

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6727106号
(P6727106)

(45) 発行日 令和2年7月22日 (2020.7.22)

(24) 登録日 令和2年7月2日 (2020.7.2)

| | | |
|---------------|---------------|------------------|
| (51) Int.Cl. | | F I |
| HO 4 N | 5/93 | (2006.01) |
| HO 4 N | 21/233 | (2011.01) |
| HO 4 N | 5/765 | (2006.01) |

H O 4 N 5/93

H O 4 N 21/233

H O 4 N 5/765

請求項の数 20 (全 18 頁)

| | | | |
|-----------|------------------------------|-----------|-------------------|
| (21) 出願番号 | 特願2016-233499 (P2016-233499) | (73) 特許権者 | 000001007 |
| (22) 出願日 | 平成28年11月30日 (2016.11.30) | | キヤノン株式会社 |
| (65) 公開番号 | 特開2018-93312 (P2018-93312A) | | 東京都大田区下丸子3丁目30番2号 |
| (43) 公開日 | 平成30年6月14日 (2018.6.14) | (74) 代理人 | 100076428 |
| 審査請求日 | 令和1年5月28日 (2019.5.28) | | 弁理士 大塚 康德 |
| | | (74) 代理人 | 100115071 |
| | | | 弁理士 大塚 康弘 |
| | | (74) 代理人 | 100112508 |
| | | | 弁理士 高柳 司郎 |
| | | (74) 代理人 | 100116894 |
| | | | 弁理士 木村 秀二 |
| | | (74) 代理人 | 100130409 |
| | | | 弁理士 下山 治 |
| | | (74) 代理人 | 100134175 |
| | | | 弁理士 永川 行光 |

最終頁に続く

(54) 【発明の名称】 画像音響処理装置および画像音響処理方法、プログラム

(57) 【特許請求の範囲】

【請求項 1】

___ 1 以上のマイクロフォンによる収音に基づく収音データであって、複数の撮影装置によりそれぞれ異なる方向から撮影される領域内の音を含む収音データを取得する取得手段と

、
前記複数の撮影装置により取得される複数の撮影画像に基づいて生成される仮想視点画像であって、同一の時点において変化させた視点から前記領域を見た場合の光景を表す動画像である仮想視点画像と共に再生される音響データとして、前記時点以前の収音期間に対応する音響データを、前記取得手段により取得された収音データに基づいて生成する生成手段とを有することを特徴とする音響処理装置。

【請求項 2】

前記収音期間は前記時点を含むことを特徴とする請求項 1 に記載の音響処理装置。

【請求項 3】

前記収音期間の開始時点は前記時点より前であり、

前記収音期間の終了時点は前記時点と一致することを特徴とする請求項 1 又は 2 に記載の音響処理装置。

【請求項 4】

前記収音期間の長さは前記仮想視点画像の再生期間の長さと一致することを特徴とする請求項 3 に記載の音響処理装置。

【請求項 5】

前記仮想視点画像は、前記時点において撮影された複数の撮影画像に基づいて生成されることを特徴とする請求項 1 乃至 4 の何れか 1 項に記載の音響処理装置。

【請求項 6】

前記仮想視点画像は、撮影時刻の進行が停止した状態で視点が変化する動画像であることを特徴とする請求項 1 乃至 5 の何れか 1 項に記載の音響処理装置。

【請求項 7】

前記音響データは、前記仮想視点画像における視点の変化に対応して仮想聴取点が変わる音のデータであることを特徴とする請求項 1 乃至 6 の何れか 1 項に記載の音響処理装置。

【請求項 8】

前記時点と前記仮想視点画像の再生期間の長さとに基づいて前記收音期間を決定する決定手段を有することを特徴とする請求項 1 乃至 7 の何れか 1 項に記載の音響処理装置。

【請求項 9】

前記仮想視点画像を取得する画像取得手段と、

前記生成手段により生成された前記音響データを、前記画像取得手段により取得された前記仮想視点画像と共に出力する出力手段とを有することを特徴とする請求項 1 乃至 8 の何れか 1 項に記載の音響処理装置。

【請求項 10】

前記生成手段により生成された音響データを含む音響コンテンツを、前記音響コンテンツと共に再生される画像コンテンツであって前記仮想視点画像を含む画像コンテンツと共に出力する出力手段を有することを特徴とする請求項 1 乃至 8 の何れか 1 項に記載の音響処理装置。

【請求項 11】

前記画像コンテンツは、前記仮想視点画像と、終了時点が前記時点に一致する撮影期間に対応する動画像とを含み、

前記音響コンテンツは、前記仮想視点画像と共に再生される第 1 の音響データと、前記撮影期間に対応する動画像と共に再生される第 2 の音響データとを含むことを特徴とする請求項 10 に記載の音響処理装置。

【請求項 12】

前記第 1 の音響データに対応する收音期間と前記第 2 の音響データに対応する收音期間とは不連続であることを特徴とする請求項 11 に記載の音響処理装置。

【請求項 13】

前記画像コンテンツは、前記仮想視点画像と、開始時点が前記時点に一致する撮影期間に対応する動画像とを含み、

前記音響コンテンツは、前記仮想視点画像と共に再生される第 1 の音響データと、前記撮影期間に対応する動画像と共に再生される第 2 の音響データとを含むことを特徴とする請求項 10 に記載の音響処理装置。

【請求項 14】

前記第 1 の音響データに対応する收音期間と前記第 2 の音響データに対応する收音期間とは連続することを特徴とする請求項 13 に記載の音響処理装置。

【請求項 15】

前記撮影期間に対応する動画像は、撮影装置により撮影された画像であることを特徴とする請求項 11 乃至 14 の何れか 1 項に記載の音響処理装置。

【請求項 16】

前記撮影期間に対応する動画像は、前記撮影期間内の複数の時点に撮影された複数の撮影画像に基づいて生成された仮想視点画像であることを特徴とする請求項 11 乃至 14 の何れか 1 項に記載の音響処理装置。

【請求項 17】

1 以上のマイクロフォンによる收音に基づく收音データであって、複数の撮影装置によりそれぞれ異なる方向から撮影される領域内の音を含む收音データを取得する取得工程と

10

20

30

40

50

、
前記複数の撮影装置により取得される複数の撮影画像に基づいて生成される仮想視点画像であって、同一の時点において変化させた視点から前記領域を見た場合の光景を表す動画像である仮想視点画像と共に再生される音響データとして、前記時点以前の收音期間に対応する音響データを、前記取得工程において取得された收音データに基づいて生成する生成工程とを有することを特徴とする音響処理方法。

【請求項 18】

前記收音期間の開始時点は前記時点より前であり、

前記收音期間の終了時点は前記時点と一致することを特徴とする請求項 17 に記載の音響処理方法。

10

【請求項 19】

前記時点と前記仮想視点画像の再生期間の長さに基づいて前記收音期間を決定する決定工程を有することを特徴とする請求項 17 又は 18 に記載の音響処理方法。

【請求項 20】

コンピュータを、請求項 1 乃至 16 の何れか 1 項に記載の音響処理装置の各手段として機能させるためのプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、画像コンテンツに対応した音響コンテンツを生成する画像音響処理装置、画像音響処理方法およびプログラムに関するものである。

20

【背景技術】

【0002】

複数のカメラを異なる位置に設置して多視点で同期撮影し、当該撮影により得られた複数視点画像を用いて仮想視点コンテンツを生成する技術が注目されている。複数視点画像から仮想視点コンテンツを生成する技術によれば、例えば、サッカーやバスケットボールのハイライトシーンを様々な角度から視聴することが出来るため、通常の画像よりもユーザに臨場感を与えることが出来る。また、画像と音響を同時に再生する場合、画像のフォーカスに応じて音量を変化させることが知られている（特許文献 1）。

【先行技術文献】

30

【特許文献】

【0003】

【特許文献 1】特開 2016 - 025633 号公報

【発明の概要】

【発明が解決しようとする課題】

【0004】

仮想視点コンテンツには、複数のカメラが 1 時刻に撮影した画像から生成する静止画仮想視点コンテンツと、複数のカメラが一定の期間に撮影した画像から生成する動画仮想視点コンテンツとがある。静止画仮想視点コンテンツと動画仮想視点コンテンツはどれも視聴時に仮想視点の切り替えを伴う多視点動画として取り扱われ、ユーザに高い臨場感を与える。

