

(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2010-529738

(P2010-529738A)

(43) 公表日 平成22年8月26日(2010.8.26)

(51) Int.Cl.		F I		テーマコード (参考)
<b>HO4N 7/15</b>	<b>(2006.01)</b>	HO4N 7/15	630Z	5C164
<b>HO4M 11/00</b>	<b>(2006.01)</b>	HO4M 11/00	302	5K201

審査請求 未請求 予備審査請求 未請求 (全 68 頁)

(21) 出願番号 特願2010-510342 (P2010-510342)  
 (86) (22) 出願日 平成20年5月29日 (2008.5.29)  
 (85) 翻訳文提出日 平成22年1月15日 (2010.1.15)  
 (86) 国際出願番号 PCT/US2008/006807  
 (87) 国際公開番号 W02008/153822  
 (87) 国際公開日 平成20年12月18日 (2008.12.18)  
 (31) 優先権主張番号 11/756,532  
 (32) 優先日 平成19年5月31日 (2007.5.31)  
 (33) 優先権主張国 米国 (US)

(71) 出願人 590000846  
 イーストマン コダック カンパニー  
 アメリカ合衆国 ニューヨーク州 ロチェ  
 スター ステート ストリート 343  
 (74) 代理人 100075258  
 弁理士 吉田 研二  
 (74) 代理人 100096976  
 弁理士 石田 純  
 (72) 発明者 カーツ アンドリュー フレデリック  
 アメリカ合衆国 ニューヨーク ロチェス  
 ター ステイト ストリート 343  
 (72) 発明者 ボーダー ジョン ノーボルド  
 アメリカ合衆国 ニューヨーク ロチェス  
 ター ステイト ストリート 343

最終頁に続く

(54) 【発明の名称】 居宅用動画通信システム

## (57) 【要約】

動画通信システム及びそれを稼働させる方法を提供する。本システム(290)の端末(300)は、画像表示装置(110)、所与の動画撮影セッティングに従いローカル環境及びそこにいる人物の動画を撮影する1個又は複数個の撮影装置(110, 120)、採音装置(323)及び鳴音装置(321)を含む音響システム、並びに1個又は複数個の場面に亘る動画をリモートサイト(364)に送信する動画通信イベントをコンテキスト伝達インタフェース(450)、プライバシー保護インタフェース(400)、イメージプロセッサ(320)及び通信コントローラ(355)と連携して実行するコンピュータ(310)を備える。インタフェース(450)は場面解析アルゴリズムに則り場面の変転を察知したとき撮影管理アルゴリズムに則り動画撮影セッティングを修正する。インタフェース(400)はプライバシー保護セッティングを提供してローカル環境動画の撮影、送信、表示又は録画を制御する。

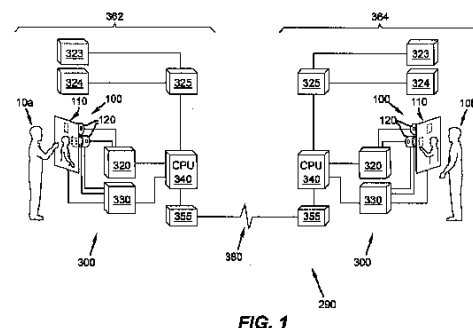


FIG. 1

**【特許請求の範囲】****【請求項 1】**

画像表示装置と、

ローカル環境及びそこにいる人物の動画像を動画撮影セッティングに従い撮影する 1 個又は複数個の撮影装置と、

採音装置及び鳴音装置を含む音響システムと、

1 個又は複数個の場面に亘る動画像を他サイトに送信する動画通信イベントをコンテキスト伝達インタフェース、プライバシー保護インタフェース、イメージプロセッサ及び通信コントローラとの連携で実行するコンピュータと、

を有する動画通信装置を備え、

10

上記コンテキスト伝達インタフェースが、場面解析アルゴリズムに則り場面の変転を察知する機能と、その変転に見合うよう撮影管理アルゴリズムに則り動画撮影セッティングを修正する機能と、を有し、

上記プライバシー保護インタフェースが、ローカル環境及びそこにいる人物の動画像の撮影、送信、表示又は録画動作に対しプライバシー保護セッティングを適用してその動作を管制する機能を有する動画通信システム。

**【請求項 2】**

請求項 1 記載の動画通信システムであって、ローカル環境及びそこにいる人物を捉えた動画像を、コンテキスト伝達インタフェース及びプライバシー保護インタフェースに従い相応に修正してから他サイトへと送信する動画通信システム。

20

**【請求項 3】**

請求項 2 記載の動画通信システムであって、その動画像修正を、撮影装置で使用される動画撮影セッティングの修正又はイメージプロセッサにおける画像処理により行う動画通信システム。

**【請求項 4】**

請求項 1 記載の動画通信システムであって、その場面解析アルゴリズムが、コンテキスト示唆キュー、場面解析ルール及び統計解析結果に基づき動画像を調べ、それにより場内変転又は場面間変転を察知するアルゴリズムである動画通信システム。

**【請求項 5】**

請求項 1 記載の動画通信システムであって、所与の動画撮影セッティングを様々な組み合わせた複数通りの動画撮影モードに従い 1 個又は複数個の撮影装置と連携稼働せうよう、またその動画撮影モードとして手動、自動、半自動、レビュー、デフォルト及びデジタルフォトフレームの諸モードのうちいずれかを使用しうよう、そのコンピュータが構成された動画通信システム。

30

**【請求項 6】**

請求項 1 記載の動画通信システムであって、その撮影装置として、狭い視野でローカル環境を撮影する第 1 撮影装置と、第 1 撮影装置の視野を包含する広い視野で第 1 撮影装置と同時にローカル環境を撮影する第 2 撮影装置と、を備える動画通信システム。

**【請求項 7】**

請求項 6 記載の動画通信システムであって、そのコンテキスト伝達インタフェースが、第 2 撮影装置の広い視野内を監視することによって、場面の変転を察知し又はプライバシー保護セッティングに抵触するコンテンツを認識する動画通信システム。

40

**【請求項 8】**

請求項 1 記載の動画通信システムであって、リモートサイトに送信されそこにいる人物に提供される動画像が、互いに別の撮影装置で撮影された二種類の動画像を示す複数の画像部分を含むコンポジット画像である動画通信システム。

**【請求項 9】**

請求項 8 記載の動画通信システムであって、リモートサイトに送信されそこにいる人物に提供される動画像が、撮影装置で撮影された画像を表す画像部分と、通信コントローラ経由で受信した画像を表す画像部分と、を含むコンポジット画像である動画通信システム

50

。

【請求項 1 0】

請求項 1 記載の動画通信システムであって、その画像表示装置が、ローカル環境を捉えた動画像と、リモートサイトから受信した動画像とを、プライバシー保護セッティングに従い同時に表示する動画通信システム。

【請求項 1 1】

請求項 1 記載の動画通信システムであって、そのコンピュータが、ローカル環境を捉えた動画像と、複数のリモートサイトから受信した複数の動画像とを同時に、或いはローカル環境内の互いに別々の個所を捉えた複数の動画像と、リモートサイトから受信した動画像とを同時に、画像表示装置によって表示させる動画通信システム。

10

【請求項 1 2】

請求項 1 記載の動画通信システムであって、その電子イメージング装置が 2 個の画面を有し、それらの画面に表示される画像が、互いに逆の方向を撮影した画像である動画通信システム。

【請求項 1 3】

請求項 1 記載の動画通信システムであって、その電子イメージング装置がローカル環境内に複数個ある動画通信システム。

【請求項 1 4】

請求項 1 2 記載の動画通信システムであって、それら複数個の電子イメージング装置が、通信イベント開催中に同時又は順繰りに使用される動画通信システム。

20

【請求項 1 5】

請求項 1 記載の動画通信システムであって、その場面解析アルゴリズムが、顔検知、音声認識、動き検知又は動き追跡の技術を利用するアルゴリズムである動画通信システム。

【請求項 1 6】

請求項 1 記載の動画通信システムであって、そのローカル環境内にいる人物をその顔又は音声の認識により認識する動画通信システム。

【請求項 1 7】

請求項 1 6 記載の動画通信システムであって、その電子イメージング装置が、認識された人物に見せるに相応しい静止画像又は動画像を表示する動画通信システム。

【請求項 1 8】

請求項 1 6 記載の動画通信システムであって、音声認識を用い話者音声を捉え、その結果を解釈して字幕を生成する動画通信システム。

30

【請求項 1 9】

請求項 1 記載の動画通信システムであって、その採音装置が指向性マイクロホンを少なくとも 1 個有する動画通信システム。

【請求項 2 0】

請求項 1 記載の動画通信システムであって、そのコンピュータが、撮影装置のうち光学的ズーム、パン又はティルト機能を有するものと連携してその機能を自動調整する動作を、その場面について察知された変転の内容、プライバシー保護セッティング、或いは動画像のコンテンツに従い実行する動画通信システム。

40

【請求項 2 1】

請求項 1 記載の動画通信システムであって、キーボード、タッチスクリーン、ジョイスティック、マウス又はその任意の組合せを含むユーザインタフェースを備える動画通信システム。

【請求項 2 2】

請求項 1 記載の動画通信システムであって、その音響システムが、ノイズキャンセル乃至ノイズリダクション機能、音源同定解析機能又はその組合せたる音響的拡張機能を有する動画通信システム。

【請求項 2 3】

請求項 1 記載の動画通信システムであって、静止画像を撮影する機能を有する動画通信

50

システム。

【請求項 24】

請求項 1 記載の動画通信システムであって、ローカル環境又はそこにいる人物の画像を解析する際に、それに先立ちそのローカル環境又は人物を撮影して取得した基準画像を、その解析の基準として使用する動画通信システム。

【請求項 25】

請求項 23 記載の動画通信システムであって、その基準画像がローカル環境を捉えた画像であり、個々の場面におけるローカル環境画像をその基準画像を手がかりにして修正する動画通信システム。

【請求項 26】

請求項 23 記載の動画通信システムであって、その基準画像がローカル環境にいる人物の外見を捉えた画像であり、後の場面におけるローカル環境内人物画像をその基準画像を手がかりにして修正する動画通信システム。

【請求項 27】

第 1 サイトに設けられた動画通信装置と、第 2 サイトに設けられた別の動画通信装置と、を備え、

個々の動画通信装置が、

a) 画像表示装置と、

b) ローカル環境及びそこにいる人物の動画像を所与の動画撮影セッティングに従い撮影する 1 個又は複数個の撮影装置と、

c) 稼働時にローカル環境にて音響を発生させる鳴音装置及び稼働時にローカル環境から音声信号を採取する採音装置を含む音響システムと、

d) 1 個又は複数個の場面に亘る動画像についての動画通信イベントをコンテキスト伝達インタフェース、プライバシー保護インタフェース及びイメージプロセッサと連携して実行するコンピュータと、

e) 通信ネットワークを介し第 1 サイトと第 2 サイトの間で動画像及び音声信号を共有させる通信コントローラと、

を有し、

上記コンテキスト伝達インタフェースが、場面の変転を場面解析アルゴリズムに則り察知する機能と、その変転に見合うよう撮影管理アルゴリズムに則り動画撮影セッティングの修正内容を決める機能と、を有し、

上記プライバシー保護インタフェースが、ローカル環境及びそこにいる人物の動画像を撮影、送信、表示又は録画する動作に対しプライバシー保護セッティングを適用しその動作を管制する機能を有する通信システム。

【請求項 28】

請求項 26 記載の通信システムであって、各動画通信装置が、自装置で決めたプライバシー保護セッティングの内容をプライバシー言語プロトコルに則り且つ通信ネットワークを介し他の動画通信装置に通知する一方、他の動画通信装置から通知された内容のプライバシー保護セッティングに従い自装置での動画像の撮影、処理、送信又は受信動作を管制する通信システム。

【請求項 29】

請求項 27 記載の通信システムであって、第 1 又は第 2 サイトにおける動画像の送信、表示又は録画動作を他サイトの動画通信装置から通知されたプライバシー保護セッティングに従い管制する通信システム。

【請求項 30】

請求項 26 記載の通信システムであって、その場面解析アルゴリズムが、コンテキスト示唆キュー、場面解析ルール及び統計解析結果に基づき動画像を調べ、それにより動画像内場面内での又は動画像内場面間の変転を察知するアルゴリズムである通信システム。

【請求項 31】

請求項 26 記載の通信システムであって、所与の動画撮影セッティングを様々に組み合

10

20

30

40

50

わせた複数通りの動画撮影モードに従い稼働するよう、またその動画撮影モードとして手動、自動、レビュー及びデフォルトの諸モードのうちいずれかを使用しうるよう、個々の動画通信装置が構成された通信システム。

【請求項 3 2】

ローカルサイトにいる第 1 の人物とリモートサイトにいる第 2 の人物との間で動画通信を行う方法であって、

ローカル及びリモートの各サイトで通信イベント毎にプライバシー保護セッティングを制定するステップと、

所与の動画撮影セッティングに従いローカルサイトで動画像を撮影するステップと、

そのローカルサイトで音声信号を採取するステップと、

ローカルサイトでの動画像の撮影と並行しその動画像における場面の変転を察知するステップと、

察知された変転に順応するよう撮影管理アルゴリズムに則り修正内容を決めて動画撮影動作を然るべく修正するステップと、

撮影された動画像及び採取された音声信号のうち少なくとも一部分をローカルサイトからリモートサイトにいる第 2 の人物へと送信するステップと、

リモートサイトを捉えた動画像及び音声信号のうち少なくとも一部分をリモートサイトからローカルサイトへと送信するステップと、

個々のサイトにおける動画像コンテンツの撮影、送信、表示又は録画動作をローカル及びリモートの各サイトで制定されたプライバシー保護セッティングに対し順応させるステップと、

を有する方法。

【請求項 3 3】

請求項 3 2 記載の方法であって、場面の変転を察知するステップが、コンテキスト示唆キュー、場面解析ルール及び統計解析結果に基づき動画像を調べて場面内変転又は場面間変転を察知するステップを含む方法。

【請求項 3 4】

請求項 3 2 記載の方法であって、所与の動画撮影セッティングを様々に組み合わせた複数通りの動画撮影モードのうち、手動、自動、レビュー及びデフォルトのいずれかのモードに従い稼働する方法。

【請求項 3 5】

請求項 3 2 記載の方法であって、個々のサイトにいる人物からの指示に従い且つ通信イベント別カスタム版プライバシー保護セッティング、デフォルト版プライバシー保護セッティング又はその組合せを用い、プライバシー保護セッティングを制定する方法。

【請求項 3 6】

請求項 3 2 記載の方法であって、ある狭い撮影視野で人物の動きを追跡しその動きによって示唆される変転に順応するよう動画撮影動作を修正する方法。

【請求項 3 7】

請求項 3 2 記載の方法であって、ローカルサイトにいる人物の動きを検知した結果に基づきリモートサイトで動画撮影セッティングの修正内容を決める方法。

【請求項 3 8】

請求項 3 7 記載の方法であって、リモートサイトにおける動画撮影動作をローカルサイトにいる人物が案内するステップを有し、その案内の手段として、ローカルサイトにいる人物を対象にした視線追跡、ジェスチャ追跡又はポインタ利用型ジェスチャ追跡を使用する方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、ローカルサイト・リモートサイト間でユーザが動画通信を行える通信システム、特に状況やユーザの多様さ、使い勝手、プライバシー上の懸念、システムコスト等の諸

10

20

30

40

50

要因が複雑に絡み合った環境である居宅にて好適に利用可能なシステムに関する。

【背景技術】

【0002】

テレビ電話(video-telephone)の概念は古くから知られている。その嚆矢は連載小説たる非特許文献14に記載のものであるが、実用的なテレビ電話システムとしては1964年のニューヨーク万国博覧会にベル研究所が出品したものが最初のものである。AT&TはこのシステムをPicturephoneなる商標の下に様々な形態で商品化したが、このシリーズは商業的な成功を見ることができなかった。それは、解像度が低い、カラー画像を扱えない、音声と動画のタイミングが合っていない等、その性能に関わる重大な技術的問題があって魅力に欠けたからである。更に、Picturephoneの撮影視野は非常に狭く、大抵は話者の顔周りしか捉えることができなかった。このことをよく示しているのは特許文献11(発明者:W. Rea)の記述、特にPicturephone用カメラの狭い撮影視野内に話者を正しく捉える手段についての記述である。撮影視野が狭いということは、背景についての情報がほとんど或いは全く得られず、コンテキスト(状況・経緯)が伝わらないということである。加えて、Picturephoneに備わっている話者プライバシー保護機能は、動画の送信を遮断するオプションだけであった。

10

【0003】

現代世界では既に様々な技術による双方向動画通信が可能となっている。なかでも広く用いられているのはカメラ付携帯電話である。昨今の携帯電話では、静止画像撮影用カメラが備わっていないものの方が珍しく、一部の携帯電話はライブ動画撮影表示機能も備えている。Fotonation Ltd.(アイルランド)のように、顔検知、顔認識、顔追跡等を含めユーザエクスペリエンス(体験的満足性)の良質化につながるライブビデオカメラ付携帯電話向け新技術を提供している企業もある。特許文献6(発明者:J.H. Lo)にもカメラ及び顔認識機能を有する携帯電話が記載されている。このように、カメラ付携帯電話はその使い勝手と可搬性が良好なため現代生活において疑いなく必須のものとなっているが、サイズ上及び価格上の制約が厳しいためその用途が限られている。

20

【0004】

これに似た機能を有する別種の装置としてはウェブカメラがある。この種のカメらはラップトップコンピュータ等のコンピュータに実装可能であり、パン、ティルト、ズーム等の機能を備えているものが多い。また、Creative Laboratories(シンガポール)、Logitech(スイス)等といった企業は、現在、ウェブカメラとして使用かのような高性能カメラをコンピュータアクセサリとして世に送り出している。それらには高度な音声信号採取機能、動き検知機能、顔追跡機能等といった高付加価値的諸機能が備わっている。特許文献10(発明者:McAlpine et al.)にも、ウェブカメラの機械的なパン、ティルト及びズーム動作でユーザの顔を追跡するシステム及び方法が記載されている。

30

【0005】

Apple Inc.(米国カリフォルニア州クペルティーノ)では、このウェブカメラの改良版に当たるiSight(登録商標;以下表記省略)、iChat(登録商標;以下表記省略)等の製品乃至サービスを提供している。iSightは、ラップトップコンピュータに装着・実装可能なカメラと、白バランス・シャープネス・色・合焦状態・露出の自動調整やノイズのフィルタリングによってその明るさ・合焦状態・発色が良好な画像を生成するオンボード画像処理手段と、からなる製品である。iChatは、iSightで生成された画像を使用し最大10人までの人々の間で1対1チャット、マルチウェイチャット、音声チャット等を行えるサービスである。ただ、こうしてビデオカメラとコンピュータを結合させてインターネット経由テレビ電話システムを構築しても、携帯電話ほどユビキタスなシステムにはなりえない。一段と値が張ることや装置が大規模になることは無論のこと、ウェブカメラで得られるユーザエクスペリエンスに多くの問題がありまだ適切に解決されていないことも、その原因となっている。とりわけ、このシステムは家庭等のようにめまぐるしく変化する環境に十分馴染んでいない。ウェブカメラを利用した

40

50

システムをその種の環境に順応させより簡便に使えるようにするには、ユーザインタフェース、画像撮影、プライバシー保護等の面での技術的改良が必要となる。

【0006】

特記すべきことに、WebEx Communications (米国カリフォルニア州サンタクララ)では、ウェブカメラ技術を応用した安価なウェブベーステレビ会議(video-conferencing)であるWebEx (登録商標; 以下表記省略)を提供している。これはミーティング、トレーニングセッション、Webinar (登録商標)、カスタマサポート等の業務で利用されるサービスであり、リアルタイム通信向けに特化した専用のウェブベースグローバルネットワーク上でアプリケーションを頒布する仕組みを採っている。このサービスでは、セキュリティを多層的に提供することで、会議出席者によるアクセスや文書保存・印刷権限を管理し、またデスクトッププライバシーを保護している。ネットワークセキュリティに関わる機能としては、認証機能、会議内容/文書暗号化機能、侵入遮断機能、データ保持回避機能(WebExサーバにデータを保存しない機能)等を提供している。特許文献29(発明者:Zhu et al.)にも、セキュア通信システム向けコラボレイティブコンピューティング方式が記載されている。こうしたWebEx方式は有益なものではあるが、ユーザが抱く様々な不安に配慮していないため、個人ベースでの動画通信に使用するには障りがある。

【0007】

これら、カメラ付携帯電話やウェブカメラに代わる装置としては更に高性能ビデオホン(enhanced video-phone)がある。これはズーム機能及び追跡機能を有する多機能カメラや高性能な音響装置を備える大画面の装置であり、複数のユーザが同時に使用可能で良質なユーザエクスペリエンスをもたらしてくれるものである。但し、この装置を設置可能な環境は、家庭、事務所、学校等のように、撮影・表示機能が良質であれば可搬性が多少劣っていても許される環境に限られている。その基本的なシステム構成は、複数のサイト(場所)間で情報を転送できるようカメラ及びテレビジョン受像機を電話回線又はインターネットに接続する、というものであり、特許文献8(発明者:AbbiEzzi et al.)に記載の一般家庭用テレビ会議システムでもそうしたシステム構成を採用している。このシステムは、居宅用遠隔会議システムに相応しい基本的な撮影素子及び表示素子が備わっているので、ウェブカメラと同じく広い視野に亘り画像を撮影して表示することができ、その原型たるPicturephoneに比べて良好にコンテキストを伝えることができる。しかしながら、このシステムにも、居宅用テレビ電話システムにまつわる種々の懸案、特に居宅のようにめまぐるしく変化する環境における周囲状況管理、プライバシー保護等に関する懸案への対策はあまり盛り込まれていない。

【0008】

より高性能なシステムとしては特許文献41(発明者:N. Chim)に記載の高性能遠隔会議システム(enhanced teleconferencing system)がある。これは居宅での使用を想定したシステムであり、マイクロホンが複数個あるので音声信号を利用した被写体追跡を好適に実行することができる。このシステムでは、更に、そのディスプレイの背後にカメラを配することで、アイコンタクト(目配せ・目交い)上のユーザエクスペリエンスを高めている。とりわけ、このシステムで実行される被写体追跡手順、即ち複数個のマイクロホンをを用いローカル環境における人物の所在場所を特定してその人物を追跡する手順は秀逸である。この手順では、オーディオプロセッサで発声させる音声追跡信号に従い、人物を追いかけるようにカメラを駆動する。更に、その人物をフレーム内に捉え続けるためそのカメラの撮影視野を最適化する。この最適化は、機械的なパン、ティルト及びズーム動作と、画像処理によるクロッピング(裁ち切り)との併用で実行される。また、そのカメラはディスプレイの後に隠れているので、ユーザの視線をカメラで真正面から捉えてアイコンタクトを成立させることができる。即ち、そのディスプレイを見ている人物のダイレクトオン画像(画面正面からの画像)を捉え、リモートサイトに送って他の人達に向けて表示させることができる。ただ、居宅環境でも使用可能とはされているが、そのプライバシー保護やコンテキスト伝達の管理面が充実していないことからすると、この文献に記載のシス

テムの実用環境は企業内会議室環境となろう。

【0009】

同じく高性能な動画通信システムとしては、テレビ会議端末を用いた高性能テレビ電話システム(enhanced video-telephony)が主として主に企業向けに市販されている。例えば Cisco Systems (米国カリフォルニア州サンノゼ)、Digital Video Enterprises (米国カリフォルニア州アーヴィン)、Destiny Conferencing (米国オハイオ州デイトン)、Teleris (英国ロンドン)等といった企業は、企業経営者向けに高性能テレビ会議端末を商品化している。それらのうち一部は、既に特許文献15及び19(共に発明者:Allen et al.)や特許文献20及び25(共に発明者:McNelley et al.)といったテレビ会議関連特許を取得している。これらの企業の商品では、画像及び音声の忠実性、環境美学及びエルゴノミクスへの配慮、アイコンタクトに配慮した撮影表示能力、ネットワーク上の大規模データストリームをシームレス且つ安全に扱う能力等が強化されている。そして、そのカメラが画面、ビームスプリッタ等の背後に隠れているので、カメラが目障りにならないようにしつつ、アイコンタクトを好適に成立させることができる。

【0010】

ただ、これらの遠隔会議乃至テレビ電話システムは、その居場所が異なる複数のユーザー間で会議を行うには便利であるが、管理の行き届いた環境での利用を前提としているので、個人の居宅や学校のようにめまぐるしい変化が珍しくない環境には不向きである。特に、これらのシステムは、中央テーブルを備えた一般的な会議室、議場風の座席を備えた高級な部屋等が既にあること又はそれを新設することを想定している。即ち、出席者同士が互いに顔見知りで、その所作(アクション)も割合に型にはまっている“きちんとした”環境を想定している。そうした環境向けのシステムは、個人の居宅のように変化が激しい環境には不向きである。更に、これらのシステムの端末では、ローカルユーザの画像から背景部分を削除してコンテキストを隠すことができるように構成されている。リモートサイトで別の出席者がその画像を見たとしても、背景が当該リモートサイトのそれ又は仮想的なそれに差し代わっているので、ローカルユーザ側のコンテキストはリモートサイトに伝わらない。また、WebEx方式での捉え方と同じく、プライバシー保護やセキュリティ提供の課題がネットワークを介したアクセス及びデータ転送可能性の問題として捉えられているので、例えば特許文献2(発明者:Le Pennec)に多ノード間セキュアテレビ会議確立方式として記載されている通り、リンクユニーク暗号鍵、暗号鍵間接続用セキュアインタフェース及びリンクユニーク暗号鍵保持用セキュアデータアーカイブを含む少なくとも三種類の暗号化デバイスが必要になる。加えて、これらのテレビ会議システムの価格は10万米ドルを上回っており、居宅向け市場で許容される水準ではない。

【0011】

留意すべきことに、そうした高性能テレビ会議システムのなかにも、複数の人物間で交わされる会話の無秩序さ想定しているものがある。とりわけ、特許文献24(発明者:Strubbe et al.)、28(発明者:Gutta et al.)、33(発明者:Strubbe et al.)等、いずれもPhilips Electronics(蘭国アイントホーフェン)を譲受人とする一連の特許で提案された動的環境下遠隔会議方式によるものである。それらのうち特許文献24に記載のテレビ会議システムはビデオロケータ及びオーディオロケータを備えている。システム稼働時には、それらのロケータ(測位器)の出力に基づき各出席者の居所を特定すること、ひいては話者に焦点を合わせることやその話者のクローズアップ画像を送信することができる。また、その話者の発言が所定時間以上に亘り続いたときや、その出席者が発言を止めてから所定時間が経過したときには、システム側でカメラの設定を調整し、発言中でない別の出席者を順繰りに表示させる。或いは、システム側でカメラを相応量ズームアウトさせ、全出席者を一覧表示させる。更に、テレビ会議セッションの最中で新たな人物が参加したときや参加中の人物が退出するときに、その人物を撮影することもできる。ただ、このシステムでは、同文献の図2に記載の通り、出席者達があたかも一つのテーブルを囲んで着座しているかのような形式でテレビ会議が行われる。そのた



め、このシステムを好適に適用できるのは、（半ば）公式的な企業内テレビ会合のように、その地位に大差がない出席者達が互いにある程度の礼儀作法を弁えて参加する場に絞られよう。発言中でない出席者を順繰りに撮影しその画像を送信するというやり方は、そうした状況想定の下でこそ採用可能且つ妥当なものであると認められる。

【 0 0 1 2 】

次に、特許文献 2 8 及び 3 3 に記載の発明は、基本的には特許文献 2 4 に記載の発明を改良し、テレビ会議のイベントをより自然に行わせるための順応手段を付加したものである。そのうち特許文献 3 3 に記載のシステムでは、カメラ、オーディオロケータ及びビデオロケータの出力に対して一群の経験則を適用することで、それまで話者であった人物を追い続けるべきか、それとも新たな話者に切り替えるべきかを、判別するようにしている。具体的には、話者間時間差、5 度コロケーションしきい値等のファクタを計測し、信頼度推定値に照らして評価することにより、別の人物に切り替えるべきかそれとも広視野撮影に切り替えるべきかをシステム側で判別している。また、特許文献 2 8 では動的環境下テレビ会議方式を更に概念的に拡張している。即ち、出席者の行動に現れているキュー（手がかり）を幾通りか認識し、それらのキューを解析して予測することで、それまでの話者から別の話者へのシームレスな動画撮影対象切替を試みている。行動に現れるキューとは、例えば抑揚パターン・音調・ラウドネス等の音響的キューや、視線・顔向け・身振り・手振り・表情等の画像的キューや、その組合せのことである。それらのキューは一般にイベント毎に異なるので上掲のように利用することができる。ただ、両文献の図 1 に記載の通り、そもそも特許文献 2 8 及び 3 3 でテレビ会議の円滑開催に相応しい場として想定しているのは、会議室の中や議場的座席配置を採る部屋の中、即ち出席者の移動や交代がほとんど生じない場である。更に、これらのシステムを好適に適用できるのは、（半ば）公式的な企業内テレビ会合のように、その地位に大差がない出席者達が互いにある程度の礼儀作法を弁えて参加する場に絞られよう。特許文献 2 8 ではその適用可能範囲を拡げており、第 1 1 カラムの表に記載の通り子供がいる状況での使用を多少は想定しているけれども、特許文献 2 4、2 8 及び 3 3 で提案されているシステムは、やはり、居宅環境に照準したシステムであるとは言い得ない。即ち、居宅における変化のめまぐるしさに順応させるには不十分で、プライバシー保護や周囲状況管理の側面も十分に行き届いていない。

【 0 0 1 3 】

特許文献 4（発明者：Girish et al.、名称：テレビ会議装置及び方法 (Video Conferencing Apparatus and Method)）に記載のテレビ会議システムでは、出席者が自分のプライバシーを守れるよう幾つかの策が講じられている。具体的には、動画撮影や音声信号採取が実行されている旨の通知をハードウェア的光インジケータから受けることができ、音声信号をミュートすることができ、またカメラ前置機械絞り（及びそのキャップ）を見ることで動画撮影が行われていないことを確認することができる。即ち、このシステムは、動画通信イベント開催中に動画像が不用意に伝送されるのを防ぐこと、特にその場にいるユーザが知らないうちに確立済ネットワークリンクを介し画像が伝送されることを、特に重視したシステムであるといえる。ただ、ユーザが自分の環境を管理し自分や家族のプライバシーを保護する上で、この文献に記載のやり方が万能なわけでは決していない。このシステムは、更に、居宅環境で有用となるであろうコンテキスト解釈及び制御の手段を欠いている。

【 0 0 1 4 】

更に、遠隔会議システムや高性能動画通信システムのなかには、会議室環境向けに限らず、オフィス環境や実験室環境における従業員間交流向けに開発されたものもある。その先駆例は、1980 年代に Xerox Palo Alto Research Center（米国カリフォルニア州パロアルト）で開発された Media Space（登録商標）である。これは、オフィス間常時リアルタイム音声付動画通信を行うシステムであった。その類例たる Video Window（商標）は、非特許文献 1 2 に記載の通り大画面を有する全二重遠隔会議システムであり、専門職従業者間非公式コラボレイティブ通信を助ける役割を担っていた。これらのシステムは、会議室向けのものより非公式な通信に

