(19)日本国特許庁(JP)　(12)公表特許公報(A)　(11)公表番号
特表2025-505569
(P2025-505569A)
(43)公表日　令和7年2月28日(2025.2.28)

(51)国際特許分類		FI		テーマコード（参考）
H04N	*21/234(2011.01)*	H04N	21/234	5C122
H04N	*23/695(2023.01)*	H04N	23/695	5C164
H04N	*23/69 (2023.01)*	H04N	23/69	

審査請求　未請求　予備審査請求　有　（全63頁）

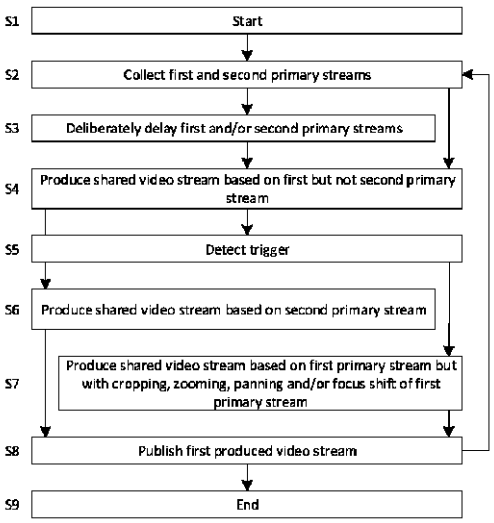
(21)出願番号　特願2024-545898(P2024-545898)
(86)(22)出願日　令和5年2月3日(2023.2.3)
(85)翻訳文提出日　令和6年10月1日(2024.10.1)
(86)国際出願番号　PCT/SE2023/050091
(87)国際公開番号　WO2023/149836
(87)国際公開日　令和5年8月10日(2023.8.10)
(31)優先権主張番号　2250113-4
(32)優先日　令和4年2月4日(2022.2.4)
(33)優先権主張国・地域又は機関
スウェーデン(SE)
(81)指定国・地域　AP(BW,CV,GH,GM,KE,LR,LS,MW,MZ,NA,RW,SD,SL,ST,SZ,TZ,UG,ZM,ZW),EA(AM,AZ,BY,KG,KZ,RU,TJ,TM),EP(AL,AT,BE,BG,CH,CY,CZ,DE,DK,EE,ES,FI,FR,GB,GR,HR,HU,IE,IS,IT,LT,LU,LV
最終頁に続く

(71)出願人　524135738
ライブアリーナ テクノロジーズ アクチエボラグ
ＬＩＶＥＡＲＥＮＡ ＴＥＣＨＮＯＬＯＧＩＥＳ ＡＢ
スウェーデン国，ストックホルム，１００ ６１，ボックス １１２０８
(74)代理人　100109380
弁理士　小西 恵
(74)代理人　100109036
弁理士　永岡 重幸
(72)発明者　ビョークマン，アンドレアス
スウェーデン国，タビー，１８７ ７７，ステンキルスヴァーゲン １３
(72)発明者　ニルソン，アンダース
最終頁に続く

(54)【発明の名称】　ビデオストリームを生成するシステムおよび方法

(57)【要約】
共有デジタルビデオストリームを提供する方法は、以下を含む：収集ステップ（Ｓ２）において、第１および第２のデジタルビデオストリームを収集し；第１の生成ステップ（Ｓ４）において、前記第１のソースが前記共有ストリームにおいて可視であり、前記第２のソースが前記共有ストリームにおいて可視でないように、前記第１のストリームに基づいて前記共有ストリームを生成し；遅延ステップ（Ａ３）において、前記第１および第２のデジタルビデオストリームに対して待ち時間を意図的に導入し；トリガ検出ステップ（Ｓ５）において、前記第１および／または第２のストリームを分析して、所定の画像および／または音声パターンの検出に基づいて、トリガを検出し；第２の生成ステップ（Ｓ６，Ｓ７）において、前記第２のストリームに基づいて、および／または前記第１のソースに基づいて、前記共有ストリームを生成するが、異なるクロッピング、異なるズーミング、異なるパンニング、または異なるフォーカスプレーン選択の少なくとも１つを行い；および、公開ステップ（Ｓ８）において、前記出力ストリームを消費者に提供す

Fig. 9



【特許請求の範囲】
【請求項１】
共有デジタルビデオストリームを提供する方法であって、該方法は以下のステップを含む方法：
収集ステップ（Ｓ２）において、第１のデジタルビデオソースから第１のデジタルビデオストリームを収集し、第２のデジタルビデオソースから第２のデジタルストリームを収集し；
第１の生成ステップ（Ｓ４）において、前記第１のデジタルビデオソースからの画像情報が前記共有デジタルビデオストリームにおいて可視であり、前記第２のデジタルビデオソースからの画像情報が前記共有デジタルビデオストリームにおいて可視でないように、前記第１のデジタルビデオストリームの連続的に考慮される複数フレームに基づいて、前記共有デジタルビデオストリームを出力デジタルビデオストリームとして生成し；
遅延ステップ（Ａ３）において、前記第１および第２のデジタルビデオストリームに対して、待ち時間を意図的に導入し；
トリガ検出ステップ（Ｓ５）において、前記第１および／または第２のデジタルビデオストリームをデジタル的に分析して、少なくとも１つのトリガを検出し、ここで、前記分析は、所定のタイプの画像および／または音声パターンの自動検出に基づいており、前記トリガ検出ステップ（Ｓ５）は、前記待ち時間を導入する前の前記第１および／または第２のデジタルビデオストリームに基づいて実行され、前記パターンの存在が前記トリガを構成し、前記トリガは、所定の生成ルールに従って、前記共有デジタルビデオストリームの生成モードを変更するよう指示するものであり；
前記トリガの検出に応答して開始される第２の生成ステップ（Ｓ６，Ｓ７）において、前記第２のデジタルビデオソースからの画像情報が前記共有デジタルビデオストリームにおいて可視であるように、前記第２のデジタルビデオストリームの連続して考慮される複数フレームに基づいて、前記共有デジタルビデオストリームを出力デジタルビデオストリームとして生成し、および／または、前記第１のデジタルビデオソースの連続的に考慮される複数フレームに基づいて、前記共有ビデオストリームを出力デジタルビデオストリームとして生成するが、前記第１の生成ステップ（Ｓ４）と比較して、前記第１のデジタルビデオストリームの異なるクロッピング、異なるズーミング、異なるパンニングまたは異なるフォーカスプレーン選択のうちの少なくとも１つを行い；および、
公開ステップ（Ｓ８）において、前記出力デジタルビデオストリームを、前記共有デジタルビデオストリームの消費者に継続的に提供し、ここで、前記待ち時間が前記共有デジタルビデオストリームに存在する。
【請求項２】
請求項１に記載の方法であって、
前記第１のデジタルビデオストリームは、第１のデジタルカメラ（４１０）によって連続的にキャプチャされ、
第２のデジタルビデオストリームは、第２のデジタルカメラ（４２０）によって連続的にキャプチャされる、方法。
【請求項３】
請求項２に記載の方法であって、
前記所定の画像および／または音声パターンは、前記第１のデジタルビデオストリームに描かれた参加者（４３０）が、前記第２のデジタルビデオストリームに描かれたオブジェクト（４４０）に向かって注視することを含み、ここで、前記所定の画像および／または音声パターンは、前記第１のカメラ（４１０）、前記参加者（４３０）、および前記オブジェクト（４４０）の相対的な向きに関する情報に基づいて検出され、前記所定の画像および／または音声パターンは、さらに、前記第１のデジタルビデオストリームに基づく、当該参加者（４３０）の身体方向、頭部方向、および視線方向の少なくとも１つのデジタル画像ベースの決定に基づいて検出される、方法。
【請求項４】

