

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5813542号  
(P5813542)

(45) 発行日 平成27年11月17日(2015.11.17)

(24) 登録日 平成27年10月2日(2015.10.2)

(51) Int.Cl. F I  
**G06T 19/00 (2011.01)** G06T 19/00 600

請求項の数 15 (全 24 頁)

(21) 出願番号	特願2012-60659 (P2012-60659)	(73) 特許権者	392026693 株式会社NTTドコモ
(22) 出願日	平成24年3月16日 (2012.3.16)		東京都千代田区永田町二丁目11番1号
(65) 公開番号	特開2013-196154 (P2013-196154A)	(73) 特許権者	594150176 日本コントロールシステム株式会社
(43) 公開日	平成25年9月30日 (2013.9.30)		東京都渋谷区恵比寿1丁目20番18号
審査請求日	平成26年7月31日 (2014.7.31)	(74) 代理人	110000752 特許業務法人朝日特許事務所
		(72) 発明者	水口 紀子 東京都千代田区永田町二丁目11番1号 株式会社エヌ・ティ・ティ・ドコモ内
		(72) 発明者	内山 健 東京都千代田区永田町二丁目11番1号 株式会社エヌ・ティ・ティ・ドコモ内

最終頁に続く

(54) 【発明の名称】 画像コミュニケーションシステム、AR (Augmented Reality) 動画生成装置及びプログラム

(57) 【特許請求の範囲】

【請求項1】

第1端末と第2端末との接続を管理する接続管理装置と、  
 前記接続管理装置により接続が管理された前記第1端末及び前記第2端末の少なくとも一方に対して、双方で撮影された映像を合成した合成動画データを送信するAR (Augmented Reality) 動画生成装置とを有し、

前記接続管理装置は、

前記第1端末と前記第2端末とを接続する通信回線に関する回線情報を前記AR動画生成装置に供給し、

前記AR動画生成装置は、

前記第1端末及び前記第2端末の一方から、所定のオブジェクトを含む被写体を撮影した第1動画データであって、当該被写体の画素毎の距離情報を含んだ第1動画データを受信する第1受信部と、

前記第1端末及び前記第2端末の他方から、所定の形状のマーカを含む背景を撮影した第2動画データを受信する第2受信部と、

前記第1受信部により受信された第1動画データの前記距離情報に基づき、前記被写体から前記所定のオブジェクトに相当する画素を抽出する抽出部と、

前記第2受信部により受信された第2動画データが表す背景中の前記マーカに基づいて定義される当該背景の3次元の座標系を特定し、前記抽出部により抽出された画素に応じた画像を当該座標系に対応するようにして当該背景に合成した合成動画データを生成する

生成部と、

前記生成部により生成された合成動画データを前記他方の端末に送信する送信部とを備え、

前記接続管理装置により供給された回線情報に応じて、前記合成動画データが表す映像を異ならせる画像コミュニケーションシステム。

【請求項 2】

第 1 端末と第 2 端末との接続を管理する接続管理装置と、

前記接続管理装置により接続が管理された前記第 1 端末及び前記第 2 端末の少なくとも一方に対して、双方で撮影された映像を合成した合成動画データを送信する A R (Augmented Reality) 動画生成装置とを有し、

前記接続管理装置は、

前記第 1 端末と前記第 2 端末とにおけるハードウェア又はソフトウェアのリソースに関するリソース情報を前記 A R 動画生成装置に供給し、

前記 A R 動画生成装置は、

前記第 1 端末及び前記第 2 端末の一方から、所定のオブジェクトを含む被写体を撮影した第 1 動画データであって、当該被写体の画素毎の距離情報を含んだ第 1 動画データを受信する第 1 受信部と、

前記第 1 端末及び前記第 2 端末の他方から、所定の形状のマーカを含む背景を撮影した第 2 動画データを受信する第 2 受信部と、

前記第 1 受信部により受信された第 1 動画データの前記距離情報に基づき、前記被写体から前記所定のオブジェクトに相当する画素を抽出する抽出部と、

前記第 2 受信部により受信された第 2 動画データが表す背景中の前記マーカに基づいて定義される当該背景の 3 次元の座標系を特定し、前記抽出部により抽出された画素に応じた画像を当該座標系に対応するようにして当該背景に合成した合成動画データを生成する生成部と、

前記生成部により生成された合成動画データを前記他方の端末に送信する送信部とを備え、

前記接続管理装置により供給されたリソース情報に応じて、前記合成動画データが表す映像を異ならせる画像コミュニケーションシステム。

【請求項 3】

前記 A R 動画生成装置は、

前記回線情報に基づいて、前記合成動画データによるコミュニケーションが可能であるか否かを前記第 1 端末と前記第 2 端末のそれぞれについて判断する判断部を備え、

前記判断部により前記合成動画データによるコミュニケーションが可能であると判断された端末について、前記合成動画データを生成及び送信する

請求項 1 に記載の画像コミュニケーションシステム。

【請求項 4】

前記 A R 動画生成装置は、

前記リソース情報に基づいて、前記合成動画データによるコミュニケーションが可能であるか否かを前記第 1 端末と前記第 2 端末のそれぞれについて判断する判断部を備え、

前記判断部により前記合成動画データによるコミュニケーションが可能であると判断された端末について、前記合成動画データを生成及び送信する

請求項 2 に記載の画像コミュニケーションシステム。

【請求項 5】

第 1 端末と第 2 端末との接続を管理する接続管理装置と、

前記接続管理装置により接続が管理された前記第 1 端末及び前記第 2 端末の少なくとも一方に対して、双方で撮影された映像を合成した合成動画データを送信する A R (Augmented Reality) 動画生成装置とを有し、

前記接続管理装置は、

前記第 1 端末と前記第 2 端末との間で音声メッセージを送受信する送受信部を備えると

10

20

30

40

50

ともに、前記送受信部により送受信される音声メッセージの音量又は音質の変化を表す情報を前記AR動画生成装置に供給し、

前記AR動画生成装置は、

前記第1端末及び前記第2端末の一方から、所定のオブジェクトを含む被写体を撮影した第1動画データであって、当該被写体の画素毎の距離情報を含んだ第1動画データを受信する第1受信部と、

前記第1端末及び前記第2端末の他方から、所定の形状のマーカを含む背景を撮影した第2動画データを受信する第2受信部と、

前記第1受信部により受信された第1動画データの距離情報に基づき、前記被写体から前記所定のオブジェクトに相当する画素を抽出する抽出部と、

前記第2受信部により受信された第2動画データが表す背景中の前記マーカに基づいて定義される当該背景の3次元の座標系を特定し、前記抽出部により抽出された画素に応じた画像を当該座標系に対応するようにして当該背景に合成した合成動画データを生成する生成部と、

前記生成部により生成された合成動画データを前記他方の端末に送信する送信部とを備え、

前記接続管理装置により供給された音量又は音質の変化を表す情報に応じて、前記合成動画データが表す映像を異ならせる画像コミュニケーションシステム。

#### 【請求項6】

第1端末と第2端末との接続を管理する接続管理装置と、

前記接続管理装置により接続が管理された前記第1端末及び前記第2端末の少なくとも一方に対して、双方で撮影された映像を合成した合成動画データを送信するAR (Augmented Reality) 動画生成装置とを有し、

前記AR動画生成装置は、

前記第1端末及び前記第2端末の一方から、所定のオブジェクトを含む被写体を撮影した第1動画データであって、当該被写体の画素毎の距離情報を含んだ第1動画データを受信する第1受信部と、

前記第1端末及び前記第2端末の他方から、所定の形状のマーカを含む背景を撮影した第2動画データを受信する第2受信部と、

前記第1受信部により受信された第1動画データの距離情報に基づき、前記被写体から前記所定のオブジェクトに相当する画素を抽出する抽出部と、

前記第2受信部により受信された第2動画データが表す背景中の前記マーカに基づいて定義される当該背景の3次元の座標系を特定し、前記抽出部により抽出された画素に応じた画像を当該座標系に対応するようにして当該背景に合成した合成動画データを生成する生成部と、

前記生成部により生成された合成動画データを前記他方の端末に送信する送信部と、

前記第1受信部により受信された第1動画データにおける前記被写体の変化を解析する解析部を備え、

前記解析部により解析された前記被写体の変化を表す情報を前記接続管理装置に供給し

、  
前記接続管理装置は、

前記第1端末と前記第2端末との間で音声メッセージを送受信する送受信部であって、前記AR動画生成装置により供給された前記被写体の変化を表す情報に応じて、送受信する前記音声メッセージの音量又は音質を異ならせる送受信部

を備える画像コミュニケーションシステム。

#### 【請求項7】

前記AR動画生成装置は、

前記第2受信部により受信された第2動画データを解析し、前記座標系を算出する算出部を備え、

前記生成部は、

10

20

30

40

50

前記算出部により算出された座標系に基づいて、前記合成動画データを生成する請求項 1 ないし 6 のいずれか 1 項に記載の画像コミュニケーションシステム。

【請求項 8】

所定のオブジェクトを含む被写体を撮影した第 1 動画データであって、当該被写体の画素毎の距離情報を含んだ第 1 動画データを受信する第 1 受信部と、

所定の形状のマーカを含む背景を撮影した第 2 動画データであって、当該マーカに基づいて 3 次元の座標系が定義される第 2 動画データを受信する第 2 受信部と、

前記第 1 受信部により受信された第 1 動画データの前記距離情報に基づき、前記被写体から前記所定のオブジェクトに相当する画素を抽出する抽出部と、

前記抽出部により抽出された画素に応じた画像を、前記第 2 受信部により受信された第 2 動画データに定義される座標系に対応するようにして当該背景に合成した合成動画データを生成する生成部と、

前記生成部により生成された合成動画データを送信する送信部と

を備え、

前記第 1 動画データを送信した端末と前記第 2 動画データを送信した端末との接続を管理する接続管理装置から、前記第 1 動画データを送信した端末と前記第 2 動画データを送信した端末とを接続する通信回線に関する回線情報が供給され、供給された前記回線情報に応じて、前記合成動画データが表す映像を異ならせる A R 動画生成装置。