適しているが、居宅環境における個人使用ではなく業務使用向けに開発されたものであるため、居宅独特の問題はその想定外となっている。

【 0 0 1 5 】

原初的な家庭用メディア空間としては、在宅勤務者・出社勤務者間で通信を行えるようにするものも既に開発されている。例えば非特許文献 3 には常時稼働型家庭用メディア空間が記載されている。この文献の筆者は、プライバシーに関し人々が抱く不安を解消することが家庭用メディア空間では業務用メディア空間よりもかなり強く求められるとの認識を示し、在宅ユーザが常時稼働の事実を忘れてたり不用意にカメラの視野内をうろついたりするとプライバシー侵害的状況が発生することを指摘した上で、そうしたリスクを軽減する手法を幾つか提案している。そのなかには、職住分離、人数カウント、身体利用制御、ジェスチャ認識、可視可聴フィードバック等の機構が含まれている。ただ、家庭内に配置されはしても、このシステムは住人による個人的な通信での利用を想定していない。即ち、一人又は複数人のユーザが行う個人的なアクションに順応させることができ、またそれらのユーザが自分のプライバシーを守ることができる居宅用通信システムにはなっていない。

【 先行技術文献 】

【 特許文献 】

【 0 0 1 6 】

- 【 特許文献 1 】 米国特許出願公開第 2 0 0 4 / 0 1 4 0 9 7 3 号明細書
- 【 特許文献 2 】 米国特許出願公開第 2 0 0 4 / 0 1 5 0 7 1 2 号明細書
- 【 特許文献 3 】 米国特許出願公開第 2 0 0 4 / 0 1 9 6 3 6 0 号明細書
- 【 特許文献 4 】 米国特許出願公開第 2 0 0 4 / 0 2 5 7 4 3 1 号明細書
- 【 特許文献 5 】 米国特許出願公開第 2 0 0 5 / 0 0 2 4 4 8 9 号明細書
- 【 特許文献 6 】 米国特許出願公開第 2 0 0 5 / 0 0 4 1 8 4 0 号明細書
- 【 特許文献 7 】 米国特許出願公開第 2 0 0 5 / 0 1 2 8 3 3 2 号明細書
- 【 特許文献 8 】 米国特許出願公開第 2 0 0 5 / 0 1 4 6 5 9 8 号明細書
- 【 特許文献 9 】 米国特許出願公開第 2 0 0 6 / 0 0 0 7 2 2 2 号明細書
- 【 特許文献 1 0 】 米国特許出願公開第 2 0 0 6 / 0 0 7 5 4 4 8 号明細書
- 【 特許文献 1 1 】 米国特許第 3 4 9 5 9 0 8 号明細書
- 【 特許文献 1 2 】 米国特許第 4 9 2 8 3 0 1 号明細書
- 【 特許文献 1 3 】 米国特許第 4 9 8 7 4 9 2 号明細書
- 【 特許文献 1 4 】 米国特許第 5 1 9 4 9 5 5 号明細書
- 【 特許文献 1 5 】 米国特許第 5 5 7 2 2 4 8 号明細書
- 【 特許文献 1 6 】 米国特許第 5 6 3 9 1 5 1 号明細書
- 【 特許文献 1 7 】 米国特許第 5 7 7 7 6 6 5 号明細書
- 【 特許文献 1 8 】 米国特許第 5 9 9 9 2 0 7 号明細書
- 【 特許文献 1 9 】 米国特許第 6 1 6 0 5 7 3 号明細書
- 【 特許文献 2 0 】 米国特許第 6 2 4 3 1 3 0 号明細書
- 【 特許文献 2 1 】 米国特許第 6 2 7 5 2 5 1 号明細書
- 【 特許文献 2 2 】 米国特許第 6 4 0 0 9 0 3 号明細書
- 【 特許文献 2 3 】 米国特許第 6 4 5 4 4 1 4 号明細書
- 【 特許文献 2 4 】 米国特許第 6 6 1 1 2 8 1 号明細書
- 【 特許文献 2 5 】 米国特許第 6 7 1 0 7 9 7 号明細書
- 【 特許文献 2 6 】 米国特許第 6 7 7 1 3 0 3 号明細書
- 【 特許文献 2 7 】 米国特許第 6 8 8 8 5 6 2 号明細書
- 【 特許文献 2 8 】 米国特許第 6 8 9 4 7 1 4 号明細書
- 【 特許文献 2 9 】 米国特許第 6 9 0 1 4 4 8 号明細書
- 【 特許文献 3 0 】 米国特許第 7 0 1 5 5 9 4 号明細書
- 【 特許文献 3 1 】 米国特許第 7 0 3 4 8 6 6 号明細書
- 【 特許文献 3 2 】 米国特許第 7 0 4 8 3 8 6 号明細書
- 【 特許文献 3 3 】 米国特許第 6 8 5 0 2 6 5 号明細書

10

20

30

40

50

【特許文献 3 4】米国特許第 7 0 5 8 2 0 9 号明細書  
 【特許文献 3 5】米国特許第 6 3 7 3 6 4 2 号明細書  
 【特許文献 3 6】米国特許第 6 4 2 4 3 7 7 号明細書  
 【特許文献 3 7】米国特許第 7 0 4 2 4 8 6 号明細書  
 【特許文献 3 8】米国特許第 6 7 2 4 4 1 7 号明細書  
 【特許文献 3 9】米国特許第 6 8 0 6 8 9 8 号明細書  
 【特許文献 4 0】米国特許第 6 9 1 9 8 9 2 号明細書  
 【特許文献 4 1】米国特許第 6 2 7 5 2 5 8 号明細書  
 【特許文献 4 2】米国特許第 5 6 7 5 3 7 6 号明細書  
 【非特許文献】

10

【0 0 1 7】

【非特許文献 1】Debby Hindus et al., "Casablanca: Designing Social Communication Devices for the Home", Interval Research Corporation, Proceedings SIGCHI'01, March 31-April 4, 2001, ACM 1-58113-327-8/01/0003, pp.325-332

【非特許文献 2】Cisco Systems, "Cisco Teleconference 3000", 2006, pp.1-5, [online] Internet URL: <http://www.cisco.com/go/telepresence>

【非特許文献 3】Carman Neustaedter and Saul Greenberg, "The Design of a Context-Aware Home Media Space for Balancing Privacy and Awareness", University of Calgary, Proceedings of the Fifth International Conference on Ubiquitous Computing (UbiComp2003), pp.1-18

20

【非特許文献 4】Digital Video Enterprises, Inc., "Telepresence Products", 2006, pp.1-6, [online] Internet URL: <http://www.dvetelepresence.com>

【非特許文献 5】Claudio Pinhanez, "The Everywhere Displays Projector: A Device to Create Ubiquitous Graphical Interfaces", IBM Thomas Watson Research Center, Proceedings of Ubiquitous Computing 2001, pp.1-18

【非特許文献 6】Emmanuel Munguia Tapia, Stephen S. Intille, John Rebola, and Steve Stoddard, "Concept and Partial Prototype Video: Ubiquitous Video Communication with the Perception of Eye Contact", Massachusetts Institute of Technology, Proceedings of UBICOMP 2003, pp.1-2

【非特許文献 7】Ken Go et al., "Familyware: Communicating with Someone You Love", Yamanashi University, Japan, Proceedings of the IFIP HOIT Conference, HOIT 2000, pp.1-10

30

【非特許文献 8】Apple Inc., "iChat AV Videoconferencing", March 2007, pp.1-4, [online] Internet URL: <http://www.apple.com>

【非特許文献 9】Xavier L.C. Brolly, Constantinos Stratelos and Jeffrey B. Mulligan, "Model-Based Head Pose Estimation for Air-Traffic Controllers", NASA Ames Research Center, California, published at ICIP 2003, the IEEE Conference on Image Processing, pp.1-4

【非特許文献 1 0】Polycom Inc., "Polycom RealPresence Experience High Definition (RPTTMHD)", 2007, pp.1-6

40

【非特許文献 1 1】Thomas Vetter, "Synthesis of Novel Views from a Single Face Image", Max-Planck-Institut, Germany, Technical Report No.26 (February 1996), [online] Internet URL: <http://www.mpik-tueb.mpg.de/projects.techreport.list.html>

【非特許文献 1 2】Robert S. Fish, Robert E. Kraut, and Barbara L. Chalfonte, "The VideoWindow System in Informal Communications", Proceedings of the 1990 ACM conference on Computer-Supported Cooperative Work, pp.1-11, October 1990

【非特許文献 1 3】WebEx Communications Inc., "WebEx Capabilities", 2005, pp.1-16, [online] Internet URL: <http://www.webex.com>

【非特許文献 1 4】"Tom Swift and His Photo Telephone" (1914)

【発明の概要】

50

## 【発明が解決しようとする課題】

## 【0018】

従って、今求められているのは、従来のシステムでは想定されていなかった需要に応え、遠隔地にいる一人又は複数人の人物との親しいテレビ会議乃至動画通信を実現するのに普く役立つ居宅用のシステムを実現することである。そのシステムには、居宅で発生する様々な状況に順応しつつほどほどシームレスに機能すること、少なくとも画像の撮影、録画及び送信に関しユーザが自分達のプライバシーを管理及び保全できるようにすること、ユーザ及びその環境に関わるコンテキストの伝達を管理して良質な通信体験を可能にすること等が求められよう。

## 【課題を解決するための手段】

## 【0019】

本願で提供するのとは動画通信システム及び方法である。本システムは、動画像を表示可能な画像表示装置と、所与の動画撮影セッティング(video capture settings)に従いローカル環境及びそこにいる人物の動画像を撮影する1個又は複数個の撮影装置と、稼働時にローカル環境にて音響を発生させる鳴音装置及び稼働時にローカル環境から音声信号を採取する採音装置を含み音響セッティング(audio settings)に従い稼働する音響システムと、それら撮影装置(群)及び音響システムのほかイメージプロセッサ及び通信コントローラと連携して動画通信イベントを実行するコンピュータと、を備える。その動画通信イベントでは、撮影済動画像に基づき撮影を行い、また動画像をリモートユーザへと送信する。

## 【0020】

そのコンピュータは更にコンテキスト伝達インタフェース(contextual interface)を提供する。コンテキスト伝達インタフェースでは、動画通信イベント開催中にその動画像に場面内変転(intra-scene transition)が生じたとき、そのことを場面内解析(intra-scene analysis)アルゴリズムに則り察知し、察知した変転に応じ且つ場面撮影管理(scene capture management)アルゴリズムに則り、動画撮影セッティングに施すべき修正の内容を特定する。

## 【0021】

コンテキスト伝達インタフェースでは、更に、人物のアクションに場面間変転(inter-scene transition)を意味すると見られる変化が生じたとき、そのことを場面間解析(inter-scene analysis)アルゴリズムに則った変転テスト(transition test)により察知し、察知した場面間変転に応じ且つ変転処理(transition process)機構を用い、動画撮影セッティングに施すべき調整の内容を特定する。なお、動画像又は音声信号に場面内変転が生じると、その動画通信イベントの通信インパクトが低下することがある。動画像又は音声信号に場面間変転が生じるのは、その動画通信イベントの性格が変わった場合等である。

## 【図面の簡単な説明】

## 【0022】

【図1】ローカル端末をリモート端末にネットワーク経由で接続するシステムの全体構成図である。

【図2】図1に示した動画通信システムが稼働するローカル環境で生じるコンテキストの一例を示す図である。

【図3A】この動画通信システムの概略機能構成を示す図である。

【図3B】この動画通信システムのより詳細な機能構成の一例を示す図である。

【図4A】一群のカメラで撮影される広狭二種類の撮影視野を示す図である。

【図4B】それらの位置関係を示す図である。

【図4C】通信イベント開催中に生じる撮影視野変転の例を示す図である。

【図5A】基準画像の一例を示す図である。

【図5B】基準画像の別例を示す図である。

【図6】この動画通信システムの一構成部分たる電子イメージング装置に画面分割画像として表示されている二種類のリモート側場面を示す図である。

10

20

30

40

50

【図 7 A】種々ある手順の一例を示すフローチャートである。

【図 7 B】その詳細を示すフローチャートである。

【図 7 C】その詳細を示すフローチャートである。

【図 7 D】その詳細を示すフローチャートである。

【図 8】この動画通信システムが稼働するローカル環境コンテキストの別例を示す図である。

【図 9 A】この動画通信システムの実施に当たり利用可能な電子イメージング装置の別例構成、特に従来技術に属する装置を示す図である。

【図 9 B】この動画通信システムの実施に当たり利用可能な電子イメージング装置の別例構成を示す図である。

【図 10】リモートユーザ用ディスプレイの方向がこの動画通信システムで使用される電子イメージング装置に及ぼす影響を示す幾何学的位置関係を示す図である。

【図 11 A】ローカルサイトリモートサイト間のやりとりで動きに追従する動作を示す図である。

【図 11 B】その続きを示す図である。

【図 11 C】その続きを示す図である。

【図 12 A】電子イメージング装置上の画像をユーザの視線に関わる視線追跡キューに応じ変化させる動作を示す図である。

【図 12 B】その続きを示す図である。

【発明を実施するための形態】

【0023】

今日、テレビ会議やテレビ電話をその性質に基づき概括するのにテレプレゼンスなる表現が使用されている。これは、自分の身の回りの環境（ローカル環境）とは別の場所にある環境（リモート環境）に何らかの体験乃至印象をもたらす、という意味である。この表現がよく用いられるのは、別の場所にいるグループ間の協業を可能にする企業内遠隔会議について述べるときである。テレビ会議システムのなかには、リモートユーザの実物大画像を表示させてアイコンタクトを図ることや、その画像の背景をローカル環境のそれに置き換えて表示することができるものがあり、そうした機能を備えるテレビ会議システムは理想的なテレプレゼンスシステムであるといえよう。ただ、それを以てテレプレゼンスと称するのはある意味で不適切というものである。何故なら、カメラのズーム機能を初めとする種々のツールを使用し相応の制御を実行することで、ある意味で“現実以上”の人為的な体験を産み出せるからである。

【0024】

テレプレゼンス概念は、居宅用テレビ電話(residential video telephony)を実現するのに有用ではあるが最適の手法とは言い切れない。従って、ローカルユーザ（群）からリモートユーザ（群）への動画通信に使用できる優れた居宅向け動画通信システムを実現するには、本願で提案するシステム、装置及び方法を用いるべきであろう。本発明の種々の実施形態に係る動画通信システムには、例えば次に示す主要機能属性のうち幾つかを持たせることができる：

- ・ 画像表示装置、例えば可搬サイズのもの；
- ・ 動画撮影用の 1 個又は複数個の撮影装置、例えばその撮影視野(field of view)が広い W F O V (wide field of view) 型撮影装置と N F O V (narrow field of view) 型撮影装置の組合せ；
- ・ 鳴音装置及び採音装置を有する音響システム；
- ・ ローカルサイト及びそこにいる人物のプライバシーを保護するプライバシー保護インタフェース(privacy interface)；
- ・ ローカルサイト及びそこにいる人物の撮影態様を取り仕切るコンテキスト伝達インタフェース、特にプライバシー保護インタフェース等と連携して種々の有用な機能を提供するもの；
- ・ キー入力、顔認識、音声認識等を通じ人物を識別する機能。

## 【0025】

図1に、一実施形態に係る双方向型の動画通信システム290のブロック構成を示す。このシステム290では、動画通信装置（端末）300が設置されている場所（ローカルサイト）362と、同様の端末300が設置されている別の場所（リモートサイト）364との間で、通話を交わすことができる。図示例の端末300はそれぞれ電子イメージング装置100を有しており、その装置100はそれぞれ画像表示装置乃至ディスプレイ110と撮影装置（カメラ）120を備えている。それらを使用し通話を交わす人物のうち、サイト362にいる方の人物10aをローカルユーザと呼び、サイト364にいる方の人物10bをリモートユーザと呼ぶ。どちらの端末300も、動画像の撮影や処理、更には通信ネットワーク360を介した動画像の送受信を、ハンドシェイクプロトコル、プライバシー保護プロトコル及び帯域的制約に従い取り仕切っている。サイト362側端末（ローカル端末）300は、図示しないリモートサーバからの支援を受け、ネットワーク360越しにサイト364側端末（リモート端末）300に接続可能であるので、ユーザ10a、10bはそれぞれ手許の装置100のディスプレイ110及びカメラ（群）120を使用し互いに通話することができる。コンピュータ340例えば制御用論理プロセッサ乃至CPUは、イメージプロセッサ320及びシステムコントローラ330の動作を調和させる。システムコントローラ330はディスプレイドライバ機能や撮影制御機能を担う部材であり、必要ならコンピュータ340に組み込むこともできる。そのコンピュータ340はローカルに、即ちローカル端末300と一体に設けるのが原則であるが、一部機能をリモートに、即ちシステム290内のリモートサーバやリモート端末300に設けることもできる。この構成では、システムコントローラ330から相応のコマンドを発することでカメラ120のビューアングル、焦点等の特性を制御することができる。また、通信コントローラ355を、無線回線、有線回線等からなるネットワーク360に接続するためのインタフェースとして用い、サイト間で画像等のデータをやりとりすることができる。

## 【0026】

この動画通信システム290によれば、例えば複数の居宅同士でテレビ会議、テレビ電話等の動画通信イベント600を好適に開催することができる。即ち、そのイベント600でやりとりされる動画像中のどの場面(video scene)620でも、サイト362・364間で動画像及び音声信号を好適にやりとりすることができる。このとき、リモートユーザ10bの姿もローカルサイト362側ディスプレイ（ローカルディスプレイ）110によって画像として表示されるので、ユーザ10aは相手の姿を見ながらユーザ10bと通話し遠隔会議を円滑に進めることができる。イメージプロセッサ320は、そうした双方向通信に役立つ種々の機能として、自サイト（例えば362）における撮影画質を高める機能、自サイトのディスプレイ110における表示画質を高める機能、他サイト（例えば364）に送信するデータを処理する（画質向上・データ圧縮・暗号化等を施す）機能等を有している。なお、図1が本発明の一実施形態についてその部材配置のあらましを示すためのものであることを重々承知されたい。カメラ120及びディスプレイ110を図示しない共通の枠乃至ハウジングに組み込んで端末300の集積性を高めることや、その枠乃至ハウジングにサイト362又は364を構成する他の部材（プロセッサ320、通信コントローラ355、コンピュータ340、システムコントローラ330等）を組み込むことも可能である。

## 【0027】

端末300は、同じく図1に示す通り、採音装置（例えばマイクロホン）323及び鳴音装置（例えばスピーカ）324や、それらが接続されているオーディオプロセッサ325を備えており、そのオーディオプロセッサ325はコンピュータ340に接続されている。マイクロホン323の個数は1個でも複数個でもよい。マイクロホンの種類は指向性マイクロホンでも無指向性マイクロホンでもよい。即ち、音声信号を採取し、その音響エネルギーを適切な形態へと変換してプロセッサ325に伝達することができればよい。そのプロセッサ325は、その音声信号を更に別の信号形態へと変換してコンピュータ340に供給する。マイクロホンには、このほかにも種々の音声通信用部材又はそれに類する補

助部材を設けることができる。音声通信の分野で習熟を積まれた方々（いわゆる当業者）にとりそれらは既知のものである。スピーカとしては既知のスピーカ又はそれに類する部材を用いることができる。プロセッサ 3 2 5 から供給される音声信号に従い音響エネルギーを輻射することができればよい。スピーカには、このほかにも種々の音声通信用部材又はそれに類する補助部材を設けることができる。音声通信の分野に係る当業者にとりそれらは既知のものである。プロセッサ 3 2 5 は、コンピュータ 3 4 0 から受け取った信号を必要に応じ相応の形態に変換し、その信号を鳴音装置 3 2 4 に供給することで可聴音や超音波を発生させ、それらを搬送波として音響エネルギーを所定方向に輻射する。加えて、マイクロホン、スピーカ、プロセッサ 3 2 5 及びコンピュータ 3 4 0 には、こうした機能のほかに、音声信号 / 音響エネルギーの採取 / 輻射時にそれらを補強する機能、例えば増幅、フィルタリング、変調等の既知手法による拡張処理機能を持たせることもできる。

10

#### 【 0 0 2 8 】

ユーザ 1 0 は、例えば図 2 に示すようなローカル環境乃至場面 4 1 5 で、自サイト（例えば 3 6 2）の端末 3 0 0 をインタラクティブに操作する。図示例ではユーザ 1 0 が台所におり、そこにある機器例えばコンロ 2 7 0 で調理をしながら、本端末 3 0 0 で通信を行っている。その端末 3 0 0 は、台所の近くにある宅内調度例えばキッチンキャビネット 2 6 5 に、カメラ 1 2 0 及びマイクロホン（いずれも図示せず）がほぼユーザ 1 0 の方を向くよう装着されている。図中破線で示されているように、カメラ 1 2 0 は角度幅（フルアングル）＝ に亘る撮影視野 4 2 0 内から、マイクロホンは採音視野 4 3 0 内からデータを採取する。

20

#### 【 0 0 2 9 】

図 3 A 及び図 3 A に、端末 3 0 0 及びその電子イメージング装置 1 0 0 の一例構成を示す。特に図 3 A に詳示の通り、この装置 1 0 0 は、表示画面 1 1 5 を有するディスプレイ 1 1 0 及びカメラ 1 2 0 のほか、ディスプレイ 1 1 0 のハウジング 1 4 6 内に収容され又は無線乃至有線で装置 1 0 0 に接続されるコンピュータ 3 4 0、システムコントローラ 3 3 0、データストレージ 3 4 5、通信コントローラ 3 5 5 等の部材を備えている。また、ディスプレイ 1 1 0 にはピクチャインピクチャ表示機能がある。これは、画面分割画像 4 1 0 を表示画面 1 1 5 の一部に表示させる機能である。この機能を担える電子式表示装置の例としては、液晶表示装置（LCD）、有機発光ダイオード（OLED）、陰極線管（CRT）、投写型ディスプレイ、光導波型ディスプレイ等がある。カメラ 1 2 0 としては、電子式カメラ乃至デジタルカメラ、即ち撮像レンズ系及びイメージセンサ（いずれも図示せず）を有するカメラを使用している。そのイメージセンサの解像度は 1 ～ 1 0 メガピクセル程度、種類は CCD 型又は CMOS 型アレイデバイスである。装置 1 0 0 は、このほか、マイクロホン（群）、スピーカ（群）、環境光検知器 1 4 0、動き検知器 1 4 2、ユーザ用コントローラ（interface control）1 9 0 等を備えている。環境光検知器 1 4 0 は、独立したセンサを装置 1 0 0 のハウジング 1 4 6 に組み込むことで実現してもよいし、カメラ 1 2 0 自体に環境光検知機能を持たせること（独立した環境光検知器 1 4 0 を不要にすること）で実現してもよい。同様に、動き検知器 1 4 2 も、独立したセンサを装置 1 0 0 のハウジング 1 4 6 に組み込むことで実現することができるほか、カメラ 1 2 0 及びマイクロホンに動き検知機能を持たせ、独立した動き検知器 1 4 2 を不要にすることもできる。図 3 A では、更に、装置 1 0 0 のハウジング 1 4 6 にユーザ用コントローラ（群）1 9 0 が組み込まれている。これらのコントローラ 1 9 0 としては、例えばボタン、ダイヤル、タッチスクリーン、その組合せ等のインタフェース部材を設けることができる。

30

40

#### 【 0 0 3 0 】

端末 3 0 0 は、他方の図 3 B に詳示の通り、更に撮影システム 3 1 0、音響システム等の電子システムを有している。撮影システム 3 1 0 はカメラ（群）1 2 0 及びイメージプロセッサ 3 2 0 で、また音響システムはマイクロホン（群）、スピーカ（群）及びオーディオプロセッサ 3 2 5 で構成されている。コンピュータ 3 4 0 は、図中破線で示す通り、撮影システム 3 1 0 乃至そのプロセッサ 3 2 0、音響システム乃至そのプロセッサ 3 2 5

50

、更にはシステムコントローラ 330 と連携して稼働する。なお、図中に破線で示した接続関係は端末 300 における主要な有線又は無線接続関係の一例に過ぎない。しかも、これ以外の接続関係にすることも可能であるし、図示した以外にも様々な接続手段、例えば電源配線、内部信号線、データ伝送路等も必要になる。コンピュータ 340 は、更にユーザ追跡(user tracking)手順 480、ユーザ識別(user identification)手順 470、視線順応(gaze adaptive)手順 495、データストレージ乃至メモリ 345、ユーザインタフェース 440 等に接続され又は連携している。これらのうち手順 480 は動き検知器 142 における動き検知結果に基づきコンピュータ 340 上で稼働する相応のアルゴリズム、手順 470 はコンピュータ 340 上で稼働する相応のアルゴリズム、手順 495 は視線補正(gaze correction)及び視線追跡(gaze tracking)の両手順を含むアルゴリズムである。10  
ストレージ 345 は 1 個又は複数個の部材、例えばコンピュータ用のハードディスク、フラッシュメモリ等や、複数の動画像フレームを保持できるフレームバッファ(動画像データの解析と調整を同時実行可能とするバッファ)で構成されている。インタフェース 440 は様々な物理的形態のユーザ用コントローラ 190、例えばキーボード、ジョイスティック、マウス、タッチスクリーン、押しボタン、グラフィカルユーザインタフェース等を備えるほか、機能的には表示画面 115 もコントローラ 190 の一種である。プライバシー保護インタフェース 400 やコンテキスト伝達インタフェース 450、更には外見修正(appearance-altering)インタフェース 490 も、インタフェース 440 の一端を担っている。インタフェース 450 の一部はキュー依拠(cue-based)インタフェースとなっており、そのキュー依拠インタフェースもインタフェース 440 の一端を担っている。キュー依20  
拠インタフェースの基本的な役目は、音声コマンド、音響キュー(抑揚、ピッチ等)、ジェスチャ、身体の姿勢等といった可解釈キュー(interpretive cue)を観測することと、それに対して端末 300 が示すべき反応乃至応答を導出乃至決定することである。これらのインタフェースの土台をなすデータベース機能、解析機能及び制御機能は、コンピュータ 340、ストレージ 345、ディスプレイ 110、カメラ 120、コントローラ 190 等、種々の装置構成部材を利用して実現されている。

#### 【0031】

ユーザインタフェース 440 上に設けられたユーザ用コントローラ 190 のうちあるものは、まずシステムオンオフスイッチ、出力音量コントローラ、表示輝度コントローラ等の部材として使用される。これらはテレビジョン受像機やコンピュータ用モニタの分野でも一般的であるので容易にご理解頂けるであろう。コントローラ 190 のうちあるものは、端末 300 に備わるテレビジョン視聴機能を制御するリモートコントローラ、即ちチャネル選択、記録媒体(DVD等)へのテレビジョン録画の制御、当該記録媒体からの動画像コンテンツ再生の制御等の手段として使用することができる。コントローラ 190 のうちあるものは、端末 300 に備わる電話機能の制御、例えば発呼、発呼者識別、電話帳管理等、電話機で一般的な諸機能の制御に使用することができる。コントローラ 190 のうちあるものは、カメラ 120 に備わるズーム、パン、ティルト等の機能の制御に使用される。そして、重要なことに、コントローラ 190 には自動機能もある。例えばユーザ識別手順 470 やユーザ追跡手順 480 を実行してカメラ(群) 120 に撮影を指令する機能である。従って、端末 300 のユーザ 10 は、シームレスで優雅な体験を享受することができる。30  
40

#### 【0032】

端末 300 は、図 3A では 1 個としたが、図 1 及び図 3B に示した通り 2 個のカメラ 120 を有している。端末 300 では、これらのカメラ 120 を所与の動画撮影セッティングに従い稼働させ動画像を撮影する。実際、端末 300 に複数個のカメラ 120 を設けること、とりわけ WFOV カメラ及び NFOV カメラを併用することは際立って有益なことである。WFOV カメラ 120 用のレンズは、カバーできる視野 420 の角度幅即ちフルアングル が一般に 60 ~ 120 ° 程度と広いカメラ用広角レンズであり、NFOV カメラ 120 用のレンズはそのフルアングル が一般に 5 ~ 15 ° 程度と狭い(固定焦点)望遠レンズである。カメラ間にビームスプリッタを配し対物空間側の光軸(及びレンズ)を50



共通化することも可能であるが、多くの場合、カメラ 120 及びそのレンズを分離させた方が有利であろう。その詳細を図 4 A に示す。この例では一方のカメラ 120 で NFOV 420 を捉え、他方のカメラ 120 で WFOV 420 を捉えている。この図では、上側に示されている NFOV カメラ 120 の撮影視野 (NFOV) 420 は、もう一方のカメラ即ち WFOV カメラ 120 の視野 (WFOV) 420 の中心に位置しているが、NFOV カメラ 120 にはパン、ティルト、ズーム等の機能が備わっているので、リモート端末 300 やコンピュータ 340 からの信号に応じパン、ティルト、ズーム等を自動調整し NFOV 420 のサイズや位置を変化させること、ひいてはローカル環境 415 における人間のアクションを追跡させることができる。なお、WFOV カメラ 120 の出力を画像処理して NFOV 420 相当の画像データを得ることも可能であるが、NFOV カメラ 120 を別体且つ専用のカメラにした方が画質が良好になる。

10

#### 【0033】

その NFOV カメラ 120 は、WFOV 420 内に複数の NFOV 420 が生じるよう複数個にすることもできる。また、図 4 A 及び図 4 B に示す通り NFOV 420 が WFOV 420 内に丸ごと収まることが多いが、NFOV カメラ 120 が複数個ある場合や、その居宅の状況がひどくめまぐるしく変化する場合等には、NFOV 420 が WFOV 420 内に収まりきらないこともある。即ち、同じローカル環境 415 で同時に使用される NFOV カメラ 120・WFOV カメラ 120 間で、それらの視野 420 間に部分的にしか重なりがないことや、全く重なりが生じないこともあり得る。

20

#### 【0034】

次に、ユーザインタフェース 440、特にその構成部分たるプライバシー保護インタフェース 400 及びコンテキスト伝達インタフェース 450 について説明する。インタフェース 440 は端末 300 の構成部材のなかでも特に重要な部材であり、そのなかでもプライバシー及びコンテキストに関わるプライバシー保護インタフェース 400 及びコンテキスト伝達インタフェース 450 は高度に工夫されている。そこで、その性質を好適に理解するため、ユーザ 10 の振る舞い方や、ユーザ 10 及びその家族の自宅での暮らしぶりについて、検討することとする。まず、端末 300 が自宅にあれば、その家の住人 (ユーザ 10) は、その端末 300 をうまく使いこなしたいと思うものである。具体的には、多くのユーザ 10 が、コンテキスト的に問題のあるコンテンツを認識及び制限するのに役立つユーザ用コントローラ (群) 190 を求めるものである。また、プライバシー管理には融通性や順応性が求められる。これは、期待されるプライバシー保護レベルが通信イベント毎に異なり、また人間毎、家族毎、世代毎、社会階層毎、文明圏毎及び宗教毎にも異なるからである。そのため、後述の通り、インタフェース 440 は、インタフェース 400 及び 450 を通じ、ユーザ 10 が種々のプライバシー保護セッティング (privacy settings) を制定して様々な洗練度のプライバシー管理を行えるようにしている。