請求項１から３のいずれか１項に記載の方法であって、
前記所定の画像および／または音声パターンは、前記第１のデジタルビデオストリームに描かれた参加者（４３０）が所定のジェスチャをすることを含む、方法。
【請求項５】
請求項４に記載の方法であって、
前記ジェスチャは、前記第２のデジタルビデオストリームに描かれたオブジェクト（４４０）に関連する、例えば、前記オブジェクト（４４０）を指し示すジェスチャであり、前記所定の画像および／または音声パターンは、前記第１のカメラ（４１０）、前記参加者（４３０）、および前記オブジェクト（４４０）の相対的な向きに関する情報に基づいて検出され、前記所定の画像および／または音声パターンは、さらに、前記第１のデジタルビデオストリームに基づく当該参加者（４３０）のジェスチャ方向のデジタル画像ベースの決定に基づいて検出される、方法。
【請求項６】
請求項１から５のいずれか１項に記載の方法であって、
前記所定の画像および／または音声パターンは、前記第２のデジタルビデオストリームに描かれた参加者（４３０）が、前記第２のデジタルビデオストリームを連続的にキャプチャしている第２のカメラ（４２０）に向かって注視していることを含む、方法。
【請求項７】
請求項１から６のいずれか１項に記載の方法であって、
前記所定の画像および／または音声パターンは、前記第１および／または第２のデジタルビデオストリームにおける相対的な動きの変化を含む、方法。
【請求項８】
請求項１から７のいずれか１項に記載の方法であって、
前記所定の画像および／または音声パターンは、例えば特定の音捕捉マイクロフォンの相対的なマイクロフォン音量によって決定される、その周波数および／または振幅および／または振幅の時間微分値および／または絶対的な振幅の変化および／または音位置によって特徴付けられる所定の音声パターンを含む、方法。
【請求項９】
請求項８に記載の方法であって、
前記所定の音声パターンは、少なくとも１つの口頭で発声された単語を含む所定のフレーズを含む、方法。
【請求項１０】
請求項１から９のいずれか１項に記載の方法であって、
前記待ち時間は、少なくとも複数ビデオフレーム、例えば少なくとも３ビデオフレーム、例えば少なくとも５ビデオフレーム、例えば少なくとも１０ビデオフレームである、方法。
【請求項１１】
請求項１から１０のいずれか１項に記載の方法であって、
前記待ち時間は、最大３０秒、例えば最大５秒、例えば最大１秒、例えば最大０．５秒である、方法。
【請求項１２】
請求項１から１１のいずれか１項に記載の方法であって、
前記画像および／または音声パターンの前記検出は、前記第１および／または第２の一次デジタルビデオストリームの特定の情報を考慮することを含み、ここで、前記特定の情報は、前記出力デジタルビデオストリームの生成にまだ使用されていないフレームよりも後のフレームに存在する、方法。
【請求項１３】
共有デジタルビデオストリームを提供するコンピュータソフトウエア製品であって、該コンピュータソフトウエアの機能は実行される際に、以下のステップを実行するコンピュータソフトウエア製品：

収集ステップ（Ｓ２）において、第１のデジタルビデオソースから第１のデジタルビデオストリームを収集し、第２のデジタルビデオソースから第２のデジタルストリームを収集し；
第１の生成ステップ（Ｓ４）において、前記第１のデジタルビデオソースからの画像情報が前記共有デジタルビデオストリームにおいて可視であり、前記第２のデジタルビデオソースからの画像情報が前記共有デジタルビデオストリームにおいて可視でないように、前記第１のデジタルビデオストリームの連続的に考慮される複数フレームに基づいて、前記共有デジタルビデオストリームを出力デジタルビデオストリームとして生成し；
遅延ステップ（Ａ３）において、第１および第２のデジタルビデオストリームに対して、待ち時間を意図的に導入し；
トリガ検出ステップ（Ｓ５）において、前記第１および／または第２のデジタルビデオストリームをデジタル的に分析して、少なくとも１つのトリガを検出し、ここで、前記分析は、所定のタイプの画像および／または音声パターンの自動検出に基づいており、前記トリガ検出ステップ（Ｓ５）は、前記待ち時間を導入する前の第１および／または第２のデジタルビデオストリームに基づいて実行され、前記パターンの存在が前記トリガを構成し、前記トリガは、所定の生成ルールに従って、前記共有デジタルビデオストリームの生成モードを変更するよう指示するものであり；
前記トリガの検出に応答して開始される第２の生成ステップ（Ｓ６，Ｓ７）において、前記第２のデジタルビデオソースからの画像情報が前記共有デジタルビデオストリームにおいて可視であるように、前記第２のデジタルビデオストリームの連続して考慮される複数フレームに基づいて、前記共有デジタルビデオストリームを出力デジタルビデオストリームとして生成し、および／または、前記第１のデジタルビデオソースの連続的に考慮される複数フレームに基づいて、前記共有ビデオストリームを出力デジタルビデオストリームとして生成するが、前記第１の生成ステップ（Ｓ４）と比較して、前記第１のデジタルビデオストリームの異なるクロッピング、異なるズーミング、異なるパンニングまたは異なるフォーカスプレーン選択のうちの少なくとも１つを行い；および、
公開ステップ（Ｓ８）において、前記出力デジタルビデオストリームを、前記共有デジタルビデオストリームの消費者に継続的に提供し、ここで、前記待ち時間が前記共有デジタルビデオストリームに存在する。

【請求項１４】

共有デジタルビデオストリームを提供するシステム（１００）であって、該システム（１００）は中央サーバ（１３０）を備え、該中央サーバ（１３０）は以下の機能を備えるシステム：
第１のデジタルビデオソースから第１のデジタルビデオストリームを収集し、第２のデジタルビデオソースから第２のデジタルストリームを収集するよう構成される収集機能；
前記第１のデジタルビデオソースからの画像情報が前記共有デジタルビデオストリームにおいて可視であり、前記第２のデジタルビデオソースからの画像情報が前記共有デジタルビデオストリームにおいて可視でないように、前記第１のデジタルビデオストリームの連続的に考慮される複数フレームに基づいて、前記共有デジタルビデオストリームを出力デジタルビデオストリームとして生成するよう構成される第１の生成機能（１３５，１３５'，１３５''，１３５'''）；
第１および第２のデジタルビデオストリームに対して、待ち時間を意図的に導入するよう構成される遅延機能；
前記第１および／または第２のデジタルビデオストリームをデジタル的に分析して、少なくとも１つのトリガを検出するよう構成されるトリガ検出機能、ここで、前記分析は、所定のタイプの画像および／または音声パターンの自動検出に基づいており、前記分析は、前記待ち時間を導入する前の第１および／または第２のデジタルビデオストリームに基づいて実行され、前記パターンの存在が前記トリガを構成し、前記トリガは、所定の生成ルールに従って、前記共有デジタルビデオストリームの生成モードを変更するよう指示するものであり；

