

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2015年12月3日(03.12.2015)



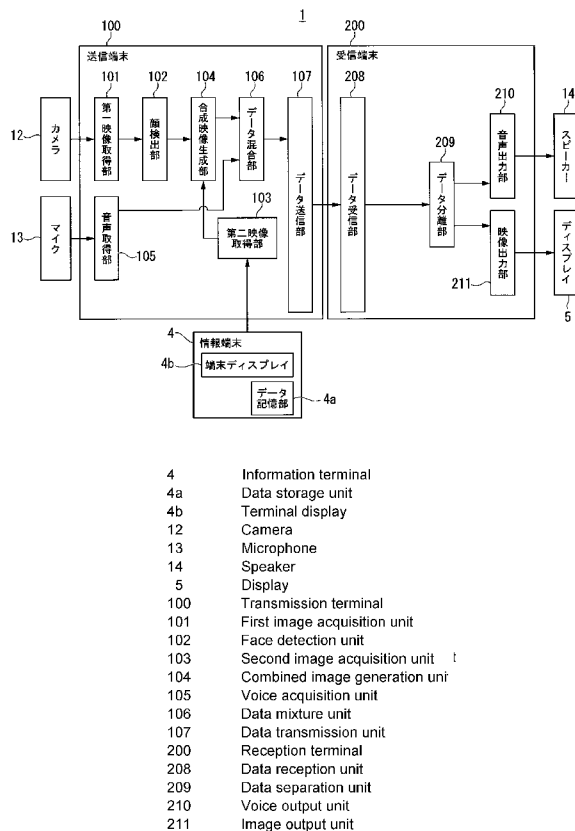
(10) 国際公開番号
WO 2015/182440 A1

- (51) 国際特許分類:
H04N 7/15 (2006.01) H04N 21/431 (2011.01)
H04N 5/225 (2006.01)
- (21) 国際出願番号: PCT/JP2015/064321
- (22) 国際出願日: 2015年5月19日(19.05.2015)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
特願 2014-111419 2014年5月29日(29.05.2014) JP
- (71) 出願人: シャープ株式会社(SHARP KABUSHIKI KAISHA) [JP/JP]; 〒5458522 大阪府大阪市阿倍野区長池町2番2号 Osaka (JP).
- (72) 発明者: 向井 成樹(MUKAI Shigeki), 若林 保孝(WAKABAYASHI Yasutaka), 田中 誠一(TANAKA Seiichi).
- (74) 代理人: 西澤 和純, 外(NISHIZAWA Kazuyoshi et al.); 〒1006620 東京都千代田区丸の内一丁目9番2号 Tokyo (JP).
- (81) 指定国(表示のない限り、全ての種類の国内保護が可能): 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, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) 指定国(表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE,

[続葉有]

(54) Title: TELEVISION CONFERENCE SYSTEM

(54) 発明の名称: テレビ会議システム



(57) Abstract: A television conference system wherein at least one or more sets of terminal apparatuses transmit and receive images via a network. In the television conference system, a terminal apparatus comprises: a face detection unit for detecting a face from a first image that is an image captured by a camera; and a generation unit for generating an image in which the image of the face detected by the face detection unit is placed, on the periphery of the image area of a second image that is an image of materials used in the conference, in accordance with a relative position of the face in the first image.

(57) 要約: 少なくとも一組以上の端末装置がネットワークを介して映像を送受信するテレビ会議システムにおいて、端末装置は、カメラにより撮影された映像である第一の映像から顔を検出する顔検出部と、会議で使用する資料の映像である第二の映像の映像領域の周囲に、顔検出部が検出した顔の映像を第一の映像における当該顔の相対位置に応じて配置した映像を生成する生成部と、を備える。

WO 2015/182440 A1



ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, 添付公開書類:

MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK,

SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ,

GW, KM, ML, MR, NE, SN, TD, TG).

— 国際調査報告 (条約第 21 条(3))

明 細 書

発明の名称： テレビ会議システム

技術分野

[0001] 本発明は、テレビ会議システムに関する。

本願は、2014年5月29日に、日本に出願された特願2014-111419号に基づき優先権を主張し、その内容をここに援用する。

背景技術

[0002] 近年、光通信や高速無線通信などを使った高速インターネットの普及により、テレビやカメラを接続して双方向通信を行うテレビ会議システムが利用されている。このテレビ会議システムでは、双方で資料を共有し、資料映像と相手側のカメラの映像（相手側の出席者の顔の映像）を同一のディスプレイで同時に見たいというニーズがある。

例えば、資料映像と相手側のカメラの映像を同一のディスプレイで同時に見るシステムでは、資料映像を親画面、カメラ映像を子画面として分け、2つの画面を並べて表示する方法が考えられるが、ディスプレイ画面の限られた表示領域内で親画面と子画面とを単に並べて表示してしまうと、資料映像の表示サイズが小さくなってしまうため、文字が読みづらくなってしまう等の問題がある。

[0003] この問題を解決する方法として、例えば、特許文献1では、音声を発した人物の顔をカメラ映像から切り取り、資料映像に重畳して表示するため、資料映像の表示サイズを変えずに相手側の出席者の顔を把握することが可能になる。また、特許文献2では、複数拠点でのカメラ映像をアイコンとして資料の周囲に表示し、それらのアイコンがユーザによって選択されると、その拠点にいる出席者の顔が資料映像に重畳されて表示されるため、出席者を把握しながら資料を見ることが可能になる。

先行技術文献

特許文献

[0004] 特許文献1：特開2013-115527号公報

特許文献2：特開2010-93583号公報

発明の概要

発明が解決しようとする課題

[0005] しかしながら、特許文献1に記載の方法では、相手側に複数の出席者がいた場合には、出席者のうちの話者のみの映像しか資料映像に重畳されないため、話者の発言内容に対する他の出席者の反応や状況が分からず、コミュニケーションが取りづらくなる可能性がある。また、資料映像の一部に人物の顔が重畳されてしまうため、資料の内容によっては文字や画像が隠れてしまい視認性が悪くなることがある。また、特許文献2に記載の方法では、選択された拠点の出席者の映像が資料映像の一部に重畳されるため、選択する拠点数が増えるほど、又は出席者が多いほど資料が隠れてしまう。このように、資料映像の視認性と、相手側の出席者の状況を把握すること、との両立が困難であった。

本発明の一態様は、上述した事情に鑑みてなされたもので、会議の資料映像の視認性を保ちつつ、相手側の出席者の状況を把握することができるテレビ会議システムを提供する。

課題を解決するための手段

[0006] 本発明の一態様は、少なくとも一組以上の端末装置がネットワークを介して映像を送受信するテレビ会議システムであって、前記端末装置は、カメラにより撮影された映像である第一の映像から顔を検出する顔検出部と、会議で使用する資料の映像である第二の映像の映像領域の周囲に、前記顔検出部が検出した顔の映像を前記第一の映像における当該顔の相対位置に応じて配置した映像を生成する生成部と、を備えるテレビ会議システムである。

発明の効果

[0007] 本発明の一態様によれば、会議の資料映像の視認性を保ちつつ、相手側の出席者の状況を把握することができる。

図面の簡単な説明

[0008] [図1]第1の実施形態によるテレビ会議システムの概略構成の一例を示すブロック図である。

[図2]第1の実施形態によるテレビ会議システムの使用環境の一例を示す模式図である。

[図3]第1の実施形態によるテレビ会議システムにおける処理の流れを示すフローチャートである。

[図4]第1の実施形態によるカメラ映像の一例を示す図である。

[図5]第1の実施形態によるカメラ映像に対する顔検出結果の一例を示す図である。

[図6]顔の回転方向の角度の説明図である。

[図7]カメラ映像から検出された顔に対する回転処理の説明図である。

[図8]背景画像の一例を示す図である。

[図9]資料映像に対する縮小処理の説明図である。

[図10]背景映像に資料映像を貼りつけた映像の一例を示す図である。

[図11]カメラ映像における顔の相対位置を示す図である。

[図12]資料映像と顔映像とを合成した合成映像を示す図である。

[図13]第2の実施形態によるテレビ会議システムの概略構成の一例を示すブロック図である。

[図14]第2の施形態によるテレビ会議システムの使用環境の一例を示す模式図である。

[図15]第2の実施形態によるテレビ会議システムにおける処理の流れを示すフローチャートである。

[図16]第2の実施形態によるカメラAが撮影した映像の一例を示す図である。

[図17]ディスプレイAに表示させるパターンの一例を示す図である。

[図18]パターンをディスプレイAに表示させた場合のカメラ映像の一例を示す図である。

[図19]テンプレートの例を示す図である。

[図20]テンプレートマッチングによる位置検出の説明図である。

[図21]第2の実施形態による回転処理の一例を説明する説明図である。

[図22]回転処理後のカメラ映像において顔が検出された状態を示す図である。

[図23]第3の実施形態によるテレビ会議システムの概略構成の一例を示すブロック図である。

[図24]第3の実施形態によるテレビ会議システムにおける処理の流れを示すフローチャートである。

[図25]送受信端末Bの第一映像取得部BがカメラBから取得したカメラ映像を示す図である。

[図26]合成映像にアイコンを重畳した映像を示す図である。

[図27]第4の実施形態によるテレビ会議システムの概略構成の一例を示すブロック図である。

[図28]カメラ映像301bにおける顔の相対位置及び顔方向を示す図である。

[図29]部屋を天井側から見下ろしたときの状態を示す模式図である。

[図30]顔方向に応じたアイコンを配置した合成映像の一例を示す図である。

発明を実施するための形態

[0009] 以下、図面を参照して、本発明の実施の形態について説明する。なお、各図面は本発明の原理に則った具体的な実施形態と実装例を示しているが、これらは本発明の理解のためのものであり、決して本発明を限定的に解釈するために用いられるものではない。また、各図面における各部の大きさ等は理解を容易にするため大小関係を誇張して描いており、実際の大きさとは異なる。

[0010] <第1の実施形態>

まず、本発明の第1の実施形態について説明する。

図1は、本発明の第1の実施形態によるテレビ会議システム1の概略構成

の一例を示すブロック図である。

[0011] テレビ会議システム 1 は、送信端末 100 と、この送信端末 100 に接続される情報端末 4、カメラ 12、及びマイク 13 と、受信端末 200 と、この受信端末 200 に接続されるディスプレイ 5、及びスピーカー 14 と、を備える。

[0012] カメラ 12 は、送信端末 100 の周辺にいる会議の出席者を撮影するための撮像装置であり、撮影した映像をケーブル等で接続された送信端末 100 に供給する。マイク 13 は、送信端末 100 の周辺の音声を集音し、音声信号をケーブル等で接続された送信端末 100 に供給する。なお、音声は、人の声に限定されるものではなく、様々な音又は声を指すものとする。

[0013] ディスプレイ 5 は、CRT (Cathode Ray Tube)、LCD (Liquid Crystal Display)、又は有機 EL (Electro-Luminescence) ディスプレイ等を備える表示装置やプロジェクタなどの映像投射装置であり、映像や情報を表示する装置であればどのようなものであってもよい。例えば、ディスプレイ 5 は、受信端末 200 から出力される映像を表示する。スピーカー 14 は、受信端末 200 から出力される音声信号を音声に変換して出力する。なお、スピーカー 14 は、1 チャンネルの音声を出力するものであってもよいし、左右 2 チャンネルの音声を出力するものであってもよい。

[0014] 情報端末 4 は、例えばパーソナルコンピュータのような固定式の情報端末装置、又はデジタルカメラ等、会議で使用する資料映像を生成する装置であればどのような情報装置であってもよい。ここでは、情報端末 4 がノート型のパーソナルコンピュータであるものとして説明する。情報端末 4 は、データ記憶部 4a と、端末ディスプレイ 4b とを備えている。データ記憶部 4a は、HDD (Hard Disk Drive)、フラッシュメモリ、又は DRAM (Dynamic Random Access Memory) といった半導体記憶装置や磁気記憶装置を備えている。このデータ記憶部 4a には、会議で使用する資料 (文書、画像、又は映像による資料、プレゼ

ンテーション資料等）のデータが記憶されている。例えば、情報端末４は、データ記憶部４aから読み出したデータに基づいて、会議で使用する資料の映像（以下、資料映像とも称する）を出力する。端末ディスプレイ４bは、上述したディスプレイ５と同様に、映像や情報を表示する。

[0015] 送信端末１００及び受信端末２００は、ネットワークを介して互いに通信する端末装置であり、例えば、セットトップボックスやパーソナルコンピュータのような固定情報端末や、ＰＤＡ（Personal Digital Assistant）やスマートフォンのような移動体通信端末、あるいはテレビ会議専用の通信装置である。

[0016] 送信端末１００は、テレビ会議システム１における送信側の端末装置の一例である。図示する送信端末１００は、第一映像取得部１０１と、顔検出部１０２と、第二映像取得部１０３と、合成映像生成部１０４（生成部）と、音声取得部１０５と、データ混合部１０６と、データ送信部１０７と、を備えている。

[0017] 第一映像取得部１０１は、カメラ１２が撮影した映像（以下、カメラ映像とも称する）を取得する。顔検出部１０２は、第一映像取得部１０１が取得したカメラ映像から顔を検出する。例えば、顔検出部１０２は、第一映像取得部１０１が取得したカメラ映像から顔の領域を検出するとともに、検出した顔の回転方向の角度を検出する。ここで、顔の回転方向の角度とは、顔を正面から見たときの中心を回転中心とする顔の回転方向の角度である。第二映像取得部１０３は、情報端末４が備えるデータ記憶部４aから会議で使用する資料映像データを読み出し、資料映像として取得する。

[0018] 合成映像生成部１０４は、顔検出部１０２が検出した顔の領域を第一映像取得部１０１が取得した映像から切り取り、切り取った顔の映像（以下、顔映像とも称する）と、第二映像取得部１０３が取得した資料映像とを合成する。また、合成映像生成部１０４は、合成した合成映像データをデータ混合部１０６に供給する。

[0019] 音声取得部１０５は、マイク１３から供給された音声信号を取得し、取得

した音声信号に基づく音声データをデータ混合部１０６に供給する。

[0020] データ混合部１０６は、合成映像生成部１０４が生成した合成映像データと、音声取得部１０５が取得した音声データとを混合し、混合した混合データをデータ送信部１０７に供給する。データ送信部１０７は、データ混合部１０６から供給された混合データを受信端末２００に送信する。

[0021] 受信端末２００は、テレビ会議システム１における受信側の端末装置の一例である。図示する受信端末２００は、データ受信部２０８と、データ分離部２０９と、音声出力部２１０と、映像出力部２１１と、を備えている。

[0022] データ受信部２０８は、送信端末１００から送信された混合データを受信する。データ分離部２０９は、データ受信部２０８が受信した混合データを、音声データと映像データとに分離する。そして、データ分離部２０９は、音声データを音声出力部２１０に供給し、映像データを映像出力部２１１に供給する。

[0023] 音声出力部２１０は、データ分離部２０９が混合データから分離した音声データを所定の音声信号に変換してスピーカ１４に出力する。映像出力部２１１は、データ分離部２０９が混合データから分離した映像データを所定の映像信号に変換してディスプレイ５に出力する。これにより、映像データに基づく映像がディスプレイ５に表示される。

[0024] このように、受信端末２００では、送信端末１００から混合データにより送信された資料映像と顔映像（送信端末１００側の会議の出席者の顔映像）との合成映像がディスプレイ５に表示され、送信端末１００から混合データにより送信された音声（送信端末１００側の会議の出席者の音声）がスピーカ１４から出力される。

[0025] なお、受信端末２００は、さらに送信端末１００の各部を備えた構成としてもよく、送信端末１００の機能も実行可能な端末装置としてもよい。また、送信端末１００は、さらに受信端末２００の各部を備えた構成としてもよく、受信端末２００の機能も実行可能な端末装置としてもよい。

[0026] （テレビ会議システム１の使用環境の具体例）

次に、図2を参照して、本実施形態によるテレビ会議システム1の使用環境の一例を説明する。

図2は、本実施形態によるテレビ会議システム1の使用環境の一例を示す模式図である。図示するテレビ会議システム1は、少なくとも一組の端末装置として、送信端末100と受信端末200とを備えている。部屋500に設置されている送信端末100と、部屋501に設置されている受信端末200とがネットワーク3を介して互いに通信を行い、テレビ会議が行われる。ネットワーク3は、インターネットや公衆回線網等であり、いずれか一方を用いて接続される場合も、両方が組み合わされて接続される場合もある。なお、送信端末100及び受信端末200のそれぞれは、LAN (Local Area Network) 15a、LAN 15bを介してネットワーク3に接続される。なお、LAN 15a、15bは、有線LANでもよいし無線LANでもよい。

[0027] 部屋500のテーブル20の略中央には、送信端末100が設置されている。また、テーブル20には、情報端末4、カメラ12、及びマイク13が設置されており、それぞれ送信端末100に接続されている。また、部屋500では、出席者A16、出席者B17、及び出席者C18の3名がテレビ会議に出席（参加）している。その3名の出席者は、テーブル20に設置された送信端末100を挟んで情報端末4の設置場所の反対側にいるものとする。つまり、出席者A16、出席者B17、及び出席者C18と、情報端末4とが、送信端末100をとり囲むような位置関係となっている。また、出席者A16、出席者B17、及び出席者C18は、情報端末4の端末ディスプレイ4bを視認できる位置におり、情報端末4の端末ディスプレイ4bの表示映像を見ながらテレビ会議を行うことができる。

[0028] カメラ12は、テーブル20の略中央（ここでは、送信端末100の上）に設置されている。例えば、カメラ12は、画角180度の円の中心からの距離が入射角に比例している等距離射影方式を採用している魚眼カメラであり、画角が広く、光軸が部屋500の天井面に対して垂直に交わるように、

上向きに設置されている。なお、部屋５００の天井面はテーブル２０のテーブル面と平行であるものとする。これにより、カメラ１２は、送信端末１００をとり囲むような位置関係にある情報端末４、出席者Ａ１６、出席者Ｂ１７、及び出席者Ｃ１８を、同時に一つの映像として撮影することができる。また、マイク１３は、無指向性のマイクを採用しており、カメラ１２の周囲で発せられた音声を集音することができる。

[0029] 一方、部屋５０１のテーブル２１には、受信端末２００が設置されている。また、テーブル２１には、ディスプレイ５、及びスピーカー１４が設置されており、それぞれ受信端末２００に接続されている。また、部屋５０１では、出席者Ｄ１９がテレビ会議に出席している。出席者Ｄ１９は、テーブル２１に設置された受信端末２００を挟んで、ディスプレイ５の設置場所の反対側にいるものとする。つまり、出席者Ｄ１９は、ディスプレイ５を視認できる位置におり、ディスプレイ５の表示映像を見ながらテレビ会議を行うことができる。

[0030] (テレビ会議システム１の処理)

次に、図３から図１２を参照して、本実施形態によるテレビ会議システム１における処理を詳しく説明する。

図３は、本実施形態によるテレビ会議システム１における処理の流れを示すフローチャートである。また、図４から図８は、テレビ会議システム１における処理に従って各部が実行する機能を説明するための説明図である。

[0031] まず、第一映像取得部１０１は、カメラ１２が撮影したカメラ映像（第一の映像）を取得する（ステップＳ１００）。図４は、カメラ１２が撮影したカメラ映像１０１ａの一例を示す図である。図示するように、カメラ映像１０１ａには、情報端末４、出席者Ａ１６、出席者Ｂ１７、及び出席者Ｃ１８が撮影されている。また、音声取得部１０５は、出席者Ａ１６、出席者Ｂ１７、及び出席者Ｃ１８の発言に応じて、音声信号をマイク１３から取得する（ステップＳ１１０）。

[0032] 第二映像取得部１０３は、情報端末４のデータ記憶部４ａに記憶されてい

るデータに基づいて資料映像（第二の映像）を取得する（ステップS120）。ここで、データ記憶部4aに記憶されているデータには、例えば、jpeg（Joint Photographic Experts Group）形式やPNG（Portable Network Graphic）形式等の画像データ、PDF（登録商標）形式やPowerPoint（登録商標）形式等のデータ等が含まれる。なお、データ記憶部4aに記憶されているデータは、上記以外の形式の各種データであってもよい。

[0033] 第一映像取得部101は、カメラ12から取得したカメラ映像101aを、顔検出部102に供給する。顔検出部102は、第一映像取得部101から供給されたカメラ映像101aから顔を検出する顔検出処理を行う。例えば、顔検出処理では、顔検出を行う映像に対して探索窓（例えば8ピクセル×8ピクセルのような判定領域）を左上から走査して順番に動かし、探索窓の領域毎に顔と認識できる特徴点を持つ領域があるか否かを判定することによって顔を検出する。この顔検出の方法としては、例えば、Viola-Jones法など、様々なアルゴリズムが提案されており、いずれの顔検出方法を用いてもよい。