40

【0005】

しかしながら、従来の技術では、静止画仮想視点コンテンツの再生において、ユーザに音響の臨場感を与えることが出来ない。なぜなら、静止画仮想視点コンテンツは 1 時刻に撮影した画像から生成され、再生期間に対応する音響が存在しないからである。その結果、静止画仮想視点コンテンツの再生中に無音になる、或は、静止画仮想視点コンテンツの再生後に画像と音響が同期しなくなる、等が発生し、ユーザに臨場感を与えることが出来ないだけでなく、違和感を与えてしまう。

【0006】

本発明は、仮想視点における画像コンテンツに関連する、高臨場感の音響コンテンツを

50

提供できるようにすることを目的とする。

【課題を解決するための手段】

【0007】

上記の目的を達成するための本発明の一態様による画像音響処理装置は以下の構成を有する。すなわち、

1以上のマイクロフォンによる収音に基づく収音データであって、複数の撮影装置によりそれぞれ異なる方向から撮影される領域内の音を含む収音データを取得する取得手段と、

前記複数の撮影装置により取得される複数の撮影画像に基づいて生成される仮想視点画像であって、同一の時点において変化させた視点から前記領域を見た場合の光景を表す動画像である仮想視点画像と共に再生される音響データとして、前記時点以前の収音期間に対応する音響データを、前記取得手段により取得された収音データに基づいて生成する生成手段とを有する。

【発明の効果】

【0008】

本発明によれば、仮想視点における画像コンテンツに関連する、高臨場感の音響コンテンツを提供できるようにすることが出来る。

【図面の簡単な説明】

【0009】

【図1A】第一実施形態に係る映像処理システムの機器構成例を示したブロック図。

【図1B】画像音響処理装置のハードウェア構成例を示したブロック図。

【図2】第一実施形態に係る画像音響処理を示すフローチャート。

【図3】第一実施形態に係る音響の再利用の判断処理を示すフローチャート。

【図4】第一実施形態に係る時刻毎の音響と画像の相関図。

【図5】第二実施形態に係る競技場でのカメラ及びマイクロフォンの設置図。

【図6】第二実施形態に係る音響のデータ構成例を説明する図。

【図7】第二実施形態に係る音響の選択処理を示すフローチャート。

【図8】第三実施形態に係る、仮想サラウンドを作成する音響の選択処理を示すフローチャート。

【発明を実施するための形態】

【0010】

以下、本発明の実施形態について添付の図面を参照して詳細に説明する。

【0011】

< 第一実施形態 >

第一実施形態では、音響コンテンツを生成する場合に、どの時刻の音響を利用するかと、どの処理を実施するか判断を行う動作について説明する。第一実施形態では、生成可能な画像コンテンツの種別を、静止画仮想視点コンテンツ、動画仮想視点コンテンツ、ライブ画像コンテンツの3つとする。ライブ画像コンテンツは、ライブ放送で使用する画像コンテンツであり、且つ、各時刻において、複数のカメラのうち、選択した1つのカメラ画像を使用する画像コンテンツである。動画仮想視点コンテンツは、複数のカメラにより取得された所定期間の多視点画像に基づいて生成される、任意視点からの動画である。静止画仮想視点コンテンツは、複数のカメラにより取得されたある時点の多視点画像に基づいて生成される、複数の任意視点からの静止画である。

【0012】

一方、本実施形態において、生成可能な音響コンテンツには、仮想サラウンドコンテンツ、ライブ音響コンテンツの2つの種別がある。仮想サラウンドコンテンツとは、画像コンテンツの被写体の位置やカメラが撮影する位置などの情報を用いて疑似サラウンド効果の処理を施した音響コンテンツである。ライブ音響コンテンツとは、ライブ放送で使用する音響コンテンツであり、たとえば選択した2つのマイクから集音された音響データに基づいて生成される2chの音響コンテンツである。

【 0 0 1 3 】

〔 機器構成と各ブロックの処理 〕

図 1 A は、第一実施形態による映像処理システムの機器構成例を示すブロック図である。映像処理システムは、画像音響処理装置 1 0 0、カメラ群 1 0 1、マイクロフォン群 1 0 2、時刻生成サーバ 1 0 3、映像配信機器 1 0 4 を備える。画像音響処理装置 1 0 0 は、カメラ群 1 0 1 からカメラの台数に応じた数の画像データと、マイクロフォン群 1 0 2 からマイクロフォンの台数に応じた数の音響データを受信する。

【 0 0 1 4 】

カメラ群 1 0 1 は、複数の方向から被写体を撮影するための複数のカメラで構成される。マイクロフォン群 1 0 2 は複数のマイクロフォンで構成される。時刻生成サーバ 1 0 3 は、時刻情報を生成し、カメラ群 1 0 1 及びマイクロフォン群 1 0 2 に送信する。カメラ群 1 0 1 及びマイクロフォン群 1 0 2 を構成するカメラ、マイクロフォンは、時刻生成サーバ 1 0 3 より時刻情報を受信し、画像及び音響に時刻情報を付与して画像音響処理装置 1 0 0 に送信する。なお、時刻生成サーバ 1 0 3 が時刻情報を画像音響処理装置 1 0 0 に送信し、画像音響処理装置 1 0 0 が時刻情報を画像および音響に付与するようにしても良い。

【 0 0 1 5 】

次に、画像音響処理装置 1 0 0 の構成について説明する。画像入力部 1 1 0 は、カメラ群 1 0 1 が撮影した画像を取得して、蓄積部 1 1 2 に伝送する。音響入力部 1 1 1 は、マイクロフォン群 1 0 2 が集音した音響を取得して、蓄積部 1 1 2 に伝送する。蓄積部 1 1 2 は、画像入力部 1 1 0 から取得した画像データ（カメラ群 1 0 1 の複数のカメラにより得られた画像データ）と音響入力部 1 1 1 が取得した音響データ（マイクロフォン群 1 0 2 の複数のマイクロフォンにより得られた音響データ）を蓄積する。蓄積部 1 1 2 に蓄積された画像と音響をそれぞれ画像データ 1 1 3、音響データ 1 1 4 と称する。映像生成制御部 1 1 8 は、画像制御部 1 1 5、同期制御部 1 1 6、音響制御部 1 1 7 に対して、映像コンテンツの生成に関わるユーザの指示や設定値を伝送する。なお、本明細書において、映像コンテンツとは、画像コンテンツと、この画像コンテンツとともに再生される音響コンテンツを含むコンテンツとする。

【 0 0 1 6 】

画像制御部 1 1 5 は、映像生成制御部 1 1 8 の指示に応じて画像データ 1 1 3 に基づいて生成する画像コンテンツの種別を決定する。静止画仮想コンテンツを生成する場合、画像制御部 1 1 5 は、蓄積部 1 1 2 から読み出した画像データ 1 1 3 を静止画仮想視点コンテンツ生成部 1 1 9 に伝送する。動画仮想視点コンテンツを生成する場合、画像制御部 1 1 5 は、蓄積部 1 1 2 から読み出した画像データ 1 1 3 を動画仮想視点コンテンツ生成部 1 2 0 に伝送する。ライブ画像コンテンツを生成する場合、画像制御部 1 1 5 は、蓄積部 1 1 2 から読み出した画像データ 1 1 3 をライブ画像コンテンツ生成部 1 2 1 に伝送する。また、画像制御部 1 1 5 は、上記した生成する画像コンテンツの種別や、読み出した画像データの時刻情報などを、同期制御部 1 1 6 に伝送する。

【 0 0 1 7 】

同期制御部 1 1 6 は、画像コンテンツの種別、画像データ 1 1 3 のタイムスタンプ、画像コンテンツ生成の開始時刻、画像データ 1 1 3 の被写体座標などを、画像制御部 1 1 5 または映像生成制御部 1 1 8 から取得し、音響制御部 1 1 7 に伝送する。音響制御部 1 1 7 は、映像生成制御部 1 1 8 からの指示と、同期制御部 1 1 6 から取得した情報に基づく音響制御部 1 1 7 内部での判断と、の何れか、または、両方に基づいて、生成する音響コンテンツの種別、使用する音響データの種類を決定する。仮想サラウンドコンテンツを生成する場合、音響制御部 1 1 7 は、蓄積部 1 1 2 から読み出した音響データ 1 1 4 を仮想サラウンドコンテンツ生成部 1 2 2 に伝送する。ライブ音響コンテンツを生成する場合、音響制御部 1 1 7 は、蓄積部 1 1 2 から読み出した音響データ 1 1 4 をライブ音響コンテンツ生成部 1 2 3 に伝送する。

