(19)日本国特許庁(JP)　(12)公表特許公報(A)　(11)特許出願公表番号
特表2004-538724
(P2004-538724A)
(43)公表日　平成16年12月24日(2004.12.24)

(51)Int.Cl.[7]	F I			テーマコード（参考）
H04N 7/15	H04N	7/15	610	5C064
H04M 3/56	H04M	3/56	C	5K015

審査請求 有　予備審査請求 有　（全 47 頁）

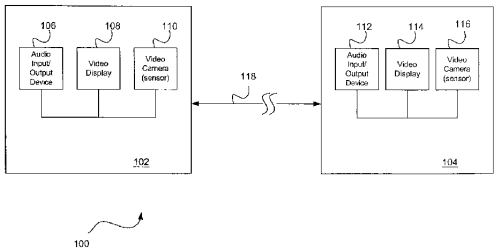
(21)出願番号	特願2003-520192(P2003-520192)
(86)(22)出願日	平成14年8月7日(2002.8.7)
(85)翻訳文提出日	平成16年2月6日(2004.2.6)
(86)国際出願番号	PCT/US2002/025477
(87)国際公開番号	WO2003/015407
(87)国際公開日	平成15年2月20日(2003.2.20)
(31)優先権主張番号	60/310,742
(32)優先日	平成13年8月7日(2001.8.7)
(33)優先権主張国	米国(US)

(71)出願人　500080720
ポリコム・インコーポレイテッド
アメリカ合衆国、９５０３５　カリフォルニア州　ミルピタス　バーバー・レーン　１５６５
(74)代理人　100070150
弁理士　伊東　忠彦
(74)代理人　100091214
弁理士　大貫　進介
(74)代理人　100107766
弁理士　伊東　忠重
(72)発明者　ケノイヤー，マイケル
アメリカ合衆国　７８７３１　テキサス州　オースティン　キャンドル・リッジ　４６０１

最終頁に続く

(54)【発明の名称】高解像度のテレビ会議のシステム及び方法

(57)【要約】
高解像度のテレビ会議のためのシステムと方法が示され、説明される。ビデオカメラ（１１０／１１６）又はセンサーと、複数のマイクロフォン（２１４）及びスピーカー（２１０ａ－２１０ｄ）と、音声（３１２）及び通信（３１８）処理エンジンを含む送信ステーション（１０２）と受信ステーション（１０４）が開示される。複数の映像ストリームが生成されることを可能にするシステムを通じて映像が処理され、転送され、テレビ会議システムに関する場所を示すことで音声が再生されることを可能にするシステムを通じて音声が処理され、転送される。



【特許請求の範囲】
【請求項１】
画像を取得するビデオセンサーと、
音源に応じて音声信号を生成する複数のマイクロフォンと、
前記ビデオセンサーと前記複数のマイクロフォンにつなげられ、少なくとも１つの映像ストリームと音源の位置を示す位置信号を生成する処理エンジンと
を有するテレビ会議装置。
【請求項２】
請求項１に記載のテレビ会議装置であって、
前記ビデオセンサーにつなげられ、ビデオセンサーと映像表示出力との間のフェーズを同期するフェーズ同期エンジンを更に有するテレビ会議装置。
【請求項３】
請求項１に記載のテレビ会議装置であって、
前記処理エンジンにつなげられ、音声信号と位置信号と少なくとも１つの映像ストリームをリモートのテレビ会議装置に送信する通信インタフェースを更に有するテレビ会議装置。
【請求項４】
請求項１に記載のテレビ会議装置であって、
前記複数のマイクロフォンから受信された電気信号又は電流信号の大きさの違いに基づいて前記位置信号が生成されるテレビ会議装置。
【請求項５】
請求項１に記載のテレビ会議装置であって、
前記処理エンジンが映像処理エンジンを更に有し、
前記映像処理エンジンが複数の画像セクションを規定し、前記複数の画像セクションに対応するそれぞれの複数の映像ストリームを生成するテレビ会議装置。
【請求項６】
請求項１に記載のテレビ会議装置であって、
音源の位置が変更すると、処理エンジンが位置の変更を反映する新たな位置信号を生成するテレビ会議装置。
【請求項７】
請求項２に記載のテレビ会議装置であって、
前記位置信号に応じてリモートのテレビ会議装置が１つ以上のスピーカーを選択的に駆動し、前記少なくとも１つの映像ストリームの画像に対応する音声信号を再生するテレビ会議装置。
【請求項８】
請求項１に記載のテレビ会議装置であって、
前記複数のマイクロフォンがｎ方向の構成で配置されるテレビ会議装置。
【請求項９】
請求項１に記載のテレビ会議装置であって、
前記複数のマイクロフォンが垂直の配列で配置されるテレビ会議装置。
【請求項１０】
請求項５に記載のテレビ会議装置であって、
前記処理エンジンが、前記複数の画像セクションのうちの第１の画像セクションを、第１の解像度を有する第１の映像ストリームに調整し、前記複数の画像セクションのうちの第２の画像セクションを、第２の解像度を有する第２の映像ストリームに調整するテレビ会議装置。
【請求項１１】
請求項１に記載のテレビ会議装置であって、
更に大きい角度のビュー角度を提供するために、前記ビデオセンサーにつなげられたパン（ｐａｎ）モーターを更に有するテレビ会議装置。

【請求項１２】
テレビ会議システムにおいて会議データを送信する方法であって、
ビデオセンサーで画像を取得し、前記画像から少なくとも１つの映像ストリームを生成し、
複数のマイクロフォンで音声データを取得し、音声信号を生成し、
前記音声データの大きさの違いに基づいて音源の位置を示す位置信号を生成し、
通信チャネルを介して位置信号と音声信号と少なくとも１つの映像ストリームを送信することを有する方法。
【請求項１３】
請求項１２に記載の方法であって、
前記位置信号に応じてリモートのテレビ会議装置の１つ以上のスピーカーを選択的に駆動し、前記少なくとも１つの映像ストリームの画像に対応する音声信号を再生することを更に有する方法。
【請求項１４】
請求項１２に記載の方法であって、
ビデオセンサーと映像表示出力との間のフェーズを同期することを更に有する方法。
【請求項１５】
請求項１２に記載の方法であって、
複数の画像セクションを規定し、前記複数の画像セクションに対応するそれぞれの複数の映像ストリームを生成することを更に有する方法。
【請求項１６】
請求項１２に記載の方法であって、
位置の変更を反映する新たな位置信号を生成することを更に有する方法。
【請求項１７】
請求項１４に記載の方法であって、
前記複数の画像セクションのうちの第１の画像セクションを、第１の解像度を有する第１の映像ストリームに調整し、前記複数の画像セクションのうちの第２の画像セクションを、第２の解像度を有する第２の映像ストリームに調整することを更に有する方法。
【請求項１８】
画像を取得し、前記画像から少なくとも１つの映像ストリームを生成する手段と、
音声を取得し、音声信号を生成する手段と、
前記音声データの大きさの違いに基づいて音源の位置を示す位置信号を生成し、前記位置信号に応じて前記位置信号がリモートのテレビ会議システムの１つ以上のスピーカーを選択的に駆動し、前記少なくとも１つの映像ストリームの画像に対応する音声信号を再生する手段と、
通信チャネルを介して位置信号と音声信号と少なくとも１つの映像ストリームを送信する手段と
を有するテレビ会議装置。
【請求項１９】
プログラムに具体化された電子的読み取り可能媒体であって、
前記プログラムが会議データを送信する方法のステップを実行する機械によって実行可能であり、
前記方法のステップが、
ビデオセンサーで画像を取得し、前記画像から少なくとも１つの映像ストリームを生成し、
複数のマイクロフォンで音声データを取得し、音声信号を生成し、
前記音声データの大きさの違いに基づいて音源の位置を示す位置信号を生成し、
通信チャネルを介して位置信号と音声信号と少なくとも１つの映像ストリームを送信することを有する電子的読み取り可能媒体。
【請求項２０】

請求項１９に記載の電子的読み取り可能媒体であって、
前記方法のステップが、前記位置信号に応じてリモートのテレビ会議システムの１つ以上のスピーカーを選択的に駆動し、前記少なくとも１つの映像ストリームの画像に対応する音声信号を再生することを更に有する電子的読み取り可能媒体。
【請求項２１】
請求項１９に記載の電子的読み取り可能媒体であって、
前記方法のステップが、複数の画像セクションを規定し、前記複数の画像セクションに対応するそれぞれの複数の映像ストリームを生成することを更に有する電子的読み取り可能媒体。
【請求項２２】
請求項１９に記載の電子的読み取り可能媒体であって、
前記方法のステップが、前記複数の画像セクションのうちの第１の画像セクションを、第１の解像度を有する第１の映像ストリームに調整し、前記複数の画像セクションのうちの第２の画像セクションを、第２の解像度を有する第２の映像ストリームに調整することを更に有する電子的読み取り可能媒体。
【発明の詳細な説明】
【０００１】
［関連出願との相互参照］
本出願は、２００１年８月７日に出願された“高解像度のテレビ会議バー（Ｈｉｇｈ　Ｒｅｓｏｌｕｔｉｏｎ　Ｖｉｄｅｏ　Ｃｏｎｆｅｒｅｎｃｉｎｇ　Ｂａｒ）”という題名の仮特許出願第６０／３１０，７４２号の優先権の利益を主張する。
［発明の背景］
１．発明の分野
本発明は概して会議システムに関するものであり、特に高解像度のテレビ会議システムに関するものである。
２．背景技術の説明
従来、テレビ会議システムは、遠隔の会議場所への送信のための会議参加者の画像を取得するためにビデオカメラを利用する。従来の（静止又は可動）ビデオカメラは、同時に特定の場所で会議場所の１つの画像又は１つのビューのみを取得することができる。同時に異なる場所で会議場所の異なる画像又はビューを取得するために、従来のビデオカメラは、カメラの回転方向を調整する装置を有利に備えている場合がある。カメラをほぼ２つの直交軸に回転するように設計された位置合わせ装置は、一般的に２つのアクチュエータを利用する。第１のアクチュエータは、ほぼ垂直軸にカメラを回転し、第２のアクチュエータは、カメラの垂直軸と直角のほぼ水平軸にカメラを回転する。従って、ほぼ水平軸へのカメラの回転は“パニング（ｐａｎｎｉｎｇ）”と称され、ほぼ垂直軸へのカメラの回転は、“チルティング（ｔｉｌｔｉｎｇ）”と称される。したがって、カメラをほぼ水平及び垂直軸に回転する装置は、一般的に“パン／チルト（ｐａｎ／ｔｉｌｔ）位置合わせ装置”と称される。更に、話し中の会議参加者の画像のような特に関心の高い画像又はビューを取得するために、従来のビデオカメラは、ズーム機能を実行する一式のズームレンズを必要とし、その結果、“パン／チルト／ズーム（ｐａｎ／ｔｉｌｔ／ｚｏｏｍ）”（“ＰＺＴ”）カメラになる。
【０００２】
不都合なことに、従来のＰＺＴカメラは多数の欠点を有する。第１に、位置合わせ装置の機械的構成要素の動きは、実質的な量のノイズを生成し得る。前記動きとノイズは会議参加者に不快であり、気を散らす。更に重要なことに、ノイズは話し中の参加者の方向にカメラを自動的に向けるために利用される音響位置測定技術に干渉し得る。第２に、位置合わせ装置の機械的構成要素は、磨耗又は手荒な取り扱いのため調整不良や破損の影響を受けやすく、それによって、位置合わせ装置が一部又は全部において動作不能になる。更なる不利点は、位置合わせ装置の製造における複雑さであり、そのため、高製造コストとその結果の高消費者価格を生ずる。