30

#### 【0035】

特に、ユーザ 10 による管制及びプライバシー保護に関する下記の要望事項は、様々な状況でユーザ 10 から頻発しうる要望であると思われる：

- ・ 自サイトから送信される画像及び音声の内容を知りたい、またその送信を自分で管制したい；
- ・ 自分のプライバシーが守られるようローカル端末 300 に自動反応機能を設けて欲しい；
- ・ 動画通信イベントに誰かが新規参加したことを自サイトに知らせてもらいたい、またその参加可否判断に自分が関与したい；
- ・ リモートユーザからの要求で自サイト内撮影視野が変更されたことを知りたい；
- ・ 撮影視野内にある被写体が写った画像の外見やコンテキストを自サイトで修正したい；
- ・ 他のユーザが画像を録画中であるか否かを自サイトに知らせてもらいたい。

40

#### 【0036】

これらの要望、ひいてはプライバシー保護インタフェース 400 及びコンテキスト伝達イ

50

ンタフェース 450 の必要性について理解するには、次のような例を考えるとよいであろう。まず、図 2 に示した例では、ユーザ 10 が、台所内環境 415 でコンロ 270 を扱いながら、端末 300 を使用し動画通信システム 290 上での動画通信イベント 600 に参加している。具体的には、遠くの知己と通話して指南を受けつつ調理を行っている。こうしたイベント 600 を開催する際、多くのユーザ 10 は、自分の身なりや自分のサイト 362 の見てくれを事前点検するものである。その事前点検で得心できない部分があった場合、無論、髪梳き、ゴミ拾い、片付け等を行い自分の身だしなみや環境 415 を綺麗にすることも可能であるが、ユーザ 10 にしてみれば、その手間を省くため端末 300 の機能でその部分をカバーして欲しいものである。ユーザ 10 は、更に、そのイベント 600 が始まったこと、特に動画撮影が始まったことを自分に知らせて欲しい、と望むであろう。ユーザ 10 は、また、イベント 600 が始まった後で生じた状況の変化、例えば自分のサイト 362 に存する環境 415 に新たな人物が現れたといった変化や、その環境 415 からそれまでいた人物が去ったといった変化に、ローカル端末 300 が速やかに順応する（プライバシーを保護しつつ撮影視野 420 を変更する）ようにして欲しい、とも望むであろう。ユーザ 10 は、そして、そのイベント 600 で撮影された動画像のデータ伝送、サイト 362 又は 364 における録画、話者追加による例えば二者間通話から三者間通話への移行等を操作乃至管制したい、と望むであろう。

10

#### 【0037】

こうした要望事項に応えるため、ユーザインタフェース 440 ではユーザ向け制御乃至管制機能を提供している。また、そのプライバシー保護インタフェース 400 でデフォルト版（所定内容）のプライバシー保護セッティングが保持されており、ユーザ 10 はインタフェース 440 によるプライバシー保護及びコンテキスト伝達の管理をそのデフォルト版プライバシー保護セッティングに委ねることもできる。従って、インタフェース 440 の操作を省くことができるので、端末 300 の日常的な使用がユーザ 10 にとり負担になることはあまりない。更に、そのインタフェース 440 上に準備されているユーザ用コントローラ 190 は簡易操作型のものであるので、ユーザ 10 は、イベント 600 の開催前に、或いは開催中でも、プライバシー保護セッティング及びコンテキスト伝達セッティング(contextual settings)を手早く修正することができる。総じて、インタフェース 400 は、各回動画通信イベントで修正されうる短期的プライバシー保護セッティングと、その修正頻度が割合に低い長期的（デフォルト版）プライバシー保護セッティングと、からなる二階層インタフェースになるので、ユーザ 10 は、居宅間個人的動画通信時に送受信される動画像コンテンツの撮影、表示及び録画を基本的にはそのインタフェース 400 の使用で管制することができる。端末 300 による画像通信をこうしてプライバシー保護セッティング及びコンテキスト伝達セッティングに従い管制することで、大筋では、現代の個人的動画通信に必要とされるエチケットに従うことができる。

20

30

#### 【0038】

端末 300 がリモートサイト 364 との接続に当たり使用するネットワーク技術は、地上伝送線（ケーブルや光ファイバ）を使用した電話接続やデータストリーミングである（このほか、無線、衛星通信、インターネット等も使用できる）。使用可能帯域幅上の制約から、ストリーミング動画像伝送技術はテレビ電話やテレビ会議の分野ではあまり使用されてこなかったが、この技術はなおも進歩を続けている。ユーザ 10 は、新規の動画通信イベント 600 をローカル端末 300 で開始させることも、既存の電話接続をイベント 600 に移行させることもできる。端末 300 には、これを支援する招待 605 及び応諾 610 の機能が備わっている。そのイベント 600 では、ユーザ 10 が修正版を使用しない限り、デフォルト版プライバシー保護セッティングが適用されることとなる。端末 300 は、そのデフォルト版プライバシー保護セッティングに従いサイト 362, 364 間の通信をまず音声伝送のみで開始し、次いで動画通信に移行する。その際、ローカル端末 300 は、リモート端末 300 との間で、プライバシー言語プロトコル(privacy language protocol)に則り通信ネットワーク 360 越しにプライバシー保護セッティングを交換する。例えば、そのユーザ 10 a, 10 b が共に画像通信開始を承認すると、それらローカル端末 30

40

50

0 及びリモート端末 300 が銘々に又は一斉に動画像伝送へと移行する（ハンドシェイク手順）。また、この動画像伝送では、ネットワーク 360 を介しサイト 362・364 間で暗号化動画像データストリームをやりとりする。そのため、それぞれ相手側がその暗号を解読することができるよう、ローカル端末 300 はリモート端末 300 向けに、またリモート端末 300 はローカル端末 300 向けに、プライバシー言語プロトコルに則り電子鍵を発行する。更に、ローカル端末 300 はローカルユーザ 10a 向けに、リモート端末 300 はリモートユーザ 10b 向けに、自サイトでの撮影が始まったことを通知する。そのためのフィードバック機構として、端末 300 は、音響や光（例えばフラッシュ）を発するインジケータを備えている。このインジケータは、動画通信イベント開始時等に、動画像が撮影及び送信されることをユーザ 10 に警告する。加えて、図示しないがカメラ 120 には物理的遮光シャッタが備わっている。ユーザ 10 は、そのシャッタを見るだけで、現在の装置状態が自サイト内を撮影可能な状態か否かを知ることができる。

10

#### 【0039】

端末 300 は、更なるプライバシー保護機能として、図 3A に示す如く画面分割画像 410 例えばピクチャインピクチャ画像を表示させる機能を有している。具体的には、ローカルディスプレイ 110 の表示画面 115 のうち主要な部分を使用し、リモートサイト 364 及びそこにいるリモートユーザ 10b の画像を表示させると共に、ローカルサイト 362 で撮影されリモートサイト 364 に送られる動画像を、画像 410 として表示させることができる。従って、自分のディスプレイ 110 上の画像 410 を見ることで、ローカルユーザ 10a は、ローカル端末 300 で撮影されサイト 364 に送信されていく画像を、自分の目で確かめることができる。即ち、そのユーザ 10a は、送信される画像のプライバシー保護レベルが所望レベルであることを確認し、必要に応じ調整する機会を得ることができる。

20

#### 【0040】

端末 300 は、動画通信イベント 600 の開始後に通信ネットワーク 360 越しの動画像伝送に不調が生じた場合、プライバシー言語プロトコルに則りそのネットワーク 360 越しに不調の事実を通知する。各端末 300 は、インジケータたる警報器 350 を作動させること、例えば画面分割画像 410 上に相応のアイコンを重畳表示させることによって、自サイトのユーザ 10 向けに伝送不調警報を発する。ローカルユーザ 10a 及びリモートユーザ 10b は、不調が生じていることを相手との音声通話で知り、或いは自サイトの端末 300 からの警報で知ることができる。それらのユーザ 10 は、自サイトの端末 300 に備わっているユーザ用コントローラ 190 のうち相応のものを使用することで、送信される音声信号を無音化（ミュート）させることや、送信される動画像を抑圧（ブランキング）させることができる。その操作は、そのイベント 600 を中断することなく迅速且つ簡便に行うことができる。更に、動画像送信抑圧中に、その動画像を別の動画像又は静止画像に差し替えることもできる。

30

#### 【0041】

ユーザ 10 は、至極もっともなことながら、動画通信イベント 600 の内容がローカルサイト 362 やリモートサイト 364 で記録されるか否か、とりわけ動画像が録画されるか否かに関して、懸念するであろう。そのため、ローカル端末 300 及びリモート端末 300 では、自端末 300 のプライバシー保護インタフェース 400 によって提供されるプライバシー保護セッティングのうち相応のものに従い、動画像のローカル録画又はリモート録画に関するユーザ 10 の承認を求め、その結果を示すステータス情報をプライバシー言語プロトコルに則りやりとりする。各端末 300 は、相手側の端末 300 から録画承認とのステータス情報を受信すると、録画（又は音声信号の記録）を開始し、そのことを相手側の端末 300 に通知する。録画実行中は、サイト 362, 364 の双方にて警報器 350 が作動する。例えば、ディスプレイ 110 によるアイコン表示やハウジング 146 外面に設けたライトの点灯によって、録画中であることが通知される。また、そのイベント 600 が二者間通話から三者間通話へと切り替わるときには、プライバシー言語プロトコルに則り所定内容のデフォルト版プライバシー保護セッティングが送信されてくる。即ち、通話への

40

50

参加意志が通知されてくる。ユーザ 10 のなかには、プライバシー保護セッティングのうちサイト 362 又は 364 での録画に関する部分や、イベント 600 の途上での多者間通話への移行に関する部分に対し、修正を施したいと望む者もいよう。また、ユーザ 10 のなかには、ローカルディスプレイ 110 による画面分割画像 410 の表示をイベント 600 の開催中は行わせないでおきたい、気が向いたら見ることができるよう録画しておきたい、と望む者もいよう。ユーザ 10 は、ユーザ用コントローラ 190 を操作することで、プライバシー保護セッティング等のセッティングを修正し、これらの望みを簡単に叶えることができる。

#### 【0042】

ユーザ 10 は、更に、ローカル端末 300 に指示して撮影視野 420 等の撮影条件を変化させることができる。こうした機能を設けたのは、図 2 に示すように、端末 300 のカメラ 120 をローカルサイト 362 例えば台所に向けたとしても、その視野 420 内に台所以外の居住スペースが部分的に入り込んでくるからである。ユーザ 10 にも色々な都合があるので、サイト 362 内の余分な個所を視野 420 に入れたくないと思うことがあろう。そうした場合、ユーザ 10 は、ドアを閉めるなり他の部屋の照明を消すなりしてローカルサイト 362 内視野 420 を制限することができるほか、NF OV カメラ 120 だけを撮影に使用するように求めるプライバシー保護セッティングを適用させることで、視野 420 を制限することができる。これに加え、NF OV カメラ 120 のパン、ティルト及びズーム動作を制限することで、端末 300 の視野 420 及び動き追跡機能を各回動画通信イベント 600 毎に制限することができる。

#### 【0043】

ユーザ 10 は、また、ローカル端末 300 で撮影された画像を電子的に修正し、撮影視野 420 の一部 422 が削除された画像を送信させることができる。ユーザ 10 が領域 422 を指定して動画像から除かせる動機は幾つかあろうが、最たるものは自分や家族のプライバシーを守りたいという想いであろう。図 4 A に、WF OV カメラ 120 の視野 (WF OV) 420 を示す。通信ネットワーク 360 経由でリモートサイト 364 に送信される画像は、修正対象として指定された部分 422 を、例えばクロッピングによって撮影済画像から削除したものである。ユーザ 10 は、プライバシー保護インタフェース 400 及びコンテキスト伝達インタフェース 450 を稼働させることで、ローカルサイト 362 を捉えた画像のうち高プライバシー部分 (privacy sensitive area) 422 の外見を、このように修正させることができる。また、ユーザ 10 は、それに適するプライバシー保護セッティングを予め制定しておくことができる。例えば、動画通信イベント 600 が始まる前にプリビューモードで動画像撮影を実行させることで、人目に付き情動を刺激する画像構成要素乃至コンテンツを調べ、それを勘案してプライバシー保護セッティングを初期設定すればよい。勿論、イベント 600 の最中でユーザ 10 が撮影済画像に対する修正内容等を指定することもできる。なお、プリビューモードにおける設定操作は、表示画面 115 上の主たる領域を使って行ってもよいし、画面分割画像 410 が表示される領域を使って行ってもよい。

#### 【0044】

コンテキスト伝達インタフェース 450 は、高プライバシー部分 422 の画像コンテンツがプライバシー保護セッティング通りに修正されるよう、例えば場面解析 (scene analysis) アルゴリズムに則りイメージプロセッサ 320 に指示し、撮影される画像からリアルタイムで修正版動画像信号を自動生成させる。このとき、その部分 422 を撮影済画像からクロッピングで削除してプライバシー保護上の問題を抑えるだけでもよいし、歪付加、ブラーリング (解像度低減)、シェーディング (輝度 / コントラスト低減) 等の視覚効果を施し隠蔽性の修正を図ってもよい。例えばシェーディングを使用すれば、天然の明暗移ろいに似た輝度勾配を付けることができる。端末 300 で決めた場面解析ルールに従い適当な隠蔽性視覚効果を推奨するようにしてもよい。状況にもよるが、推奨された視覚効果又はそれに類する隠蔽性視覚効果を適用することで、部分 422 に対する隠蔽効果を発生させることができる。即ち、部分 422 にどのようなアクション、どのような物体が写っている

かを、リモートユーザ 10b が容易に視認乃至推察できない修正版画像を得ることができる。更に、修正される部分 422 の形状を任意に指定することで、撮影視野 420 内の所望部分を修正例えば除去して画像を送信させることができる。

#### 【0045】

ローカルディスプレイ 110 は、こうしてプライバシー保護セッティングに従い削除乃至視覚効果が施された画像や、その部分を含めローカル環境 415 内をより広範囲に捉えた画像を、画面分割画像 410 として表示する。例えば図 4B に示す例では、環境 415 のうち WFOV カメラ 120 の視野 (WFOV) 420 から捉えられた画像が、高プライバシー部分 422 として指定されたため送信対象域から除外されている部分を含め画像 410 としてローカル表示されている。カメラ 120 の可動範囲例えばパン、ティルト又はズームの可動範囲も、プライバシー保護の観点から制限されている。WFOV カメラ 120 で捉えた画像をこうしてローカルディスプレイ 110 により画像 410 として表示させる場合、部分 422 に目印例えば陰影を付けて WFOV 420 全域に亘る画像を表示させる。その目印を見ることで、ローカルユーザ 10a は、現在どの画像領域及びコンテンツが隠蔽乃至除外されているのかを即座に理解することができる。また、これに代え、実際に送信される画像、即ち WFOV 420 から部分 422 を除いた部分の画像を画像 410 として表示させてもよい。画像 410 としては、このほか、WFOV 420 のうちリモートユーザ 10b が現在見ている NFOV 420 内部分だけを表示させることや、当該 NFOV 420 内部分を含む大きな WFOV 420 内画像を NFOV 420 内部分指示目印 (図示例では子供達を囲んでいる鎖線) と共に表示させることができる。

#### 【0046】

プライバシー保護セッティングは、ユーザ 10 によって制定される段階では人知可能な画像コンテンツの形態をとっているが、端末 300 を動作させる際には、端末 300 の動作に適した形態、即ち動画像乃至音声データに準じた形態に変換する必要がある。この変換で得られるのは、画像中の隠蔽すべき部分 422 (その広さと場所)、その隠蔽を担う視覚効果 (種類)、その隠蔽の規模・程度、身体的特徴、音声、ジェスチャ、ローカルユーザ 10a 及びリモートユーザ 10b の身元、物品の種類等々、撮影及び送信される画像のコンテンツ及びコンテキストに関わるプライバシー保護セッティングパラメタである。これらは各回動画通信イベント 600 毎に変わることもあれば変わらないこともある。図 2 に示した端末 300 のように、そのローカル環境 415 にある宅内調度例えばキャビネット 265 に装着等して常用されるのであれば、それら部分的画像修正に関わるプライバシー保護セッティング及び対応するプライバシー保護セッティングパラメタをまとめてデフォルト設定しておけば、各回使用時に、それらのデフォルト条件を迅速に読み込み自動適用させることができる。また、端末 300 の典型的な使用環境は判っているので、現存の環境 415 がどのような環境であるかを、コンテキスト伝達インタフェース 450 で場面解析アルゴリズムに則り認識し、所定のデフォルト版プライバシー保護セッティングのうちその認識の結果に相応するものを適用することができる。

#### 【0047】

こうしたプライバシー保護セッティングは、動画通信イベント 600 又は場面 620 の切り替わりに対して割合に安定であるが、カメラ 120 や注目被写体の動きに対応するには、画像コンテンツへの適用形態を随時変化させる必要がある。そのため、プライバシー保護セッティングのうちユーザ 10 の身体的プライバシーを保護するためのもの、例えば撮影視野 420 内を動き回る裸の子供を隠すためのものを、その隠蔽の対象となる任意形状の修正対象部分 422 を適宜移動及び伸縮させてその部分 422 を衆目から守る、というセッティングにしてある。カメラ 120 が固定型である場合、こうしたプライバシー関連画像部分隠蔽処理を実行するには、視野 420 の形状・寸法に依拠しつつ、被写体に対するカメラ 120 の動きを補償する必要がある。そのため、コンテキスト伝達インタフェース 450 上の場面解析アルゴリズムでは被写体本位法を使用している。被写体本位法とは、プライバシー上問題になる部分としてユーザ 10 がプライバシー保護セッティングで指定している部分 422 の存否及び位置を、被写体の画像を調べることにより認識し、認識した部分 4

22をプライバシー保護セッティングパラメタにより被写体と関連付ける手法である。

【0048】

加えて、ユーザ10は、ローカル端末300に対し充実したプライバシー保護機能を求めるものである。そのため、この端末300では、事態の展開に相応するプライバシー保護セッティングを場面解析アルゴリズムに則り自動的に選定し、それを適用するようにしている。その際には、プライバシー保護セッティングパラメタに照らし且つプライバシー保護セッティング指標を用い、場面解析アルゴリズムに則って動画像を自動精査する。プライバシー保護セッティング指標は潜在するプライバシー問題を示唆する指標であり、画像の大きさ・形状・表面起伏等といった画像コンテンツに関するものも、また不適切な服装等といった画像コンテキストキューに関するものもある。場面解析アルゴリズムでは、プライバシー保護セッティング指標を用い撮影視野420例えば個々の動画像フレーム内の画像を解析することで、隠蔽対象となりうる修正対象部分422の所在及び位置を認識する。1個又は複数個のカメラ120から得られる直近の動画像フレームを評価し、従前の動画像フレームとの関係を精査するので、場面解析アルゴリズムによる解析処理の流れはスムーズで、その結果も良好なものになる。必要なら、プライバシー補正自信度を統計的に算出し、それに基づき部分422の認識結果を点検乃至評価した上で、動画像にその修正を施すようにしてもよい。場面解析アルゴリズムによる処理は、画像データを(プライバシー保護セッティング指標を介し)プライバシー保護セッティング(又は対応するパラメタ)と照合し又は連関させることで進行していく。それにより状況が整ったら、撮影された動画像を場面解析ルールの適用結果に従い直ちに自動修正する。これにより、ユーザからの指示(それを表すプライバシー保護セッティングパラメタ)、撮影条件、場面内コンテキスト等の要素に従い、認識済の修正対象部分422が隠蔽されることとなる。

【0049】

また、端末300は、主として個人の居宅で使用される端末である。個人の居宅では、場の静けさ、寛ぎ方、作法等が動画通信イベント600の最中にしばしば変化する。例えば、端末300の撮影視野420に新たな人物が入ってくる、といったことが起こる。端末300は、その場合、ローカル環境415にその人物がいることを示す画像をリモートサイト364に送ってもよいが、コンテキストを踏まえて適切に対処しなければならない。また、新たな人物の出現に伴い未適用のプライバシー保護セッティングが幾つか急に適用可能になった場合、端末300はその新たなプライバシー保護セッティングを認識、読込及び適用しなければならない。これらの応答、即ち動画撮影対象の新旧変転に対する応答の具体的な内容は、どのような状況で誰が入ってきたのかによって変わってくる。例えば、図4Cに示す例では、イベント600が催されている環境415にローカルユーザ10aがいる。図示されている場面のうち第1の場面620では、通話を行っているユーザのうち一方(ユーザ10a)がテーブルに向かって着席している。このユーザ10aは、ローカルディスプレイ110に表示されているリモートユーザ10bの画像を見つめながら、大切な相談事についてそのユーザ10bと有目的な会話を交わしている。こうした場面620では、ユーザ10aを捉える視野420は狭めでもよい。これに対し、そのイベント600が更に進んでユーザ10a、10bが寛ぎ始め、より取り留めなく話をするようになった場面620'では、ユーザ10aが席を立て鷹揚に振る舞うので、視野420はより広い方がよい。総じて、本発明の実施形態では、端末300の対人対応動作、特にその端末300が自動稼働しているときのそれを、その端末300のコンテキスト伝達インタフェース450によって取り仕切ることができる。

【0050】

端末300は、図4Cに例示した一般的な筋書き以外の様々な筋書きでも使用されうる。寧ろ、動画通信イベント600の初っ端からリラックスして取り留めのない話が交わされる筋書きの方が多くであろう。そのことを考えると、当初から広めの撮影視野420を使用する方がよい。そのため、動画通信イベント開始時のカメラ制御に使用されることが多いデフォルト版動画撮影セッティングは、視野420の全域又は大半をWFOVカメラ120で捉える内容にしておく。従って、リモートユーザ10bは、少なくとも通信開始

10

20

30

40

50

当初は、ローカルユーザ 10 a を取り巻くローカル環境 415 のコンテキストをよく知ることができる。その後イベント 600 が進行すると、ユーザ 10 a 又は 10 b が、ローカル環境 415 内のある限られた部分 (NF OV) 内を接写せよとの指示を発する局面も発生してくる。また、ユーザ 10 b が、それまで撮影されていた個所から別の個所へと NF OV 420 を移すよう (例えば図 4 B 中でテーブル近辺から子供の近辺へと NF OV 420 を移動させるよう) 指示を発することもある。これらの指示は、ユーザ 10 a, 10 b のいずれからでも、ユーザ用コントローラ 190 の操作で発することができる。視野 420 がこれらの指示に応じ変更された場合も、ユーザ 10 a は、画面分割画像 410 を見ることで、どのような動画像がリモートサイト 364 に送信されるのかを知ることができる。また、プライバシー保護インタフェース 400 が稼働するので、ユーザ 10 a は、視野 420 を狭めよといったユーザ 10 b からの指示を拒否、改変等することができる。視野 420 を狭める処理は、通信ネットワーク 360 経由でサイト 364 に送信される画像をイメージプロセッサ 320 でクロッピング又はズームングすることで実行してもよいが、こうしたデジタルズームで得られる画像の質、特に解像度は、ある限界を超えると途端に低下する。従って、光学ズーム機能付のカメラ、NF OV カメラ 120 又はその双方を使用した方がよい。なお、光学ズーム機能付のカメラとは、一般に、そのレンズアセンブリ内に何個かの可動レンズが配されたカメラのことである。この端末 300 では、また、環境 415 内又はその付近にいるユーザ 10 a に対し、その視野 420 が及ぶ限りにおいて照準追跡するユーザ追跡手順 480 を実行することもできる。

10

20

30

40

50

#### 【0051】

端末 300 は、こうして動画通信イベント 600 が開催されている間、ローカルユーザ 10 a 及びリモートユーザ 10 b によるインタラクティブな操作に応じ、その撮影視野 420 を適宜修正していく。更に、端末 300 では、様々な状況で撮影動作を自動制御することで、ユーザ 10 からの期待に応えるようにしている。即ち、コンテキスト伝達インタフェース 450 では、種々のアルゴリズム及びデフォルト条件を適用することで、その人が今起こしているアクションを解釈して動画通信コンテキストを調べ、端末 300 による動画撮影のモード (video capture mode) をその結果に応じ従前のモードから新たなモードへと適宜変更させる。例えば、イベント 600 の開始時にローカル環境 415 内を WF OV カメラ 120 で撮影する場合、端末 300 は、その WF OV 420 内のどこかにユーザ 10 a の姿を捉えることができよう。コンテキスト伝達インタフェース 450 は、そのユーザ 10 a に対し時間制限及び空間制限を適用することで、ユーザ 10 a がその場所に所定時間 (例えば 30 秒) 以上に亘り居続けているか否かを判別する。インタフェース 450 にて「居続けている」と判別された場合、この端末 300 では、例えば表示画面 115 上で約 30 % 又はそれ以上の面積をその姿が占めるに至るように、ユーザ 10 a に向けて NF OV カメラ 120 をズームインさせる。インタフェース 450 では、更に、動き検知器 142 及びユーザ追跡手順 480 を稼働・実行させることで、環境 415 内又はその付近にいるユーザ 10 a に照準してその人物を追跡する (照準追跡 (lock onto and follow))。照準追跡する人物をユーザ種別、ユーザ身元、アクション種別等に基づき特定人物に絞ることも可能である。また、ユーザ 10 a の動きや居場所が広がってきたら、インタフェース 450 では画面 115 に占めるユーザ 10 a の姿の面積比を変化させる。そして、端末 300 は、これらの中に相当する半自動モードで動画撮影を行うこともできる。これは、ローカル又はリモートユーザ 10 とインタフェース 450 が、撮影及び送信の対象となる場所を変化させることができ、且つそのユーザ 10 にオーバーライド特権があるモードである。

#### 【0052】

端末 300 は、動画通信イベント 600 の開催中に誰かがそのローカル環境 415 の撮影視野 420 に出入りしたときには、より複雑な動作を実行する。例えば、その環境 415 からローカルユーザ 10 a (特にその時点でたった一人のローカルユーザ 10 a であった人物) が退出した瞬間に、それに伴うローカル画像コンテンツの変化に直ちに順応するための動作、具体的にはその撮影視野 420 を WF OV 又は NF OV 420 に変化させる

動作を実行する。端末300では、このように視野420を変転させユーザ又はそのアクションの変化に対応する動作を、ユーザ10による直接入力に応じて、或いはユーザ10が制定した動画撮影セッティングに従い実行する。端末300は、或いは、ユーザ10aが戻ってくるまでの間、送信する画像を静止画像やローカル環境415以外の画像に差し替える。また、その撮影視野420に新たな人物が入ってくる状況にも、端末300は順応することができる。即ち、ローカルサイト362でそれ以前からそのイベント600に参加しているユーザ10aが通常はそのことに気づくので、そのユーザ10aから発せられる指令に応じ視野420を拡張、縮小又は移動させることで、その人物を視野420に取り込み又はその視野420から外すことができる。

#### 【0053】

端末300は、更に、コンテンツに変転が生じたときに、そのコンテキスト伝達インタフェース450の動作で撮影視野420を順応的に自動変転させる。例えば、動画通信イベント600への人物の出入りを、それらの人々がアクションを起こしている場所及びそのアクションの持続時間(location and duration of activity)に基づき認否する。即ち、そのアクションの場所及び持続時間にまつわる種々のキューから動画通信のコンテキストを認識するので、インタフェース450では、動画像、音声、プライバシー、場面内又は場面間変転管理等の諸条件を充足させうるように、動画撮影モードを決めることができる。特に、アクションの場所及び持続時間を調べて場面の変転を認識するには、そのアクションの範囲、規模及び速度に関する量的指標があればよく、そのアクションの種別がわからなくてもよい。

#### 【0054】

端末300は、アクションの場所及び持続時間についてのデータを様々な形態で利用する。例えば自動モードで動画像撮影を行いながら動画通信イベント600に参加中の端末300では、コンテキスト伝達インタフェース450による決定に基づき、NF OVカメラ120でNF OV420内を撮影して得られる動画像(例えば図4B中の少年達の画像)を送信する動作と、WF OVカメラ120でより広い視野(例えば図4B中のWF OV420)内を撮影しローカル環境415にいる人々のアクションを広範囲に監視する動作とを、同時並行的に実行することができる。このとき、インタフェース450では、そのアクションの場所及び持続時間にまつわる種々のキューを利用することで、WF OV420から得られた画像のうちその時点で動画像として送信されていない外側部分を解析し、リモートユーザ10bの関心を惹きそうな画像を探す。インタフェース450がこの監視に当たり依拠する場面解析アルゴリズムは、WF OV420から得られる画像を相応の指標及びルールに基づき評価し、発生している変転の性質をユーザ10aのアクションに基づき検知するアルゴリズムである。なお、送信される動画像は、NF OV420又はその一部を捉えた画像、WF OV420又はその一部を捉えた画像、それらを任意に組み合わせたコンポジット画像等のなかから、プライバシー保護インタフェース400から課される制限の下に、インタフェース450で選定したものである。

#### 【0055】

コンテキスト伝達インタフェース450では、WF OV420のうち動画像として送信されない外側部分での人々のアクションを監視するのに、カメラ120で捉えた画像だけでなく、マイクロホン144で捉えた音声データも使用することができる。マイクロホン144の採音視野430はローカル環境415の物理的形状と相俟って決まるので、撮影視野420とは広狭の差がある。一般には視野430の方が広くなるので、主には撮影視野420からのデータに従い端末300を稼働させることにしつつ視野430も監視することで、人物の出入り等に早期対応することができる。

#### 【0056】

即ち、この機能を利用しWF OV420及び採音視野430の双方を対象に人々のアクションを監視することで、ローカル環境415への人々の出入りをより好適に検知することができる。具体的には、コンテキスト伝達インタフェース450上で種々の場面解析アルゴリズムを稼働させることで、新たな人物(潜在的な被写体乃至ユーザ10)がローカ

10

20

30

40

50



ル環境 4 1 5 の主要部例えば中央部に入ってくるのか、それともその環境 4 1 5 の辺縁部を過ぎるだけかを、予測、判別することができる。その際に使用するデータとしては、まず最新の動画像フレームに係る画像データ及び従前の動画像フレームに係る画像データがある。場面解析アルゴリズムでは、それらのデータのうち辺縁部でアクションを起こしている人々のいる場所及びそのアクションの持続時間にまつわるものを、画像エリア、時間しきい値及び動き追跡データを参照して調べる。参照されるデータとしては、更に、動画通信コンテキストに関連するコンテキスト伝達データ(contextual data)がある。コンテキスト伝達データとはユーザ種別(user classification)、ユーザ身元(user identity)、イベント種別(event classification)、ユーザ 1 0 のアクション・ジェスチャ・挙動のコンテキスト種別(contextual classification)等といったデータのことである。場面解析アルゴリズムでは、後述の如く、現在 N F O V 4 2 0 外にどのようなアクションが見られるかを監視・解析する際、このデータを使用する。即ち、場面解析アルゴリズムでは、それら画像データ及びコンテキスト伝達データに基づき、且つ相応する場面解析ルールに従い、辺縁部でのアクションを動画撮影対象に含めるべきか否かを判別する。例えば、その辺縁部から検知されたアクションが動画像に係る所定のコンテキスト、例えば一時的なアクションに該当する場合、そのアクションがプライバシー保護セッティングに抵触しないものであれば、送信する動画像に含めてもかまわない。この後は、場面内調整(scene adjustment)ルールに従い撮影形態修正(capture transition)アルゴリズムを稼働させ、撮影視野 4 2 0 を変化させる速度及びタイミングを決定する。