【 0 0 1 8 】

10

20

30

40

50

静止画仮想視点コンテンツ生成部 119 は、画像制御部 115 から取得した画像データ 113 を用いて静止画仮想視点コンテンツを生成し、多重部 124 に伝送する。動画仮想視点コンテンツ生成部 120 は、画像制御部 115 から取得した画像データ 113 を用いて動画仮想視点コンテンツを生成し、多重部 124 に伝送する。ライブ画像コンテンツ生成部 121 は、画像制御部 115 から取得した画像データ 113 を用いてライブ画像コンテンツを生成し、多重部 124 に伝送する。仮想サウンドコンテンツ生成部 122 は、音響制御部 117 から取得した音響データ 114 を用いて仮想サウンドコンテンツを生成し、多重部 124 に伝送する。ライブ音響コンテンツ生成部 123 は、音響制御部 117 から取得した音響データ 114 を用いてライブ音響コンテンツを生成し、多重部 124 に伝送する。

10

【0019】

多重部 124 と映像出力部 125 は、音響コンテンツを画像コンテンツとともに再生させるための制御を実行する。多重部 124 は、静止画仮想視点コンテンツ生成部 119、動画仮想視点コンテンツ生成部 120、ライブ画像コンテンツ生成部 121 から画像コンテンツを取得する。また、多重部 124 は、仮想サウンドコンテンツ生成部 122、ライブ音響コンテンツ生成部 123 から音響コンテンツを取得する。そして、多重部 124 は、画像コンテンツと音響コンテンツを 1 つのデータ列に多重化して映像コンテンツを生成し、映像出力部 125 に出力する。映像出力部 125 は、多重部 124 から取得した映像コンテンツを映像配信機器 104 に出力する。

【0020】

20

図 1B は、第一実施形態による画像音響処理装置 100 のハードウェア構成例を示す図である。CPU 11 は、読み出し専用メモリである ROM 12 または随時読み書きが可能な RAM 13 に格納されたプログラムを実行することにより、上述した各機能部を実現する。インターフェース 14 は、カメラ群 101、マイクロフォン群 102、映像配信機器 104 等の外部機器と画像音響処理装置 100 を接続する。入力部 15 は、ユーザによる各種入力を受け付ける。表示部 16 は、たとえば液晶表示器であり、CPU 11 の制御下で種々の表示を行う。記憶装置 17 は、たとえばハードディスクドライブで構成され、蓄積部 112 を実現する。また、記憶装置 17 は種々のプログラムを格納し、必要なプログラムが、CPU 11 による実行のために、適宜に RAM 13 に展開される。上述の各構成は、内部バス 18 を介して相互に通信可能に接続されている。

30

【0021】

〔動作〕

以上のような構成を備えた画像音響処理装置 100 の動作について説明する。図 2 は、第一実施形態に係る、画像音響処理を説明するフローチャートである。本実施形態では、音響データのうちの、画像データに基づいて生成される仮想視点の画像コンテンツに対応する撮影時刻よりも前の時刻に集音された音響データであるところの以前の音響データを用いて、画像コンテンツと共に再生される音響コンテンツが生成される。なお、本実施形態において、以前の音響データを用いて生成された音響コンテンツが関連付けられる仮想視点の画像コンテンツとは、静止画仮想視点コンテンツである。

【0022】

40

ステップ S20 において、映像生成制御部 118 は音響制御部 117 に対して再利用モードの設定を行う。本実施形態では、設定が可能な再利用モードは、ユーザ指示モード、タイムスタンプ自動判断モード、コンテンツ自動判断モードの 3 種類である。なお、再利用モードは、所定のユーザインターフェース（入力部 15）を介してユーザにより指定される。再利用モードは、任意のタイミングで設定し直すことができるようにしてもよい。各モードの具体的な動作は、図 3 を用いて後述する。また、本モードは全ての音響データ 114 を対象としているため、必ずしも、再利用するデータであるとは限らない。つまり、過去に利用されていない音響データが利用される場合もありうる。

【0023】

ステップ S21 において、映像生成制御部 118 は画像制御部 115 に対して画像コン

50

テンツ設定を行う。画像コンテンツ設定には、生成するコンテンツの種別や、生成を開始する時刻、生成を終了する時刻、及びカメラ群 101 の設定値などが含まれる。なお、画像コンテンツ設定は、同期制御部 116 にも伝えられる。ステップ S 22 において、映像生成制御部 118 は音響制御部 117 に対して音響コンテンツ設定を行う。音響コンテンツ設定には、生成するコンテンツのチャンネル数、マイクロフォン群 102 の設定値などが含まれる。

【0024】

ステップ S 23 において、映像生成制御部 118 はカメラ群 101、マイクロフォン群 102、及び画像音響処理装置 100 の各ブロックに対して撮影及び集音の開始を指示する。この指示により、画像入力部 110 がカメラ群 101 から受信した画像を蓄積部 112 に蓄積することと、音響入力部 111 がマイクロフォン群 102 から受信した音響を蓄積部 112 に蓄積することとが開始される。

【0025】

ステップ S 211 ~ S 215 は、蓄積部 112 に蓄積されている画像データ 113 や音響データ 114 を用いて画像コンテンツと音響コンテンツを含む映像コンテンツを生成するための編集処理である。映像編集の開始が指示されると、映像生成制御部 118 は映像編集処理を開始する（ステップ S 211、YES）。映像編集処理は、画像制御部 115 による画像コンテンツの生成（ステップ S 212）、音響制御部 117 による音響コンテンツの生成（ステップ S 213）、多重部 124 による映像コンテンツの生成（ステップ S 214）を含む。画像コンテンツの生成においては、画像制御部 115 が、画像コンテンツの生成に使用する画像データを蓄積部 112 から読み出し、ユーザが指定した画像コンテンツの種別に応じて、119 ~ 121 の何れかの生成部へ伝送する。音響コンテンツの生成においては、音響制御部 117 が、画像コンテンツの生成で用いられた画像データのタイムスタンプに基づいて特定された音響データを蓄積部 112 から読み出し、122 ~ 123 の何れかの生成部へ伝送する。映像コンテンツの生成においては、多重部 124 が、各コンテンツ生成部から受け取った画像コンテンツと音響コンテンツを 1 つのデータ列に多重化して出力する。ユーザは、編集結果としての映像コンテンツを直ちに視聴できる。ユーザによる編集終了の指示があるまで、ステップ S 212 ~ S 214 の処理が繰り返される（ステップ S 215）。

【0026】

たとえば、図 4 において、ユーザは、タイムスタンプ t_n0 から画像データを再生させて画像コンテンツを生成していく。この間、画像データがライブ画像コンテンツ生成部 121 に伝送され、ライブ画像コンテンツが生成される。タイムスタンプ t_n1 （時刻 $T1$ ）の画像が再生されるタイミングにおいて、ユーザは一時停止して静止画表示の状態とし、更に視点を移動させる。この間、必要な画像データが静止画仮想視点コンテンツ生成部 119 へ伝送され、静止画仮想視点コンテンツが生成される。その後、動画再生を再開することにより（時刻 $T2$ ）、タイムスタンプ t_n1 以降の画像が再生され、ライブ画像コンテンツが生成される。タイムスタンプ t_n2 の画像が再生されるタイミング（時刻 $T3$ ）で映像編集を終えると、図 4 に示すような構成の画像コンテンツが生成されることになる。

【0027】

上述したような画像コンテンツの生成において、音響制御部 117 は、通常は画像のタイムスタンプに基づいて、音響コンテンツの生成に用いる音響を選択する。例えば、音響制御部 117 は、ライブ画像コンテンツ 44（図 4）に対応する音響データとしてライブ音響コンテンツ 40、41 を生成する。このとき、音響制御部 117 は、ライブ画像コンテンツ 44 の生成に用いられた画像データに対応する音響データをライブ音響コンテンツ生成部 123 へ送信することにより音響コンテンツを生成する。静止画仮想視点コンテンツの場合、タイムスタンプが 1 つになるため、音響コンテンツを生成するための音響が得られない。そこで、音響制御部 117 は、再利用指示を仮想サラウンドコンテンツ生成部 122 に送ることで、当該静止画仮想コンテンツの音響コンテンツを生成させる。