【０００３】
技術開発に伴い、テレビ会議システムの表示スクリーンのサイズは大きくなってきている。その結果、表示スクリーンの参加者の話す人の位置が広範囲の領域で変わり得る。しかし、不都合なことに、位置が広範囲の領域で変わると、従来のテレビ会議システムは新たな参加者の話す人の位置に調整することができない。
【０００４】
従って、複雑な機械的構造を有することなく、会議場所の複数のビューを取得するテレビ会議システム及び方法についての必要性が存在する。話す人の位置に関連する音響を調整するテレビ会議システム及び方法についての更なる必要性も存在する。
［発明の概要］
本発明は、音源に応じて音声信号を生成する複数のマイクロフォンと、音源の位置を示す位置信号を生成し、音声信号を処理する音声処理エンジンと、通信チャネルに音声信号及び位置信号を送信する通信インタフェースとを含み、第１の場所に設置された送信ステーションを有するテレビ会議システムを提供する。テレビ会議システムの複数のマイクロフォンは、垂直の配列及びｎ方向の構成で配置され得る。テレビ会議システムはまた、通信チャネルから音声信号及び位置信号を受信する通信インタフェースと、音声信号を再生する複数のスピーカーと、位置信号に応じてスピーカーの１つを選択的に駆動し、選択されたスピーカーで音声信号を再生する音声処理エンジンとを含み、第２の場所に設置された受信ステーションを有し得る。
【０００５】
テレビ会議システムで生成された位置信号は、複数のマイクロフォンから受信された電気信号又は電流信号の大きさの違いに基づく。音源の位置が変更すると、音声処理エンジンは位置の変更を反映する新たな位置信号を生成する。
【０００６】
送信ステーションの通信インタフェースは、音声信号と位置信号をコード化し、圧縮する通信処理エンジンと、通信チャネルを通じて音声信号及び位置信号を送信する送受信装置とを含む。逆に、受信ステーションの通信インタフェースは、通信チャネルを通じて音声信号及び位置信号を受信する送受信装置と、音声信号と位置信号をデコードし、解凍する通信処理エンジンとを含む。
【０００７】
他の実施例において、テレビ会議システムは、画像を生成する高解像度のビデオセンサーと、高解像度の画像を保存する映像メモリと、ビデオセンサーから映像メモリに画像データをロードするデータロードエンジンとを含み、第１の場所に設置された送信ステーションを有する。更に、フィールド・プログラマブル・ゲート・アレイ（Ｆｉｅｌｄ　Ｐｒｏｇｒａｍｍａｂｌｅ　Ｇａｔｅ　Ａｒｒａｙ）／特定用途向け集積回路（ＦＰＧＡ／ＡＳＩＣ）が映像メモリとデータロードエンジンにつなげられる。ＦＰＧＡ／ＡＳＩＣは映像メモリに保存された高解像度画像内で第１の画像セクションと第２の画像セクションを規定する。更に、ＦＰＧＡ／ＡＳＩＣは第１の解像度で第１の映像ストリームに合わせて第１の画像セクションを調整し、第２の解像度で第２の映像ストリームに合わせて第２の画像セクションを調整し得る。ＦＰＧＡ／ＡＳＩＣにつなげられた通信インタフェースは第１の映像ストリームと第２の映像ストリームを通信チャネルに送信する。テレビ会議システムはまた、通信チャネルから第１の映像ストリームと第２の映像ストリームを受信する通信インタフェースを含み、第２の場所に設置された受信ステーションを有し得る。受信ステーションは、第１の映像ストリームと第２の映像ストリームを処理し、第１の解像度で第１の画像として第１の映像ストリームを表示し、第２の解像度で第２の画像として第２の映像ストリームを表示する映像処理エンジンを更に含み、通信インタフェースにつなげられる。
【０００８】
この実施例における送信ステーションの通信インタフェースは、第１と第２の映像ストリームをコード化し、圧縮する通信処理エンジンと、通信チャネルを通じて第１と第２の映

像ストリームを送信する送受信装置とを有する。逆に、本実施例の受信ステーションの映像処理エンジンは、第１の映像ストリームと第２の映像ストリームを保存する映像メモリと、受信ステーションの通信インタフェースから第１の映像ストリームと第２の映像ストリームをロードするデータロードエンジンと、映像メモリに保存された高解像度の画像に基づいて第１と第２の画像データストリームを表示するＦＰＧＡ／ＡＳＩＣとを有する。

【０００９】

更に他の実施例において、テレビ会議システムは、通信チャネルから映像信号を受信する通信インタフェースと、映像信号に応じて映像表示出力を生成する映像処理エンジンと、映像表示出力を表示する映像ディスプレイとを有し、第１の場所に設置された受信ステーションを有する。テレビ会議システムは、映像信号を生成するビデオカメラと、映像信号を処理する映像処理エンジンと、送信ステーションのビデオカメラと受信ステーションの映像表示出力の間のフェーズを同期するフェーズ同期エンジンと、通信チャネルに映像信号を送信する通信インタフェースとを有し、第２の場所に設置された送信ステーションを更に有し得る。

［発明の説明］

図１は、本発明による例示的なテレビ会議システム１００を示したものである。テレビ会議システム１００は、第１の会議ステーションと第２の会議ステーションを有する。第１の会議ステーション１０２は、音声入出力装置１０６と、１０８と、ビデオカメラ（又はビデオセンサー）１１０を有する。同様に、第２の会議ステーション１０４は、音声入出力装置１１２と、映像ディスプレイ１１４と、ビデオカメラ（又はビデオセンサー）１１６を有する。第１の会議ステーション１０２は、通信チャネル１１８を通じて第２の会議ステーション１０４と通信する。通信チャネル１１８は、インターネット、ＬＡＮ、ＷＡＮ、又は何らかの他の形式のネットワーク通信手段である可能性がある。図１は２つの会議ステーション１０２と１０４のみを示しているが、追加の会議ステーションがテレビ会議システム１００につなげられ得ることがその技術に熟練した人は認識するであろう。

【００１０】

図２は、本発明の一実施例により、図１の会議ステーション１０２と１０４と同様の例示的な会議ステーション２００を示したものである。会議ステーション２００は、ディスプレイ２０２と、高解像度会議バー２０４と、映像処理ユニット２０６とを有する。好ましくは、ディスプレイ２０２は、１６：９の視聴可能領域を備えた比較的大きいサイズのフラットスクリーン２０８を有する高解像度（“ＨＤ”）モニタである。その他、他の視聴領域の比率や、他の形式のディスプレイ２０２が検討され、用いられ得る。

【００１１】

好ましくは、高解像度のテレビ会議バー２０４は、複数のスピーカー２１０ａ－２１０ｄと、ビデオセンサー（例えばＣＭＯＳビデオセンサーのような高解像度デジタルビデオ画像センサー）２１２と、複数のマイクロフォン２１４とを含む。スピーカー２１０ａ－２１０ｄは、好ましくは２５０Ｈｚを越える周波数で動作する。しかし、スピーカー２１０ａ－２１０ｄは、本発明の多様な実施例と互換性がある何らかの他の周波数で動作し得る。一実施例において、会議バー２０４はおよそ幅が３６インチ、高さが２インチ、奥行が４インチであるが、会議バー２０４は何らかの他の寸法であってもよい。一般的に、会議バー２０４はディスプレイ２０２の先端より少し小さい幅の前面部２１８でディスプレイ２０２の上に設置されるように設計される。会議バー２０４の位置は、スピーカー２１０ａ－２１０ｄと、ビデオセンサー２１２と、複数のマイクロフォン２１４とをスクリーン２０８の近くにもたらし、ディスプレイ２０２の先端で位置の基準を提供する。他の会議バー２０４の位置も、本発明の範囲と目的と調和するように利用され得る。更に、図２には４つのスピーカーのみが示されているが、本発明においてそれよい多い又はそれより少ないスピーカーが利用され得る。

【００１２】

ビデオセンサー２１２は、７２０ｉ（すなわち、毎秒６０フィールドでインタレースされた１２８０×７２０）以上の好ましい解像度で、リアルタイムで複数の画像を出力するこ

とが可能であるが、本発明により他の解像度も考えられる。全会議場所を取得する約６５度のビューに基づき、ビデオセンサー２１２の解像度は十分である。更に広いビュー（９０度のビュー等）のために、限られた水平のパン（ｐａｎ）モーターが提供され得る。前記限られた水平のパン（ｐａｎ）モーターを提供することにより、高価で複雑な全ての機械的なパン／チルト／ズーム（ｐａｎ／ｔｉｌｔ／ｚｏｏｍ）カメラとレンズシステムを避けることができる。更に、純粋なデジタルズームが固定レンズに備えられ、最小のＣＩＦ（３５２×２８８）の解像度の画像を維持する一方で、８倍以上の有効なズームまでに対応し得る。

【００１３】

複数のマイクロフォン２１４は、会議バー２０４のビデオセンサー２１２の両側に設置され、図２に示すように、より良い順方向の特性を提供するｎ方向の構成で配置され得る。垂直のマイクロフォンが、ディスプレイ２０２の側面にオプションで配置され、垂直の位置基準を提供し得る。

【００１４】

会議バー２０４は、高速デジタルリンク２０５を介して処理ユニット２０６につなげられる。処理ユニット２０６は、好ましくは２５０Ｈｚ未満から５０－１００Ｈｚの周波数で動作するサブウーファー（ｓｕｂ－ｗｏｏｆｅｒ）装置を含み得る。処理ユニット２０６は、図３と共に更に詳細に説明される。処理ユニット２０６は会議バー２０４から分離して示されているが、その代わりに処理ユニット２０６は会議バー２０４に含まれ得る。

【００１５】

会議の参加者はビデオセンサー２１２を見ていると、又はその動きを見ていると不快に思う場合があるため、ビデオセンサー２１２の前面及び／又は会議バー２０４の他の部分にスモークガラス又は他のカバーが設置される場合があり、それによって会議の参加者がビデオセンサー２１２及び／又はスピーカー２１０ａ－２１０ｄ並びに複数のマイクロフォン２１４を見ることができなくなる。

【００１６】

図３は、本発明の一実施例により、図２の処理ユニット２０６を更に詳細に示した例示的なブロック図である。処理ユニット２０６は、好ましくは処理エンジン３０２と通信インタフェース２０４とサブウーファー（ｓｕｂ－ｗｏｏｆｅｒ）装置とを有する。処理エンジン３０２は、フェーズ同期エンジン３０８と映像処理エンジン３１０と音声処理エンジン３１２とを更に有する。フェーズ同期エンジン３０８は、送信遅延によって生じた悪影響を減少する又は最小限にすることが可能である。特に、ローカルの（又は第１の）会議ステーション１０２（図１）のビデオカメラ１１０（図１）は、リモートの（又は第２の）会議ステーション１０４（図１）の映像表示出力に関して不定のフェーズを有する。従って、リモートの会議ステーション１０４の映像表示出力は、ローカルの会議ステーション１０２に設置されたビデオカメラ１１０とフェーズの不一致がある場合がある。

【００１７】

更に、ローカルの会議ステーション１０２からリモートの会議ステーション１０４に供給側の映像信号を送信する際に、供給側の映像信号がローカルの会議ステーション１０２で生成される時間と供給側の映像信号がリモートの会議ステーション１０４で表示される時間との間の送信遅延が存在する。リモートの会議ステーション１０４の映像表示出力がローカルの会議ステーション１０２に配置されたビデオカメラ１１０とフェーズの不一致があると、送信遅延は補正できない。その結果、送信遅延がリモートの会議ステーション１０４の映像表示出力に加えられ、双方向テレビ会議にマイナスの効果を生じ得る。例えば、ローカルの会議ステーション１０２のユーザが一時停止後に話し始めると、送信遅延のためにリモートの会議ステーション１０４の参加者は一時停止中のユーザを依然として見る場合がある。リモートの会議ステーション１０４の何らかの参加者がこの時点でユーザを割り込むと、リモートの参加者とユーザがお互いに話すことになる。

【００１８】

有利には、本発明は、映像出力で送信遅延が補正又は減少され得るように、ローカルの会

議ステーション１０２に設置されたビデオカメラ１１０とリモートの会議ステーション１０４の映像表示出力との間のフェーズを同期する。特に、テレビ会議中にローカルの会議ステーション１０２のビデオカメラ１１０が特定の頻度と速度で動き、リモートの会議ステーション１０４の映像表示出力に関してフェーズのずれを引き起こす。ローカルの会議ステーション１０２のビデオカメラ１１０の動きは、ビデオカメラ１１０と映像表示出力との間のフェーズを同期する基準として測定され、用いられ得る。フェーズ同期エンジン３０８は、フェーズ同期又は固定機能を実行するフェーズ同期モジュールを保存するメモリ装置３１４を有する。

【００１９】

動作中に、供給源の映像信号を送信するために、映像処理エンジン３１０はまずビデオセンサー２１２（又はビデオカメラ１１０）から高解像度の画像を受信し、ビデオメモリ（図示なし）に画像を保存する。映像処理エンジン３１０は、好ましくはビデオメモリに保存された高解像度画像内で２つの画像セクション（ビュー）を規定し、２つの画像セクション（ビュー）について２つのそれぞれの映像ストリームを生成する。その他、それより多い又は少ない画像セクションと対応する映像ストリームが考えられる。その後、映像処理エンジン３１０は２つの映像ストリームを通信インタフェース３０４に送信する。逆に、リモートの場所からリモートの映像信号を表示するために、映像処理エンジン３１０は通信インタフェース３０４から少なくとも２つの映像ストリーム（すなわち、映像ストリームＡ及びＢ）を受信する。その後、映像処理エンジン３１０は、映像ストリームＡ及びＢを処理し、２つの映像ストリームＡ及びＢについてスクリーン２０８に２つの画像のビューをそれぞれ表示する。

【００２０】

供給側の音声信号を送信するために、会議バー２０４の複数のマイクロフォン２１４（図２）のそれぞれが、音源（例えば話し中の参加者）から音を受信し、受信音を電気又は電流信号に変換する。音源と会議バー２０４に関して異なる位置に複数のマイクロフォン２１４が設置されているため、複数のマイクロフォン２１４の電気信号又は電流信号が異なる大きさを有する。電気信号又は電流信号の大きさの違いは音源の位置を示す。複数のマイクロフォン２１４から電気信号又は電流信号を受信すると、音声処理エンジン３１２は音声信号と位置信号を生成する。位置信号は会議バー２０４に関する話す人の位置を示す情報を有し得る。音源の位置が変わると、音声処理エンジン３１２は新しい位置信号を生成し、位置の変更を反映する。その後、音声処理エンジン３１２は音声信号と位置信号を通信インタフェース３０４に送信する。