[0034] 上述の顔検出処理において、顔検出部102は、カメラ映像101aから顔が検出されたか否かを判定する（ステップS130）。カメラ映像101aから顔が検出されなかった場合（ステップS130：No）、顔検出部102は、顔検出処理を終了して、ステップS150に処理を進める。一方、カメラ映像101aから顔が検出された場合（ステップS130：Yes）、顔検出部102は、検出した顔の領域の座標と、検出した際の顔の回転方向の角度を合成映像生成部104に供給する。

[0035] 図5は、カメラ映像101aに対する顔検出結果の一例を示す図である。例えば、顔検出部102は、図4に示すカメラ映像101aに対して顔検出処理を行った場合、図5に示すように、それぞれ出席者A16の顔領域40、出席者B17の顔領域41、及び出席者C18の顔領域42を検出するとともに、それぞれ検出された顔の回転方向の角度を検出する。なお、顔の回

転方向とは、顔を正面から見たときの中心を回転中心とする顔の回転方向である。

[0036] 図6は、顔の回転方向の角度の説明図である。この図を参照して、顔の回転方向の角度について説明する。図示するように、カメラ映像101aの映像領域の中心座標43とカメラ映像101aの下辺46aとを結ぶ直線44が下辺46aと直交するとき、中心座標43を回転中心とする直線44の回転方向の角度を0度とする。この直線44と、顔を左半分と右半分とに分ける境界線となる顔の中心線45とが平行な関係にあるときに、顔の回転方向の角度を0度とする。顔検出部102は、顔の回転方向の角度を、反時計回りに0度から359度までの範囲で検出できるものとする。

[0037] 例えば、中心座標43を回転中心として直線44を回転させたとき、カメラ映像101aの右辺46bと直交するときの直線44の回転方向の角度が90度、上辺46cと直交するときの直線44の回転方向の角度が180度、直線44が左辺46dと直交するときの直線44の回転方向の角度が270度となる。回転方向の角度が90度、180度、270度になるそれぞれの直線44と顔の中心線45とが平行な関係にあるとき、それぞれの顔の回転方向の角度が90度、180度、270度となる。

[0038] 次に、合成映像生成部104は、顔検出部102が検出した顔領域の座標情報に基づいて、第一映像取得部101が取得したカメラ映像101aから顔の領域を切り出すとともに、検出した顔の回転方向の角度に基づいて、それぞれの顔の回転方向の角度が0度になるように、切り出した顔映像の回転処理を行う（ステップS140）。また、合成映像生成部104は、第二映像取得部103が取得した資料映像に対し、映像のサイズを変換（縮小）する処理を行う（ステップS150）。

[0039] 続いて、合成映像生成部104は、サイズを変換した資料映像の映像領域の周囲に、当該資料映像に重ならないように、切り出して回転処理を行った顔映像を配置する。例えば、合成映像生成部104は、資料映像の映像領域の周囲に、切り出して回転処理を行った顔映像をカメラ映像101aにおけ

る顔の相対位置に応じて配置した合成映像を生成する。具体的には、合成映像生成部104は、まず黒で塗りつぶされた背景映像を生成して、その背景映像の中央に、サイズ変換後（縮小後）の資料映像を貼りつける。次に、合成映像生成部104は、サイズ変換後（縮小後）の資料映像の映像領域の周囲の背景映像の領域に、切り出して回転処理を行った顔映像を、カメラ映像101aにおいて検出された顔領域の中心座標を基に顔領域の相対位置関係を保持したまま貼りつけ、合成映像を生成する（ステップS160）。

[0040] 以下、図7から図12を参照して、ステップS140からステップS160の合成映像生成部104の処理を具体的に説明する。なお、ステップS100において第一映像取得部101が取得したカメラ映像101a（図4参照）の解像度を横1280×縦720画素とする。また、ステップS130において、顔検出部102が、出席者A16の顔領域40（横100×縦100画素）、出席者B17の顔領域41（横100×縦100画素）、出席者C18の顔領域42（横100×縦100画素）のそれぞれの領域を顔領域として検出するものとする（図5参照）。

[0041] ステップS140において、合成映像生成部104は、第一映像取得部101が取得したカメラ映像101aから、顔検出部102が検出した顔の領域を切り出すとともに、切り出した顔映像の回転処理を行う。

図7は、カメラ映像101aから検出された顔に対する回転処理の説明図である。図7（a）は、カメラ映像101aから切り出された顔映像を示している。ここでは、検出された顔領域40、顔領域41、及び顔領域42のそれぞれの顔の回転方向の角度が、270度、10度、80度である場合の顔映像の例を示している。合成映像生成部104は、それぞれの顔の回転方向が0度になるように回転処理を行う。図7（b）は、回転処理後の顔領域40、顔領域41、及び顔領域42のそれぞれの顔映像を示している。

[0042] 次に、ステップS150において、合成映像生成部104は、第二映像取得部103が取得した資料映像を縮小して、背景映像に貼りつける。図8は、背景画像の一例を示す図である。例えば、合成映像生成部104は、図示

するような黒色で塗られた映像（横1280×縦720画素）を背景映像101bとして生成する。なお、ここでは、黒色で塗られた映像を背景映像101bとしたが、他の色の映像、模様のある映像、又は写真映像等のようにいずれの映像を背景映像としてもよい。符号47は、背景映像101bの映像領域の中心座標を示す。

[0043] また、第二映像取得部103が取得した資料映像が横1280×縦720画素であったとする。この場合、合成映像生成部104は、この資料映像を横1080×縦520画素となるように縮小処理を行う。図9は、資料映像に対する縮小処理の説明図である。この図では、第二映像取得部103が取得した資料映像101c（縮小処理前）が、資料映像101dに縮小されることを示している。符号48は、縮小処理後の資料映像101dの映像領域の中心座標を示す。

[0044] ここで、縮小処理には、ニアレストネイバー法、バイリニア法、バイキュービック法等のような手法が用いられてもよい。また、縮小処理は、アスペクト比を固定したまま縮小させる処理としてもよい。アスペクト比を固定した縮小処理とする場合、横と縦の長さのうち短い方を基準に縮小させる必要があり、上述した解像度の場合には横944×縦520画素となるように縮小させてもよい。

[0045] 次に、合成映像生成部104は、背景映像101bの中心座標47（図8参照）と、縮小処理後の資料映像101dの中心座標48（図9参照）とが一致するように、背景映像101bに資料映像101dを貼りつける。図10は、背景映像101bに資料映像101dを貼りつけた映像の一例を示す図である。

[0046] そして、ステップS160において、合成映像生成部104は、図7（b）に示す回転処理後の顔映像を、図10に示す背景映像101bに資料映像101dを貼りつけた映像の、資料映像101dの周囲（外側）に貼りつけた合成映像を生成する。このとき、合成映像生成部104は、カメラ映像101aにおいて検出された顔領域の中心座標を基に顔領域の相対位置関係を

保持したまま、顔映像を貼りつける。

[0047] 図11は、カメラ映像101aにおける顔の相対位置を示す図である。カメラ映像101aから検出された顔領域40、顔領域41、及び顔領域42のそれぞれの中心の座標を、座標50、座標51、及び座標52とする。また、カメラ映像101aの中心座標43と、座標50、座標51、及び座標52のそれぞれとを結んだ直線を、直線53、直線54、及び直線55とする。この場合、カメラ映像101aにおいて検出されたそれぞれの顔の相対位置は、カメラ映像101aの中心座標43を基準として、直線53、直線54、及び直線55のそれぞれの線上の位置として表すことができる。

[0048] 図12は、資料映像101dと顔映像とを合成した合成映像101eの一例を示す図である。合成映像生成部104は、図10に示す背景映像101bに資料映像101dを貼りつけた映像に対して顔映像を貼りつけて合成映像101eを生成する。このとき、合成映像生成部104は、図11に示すカメラ映像101aの中心座標43に代えて資料映像101dの中心座標48を基準とし、中心座標48を基準としたときの直線53、直線54、及び直線55のそれぞれの直線上、且つ資料映像101dの周囲（外側）に、図7（b）に示す回転処理後の顔領域40、顔領域41、及び顔領域42のそれぞれの顔映像を貼りつける。

[0049] このようにすることで、合成映像101e上に表示される顔映像が、カメラ映像101aにおいて検出された顔の相対位置関係を保持したまま表示されることになる。なお、上述したカメラ映像101a、顔領域40、顔領域41、顔領域42、背景映像101b、資料映像101c、及び資料映像101dのそれぞれの画素は、一例であってこれに限定されるものではない。

[0050] 図3に戻り、ステップS170以降の処理を説明する。

データ混合部106は、合成映像生成部104が生成した合成映像データと、音声取得部105が取得した音声データとの同期を取り、多重化したストリームに変換することで混合を行う（ステップS170）。そして、データ混合部106は、混合した混合データをデータ送信部107に供給する。

- [0051] データ送信部１０７は、データ混合部１０６から供給された混合データをＬＡＮ１５経由で、受信端末２００に送信する（ステップＳ１８０）。ここで、混合データの送受信の一例として、Ｈ．３２０の通信プロトコルを用いた通信が行われるものとするが、これに限られるものではなく、Ｈ．３２３、Ｈ．３２４等のようにいずれの通信プロトコルを用いた通信が行われてもよい。
- [0052] 一方、受信端末２００は、送信端末１００から送信された、混合データを受信する。例えば、受信端末２００のデータ受信部２０８は、送信端末１００から送信された混合データを受信し、データ分離部２０９に供給する（ステップＳ１９０）。データ分離部２０９は、混合データの多重化されたストリームを映像データと音声データとに分離する。そして、データ分離部２０９は、分離した映像データを映像出力部２１１に供給し、分離した音声データを音声出力部２１０に供給する（ステップＳ２００）。
- [0053] 映像出力部２１１は、映像データに基づく映像をディスプレイ５に表示させる。つまり、映像出力部２１１は、送信端末１００で合成された合成映像１０１ｅ（図１２参照）をディスプレイ５に表示させる。（ステップＳ２１０）。また、音声出力部２１０は、音声データに基づく音声をスピーカ１４から出力させる（ステップＳ２２０）。
- [0054] 以上説明したように、本実施形態によるテレビ会議システム１は、少なくとも一組以上の端末装置（例えば、送信端末１００及び受信端末２００）がネットワーク３を介して映像を送受信する。送信端末１００は、顔検出部１０２と、合成映像生成部１０４（生成部）と、を備えている。顔検出部１０２は、カメラ１２により撮影されたカメラ映像１０１ａ（第一の映像）から顔を検出する。そして、合成映像生成部１０４は、会議で使用する資料の資料映像１０１ｄ（第二の映像）の映像領域の周囲に、顔検出部１０２が検出した顔映像をカメラ映像１０１ａにおける当該顔の相対位置に応じて配置した合成映像を生成する。
- [0055] このように、本実施形態によるテレビ会議システム１は、顔を検出した座

標の相対位置関係を保持したまま、顔映像を資料映像の周囲に表示させるため、会議の資料映像と相手側の出席者の顔映像との両方を、視認性を保ちつつ同時に見ながらテレビ会議を行うことができる。また、テレビ会議システム 1 は、相手側の出席者同士の相対位置関係を保持して表示することができるため、相手側の会議の状況や出席者の反応を限られた表示領域の中だけで、直観的に把握することができる。よって、本実施形態によれば、会議の資料映像の視認性を保ちつつ、相手側の出席者の状況を把握することができる。

[0056] なお、合成映像生成部 104 が、顔映像を資料映像の周囲に配置する場合、顔映像が資料映像に重ならないよう資料映像の外側に配置するのが望ましいが、資料映像の視認性を著しく損なわない程度に、資料映像の一部に重なるように配置してもよい。

[0057] また、カメラ 12 に画角 180 度の魚眼カメラを用いた例を説明したが、カメラ 12 が設置された部屋 500 で会議に出席している出席者のほぼ全員を映すための画角のカメラであれば、画角 180 度の魚眼カメラに限られるものではない。ここで、ほぼ全員と記述しているのは、出席者のいる場所によっては、画角 180 度の魚眼カメラを用いたとしても出席者の全員が映ることが保障されるものではないためである。例えば、カメラ 12 は、テーブル 20 の周囲近傍にいる出席者の全員を映すことが可能な程度の画角を有するカメラであることが望ましい。また、カメラ 12 は、複数のカメラを組み合わせ、パノラマ映像、全周囲映像を生成して映像を出力するカメラであってもよい。

[0058] また、マイク 13 に無指向性のマイクを用いた例を説明したが、カメラ 12 の周囲で発せられた音声を入力することができれば、いずれのマイクを用いてもよい。

[0059] <第 2 の実施形態>

次に、本発明の第 2 の実施形態について説明する。

第 1 の実施形態では、縮小後の資料映像 101 d の映像領域の周囲に、図

4に示すカメラ映像101aから検出した顔映像を合成することにより、図12に示す合成映像101eを生成する例を説明した。この例の場合には、カメラ12の設置方向によって、カメラ映像101a内の情報端末4の端末ディスプレイ4bの位置が変わり、検出される出席者の位置も変化する。したがって、第1の実施形態では、合成映像において、出席者の顔映像どうしの相対位置関係が保持されたとしても、カメラ12の設置方向によって、いずれの方向から見たときの出席者の位置になるかが変化することになる。例えば、図4に示すように情報端末4の端末ディスプレイ4bの位置が垂直方向で上部且つ水平方向で中央となるようなカメラ映像101aから検出された顔映像の相対位置関係が保持されれば、送信側の出席者と受信側の出席者とが向かい合った関係で合成映像を見ながら会議を行うことができる。そのためには、第1の実施形態では、カメラ12を設置する際にカメラ12の向きを合わせておく必要があるが、本実施形態では、向きを合わせなくてもよい形態について説明する。

[0060] 図13は、本実施形態によるテレビ会議システム1aの概略構成の一例を示すブロック図である。テレビ会議システム1aは、送受信端末A300と、この送受信端末A300に接続されるカメラA60、マイクA61、スピーカーA62、ディスプレイA63、及び情報端末A68と、送受信端末B400と、この送受信端末B400に接続されるカメラB64、マイクB65、ディスプレイB66、及びスピーカーB67を備える。

[0061] カメラA60及びカメラB64は、図1のカメラ12に相当し、マイクA61及びマイクB65は、図1のマイク13に相当する。また、スピーカーA62及びスピーカーB67は、図1のスピーカー14に相当し、ディスプレイA63及びディスプレイB66は、図1のディスプレイ5に相当する。また、情報端末A68は、図1の情報端末4に相当し、会議で使用される資料や情報等のデータが記憶されるデータ記憶部A68aを備えている。なお、ディスプレイA63は、図1の情報端末4が備える端末ディスプレイ4bに相当する構成としてもよく、ここでは、情報端末A68に備えられている

ものとして説明する。

[0062] 送受信端末A300及び送受信端末B400は、図1の送信端末100及び受信端末200と同様にネットワーク3を介して互いに通信する端末装置である。

送受信端末A300は、第一映像取得部A301と、ディスプレイ検出部A302（位置検出部）と、回転処理部A303と、顔検出部A304と、第二映像取得部A305と、合成映像生成部A306（生成部）と、音声取得部A307と、データ混合部A308と、データ送信部A309と、データ受信部A310と、データ分離部A311と、音声出力部A312と、映像出力部A314と、を備えている。

[0063] 第一映像取得部A301は、カメラA60が撮影した映像を取得する。ディスプレイ検出部A302は、第一映像取得部A301が取得した映像からディスプレイA63の位置を検出する。回転処理部A303は、ディスプレイ検出部A302が検出したディスプレイA63の位置に応じて、第一映像取得部A301が取得した映像に回転処理を行う。

[0064] 顔検出部A304は、回転処理部A303が回転処理を行った映像から顔の領域と、顔の回転方向の角度を検出する。第二映像取得部A305は、情報端末A68が備えるデータ記憶部A68aから会議で使用する資料映像データを読み出し、資料映像として取得する。

[0065] 合成映像生成部A306は、顔検出部A304が検出した顔の領域を第一映像取得部A301が取得した映像から切り取り、切り取った顔映像と、第二映像取得部A305が取得した資料映像とを合成する。また、合成映像生成部A306は、合成した合成映像データをデータ混合部A308に供給する。音声取得部A307は、マイクA61から供給された音声信号を取得し、取得した音声信号に基づく音声データをデータ混合部A308に供給する。

[0066] データ混合部A308は、合成映像生成部A306が生成した合成映像データと、音声取得部A307が取得した音声データとを混合し、混合した混

合データをデータ送信部A309に供給する。データ送信部A309は、データ混合部A308から供給された混合データを送受信端末B400に送信する。

[0067] 次に受信側の機能として、データ受信部A310は、送受信端末B400から送信された混合データを受信する。データ分離部A311は、データ受信部A310が受信した混合データを、音声データと映像データとに分離する。そして、データ分離部A311は、音声データを音声出力部A312に供給し、映像データを映像出力部A314に供給する。

[0068] 音声出力部A312は、データ分離部A311が混合データから分離した音声データを所定の音声信号に変換してスピーカーA62に出力する。映像出力部A314は、データ分離部A311が混合データから分離した映像データを所定の映像信号に変換してディスプレイA63に出力する。これにより、映像データに基づく映像が情報端末A68のディスプレイA63に表示される。

[0069] なお、送受信端末B400は、送受信端末A300の各部と同様の構成を備え、同様の機能を有しているものとする。例えば、送受信端末B400は、第一映像取得部B401と、ディスプレイ検出部B402（位置検出部）と、回転処理部B403と、顔検出部B404と、第二映像取得部B405と、合成映像生成部B406（生成部）と、音声取得部B407と、データ混合部B408と、データ送信部B409と、データ受信部B410と、データ分離部B411と、音声出力部B412と、映像出力部B414と、を備えており、これらの各部が送受信端末A300の各部に対応する。なお、この図13では、送受信端末B400に情報端末A68に相当する情報端末の図示を省略しているが、送受信端末B400には情報端末が接続されてもよいし接続されなくてもよい。

[0070] （テレビ会議システム1aの使用環境の具体例）

次に、図14を参照して、本実施形態によるテレビ会議システム1aの使用環境の一例を説明する。図14は、本実施形態によるテレビ会議システム

1 a の使用環境の一例を示す模式図である。

図示するテレビ会議システム 1 a は、少なくとも一組の端末装置として、送受信端末 A 3 0 0 と送受信端末 B 4 0 0 とを備えている。部屋 5 1 0 に設置されている送受信端末 A 3 0 0 と、部屋 5 1 1 に設置されている送受信端末 B 4 0 0 とがネットワーク 3 を介して互いに通信を行い、テレビ会議が行われる。本例では、送受信端末 A 3 0 0 及び送受信端末 B 4 0 0 のそれぞれは、Wi-Fi（登録商標：Wireless Fidelity）7 3、Wi-Fi 7 5 を介してネットワーク 3 に接続される。なお、送受信端末 A 3 0 0 及び送受信端末 B 4 0 0 のそれぞれは、有線 LAN を介してネットワーク 3 に接続されてもよい。

[0071] 部屋 5 1 0 のテーブル 7 0 の略中央には、送受信端末 A 3 0 0 が設置されている。また、テーブル 7 0 には、情報端末 A 6 8、カメラ A 6 0、マイク A 6 1、及びスピーカー A 6 2 が設置されており、それぞれ送受信端末 A 3 0 0 に接続されている。また、部屋 5 1 0 では、出席者 E 8 0、出席者 F 8 1、及び出席者 G 8 2 の 3 名がテレビ会議に出席（参加）している。その 3 名の出席者は、テーブル 7 0 に設置された送受信端末 A 3 0 0 を挟んで情報端末 A 6 8 の設置場所の反対側にいるものとする。つまり、出席者 E 8 0、出席者 F 8 1、及び出席者 G 8 2 と、情報端末 A 6 8 とが、送受信端末 A 3 0 0 をとり囲むような位置関係となっている。また、出席者 E 8 0、出席者 F 8 1、及び出席者 G 8 2 は、情報端末 A 6 8 のディスプレイ A 6 3 を視認できる位置におり、情報端末 A 6 8 のディスプレイ A 6 3 の表示映像を見ながらテレビ会議を行うことができる。

[0072] 一方、部屋 5 1 1 のテーブル 7 1 には、送受信端末 B 4 0 0 が設置されている。また、テーブル 7 1 には、カメラ B 6 4、マイク B 6 5、ディスプレイ B 6 6、及びスピーカー B 6 7 が設置されており、それぞれ送受信端末 B 4 0 0 に接続されている。また、部屋 5 1 1 では、出席者 H 8 3 がテレビ会議に出席している。出席者 H 8 3 は、テーブル 7 1 に設置された送受信端末 B 4 0 0 を挟んで、ディスプレイ B 6 6 の設置場所の反対側にいるものとする。