前記トリガの検出に応答して開始されるよう構成され、および、前記第２のデジタルビデオソースからの画像情報が前記共有デジタルビデオストリームにおいて可視であるように、前記第２のデジタルビデオストリームの連続して考慮される複数フレームに基づいて、前記共有デジタルビデオストリームを出力デジタルビデオストリームとして生成し、および／または、前記第１のデジタルビデオソースの連続的に考慮される複数フレームに基づいて、前記共有ビデオストリームを出力デジタルビデオストリームとして生成するが、前記第１の生成機能（１３５，１３５'，１３５''，１３５'''）と比較して、前記第１のデジタルビデオストリームの異なるクロッピング、異なるズーミング、異なるパンニングまたは異なるフォーカスプレーン選択のうちの少なくとも１つを行うよう構成される第２の生成機能（１３５，１３５'，１３５''，１３５'''）；および、

前記出力デジタルビデオストリームを、前記共有デジタルビデオストリームの消費者に継続的に提供するよう構成される公開機能、ここで、前記待ち時間が前記共有デジタルビデオストリームに存在する。

【発明の詳細な説明】

【技術分野】

【０００１】

本発明は、デジタルビデオストリームを生成するための、特に、２つ以上の異なるデジタル入力ビデオストリームに基づいてデジタルビデオストリームを生成するための、システム、コンピュータソフトウエア製品および方法に関する。好ましい実施形態において、デジタルビデオストリームは、特に複数の異なる同時接続ユーザが参加するデジタルビデオ会議、またはデジタルビデオ会議ないしミーティングシステムのコンテキストにおいて生成される。生成されたデジタルビデオストリームは、外部に公開(published)されてもよいし、デジタルビデオ会議またはデジタルビデオ会議システム内で公開されてもよい。

【０００２】

他の実施形態では、本発明は、デジタルビデオ会議ではないが、複数のデジタルビデオ入力ストリームが同時に処理され、生成されるべきデジタルビデオストリームに結合されるコンテキストに適用される。例えば、このようなコンテキストは、教育的または指導的なものであってよい。

【背景技術】

【０００３】

Microsoft（登録商標）Teams（登録商標）、Zoom（登録商標）、Google（登録商標） Meet（登録商標）など、多くのデジタルビデオ会議システムが知られており、２人以上の参加者が、ローカルで録画されたデジタルビデオとオーディオを使用してバーチャルに会議を行い、参加者全員にブロードキャストすることで、物理的な会議をエミュレートすることができる。

【０００４】

このようなデジタルビデオ会議ソリューションを改善する一般的なニーズがあり、特に、視聴コンテンツの生成（制作）に関して、どのようなコンテンツを、どのような時間に、誰に、どのような配信チャネルを通じて見せるかといったことが重要である。

【０００５】

たとえば、現在発話中の参加者を自動的に検出し、その発話中の参加者の対応するビデオフィード(video feed)を他の参加者に表示するシステムもある。多くのシステムでは、現在表示されている画面、ビューイングウィンドウ、デジタルプレゼンテーションなどのグラフィックを共有することが可能である。しかし、バーチャルミーティングがより複雑になるにつれ、現在利用可能なすべての情報のうち、どの情報を各時点で各参加者に表示すべきかをサービス側が把握することは、すぐに難しくなるであろう。

【０００６】

他の例では、プレゼンテーションしている参加者がデジタルプレゼンテーションのスライドについて話しながらステージ上を動き回る。この場合、システムは、プレゼンテーシ

ョンを表示するか、発表者を表示するか、あるいはその両方を表示するか、あるいはその２つを切り替えるかを決定する必要がある。

【０００７】

自動生成プロセスによって、多数の入力デジタルビデオストリームに基づいて１つまたは複数の出力デジタルビデオストリームを生成し、そのような生成されたデジタルビデオストリームまたはストリーム群を１人または複数の消費エンティティに提供することが望ましい場合がある。

【発明の開示】

【発明が解決しようとする課題】

【０００８】

しかしながら多くの場合、このようなデジタルビデオ会議システムが直面する多くの技術的な困難のために、動的な会議画面のレイアウトマネージャや他の自動化された生成機能が、どのような情報を表示すべきかを選択することは困難である。

【０００９】

第１に、デジタルビデオ会議はリアルタイム性を重視するため、遅延が小さいことが重要である。異なる参加者が異なるハードウエアを使用して参加するなど、異なる受信デジタルビデオストリームが、異なる遅延、異なるフレームレート、異なるアスペクト比、または異なる解像度と関連付けられる場合、これは問題となる。多くの場合、このような受信デジタルビデオストリームは、整形された(well-formed)ユーザ体験のための処理が必要とする。

【００１０】

第２に、時間同期の問題がある。外部デジタルビデオストリームや参加者が提供するデジタルビデオストリームなどの、多様な入力デジタルビデオストリームは、典型的には中央サーバ等に供給されるため、そのようなデジタルビデオフィードをそれぞれ同期させる絶対的な時間は存在しない。大き過ぎる遅延と同様に、同期されていないデジタルビデオフィードは、ユーザ体験の低下につながる。

【００１１】

第３に、複数参加者間のデジタルビデオ会議は、異なるエンコーディングまたはフォーマットを持つ異なるデジタルビデオストリームを含む可能性があり、これはデコードと再エンコーディングを必要とし、遅延と同期の点で問題が生じる。また、このようなエンコーディングは、計算負荷が高く、ハードウエア要件の点で高価となる。

【００１２】

第４に、異なるデジタルビデオソースが異なるフレームレート、異なるアスペクト比、異なる解像度に関連付けられる可能性があるという事実は、メモリ割り当てのニーズが予測不可能に変化すること、継続的なバランスを必要とし得る。その結果、遅延や同期の問題が追加的に発生する可能性がある。その結果、大きなバッファが必要となる。

【００１３】

第５に、参加者は、接続性の変動、離脱／再接続などの点で様々な困難を経験する可能性があり、整形されたユーザ体験を自動的に生成する上でさらなる困難をもたらす。

【００１４】

これらの問題は、多数の参加者がいる；参加者が異なるハードウエアおよび／またはソフトウエアを使用して接続している；外部から提供されたデジタルビデオストリームを使用している；画面共有をしている；複数のホストを使用しているなど、より複雑な会議の状況で増幅される。

【００１５】

教育や指導のためのデジタルビデオ生成システムのように、複数の入力デジタルビデオストリームに基づいて出力デジタルビデオストリームを生成すべき場合、対応する問題が他のコンテキストで生じる。

【００１６】

スウェーデン特許出願ＳＥ２１５１２６７－８（本出願の有効優先日時点では未公開）

は、上述の問題に対する様々な解決策を開示している。

【００１７】

スウェーデン特許出願ＳＥ２１５１４６１－７は、本出願の有効日時点では未公開であるが、異なる参加者グループが異なる一般的な待ち時間と関連している場合など、多数参加者デジタルビデオ環境における待ち時間の処理に特化した多様な解決手段を開示している。