【００２１】

逆に、リモートの場所からリモートの音声信号を再生するために、音声処理エンジン３１２は、まず通信インタフェース３０４から音声信号と位置信号を受信する。その後、音声処理エンジン３１２が位置信号に応じて会議バー２０４の１つ以上のスピーカー２１０ａ－２１０ｄ（図２）を駆動し、映像処理エンジン３１０がスクリーン２０８に１つ以上の画像を表示する。会議バー２０４のスピーカー２１０ａ－２１０ｄは、スクリーン２０８に表示された話し中の参加者の位置に基づいて選択される。スクリーン２０８が比較的大きいサイズを有しているため、音が話し中の参加者の位置から来ることがわかるようにすることによって本発明がテレビ会議を改善する。２５０Ｈｚより上の周波数内の音は指向的な特性を有するため、会議バー２０４のスピーカーの配列におけるスピーカー２１０ａ－２１０ｄは一般的に２５０Ｈｚより上の周波数で動作することに留意すべきである。従って、映像処理ユニット２０６に設置されたサブウーファー（ｓｕｂ－ｗｏｏｆｅｒ）装置３０６（図３）は、２５０Ｈｚ未満から５０－１００Ｈｚまでの周波数内の音が指向性を有さないため、好ましくはその周波数で動作する。本発明はサブウーファー（ｓｕｂ－ｗｏｏｆｅｒ）装置３０６を有するものとして説明されるが、その技術に熟練した人はサブウーファー（ｓｕｂ－ｗｏｏｆｅｒ）装置３０６が本発明の動作及び機能に必要ないことがわかる。本発明において何らかの周波数帯の音が利用され得ることもまた、その技術に熟練した人がわかる。例えば、更に低い周波数が会議バー２０４のスピーカーの配列に

おけるスピーカー２１０ａ－２１０ｄに用いられ得る。

【００２２】

通信インタフェース３０４は、送受信装置３１６と通信処理エンジン３１８とを有する。音声信号と位置信号と２つの映像ストリームＡ及びＢとを含む通信信号の送信は、通信処理エンジン３１８を必要とし、音声処理エンジン３１２から音声信号と位置信号を受信し、映像処理エンジン３１０から２つの映像ストリームＡ及びＢを受信する。その結果、通信処理エンジン３１８は、通信信号をコード化し、圧縮して、それを送受信装置３１６に送信する。通信信号を受信すると、送受信装置３１６は通信チャネル１１８を通じて通信信号をリモートの場所に転送する。

【００２３】

逆に、音声信号と位置信号と２つの映像ストリームＡ及びＢを含む通信信号を受信するために、送受信装置３１６は通信チャネル１１８から通信信号を受信し、通信信号を通信処理エンジン３１８に転送する。その後、通信処理エンジン３１８は通信信号を回答し、デコードして、音声信号と位置信号と２つの映像データストリームを回復する。

【００２４】

図４は、図３の映像処理エンジン３１０の構成要素を示した例示的なブロック図である。映像処理エンジン３１０は、ビデオセンサー２１２（図２）につなげられたデータロードエンジン４０２と、映像メモリ４０４と、ＦＰＧＡ／ＡＳＩＣ４０６とを有する。データロードエンジン４０２がビデオセンサー２１２からビデオ画像データを受信し、映像メモリ４０４に保存し、ＦＰＧＡ／ＡＳＩＣ４０６がデータロードエンジン４０２と映像メモリ４０４を制御する。ビデオセンサー２１２は好ましくは高解像度のデジタル画像センサーであるため、ビデオセンサー２１２は大量の画像データを生成し得る。例えば、３，０００×２，０００の解像度でビデオセンサー２１２は１つの画像について６，０００，０００ピクセルを生成する。入力処理能力を増加させるために、データロードエンジン４０２は好ましくは６個の並列データチャネル１－６を有する。ＦＰＧＡ／ＡＳＩＣ４０６は前記６個の並列データチャネル１－６を通じて映像メモリ４０４に全画像ピクセルを供給するようにプログラムされる。ＦＰＧＡ／ＡＳＩＣ４０６はまた、選択可能な解像度で映像メモリ４０４に保存された画像上で少なくとも２つの画像セクション（ビュー）を規定し、２つの画像セクション（ビュー）について２つの映像ストリームをそれぞれ生成するようにプログラムされる。本発明の実施例は６個のデータチャネルを利用することを考えるが、何らかの数のデータチャネルが本発明によって使用され得る。更に、何らかの数の画像セクションと対応する映像ストリームが本発明で利用され得る。

【００２５】

図５は、ＦＰＧＡ／ＡＳＩＣ４０６（図４）によって規定され、ディスプレイ２０２（図２）で見られる本発明の一実施例による例示的な画像セクション（又はビュー）の構成である。図５において大きいセクションＡ５０２は７００×４００の解像度を有する画像の全てのビューを規定し、小さいセクションＢ５０４は、リモートの会議ステーションから話し中の参加者が表示される３００×２００の解像度を有するビューを規定する。映像メモリ４０４（図４）に保存された画像に基づき、ＦＰＧＡ／ＡＳＩＣ４０６は全画像を７００×４００の解像度に縮小し、大きいセクションＡ５０２のための映像ストリームＡ（図３）を作る。その後、ＦＰＧＡ／ＡＳＩＣ４０６はセクションＢ５０４の画像を３００×２００の解像度に縮小し、映像ストリームＢ（図３）を作る。映像メモリ４０２に保存された画像は、比較的高解像度を有するため、２つの縮小された画像は依然として良い解像度の質を示す。本発明において他の解像度が利用され得ることがその技術に熟練した人は認識するであろう。

【００２６】

有利には、本発明は会議場所の全画像を生成し、全画像のうちの何らかの任意のセクションからビューをズームすることが可能である。更に、１つの画像について少なくとも２つの映像ストリームが生成されるため、特定の話し中の参加者を示すはめ込みのズームされたビュー（例えばセクションＢ５０４）と共に、会議場所の全参加者を含む広角度の高解

像度の画像（例えばセクションＡ５０２）を送信することが可能である。その他、単一の画像からそれより多い又は少ないストリームが作られ、その結果それより多い又は少ないビューが表示され得る。従って、本発明は従来の機械的なパン／チルト／ズーム（ｐａｎ／ｔｉｌｔ／ｚｏｏｍ）カメラに代わって用いられ得る。

【００２７】

現在の技術で、一般的なＣＭＯＳビデオセンサーは、およそ６５度のビューの角度を有効に提供し得る。実際には、９０度のビューの角度が必要になる場合がある。従って、小さく安価なパン（ｐａｎ）モーターが水平方向にＣＭＯＳビデオセンサーを動かすために用いられ得る。しかし、ＣＭＯＳビデオセンサーの動きと結果として生じるノイズが比較的小さいため、その動きと結果として生じるノイズは会議の参加者にほとんど目立たない。技術の発達でＣＭＯＳビデオセンサーはコスト効率の良い９０度のビュー角度を提供することができるであろう。

【００２８】

図６において、テレビ会議システムにおいて音声データを送信する処理を示した例示的なフローチャート６００が示されている。ステップ６１０において、受信音を電気信号又は電流信号に変換することによって音源に応じて、第１の場所の送信ステーションで複数のマイクロフォン２１４（図２）によって音声信号が生成される。次に、ステップ６２０で音源の位置を示す位置信号が生成される。送信ステーションからの音源の位置に応じて、電流信号は特定の大きさを有する。電流信号の大きさに基づいて音声処理エンジン３１２（図３）が位置信号を規定する。その後、音声信号及び位置信号が通信インタフェース３０４（図３）に送信され、ステップ６３０で通信処理エンジン３１８（図３）によって処理される。前記処理は、送信のために音声信号及び位置信号を圧縮し、コード化することを含み得る。その後、ステップ６４０において、音声信号及び位置信号が、インターネットや、ＬＡＮや、ＷＡＮや、何らかの形式のネットワーク通信手段のような通信チャネルを通じて送受信装置によって送信される。ステップ６５０において、第２の場所の受信ステーションの送受信装置が音声信号及び位置信号を受信する。ステップ６６０で通信処理エンジンが音声信号及び位置信号を処理し、そのことは再生のための音声信号及び位置信号を解凍し、デコードすることを有し得る。その後、ステップ６７０において、位置信号に基づいて受信ステーションの１つ以上のスピーカーが音声信号を再生するために駆動される。スピーカーのうちの１つの音声信号の再生により音源の位置から音声信号が来ることがわかるようになるため、音声処理エンジンによって生成された位置信号が更に現実的なテレビ会議の状況を作り出す。その後、ステップ６８０において、更なるテレビ会議が生じているかどうかをシステムが決定する。会議が続く場合には、システムがステップ６１０から６７０を繰り返す。

【００２９】

図７において、テレビ会議システムで高解像度の画像を送信する処理を示した例示的なフローチャート７００が示されている。ステップ７１０において、ビデオカメラ又はビデオセンサーが高解像度の画像を取得する。その後、高解像度の画像がロードされ、ビデオカメラ又はビデオセンサーから映像メモリに保存される。次にステップ７２０で画像が映像ストリームに変換される。映像メモリに保存された高解像度の画像内で、第１と第２の画像セクションが最初に送信ステーションの映像処理エンジンによって規定される。その後、第１と第２の画像セクションが第１の解像度を有する第１の映像ストリームと第２の解像度を有する第２の映像ストリームとに変換される。変換は映像処理エンジン３１０（図３）のＦＰＧＡ／ＡＳＩＣ４０６（図４）によって実行され、第１の画像セクションを７００×４００の解像度を有する第１の映像ストリームに変換し、第２の画像セクションを３００×２００の解像度を有する第２の映像ストリームに変換する。本発明において他の解像度も利用される場合があり、２つより多い又は少ない画像セクションと２つより多い又は少ない映像ストリームもまた利用され得ることが、その技術に熟練した人は認識するであろう。

【００３０】

ステップ７３０において、映像ストリームは送信ステーションの通信処理エンジンによって処理される。前記処理は送信のためにストリームのコード化と圧縮を有し得る。一般的に映像データの更に高速な送信を可能にするために、映像ストリームがコード化され、圧縮される。次にステップ７４０において、処理された映像ストリームが通信チャネルを通じて受信ステーションに送信される。通信チャネルは何らかのパケット交換網、（非同期転送モード（“ＡＴＭ”）ネットワークのような）回線交換網、又は周知のインターネットを含むデータを運ぶ何らかの他のネットワークである場合がある。通信チャネルはまた、インターネット、エクストラネット、ローカルエリアネットワーク、又はその技術において周知の他のネットワークである場合がある。ステップ７５０において、映像ストリームが受信ステーションの映像処理エンジンによってデコードされ、解凍され、受信ステーションの映像ディスプレイに表示される。ステップ７６０において、更にテレビ会議が生じているかどうかをシステムが決定する。会議が続く場合には、システムはステップ７１０から７５０を繰り返す。音声と位置と映像のデータの送信を別のフローチャートと方法で説明したが、本発明は前記データの同時又はほぼ同時の送信を考慮する。

【００３１】

図８において、テレビ会議システムで映像信号を送信する他の処理を示した例示的なフローチャート８００が示されている。ステップ８１０において、ビデオカメラ又はビデオセンサーがビデオ画像を取得する。次に、ステップ８２０で映像信号が送信ステーションの通信エンジンで処理される。前記処理は映像信号のコード化と圧縮を含み得る。一般的に映像データの更に高速な送信を可能にするために、映像ストリームがコード化され、圧縮される。ステップ８３０において、フェーズ同期エンジンがビデオカメラと映像表示出力との間のフェーズを同期する。ビデオカメラと映像表示出力との間のフェーズの同期は、送信遅延によって引き起こされる悪影響を最小にすることを可能にする。特に、ビデオカメラが映像表示出力とフェーズの不一致がある場合には、送信ステーションのユーザが話し始めた後でもなお、受信ステーションの参加者が送信ステーションの一時停止中のユーザを依然として見る場合がある。

【００３２】

次に、ステップ８４０で映像信号が通信チャネルを介して受信ステーションに送信される。通信チャネルは何らかのパケット交換網、（非同期転送モード（“ＡＴＭ”）ネットワークのような）回線交換網、又は周知のインターネットを含むデータを運ぶ何らかの他のネットワークである場合がある。通信チャネルはまた、インターネット、エクストラネット、ローカルエリアネットワーク、又はその技術において周知の他のネットワークである場合がある。その後、ステップ８５０において、映像信号が受信ステーションの通信処理エンジンによって映像表示出力での表示のために処理される。前記処理は、映像信号のデコードと解凍を含み得る。デコードされ、解凍された映像信号に応じて映像表示出力が生成され、受信ステーションの映像ディスプレイに表示される。ステップ８６０において、更にテレビ会議が生じているかどうかをシステムが決定する。会議が続く場合には、システムはステップ８１０から８５０を繰り返す。