る。つまり、出席者H 8 3は、ディスプレイB 6 6を視認できる位置におり、ディスプレイB 6 6の表示映像を見ながらテレビ会議を行うことができる。

[0073] カメラA 6 0は、送受信端末A 3 0 0の上に設置されている。一方、カメラB 6 4は、送受信端末B 4 0 0の上に設置されている。例えば、カメラA 6 0及びカメラB 6 4は、画角1 8 0度の映像上の面積が立体角に比例する等立体角射影方式を採用している魚眼カメラであり、画角が広く、それぞれの光軸が部屋5 1 0の天井面と部屋5 1 1の天井面とのそれぞれに対して垂直に交わるように、上向きに設置されている。なお、部屋5 1 0及び部屋5 1 1の天井面のそれぞれは、テーブル7 0及びテーブル7 1のそれぞれのテーブル面と平行であるものとする。例えば、カメラA 6 0は、送受信端末A 3 0 0をとり囲むような位置関係にある情報端末A 6 8、出席者E 8 0、出席者F 8 1、及び出席者G 8 2を、同時に一つの映像として撮影することができる。また、カメラB 6 4は、ディスプレイB 6 6及び出席者H 8 3を、同時に一つの映像として撮影することができる。また、マイクA 6 1及びマイクB 6 5のそれぞれは、無指向性のマイクを採用しており、カメラA 6 0とカメラB 6 4とのそれぞれの周囲で発せられた音声を集音することができる。

[0074] (テレビ会議システム1 aの処理)

次に、図1 5から図2 2を参照して、本実施形態によるテレビ会議システム1 aにおける処理を詳しく説明する。

図1 5は、本実施形態によるテレビ会議システム1 aにおける処理の流れを示すフローチャートである。また、図1 6から図2 2は、テレビ会議システム1 aにおける処理に従って各部が実行する機能を説明するための説明図である。

[0075] なお、図1 5に示す処理の流れにおいて、ステップS 3 0 0からステップS 3 2 0までの各処理は、図3に示すステップS 1 0 0からステップS 1 2 0までの各処理と同様の処理であり、適宜説明を省略する。また、図1 5に

示す処理の流れにおいて、ステップS 3 3 0からステップS 4 2 0までの各処理は、図3に示すステップS 1 3 0からステップS 4 2 0までの各処理と同様の処理であり、適宜説明を省略する。

[0076] まず、第一映像取得部A 3 0 1は、カメラA 6 0が撮影したカメラ映像（第一の映像）を取得する（ステップS 3 0 0）。図16は、カメラA 6 0が撮影したカメラ映像3 0 1 aの一例を示す図である。本例のカメラ映像3 0 1 aには、情報端末A 6 8、出席者E 8 0、出席者F 8 1、及び出席者G 8 2が撮影されているが、それらの被写体が撮影されている位置が図4に示す例と異なる。例えば、図4に示すように情報端末4の端末ディスプレイ4 b（図16では情報端末A 6 8のディスプレイA 6 3）の位置が垂直方向で上部且つ水平方向で中央となるような映像を得るためには、カメラA 6 0を設置する際にカメラA 6 0の向きを合わせておく必要がある。この図16に示すカメラ映像3 0 1 aの例は、カメラA 6 0を設置する際にカメラA 6 0の向きを合わせていない場合の例であり、情報端末A 6 8のディスプレイA 6 3の位置が垂直方向で上部且つ水平方向で中央となる位置にない。

[0077] また、音声取得部A 3 0 7は、出席者E 8 0、出席者F 8 1、及び出席者G 8 2の発言に応じて、音声信号をマイクA 6 1から取得する（ステップS 3 1 0）。第二映像取得部A 3 0 5は、情報端末A 6 8のデータ記憶部A 6 8 aに記憶されているデータに基づいて資料映像（第二の映像）を取得する（ステップS 3 2 0）。

[0078] 次に、ディスプレイ検出部A 3 0 2は、第一映像取得部A 3 0 1が取得したカメラ映像3 0 1 aからディスプレイA 6 3を検出する。例えば、ディスプレイ検出部A 3 0 2は、ディスプレイA 6 3に一定のパターンを表示させることにより、そのパターンを認識してディスプレイA 6 3の位置を検出する。図17は、ディスプレイA 6 3に表示させるパターン9 0の一例を示す図である。図示するパターン9 0は、横2×縦2の白黒の市松模様のパターンである。図18は、図17のパターン9 0（例えば、横1 2 8 0×縦7 2 0画素）をディスプレイA 6 3に表示させた場合に、第一映像取得部A 3 0

1 が取得するカメラ映像 301 a の一例を示す図である。

[0079] ディスプレイ検出部 A 302 は、パターン 90 を縮小させた画像をテンプレート 91（例えば、横 80×縦 45 画素）として、テンプレートマッチングを行うことによりカメラ映像 301 a からディスプレイ A 63 の位置を検出する。図 19 は、テンプレート 91 の例を示す図である。例えば、パターン 90 の縮小画像を、縮小画像の中央を回転中心として 0 度、45 度（225 度）、90 度（270 度）、135 度（315 度）のそれぞれの回転角度に回転させた 4 種類の画像をテンプレート 91 として用いる。

[0080] 図 20 は、テンプレートマッチングによる位置検出の説明図である。ディスプレイ検出部 A 302 は、4 種類の回転方向のテンプレート 91 を順番に用いて 4 方向探索のテンプレートマッチングを行い、最も一致度が高くなったときのテンプレート 91 の中心座標を、カメラ映像 301 a におけるディスプレイ A 63 の位置として検出する。

[0081] なお、ディスプレイ検出部 A 302 は、図 17 に示すパターン 90 に限らず、赤や青色等の単色のパターンをディスプレイ A 63 に表示させ、表示させたパターンと同様な色の領域の面積が最も大きくなる領域の中心座標をディスプレイ A 63 の位置として検出してもよい。

[0082] 次に、回転処理部 A 303 は、ディスプレイ検出部 A 302 によりカメラ映像 301 a からディスプレイ A 63 が検出されたか否かを判定する（ステップ S 322）。カメラ映像 301 a からディスプレイ A 63 が検出されなかったと判定された場合（ステップ S 322：No）、回転処理部 A 303 は、回転処理を行わず、ステップ S 300 に処理を戻す。

[0083] 一方、カメラ映像 301 a からディスプレイ A 63 が検出されたと判定された場合（ステップ S 322：Yes）、回転処理部 A 303 は、ディスプレイ検出部 A 302 が検出したディスプレイ A 63 の位置に応じて、カメラ映像 301 a の回転処理を行う。例えば、回転処理部 A 303 は、ディスプレイ検出部 A 302 が検出したディスプレイ A 63（表示部）の位置が垂直方向で上部且つ水平方向で中央となるようにカメラ映像 301 a を回転させ

る（ステップS 3 2 4）。

[0084] 図 2 1 は、本実施形態による回転処理の一例を説明する説明図である。図 2 1（a）は回転処理前のカメラ映像 3 0 1 a を示しており、図 2 1（b）は回転処理後のカメラ映像 3 0 1 b を示している。図 2 1（a）において、ディスプレイ A 6 3 の画面の中心座標 9 3 とカメラ映像 3 0 1 a の映像領域の中心座標 9 2 とを結ぶ直線を直線 9 4 とする。回転処理部 A 3 0 3 は、直線 9 4 がカメラ映像 3 0 1 a の上辺 9 5 と直交するように、中心座標 9 2 を回転中心としてカメラ映像 3 0 1 a を回転させる。図 2 1（b）に示すように、回転処理後のカメラ映像 3 0 1 b では、直線 9 4 がカメラ映像 3 0 1 b の上辺 9 5 と直交するように回転された映像となる。即ち、図 2 1（b）に示すカメラ映像 3 0 1 b は、ディスプレイ A 6 3 の位置が垂直方向で上部且つ水平方向で中央となるように回転された映像である。

[0085] 次に、顔検出部 A 3 0 4 は、回転処理部 A 3 0 3 が回転処理を行ったカメラ映像 3 0 1 b に対して顔検出処理を行い、顔が検出されたか否かを判定する（ステップS 3 3 0）。

[0086] 顔が検出されなかった場合（ステップS 3 3 0：N o）、顔検出部 A 3 0 4 は、顔検出処理を終了して、ステップS 3 5 0 に処理を進める。一方、顔が検出された場合（ステップS 3 3 0：Y e s）、顔検出部 A 3 0 4 は、検出した顔の領域の座標と、検出した際の顔の回転方向の角度を合成映像生成部 A 3 0 6 に供給する。図 2 2 は、回転処理後のカメラ映像 3 0 1 b において顔が検出された状態を示す図である。顔検出部 A 3 0 4 は、出席者 E 8 0 の顔領域 9 6、出席者 F 8 1 の顔領域 9 7、出席者 G 8 2 の顔領域 9 8 のそれぞれの領域を顔領域として検出する。この図に示す顔が検出された状態は、図 5 に示す顔が検出された状態と同様である。そのため、ステップS 3 3 0 より後の各処理は、第 1 の実施形態で説明した、図 3 のステップS 1 3 0 より後の各処理と基本的に同様の処理とすることができる。以下、ステップS 3 3 0 以降について詳細を省略して説明する。

[0087] 合成映像生成部 A 3 0 6 は、顔検出部 A 3 0 4 が検出した顔領域の座標情

報に基づいて、回転処理部A303が回転処理を行ったカメラ映像301bから顔の領域を切り出すとともに、検出した顔の回転方向の角度に基づいて、それぞれの顔の回転方向の角度が0度になるように、切り出した顔映像の回転処理を行う（ステップS340：図7参照）。また、合成映像生成部A306は、第二映像取得部A305が取得した資料映像に対し、映像のサイズを変換（縮小）する処理を行う（ステップS350：図9参照）。

[0088] 次に、合成映像生成部A306は、サイズを変換した資料映像の映像領域の周囲に、切り出して回転処理を行った顔映像を、回転処理後のカメラ映像301bにおける顔の相対位置に応じて配置した合成映像を生成する（ステップS360：図12の合成映像101e参照）。

[0089] データ混合部A308は、合成映像生成部A306が生成した合成映像データと、音声取得部A307が取得した音声データとを混合し、混合した混合データをデータ送信部A309に供給する（ステップS370）。データ送信部A309は、データ混合部A308から供給された混合データを、送受信端末B400に送信する（ステップS380）。

[0090] 一方、送受信端末B400は、送受信端末A300から送信された、混合データを受信する。例えば、送受信端末B400のデータ受信部B410は、送受信端末A300から送信された混合データを受信し、データ分離部B411に供給する（ステップS390）。データ分離部B411は、混合データを映像データと音声データとに分離する。そして、データ分離部B411は、分離した映像データを映像出力部B414に供給し、分離した音声データを音声出力部B412に供給する（ステップS400）。

[0091] 映像出力部B414は、映像データに基づく映像をディスプレイB66に表示させる（ステップS410）。また、音声出力部B412は、音声データに基づく音声をスピーカーB67から出力させる（ステップS420）。

[0092] なお、本実施形態では、送受信端末A300と送受信端末B400とが同様の機能を持ち互いに映像を送受信することができる。そのため、上述した送受信端末A300がステップS300からステップS380の処理を行う

のと平行して送受信端末B 4 0 0も同様の処理を行い、送受信端末B 4 0 0がステップS 3 9 0からステップS 4 2 0の処理を行うのと平行して送受信端末A 3 0 0も同様の処理を行うことができる。

[0093] 以上説明したように、本実施形態によるテレビ会議システム1 aは、少なくとも一組以上の端末装置（例えば、送受信端末A 3 0 0及び送受信端末4 0 0）がネットワーク3を介して映像を送受信する。

例えば、送受信端末A 3 0 0が備えるカメラA 6 0は、送受信端末A 3 0 0を利用して会議に出席している出席者と、送受信端末4 0 0（他の端末装置）から取得した映像を表示可能なディスプレイA 6 3（表示部）と、を含むカメラ映像3 0 1 a（第一の映像）を撮影する。また、送受信端末A 3 0 0は、ディスプレイ検出部A 3 0 2（位置検出部）と、回転処理部A 3 0 3と、顔検出部A 3 0 4と、合成映像生成部A 3 0 6（生成部）と、を備えている。

ディスプレイ検出部A 3 0 2は、カメラA 6 0により撮影されたカメラ映像3 0 1 aからディスプレイA 6 3の位置を検出する。回転処理部A 3 0 3は、ディスプレイ検出部A 3 0 2が検出したディスプレイA 6 3の位置が垂直方向で上部且つ水平方向で中央となるようにカメラ映像3 0 1 aを回転させる。顔検出部A 3 0 4は、回転処理部A 3 0 3が回転させたカメラ映像3 0 1 bから顔を検出する。そして、合成映像生成部A 3 0 6は、顔検出部A 3 0 4が検出した顔の映像を、会議で使用する資料の資料映像（第二の映像）の映像領域の周囲に配置する際に、回転処理部A 3 0 3が回転させたカメラ映像3 0 1 bにおける当該顔の相対位置に応じて配置する。

[0094] このように、本実施形態によるテレビ会議システム1 bは、画角の広いカメラA 6 0を天井上部方向に向けて設置し、ディスプレイA 6 3の位置が垂直方向で上部且つ水平方向で中央となるようにカメラA 6 0により撮影されたカメラ映像3 0 1 aを回転させ、顔の相対位置関係を保持したまま顔映像を資料映像の周囲に表示させる。これにより、送受信端末A 3 0 0と送受信端末B 4 0 0とが向かい合った関係の映像を送受できるため、会議の出席者

が相手側の出席者との位置関係を直観的に認識して対話することが可能となるとともに、会議の資料映像の視認性も確保できる。よって、本実施形態によれば、会議の資料映像の視認性を保ちつつ、相手側の出席者の状況を把握することができる。

[0095] また、本実施形態では、送受信端末A300及び送受信端末B400が、ディスプレイA63の位置が垂直方向で上部且つ水平方向で中央となるようにカメラ映像301aを回転させるため、カメラA60を設置する際にカメラA60の向きを合わせる必要がなく、設置の際の手間を省くことができる。

[0096] なお、本実施形態では、送信側の機能と受信側の機能との両方を有する送受信端末A300及び送受信端末B400がネットワーク3を介して映像を送受信するテレビ会議システム1aについて説明したが、これに限られるものではなく、第1の実施形態で説明したような送信側の機能を有する端末装置と受信側の機能を有する端末装置とによるテレビ会議システム1aとしてもよい。例えば、送受信端末A300は、図13に示す送受信端末A300が備える構成のうちの送信側の機能を有する構成のみを備えてもよいし、受信側の機能を有する構成のみを備えてもよい。一方、送受信端末B400は、受信側の機能を有する構成のみを備えてもよいし、送信側の機能を有する構成のみを備えてもよい。

[0097] また、カメラA60に画角180度の魚眼カメラを用いた例を説明したが、カメラA60が設置された部屋で、ディスプレイA63と、会議に出席している出席者のほぼ全員とを映すための画角のカメラであれば、画角180度の魚眼カメラに限られるものではない。ここで、ほぼ全員と記述しているのは、出席者のいる場所によっては、画角180度の魚眼カメラを用いたとしても出席者の全員が映ることが保障されるものではないためである。例えば、カメラA60は、テーブル70に設置されているディスプレイA63と、テーブル70の周囲近傍にいる出席者の全員を映すことが可能な程度の画角を有するカメラであることが望ましい。また、複数のカメラを組み合わせ

て、パノラマ映像、全周囲映像を生成して映像を出力するカメラであってもよい。また、カメラB 6 4もカメラA 6 0と同様である。

[0098] また、マイクA 6 1に無指向性のマイクを用いた例を説明したが、カメラA 6 0の周囲で発せられた音声を入力することができれば、いずれのマイクを用いてもよい。また、マイクB 6 5マイクA 6 1と同様である。

[0099] <第3の実施形態>

次に、本発明の第3の実施形態について説明する。

本実施形態では、テレビ会議の出席者自身が顔検出されているか否かを容易に判断できるようにする形態を説明する。

[0100] 図23は、本実施形態によるテレビ会議システム1bの概略構成の一例を示すブロック図である。この図に示すテレビ会議システム1bの構成は、図8に示すテレビ会議システム1aの送受信端末A 3 0 0及び送受信端末B 4 0 0に代えて、送受信端末A 3 5 0及び送受信端末B 4 5 0を備えている。送受信端末A 3 5 0は、図8の送受信端末A 3 0 0の構成に加えて、アイコン重畳部A 3 1 3を備えている。また、送受信端末B 4 5 0は、図8の送受信端末B 4 0 0に加えて、アイコン重畳部B 4 1 3を備えている。

[0101] アイコン重畳部A 3 1 3は、データ分離部A 3 1 1が分離した映像データの映像に対して、顔検出部A 3 0 4が検出した顔の位置にアイコンを重畳する。そして、映像出力部A 3 1 4は、アイコン重畳部A 3 1 3がアイコンを重畳した映像をディスプレイA 6 3に表示させる。同様に、アイコン重畳部B 4 1 3は、データ分離部B 4 1 1が分離した映像データの映像に対して、顔検出部B 4 0 4が検出した顔の位置にアイコンを重畳する。そして、映像出力部B 4 1 4は、アイコン重畳部B 4 1 3がアイコンを重畳した映像をディスプレイB 6 6に表示させる。つまり、アイコン重畳部A 3 1 3及びアイコン重畳部B 4 1 3は、テレビ会議に出席している出席者が見ている映像に、その出席者自身が顔検出されている場合にはアイコンが表示されるようにする。

[0102] (テレビ会議システム1bの処理)

次に、図 2 4 から図 2 6 を参照して、本実施形態によるテレビ会議システム 1 b における処理を詳しく説明する。なお、本実施形態のテレビ会議システム 1 b の使用環境は、第 2 の実施形態で説明した図 1 4 に示す例と同様であるものとして説明する。

[0103] 図 2 4 は、本実施形態によるテレビ会議システム 1 b における処理の流れを示すフローチャートである。図 2 4 に示す処理は、図 1 5 に示す処理に対して、ステップ S 4 0 5 の処理が追加された点が異なる。なお、図 2 4 において、図 1 5 の各処理に対応する処理には同一の符号を付けており、その説明を省略し、ここでは、相違点であるステップ S 4 0 5 の処理について説明する。

[0104] 送受信端末 B 4 5 0 のアイコン重畳部 B 4 1 3 は、データ分離部 B 4 1 1 が分離した映像データの映像に対して、送受信端末 B 4 5 0 の顔検出部 B 4 0 4 が検出した出席者 H 8 3 の顔に対応するアイコンを重畳する（ステップ S 4 0 5）。このアイコンを重畳する処理について、以下に詳しく説明する。

[0105] 図 2 5 は、送受信端末 B 4 5 0 の第一映像取得部 B 4 0 1 がカメラ B 6 4 から取得したカメラ映像 4 0 1 a を示す図である。このカメラ映像 4 0 1 a は、データ分離部 B 4 1 1 が分離した映像データの映像である。カメラ映像 4 0 1 a には、ディスプレイ B 6 6 と、出席者 H 8 3 とが撮影されており、顔検出部 B 4 0 4 により出席者 H 8 3 の顔領域 6 0 0 が検出されている。

[0106] なお、このカメラ映像 4 0 1 a では、ディスプレイ B 6 6 の位置が垂直方向で上部且つ水平方向で中央となっている。ディスプレイ B 6 6 の位置が垂直方向で上部且つ水平方向で中央となっていない場合には、回転処理部 B 4 0 3 が回転処理を行った後のカメラ映像 4 0 1 b を用いて顔検出部 B 4 0 4 による顔検出処理が行われてもよい。

[0107] ここで、カメラ映像 4 0 1 a（4 0 1 b）の中心座標 6 0 2 から出席者 H 8 3 の顔領域 6 0 0 の中心座標 6 0 1 を通る直線 6 0 5 を、中心座標 6 0 2 を通るカメラ映像 4 0 1 a（4 0 1 b）の上辺 6 0 3 に直交する直線 6 0 4

を基準に左右反転させた直線を直線 606 とする。この場合、カメラ B64 のレンズ中心を基準とした実際の出席者 H83 の顔の相対位置は、カメラ映像 401a (401b) の中心座標 602 を基準とした直線 606 の線上の位置として表すことができる。

[0108] 図 26 は、資料映像と顔映像とを合成した合成映像にアイコン 99 を重畳した映像を示す図である。アイコン重畳部 B413 は、資料映像 101d と顔映像とを合成した合成映像（図 10 の合成映像 101e 参照）に対して、出席者 H83 の顔に対応するアイコン 99 を重畳する。この場合、アイコン重畳部 B413 は、図 25 に示すカメラ映像 401a (401b) の中心座標 602 に代えて資料映像 101d の映像領域の中心座標 48 を基準とし、中心座標 48 を基準としたときの直線 606 の線上、且つ資料映像 101d の周囲（外側）に、出席者 H83 の顔に対応するアイコン 99 を重畳する。

[0109] ここで、アイコン重畳部 B413 が重畳するアイコン 99 は、送受信端末 B450 の合成映像生成部 B406 が切り出して回転させた顔映像をアイコンとしたものでもよいし、顔映像のアイコン以外に、絵柄、記号、印等のアイコンとしてもよい。また、出席者が予め作成して登録しておいたアイコンを用いるようにしてもよい。