【００１８】

上述のようなタイプの会議の自動生成を実現するには、まだ問題がある。特に、このような自動化された生成で複数のカメラが利用可能な場合、会議の参加者にとって自然かつ直感的な方法で、そのようなカメラのそれぞれの画像出力を使用することは困難であることが判明している。

【００１９】

本発明は、上述した問題の１つまたは複数を解決するものである。

【課題を解決するための手段】

【００２０】

したがって、本発明は、第２のデジタルビデオストリームを提供する方法に関し、該方法は以下のステップを含む：収集ステップにおいて、第１の参加者クライアントから第１の一次デジタルビデオストリームを収集し、第２の参加者クライアントから第２の一次デジタルビデオストリームを収集し、第３の参加者クライアントから第３の一次デジタルビデオストリームを収集し；公開ステップにおいて、前記第１の参加者クライアントおよび前記第２の参加者クライアントの少なくとも１つに対して、前記第１の一次デジタルビデオストリーム、前記第２の一次デジタルビデオストリーム、および、前記第１および第２の一次ビデオストリームの少なくとも１つに基づいて生成された第１の生成ビデオストリームの少なくとも１つを提供し；第２の生成ステップにおいて、前記第１の一次デジタルビデオストリーム、前記第２の一次デジタルビデオストリーム、および前記第３の一次デジタルビデオストリームに基づいて、前記第２の生成ビデオストリームをデジタルビデオストリームとして生成し、ここで、記第２の生成ステップは、前記第２の生成ビデオストリームが、前記公開ステップにおいて前記第１または第２の参加者クライアントに提供されるいずれのビデオストリームとも時間的に非同期であるように、時間遅延を導入し；ここで、前記公開ステップはさらに、前記第２の生成ビデオストリームを、前記第１または第２の参加者クライアントではない少なくとも１つの消費者クライアントに連続的に提供することを含む。

【００２１】

本発明はまた、第２のデジタルビデオストリームを提供する方法に関し、該方法は以下のステップを含む：収集ステップにおいて、第１のデジタルビデオソースから第１のデジタルビデオストリームを収集し、第２のデジタルビデオソースから第２のデジタルストリームを収集し；第１の生成ステップにおいて、前記第１のデジタルビデオソースからの画像情報が前記共有デジタルビデオストリームにおいて可視であり、前記第２のデジタルビデオソースからの画像情報が前記共有デジタルビデオストリームにおいて可視でないように、前記第１のデジタルビデオストリームの連続的に考慮される複数フレームに基づいて、前記共有デジタルビデオストリームを出力デジタルビデオストリームとして生成し；遅延ステップにおいて、前記第１および第２のデジタルビデオストリームに対して、待ち時間を意図的に導入し；トリガ検出ステップにおいて、前記第１および／または第２のデジタルビデオストリームをデジタル的に分析して、少なくとも１つのトリガを検出し、ここで、前記分析は、所定のタイプの画像および／または音声パターンの自動検出に基づいており、前記トリガ検出ステップは、前記待ち時間を導入する前の前記第１および／または第２のデジタルビデオストリームに基づいて実行され、前記パターンの存在が前記トリガを構成し、前記トリガは、所定の生成ルールに従って、前記共有デジタルビデオストリームの生成モードを変更するよう指示するものであり；前記トリガの検出に応答して開始される第２の生成ステップにおいて、前記第２のデジタルビデオソースからの画像情報が前

記共有デジタルビデオストリームにおいて可視であるように、前記第２のデジタルビデオストリームの連続して考慮される複数フレームに基づいて、前記共有デジタルビデオストリームを出力デジタルビデオストリームとして生成し、および／または、前記第１のデジタルビデオソースの連続的に考慮される複数フレームに基づいて、前記共有ビデオストリームを出力デジタルビデオストリームとして生成するが、前記第１の生成ステップと比較して、前記第１のデジタルビデオストリームの異なるクロッピング、異なるズーミング、異なるパンニングまたは異なるフォーカスプレーン選択のうちの少なくとも１つを行い；および、公開ステップにおいて、前記出力デジタルビデオストリームを、前記共有デジタルビデオストリームの消費者に継続的に提供し、ここで、前記待ち時間が前記共有デジタルビデオストリームに存在する。

【００２２】

本発明はまた、共有デジタルビデオストリームを提供するコンピュータソフトウエア製品に関し、該コンピュータソフトウエアの機能は実行される際に、以下のステップを実行する：収集ステップにおいて、第１のデジタルビデオソースから第１のデジタルビデオストリームを収集し、第２のデジタルビデオソースから第２のデジタルストリームを収集し；第１の生成ステップにおいて、前記第１のデジタルビデオソースからの画像情報が前記共有デジタルビデオストリームにおいて可視であり、前記第２のデジタルビデオソースからの画像情報が前記共有デジタルビデオストリームにおいて可視でないように、前記第１のデジタルビデオストリームの連続的に考慮される複数フレームに基づいて、前記共有デジタルビデオストリームを出力デジタルビデオストリームとして生成し；遅延ステップにおいて、第１および第２のデジタルビデオストリームに対して、待ち時間を意図的に導入し；トリガ検出ステップにおいて、前記第１および／または第２のデジタルビデオストリームをデジタル的に分析して、少なくとも１つのトリガを検出し、ここで、前記分析は、所定のタイプの画像および／または音声パターンの自動検出に基づいており、前記トリガ検出ステップは、前記待ち時間を導入する前の第１および／または第２のデジタルビデオストリームに基づいて実行され、前記パターンの存在が前記トリガを構成し、前記トリガは、所定の生成ルールに従って、前記共有デジタルビデオストリームの生成モードを変更するよう指示するものであり；前記トリガの検出に応答して開始される第２の生成ステップにおいて、前記第２のデジタルビデオソースからの画像情報が前記共有デジタルビデオストリームにおいて可視であるように、前記第２のデジタルビデオストリームの連続して考慮される複数フレームに基づいて、前記共有デジタルビデオストリームを出力デジタルビデオストリームとして生成し、および／または、前記第１のデジタルビデオソースの連続的に考慮される複数フレームに基づいて、前記共有ビデオストリームを出力デジタルビデオストリームとして生成するが、前記第１の生成ステップと比較して、前記第１のデジタルビデオストリームの異なるクロッピング、異なるズーミング、異なるパンニングまたは異なるフォーカスプレーン選択のうちの少なくとも１つを行い；および、公開ステップにおいて、前記出力デジタルビデオストリームを、前記共有デジタルビデオストリームの消費者に継続的に提供し、ここで、前記待ち時間が前記共有デジタルビデオストリームに存在する。