【００３３】

本発明は、例示的な実施例を参照して述べられた。実施例と共に開示された多様な特徴が別々に又は一緒に用いられる場合があり、多様な改良が行われる場合があり、本発明の更に広い範囲を逸脱することなく他の実施例が用いられ得ることが、その技術に熟練した人は認識するであろう。例えば、本発明の位置合わせ装置が好ましい実施例を参照して説明されたが、いくつもの環境と実施において本発明が有利に利用され得ることをその技術に通常に熟練した人が認識することがわかる。従って、ここで開示された発明の全範囲と意図を考慮して特許請求の範囲が解釈されるべきである。

【図面の簡単な説明】

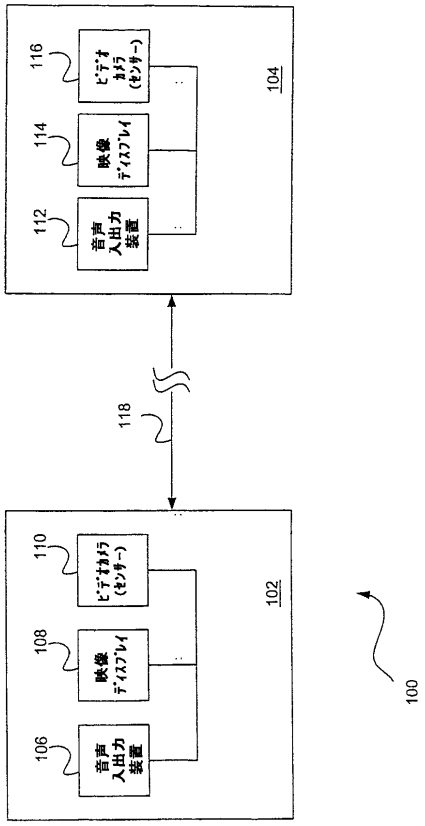
【００３４】

【図１】本発明による例示的なテレビ会議システムを示したものである。

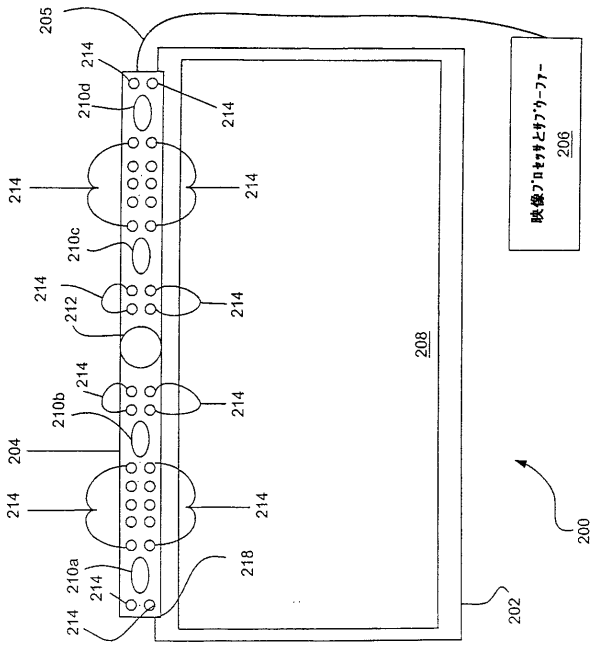
【図２】例示的な会議ステーションを示したものである。

【図３】図２の処理ユニットを更に詳細に示す例示的なブロック図である。
【図４】図３の映像処理エンジンの構成要素を示した例示的なブロック図である。
【図５】本発明による例示的なセクション（又はビュー）の構成である。
【図６】テレビ会議システムで音声を送信する例示的な処理を示したフローチャートである。
【図７】テレビ会議システムで高解像度の画像を送信する例示的な処理を示したフローチャートである。
【図８】テレビ会議システムで映像信号を送信する例示的な処理を示したフローチャートである。

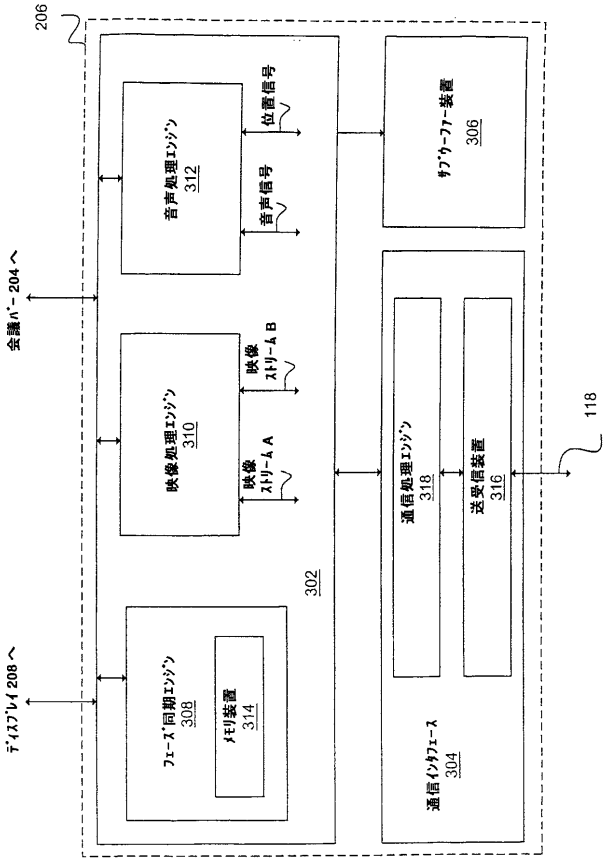
【図１】



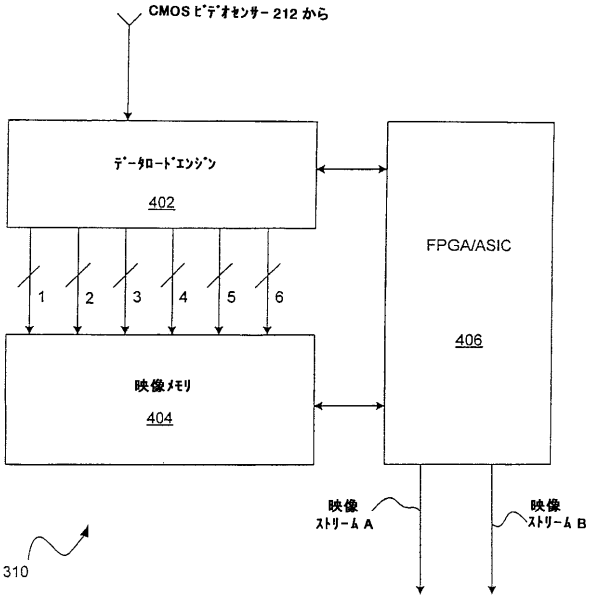
【図２】



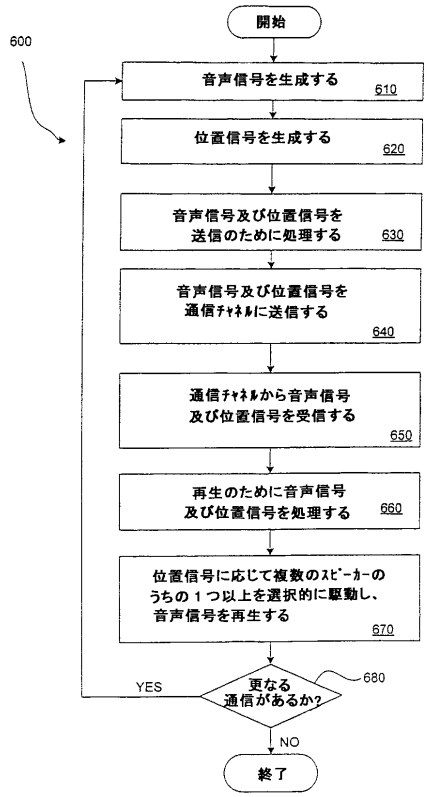
【図３】


206
音声処理エンジン 312
位置信号
音声信号
バッファー装置 306
会議バー 204へ
映像処理エンジン 310
映像ストリーム B
映像ストリーム A
通信処理エンジン 318
送受信装置 316
118
302
フェーズ同期エンジン 308
メモリ装置 314
通信インタフェース 304
ディスプレイ 208へ

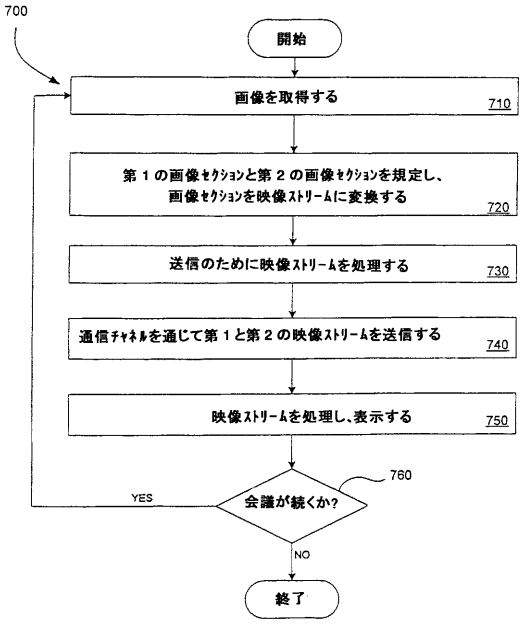
【図４】


CMOS ビデオセンサー 212 から
データロードエンジン 402
FPGA/ASIC 406
1
2
3
4
5
6
映像メモリ 404
映像ストリーム A
映像ストリーム B
310

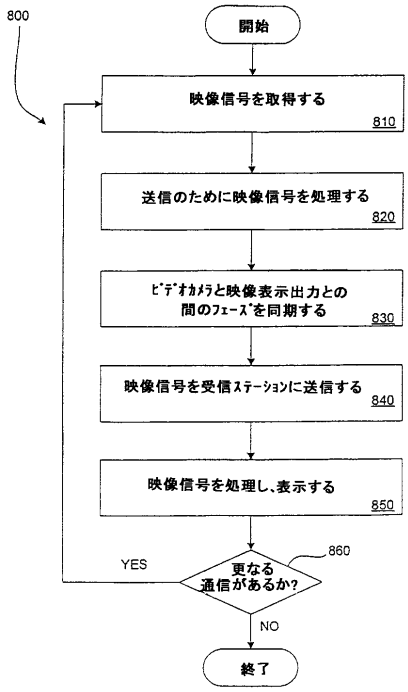
【図６】


600
開始
音声信号を生成する 610
位置信号を生成する 620
音声信号及び位置信号を送信のために処理する 630
音声信号及び位置信号を通信チャネルに送信する 640
通信チャネルから音声信号及び位置信号を受信する 650
再生のために音声信号及び位置信号を処理する 660
位置信号に応じて複数のスピーカーのうちの１つ以上を選択的に駆動し、音声信号を再生する 670
更なる通信があるか? 680
YES
NO
終了

【図７】


700
開始
画像を取得する 710
第１の画像セクションと第２の画像セクションを規定し、画像セクションを映像ストリームに変換する 720
送信のために映像ストリームを処理する 730
通信チャネルを通じて第１と第２の映像ストリームを送信する 740
映像ストリームを処理し、表示する 750
会議が続くか? 760
YES
NO
終了

【図８】


800
開始
映像信号を取得する
810
送信のために映像信号を処理する
820
ﾋﾞﾃﾞｵｶﾒﾗと映像表示出力との
間のﾌｪｰｽﾞを同期する
830
映像信号を受信ｽﾃｰｼｮﾝに送信する
840
映像信号を処理し、表示する
850
860
更なる
通信があるか?
YES
NO
終了

【国際公開パンフレット】

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
20 February 2003 (20.02.2003)

PCT

(10) International Publication Number
WO 03/015407 A1

(51) International Patent Classification⁷: H04N 7/14

(21) International Application Number: PCT/US02/25477

(22) International Filing Date: 7 August 2002 (07.08.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/310,742 7 August 2001 (07.08.2001) US

(71) Applicant: **POLYCOM, INC.** [US/US]; 1565 Barber Lane, Milpitas, CA 95035 (US).

(72) Inventors: **KENOYER, Michael**; 4601 Candle Ridge, Austin, TX 78731 (US). **MALLOY, Craig**; 8307 Club Ridge Dr., Austin, TX 78735 (US). **WASHINGTON, Richard**; 24912 Singleton Bend Road #7, Marble Falls, TX 78735 (US). **CHU, Peter**; 7 Hadley Road, Lexington, MA 02420 (US).

(74) Agents: **YEE, Susan** et al.; Carr & Ferrell, LLP, 2225 East Bayshore Suite #200, Palo Alto, CA 94303 (US).