[0110] 以上説明したように、本実施形態によるテレビ会議システム 1b おいて、送受信端末 B450 は、顔検出部 B404 と、データ受信部 B410（受信部）と、アイコン重畳部 B413（重畳部）と、を備えている。顔検出部 B404 は、カメラ B64 が撮影したカメラ映像 401a（又は回転処理後のカメラ映像 401b）から顔を検出する。データ受信部 B410 は、送受信端末 A350（他の端末装置）の合成映像生成部 A306（生成部）が生成した映像を受信する。そして、アイコン重畳部 B413 は、データ受信部 B410 が受信した映像に、顔検出部 B404 が検出した顔に応じてアイコン 99（第一情報）を重畳する。

[0111] このように、本実施形態によるテレビ会議システム 1b は、会議の出席者の顔が検出されているか否かを示すアイコン 99 を、出席者自身が見るディ

スプレイに表示するため、出席者が確認画面に表示を切り替えることなく、資料映像及び相手側の出席者の顔映像が表示されている限られた表示領域の中で、出席者自身の顔検出の状態を容易に確認することができる。

[0112] また、アイコン重畳部 B 4 1 3 は、データ受信部 B 4 1 0 が受信した映像に、顔検出部 B 4 0 4 が検出した顔の位置に応じてアイコン 9 9 を重畳してもよい。これにより、テレビ会議システム 1 c は、会議の出席者が複数人の場合であっても、各出席者が自身のアイコン 9 9 を区別することができるため、出席者自身の顔検出の状態を容易に確認することができる。

[0113] また、アイコン重畳部 B 4 1 3 は、アイコン 9 9 に用いる映像および作成した映像の α チャンネルを設定し、アイコン 9 9 の透明度をデータ分離部 B 4 1 1 が分離した映像データより、下げて重畳してもよい。

これにより、アイコン 9 9 が顔映像の合成位置と重なってしまった場合でも、顔映像の視認性低下を抑制することができる。

[0114] なお、本実施形態では、送受信端末 A 3 5 0 と送受信端末 B 4 5 0 とが同様の機能を持ち互いに映像を送受信することができる。そのため、上述した送受信端末 B 4 5 0 のアイコン重畳部 B 4 1 3 行う処理と同様の処理を、送受信端末 A 3 5 0 のアイコン重畳部 A 1 1 3 も行うことができる。

[0115] <第 4 の実施形態>

次に、本発明の第 4 の実施形態について説明する。

本実施形態では、テレビ会議において、相手側の出席者がディスプレイに表示されている資料を見ている否かを容易に判断できるようにする形態を説明する。

[0116] 図 2 7 は、本実施形態によるテレビ会議システム 1 c の概略構成の一例を示すブロック図である。この図において、図 2 3 の各部に対応する構成には同一の符号を付け、その説明を省略する。

[0117] 図 2 7 に示すテレビ会議システム 1 c の構成は、図 2 3 に示すテレビ会議システム 1 b の送受信端末 A 3 5 0 及び送受信端末 B 4 5 0 に代えて、送受信端末 A 3 6 0 及び送受信端末 B 4 6 0 を備えている。送受信端末 A 3 6 0

は、図 23 の送受信端末 A 350 の構成に対して、顔方向推定部 A 361 を更に備えている点と、合成映像生成部 A 362 が合成映像生成部 A 306 の機能に加えて出席者の顔方向に応じた情報を合成する点とが異なる。同様に、送受信端末 B 460 は、図 23 の送受信端末 B 450 の構成に対して、顔方向推定部 B 461 を更に備えている点と、合成映像生成部 B 462 が合成映像生成部 B 406 の機能に加えて出席者の顔方向に応じた情報を合成する点とが異なる。

[0118] 顔方向推定部 A 361 は、顔検出部 A 304 が検出した顔領域の映像に基づいて、顔が向いている方向（顔方向）を推定する。例えば、顔方向推定部 A 361 は、顔検出部 A 304 が検出した顔領域から顔の特徴点である鼻や目、口の位置を抽出し、抽出した特徴点から求めた特徴量に基づいて、顔方向が顔正面から左右方向に何度の方角に向いているかの角度を推定する。

[0119] ここで特徴点とは、鼻の頂点、目の端点、口の端点等の座標のことを指し、特徴量とは、これらの座標を基に算出した距離や面積等を指す。例えば、顔方向推定部 A 361 は、抽出した複数の特徴点から、特徴点間の距離、3 つの特徴点で囲まれる面積、輝度分布等の特徴量を求める。そして、顔方向推定部 A 361 は、予め複数人の顔を用いて取得しておいた、顔方向の角度に対応する特徴点から求めた特徴量が集約されたデータベースを参照することにより、顔方向の角度を推定する。

[0120] 合成映像生成部 A 362 は、顔検出部 102 が検出した顔の領域を第一映像取得部 101 が取得した映像から切り取り、切り取った顔映像と、第二映像取得部 A 305 が取得した資料映像とを合成するとともに、合成した顔映像に対応する位置に、顔方向推定部 A 361 が推定した当該顔の顔方向に応じたアイコンを重畳した合成映像を生成する。

[0121] なお、送受信端末 B 460 の顔方向推定部 B 461 及び合成映像生成部 B 462 のそれぞれは、送受信端末 A 360 の顔方向推定部 A 361 及び合成映像生成部 A 362 のそれぞれと同様の機能を有している。

[0122] （テレビ会議システム 1c の処理）

次に、本実施形態によるテレビ会議システム 1 c における処理について詳しく説明する。なお、本実施形態のテレビ会議システム 1 c の使用環境は、第 2 の実施形態で説明した図 1 4 に示す例と同様であるものとして説明する。

[0123] 本実施形態による処理の流れは、図 1 5 又は図 2 4 に示す処理において以下の処理を変更すればよい。例えば、図 1 5 又は図 2 4 のステップ S 3 4 0 において、合成映像生成部 A 3 6 2 がカメラ映像 3 0 1 b から顔の領域を切り出して回転を行うとともに、顔方向推定部 A 3 6 1 が当該顔の顔方向を推定すればよい。また、図 1 5 又は図 2 4 のステップ S 3 6 0 において、合成映像生成部 A 3 6 2 が、顔映像と資料映像とを合成するとともに、顔方向に応じたアイコンを重畳した合成映像を生成すればよい。

[0124] 例えば、合成映像生成部 A 3 6 2 は、検出された顔からディスプレイ A 6 3 への方向と、顔方向推定部 A 3 6 1 が推定した当該顔の顔方向とが一致している場合、その旨を示すアイコンを当該顔映像に対応する位置に配置する。即ち、合成映像生成部 A 3 6 2 は、検出された顔がディスプレイ A 6 3 の方向を向いているか否かを示す情報としてのアイコンを重畳した映像を生成する。

[0125] 以下、図 2 8 及び図 2 9 を参照して検出された顔からディスプレイ A 6 3 への方向と、顔方向推定部 A 3 6 1 が推定した当該顔の顔方向との関係について説明する。

[0126] 図 2 8 は、カメラ映像 3 0 1 b における顔の相対位置及び顔方向を示す図である。この図において、回転処理部 A 3 0 3 による回転処理後のカメラ映像 3 0 1 b から検出されたディスプレイ A 6 3 の画面の中心座標を座標 7 0 1 とする。また、カメラ映像 3 0 1 b から検出された顔領域 9 6、顔領域 9 7、及び顔領域 9 8 のそれぞれの中心の座標を、座標 7 0 2、座標 7 0 3、及び座標 7 0 4 とする。そして、カメラ映像 3 0 1 b の映像領域の中心座標 9 2 から、座標 7 0 1、座標 7 0 2、座標 7 0 3、及び座標 7 0 4 のそれぞれを通る直線を、直線 7 0 5、直線 7 0 6、直線 7 0 7、及び直線 7 0 8 と

する。

[0127] 例えば、顔検出部 A 3 0 4 により、出席者 E 8 0、出席者 F 8 1、出席者 G 8 2 のそれぞれの顔が、中心座標 9 2 を回転中心とした回転方向の角度が 2 7 0 度、1 0 度、8 0 度の位置に検出されたとする。また、中心座標 9 2 を回転中心としたディスプレイ A 6 3 の回転方向の角度は 1 8 0 度である。この場合、直線 7 0 5、直線 7 0 6、直線 7 0 7、及び直線 7 0 8 のそれぞれの回転方向の角度は、1 8 0 度、2 7 0 度、1 0 度、8 0 度となる。なお、カメラ映像 3 0 1 b における回転方向の角度は、中心座標 9 2 からカメラ映像 3 0 1 b の下辺 7 0 9 に直交する直線上の位置を回転方向の角度を 0 度とし、反時計まわりに 0 度から 3 5 9 度まで検出されるものとする。

[0128] また、顔方向については、カメラ映像 3 0 1 b において顔正面に向かって右向き（出席者自身からみて左向き）を正（+）の角度、左向き（出席者からみて右向き）を負（-）の角度で表す。例えば、顔方向推定部 A 3 6 1 は、正負それぞれ 6 0 度までの角度範囲まで推定できるものとする。ここでは、顔方向推定部 A 3 6 1 により、出席者 E 8 0、出席者 F 8 1、及び出席者 G 8 2 のそれぞれの顔方向が、0 度、3 0 度、4 0 度であると推定されたとする。

[0129] 図 2 9 は、部屋 5 0 1 を天井側から見下ろしたときの状態を示す模式図である。この図には、図 2 8 に示すカメラ映像 3 0 1 b に撮影されている各位置関係に対応させて、同一の符号を用いて表している。なお、説明を容易にするため、カメラ A 6 0 から、ディスプレイ A 6 3、出席者 E 8 0 の顔、出席者 F 8 1 の顔、及び出席者 G 8 2 の顔のそれぞれまでの距離が等しいものとして説明する。

[0130] 図 2 9 のカメラ A 6 0 のレンズの中心座標 9 2 は、図 2 8 のカメラ映像 3 0 1 b の映像領域の中心座標 9 2 に対応する。この図 2 9 では、出席者 E 8 0、出席者 F 8 1、及び出席者 G 8 2 のそれぞれの顔の回転方向の角度は、時計回りでそれぞれ 2 7 0 度、1 0 度、8 0 度となる位置になる。また、ディスプレイ A 6 3 の回転方向の角度は、時計回りで 1 8 0 度となる位置にな

る。座標 702、座標 703、及び座標 704 のそれぞれは、出席者 E80、出席者 F81、及び出席者 G82 のそれぞれの顔中心に対応する。

[0131] 出席者 E80 の顔中心の座標 702 と中心座標 92 とを結ぶ直線 706 は、ディスプレイ A63 の中心の座標 701 と中心座標 92 とを結ぶ直線 705 に直交する。また、ディスプレイ A63 の画面の中心座標 701 と出席者 E80 の顔中心の座標 702 とを結ぶ直線を直線 711 とすると、直線 711 と直線 706 とのなす角（鋭角側）は 45 度（出席者 E80 の顔方向としては負方向に 45 度）である。よって、出席者 E80 の顔方向が -45 度の場合に、出席者 E80 の顔がディスプレイ A63 の方向を向いていることになる。ここで、顔方向推定部 A361 により、出席者 E80 の顔方向が 0 度であると推定されたため、出席者 E80 の顔がディスプレイ A63 の方向を向いていないと判断できる。

[0132] また、出席者 F81 の顔中心の座標 703 と中心座標 92 とを結ぶ直線 707 と、ディスプレイ A63 の中心の座標 701 と中心座標 92 とを結ぶ直線 705 と、のなす角（鋭角側）は 170 度である。また、ディスプレイ A63 の画面の中心座標 701 と出席者 F81 の顔中心の座標 703 とを結ぶ直線を直線 712 とすると、直線 712 と直線 707 とのなす角（鋭角側）は 5 度（出席者 F81 の顔方向として正方向に 5 度）である。よって、出席者 F81 の顔方向が 5 度の場合に、出席者 F81 の顔がディスプレイ A63 の方向を向いていることになる。ここで、顔方向推定部 A361 により、出席者 F81 の顔方向が 30 度であると推定されたため、出席者 F81 の顔がディスプレイ A63 の方向を向いていないと判断できる。

[0133] また、出席者 G82 の顔中心の座標 704 と中心座標 92 とを結ぶ直線 708 と、ディスプレイ A63 の中心の座標 701 と中心座標 92 とを結ぶ直線 705 とのなす角（鋭角側）は 100 度である。また、ディスプレイ A63 の画面の中心座標 701 と出席者 G82 の顔中心の座標 704 とを結ぶ直線を直線 713 とすると、直線 713 と直線 708 とのなす角（鋭角側）は 40 度（出席者 G82 の顔方向として正方向に 40 度）である。よって、出

席者G 8 2の顔方向が4 0度の場合に、出席者G 8 2の顔がディスプレイA 6 3の方向を向いていることになる。ここで、顔方向推定部A 3 6 1により、出席者G 8 2の顔方向が4 0度であると推定されたため、推定された顔方向と出席者G 8 2の顔からディスプレイA 6 3への方向とが一致し、出席者G 8 2の顔がディスプレイA 6 3の方向を向いていると判断できる。

[0134] 例えば、合成映像生成部A 3 6 2は、検出された顔からディスプレイA 6 3への方向と、顔方向推定部A 3 6 1が推定した当該顔の顔方向とが一致している出席者G 8 2の顔映像に対応する位置に、顔方向がディスプレイA 6 3の方向を向いていることを示すアイコンを重畳する。一方、合成映像生成部A 3 6 2は、検出された顔からディスプレイA 6 3への方向と、顔方向推定部A 3 6 1が推定した当該顔の顔方向とが一致していない出席者E 8 0及び出席者F 8 1の顔映像に対応する位置には、顔方向がディスプレイA 6 3の方向を向いていないことを示すアイコンを重畳する。

[0135] 図3 0は、顔方向に応じたアイコンを配置した合成映像3 0 1 dの一例を示す図である。図示する例では、アイコン8 0 0、アイコン8 0 1、アイコン8 0 2のそれぞれが、出席者E 8 0、出席者F 8 1、及び出席者G 8 2のそれぞれの顔領域（横1 0 0×縦1 0 0画素）に対応する位置に重畳される。例えば、資料映像1 0 1 dと、顔領域9 6、顔領域9 7、及び顔領域9 8のそれぞれとが接している辺上にそれぞれのアイコンが重畳される。

[0136] 例えば、出席者の顔方向がディスプレイA 6 3の方向と一致している場合と一致していない場合とで、色又は濃度の異なるアイコンが表示される。ここでは、出席者G 8 2の顔方向がディスプレイA 6 3の方向と一致しており、アイコン8 0 2は、一致していることを示すアイコン（例えば、黒丸）である。一方、アイコン8 0 0及びアイコン8 0 1は、出席者の顔方向がディスプレイA 6 3の方向と一致していないことを示すアイコン（例えば、白丸）である。

[0137] なお、出席者の顔方向がディスプレイA 6 3の方向と一致している場合と一致していない場合とで、アイコンの種類を変えてもよい。

- [0138] また、出席者の顔方向に応じてアイコンの位置を変えてもよい。図30において、合成映像301dの映像領域の中心座標803は、図29のカメラA60のレンズの中心座標92に対応する。また、この中心座標803を基準としたときの直線705の180度側の直線上、且つ資料映像101dの周囲（外側）にある座標804を、図29のディスプレイA63の画面の中心座標701に対応する位置とする。
- [0139] つまり、図30に示す例では、各出席者の顔領域からディスプレイA63の中心座標に相当する座標804への方が、各出席者の顔に対するディスプレイA63の方向（顔方向0度）となる。そして、資料映像101dと各顔領域とが接している辺上において、各出席者の顔方向と各出席者の顔からディスプレイA63への方角との差分に応じた位置にそれぞれのアイコンが重畳される。
- [0140] 例えば、出席者G82の顔方向がディスプレイA63の方向と一致しているため、資料映像101dと出席者G82の顔領域98とが接する辺上において、顔領域98の中心と座標804とを結ぶ直線713に交わる位置に、アイコン802が重畳される。
- [0141] 一方、出席者E80の顔方向はディスプレイA63の方向と一致していない。そのため、アイコン800は、資料映像101dと出席者E80の顔領域96とが接する辺上において、顔領域96の中心と座標804とを結ぶ直線711に交わる位置に対して、顔方向に応じて離れた位置に重畳される。例えば、図示するように、アイコン800は、資料映像101dと顔領域96とが接する辺上において、直線711に交わる位置から出席者F80の顔の向いている側に離れた位置に重畳される。なお、直線711に交わる位置からアイコン800が重畳される位置までの距離は、例えば、顔方向とディスプレイA63の方向との差の大きさに応じて定まる。
- [0142] 同様に、出席者F81の顔方向はディスプレイA63の方向と一致していない。そのため、アイコン801は、資料映像101dと出席者F81の顔領域97とが接する辺上において、顔領域97の中心と座標804とを結ぶ

直線 7 1 2 に交わる位置に対して、顔方向に応じて離れた位置に重畳される。

[0143] 以上説明したように、本実施形態によるテレビ会議システム 1 c において、例えば、送受信端末 A 3 6 0 が備えるカメラ A 6 0 は、送受信端末 A 3 6 0 を利用して会議に出席している出席者と、送受信端末 B 4 6 0（他の端末装置）から取得した映像を表示可能なディスプレイ A 6 3（表示部）と、を含むカメラ映像 3 0 1 a（第一の映像）を撮影する。

送受信端末 A 3 6 0 は、ディスプレイ検出部 A 3 0 2（位置検出部）と、顔検出部 A 3 0 4 と、顔方向推定部 A 3 6 1 と、合成映像生成部 A 3 6 2（生成部）と、を備えている。ディスプレイ検出部 A 3 0 2 は、カメラ A 6 0 により撮影されたカメラ映像 3 0 1 a からディスプレイ A 6 3 の位置を検出する。顔検出部 A 3 0 4 は、カメラ A 6 0 により撮影されたカメラ映像 3 0 1 a（又は、回転処理部 A 3 0 3 が回転させたカメラ映像 3 0 1 b でもよい）から顔を検出する。顔方向推定部 A 3 6 1 は、顔検出部 A 3 0 4 が検出した顔に基づいて、当該顔の顔方向を推定する。そして、合成映像生成部 A 3 6 2 は、顔検出部 A 3 0 4 が検出した顔の位置からディスプレイ検出部 A 3 0 2 が検出したディスプレイ A 6 3 の位置への方向と、顔方向推定部 A 3 6 1 が推定した顔方向との関係に応じて、顔検出部 A 3 0 4 が検出した顔の映像に対応する位置にアイコン（第二情報）を重畳した映像を生成する。

[0144] このように、本実施形態によるテレビ会議システム 1 c は、例えば、送受信端末 A 3 6 0 を利用して会議に出席している出席者 E 8 0、出席者 F 8 1、及び出席者 G 8 2 がディスプレイ A 6 3 の方向を向いているか否かを示すアイコンを、送受信端末 B 4 6 0 を利用して会議に出席している出席者 H 8 3 が見るディスプレイ B 6 6 に表示させる。これにより、送受信端末 B 4 6 0 のディスプレイ B 6 6 を見ている出席者 H 8 3 は、相手側の出席者 E 8 0、出席者 F 8 1、及び出席者 G 8 2 が資料を見ているか否かがわかるため、より相手側の出席者の状況を知ることができる。また、出席者 H 8 3 が資料を使用して説明している際に、相手側の出席者 E 8 0、出席者 F 8 1、及び

出席者 G 8 2 から注目されているか否かを知ることができる。よって、本実施形態によれば、相手側の出席者の状況を詳しく把握することができる。

[0145] なお、上記説明では、各出席者の顔方向がディスプレイ A 6 3 の方向と一致している場合と一致していない場合とで、アイコンの色又は種類を変えてもよいことを説明したが、完全一致の場合に限られるものではなく、一定範囲内（例えば、 -10 度 ~ 10 度の範囲）であれば一致していると判定されてもよい。また、合成映像生成部 A 3 6 2 は、顔方向がディスプレイ A 6 3 の方向と一致している場合のみ、アイコンを描画するようにしてもよい。

[0146] 以上、この発明の第 1 から第 4 の実施形態を図面を参照して詳述してきたが、具体的な構成はこの実施形態によって限定的に解釈されるものではなく、この発明の要旨を逸脱しない範囲の設計等も含まれ種々の変更が可能である。例えば、上述の第 1 から第 4 の実施形態において説明した各機能は、任意に組み合わせることができる。また、上記実施形態の各構成要素は、任意に取捨選択することができ、取捨選択した構成を具備する形態も本発明に含まれるものである。

[0147] なお、上述した実施形態における送信端末 1 0 0、受信端末 2 0 0、送受信端末 A 3 0 0（A 3 5 0、A 3 6 0）、又は送受信端末 B 4 0 0（4 5 0、4 6 0）の一部または全部の機能をコンピュータで実現するようにしてもよい。その場合、上述の機能を実現するためのプログラムをコンピュータ読み取り可能な記録媒体に記録して、この記録媒体に記録されたプログラムをコンピュータシステムに読み込ませ、実行することによって上述の機能を実現してもよい。なお、ここでいう「コンピュータシステム」とは、送信端末 1 0 0、受信端末 2 0 0、送受信端末 A 3 0 0（A 3 5 0、A 3 6 0）、又は送受信端末 B 4 0 0（4 5 0、4 6 0）に内蔵されたコンピュータシステムであって、OS や周辺機器等のハードウェアを含むものとする。また、「コンピュータ読み取り可能な記録媒体」とは、フレキシブルディスク、光磁気ディスク、ROM、CD-ROM 等の可搬媒体、コンピュータシステムに内蔵されるハードディスク等の記憶装置のことをいう。さらに「コンピュー