10

20

30

40

50

【 0 0 2 8 】

ステップ S 2 4 ~ S 2 6 は、音響制御部 1 1 7 がこのような再利用指示を仮想サラウンドコンテンツ生成部 1 2 2 へ出力するための処理である。ステップ S 2 4 ~ S 2 6 の処理はステップ S 2 1 1 ~ S 2 1 5 の処理と実質的に並列に実行が可能である。ステップ S 2 4 において、音響制御部 1 1 7 はステップ S 2 0 で映像生成制御部 1 1 8 が設定した再利用モード設定に従い、音響データを再利用するか否かを判断する。ステップ S 2 4 の詳細については図 3 のフローチャートの参照により後述する。再利用すると判断された場合、ステップ S 2 5 において、音響制御部 1 1 7 は同期制御部 1 1 6 から再利用に必要な、画像コンテンツの開始時刻と、画像コンテンツの再生期間を取得する。そして、ステップ S 2 6 において、音響制御部 1 1 7 は仮想サラウンドコンテンツ生成部 1 2 2 に対して、音響データ、画像コンテンツの開始時刻、画像コンテンツの再生期間を伝送する。本ステップで、音響制御部 1 1 7 は映像生成制御部 1 1 8 に対して、指示を完了したことを伝えても良い。ステップ S 2 4 において再利用しないと判断された場合、上述のステップ S 2 5 と S 2 6 はスキップされる。また、ステップ S 2 5 で取得された再生期間の間は音響データの再利用が実行されるので、ステップ S 2 6 では、この再生期間にわたって処理が待機状態となり、再生期間が経過すると処理はステップ S 2 4 へ戻る。

10

【 0 0 2 9 】

ステップ S 2 7 において、映像生成制御部 1 1 8 は、撮影及び集音を継続するか判断を行うため、ユーザの指示を確認する。撮影及び集音を継続する場合、処理はステップ S 2 4 に戻り、撮影及び集音を継続しない場合、処理はステップ S 2 8 に進む。ステップ S 2 8 において、映像生成制御部 1 1 8 は、撮影及び集音の終了処理を行う。

20

【 0 0 3 0 】

次に、ステップ S 2 4 で実行される再利用の判断について図 3 を用いて説明する。図 3 のフローチャートは、ステップ S 2 4 の詳細を示したものである。

【 0 0 3 1 】

ステップ S 3 0 において、音響制御部 1 1 7 は、設定されている再利用モードを確認する。再利用モードは、ユーザ指示に従って上述の再利用の実行を判断するユーザ指示モードと画像コンテンツに基づいて上述の再利用の実行を自動的に判断する自動判断モードがある。自動判断モードには、画像コンテンツの生成種別から音響制御部 1 1 7 が再利用の判断を実行するコンテンツ自動判断モードと、画像コンテンツのタイムスタンプ情報から音響制御部 1 1 7 が再利用の判断を実行するタイムスタンプ自動判断モードがある。ユーザ指示モードの場合、処理はステップ S 3 1 に進む。コンテンツ自動判断モードの場合、処理はステップ S 3 2 に進む。タイムスタンプ自動判断モードの場合、処理はステップ S 3 3 に進む。

30

【 0 0 3 2 】

ユーザ指示モードでは、ユーザからの指示に応じて音響データの再利用を実行するか否かを判定する。ステップ S 3 1 において、音響制御部 1 1 7 はユーザからの再利用の指示の有無を同期制御部 1 1 6 から取得して確認する。ユーザからの再利用の指示がある場合、処理はステップ S 3 6 に進み、ユーザからの再利用の指示がない場合、処理はステップ S 3 5 に進む。

40

【 0 0 3 3 】

コンテンツ自動判断モードでは、画像コンテンツが仮想視点からの静止画であると判定した場合、すなわち静止画仮想視点コンテンツであると判定した場合に、音響データの再利用を実行すると判断する。ステップ S 3 2 において、音響制御部 1 1 7 は、生成中の画像コンテンツの種別を同期制御部 1 1 6 から取得して確認する。画像コンテンツが静止画仮想視点コンテンツの場合、処理はステップ S 3 6 に進み、動画仮想視点コンテンツまたはライブ画像コンテンツの場合、処理はステップ S 3 5 に進む。

【 0 0 3 4 】

タイムスタンプ自動判断モードでは、画像コンテンツの生成に用いられている画像データのタイムスタンプと再生時刻に基づいて、音響データの再利用を実行するか否かを判定

50

する。ステップS33において、音響制御部117は画像コンテンツの作成に使用する画像データ113のタイムスタンプと、映像生成制御部118から取得した再生時刻とを比較する。不一致の場合、処理はステップS36に進み、音響制御部117は、音響データ114を再利用すると判断する。比較した結果が一致する場合、処理はステップS34に進む。なお、タイムスタンプと再生時刻は一般には同一の値にはならない。例えば、図4において、時刻 T_0 タイムスタンプ t_{n0} 、時刻 T_1 タイムスタンプ t_{n1} である。本実施形態では、画像コンテンツに対応する画像データの撮影時刻と画像コンテンツの再生時刻との差分が変化した場合に、再利用すると判断する。たとえば、時刻 T_0 とタイムスタンプ t_{n0} との差分を基準とする。時刻 $T_1 + T$ において、静止画仮想視点コンテンツに用いられる画像データのタイムスタンプは t_{n1} のままであり、画像データの撮影時刻と画像コンテンツの再生時刻との差分が変化する。ステップS33では、この変化が検出される。また、ステップS26での待機の終了（再生期間の経過）に応じて差分の変化の監視が再開される場合、その時点における再生時刻とタイムスタンプとの差分が基準となる。

【0035】

ステップS34において、音響制御部117は画像コンテンツの作成に使用する画像データ113のタイムスタンプを同期制御部116から取得する。そして、前の再生時刻の画像コンテンツの作成に使用した画像データのタイムスタンプと比較して時刻が連続しているかを判断する。時刻が連続している場合、処理はステップS35に進み、音響制御部117は、音響データ114を再利用しないと判断する。時刻が連続していない場合、処理はステップS36に進み、音響制御部117は、音響データ114を再利用すると判断する。

【0036】

[処理結果の例]

次に、音響データ114を再利用すると判断し、静止画仮想視点コンテンツを閲覧中に仮想サラウンドコンテンツを視聴した場合のデータ推移を、図4を用いて説明する。図4は、時刻 T_0 から時刻 T_3 の間に音響コンテンツ及び画像コンテンツを再生する場合を示している。

【0037】

時刻 T_0 から時刻 T_1 にかけて、ライブ画像コンテンツ生成部121はライブ画像コンテンツ44を生成し、ライブ音響コンテンツ生成部123はライブ音響コンテンツ40及びライブ音響コンテンツ41を生成する。これらのコンテンツは多重部124において多重化され、映像コンテンツとして映像出力部125から出力される。ライブ画像コンテンツ44を作成するために用いられている画像データのタイムスタンプを $t_{n0} \sim t_{n1}$ とする。 $T_0 \sim T_1$ の間、ライブ画像コンテンツを生成するために用いられた画像データのタイムスタンプと再生時刻との差が、再生開始時の時刻とタイムスタンプとの差 $t = T_0 - t_{n0}$ に維持される。そのため、タイムスタンプ自動判断モードにおいて、音響データを再利用しないと判断される。また、同期制御部116はライブ画像コンテンツを生成することを音響制御部117に通知する。したがって、コンテンツ自動判断モードの場合、音響制御部117はこの通知に基づいて音響データを再利用しないと判断する。

【0038】

時刻 T_1 から時刻 T_2 にかけて、静止画仮想視点コンテンツ生成部119は静止画仮想視点コンテンツ45を生成し映像出力部125から出力する。たとえば、時刻 T_1 においてユーザが所定の操作により静止画仮想視点コンテンツの生成を指示すると、その指示は映像生成制御部118から画像制御部115に通知される。その指示の通知を受けた画像制御部115は、静止画仮想視点コンテンツ生成部119による静止画仮想視点コンテンツの生成を開始させ、ライブ画像コンテンツ生成部121によるライブ画像の生成を停止させる。また、ライブ画像の生成の停止に伴って音響コンテンツの生成に用いる音響データが無くなるため、ライブ音響コンテンツ生成部123においても音響コンテンツの生成が停止する。