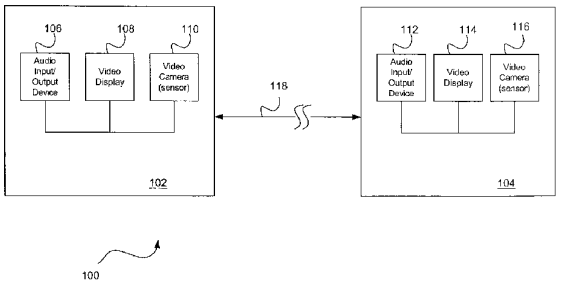
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
with international search report
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: SYSTEM AND METHOD FOR HIGH RESOLUTION VIDEOCONFERENCING



(57) Abstract: A system and method for high resolution video conferencing is shown and described. A transmitting station (102) and a receiving station (104) including video cameras (110/116) or sensors, a plurality of microphones (214) and speakers (210a-210d), video (310), audio (312), and communication (318) processing engines are disclosed. Video is processed and transferred through the system allowing for multiple video streams to be produced and audio is processed and transferred through the system allowing for sound to be played back with an indication of position in relation to the videoconferencing system.

WO 03/015407 A1

SYSTEM AND METHOD FOR HIGH RESOLUTION VIDEOCONFERENCING

CROSS-REFERENCE TO RELATED APPLICATIONS

The present application claims the benefit of priority from U.S. Provisional Patent Application No. 60/310,742, entitled "High Resolution Video Conferencing Bar" filed on August 7, 2001, which is herein incorporated by reference.

BACKGROUND OF THE INVENTION

1. Field of the Invention

[0002] The present invention relates generally to conferencing systems, and more particularly to a high resolution videoconferencing system.

2. Description of the Background Art

[0003] Conventionally, videoconferencing systems utilize video cameras to capture an image of the conference participants for transmission to a remote conferencing site. A conventional (stationary or movable) video camera can only capture one image or one view of a conferencing site at a certain point in time. In order to capture different images or views of a conferencing site at different points in time, a conventional video camera may be beneficially provided with a device for adjusting a rotational orientation of the camera. Positioning devices designed to rotate the camera about two orthogonal axes typically utilize two actuators: a first actuator rotates the camera about a vertical axis and a second actuator rotates the camera about a horizontal axis perpendicular to the camera's vertical axis. Rotation of the camera about the horizontal axis is referred to as "panning", while rotation about the vertical axis is referred to as "tilting." As such, devices for rotating the camera about the horizontal and vertical axis are commonly referred to as "pan/tilt positioning devices." Further, to capture an image or view that is of a particular interest, such as the

image of a speaking conference participant, a conventional video camera would require a set of zoom lenses for performing zooming functions, resulting in a "pan/tilt/zoom" ("PZT") camera.

[0004] Disadvantageously, conventional PZT cameras have many shortcomings. First, movement of mechanical components in the positioning device can generate a substantial amount of noise. These movements and noise can be annoying and distracting to the conference participants. More importantly, the noise can interfere with acoustic localization techniques utilized to automatically orient the camera in a direction of the speaking participant. Secondly, the mechanical components in the positioning device may be susceptible to misalignment or breakage due to wear or rough handling, thereby rendering the positioning device partially or fully inoperative. A further disadvantage is complexity in manufacturing of the positioning device; thus resulting in high manufacturing costs and, subsequently, high consumer prices.

[0005] With the development of technology, sizes of display screens in videoconferencing systems are getting larger and larger. Consequently, positions of participant speakers on the display screen can change over a large span area. Disadvantageously, however, conventional videoconferencing systems are unable to adjust to a new participant speaker position as the position changes over the large span area.

[0006] Therefore, there is a need for a videoconferencing system and method which captures multiple views of a conferencing site without involving a complex mechanical structure. There is another need for a videoconferencing system and method which adjusts acoustics relative to a speaker's position.

SUMMARY OF THE INVENTION

[0007] The present invention provides for a videoconferencing system comprising a transmitting station located at a first site, including a plurality of microphones for generating an audio signal in response to a sound source; an audio processing engine for generating a position signal that indicates the position of the sound source and for processing the audio signal; and a communication interface for transmitting the audio and position signals to a communication channel. The plurality of microphones of the videoconferencing system can be arranged in an n-fire configuration as well as a vertical array. The videoconferencing system may also comprise a receiving station located at a second site, including a communication interface for receiving the audio and position signals from the communication channel, a plurality of speakers for playing the audio signal, and an audio processing engine for selectively driving one of the speakers in response to the position signal to play the audio signal on the selected speaker.

[0008] The position signal generated by the videoconferencing system is based upon magnitude differences of electric or current signals received from the plurality of microphones. Whereas, if the position of the sound source changes, the audio processing engine generates a new position signal to reflect a position change.

[0009] The transmitting station communication interface includes a communication processing engine for encoding and compressing the audio signal and the position signal, and a transceiver device for transmitting the audio and position signals through the communication channel. Conversely, the receiving station communication interface includes a transceiver device, for receiving the audio and position signals through the communication channel, and a communication processing engine for decoding and decompressing the audio signal and the position signal.

[00010] In another embodiment, a videoconferencing system comprises a transmitting station located at a first site, including a high resolution video sensor for generating an image, a video memory for storing the high resolution image, a data loading engine for loading image data from the video sensor to the video memory. Additionally, a Field Programmable Gate Array/Application Specific Integrated Circuit (FPGA/ASIC) is coupled to the video memory and data loading engine. The FPGA/ASIC defines a first image section and a second image section within the high resolution image stored in the video memory. Further

the FPGA/ASIC can scale the first image section into a first video stream with a first resolution and scale the second image section into a second video stream with a second resolution. A communication interface coupled to the FPGA/ASIC transmits the first video stream and the second video stream to a communication channel. The videoconferencing system may also comprise a receiving station located at a second site, including a communication interface for receiving the first video stream and the second video stream from the communication channel. The receiving station further includes a video processing engine for processing the first video stream and the second video stream and for displaying the first video stream as a first image with a first resolution and displaying the second video stream as a second image with a second resolution, is coupled to the communication interface.

[00011] The transmitting station communication interface in this embodiment comprises a communication processing engine for encoding and compressing the first and second video stream, and a transceiver device for transmitting the first and second video stream through the communication channel. Conversely, the receiving station video processing engine of the present embodiment comprises a video memory for storing the first video stream and the second video stream, a data loading engine for loading the first video stream and the second video stream from the receiving station communication interface and an FPGA/ASIC for displaying the first and second image data stream based on the high resolution image stored in the video memory.

[00012] In yet another embodiment, a videoconferencing system comprises a receiving station located at a first site having a communication interface for receiving a video signal from a communication channel, a video processing engine for generating a video display output in response to the video signal, and a video display for displaying the video display output. The videoconferencing system may further comprise a transmitting station located at a second site, having a video camera for generating the video signal, a video processing engine for processing the video signal, a phase synchronization engine for synchronizing a phase between the video camera at the transmitting station and the video display output at the receiving station, and a communication interface for transmitting the video signal to the communication channel.

BRIEF DESCRIPTION OF THE FIGURES

[00013] FIG. 1 shows an exemplary videoconferencing system in accordance with the present invention;

[00014] FIG. 2 shows an exemplary conferencing station;

[00015] FIG. 3 is an exemplary block diagram illustrating the processing unit of FIG. 2 in greater detail;

[00016] FIG. 4 is an exemplary block diagram illustrating components in the video processing engine of FIG. 3;

[00017] FIG. 5 is an exemplary section (or view) configuration in accordance with the present invention;

[00018] FIG. 6 is a flowchart illustrating an exemplary process for transmitting audio in a videoconferencing system;

[00019] FIG. 7 is a flowchart illustrating an exemplary process for transmitting high resolution images in a videoconferencing system; and

[00020] FIG. 8 is a flowchart illustrating an exemplary process for transmitting a video signal in a videoconferencing system.

DESCRIPTION OF THE INVENTION

[00021] FIG. 1 shows an exemplary videoconferencing system 100 in accordance with the present invention. The videoconferencing system 100 includes a first conferencing station 102 and a second conferencing station 104. The first conferencing station 102 includes an audio input/output device 106, a video display 108 and a video camera (or video sensor) 110. Similarly, the second conferencing station 104 includes an audio input/output device 112, a video display 114 and a video camera (or a video sensor) 116. The first conferencing station 102 communicates with the second conferencing station 104 through a communication channel 118. The communication channel 118 can be an Internet, a LAN, a WAN, or any other type of network communication means. Although FIG. 1 only shows two conferencing stations 102 and 104, those skilled in the art will recognize that additional conferencing stations may be coupled to the videoconferencing system 100.

[00022] FIG. 2 shows an exemplary conferencing station 200, similar to the conferencing stations 102 and 104 of FIG. 1, in accordance with one embodiment of the present invention. The conferencing station 200 includes a display 202, a high resolution conferencing bar 204, and a video processing unit 206. Preferably, the display 202 is a High Definition ("HD") monitor having a relatively large-size flat screen 208 with a 16:9 viewable area. Alternatively, other view area proportions and other types of displays 202 are contemplated and may be used.

[00023] Preferably, the high resolution video conferencing bar 204 contains multiple speakers 210a to 210d, a video sensor (e.g., a high resolution digital video image sensor such as a CMOS video sensor) 212, and a plurality of microphones 214. The speakers 210a to 210d preferably operate at frequencies above 250 Hz. However, the speakers 210a to 210d may operate at any other frequency compatible with various embodiments of the present invention. In one embodiment, the conferencing bar 204 is approximately 36 inch wide by 2 inch high and by 4 inch deep, although the conferencing bar 204 may comprise any other dimension. Typically, the conferencing bar 204 is designed to sit atop the display 202 with a front portion 218 extending slightly below a front edge of the display 202. The positioning of the conferencing bar 204 brings the speakers 210a to 210d, the video sensor 212, and the plurality of microphones 214 closer to the screen 208, and provides a positioning reference at

the front edge of the display 202. Other conference bar 204 positions may be utilized in keeping with the scope and objects of the present invention. Further, although only four speakers are shown in FIG. 2, more or less speakers may be utilized in the present invention.

[00024] The video sensor 212 has the capability to output multiple images in real-time at a preferred resolution of 720i (i.e., 1280x720 interlaced at 60 fields per second) or higher, although other resolutions are contemplated by the present invention. The resolution of the video sensor 212 is sufficient based on approximately a 65 degree field of view to capture an entire conferencing site. For a wider degree field of view (such as a 90 degree field of view), a limited horizontal pan motor may be provided. Providing this limited horizontal pan motor results in the avoidance of a costly and complicated full mechanical pan/tilt/zoom camera and lens system. Further, a pure digital zoom may be provided with a fixed lens to accommodate up to an 8x or higher effective zoom while maintaining a minimum Full CIF (352x288) resolution image.

[00025] The plurality of microphones 214 are located on both sides of the video sensor 212 on the conferencing bar 204, and can be arranged in an n-fire configuration, as shown in FIG. 2, which provides a better forward directional feature. A vertical microphone array can be optionally arranged along a side of the display 202 to provide vertical positioning references.

[00026] The conferencing bar 204 is coupled to the processing unit 206 via a high speed digital link 205. The processing unit 206 may contain a sub-woofer device that, preferably, operates from 250 Hz down to 50-100 Hz frequencies. The processing unit 206 will be discussed in more details in connection with FIG. 3. Although the processing unit 206 is shown as being separate from the conferencing bar 204, alternatively, the processing unit 206 may be encompassed within the conferencing bar 204.

[00027] Because conference participants may not feel comfortable in view of, or seeing the movement of, the video sensor 212, a smoked glass or similar covering can be installed in front of the video senor 212 and/or other portions of the conferencing bar 204 so that the conference participants cannot view the video sensor 212, and/or the speakers 210a to 210d and the plurality of microphones 214.

[00028] FIG. 3 is an exemplary block diagram illustrating the processing unit 206 of FIG. 2 in greater detail in accordance with one embodiment of the present invention. The processing unit 206 preferably includes a processing engine 302, a communication interface 304, and a sub-woofer device 306. The processing engine 302 further comprises a phase synchronization engine 308, a video processing engine 310, and an audio processing engine 312. The phase synchronization engine 308 is able to reduce or minimize negative impact caused by transmission delay. Specifically, the video camera 110 (FIG. 1) at the local (or first) conferencing station 102 (FIG. 1) has an arbitrary phase relative to a video display output at a remote (or second) conferencing station 104 (FIG. 1). Thus, the video display output at the remote conferencing station 104 may be out of phase with the video camera 110 located at the local conferencing station 102.

[00029] Further, in transmitting a source video signal from the local conferencing station 102 to the remote conferencing station 104, there is a transmission delay between a time when a source video signal is being generated at the local conferencing station 102 and a time when the source video signal is displayed at the remote conferencing station 104. The transmission delay cannot be compensated for when the video display output at the remote conferencing station 104 is out of phase with the video camera 110 located at the local conferencing station 102. As a result, the transmission delay is added to the video display output at the remote conferencing station 104, which may generate a negative effect in an interactive video conference. For example, when a user at the local conferencing station 102 starts to speak after a pause, participants at the remote conferencing station 104 may still see the user in pause due to the transmission delay. If any of the participants at the remote conferencing station 104 interrupts the user at this moment, the remote participant and the user will talk over each other.