10

20

30

40

50

#### 【 0 0 5 7 】

例えば、新たな人物がローカル環境 4 1 5 に入ってきたそうであるとする。その場合、コンテキスト伝達インタフェース 4 5 0 では場面解析アルゴリズムに従い予測を行い、本当にその人物がその環境 4 1 5 の主要部に入ってくるのかそれとも辺縁部を過ぎるだけかを判別する。辺縁部を過ぎるだけ、という判別結果なら、その端末 3 0 0 ではその人物を単に無視する。これに対し、実際に入ってくる見込み、という判別結果であれば、それまで捉えていた狭い範囲よりも広い範囲を捉えるよう、即ち元々そこにいたローカルユーザ 1 0 a だけでなく新たにやってきた人物（新たにユーザ 1 0 a の一人となった人物）も捉えることができるよう、端末 3 0 0 の視野 4 2 0 を拡張させる。インタフェース 4 5 0 では、それらのユーザ 1 0 a が暫しその場所に一緒にいると、場面解析アルゴリズムに従い端末 3 0 0 の視野 4 2 0 をユーザ 1 0 a の居所へと狭めるよう指示し、カメラ 1 2 0 又はイメージプロセッサ 3 2 0 がその指示に従い動作する。インタフェース 4 5 0 では、それまでその視野 4 2 0 に捉えていたユーザ 1 0 a のうち一人又は複数人が立ち去っていく状況にも、同様にして順応することができる。例えば、それらのユーザ 1 0 a が立ち去っていくことが判るように、視野 4 2 0 を拡張させる。また、現在 N F O V 4 2 0 内にいる人物が N F O V 4 2 0 の辺縁部又は外部にいる別の人物と会話を始めたことを認識したときも、インタフェース 4 5 0 では必要であると判断し、後者の人物まで包括されるように端末 3 0 0 の視野 4 2 0 を拡張させる。なお、ユーザ 1 0 は、音声コンテキストを非常に厳しく制限する内容のプライバシー保護セッティングを制定することができる。例えば、同じ動画通信イベント 6 0 0 又は場面 6 2 0 が続いている限り、誰が関心を持ちそうでも、そのときの視野 4 2 0 外でのアクションを排除し続ける、という内容にすることができる。

#### 【 0 0 5 8 】

端末 3 0 0 のコンテキスト伝達インタフェース 4 5 0 は、更に、ユーザ種別に依拠するタグをキューとして用いつつコンテンツ撮影及びコンテキスト変転を管理し、動画像コンテキスト及び相応しい動画撮影モードを決定する、という手法でも変転を管理する。例えば、動画通信イベント開始時に、それに参画しているローカルユーザ 1 0 a に対しその動画通信イベント 6 0 0 について「一次ユーザ」なるタグを付与する。具体的には、顔認識手順乃至アルゴリズムと音声認識手順乃至アルゴリズムとを併含するユーザ識別手順 4 7 0 を、インタフェース 4 5 0 と共に稼働させることによって、端末 3 0 0 は、ユーザ 1 0 a に対し、「一次ユーザ（初期ユーザ）」「二次ユーザ」「第 1 成年ユーザ」「第 2 成年ユーザ」「第 1 未成年ユーザ」「第 2 未成年ユーザ」等の含意的な分類子乃至識別子を付

与する。このとき、端末300のインタフェース450でユーザ10aの真正な身元（名前、関係、年齢等）を知る必要はなく、タグ付け即ちユーザ分類に十分なデータがあればよい。付与されるタグの有効期間は通信イベント1回限り又は複数回である。ユーザ10aのうち動画通信イベント開始時に参画していた人（初期ユーザ）、成年に達している人、以前に使用したことがある人等には優先度の高いタグを付与する。例えばローカル環境415の撮影視野420内にタグ「一次ユーザ」「二次ユーザ」「第1未成年ユーザ」を有する三者が収まっている場面620から、そのうちの「一次ユーザ」が立ち去りつつあることを察知した場合、インタフェース450では、そのユーザ10aの立ち去る姿がカメラ120で捉えられるよう、暫くの間端末300の視野420をWFOV420へと広げさせる。その後、インタフェース450では、後に残った「二次ユーザ」及び「第1未成年ユーザ」のアクションがより好適に捉えられるように視野420を狭めさせる。インタフェース450では、そのイベント600が続いている間は、WFOV420及び採音視野430の監視及びそれにより得られる画像データに基づく辺縁部内アクションの解釈を続ける。即ち、インタフェース450では、辺縁部にいる人物及びそのアクションの重要性を場面解析アルゴリズムに則り判別する。例えば、上述の階層的タグ体系に見合った場面解析ルールに従い重要性を判別する。重要と見られる人物又はアクションを特定した場合、インタフェース450では、WFOVカメラ120が複数個のカメラを稼働させることで、その人物が包括されるように視野420を広げて動画像撮影を継続させる。単独画像又は画面分割画像410を用いることで、ローカルディスプレイ110による表示も然るべく広げる。

10

20

#### 【0059】

更に、この端末300で想定している主な使用場所は居宅である。なかでも所帯持ちの居宅では状況変動が生じやすく、例えばパーティ、家族夕食会、ゲーム大会等といった様々なイベントが開かれる。ローカルユーザ10aは、そのイベントに相応しいものとなるよう撮影視野420を設定し、その全域を端末300によって撮影及び送信させることができる。また、コンテキスト伝達インタフェース450には、開かれるイベントに自動順応する機能がある。その際、インタフェース450では、一群のイベント種別データをキューとして用いつつそのイベント乃至アクションの種別を認識し、動画通信コンテキスト及び適切な動画撮影モードを決定する。例えば、開催されているイベントの種別（イベント種別）がインタフェース450にて“パーティ”と認識されると、端末300はそれに

30

40

50

#### 【0060】

そのコンテキスト伝達インタフェース450は、手始めに、そこにいる人々の身元、アクションの内容、アクションの場所等に関わりなく、端末300の撮影視野420を広めにして撮影を開始させる。インタフェース450は、また、イベント種別や場所種別を被写体認識を通じて認識する。例えば椅子、テーブル、机、ソファ、自動車、芝刈り機、煉瓦、窓、絨毯、木目床、コンクリート床等の可認識被写体は、通常、ある特定の種類の部屋及びそれが使用されるイベントを示唆しているので、それらを認識することでそのイベントの開催場所を認識し、そこで開かれているイベントの種別（家族夕食会・パーティ・カードゲーム会・お茶会・寄り合い・井戸端会議等の別）を認識することができる。インタフェース450は、更に、そのイベントに相応しい一群のイベント撮影ガイドラインを適用する。このイベント撮影ガイドラインに組み込まれるのは、撮影視野420の設定、音声処理、所要照明、プライバシー要素、視線補正、追跡等の要素である。インタフェース450は、また、そのイベントでのアクションに見られる変化、例えば家族夕食会の後片付けが済んでボードゲーム遊びが始まったこと等を監視によって察知し、それらの変化の過程を通じ随時暫定版の動画撮影セッティングを適用して動画像の撮影及び表示動作、採

音動作等に係る変転を取り仕切る。更に、インタフェース４５０は、キュー依拠評価アルゴリズムを手助けとしてこのイベント解釈を実行する。このアルゴリズムは、諸音声要素（抑揚、ピッチ、ラウドネス、拍等）、ジェスチャ、姿勢等の可解釈キューを観測し、その結果に基づきアクションの性質を判別し、然るべき撮影動作や採音動作を指示するアルゴリズムである。そして、他の場合と同様、ユーザ１０は、自動モード下でインタフェース４５０が発した撮影関連の指示をオーバーライドすることができる。

#### 【００６１】

また、一般の居宅では、子供達や愛玩動物や隣人達が動き回り動画通信イベント６００の最中に撮影視野４２０に入り込んでくることが、例えばよそ様に見せたくないほど散らかった家の中を裸の子供達がうろつき回することも珍しくない。コンテキスト伝達インタフェース４５０では、そうした状況が発生したことを速やかに認識して相応の指示をイメージプロセッサ３２０宛に発し、プライバシー保護上問題となる部分にブラーリング（ぼかし）又はクロッピング（削除）を施させる。実際、プライバシー保護インタフェース４００にデフォルト版のプライバシー保護セッティングを組み込む際には、この種のブラーリングやクロッピングについて規定することが必須となろう。

#### 【００６２】

更に、居宅内で家族員等が好適に共用できるようにするため、端末３００には、ユーザ身元データをキューとして用いつつ個々の人物の身元を認識し、動画像コンテキスト及び適切な動画撮影モードを決定する機能が備わっている。前述の通り、この端末３００には顔認識、音声認識、バイオメトリクス認識（例えば指紋認識）、その任意の組合せ等の認識手法に基づくユーザ識別手順４７０が実装されているので、それを使用することによって、撮影視野４２０内にいる個々の人物の身元を認識することができる。例えば、顔認識を利用し家族員同士で端末３００を共用する場合、個々の家族員の良質な顔画像を端末３００に登録するため、まずその端末３００に初期設定動作を実行させる。その初期設定動作では、その家に住んでいる個々の家族員（ユーザ１０）の顔画像を端末３００に撮影させ、次いでその顔画像を所定の解析アルゴリズムに従い処理させる。この解析は、その顔画像に写っているユーザ１０の姿勢及び照明状態がロバスト且つリアルタイムな個人識別に耐えうる程度に良好か否かを調べる解析である。例えば特許文献３４（発明者：Shoupu et al.、譲受人：本願出願人、名称：顔面起伏位置特定方法及びそのためのコンピュータプログラム(Method and Computer Program Product for Locating Facial Features)）に記載の顔認識アルゴリズムはこの処理に利用することができる。この初期設定動作では、こうして各人毎に１枚又は複数枚の高品質顔画像を生成し、それらの画像を基準画像４６０（図５Ａ参照）として保存する。また、ローカルユーザ１０aは、その画像４６０に写っている人物の身元データ（例えば名前）をテキスト入力や音声認識で端末３００に入力することができる。ユーザ１０aは、更に、同居はしていないが自分達にとっては近親者で時々往來のある人物の身元を端末３００に登録すること、例えば別宅に住んでいて時折来訪する祖父母達の身元を登録することができる。再来者に対してはコンテキスト伝達インタフェース４５０でも自動的にラベルを付して追跡を行う。人物名付画像データ、例えばスキャン画像データを入力してもよい。身元や続柄（家族か友人か等）を表すこうした身元データは、入力後はインタフェース４５０によって管理されることとなる。

#### 【００６３】

これらのユーザ身元データは、端末３００に実装されている種々のプライバシー乃至コンテキスト関連機能で活用できる。まず、この端末３００には、パスワード、音声認識、顔認識等のアクセス制御手段を用いユーザ１０を認証する機能がある。これは、プライバシー保護インタフェース４００上のプライバシー保護セッティングを改変する権限のある筆頭ユーザや、プライバシー保護セッティングにアクセスする権限はないがプライバシー保護セッティング及び動作オプション制限の許に端末を利用することはできる他種ユーザを認証する機能である。端末３００に動画撮影や送信を行わせる権限を筆頭ユーザだけに認め他種ユーザには端末３００へのアクセス自体を認めないようにすることもできる。端末３００を家族員で共用する場合、家族員のなかでも年長者が筆頭ユーザとされ、同時に前掲の一次

10

20

30

40

50

ユーザ（初期ユーザ）ともされることが多いであろう。コンテキスト伝達インタフェース 450では、プライバシー保護インタフェース400と連携しつつユーザ識別手順470を実行させることで、撮影視野420内の人物のうち誰がプライバシー保護セッティング改変権限を持ち誰が持たないのかを把握する。また、端末300はユーザ名及びパスワードの入力も受け付けるので、手順470では筆頭ユーザと認識され得ない人物が筆頭ユーザ名及びそのパスワードを入力し筆頭ユーザとして認証を受けること、例えば筆頭ユーザからプライバシー保護セッティング改変権限を一時的に譲り受けることもある。更に、筆頭ユーザは、インタフェース400を用い人物毎にプライバシー保護セッティングを制定することができる。インタフェース450では、動画通信イベント600の開催中に手順470に従い既知の人物を捜し、認識できた人物を相応のプライバシー保護セッティングに対応付ける。そのプライバシー保護セッティングには、その人物が使用できる視野、その人物に連絡できるアドレス乃至電話番号、記録オプション等の事項に関する制限規定を含めることができる。

10

#### 【0064】

ユーザ身元データは、更に、動画通信イベント600に自動モードで参加中のローカル環境415にて、コンテキスト伝達インタフェース450によるイベント解釈にも使用される。例えば、WFOV420の辺縁部等、WFOVカメラ120の視野(WFOV)420内だがNFOV420外の位置に既知の被写体乃至ユーザ10が長々と居座っていることをユーザ識別手順470の実行を通じて認識すると、インタフェース450は、ユーザ身元データを動画撮影優先順位決定用階層ツールとして有効活用しながら、その人物がローカルユーザ10aにとりどの程度重要な人物であるのかを判別する。インタフェース450は、その辺縁部居座りユーザ10が一次ユーザ（当初ユーザ）10にとり相応に重要な人物であると認めた場合、一次ユーザ10及び辺縁部居座りユーザ10のアクションが共に撮影されるよう視野420を拡張させる。辺縁部居座りユーザ10のアクションを捉え、ローカルディスプレイ110上に第2画面分割画像410'として表示させることや、一次ユーザ10及び辺縁部居座りユーザ10の画像を、画面分割画像としてリモートユーザ10bに届けることもできる。図6に、リモートサイト364でそのユーザ10b向けに表示画面115上に表示される画像の例を示す。この例では、そのサイト364内の環境を示す画像を画面分割画像410として表示する一方、画面115のより広い部分を利用し、他のサイト362内の環境を示す2個の画面分割画像410'及び410"（互いに異なる部位におけるアクションを捉えた画像）を表示している。サイト364でこうした表示が行われているとき、ローカルサイト362では、例えば、サイト364で表示中の画像410'及び410"に相当する画像か、WFOV420内の全てのアクションを捉えた画像を、画像410として表示させる。より一般的な表現でいえば、ローカルディスプレイ110の画面115に表示される複数個の画面分割画像、例えばローカル撮影コンテンツに係る画像410、サイト364側WFOV420内を捉えた画像410'、サイト364側NFOV420内を捉えた画像410"等をローカルユーザ10aは看取することができる。画面分割画像やピックアップ画像を表示させるのではなく、複数個のディスプレイを用い複数個の画像を表示させてもよい。

20

30

#### 【0065】

ローカルユーザ10aは、当然のことながら、辺縁部居座りユーザ10を撮影しその画像をリモートサイト364に送信する機能をオーバーライドすることができる。また、辺縁部居座りユーザ10の姿を撮影、送信し画面分割表示させる機能を、手動モードで稼働中の端末300ですることもできる。更に、リモートユーザ10bに見せる画面分割画像にローカル環境415内の様々な（恐らくは互いに無関係な）アクションが写っている場合に、ユーザ10bがそれらのうち任意のアクションに係る音を聞き取れるようにすることができる。例えば、電子イメージング装置100の音響システム部分を構成する指向性マイクロホン144及び指向性スピーカ125を使用しユーザ10bによる聞き分けを容易にすることや、ミュート機能を使用して不要な音声信号を抑圧し聞き取りたい音声を残すことや、音声認識ソフトウェアを音声信号に適用して話者音声を解釈しテキストを生

40

50

成することができる。生成したテキストは、ユーザ 10b 向けに字幕 412 (図 6 参照) 又は話者ラベルとして表示させることができる。

#### 【0066】

コンテキスト伝達インタフェース 450 では、更に、ローカルユーザ 10a とリモートユーザ 10b の関係性 / 親密性を示すデータを利用する。例えば、顔画像データ、音声認識、リモートユーザ位置 (電話番号や IP アドレスの類) 等といった含意データ (semantic data) を利用することで、ユーザ 10b のなかにユーザ 10a にとり親密な人物がいた場合にその人物を認識する。即ち、インタフェース 450 は、ユーザ 10a にとり親、祖父母、友人、家族等に該当するユーザ 10b を認識する。インタフェース 450 は、認識したユーザ 10b の興味関心について認識結果に基づき判別する。例えば、そのユーザ 10b がユーザ 10a の親であるのなら、ユーザ 10a がいる主要部とは別の場所 (但し同じローカル環境 415 内) にいる子供又は子供達 (そのユーザ 10b から見て孫) がそこで何をしているのか、そのユーザ 10b が関心を抱くものと判別する。インタフェース 450 は、こうした判別の結果に基づきローカル端末 362 及びリモート端末 364 に要求し、ローカルサイト 362 及びリモートサイト 364 双方のディスプレイ 110 に、ローカル環境 415 内主要部の画像と子供所在部の画像を、共に画面分割画像として表示させる。その際には、ローカル端末 300 のプライバシー保護インタフェース 400 によって保持されているプライバシー保護セッティングも適用されるが、その形態はユーザ 10b の身元によって変わってくる。例えば、ユーザ 10b がユーザ 10a の親であると認識された場合、ユーザ 10a の子供はユーザ 10b の孫であるから、脱衣した子供の裸を隠すように定めたプライバシー保護セッティングを適用する必要性は低い。また、動画通信イベント 600 の途中でユーザ 10b 又はその顔ぶれに変化が生じたことを例えば顔認識により認識した場合、ローカル端末 300 は、適用するプライバシー保護セッティングを自動的に切り替える。プライバシー保護セッティングの切替方は状況によって変えるので、速やかな場合も、ゆっくりな場合も、遅らせる場合もあり得る。インタフェース 450 は、更に、受信データ例えばサイト 364 発の動画画像データを、サイト 362 のプライバシー保護セッティングに合致するよう編集する。この編集で、例えば裸体が隠され或いは罰当たりな言葉が削除されるので、ユーザ 10a の意志に反しローカル環境 415 内にそれらが侵入することを防ぐことができる。

#### 【0067】

コンテキスト伝達インタフェース 450 では、また、前述した幾通りかの動画撮影モードを提供する。ユーザ 10 がそれらの動画撮影モードのうちいずれかを指定すると端末 300 がその動画撮影モードで稼働するので、ユーザ 10 は自分達のアクションを撮影しリモートサイト 364 に送信する動作をそのモードで行わせることができる。動画撮影モードとしては WFOV 撮影専用モードのほかに追跡機能のあるモード群が準備されている。後者は、端末 300 に備わるズーム、パン及びティルト機能を活かし、そのローカル環境 415 で人々がアクションを起こしている場所の特定や、そのアクションを端末 300 に追跡させるモード群であり、ユーザ 10 の多くはこちらの方を好むであろう。コンテキスト伝達インタフェース 450 が重要であるのは、人物やそのアクションの重要性をローカル環境 415 との関係で認識できるためである。インタフェース 450 には、更に、その WFOV 420 内を監視しその環境 415 内で何が行われているのかを調べさせる動作と、その WFOV 420 とは大分違う視野 420 内を撮影させる動作とを、同時並行的に実行できる、という重要な特徴がある。しかも、人物やそのアクションの変転 (移り変わり) 等に応じそれらの動作を自動実行させることができる。また、この端末 300 で使用できる動画撮影モードには前述の通り幾通りかのモードがある。それらを表 1 に示す。この表では、自動モードに幾通りかの種類があることや、それらの間に定義の違いがあることも記されている。その違いの基礎となる動画通信のコンテキスト及びコンテンツとしては、1) そのユーザの種別、2) そのユーザの身元、3) ユーザに対する照準追跡の結果、4) その環境乃至領域内で人々がアクションを起こしている場所及びそのアクションの持続時間、5) イベント種別等が示されている。例えば、ユーザ種別データやユーザ身元デ

10

20

30

40

50

ータに基づき優先順位を付けて動画像撮影を行う際には、前述の通り階層的な手法で動画像のコンテンツを選別することができる。また、ユーザに対する照準追跡を行いつつ動画像を撮影する際には、ユーザ身元データに基づき又はイベントをより重視し、例えばピエロの追跡等を行うことができる。こうした自動モードでは、アクションそのものではなくアクションの場所の認識結果に基づき、或いはジェスチャ、ユーザ音声キュー、ユーザ挙動等についての認識結果に基づき、動画像撮影を行って動画像の撮影／処理動作に施すべき修正の内容を決めることができる。更に、動画撮影モード間の定義の違いは、そのもとになるイベントの種別によってももたらされうる。例えば大勢でパーティを行っている場面で使用される動画撮影セッティングは、相談事があって個人的に通話している場面で使用される動画撮影セッティングとは、大きく異なるものになる。

【 0 0 6 8 】

【 表 1 】

表1 … 動画撮影モード
<p>動画撮影モードは、種々の撮影動作属性について規定する一組のセッティングによって定義される。定義に使用される属性としては、WFOVカメラの動作、NFOVカメラの動作、パン、ティルト及びズーム機能の動作、プライバシー保護セッティング、音響セッティング及び処理の適用、動画像処理の利用、場面内変転及び場面間変転に対する応答(タイミングや大きさ)等に関わる属性がある。</p>
<ul style="list-style-type: none"> <li>・ 手動モード</li> <li>・ プレビューモード</li> <li>・ デフォルトモード群 … WFOV撮影専用、採音専用、初期撮影専用等</li> <li>・ 自動モード群 … コンテキスト示唆キュー、例えばユーザ種別、ユーザ身元、ユーザジェスチャ、画像／音声キュー、ユーザ挙動、イベント種別等のキューに依拠</li> <li>・ 中間即ち半自動モード群 … 手動モードと自動モードの複合物</li> <li>・ ポータブルモード … アウトドアモードもその一種</li> </ul>

【 0 0 6 9 】

使用される動画撮影モードは、動画通信イベント 6 0 0 毎に切り替わることもあれば、同じイベント 6 0 0 内で切り替わることもある。その切り替わりが端末 3 0 0 の自動動作で引き起こされることもあれば、ユーザ 1 0 からの指示で行われることもある。例えば、ユーザ用コントローラ 1 9 0 を操作することで、ユーザ 1 0 は、コンテキスト示唆キュー(contextual cue)を利用しイベント相応の撮影及び変転管理を行うモードや、端末 3 0 0 の移動に順応しつつ撮影及び変転管理を行うポータブルモードや、照明条件等の環境条件を認識しそれに順応するアウトドアモード等といった動画撮影モードを指示することができる。これらの動画撮影モードには、撮影視野 4 2 0 内にいる人物やそのアクションに反応する機能だけでなく、視野 4 2 0 内にいる人物又はそのアクションを解釈及び予測する機能もある。コンテキスト伝達インタフェース 4 5 0 では、コンテキスト及びプライバシーとの関連でそうした解釈予測機能を担うべく様々なやり方で含意データを利用し、今の動画像コンテキストに代わる新たな動画像コンテキストや、それに相応する動画撮影モードを特定する。なお、含意データとは人知及び体験に関わるデータ、例えばイベント、人物のアクション、人物の異同、場所、被写体、音楽ジャンル等に関するデータである。含意データに関連付けうる情報資源としては、画像レコード、音声レコード、データファイル等がある。

【 0 0 7 0 】

そのコンテキスト伝達インタフェース 4 5 0 は、これまで詳細に説明してきた画像認識／解釈関連の諸動作のほかに、音声に関しても重要な動作を実行する。前述した音声認識

に関わる動作、音声コマンド及びキューを利用した動作、採音視野 430 を監視する動作、音響を利用した動き追跡及び測位動作等々である。これにまつわる問題は数多いが、そのなかでも問題なのは音質である。音質が低いと、良質な音響的ユーザエクスペリエンスを端末 300 からリモートユーザ 10b にもたすことができないので、音響システムには幾つかの工夫が施されている。まず、ユーザ 10 は、相応のユーザ用コントローラ 190 を操作することで、音響システムのトーン感度を修正（バス/トレブル調整）することができる。また、端末 300 に複数個備わる前述のマイクロホンのうち指向性が鋭い 1 個又は複数個を利用することで、雑音源の所在を突き止めることや音源乃至音声源の位置を個別に認識することができる。更に、オーディオプロセッサ 325 に相応の処理を実行させることにより、音響システムにノイズキャンセル機能やノイズリダクション機能を持たせ、音響的ユーザエクスペリエンスを向上させることができる。また、ユーザ 10b による聞き取りを邪魔する音響、例えばローカル環境 415 の周囲で発生する車両騒音、犬の吠え声等の環境音や、電子雑音等の雑音を減らすことが重要であるが、流れている音楽等の環境音をユーザ 10b に聞かせたい場合もあるので、音質調整用のユーザ用コントローラ 190 も設けてある。そして、インタフェース 450 では、拍手の音、笑い声、鳴き声、犬の吠え声、猫の鳴き声等が登録された非音声音響キューデータベース乃至ライブラリも使用する。このデータベース乃至ライブラリには、汎用性のある音響キューだけでなく、その端末 300 が使用される居宅、使用する人物又は使用される場所に固有の音響キューも登録しておくことができる。従って、動物達の基準画像 460 や基準音声を作成し、その端末 300 に登録しておくことができる。

10

20

#### 【0071】

それら、音響キューの利用がとりわけ役に立つのは、そのコンテキスト伝達インタフェース 450 でペットを認識する際である。これは、ペットを外観で認識することができない場合があるからである。即ち、大抵のペットは床の近くにいたので、その姿がカメラ 120 の撮影視野 420 から外れてしまう。端末 300 によるペットの認識及び追跡に音響キューを用いることで、そのペットが写るよう、写らないよう或いは目立たないよう、動画像を撮影して送信することができる。

#### 【0072】

また、居宅向けの動画通信システム 290 を構築する上で重要なのは、プライバシー保護プロトコル、パスワード認証、録画/録音制御、暗号化等の電子的手段と、撮影視野制御、撮影時警告、身元確認、美観配慮、イベント対応等のコンテキスト的手段とを併用して、ユーザ 10 のプライバシーを十分に管理することである。表 2 に、システム 290 を構成する個々の端末 300、特にそのプライバシー保護インタフェース 400 及びコンテキスト伝達インタフェース 450 に備わるべき主要属性のあらましを示す。

30

#### 【0073】

【表 2】

表2	
プライバシー保護インタフェースの機能及び プライバシー保護セッティング:	コンテキスト伝達インタフェースの機能及び コンテキスト伝達セッティング:
<ul style="list-style-type: none"> <li>・プライバシー言語プロトコル…動画像送信開始、ネットワーク通信暗号化／解読、録画／録音制御、多者間通話管制、伝送障害通知等を担当</li> <li>・撮影インジケータ…音調変化、可視表示、シャッタ開放等がこれに該当</li> <li>・画面分割画像の表示…ピクチャインピクチャ画像等</li> <li>・撮影視野を制限又は修正…カメラ選択、カメラ制御、撮影後画像処理等で実行</li> <li>・リモートユーザからの指示で撮影視野限定</li> <li>・プライバシー保護上重要な身体部分に関しプライバシー上のセッティング</li> <li>・プライバシー保護セッティングの参照及び改変を制限…パスワード等による</li> </ul>	<ul style="list-style-type: none"> <li>・場面解析／調整アルゴリズム(及びルール)</li> <li>・ユーザ指定に基づく撮影視野の制限又は修正を管理 … クロッピング、ブラーリング、シェーディング等で実行</li> <li>・部位又は人物に係るプライバシーが保護されるよう、画像を能動的に修正するタイミングを決定…例えば裸体が隠されるよう視野制限、ブラーリング又はクロッピング</li> <li>・ローカル環境内視野サイズ／位置を修正して人物を追跡</li> <li>・ローカル環境内視野サイズ／位置を修正して複数人を同時撮影</li> <li>・キュー例えばジェスチャを解釈</li> <li>・イベント種別を認識し撮影及び送信動作を然るべく順応</li> <li>・環境内にいるユーザを認識しその種別(大人か子供か等)を特定</li> <li>・環境内ユーザの身元を認識</li> <li>・動物の存否を認識</li> <li>・身元、続柄、イベント種別等のコンテキスト伝達データを保持</li> <li>・ローカルユーザと対応するプライバシー保護セッティングアクセス又はプライバシー保護セッティングに関連付け</li> <li>・ローカルユーザ及びリモートユーザを対応するローカル環境の動画像コンテキストに関連付け…身元及び続柄の重要性</li> <li>・ユーザ指示に従い動画像コンテキストを修正</li> <li>・受信した動画像に従い動画像コンテキストを修正</li> </ul>

10

20

30

## 【 0 0 7 4 】

40

但し、表 2 に示したプライバシー乃至コンテキスト関連の諸機能及び諸セッティングを動画通信システム 2 9 0 及びその端末 3 0 0 で漏れなく使用することも、それらを全て搭載・制定することも必要でない。例えばその端末 3 0 0 の動画撮影モードが手動モードになっている状態では、コンテキスト伝達セッティングやコンテキスト伝達インタフェース 4 5 0 に備わる機能の多くはほとんど必要ない。従って、コンテキスト管理用の機構を然るべく削減することで、ユーザインタフェース 4 4 0 を全体的に簡素化することができる。例えば、コンテキスト伝達インタフェース 4 5 0 の機能を、撮影視野 4 2 0 及び修正対象部分 4 2 2 を利用し身体プライバシー保護機能に絞ってもよい。その逆に、人物(達)に対する視野 4 2 0 のサイズ及び位置をインタフェース 4 5 0 で自動的に決める機能を付加することにより、その動作の自動性を高めることもできる。同様に、可解釈な一群の含意キ

50



ユー、例えば人間のアクション（笑い・拍手・踊り・ジャンプ等）、イベント（パーティ・夕食会・遊技会等）或いは付き添い（動物等）を認識し、それに応じた動作を実行させる機能を追加することでも、その動作の自動性を高めることができる。更に、ローカルユーザ10a又はリモートユーザ10bにまつわる人物の身元及び続柄に関わる含意データを、プライバシー又はコンテキストに関わるデータに基づき保存及び適用することで、インタフェース450に価値ある機能を追加することができる。但し、その実行には多くの個人情報が必要になる。

