる。

40

【0029】

もしくは、ユーザが出力装置6に表示された画像に対して、被写体を示す対応点を入力してもよい。この対応点に該当する被写体を初期値とする。この初期値としては、オブジェクトの位置、動きベクトルの大きさ、オブジェクトの輪郭等がある。初期値が設定されると、オブジェクト（例えば、ボール、人物像、背番号）の動きベクトルやユーザの操作入力によって画像の特定部分を検出し、動きを追尾できる。被写体の特徴量をマッチングする処理としては、同一フレームにおけるブロックマッチングや、動き量のマッチングによる推定等がある。

【0030】

50

被写体を示す初期値が設定されると、この被写体を注目オブジェクトとする。特徴量検出部 2 1 は、画像毎に注目オブジェクトの動きを検出して、各画像におけるオブジェクトの位置を追尾する。

【 0 0 3 1 】

次に、画像処理部 2 2 は、図 5 に示す所定の画像処理を行う（ステップ S 4）。特徴量が検出されると、画像処理部 2 2 によって、画像が処理される。この処理は、ユーザの操作入力による要求や画像を出力する環境に適するように行われる。

【 0 0 3 2 】

例えば、出力装置 6 にディスプレイ画面が一つしかないときは、複数の画像が重畳して提示される。ただし、画像処理によって作成される画像は、複数のディスプレイ画面に提示されることもある。

10

【 0 0 3 3 】

次に、音声処理部 2 3 は所定の音声処理を行う（ステップ S 5）。ここでは、注目オブジェクトが検出された画像に示されるオブジェクトの位置関係に基づいて、音像の定位位置を決定する処理が行われる。

【 0 0 3 4 】

例えば、一つの画面内で同じ被写体が同じ位置となるよう制御されている場合、被写体が映る位置に音像が定位される。また、ユーザが注目する被写体が画面内で同じ位置にない場合は、一方の音像を他方の注目する被写体の位置に定位させる。音像定位とは、仮想的な音像を空間に位置させることによって、ユーザは音像から音が出ているように感じる  
ことができる技術である。例えば、左右スピーカを同じ音量、かつ、同じ時間に放音すると、左右スピーカの真ん中に音像定位される。しかし、左右スピーカの音量を変えたり、放音する時間を遅らせたりすることで、定位される音像は移動する。このため、ユーザは、音像の動きを感じることができる。

20

【 0 0 3 5 】

そして、画像処理部 2 2 は、出力装置 6 に処理された画像を出力する（ステップ S 6）。このとき、出力装置 6 は、処理された画像をディスプレイ画面に表示する。

【 0 0 3 6 】

一方、音声処理部 2 3 は、画像処理部 2 2 で処理された画像に基づいて音声を処理し、出力装置 6 に処理された音声を出力する（ステップ S 7）。このとき、出力装置 6 は、処理された音声をスピーカから放音する。出力装置 6 によって放音される音声は、パン画像における被写体の位置に音像定位される。

30

【 0 0 3 7 】

画面に含まれる被写体の動きは、特にパン画像の動きに基づいて音像定位される。このため、ユーザはズーム画像の被写体を見つつ、パン画像における被写体の位置に音像定位された音声を聴きながら、この被写体がパン画像のどの位置で動いているかを違和感なく理解できる。

【 0 0 3 8 】

図 5 は、画像処理の例を示すフローチャートである。

始めに、画像処理部 2 2 は、複数の画像をどのように合成し、出力装置 6 に表示するかを決定する（ステップ S 1 1）。この決定は、外部信号入力部 1 2 を介して入力されたユーザ操作やユーザ環境の状態に基づいて行われる。

40

【 0 0 3 9 】

そして、画像処理部 2 2 は、特徴量検出部 2 1 から画像を読み込む（ステップ S 1 2）。そして、読み込んだ複数の画像の画素毎に色域を検出する（ステップ S 1 3）。

【 0 0 4 0 】

除去領域検出部 2 4 は、読み込んだ複数の画像から最頻色域を検出する（ステップ S 1 4）。そして、第 2 の画像から除去する最頻色域が含まれる領域を決定し（ステップ S 1 5）、決定された領域から最頻色域を除去する。

【 0 0 4 1 】

50

画像合成部 25 は、最頻色域が除去された第 2 の画像を第 1 の画像に重ねて合成する（ステップ S 16）。そして、画像合成部 25 は、合成した画像を画像表示装置 6 に出力する（ステップ S 17）。

【0042】

そして、画像合成部 32 は、画像表示装置 6 に重畳した合成画像を出力するか否かを判断する（ステップ S 18）。出力された映像が所望の映像でなければ、最頻度数を変更し、色域を再決定する（ステップ S 19）。

【0043】

このとき、ユーザ操作に基づいて、もしくは、操作ログなどを用いて、除去する色域の領域を変化させたり、例外処理をしたりする等の変更を行う。例外処理とは、除去領域検出部 24 によって除去された領域が想定より大きい場合に、色を戻す処理である。そして、ステップ S 13 に処理を戻す。

10

一方、ユーザが所望する画像が得られていれば画像合成処理を終了する。

【0044】

図 6 は、音声処理の例を示すフローチャートである。

【0045】

初めに、音声処理部 23 は、画像処理部 22 で画像処理が行われた画像を読み込む（ステップ S 21）。次に、画像処理部 22 によって処理された画像がどのように出力されるかを解析する（ステップ S 22）。このとき、音声処理部 23 は、画像の提示方法やディスプレイ数を検出し、音声と画像の関係性を解析する（ステップ S 23）。

20

【0046】

そして、音声処理部 23 は、音源がどのように入力されるかを解析する（ステップ S 24）。このとき、画像信号の特徴量と音声信号の特徴量を交換することで、音像の定位位置を入れ替える。この処理では、パン画像に含まれるオブジェクトの動きに合わせて、ズーム画像に含まれるオブジェクトの音像を定位している。このように、ズーム画像の音声信号の動きをパン画像に含まれるオブジェクトの動きと交換することを、「特徴量を交換する」と称している。そして、処理された音声は、出力装置 6 のスピーカ等で放音される（ステップ S 25）。

