

(19)日本国特許庁(JP)

(12)公表特許公報(A)

(11)公表番号
特表2024-513082
(P2024-513082A)

(43)公表日 令和6年3月21日(2024.3.21)

(51)国際特許分類	F I	テーマコード (参考)
G 0 6 T 19/00 (2011.01)	G 0 6 T 19/00 6 0 0	5 B 0 5 0
G 0 6 F 3/01 (2006.01)	G 0 6 T 19/00 C	5 E 5 5 5
	G 0 6 F 3/01 5 1 0	

審査請求 未請求 予備審査請求 未請求 (全30頁)

(21)出願番号	特願2023-561173(P2023-561173)	(71)出願人	590000248
(86)(22)出願日	令和4年3月29日(2022.3.29)		コーニンクレッカ フィリップス エヌ
(85)翻訳文提出日	令和5年10月4日(2023.10.4)		ヴェ
(86)国際出願番号	PCT/EP2022/058273		Koninklijke Philips
(87)国際公開番号	WO2022/214357		N. V.
(87)国際公開日	令和4年10月13日(2022.10.13)		オランダ国 5 6 5 6 アーヘー アイン
(31)優先権主張番号	21167514.5		ドーフエン ハイテック キャンパス 5 2
(32)優先日	令和3年4月8日(2021.4.8)		High Tech Campus 52,
(33)優先権主張国・地域又は機関	欧州特許庁(EP)		5 6 5 6 AG Eindhoven, N
			etherlands
(81)指定国・地域	AP(BW,GH,GM,KE,LR,LS,MW,MZ,NA	(74)代理人	110001690
	,RW,SD,SL,ST,SZ,TZ,UG,ZM,ZW),EA(		弁理士法人M&Sパートナーズ
	AM,AZ,BY,KG,KZ,RU,TJ,TM),EP(AL,A	(72)発明者	ヴァーエカンブ クリスティアン
	T,BE,BG,CH,CY,CZ,DE,DK,EE,ES,FI,FR		オランダ国 5 6 5 6 アーヘー アイン
	,GB,GR,HR,HU,IE,IS,IT,LT,LU,LV,MC,		ドーフエン ハイ テック キャンパス 5
	最終頁に続く		最終頁に続く

(54)【発明の名称】 オーディオ信号を生成する装置及び方法

(57)【要約】

装置は、画像センサから深度及び／又は視覚データのリアルタイム画像シーケンスを受信する第1の受信器201を含む。第2の受信器209は、オーディオオブジェクトのセット、及びオーディオオブジェクトと材料特性との関連性を示すメタデータを受信する。画像生成器203は、現実世界のシーンの仮想シーンオブジェクトに対応する画像オブジェクトを含む出力画像シーケンスを生成し、検出器205は、仮想シーンオブジェクトと現実世界のシーンの現実世界シーンオブジェクトとの相互作用を検出する。推定器207は、画像フレームのシーケンスの画像データに応じて現実世界シーンオブジェクトの材料特性を決定し、セクタ211は、材料特性、及びオーディオオブジェクトのセットに関連付けられた材料特徴に応じてオーディオオブジェクトを選択する。出力回路213は、第1のオーディオオブジェクトを含む出力オーディオ信号を生成する。

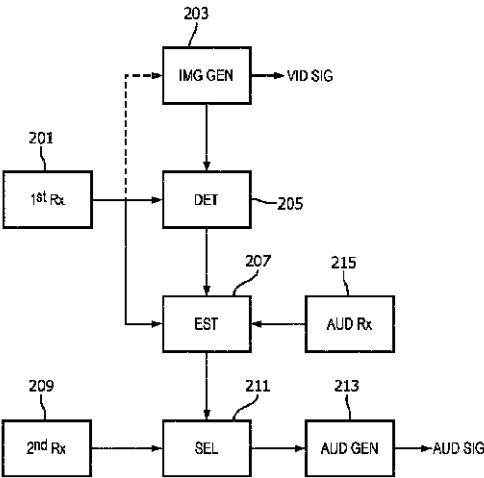


FIG. 2

【特許請求の範囲】

【請求項 1】

出力オーディオ信号を生成する装置であって、前記装置は、

画像センサから現実世界のシーンのリアルタイム画像シーケンスを受信する第 1 の受信器であって、前記リアルタイム画像シーケンスは画像フレームのシーケンスを含み、各画像フレームは視覚画像データ及び深度画像データのうちの少なくとも一方を含む、第 1 の受信器と、

オーディオオブジェクトのセット、及び前記オーディオオブジェクトの前記セットのオーディオオブジェクトのメタデータを受信する第 2 の受信器であって、前記メタデータは、前記オーディオオブジェクトの前記セットのオーディオオブジェクトと材料特徴との関連性を示す、第 2 の受信器と、

前記現実世界のシーンの仮想シーンオブジェクトに対応する画像オブジェクトを含む出力画像シーケンスを生成する画像生成器と、

前記仮想シーンオブジェクトと前記現実世界のシーンの現実世界シーンオブジェクトとの相互作用を、前記仮想シーンオブジェクトと前記現実世界シーンオブジェクトとの近接の検出に応じて検出する検出器と、

前記画像フレームの前記シーケンスの画像データに応じて、前記現実世界シーンオブジェクトの材料特性を決定する推定器と、

前記材料特性、及び前記オーディオオブジェクトのセットのオーディオオブジェクトに関連付けられた前記材料特徴に応じて、前記オーディオオブジェクトの前記セットの第 1 のオーディオオブジェクトを選択するセレクトと、

前記第 1 のオーディオオブジェクトを含む前記出力オーディオ信号を生成する出力回路と、

を含む、装置。

【請求項 2】

前記推定器は、

前記画像フレームの前記シーケンスの少なくとも 1 つの画像フレーム内の相互作用画像領域を決定することであって、前記相互作用画像領域は、前記相互作用が発生する前記少なくとも 1 つの画像フレームの画像領域である、決定することと、

前記相互作用画像領域の画像データに応じて、前記シーンオブジェクトの前記材料特性を決定することと、

を実行する、請求項 1 に記載の装置。

【請求項 3】

前記第 2 の受信器は、リモートサーバから前記メタデータを受信する、請求項 1 又は 2 に記載の装置。

【請求項 4】

少なくともいくつかのオーディオオブジェクトの前記メタデータは、前記少なくともいくつかのオーディオオブジェクトと前記現実世界シーンオブジェクトの前記材料特徴との関連性の指示と、前記少なくともいくつかのオーディオオブジェクトと前記仮想シーンオブジェクトの材料特徴との関連性の指示とを含み、前記セレクトは、前記材料特性、及び前記オーディオオブジェクトの前記セットに関連付けられた前記現実世界シーンオブジェクトの前記材料特徴に応じて、且つ前記仮想シーンオブジェクトの材料特性及び前記オーディオオブジェクトの前記セットに関連付けられた前記仮想シーンオブジェクトの前記材料特徴に応じて、前記第 1 のオーディオオブジェクトを選択する、請求項 1 から 3 のいずれか一項に記載の装置。

【請求項 5】

前記セレクトは、前記仮想シーンオブジェクトの動的特性に応じて前記第 1 のオーディオオブジェクトを選択する、請求項 1 から 4 のいずれか一項に記載の装置。

【請求項 6】

前記検出器は前記相互作用の特性を決定し、前記セレクトは前記相互作用の前記特性に

10

20

30

40

50

応じて前記第 1 のオーディオオブジェクトを選択する、請求項 1 から 5 のいずれか一項に記載の装置。

【請求項 7】

前記相互作用の前記特性は、

前記相互作用の速度、

前記仮想シーンオブジェクトと前記現実世界シーンオブジェクトとの衝突の力、

前記仮想シーンオブジェクトと前記現実世界シーンオブジェクトとの衝突の弾性、

前記相互作用の持続時間、及び

前記現実世界シーンオブジェクトに対する前記仮想シーンオブジェクトの動きの方向、の群から選択される少なくとも 1 つの特性である、請求項 6 に記載の装置。

10

【請求項 8】

前記セクタは、前記現実世界シーンオブジェクトに対する前記仮想オブジェクトの向きに応じて前記第 1 のオーディオオブジェクトを選択する、請求項 1 から 7 のいずれか一項に記載の装置。

【請求項 9】

前記推定器は、複数のオブジェクトカテゴリの中の少なくとも第 1 のカテゴリに対する前記現実世界シーンオブジェクトの一致指示を決定し、且つ前記一致指示、及び前記オブジェクトカテゴリに関連付けられている前記材料特性に応じて前記材料特性を決定する、請求項 1 から 8 のいずれか一項に記載の装置。

【請求項 10】

前記現実世界のシーンでキャプチャされたリアルタイムオーディオのオーディオ信号を受信するオーディオ受信器を更に含み、前記推定器は、前記オーディオ信号に応じて前記一致指示を決定する、請求項 9 に記載の装置。

20

【請求項 11】

前記セクタは、選択基準を満たすオーディオオブジェクトが検出されなかった場合に、前記第 1 のオーディオオブジェクトをデフォルトのオーディオオブジェクトとして選択する、請求項 1 から 10 のいずれか一項に記載の装置。

【請求項 12】

少なくとも 1 つの画像フレームが、深度画像データを含み、前記推定器は、前記現実世界シーンオブジェクトを表す前記少なくとも 1 つの画像フレームの画像領域の少なくとも一部が、閾値を超えない、深度画像データ用の信頼レベルを有するとの検出に応じて、前記現実世界シーンオブジェクトの前記材料特性を決定する、請求項 1 から 11 のいずれか一項に記載の装置。

30

【請求項 13】

出力オーディオ信号を生成する方法であって、前記方法は、

画像センサから現実世界のシーンのリアルタイム画像シーケンスを受信するステップであって、前記リアルタイム画像シーケンスは画像フレームのシーケンスを含み、各画像フレームは視覚画像データ及び深度画像データのうちの少なくとも一方を含む、受信するステップと、

オーディオオブジェクトのセット、及び前記オーディオオブジェクトの前記セットのオーディオオブジェクトのメタデータを受信するステップであって、前記メタデータは、前記オーディオオブジェクトの前記セットの前記オーディオオブジェクトと材料特徴との関連性を示す、受信するステップと、

40

前記現実世界のシーンの仮想シーンオブジェクトに対応する画像オブジェクトを含む出力画像シーケンスを生成するステップと、

前記仮想シーンオブジェクトと前記現実世界のシーンの現実世界シーンオブジェクトとの相互作用を、前記仮想シーンオブジェクトと前記現実世界シーンオブジェクトとの近接の検出に応じて検出するステップと、

前記画像フレームの前記シーケンスの画像データに応じて、前記現実世界シーンオブジェクトの材料特性を決定するステップと、

50

前記材料特性、及び前記オーディオオブジェクトの前記セットのオーディオオブジェクトに関連付けられた前記材料特徴に応じて、前記オーディオオブジェクトの前記セットの第1のオーディオオブジェクトを選択するステップと、

前記第1のオーディオオブジェクトを含む前記出力オーディオ信号を生成するステップと、

を含む、方法。

【請求項14】

コンピュータプログラムコード手段を含むコンピュータプログラムであって、前記コンピュータプログラムコード手段は、前記プログラムがコンピュータ上で実行されると、請求項13に記載の方法の全てのステップを行う、コンピュータプログラム。