【請求項 9】

所定のオブジェクトを含む被写体を撮影した第 1 動画データであって、当該被写体の画素毎の距離情報を含んだ第 1 動画データを受信する第 1 受信部と、

所定の形状のマーカを含む背景を撮影した第 2 動画データであって、当該マーカに基づいて 3 次元の座標系が定義される第 2 動画データを受信する第 2 受信部と、

前記第 1 受信部により受信された第 1 動画データの前記距離情報に基づき、前記被写体から前記所定のオブジェクトに相当する画素を抽出する抽出部と、

前記抽出部により抽出された画素に応じた画像を、前記第 2 受信部により受信された第 2 動画データに定義される座標系に対応するようにして当該背景に合成した合成動画データを生成する生成部と、

前記生成部により生成された合成動画データを送信する送信部と

を備え、

前記第 1 動画データを送信した端末と前記第 2 動画データを送信した端末とを接続する接続管理装置から、前記端末におけるハードウェア又はソフトウェアのリソースに関するリソース情報が供給され、供給された前記リソース情報に応じて、前記合成動画データが表す映像を異ならせる A R 動画生成装置。

【請求項 10】

所定のオブジェクトを含む被写体を撮影した第 1 動画データであって、当該被写体の画素毎の距離情報を含んだ第 1 動画データを受信する第 1 受信部と、

所定の形状のマーカを含む背景を撮影した第 2 動画データであって、当該マーカに基づいて 3 次元の座標系が定義される第 2 動画データを受信する第 2 受信部と、

前記第 1 受信部により受信された第 1 動画データの前記距離情報に基づき、前記被写体から前記所定のオブジェクトに相当する画素を抽出する抽出部と、

前記抽出部により抽出された画素に応じた画像を、前記第 2 受信部により受信された第 2 動画データに定義される座標系に対応するようにして当該背景に合成した合成動画データを生成する生成部と、

前記生成部により生成された合成動画データを送信する送信部と

を備え、

前記第 1 動画データを送信した端末と前記第 2 動画データを送信した端末との間で音声メッセージを送受信する送受信部を備えるとともに、前記送受信部により送受信される音声メッセージの音量又は音質の変化を表す情報を供給する接続管理装置から、前記送受信部により送受信される音声メッセージの音量又は音質の変化を表す情報が供給

された情報に応じて、前記合成動画データが表す映像を異ならせる A R 動画生成装置。

【請求項 1 1】

所定のオブジェクトを含む被写体を撮影した第 1 動画データであって、当該被写体の画素毎の距離情報を含んだ第 1 動画データを受信する第 1 受信部と、

所定の形状のマーカを含む背景を撮影した第 2 動画データであって、当該マーカに基づいて 3 次元の座標系が定義される第 2 動画データを受信する第 2 受信部と、

前記第 1 受信部により受信された第 1 動画データの前記距離情報に基づき、前記被写体から前記所定のオブジェクトに相当する画素を抽出する抽出部と、

前記抽出部により抽出された画素に応じた画像を、前記第 2 受信部により受信された第 2 動画データに定義される座標系に対応するようにして当該背景に合成した合成動画データを生成する生成部と、

前記生成部により生成された合成動画データを送信する送信部と、

前記第 1 受信部により受信された第 1 動画データにおける前記被写体の変化を解析する解析部を備え、

前記第 1 動画データを送信した端末と前記第 2 動画データを送信した端末との間で音声メッセージを送受信する送受信部であって、供給される前記被写体の変化を表す情報に応じて、送受信する前記音声メッセージの音量又は音質を異ならせる送受信部を備える接続管理装置へ、前記解析部により解析された前記被写体の変化を表す情報を供給する

A R 動画生成装置。

【請求項 1 2】

コンピュータに、

所定のオブジェクトを含む被写体を撮影した第 1 動画データであって、当該被写体の画素毎の距離情報を含んだ第 1 動画データを受信する第 1 ステップと、

所定の形状のマーカを含む背景を撮影した第 2 動画データであって、当該マーカに基づいて 3 次元の座標系が定義される第 2 動画データを受信する第 2 ステップと、

前記第 1 ステップにより受信された第 1 動画データの前記距離情報に基づき、前記被写体から前記所定のオブジェクトに相当する画素を抽出する第 3 ステップと、

前記第 3 ステップにより抽出された画素に応じた画像を、前記第 2 ステップにより受信された第 2 動画データに定義される座標系に対応するようにして当該背景に合成した合成動画データを生成する第 4 ステップと、

前記第 4 ステップにより生成された合成動画データを送信する第 5 ステップと、

前記第 1 動画データを送信した端末と前記第 2 動画データを送信した端末との接続を管理する接続管理装置から、前記第 1 動画データを送信した端末と前記第 2 動画データを送信した端末とを接続する通信回線に関する回線情報が供給され、供給された前記回線情報に応じて、前記合成動画データが表す映像を異ならせる第 6 ステップと

を実行させるためのプログラム。

【請求項 1 3】

コンピュータに、

所定のオブジェクトを含む被写体を撮影した第 1 動画データであって、当該被写体の画素毎の距離情報を含んだ第 1 動画データを受信する第 1 ステップと、

所定の形状のマーカを含む背景を撮影した第 2 動画データであって、当該マーカに基づいて 3 次元の座標系が定義される第 2 動画データを受信する第 2 ステップと、

前記第 1 ステップにより受信された第 1 動画データの前記距離情報に基づき、前記被写体から前記所定のオブジェクトに相当する画素を抽出する第 3 ステップと、

前記第 3 ステップにより抽出された画素に応じた画像を、前記第 2 ステップにより受信された第 2 動画データに定義される座標系に対応するようにして当該背景に合成した合成動画データを生成する第 4 ステップと、

前記第 4 ステップにより生成された合成動画データを送信する第 5 ステップと、

前記第 1 動画データを送信した端末と前記第 2 動画データを送信した端末とを接続する接続管理装置から、前記端末におけるハードウェア又はソフトウェアのリソースに関する

10

20

30

40

50

リソース情報が供給され、供給された前記リソース情報に応じて、前記合成動画データが表す映像を異ならせる第6ステップと

を実行させるためのプログラム。

【請求項14】

コンピュータに、

所定のオブジェクトを含む被写体を撮影した第1動画データであって、当該被写体の画素毎の距離情報を含んだ第1動画データを受信する第1ステップと、

所定の形状のマーカを含む背景を撮影した第2動画データであって、当該マーカに基づいて3次元の座標系が定義される第2動画データを受信する第2ステップと、

前記第1ステップにより受信された第1動画データの前記距離情報に基づき、前記被写体から前記所定のオブジェクトに相当する画素を抽出する第3ステップと、

前記第3ステップにより抽出された画素に応じた画像を、前記第2ステップにより受信された第2動画データに定義される座標系に対応するようにして当該背景に合成した合成動画データを生成する第4ステップと、

前記第4ステップにより生成された合成動画データを送信する第5ステップと、

前記第1動画データを送信した端末と前記第2動画データを送信した端末との間で音声メッセージを送受信する送受信部を備えるとともに、前記送受信部により送受信される音声メッセージの音量又は音質の変化を表す情報を供給する接続管理装置から、前記送受信部により送受信される音声メッセージの音量又は音質の変化を表す情報が供給され、供給された情報に応じて、前記合成動画データが表す映像を異ならせる第6ステップと

を実行させるためのプログラム。

【請求項15】

コンピュータに、

所定のオブジェクトを含む被写体を撮影した第1動画データであって、当該被写体の画素毎の距離情報を含んだ第1動画データを受信する第1ステップと、

所定の形状のマーカを含む背景を撮影した第2動画データであって、当該マーカに基づいて3次元の座標系が定義される第2動画データを受信する第2ステップと、

前記第1ステップにより受信された第1動画データの前記距離情報に基づき、前記被写体から前記所定のオブジェクトに相当する画素を抽出する第3ステップと、

前記第3ステップにより抽出された画素に応じた画像を、前記第2ステップにより受信された第2動画データに定義される座標系に対応するようにして当該背景に合成した合成動画データを生成する第4ステップと、

前記第4ステップにより生成された合成動画データを送信する第5ステップと、

前記第1ステップにより受信された第1動画データにおける前記被写体の変化を解析する第6ステップと、

前記第1動画データを送信した端末と前記第2動画データを送信した端末との間で音声メッセージを送受信する送受信部であって、供給される前記被写体の変化を表す情報に応じて、送受信する前記音声メッセージの音量又は音質を異ならせる送受信部を備える接続管理装置へ、前記第6ステップにより解析された前記被写体の変化を表す情報を供給する第7ステップと

を実行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、拡張現実（AR）を用いたコミュニケーションに関する。

【背景技術】

【0002】

AR技術による合成画像を通信ネットワークを介して配信する技術がある（例えば、特許文献1参照）。合成画像を得るための方法としては、クロマキー合成（ブルーバック合成）のように特定色の背景を利用して撮影を行うものや、CG（Computer Graphics）に

10

20

30

40

50

よって得られた実在しない画像（アバタなど）を合成するものなどが知られている。

【先行技術文献】

【特許文献】

【0003】

【特許文献1】特開2004-341642号公報

【発明の概要】

【発明が解決しようとする課題】

【0004】

ところで、AR技術を利用した従来の合成画像は、ユーザが撮影した映像に対して、あらかじめ用意された画像を重ね合わせるものが一般的であった。したがって、いわゆるテレビ電話のような、ユーザ同士の実際の映像（すなわち、あらかじめ用意されたものではない画像）を利用したリアルタイムなコミュニケーションにそのまま適用することはできなかった。