【0047】

図 7 ~ 図 15 は、重畳される画像とそれらの色域の例を示す。

30

図 7 は、第 1 のカメラ 3 が撮像した第 1 の画像 31 の例である。

図 8 は、第 2 のカメラ 4 が撮像した第 2 の画像 32 の例である。

情報処理装置 1 に入力される画像は、同一情景を撮影した画像である。本例では、第 1 の画像 31 は被写体 35（ラグビー競技におけるラック状態）のパン画像である。第 2 の画像 32 は第 1 の画像に含まれる被写体 35 のズーム画像である。保存部 11 から読み出される複数の画像には、例えば、原画像と原画像の一部をズームした画像、又は、アングルが異なる複数の加工画像がある。ただし、これらの画像には同一の被写体が含まれる。

【0048】

画像の重畳処理が行われる場合、ユーザは複数の画像から主画像となる画像を選択する。ここでは、第 2 の画像 32 が主画像として選択されたものとする。ただし、本例では、主画像としてズーム画像を用い、副画像として加工されたパン画像を用いる。

40

【0049】

図 9 は、第 1 の画像 31 の色域分布の例である。

図 10 は、第 2 の画像 32 の色域分布の例である。

以下の説明において、色域分布の図は、YUV 座標系における  $u$ 、 $v$  と、各色域の度数を示す。撮像したカメラの画素数が異なる場合、色域分布図における度数の範囲は異なる。このため、共通部分の色域を求める際には、正規化を行い、度数を合わせる必要がある。正規化とは、2 つの画像の度数の最大値を合わせる処理である。

【0050】

図 9 と図 10 には、それぞれ最頻色域を決定するための閾値 37、37 を示している

50

。この閾値を超える色域が画像毎に多く含まれる色であると言える。特徴量検出部 2 1 は、各入力画像に対して、全画面の色域の分布をとる。そして、各色域の分布を正規化し、複数の入力画像信号から色域が共通する領域を検出する。このとき、第 1 のカメラ 3 と第 2 のカメラ 4 のパラメータを補正し、頻出色域をどちらかに合わせる。色域の変換テーブルを用いてもよい。そして、重畳する画像から除去する色域を決定する。ただし、検出した色域が含まれる領域と、除去する領域は、必ずしも 1 対 1 に対応しない。

【 0 0 5 1 】

図 1 1 は、閾値を超える第 1 の画像 3 1 の色域の例である。

図 1 2 は、閾値を超える第 2 の画像 3 2 の色域の例である。

図 1 1 と図 1 2 は、図 9 と図 1 0 に示された色域分布から所定の閾値 3 7 , 3 7 を超える色域を示している。閾値は、ユーザが任意に変更できる。そして、第 1 の画像 3 1 と第 2 の画像 3 2 に共通する色域は、以下のように求められる。

【 0 0 5 2 】

図 1 3 は、閾値を超える第 1 の画像 3 1 と第 2 の画像 3 2 の共通部分の色域の例である。

。

図 1 4 は、第 1 の画像 3 1 と第 2 の画像 3 2 の共通部分の色域を 2 次元表示した例である。

第 1 の画像 3 1 と第 2 の画像 3 2 に共通して最も多く見られる色は、グラウンドの色（例えば、茶色）である。このため、図 9 と図 1 0 で示した色域分布から、所定の閾値を超える色域分布を抽出し、各画像に共通する色域を求める。共通する色域がない場合、閾値を下げ、色域分布を広げて共通する色域を求める。そして、除去領域検出部 2 4 によって、該当する領域の色域が除去される。

【 0 0 5 3 】

図 1 5 は、第 1 の画像 3 1 に画像処理が施された第 1 の画像 3 1 の例を示す。

図 1 5 に示す画像は、二つの画像（第 1 の画像 3 1 と第 2 の画像 3 2）から冗長と判定される最頻の共通部分を抽出したものである。この共通部分の抽出は、複数の画像から背景を抽出する方法としても用いられる。

【 0 0 5 4 】

主画像（第 2 の画像 3 2）が決定されると、副画像（第 1 の画像 3 1）と共通し冗長な共通部分として抽出される最頻色域が求められる。第 1 の画像 3 1 で最も高い色域は、画面のほぼ半分を占めるグラウンドの色である。同様に、第 2 の画像 3 2 で最も高い色域は、画面のほぼ半分を占めるグラウンドの色である。このため、第 1 の画像 3 1 から最頻色域が除かれると第 1 の画像 3 1 が生成される。

【 0 0 5 5 】

また、シーンによっては、色域の最頻部分を判定するために、それぞれのカメラの動きをキャンセルする。このとき、動きベクトルの最頻値を複数の画像の共通部分として用いる。複数のカメラで撮像する場合、パンやチルトする個々のカメラの動きによって、画像のマッチングをとることが難しい場合がある。そのため、画像のマッチング前にカメラの動き量をキャンセルすることでマッチング精度を上げる処理を行っている。この処理を、「カメラの動きをキャンセルする。」とも称する。なお、画像に含まれる色の判断をするときには、カメラの動きをキャンセルする処理は不要である。一方、最頻の動きベクトルを検出してオブジェクトを特定する場合には、カメラの動きをキャンセルする処理が必要である。

【 0 0 5 6 】

図 1 6 は、第 2 の画像 3 2 を主画像として、第 1 の画像 3 1 を副画像として合成した合成画像 3 3 の例を示す。

図 1 6 ( a ) は、共通色域を除去した場合の合成画像の例を示す。

図 1 6 ( b ) は、共通色域を除去しない場合の合成画像の例を示す。

【 0 0 5 7 】

ここでは、画像処理された第 1 の画像 3 1（第 1 の画像 3 1）を第 2 の画像 3 2 に重

10

20

30

40

50

ねる。第1の画像31に施される画像処理は、上述した図15で説明した処理と同様である。ただし、副画像となる第1の画像31の画像はズーム率を変えて重畳される。このように画像合成の倍率、形状は、ユーザの操作によって変えられる。例えば、動画像である場合、個々の人物の動きが明確になる。このため、ユーザは、違和感なく重畳された画像を見ることができる。また、共通部分(本例では、グラウンドの色域)を除去した上で、複数の画像を重畳できる。これにより、ユーザが画像を觀賞する際に視点移動が少なくなる。