【請求項15】

オーディオオブジェクトのセットと、前記オーディオオブジェクトのメタデータとを含み、前記メタデータは、前記オーディオオブジェクトと現実世界オブジェクトの材料特徴との関連性を示す、オーディオ信号。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、仮想世界シーンオブジェクトと現実世界シーンオブジェクトとの相互作用のためのオーディオ信号を生成する装置及び方法に関し、特に、排他的ではないが、拡張現実アプリケーション用のオーディオ信号を生成する装置及び方法に関する。

【背景技術】

【0002】

近年では、オーディオビジュアルコンテンツに基づいたエクスペリエンスの多様性や範囲が大幅に拡大しており、そのようなコンテンツを利用し消費する新しいサービスや手法が継続的に開発され紹介されている。特に、より複雑で没入感のあるエクスペリエンスをユーザに提供するために、多くの空間的及びインタラクティブなサービス、アプリケーション、及びエクスペリエンスが開発されている。

【0003】

このようなアプリケーションの例として、急速に主流になりつつある仮想現実（VR）アプリケーション、拡張現実（AR）アプリケーション、及び複合現実（MR）アプリケーションがあり、コンシューマ市場向けのソリューションが多数ある。また、多くの標準化団体によっていくつかの標準が開発中である。このような標準化活動では、ストリーミング、ブロードキャスト、レンダリングなど、VR／AR／MRシステムの様々な側面の標準化が積極的に行われている。

【0004】

VRアプリケーションは、ユーザが異なる世界／環境／シーンにいることに対応したユーザエクスペリエンスを提供する傾向があるが、AR（混合現実MRを含む）アプリケーションは、ユーザが実際のローカル環境であるが、追加情報又は仮想オブジェクト若しくは情報が追加された環境にいることに対応したユーザエクスペリエンスを提供する傾向がある。したがって、VRアプリケーションは完全に没入感のある合成的に生成された世界／シーンを提供する傾向があるが、ARアプリケーションは、ユーザが物理的に存在する実際のシーンにオーバーレイされる部分的に合成された世界／シーンを提供する傾向がある。しかし、これらの用語はしばしば同じ意味で使用され、重複の度合いが高い。以下では、拡張現実／ARという用語は、拡張現実と複合現実の両方を示すために使用する（また、いくつかのバリエーションは仮想現実と呼ぶことがある）。

【0005】

例として、ハンドヘルドデバイスを使用した拡張現実用のサービス及びアプリケーションの普及が進み、スマートフォンやタブレットでの拡張現実アプリケーションをサポートするために、ARKit（Apple社が開発）やARCore（Google社が開発）などのソフトウェアAPI（アプリケーションプログラミングインターフェース）やツ

10

20

30

40

50

ールキットが導入されている。これらのアプリケーションでは、内蔵カメラやデバイスの他のセンサを使用して、環境のリアルタイム画像を、その提示された画像にオーバーレイする仮想グラフィックスと共に生成する。例えばアプリケーションは、ライブビデオフィードを、そのライブビデオにオーバーレイするグラフィックオブジェクトと共に生成する。このようなグラフィックオブジェクトを使用して、例えば現実世界のシーンに存在すると認識されるように仮想オブジェクトを配置できる。

【0006】

別の例として、ARメガネを通して現実世界のシーンを直接見ることができるヘッドセットやメガネが開発されているが、これらはユーザがメガネを通して見た画像を生成することもできる。これは、見ている現実世界のシーンの一部であると認識される仮想画像を提示するためにも使用できる。モーションセンサを使用して頭の動きを追跡し、提示される仮想オブジェクトは、仮想オブジェクトが現実世界で見られる現実世界のオブジェクトであるという印象を与えるために、それに応じて適応される。

【0007】

これらのアプローチはそれぞれパススルーとシースルーとして知られており、どちらも斬新でエキサイティングなユーザエクスペリエンスを提供できる。

【0008】

視覚的レンダリングに加えて、いくつかのARアプリケーションでは、対応するオーディオエクスペリエンスを提供できる。更に、視覚的エクスペリエンスに対して、仮想オブジェクトに対応するオーディオを提供することが提案されている。例えば、仮想オブジェクトがノイズを発生するオブジェクトである場合、オブジェクトのビジュアル表現は、対応する音が生成されることによって補われる。場合によっては、仮想オブジェクトのアクションを反映するために音が生成されることもある。

【0009】

通常、事前に録音されたオーディオクリップを適切なタイミングでレンダリングすることで、音はデフォルトの音として生成される。場合によっては、音は、例えば、現在の環境に応じて認識される残響を適応させるなど、現在の環境を反映するように適応されるか、又は、例えば、現実世界における仮想オブジェクトの認識位置に対応する位置から到達するように認識されるように処理される。多くの場合、このようなポジショニングは、適切なヘッドフォンオーディオ出力を生成するためにバイノーラル処理によって達成される。

【0010】

しかし、このようなアプローチは、多くの実施形態において興味深いアプリケーション及びユーザエクスペリエンスを提供する可能性があるが、従来のアプローチは準最適である傾向があり、実装が困難である傾向があり、並びに／又は準最適なパフォーマンス及びユーザエクスペリエンスを提供する傾向がある。

【0011】

そのため、改良されたアプローチが有利である。特に、操作の改善、柔軟性の向上、複雑さの軽減、実装の容易化、オーディオエクスペリエンスの向上、オーディオ品質の向上、計算負荷の軽減、複合現実／拡張現実アプリケーションの適合性やパフォーマンスの向上、ユーザの没入感の向上、並びに／又はパフォーマンスや操作の改善を可能にするオーディオ信号を生成するアプローチが有利である。

【発明の概要】

【発明が解決しようとする課題】

【0012】

したがって、本発明は、好ましくは、上記の欠点のうちの1つ以上を、単独又は任意の組み合わせで軽減、緩和、又は排除しようとする務めるものである。

【課題を解決するための手段】

【0013】

本発明の一態様によれば、出力オーディオ信号を生成する装置が提供される。この装置

10

20

30

40

50

は、画像センサから現実世界のシーンのリアルタイム画像シーケンスを受信する第1の受信器であって、リアルタイム画像シーケンスは画像フレームのシーケンスを含み、各画像フレームは視覚画像データ及び深度画像データのうちの少なくとも一方を含む、第1の受信器と、オーディオオブジェクトのセット、及びオーディオオブジェクトのセットのオーディオオブジェクトのメタデータを受信する第2の受信器であって、メタデータは、オーディオオブジェクトのセットのオーディオオブジェクトと材料特徴との関連性を示す、第2の受信器と、現実世界のシーンの仮想シーンオブジェクトに対応する画像オブジェクトを含む出力画像シーケンスを生成する画像生成器と、仮想シーンオブジェクトと現実世界のシーンの現実世界シーンオブジェクトとの相互作用を、仮想シーンオブジェクトと現実世界シーンオブジェクトとの近接の検出に応じて検出する検出器と、画像フレームのシーケンス画像データに応じて、現実世界シーンオブジェクトの材料特性を決定する推定器と、材料特性、及びオーディオオブジェクトのセットのオーディオオブジェクトに関連付けられた材料特徴に応じて、オーディオオブジェクトのセットの第1のオーディオオブジェクトを選択するセレクトと、第1のオーディオオブジェクトを含む出力オーディオ信号を生成する出力回路とを含む。

10

【0014】

本発明は、多くの実施形態において、ユーザエクスペリエンスを向上させ、また、特に、多くの実施形態において、向上されたより没入感のある拡張現実エクスペリエンスを提供する。このアプローチは、多くの実施形態において、複雑さや実施を低く維持しながら達成される。このアプローチは、オーディオとサポートするメタデータとがリモートサーバから提供されるARシステムに非常に適している。このアプローチは、集中型サーバが、多数のリモートクライアントへのサポートを提供し、リモートクライアントの実施を実質的に容易にするシステムを容易及び／又はサポートする。このアプローチは、ARアプリケーション及びエクスペリエンスを高めるためのオーディオの集中型生成及び管理をサポートする。

20

【0015】

オーディオオブジェクトは、オーディオクリップ／フラグメント／などであり、任意の適切なやり方で提示される。多くの実施形態では、各オーディオオブジェクトは、ある時間間隔における音を表す。多くの実施形態では、この時間間隔は、いずれのオーディオオブジェクトについても5秒、10秒、又は20秒未満である。

30

【0016】

材料特徴は、現実世界のオブジェクトの材料特徴である。

【0017】

出力画像シーケンスは、視覚画像を含む。

【0018】

相互作用は、現実世界のシーンを表す3次元座標系において検出される。相互作用とは、仮想シーンオブジェクトと現実世界シーンオブジェクトとの近接／衝突／接触である。

【0019】

装置は拡張現実装置である。この装置は、拡張現実アプリケーション用の出力画像シーケンス及び出力オーディオ信号を提供する装置である。拡張現実アプリケーションは、現実世界のシーン内に仮想シーンオブジェクトを提示する。

40

【0020】

本発明の任意選択の特徴によれば、推定器は、画像フレームのシーケンスの少なくとも1つの画像フレーム内の相互作用画像領域を決定することであって、相互作用画像領域は、相互作用が発生する少なくとも1つの画像フレームの画像領域である、決定することと、相互作用画像領域の画像データに応じて、シーンオブジェクトの材料特性を決定することとを実行する。

【0021】

これにより、多くの実施形態において、特に効率的で有利な材料特性推定が提供され、特に、多くの実施形態において、より正確な材料特性推定が可能になる。このアプローチ

50

は、結果として、ユーザエクスペリエンスを向上させる。

【0022】

本発明の任意選択の特徴によれば、第2の受信器は、リモートサーバからメタデータを受信する。

【0023】

このアプローチにより、オーディオが、適切な現在の条件にローカルで効果的に適応されつつ、リモートで、場合によっては、中央で生成されて管理される特に効率的なアプリケーションが提供される。

【0024】

本発明の任意選択の特徴によれば、少なくともいくつかのオーディオオブジェクトのメタデータは、少なくともいくつかのオーディオオブジェクトと現実世界シーンオブジェクトの材料特徴との関連性の指示と、少なくともいくつかのオーディオオブジェクトと仮想シーンオブジェクトの材料特徴との関連性の指示とを含み、セレクトは、材料特性、及びオーディオオブジェクトのセットに関連付けられた現実世界のオブジェクトの材料特徴に応じて、且つ仮想シーンオブジェクトの材料特性及びオーディオオブジェクトのセットに関連付けられた仮想シーンオブジェクトの材料特徴に応じて、第1のオーディオオブジェクトを選択する。

10

【0025】

これにより、多くの実施形態において、性能が向上され、また、特に、特定の相互作用への適応が向上される。大抵の場合、より没入感のあるユーザエクスペリエンスが達成される。

20

【0026】

本発明の任意選択の特徴によれば、セレクトは、仮想シーンオブジェクトの動的特性に応じて第1のオーディオオブジェクトを選択する。

【0027】

これにより、多くの実施形態において、性能が向上され、また、特に、特定の相互作用への適応が向上される。大抵の場合、より没入感のあるユーザエクスペリエンスが達成される。

【0028】

本発明の任意選択の特徴によれば、検出器は相互作用の特性を決定し、セレクトは相互作用の特性に応じて第1のオーディオオブジェクトを選択する。

30

【0029】

これにより、多くの実施形態において、性能が向上され、また、特に、特定の相互作用への適応が向上される。大抵の場合、より没入感のあるユーザエクスペリエンスが達成される。

【0030】

本発明の任意選択の特徴によれば、相互作用の特性は、相互作用の速度、仮想シーンオブジェクトと現実世界シーンオブジェクトとの衝突の力、仮想シーンオブジェクトと現実世界シーンオブジェクトとの衝突の弾性、相互作用の持続時間、及び現実世界シーンオブジェクトに対する仮想シーンオブジェクトの動きの方向の群から選択される少なくとも1つの特性である。