#### 【0075】

ユーザインタフェース440は、これらプライバシー保護インタフェース400及びコンテキスト伝達インタフェース450と、それらから基本的に独立している前述したオンオフスイッチ、出力音量コントローラ、表示輝度コントローラ等のユーザ用コントローラ190と、を有している。それらのユーザインタフェース構成要素中で重きをなすのはやはりインタフェース400及び450である。その構成要素としては、まず、種々のセッティング及び制御オプションや、グラフィカルユーザインタフェース、メニュー型ユーザインタフェース等の直接可操作インタフェースがある。ユーザ10は、インタフェース400, 450が銘々に提供する直接可操作インタフェースを使用し、それらのセッティング及び制御オプションにアクセスすることができる。各インタフェース400, 450の構成要素としては、更に、種々の内部インタフェース又は制御セッティング、助力となる種々のデータベース、並びに端末300の動作や動画通信システム290内の端末300間での通信動作を規定する種々のアルゴリズムがある。ユーザ10がセッティング及び制御オプションに関する入力を行うと、その内容が内部インタフェースや制御セッティング、更にはそれに関連する内部的なアルゴリズムに反映するので、そのアルゴリズムの適用により、そのユーザ入力が端末300やシステム290の動作に反映することとなる。また、使用されている動画撮影モードが手動モードであれ自動モードであれ、ユーザインタフェース440は、動画通信イベント600の開始又はそれへの参加を求めるユーザ10からの指示に対し遅滞なく、即ち種々のオプションで極端に煩わせることなく対処しなければならない。この条件を満たすには、例えば両インタフェース400, 450を二階層構造、即ち過渡的な性格が強いセッティングからなる階層及び安定的な性格が強い（例えばデフォルト版の）セッティングからなる別の階層からなる構造にすればよい。更に、インタフェース440には、一群の稼働形態のうちいずれかが指定されたときに、その稼働形態に対応する動画撮影モードでの動作を、所定内容のプライバシー保護セッティング及びコンテキスト伝達セッティングに従い即座に実行させる機能がある。ユーザ10は、そのイベント600に相応しい稼働形態、例えば「夕食会」「祖父母」「ボードゲーム」「台所」「仕事場」「友人達」等のうちいずれかを指定するだけでよい。端末300に種々の稼働形態を設定する操作は、ユーザ10が行ってもよいし、ユーザ10以外の人物例えばシステム製造者が行ってもよい。

#### 【0076】

以上、端末300並びにその主要構成要素であるプライバシー保護インタフェース400及びコンテキスト伝達インタフェース450に関し、人間同士の個人的なやりとりや個人的な関係への関わり方に注視し定性的に説明した。次に、端末300やそのインタフェース400, 450の動作について理解を深めるため、動画通信イベント600における動作の流れを、一群のフローチャートを参照して説明することとする。まず、図7Aに、イベント600への参加に際し端末300で順を追って実行される処理全体のあらましを示す。図示の通り、通常のイベント600ではまず招待605及び応諾610の手順を実行する。それにより、図1に示した通り、ローカル端末300・リモート端末300間が通信ネットワーク360経由で接続される。招待605及び応諾610の手順は、システムの観点からすると、ネットワークプロトコルに則り接続先アドレスの探索及び特定並びに話者間接続を実行させる手順である。この手順では、通話に参加する2個の端末300間を接続するため電子的ハンドシェイク手順365が実行される。その手順365では、プライバシー言語プロトコルに則り、通話に参加する2個の端末300でネットワーク360

越しにプライバシー保護セッティングを交換させる。これは、（音声及び）動画像の送信、暗号化及び暗号解読、録画／録音等といったプライバシー関連機能に関する認証ステータスの交換によって行う。また、ユーザ 10 に対しては、招待 605 及び応諾 610 が実行されているときに、ユーザインタフェース 440 によって発着呼手段が提供される。ユーザ 10 は、その発着呼手段に組み込まれているアドレス帳、発呼者識別、着呼／接続警報等の機能を利用することができる。ユーザ 10 は、更に、招待 605 及び応諾 610 で使用される初期的なプライバシー保護セッティング又はコンテキスト伝達セッティングを、インタフェース 440 を使用して指定乃至制定することができる。インタフェース 400, 450 はこの動作にも随時関与してくる。また、この初期的な指定乃至制定に際しては、レビューモードを使用することができる。そのレビューモードで稼働している間に、動画像及び音声の採取動作やプライバシー保護セッティングを修正することができる。そして、招待 605 及び応諾 610 の手順が終わったら、通常は動画像伝送 615 が始まる。招待 605 及び応諾 610 の終了から伝送 615 の開始までに間がないこともあれば間が空くこともある。そのいずれになるかは、ユーザ 10 が初期音声ミュート機能や動画像ランキング機能を使用しているか否かで決まる。

10

#### 【0077】

動画像伝送 615 が始まると、動画像中の最初の場面 620 で動画像による通話が始まる。このとき、個々の環境（415 等）に設けられているカメラ（群）120 は、その環境の初期的な動画像コンテキストに従い稼働される。動画像及び音声の採取及び処理に関するセッティングのため初期的に使用される動画像コンテキスト（例えばイベント種別）やそれに対応する動画撮影モードは、ユーザ 10 が制定したものでよいし、端末 300 で保持している（例えばデフォルト版の）ものでよい。最初の場面 620 は、ローカルユーザ 10a がいる環境 415 及びリモートユーザ 10b がいる環境それぞれを捉えたものになるのが普通である。そのまま場面 620 に変転が生じないで動画通信イベント 600 が伝送終了 617 に至ることもあれば、成り行きに応じた変転が生じることもある。変転の発生に対処するため、端末 300 では同時並行的に変転テスト 630 を実行する。特に、自動モードで動画像撮影を行う場合にこれが役に立つ。

20

#### 【0078】

本願で提案する手法では、発生しうる変転、例えばユーザ 10 のアクションが変化しイベント種別等に係る動画撮影セッティングが不相応になるといった前掲の変転を、場面内変転と場面間変転に区分することで、それらの変転に好適に対処している。例えば図 4C に示す動画通信イベント 600 で生じている何種類かのコンテンツ的及びコンテキスト的変転のうち、場面 620 における変転はローカルユーザ 10a の小さな動き、例えば姿勢の変更等に伴うものである。こうした変転に対しては、動画像を撮影する動作や撮影された動画像の処理を小幅に修正すること、例えば焦点、コントラスト、視線、視野等に関わる小規模な修正を施すことで対処することができる。場面内変転とは、このようにユーザ 10 のアクションにおける小規模な変化に伴い生じる変転であり、その動画通信イベント又は場面に係る動画像コンテキストが大きく変化しないため、動画撮影セッティングに大きな修正を施さなくても対処することができる。即ち、場面内変転を示唆する変化は、開催中のイベント 600 又は場面 620 の性質に影響しない程度に小規模な、ユーザ 10 のアクションにおける変化である。どの程度のものを「小規模」な変化と呼べるかはコンテキストによって違い、例えば図 4C 中の場面 620 のように有目的な（即ちはっきりとした通話目的のある）場面での「小規模」と、図 2 に示した場面のように無目的な（即ちはっきりとした通話目的がなく寛いだ）場面での「小規模」との間には、かなりの違いがある。また、環境 415 又はその撮影視野 420 内の諸部分で発生し場面内変転をもたらす小規模な変化のうち、場面の構図又は音響に関わる変化は、補正しないと動画通信の質乃至通信インパクトが損なわれることが多いが、新たにやってきた人物をカメラ 120 のパンニングにより撮影視野 420 内に取り込む等といった動画像撮影上の変化は、単純に画質損失乃至音質損失を補償するのに比べたらリモートユーザ 10b にとり有益なものである。

30

40

50

## 【 0 0 7 9 】

これに対し、場面間変転はその動画通信イベントの性質に関わる変転であり、動画像コンテキストが従前のものから新たなものへと変化するため、動画撮影セッティングを大きく変化させないと対処することができない。例えば図 4 C では、ローカルユーザ 1 0 a のいる場面が第 1 の場面 6 2 0 から第 2 の場面 6 2 0 ' へと切り替わる過程で、そのユーザ 1 0 a のアクションが有目的なものから無目的なものへと変化していく。動画像コンテキストにこのような移り変わりが生じるときには、動画撮影セッティングを大きく変化させる必要がある。具体的には、有目的な場面 6 2 0 で使用していたセッティング（緻密な視線補正でアイコンタクトを成立させるセッティング）から、無目的な場面 6 2 0 ' に相応しい別のセッティング（撮影視野 4 2 0 は広めにまた視線補正は間欠実行にし必要に応じ照準追尾による撮影を実行するセッティング）へと切り替えていく必要がある。このように、そのイベント 6 0 0 で生じているコンテンツ及びアクションの変化が、従前の動画像コンテキスト又は対応する動画撮影モードに照らし動画像コンテキストの交替といえるほどに顕著であるので、図 4 C における場面 6 2 0 から場面 6 2 0 ' への移り変わりは場面間変転を表しているといえる。

## 【 0 0 8 0 】

変転テスト 6 3 0 は、大まかにいえば、送受信中の動画像データストリームや撮影視野 4 2 0 内送信対象外部分に発する動画像及び音声を監視することで、例えば今の場面 6 2 0 から次の場面 6 2 0 ' へ移行する際の動画像コンテキストの交替による変転等といった場面間変転を察知する処理である。このテスト 6 3 0 を実行することで、例えば、動画像コンテキストの交替による次の場面 6 2 0 ' への場面間変転が生じつつあることを、好適に察知することができる。このテスト 6 3 0 は、アクションの範囲、アクションの速度、撮影視野 4 2 0、自信度等に関わる指標や、コンテキスト示唆キュー、従前の動画像コンテキスト（又はそこからの乖離）等に基づき、従前のデータとの関連で今の場面に係る動画像を調べること、行うことができる。更に、このテスト 6 3 0 で場面間変転示唆と觉しき変化が察知された場合は変転処理 6 4 0 が実行される。その処理 6 4 0 では、その変転の性質に応じた応答の仕方であって端末 3 0 0 の動作を変化させる。逆に、変転テスト 6 3 0 で調べた場面の变化が局所的で小規模なものである場合は場面撮影管理 6 5 0 が実行される。

## 【 0 0 8 1 】

これら、図 7 A に示した変転テスト 6 3 0、変転処理 6 4 0 及び場面撮影管理 6 5 0 は、図 7 B に示す通信イベント解析 (communication event analysis) 6 5 5 と共に、前述したコンテキスト伝達インタフェース 4 5 0 によって実行される。そのインタフェース 4 5 0 は、見方によっては、ユーザ 1 0 のアクションに対するこうした順応的乃至予測的応答を一群のアルゴリズムに従い実行し、画像的ユーザエクスペリエンスが高まるように動画撮影手順を修正する知的エージェント乃至人工知能 (AI) である。インタフェース 4 5 0 でこうしたシステム知性が提供されることは、ステップ 6 3 0 ~ 6 5 5 における処理内容からも窺える。ユーザ 1 0 のアクションや通信ニーズを徐々に理解していくので、インタフェース 4 5 0 は学習システムであるともいえる。インタフェース 4 5 0 では、表 3 に列挙したシステムデータ及びアルゴリズムを使用し図 7 A に示した手順を実行し、動画通信イベント 6 0 0 に参加する。図 7 B 及び図 7 C は、それらのうち変転処理 6 4 0 及び場面撮影管理 6 5 0 に関わる動作をより詳細に示したものである。図 7 B に示すように、端末 3 0 0 は同じ場面 6 2 0 に関する通信イベント解析 6 5 5 及びテスト 6 3 0 を同時並行的に実行することができる。解析 6 5 5 では場面内変転を調べテスト 6 3 0 では場面間変転を調べる、という違いはあるものの、いずれも同じ場面解析アルゴリズムに則った処理であり、動画像（及び音声）データストリームから直接導出されたデータを解析してその場面に生じたコンテンツ的又はコンテキスト的化の顕著性を識別する点で共通している。場面コンテンツ指標、場面解析ルール、コンテキスト示唆キュー、統計解析結果等を併用するので、その解析手法は多変数的である。管理 6 5 0 及び処理 6 4 0（特に図 7 C 中のステップ 6 4 4）も、共に撮影形態修正アルゴリズム (video capture scene adjustment

algorithm)に則った処理であり、今適用中の動画撮影セッティングを調整して進行中の動画撮影動作を修正する点で共通している。図7Dに詳示する通り、図7Aに示した手順に従い動画通信イベント600を開催する際には、動画通信システム290内の2端末300間で、通信ネットワーク360上の動画通信リンクを介したデータ交換やデータ伝送（動画像や音声信号の伝送等）等のやりとりが実行される。

【0082】

【表 3】

表3ーコンテキスト伝達インタフェース450用の主な手順及びデータ	
システムデータの種類:	
・ アルゴリズム的ルール…イベント種別、プライバシー保護セッティング、期待タイミング(変化の速度・頻度)等の情報を利用する。	
・ アルゴリズム的定量指標…しきい値、加重係数、確率値、場面内／場面間変転自信度情報等がこれに該当。	
・ コンテンツ指標(撮影条件)…視野(FOV)の広さ及び方向、画質(焦点解像度・発色・輝度・コントラスト等)、音質等がこれに該当。	10
・ コンテンツ指標(画像)…表面凹凸を示唆する種々の画像パターン及びデータ、アクションの場所及び持続時間を示唆する画像データ、動きの速度、視線等がこれに該当。	
・ 動画通信コンテキスト(動画画像コンテキスト)…それぞれ自動の動画撮影モードに関連付けうるコンテキスト又はコンテンツベースの分類。その基礎となる情報には、アクションの場所及び持続時間の拡がり(動き検知域)を示す情報や、大集団(パーティ・夕食・ゲーム)／小集団／個人の別、無目的的／有目的の別、過渡性の有無、場所移動性の有無、照準追跡の要否、(ユーザ種別／ユーザ身元に基づく)階層の別等を示すイベント種別データがある。	
・ キュー／含意情報サポート性コンテキスト … 人々の属性、身元、続柄、ジェスチャ(指さし・身震い・眼差し等)、身動き(笑い・拍手・踊り・跳躍等)、プライバシー示唆キュー、物体やペットに関する知識、暦やスケジュールに関する知識等がこれに該当。	20
・ 基準画像460	
表3(続き)ーコンテキスト伝達インタフェース450用の主な手順及びデータの続き	
今進行中の通信イベントに関して行われる処理(図7B参照):	
・ 通信イベント解析655…今進行中の通信イベント600又は場面620で得られる動画画像及び音声場面内解析アルゴリズムに則り且つ種々の場面解析ルール及び場面コンテンツ指標に従い調べること、コンテンツ的又はコンテキスト的な場面内変転を察知し、動画(及び音響)撮影(及び処理)セッティングに対する小規模な修正、ひいてはユーザエクスペリエンス向上に役立てる。	
・ 場面撮影管理650…通信イベント解析の結果を受けて、撮影形態修正アルゴリズムに則り且つ種々の場面内調整ルールに従い、現場面で進行中の動画撮影及び処理動作にどのような場面内調整をどのように施すべきかを、タイミング、規模等の要素について判別する。	
変転テスト630(図7B及び図7C参照):	30
・ 変転テストのアルゴリズム…今進行中の動画通信イベント600又は場面620から得られる動画画像及び音声を場面間解析アルゴリズムに則り且つ種々の場面解析ルール及び場面コンテンツに従い調べること、コンテンツ的又はコンテキスト的な場面間変転を察知する。新たな場面では、ユーザのアクションに現れた大規模な変化に順応するよう、使用する動画撮影モードをそれまでのものから変化させる。	
変転処理640(図7C参照):	
・ 変転処理…新たな動画画像コンテキストを識別して適切な動画撮影セッティングを求め、その結果について自信度情報を生成し、有効性テストを実施してからそれらのセッティングを適用し、必要なら暫定版の動画撮影セッティングを適用する、という一群のステップからなる組織的な処理。	
・ 動画画像コンテキスト(イベント種別)特定641のアルゴリズム。	
・ 新規動画撮影モード選定642のアルゴリズム。	40
・ 有効性評価643のアルゴリズム。	
・ 暫定セッティング適用645のアルゴリズム … 暫定版の動画撮影セッティングを適用。	
・ 撮影形態修正644のアルゴリズム … それまでの動画画像コンテキストから新たな動画画像コンテキストへの移行のため動画撮影セッティングを修正。	
外見修正インタフェース490:	
・ 視線順応手順495や視線方向自信度情報。	

【0083】

それらのうち通信イベント解析655では、動画画像中の個々の場面620から得た画像を調べて場面内変転を察知する。その結果に応じ動画撮影(及び動画画像処理;以下同様)

セッティングや音響（及び音声処理；以下同様）セッティング、ひいてはそれに基づく動作を調整することで、ユーザエクスペリエンスを高めることができる。具体的には、その動画像の新旧両場面のコンテンツ又はコンテキストに対し種々の場面解析ルールを適用し、プライバシー、撮影視野420、被写体の動き、人々のアクション、画質、音質、視線、コンテキスト、諸キュー（音声・ジェスチャ・情動等）等といった属性を解析655で調べることで、場面620内で生じる小規模なコンテンツ変化を探索する。この探索で見つかる変化は、それまでの動画像コンテキスト及びそれに対応する動画撮影モードと違背しない（或いは関連を保つ）性質の変化であるので、端末300の動画撮影セッティングを小幅に変化させることで対処することができる。例えば図4Cに示した場面620では、場面解析アルゴリズムに則った通信イベント解析655によって、着席しているローカルユーザ10aの身じろぎ等が検知される。それに対しては、焦点、視野420、輝度等を改善方向に修正し対処することができる。更に、解析655では、今の視野420から得られる動画像（及び音声）だけでなく、その視野420の外側から得られる動画像、例えばWFOVカメラ120で得られる動画像（図4A及び図4B参照）等も調べることができる。そして、解析655では、動画撮影セッティングに施すべき修正を特定する際に、その特定結果の信頼性を量る場面内変転自信度を求める。更に、この解析655の結果を受けて実行される場面撮影管理650では、その場面を撮影する動作に対しどのような調整をどのようにして施すべきかを判別し、今行われている動画撮影／処理動作に対し、コンピュータ340、イメージプロセッサ320等の部材を通じてその調整を適用させる。

10

20

30

40

50

#### 【0084】

対するに、変転テスト630では、新旧両場面から得られるデータ（動画像や音声）を場面解析アルゴリズムに則り且つ種々の場面解析ルール及び種々の場面指標に従い調べて場面間変転を察知する。具体的には、所与の動画像コンテキスト及び動画撮影モードが使用されている現在の場面620から別の場面620'へと場面が変転することを示唆する場面コンテンツを検知する。テスト630では、そのため、その時点でNF OV 420から得られるデータに加え、その外側、例えばWFOV 420や採音視野430から収集されたデータをも使用して、ユーザ10のアクションに現れている変化を調べる。その変化がそれまでの動画通信コンテキストに照らして大規模なものであった場合、テスト630では、今依拠しているイベント種別又は今使用している動画撮影モードがもはや動画撮影動作に適さないものになっている又はなりつつあると判断する。例えば、ローカルユーザ10aのアクションが、図4Cに示した如く、場面620相応のアクションから場面620'相応のアクションに変化したとする。このとき、テスト630では、場面間変転を示すと覚しき変化がユーザ10aのアクションに現れた、今の動画撮影セッティングではその変転に対処できそうにない、と判断する。こうして場面間変転と疑わしい場面コンテンツ変化が認識されると、それを受けて実行される変転処理640にて、その場面間変転の性質が子細に調べられ、今使用中の動画撮影セッティングに施すべき修正が決定され、その修正が適用される。

#### 【0085】

その変転処理640の機構は、図7Cに詳示する通り一連の変転処理ステップ641～645で形成されている。そのうちの最初のステップ641では、新旧両場面の動画像を動画像コンテキスト特定(video context selection)アルゴリズムに則り且つ統計指標等に従い調べることで、場面のコンテンツに場面を跨いで生じたコンテキスト的变化、ひいては新たな場面に相応しい動画像コンテキストの種別が特定される。例えば、人々のアクションがパーティでよくあるアクションからカードゲームに多いアクションへと切り替わりつつあること、有目的な通話からより無目的な通話に切り替わりつつあること（図4C参照）、子供達の集まりからその親達の集まりに切り替わりつつあること等を認識した場合には、その場面のコンテンツに係るイベントがそれまでとは別の種別に係るイベントに切り替わりつつあるということであるので、それまでの動画像コンテキストに代わるべき新たな動画像コンテキストの種別が特定されることになる。その次のステップ642では、こうして特定された新たな（イベント）種別に係る動画像コンテキストに対応する

新たな動画撮影モードが、動画撮影／音響セッティングへの修正内容を然るべく決定することによって選定される。動画撮影セッティングに施されうる修正としては、例えば視野の調整、プライバシー保護セッティングの適用、ユーザの動き・アクション・ジェスチャへの応答、画像の外見的修正（視線等）、画質の向上等がある。また、既に用意されている別の動画撮影セッティングと同じ内容の修正なら当該別の動画撮影セッティングを使用することができ、そうでない場合もその動画通信イベント600又は場面620に相応しいカスタム版の動画撮影セッティングを制定して使用することができる。三番目のステップ643では、場面間変転の認識結果、それに基づき特定された動画像コンテキスト種別、それに基づき選定された動画撮影モード、並びにそれに対応する動画撮影セッティング及び音響セッティングの有効性が（統計的に）確認される。肯定的な結果が得られた場合、即ち有効性が確認された場合には、その次のステップ644にて、動画像コンテキストが新たな動画像コンテキストに切り替わるよう動画撮影セッティングを変更して動画像の撮影を継続させる。逆に、否定的な結果が得られた場合、即ち有効性が確認できないとの結果になった場合、この端末300では、有効性確認との結果が得られるに至るまで、従前の動画撮影モードのままで場面撮影管理650が実行され、或いはステップ645に従い暫定的な動画撮影セッティングが適用され始める。

10

20

30

40

50

#### 【0086】

コンテキスト伝達インタフェース450がこの動作で選定しうる動画撮影モードとしては、ユーザ10からの指示に応じ動画撮影セッティングが制御される手動モードのほかに、良質な画像的ユーザエクスペリエンスがもたらされるようインタフェース450自身がその端末300を制御する幾種類かの自動モードがある。ユーザ10が良質な画像的ユーザエクスペリエンスを味わえるのは、まさに、個々の端末300が、個々の動画通信イベント600でそのローカル環境415に生じる変化に対し、間断なく且つ円滑に応答して動画像の画質を高めるように構成されているからである。そのためには、端末300が動画撮影セッティングに修正を施すタイミング及び頻度が適切であること、即ちユーザ10を煩わせず画像的ユーザエクスペリエンスも良質になりうるタイミング及び頻度であることが求められる。次に、図7A～図7Cを参照して概述済の動作のうち、解析乃至評価に関わる手順乃至ステップについて、より詳細に説明する。

#### 【0087】

まず、図7Aに通信イベント解析655を描くとしたら変転テスト630の後であるが、実際にはそれら二種類の場面解析手順を図7Bの如く同時並行的に実行することができる。両者は共に今の場面におけるユーザ10のアクションを調べる手順であるが、調べる対象が異なっている。即ち、解析655で調べるのはそのアクションが今の動画撮影モードやそれに対応する動画像コンテキスト乃至イベントの種別に整合するか否かであり、テスト630で調べるのはそのアクションが場面間変転の進行又はその前段階を示しているか否かである。その変転後の場面がそれまでとは別の動画像コンテキストで表すべきもの、即ち別の動画撮影モードで撮影すべきものであるかもしれないので、テスト630の後には新たな動画像コンテキストを識別する処理を実行する。この処理では、ユーザの種別、身元、ジェスチャ、挙動（笑い・跳躍等）乃至音声キュー、プライバシー保護セッティング、含意的暦データ、被写体、被服等といった種々のコンテキスト示唆キューや、アクションの場所及び持続時間、照準追跡の要否、大集団／小集団／個人の別、無目的／有目的の別、過渡性の有無、場所移動性の有無等に基づく種々のコンテキスト依存型イベント種別データに対し、場面620又は動画通信イベント600を分類するためのルール群を適用することで、その新たな動画像コンテキストの種別を特定する。新たな動画像コンテキストの種別を特定したら、それに相応する動画撮影モードを選定することで、動画撮影セッティングに対する修正内容を決める。なお、動画像コンテキストの種別を求める際には、人々のアクションを示す幾通りかの量的属性、例えば人数（顔の個数）、目立ったアクションが見られる領域の拡がり、人々が集まっている領域の拡がり、撮影視野420、動きの速度乃至パターン等も使用することができるほか、予め取得済の基準画像が登録されているライブラリや、それに関連する参照データ及び統計量も、使用することもでき

る。例えば、有目的に通話に参加している一人又は二人のローカルユーザ 10a についての指標値と、パーティに参加している大勢（例えば 15 人）の人々についての指標値との間には、大きな違いがあるのが普通である。

#### 【0088】

コンテキスト伝達インタフェース 450 がこれら場面解析処理（変転テスト 630 及び通信イベント解析 655）及び動画像コンテキスト特定を実行する際に使用するアルゴリズムは、統計解析的な手法を使用しその場面のコンテンツ及びコンテキストに関する結論を導出するアルゴリズムである。例えば、場面解析処理の際には、ユーザのアクションのアクションに見られる変化が場面間変転（テスト 630）又は場面内変転（解析 655）を示唆している確率を算出し、その値が低いか、中庸か、それとも高いかを調べる、という統計的手法に従い、撮影された動画像を直ちに解析して変転発生の有無を判別する。これに加え、変転自信度又はそのレベルを導出するようにしてもよい。変転自信度は、調べた結果（ここでは変転示唆との判断）にどの程度の自信があるかを示す物差しであり、通常は百分率値（0～100%）又は確率値（0～1）で表される。同様に、場面間変転示唆と覚しき変化が認識されたときに稼働する動画像コンテキスト特定アルゴリズムでも、種々のコンテキスト示唆キュー及び判別ルールに従い新たな動画像コンテキストの種別を求め、その結果にどの程度の自信があるかを示す動画像コンテキスト自信度を算出する。その動画像コンテキスト自信度は、場面間変転を示唆していると覚しき変化を見つけた解析の有効性を評価する際に使用される（ステップ 643）。場面解析ルールとしては、これらの手順を実行できるよう、解析の条件、変転示唆と見なされる確率及び自信度のレベルを動画像コンテキスト種別毎に定めたものを使用する。

#### 【0089】

まず、それらのうち通信イベント解析 655 における処理、即ち場面解析アルゴリズムに則り且つ場面解析ルール及び指標に従い今の場面 620 を解析して場面間変転を察知する処理について、より詳細に説明する。一例として、図 4C に示した第 1 の場面 620、即ちローカルユーザ 10a が動画通信イベント 600 に参加しリモートユーザ 10b との通話を有目的に行っている場面を考えることとする。プライバシー保護セッティング、音響セッティング、動画撮影セッティング、画質等の撮影関連セッティングについては、ユーザ 10 からの指示、端末 300 による決定又はその併用によってそれ以前に設定されているものとする。ユーザ 10a はこの場面 620 での通話を身じろぎせずに進めるわけではなく、撮影視野 420 内で背伸び、貧乏揺すり、よそ見等の仕草を示すこともあれば、ユーザ 10b との通信を効果的に進めるため電子メール、文書スキャナ、インターネット等の使用に及ぶこともある。これらのアクションであれば、変転テスト 630 の結果は、条件不備で場面間変転を示唆するものと認められない（場面間変転自信度が低い）との結果になる。これに対し、解析 655 の結果は、今の場面 620 向けの指標及びセッティングやそれに対応する動画撮影モードと整合又は関連している（場面内変転自信度が高い）との結果になりうる。そうした結果になった場合、解析 655 では、動画撮影/音響セッティングに施せる修正のなかから、今の場面 620 と齟齬しておらず且つそのユーザ 10a のアクション乃至対応する場面間変転に相応しいものを選定乃至特定する。具体的には、まず場面解析アルゴリズムに則りイベント種別相応の場面解析ルールを適用する。イベント種別に見合った場面解析ルールを適用するのは、例えば図 4C に示した場面 620 で有目的に通話を行っている一人物向けの修正と、パーティ等の大規模なイベントに参加している人々向けの修正等とは、その内容を大きく違えるべきであるからである。同様に、イベント種別、ユーザ 10a の動き（視野 420 を基準とした動きの規模及び速度）等に基づき、場面解析アルゴリズムで参照される動画像フレームの個数も変える。更に、場面解析アルゴリズムに則りそれらの動画像フレームを参照して例えばユーザ 10a の動きを追跡する。そして、その動きが撮影視野 420 や焦点（画質）の小規模修正で対処するに相応しい動きであるか否かを場面内調整自信度の算出値等に基づき判別する。例えば、図 4C 中の場面 620 におけるユーザ 10a のアクションが“椅子への腰掛け直し”であれば、解析 655 では当該小規模修正で対処するに相応しいと判別されるが、“貧乏揺



すり”であれば、そうした修正で対処する条件が備わっていないとして無視されるであろう。また、そのアクションが“一時的なよそ見”であるなら、後述する視線順応手順495を一時的に停止させ後刻復帰させる。

#### 【0090】

場面撮影管理650では、通信イベント解析655にて選定乃至特定された修正の内容や対応する場面内調整自信度等を示すデータを参照して撮影形態修正アルゴリズムを適用することにより、その場面における動画撮影セッティングの調整をその端末300でどのように実行するかを決める。この撮影形態修正アルゴリズムで使用される一群の場面内調整ルール(scene adjustment rule)のなかには、イベント種別、プライバシー保護セッティング、時間的要素(ローカルユーザ10aのアクションが変化する速度やリモートユーザ10bがその変化を知覚する速度に比した動画撮影/処理動作修正の速度及び頻度)、修正の規模、場面内調整自信度等の要素に依拠するものが含まれている。そのため、例えば図4C中の場面620にてユーザ10aが着座状態で身じろぎした場合に、場面撮影管理650では、その動きをユーザ10bにどの程度気づかせたいかに応じて時間を(数フレーム周期分、数十秒等々)決め、その時間をかけて端末300の撮影視野420及び焦点を変化させることができる。また例えば、その動画通信イベント600で照準追跡を行っている場合に、場面内調整ルールに基づき決めた加重係数によって優先順位付けを行うことで、撮影視野420及び焦点の補正を、(発色)画質や視線等の修正乃至補正より先に行わせることができる。

#### 【0091】

次に、変転テスト630及び変転処理640における処理内容について同じくより詳細に説明する。このテスト630で使用されるアルゴリズムは、前述の通り、現在の場面620で得られる動画像及び音声に基づき、現在の動画撮影モードに相応しくないほど大きく場面間変転示唆と覚しき変化が、ユーザ10のアクションに現れているか否かを調べる、というアルゴリズムである。このアルゴリズムでは、場面解析ルール及び判別ルール、例えば動画像コンテキスト、イベント種別データ、最新の撮影条件データ、人的コンテキストデータ、加重係数、しきい値条件、場面間変転自信度等の項目に関するルールを使用することで、そうした解析を実行する。これについて理解するには、図4C中の場面620から場面620'にかけてローカルユーザ10aが示すアクションを考えるとよい。こうした状況では、ユーザ10aは、テーブル上を片付けるにせよお茶を飲むにせよ、まずは椅子から立ち上がるアクション、即ち着座したまま体勢を整えてから椅子を後に引き腰を上げるという一連のアクションをすることとなる。その場合、ユーザ10aが着座したまま体勢を整えている段階では、テスト630における解析結果は、場面間変転と認めるためのしきい値条件に達していない、場面間変転自信度が低いので統計的に見て変転発生とは認めがたい、という判断になる。そのため、動画撮影セッティングを適切なものに修正する処理は通信イベント解析655及び場面撮影管理650に委ねられることとなる。次の段階、即ちユーザ10aが椅子を後に引いている段階では、テスト630における解析結果は、場面間変転と認めるためのしきい値条件に相変わらず達していない、しかし場面間変転自信度は若干上がって中庸に達している、という結果になる。動画撮影セッティングは、これを受けて暫定版のものに修正される(645)。そして、ユーザ10aが腰を上げてテーブル上の片付け等を始めた段階では、テスト630における解析結果は、場面間変転と認めるためのしきい値条件に達した、場面間変転自信度も十分に高い、という結果になる。