【0058】

第2の画像32の音声の定位位置は、第1の画像31内における被写体の位置とオブジェクトの変化によって決定される。例えば、図16(a)に示される審判36にピン型のマイクホンを取り付けて音声を收音する場合、第1の画像31のズーム率を変えたパン画像34に含まれる被写体35によって音像の定位位置が決定される。このため、第2の画像32が主画像として表示されると、パン画像34が副画像として表示されるため、被写体35の位置(左側)に定位される。

10

【0059】

合成画像33は、第2の画像32の下部に第1の画像31の一部が重ねて合成される。ただし、第1の画像31が重ねられる位置は、第2の画像32の上部、中部等に限られない。また、図16(b)に示す合成画像37のように、主画像の一部に副画像をはめ込んで提示してもよい。

【0060】

図17は、時間経過によって変化するズーム画像の音声の定位位置の例である。

図17(a)は、時刻 $t$ における合成画像と音像の例を示す。

図17(b)は、時刻 $t+1$ における合成画像と音像の例を示す。

【0061】

図17で示すように、第1の画像41に含まれる被写体46の位置を音像として音像定位させる。音像の定位位置は、上下・左右・前後の3次元空間方向に移動する。被写体46が含まれるパン画像41とズーム画像42が重ねて提示される合成画像43がスクリーンに投影される。ユーザ50は、音像45の定位を感じられるため、ズーム画像42の被写体46を見ながらパン画像41における被写体46の動きを理解できる。

20

【0062】

例えば、図17(a)において、パン画像41に含まれる被写体46の動きがユーザ50に近づくように撮像されている場合を検討する。この場合、被写体46の音像に該当する音像45についても、図17(b)に示すように、被写体46の移動に伴って定位する。被写体46の移動は被写体46の輪郭の大きさの変化や動きベクトルの変化量から求められる。このため、ユーザ50は、被写体46が自分に近づいてくるように感じられる。

30

【0063】

音像の定位は、両壁面に設置されたスピーカ51, 52によって行われる。例えば、スピーカ51, 52で放音される音声を遅延して出力するとユーザ50は音像定位を感じられる。音像の定位を変える処理としては、オブジェクトの純粋な動き量を用いるのではなく、動き量の変化量から音像の動きを求めることが考えられる。ここで、「動き量」とは、純粋に前後のフレーム差分から求められる移動量を表す。また、「動き量の変化量」とは、動き量を微分した値である。例えば、被写体が等速運動をしている場合について検討する。オブジェクトの純粋な動き量を用いると、この動き量に合わせて音像が移動する。一方、動き量の変化量を用いると、音像は移動しない。この場合、オブジェクトが加速した場合や、方向を変えた時に動き量の変化量が大きくなるため、音像の定位が移動する。

40

【0064】

図18は、時間経過によって変化するパン画像の音声の定位位置の例である。ただし、図18において、既に説明した図17に対応する部分には同一符号を付し、詳細な説明を省略する。

図18(a)は、時刻 $t$ における合成画像と音像の例を示す。

50

図18(b)は、時刻  $t + 1$  における合成画像と音像の例を示す。

【0065】

パン画像が撮像された際に、音声サラウンド録音された場合、ズーム画像の変化に応じて音量などを変える等の処理も行う。

本実施の形態例では、パン画像41の音声の定位位置が変わる。例えば、ユーザ50が注目する被写体46が左から右へ移動するシーンであれば、左のスピーカ音量が小さくなり、右のスピーカの音量が大きくなる。このため、ユーザ50は、音像の定位を感じられる。

【0066】

以上説明した本実施の形態例に係る情報処理装置10では、複数のカメラで撮像された画像から、主画像と副画像を選択して自由に画像を重ねて合成できる。このとき、どの画像を重ねて合成するかを選択したり、ズーム率や選択画像の切り出し方を変えたりする操作が可能である。このため、ユーザは、複数の画像に含まれる被写体を見る際における、視線移動が容易となるという効果がある。また、音声による視線の誘導によって複数のアングルの画像を効率よく視聴することが可能になる。

【0067】

ユーザは、画像だけではなく、音の効果によっても視線が誘導される。このため、同じオブジェクトが離れた場所に提示されていても、ユーザは視線の移動が容易であり、オブジェクトを認識することが容易になる。また、ユーザに提示される情報が画像である場合、ディスプレイパネルのように、平面上にしか情報(画像)を提示できない。しかし、ユーザに提示される情報が音声である場合、3次元空間内で音像を定位させることで位置を表現できる。従来は、ズーム画像と空間内を移動する音像だけでは、音と映像の関連情報が少ないため、音と映像の関係を同時に認識することが困難であった。この課題を解消するためには、パン画像を提示しながらパン画像の音声に基づいて音像を定位させ、ズーム画像を重ねて提示することで、複数映像を同時に視聴することができる。これによって、ユーザは複数の映像を同時に観賞しながら、画像毎に提示されていたときには分からなかった注目したいズーム画像の状況とパン画像に含まれる被写体との関係性を同時に理解できる。

【0068】

また、ユーザは主画像として提示されるズーム画像を観賞しながら、副画像として提示されるパン画像に基づいて音像定位された音声を聴くことができる。このため、ユーザはパン画像に含まれる被写体の位置関係を容易に理解できる。

【0069】

また、複数の画像を提示する場合に、従来のように複数の画面を並べると、各画像は元の画サイズより小さいサイズでしか提示されない。しかし、本実施の形態例のように画像を重ねて提示することで、主画像については画サイズを維持できる。このため、複数の画像があっても、主画像の迫力等が損なわれないという効果がある。

【0070】

また、複数の画像間で冗長な共通部分を排除して重畳することで、複数の画像を重ねても、各画像の内容を直感的に理解できる。また、重要部分は重畳した画像の最上面に配置するため、重畳される側の画像も大きく隠されることなく視聴できるという効果がある。

【0071】

また、主画像のうちユーザが指定する部分(あるいは、情報処理装置10が指定する部分)のみに副画像が重畳される。このため、従来のように二つの画像を並べて見る場合に比べて、異なるアングルから撮像された複数の画像を同時に比較することが容易となる。また、画像合成部23によって合成される第1の画像に対する第2の画像は、第1の画像の時間変化によって異なる画像を重ねてもよい。例えば、ユーザが画枠をズームして重畳画像を加工する処理として、単位時間毎に撮像した時間をずらして複数の画像を重ねてもよい。この場合、ユーザは、同じ被写体の時間による変化を明瞭に理解できるという効