タ読み取り可能な記録媒体」とは、インターネット等のネットワークや電話回線等の通信回線を介してプログラムが送信された場合のサーバやクライアントとなるコンピュータシステム内部の揮発性メモリ（RAM）のように、一定時間プログラムを保持しているものも含むものとする。また上記プログラムは、前述した機能の一部を実現するためのものであってもよく、さらに前述した機能をコンピュータシステムにすでに記録されているプログラムとの組み合わせで実現できるものであってもよい。

[0148] また、上述した実施形態における送信端末100、受信端末200、送受信端末A300（A350、A360）、又は送受信端末B400（450、460）の一部、または全部を、LSI（Large Scale Integration）等の集積回路として実現してもよい。送信端末100、受信端末200、送受信端末A300（A350、A360）、又は送受信端末B400（450、460）の各機能ブロックは個別にプロセッサ化してもよいし、一部、または全部を集積してプロセッサ化してもよい。また、集積回路化の手法はLSIに限らず専用回路、または汎用プロセッサで実現してもよい。また、半導体技術の進歩によりLSIに代替する集積回路化の技術が出現した場合、当該技術による集積回路を用いてもよい。

産業上の利用可能性

[0149] 本発明の一態様は、会議の資料映像の視認性を保ちつつ、相手側の出席者の状況を把握することが必要なテレビ会議システムなどに適用することができる。

符号の説明

[0150] 1、1a、1b、1c テレビ会議システム、3、ネットワーク、4 情報端末、4a データ記憶部、4b 端末ディスプレイ、5 ディスプレイ、12 カメラ、13 マイク、14 スピーカー、60 カメラA、61 マイクA、62 スピーカーA、63 ディスプレイA、64 カメラB、65 マイクB、66 ディスプレイB、67 スピーカーB、68 情報端末A、68a データ記憶部A、100 送信端末、101 第一映像取得

部、１０２ 顔検出部、１０３ 第二映像取得部、１０４ 合成映像生成部（生成部）、１０５ 音声取得部、１０６ データ混合部、１０７ データ送信部、２００受信端末、２０８ データ受信部、２０９ データ分離部、２１０ 音声出力部、２１１ 映像出力部、３００、３５０、３６０ 送受信端末Ａ、３０１ 第一映像取得部Ａ、３０２ ディスプレイ検出部Ａ（位置検出部）、３０３ 回転処理部Ａ、３０４ 顔検出部Ａ、３０５ 第二映像取得部Ａ、３０６、３６２ 合成映像生成部Ａ（生成部）、３０７ 音声取得部Ａ、３０８ データ混合部Ａ、３０９ データ送信部Ａ、３１０ データ受信部Ａ、３１１ データ分離部Ａ、３１２ 音声出力部Ａ、３１３ アイコン重畳部Ａ、３１４ 映像出力部Ａ、３６１ 顔方向推定部Ａ、４００、４５０、４６０ 送受信端末Ｂ、４０１ 第一映像取得部Ｂ、４０２ ディスプレイ検出部Ｂ（位置検出部）、４０３ 回転処理部Ｂ、４０４ 顔検出部Ｂ、４０５ 第二映像取得部Ｂ、４０６、４６２ 合成映像生成部Ｂ（生成部）、４０７ 音声取得部Ｂ、４０８ データ混合部Ｂ、４０９ データ送信部Ｂ、４１０ データ受信部Ｂ、４１１ データ分離部Ｂ、４１２ 音声出力部Ｂ、４１３ アイコン重畳部Ｂ、４１４ 映像出力部Ｂ、４６１ 顔方向推定部Ｂ

請求の範囲

- [請求項1] 少なくとも一組以上の端末装置がネットワークを介して映像を送受信するテレビ会議システムであって、
- 前記端末装置は、
- カメラにより撮影された映像である第一の映像から顔を検出する顔検出部と、
- 会議で使用する資料の映像である第二の映像の映像領域の周囲に、前記顔検出部が検出した顔の映像を前記第一の映像における当該顔の相対位置に応じて配置した映像を生成する生成部と、
- を備えるテレビ会議システム。
- [請求項2] 前記カメラは、前記端末装置を利用して前記会議に出席している出席者と、他の端末装置から取得した映像を表示可能な表示部と、を含む前記第一の映像を撮影し、
- 前記端末装置は、
- 前記カメラにより撮影された前記第一の映像から前記表示部の位置を検出する位置検出部と、
- 前記位置検出部が検出した前記表示部の位置が垂直方向で上部且つ水平方向で中央となるように前記第一の映像を回転させる回転処理部と、
- を備え、
- 前記生成部は、
- 前記顔検出部が検出した顔の映像を、前記回転処理部が回転させた前記第一の映像における当該顔の相対位置に応じて配置する、
- 請求項1に記載のテレビ会議システム。
- [請求項3] 前記端末装置は、
- 他の端末装置の前記生成部が生成した映像を受信する受信部と、
- 前記受信部が受信した映像に、前記顔検出部が検出した顔に応じて第一情報を重畳する重畳部と、

を備える請求項 1 又は 2 に記載のテレビ会議システム。

[請求項4]

前記重畳部は、

重畳する第一情報の透明度を、前記受信部が受信した映像より下げて重畳する

請求項 3 に記載のテレビ会議システム。

[請求項5]

前記カメラは、前記端末装置を利用して前記会議に出席している出席者と、他の前記端末装置から取得した映像を表示可能な表示部と、を含む前記第一の映像を撮影し、

前記端末装置は、

前記カメラにより撮影された前記第一の映像から前記表示部の位置を検出する位置検出部と、

前記顔検出部が検出した顔に基づいて、当該顔の顔方向を推定する顔方向推定部と、

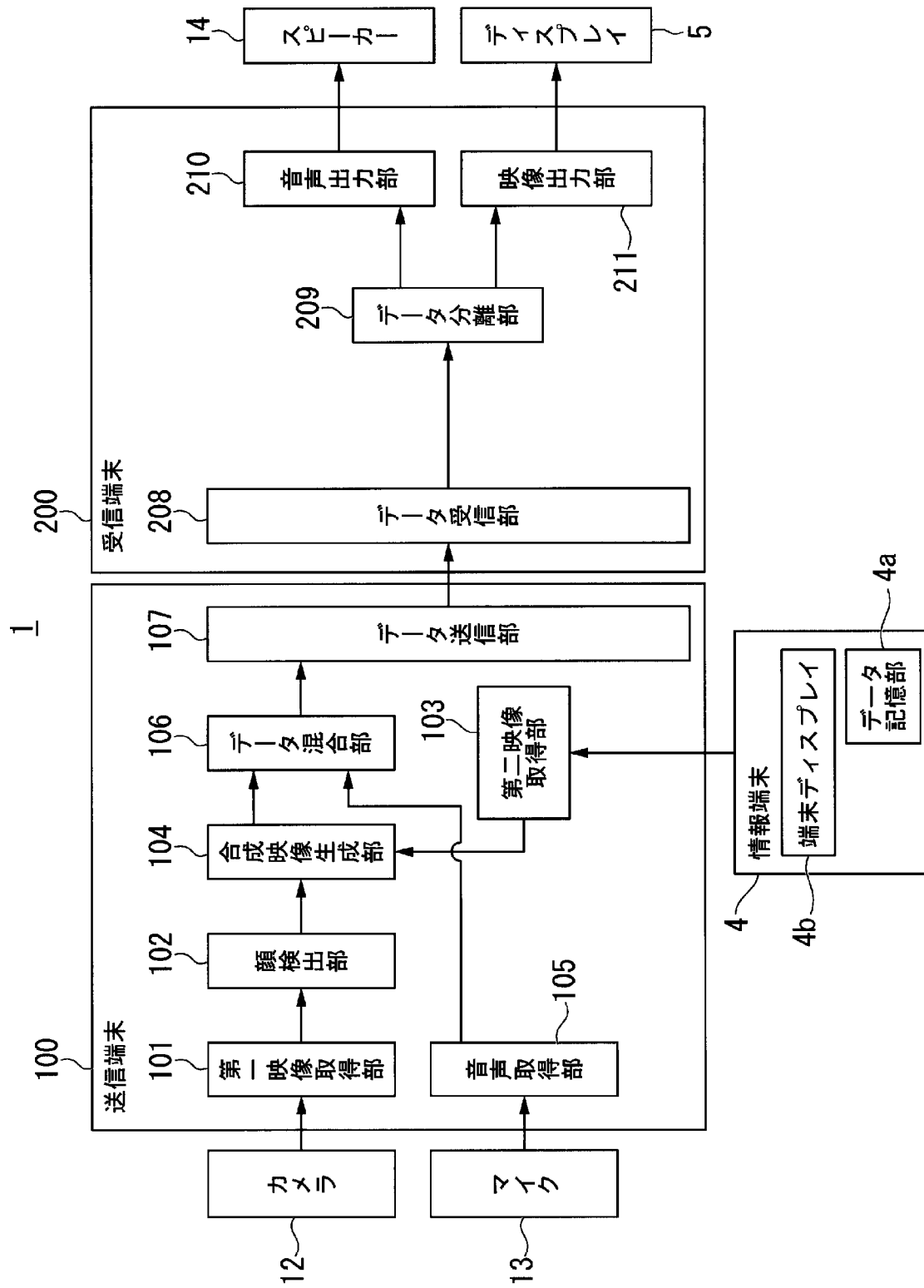
を備え、

前記生成部は、

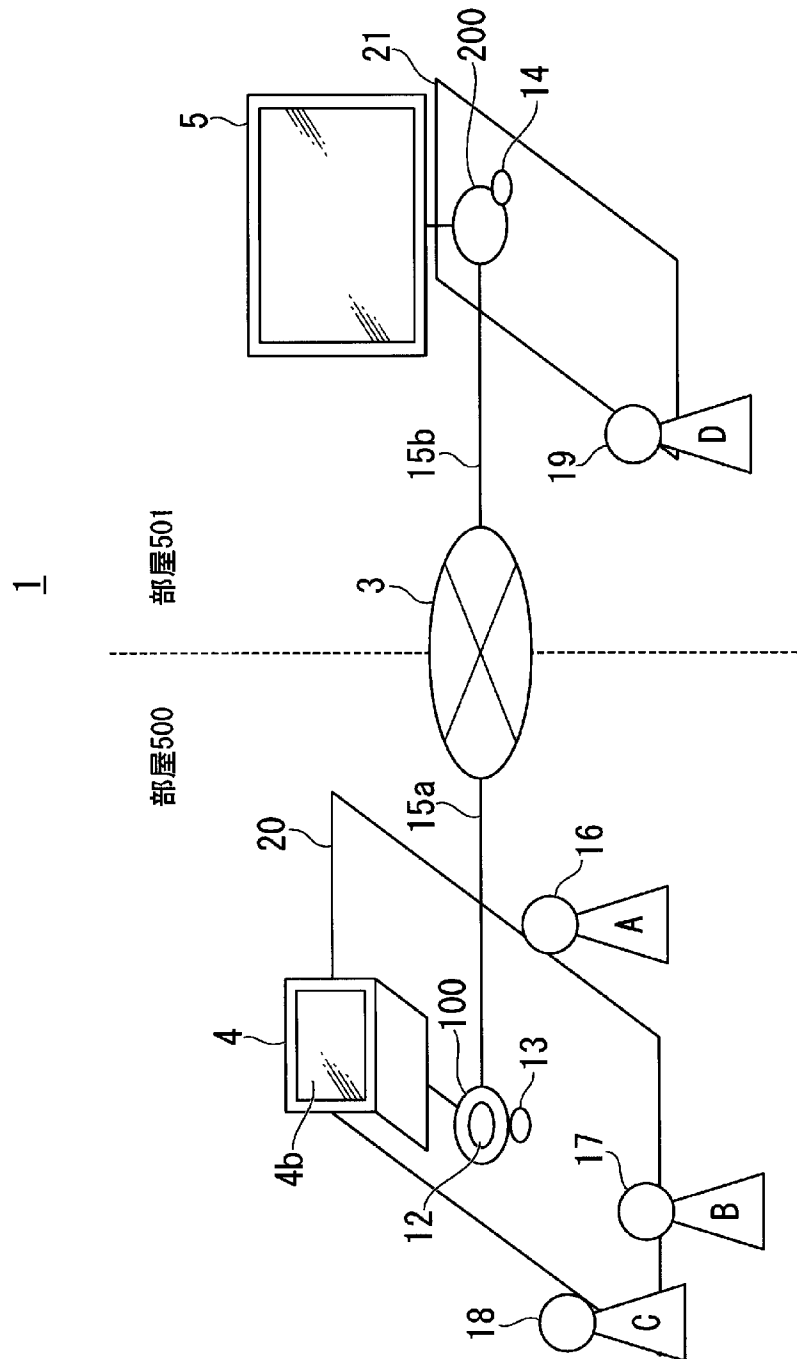
前記顔検出部が検出した顔の位置から前記位置検出部が検出した前記表示部の位置への方向と、前記顔方向推定部が推定した前記顔方向との関係に応じて、前記顔検出部が検出した顔の映像に対応する位置に第二情報を重畳した映像を生成する、

請求項 1 から 4 の何れか一項に記載のテレビ会議システム。

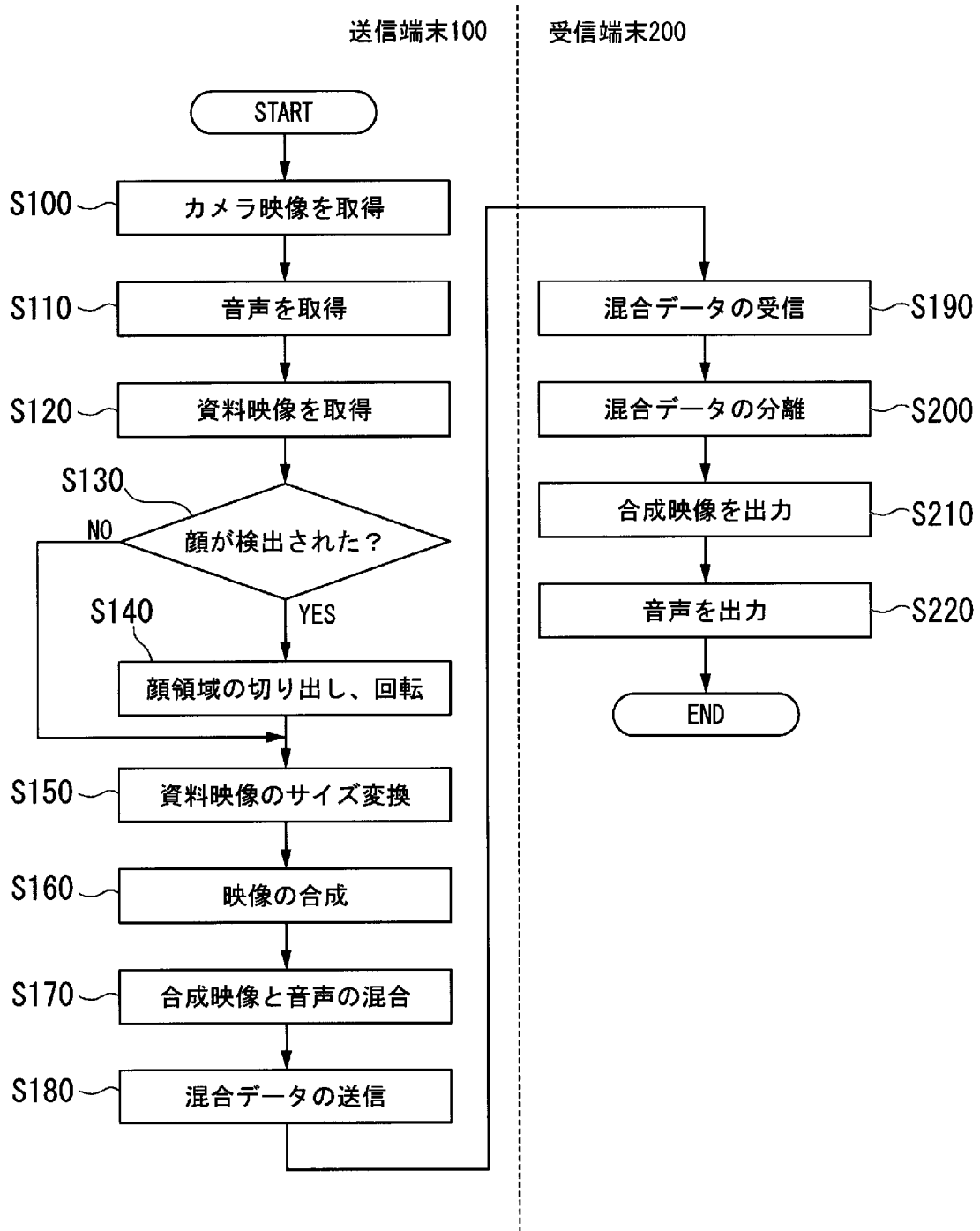
[図1]



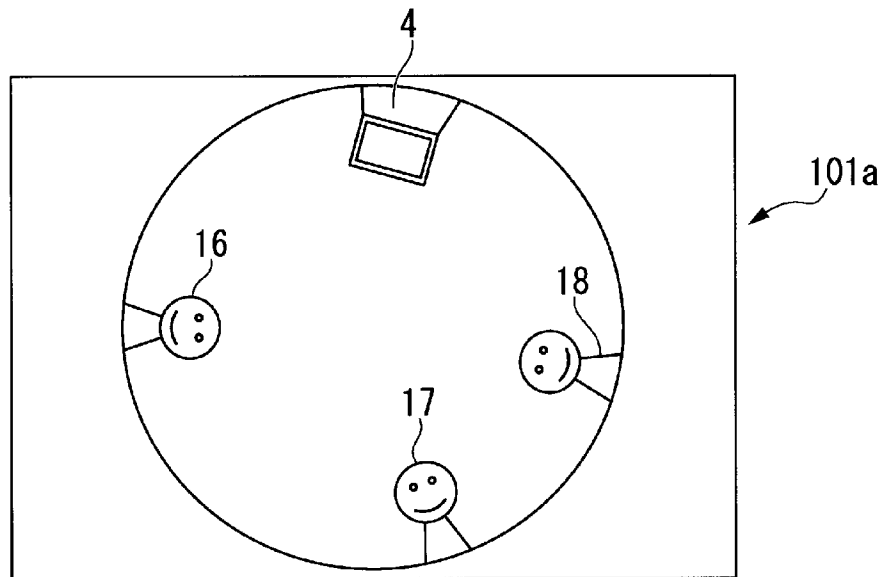
[図2]



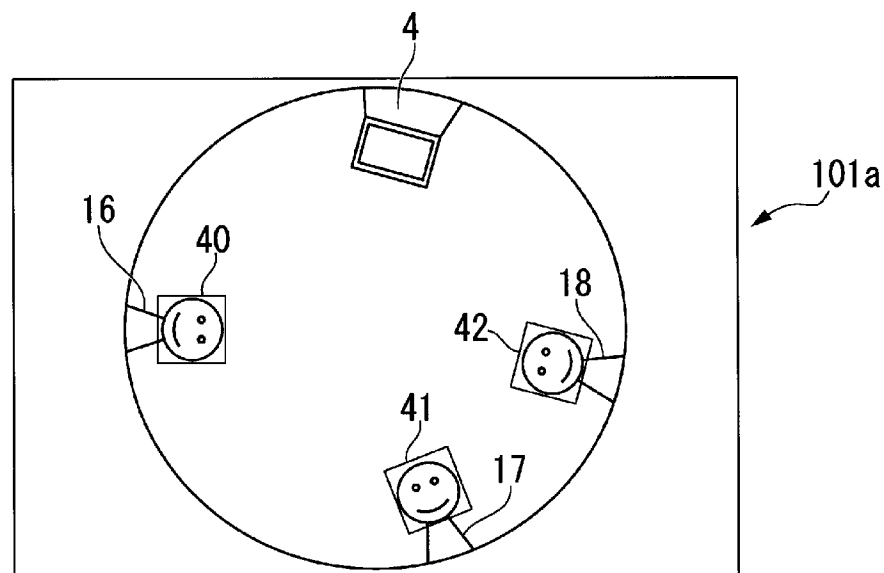
[図3]