【 0 0 3 9 】

音響制御部 1 1 7 は、音響データの再利用をするか否かの判断（ステップ S 2 4）を繰り返している。再利用モードがコンテンツ自動判断の場合、時刻 T 1 においてコンテンツの生成種別が静止画仮想視点コンテンツになったことを検出し、音響データを再利用すると判断する（ステップ S 3 2（YES） ステップ S 3 6）。静止画仮想視点コンテンツになったことの検出は、たとえば、音響制御部 1 1 7 が同期制御部 1 1 6 からコンテンツ種別の通知を受けることによりなされる。或いは、静止画仮想視点コンテンツ生成部 1 1 9 が出力するコンテンツ（画像データ）のメタデータから音響制御部 1 1 7 がコンテンツ種別を判定するようにしてもよい。また、静止画仮想視点コンテンツに用いられる画像データのタイムスタンプの進行が停止し、不連続となる（図 4 では、 t_{n1} に維持される）ため、再生時刻と画像データのタイムスタンプとの差が、上述の t を維持できなくなる。再利用モードがタイムスタンプ自動判断の場合は、これらの現象（ t を維持できない（ステップ S 3 3）、または、タイムスタンプの進行が停止した（不連続になった）こと（ステップ S 3 4））を検出することで音響データを再利用すると判断する。

10

【 0 0 4 0 】

音響データを再利用すると判断した場合、音響制御部 1 1 7 は仮想サラウンドコンテンツ生成部 1 2 2 に仮想サラウンドコンテンツ生成の開始を指示するとともに、開始時刻 T 1 と再生期間（ $T = T_2 - T_1$ ）を伝える。また、音響制御部 1 1 7 は、再生時刻 T 1 - $T = 2T_1 - T_2$ から T 1 に対応する音響データ（タイムスタンプが $t_{n1} - T$ から t_{n1} の音響データ）を蓄積部 1 1 2 から読み出し、仮想サラウンドコンテンツ生成部 1 2 2 に伝送する。仮想サラウンドコンテンツ生成部 1 2 2 は、時刻 $2T_1 - T_2$ から時刻 T 1 に対応する音響データを用いて、時刻 T 1 から時刻 T 2 の間に視聴する音響コンテンツ（仮想サラウンドコンテンツ）を生成し、映像出力部 1 2 5 に出力する。なお、本実施形態では、静止画仮想視点コンテンツの仮想視点位置に基づいて仮想サラウンドコンテンツを生成するための音響データが選択される。たとえば、静止画仮想視点コンテンツの生成開始時（時刻 T 1）において用いられている画像データを提供するカメラ（視点）に近いマイクロフォンから集音された音響データが仮想サラウンドコンテンツの生成に用いられる。もちろん、使用される音響データの選択は、これに限られるものではない。たとえば、再生時刻が $2T_1 - T_2$ から T 1 までの間の音響コンテンツの生成に用いられた音響データが用いられるようにしてもよい。

20

30

【 0 0 4 1 】

静止画仮想視点コンテンツの挿入を終えると（時刻 T 2 に到達すると）、画像制御部 1 1 5 は、静止画仮想視点コンテンツ生成部 1 1 9 によるコンテンツの生成を停止させ、ライブ画像コンテンツ生成部 1 2 1 によるライブ画像コンテンツの生成を再開させる。図 4 の例では、静止画仮想視点コンテンツの生成を開始した時刻 T 1 において中断されたライブ画像コンテンツの生成が再開されるようにする。すなわち、タイムスタンプ t_{n1} 以降の画像データを用いてライブ画像コンテンツが生成される。ただし、静止画仮想視点コンテンツにおいて視点が移動した場合には、視点の最終位置に近いカメラからの画像データを用いる。こうすることにより、画像の連続性が維持され、自然な再生画像となる。以上のように、静止画仮想視点コンテンツから他のコンテンツ（図 4 ではライブ画像コンテンツ）に切り替わる時に、連続した音響を視聴することが出来る。そのため、ユーザは違和感なく、臨場感のある音響を視聴することができる。

40

【 0 0 4 2 】

時刻 T 2 から時刻 T 3 にかけて、ライブ画像コンテンツ生成部 1 2 1 はライブ画像コンテンツ 4 6 を生成し、映像出力部 1 2 5 へ出力する。また、これに伴って、ライブ音響コンテンツ生成部 1 2 3 はライブ音響コンテンツ 4 3 を生成し、映像出力部 1 2 5 へ出力する。

【 0 0 4 3 】

なお、本実施形態において、時刻 T 1 において静止画仮想視点コンテンツの生成が開始されるが、その再生期間である T は、予め設定されているものとする。また、再生期間

50

を静止画仮想視点コンテンツの生成の指示とともに指定できるようにしてもよい。たとえば、5秒間の静止画仮想視点コンテンツを生成するボタン、10秒間の静止画仮想視点コンテンツを生成するボタンなどを設けておき、いずれのボタンが指示されたかにより再生期間 T が決定されるようにしてもよい。或いは、静止画の仮想視点を被写体に対して360度回転させることが可能な構成において、360度の回転に要する時間を予め設定しておき、指定された回転量に応じて再生期間が設定されるようにしてもよい。この場合、たとえば、静止画仮想視点コンテンツの生成とともに被写体に対して180度回転させることが指示されると、360度の回転に要する時間の半分が静止画仮想視点コンテンツの再生期間として設定される。

【0044】

また、上記実施形態において、音響データの再利用において、静止画仮想視点コンテンツの直前の T の期間の音響データが用いられたがこれに限られるものではない。画像データ113とともに蓄積部112に蓄積された音響データ114のうち、静止画仮想視点コンテンツのタイムスタンプ以前の任意のタイミングの音響データを用いるようにしてもよい。すなわち、静止画仮想視点コンテンツに対応する撮影時刻よりも前の時刻に集音された音響データを用いて静止画仮想視点コンテンツとともに再生される音響コンテンツを生成するようにすればよい。但し、図4のような画像コンテンツを生成した場合には時刻 $2T1 - T2 \sim$ 時刻 $T1$ の音響データを利用することが好ましい。時刻 $T2$ において良好な音響の連続性が得られるからである。

【0045】

以上のように、第1実施形態によれば、画像コンテンツの生成に使用した画像データの撮影時刻と再生時刻とに基づいて、当該撮影時刻よりも以前の音響データが選択され、当該画像コンテンツのための音響コンテンツが生成される。したがって、静止画仮想視点コンテンツのように画像に対応する音響データが存在しない場合でも、ユーザは音響を視聴することが出来る。また、その音響の生成に用いられる音響データは、映像コンテンツの生成に用いられた画像データに関連した音響データであり、ユーザは違和感なく視聴を続けることができる。特に、図4に示したように、静止画仮想視点コンテンツの再生期間である時刻 $T1 \sim T2$ に対応する音響コンテンツを、時刻 $2T2 - T1 \sim T1$ の期間に対応する音響データを用いて生成することにより、音響の連続性がより向上する。

【0046】

< 第二実施形態 >

第一実施形態では、静止画仮想視点コンテンツの生成時の音響コンテンツを生成するために、静止画仮想視点コンテンツの生成開始時における視点に近いマイクロフォンからの音響データを用いる構成を説明した。しかしながら、静止画仮想視点コンテンツの生成時における音響データの選択方法は、これに限られるものではない。第二の実施形態では、静止画仮想視点コンテンツに対応する仮想サラウンドコンテンツを生成する場合に、被写体の位置に基づいて音響データを選択する構成について説明する。なお、映像処理システムおよび画像音響処理装置100の構成は第一実施形態(図1A、図1B)と同様である。

【0047】

[競技場でのカメラ及びマイクロフォンの設置例]

図5は、競技場でのカメラ及びマイクロフォンの設置例である。競技場500は、スタンド501、フィールド502、実況室503を有している。また、フィールド502を点線で示す4つの領域に分割した場合の第一象限を分割エリア560、第二象限を分割エリア561、第三象限を分割エリア562、第四象限を分割エリア563と称する。マイクロフォン510～517はフィールド502の内部に設置されたマイクロフォンである。図5の例では、分割エリア560～563毎に2本ずつのマイクロフォンが設置されている。マイクロフォン520～523はスタンド501に設置されたマイクロフォンである。マイクロフォン530は実況室503に設置されたマイクロフォンである。マイクロフォン510～517、520～523は、マイクロフォン群102を構成している。カ