10

20

30

40

50

果がある。

【 0 0 7 2 】

なお、本発明は上述した実施の形態例に限られない。以下、図 1 9 ~ 図 2 1 を参照して他の実施の形態例について説明する。ただし、図 1 9 ~ 図 2 1 において、既に説明した図 1 7 に対応する部分には同一符号を付し、詳細な説明を省略する。

【 0 0 7 3 】

図 1 9 は、床に設置された表示装置である床面ディスプレイ装置 5 3 にパン画像 4 1 が提示される例である。

本実施の形態例では、複数の画像を二つの画面に出力するものである。このとき、ズーム画像 4 2 を壁面のディスプレイ装置に提示する。そして、パン画像 4 1 を床面ディスプレイ装置 5 3 に提示する。音声処理部 2 3 は、パン画像 4 1 に含まれる被写体 4 6 が遠ざかる動作に応じて、音声の高周波成分を除いて音像を定位する。このため、被写体 4 6 が遠ざかると音声はこもる。このため、ユーザは、パン画像 4 1 に表示された被写体 4 6 の動きに合わせて、音像 4 5 の移動を感じられる。

【 0 0 7 4 】

図 2 0 は、前後左右に移動可能な可動ディスプレイ装置 5 4 にズーム画像 4 2 が提示される例である。

本実施の形態例では、ズーム画像 4 2 を可動ディスプレイ装置 5 4 に表示する。このとき、パン画像 4 1 の移動に合わせて音像を定位する。このとき、ユーザ 5 0 は可動ディスプレイ装置 5 4 に提示された被写体 4 6 を見ながらパン画像 4 1 に含まれる被写体の位置を知ることができる。

【 0 0 7 5 】

図 2 1 は、3 つ以上の画像を合成した画像が提示される例である。

本実施の形態例では、パン画像 4 1 に含まれる被写体のそれぞれに対して、ズーム画像 4 2 a , 4 2 b が重ねて合成される。ズーム画像 4 2 a , 4 2 b のそれぞれに含まれる被写体 4 6 a , 4 6 b に対応して音像 4 5 a , 4 5 b が定位される。ただし、音像 4 5 a , 4 5 b は、パン画像 4 1 に含まれる被写体の位置関係に基づいて定位される。このため、ユーザ 5 0 は、3 つ以上の画像が一画面に同時に提示された場合であっても、各画像の関係を理解しやすい。このため、提示される画像の数は限定されない。

【 0 0 7 6 】

また、図 1 9 ~ 図 2 1 に示した画像と音声の出力形態はそれぞれを組み合わせてもよい。このとき、画像の重畳の仕方をオブジェクトの位置に合わせたり、ピクチャ・イン・ピクチャで合成画像を提示したり、別の表示装置に画像を提示したりする等、様々な方式で提示する方法を選択できるようにしてもよい。

【 0 0 7 7 】

また、上述した実施の形態例における一連の処理は、ハードウェアにより実行することができるが、ソフトウェアにより実行させることもできる。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を、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汎用のパーソナルコンピュータなどに所望のソフトウェアを構成するプログラムをインストールして実行させる。

【 0 0 7 8 】

また、上述した実施の形態例の機能を実現するソフトウェアのプログラムコードを記録した記録媒体を、システムあるいは装置に供給し、そのシステムあるいは装置のコンピュータ（または CPU 等の制御装置）が記録媒体に格納されたプログラムコードを読み出し実行することによっても、達成されることは言うまでもない。

【 0 0 7 9 】

この場合のプログラムコードを供給するための記録媒体としては、例えば、フロッピディスク、ハードディスク、光ディスク、光磁気ディスク、CD-ROM、CD-R、磁気テープ、不揮発性のメモリカード、ROMなどを用いることができる。

10

20

30

40

50

## 【0080】

また、コンピュータが読み出したプログラムコードを実行することにより、上述した実施の形態例の機能が実現されるだけでなく、そのプログラムコードの指示に基づき、コンピュータ上で稼動しているOSなどが実際の処理の一部又は全部を行い、その処理によって上述した実施の形態例の機能が実現される場合も含まれる。

## 【0081】

また、本明細書において、ソフトウェアを構成するプログラムを記述するステップは、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。