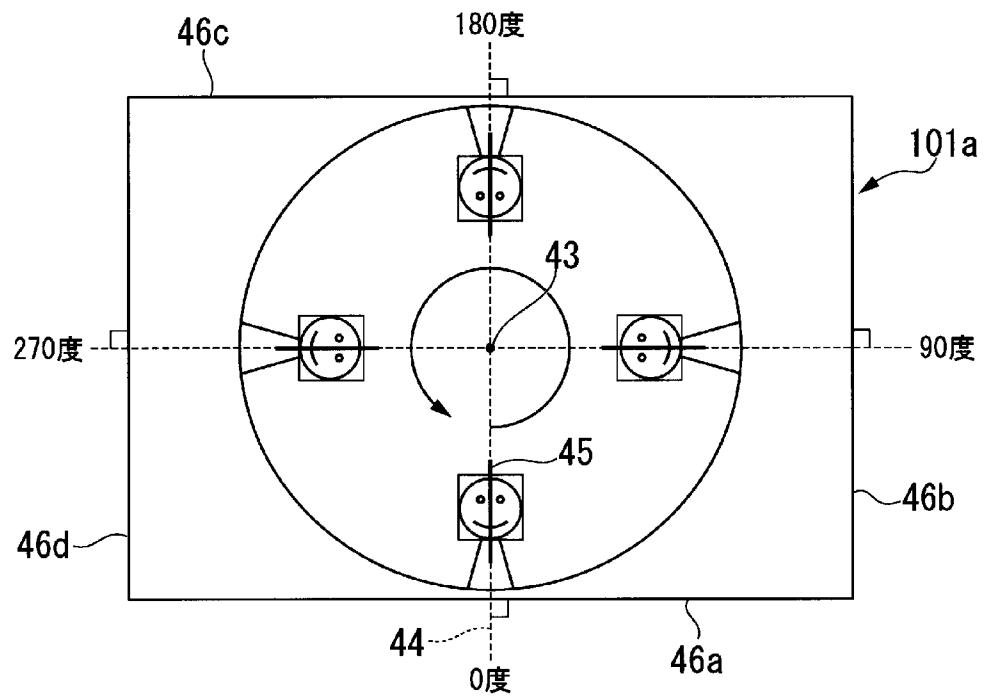
[図4]



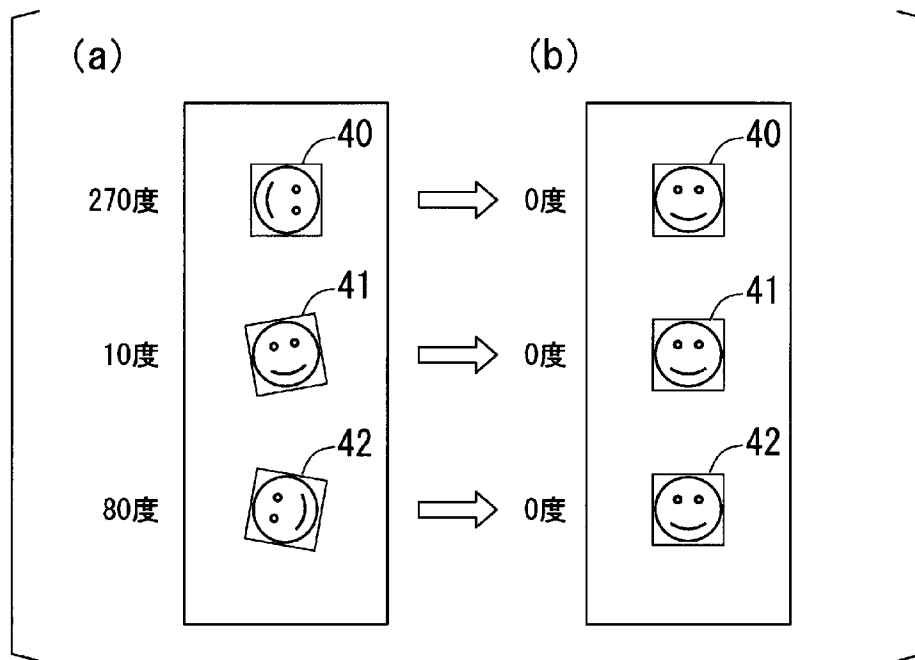
[図5]



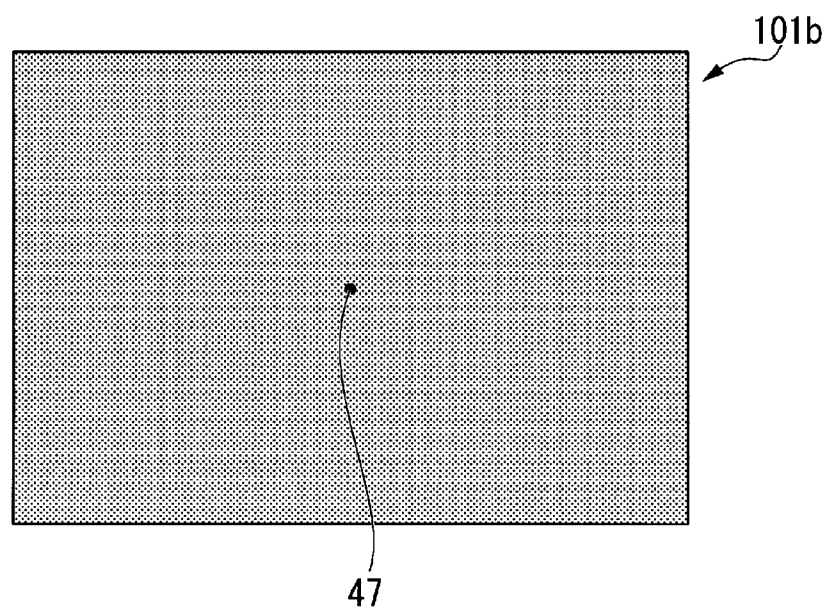
[図6]



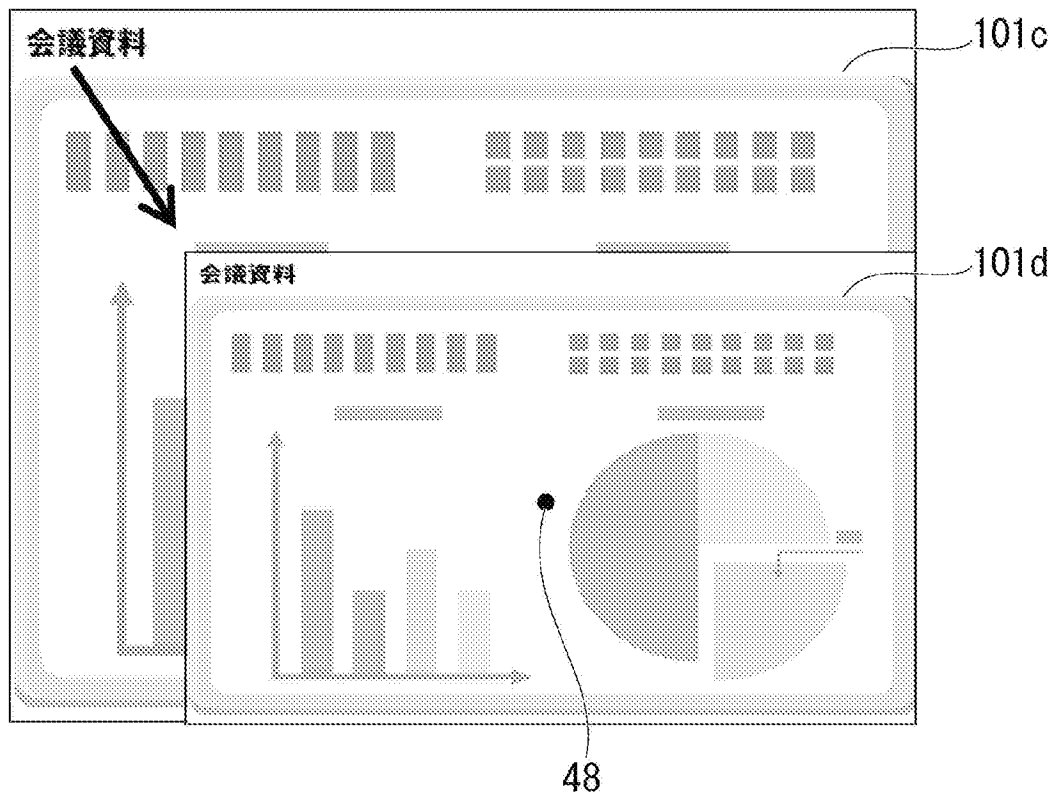
[図7]



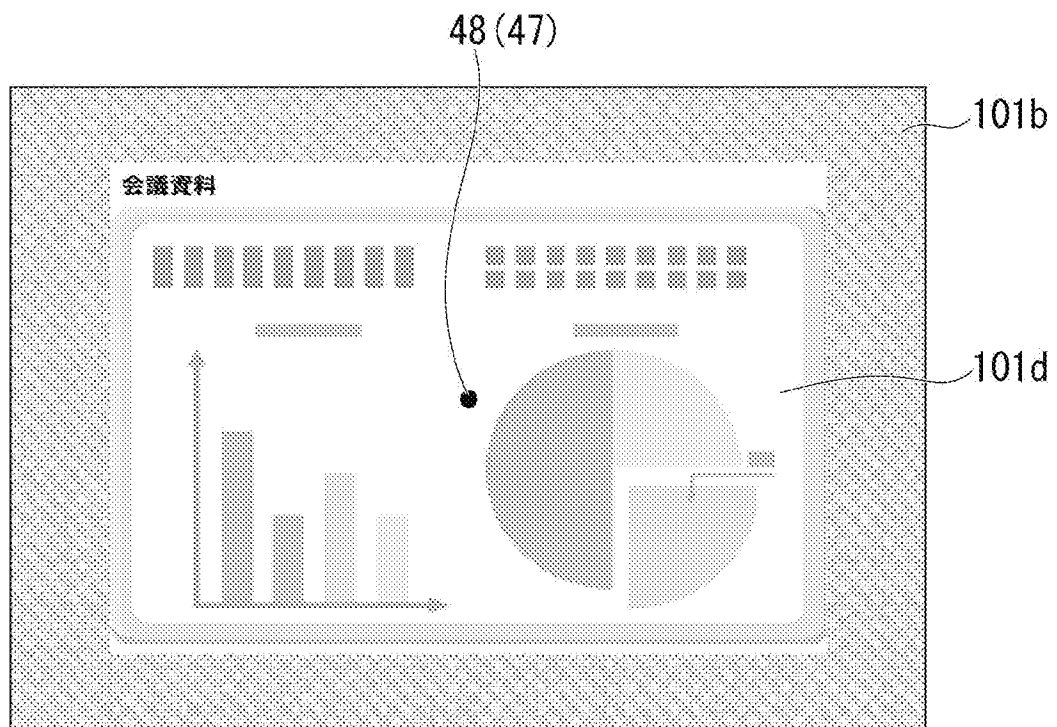
[図8]



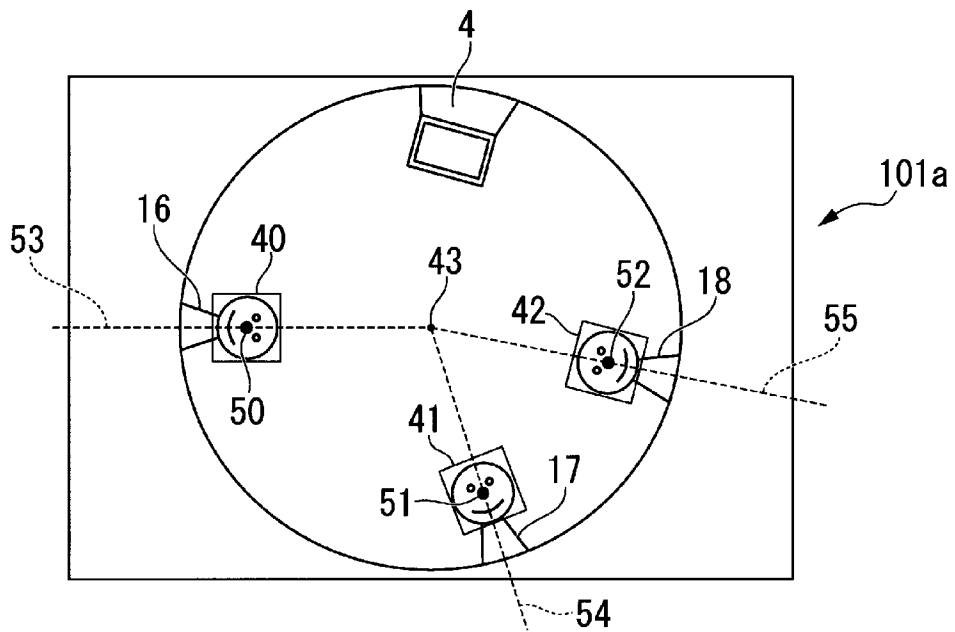
[図9]



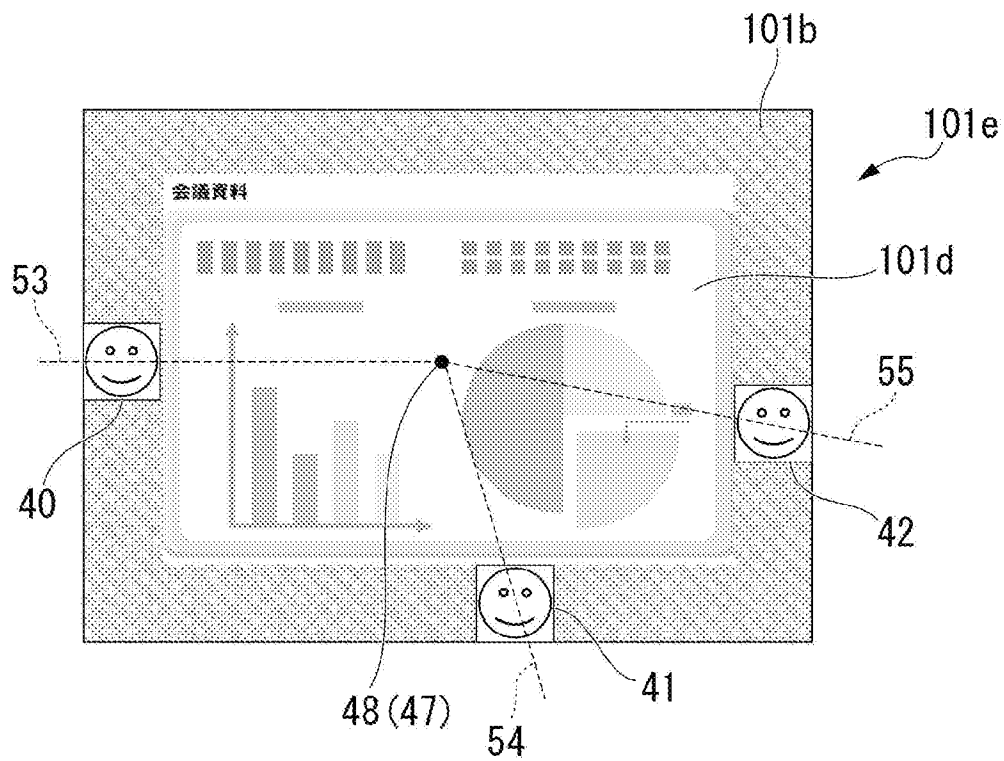
[図10]



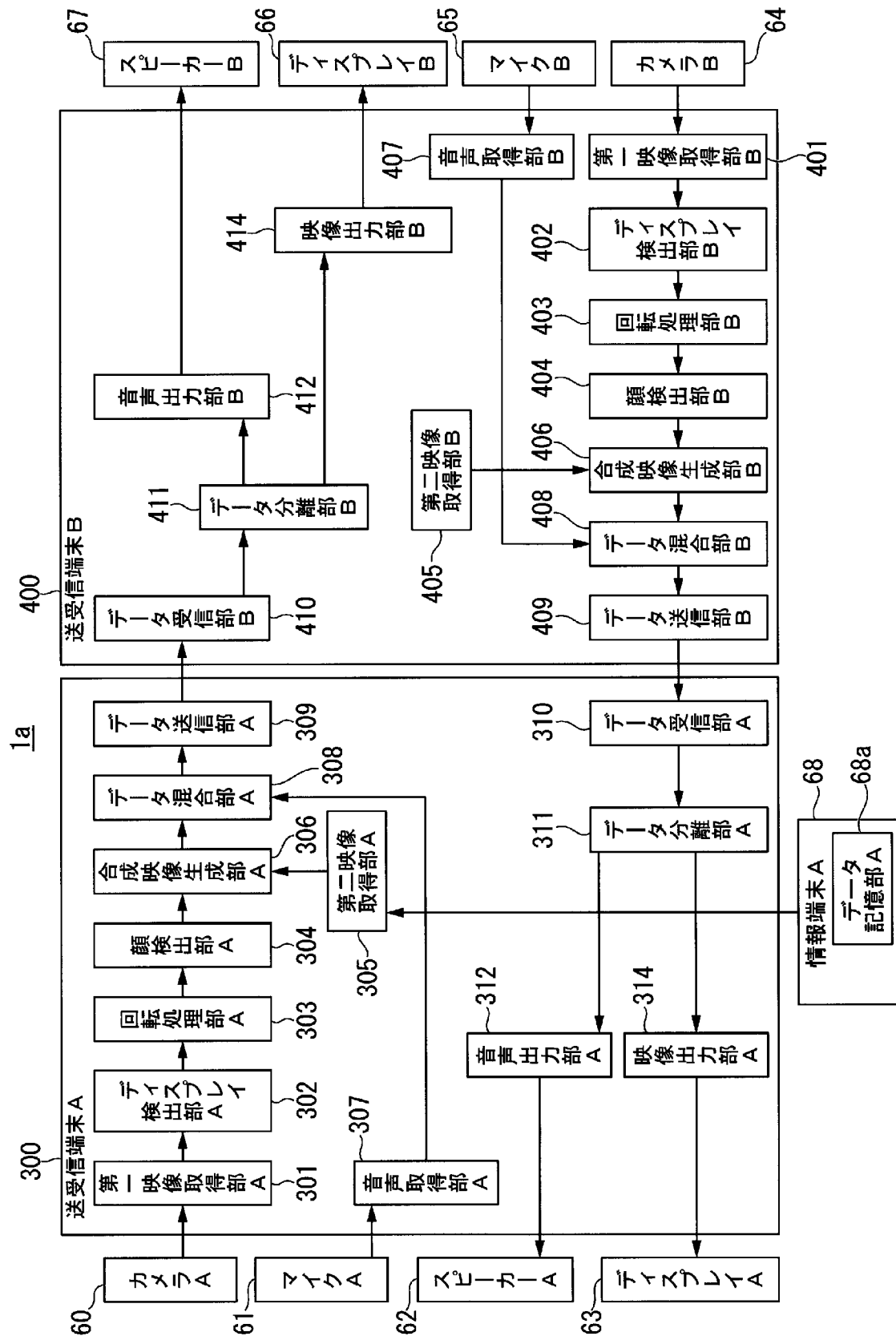
[図11]



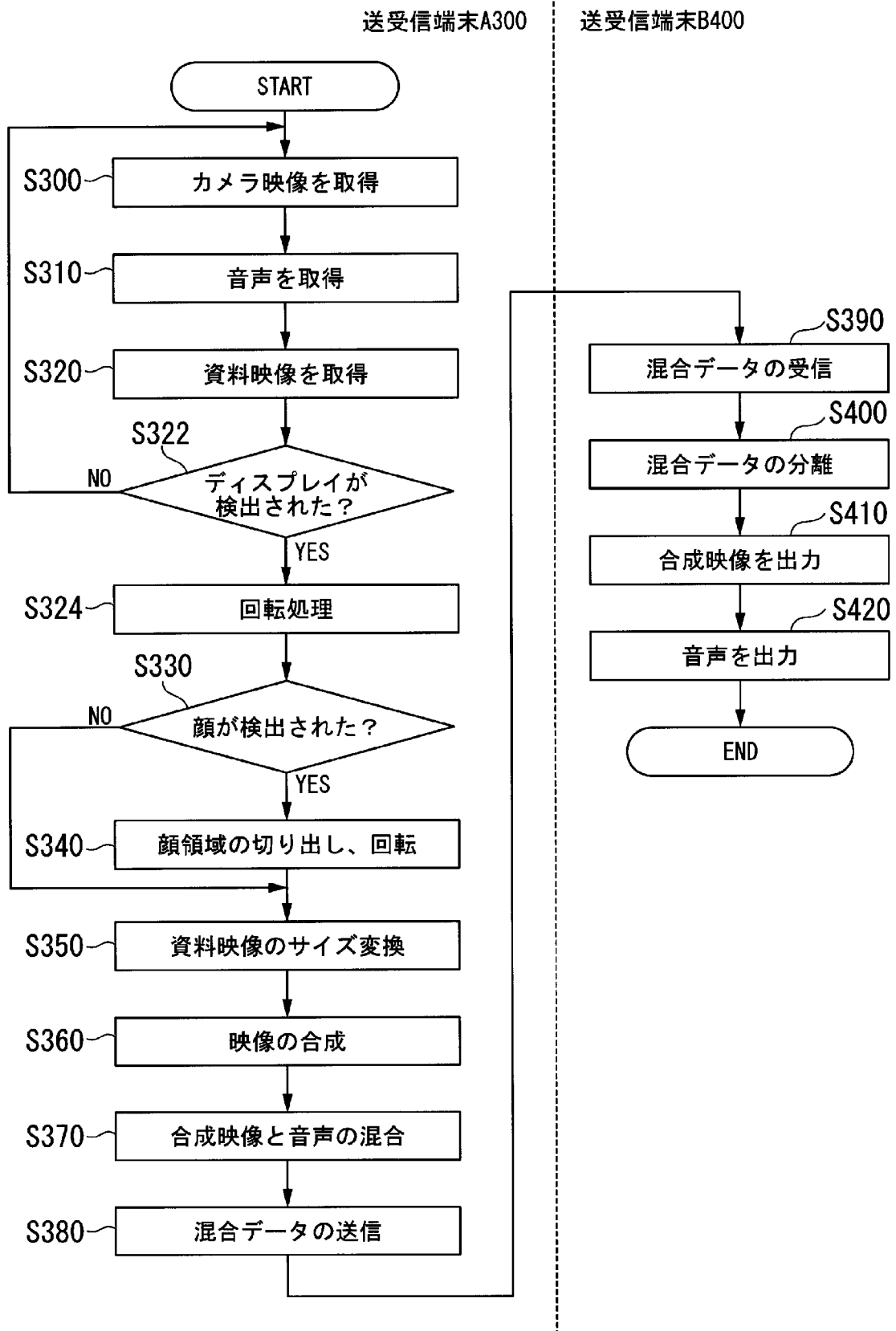
[図12]