#### 【0092】

変転テスト630でこのような肯定的な結果が得られると変転処理640が実行される。この処理640では、まず、動画像コンテキスト特定アルゴリズムに則り新たな動画像コンテキストを特定する(641)。この処理では、撮影視野420内におけるローカルユーザ10aのアクションを示すシステムデータや、イベント種別、ユーザ10aの身振り・手振り・種別・身元、音声/被写体認識結果、ユーザ10aによる入力等のデータに基づき、その又はそれらのユーザ10aが関わっているイベントの種別を識別する。併せ

て、そのイベント種別識別結果がどの程度の統計的確率で正鵠を射ているかを示すべく動画像コンテキスト自信度を算出する。これに次いで、そのイベント種別識別結果に相応しい動画撮影セッティングを特定する(642)。この処理では、予め動画撮影モード毎に制定されている標準的な動画撮影セッティングのなかに相応しいものがあれば、それを選ぶこともできる。例えば、図4Cに示した有目的な場面620からより無目的な場面620'への変転を察知したときには、ユーザ10bがより良質な画像的エクスペリエンスを味わえるよう、緩慢なパン、ティルト又はズーム動作で撮影視野を拡げたり新たな場面620'に相応しく輝度及びコントラストを補正したりするセッティングを、選択するのが望ましい。その次は、動画撮影セッティング変更有効性評価アルゴリズムに則り、且つ動画像コンテキスト自信度を指標として用い、新たな動画撮影セッティング及び音響セッティングが適切であるか否かをその適用に先立ち判別する(643)。その結果が肯定的であれば、撮影形態修正アルゴリズムに則り、その新たなセッティングを適用して端末300による動画撮影/処理動作を然るべく修正させる(644)。この処理では、前述した場面撮影管理650での処理を連想させるが如く、一群の場面内調整ルールに従い動画撮影セッティングの変更速度及び変更タイミングを決めてそれを管理する。例えば、そのローカル環境415におけるアクションの拡がり方が速い場合は速め(例えば数フレーム以内)に、遅い場合は遅めに、動画撮影セッティングを変化させるようにする。

#### 【0093】

無論、場面間変転自信度が中庸水準又は高水準に達しているにもかかわらず、ユーザ10が今行っているアクションがどのような種類のイベントに係るアクションなのか、変転処理640でうまく特定できない(動画像コンテキスト自信度が低い)場合もあり得る。そうした場合には、デフォルト版乃至過渡版の動画像コンテキスト乃至イベント種別を一時的に適用し、それに応じた動画撮影セッティングを暫定し(645)、後刻より相応しい動画撮影セッティングに移行させるようにする。図4Cに示した例なら、場面620と場面620'の間に挟まる期間にローカルユーザ10aが見せるアクション、例えばテーブル上を片付けるアクションや茶を飲むアクションに対して、暫定版の動画撮影セッティングが適用されることとなろう。

#### 【0094】

なお、先の場面(例えば620)の動画像、後の場面(例えば620')の動画像といえは概ねその場面で撮影された動画像フレームの集まりことである。但し、後の場面(例えば620')に係る動画像として扱われる動画像フレームのなかに、先の場面(例えば620)で撮影された動画像フレームが何個が含まれることもあるので、その点に留意すべきである。

#### 【0095】

また、上掲の構成では、場面解析アルゴリズムに則り且つ新旧双方の動画像に基づき統計解析を行い、その解析で場面内変転又は場面間変転を察知したときに、動画撮影セッティングを修正している。そのため、動画撮影セッティングの修正が後追いのになっている。この問題点には、先取りの乃至予測的統計解析法を用い場面の變転を事前に察知し、動画撮影セッティングを先取りのに修正することで対処可能である。また、それは、ベイズネットワーク(Bayesian network)技術等を用いることで実現することができる。ベイズネットワークは変数間の確率的関係を表すグラフィカルモデルに依拠するネットワークである。そのモデルには、事象Xの生起確率を古典的且つ物理的な生起確率ではなく主観確率(person's degree of belief)で表すという特徴がある。その主観確率は個々の命題に関連付けられており、確率則に則り命題間で結合させて操作することができるので、ベイズネットワークは、先行事象発生確率(予備知識乃至事前分布)によって条件付けられる事象発生確率(条件付確率)を扱う際にとりわけ便利である。また、ベイズネットワークでは、推論に際し、そのインタフェース越しに与えられる仮説に整合しそうな証拠や、その仮説に違背しそうな証拠が収集される。証拠収集が進むにつれその仮説についての主観確率は変動していき、証拠が十分に集まると通常はかなり高い又は低い値になる。言い換えれば、ベイズネットワークでは、そのネットワーク上にある個々のノードについて条件付

10

20

30

40

50

確率を算出し、それらのノードのうち幾つかのノードについて条件付確率の値を求めることができる。例えば、専門家の許でベイズネットワークに学習を施し、そのベイズネットワークを推論に使用することができる。また、ネットワーク構造や個々のローカル端末の諸元を表すデータに従いベイズネットワークに学習を施し、面倒すぎて人手に負えないようなネットワーク構築作業にそのベイズネットワークを使用することもできる。

#### 【0096】

ここに、ベイズの定理によれば、ランダムな事象 A 及び B に係る条件付確率・周辺確率間の関係は、両事象間のつながり乃至条件付確率を考慮に入れると、

$$P(A/B) = [P(B/A) * P(A)] / P(B) \quad (1)$$

と表される。式(1)中、 $P(A/B)$  は事象 B が生起した後に事象 A が生起する条件付確率(事象 A の事後確率)、 $P(B/A)$  は事象 A が生起した後に事象 B が生起する条件付確率(事象 B の事後確率)であり、 $P(A)$ 、 $P(B)$  は先行事象(先行体験)から独立に生起するときの事象 A、B の周辺確率(アプリアリ確率)である(符号同順)。通常、事象 B に対する事象 A の事後確率と事象 A に対する事象 B の事後確率は一致しないが、ベイズの定理で記述される厳密な関係が両者間に存在する点に、留意すべきである。

#### 【0097】

こうしたベイズネットワーク技術は、コンテキスト伝達インタフェース 450 によるイベント種別識別動作や判別動作に利用することができる。それは、大勢が参加するイベント、無目的的なイベント、有目的的な通話等、発生しうる種々のイベントについて個別に周辺確率 P を決めうるからである。また、有目的な通話から無目的な通話へと場面が変転する際(図 4C 参照)と、その逆に無目的な通話から有目的な通話へと場面が変転する際とで、条件付確率の値が異なってくるからでもある。ある種類のイベントから別の種類のイベントへと場面が変転する確率は、物理的事象の古典的生起確率(例えばコイントスで表又は裏が出る確率)ではなく人間の起居作法についての知識に基づくものである。ベイズネットワーク技術による解析はかなり有益な手段である。

#### 【0098】

また、動画通信イベント 600 中である場面 620 が生じたときに、その場面 620 の次の場面 620' で現れる動画像コンテキスト及びイベントの種類、即ちその場面 620' で実際に生起する可能性のある事象は、無論のこと多様である。このように 1 個の事象からランダムに分岐して様々な事象が発生する現象を取り扱うには、マルコフ的確率モデル(Markov probability model)を使用すればよい。ただ、端末 300 の主たる稼働環境は居宅であるので、それらの事象への分岐確率は今の場面でのイベントによって条件付けられるものと考えべきである。例えば、有目的な場面の次にくる可能性が高いのは、少人数で無目的に通話している場面が誰かを照準追跡している場面であり、大勢で通話している場面となる可能性は低い。この種の統計的問題を取り扱うには、有向性のあるマルコフ的確率モデル、例えば一次マルコフソースモデルを使用するのが理に適っている。

#### 【0099】

これら、ベイズ的確率推論法、マルコフ的確率推論法又はそれらを組み合わせた複合的推論法は、コンテキスト伝達インタフェース 450 における場面変転管理手法として有効な方法である。例えば、ベイズ的確率推論法に則り新旧双方の動画像の場面解析を実行することで、アクションの変化が場面間変転を示唆しているか否かを弁別することができる(変転テスト 630)。その変化が場面間変転示唆として識別された場合は、それに対し有向マルコフモデルを適用することで、新たな場面 620' に相応しかろうイベント種別を推定する。更に、前述の諸例に倣い自信度を導出し、それらベイズ的確率推論法による場面間変転識別及びマルコフ的確率推論法によるイベント種別識別の確からしさを量る物差しとすること、例えばその自信度を参照して有効性評価を実行し必要時にイベント種別を暫定することもできる。このようにベイズモデルを使用する手法は、何が発生するかについての信念モデル(belief model)に基づき後の場面におけるイベント種別(及び動画撮影モード)を予測しようとする点で、予測的乃至先取りのであるといえる。但し、統計的手法で後追いする前掲の手法に比べると、その実現はやや難しい。

## 【 0 1 0 0 】

次に、ユーザ 1 0 の意向に反することなく端末 3 0 0 又はそのコンテキスト伝達インタフェース 4 5 0 による自動処理に委ねうる範囲が、動画通信イベント 6 0 0 が辿る一連の場面毎に且つその場面に先行する場面に応じて変化していく点にも、留意すべきである。例えば、図 4 C 中の場面 6 2 0 が最初の場面、場面 6 2 0 ' がその次の場面であるとする。この場合、最初の場面 6 2 0 では、ローカルユーザ 1 0 a が手許の電子イメージング装置 1 0 0 を使用し有目的に通話しているので、端末 3 0 0 相手の手動操作をユーザ 1 0 a にかんがりの程度行わせても、ユーザ 1 0 a としては期待乃至許容の範囲内となろう。先行する場面がこうした場面 6 2 0 であるので、その次の場面 6 2 0 '、即ちユーザ 1 0 a がより無目的に通話する場面では、撮影視野 4 2 0、焦点、輝度、プライバシー保護等といった属性の管理動作を自動実行するだけで、端末 3 0 0 としては、そのユーザ 1 0 a の期待におおよそ応えたことになる。しかも、その管理動作はインタフェース 4 5 0 による場面解析を通じ好適に実行することができ、その場面解析に必要なキューも個々の領域におけるアクションの場所及び持続時間や顔追跡結果だけで足りる。また、これと対照的な例として、ユーザ 1 0 a が気楽なパーティに参加している場面（図示せず）が最初の場面、ユーザ 1 0 a がより有目的に通話する場面 6 2 0（図 4 C 参照）がその次の場面となる例を考える。この例の最初の場面（気楽なパーティの場面）では、数多くの場面解析ルール及びコンテキスト対応キュー（例えばユーザ 1 0 a の種別・身元・ジェスチャ・アクション等）に従い様々なイベント種別のなかからイベントを特定する高度な場面解析／変転テスト方式を採らないと、ユーザ 1 0 a が抱く期待に応えることができない。先行する場面がそうした場面であるので、その次の場面 6 2 0、即ちユーザ 1 0 a がより有目的に通話する場面では、先の例と同じく自動動作の範囲を限定しユーザ 1 0 a による手動操作に任せることも可能ではあるが、大抵のユーザ 1 0 a は、先の場面における高度に自動的な動作をその場面 6 2 0 でも引き続き実行するよう期待するはずである。

10

20

## 【 0 1 0 1 】

そうしたコンテキスト的問題の多くは、個々人のアクションを検討して初めて判る類の問題である。例えば、来客があるときに予め自宅や居室や接客場所を清掃しておく、といった通例的な配慮と同様の配慮は、前述の通り動画通信システム 2 9 0 のユーザ 1 0 もその使用に際して行うものと思われる。ユーザ 1 0 のなかには、更に、自サイトの環境 4 1 5 がより綺麗な環境又はより整頓の行き届いた環境に見えるよう、端末 3 0 0 側で助力して欲しいと期待する者もあろう。端末 3 0 0 のコンテキスト伝達インタフェース 4 5 0 に（必須ではないが）設けられている外見修正インタフェース 4 9 0 は、そうした期待に応えるのに役立っている。例えば、端末 3 0 0 に初期設定動作を実行させると、図 5 B に示すようにその環境 4 1 5 の基準画像 4 6 0 が何枚か撮影される。それらの画像 4 6 0 は一種の校正基準を表しているので、ユーザ 1 0 は、それらの画像 4 6 0 を参照しながら自サイトの環境 4 1 5 を整え、清掃及び整頓が行き届いた適度に明るい環境にすることができる。そのユーザ 1 0 は、更に、環境 4 1 5 の見栄えを高めよとの指令を端末 3 0 0 に与えることができる。端末 3 0 0 では、その指令に応じ画像 4 6 0 のうち適当なものにアクセスして合成画像を作成し、その環境 4 1 5 の撮影で得られた画像をその合成画像に差し替える。この機能は、例えばユーザ用コントローラ 1 9 0 に設けられているデクラッタボタンをローカルユーザ 1 0 a が操作するたびに 1 回又は複数回実行される。1 回目の実行で塵埃や小さなゴミが画面に現れなくなり、2 回目以降の実行でより大きな物体 2 6 0 等が現れなくなる、といった具合である。また、画面に現れていない物体、例えば図 2 中の照明ランプ 2 1 5 等を、ユーザ 1 0 a が動画通信イベント 6 0 0 の最中に操作することも無論あり得る。そうした操作が行われると、リモートユーザ 1 0 b 向けに表示される画像が、ちぐはぐで滑稽なものになってしまう。そのため、画面に現れていない物体が操作されるときにその物体を画面上に表示させてその時点以後のイベント 6 0 0 を続けられるようにしてある。即ち、そのユーザインタフェース部分を介したユーザ 1 0 a からの指示等でそうした操作を察知すると、インタフェース 4 9 0 から相応の通知乃至指示が発せられるため、撮影で得られる動画像に対し場面撮影管理 6 5 0 にて施される画像処理の内容が然

30

40

50

るべく修正されることとなる。

#### 【0102】

このほか、ユーザ10は、ローカル環境415の外見を修正する機能を以下の用途にも使用することができる。まず、外見修正インタフェース490による修正で照明状態をブラック照明類似の状態にすることができる。更に、インタフェース490により自分の姿を環境415から抽出させ、別途準備してある仮想環境画像（例えば浜辺の画像）にそれをはめ込ませることができる。その仮想環境画像は静止画像でも動画像でもよい。端末300内に或いはネットワーク経由でアクセス可能な他サイトに仮想環境画像ライブラリを準備しておき、そこから仮想環境画像を取得する構成にすることも可能である。また、インタフェース490による修正で自分の容貌、例えば肌、髪、衣服等を別物にして表示させることもできる。具体的には、自分の端末300のコンテキスト伝達インタフェース450に備わる外見修正インタフェース490を通じ、自分の姿がより美しく見えるよう、顔25や髪40の表示を修正させたり衣服の色を見かけ上変化させたりすることができる。基準画像460のなかにそのユーザ10の姿を捉えているものがあれば、そのユーザ10の最新の姿を捉えた画像と併せて利用することで、この修正をより効果的に行うことができる。なお、ユーザ10の音声特性についても、外見についてのものと似た修正機能を使用することができる。

10

#### 【0103】

そうして得られる画質は、動画通信システム290の稼働品質及びユーザ満足度を左右する。撮影で得られる画像の色を真色にする必要こそないが、発色上の画質を許容水準以上にするには重要な記憶色を好適に発現させること、例えば肌色を好適に発現させることが必要となろう。そのためには、照明光のスペクトル品質を高めることだけでなく、その照明光を十分な光量及び均一度にすることが必要である。その際、照明が設置される部屋の種類、照明される家具の種類、日照の取り込まれ方等によって室内環境光条件が大きく変わることや、日照が天候でしばしば左右されることにも注意した方がよい。

20

#### 【0104】

そのため、上述の構成では、環境光検知器140から得られる信号に基づき、環境光条件が許容水準に達しているか否かを調べて画質を管理している。その検知器140としては、まず、ローカル環境415における環境光200のレベルを計測するタイプのものを使用することができる。検知器140としては、更に、スペクトル選別的濾波/分散装置（例えばダイクロイックフィルタや回折格子）が付設されていて環境光スペクトラムを計測可能なものも使用することができる。その場合、端末300では、検知器140からスペクトル計測結果を取得し、室内照明の近似的黒体放射色温度（普通の日照下なら5900K）を推定することができる。同様に、端末300では、検知器140から数通りの指定波長におけるスペクトル計測結果を取得し、環境光200の一般的な発生源（日照・白熱灯・蛍光灯・LEDランプ等）及びその寄与比率を1個又は複数個の発生源について調べることができる。ある時点での環境光発生源光源及びその寄与比率が例えば日照＝約25%且つ白熱灯＝約75%と判明すれば、その結果に基づき有効な光源モデルを合成することができる。そして、検知器140としては、分光計又は分光放射計を備えていて詳細なスペクトル計測結果が得られるものも使用することができる。その場合、逐次得られるスペクトル計測結果又はその光源モデルを、以前に取得して保存しておいたスペクトル計測結果や以前に生成して保存しておいた色補正データと比較することで、対応する電子イメージング装置100での撮影に使用可能な色補正データを生成又は更新することができる。加えて、基準画像460も画質調整の目標として使用することができる。画像460上での色を本来の色と見なし、それに近づくように現在の色信号値を調整すればよい。色の変化は適当な色空間モデル、例えばCIELABに従い追跡することができる。

30

40

#### 【0105】

こうした動作に次いで、端末300は、環境光検知器140から得られる信号を画質評価アルゴリズムに則り処理することで、今の環境光条件が照明として十分な水準に達しているか否かを調べる。達していない場合、その端末300は、ローカル環境415の照明

50

状態が全体的又は部分的に不備である旨通知する。多くの場合、ユーザ 10 は、これを受けて新たな照明光源を点灯させ、或いはその環境 415 への日照採光量を加増させる。端末 300 は、その一方で、照明ランプ 215 等の室内照明光源乃至採光源のうち無線等で制御可能なものを直接に自動制御して状況を改善させる。そして、イメージプロセッサ 320 では、強度補正データ及び色補正データを生成し、それらのデータに基づく補正を撮影済画像に施すことで、送信用の画像を生成する。なお、色補正データは、その端末 300 で使用されている個々のカメラ毎に、生成して保存することができる。

#### 【0106】

更に、一軒の居宅に電子イメージング装置 100 を複数個設けることもできる。それらを構成する都合各複数個のディスプレイ 110 及びカメラ 120 を内部ネットワークで相互接続すれば、それらを同じローカル端末 300 の一部として機能させることができる。装置 100 が複数個あると、例えば複数のユーザ 10 が同時並行的に装置 100 を使用することができるほか、ユーザ 10 が動画通信イベント 600 の途中で部屋から部屋へと移動しつつ個々の部屋の装置 100 を順繰りに使用することができる。後者の場合、ユーザ 10 が部屋から部屋へと（装置 100 から別の装置 100 へと）動画通信イベント開催中に移動するので、部屋から部屋への移動というユーザ 10 のアクションに追従するよう動画撮影を行うようにする。そのアクション追従的動画撮影動作は自動モードで起動させてもよいし、ローカルユーザ 10a の管制等による手動モードで起動させてもよい。なお、ユーザ 10 は自分の居宅の部屋毎（環境 415 毎）に別々のプライバシー保護セッティングを制定することができる。従って、部屋から部屋へと移動するユーザ 10 の姿を追って撮影しているときの端末 300 の反応が、部屋毎に異なるものになることがある。また、内部ネットワーク経由で相互に接続されている複数の装置 100 のうちいずれかでイベント 600 を実行しているときに、その装置 100 とは別の部屋にある別の装置 100 を稼働させてその部屋の動画像や音声を採取することは、必要とまではいえない。しかしながら、それらの部屋の装置 100 を同時に稼働させて動画像や音声の採取を行わせれば、それらの部屋のローカルディスプレイ 110 それぞれに、自他の部屋の様子を示す何個かの画面分割画像 410 を表示させることができる。その場合も、コンテキスト伝達インタフェース 450 では、アクションやイベントの種別、ユーザ 10a の種別や身元、リモートユーザの種別や身元等といった動画像コンテキスト関連知識を適用することで、撮影及び送信すべきコンテンツを特定することができる。

#### 【0107】

端末 300 は、また、使い勝手がよく様々な形態で利用できる装置になるよう、（少なくとも部分的に）可搬な装置として構成されている。そのため、端末 300（少なくともその電子イメージング装置 100）をキッチンキャビネット 265 上に固定して使用することも（図 2 参照）、装置 100 を仕事場や車庫で使用することも（図 8 参照）できる。後者の用法では、リモートユーザ 10b とのやりとりを通じローカルユーザ 10a が自動車修理等を助けてもらうことができる。また、いつもは図 2 のローカル環境 415 にあるキャビネット 265 に載せていき、必要なときにそこから取り外して図 8 のローカル環境 415（車庫等）に持ち込む、という拡張的な形態で装置 100 を使用することも、図 10 に示すように装置 100 を居宅内の別の場所（例えばテーブル上）に移動させて使用することもできる。また、このようにユーザ 10 が自分の目的用途に応じ端末 300 を持ち運べるようにするため、画面サイズをそれに相応しいサイズにしてある。即ち、ディスプレイ 110 のサイズが大画面テレビジョン受像機並であると端末 300 の可搬性がかなり低くなるので、より小さなラップトップコンピュータ並又は電子書籍リーダ並のサイズにして可搬性を高め、或いは更に小さな携帯電話並又は携帯情報端末（PDA）並のサイズにして更に可搬性を高めてある。携帯電話や PDA 並のサイズだと、端末 300 及びその画面が小さいのでプライバシー保護やコンテキスト対応に関わる機能が一部制約されるが、GPS（汎地球測位システム）測位機能の付加で機能及び性能が高まる面もある。また、画面が小さいとその上に画面分割画像 410 を表示させることが難しいが、マイクロプロジェクタ技術が近年とみに進歩しているので、携帯電話等の一般消費者向け小型電子機器

にLEDプロジェクタが装備され、画面サイズが実質的に拡張されるのも遠い未来のことではなからう。そうした機器は本発明の実施にひときわ適している。

【0108】

端末300の機能は、更に、その端末300で使用される撮影装置乃至カメラ120の種類及び配置によっても左右される。例えば、端末300におけるカメラ120の使用個数は、図3A及び図6の如く1個にすることも、図1、図3B及び図4Aの如く2個にすることも、或いはそれ以上にすることもできるので、前述した図4Aの如く、WFOVカメラ120と1個又は複数個のパン、ティルト及びズーム機能付NFOVカメラ120とを対にして設けて利便性を高めることができるほか、互いに異なる方向を向くようNFOVカメラ120を複数個設けて複数通りのNFOV420を捉えうる構成にすることもできる。また、前述した図3Aの如く、電子イメージング装置100の構成をそのカメラ120及びディスプレイ110が共にほぼ+Z方向を向く前睨み型の構成にすることができるほか、ディスプレイ110、カメラ120又はその双方が+Z方向だけでなくその逆の-Z方向にも向く（或いは-Z方向側にもある）両睨み型の構成にし、二方向からその部屋乃至ローカル環境415を撮影してその画像を表示することも可能である。更に、電子イメージング装置100を構成するカメラ120をそのフルアングルが約180°又はそれ以上にもなる超広角カメラ、例えば魚眼レンズカメラや360°パノラマカメラにすることができる。例えば特許文献35（発明者：Wallerstein et al.）又は特許文献36（発明者：Driscoll et al.）に記載の360°パノラマカメラ、例えばカメラを使用すれば、ローカル環境415を取り巻くエリアにおけるアクションを、その端末300でより素早く撮影乃至監視することができる。但し、超広角レンズで撮影された画像は歪んでいることが多いので、ローカルユーザ10a又はリモートユーザ10bに示される画像に対し、イメージプロセッサ320で歪補正処理を施すことが必要になる。そして、端末300の電子イメージング装置100が設けられている場所から物理的に離れた場所を占めるよう、ローカル環境415内に更に1個又は複数個のカメラ120を設けることもできる。例えば、装置100が設けられている部屋とは別の部屋に、新たなカメラ120（又はそれにディスプレイ110を付加した新たな電子イメージング装置100）を追加配置することができる。追加したカメラ120（又は装置100）を内部ネットワーク経由で端末300と接続すれば、ユーザ10の動きをより広い範囲に亘りその端末300で捉えることが可能になる。しかも、元々あった装置100を構成しているカメラ120だけではうまく捉えられない方向から、その新たなカメラ120で撮影を行える分、撮影能力が高まることとなる。プロセッサ320では、それらの画像から秀逸なコンポジット画像を生成することができる。

【0109】

ただ、ユーザ10がそのローカル環境415内又は居宅内を動き回ると、いずれはアイコンタクトが成立しなくなる。即ち、電子イメージング装置100のカメラ120を真正面から見られない状況や、その装置100に表示されるリモートユーザ画像と視線が一致しない状況に立ち至る。例えば、図4C中の場面620ではローカルユーザ10aの視線が真正面から装置100に向いているのに対し、場面620'では反れている。図2中のユーザ10の視線も真正面から装置100に向いているわけではない。この問題に対処する手段としては、非特許文献6に記載の興味深い手法を使用することができる。具体的には、複数個のカメラからなるセンサアレイを壁や家具に設ける一方、指向性ビデオプロジェクタ（いわゆるエブリフェアディスプレイプロジェクタ）を好ましくは天井付近に設けておく。更に、動画通信イベント開催中にユーザ10aが壁又は家具上のカメラを注視したら、指向性ビデオプロジェクタがリモートユーザ10bの姿をそのカメラの配設面（壁又は家具の外面）に映写するように、システムを構成すればよい。この手法ならアイコンタクト問題を好適に解消することができる。反面、この手法は、ユビキタスコンピューティング又はパーベイシブコンピューティングがある程度普及していないと実施できない、システム構築に当たり相当なインフラストラクチャが必要になる、普及が進むか否かが懸念される、ユーザ10bの頭部が体から切り離されて壁上に浮かび上がるので見た目に気

10

20

30

40

50

味が悪い、等といった難題も抱えている。

【 0 1 1 0 】

また、テレビ会議の分野では、従来から、できるだけ小さな視差でアイコンタクトを成立させることがひときわ重要であるとされている。これは、互いに実際に会って会話しているかのような感覚を、ローカルユーザ 1 0 a 及びリモートユーザ 1 0 b に抱かせるためである。これを達成するには、既存の多くの特許文献の記載に倣い、電子イメージング装置を構成するディスプレイの背後にカメラを仕込んだ構造にするとよい。そうした構造では、そのディスプレイを真正面から見つめているユーザ 1 0 a の姿を、ディスプレイ越しにカメラで撮影し、その撮影で得られた画像（ダイレクトオン画像）をユーザ 1 0 b に送信することができる。図 9 A に、そうしたディスプレイ越し撮影装置の一従来例として、特許文献 3 7（発明者：Manico et al.、譲受人：本願出願人、名称：画像撮影表示装置（Image capture and display device））に記載の装置を示す。図示例の電子イメージング装置 1 0 0 は、カメラ 1 2 0 及びプロジェクタ 1 8 0 に加え、散光状態・透光状態間でスイッチングする明滅画面 1 3 2 や、カメラ 1 2 0 及びプロジェクタ 1 8 0 から画面 1 3 2 に至る光軸を部分的に共通光軸 1 7 0 にするビームスプリッタ（例えば部分的に銀色をした半透明のミラー）1 3 4 や、個々のフレーム周期のうちカメラ 1 2 0 による撮影に割り当てられている期間だけ遮光されるようプロジェクタ 1 8 0 からの光を変調するシャッタ 1 8 4 を備えている。そのシャッタ 1 8 4 は、画面 1 3 2 が散光状態になるときに透光状態になるよう（プロジェクタ 1 8 0 からの画像が画面 1 3 2 上に表示されるよう）、また画面 1 3 2 が透光状態になるときに透光状態になるよう（カメラ 1 2 0 が画面 1 3 2 越しにユーザ 1 0 を撮影できるよう）、画面 1 3 2 のスイッチングと同期して駆動される。

10

20

【 0 1 1 1 】

この種のアイコンタクト型遠隔会議装置はこれまでも数多くの特許文献で提案されている。例えば、上掲の特許文献 3 7 のほかに、特許文献 1 4（発明者：Yoneta et al.、名称：テレビ電話機（Video Telephone））、特許文献 1 6（発明者：McNelley, et al.、名称：パススルー反射型投射表示装置（Pass-Through Reflective Projection Display））、特許文献 1 7（発明者：McNelley, et al.、名称：画像ブロック型遠隔会議用アイコンタクト端末（Image Blocking Teleconferencing Eye Contact Terminal））等がある。ただ、ビームスプリッタ（例えば部分的に銀色をしたミラー）を使用するため、この種の従来型テレビ会議装置は図 9 A に示す如く（特に奥行き方向に）大きくなりやすく、現在この種の仕組みによる製品は一般消費者向け市場ではなく企業経営者向け市場で販売されるに留まっている。

30

【 0 1 1 2 】

そのため、ディスプレイの背後にカメラを設置する方法の改良案として、表示部材と撮影部材が密着している装置を様々な企業が提案している。その一例として、図 9 B に、米国特許出願第 1 1 / 5 5 5 8 2 2 号（発明者：Kurtz et al.、譲受人：本願出願人、名称：撮影装置を複数個備える集積型表示装置（An Integrated Display Having Multiple Capture Devices））に記載の構成を採る電子イメージング装置 1 0 0、特にその個別画素統合型画像表示兼撮影部の構成を示す。この構成では、基本的に、複数個の部分透明画素 1 5 4 によって形成されるディスプレイ 1 1 0 越しにカメラ 1 2 0 で撮影を行うことができる。具体的には、この装置 1 0 0 の基板 1 5 8 上には通常表示画素 1 5 0 及び部分透明画素 1 5 4 が各複数個形成されており、それらの画素 1 5 0、1 5 4 に対する制御信号及び駆動電流の供給が薄膜電子部品 1 6 2 のパターンによって、或いは更に反射電極 1 6 8 及び透明電極 1 6 6 のパターンによって行われている。画素 1 5 4 たる窓素子の上下両面に電極 1 6 6 が配されているのに対し、画素 1 5 0 ではその下面に電極 1 6 8 が配されている、という違いはあるが、それらの画素 1 5 0、1 5 4 は、いずれも OLED 技術、ポリマ発光ダイオード（PLED）等を用い白色或いは特定色（レッド、グリーン、ブルー等）の発光素子として形成されている。こうした構成では、複数個の画素（窓素子）1 5 4 によって開口 A が何個所かに形成されるので、カメラ 1 2 0 がそこから前方を覗くことができる。また、スクリーンドア効果（網扉効果）等の偽像が発生して撮影画質が低くなる