[00030] Advantageously, the present invention synchronizes the phase between the video camera 110 located at the local conferencing station 102 and the video display output at the remote conferencing station 104 so that the transmission delay can be compensated for or reduced in the video display output. Specifically, during a video conference, the video camera 110 at the local conferencing station 102 moves at a certain frequency and speed which causes phase shifting relative to the video display output at the remote conferencing station 104. The movement of the video camera 110 at the local conferencing station 102 can

be measured and used as a reference to synchronize the phase between the video camera 110 and the video display output. The phase synchronization engine 308 includes a memory device 314 for storing a phase synchronization module for performing the phase synchronization or locking function.

[00031] In operation, to transmit a source video signal, the video processing engine 310 first receives a high resolution image from the video sensor 212 (or video camera 110) and stores the image into a video memory (not shown). The video processing engine 310 then, preferably, defines two image sections (views) within the high resolution image stored in the video memory, and generates two respective video streams for the two image sections (views). Alternatively, more or less image sections and corresponding video streams are contemplated. The video processing engine 310 then sends the two video streams to the communication interface 304. Conversely, to display a remote video signal from a remote site, the video processing engine 310 receives at least two video streams (i.e., Video Streams A and B) from the communication interface 304. The video processing engine 310 then processes the video streams A and B and displays two image views on the screen 208 for the two video streams A and B, respectively.

[00032] To transmit a source audio signal, each of the plurality of microphones 214 (FIG. 2) in the conferencing bar 204 receives a sound from an acoustic source (e.g., from a speaking participant) and converts the received sound to an electric or current signal. Because the plurality of microphones 214 are located at different positions in reference to the conferencing bar 204 and the acoustic source, the electric or current signals in the plurality of microphones 214 have different magnitudes. The magnitude differences in the electric or current signals indicate a position of the acoustic source. Upon receiving the electric or current signals from the plurality of microphones 214, the audio processing engine 312 generates an audio signal and a position signal. The position signal may contain information indicating a speaker's position relative to the conferencing bar 204. If the position of the acoustic source changes, the audio processing engine 312 generates a new position signal to reflect the position change. The audio processing engine 312 then sends the audio and position signals to the communication interface 304.

[00033] Conversely, to play a remote audio signal from a remote site, the audio processing engine 312 first receives the audio signal and position signal from the communication

interface 304. The audio processing engine 312 then drives one or more of the speakers 210a to 210d (FIG. 2) in the conferencing bar 204 according to the position signal, while the video processing engine 310 is displaying one or more views of an image on the screen 208. The speakers 210a to 210d in the conferencing bar 204 are selected based on the position of the speaking participant displayed on the screen 208. Because the screen 208 has a relatively large size, the present invention improves video conference by making it appear as if the sound is coming from the location of the speaking participant. It should be noted that the speakers 210a to 210d in the speaker array of the conferencing bar 204 operate, typically, at frequencies above 250 Hz, because the sounds within this frequency range have directional characteristics. Consequently, the sub-woofer device 306 (FIG. 3) installed within the video processing unit 206 operates, preferably, at frequencies from 250 Hz down to 50-100 Hz, because the sounds within this frequency range are not directional. Although the present invention is described as including the sub-woofer device 306, those skilled in the art will recognize that the sub-woofer device 306 is not required for operation and function of the present invention. Those skilled in the art will also recognize that any frequency range of acoustics may be utilized in the present invention. For example, lower frequencies may be used for the speakers 210a to 210d in the speaker array of the conferencing bar 204.

[00034] The communication interface 304 includes a transceiver device 316 and a communication processing engine 318. The transmission of a communication signal containing an audio signal, a position signal, and two video streams A and B requires the communication processing engine 318 to receive the audio and position signals from the audio processing engine 312 and the two video streams A and B from the video processing engine 310. Subsequently, the communication processing engine 318 encodes and compresses this communication signal and sends it to the transceiver device 316. Upon receiving the communication signal, the transceiver device 316 forwards the communication signal to a remote site through the communication channel 118.

[00035] Conversely, to receive a communication signal containing an audio signal, a position signal, and two video streams A and B, the transceiver device 316 receives the communication signal from the communication channel 118 and forwards the communication signal to the communication processing engine 318. The communication processing engine

318 then decompresses and decodes the communication signal to recover the audio signal, position signal, and two video data streams.

[00036] FIG. 4 is an exemplary block diagram illustrating components of the video processing engine 310 of FIG. 3. The video processing engine 310 includes a data loading engine 402 coupled to the video sensor 212 (FIG. 2), a video memory 404, and an FPGA/ASIC 406. The data loading engine 402 receives video image data from the video sensor 212 and stores it into the video memory 404, while the FPGA/ASIC 406 controls the data loading engine 402 and the video memory 404. Because the video sensor 212 is, preferably, a high resolution digital image sensor, the video sensor 212 can generate a large amount of image data. For example, with a 3,000x2000 resolution, the video sensor 212 generates 6,000,000 pixels for an image. To increase input bandwidth, the data loading engine 402, preferably, has six parallel data channels 1-6. The FPGA/ASIC 406 is programmed to feed entire image pixels to the video memory 404 through these six parallel data channels 1-6. The FPGA/ASIC 406 is also programmed to define at least two image sections (views) over the image stored in the video memory 404 with selectable resolutions, and to produce two video streams for the two image sections (views), respectively. Although the present embodiment contemplates utilizing six data channels, any number of data channels may be used by the present invention. Further, any number of image sections and corresponding video streams may be utilized in the present invention.

[00037] FIG. 5 is an exemplary image section (or view) configuration in accordance with one embodiment of the present invention defined by the FPGA/ASIC 406 (FIG. 4) and viewed on the display 202 (FIG. 2). In FIG. 5, a large section A 502 defines an entire view of an image having a 700x400 resolution, while a small section B 504 defines a view having a 300x200 resolution in which a speaking participant from a remote conferencing station is displayed. Based on the image stored in the video memory 404 (FIG. 4), the FPGA/ASIC 406 scales the entire image down to a 700x400 resolution image to produce the video stream A (FIG. 3) for the large section A 502. Subsequently, the FPGA/ASIC 406 scales the section B 504 image down to 300x200 resolution to produce the video stream B (FIG. 3). Because the image stored in the video memory 402 has a relatively high resolution, the two scaled

images still present good resolution quality. Those skilled in the art will recognize that other resolutions may be utilized in the present invention.

[00038] Advantageously, the present invention has the ability to generate a whole image of a conferencing site while zooming a view from any arbitrary section of the whole image. Further, because at least two video streams are produced for an image, it is possible to transmit a wide angle high resolution image including all participants at a conferencing site (e.g., section A 502) along with an inset zoomed view (e.g., section B 504) showing a particular speaking participant. Alternatively, more or fewer streams may be produced from a single image and consequently more or fewer views displayed. Therefore, the present invention can be used to replace conventional mechanical pan/tilt/zoom cameras.

[00039] With current technology, a typical COMS video sensor can effectively provide approximately 65 degree view angle. In reality, a 90 degree view angle may be required. Therefore, a small, inexpensive pan motor can be used to move the COMS video sensor in the horizontal direction. However, because the movement and the resulting noise of the CMOS video sensor are relatively small, such movement and resulting noise are hardly noticeable to the conferencing participants. With the development of technology, the COMS video sensor may be able to provide a cost effective 90 degree view angle.

[00040] In FIG. 6, an exemplary flowchart 600 illustrating a process for transmitting audio data in a videoconferencing system is shown. At step 610, an audio signal is generated at a transmitting station of a first site by the plurality of microphones 214 (FIG. 2) in response to an acoustic source by converting the received sound into an electric or current signal. Next, a position signal is generated at step 620 that indicates a position of the acoustic source. Depending upon the position of the acoustic source from the transmitting station, the current signal will have a particular magnitude. The audio processing engine 312 (FIG. 3) determines the position signal based on the magnitude of the current signal. The audio and position signals are then transmitted to the communication interface 304 (FIG. 3) and then processed at step 630 by the communications processing engine 318 (FIG. 3). This processing can include compressing and encoding the audio and position signals for transmission. The audio and position signals are then transmitted through a communication channel such as an Internet, a LAN, a WAN, or any other type of network communication

means at step 640 by a transceiver device. In step 650, a transceiver device at a receiving station of a second site receives the audio and position signals. A communications processing engine processes the audio and position signals at step 660, which may include decompressing and decoding the audio and position signals for playback. Subsequently, at step 670, based on the position signal, one or more speakers at the receiving station are driven to play the audio signal. The position signal generated by the audio processing engine creates a more realistic video conference situation because the playback of the audio signal on one of the speakers makes it appear as if the audio signal is coming from a location of the acoustic source. The system then determines whether more video conferencing is occurring in step 680. If the conference continues, the system repeats steps 610 though 670.

[00041] In FIG. 7, an exemplary flowchart 700 illustrating a process for transmitting high resolution images in a videoconferencing system is shown. At step 710, a video camera or video sensor captures a high resolution image. The high resolution image is then loaded and stored from the video camera or video sensor to a video memory. Next, the images are converted to video streams in step 720. Within the high resolution image stored in the video memory, a first and a second image section are initially defined by the transmitting station video processing engine. Subsequently, the first and second image sections are scaled to a first video stream having a first resolution and a second video stream having a second resolution. Scaling is implemented by the FPGA/ASIC 406 (FIG. 4) of the video processing engine 310 (FIG. 3), which scales the first image section to a first video stream having a 700x400 resolution and scales the second image section to a second video stream having a 300x200 resolution. Those skilled in the art will recognize that other resolutions may be utilized in the present invention, and that more or less than two image sections, and subsequently more or less than two video streams can also be utilized.

[00042] At step 730, the video streams are processed by a transmitting station communication processing engine. This processing can include encoding and compressing of the streams for transmission. Typically, the video streams are encoded and compressed to allow for faster transmission of the video data. Next, the processed video streams are sent to a receiving station through a communication channel in step 740. The communication channel may be any packet-switched network, a circuit-switched network (such as an

Asynchronous Transfer Mode ("ATM") network), or any other network for carrying data including the well-known Internet. The communication channel may also be the Internet, an extranet, a local area network, or other networks known in the art. The video streams are then decoded and decompressed by the receiving station video processing engine and displayed on a video display of the receiving station at step 750. The system then determines whether more video conferencing is occurring in step 760. If the conference continues, the system repeats steps 710 though 750. Although the transmission of audio, position, and video data are described in separate flowcharts and methods, the present invention contemplates the simultaneous or near simultaneous transmission of these data.

[00043] In FIG. 8, an exemplary flowchart 800 illustrating an alternative process for transmitting a video signal in a videoconferencing system is shown. At step 810, a video camera or video sensor captures a video image. Next, the video signal is processed by a transmitting station communication engine at step 820. This processing can include encoding and compressing the video signal. Typically, the video streams are encoded and compressed to allow for faster transmission of the video data. At step 830, a phase synchronization engine synchronizes a phase between the video camera and a video display output. The synchronizing of the phase between the video camera and the video display output allows for a minimization of a negative impact that can be caused by transmission delay. Specifically, if the video camera is out of phase with the video display output, participants at a receiving station may still see a user in pause at the transmitting station, even after the user at the transmitting station has begun to speak again.

[00044] Next, the video signal is transmitted to the receiving station at step 840 via a communication channel. The communication channel may be any packet-switched network, a circuit-switched network (such as an Asynchronous Transfer Mode ("ATM") network), or any other network for carrying data including the well-known Internet. The communication channel may also be the Internet, an extranet, a local area network, or other networks known in the art. Subsequently, at step 850, the video signal is processed for display on the video display output by a receiving station communication processing engine. This processing can include decoding and decompressing the video signal. The video display output is generated

in response to the decoded and decompressed video signal and displayed on a receiving station video display. The system then determines whether more video conferencing is occurring in step 860. If the conference continues, the system repeats steps 810 though 850.

[00045] The invention has been described with reference to exemplary embodiments. Those skilled in the art will recognize that various features disclosed in connection with the embodiments may be used either individually or jointly, and that various modifications may be made and other embodiments can be used without departing from the broader scope of the invention. For example, it is to be appreciated that while the positioning apparatus of the present invention has been described with reference to a preferred implementation, those having ordinary skill in the art will recognize that the present invention may be beneficially utilized in any number of environments and implementations. Accordingly, the claims set forth below should be construed in view of the full breadth and spirit of the invention as disclosed herein.

What is claimed is:

1. A videoconferencing device comprising:
a video sensor for capturing an image;
a plurality of microphones for generating an audio signal in response to an acoustic source; and
a processing engine coupled to the video sensor and the plurality of microphones for generating at least one video stream and a position signal indicating a position of the acoustic source.

2. The videoconferencing device of claim 1, further comprising a phase synchronization engine coupled to the video sensor for synchronizing a phase between the video sensor and a video display output.

3. The videoconferencing device of claim 1, further comprising a communication interface coupled to the processing engine for transmitting the audio signal, position signal, and at least one video stream to a remote videoconferencing device.

4. The videoconferencing device of claim 1, wherein the position signal is generated based upon magnitude differences of electric or current signals received from the plurality of microphones.