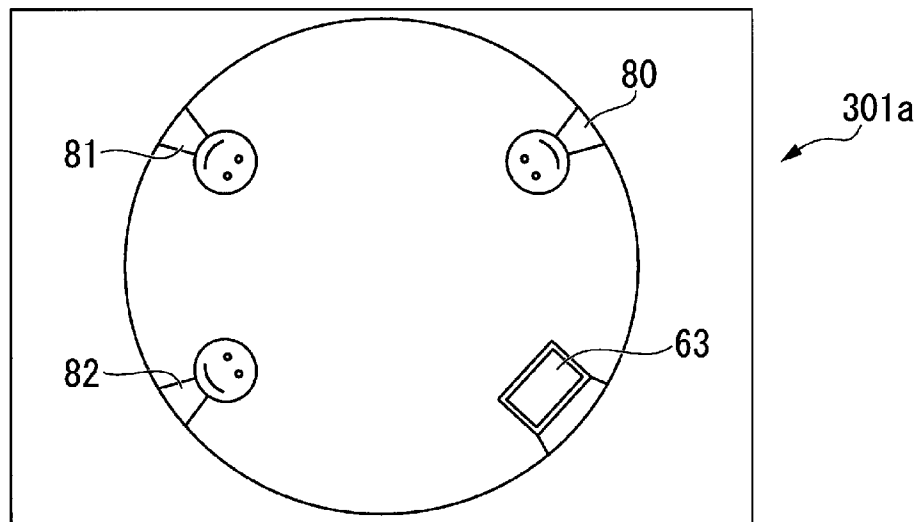
[図13]



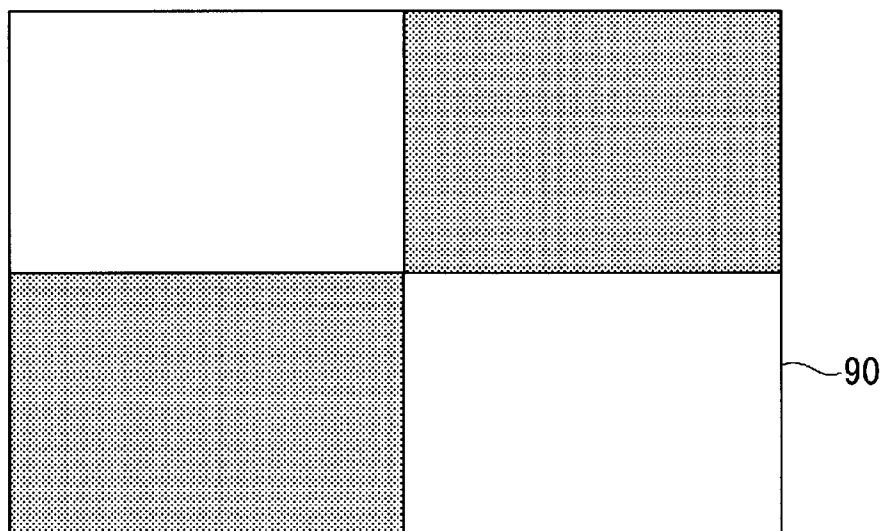
[図15]



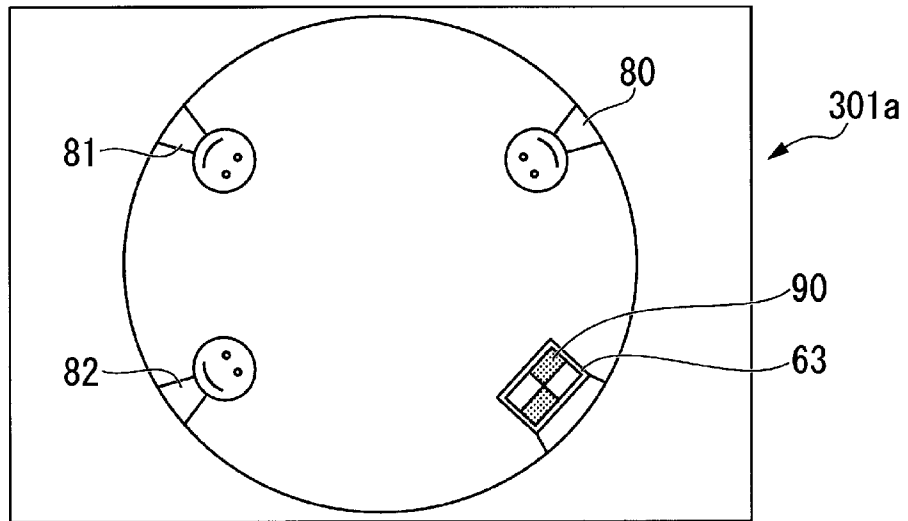
[図16]



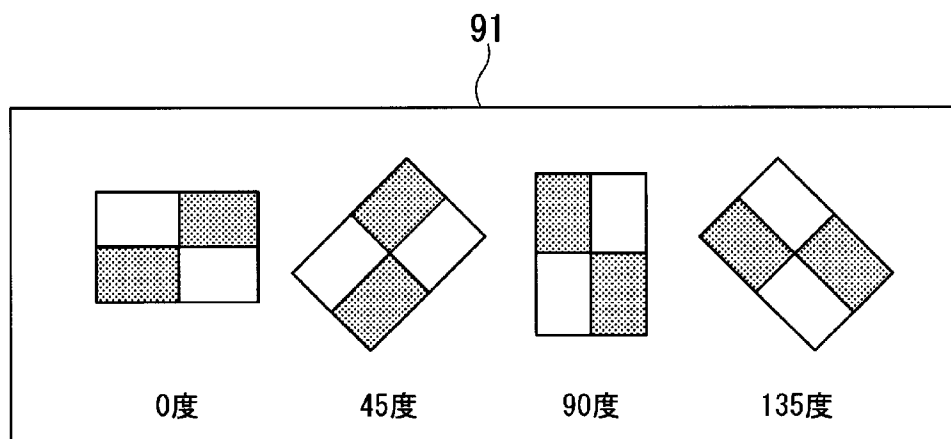
[図17]



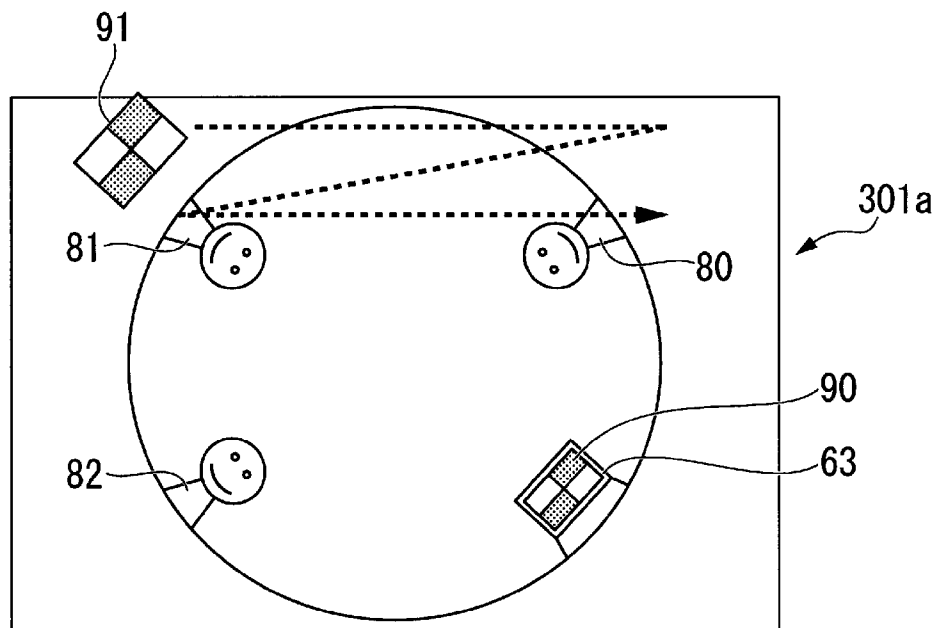
[図18]



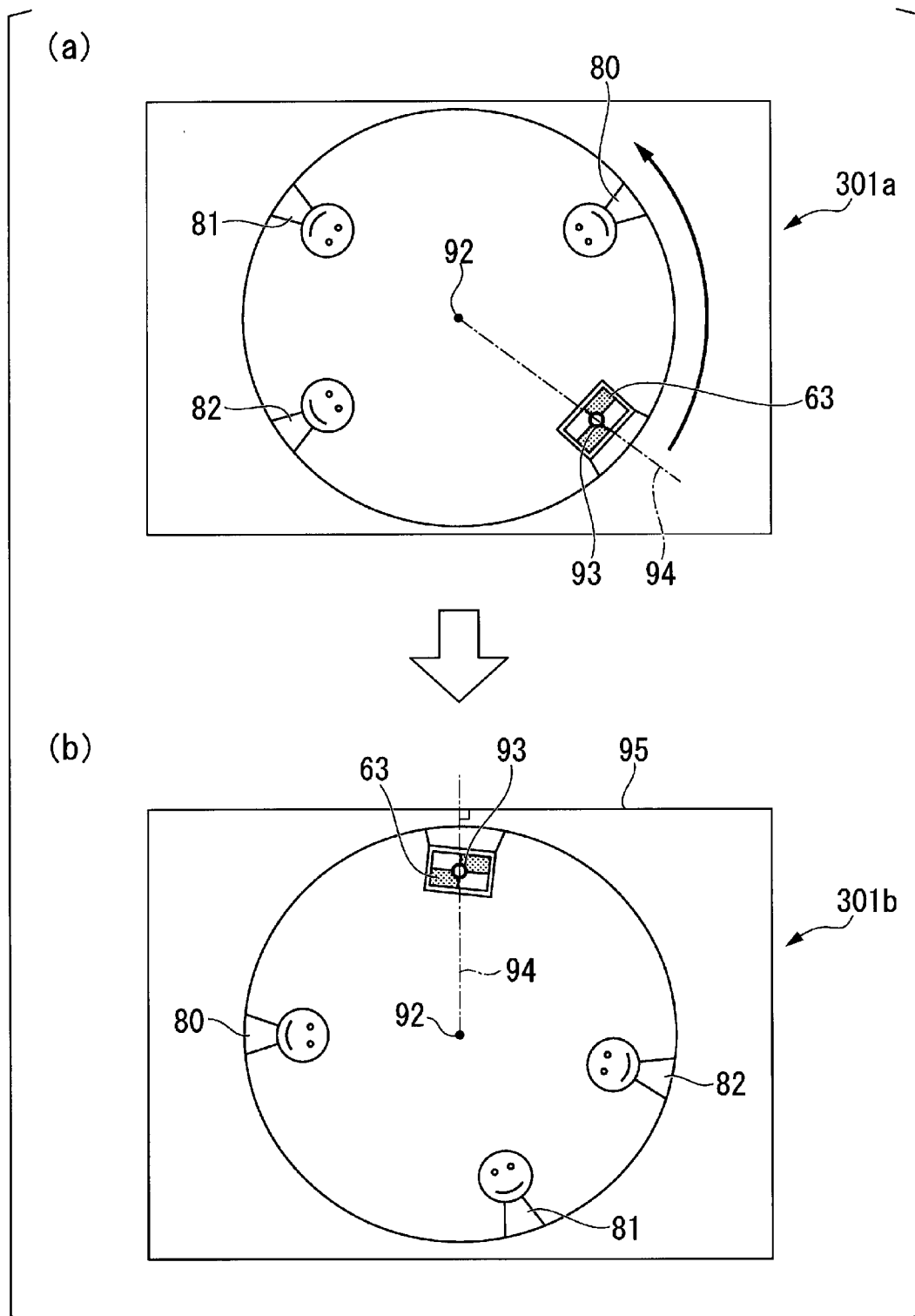
[図19]



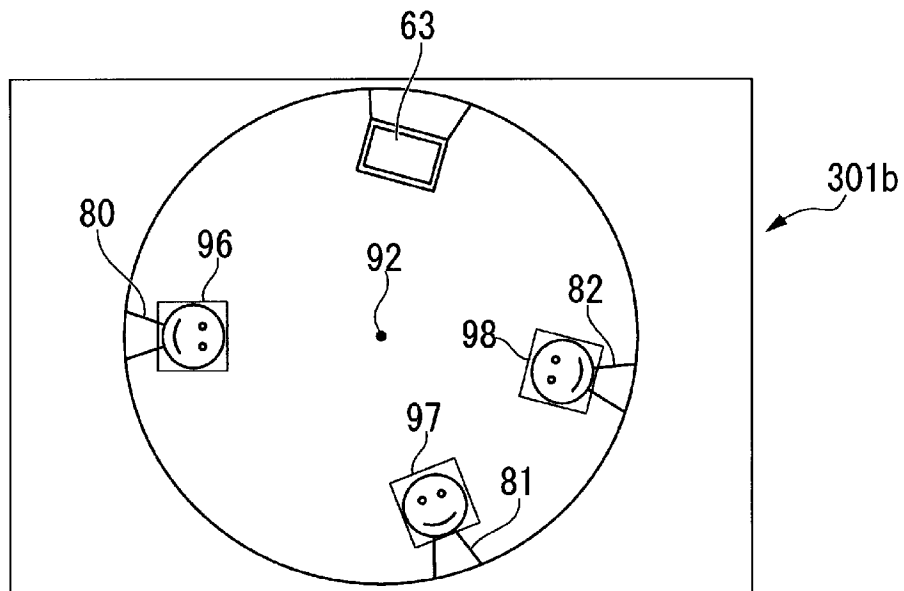
[図20]



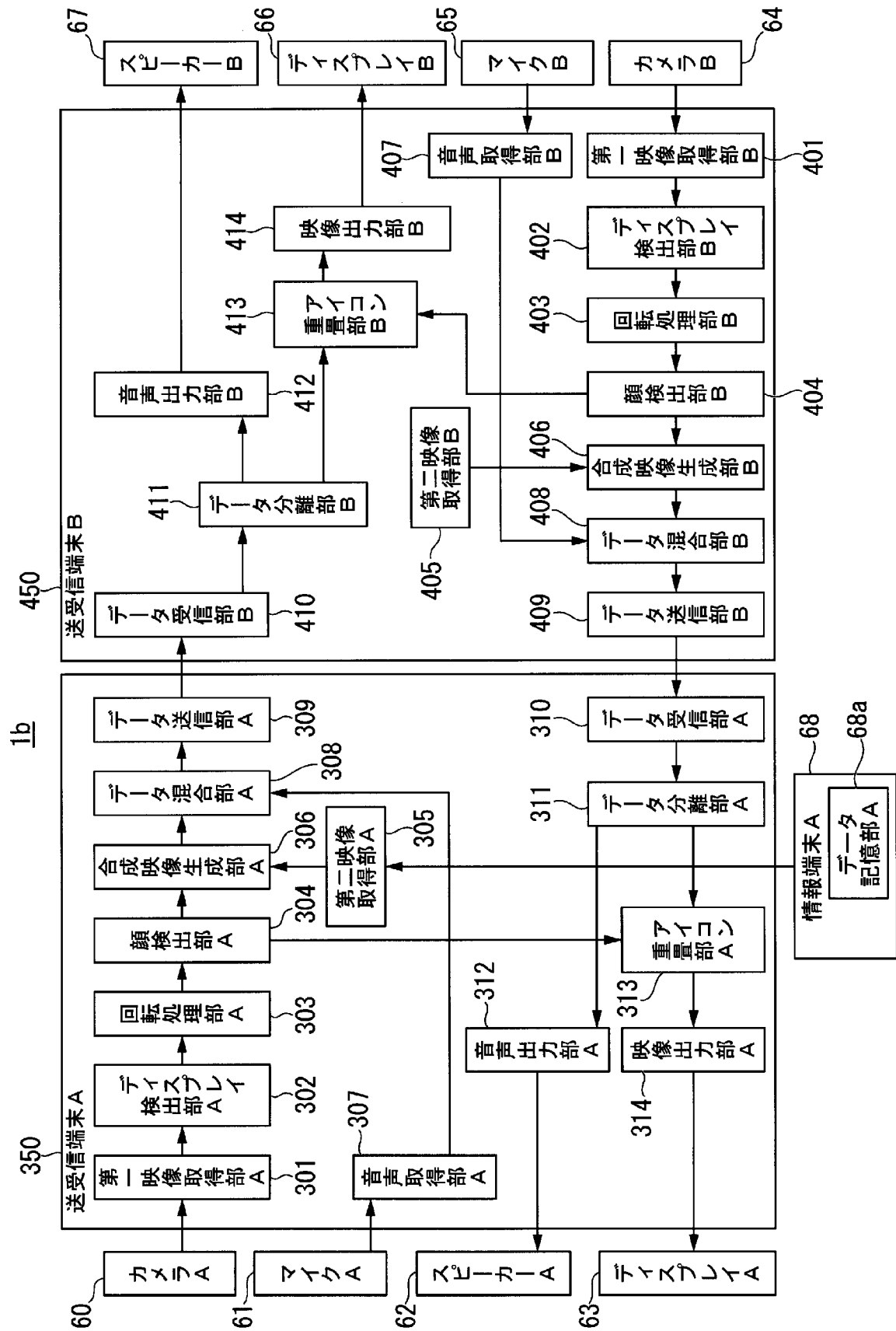
[図21]



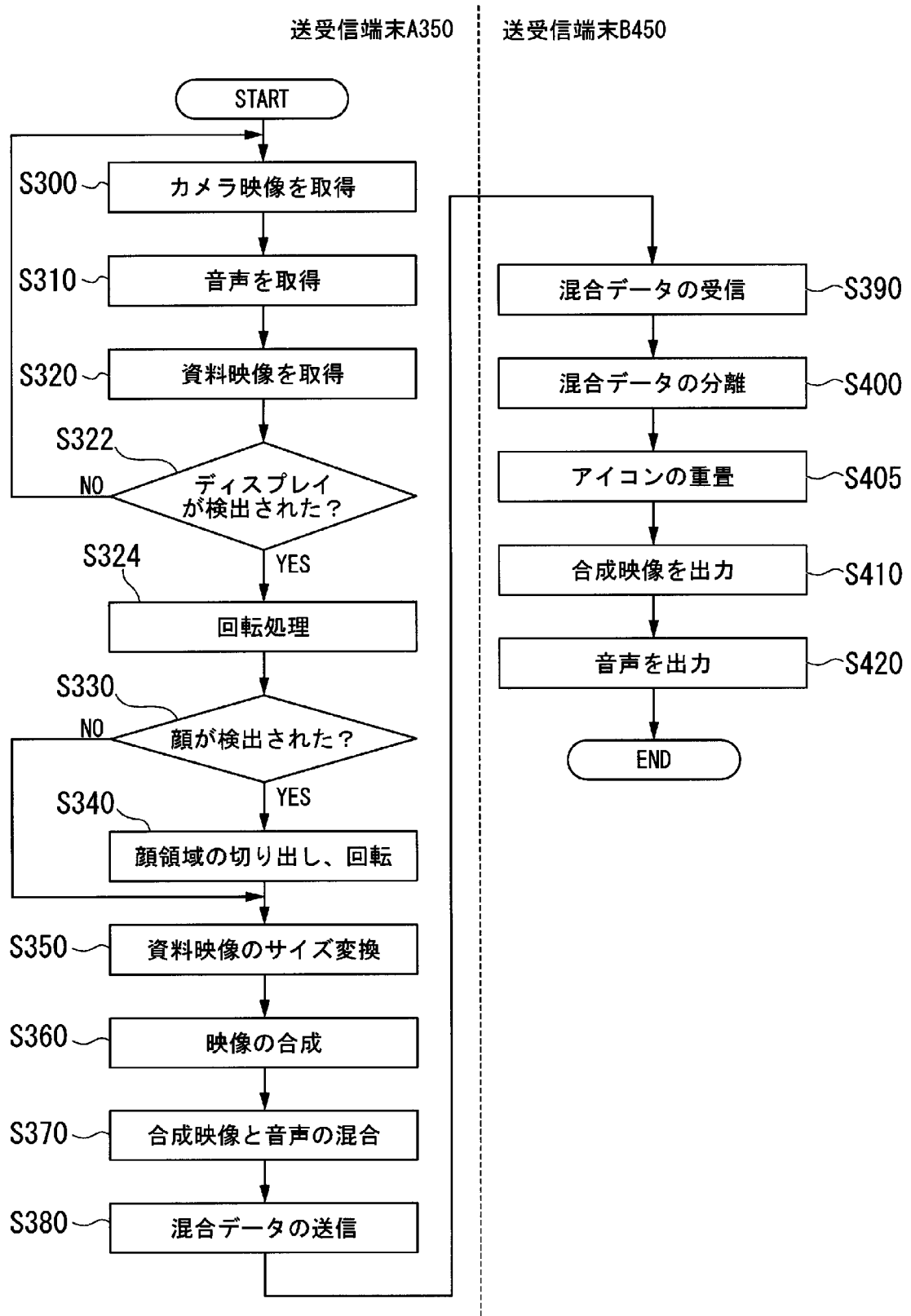
[図22]