【００２３】

本発明はまた、共有デジタルビデオストリームを提供するシステムに関し、該システムは中央サーバを備え、該中央サーバは以下の機能を備える：第１のデジタルビデオソースから第１のデジタルビデオストリームを収集し、第２のデジタルビデオソースから第２のデジタルストリームを収集するよう構成される収集機能；前記第１のデジタルビデオソースからの画像情報が前記共有デジタルビデオストリームにおいて可視であり、前記第２のデジタルビデオソースからの画像情報が前記共有デジタルビデオストリームにおいて可視でないように、前記第１のデジタルビデオストリームの連続的に考慮される複数フレームに基づいて、前記共有デジタルビデオストリームを出力デジタルビデオストリームとして生成するよう構成される第１の生成機能；第１および第２のデジタルビデオストリームに対して、待ち時間を意図的に導入するよう構成される遅延機能；前記第１および／または

第２のデジタルビデオストリームをデジタル的に分析して、少なくとも１つのトリガを検出するよう構成されるトリガ検出機能、ここで、前記分析は、所定のタイプの画像および／または音声パターンの自動検出に基づいており、前記待ち時間を導入する前の第１および／または第２のデジタルビデオストリームに基づいて実行され、前記パターンの存在が前記トリガを構成し、前記トリガは、所定の生成ルールに従って、前記共有デジタルビデオストリームの生成モードを変更するよう指示するものであり；前記トリガの検出に応答して開始されるよう構成され、および、前記第２のデジタルビデオソースからの画像情報が前記共有デジタルビデオストリームにおいて可視であるように、前記第２のデジタルビデオストリームの連続して考慮される複数フレームに基づいて、前記共有デジタルビデオストリームを出力デジタルビデオストリームとして生成し、および／または、前記第１のデジタルビデオソースの連続的に考慮される複数フレームに基づいて、前記共有ビデオストリームを出力デジタルビデオストリームとして生成するが、前記第１の生成機能と比較して、前記第１のデジタルビデオストリームの異なるクロッピング、異なるズーミング、異なるパンニングまたは異なるフォーカスプレーン選択のうちの少なくとも１つを行うよう構成される第２の生成機能；および、前記出力デジタルビデオストリームを、前記共有デジタルビデオストリームの消費者に継続的に提供するよう構成される公開機能、ここで、前記待ち時間が前記共有デジタルビデオストリームに存在する。

【００２４】

以下では、本発明の例示的な実施形態および添付の図面を参照しながら、本発明を詳細に説明する。

【図面の簡単な説明】

【００２５】

【図１】図１は、第１の例示的システムを示す図である。

【図２】図２は、第２の例示的システムを示す図である。

【図３】図３は、第３の例示的システムを示す図である。

【図４】図４は、中央サーバを示す図である。

【図５】図５は、第１の方法を示す図である。

【図６ａ】図６ａは、図５に示す方法における異なる方法ステップに関連した後続する状態を示す図である。

【図６ｂ】図６ｂは、図５に示す方法における異なる方法ステップに関連した後続する状態を示す図である。

【図６ｃ】図６ｃは、図５に示す方法における異なる方法ステップに関連した後続する状態を示す図である。

【図６ｄ】図６ｄは、図５に示す方法における異なる方法ステップに関連した後続する状態を示す図である。

【図６ｅ】図６ｅは、図５に示す方法における異なる方法ステップに関連した後続する状態を示す図である。

【図６ｆ】図６ｆは、図５に示す方法における異なる方法ステップに関連した後続する状態を示す図である。

【図７】図７は、共通プロトコルを概念的に示す図である。

【図８】図８は、第４の例示的システムを示す図である。

【図９】図９は、第２の方法を示す図である。

【図１０ａ】図１０ａは、上記および多様な構成におけるカメラと参加者を示す図である。

【図１０ｂ】図１０ｂは、上記および多様な構成におけるカメラと参加者を示す図である。

【図１０ｃ】図１０ｃは、上記および多様な構成におけるカメラと参加者を示す図である。

【図１１ａ】図１１ａは、上記および多様な構成におけるカメラと参加者を示す図である。

【図１１ｂ】図１１ｂは、上記および多様な構成におけるカメラと参加者を示す図である。

【発明を実施するための形態】

【００２６】

すべての図は、同一または対応する部分の参照符号を共有するものである。

【００２７】

図１は、デジタルビデオストリーム、例えば共有デジタルビデオストリーム、を提供するための本発明に係る方法を実行するように構成された、本発明に係るシステム１００を示す。

【００２８】

システム１００は、ビデオ通信サービス１１０を備えてもよいが、ビデオ通信サービス１１０は、いくつかの実施形態では、システム１００の外部にあってもよい。以下で説明するように、複数のビデオ通信サービス１１０が備えられてよい。

【００２９】

システム１００は、１つまたは複数の参加者クライアント１２１を備えてよいが、１つ、いくつか、またはすべての参加者クライアント１２１は、いくつかの実施形態では、システム１００の外部であってもよい。

【００３０】

システム１００は中央サーバ１３０を備える。

【００３１】

本明細書で使用する場合、「中央サーバ」という用語は、明確に定義されたＡＰＩ（アプリケーション・プログラミング・インタフェース）を介するなど、論理的に一元化された方法でアクセスできるように構成された、コンピュータに実装された機能である。このような中央サーバの機能は、純粋にコンピュータソフトウエアに実装されてもよいし、ソフトウエアと仮想ハードウエアおよび／または物理ハードウエアの組み合わせに実装されてもよい。また、スタンドアロンの物理的または仮想的なサーバコンピュータに実装してもよいし、相互接続された複数の物理的および／または仮想的なサーバコンピュータに分散させて実装してもよい。

【００３２】

中央サーバ１３０が実行する物理的または仮想的ハードウエア、言い換えれば、中央サーバ１３０の機能を定義するコンピュータソフトウエアは、それ自体従来のＣＰＵ、それ自体従来のＧＰＵ、それ自体従来のＲＡＭ／ＲＯＭメモリ、それ自体従来のコンピュータバス、およびインターネット接続のようなそれ自体従来の外部通信機能から構成され得る。

【００３３】

ビデオ通信サービス１１０は、それが使用される限りにおいて、上記意味での中央サーバでもあり、それは中央サーバ１３０とは異なる中央サーバであってもよいし、中央サーバ１３０の一部であってもよい。

【００３４】

これに対応して、参加者クライアント１２１のそれぞれは、対応する解釈で、各参加者クライアント１２１が実行する物理的または仮想的ハードウエア、言い換えれば、参加者クライアント１２１の機能を定義するコンピュータソフトウエアが、それ自体従来のＣＰＵ／ＧＰＵ、それ自体従来のＲＡＭ／ＲＯＭメモリ、それ自体従来のコンピュータバス、およびインターネット接続のようなそれ自体従来の外部通信機能を備える、上記意味での中央サーバであってもよい。

【００３５】

各参加者クライアント１２１はまた、典型的には、進行中のビデオ通信の一部として参加者クライアント１２１に提供されるビデオコンテンツを表示するように配置されたコンピュータスクリーンと、ビデオ通信の一部として参加者クライアント１２１に提供されるサウンドコンテンツを発するように配置されたスピーカと、ビデオカメラと、ビデオ通信

に対する人間の参加者１２２にローカルにサウンドを記録するように配置されたマイクロフォンとを備えるか、またはそれらと通信しており、参加者１２２は、当該参加者クライアント１２１を使用してビデオ通信に参加する。