## 【0082】

さらに、本発明は上述した実施の形態例に限られるものではなく、本発明の要旨を逸脱することなくその他種々の構成を取り得ることは勿論である。

## 【図面の簡単な説明】

## 【0083】

【図1】本発明の一実施の形態例における情報処理装置の配置例を示す説明図である。

【図2】本発明の一実施の形態例における情報処理装置の内部構成例を示すブロック図である。

【図3】本発明の一実施の形態例における画像処理部の内部構成例を示すブロック図である。

【図4】本発明の一実施の形態例における画像の色域検出処理と画像合成処理の例を示すフローチャートである。

【図5】本発明の一実施の形態例における画像処理の例を示すフローチャートである。

【図6】本発明の一実施の形態例における音声処理の例を示すフローチャートである。

【図7】本発明の一実施の形態例における第1の画像の例を示す説明図である。

【図8】本発明の一実施の形態例における第2の画像の例を示す説明図である。

【図9】本発明の一実施の形態例における第1の画像の色域分布の例を示す説明図である。

【図10】本発明の一実施の形態例における第2の画像の色域分布の例を示す説明図である。

【図11】本発明の一実施の形態例における閾値を超える第1の画像の色域の例を示す説明図である。

【図12】本発明の一実施の形態例における閾値を超える第2の画像の色域の例を示す説明図である。

【図13】本発明の一実施の形態例における第1及び第2の画像の色域分布のうち、閾値を超える共通部分の色域を3次元表示した例を示す説明図である。

【図14】本発明の一実施の形態例における第1及び第2の画像の色域分布のうち、閾値を超える共通部分の色域を2次元表示した例を示す説明図である。

【図15】本発明の一実施の形態例における第2の画像から所定の色域（グラウンドの色）を除いた例を示す説明図である。

【図16】本発明の一実施の形態例における合成画像の例を示す説明図である。

【図17】本発明の一実施の形態例における時間経過によって変化するズーム画像の音声の定位位置の例を示す説明図である。

【図18】本発明の一実施の形態例における時間経過によって変化するパン画像の音声の定位位置の例を示す説明図である。

【図19】本発明の他の実施の形態例における複数の画面に提示される例を示す説明図である。

【図20】本発明の他の実施の形態例における可動ディスプレイ装置を用いて画面に提示される例を示す説明図である。

【図21】本発明の他の実施の形態例における3つ以上の画像を同時に提示する例を示す説明図である。

10

20

30

40

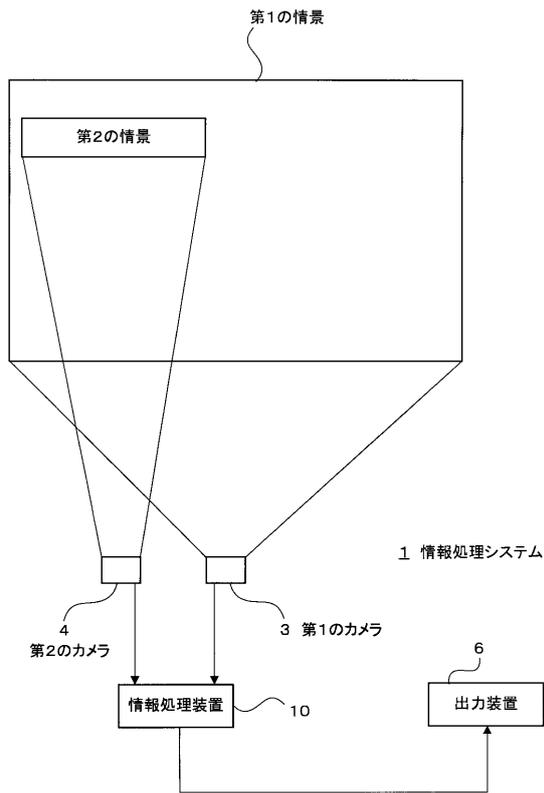
50

【符号の説明】

【0084】

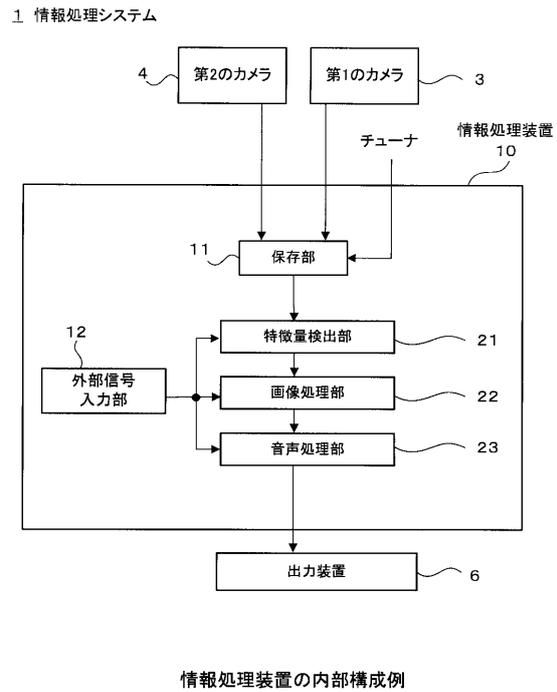
1 ... 情報処理システム、3 ... 第1のカメラ、4 ... 第2のカメラ、6 ... 画像表示装置、10 ... 情報処理装置、11 ... 保存部、12 ... 外部信号入力部、21 ... 特徴量検出部、22 ... 画像処理部、23 ... 音声処理部、24 ... 除去領域検出部、25 ... 画像合成部