10

そこで、本発明は、相互のユーザによって撮影された映像に基づいて合成された動画を用いたリアルタイムなコミュニケーションを可能にする技術を提供する。

【課題を解決するための手段】

【0005】

本発明の一態様に係る画像コミュニケーションシステムは、第1端末と第2端末との接続を管理する接続管理装置と、前記接続管理装置により接続が管理された前記第1端末及び前記第2端末の少なくとも一方に対して、双方で撮影された映像を合成した合成動画データを送信するAR（Augmented Reality）動画生成装置とを有し、前記接続管理装置は、前記第1端末と前記第2端末とを接続する通信回線に関する回線情報を前記AR動画生成装置に供給し、前記AR動画生成装置は、前記第1端末及び前記第2端末の一方から、所定のオブジェクトを含む被写体を撮影した第1動画データであって、当該被写体の画素毎の距離情報を含んだ第1動画データを受信する第1受信部と、前記第1端末及び前記第2端末の他方から、所定の形状のマーカを含む背景を撮影した第2動画データを受信する第2受信部と、前記第1受信部により受信された第1動画データの前記距離情報に基づき、前記被写体から前記所定のオブジェクトに相当する画素を抽出する抽出部と、前記第2受信部により受信された第2動画データが表す背景中の前記マーカに基づいて定義される当該背景の3次元の座標系を特定し、前記抽出部により抽出された画素に応じた画像を当該座標系に対応するようにして当該背景に合成した合成動画データを生成する生成部と、前記生成部により生成された合成動画データを前記他方の端末に送信する送信部とを備え、前記接続管理装置により供給された回線情報に応じて、前記合成動画データが表す映像を異ならせる。

20

30

好ましい態様において、前記AR動画生成装置は、前記回線情報に基づいて、前記合成動画データによるコミュニケーションが可能であるか否かを前記第1端末と前記第2端末のそれぞれについて判断する判断部を備え、前記判断部により前記合成動画データによるコミュニケーションが可能であると判断された端末について、前記合成動画データを生成及び送信する。

また、本発明の一態様に係る画像コミュニケーションシステムは、第1端末と第2端末との接続を管理する接続管理装置と、前記接続管理装置により接続が管理された前記第1端末及び前記第2端末の少なくとも一方に対して、双方で撮影された映像を合成した合成動画データを送信するAR（Augmented Reality）動画生成装置とを有し、前記接続管理装置は、前記第1端末と前記第2端末とにおけるハードウェア又はソフトウェアのリソースに関するリソース情報を前記AR動画生成装置に供給し、前記AR動画生成装置は、前記第1端末及び前記第2端末の一方から、所定のオブジェクトを含む被写体を撮影した第1動画データであって、当該被写体の画素毎の距離情報を含んだ第1動画データを受信する第1受信部と、前記第1端末及び前記第2端末の他方から、所定の形状のマーカを含む背景を撮影した第2動画データを受信する第2受信部と、前記第1受信部により受信された第1動画データの前記距離情報に基づき、前記被写体から前記所定のオブジェクトに相

40

50

当する画素を抽出する抽出部と、前記第2受信部により受信された第2動画データが表す背景中の前記マーカに基づいて定義される当該背景の3次元の座標系を特定し、前記抽出部により抽出された画素に応じた画像を当該座標系に対応するようにして当該背景に合成した合成動画データを生成する生成部と、前記生成部により生成された合成動画データを前記他方の端末に送信する送信部とを備え、前記接続管理装置により供給されたリソース情報に応じて、前記合成動画データが表す映像を異ならせる。

好ましい態様において、前記AR動画生成装置は、前記リソース情報に基づいて、前記合成動画データによるコミュニケーションが可能であるか否かを前記第1端末と前記第2端末のそれぞれについて判断する判断部を備え、前記判断部により前記合成動画データによるコミュニケーションが可能であると判断された端末について、前記合成動画データを生成及び送信する。

また、本発明の一態様に係る画像コミュニケーションシステムは、第1端末と第2端末との接続を管理する接続管理装置と、前記接続管理装置により接続が管理された前記第1端末及び前記第2端末の少なくとも一方に対して、双方で撮影された映像を合成した合成動画データを送信するAR (Augmented Reality) 動画生成装置とを有し、前記接続管理装置は、前記第1端末と前記第2端末との間で音声メッセージを送受信する送受信部を備え、前記送受信部により送受信される音声メッセージの音量又は音質の変化を表す情報を前記AR動画生成装置に供給し、前記AR動画生成装置は、前記第1端末及び前記第2端末の一方から、所定のオブジェクトを含む被写体を撮影した第1動画データであって、当該被写体の画素毎の距離情報を含んだ第1動画データを受信する第1受信部と、前記第1端末及び前記第2端末の他方から、所定の形状のマーカを含む背景を撮影した第2動画データを受信する第2受信部と、前記第1受信部により受信された第1動画データの前記距離情報に基づき、前記被写体から前記所定のオブジェクトに相当する画素を抽出する抽出部と、前記第2受信部により受信された第2動画データが表す背景中の前記マーカに基づいて定義される当該背景の3次元の座標系を特定し、前記抽出部により抽出された画素に応じた画像を当該座標系に対応するようにして当該背景に合成した合成動画データを生成する生成部と、前記生成部により生成された合成動画データを前記他方の端末に送信する送信部とを備え、前記接続管理装置により供給された音量又は音質の変化を表す情報に応じて、前記合成動画データが表す映像を異ならせる。

また、本発明の一態様に係る画像コミュニケーションシステムは、第1端末と第2端末との接続を管理する接続管理装置と、前記接続管理装置により接続が管理された前記第1端末及び前記第2端末の少なくとも一方に対して、双方で撮影された映像を合成した合成動画データを送信するAR (Augmented Reality) 動画生成装置とを有し、前記AR動画生成装置は、前記第1端末及び前記第2端末の一方から、所定のオブジェクトを含む被写体を撮影した第1動画データであって、当該被写体の画素毎の距離情報を含んだ第1動画データを受信する第1受信部と、前記第1端末及び前記第2端末の他方から、所定の形状のマーカを含む背景を撮影した第2動画データを受信する第2受信部と、前記第1受信部により受信された第1動画データの前記距離情報に基づき、前記被写体から前記所定のオブジェクトに相当する画素を抽出する抽出部と、前記第2受信部により受信された第2動画データが表す背景中の前記マーカに基づいて定義される当該背景の3次元の座標系を特定し、前記抽出部により抽出された画素に応じた画像を当該座標系に対応するようにして当該背景に合成した合成動画データを生成する生成部と、前記生成部により生成された合成動画データを前記他方の端末に送信する送信部と、前記第1受信部により受信された第1動画データにおける前記被写体の変化を解析する解析部を備え、前記解析部により解析された前記被写体の変化を表す情報を前記接続管理装置に供給し、前記接続管理装置は、前記第1端末と前記第2端末との間で音声メッセージを送受信する送受信部であって、前記AR動画生成装置により供給された前記被写体の変化を表す情報に応じて、送受信する前記音声メッセージの音量又は音質を異ならせる送受信部を備える。

【0006】

好ましい態様において、前記AR動画生成装置は、前記第2受信部により受信された第

10

20

30

40

50

2 動画データを解析し、前記座標系を算出する算出部を備え、前記生成部は、前記算出部により算出された座標系に基づいて、前記合成動画データを生成する。

【0007】

本発明の他の態様に係るAR動画生成装置は、所定のオブジェクトを含む被写体を撮影した第1動画データであって、当該被写体の画素毎の距離情報を含んだ第1動画データを受信する第1受信部と、所定の形状のマーカを含む背景を撮影した第2動画データであって、当該マーカに基づいて3次元の座標系が定義される第2動画データを受信する第2受信部と、前記第1受信部により受信された第1動画データの前記距離情報に基づき、前記被写体から前記所定のオブジェクトに相当する画素を抽出する抽出部と、前記抽出部により抽出された画素に応じた画像を、前記第2受信部により受信された第2動画データに定義される座標系に対応するようにして当該背景に合成した合成動画データを生成する生成部と、前記生成部により生成された合成動画データを送信する送信部とを備え、前記第1動画データを送信した端末と前記第2動画データを送信した端末との接続を管理する接続管理装置から、前記第1動画データを送信した端末と前記第2動画データを送信した端末とを接続する通信回線に関する回線情報が供給され、供給された前記回線情報に応じて、前記合成動画データが表す映像を異ならせる。

10

また、本発明の他の態様に係るAR動画生成装置は、所定のオブジェクトを含む被写体を撮影した第1動画データであって、当該被写体の画素毎の距離情報を含んだ第1動画データを受信する第1受信部と、所定の形状のマーカを含む背景を撮影した第2動画データであって、当該マーカに基づいて3次元の座標系が定義される第2動画データを受信する第2受信部と、前記第1受信部により受信された第1動画データの前記距離情報に基づき、前記被写体から前記所定のオブジェクトに相当する画素を抽出する抽出部と、前記抽出部により抽出された画素に応じた画像を、前記第2受信部により受信された第2動画データに定義される座標系に対応するようにして当該背景に合成した合成動画データを生成する生成部と、前記生成部により生成された合成動画データを送信する送信部とを備え、前記第1動画データを送信した端末と前記第2動画データを送信した端末とを接続する接続管理装置から、前記端末におけるハードウェア又はソフトウェアのリソースに関するリソース情報が供給され、供給された前記リソース情報に応じて、前記合成動画データが表す映像を異ならせる。

20

また、本発明の他の態様に係るAR動画生成装置は、所定のオブジェクトを含む被写体を撮影した第1動画データであって、当該被写体の画素毎の距離情報を含んだ第1動画データを受信する第1受信部と、所定の形状のマーカを含む背景を撮影した第2動画データであって、当該マーカに基づいて3次元の座標系が定義される第2動画データを受信する第2受信部と、前記第1受信部により受信された第1動画データの前記距離情報に基づき、前記被写体から前記所定のオブジェクトに相当する画素を抽出する抽出部と、前記抽出部により抽出された画素に応じた画像を、前記第2受信部により受信された第2動画データに定義される座標系に対応するようにして当該背景に合成した合成動画データを生成する生成部と、前記生成部により生成された合成動画データを送信する送信部とを備え、前記第1動画データを送信した端末と前記第2動画データを送信した端末との間で音声メッセージを送受信する送受信部を備え、前記送受信部により送受信される音声メッセージの音量又は音質の変化を表す情報を供給する接続管理装置から、前記送受信部により送受信される音声メッセージの音量又は音質の変化を表す情報が供給され、供給された情報に応じて、前記合成動画データが表す映像を異ならせる。

30

40

また、本発明の他の態様に係るAR動画生成装置は、所定のオブジェクトを含む被写体を撮影した第1動画データであって、当該被写体の画素毎の距離情報を含んだ第1動画データを受信する第1受信部と、所定の形状のマーカを含む背景を撮影した第2動画データであって、当該マーカに基づいて3次元の座標系が定義される第2動画データを受信する第2受信部と、前記第1受信部により受信された第1動画データの前記距離情報に基づき、前記被写体から前記所定のオブジェクトに相当する画素を抽出する抽出部と、前記抽出部により抽出された画素に応じた画像を、前記第2受信部により受信された第2動画デー