【００３６】

言い換えれば、各参加者クライアント１２１のそれぞれのヒューマン・マシン・インタフェースにより、それぞれの参加者１２２は、ビデオ通信において、他の参加者、および／または多様なソースから提供されるオーディオ／ビデオストリームと、当該クライアント１２１で対話することができる。

【００３７】

一般に、各参加者クライアント１２１は、デジタルビデオストリーム、デジタルオーディオストリームおよび／または他のデジタルデータを受信するための、上記ビデオカメラ；上記マイクロフォン；キーボード；コンピュータマウスまたはトラックパッド；および／またはＡＰＩから構成され得る、それぞれの入力手段１２３を備える。入力手段１２３は、特に、ビデオ通信サービス１１０および／または中央サーバ１３０などの中央サーバからビデオストリームおよび／またはオーディオストリームを受信するように構成され、そのようなビデオストリームおよび／またはオーディオストリームは、ビデオ通信の一部として提供され、好ましくは、そのようなデジタルデータ入力ストリームの少なくとも２つのソース、たとえば参加者クライアント１２１および／または外部ソース（後述）から上記中央サーバに提供される対応するデジタルデータ入力ストリームに基づいて生成される。

【００３８】

さらに一般的に、各参加者クライアント１２１は、上記コンピュータスクリーン；上記スピーカ；およびデジタルビデオおよび／またはオーディオストリームを発するＡＰＩから構成され得るそれぞれの出力手段１２４を備える。こうしたストリームは、当該参加者クライアント１２１を使用した、参加者１２２に対するローカルにキャプチャされたビデオおよび／またはオーディオを代表するものである。

【００３９】

実際には、各参加者クライアント１２１は、スクリーン、スピーカ、マイク、およびインターネット接続を備える携帯電話などのモバイルデバイスであってもよく、モバイルデバイスは、ローカルでコンピュータソフトウエアを実行するか、リモートで実行されるコンピュータソフトウエアにアクセスして、当該参加者クライアント１２１の機能を実行する。これに対応して、参加者クライアント１２１は、厚型または薄型のラップトップまたは据え置き型コンピュータであってもよく、ローカルにインストールされたアプリケーションを実行したり、ウェブブラウザを介してリモートアクセスされた機能を使用したりすることもできる。

【００４０】

本実施形態の１つの同じビデオ通信で使用される参加者クライアント１２１は、１つ以上、例えば少なくとも３つ、または少なくとも４つ存在してよい。

【００４１】

参加クライアントの少なくとも２つの異なるグループが存在してよい。参加クライアントの各々は、そのような各グループに割り当てられてよい。グループは、参加クライアントの異なる役割、参加クライアントの異なる仮想的または物理的な場所、および／または参加クライアントの異なる対話権限(interaction rights)を反映してよい。

【００４２】

そのような役割としては、多様なものが利用可能であり、例えば、「リーダ」または「会議主催者」、「スピーカ」、「パネリスト」、「対話する聴衆」、「遠隔リスナ」であってよい。

【００４３】

そのような物理的な場所としては、多様なものが利用可能であり、例えば「ステージ上」、「パネル内」、「物理的に存在する聴衆」、「物理的に離れた聴衆」であってよい。

【００４４】
仮想的な場所は、物理的な場所という観点から定義されてよいが、物理的な場所と部分的に重複し得る仮想的なグルーピングを含んでもよい。例えば、物理的に存在する聴衆参加者を第１の仮想グループと第２の仮想グループとに分け、物理的に存在する一部の聴衆参加者を、物理的に離隔する一部の聴衆参加者とともに１つの同じ仮想グループにグルーピングしてよい。
【００４５】
このような対話権限としては、多様なものが利用可能であり、例えば、「完全な対話」（制限なし）、「話すことはできるが、マイクを要求した後でなければ話せない」（ビデオ会議サービスで仮想の手を挙げるなど）、「話すことはできないが、共通のチャットに書き込むことができる」、「見る／聞くだけ」であってよい。
【００４６】
ある実施態様では、定義された各役割、および／または物理的／仮想的な場所は、ある所定の対話権限に関して定義されるかもしれない。他の例では、同じ対話権限を持つすべての参加者がグループを形成する。したがって、定義された役割、場所、および／または対話権限は、多様なグループ割当てを反映することができ、異なるグループは、必要に応じて相互に異なっても、または重複してもよい。
【００４７】
これは以下に例示的に説明される。
【００４８】
ビデオ通信は、本明細書で説明し例示するように、少なくとも一部がビデオ通信サービス１１０によって提供され、少なくとも一部が中央サーバ１３０によって提供され得る。
【００４９】
この用語が本明細書で使用される場合、「ビデオ通信」とは、少なくとも２つ、好ましくは少なくとも３つまたは少なくとも４つのビデオストリームを含む双方向デジタル通信セッションであり、好ましくは、１つまたは複数の混合または共同デジタルビデオ／オーディオストリームを生成するために使用され、ビデオおよび／またはオーディオを介してビデオ通信に貢献してもしなくてもよい１人または複数の消費者（例えば、上述のタイプの参加者クライアント）により消費されるオーディオストリームとも一致する。このようなビデオ通信は、リアルタイムであり、一定の待ち時間または遅延があってもなくてもよい。このようなビデオ通信に参加する少なくとも１人、好ましくは少なくとも２人、または少なくとも４人の参加者１２２は、ビデオ／オーディオ情報を提供および消費する双方向の方法でビデオ通信に関与する。
【００５０】
参加者クライアント１２１の少なくとも１つ、または参加者クライアント１２１のすべては、ローカル同期ソフトウエア機能１２５を備えており、その詳細は後述する。
【００５１】
ビデオ通信サービス１１０は、以下により詳細に説明するように、共通時間基準を備えるか、または共通時間基準にアクセスすることができる。
【００５２】
少なくとも１つの中央サーバ１３０のそれぞれは、当該中央サーバ１３０の外部のエンティティとデジタル通信するためのＡＰＩ１３７を含んでいてもよい。このような通信は、入力と出力の両方を含んでよい。
【００５３】
中央サーバ１３０などのシステム１００は、外部から提供されるビデオストリームなどの外部情報ソース３００とデジタル通信し、特に外部情報ソース３００からオーディオおよび／またはビデオストリームデータなどのデジタル情報を受信するように構成され得る。情報ソース３００が「外部」であるということは、中央サーバ１３０から、または中央サーバ１３０の一部として提供されないことを意味する。好ましくは、外部情報ソース３００によって提供されるデジタルデータは、中央サーバ１３０から独立しており、中央サ

ーバ１３０はその情報内容に影響を与えることができない。例えば、外部情報ソース３００は、公のスポーツイベントや進行中のニュースイベントや報道のような、ライブキャプチャされたビデオおよび／またはオーディオであってもよい。また、外部情報ソース３００は、参加者クライアント１２１のいずれかによってではなく、ウェブカメラなどによってキャプチャされてもよい。したがって、そのようなキャプチャされたビデオは、参加者クライアント１２１のいずれか１つと同じ局所（地域）を描写することができるが、参加者クライアント１２１の活動の一部としてキャプチャされることはない。外部から提供される情報ソース３００と内部から提供される情報ソース１２０との間の１つのあり得る相違点は、内部から提供される情報ソースは、上記定義されたタイプのビデオ通信への参加者として、その能力において提供され得るが、外部から提供される情報ソース３００はそうではなく、代わりに、当該ビデオ会議の外部であるコンテキストの一部として提供されることである。