40

【0031】

本発明の任意選択の特徴によれば、セレクトは、現実世界シーンオブジェクトに対する仮想オブジェクトの向きに応じて第1のオーディオオブジェクトを選択する。

【0032】

これにより、多くの実施形態において、性能が向上され、また、特に、特定の相互作用への適応が向上される。大抵の場合、より没入感のあるユーザエクスペリエンスが達成される。

【0033】

本発明の任意選択の特徴によれば、推定器は、複数のオブジェクトカテゴリの中の少な

50

くとも第1のカテゴリに対する現実世界シーンオブジェクトの一致指示を決定し、且つ一致指示、及びオブジェクトカテゴリに関連付けられている材料特性に応じて材料特性を決定する。

【0034】

これにより、特に有利で、大抵の場合は、複雑さの低い材料特性の決定が提供される。この決定は、しかしながら、依然として精度は高い。多くの実施形態では、カテゴリ化／分類は、ニューラルネットワークを使用して有利に達成される。

【0035】

本発明の任意選択の特徴によれば、本装置は、現実世界のシーンでキャプチャされたりリアルタイムオーディオのオーディオ信号を受信するオーディオ受信器を更に含み、推定器は、オーディオ信号に応じて一致指示を決定する。

10

【0036】

このようなアプローチは、多くの実施形態では、材料特性推定の精度を実質的に向上させ、性能全体が向上される。

【0037】

本発明の任意選択の特徴によれば、セレクトは、選択基準を満たすオーディオオブジェクトが検出されなかった場合に、第1のオーディオオブジェクトをデフォルトのオーディオオブジェクトとして選択する。

【0038】

本発明の任意選択の特徴によれば、少なくとも1つの画像フレームが、深度画像データを含み、推定器は、現実世界シーンオブジェクトを表す少なくとも1つの画像フレームの画像領域の少なくとも一部が、閾値を超えない、深度画像データ用の信頼レベルを有するとの検出に応じて、現実世界シーンオブジェクトの材料特性を決定する。

20

【0039】

いくつかの実施形態では、画像フレームは視覚画像データと深度画像データとを含み、推定器は、画像領域の少なくとも一部について、視覚画像データの輝度が閾値を超え、深度画像データの信頼レベルが閾値を超えないとの検出に応じて、現実世界シーンオブジェクトが金属成分を有すると決定する。

【0040】

本発明の別の態様によれば、出力オーディオ信号を生成する方法が提供される。この方法は、画像センサから現実世界のシーンのリアルタイム画像シーケンスを受信するステップであって、リアルタイム画像シーケンスは画像フレームのシーケンスを含み、各画像フレームは視覚画像データ及び深度画像データのうちの少なくとも一方を含む、受信するステップと、オーディオオブジェクトのセット、及びオーディオオブジェクトのセットのオーディオオブジェクトのメタデータを受信するステップであって、メタデータは、オーディオオブジェクトのセットのオーディオオブジェクトと材料特徴との関連性を示す、受信するステップと、現実世界のシーンの仮想シーンオブジェクトに対応する画像オブジェクトを含む出力画像シーケンスを生成するステップと、仮想シーンオブジェクトと現実世界のシーンの現実世界シーンオブジェクトとの相互作用を、仮想シーンオブジェクトと現実世界のシーンオブジェクトとの近接の検出に応じて検出するステップと、画像フレームのシーケンスの画像データに応じて、現実世界シーンオブジェクトの材料特性を決定するステップと、材料特性、及びオーディオオブジェクトのセットのオーディオオブジェクトに関連付けられた材料特徴に応じて、オーディオオブジェクトのセットの第1のオーディオオブジェクトを選択するステップと、第1のオーディオオブジェクトを含む出力オーディオ信号を生成するステップとを含む。

30

40

【0041】

この方法は、出力画像シーケンスを表示するステップや、出力オーディオ信号をレンダリングするステップを含んでいてもよい。

【0042】

本発明のこれらの及び他の態様、特徴、及び利点は、以下に説明される実施形態から明

50

らかになり、また、当該実施形態を参照して説明される。

【図面の簡単な説明】

【0043】

本発明の実施形態を、ほんの一例として図面を参照して以下に説明する。

【0044】

【図1】図1は、拡張現実システムの要素の一例を示す。

【図2】図2は、本発明のいくつかの実施形態に従って、出力オーディオ信号を生成するオーディオ装置の一例を示す。

【図3】図3は、仮想シーンオブジェクトを有する現実世界のシーンの画像の一例を示す。

【図4】図4は、本発明のいくつかの実施形態に従って、出力オーディオ信号を生成するアプローチの一例を示す。

【発明を実施するための形態】

【0045】

以下の説明では、拡張現実アプリケーションの一部として、現実世界のシーンにおける仮想オブジェクトの画像生成を補完するオーディオ信号の生成に焦点を当てる。しかし、説明する原理及び概念は、他の多くのアプリケーションや実施形態で使用されてもよいことが理解されるであろう。

【0046】

現実世界の環境を補完するために仮想情報やオブジェクトを提示できる拡張現実エクスペリエンスがますます普及しており、その需要を満たすためのサービスが開発されている。

【0047】

多くのアプローチでは、ARアプリケーションは、リモートARサーバを使用しない、又はアクセスさえしないスタンドアロンデバイスなどによって、視聴者のローカルに提供される。ただし、他のアプリケーションでは、ARアプリケーションはリモートサーバ又は中央サーバから受信したデータに基づいている場合がある。例えば、リモートの中央サーバからARデバイスにオーディオデータやグラフィックデータが提供され、ローカルで処理されて所望のARエクスペリエンスが生成される。

【0048】

図1は、リモートARクライアントデバイス101が、ARサーバ103とインターネットなどのネットワーク105を介して通信するARシステムの一例を示す。サーバ103は、同時に多数のクライアントデバイス101をサポートできるように構成される。

【0049】

例えば、ARサーバ103は、仮想環境の要素とオブジェクトを定義するデータをクライアントデバイス101に送信することで、拡張エクスペリエンスをサポートする。このデータは、クライアントデバイス101がユーザに提示できるオーバーレイグラフィックスを生成するために使用する、多数の仮想オブジェクトの視覚的特徴や幾何学的特性を具体的に記述する。いくつかの実施形態では、データにはユーザに提示できる様々な情報も含まれている。更に、サーバ103は、ユーザエクスペリエンス、特に没入感を更に高め得る仮想音／オーディオをローカルに生成するために使用できるオーディオデータをクライアントデバイス103に提供する。

【0050】

図2は、本発明のいくつかの実施形態に従うデバイスを示す。このデバイスは、具体的には図1のクライアントデバイス101であり、そのような実施形態を参照して説明する。

【0051】

装置は、1つ以上の画像センサから画像データを受信する第1の受信器201を含む。画像センサからの画像データは、具体的には実世界のシーンのフレーム／画像のリアルタイム画像シーケンスを含む。

10

20

30

40

50

【0052】

多くの実施形態では、フレームは、視覚画像カメラからの視覚画像データを含む。フレーム／画像データは、各ピクセルがピクセルのビュー方向からの受光強度を表す視覚画像のリアルタイムシーケンスを含み得る。例えば、各ピクセルには、可視スペクトルの（場合によっては重み付けされた）間隔に対する光の輝度／強度値のセットが含まれている。例えば、視覚画像のピクセル値は、例えばカラーチャンネルの輝度／強度値のセットなど、1つ以上の輝度レベルを表す。例えば、画像はRGB画像である。多くの実施形態では、視覚画像はカラー画像及び／又はRGB画像である（そして、実施形態によっては、視覚画像への参照は、カラー画像及び／又はRGB画像への参照に置き換えられる）。

【0053】

或いは又は更に、画像センサは、深度画像／フレームのリアルタイムシーケンスを提供する深度センサであってもよい。このような深度画像の場合、各ピクセル値は、ピクセルのビュー方向のオブジェクトへの深度／距離を表す。例えば、各ピクセル値は、視差値又は深度値であってもよい。深度画像は、深度マップとも呼ばれる。

【0054】

したがって、異なる実施形態において、画像センサから受信されるフレーム／画像は、例えば、視覚／カラー画像、赤外線画像、深度画像、レーダー画像、マルチスペクトル画像、ソナー画像、位相画像、位相差画像、強度画像、コヒーレンスマグニチュード画像、及び／又は信頼画像などである。したがって、画像／フレームのシーケンスは、ピクセルのビュー方向における現実世界の特性を表す値を含むピクセルの2次元構造である。

【0055】

多くの実施形態では、装置は視覚画像及び深度画像の両方を処理し、第1の受信器201は、現実世界のシーンのリアルタイムの視覚画像のシーケンスと、現実世界のシーンのリアルタイムの深度画像のシーケンスとの両方を受信できる。多くの実施形態では、画像は視覚画像と深度画像との複合画像であり、第1の受信器201は、現実世界のシーンのリアルタイムの深度画像及び視覚画像のシーケンスを受信する。

【0056】

視覚画像や深度画像を生成するための多くの異なるセンサ及びアプローチが知られており、任意の適切なアプローチ及び画像センサを使用してもよいことが理解されるであろう。例えば、従来のビデオカメラを、視覚画像を生成するための画像センサとして使用してもよい。例えば、深度画像は、赤外線レンジカメラなどの専用の深度カメラを使用して生成されてもよいし、例えば、既知の物理オフセットを有する2つの視覚カメラなどに基づく視差推定によって生成されてもよい。

【0057】

画像センサは現実世界のシーンをキャプチャしているため、受信された画像はこの現実世界のシーンのキャプチャを含む。例えば、多くのARアプリケーションでは、受信した画像はユーザの環境に対応した現実世界のシーンの画像である。具体例として、ユーザは、視聴者が見ている方向の現実世界のシーンをキャプチャする1つ以上の画像センサを含むARヘッドセット又はメガネを着用する。

【0058】

装置は更に、仮想シーンオブジェクトを表す画像オブジェクトを含む視覚画像シーケンスを生成して出力する画像生成器203を含む。出力された視覚画像シーケンスは、ARヘッドセット又はARメガネのディスプレイ（又は複数のディスプレイ）などの適切なディスプレイを使用してユーザに提示される。したがって、出力された視覚画像シーケンスが提示されると、ユーザは、画像オブジェクトが現実世界のシーンに存在する仮想オブジェクトに対応するように認識する。

【0059】

画像生成器203は、例えば、適切な動きセンサから受信したデータによって提供されるユーザの姿勢（位置及び／又は向き）の変化に応じて、画像オブジェクト（画像内の向き、画像内の位置、ビュー方向など）を適応させる。したがって、画像オブジェクトは、

10

20

30

40

50

見ている現実世界のシーンに仮想オブジェクトが存在している印象を与えるように提示される。

【0060】

いくつかの実施形態では、生成された出力視覚画像シーケンスは、仮想オブジェクト（又は複数ある場合は複数の仮想オブジェクト）のみを含み、例えばシースルーメガネを使用して提示されて、現実世界のシーンをメガネを通して見るようにできることにされる。このようなアプローチはARシースルーアプリケーションとして知られている。

【0061】

他の実施形態では、出力視覚画像シーケンスは、画像センサがキャプチャした現実世界のシーンの表現も含むように生成される。したがって、このようなアプローチでは、現実世界のシーンは、生成された画像を介しても（また、介してのみ）見るようにできる。このようなアプローチはARパースルーアプリケーションとして知られている。