5. The videoconferencing system of claim 1, wherein the processing engine further comprises a video processing engine, the video processing engine defining a plurality of image sections and generating a respective plurality of video streams corresponding to the plurality of image sections.

6. The videoconferencing system of claim 1, wherein if the position of the sound source changes, the processing engine generates a new position signal to reflect a position change.

7. The videoconferencing device of claim 2, wherein the remote videoconferencing device selectively drives one or more speakers in response to the position signal to play the audio signal corresponding to the image of the at least one video stream.

8. The videoconferencing device of claim 1, wherein the plurality of microphones are arranged in an n-fire configuration.

9. The videoconferencing device of claim 1, wherein the plurality of microphones are arranged in a vertical array.

10. The videoconferencing device of claim 5, wherein the processing engine scale a first image section of the plurality of image sections into a first video stream having a first resolution and scales a second image section of the plurality of image sections into a second video stream having a second resolution.

11. The videoconferencing system of claim 1, further comprising a pan motor coupled to the video sensor for providing a larger degree view angle.

12. A method for transmitting conferencing data in a video conferencing system, comprising:

capturing an image with a video sensor and generating at least one video stream from the image;

capturing audio data with a plurality of microphones and generating an audio signal;

generating a position signal indicating a position of an acoustic source based upon magnitude differences of the audio data; and

transmitting the position signal, audio signals, and the at least one video streams via a communication channel.

13. The method of claim 12, further comprising selectively driving one or more speakers of a remote video conferencing system in response to the position signal to play the audio signal corresponding to the image of the at least one video stream.

14. The method of claim 12, further comprising synchronizing a phase between the video sensor and a video display output.

15. The method of claim 12, further comprising defining a plurality of image sections and generating a respective plurality of video streams corresponding to the plurality of image sections.

16. The method of claim 12, further comprising generating a new position signal to reflect a position change.

17. The method of claim 14, further comprising scaling a first image section of the plurality of image sections into a first video stream having a first resolution and scaling a second image section of the plurality of image sections into a second video stream having a second resolution.

18. A videoconferencing device comprising:

means for capturing an image and generating at least one video stream from the image;

means for capturing audio and generating an audio signal;

means for generating a position signal indicating a position of an acoustic source based upon magnitude differences of the audio data, the position signal selectively driving one or more speakers of a remote videoconferencing system in response to the position signal to play the audio signal corresponding to the image of the at least one video stream; and

means for transmitting the position signal, audio signals, and the at least one video streams via a communication channel.

19. An electronically-readable medium having embodied thereon a program, the program being executable by a machine to perform method steps for transmitting conferencing data, the method steps comprising:

capturing an image with a video sensor and generating at least one video stream from the image;

capturing audio data with a plurality of microphones and generating an audio signal;

generating a position signal indicating a position of an acoustic source based upon magnitude differences of the audio data; and

transmitting the position signal, audio signals, and the at least one video streams via a communication channel.

20. The electronically-readable medium of claim 19, wherein the method steps further comprise selectively driving one or more speakers of a remote videoconferencing system in response to the position signal to play the audio signal corresponding to the image of the at least one video stream.

21. The electronically-readable medium of claim 19, wherein the method steps further comprise defining a plurality of image sections and generating a respective plurality of video streams corresponding to the plurality of image sections.

22. The electronically-readable medium of claim 19, wherein the method steps further comprise scaling a first image section of the plurality of image sections into a first video stream having a first resolution and scaling a second image section of the plurality of image sections into a second video stream having a second resolution.

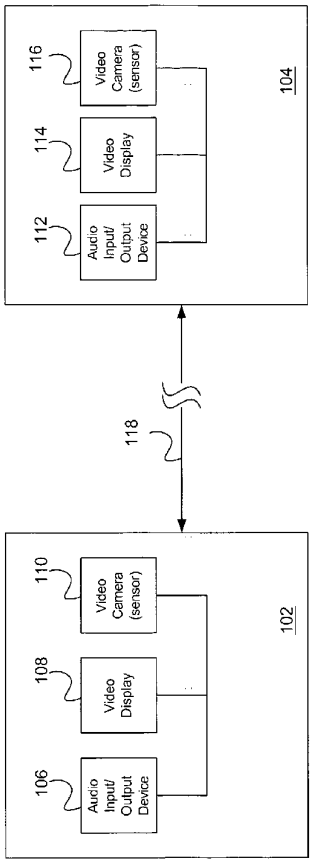

116
Video Camera (sensor)
114
Video Display
112
Audio Input/ Output Device
104
118
110
Video Camera (sensor)
108
Video Display
106
Audio Input/ Output Device
102
100

FIG. 1

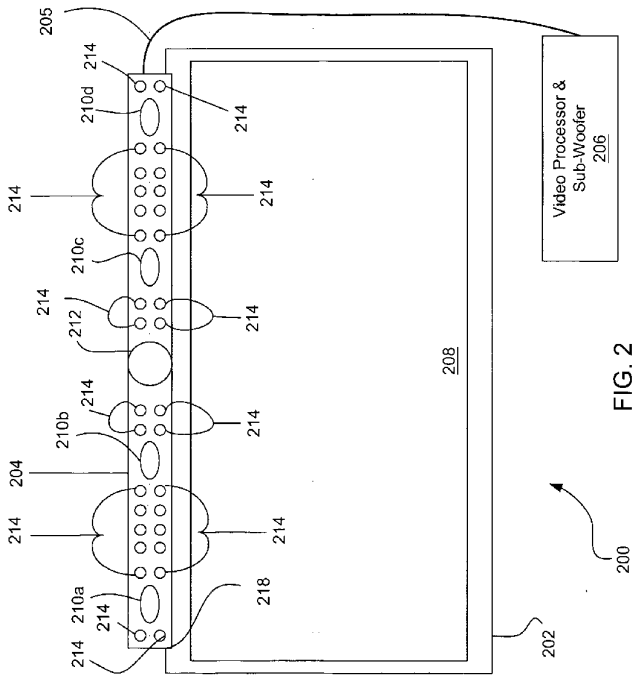

205
214
210d
214
214
214
210c
214
212
214
208
214
210b
214
204
214
214
210a
214
214
218
202
200
Video Processor & Sub-Woofer 206

FIG. 2

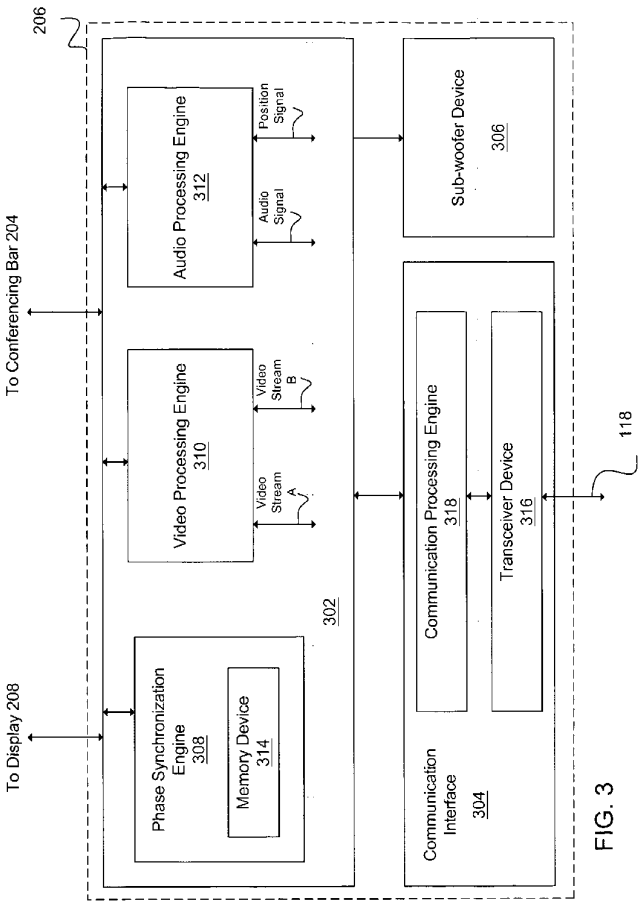

206
To Conferencing Bar 204
To Display 208
Audio Processing Engine
312
Position Signal
Audio Signal
Video Processing Engine
310
Video Stream B
Video Stream A
Phase Synchronization Engine
308
Memory Device
314
302
Sub-woofer Device
306
Communication Processing Engine
318
Transceiver Device
316
Communication Interface
304
118

FIG. 3

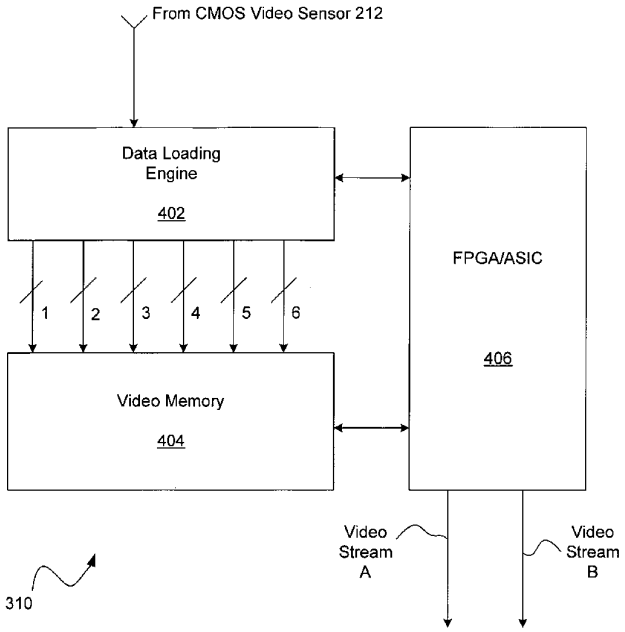

From CMOS Video Sensor 212
Data Loading Engine
402
FPGA/ASIC
406
1
2
3
4
5
6
Video Memory
404
Video Stream A
Video Stream B
310

FIG. 4

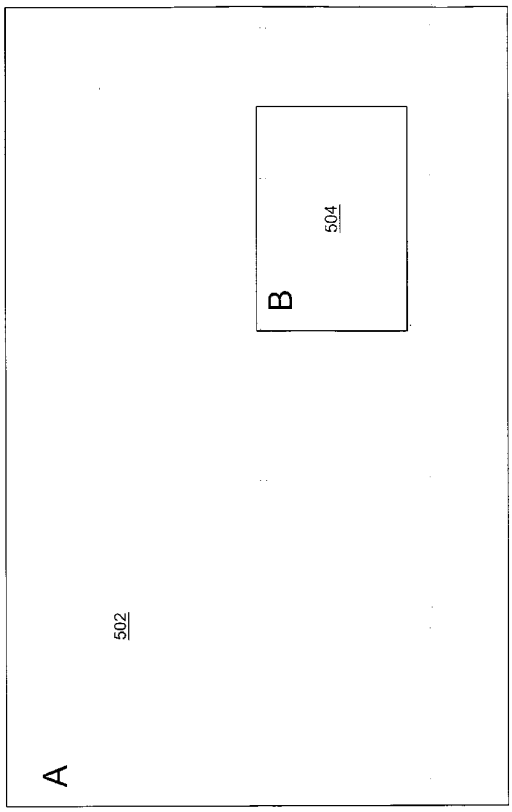


FIG. 5

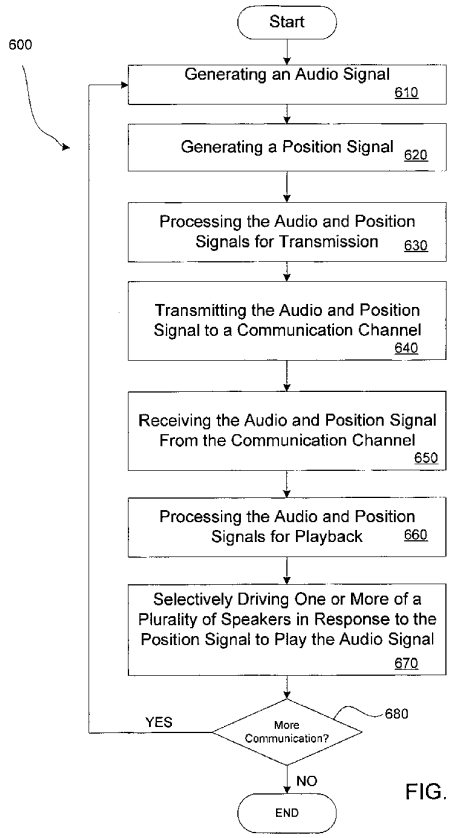

600
Start
Generating an Audio Signal 610
Generating a Position Signal 620
Processing the Audio and Position Signals for Transmission 630
Transmitting the Audio and Position Signal to a Communication Channel 640
Receiving the Audio and Position Signal From the Communication Channel 650
Processing the Audio and Position Signals for Playback 660
Selectively Driving One or More of a Plurality of Speakers in Response to the Position Signal to Play the Audio Signal 670
680
More Communication?
YES
NO
END