【図1】



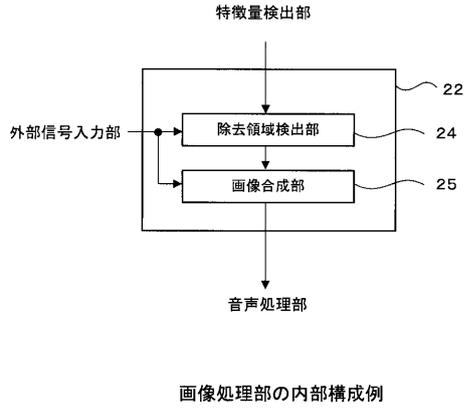
情報処理システムの配置例

【図2】

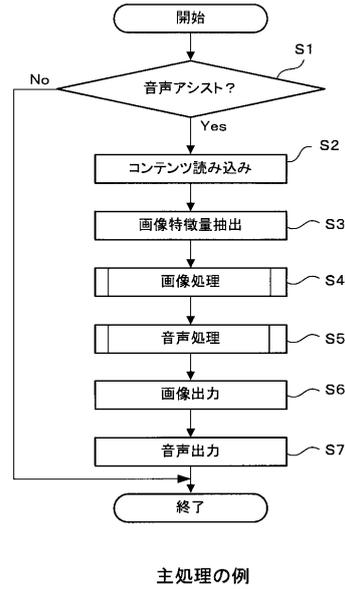


情報処理装置の内部構成例

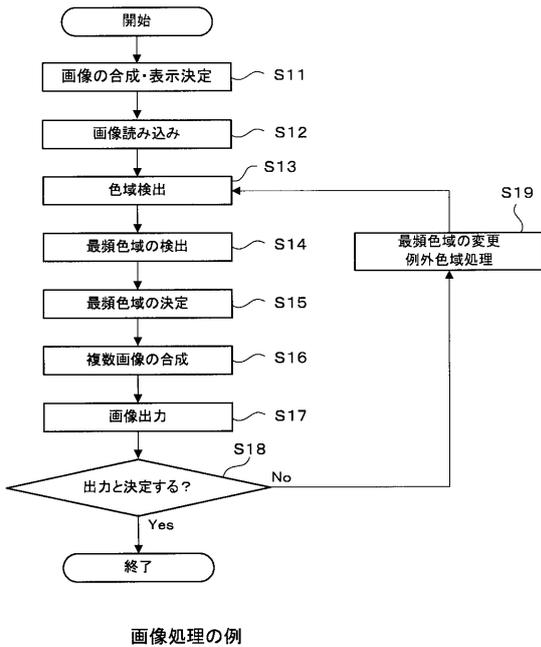
【図3】



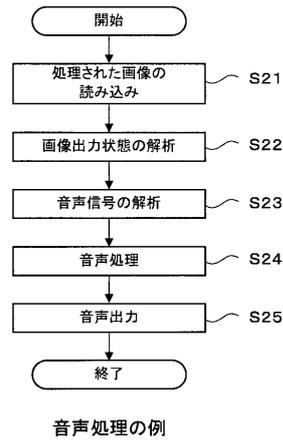
【図4】



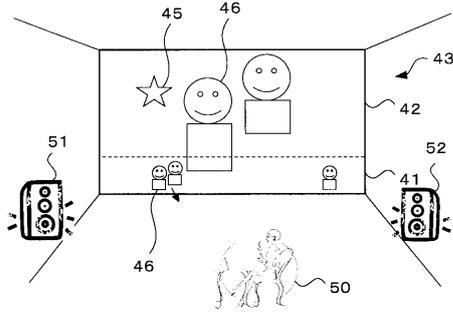
【図5】



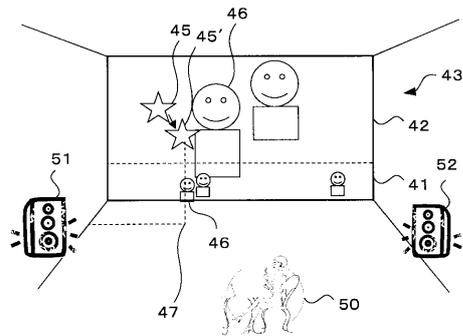
【図6】



【図17】



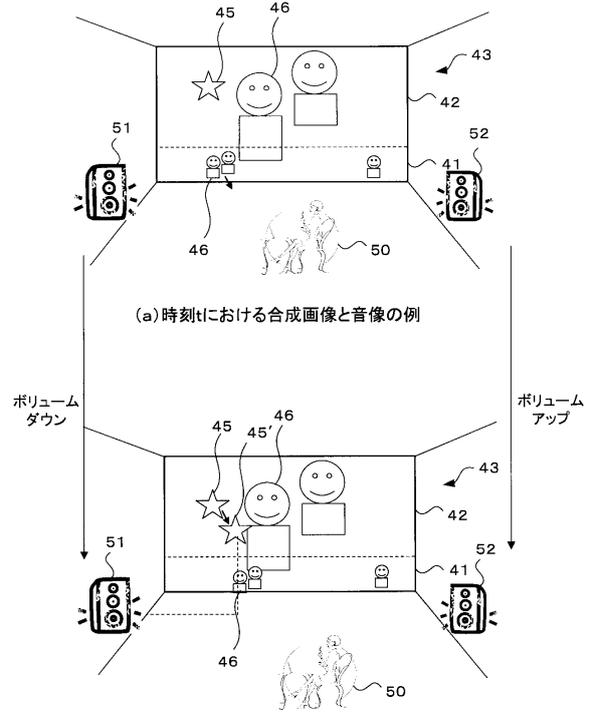
(a)時刻tにおける合成画像と音像の例



(b)時刻t+1における合成画像と音像の例

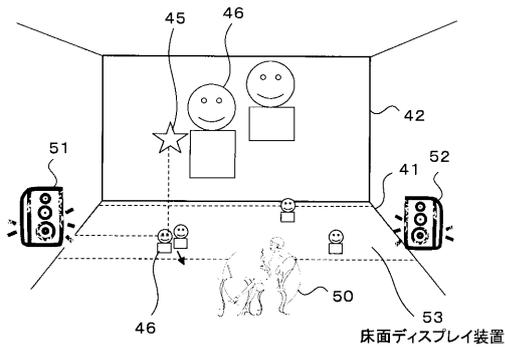
時間経過によって変化するズーム画像の音声の定位位置の例

【図18】



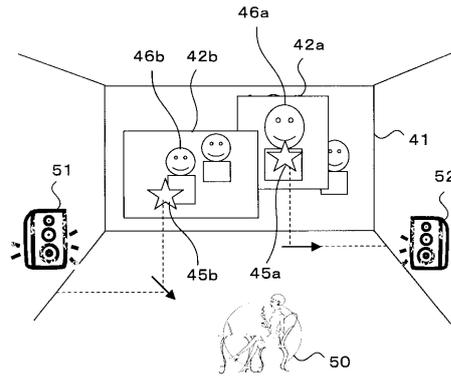
時間経過によって変化するパン画像の音声の定位位置の例

【図19】



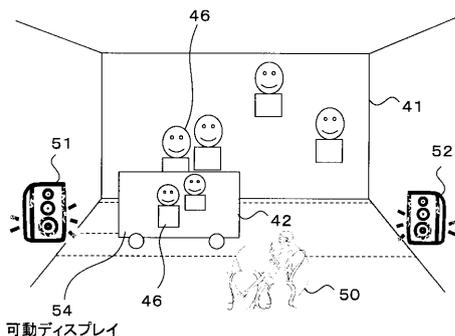