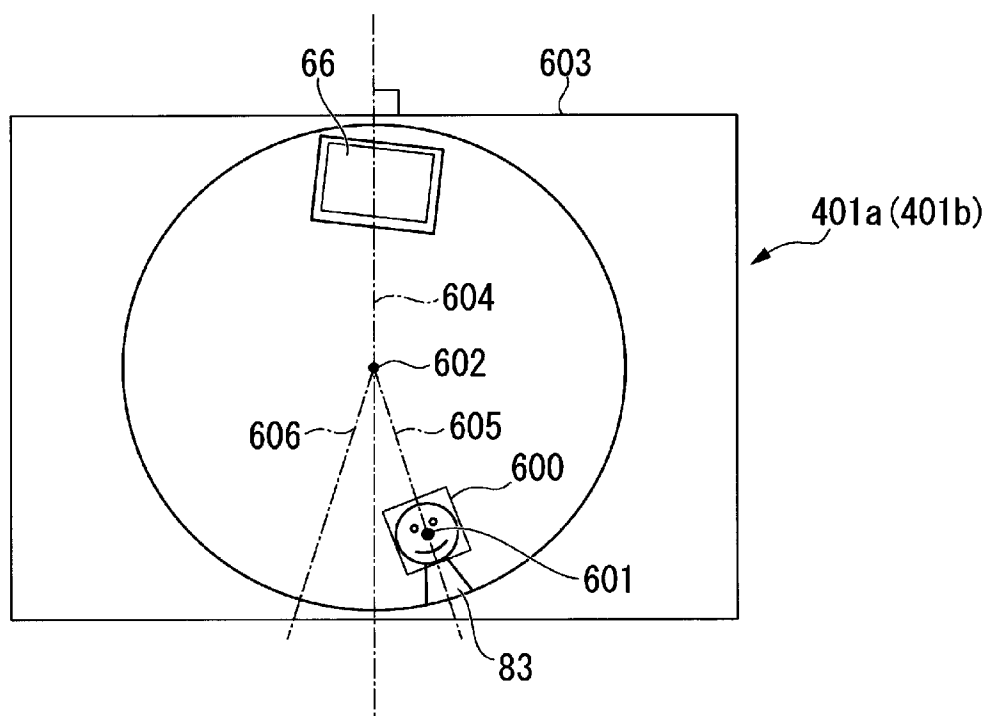
[図23]



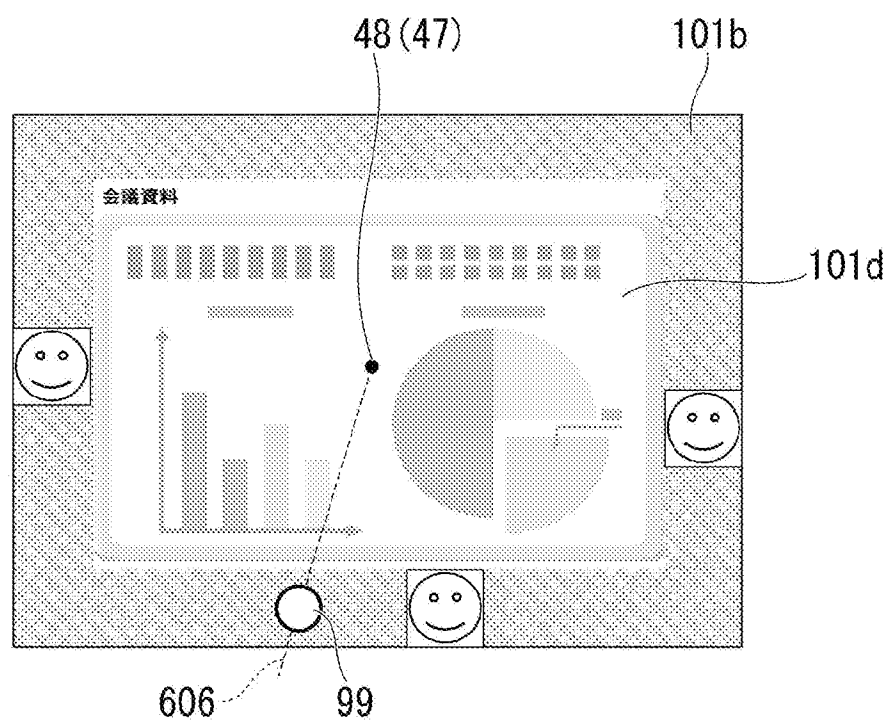
[図24]



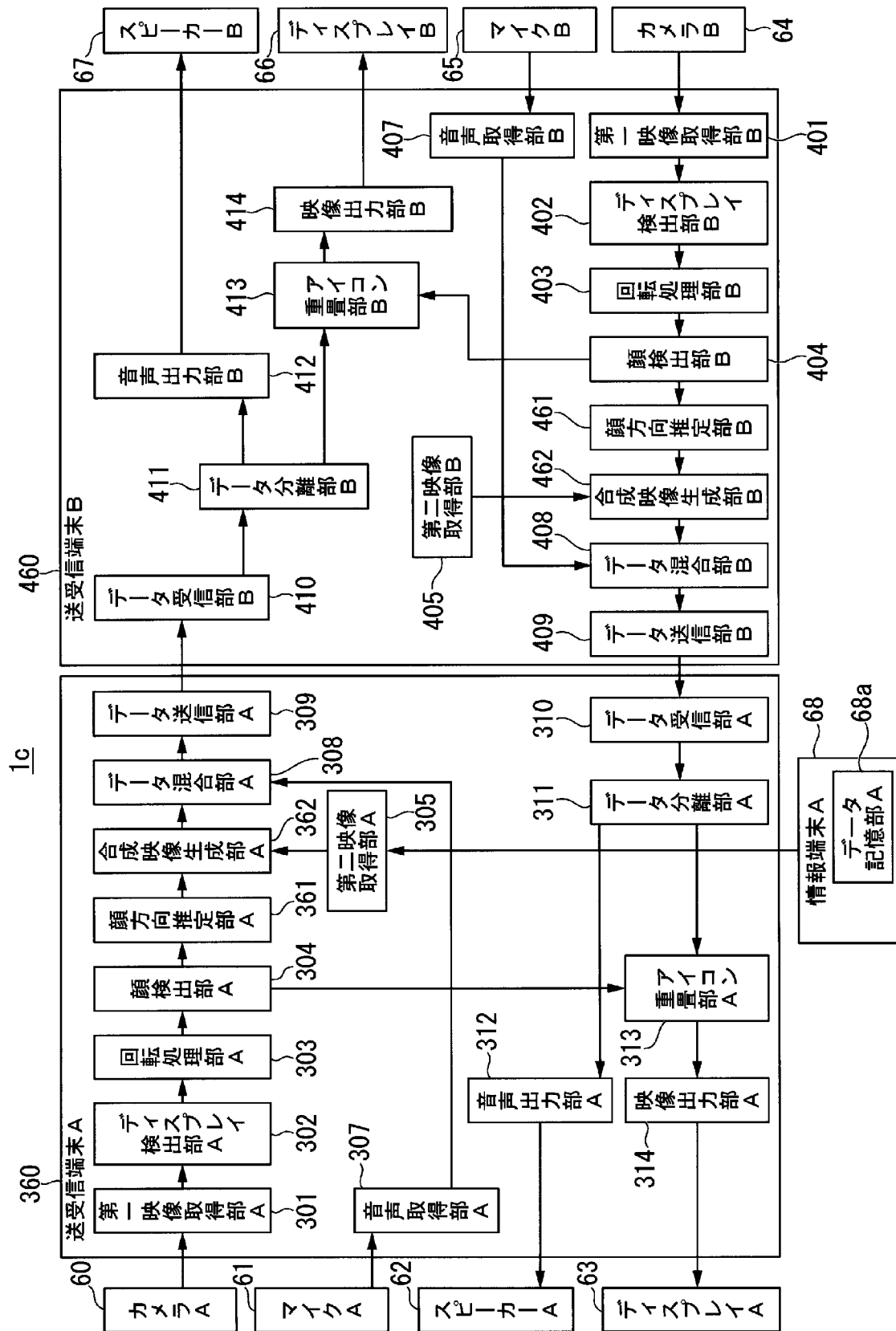
[図25]



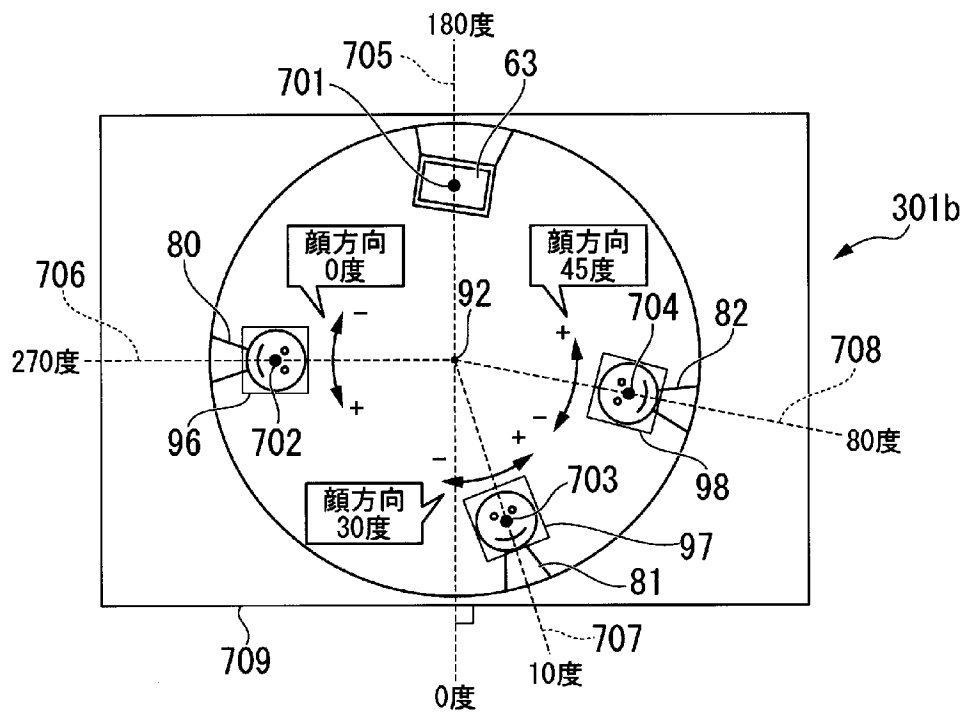
[図26]



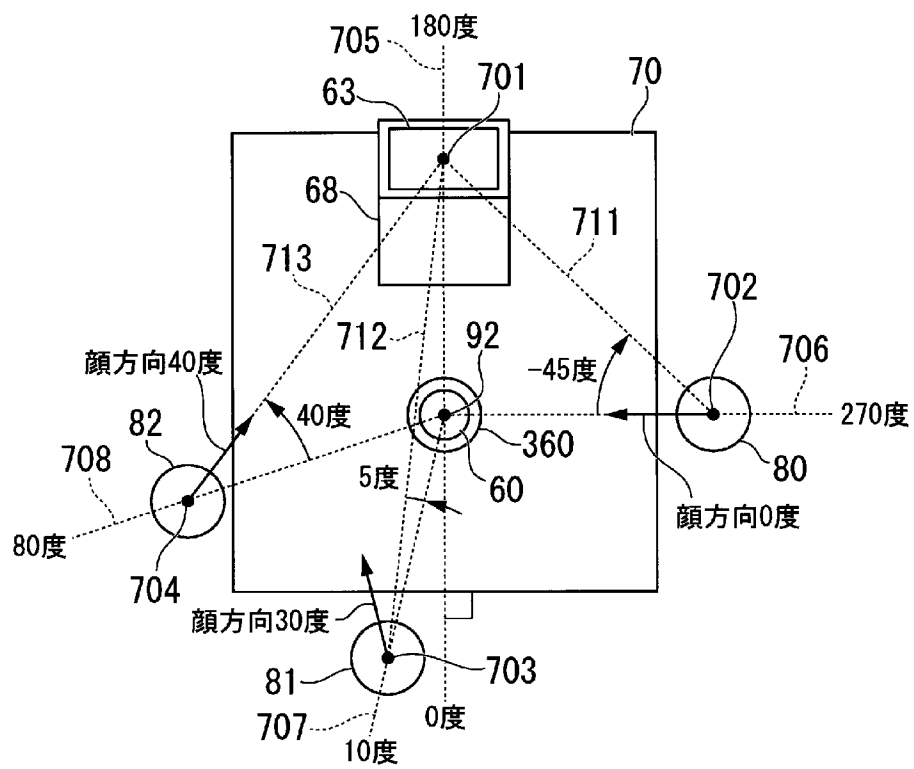
[図27]



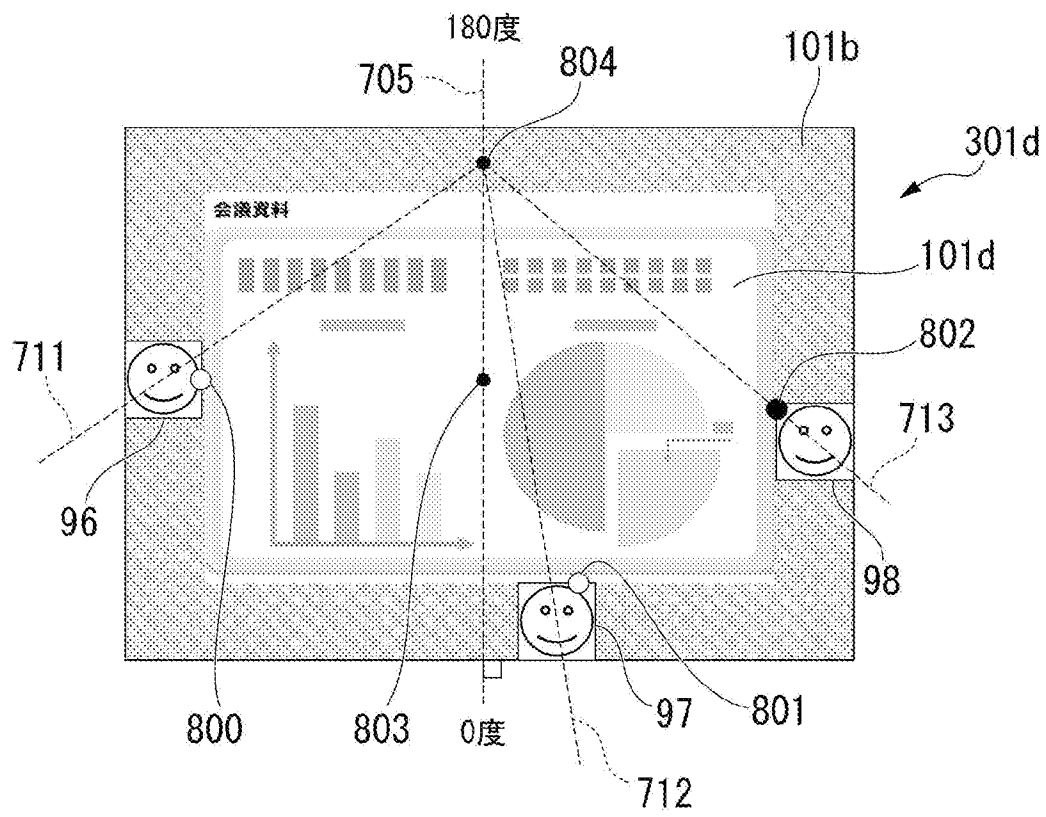
[図28]



[図29]



[図30]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2015/064321

A. CLASSIFICATION OF SUBJECT MATTER

H04N7/15(2006.01)i, H04N5/225(2006.01)i, H04N21/431(2011.01)i

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)

H04N7/15, H04N5/225, H04N21/431

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2015
Kokai Jitsuyo Shinan Koho	1971-2015	Toroku Jitsuyo Shinan Koho	1994-2015

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y A	JP 2011-066745 A (Brother Industries, Ltd.), 31 March 2011 (31.03.2011), paragraphs [0047] to [0067]; fig. 3 to 5 (Family: none)	1 3, 4 2, 5
Y	JP 2012-085009 A (Sony Corp.), 26 April 2012 (26.04.2012), paragraphs [0107] to [0130]; fig. 6 to 12 & US 2013/0182914 A1 & WO 2012/046425 A1 & EP 2625849 A & CN 103141085 A & KR 10-2013-0116860 A	3, 4
A	JP 2012-142727 A (Brother Industries, Ltd.), 26 July 2012 (26.07.2012), paragraphs [0059] to [0096]; fig. 9 to 17 (Family: none)	2

☒ 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
29 June 2015 (29.06.15)

Date of mailing of the international search report
07 July 2015 (07.07.15)

Name and mailing address of the ISA/
Japan Patent Office
3-4-3, Kasumigaseki, Chiyoda-ku,
Tokyo 100-8915, Japan

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2015/064321

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 8-256316 A (Toshiba Corp.), 01 October 1996 (01.10.1996), paragraphs [0029] to [0047]; fig. 5 (Family: none)	5
A	US 2014/0139620 A1 (THOMSON LICENSING), 22 May 2014 (22.05.2014), paragraphs [0038] to [0045]; fig. 5, 6 & WO 2013/019259 A1 & EP 2740266 A	5

A. 発明の属する分野の分類（国際特許分類（I P C））

Int.Cl. H04N7/15(2006.01)i, H04N5/225(2006.01)i, H04N21/431(2011.01)i

B. 調査を行った分野

調査を行った最小限資料（国際特許分類（I P C））

Int.Cl. H04N7/15, H04N5/225, H04N21/431

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1 9 2 2 - 1 9 9 6 年
日本国公開実用新案公報	1 9 7 1 - 2 0 1 5 年
日本国実用新案登録公報	1 9 9 6 - 2 0 1 5 年
日本国登録実用新案公報	1 9 9 4 - 2 0 1 5 年

国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
X Y A	JP 2011-066745 A（ブラザー工業株式会社）2011.03.31, 段落 [0047]-[0067], 図 3-5（ファミリーなし）	1 3, 4 2, 5
Y	JP 2012-085009 A（ソニー株式会社）2012.04.26, 段落 [0107]-[0130], 図 6-12 & US 2013/0182914 A1 & WO 2012/046425 A1 & EP 2625849 A & CN 103141085 A & KR 10-2013-0116860 A	3, 4

☒ C 欄の続きにも文献が列挙されている。☐ パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

「A」特に関連のある文献ではなく、一般的技術水準を示すもの
「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）
「O」口頭による開示、使用、展示等に言及する文献
「P」国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
「Y」特に関連のある文献であって、当該文献と他の 1 以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
「&」同一パテントファミリー文献

国際調査を完了した日

2 9 . 0 6 . 2 0 1 5

国際調査報告の発送日

0 7 . 0 7 . 2 0 1 5

国際調査機関の名称及びあて先

日本国特許庁（I S A / J P）
郵便番号 1 0 0 - 8 9 1 5
東京都千代田区霞が関三丁目 4 番 3 号

特許庁審査官（権限のある職員）

鋏 利孝

5 C

5 0 9 4

電話番号 0 3 - 3 5 8 1 - 1 1 0 1 内線 3 5 4 1

C (続き) . 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	JP 2012-142727 A (ブラザー工業株式会社) 2012. 07. 26, 段落 [0059]-[0096], 図 9-17 (ファミリーなし)	2
A	JP 8-256316 A (株式会社東芝) 1996. 10. 01, 段落[0029]-[0047], 図 5 (ファミリーなし)	5
A	US 2014/0139620 A1 (THOMSON LICENSING) 2014. 05. 22, 段落 [0038]-[0045], 図 5, 6 & WO 2013/019259 A1 & EP 2740266 A	5