10

20

30

40

50

メラ 5 4 0 ~ 5 5 7 はスタンド 5 0 1 に設置されたカメラである。カメラ 5 4 0 ~ 5 5 7 はカメラ群 1 0 1 を構成している。

【 0 0 4 8 】

[音響データの構成]

図 6 は、蓄積部 1 1 2 に蓄積される音響データ 1 1 4 の内部構造を示す図である。音響データ 1 1 4 はマイクロフォン 5 1 0 ~ 5 1 7 が集音したフィールド音響 6 0 と、マイクロフォン 5 2 0 ~ 5 2 3 が集音したスタンド音響 6 1 と、マイクロフォン 5 3 0 が集音した実況音響 6 2 の複数のカテゴリを含む。フィールド音響 6 0 は競技場のフィールドで集音された音響データである。スタンド音響 6 1 は競技場のスタンドで集音された音響データである。実況音響 6 2 は実況者の音声を集音することにより得られた音響データである。音響制御部 1 1 7 は、動画コンテンツのタイムスタンプや視点に基づいて音響データ 1 1 4 から必要なフィールド音響 6 0、スタンド音響 6 1、実況音響 6 2 の少なくとも何れかを選択して蓄積部 1 1 2 から読み出す。音響制御部 1 1 7 は、仮想サラウンドコンテンツ生成部 1 2 2 とライブ音響コンテンツ生成部 1 2 3 のうち読み出した音響データの送信先を、生成中の動画コンテンツの種別に応じて決定する。音響制御部 1 1 7 は、読み出した音響データを決定されたコンテンツ生成部へ送る。

10

【 0 0 4 9 】

[動作]

音響制御部 1 1 7 が、静止画仮想視点コンテンツのための音響コンテンツの生成に用いる音響データを音響データ 1 1 4 から選択する処理について、図 7 のフローチャートを用いて説明する。図 7 のフローチャートにより示される処理は、図 2 のステップ S 2 5、S 2 6 の処理の詳細の一例である。

20

【 0 0 5 0 】

ステップ S 7 1 において、音響制御部 1 1 7 は同期制御部 1 1 6 から静止画仮想視点コンテンツの生成に用いられた画像データのタイムスタンプと静止画仮想視点コンテンツの再生期間（たとえば図 4 の T）に基づいてタイムスタンプ範囲情報を取得する。タイムスタンプ範囲情報は、音響データを取得するためのタイムスタンプの範囲であり、たとえば、図 4 の「T 1 - T 2」から「T 1」の範囲に対応するタイムスタンプの範囲（t n 1 - T ~ t n 1）である。ステップ S 7 2 において、音響制御部 1 1 7 は、同期制御部 1 1 6 から画像コンテンツの被写体が存在するエリア情報を取得する。エリア情報は 2 次元または 3 次元の座標でも良い。

30

【 0 0 5 1 】

ステップ S 7 3 において、音響制御部 1 1 7 は、ステップ S 7 1 で取得したタイムスタンプ範囲情報とステップ S 7 2 で取得したエリア情報に基づいて仮想サラウンドコンテンツ生成部 1 2 2 に伝送する音響データを決定する。例えば、ステップ S 7 2 で取得されたエリア情報に基づいて特定されたマイクロフォンから得られた音響データのうち、ステップ S 7 1 で取得されたタイムスタンプ範囲情報で示される範囲の音響データが、伝送する音響データとして決定される。ステップ S 7 4 において、音響制御部 1 1 7 は、音響データ 1 1 4 のうち、ステップ S 7 3 で決定した音響データを取得する。たとえば、ステップ S 7 1 で取得されたタイムスタンプ範囲情報が示す範囲のタイムスタンプを有し、ステップ S 7 2 で取得されたエリア情報が示すエリアに対応するマイクロフォンから集音された音響データが取得される。ステップ S 7 5 において、音響制御部 1 1 7 は、仮想サラウンドコンテンツ生成部 1 2 2 へステップ S 7 4 で取得した音響データを伝送する。

40

【 0 0 5 2 】

[音響データの選択の具体例]

以下、図 5 に示した競技場 5 0 0 の設置例と図 6 に示した音響データ 1 1 4 の構成例を用いて、音響制御部 1 1 7 による音響データの選択の処理（ステップ S 7 3）の例を示す。生成される画像コンテンツは静止画仮想視点コンテンツであり、被写体は分割エリア 5 6 1 に存在するものとする。静止画仮想視点コンテンツの場合には対応する音響データが無い場合、仮想サラウンドが生成される。

50

【 0 0 5 3 】

音響データの選択方法にはたとえば次の２種類があげられる。１つ目は、スタンド音響 6 1 のみを選択する方法であり、２つ目はフィールド音響 6 0 のみを選択する方法である。スタンド音響 6 1 のみを選択する方法では、被写体の存在する分割エリア 5 6 1 に対応するマイクロフォン 5 2 1 から得られた音響データのうち、タイムスタンプがタイムスタンプ範囲情報で示される範囲にある音響データが選択される。フィールド音響 6 0 のみを選択する方法では、分割エリア 5 6 1 に存在するマイクロフォン 5 2 1 以外のマイクロフォン 5 1 1、5 1 2 から得られた音響データのうち、タイムスタンプがタイムスタンプ範囲情報で示される範囲にある音響データが選択される。どちらの選択方法を採用するかはユーザ設定により決定されるものとする。

10

【 0 0 5 4 】

もちろん、音響データの選択方法は上記に限られるものではない。たとえば、画像コンテンツ（静止画仮想視点コンテンツ）の生成に用いた画像データの被写体の位置から一番近い位置にあるマイクロフォンで集音した音響データを除いて音響コンテンツを生成するようにしてもよい。なお、実況音声は過去の音響データを再利用すると不連続性が目立つので、本実施形態では、選択の対象としない。

【 0 0 5 5 】

以上、説明したように、第二実施形態によれば、静止画仮想視点コンテンツのための仮想サラウンドを生成する場合の音響データの選択方法をユーザ指定により選択することが出来る。そのため、ユーザは静止画仮想視点コンテンツを閲覧する時に、様々な音響を視聴でき、様々な臨場感を体験できる。

20

【 0 0 5 6 】

< 第三実施形態 >

第一実施形態では、仮想視点の位置に基づいて音響データして音響コンテンツを生成する構成を、第二実施形態では、静止画仮想視点コンテンツの被写体の位置に基づいて音響データを選択して音響コンテンツを生成する構成を示した。第三実施形態では、第一実施形態と同様に静止画仮想視点コンテンツの仮想視点に基づいて音響コンテンツを生成する構成を説明する。但し、第三実施形態では、仮想視点の位置を用いた音響コンテンツの生成方法として、複数の方法を切り替え可能な構成について説明する。特に生成方法の一つとして、仮想視点の移動に応じて音響が移動（たとえば回転）するように仮想サラウンドコンテンツを生成する方法が含まれる場合について説明する。なお、音響が移動するとは、仮想の集音位置が移動することである。映像処理システムおよび画像音響処理装置 1 0 0 の構成は第一実施形態（図 1 A、図 1 B）と同様である。

30

【 0 0 5 7 】

〔 仮想サラウンドの種類 〕

本実施形態では、仮想サラウンドの生成方法として３種類の生成方法を有する構成を説明する。１つ目の生成方法は、撮影座標基準による生成方法であり、被写体を撮影するカメラ、すなわち仮想視点の位置（以下、撮影座標）の移動に基づいて音響コンテンツの集音位置を移動（回転）する。撮影座標基準では複数の撮影座標を取得する必要がある。２つ目の生成方法は、終了基準による生成方法であり、静止画仮想視点コンテンツにおける最後の撮影座標に基づいて音響コンテンツを生成する。３つ目の生成方法は、開始基準による生成方法であり、静止画仮想視点コンテンツにおける最初の撮影座標に基づいて音響コンテンツを生成する。

40

【 0 0 5 8 】

なお、撮影座標基準では、仮想視点の位置の移動に伴って撮影座標を取得する必要がある。また、終了基準による生成方法では、静止画仮想視点コンテンツの最後の撮影座標を取得した後に仮想サラウンドコンテンツを生成することになる。但し、静止画仮想視点コンテンツの生成における仮想視点の移動経路が予め分かっている場合、すなわち、必要な複数の撮影座標あるいは最後の撮影座標が予め分かっている場合は、直ちに複数の撮影座標あるいは最後の撮影座標を取得することができる。