【0062】

このようなAR画像を生成したり、生成された画像オブジェクトを適応、修正して仮想オブジェクトが現実世界のシーンに存在する印象を与えるための多くの異なるアプローチが知られていることが理解されるであろう。簡潔にするために、このような特徴については詳細には説明しない。本発明からそれることなく、任意の適切なアプローチを使用してもよいことが理解されるであろう。例えば、ARKitやARCoreのプロセスなど、開発されたARツールキットのアルゴリズムやアプローチを使用できる。

【0063】

したがって、画像生成器203は、仮想シーンオブジェクトが現実世界のシーンに追加されたと認識されるような画像オブジェクトを生成する。多くの実施形態では、仮想シーンオブジェクトは少なくとも1つの現実世界シーンオブジェクトに対して移動可能であり、通常、仮想シーンオブジェクトは現実世界のシーン全体に対して移動可能である。例えば、仮想シーンオブジェクトは、ユーザアクション、現実世界のシーンの特性の変化、所定の効果をもたらすための所定の動きなどに応じて移動する。

【0064】

装置は更に、仮想シーンオブジェクトと現実世界のシーンの現実世界シーンオブジェクトとの相互作用を検出する検出器205を含む。検出器205は、具体的には仮想シーンオブジェクトと現実世界シーンオブジェクトとの間の衝突又は接触を検出する。他の可能な相互作用としては、例えば、仮想シーンオブジェクトと現実世界シーンオブジェクトとの近接、又は仮想オブジェクトと現実世界オブジェクトとが互いに接触したことが確実である時間間隔である接触の持続時間の検出が考えられる。相互作用の別の例は、現実世界で雨が降り、雨滴が仮想オブジェクトに落ちる場合である。別の例は、現実世界で空気流が発生し、空気が仮想オブジェクトに向かって流れている場合である。

【0065】

したがって、検出器205は、仮想シーンオブジェクトと現実世界のシーンの現実世界シーンオブジェクトとの相互作用を、仮想シーンオブジェクトと現実世界シーンオブジェクトとの近接の検出に応じて検出する。近接の検出は、任意の適切な近接基準に基づくものであってよい。

【0066】

仮想シーンオブジェクトと現実世界シーンオブジェクトとの近接の検出は、現実世界シーンオブジェクトの位置と仮想シーンオブジェクトの位置との間の距離が近接基準を満たしていることの検出である。仮想シーンオブジェクトと現実世界シーンオブジェクトとの近接の検出は、現実世界シーンオブジェクトの位置と仮想シーンオブジェクトの位置との間の距離が閾値未満であることの検出である。位置及び距離は、現実世界のシーンのシーン座標系で決定される。仮想シーンオブジェクトは、現実世界のシーンに存在する現実世界シーンオブジェクトであるかのように提示／表示されるオブジェクトである。

【0067】

いくつかの実施形態では、検出器205は、（シーン用の）（シーン）座標系における

10

20

30

40

50

現実世界シーンオブジェクトの（第１の）位置と、上記座標系における仮想シーンオブジェクトの（第２の）位置を決定する。検出器２０５は、（第１及び第２の）位置が近接基準を満たしていることに応じて、また、特に、満たす場合に、相互作用を検出する。具体的には、検出器２０５は、（第１と第２の）位置間の（任意の適切な距離尺度に従う）距離が距離近接基準を満たしていることに応じて、また、特に、満たす場合に、また、特に、当該距離が閾値未満である場合に、相互作用を検出する。

【００６８】

画像生成器は、仮想シーンオブジェクトを、現実世界のシーン／シーン座標系において所与の姿勢／空間的存在を有するオブジェクトとして表現するために画像オブジェクトを生成する。近接は、現実世界のシーン／シーン座標系におけるこの姿勢／空間的存在を示す位置と、現実世界のシーン／シーン座標系における現実世界シーンオブジェクトの位置とに依拠して決定される。近接は、現実世界のシーン／シーン座標系におけるこの仮想シーンオブジェクト／空間的処理の位置と、現実世界のシーン／シーン座標系における現実世界シーンオブジェクトの位置とに依拠して決定される。

10

【００６９】

特定の実施形態の所望の操作及び性能に依拠して異なる相互作用を検出するためのアプローチが異なる実施形態で使用され、様々なアプローチについて後述する。

【００７０】

検出器２０５は、仮想シーンオブジェクトが相互作用するものと検出されたシーンオブジェクトの材料特性を決定する推定器２０７に結合されている。材料特性は、リアルタイムのオブジェクトができている材料／物質／物（オブジェクトが様々な材料を含む場合の材料組成を含む）を示す。材料特性は、例えば、現実世界のオブジェクトが複数のカテゴリのうちの１つでできているかどうか（例えば、木材、布、金属、プラスチックなどでできているかどうか）を示すために生成される。

20

【００７１】

推定器２０７は、画像センサから受信した画像データに依拠して材料特性を決定する。即ち、材料特性は、センサから受信した画像フレームのシーケンスのデータに依拠して決定される。多くの実施形態では、画像フレームのシーケンスは深度画像データと視覚画像データの両方を含み、推定器２０７は、深度画像データと視覚画像データの両方に依拠して材料特性を推定する。他の実施形態では、推定器２０７は視覚データのみ、又は深度データのみを考慮する。

30

【００７２】

異なる実施形態で材料特性を推定するために、異なるアルゴリズム及びアプローチを使用してもよいことが理解されるであろう。また、多くの実施形態では、材料特性の非常に不正確で信頼性の低い推定であっても、有用であり、ユーザエクスペリエンスの向上をもたらす得ることが理解されるであろう。

【００７３】

複雑さの低い実施例では、推定器２０７は、例えば、受信した視覚画像に色ベースのセグメンテーションを適用して材料特性を決定し、リアルタイム画像オブジェクトと一致すると考えられる画像オブジェクトを決定する。次に、平均色を事前定義されたカテゴリのセットと比較して、最も一致するカテゴリを見つけることができる。リアルタイム画像オブジェクトの材料特性は、最も近いカテゴリに関連付けられた材料特性に設定される。単純な例として、現実世界の画像オブジェクトが主に茶色であると決定された場合、それは木でできていると推定されたり、主に銀色である場合、それは金属であると推定されたり、主に明るい原色である場合、プラスチックであると推定されたりなどする。

40

【００７４】

ほとんどの実施形態では、材料特性を推定するためのより複雑で、且つより正確なアプローチが使用され得ることが理解されるであろう。更なる実施例については後述する。

【００７５】

装置は更に、多数のオーディオオブジェクト及びオーディオオブジェクトのメタデータ

50

を受信する第2の受信器209を含む。オーディオオブジェクトは、音／オーディオクリップ／オーディオフラグメントであり、具体的には限られた持続時間のオーディオ信号である。オーディオオブジェクトが音を表現するために、任意の表現及び形式を使用してもよいことが理解されるであろう。したがって、オーディオオブジェクトは、装置によって選択され、且つ様々なアクションが発生したときに音出力を提供するためにレンダリングされる音に対応する。

【0076】

具体的には、装置は、仮想シーンオブジェクトと現実世界のオーディオオブジェクトとの相互作用が検出されたときに、受信したオーディオオブジェクトの中からオーディオオブジェクトを選択するセレクタ211を含む。選択されたオーディオオブジェクトは、第1のオーディオオブジェクトを含む出力オーディオ信号を生成する出力回路213に供給される。したがって、出力回路213は、オーディオオブジェクトのレンダリングを含む出力オーディオ信号を生成することによって、選択されたオーディオオブジェクトをレンダリングする。

【0077】

出力回路213は、オーディオオブジェクトの形式に依存して、場合によっては、出力信号をオーディオオブジェクトのオーディオとして生成するだけか、いくつかの実施形態では、オーディオオブジェクトの復号化を含むか、オーディオオブジェクトと組み合わせられる様々なオーディオ成分（周囲音、ナレーターオーディオなど）間の混合を含むために出力オーディオ信号を生成し得ることが理解されるであろう。また、出力回路213には、例えば、直接スピーカを駆動するために、出力回路213がアナログ出力オーディオ信号を生成する実施形態では、デジタル－アナログ変換及びアナログ信号の増幅が含まれ得ることが理解されるであろう。

【0078】

出力回路213は、例えば、頭部インパルス応答（HRI R）、頭部伝達関数（HRT F）、若しくは両耳室内インパルス応答（BRI R）を用いたバイノーラル処理によるオーディオオブジェクトのレンダリング処理、又は、例えば、ベクトルベース振幅パニング（VBAP）や、例えば、オクルージョン、回折、反射、残響、ソース範囲などの更なる音響刺激によるスピーカ構成へのレンダリングを含む。具体的には、レンダリング処理は、オーディオオブジェクトの音が、検出された相互作用の場所から発生していると認識されるように構成される。

【0079】

いくつかの実施形態では、出力回路213には、特に可変遅延など、選択されたオーディオオブジェクトのレンダリングのタイミングを制御するためのタイミング適応が含まれている。例えば、「視覚」による接触／相互作用の時間と、音をレンダリングする時間との間の遅延が含まれる。遅延は、例えば、音速や仮想オブジェクトと観察者との間の距離に基づいて調整されて、例えば、生成された音が視覚的な相互作用と同時に認識されるようにする。いくつかの実施形態では、このような時間調整は、出力回路213以外の装置の他の部分で行われてもよい。

【0080】

第2の受信器209が受信するメタデータには、少なくとも一部のオーディオオブジェクトと材料特徴との関連性を示すメタデータが含まれている。例えば、メタデータは、各オーディオオブジェクトについて、関連付けられている現実世界のオブジェクトの材料を示す。例えば、メタデータは、あるオーディオオブジェクトが木でできている現実世界のオブジェクトに関連付けられており、別のオーディオオブジェクトが金属でできている現実世界のオブジェクトに関連付けられており、別のオーディオオブジェクトがプラスチックでできている現実世界のオブジェクトに関連付けられていることを示す。

【0081】

後で詳しく説明するように、多くの実施形態では、メタデータには、オーディオオブジェクトと仮想シーンオブジェクトとの関連付け、オーディオオブジェクトと仮想シーンオ

10

20

30

40

50

プロジェクトの材料特徴との関連付け、相互作用のタイプなど、他の更なる関連性が含まれていてもよい。

【0082】

セレクタ211は、検出器205が仮想シーンオブジェクトと現実世界シーンオブジェクトとの相互作用を検出したことに応じて、現実世界シーンオブジェクトの推定材料特性、及びオーディオオブジェクトと材料特徴との関連性を示すメタデータに基づいて、オーディオオブジェクトを選択する。

【0083】

複雑さの低い実施形態では、セレクタ211は、推定された材料特性と関連付けられた材料特徴とが一致するオーディオオブジェクトを単に選択する。例えば、推定された特性が「木材」の場合、「木材」に関連付けられたオーディオオブジェクトが選択され、対応する出力オーディオ信号が生成される。

【0084】

装置は、多くの実施形態で、ユーザエクスペリエンスを向上させることができ、特に、認識するオーディオが仮想シーンオブジェクトと現実世界のオブジェクトとの相互作用をより厳密に反映するより没入感のある自然な音エクスペリエンスを提供することができる。例えば、仮想シーンオブジェクトと現実世界のオブジェクトとの相互作用（衝突など）に画一的なオーディオを単に付随させるのではなく、装置は、ユーザの視覚的認識により厳密に一致したオーディオ出力を提供し、ユーザが認識する視覚入力とオーディオ入力とのより厳密な対応を提供できる。これは、装置が相互作用により厳密に一致したオーディオを適応させ、生成することによって達成される。