50

々に定義される座標系に対応するようにして当該背景に合成した合成動画データを生成する生成部と、前記生成部により生成された合成動画データを送信する送信部と、前記第1受信部により受信された第1動画データにおける前記被写体の変化を解析する解析部を備え、前記第1動画データを送信した端末と前記第2動画データを送信した端末との間で音声メッセージを送受信する送受信部であって、供給される前記被写体の変化を表す情報に応じて、送受信する前記音声メッセージの音量又は音質を異ならせる送受信部を備える接続管理装置へ、前記解析部により解析された前記被写体の変化を表す情報を供給する。

【0008】

本発明の他の態様に係るプログラムは、コンピュータに、所定のオブジェクトを含む被写体を撮影した第1動画データであって、当該被写体の画素毎の距離情報を含んだ第1動画データを受信する第1ステップと、所定の形状のマーカを含む背景を撮影した第2動画データであって、当該マーカに基づいて3次元の座標系が定義される第2動画データを受信する第2ステップと、前記第1ステップにより受信された第1動画データの前記距離情報に基づき、前記被写体から前記所定のオブジェクトに相当する画素を抽出する第3ステップと、前記第3ステップにより抽出された画素に応じた画像を、前記第2ステップにより受信された第2動画データに定義される座標系に対応するようにして当該背景に合成した合成動画データを生成する第4ステップと、前記第4ステップにより生成された合成動画データを送信する第5ステップと、前記第1動画データを送信した端末と前記第2動画データを送信した端末との接続を管理する接続管理装置から、前記第1動画データを送信した端末と前記第2動画データを送信した端末とを接続する通信回線に関する回線情報が供給され、供給された前記回線情報に応じて、前記合成動画データが表す映像を異ならせる第6ステップとを実行させるためのプログラムである。

また、本発明の他の態様に係るプログラムは、コンピュータに、所定のオブジェクトを含む被写体を撮影した第1動画データであって、当該被写体の画素毎の距離情報を含んだ第1動画データを受信する第1ステップと、所定の形状のマーカを含む背景を撮影した第2動画データであって、当該マーカに基づいて3次元の座標系が定義される第2動画データを受信する第2ステップと、前記第1ステップにより受信された第1動画データの前記距離情報に基づき、前記被写体から前記所定のオブジェクトに相当する画素を抽出する第3ステップと、前記第3ステップにより抽出された画素に応じた画像を、前記第2ステップにより受信された第2動画データに定義される座標系に対応するようにして当該背景に合成した合成動画データを生成する第4ステップと、前記第4ステップにより生成された合成動画データを送信する第5ステップと、前記第1動画データを送信した端末と前記第2動画データを送信した端末とを接続する接続管理装置から、前記端末におけるハードウェア又はソフトウェアのリソースに関するリソース情報が供給され、供給された前記リソース情報に応じて、前記合成動画データが表す映像を異ならせる第6ステップとを実行させるためのプログラムである。

また、本発明の他の態様に係るプログラムは、コンピュータに、所定のオブジェクトを含む被写体を撮影した第1動画データであって、当該被写体の画素毎の距離情報を含んだ第1動画データを受信する第1ステップと、所定の形状のマーカを含む背景を撮影した第2動画データであって、当該マーカに基づいて3次元の座標系が定義される第2動画データを受信する第2ステップと、前記第1ステップにより受信された第1動画データの前記距離情報に基づき、前記被写体から前記所定のオブジェクトに相当する画素を抽出する第3ステップと、前記第3ステップにより抽出された画素に応じた画像を、前記第2ステップにより受信された第2動画データに定義される座標系に対応するようにして当該背景に合成した合成動画データを生成する第4ステップと、前記第4ステップにより生成された合成動画データを送信する第5ステップと、前記第1動画データを送信した端末と前記第2動画データを送信した端末との間で音声メッセージを送受信する送受信部を備えるとともに、前記送受信部により送受信される音声メッセージの音量又は音質の変化を表す情報を供給する接続管理装置から、前記送受信部により送受信される音声メッセージの音量又は音質の変化を表す情報が供給され、供給された情報に応じて、前記合成動画データが表

10

20

30

40

50

す映像を異ならせる第6ステップとを実行させるためのプログラムである。

また、本発明の他の態様に係るプログラムは、コンピュータに、所定のオブジェクトを含む被写体を撮影した第1動画データであって、当該被写体の画素毎の距離情報を含んだ第1動画データを受信する第1ステップと、所定の形状のマーカーを含む背景を撮影した第2動画データであって、当該マーカーに基づいて3次元の座標系が定義される第2動画データを受信する第2ステップと、前記第1ステップにより受信された第1動画データの前記距離情報に基づき、前記被写体から前記所定のオブジェクトに相当する画素を抽出する第3ステップと、前記第3ステップにより抽出された画素に応じた画像を、前記第2ステップにより受信された第2動画データに定義される座標系に対応するようにして当該背景に合成した合成動画データを生成する第4ステップと、前記第4ステップにより生成された合成動画データを送信する第5ステップと、前記第1ステップにより受信された第1動画データにおける前記被写体の変化を解析する第6ステップと、前記第1動画データを送信した端末と前記第2動画データを送信した端末との間で音声メッセージを送受信する送受信部であって、供給される前記被写体の変化を表す情報に応じて、送受信する前記音声メッセージの音量又は音質を異ならせる送受信部を備える接続管理装置へ、前記第6ステップにより解析された前記被写体の変化を表す情報を供給する第7ステップとを実行させるためのプログラムである。

10

【発明の効果】

【0009】

本発明によれば、相互のユーザによって撮影された映像に基づいて合成された動画を用いたリアルタイムなコミュニケーションが可能である。

20

【図面の簡単な説明】

【0010】

【図1】通信システムの全体構成を示すブロック図

【図2】ユーザ端末のハードウェア構成を示すブロック図

【図3】AR動画生成装置及び接続管理装置のハードウェア構成を示すブロック図

【図4】AR動画生成装置及び接続管理装置の機能構成を示すブロック図

【図5】動画データが表す映像を説明するための模式図

【図6】オブジェクトの抽出原理を説明するための図

【図7】AR動画生成装置が実行する処理を示すフローチャート

30

【図8】通信システムの各装置における処理を示すシーケンスチャート

【発明を実施するための形態】

【0011】

[実施形態]

図1は、本発明の一実施形態である通信システム10の全体構成を示すブロック図である。通信システム10は、ユーザ間で音声とAR動画によるリアルタイムなコミュニケーションを実現するための情報処理システムであり、本発明に係る画像コミュニケーションシステムの一例である。ここにおいて、AR動画とは、複数のユーザによって撮影された映像を互いに合成した動画をいい、AR技術を用いて生成されるものである。なお、ここでいうリアルタイムとは、目的とする処理があらかじめ決められた時間までに終了することを意味し、当該時間が比較的短時間であるものを意味する。通信システム10は、AR動画生成装置100と、接続管理装置200と、ユーザ端末300と、ネットワーク400とを備える。

40

【0012】

AR動画生成装置100は、複数のユーザ端末300から動画データを受信し、これらを合成した合成動画データを生成及び送信するサーバ装置である。接続管理装置200は、ユーザ端末300間の接続(ここでは呼接続)を管理し、発話や終話を制御するサーバ装置である。接続管理装置200による管理には、コミュニケーションのリアルタイム性を確保するための処理が含まれる。なお、AR動画生成装置100と接続管理装置200とは、ネットワーク400によらずに、これとは異なる通信回線で互いに接続されていて

50

もよい。

【0013】

ユーザ端末300は、通信システム10によってコミュニケーションを行うユーザが使用する通信端末である。ユーザ端末300は、通信システム10のための専用の通信端末であってもよいし、本実施形態のコミュニケーションを行うために必要な周辺機器をスマートフォンやパーソナルコンピュータに取り付けたものであってもよい。なお、ユーザ端末300は、実際にはネットワーク400に多数接続されていてもよいが、図1ではコミュニケーションを行う二者の端末のみが図示されている。以下においては、これらの端末を区別する必要がある場合には、「ユーザ端末300a」、「ユーザ端末300b」と表記するものとする。

10

【0014】

ネットワーク400は、音声や動画などのデータを送受信するためのネットワークである。ネットワーク400は、音声を送受信するための通信回線と動画を送受信するための通信回線を別異に構成したものであってもよいが、そうでなくともよい。また、ネットワーク400は、パケット交換方式と回線交換方式のいずれの通信回線のいずれを用いたものであってもよい。

【0015】

図2は、ユーザ端末300のハードウェア構成を示すブロック図である。ユーザ端末300は、制御部310と、記憶部320と、通信部330と、第1撮影部340と、第2撮影部350と、表示部360と、音声入出力部370と、操作部380とを備える。図2に示す構成のことを、以下においては「標準構成」という。

20

【0016】

制御部310は、ユーザ端末300の動作を制御する手段である。制御部310は、CPU (Central Processing Unit) などの演算処理装置やメモリを備え、所定のプログラムを実行することによって動画の撮影やデータ通信などを制御する。また、制御部310は、後述する3次元の座標系を算出する算出部311を備える。なお、算出部311による座標系の算出方法は、周知の方法と同様であってよい。例えば、算出部311は、撮影された映像の中から所定の形状のマーカを認識し、当該認識された形状と実際の形状の相違(すなわち歪み)に基づいて傾きを特定して傾き情報を算出するとともに、特定した傾きとマーカの位置とに基づいて3次元の直交座標系を定義し、座標を表す座標情報を算出する。

30

【0017】

記憶部320は、データを記憶する手段である。記憶部320は、補助記憶装置に相当し、例えば、ハードディスクやフラッシュメモリを含んで構成される。記憶部320は、ユーザ端末300(自装置)のリソースに関するリソース情報を記憶している。ここにおいて、リソース情報とは、ユーザ端末300がどのようなリソースを有しているかを示すデータをいう。ここでいうリソースは、ハードウェア的なリソースとソフトウェア的なリソースの双方を含み得る。リソース情報は、例えば、ユーザ端末300の識別情報(機種名など)、CPUの処理能力、第1撮影部340や第2撮影部350の性能(画素数など)、表示部360の性能(表示解像度など)、対応しているコーデック、コミュニケーションに必要なソフトウェアのバージョン情報などを示す。