複数の画面に提示される例

【図21】



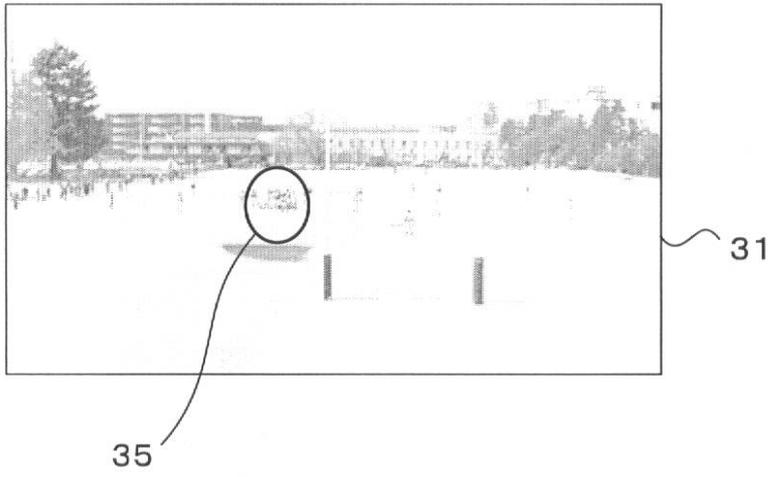
3つ以上の画像を同時に提示する例

【図20】



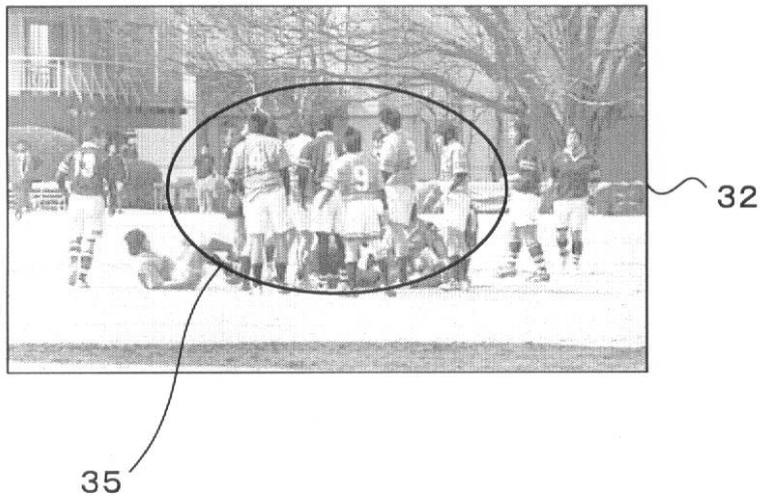
可動ディスプレイ装置を用いた画像提示の例

【図7】



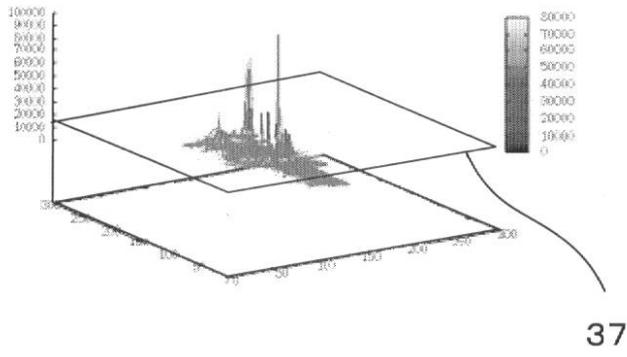
第1の画像の例

【図8】



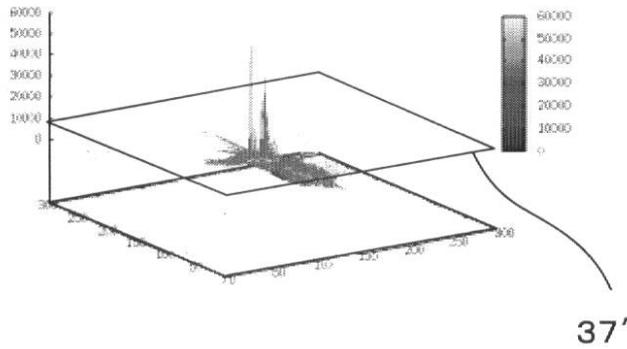
第2の画像の例

【図9】



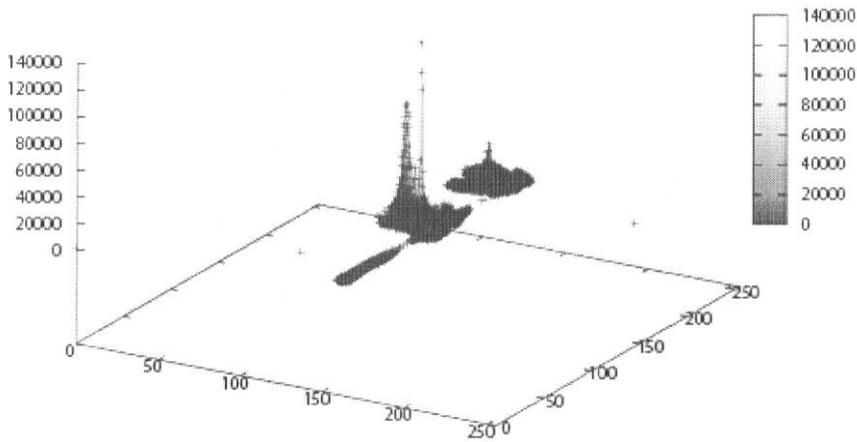
第1の画像の色域の例

【図10】



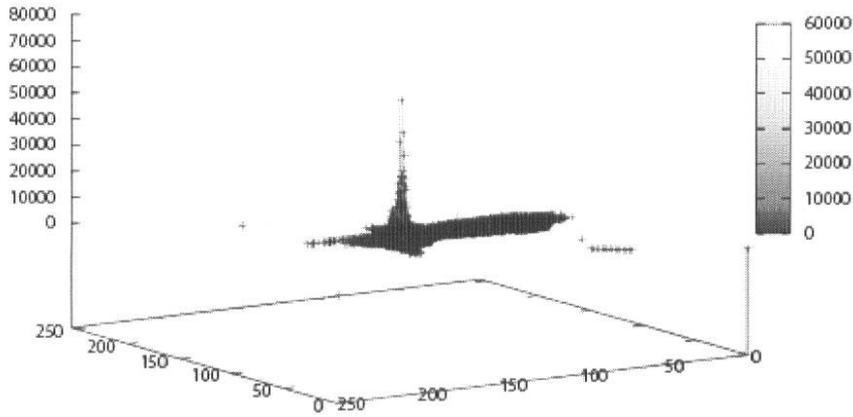
第2の画像の色域の例

【図11】



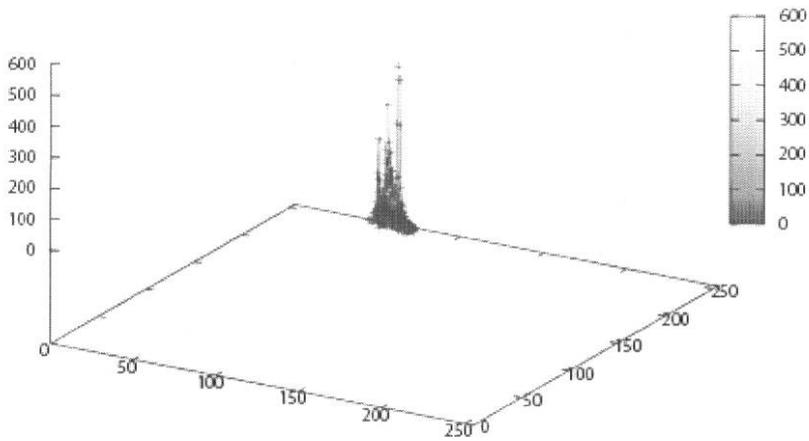
閾値を超える第1の画像の色域の例

【図12】



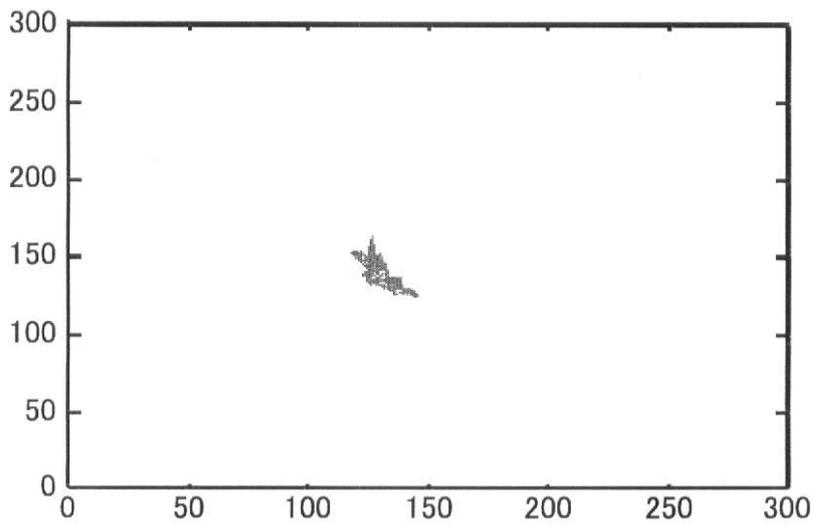
閾値を超える第2の画像の色域の例

【図13】



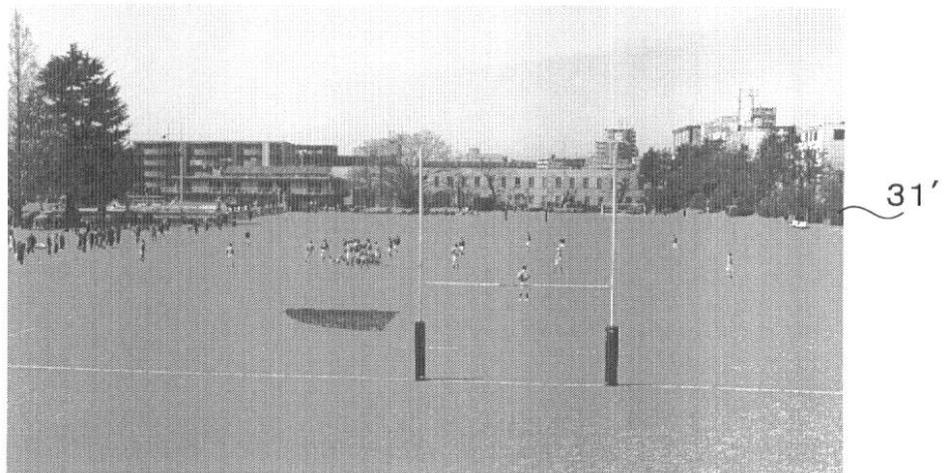
閾値を超える共通部分の色域の例

【図14】