【0085】

システムは、例えば、仮想シーンオブジェクト用に提供された複数の記録又は合成されたオーディオオブジェクト／クリップ／フラグメントを使用する。各オーディオフラグメントは、シーン内で発生する可能性のある特定のクラスの現実世界のオブジェクトとの相互作用をモデル化していてもよい。実行時に、装置は、内蔵カメラを使用して、例えば、仮想オブジェクトが置かれた表面を分類し、この分類結果に基づいて、特定の相互作用に的確なオーディオフラグメントを再生する。

【0086】

いくつかの実施形態では、第2の受信器209は、内部ソースからオーディオオブジェクト及び／又はメタデータを受信する。しかし、多くの実施形態では、第2の受信器209は、メタデータ、及び、多くの場合はオーディオオブジェクトを、リモートサーバ（特に図1のサーバ103）から受信する。

【0087】

例えば、リモートサーバは、オブジェクト間の様々な可能な相互作用に対応するオーディオオブジェクトの大規模なライブラリを維持し、更に、相互作用に関与するオブジェクトの材料特性を定義するデータ、相互作用の特性などを含むメタデータを維持する。

【0088】

アプリケーションの初期化時、又は繰り返し率では、第2の受信器209は、全てのオーディオオブジェクトのメタデータを取得してもよい。検出器205によって相互作用が検出されると、推定器207は現実世界シーンオブジェクトの材料特性を推定し、セレクタ211はメタデータを評価して一致するオーディオオブジェクトを見つける。次に、第2の受信器209を制御して、選択したオーディオオブジェクトをサーバ103から取得する。受信すると、オーディオ装置が、受信したオーディオオブジェクトを含むオーディオ信号を生成する。

【0089】

いくつかの実施形態では、このようなアプローチはリアルタイム操作には遅すぎる場合があり、いくつかの実施形態では、第2の受信器209は、リモートサーバ103からサブセット又は全てのオーディオオブジェクトを取得し、高速取得のためにローカルに保存する。例えば、いくつかの実施形態では、装置が可能な限り多くのオブジェクトを検出し

10

20

30

40

50

て、起動時又は一定の間隔で現実世界のシーンの分析が行われる。その後、装置は、現実世界のシーンに検出されたオブジェクト又はオブジェクトに対応する材料や、ユーザ位置の近くでアクティブな又は近くにある仮想シーンオブジェクトに対応する全てのオーディオオブジェクトを取得する。他の実施形態では、ユーザは初期化時に現実世界のシーン（居間、スポーツアリーナなど）の入力を提供し、装置はそのような環境で可能なオブジェクトとして装置に保存されているオーディオオブジェクトのセットを取得する。

【0090】

したがって、このアプローチは、ARアプリケーションのオーディオサポートを提供するために、中央サーバがリモートデバイス装置と相互作用できるようにするための効率的なアプローチを提供する。上記のような中央サーバは、5Gエッジサーバなどの比較的ローカルなサーバでも表現できることが理解されるであろう。

【0091】

このアプローチでは、仮想シーンオブジェクトは、現実世界のシーンに存在していると認識されるようにユーザに提示されるため、現実世界のシーンの空間的な（通常は3次元の）幾何学的配置／形状と関連付けられている。検出器205は、3次元の現実世界空間における仮想シーンオブジェクトと現実世界シーンオブジェクトとの間の距離が所与の閾値よりも小さいことの検出、具体的にはそれらが互いに接触することの検出として相互作用を検出する。場合によっては、オブジェクトが互いに向かって移動する速度など、更なる考慮事項が考慮される。

【0092】

例えば、ARアプリケーションをサポートするために開発されたツールキットの中には、3D平面表面検出機能を提供するものもある。この機能では、時間の経過に伴う特徴点の追跡を介して、時間の経過と共に特徴点のクラスタがたどる特定のモーションによって平面領域が検出される。

【0093】

いくつかの実施形態では、検出器205は、仮想グラフィックスシーンオブジェクトの境界メッシュ内の任意の頂点と検出されたいずれかの平面との間の最短距離を追跡することによって、相互作用を検出する。

【数1】

$$d(t) \equiv \min_{i \in V} \left(\min_{j \in P} \left(\text{dist} \left(\mathbf{x}_j(t), \mathbf{x}_i(t) \right) \right) \right)$$

ここで、 $i \in V$ は仮想シーンオブジェクトの境界メッシュに存在する全ての頂点 V の集合からの頂点 i を表し、 $j \in P$ は最短距離が計算された平面であり、関数 $\text{dist}(\mathbf{x}_j(t), \mathbf{x}_i(t))$ はこの距離を評価する。仮想シーンオブジェクトの境界メッシュと画像センサの両方は一般的に動いていることに留意されたい。つまり、メッシュ内の頂点位置と検出された平面上の位置、そして、最終的な最短距離 $d(t)$ も全て時間と共に変化する。

【0094】

距離が閾値未満の場合、相互作用が発生したと検出される。具体的には、距離が小さな閾値未満の場合、又は、例えば、ゼロに達した場合、衝突が発生したと検出される。

【0095】

いくつかの実施形態では、頂点から平面までの最小距離が最小閾値距離 $\Delta_{\text{collision}}$ 内に減少し、瞬間オブジェクト速度 $v(t)$ が所与の最小閾値 v_{min} を超えたときに、相互作用、具体的には衝突が発生したと検出される。

$$(d(t) \leq \Delta_{\text{collision}}) \wedge (v(t) > v_{\text{min}})$$

【0096】

10

20

30

40

50

速度条件の背後にある理由は、仮想オブジェクトが検出された表面のすぐ前又はそれ以外で停止する可能性があるため、オブジェクトの速度が低いときは実際に衝突が発生したかどうかを判断することが実際には困難になる傾向があることである。有用なパラメータ値は、例えば、 $\Delta \text{collision} = 5 \text{ mm}$ 、 $v_{\text{min}} = 0.05 \text{ m/s}$ である。

【0097】

次に、推定器207は、相互作用が発生したことが検出された現実世界シーンオブジェクトの材料特性を決定する。多くの実施形態では、検出は、推定器207が少なくとも1つの入力画像フレーム内の相互作用画像領域を決定することに基づいている。この相互作用画像領域は相互作用が発生した画像フレームの画像領域である。したがって、推定器207は、検出器205によって検出されたオブジェクト間の接触点を含む画像フレーム内の2次元領域を決定する。例えば、2次元画像フレーム上の接触点の投影が決定され、その周囲の領域が特定される。

10

【0098】

いくつかの実施形態では、例えば、画像領域は、投影された接触点を中心にした所定の形状である。他の実施形態では、画像領域は画像特性に適応され得る。例えば、画像領域は、例えば、視覚的特性や深度が、投影された接触点の特性から所定量を超えて逸脱しないように類似性基準を満たす投影された接触点の周囲の領域として特定される。

【0099】

この場合、推定器207は、相互作用画像領域の画像データに応じてシーンオブジェクトの材料特性を決定する。例えば、色やテクスチャの変動が、所定セットのクラスの対応する特性と比較され、材料特性は、最も近い一致するクラスの特性に設定される。

20

【0100】

多くの実施形態では、画像データについて決定された画像特性を複数の画像特性参照と比較することによって材料特性が決定される。各画像特性参照は、材料特性値に関連付けられている。画像特性は、複数の画像特性参照のうちの一一致する画像特性参照に応じて／一一致する画像特性参照に関連付けられた材料特性値として決定される。一致する画像特性参照は、画像特性と一致する画像特性参照が一一致基準を満たすものとして決定される。画像特性は、輝度、色、及び／又はテクスチャのいずれか又は全ての特性である。

【0101】

具体的には、衝突が検出されると、画像の相互作用領域が決定される。これは、最初に、カメラビューマトリックス及びカメラ投影マトリックスを使用して、最も近いメッシュポイント x_i を画像に投影することによって行われる。次に、結果として得られる2D画像点 (u_i, v_i) を使用して、カメラ画像から (u_i, v_i) を中心とした固定サイズのサブ画像を決定する。図3は、ボールの形の仮想シーンオブジェクト301が、現実世界のテーブルの形の現実世界シーンオブジェクト303と相互作用し、結果として画像領域305が検出される画像フレームの例を示す。

30

【0102】

多くの実施形態では、推定器207は、複数のオブジェクトカテゴリ／クラスの中の少なくとも第1のカテゴリ／クラスに対する現実世界シーンオブジェクトの一致指示を決定する。例えば、推定器207は、色、テクスチャの変化、深度の変化など、特定の材料に対応する画像領域について保存された多数の特性を有する。推定器207は画像領域の対応する特性を決定し、これらを全てのカテゴリについて保存された特性と比較する。一致指示は、特性がどれだけ一致するかを反映するように決定され、カテゴリは、一致指示が最も高い一致を示すカテゴリとして選択される。

40

【0103】

次に、現実世界シーンオブジェクトの材料特性を、選択したカテゴリ又はクラスの保存されている特性として選択される。したがって、各カテゴリ／クラスは1つ以上の材料特性に関連付けられている。例えば、オブジェクトが作られている材料の単なる指示である。また、現実世界シーンオブジェクトの材料特性は、選択したカテゴリで保存されているものに設定される。

50

【0104】

多くの実施形態では、推定器207は現実世界シーンオブジェクトの材料特性を決定するために、特に、画像領域に対応する画像領域の特性に最も厳密に一致すると考えられるクラス又はカテゴリを特定するために、実質的により複雑なアプローチを使用してもよいことが理解されるであろう。

【0105】

実際、多くの実施形態では、推定器207は、相互作用画像領域の画像データに基づいて現実世界シーンオブジェクトの材料特性を推定するニューラルネットワークを含む。ニューラルネットワークは、特に、現実世界シーンオブジェクトのカテゴリ又はクラスを特定するために使用される。

10

【0106】

多くの実施形態では、抽出された（画像）相互作用領域は、完全な画像よりもはるかに小さいサイズである。完全な画像のサイズは2K又は4Kであることが多いが、相互作用サブ画像は256×256ピクセルなどの一定の小さいサイズを有する。

【0107】

例えば、256×256ピクセルのカラー画像は、次元3×256×256のテンソルとして表現され、トレーニング済みの分類器ニューラルネットワークに供給される。ニューラルネットワークはベクトルを出力し、事前に定義された材料クラスについて、画像が所与のクラスに属する確率を提供する。

【0108】

20

多くの実施形態において、次のオブジェクト（表面）クラスが特に有利であり得る：

地面がカーペット（材料の短い高さの変化）

地面がカーペット（材料の高い高さの変化は衝突音を抑制する）

地面が石

地面が木材

テーブル表面がガラス

テーブル表面が木材

壁

カウチ

クッション

30

家庭用オブジェクト（クッカー、キーボード、ラップトップコンピュータ）のプラスチック表面

【0109】

これらの各クラスには、特定の色やテクスチャの統計値があるか、又は、境界の幾何学的配置（テーブルやクッションなど）によって区別される。

【0110】

トレーニングデータセットとして、異なるオブジェクト表面からの画像がキャプチャされ、グラウンドトゥールスクラスラベルを使用して手動で注釈が付けられる。

【0111】

ニューラルネットワークアーキテクチャでは、オプションは、チャンネル数を2倍にする2D畳み込み層を基本ブロックとして使用し、続いて2D平均プーリング層（ストライド=2）と、活性化として整流した線形とを使用することである。各ブロックは、係数2でテンソルの空間分解能を縮小するが、チャンネル数は2倍にする。このアプローチは、テンソルが分解能Nチャンネル×1×1になるまでこれらのブロックを結合し続ける。次に、2つの全結合層を追加し、続いてシグモイド活性化関数が追加される。