40

【0018】

通信部330は、ネットワーク400を介してデータを送受信する手段である。通信部330による通信は、無線・有線のいずれであってもよい。また、通信部330は、ユーザ端末300の一部が外付けの周辺機器によって構成される場合に当該周辺機器と通信を行う構成を含んでもよい。

【0019】

第1撮影部340及び第2撮影部350は、いずれも被写体を撮影して動画データを出力する手段である。第1撮影部340及び第2撮影部350により出力される動画データは、被写体の各画素の色を表す色情報を少なくとも含んでいる。また、第1撮影部340

50

は、ユーザを含む被写体を撮影するために用いられ、第2撮影部350は、マーカ（ARマーカ）を含む被写体を撮影するために用いられる。さらに、第1撮影部340は、距離画像センサを含んで構成される。一方、第2撮影部350は、通常のイメージセンサ、すなわち距離情報を出力しないイメージセンサであってもよい。

【0020】

距離画像センサとは、各画素の距離情報を出力することができるイメージセンサである。距離画像センサには、あらかじめ決められたパターンの光を照射し、その反射光の歪みによってセンサと被写体との距離を測定する「パターン照射方式」と、被写体に対して光を照射し、光が被写体に反射して戻ってくるまでの所要時間によってセンサと被写体との距離を測定する「TOF（Time of Flight）方式」とがあるが、第1撮影部340にはい

10

【0021】

第1撮影部340は、このような距離画像センサと通常のイメージセンサとを備え、色情報と距離情報とを出力する。距離画像センサとイメージセンサは、同じ方向を撮影し、距離画像センサの画素とイメージセンサの画素との対応付けがあらかじめ得られているものとする。

【0022】

表示部360は、画像を表示する手段である。表示部360は、ユーザ端末300に備わる液晶等のディスプレイであってもよいし、ユーザ端末300とは別体のテレビのようなディスプレイであってもよい。また、表示部360は、HMD（Head Mounted Display）のようにユーザに装着される構成であってもよい。

20

【0023】

音声入出力部370は、音声を入力及び出力する手段である。音声入出力部370は、スピーカやマイクロホンを備える。音声入出力部370は、音声の入出力をステレオ・モノラルのいずれで行ってもよい。

操作部380は、ユーザの操作を受け付ける手段である。操作部380は、キーボード（キーボード）や、表示部360のディスプレイに重ねて設けられるタッチスクリーンを備える。

【0024】

なお、ユーザ端末300は、これらの構成を全て備えるものを標準構成とするが、後述するように、当該構成の一部を有しない端末であっても、本実施形態のコミュニケーションを（制限付きで）行うことが可能である。また、ユーザ端末300は、表示解像度などの個々のリソースが一致していなくてもよい。

30

【0025】

図3は、AR動画生成装置100及び接続管理装置200のハードウェア構成を示すブロック図である。AR動画生成装置100は、制御部110と、記憶部120と、通信部130とを備える。また、接続管理装置200は、制御部210と、記憶部220と、通信部130とを備える。制御部110、120は、自装置（AR動画生成装置100又は接続管理装置200）の動作を制御する手段である。また、記憶部120、220は、データを記憶する手段であり、通信部130、230は、データを送受信する手段である。通信部130、230は、ネットワーク400と通信するほか、ネットワーク400を介さずに互いに通信することも可能である。

40

【0026】

図4は、AR動画生成装置100及び接続管理装置200の機能構成を示すブロック図である。AR動画生成装置100及び接続管理装置200は、所定のプログラムを実行することにより、図4に示す機能を実現させる。接続管理装置200の制御部210は、送受信部211及びデータ授受部212として機能する。また、AR動画生成装置100の制御部110は、合成制御部110a及び同期制御部110bとして機能する。

【0027】

送受信部211は、音声メッセージを送受信する手段である。送受信部211は、ユー

50

ザ端末300aから送信された音声メッセージをユーザ端末300bに送信し、ユーザ端末300bから送信された音声メッセージをユーザ端末300aに送信する。また、送受信部211は、音声メッセージを送受信するために必要な接続管理を行い、必要に応じて、音量や音質の制御を行う。

【0028】

データ授受部212は、AR動画生成装置100との間でデータを授受する手段である。データ授受部212は、AR動画生成装置100からリソース情報を取得し、AR動画生成装置100に回線情報を供給する。回線情報は、ユーザ端末300a、300bのそれぞれが使用している通信回線に関する情報であって、通信回線の伝達能力(転送帯域など)を示す。回線情報は、ユーザの通信事業者との契約内容によって異なる場合もあるし、通信回線の実際の利用状況(混雑の度合い)などによっても異なり得る。

10

【0029】

合成制御部110aは、映像の合成を制御する手段である。合成制御部110aは、より詳細には、第1受信部111a、111bと、第2受信部112a、112bと、抽出部113a、113bと、生成部114a、114bと、送信部115a、115bとを備える。

【0030】

なお、第1受信部111a、第2受信部112a、抽出部113a、生成部114a及び送信部115aは、それぞれ、ユーザ端末300aに合成動画データを送信するための構成である。一方、第1受信部111b、第2受信部112b、抽出部113b、生成部114b及び送信部115bは、それぞれ、ユーザ端末300bに合成動画データを送信するための構成である。例えば、第1受信部111aと第1受信部111bとは、動画データを受信する端末が異なるだけであって、実行する動作自体には相違がない。同様に、第2受信部112a、112b、抽出部113a、113b、生成部114a、114b及び送信部115a、115bの各組み合わせも、処理対象とするデータが異なるだけで、実行可能な処理には相違がないものである。

20

【0031】

第1受信部111a、111bは、第1撮影部340により撮影された動画データを受信する手段である。すなわち、第1受信部111a、111bは、通信部130を介して受信されるデータのうち、第1撮影部340により撮影された動画データを選択的に取得する。第1受信部111aは、この動画データをユーザ端末300bから取得し、第1受信部111bは、この動画データをユーザ端末300aから取得する。第1受信部111a、111bが受信する動画データは、距離情報を含む動画データである。

30

【0032】

第2受信部112a、112bは、第2撮影部350により撮影された動画データを受信する手段である。すなわち、第2受信部112a、112bは、通信部130を介して受信されるデータのうち、第2撮影部350により撮影された動画データを選択的に取得する。第2受信部112aは、この動画データをユーザ端末300aから取得し、第2受信部112bは、この動画データをユーザ端末300bから取得する。第2受信部112a、112bが受信する動画データは、その映像中にマーカを少なくとも含み、当該マーカによって定義される3次元の座標系の情報を含む動画データである。

40

【0033】

以下においては、説明の便宜上、第1撮影部340により撮影された動画データを「第1動画データ」といい、第2撮影部350により撮影された動画データを「第2動画データ」という。つまり、第1動画データは距離情報を含み、第2動画データは(座標系を特定可能な)マーカの画像を含む。

【0034】

抽出部113a、113bは、第1受信部111a、111bにより受信された第1動画データから、所定のオブジェクトに相当する画素を抽出する手段である。抽出部113a、113bは、第1動画データの距離情報に基づき、センサとの距離が所定の条件を満

50

たす画素を抽出し、当該画素の色情報と距離情報を特定する。ここにおいて、所定の条件は、センサとの距離について設定された閾値によって定まる。閾値は、距離の上限のみが設定されていてもよいし、上限と下限の双方が設定されていてもよい。

**【0035】**

生成部114a、114bは、第1受信部111a、111bにより受信された第1動画データと第2受信部112a、112bにより受信された第2動画データとに基づいて、合成動画データを生成する手段である。生成部114a、114bは、第2受信部112a、112bにより受信された第2動画データが表す映像から背景中にあるマーカを検出し、検出したマーカに基づいて当該背景の3次元の座標系を特定し、抽出部113a、113bにより抽出された画素に応じた画像を当該座標系に対応するようにして当該背景に合成する。本実施形態において、生成部114a、114bは、ユーザ端末300の算出部311により算出された座標系を用いて合成動画データを生成する。

10

**【0036】**

生成部114a、114bは、第1動画データが表す映像の座標系と第2動画データが表す映像の座標系を対応するようにして合成動画データを生成する。具体的には、生成部114a、114bは、第1動画データについて、画素の配列に基づいてx軸とy軸を定義し、距離情報の方向にz軸を定義すると、3次元の直交座標系を定義することができる。そして、生成部114a、114bは、第2動画データについてマーカによって定義される座標系と第1動画データに定義された座標系とを対応付けるようにしてこれらの映像を合成する。生成部114a、114bは、いったん両者の座標系を対応付けたら、その後はその対応付けが維持されるようにして映像の合成を続ける。このとき、生成部114a、114bは、一方の座標系のある座標と他方の座標系のある座標とが一致するように合成を行うなどして、これらの映像の位置合わせを行う。

20

**【0037】**

また、生成部114a、114bは、これらの映像を合成するとき、タイミングの調整やエフェクト等の画像処理を行ってもよい。例えば、生成部114a、114bは、第1動画データと第2動画データの受信タイミングに時間差がある場合に、その時間差が小さくなるように合成時のタイミングを調整する。また、ここでいう画像処理は、映像に対して何らかの画像を重畳する処理であってもよいし、合成する2つの映像の明るさや色合いを合わせる処理であってもよい。

30

**【0038】**

送信部115a、115bは、生成部114a、114bにより生成された合成動画データをユーザ端末300a、300bに送信する手段である。送信部115aは、合成動画データをユーザ端末300aに送信し、送信部115bは、合成動画データをユーザ端末300bに送信する。このようにすることで、ユーザ端末300a、300bのユーザは、自身が送信した第2動画データ(マーカを撮影した動画データ)に対して相手方のユーザのオブジェクト(顔、上半身など)が合成された動画を見ることができるようになる。