【００５４】

また、オーディオおよび／またはビデオストリームなどの当該タイプのデジタル情報を中央サーバ１３０に並行して提供する、複数の外部情報ソース３００が存在してもよい。

【００５５】

図１に示すように、各参加者クライアント１２１は、説明したように、当該参加者クライアント１２１によってビデオ通信サービス１１０に提供されるそれぞれの情報（ビデオおよび／またはオーディオ）ストリーム１２０のソースを構成する。

【００５６】

中央サーバ１３０などのシステム１００は、外部消費者１５０とデジタル通信し、特に外部消費者１５０にデジタル情報を発するようにさらに構成されてもよい。例えば、中央サーバ１３０によって生成されたデジタルビデオおよび／またはオーディオストリームは、上記ＡＰＩ１３７を介して、１つまたは複数の外部消費者１５０に、リアルタイムまたはほぼリアルタイムで、連続的に提供され得る。ここでも、消費者１５０が「外部」であるということは、消費者１５０が中央サーバ１３０の一部として提供されないこと、および／または当該ビデオ通信の当事者ではないことを意味する。

【００５７】

特に断りのない限り、本明細書におけるすべての機能および通信は、適切なコンピュータハードウエア上で実行されるコンピュータソフトウエアによって実現され、インターネットのようなデジタル通信ネットワークまたはチャネルを介して通信されて、デジタル的かつ電子的に提供される。

【００５８】

したがって、図１に示されるシステム１００の構成では、多数の参加者クライアント１２１が、ビデオ通信サービス１１０によって提供されるデジタルビデオ通信に参加する。各参加者クライアント１２１は、それゆえ、ビデオ通信サービス１１０に対する進行中のログイン、セッション、または同様のものを有し、ビデオ通信サービス１１０によって提供される１つの同じ進行中のビデオ通信に参加することができる。言い換えれば、ビデオ通信は、参加者クライアント１２１の間で「共有」され、したがって、対応する人間の参加者１２２によっても「共有」される。

【００５９】

図１では、中央サーバ１３０は、自動参加者クライアント１４０を備え、これは参加者クライアント１２１に対応する自動クライアントであるが、人間の参加者１２２には関連付けられない。その代わりに、自動参加者クライアント１４０は、参加者クライアント１２１と同じ共有ビデオ通信に参加するために、ビデオ通信サービス１１０に参加者クライアントとして追加される。このような参加者クライアントとして、自動参加者クライアント１４０には、ビデオ通信サービス１１０によって進行中のビデオ通信の一部として提供される、連続的に生成されるデジタルビデオおよび／またはオーディオストリーム（複数可）へのアクセスが付与され、こうしたストリームは自動参加者クライアント１４０を介して中央サーバ１３０によって消費されることができる。好ましくは、自動参加者クライ

アント１４０は、ビデオ通信サービス１１０から、各参加者クライアント１２１に配信される、または配信され得る共通のビデオおよび／またはオーディオストリーム；１つまたは複数の参加者クライアント１２１の各々からビデオ通信サービス１１０に提供され、ビデオ通信サービス１１０によってすべての参加者クライアント１２１または要求している参加者クライアント１２１に生データまたは修正された形式で中継されるそれぞれのビデオおよび／またはオーディオストリーム；および／または共通時間基準、を受信する。

【００６０】

中央サーバ１３０は、後述するように処理するために、自動参加者クライアント１４０から、場合によっては上記外部情報ソース（複数可）３００からも、上記タイプのビデオストリームおよび／またはオーディオストリームを複数受信し、その後、ＡＰＩ１３７を介して生成されたビデオストリーム、例えば共有ビデオストリームを提供するように構成された収集機能１３１を備えてよい。たとえば、この生成されたビデオストリームは、外部の消費者１５０によって、および／またはビデオ通信サービス１１０によって消費され、ビデオ通信サービス１１０によって、参加者クライアント１２１のすべてまたは要求する任意の１つに配信される。

【００６１】

図２は図１と同様であるが、自動参加者クライアント１４０を使用する代わりに、中央サーバ１３０は、ビデオ通信サービス１１０のＡＰＩ１１２を介して、進行中のビデオ通信からビデオおよび／またはオーディオストリームデータを受信する。

【００６２】

図３も図１と同様であるが、ビデオ通信サービス１１０は示されていない。この場合、参加者クライアント１２１は、中央サーバ１３０のＡＰＩ１３７と直接通信し、例えば、中央サーバ１３０にビデオおよび／またはオーディオストリームデータを提供し、および／または中央サーバ１３０からビデオおよび／またはオーディオストリームデータを受信する。次に、生成された共有ストリームは、外部消費者１５０に、および／またはクライアント参加者１２１のうちの１つまたは複数に提供され得る。

【００６３】

図４は、中央サーバ１３０をより詳細に示している。図示されるように、収集機能１３１は、１つまたは好ましくは複数の、フォーマット固有の収集機能１３１ａから構成され得る。フォーマット固有の収集機能１３１ａの各々は、所定のバイナリエンコーディングフォーマットおよび／または所定のストリームデータコンテナのような所定のフォーマットを有するビデオおよび／またはオーディオストリームを受信するように構成され、具体的には、上記フォーマットのバイナリビデオおよび／またはオーディオデータを解析(parse)して、個々のビデオフレーム、ビデオフレームのシーケンスおよび／またはタイムスロットに分類するように構成される。

【００６４】

中央サーバ１３０は、収集機能１３１からバイナリストリームデータなどのビデオおよび／またはオーディオストリームデータを受信し、受信した複数のデータストリームの個々の１つに対してそれぞれのイベント検出を実行するように構成された、イベント検出機能１３２をさらに備える。イベント検出機能１３２は、イベント検出を実行するためのＡＩ（人工知能）コンポーネント１３２ａを備えてよい。イベント検出は、収集された個別の複数のストリームを最初に時間同期させることなく実行されてよい。

【００６５】

中央サーバ１３０は、収集機能１３１によって提供され、イベント検出機能１３２によって処理されてよい複数のデータストリームを時間同期させるように構成された同期機能１３３をさらに備える。同期機能１３３は、時間同期を実行するためのＡＩコンポーネント１３３ａを備えてよい。

【００６６】

中央サーバ１３０は、受信した複数のデータストリームの少なくとも１つでもよいが、多くの場合は少なくとも２つ、例えば少なくとも３つまたは少なくとも４つ、例えばすべ

て、の組み合わせに基づいてパターン検出を実行するように構成されたパターン検出機能１３４をさらに備えてよい。パターン検出はさらに、イベント検出機能１３２によって上記複数のデータストリームの個々の１つごとに検出された１つ、場合によっては少なくとも２つ以上のイベントに基づいてもよい。パターン検出機能１３４によって考慮されるそのような検出されたイベントは、個々の収集されたストリームに関して時間に亘って分布し得る。パターン検出機能１３４は、パターン検出を実行するためのＡＩコンポーネント１３４ａを備えてよい。パターン検出はさらに、上述したグルーピングに基づいてもよく、特に、１つのグループに関してのみ発生する特定のパターンを検出するように構成されてもよく、すべてのグループではなくいくつかのグループに関してのみ発生する特定のパターンを検出するように構成されてもよく、すべてのグループに関して発生する特定のパターンを検出するように構成されてもよい。