40

【0112】

全てのクラス確率の最大値が所与の閾値（多くの実施形態では0.5が適切な値であると思われる）を下回っている場合、デフォルト（ニュートラル）の相互作用の音が再生される。即ち、推定器207が現実世界シーンオブジェクトの材料を十分に正確に決定できない場合、デフォルトのオーディオオブジェクトが選択される。それ以外の全ての場合、

50

セクタ 2 1 1 によってオーディオオブジェクト／音ファイルが選択される。具体的には、カテゴリ化によって確率が最も高いと決定された材料に対応するオーディオオブジェクトが選択される。

【0 1 1 3】

いくつかの実施形態では、装置は更に、現実世界のシーンでキャプチャされたリアルタイムオーディオのオーディオ信号を受信するオーディオ受信器 2 1 5 を含み、推定器 2 0 7 は、そのオーディオ信号に応じて一致指示を決定してもよい。

【0 1 1 4】

したがって、材料の分類は、マイクを使って物理空間内の音を記録するオーディオベースの分類によって支援される。物理オブジェクト間で相互作用が発生した場合、結果として生じる音は、材料の分類に貴重な情報を与えるか、より具体的な特性に与える。これは、より適切に材料を分類したり、材料の更なる属性（中実／中空、薄い／厚い、張力／たわみ、大／小）を与えたりするために使用できる。

【0 1 1 5】

例えば、現在及び最近の過去の環境音 S_{env} が、短い時間ウィンドウ（例えば 1 0 秒）にわたってキャプチャされ、材料分類器に供給される。例えば、スマートフォンを持っているユーザや AR ヘッドセットを着用しているユーザ、又は他の人が部屋の中を歩いていて、地面上の靴の音が地面の材料に関する情報を与える。材料分類器は、様々な材料についての画像及び音を収集することで生成される。上述したより高度なニューラルネットワークは、オーディオフラグメントの断片をキャプチャした画像と結び付けて、それらをトレーニング済みネットワークの入力として使用する。

【0 1 1 6】

具体例として、いくつかの実施形態では、材料特性の決定は深度画像データに応じて行われる。具体的には、推定器は、現実世界シーンオブジェクトを表す少なくとも 1 つの画像フレームの画像領域の少なくとも一部が、閾値を超えない、深度画像データ用の信頼レベルを有するとの検出に応じて、現実世界シーンオブジェクトの材料特性を決定できる。

【0 1 1 7】

深度推定値は、生成された深度値／推定値の信頼性を示す信頼データと共に生成されることがよくある。例えば、視差推定の場合、信頼レベルは、視差推定の根拠を形成する画像領域がどれだけ厳密に画像内で一致するかを反映するように決定される。別の例として、例えば赤外光の反射に基づく深度カメラの場合、信頼レベルは、所与のピクセルが受けた赤外光の量を反映するように生成される。少量の光しか受けていない場合、距離推定／距離測定のプロセスは、大量の光を受けた場合ほど正確ではなくなる。場合によっては、深度画像／マップの各ピクセルは、推定深度値と、ピクセルの深度値の信頼性を示す信頼値／レベルの両方を含む。

【0 1 1 8】

いくつかの実施形態では、材料特性の決定の際に、この深度信頼データが考慮される。例えば、いくつかの材料は、他の材料と比較して信頼レベルが低いことが知られている。例えば、いくつかの材料は局所的な視覚的外観が大きく変化するため、視差推定の信頼性は低い。

【0 1 1 9】

別の例として、センサから放出される反射光、特に反射赤外線に基づく深度推定では、いくつかの材料は、反射されてセンサに戻る光の量が大幅に減少することにより、信頼度が大幅に低下する。これは、例えば、反射しないが、赤外線の吸収が高い材料の場合である。また、これは、特に、鏡面反射を示す材料（金属オブジェクトなど）の場合でもある。この場合、反射されて送信器及びセンサに戻る赤外線はほんの少量である。なお、飛行時間又は構造化光に基づくアクティブ深度センサの場合、赤外光センサ（CMOS など）は、光送信器／エミッタと同じ場所に配置されていることが多い。同時に、金属表面から受光される視覚光は、表面による周囲光の反射のために、かなり実質的であり得る。多くの実施形態では、例えば、推定器 2 0 7 は相互作用画像領域のある部分の深度信頼度は低

10

20

30

40

50

いが、かなり明るい領域を検出する。このような部分は、金属表面といった反射率が高いが分散性のある表面を示している場合がある。

【0120】

いくつかの実施形態では、画像フレームは視覚画像データと深度画像データとを含み、推定器207は、画像領域の少なくとも一部について、視覚画像データの輝度が閾値を超え、深度画像データの信頼レベルが閾値を超えないとの検出に応じて、現実世界シーンオブジェクトが金属成分を有すると決定する。

【0121】

前述したように、メタデータはリモートソースから受信され、オーディオオブジェクトとオブジェクト（具体的には、現実世界シーンオブジェクト）の材料特性との関連性を含む。しかし、いくつかの実施形態では、メタデータは2つのオブジェクトの材料特性への関連性を含み、実際には、仮想シーンオブジェクトと現実世界シーンオブジェクトを区別する場合がある。このような場合、セクタ211は、現実世界シーンオブジェクトの推定材料特性と、仮想シーンオブジェクトの材料特性との両方に基づいてオーディオオブジェクトを選択できる。仮想シーンオブジェクトは、装置によって生成された仮想オブジェクトであるため、このオブジェクトが作られていると認識される材料は、通常は既知である。したがって、この場合、オーディオオブジェクトは、両方の特性をできるだけ厳密に一致させるように選択される。

【0122】

したがって、いくつかの実施形態では、少なくともいくつかのオーディオオブジェクトのメタデータは、少なくともいくつかのオーディオオブジェクトと現実世界シーンオブジェクトの材料特徴との関連性の指示と、少なくともいくつかのオーディオオブジェクトと仮想シーンオブジェクトの材料特徴との関連性の指示とを含む。このようなシステムでは、セクタ211は、材料特性、及びオーディオオブジェクトのセットに関連付けられた現実世界のオブジェクトの材料特徴に応じて、且つ仮想シーンオブジェクトの材料特性及びオーディオオブジェクトとのセットに関連付けられた仮想シーンオブジェクトの材料特徴に応じて、オーディオオブジェクトを選択する。

【0123】

例として、リモートサーバ103は、オーディオオブジェクトを材料特性に関連付ける2次元構造を提供する。このような構造の例として、次の2次元ルックアップテーブルがある。

【表1】

	検出された現実世界の材料			
仮想オブジェクト材料	木材	金属	コンクリート	不明
木材	木材上の木材	金属上の木材	コンクリート上の木材	不明な物の上の木材
金属	木材上の金属	金属上の金属	コンクリート上の金属	不明な物の上の金属
ゴム	木材上のゴム	金属上のゴム	コンクリート上のゴム	不明な物の上のゴム

【0124】

他の実施形態では、メタデータは、他の特徴への関連性を提供し、オーディオオブジェクトの選択では、そのような追加特徴が考慮に入れられる。

【0125】

例えば、多くの実施形態では、仮想シーンオブジェクトの動的特性が更に考慮される。このような実施形態では、メタデータは、異なるオーディオオブジェクトが仮想シーンオブジェクトの異なる状態／特性に関連付けられていることを示す。仮想シーンオブジェクトの状態／特性は動的に変化する。そして、一致するオーディオオブジェクトを見つける

際に、現在の状態／特性を更に考慮して選択が行われる。

【0126】

例えば、しぼんだ仮想ボールは、膨らんだ仮想ボールとは異なる音を発生する。いくつかの実施形態では、ボールの仮想シーンオブジェクトは、しぼんだときとふくらんだときとで2回ルックアップテーブルに含まれ、それぞれが異なるオーディオオブジェクト／音効果に関連付けられる。

【0127】

仮想オブジェクトの更なる属性が、適切な音効果の選択又は生成に影響を与える場合がある。例えば、仮想オブジェクトの向きによって異なる効果音が発生する場合がある。

【0128】

仮想シーンオブジェクトは、例えば、異なる側面に異なる表面特性を有する立方体など、異なるタイプの表面を有している場合がある。このような場合、仮想シーンオブジェクトの向きによって、どの表面が現実世界シーンオブジェクトと衝突するかが決定され、オーディオオブジェクトの選択ではこの向きが考慮される。

【0129】

具体例として、ルックアップテーブルは異なる向きの異なるエントリを含み、したがって、オーディオオブジェクトの選択には、現在の向きに対応するエントリの選択が含まれる。仮想シーンオブジェクトはメッシュで表すことができ、各フェース又はフェースグループは、オブジェクトの向き範囲、仮想材料特性、又はオーディオオブジェクトのサブセットに関連付けられている。

【0130】

最後に、現実世界シーンオブジェクトの更なる属性を検出して、音効果の選択又は生成に使用できる。開いたドアは閉じたドアとは音が異なる。ガラスは、液体が入っていると音が異なる。

【0131】

いくつかの実施形態では、オーディオオブジェクトの選択は更に、相互作用の特性、特に、相互作用のタイプに依存して応じたものであってもよい。したがって、いくつかの実施形態では、検出器205は相互作用の特性を決定し、セクタ211は相互作用の特性に応じて第1のオーディオオブジェクトを選択する。

【0132】

したがって、いくつかの実施形態では、相互作用自体の属性が音効果の選択又は生成に影響を与える場合がある。物理的な衝突は衝突の強度と非線形である場合があるため、相互作用の速度や強さは異なる音を引き起こす（例えば、ボールの跳ね返り、ガラスは特定の衝突強度で割れる）。他の例としては、衝突の方向や、物理表面上を仮想オブジェクトが滑ったり／擦れたりすることなどの非衝突相互作用がある。

【0133】

具体例として、オーディオオブジェクトの選択で更に考慮される相互作用の特性は、次のうちの1つ以上である。

【0134】

相互作用の速度：例えば、検出器205は、仮想シーンオブジェクトと現実世界シーンオブジェクトとの相対速度を決定し、これに基づいてオーディオオブジェクトを選択する。例えば、ゆっくりと床にぶつかるボールは、はるかに高い速度で床にぶつかるボールとは異なる音を有する。例えば、ルックアップテーブルには、異なる速度範囲に対して異なるエントリがある。各エントリは、異なる速度で適切な床材にぶつかるボールの記録を表すオーディオオブジェクトに関連付けられている。例として、相互作用検出には、物理的表面の向きと比較される速度ベクトルが含まれ、任意選択で相互作用の直後のその挙動により相互作用タイプの検出が可能になる。例えば、オブジェクトの速度ベクトルが物理的表面に垂直であり、且つ物理的表面に向かう場合、相互作用は「衝撃」又は「付着」の相互作用タイプと見なされる。オブジェクト表面に沿った速度ベクトルは、「滑る」相互作用タイプになる。このような異なる相互作用タイプが、異なるオーディオオブジェクトに関

10

20

30

40

50

連付けられる。

【0135】

仮想シーンオブジェクトと現実世界シーンオブジェクトとの間の衝突の力：例えば、力は、現在の速度と、仮想シーンオブジェクト及び現実世界シーンオブジェクトのうちの少なくとも一方の推定又は想定重量との関数として決定される。異なる力の衝撃が異なる音につながり、この力に基づいて異なるオーディオオブジェクトが提供されたり選択されたりする。