FIG. 6

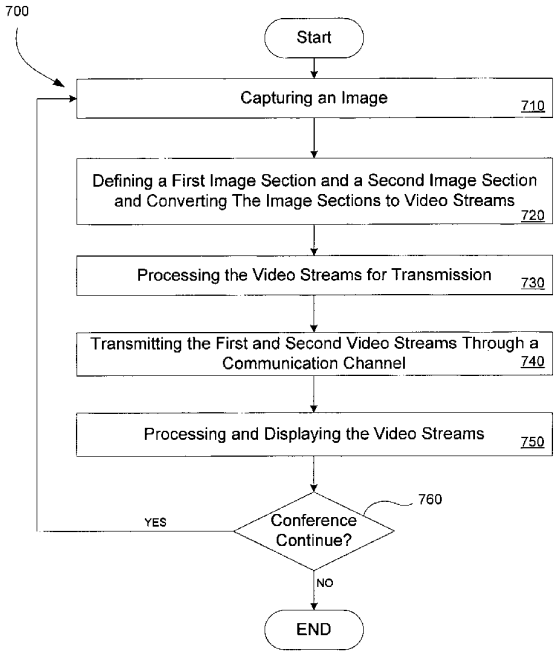

700
Start
Capturing an Image
710
Defining a First Image Section and a Second Image Section and Converting The Image Sections to Video Streams
720
Processing the Video Streams for Transmission
730
Transmitting the First and Second Video Streams Through a Communication Channel
740
Processing and Displaying the Video Streams
750
760
Conference Continue?
YES
NO
END

FIG. 7

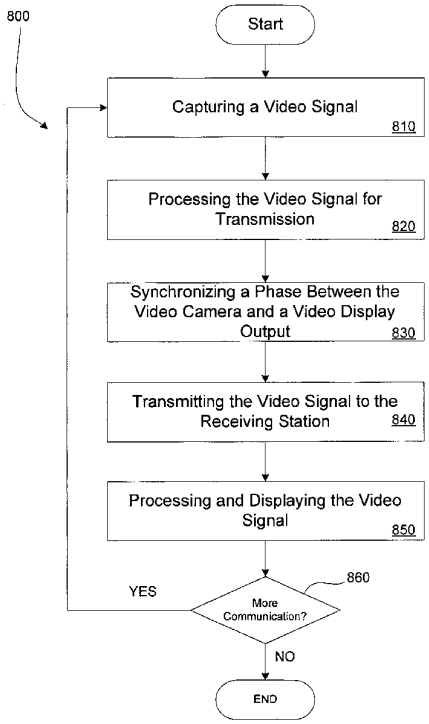

800
Start
Capturing a Video Signal
810
Processing the Video Signal for Transmission
820
Synchronizing a Phase Between the Video Camera and a Video Display Output
830
Transmitting the Video Signal to the Receiving Station
840
Processing and Displaying the Video Signal
850
860
YES
More Communication?
NO
END

FIG. 8

【手続補正書】
【提出日】平成15年7月21日(2003.7.21)
【手続補正１】
【補正対象書類名】特許請求の範囲
【補正対象項目名】全文
【補正方法】変更
【補正の内容】
【特許請求の範囲】
【請求項１】
映像ディスプレイを備えたローカルのテレビ会議装置と、ネットワークを通じて相互接続された映像ディスプレイを備えたリモートのテレビ会議装置とを有するテレビ会議システムのためのローカルのテレビ会議装置であって、
画像を取得するビデオセンサーと、
音声を取得する複数のマイクロフォンと、
音声を生成する複数のスピーカーと
を有するテレビ会議バーであって、前記ビデオセンサーと前記マイクロフォンと前記スピーカーがテレビ会議バーの固定位置に配置されたテレビ会議バーと、
前記テレビ会議バーにつなげられた処理ユニットと、
前記処理ユニットとネットワークを介して他のテレビ会議装置とにつなげられた通信インタフェースとを有し、
前記処理ユニットが、前記ビデオセンサーから受信された信号から少なくとも１つの映像ストリームと、前記マイクロフォンから受信された信号から音声ストリーム及び音源の位置信号とを作るように動作し、
前記処理ユニットが、リモートのテレビ会議装置から少なくとも１つの映像ストリームと１つの音声ストリームと１つの音源の位置信号とを受信するように動作し、
前記処理ユニットが、前記受信された音声ストリームと音源の位置情報に従って前記複数のスピーカーを駆動し、音声を再生するように動作するテレビ会議装置。
【請求項２】
請求項１に記載のテレビ会議装置であって、
前記ビデオセンサーが高解像度の映像ストリームを作るように動作し、
前記第１の映像ストリームが第１の解像度であり、
前記処理ユニットが、第２の映像ストリームを作るように動作し、
前記第２の映像ストリームが第２の解像度であり、前記第１の映像ストリームの中の領域を表すテレビ会議装置。
【請求項３】
請求項２に記載のテレビ会議装置であって、
前記第１の映像ストリームの第１の解像度が７００×４００ピクセルであり、
前記第２の映像ストリームの第２の解像度が３００×２００ピクセルであるテレビ会議装置。
【請求項４】
請求項２に記載のテレビ会議装置であって、
前記ビデオセンサーの最大の解像度が３０００×２０００ピクセルであるテレビ会議装置。
【請求項５】
請求項２に記載のテレビ会議装置であって、
前記第２の映像ストリームが話し中のテレビ会議の参加者の画像を表すテレビ会議装置。
【請求項６】
請求項５に記載のテレビ会議装置であって、
前記話し中のテレビ会議の参加者が変更すると、前記第２の映像ストリームが話し中のテレビ会議の参加者を追い、変更するテレビ会議装置。

【請求項７】
請求項１に記載のテレビ会議装置であって、
前記処理ユニットが、前記複数のマイクロフォンから受信された音声信号の大きさの違いに基づいて位置信号を生成するように動作するテレビ会議装置。
【請求項８】
請求項１に記載のテレビ会議装置であって、
前記処理ユニットが、前記ビデオセンサーからの信号のフェーズと、リモートの映像ディスプレイに表示するためのリモートのテレビ会議装置による映像ストリームの出力とを同期するように動作するテレビ会議装置。
【請求項９】
請求項１に記載のテレビ会議装置であって、
前記処理ユニットが、前記リモートのテレビ会議装置から受信された位置信号に応じて１つ以上のスピーカーを選択的に駆動し、前記少なくとも１つの映像ストリームの画像に対応する音声信号を再生することによって、前記受信された音声信号と音源の位置信号に従って前記複数のスピーカーを駆動し、音声を再生するように動作するテレビ会議装置。
【請求項１０】
請求項１に記載のテレビ会議装置であって、
前記複数のマイクロフォンがテレビ会議バーにおいてｎ方向の構成で配置されるテレビ会議装置。
【請求項１１】
請求項１に記載のテレビ会議装置であって、
テレビ会議バーが、その中身がテレビ会議の参加者に見えず、映像ディスプレイの上に置かれるように動作するように水平方向に囲まれたテレビ会議装置。
【請求項１２】
請求項１１に記載のテレビ会議装置であって、
複数のマイクロフォン及びスピーカーを有する２つの側面のバーを更に有し、
前記２つの側面のバーが垂直方向であり、映像ディスプレイの両側に配置されるように動作するテレビ会議装置。
【請求項１３】
請求項１に記載のテレビ会議装置であって、
前記ビデオセンサーが広い視野の角度を有するテレビ会議装置。
【請求項１４】
請求項１３に記載のテレビ会議装置であって、
前記広い視野の角度が６５度であるテレビ会議装置。
【請求項１５】
請求項１３に記載のテレビ会議装置であって、
前記ビデオセンサーの視界の角度を増加させるためにパン（ｐａｎ）モーターを更に有するテレビ会議装置。
【請求項１６】
テレビ会議のための方法であって、
複数のテレビ会議装置がネットワークを通じて相互接続され、
各テレビ会議装置が、ビデオセンサーと、複数のマイクロフォン及びスピーカーと、処理ユニットと、映像ディスプレイと、ネットワークインタフェースとを有するテレビ会議バーを有し、
前記テレビ会議バーのビデオセンサーでビデオ画像を取得し、
前記テレビ会議バーのマイクロフォンで音声信号を取得することを有する方法であって、
前記処理ユニットが前記ビデオ画像と前記音声信号を受信し、
前記ビデオ画像から第１の映像ストリームと、前記音声信号から音声ストリーム及び音声位置信号とを生成し、
前記第１の映像ストリームと音声ストリームと音声位置信号をリモートの会議装置に送信

し、
リモートの会議装置からリモートの映像ストリームとリモートの音声ストリームとリモートの音声位置信号を受信し、
映像ディスプレイに表示するために前記リモートの映像ストリームの出力を提供し、
前記スピーカーを駆動し、前記リモートの音声ストリームと前記リモートの音声位置信号に従って音声を再生する方法。
【請求項１７】
請求項１６に記載の方法であって、
前記ビデオ画像が高解像度であり、
前記第１の映像ストリームが第１の解像度である方法。
【請求項１８】
請求項１７に記載の方法であって、
前記処理ユニットが第２の映像ストリームを生成し、
前記第２の映像ストリームが第２の解像度であり、前記第１の映像ストリーム中の領域を表すことを更に有する方法。
【請求項１９】
請求項１８に記載の方法であって、
前記第２の映像ストリームが話し中のテレビ会議の参加者の画像を表す方法。
【請求項２０】
請求項１６に記載の方法であって、
前記音声位置信号が、前記複数のマイクロフォンから受信された音声信号の大きさの違いに基づいて生成される方法。
【請求項２１】
請求項１６に記載の方法であって、
前記処理ユニットが、前記ビデオセンサーからの信号のフェーズと、リモートの映像ディスプレイに表示するためのリモートのテレビ会議装置による映像ストリームの出力とを同期する方法。
【請求項２２】
請求項１６に記載の方法であって、
前記処理ユニットが、前記リモートのテレビ会議装置から受信された位置信号に応じて１つ以上のスピーカーを選択的に駆動し、前記少なくとも１つの映像ストリームの画像に対応する音声信号を再生することによって、前記受信された音声ストリームと音声位置信号に従って前記複数のスピーカーを駆動し、音声を再生する方法。

【国際調査報告】

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US02/25477

A. **CLASSIFICATION OF SUBJECT MATTER**
IPC(7) :H04N 7/14
US CL :348/14.08
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)
U.S. : 348/14.08, 14.01-14.07, 14.09; 370/260, 261; 709/204

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)
WEST

C. **DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	JP 2000-138913 (KONDO et al.) 16 MAY 2000, (entire document)	1,3-8,10,12-13,15-18 --- 2,9,11,14
Y	JP405276510A (MURAMASTU), 22 OCTOBER 1993, (fig. 1, see abstract)	2,14
Y	JP409140000A (MIZUSHIMA et al.) 27 MAY 1997, (fig. 2, see abstract)	9
Y	US 6,020,914 (DOUCHET), 01 FEBRUARY 2000, (fig. 1, col. 2 lines 26-36)	11

☐ 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 document 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	Date of mailing of the international search report
21 OCTOBER 2002	13 DEC 2002
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	Authorized officer MELUR RAMAKRISHNAIAH
Facsimile No. (703) 305-3230	Telephone No. (703) 305-4461

Form PCT/ISA/210 (second sheet) (July 1998)★

フロントページの続き

(81)指定国 AP(GH,GM,KE,LS,MW,MZ,SD,SL,SZ,TZ,UG,ZM,ZW),EA(AM,AZ,BY,KG,KZ,MD,RU,TJ,TM),EP(AT,BE,BG,CH,CY,CZ,DE,DK,EE,ES,FI,FR,GB,GR,IE,IT,LU,MC,NL,PT,SE,SK,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GQ,GW,ML,MR,NE,SN,TD,TG),AE,AG,AL,AM,AT,AU,AZ,BA,BB,BG,BR,BY,BZ,CA,CH,CN,CO,CR,CU,CZ,DE,DK,DM,DZ,EC,EE,ES,FI,GB,GD,GE,GH,GM,HR,HU,ID,IL,IN,IS,JP,KE,KG,KP,KR,KZ,LC,LK,LR,LS,LT,LU,LV,MA,MD,MG,MK,MN,MW,MX,MZ,NO,NZ,OM,PH,PL,PT,RO,RU,SD,SE,SG,SI,SK,SL,TJ,TM,TN,TR,TT,TZ,UA,UG,UZ,VC,VN,YU,ZA,ZM,ZW

(72)発明者 マロイ，クレイグ
アメリカ合衆国 ７８７３５ テキサス州 オースティン クラブ・リッジ・ドライヴ ８３０７
(72)発明者 ワシントン，リチャード
アメリカ合衆国 ７８７３５ テキサス州 マーブル・フォールズ シングルトン・ベンド・ロード ２４９１２ ７号
(72)発明者 チュ，ピーター
アメリカ合衆国 ０２４２０ マサチューセッツ州 レキシントン ハッドレイ・ロード ７

Ｆターム(参考) 5C064 AA02 AC04 AC07 AC13 AC16 AD06
5K015 AB01 JA00 JA10