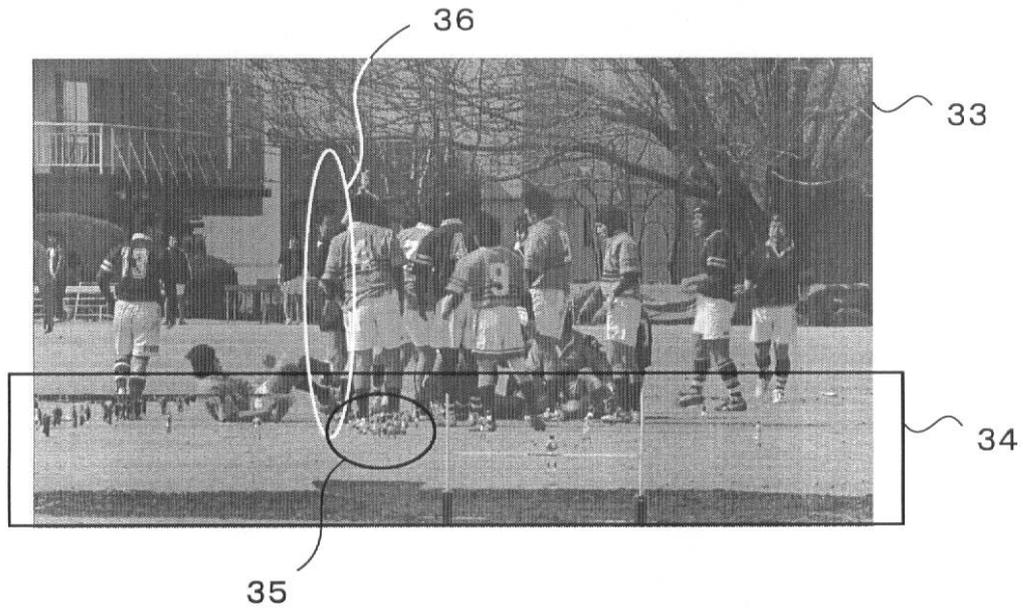
共通部分の色域の例

【図15】



第1及び第2の画像から所定の色域(グラウンドの色)を除いた例

【図16】



(a) 共通色域を除去した場合の合成画像の例



(b) 共通色域を除去しない場合の合成画像の例

合成画像の例

## フロントページの続き

- (72)発明者 田中 健司  
東京都港区港南1丁目7番1号 ソニー株式会社内
- (72)発明者 向井 仁志  
東京都港区港南1丁目7番1号 ソニー株式会社内
- (72)発明者 日比 啓文  
東京都港区港南1丁目7番1号 ソニー株式会社内
- (72)発明者 田中 和政  
東京都港区港南1丁目7番1号 ソニー株式会社内
- (72)発明者 江間 拓郎  
東京都港区港南1丁目7番1号 ソニー株式会社内
- (72)発明者 森崎 裕之  
東京都港区港南1丁目7番1号 ソニー株式会社内

審査官 菊池 充

- (56)参考文献 特開2005-295181(JP,A)  
特開2005-165684(JP,A)  
特開2004-173300(JP,A)  
特開平06-284330(JP,A)  
特開2004-328052(JP,A)

## (58)調査した分野(Int.Cl., DB名)

H04S 1/00 - 7/00  
H04N 5/222 - 5/257