【0136】

仮想シーンオブジェクトと現実世界シーンオブジェクトとの衝突の弾性：例えば、床にぶつかる時、高圧に膨らんだボールは、部分的にしぼんだボールとは異なる音がする。ボールの弾性が異なることから非常に異なる音になる。特に、いくつかの実施形態では、仮想シーンオブジェクトの異なる弾性に対して異なるオーディオオブジェクトが提供され、メタデータにはそのような弾性に基づいてオーディオオブジェクトを区別するためのデータが含まれていてもよい。

【0137】

相互作用の持続時間：例えば、装置は、現実世界のオブジェクトとの相互作用が検出された後の仮想シーンオブジェクトの更新された軌道を予測する物理エンジンを含むか、又は物理エンジンと通信する。予測された軌道から、相互作用、若しくは相互作用に関与した仮想世界及び現実世界の材料の推進力、慣性、方向又は弾性による相互作用の持続時間が得られる。相互作用の持続時間が長いと、長い若しくは公称持続時間を有するオーディオオブジェクトがもたらされ、持続時間がより短いと、より短い持続時間を有するオーディオオブジェクトか、若しくは公称オーディオオブジェクトのより短いレンダリングがもたらされる。他の実施例では、相互作用は、ユーザが、仮想シーンオブジェクトを現実世界の環境でドラッグするなど、仮想コンテンツと相互作用することによってもたらされる。この相互作用は、ユーザが仮想シーンオブジェクトを現実世界のオブジェクトと近接させているか、又は（仮想的な）接触をしている限り継続される。

【0138】

現実世界シーンオブジェクトに対する仮想シーンオブジェクトの動きの方向：例えば、現実世界の木材表面に沿って触れながら仮想オブジェクトを滑らせた場合、一般的に、現実世界の木材表面を表面法線と平行な方向で局所的にぶつかる場合よりも高いピッチの音が生成される。

【0139】

多くの実施形態では、メタデータはルックアップテーブルの形式で提供される。ルックアップテーブルは、オーディオオブジェクトを選択する際に考慮する／すべき特性の数に応じて多次元的であってもよい。例えば、次の次元を有するルックアップテーブルが提供される。

次元1：仮想材料

次元2：検出された現実世界の材料

次元3：相互作用のタイプ（衝撃／滑る／跳ねる／付着／脱付着など）

次元4：相互作用の速度

【0140】

いくつかの実施形態では、基本的に2つの材料タイプを各音効果に関連付けるメタデータテーブルが提供される。しかしながら、同じ材料、更には相互作用タイプ（材料1、材料2、及び相互作用のタイプ）に多くの音効果がある場合がある。この情報に基づいて、これらは全て特定の相互作用を表すのに適している可能性がある。1つのアプローチは、常に第1のものを選ぶことであるが、音効果をランダムに選択した場合、エクスペリエンスはより説得力のあるものになる。反復的な相互作用では、オーディオの小さな違いにより、あまり人工的に聞えなくすることができる。音クリップの選択は、ライブラリのメタデータに網羅されていない相互作用の属性（速度、衝撃の方向、ボールのn回目の跳ねなど）に基づいて変更することさえもできる。

【0141】

図2の装置の具体的なアプローチは、図4のフロー図に示す。この実施例では、現在のビデオ画像Iを解析して、相互作用が発生する可能性のある画像領域が予測される。この解析は、仮想シーンオブジェクト／グラフィックスオブジェクトGの3D位置情報に基づいて行われる。仮想シーンオブジェクトが現実環境の現実世界のオブジェクトと衝突する可能性が高い結果として得られる空間画像領域Rの画像部分は、材料クラスMを決定する材料分類器（推定器207）に供給される。最後に、材料クラスMと仮想シーンオブジェクト／グラフィックスオブジェクトGに依存して、音が生成される。

【0142】

視覚情報のみに基づいた表面材料分類は、多くの実用的な実装形態や用途での応用に十分である可能性が高い。更に、一致するカテゴリや音が特定されない場合は、現実世界のオブジェクトの材料に固有のものではなく、仮想シーンオブジェクトGのデフォルトの相互作用音を使用できる。任意選択で、現在及び最近の過去の環境音 S_{env} を材料分類器に入力できる。

【0143】

例えば、このアプローチは、仮想シーンオブジェクトと現実世界シーンオブジェクトとの衝突が発生するとすぐに、事前に録音された音の再生を開始するアプリケーションを可能にする。仮想オブジェクトごとに全ての可能な現実世界の材料との相互作用音指定するのではなく、各仮想オブジェクトには材料特性が与えられ、適切な音を選択するための基準として単一のルックアップテーブルが使用される。材料分類器が事前定義されたクラスのいずれに対しても十分に高い確率を生成しない場合でも、「不明」クラスを使用して、ほぼ正確な音を生成することができる。例えば、仮想のゴムボールは、ぶつかる現実世界の表面が異なっても、多かれ少なかれ同じ音を出す。

【0144】

AR標準は、特定の物理領域を拡張するために、仮想要素のオーディオデータ、視覚データ、及びメタデータを有するビットストリームのシンタックスを提供する。この標準は、ユーザが物理環境でこれらの仮想要素と相互作用できるようにする。これらの相互作用のオーディオを拡張するために、標準は、多くの音効果クリップをデコーダに送信する手段を提供する。これらの音効果クリップは、トリガされない限りユーザに対してレンダリングされることはない。

【0145】

提供されたビットストリームには、多次元ルックアップテーブルを示すメタデータが含まれている。このルックアップテーブルは、全ての音効果クリップをそのルックアップテーブルのエントリに割り当てる。テーブルの各次元は、仮想要素と物理要素との相互作用の側面に対応している。ある相互作用が、このテーブル内に存在する、特定のテーブルエントリを共同で特定する側面（例えば、分類及び／又は論理アルゴリズムによって検出される）を有する場合、このテーブルエントリに関連付けられた音効果クリップが、物理要素と仮想要素との相互作用の位置で再生される。

【0146】

明確にするための上記の説明は、様々な機能回路、ユニット、及びプロセッサを参照して本発明の実施形態を説明していることが理解されるであろう。しかしながら、本発明を損なうことなく、様々な機能回路、ユニット、又はプロセッサ間で適切に機能を分配することは明らかである。例えば、別々のプロセッサ又はコントローラによって実行されるものと説明される機能が、同じプロセッサ又はコントローラによって実行されてもよい。したがって、特定の機能ユニット又は回路への参照は、厳密な論理若しくは物理構造又は組織を示すのではなく、説明された機能を提供するための適切な手段への参照としてのみ見なされる。

【0147】

本発明は、ハードウェア、ソフトウェア、ファームウェア、又はこれらの任意の組み合わせを含む、任意の適切な形式で実装できる。本発明は、任意選択で、1つ以上のデータ

10

20

30

40

50

プロセッサ及び／又はデジタル信号プロセッサ上で動作するコンピュータソフトウェアとして少なくとも部分的に実装されてもよい。本発明の実施形態の要素及び構成要素は、任意の適切なやり方で物理的、機能的、及び論理的に実装できる。実際に、機能は、1つのユニット、複数のユニット、又は他の機能ユニットの一部として実装できる。したがって、本発明は、1つのユニットに実装することも、異なるユニット、回路、及びプロセッサ間で物理的且つ機能的に分散させることもできる。

【0148】

本発明は、いくつかの実施形態に関連して説明されているが、本明細書に記載される特定の形態に限定されることを意図していない。むしろ、本発明の範囲は、添付の特許請求の範囲によってのみ限定されるものである。更に、ある特徴が特定の実施形態に関連して説明されているように見える場合もあるが、当業者であれば、説明される実施形態の様々な特徴を本発明に従って組み合わせてもよいことを認識するであろう。特許請求の範囲では、「含む」という用語は、他の要素やステップの存在を排除するものではない。

10

【0149】

更に、個別にリストされているが、複数の手段、要素、回路又は方法ステップは、例えば1つの回路、ユニット、又はプロセッサによって実装できる。更に、個々の特徴が異なる請求項に含まれている場合があるが、これらの特徴を有利に組み合わせることもでき、様々な請求項における包含は、特徴の組み合わせが実現可能ではない及び／又は有利ではないことを示唆するものではない。また、請求項の1つのカテゴリにおける特徴の包含は、このカテゴリの限定を示唆するものではなく、むしろ、必要に応じて、特徴が他の請求項カテゴリにも同様に適用できることを示している。更に、請求項における特徴の順序は、特徴が機能する必要がある特定の順序を示唆するものではなく、特に、方法の請求項における個々のステップの順序は、この順序でステップを実行する必要があることを示唆するものではない。むしろ、ステップは、任意の適切な順序で実行できる。また、単数形の参照は、複数形の参照を排除するものではない。したがって、「第1の」、「第2の」などの参照は、複数形の参照を排除するものではない。特許請求の範囲における参照符号は、明確にするための例としてのみ提供されており、これらの例は、いかようにも特許請求の範囲を限定するものと解釈されるべきではない。

20

【0150】

命令を含むコンピュータプログラム製品が提供される。命令は、プログラムがコンピュータによって実行されると、上記コンピュータに、請求項13の方法[のステップ]を実行させる。

30

【0151】

実施形態：

[請求項1]

出力オーディオ信号を生成する装置であって、

画像センサから現実世界のシーンのリアルタイム画像シーケンスを受信する第1の受信器(201)であって、上記リアルタイム画像シーケンスは画像フレームのシーケンスを含み、各画像フレームは視覚画像データ及び深度画像データのうちの少なくとも一方を含む、第1の受信器(201)と、

40

オーディオオブジェクトのセット、及びオーディオオブジェクトの上記セットのオーディオオブジェクトのメタデータを受信する第2の受信器(209)であって、上記メタデータは、オーディオオブジェクトの上記セットのオーディオオブジェクトと材料特徴との関連性を示す、第2の受信器(209)と、

上記現実世界のシーンの仮想シーンオブジェクトに対応する画像オブジェクトを含む出力画像シーケンスを生成する画像生成器(203)と、

上記仮想シーンオブジェクトと上記現実世界のシーンの現実世界シーンオブジェクトとの相互作用を検出する検出器(205)と、

画像フレームの上記シーケンスの画像データに応じて、上記現実世界シーンオブジェクトの材料特性を決定する推定器(207)と、

50

上記材料特性、及びオーディオオブジェクトの上記セットのオーディオオブジェクトに関連付けられた上記材料特徴に応じて、オーディオオブジェクトの上記セットの第1のオーディオオブジェクトを選択するセクタ(211)と、

上記第1のオーディオオブジェクトを含む上記出力オーディオ信号を生成する出力回路(213)と、

を含む、装置。

[請求項2]

上記推定器(207)は、

画像フレームの上記シーケンスの少なくとも1つの画像フレーム内の相互作用画像領域を決定することであって、上記相互作用画像領域は、上記相互作用が発生する上記少なくとも1つの画像フレームの画像領域である、決定することと、

上記相互作用画像領域の画像データに応じて、上記シーンオブジェクトの上記材料特性を決定することと、

を実行する、請求項1に記載の装置。

[請求項3]

上記第2の受信器(209)は、リモートサーバから上記メタデータを受信する、請求項1又は2に記載の装置。

[請求項4]