**【0039】**

同期制御部110bは、合成動画データの再生と音声メッセージの再生とを同期させるための処理を実行する手段である。例えば、同期制御部110bは、回線情報の変化に追従するように、合成動画データの圧縮方式を異ならせることができる。また、同期制御部110bは、送受信部211が音声メッセージを合成動画データに合わせて再生するために必要なデータを接続管理装置200に供給することもできる。同期制御部110bは、より詳細には、データ授受部116と、判断部117と、解析部118とを含んで構成される。

40

**【0040】**

データ授受部116は、接続管理装置200との間でデータを授受する手段である。データ授受部116は、接続管理装置200から回線情報を取得し、接続管理装置200にリソース情報を供給する。

50

## 【 0 0 4 1 】

判断部 1 1 7 は、回線情報又はリソース情報に基づいて、ユーザ端末 3 0 0 a とユーザ端末 3 0 0 b とが合成動画データによるコミュニケーションを行うことができるか否かを端末毎に判断する手段である。例えば、合成制御部 1 1 0 a は、ユーザ端末 3 0 0 a とユーザ端末 3 0 0 b の一方の通信回線が合成動画データのリアルタイムな再生に必要な品質を満たしていない場合には、当該一方についての合成動画データを生成せずに、他方の合成動画データのみを生成及び送信する、といった処理を、判断部 1 1 7 による判断結果に基づいて行うことができる。また、合成制御部 1 1 0 a は、リソース情報による判断結果に基づき、ユーザ端末 3 0 0 が第 1 撮影部 3 4 0 を備えていない場合には合成動画データの生成を行わないようにすることも可能である。

10

## 【 0 0 4 2 】

解析部 1 1 8 は、第 1 受信部 1 1 1 a、1 1 1 b により受信された第 1 動画データにおける被写体の変化を解析し、その解析結果を表す情報を生成する手段である。同期制御部 1 1 0 b は、解析部 1 1 8 により生成された情報を接続管理装置 2 0 0 に供給する。解析部 1 1 8 は、例えば、被写体であるユーザの移動を解析する。あるいは、解析部 1 1 8 は、周知の顔認識技術（例えば、笑顔を検出する技術）を用いて、被写体であるユーザの表情の変化を解析してもよい。また、解析部 1 1 8 は、ユーザの変化に限らず、被写体全体の変化（例えば、明るさの変化）を解析により求めてもよい。

## 【 0 0 4 3 】

通信システム 1 0 の構成は、以上のとおりである。この構成のもと、ユーザは、ユーザ端末 3 0 0 を用いて他のユーザとリアルタイムなコミュニケーションを行う。ここでいうコミュニケーションは、音声と映像とを用いたものであり、例えば、いわゆるテレビ電話のようなものである。ただし、本実施形態のコミュニケーションは、AR 技術によって合成された映像を利用可能である点において従来のテレビ電話と異なっている。

20

## 【 0 0 4 4 】

このようなコミュニケーションを実現するために、ユーザ端末 3 0 0 においては、被写体が異なる 2 種類の動画データが撮影及び送信される。動画データの一つは、上述した第 1 動画データであり、ユーザが自身を撮影して得られるものである。また、もう一つの動画データは、上述した第 2 動画データであり、ユーザがマーカを撮影して得られるものである。

30

## 【 0 0 4 5 】

AR 動画生成装置 1 0 0 は、これらの動画データを受信し、合成動画データを生成及び送信する。AR 動画生成装置 1 0 0 は、一方のユーザ端末 3 0 0 から送信され、ユーザを被写体に含む第 2 動画データと、他方のユーザ端末 3 0 0 から送信され、マーカを被写体に含む第 1 動画データとを用いて合成動画データを生成し、これを当該一方のユーザ端末 3 0 0 に送信する。すなわち、AR 動画生成装置 1 0 0 は、マーカを映した動画のあるユーザから受信し、これを背景に用いて、他のユーザを映した映像をここに合成する。このとき、AR 動画生成装置 1 0 0 は、第 2 動画データの映像からユーザに相当する所定のオブジェクトを抽出し、これを背景に重ね合わせる。

## 【 0 0 4 6 】

40

図 5 は、本実施形態において送受信される動画データが表す映像を説明するための模式図である。ここにおいて、映像 V 1 a、V 2 a、V 3 a は、それぞれ、ユーザ端末 3 0 0 a により送受信される動画データが表す映像であり、映像 V 1 b、V 2 b、V 3 b は、それぞれ、ユーザ端末 3 0 0 b により送受信される動画データが表す映像であるとする。また、映像 V 1 a、V 1 b が第 1 動画データ、映像 V 2 a、V 2 b が第 2 動画データ、映像 V 3 a、V 3 b が合成動画データに、それぞれ相当する。

## 【 0 0 4 7 】

なお、図 5 において、M a、M b は、それぞれマーカを示している。マーカ M a、M b は、あらかじめ決められた形状の画像を印刷等によって形成した小片であり、ユーザが好きな場所に配置することが可能である。マーカ M a、M b に形成された画像は、他の背景

50

部分との識別が容易であり、かつ、座標系や傾きの特定が容易なように構成されていれば、どのようなパターンの画像であってもよい。

【0048】

図5に示すように、ユーザ端末300aにおいて表示される映像V3aは、ユーザ端末300bからの第1動画データ(V1b)とユーザ端末300aからの第2動画データ(V2a)とを合成して得られる映像である。一方、ユーザ端末300bにおいて表示される映像V3bは、ユーザ端末300aからの第1動画データ(V1a)とユーザ端末300bからの第2動画データ(V2b)とを合成して得られる映像である。

【0049】

なお、図5に示すように、映像V3a、V3bに合成されるオブジェクト(この場合、ユーザの上半身)は、映像V1a、V1bに含まれるオブジェクトと等しい倍率である必要はなく、拡大・縮小などを適宜に行ってもよい。また、映像V3a、V3bに合成されるオブジェクトは、距離情報に基づいて凹凸が表現され、色情報に基づいて着色された立体的な画像(ポリゴン)であるとするが、平面的な画像であってもよい。

【0050】

図6は、オブジェクトの抽出原理を説明するための図である。距離画像センサによって得られる距離情報は、図中の破線の矢印によって示すように、各画素について得られる。ここにおいて、図中のThで示した閾値を設定し、距離情報により表される距離がこの閾値よりも小さい画素を抽出するようにすれば、被写体のうちのユーザに相当する部分の画素が選択され、ユーザ以外の部分(ユーザの背後の壁など)が除外されたオブジェクトが得られる。

【0051】

なお、この閾値は、あらかじめ設定されており、例えばユーザが距離画像センサから1m以内の場所で会話するようにしてもよいし、複数の選択肢の中からユーザが選択できるようになっているもよい。あるいは、閾値は、ユーザ端末300によって動的に設定されてもよい。例えば、ユーザ端末300は、色情報に基づいてユーザの位置(顔など)を推定し、この推定結果に基づいて閾値を設定することも可能である。

【0052】

ユーザは、このようにして合成された映像を見ながら相手と会話することで、あたかも自分のそば(背景として撮影している位置)に相手がいるような感覚でコミュニケーションを行うことができる。これにより、ユーザは、リアリティのあるコミュニケーションをリアルタイムに行うことが可能になる。

【0053】

通信システム10においては、このような映像をユーザ端末300に表示可能にするために、AR動画生成装置100が合成動画データを生成及び送信する。AR動画生成装置100は、リソース情報や回線情報を参照し、必要な合成動画データを生成する。例えば、AR動画生成装置100は、ユーザ端末300のリソースや通信回線の状況に応じて、合成動画データの圧縮方式や転送レートを決定する。また、AR動画生成装置100は、リソース情報や回線情報に基づいて、合成動画データを生成するか否かを判断することも可能である。例えば、AR動画生成装置100は、合成動画データの送信に十分な転送帯域が確保されていない場合や、そもそもユーザ端末300が合成動画データを再生する機能を有しない場合などには、合成動画データを送信しないか、合成動画データよりもデータ量が少ない代替的なデータを送信する、といったことが可能である。

【0054】

図7は、AR動画生成装置100が合成動画データを生成するときに行う処理を示すフローチャートである。AR動画生成装置100の制御部110は、まず最初に、リソース情報や回線情報の少なくともいずれかを用いて、合成動画データの生成態様を判定する(ステップSa1)。このとき、制御部110は、合成動画データの生成の可否や、合成動画データを生成するときの圧縮方式や転送レートなどを、合成動画データの送信対象であるユーザ端末300のそれぞれについて判定する。

## 【 0 0 5 5 】

その後、制御部 1 1 0 は、ステップ S a 1 の判定結果に応じた処理を実行する。制御部 1 1 0 は、一方のユーザ端末 3 0 0 であるユーザ端末 3 0 0 a について、合成動画データを生成するか否かを判断し（ステップ S a 2）、合成動画データを生成すると判定した場合には合成動画データを生成する一方、そうでなければ生成を行わない（ステップ S a 3）。また、合成動画データを生成する場合、制御部 1 1 0 は、ステップ S a 1 の判定結果に応じて（すなわち回線情報やリソース情報に応じて）、合成動画データの圧縮方式や転送レートなどを異ならせる。

## 【 0 0 5 6 】

続いて、制御部 1 1 0 は、他方のユーザ端末 3 0 0 であるユーザ端末 3 0 0 b についても同様に、合成動画データを生成するか否かの判断（ステップ S a 4）と合成動画データの生成（ステップ S a 5）とを実行する。そして、制御部 1 1 0 は、生成した合成動画データを通信部 1 3 0 に供給することにより、合成動画データをユーザ端末 3 0 0 a、3 0 0 b のそれぞれに送信する（ステップ S a 6）。

10

## 【 0 0 5 7 】

A R 動画生成装置 1 0 0 は、このような処理を実行することで、ユーザ端末 3 0 0 a とユーザ端末 3 0 0 b のそれぞれに応じた品質の合成動画データを生成することが可能である。また、A R 動画生成装置 1 0 0 は、送信する必要がない合成動画データの生成を省略することが可能であり、自装置のリソースを効率的に使用して処理を進めることができる。