50

【 0 0 5 9 】

〔 動作 〕

図 8 は、第三実施形態による、静止画仮想視点コンテンツのための仮想サラウンドの生成に必要な音響データを蓄積部 1 1 2 に蓄積されている音響データ 1 1 4の中から選択する処理を示すフローチャートである。

【 0 0 6 0 】

ステップ S 8 0 において、音響制御部 1 1 7 は同期制御部 1 1 6 または映像生成制御部 1 1 8 から、ユーザにより指定された仮想サラウンドの生成方法を取得する。指定された生成方法が撮影座標基準の場合に処理はステップ S 8 1 に進み、終了座標基準の場合に処理はステップ S 8 3 に進み、開始座標基準の場合に処理はステップ S 8 2 に進む。ステップ S 8 1 では、音響制御部 1 1 7 が同期制御部 1 1 6 からカメラの撮影座標を取得する。ステップ S 8 2 では、音響制御部 1 1 7 が同期制御部 1 1 6 から静止画仮想視点コンテンツの開始時の撮影座標を取得する。ステップ S 8 3 では、音響制御部 1 1 7 が同期制御部 1 1 6 から仮想視点コンテンツの終了時の撮影座標を取得する。

10

【 0 0 6 1 】

ステップ S 8 1 ~ S 8 3 の後、処理はステップ S 8 4 に進む。ステップ S 8 4 において、音響制御部 1 1 7 はステップ S 8 1、ステップ S 8 2、またはステップ S 8 3 で取得した撮影座標に基づいて分割エリアを選択する。例えば、図 5 に示した分割エリア 5 6 0 ~ 5 6 3 のうち、撮影座標が存在している分割エリアが選択される。仮想サラウンドの作成に必要なマイクロフォンが存在する分割エリアを決定する。分割エリアを選択する代わりに、マイクロフォンを選択する構成を採用しても良い。

20

【 0 0 6 2 】

ステップ S 8 5 において、音響制御部 1 1 7 はステップ S 8 4 で選択した分割エリアからマイクロフォンを決定し、使用する音響データを決定する。ステップ S 7 4、S 7 5 の処理は、図 7 で説明したとおりである。撮影座標基準の場合、複数の撮影座標に対応する複数の音響データが仮想サラウンドコンテンツ生成部 1 2 2 に伝送され、仮想サラウンドコンテンツ生成部 1 2 2 は、仮想視点の移動に応じて移動する音響コンテンツを生成する。開始基準の場合、開始座標に対応する音響データが仮想サラウンドコンテンツ生成部 1 2 2 に伝送され、仮想サラウンドコンテンツ生成部 1 2 2 は、この音響データを加工して音響コンテンツを生成する。たとえば、時間の経過とともに音をぼかしていくような加工がなされる。終了基準の場合、終了座標に対応する音響データが仮想サラウンドコンテンツ生成部 1 2 2 に伝送され、仮想サラウンドコンテンツ生成部 1 2 2 は、この音響データを加工して音響コンテンツを生成する。たとえば、開始基準の場合とは逆に、時間の経過とともにぼかした音からシャープな音へ変化していくような加工がなされる。

30

【 0 0 6 3 】

以上、説明したように、本実施形態においては、仮想視点コンテンツに応じて仮想サラウンドを作成するための入力音響を決定し、仮想視点と同様に音声を回転することで、ユーザに高臨場感を与えることが出来る。また、複数の基準座標を設けることで、ユーザに様々な臨場感を選択できる余地を与えることが出来る。

【 0 0 6 4 】

なお、上述の各実施形態では、画像コンテンツの生成と音響コンテンツの生成を略並列に行う構成を示したがこれに限られるものではない。たとえば、図 4 において、T 1 ~ T 3 の画像コンテンツが予め生成された状況で、予め生成された画像コンテンツに対して音響コンテンツを後付けするように生成する処理であっても、上述した音響コンテンツの生成を適用可能である。また、再利用モードがユーザ指示モードの場合には、画像コンテンツの種別に関わらず音響コンテンツの再利用が実行される。たとえば、ライブ画像コンテンツの生成中にユーザから再利用指示があった場合には、その指示の直後からあらかじめ決められた期間（再生期間）にわたって以前の音響データを用いた音響コンテンツの生成が行われる。

40

【 0 0 6 5 】

50

(その他の実施例)

本発明は、上述の実施形態の１以上の機能を実現するプログラムを、ネットワーク又は記憶媒体を介してシステム又は装置に供給し、そのシステム又は装置のコンピュータにおける１つ以上のプロセッサがプログラムを読み出し実行する処理でも実現可能である。また、１以上の機能を実現する回路（例えば、ＡＳＩＣ）によっても実現可能である。

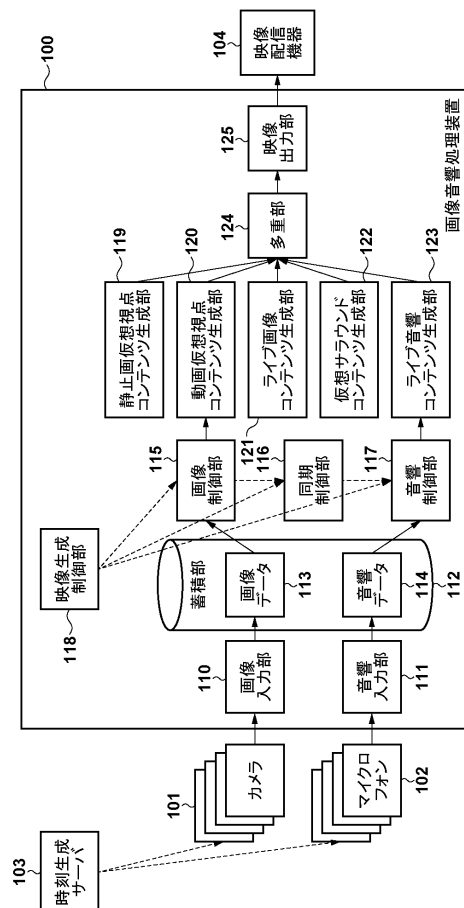
【符号の説明】

【００６６】

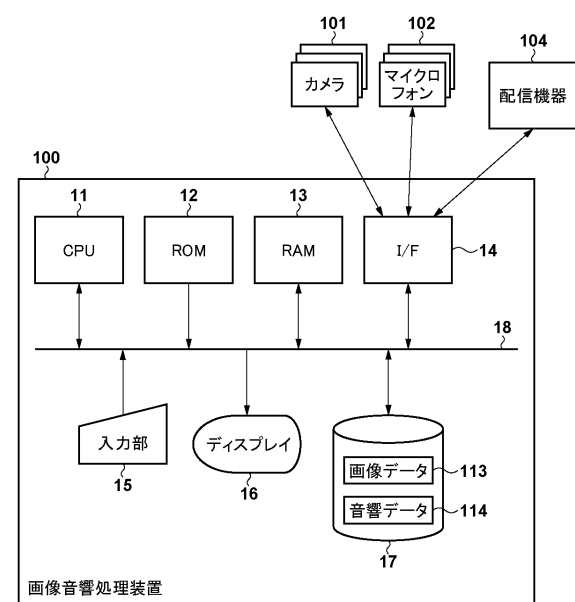
１００：画像音響処理装置、１１０：画像入力部、１１１：音響入力部、１１２：蓄積部、１１５：画像制御部、１１６：同期制御部、１１７：音響制御部、１１８：映像生成制御部、１１９：静止画仮想視点コンテンツ生成部、１２０：動画仮想視点コンテンツ生成部、１２１：ライブ画像コンテンツ生成部、１２２：仮想サラウンドコンテンツ生成部、１２３：ライブ音響コンテンツ生成部、１２４：多重部、１２５：映像出力部、１２６：音響出力部、１０１：カメラ、１０２：マイク、１０３：時刻生成サーバ、１０４：配信機器