40

50



恐れもあるものの、カメラ１２０は概ね通常通りに合焦及び機能させることができる。そのカメラ１２０からディスプレイ前面までの距離が図９Ａの構成に比べて短いので、この装置１００は図９Ａのそれに比べコンパクトで（特に奥行き方向寸法が小さく）、構造上は図３Ａ及び図３Ｂのそれにより近いものになる。また、コンピュータ用モニタ又はテレビジョン受像機としても機能するディスプレイ１１０（図１に関する説明を参照）に対しこの集積手法を適用すると、カメラ１２０の位置がディスプレイ１１０の側方ではなく背後になる。その結果、アイコンタクトが成立しやすくしかもコンパクトな集積型の装置１００が得られるので、この集積手法はとりわけその種のディスプレイ１１０に適しているといえる。

#### 【０１１３】

こうしてアイコンタクトを成立させることは、全ての用途で明らかに有益なこととはいえないし、あらゆる動画撮影モードで必要とされているわけでもない。しかし、前述した従来のテレビ電話装置や遠隔会議装置のように、ローカルユーザ及びリモートユーザが共に会議室風の部屋にいる場合には、ダイレクトオン画像を表示させることによって、ローカルユーザ・リモートユーザ間のアイコンタクトを成立させることができる。即ち、ローカルユーザ及びリモートユーザが自分の顔を自分側のディスプレイの画面と同じ高さにし、その画面に真正面から向かい合って画面中央を見つめるものと想定できる用途なら、ユーザの視線がくるはずの位置（画面の中央）から画面越しに覗けるよう端末のカメラを配置することで、アイコンタクトを好適に成立させてユーザエクスペリエンスを高めることができる。

#### 【０１１４】

それとは違い、画面を真正面から見つめているユーザの姿を斜め方向から撮影するようカメラが配置されている場合でも、ソフトウェア的な処理を利用してアイコンタクトを成立させることができる。その手法としては既提案のものを使用することができる。例えば、特許文献４２（発明者：Andersson et al.）に記載のテレビ会議システムでは、リモートユーザの姿を捉えている画像データから目（瞳と瞼）のデータを取り出し、それをライブ画像データに編入するソフトウェアを使用している。このソフトウェアを使用すれば、リモートサイト側被写体がこちらをまっすぐ向いているかの如く見えるよう、リモートユーザ画像を外見的に修正して（カメラの位置オフセット分を補正して）表示させることができる。また、特許文献３８（発明者：Hillis et al.）に記載のテレビ遠隔会議システムでは、複数個の動画撮影装置で撮影される複数通りの動画像に対し同時に画像モーフィングアルゴリズムを適用する、という手法を使用している。この手法を使用すれば、複数通り得られる動画像間の補間でローカルユーザ画像を合成し、そのコンポジット画像をリモートユーザ向けに表示して、アイコンタクトを成立させることができる。

#### 【０１１５】

また、遠隔会議用のものと違い、居宅向けの動画通信システム２９０及びその端末３００は様々な流れで使用される。そのなかには、ダイレクトオン画像の撮影及び表示でアイコンタクトを成立させることができない流れも数多く含まれる。例えば、図４Ｃ中の場面６２０でローカルユーザ１０ａのダイレクトオン画像が得られるのに対し、同図中の別の場面６２０'や図２の場面では、ユーザ１０又は１０ａの視線が電子イメージング装置１００に対し上下又は左右に反れているためそのダイレクトオン画像を得ることができない。この点についてより詳細に検討するため、図１０に、ユーザ１０が自サイトの装置１００及びそのカメラ１２０を見下ろしている状況を示す。カメラ１２０がディスプレイ１１０の背後に仕込まれている場合、ユーザ１０がテーブルに向かって着席し、真正面からディスプレイ１１０の中央を見据えれば、即ちディスプレイ１１０の中央から真正面に延びる光軸１７０に沿ってユーザ１０、ディスプレイ１１０及びその背後のカメラ１２０が並べば、そのカメラ１２０でユーザ１０を撮影してアイコンタクトを成立させることが可能である。しかし、図示状況ではユーザ１０が装置１００を見下ろしている。この状況では、ディスプレイ１１０の中央からユーザ１０の目に至る光軸が、軸１７０に対し傾斜した別の軸１７０'になる。そのため、ディスプレイ１１０の背後にあるカメラ１２０で下側

10

20

30

40

50

の軸 170 沿いにユーザ 10 を撮影しても、そのユーザ 10 の顔画像が得られるとは限らない。得られるにしても、リモートユーザがそれを見てユーザ 10 とアイコンタクトできるような顔画像ではないだろう。また、この種の動的な環境では、リモートユーザ画像を表示する際にも、これと似たアイコンタクト問題が発生する。具体的には、ローカルディスプレイ 110 上にリモートユーザ 10 b の画像を表示させる際に、その視線をユーザ 10 a にどうやって向けるか、という問題が発生する。例えば、図 10 に示したディスプレイ 110 上にリモートユーザ画像を表示させたとしても、その画像の視線方向が下側の軸 170 に沿っているのであれば、ディスプレイ 110 を軸 170' 沿いに見ているユーザ 10 a は、恐らく、ユーザ 10 b とのアイコンタクト感を味わうことができない。

#### 【0116】

10

このように、リモートユーザがローカルユーザの視線を感じられるようローカルユーザを撮影することや、ローカルユーザがリモートユーザの視線を感じられるよう視線補正してリモートユーザ画像を表示させることは、決して簡単なことではない。実際、ユーザ 10、ディスプレイ 110 及びカメラ 120 三者間の幾何学的配置に生じる変動に抗し、完全にアイコンタクトを成立させるとしたら、ユーザ 10 の移動に応じダイナミックな修正が施されるよう撮影、表示の両局面でそれぞれ視線追跡処理を行うことが必須となろう。無論、この問題の解決をそこまで絶対視する必要はない。自分が動画通信イベントに参加しているという事実をユーザ 10 がどの程度まで許容できるか、またその許容度が“ちょうど窓から外を見ているかのような擬似臨所体験を味わいたい”というユーザ 10 の想いに比し如何ほどのものかに応じて解決を図ればよい。ユーザ 10 がアイコンタクトを望む度合いはイベントの種類によっても変わってくるので、この問題の重要度は動画像コンテキスト、特にイベント種別によっても左右される。

20

#### 【0117】

アイコンタクトが成立しやすい画像が得られるよう撮影を行う上で重要なのは、使用している端末 300 の構成を十分に利用することである。例えば、その端末 300 で使用される撮影装置 120 のうち第 1 のものが広角カメラであり、第 2 のものが表示画面 115 の背後に仕込まれたズーム機能付 NFOV カメラ（図 9 B に示した構成のもの等）であり、第 3 のものがどこか画面 115 以外の場所に実装されたズーム機能付 NFOV カメラであるとする。この場合、ディスプレイ 110 の構成上パン動作及びティルト動作が制約されるので、第 2 のカメラ 120 で好適に撮影できる方向は、例えば図 10 中の軸 170' 沿いに限られる。他方、第 3 のカメラ 120 は、画面 115 の背後に仕込まれていないためそのパン動作及びティルト動作にあまり制約が課されない。従って、第 3 のカメラ 120 で高品質な画像が得られる方向はより多様である。図 10 中の軸 170" 以外にも様々な方向沿いにローカル環境 415 を高品質撮影することができるので、第 3 のカメラ 120 をうまく利用すれば、ローカルユーザ 10 a や物体 260 のダイレクトオン画像を多様な方向から得ることができる。また、特許文献 38 の記載に倣い、非仕込みタイプでパン及びティルト可能なカメラ 120 を複数個用い、それらで捉えたユーザ 10 の顔画像を合成してよりアイコンタクトが成立しやすいコンボジット画像を生成することも可能であろう。但し、ユーザ 10 がいつもディスプレイ 110 やカメラ 120 を見ているとは限らない。よそ見が甚だしくなると、アイコンタクト性の高い画像を撮影することがコンテキスト的に重要でなくなり或いは技術的に不可能になる。

30

40

#### 【0118】

同様に、アイコンタクトが成立しやすいように画像を表示させる処理で重要なのは、以下の要領で幾何学的配置の変化に対処することである。まず、図 4 C 中の場面 620 で表示されているのと同様のリモートユーザ画像を、図 10 に示したサイトのディスプレイ 110 によって表示させ、そのサイトのユーザ 10 に見せる場合を考える。通常のテレビ会議システムであれば、軸 170 に沿ってリモートユーザ画像を眺めているユーザ 10 と、表示されているリモートユーザとの間にアイコンタクトが成立するよう、真正面を見つめているリモートユーザ画像を画面の中央に表示させるところである。しかし、図 10 に示す例では、ユーザ 10（特にその顔 25）の位置が軸 170 から外れているので、そのよ

50

うなりリモートユーザ画像を表示させたのでは、ユーザ 10 の視線とリモートユーザの視線とを一致させることができない。この問題を抑えるため、端末 300 では、まず、そのローカル環境 415 におけるユーザ 10 の方向又は位置をディスプレイ 110 を基準にして推定する。その推定のアルゴリズムとしては、ディスプレイ 110 又はその中央を原点にした三次元空間でそのユーザ 10 の顔の向き、位置等を追跡する姿勢追跡アルゴリズムや、同じ空間でそのユーザ 10 の目の向き、位置等を追跡する視線追跡アルゴリズムを使用する。ユーザ 10 の位置を調べる手法としては、ステレオカメライメージングや、光波又は音波（ソナー）による測距及びマッピング等、様々な手法を使用することができる。端末 300 では、次いで、そのイメージプロセッサ 320 に視線順応手順 495 を実行させる。手順 495 としては、例えば、リモートユーザがユーザ 10 の方を向いているように見せるためリモートユーザ画像中の目の部分を修正する手順を、使用することもできる。ディスプレイ 110 上に表示されるのはこの修正を経たリモートユーザ画像である。目の部分を修正する手法としては、例えば特許文献 42 に記載の眼差し修正法を使用する。この手法では、画像データのうち瞳又は瞼を表している部分を修正し、リモートユーザがあたかも適切な方向（軸 170 沿いの正面方向でもよい）を向いているかのような外見に変化させる。

#### 【0119】

ただ、実際には、自分の視線を左右（横方向）又は上下（縦方向）に動かす際に、人間というものは、目の動きだけで済ますのではなく、いっとき目を動かして対処した後素早くその方向へと頭を巡らしそれにつれて目の位置を元に戻すはずである。そのため、視線順応手順 495 としては、より進んだ手順、即ちより自然にアイコンタクトが成立するよう少なくとも頭の動きを勘案して画像を表示させる手順を使用する。この手順では、リモートユーザ画像を傾斜及び回転させてローカルユーザの方を向かせる処理を実行する。その処理は、例えば特許文献 38 に記載のデュアルカメラ法の拡張で、即ち非特許文献 11 に記載の眼差し修正法等に則り顔の向きを修正しローカルユーザの方を向かせる手法の併用で実行される。また、特許文献 39（発明者：Toyama et al.）に記載のテレビ会議システムでは、カメラによる撮影で得られたローカルユーザ画像に対し、バーチャルな三次元画像に変換した上で修正を施している。その修正は、ローカルユーザの頭や目の向きを画像上で改変して画面の真正面に向けることで、リモートユーザがアイコンタクト感を得られるようにする、というものである。従って、頭や目を画像修正で様々な方向に向けられるようこの手法を拡張し、それを手順 495 で実行するようにしてもよい。例えば、あるリモートサイト 364 の端末 300 が、その環境内の様々な場所に配置された都合複数個のカメラ 120 を備えているのであれば、それらのカメラ 120 のうちリモートユーザ 10b を最善の向きから捉えているもの（手順 495 による後処理ができるだけ少なくなるもの）を選び、そのカメラ 120 から得られる一群の画像をローカルサイト 362 にいるユーザ 10 向けに表示させるようにすればよい。

#### 【0120】

また、ローカル環境にリモートユーザ画像を表示させてアイコンタクトを図ることの可否及びその画像の表示形態は、前述の通り、ローカル、リモートの各環境で生起しているイベント及びアクション双方のコンテキストによって場面毎に変わってくる。例えば、ローカル環境 415 に複数の人物がいる場合、それらローカルユーザ 10a 全員の平均的な居場所を見やる画像になるようリモートユーザ画像を修正すべきか、それともそれらのユーザ 10a のうち現在の話者を見やる画像になるようリモートユーザ画像を修正すべきか、といった興味深い問題が発生する。この問題への対処が重要になるのは、通常、ユーザ 10a 及びそれを取り巻く環境 415 を捉えるための撮影視野 420 があまり広くないときである。

#### 【0121】

より詳細に述べると、表示される画像に視線補正を施しアイコンタクトの成立を図る処理は、視線指標やコンテキストに応じその内容が切り替わるよう、視線又は動画像コンテキストに基づく一群の視線コンテキストルールに従い実行される。例えば図 2 では、ユー

10

20

30

40

50

ザ 10 が宅内機器たるコンロ 270 を見ながらローカル環境 415 たる台所で作業している。コンロ 270 に限らず、ユーザ 10 が見ているのが端末 300 以外の物体であるときには、その環境 415 がどのような様子（コンテキスト）であるかが色々と伝わるように、そのユーザ 10 の画像を視線補正無しでリモートユーザ向けに表示する。逆に、ユーザ 10 が端末 300、特にそのカメラ 120 を見ているときには、コンテキスト伝達インタフェース 450 にて時間的及び方向的な視線指標に基づきそのことを検知し、視線補正アルゴリズムをそのユーザ 10 の画像に適用することで、そのユーザ 10 のダイレクトオン画像をよりアイコンタクトが成立しやすい画像に修正する。リモートユーザが見るのはこの修正を経た画像である。

#### 【0122】

更に、居宅間動画通信ではユーザの位置やコンテキストがめまぐるしく変化する。視線配慮型の撮影動作（リモートユーザとのアイコンタクトが成立しやすい画像が得られるタイミング及び形態でのローカルユーザ画像撮影）や、視線配慮型の表示動作（ローカルユーザとのアイコンタクトが成立しやすい向き及び形態でのリモートユーザ画像表示）は、こうした制約条件の下に実行される動作である。その際踏まえるべきユーザ位置及びコンテキストは、前述の通り、ローカルサイト、リモートサイト双方のそれである。この点については、不完全ながら、動画通信システム 290 に属する二端末 300 間で動画通信イベント 600 の最中に交わされるやりとり、という形態で図 7D にも示されている。こうした制約条件を充足させるため、通信イベント解析 655 では、その時々場面（例えば 620）に場面解析アルゴリズムを適用する際、一群の視線コンテキストルールに従いその場面 620 の特性に関わる種々の情報を調べる。その対象となるのは、頭又は目の動き・その速度・（カメラ 120 から見た）方向等の視線指標、撮影視野設定、人々のアクション・会話等の人的キュー、イベント種別等の動画像コンテキスト等々、撮影時の視線追跡や視線補正（視線対応）を行うべきか否かを判別するのに役立つ情報である。もし、ローカルユーザ 10a の参加先イベントが有目的イベントであれば、視線コンテキストルール群を適用したときに一群の視線指標に基づき視線対応の必要性が認められるのに対し、ユーザ 10a の参加先イベントが多数イベント、即ちディスプレイ 110 を眺めはするがリモートユーザ 10b との通話はあまり行われないイベントであれば、そうした視線対応は概ね不要であるということになる。使用される視線コンテキストルール群は、例えば、撮影視野が狭く捉えている人数が少ないほど視線対応の必要性及び重要性が高いと判別されるようなルールである。また、視線対応が不要な状況でも、場面解析アルゴリズムに則り例えば推定姿勢角を求め、その推定姿勢をトリガ値乃至しきい値に照らすことで、ユーザ 10a がディスプレイ 110 を見ようとしているか否かを判別する。推定姿勢角は、ユーザ 10a の視線方向を追跡する視線追跡処理で得られる情報であり、そのユーザ 10a の注意を惹いているものの方向を表している。これを導出するには実際に目（瞳）の動きを追跡した方がよいが、非特許文献 9 に記載の如く頭の姿勢から視線方向を推論するだけで十分であることも多い。

#### 【0123】

その場面解析アルゴリズムで視線対応の要否判別に使用される視線コンテキストルール、即ち現場面の動画像コンテキスト（イベント種別や撮影視野設定）を視線対応の必要性乃至重要性に関連付けるルールには、視線配慮型撮影時動作に関わるローカルユーザ特性（視線指標等）の変動範囲を現場面の動画像コンテキストに基づき定めるルールや、視線配慮型表示時動作に関わるリモートユーザ特性（視線指標等）の変動範囲を定めるルールが含まれている。従って、通信イベント解析 655 における視線追跡／補正の要否判別は、現場面での視線指標を導出し、現場面の動画像コンテキストに対応する視線指標変動範囲と比較することで、行うことができる。視線指標の算出は、場面解析アルゴリズムに関し前述した別種処理に倣い行い、併せてその統計的品質評価指標たる視線方向自信度も算出する。コンテキスト伝達インタフェース 450 では、こうした解析 655 の実行を通じ、ローカルユーザの姿を視線に応じ撮影又は画像処理すべきことが、高い視線方向自信度で判明した場合、視線順応手順 495 の実行を通じ指令を発することで、カメラ 120 に

10

20

30

40

50

よる撮影動作を調整する。即ち、使用するカメラ120を指定し、或いはズーム、パン、ティルト等のセッティングを調整して撮影を行わせることで、リモートユーザがアイコンタクト感を得やすい画像を撮るようにする。必要ならば、その視線順応手順495から場面撮影管理650経由でイメージプロセッサ320にデータを供給することで、撮影済の画像データを修正し、それによりリモートユーザとのアイコンタクトが成立しやすい動画画像を生成する。

#### 【0124】

その画像修正では、アイコンタクト感を高めるため、前述の通り画像データのうち瞳又は顔を表す部分や頭の向きを修正する。同様の目的で身体の姿勢や場面の向きを修正することもある。視線順応手順495では、更に、先方（リモート環境）の動画画像コンテキスト、場面特性及び視線指標が判る場合は、先方のサイト364でローカルユーザ10aの画像が表示される際、目又は頭の向きをそれらに基づき指定又は修正する。例えば、先方のサイト364の動画画像コンテキストが複数の人物をWFOV420で捉える性格のコンテキストであるのなら、先方で表示される画像におけるユーザ10aの目又は頭の向きを、先方のサイト364で捉えている人物群の平均的な居場所を見やる向きになるよう指定又は修正することで、アイコンタクトを成立させやすくする。リモートディスプレイ110から見たリモートユーザ10bの位置又は方向が判明している場合は、先方で表示される画像におけるユーザ10aの目又は頭の向きを、そのユーザ10bの方向になるよう指定又は修正することで、そのユーザ10bとのアイコンタクトを図ることもできる。これを実現するため、端末300は、動画画像コンテキスト（イベント種別）、撮影視野設定、話者位置、視線指標等のデータを収集し、ネットワーク365経由で他の端末300へとそのデータを送信する。その一方、端末300は、他の端末300からその種のデータを受信して手順495を実行する。例えば、そのデータに基づき画像処理を施すことで、表示される画像における目又は頭の向きを修正する。また、手順495では、前述した場面内調整アルゴリズムに倣い、且つこうした視線補正の必要性を示す指標として視線方向自信度を用いつつ、その視線補正の要否を判別し又は予め定められている処理のうちどれを今回の視線補正で使用するのかを指定する。その手順495では、更にタイミングの制御も実行する。例えば、ユーザの視線が引き続きディスプレイに達しているが、その視線がそのディスプレイの仕込みカメラの真正面から外れているため視線補正が必要な場合、視線補正が施されていない状態から所要内容の視線補正が施されている状態へとスムーズに遷移するよう、視線補正開始までに費やす時間及び視線補正開始から遷移完了までに費やす時間双方の長さを制御する。

#### 【0125】

この画像処理、即ちリモートサイト364で表示される画像に視線補正を施しリモートユーザ10bのアイコンタクト感を高める処理は、そのサイト364から供給される視線関連データに基づきローカルサイト362で動画画像データを修正してサイト364に送信する、という段取りでも、動画画像データの修正を含む仕上げ部分をサイト364で行う、という段取りでも行うことができる。どちらもネットワーク365経由でのデータ交換が必要になるが、使用帯域幅を抑えるには前者の方がよい。同じ端末300で両者を併用することもできる。このほかには、サイト364から供給される動画画像データを独自に解析することで、視線指標、動画画像コンテキスト等の視線関連場面特性をサイト362の端末300が独自に導出し、その場面特性に従いローカルユーザ画像に視線補正を施してサイト364に送信する、という段取りも採用することができるが、関連データをネットワーク越しに交換する前掲の手法に比べると低効率なものになる。

#### 【0126】

関連する事項として、ローカル環境415におけるローカルユーザ10aのアクションに対し、リモートユーザ10bが受動的にも能動的にも関わり合えることに、留意すべきである。即ち、ユーザ10bは、ユーザ10a又はその端末300から送られてくる画像を見ることで受動的に、またローカルサイト362における撮影視野変更を要求又は指令することで能動的に、関わり合うことができる。能動的な例では、図4Bに示す如く視野

420を縮小することや、NF OV 420を第1の場所(テーブル)から第2の場所(子供達)へと移動させることを、ユーザ用コントローラ190の操作等で要求する。ユーザ10bからのこうした要求乃至指令に応じ、ユーザ10a側のディスプレイ110では、例えば、WF OV 420におけるNF OV 420の動き又はそこから得られる画像を、画面分割画像410として表示する。ユーザ10bが見せるジェスチャに反応してカメラ120に指令を発する意向察知応答手順(prompt and response sequence)を実行し、能動性を更に高めることもできる。その場合、ジェスチャに反応するタイプのコントローラ190を、端末300のユーザインタフェース440に何個か設けるとよい。その種のコントローラ190は、システム稼働だけでなく、ユーザ10a・10b間のやりとりにも使用することができる。

10

#### 【0127】

図11A及び図11Bに、この動画通信システム290で使用可能な意向察知応答手順の一例を模式的に示す。この例は、ユーザ10a及び10bが自分の端末300を使用しネットワーク360経由で通話しており、またその端末300のディスプレイ110に通話相手の姿が映し出されている点で前掲の例と同様であるが、カメラ120で捉えた画像からユーザ10aの意向(prompt)を察知する機能、例えば非特許文献9に記載の如く視線追跡結果をジェスチャキューとして使用し頭部姿勢等から意向を察知する機能がローカルサイト362側の電子イメージング装置に備わっている点で異なっている。そうした機能があるため、図11Aに示すように、ローカルユーザ10aがリモートユーザ画像を見つめると、リモートサイト364側のカメラ120がその画像の相対位置Mに対応する位置M'を向く。位置M'にはリモートユーザ10bがいるので、ユーザ10a及び10bはアイコンタクトしながら通話することができる。更に、その通話が進展してユーザ10aの注意が例えば上の方にずれると、ユーザ10aの見つめる先が相対位置Nに移動した旨、視線検知アルゴリズムに従い検知される。サイト362側のコンピュータ340は、こうしてユーザ10aが関心を示した方向に基づく指令を、通信コントローラ355及びネットワーク360を介しサイト364側のコンピュータ340に送信する。サイト364側のコンピュータ340は、図11Bに示す如く、その位置Nに対応する位置N'に向くよう自サイトにあるカメラ120の向きを修正する。

20

#### 【0128】

ご理解頂けるように、こうしたジェスチャ追跡インタフェースとしての視線追跡は、アイコンタクトしづらい画像を表示させるための視線追跡及び視線補正との連携で進行する。例えば、図11Aではリモートユーザ10bの顔がサイト364内の位置M'にあり、ローカルユーザ10aがその位置M'に対応する位置Mを見つめている。また、サイト362側のカメラ120は、例えばディスプレイ110越しにユーザ10aのダイレクトオン画像を捉えている。その画像はサイト364のディスプレイ110に表示されるので、ユーザ10aと、その画像を見ているユーザ10bとの間に、アイコンタクトが成立する。この状態からユーザ10aの見つめる先が位置Nへと移動すると、ユーザ10aの視線はいつときそのサイト362のカメラ120から外れる。しかし、サイト364にあるカメラ120の向きがその位置Nに対応する位置N'の方向に修正されるため、位置N'にある物体260の画像がユーザ10a用のディスプレイ110の中央に表示されることとなり、ユーザ10aの視線が再び自サイトのカメラ120の方に寄せられる。この状態では、位置N'にある物体260の画像がユーザ10aの目に映る(位置M'にあるユーザ10bの顔は映らないこともある)一方で、ユーザ10b向けに表示される画像がユーザ10aのダイレクトオン画像になる。即ち、非対称ながら、ユーザ10bだけがアイコンタクト感を味わうことができる。なお、視線補正アルゴリズムがユーザ10aの画像に適用されるため、ユーザ10aが位置N(N')を見ているか否かについて、ユーザ10b側に混乱が生じる可能性がある。

30

40

#### 【0129】

ジェスチャ追跡インタフェースとしての視線追跡は、更に、ユーザ10が表示画面115を見ながら(ちょうど窓の外を眺めて隅から隅へと視線を走らせるときのように)視線

50

や姿勢を横方向にずらし、そのことがコンテキスト伝達インタフェース 450 により検知されたときに実行される。例えば図 12A では、ユーザ 10 が端末 300 に備わるディスプレイ 110 の画面 115 を眺めており、その画面 115 には子供二人と犬一匹がいる場面が映っており、犬の姿はその画面 115 の縁の方であって体の一部しか見えていない。その犬の姿をもっとよく見たい場合、ユーザ 10 は、窓の縁の方へと視線を移していくときとちょうど同じように、自分の視線を画面 115 の縁の方へと動かしていく。即ち、自分の頭、顔 25、或いは更に身体 50 を動かすことで、自分の姿勢ひいては視線方向を変えていく。端末 300 は、こうしたユーザ 10 の姿勢又は位置変化を検知し、場面の向き及び位置を変化させる。即ち、図 12B に示すように、それまで縁の方に位置していた画像コンテンツが画面 115 の中央寄りに表示されるよう表示を調整する。この調整は、例えば、インタフェース 450 からの求めでリモートサイト 364 における撮影動作に修正を加え、ローカルサイト 362 で表示される画像が窓の外を眺めているときと同様に变化するよう、そのサイト 364 における撮影視野 420 (その撮影対象部分) を横方向に移動させることで行う。

10

20

30

40

50

#### 【0130】

そのコンテキスト伝達インタフェース 450 には、更に、視線追跡及び視線補正のどちらを優先的に適用するかを決定する役割がある。例えば、ジェスチャ追跡時の視線追跡では、ユーザ側の混乱を抑えるため視線補正付の視線追跡よりも視線補正抜きで視線追跡を優先させ、またコンテキストの混乱を抑えるため単なる視線追跡より撮影視野ズームアウト付の視線追跡を優先させる。撮影視野ズームアウト付での実行を優先させるのは、ローカルユーザ 10a やリモートユーザ 10b のポートレイトを接写するより、ユーザ 10a やユーザ 10b を取り巻く広い空間を撮影した方が、前述の通りより豊富にコンテキストを伝達して視線方向関連の抵触問題を抑えられるためである。

#### 【0131】

図 11C に、端末 300 で使用できるジェスチャ追跡インタフェースの別例を示す。この例では、ユーザ 10a がレーザポインタ等のポインタ装置 192 を手にしており、自分の好きな場所例えば位置 M をポイントすることができる。また、そのポインタから発せられる光ビームの射突位置を検知するプロンプトディテクタ 194 例えば二次元センサが、別途設けられている。コンピュータ 340 は、このポインタ利用型ディテクタ 194 の出力に基づく指令を、ネットワーク 360 経由でリモートサイト 364 側のカメラ 120 に送信する。これら、ポインタ装置 192 及びディテクタ 194 は、基本的にジェスチャ追跡用ツールの一種であるといえよう。端末 300 では、このほかにも様々な形態でジェスチャ追跡を行うことができる。例えば、ジェスチャ解析アルゴリズムに則り画像を解析し、それにより識別される手、腕、頭、身体等の動きに追従するようにしてもよい。

#### 【0132】

こうしたジェスチャ追跡インタフェース、例えば図 11C に示すようにポインタを利用するタイプのそれには、図 11A、図 11B、図 12A 及び図 12B に示した単純な視線追跡インタフェースに比べてロバスタな面がある。例えば、ユーザ 10 がジェスチャ言語を習得する必要があるものの、ユーザ 10 の意図が誤解されることが少なくなる。更には、ユーザアクションの解釈を通じ人間同士のやりとりの内容を察知することができる様々な他種インタフェース乃至アルゴリズムを、コンテキスト伝達インタフェース 450 で利用可能であることも、重要なことである。例えば、表情、言葉、対話内容等のキューを手がかりに人間同士のやりとりを察知するインタフェースや、やりとりの状況と関連しており大部分無意識的に生まれてくる非言語キューを手がかりに人間同士のやりとりの内容を察知するインタフェースである。非言語キューとは、話者が話し相手に指図するときや話者が自分の気持ちを仄めかすときに使用する手段のことであり、ジェスチャ等も含めて韻律と総称されている。その韻律には、声の音調、拍子取り、ラウドネス等、意識的に発現され又は無意識的に発生する様々な発声要素が含まれる。本発明のシステムでそうした非言語キューを使用する際には、ある撮影視野又は動画像コンテキストの撮影から別の撮影視野又は動画像コンテキストへの変転に対応するための案内手段として使用するのが基本

であるが、プライバシー管理やジェスチャ追跡でもそうした非言語キューを使用することができる。

#### 【 0 1 3 3 】

更に、個々の人物の動画像をローカルサイトリモートサイト間で交換する際にアニメーション技術を使用できることにも留意すべきである。例えば、特許文献 4 0 ( 発明者 : Ch e i k y e t a l . 、 名 称 : フォトリアリスティックトーキングヘッド生成システム及び方法 ( P h o t o R e a l i s t i c T a l k i n g H e a d C r e a t i o n S y s t e m a n d M e t h o d ) ) に記載の方法では、人物 ( 被写体 ) の姿を複数通り撮影して画像ライブラリを構築する。具体的には、その被写体の姿を様々な方向から撮影することで、目や口の外見的特徴を様々な方向から捉えた何枚かの画像を取得し、それらの画像を用い通話用頭部画像 ( トーキングヘッド ) のフォトリアリスティック画像ライブラリを構築する。次いで、そのフォトリアリスティック画像ライブラリを参照することで、フォトリアリスティックな画像になるよう、即ち誰が何をどのような気持ちで伝えようとしているのかを読み取れる画像になるよう、その被写体のアニメーション画像即ちアバタを生成する。そのアバタ、即ち被写体を模した動画像類似の画像はインターネット等のネットワークを介してよそに送信することができ、それに必要な帯域幅は狭く、しかも圧縮由来の偽像例えばブラーが少ないため、最終的な画質が良好になる。この文献に記載の方法の想定用途はニュースキャスティング等であるが、居宅用の動画通信システム 2 9 0 でもこうしたアバタの使用は可能である。