20

## 【 0 0 5 8 】

図 8 は、通信システム 1 0 の各装置における処理を示すシーケンスチャートである。なお、図 8 に示す例は、ユーザ端末 3 0 0 a、3 0 0 b の双方に合成動画データを送信する場合、すなわち、図 7 のステップ S a 2、S a 4 のいずれの判断も「Y E S」となる場合のものである。また、図 8 においては、図示の便宜上、A R 動画生成装置 1 0 0 を 2 つ示し、ユーザ端末 3 0 0 a に合成動画データを送信するための構成とユーザ端末 3 0 0 b に合成動画データを送信するための構成とを分けているが、実際には同一の装置で同時並行的に処理が行われてもよい。

## 【 0 0 5 9 】

まず、ユーザ端末 3 0 0 a とユーザ端末 3 0 0 b とは、接続管理装置 2 0 を介して、呼接続を確立する（ステップ S b 1）。例えば、ユーザは、電話番号やこれに代替するもの（ユーザアカウントなど）を用いてコミュニケーションの相手を指定し、発話を開始することができる。

30

## 【 0 0 6 0 】

呼接続が確立すると、接続管理装置 2 0 は、ユーザ端末 3 0 0 a、3 0 0 b に接続情報を送信する（ステップ S b 2）。接続情報は、ユーザ端末 3 0 0 a とユーザ端末 3 0 0 b とが接続されたときに送信される情報であって、コミュニケーションの相手に関する情報などを含んでいる。例えば、接続情報には、呼接続以降にシステム内部で呼を識別するために用いられる呼識別情報などが含まれる。また、接続情報は、上述した回線情報やリソース情報を含んでもよい。なお、回線情報やリソース情報を接続情報に含む場合、ユーザ端末 3 0 0 は、第 1 動画データや第 2 動画データを送信するか否かを接続情報に基づいて判断してもよい。このようにすれば、図 7 に示した判定の一部をユーザ端末 3 0 0 で行うことが可能になる。

40

## 【 0 0 6 1 】

また、ユーザ端末 3 0 0 a、3 0 0 b は、呼接続が確立すると、映像の撮影を開始する（ステップ S b 3）。すなわち、ユーザ端末 3 0 0 a、3 0 0 b は、第 1 撮影部 3 4 0 及び第 2 撮影部 3 5 0 を起動し、動画データを送信できる状態にする。そして、ユーザ端末 3 0 0 a、3 0 0 b は、第 1 動画データと第 2 動画データを、A R 動画生成装置 1 0 0 にそれぞれ送信する（ステップ S b 4、S b 5）。A R 動画生成装置 1 0 0 は、第 1 動画データお第 2 動画データを受信すると、合成動画データを生成し（ステップ S b 6）、ユー

50

ザ端末300a、300bに送信する(ステップSb7)。

【0062】

以上のとおり、本実施形態によれば、一方のユーザが撮影した背景に対して他方のユーザの映像を合成し、合成した映像を用いながら音声メッセージをやり取りすることが可能になる。また、本実施形態においては、動画データの合成をAR動画生成装置100が行うため、ユーザ端末300においてポリゴンデータのレンダリングのような比較的負荷が高い処理を実行する必要がない。

【0063】

また、本実施形態において、AR動画生成装置100は、合成動画データの生成態様をリソース情報や回線情報に応じて異ならせることができる。これにより、AR動画生成装置100は、遅延するおそれがより少ない態様でデータ転送を行ったり、無駄な処理の実行を省略したりすることが可能である。また、AR動画生成装置100は、一方のユーザ端末300のみが距離画像センサを備え、他方のユーザ端末300には距離画像センサが備わっていない場合であれば、当該他方のユーザ端末300に送信する合成動画データのみを生成することができ、このような態様でのコミュニケーションも実現可能である。

【0064】

[変形例]

本発明は、上述した実施形態の態様に限らず、他の態様でも実施することができる。以下に示すのは、本発明の他の態様の一例である。なお、これらの変形例は、必要に応じて、各々を適宜組み合わせてもよい。

【0065】

(1)本発明は、二者に限らず、三者以上のコミュニケーションにも適用可能である。この場合において、各ユーザのユーザ端末300が距離画像センサを備えていてもよいが、特定の一のユーザのみが距離画像センサで自身を撮影する態様も可能である。例えば、第1ユーザ、第2ユーザ、第3ユーザという3名のユーザがコミュニケーションを行う場合において、第1ユーザのみが距離画像センサで自身を撮影し、第2ユーザと第3ユーザは背景のみを撮影する、といった態様でコミュニケーションを行うことも可能である。この場合、AR動画生成装置100は、撮影された第1ユーザに相当する画像を抽出し、これを第2ユーザ側で撮影された第2動画データに合成する処理と、第3ユーザ側で撮影された第2動画データに合成する処理とを実行し、第2ユーザと第3ユーザのそれぞれに応じた合成動画データを生成する。

【0066】

このようにすれば、第2ユーザにあっては、自身が撮影している背景に第1ユーザの映像が重畳された映像を視聴可能である一方、第3ユーザにあっては、自身が撮影している背景に第1ユーザの映像が重畳された映像(すなわち、第1ユーザの映像は重畳されているが、第2ユーザが視聴している映像とは異なる映像)を視聴可能である。

なお、このような態様においては、第1ユーザのユーザ端末300は第2撮影部350を備えていなくてもよく、また、第2ユーザ及び第3ユーザのユーザ端末300は第1撮影部340を備えていなくてもよい。

【0067】

(2)上述したように、通信システム10においては、音声メッセージの音量又は音質の変化に応じて合成動画データの映像を変化させたり、あるいは第1動画データの映像の変化に応じて音声メッセージの音量又は音質を変化させることが可能である。具体的な例としては、以下のようなものがある。

【0068】

例えば、AR動画生成装置100は、ユーザ端末300aに対して送信する合成動画データに関して、抽出されたオブジェクトを合成する場合において、ユーザ端末300bから送信された音声メッセージの音量が大きくなったときには、当該オブジェクトを拡大する一方、当該音声メッセージの音量が小さくなったときには、当該オブジェクトを縮小するようにしてもよい。また、AR動画生成装置100は、音声メッセージにノイズが生じ

10

20

30

40

50

るなどして音質が悪化した場合に、合成動画データが表す映像にも（意図的な）ノイズを付加したり、あるいは画質を劣化させたりしてもよい。このようにすれば、合成される映像と音声との間に感覚的な連動性を与えることが可能である。

【 0 0 6 9 】

また、接続管理装置 2 0 0 は、A R 動画生成装置 1 0 0 から被写体の変化を表す情報に基づいて、抽出されたオブジェクトの大きさの変化や、距離情報の変化や、あるいは被写体全体の明るさの変化などを認識することが可能である。接続管理装置 2 0 0 は、このようにして認識された映像の変化に応じて、例えば、オブジェクトの大きさやその距離の変化に合わせて音量を変化させたり、被写体が明るいときと暗いときとで音質を異ならせたりしてもよい。また、接続管理装置 2 0 0 は、被写体であるユーザの表情の変化に応じて音質や音量を異ならせてもよい。さらに、接続管理装置 2 0 0 は、音声メッセージがマルチチャンネル（例えばステレオ）のデータの場合であれば、所定のオブジェクトが抽出された位置の変化に応じて音像定位を制御し、例えば、被写体であるユーザの移動に追従するように音量やディレイを調整してもよい。

10

【 0 0 7 0 】

（ 3 ）算出部 3 1 1 に相当する機能は、ユーザ端末 3 0 0 ではなく、A R 動画生成装置 1 0 0 に備わっていてもよい。すなわち、A R 動画生成装置 1 0 0 は、ユーザ端末 3 0 0 から送信された動画データに基づいて座標系を算出するように構成されてもよい。算出部 3 1 1 に相当する機能は、例えば、第 2 受信部 1 1 2 a、1 1 2 b が有していてもよい。

【 0 0 7 1 】

（ 4 ）A R 動画生成装置 1 0 0 は、ユーザ端末 3 0 0 の動き（より詳細には、第 2 撮影部 3 5 0 の動き）に応じて、抽出したオブジェクトを回転させて合成することが可能である。しかし、距離画像センサで撮影ができるのは、ユーザが当該センサに向けている側（正面側）のみであって、反対側（背面側）を撮影することはできない。したがって、抽出されたオブジェクトのうち、距離画像センサで撮影できない部分については、これを表示させないか、あるいはあらかじめ決められた適当な代替的な映像を表示させるようにすればよい。

20

【 0 0 7 2 】

なお、オブジェクトの回転は、ユーザの操作によって行われてもよい。すなわち、A R 動画生成装置 1 0 0 は、ユーザの操作を受け付け、受け付けた操作に応じてオブジェクトを回転させてから、これを背景の映像に合成して合成動画データを生成することが可能である。

30

【 0 0 7 3 】

（ 5 ）A R 動画生成装置 1 0 0 は、抽出したオブジェクトに所定の画像を付加し、これを背景に合成することも可能である。ここでいう所定の画像とは、例えば、ユーザの顔や身体に装着する装飾品を模した画像（サングラス、洋服など）である。このようにすれば、ユーザの顔や身体の一部を隠した映像を表示することが可能になる。あるいは、A R 動画生成装置 1 0 0 は、抽出したオブジェクトを所定の画像（アバタなど）に置き換えて合成を行ってもよい。A R 動画生成装置 1 0 0 は、第 2 動画データが送信されてこなかった場合や、第 2 動画データの転送に遅延が生じている場合を判断し、このような場合に置き換えを行うようにしてもよい。なお、これらの画像を表示するための画像データは、あらかじめ A R 動画生成装置 1 0 0 や接続管理装置 2 0 0 に記憶されていてもよいし、ユーザ端末 3 0 0 がコミュニケーションを開始するときに接続管理装置 2 0 0 に送信してもよい。

40

【 0 0 7 4 】

（ 6 ）接続情報には、上述した変形例（ 4 ）に示した画像データが含まれてもよい。また、接続情報は、このほかに、ユーザがあらかじめ設定した属性情報（性別、趣味）などを含み得る。変形例（ 4 ）において、A R 動画生成装置 1 0 0 は、抽出したオブジェクトを所定の画像に置き換え、又は当該オブジェクトに所定の画像を付加する場合に、表示すべき画像を属性情報に基づいて決定してもよい。