少なくともいくつかのオーディオオブジェクトの上記メタデータは、上記少なくともいくつかのオーディオオブジェクトと現実世界シーンオブジェクトの材料特徴との関連性の指示と、上記少なくともいくつかのオーディオオブジェクトと仮想シーンオブジェクトの材料特徴との関連性の指示とを含み、上記セクタ(211)は、上記材料特性、及びオーディオオブジェクトの上記セットに関連付けられた現実世界のオブジェクトの材料特徴に応じて、且つ上記仮想シーンオブジェクトの材料特性及びオーディオオブジェクトの上記セットに関連付けられた仮想シーンオブジェクトの材料特徴に応じて、上記第1のオーディオオブジェクトを選択する、請求項1から3のいずれか一項に記載の装置。

[請求項5]

上記セクタ(211)は、上記仮想シーンオブジェクトの動的特性に応じて上記第1のオーディオオブジェクトを選択する、請求項1から4のいずれか一項に記載の装置。

[請求項6]

上記検出器(205)は上記相互作用の特性を決定し、上記セクタ(211)は上記相互作用の上記特性に応じて上記第1のオーディオオブジェクトを選択する、請求項1から5のいずれか一項に記載の装置。

[請求項7]

上記相互作用の上記特性は、

上記相互作用の速度、

上記仮想シーンオブジェクトと上記現実世界シーンオブジェクトとの衝突の力、

上記仮想シーンオブジェクトと上記現実世界シーンオブジェクトとの衝突の弾性、

上記相互作用の持続時間、及び

上記現実世界シーンオブジェクトに対する上記仮想シーンオブジェクトの動きの方向、の群から選択される少なくとも1つの特性である、請求項6に記載の装置。

[請求項8]

上記セクタ(211)は、上記現実世界シーンオブジェクトに対する上記仮想オブジェクトの向きに応じて上記第1のオーディオオブジェクトを選択する、請求項1から7のいずれか一項に記載の装置。

[請求項9]

上記推定器(207)は、複数のオブジェクトカテゴリの中の少なくとも第1のカテゴリに対する上記現実世界シーンオブジェクトの一致指示を決定し、且つ上記一致指示、及び上記オブジェクトカテゴリに関連付けられている材料特性に応じて上記材料特性を決定する、請求項1から8のいずれか一項に記載の装置。

[請求項 10]

上記現実世界のシーンでキャプチャされたリアルタイムオーディオのオーディオ信号を受信するオーディオ受信器(215)を更に含み、上記推定器は、上記オーディオ信号に応じて上記一致指示を決定する、請求項9に記載の装置。

[請求項 11]

上記セクタ(211)は、選択基準を満たすオーディオオブジェクトが検出されなかった場合に、上記第1のオーディオオブジェクトをデフォルトのオーディオオブジェクトとして選択する、請求項1から10のいずれか一項に記載の装置。

[請求項 12]

少なくとも1つの画像フレームが、深度画像データを含み、上記推定器(207)は、上記現実世界シーンオブジェクトを表す上記少なくとも1つの画像フレームの画像領域の少なくとも一部が、閾値を超えない、深度画像データ用の信頼レベルを有するとの検出に応じて、上記現実世界シーンオブジェクトの上記材料特性を決定する、請求項1から11のいずれか一項に記載の装置。

[請求項 13]

出力オーディオ信号を生成する方法であって、

画像センサから現実世界のシーンのリアルタイム画像シーケンスを受信するステップであって、上記リアルタイム画像シーケンスは画像フレームのシーケンスを含み、各画像フレームは視覚画像データ及び深度画像データのうちの少なくとも一方を含む、受信するステップと、

オーディオオブジェクトのセット、及びオーディオオブジェクトの上記セットのオーディオオブジェクトのメタデータを受信するステップであって、上記メタデータは、オーディオオブジェクトの上記セットのオーディオオブジェクトと材料特徴との関連性を示す、受信するステップと、

上記現実世界のシーンの仮想シーンオブジェクトに対応する画像オブジェクトを含む出力画像シーケンスを生成するステップと、

上記仮想シーンオブジェクトと上記現実世界のシーンの現実世界シーンオブジェクトとの相互作用を検出するステップと、

画像フレームの上記シーケンスの画像データに応じて、上記現実世界シーンオブジェクトの材料特性を決定するステップと、

上記材料特性、及びオーディオオブジェクトの上記セットのオーディオオブジェクトに関連付けられた上記材料特徴に応じて、オーディオオブジェクトの上記セットの第1のオーディオオブジェクトを選択するステップと、

上記第1のオーディオオブジェクトを含む上記出力オーディオ信号を生成するステップと、

を含む、方法。

[請求項 14]

コンピュータプログラムコード手段を含むコンピュータプログラム製品であって、上記コンピュータプログラムコード手段は、上記プログラムがコンピュータ上で実行されると、請求項13に記載の全てのステップを行うように適応されている、コンピュータプログラム製品。

[請求項 15]

オーディオオブジェクトのセットと、上記オーディオオブジェクトのメタデータと、を含み、上記メタデータは、オーディオオブジェクトと現実世界のオブジェクトの材料特徴との関連性を示す、オーディオ信号。

10

20

30

40

50

【図面】
【図 1】

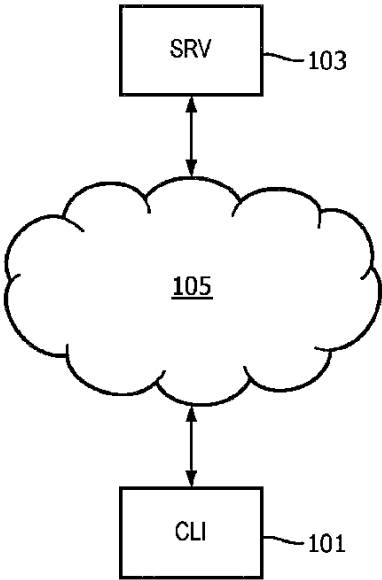


FIG. 1

【図 2】

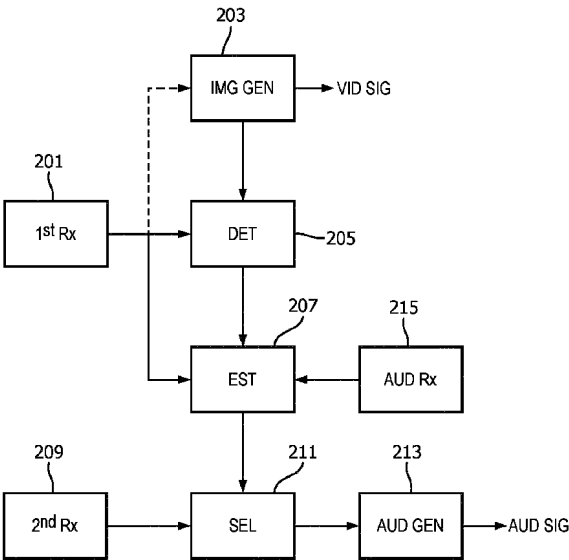


FIG. 2

【図 3】

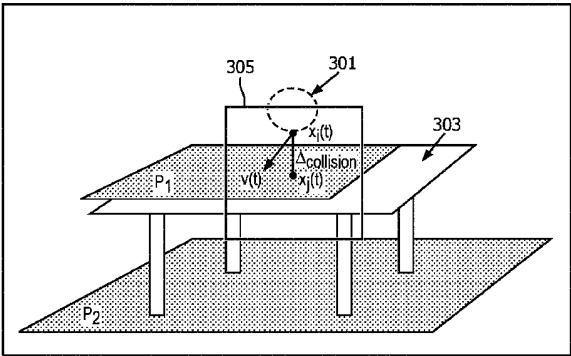


FIG. 3

【図 4】

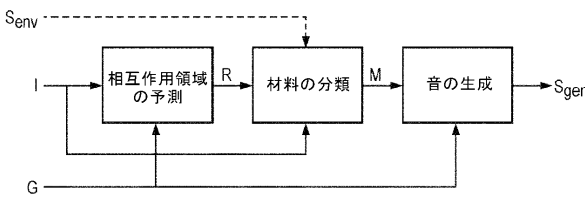


図 4

10

20

30

40

50

【国際調査報告】

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2022/058273

A. CLASSIFICATION OF SUBJECT MATTER INV. G06F3/01 G06F3/16 ADD.		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED		
Minimum documentation searched (classification system followed by classification symbols) G06F		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) EPO-Internal, WPI Data		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2013/236040 A1 (CRAWFORD DAVID W [US] ET AL) 12 September 2013 (2013-09-12) paragraphs [0048] - [0056]; figures 1, 2, 3, 4 -----	1-15
X	US 2021/082191 A1 (TAJIK ANASTASIA ANDREYEVNA [US]) 18 March 2021 (2021-03-18) paragraphs [0042], [0043], [0062] - [0067], [0074], [0077] - [0090]; figures 2A-10c ----- -/--	1-15
☒ Further documents are listed in the continuation of Box C. ☒ See patent family annex.		
* Special categories of cited documents : "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search 19 May 2022		Date of mailing of the international search report 30/05/2022
Name and mailing address of the ISA/ European Patent Office, P.B. 5618 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016		Authorized officer Seifert, J

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2022/058273

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CORREA ANA ET AL: "GenVirtual: An Augmented Reality Musical Game for Cognitive and Motor Rehabilitation", VIRTUAL REHABILITATION, 2007, 1 September 2007 (2007-09-01), pages 1-6, XP055922648, Pi DOI: 10.1109/ICVR.2007.4362120 ISBN: 978-1-4244-1203-7 the whole document -----	1-14

10

20

30

40

50

INTERNATIONAL SEARCH REPORT				International application No PCT/EP2022/058273	
Information on patent family members					
Patent document cited in search report		Publication date	Patent family member(s)		Publication date
US 2013236040	A1	12-09-2013	NONE		

US 2021082191	A1	18-03-2021	CA 3090178 A1		22-08-2019
			CN 111713090 A		25-09-2020
			JP 2021514082 A		03-06-2021
			US 2021082191 A1		18-03-2021
			WO 2019161312 A1		22-08-2019

10

20

30

40

50

フロントページの続き

MK,MT,NL,NO,PL,PT,RO,RS,SE,SI,SK,SM,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GQ,GW,KM,ML,MR,N
E,SN,TD,TG),AE,AG,AL,AM,AO,AT,AU,AZ,BA,BB,BG,BH,BN,BR,BW,BY,BZ,CA,CH,CL,CN,CO,CR,CU,
CZ,DE,DJ,DK,DM,DO,DZ,EC,EE,EG,ES,FI,GB,GD,GE,GH,GM,GT,HN,HR,HU,ID,IL,IN,IR,IS,IT,JM,JO,J
P,KE,KG,KH,KN,KP,KR,KW,KZ,LA,LC,LK,LR,LS,LU,LY,MA,MD,ME,MG,MK,MN,MW,MX,MY,MZ,N
A,NG,NI,NO,NZ,OM,PA,PE,PG,PH,PL,PT,QA,RO,RS,RU,RW,SA,SC,SD,SE,SG,SK,SL,ST,SV,SY,TH,TJ,
TM,TN,TR,TT,TZ,UA,UG,US,UZ,VC,VN,WS,ZA,ZM,ZW

2 フィリップス インターナショナル ビー． ヴィ． インテレクチュアル プロパティー アンド
スタンダーズ

(72)発明者 コッペンス イェルーン ジェラルドゥス ヘンリクス

オランダ国 5 6 5 6 アーヘー アインドーフェン ハイ テック キャンパス 5 2 フィリップス
インターナショナル ビー． ヴィ． インテレクチュアル プロパティー アンド スタンダーズ

F ターム (参考) 5B050 BA09 BA11 CA08 DA01 EA07 EA19 FA02 FA10
5E555 AA76 BA16 BB16 BC01 BE17 DA23 FA00