10

#### 【 0 1 3 4 】

なお、以上の説明では、実施形態に係る端末 3 0 0 で動画像 ( 基本的にはテレビジョン信号中の画像部分 ) 及びそれに付随する音声生成されることを示したが、端末 3 0 0 でデジタルスチルカメラを利用することや、画像処理で動画像データストリームから静止画像を抽出する構成にすることも可能であるので、その点をご理解頂きたい。例えば、動画通信システム 2 9 0 から出力される動画像にキーフレーム抽出アルゴリズムを適用することで、そのなかで最も構図及び顔の映り方がよいフレームを識別し、そのフレームを静止画像として抽出することができる。そのシステム 2 9 0 乃至端末 3 0 0 では、更に、動画像又は静止画像の画像データに付随又はリンクするかたちで、含意データ等のメタデータを生成することができる。このメタデータに含めうる情報としては、例えば日時、ローカルユーザ及びリモートユーザの身元、イベント種別、音声認識ソフトウェアで抽出したキーワード、その通信イベントに係るプライバシー保護セッティング、ユーザが入力した注釈やタイトル等々といったデータがある。そうしたメタデータがあると、システム 2 9 0 又はその端末 3 0 0 で生成された動画像、静止画像、音声等のデータを書庫に保存するときや、そのデータを書庫から読み出すときに便利である。

20

30

#### 【 0 1 3 5 】

また、動画通信システム 2 9 0 が基本的には個人居宅向けのテレビ電話システムであるとして説明したが、このシステム 2 9 0 はそれ以外の種類の居宅間通信、とりわけプライバシー保護やコンテキスト伝達が画像管理上の問題になるものにも拡張することができる。例えば、在宅勤務中のローカルユーザ 1 0 は、遠隔会議の開催にこのシステム 2 9 0 を利用することができる。そうすれば、私宅の状況をあからさまにすることなく自分の業務を遂行することができる。また、ローカルユーザ 1 0 は、ホームセキュリティ管理を支援する手段としてこのシステム 2 9 0 及びその端末 3 0 0 を利用することができる。例えば、ローカルユーザ 1 0 は、自宅とは別の場所にいるときでも、システム 2 9 0 及び端末 3 0 0 に接続し、パスワード等によるセキュリティチェックを受け、カメラ 1 2 0 で撮影された画像をシステム 2 9 0 に要求することで、自宅で何が起こったのか或いは今何が起きているのかを知ることができる。

40

#### 【 0 1 3 6 】

更に、動画通信システム 2 9 0 及びその端末 3 0 0 による個人的通信を拡張して新出の通信形態で利用すること、例えばユーザがワールドワイドウェブ上に動画像データをエクスポートする際に利用することもできる。具体的には、システム 2 9 0 を利用し、メタベース ( 登録商標 ) 、 S e c o n d L i f e ( 登録商標 ) 等のオンラインコミュニティや、

50



YouTube（登録商標）等のビデオブログを初め、様々な場所に画像を送り込むことができる。自明な通り、そうしたコミュニティ類を現在利用しているオンラインユーザの多くは、プライバシー保護及びコンテキスト伝達の機能に多くの期待を寄せている。その期待は、その種のインターネット技術の採用にあまり積極的でない他の多くの人々が抱いているものとはまた違った水準のものである。然るに、システム 290 及び端末 300 にはプライバシー保護及びコンテキスト伝達を管制する機能が備わっている。ユーザ 10 は、インターネット向け動画像コンテンツの生成を管制する手段として、この機能を利用することができる。その管制の内容は、これまでは得ることができなかった水準のものである。

#### 【0137】

10

また、端末 300 の使い途はほかにも色々ある。まず、ユーザ 10 は自サイトの端末 300 を前述の如くテレビジョン受像機として使用することができる。動画通信に使用されていないときに、端末 300 を一種のデジタルフォトフレーム(electronic picture frame)として稼働させ、ローカル環境にいる人々の個人的関心を惹くような静止画像乃至動画像を表示させることもできる。そのデジタルフォトフレームモードは受動的な形態でも受動的な形態でも実行することができる。能動的に実行する場合は、端末 300 が自サイトにいるユーザ 10 を検知及び識別し、識別したユーザ 10 の（既知の）興味関心事項に応じた静止画像又は動画像を表示させる。居宅内の様々な場所に都合複数個の電子イメージング装置 100 が配されている端末 300 で能動的なデジタルフォトフレームモードを実行する場合は、そのサイトのユーザ 10 が居宅内を動き回りながら画像を見ることができるようそれらの装置 100 で画像を表示させる。

20

#### 【0138】

更に、より興味深い使い途としては、住人（特に老人）がその住宅に居ながらにして健康状態管理を受けるネット見守り住宅(smart medical home)がある。この種の住宅では、その内部の様々な場所、例えば家具、床、電化製品、薬棚等にセンサを配すると共に、その住人を見守るためのカメラを配する。カメラで撮られるのは住人にとりかなり煩わしいことであるので、そのカメラは、例えばデジタルフォトフレームの背後に隠れるように配される。ただ、このように目立たないかたちでカメラを配しても、自分を“監視”するカメラが存在していること自体が、住人にとっては不快なことである。従って、こうした構成を採るよりは、実施形態に係る端末 300 をそうした住宅に配した方がよい。そうすれば、その住人はリモートユーザ、例えば遠地にいる友人や親戚、介護担当者、看護師等と通話することができる。その介護担当者や看護師は、画像を見る必要が生じたときに、住人側の端末 300 のプライバシー保護インタフェース 400 に係るローカルなプライバシー保護セッティングを、一時的にオーバーライドすることができる。更に、住人たるローカルユーザ 10 は撮影に関する警告を受けることができる。その警告は、例えば、介護担当者や看護師の姿を示す画面分割画像 410 の表示と、ローカルユーザ 10 の姿やローカル環境 415 の状況を示す画面分割画像 410 の表示とにより行われる。

30

#### 【0139】

また、動画通信システム 290 及びその端末 300 は、プライバシー保護やコンテキスト伝達の管理が重視される用途であれば、居宅間通信以外の用途でも使用することができる。例えば、教育用にシステム 290 を使用すれば、生徒と教師がシステム越しにやりとりすることができ、ビジネス環境用に使用すれば、別々の場所にいる同僚達が動画通信を通じ仮想的に一つのデスクを共有して共用作業空間を作り出すことができる。更に、以上の説明では、基本的に、ローカルユーザ 10 a 及びその環境 415 を示す画面分割画像 410 と共に何人かのリモートユーザ 10 b の姿をローカルディスプレイ 110 に表示させるものとしてシステム 290 を説明してきたが、システム 290 では、ユーザ 10 b がいないときのリモートサイト 364 の環境、例えばユーザ 10 b が一時的にいないときのリモート環境や、リモート環境のなかで元々ユーザ 10 b がいない個所の状況を、表示させることもできる。そして、システム 290 の更なる用途としては、リモートサイトから継続的に送られてくる動画像をディスプレイに表示させ、それによって仮想的な窓

40

50

を作り出す、という用途もある。その“窓”には、例えば休日をよく過ごす場所等の風景等、リモートサイトの状況を映し出すことができる。

#### 【0140】

そうした幻想的な窓には、他に、自然界や舞踏会の光景をテーマにした番組等も映し出すことができる。端末300は、そうした番組を見ているローカルユーザ10aの姿や、その番組に対するユーザ10aの反応を、予め定められているプライバシー保護セッティング及びコンテキスト伝達セッティングに従いながら撮影する。この種の稼働モデルは、教育、市場調査、動向調査、薬物調査等といった様々な分野で、また様々な用途で使うことができよう。更に、その動画通信システム290及び端末300を用い多者間通信イベントを開催する際、個々のサイトで、表示される画像のうち1個を共通の動画像（例えばテレビジョン番組の画像）とし、他の何個かをユーザ10a又はリモートユーザ10bの画像にすることもできる。例えば、図6に示したディスプレイ110に表示される三種類の画像のうち、左側の画像410'をテレビジョン番組の画像、右側の画像410''をユーザ10bの画像、挿入される画面分割画像410をユーザ10aの画像にする、といった具合である。このようにすれば、ユーザ10a及び10bが、面白い番組（例えばスポーツ番組）を一緒に見たかのような感覚を共有することができる。

10

#### 【0141】

更に、本発明を実施する形態がその用途によって様々に変わることをご理解頂けよう。即ち、その双方向動画通信システムに備わるべき撮影機能、画像処理機能、音声機能、デジタル信号プロセッサ、送受信機能、表示機能等の機能の種類は、そのシステムの用途次第で様々に変わってくるし、備わる機能の種類が違えば画像ベースでの入力を利用できる度合いも違ってくる。例えば大型の壁掛けユニットなら、高解像度の撮影装置や強力なデジタル信号プロセッサを搭載できるため、身振り、手振り、音声コマンド等による入力を検知可能な構成にすることができるが、携帯電話では、低解像度の撮影装置や非力なデジタル信号プロセッサしか搭載できないため、身振りや手振りを検知できず恐らくは音声コマンドの検知や動き検知に限られるであろう。ご理解頂けるように、最善の性能が発揮されるのは、ローカルサイト362の端末300とリモートサイト364の端末300が同一又は同等のものである場合である。

20

#### 【0142】

また、いわゆる当業者には自明な通り、好適な実施形態に係る双方向動画通信システムでは、前述した情報以外に、静止画像、テキスト、スプレッドシート、録音済音声等の情報乃至画像を双方向通信に供し、通話に参加している何人かの間でそれらを共有することができる。例えば、図6中の画面分割画像410'と同様にして、テキスト文書を別途表示させることができる。

30

#### 【0143】

以上、その好適な実施形態のうち幾つかを具体的に参照しつつ、本発明に関して詳細に説明したが、それらに変形や改良を施すことが可能であることや、その変形乃至改良を本発明の本質的な技術的範囲で行うことが可能であることを了解されたい。強調すべきことに、本願記載の装置乃至方法は様々なシステム構成で実施することができ、またそのシステムで使えるハードウェア及びソフトウェアの種類も非常に多様である。特に、動画通信システム290及びその端末300で場面解析アルゴリズム、場面内調整アルゴリズム、場面解析ルール及びコンテキスト種別が使用されること、それらがコンテキスト伝達インタフェース450の自動稼働要素であること、そしてそれらの要素による動画像の撮影、解析及び調整機構が一連の手順（図7A～図7D参照）に従い連携して動作することを説明してきたが、これらに比肩しうる別の手法、機構及び連携動作で、或いは時間的順序の入替で、別紙特許請求の範囲に記載の機能を実現することもできる。そして、別紙図面は、本発明の好適な実施形態で使われる部材及び原理のうち主立ったものを説明するためのものであり、実物を均等に縮尺したものではない。

40



【図 4 A】

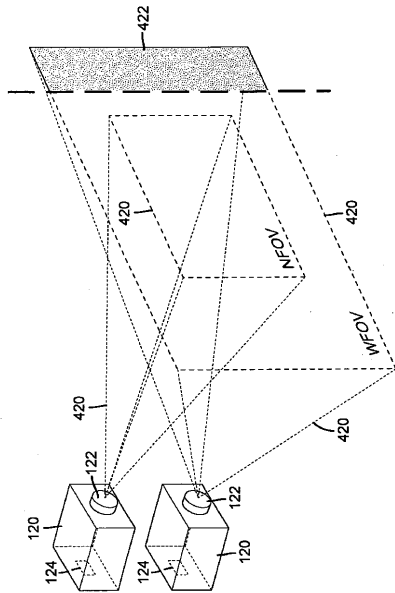


FIG. 4A

【図 4 B】

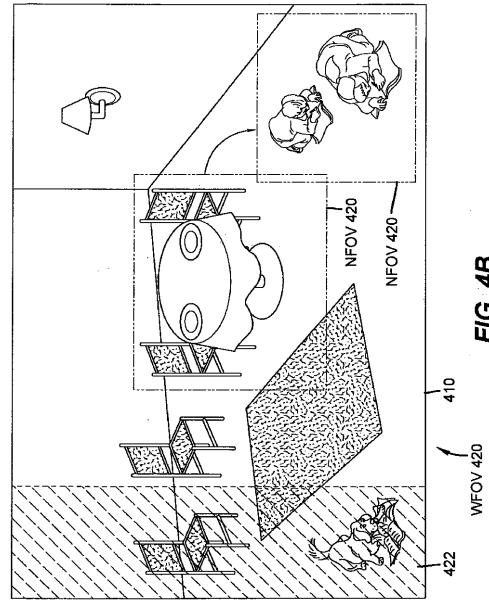


FIG. 4B

【図 4 C】

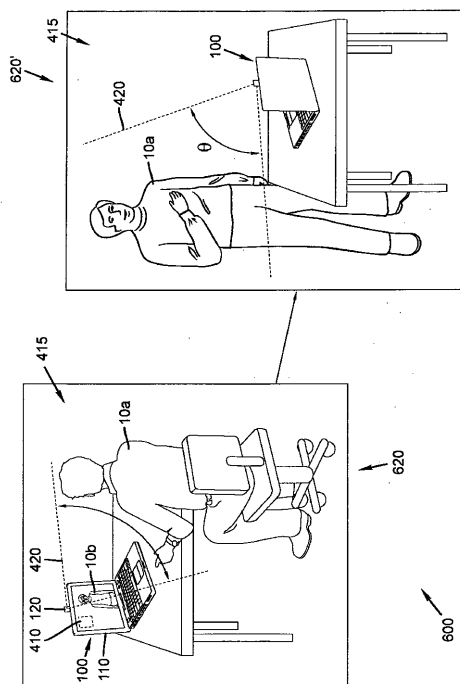


FIG. 4C

【図 5 A】

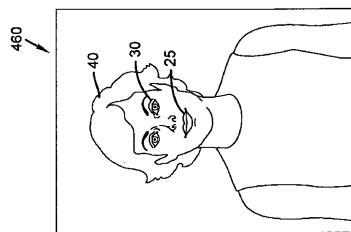


FIG. 5A

【図 5 B】

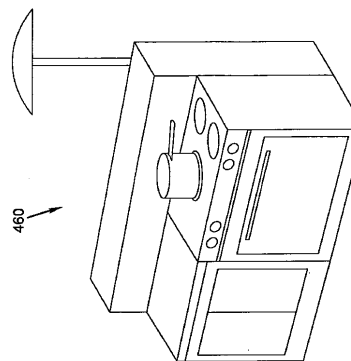


FIG. 5B

【図 6】

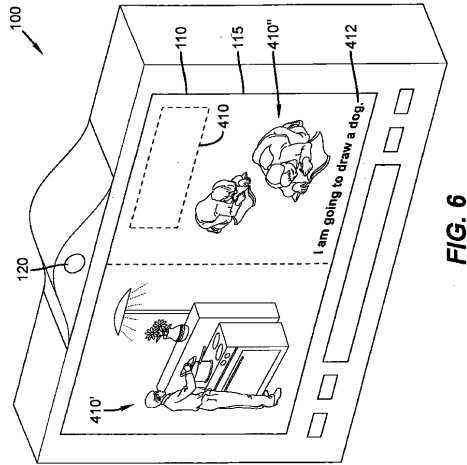
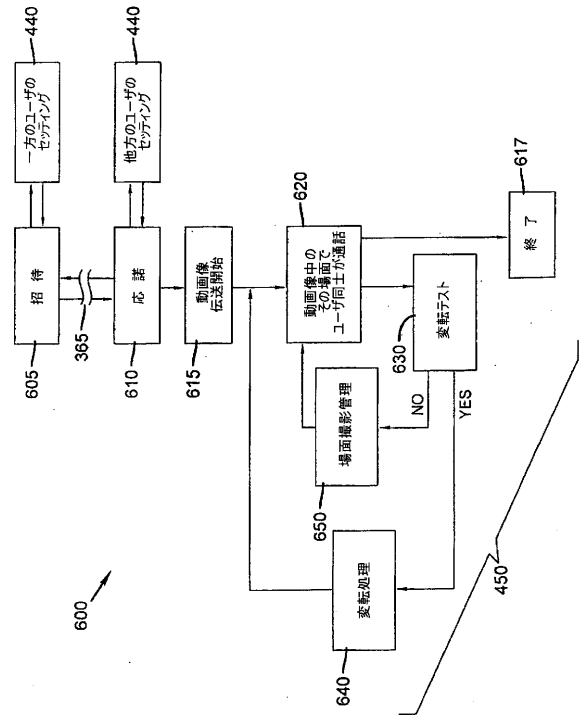
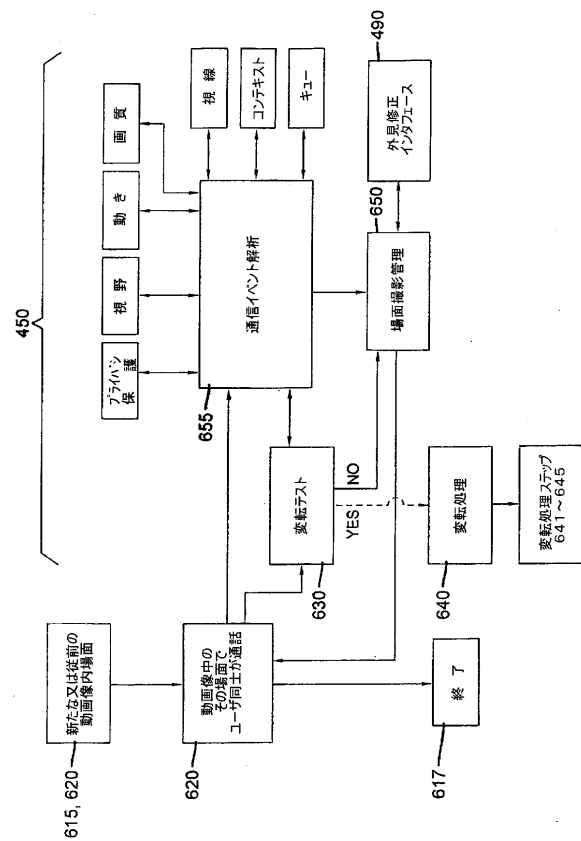


FIG. 6

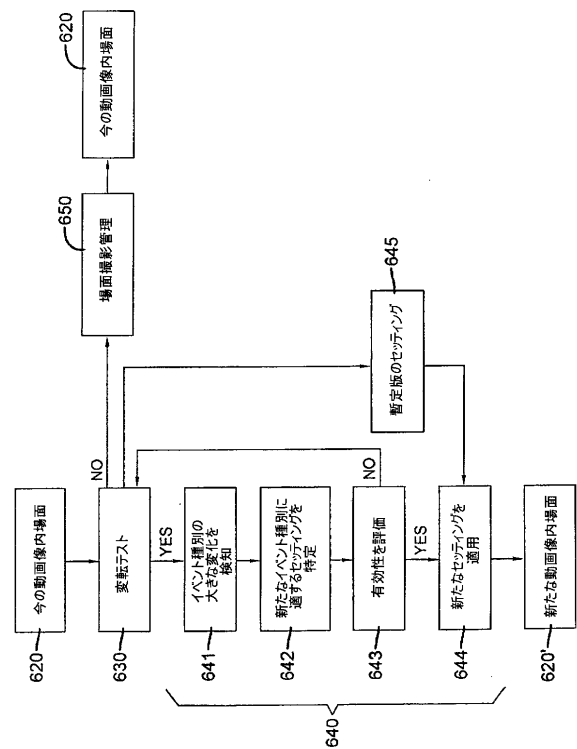
【図 7 A】



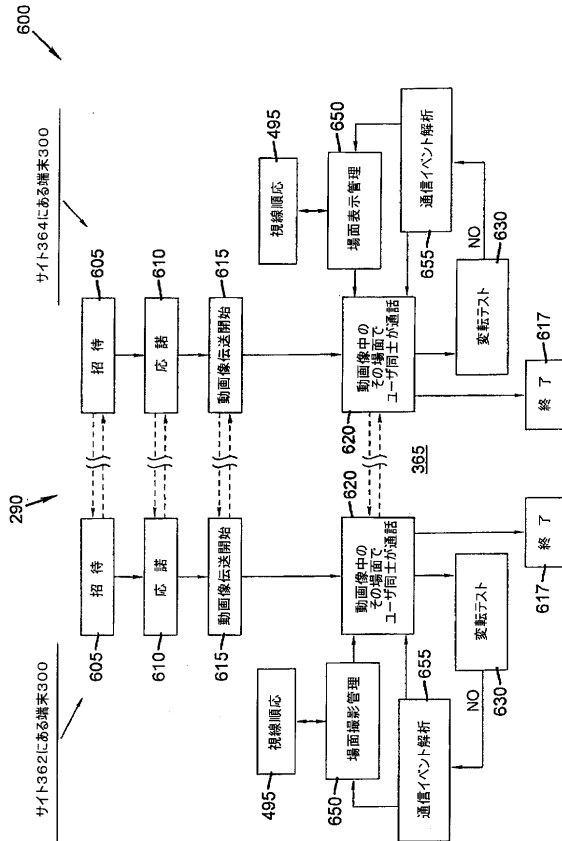
【図 7 B】



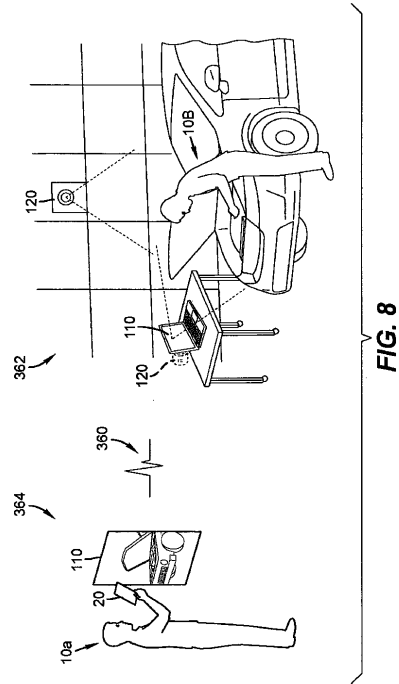
【図 7 C】



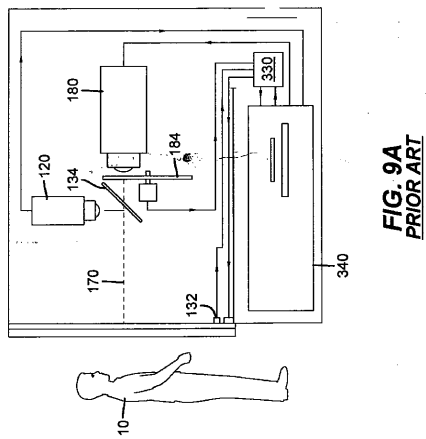
【図 7 D】



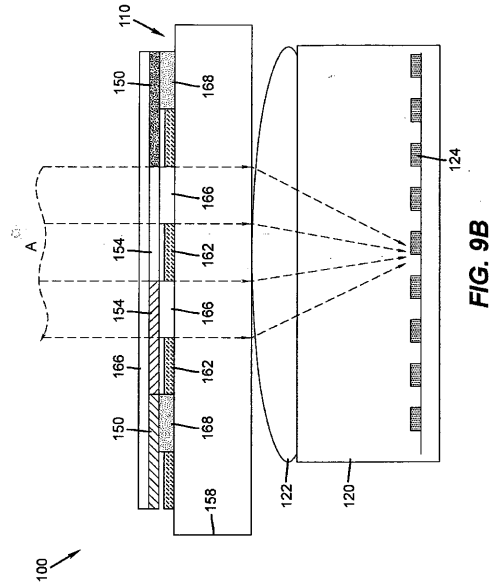
【図 8】



【図 9 A】



【図 9 B】



【図 10】

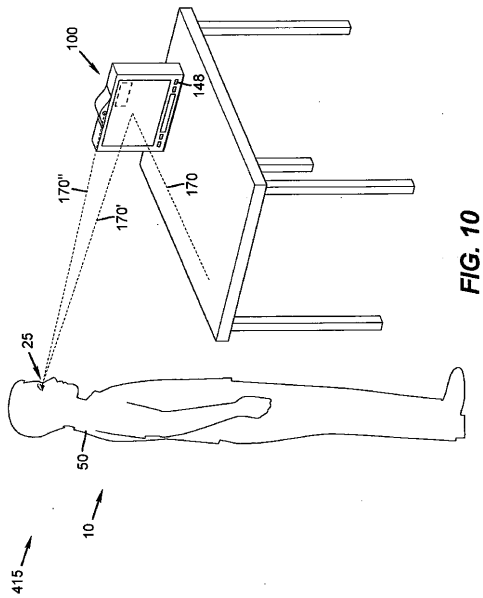


FIG. 10

【図 11A】

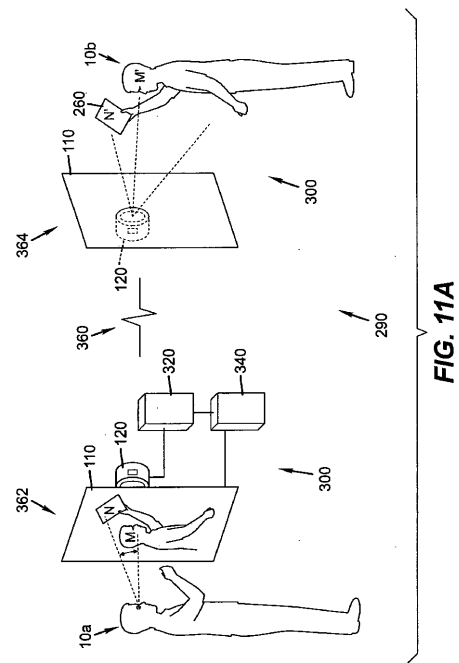


FIG. 11A

【図 11B】

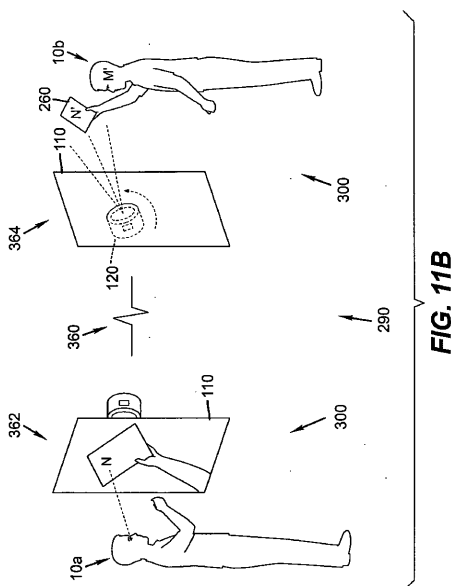


FIG. 11B

【図 11C】

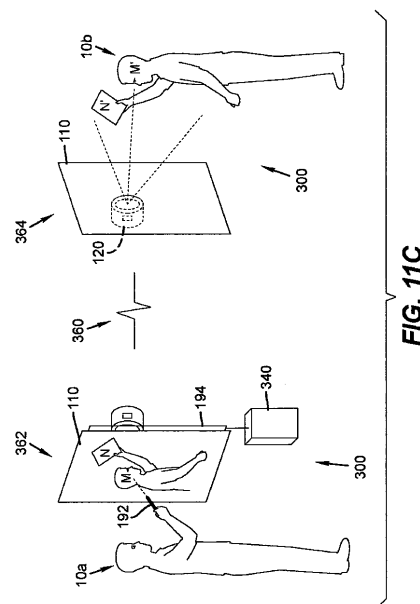
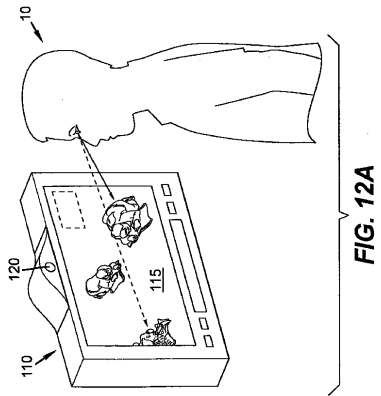
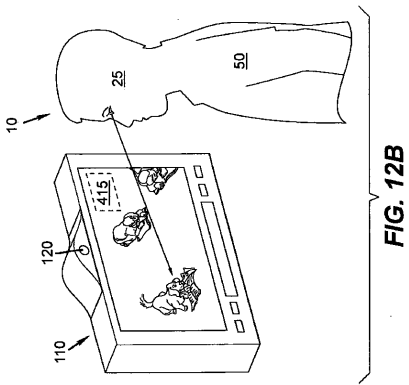


FIG. 11C

【図 12A】



【図 12B】





## 【国際調査報告】

## INTERNATIONAL SEARCH REPORT

International application No  
PCT/US2008/006807

<b>A. CLASSIFICATION OF SUBJECT MATTER</b> INV. H04N7/14 H04N7/15		
According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b> Minimum documentation searched (classification system followed by classification symbols) H04N		
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 practical, search terms used) EP0-Internal		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CARMAN NEUSTAEDTER AND SAUL GREENBERG: "The Design of a Context-Aware Home Media Space for Balancing Privacy and Awareness" UBICOMP 2003: UBIQUITOUS COMPUTING, vol. 2864/2003, 30 October 2003 (2003-10-30), pages 297-314, XP002531253 ISSN: 1611-3349 ISBN: 978-3-540-20301-8 the whole document	1-38
A	WO 00/38414 A (KONINKL PHILIPS ELECTRONICS NV [NL]) 29 June 2000 (2000-06-29) the whole document	1-38
A	WO 99/51023 A (3DV SYSTEMS LTD [IL]; BRAUN ORI J [IL]) 7 October 1999 (1999-10-07) the whole document	1-38
-/-		
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> 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 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. "Z" document member of the same patent family		
Date of the actual completion of the international search  9 June 2009		Date of mailing of the international search report  19/06/2009
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040 Fax (+31-70) 340-3016		Authorized officer  Blais, Denis

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

## INTERNATIONAL SEARCH REPORT

International application No  
PCT/US2008/006807

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EP 0 644 694 A (CANON KK [JP]) 22 March 1995 (1995-03-22) column 22, line 9 - line 45 column 24, line 25 - line 42	1-38

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

**INTERNATIONAL SEARCH REPORT**

Information on patent family members

International application No  
PCT/US2008/006807

Patent document cited in search report		Publication date		Patent family member(s)		Publication date
WO 0038414	A	29-06-2000	CN	1298601 A		06-06-2001
			DE	69938173 T2		19-02-2009
			EP	1057326 A1		06-12-2000
			ES	2301256 T3		16-06-2008
			JP	2002534009 T		08-10-2002
			US	7057636 B1		06-06-2006
WO 9951023	A	07-10-1999	AU	6516698 A		18-10-1999
EP 0644694	A	22-03-1995	DE	69424109 D1		31-05-2000
			DE	69424109 T2		12-10-2000
			DE	69434779 T2		14-06-2007

## フロントページの続き

(81)指定国 AP(BW, GH, GM, KE, LS, MW, MZ, NA, 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, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OA(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG), AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, 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, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW

(72)発明者 コステロ キャスリーン メアリー

アメリカ合衆国 ニューヨーク ロチェスター ステイト ストリート 3 4 3

(72)発明者 ジャコビィ キース エイ

アメリカ合衆国 ニューヨーク ロチェスター ステイト ストリート 3 4 3

(72)発明者 パラダ ロバート ジョン ジュニア

アメリカ合衆国 ニューヨーク ロチェスター ステイト ストリート 3 4 3

Fターム(参考) 5C164 FA10 PA43 UB08S UB41P UB89S VA04P VA07S VA35S YA12

5K201 CA01 CA06 CA10 CD06 DC05 EF04 EF10