【 0 0 7 5 】

50

(7) 色情報の画素のサイズと距離情報の画素のサイズ、すなわち両者の解像度は、必ずしも一致していなくてもよい。例えば、距離情報の1画素は、色情報の4画素(縦2画素×横2画素)に相当していてもよい。この場合の距離情報は、色情報の4画素分について同一の値であるとしてもよいが、隣り合う画素の距離情報を参照して適当な補間処理を実行して算出されてもよい。

【0076】

(8) 本発明において抽出されるオブジェクトは、距離情報が所定の条件を満たすものであればどのようなものであってもよく、必ずしもユーザである必要はない。例えば、ユーザは、自身に代えて、人形や、飼育しているペットを撮影し、これを合成対象の映像とすることも可能である。

10

【0077】

(9) AR動画生成装置100及び接続管理装置200は、別体の装置として構成されるのではなく、一体に構成されてもよい。また、本発明は、音声メッセージに代えて、テキストメッセージを交換するものであってもよい。このようにすれば、チャットのようなコミュニケーションにも本発明を適用することが可能である。この場合において、AR動画生成装置100は、音量や音質に代えて、文字のサイズや文字装飾(フォント、下線、色など)に応じて、あるいは特定の文字(感嘆符など)の有無に応じて映像を異ならせることも可能である。

【0078】

(10) 本発明は、AR動画生成装置やこれを備える画像コミュニケーションシステムとしてだけでなく、AR動画生成装置が実行するプログラムや、合成した動画を用いたコミュニケーション方法として把握することも可能である。また、このプログラムは、光ディスクなどの記録媒体に記録した形態や、インターネットなどのネットワークを介して、コンピュータにダウンロードさせ、これをインストールして利用可能にする形態などでも提供することができる。

20

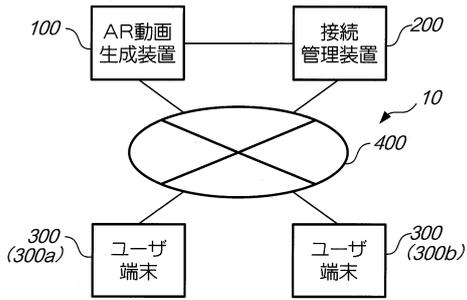
【符号の説明】

【0079】

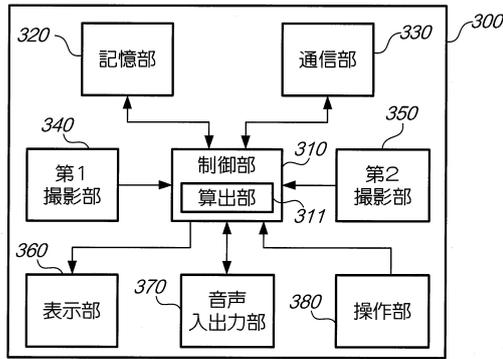
10...通信システム、100...AR動画生成装置、110...制御部、111a、111b...第1受信部、112a、112b...第2受信部、113a、113b...抽出部、114a、114b...生成部、115a、115b...送信部、116...データ授受部、117...判断部、118...解析部、記憶部...120、通信部...130、200...接続管理装置、210...制御部、211...送受信部、212...データ授受部、220...記憶部、130...通信部、300、300a、300b...ユーザ端末、310...制御部、320...記憶部、330...通信部、340...第1撮影部、350...第2撮影部、360...表示部、370...音声入出力部、380...操作部、400...ネットワーク

30

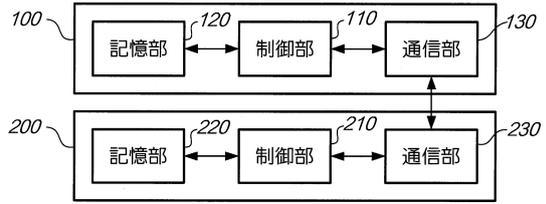
【図1】



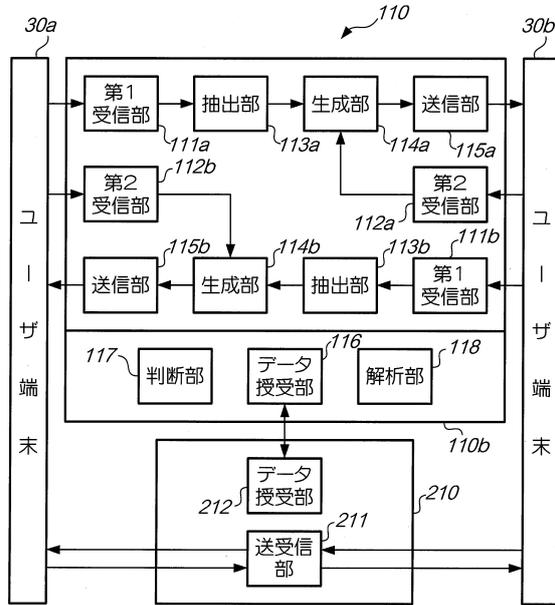
【図2】



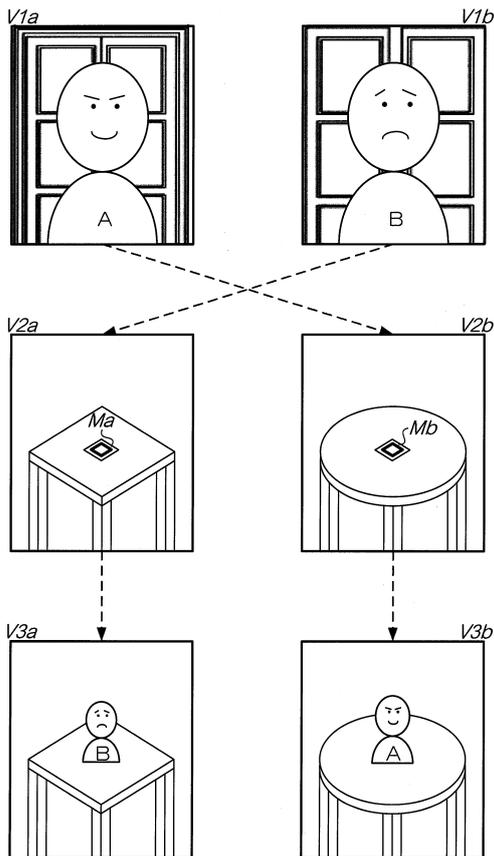
【図3】



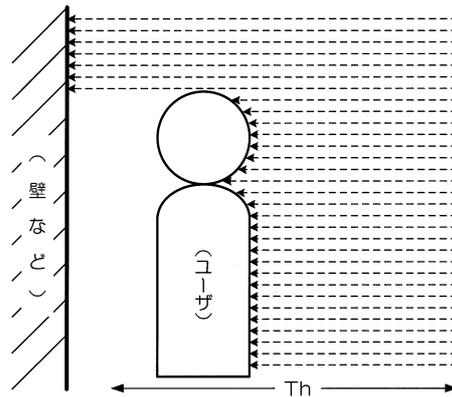
【図4】



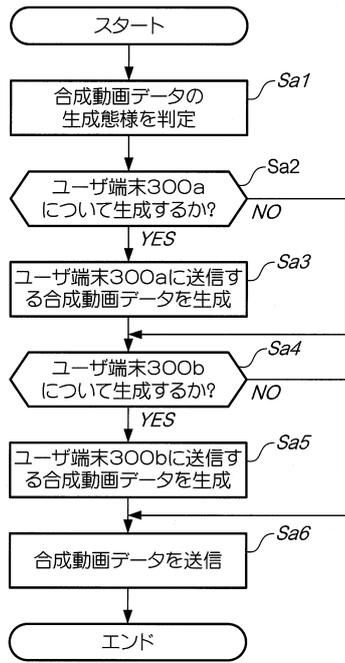
【図5】



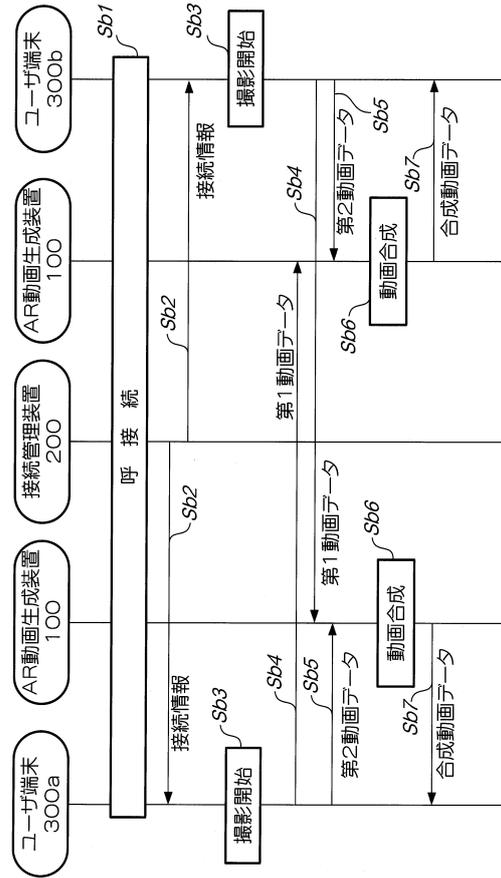
【図6】



【 図 7 】



【 図 8 】



## フロントページの続き

- (72)発明者 阿久澤 まり  
東京都千代田区永田町二丁目11番1号 株式会社エヌ・ティ・ティ・ドコモ内
- (72)発明者 松村 択磨  
東京都港区赤坂二丁目4番5号 ドコモ・テクノロジー株式会社内
- (72)発明者 清水 大樹  
東京都渋谷区恵比寿一丁目19番15号 日本コントロールシステム株式会社内
- (72)発明者 野中 洋志  
東京都渋谷区恵比寿一丁目19番15号 日本コントロールシステム株式会社内

審査官 村松 貴士

- (56)参考文献 加藤博一, 外2名, “拡張現実感技術を用いたビデオ会議システムとその評価”, 映像情報メディア学会技術報告, (社)映像情報メディア学会, 2001年 7月26日, Vol. 25, No. 48, p. 29 - 34  
小木哲朗, 外1名, “ビデオアバタと仮想空間コミュニケーション”, 計測と制御, 社団法人計測自動制御学会, 2002年 9月10日, 第41巻, 第9号, p. 649 - 652

(58)調査した分野(Int.Cl., DB名)

G06T 19/00

G06F 3/048 - 3/0489