10

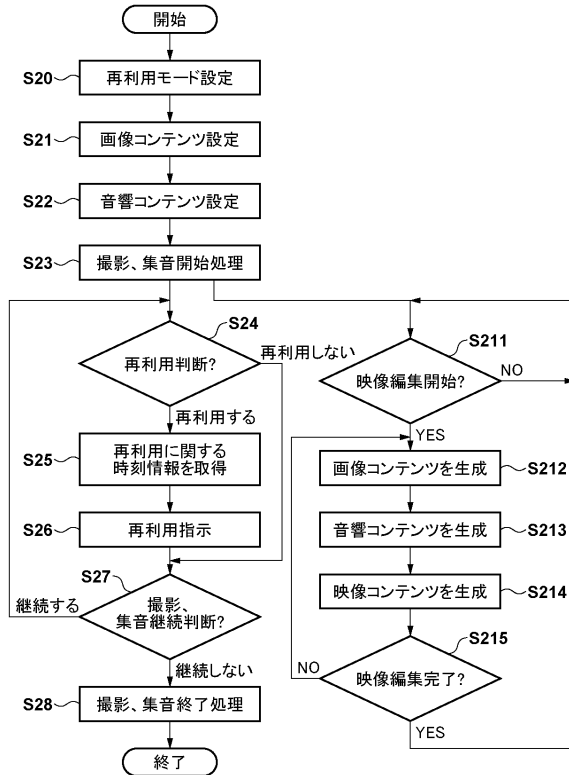
【図１Ａ】



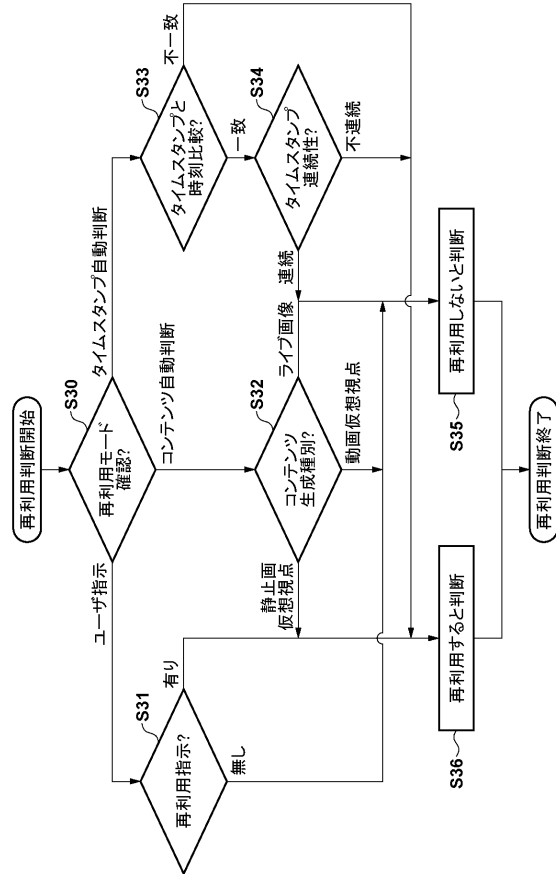
【図１Ｂ】



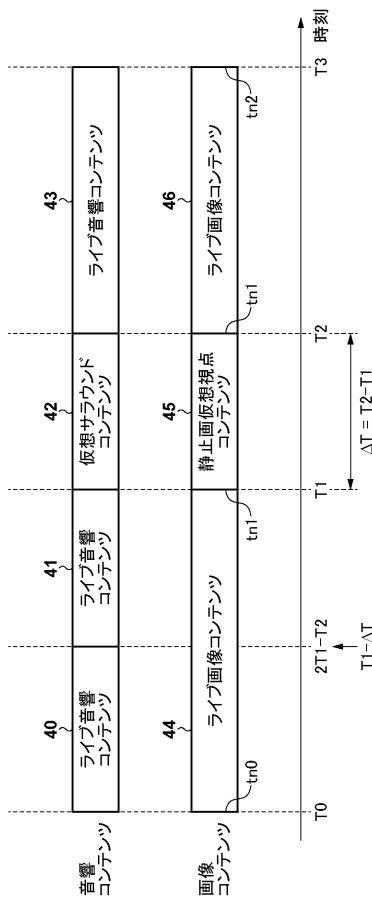
【図 2】



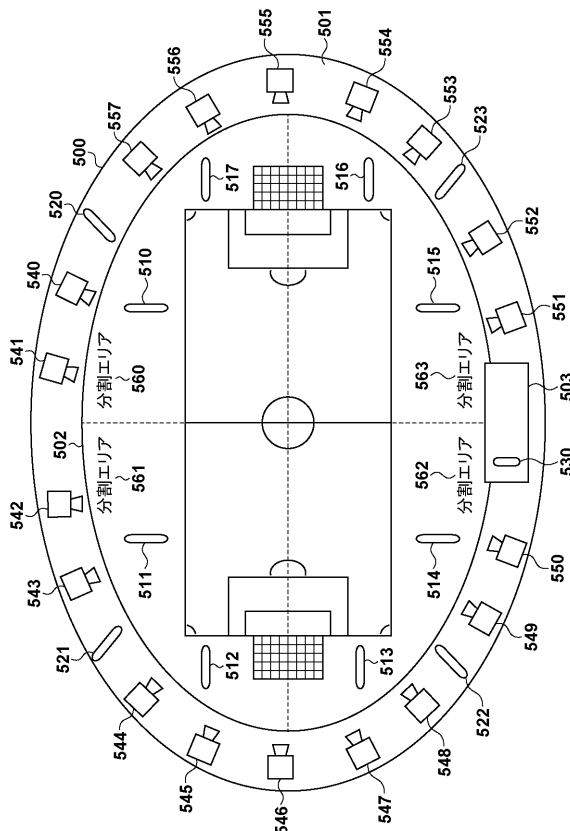
【図 3】



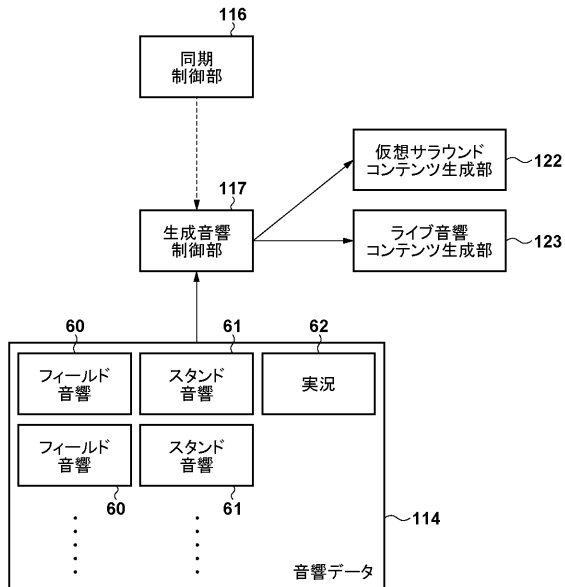
【図 4】



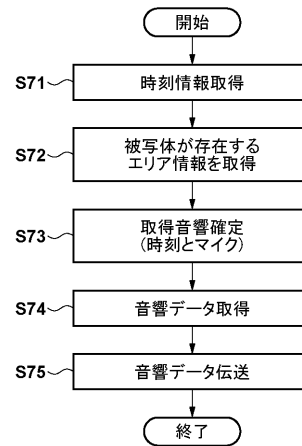
【図 5】



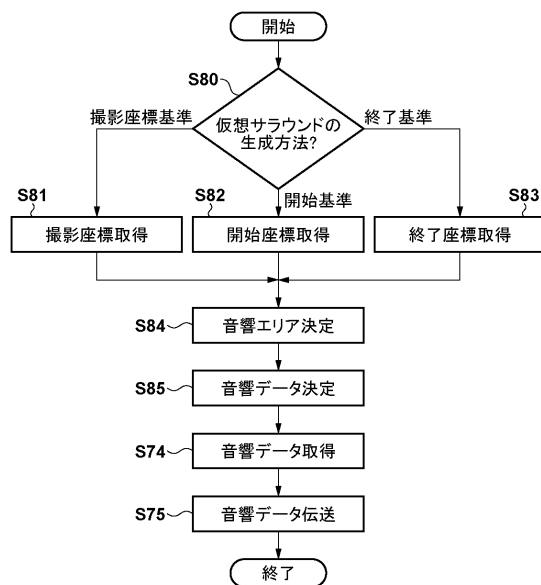
【図 6】



【図 7】



【図 8】



フロントページの続き

(72)発明者 田中 克昌
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

審査官 鈴木 順三

(56)参考文献 国際公開第2015/162947(WO, A1)
特開2013-135465(JP, A)

(58)調査した分野(Int.Cl., DB名)
H04N 5/76 - 5/956
H04N 21/00 - 21/858