【００６７】

中央サーバ１３０は、収集機能１３１から提供された複数のデータストリームに基づいて、さらに、場合によっては検出されたあらゆるイベントおよび／またはパターンに基づいて、例えば共有デジタルビデオストリームなどの生成されたデジタルビデオストリームを生成するように構成された、生成機能１３５をさらに備える。生成されたビデオストリームは、少なくとも、収集機能１３１によって提供された、生データの、再フォーマットされた、または変換されたビデオストリームの１つまたは複数を備えるように生成されたビデオストリームを含み、対応するオーディオストリームデータを含んでもよい。以下に例示するように、複数の生成されたビデオストリームが存在してよく、そのような生成されたビデオストリームの１つは、上述の方法で生成されてよいが、別の既に生成されたビデオストリームにさらに基づいて生成されてよい。

【００６８】

すべての生成されたビデオストリームは、好ましくは連続的に、好ましくはほぼリアルタイムで生成される（本明細書で後述するタイプの待ち時間や遅延を差し引いた後）。

【００６９】

中央サーバ１３０はさらに、上述のＡＰＩ１３７を介するなどして、生成された共有デジタルビデオストリームを公開する(publish)ように構成された公開機能１３６を備えてよい。

【００７０】

図１、図２および図３は、本明細書で説明する原理を実施するために、特に本発明による方法を提供するために、中央サーバ１３０をどのように使用できるかの３つの異なる例を示しているが、１つまたは複数のビデオ通信サービス１１０を使用する、または使用しない他の構成も可能であることに留意されたい。

【００７１】

したがって、図５は、生成されたデジタルビデオストリームを提供するための方法を図示する。図６ａから図６ｆは、図５に示される方法ステップの結果として生じる異なるデジタルビデオ／オーディオデータストリームの状態を示す。

【００７２】

最初のステップで、本方法は始まる。

【００７３】

続く収集ステップにおいて、それぞれの複数の一次デジタルビデオストリーム２１０、３０１が、例えば収集機能１３１によって、上記デジタルビデオソース１２０、３００のうちの少なくとも２つから収集される。このような複数の一次データストリーム２１０、３０１はそれぞれ、オーディオ部分２１４および／またはビデオ部分２１５を備えてよい。このコンテキストにおける「ビデオ」は、かかるデータストリームの動画および／または静止画像コンテンツを示すことが理解される。各一次データストリーム２１０、３０１は、（当該一次ストリーム２１０、３０１を提供するエンティティによって使用されるそれぞれのコーデックを使用して）任意のビデオ／オーディオ符号化標準に従ってエンコードされてよく、符号化フォーマットは、１つの同じビデオ通信で同時に使用される当該複

数の一次ストリーム２１０、３０１の異なるもの間で異なってよい。複数の一次データストリーム２１０、３０１の少なくとも１つ、例えば全ては、バイナリデータのストリームとして提供されることが好ましく、場合によっては、それ自体従来のデータコンテナデータ構造で提供される。複数の一次データストリーム２１０、３０１の少なくとも１つ、例えば少なくとも２つ、あるいはすべてが、それぞれのライブビデオ録画として提供されることが好ましい。

【００７４】

複数の一次データストリーム２１０、３０１は、収集機能１３１によって受信されるとき、時間的に同期していない可能性があることに留意されたい。これは、それらが互いに対して異なる待ち時間または遅延に関連付けられていることを意味し得る。例えば、２つの一次ビデオストリーム２１０、３０１がライブ録画である場合、これは、収集機能１３１によって受信されたときに、それらが録画時間に関して異なる待ち時間に関連付けられていることを意味し得る。

【００７５】

また、複数の一次データストリーム２１０、３０１は、それ自体が、ウェブカメラからのそれぞれのライブカメラフィード；現在共有されている画面やプレゼンテーション；視聴されているフィルムクリップ；あるいは、１つの同じ画面内に様々な方法で配置されたこれらの任意の組み合わせであったりすることに留意されたい。

【００７６】

収集ステップは、図６ａおよび図６ｂに示されている。図６ｂでは、収集機能１３１が、各一次ビデオストリーム２１０、３０１を、バンドルされたオーディオ／ビデオ情報として、または関連するビデオストリームデータから分離されたオーディオストリームデータとして、どのように格納できるかも示されている。図６ｂは、一次ビデオストリーム２１０、３０１のデータが、個々のフレーム２１３またはフレームの集積／クラスタとしてどのように格納されるかを図示し、「フレーム」は、ここでは、画像データおよび／または任意の関連するオーディオデータの時間制限された部分を指し、たとえば、各フレームは、個々の静止画像であるか、または連続する一連の画像（たとえば、最大で１秒の動画像を構成するような一連の画像）であり、一緒に動画像ビデオコンテンツを形成する。

【００７７】

イベント検出機能１３２によって実行される後続のイベント検出ステップにおいて、複数の一次デジタルビデオストリーム２１０、３０１は、イベント検出機能１３２、特にＡＩコンポーネント１３２ａなどによって分析され、第１のイベントセットから選択された少なくとも１つのイベント２１１を検出する。これを図６ｃに示す。

【００７８】

このイベント検出ステップは、少なくとも１つの、例えば少なくとも２つの、例えば全ての、一次ビデオストリーム２１０、３０１に対して行われ、当該一次ビデオストリーム２１０、３０１の各々に対して個別に行われることが好ましい。換言すれば、イベント検出ステップは、好ましくは、当該特定の一次ビデオストリーム２１０、３０１の一部として含まれる情報のみを考慮し、特に、他の一次ビデオストリームの一部として含まれる情報を考慮することなく、当該個々の一次ビデオストリーム２１０、３０１に対して行われる。さらに、イベント検出は、好ましくは、複数の一次ビデオストリーム２１０、３０１に関連付けられるあらゆる共通の時間基準２６０を考慮することなく行われる。

【００７９】

一方、好ましくは、イベント検出は、ある時間間隔に亘って、例えば、０秒より長い、例えば、少なくとも０．１秒、例えば、少なくとも１秒のような一次ビデオストリームの履歴時間間隔に亘って、当該個別に分析された一次ビデオストリームの一部として含まれる情報を考慮する。

【００８０】

イベント検出は、一次ビデオストリーム２１０、３０１の一部として含まれるオーディオおよび／またはビデオデータに含まれる情報を考慮してもよい。

【００８１】
上記第１のイベントセットには、当該一次ビデオストリーム２１０、３０１を構成する、または当該一次ビデオストリーム２１０、３０１の一部であるスライドプレゼンテーションにおけるスライドの変化；当該一次ビデオストリーム２１０、３０１を提供するソース１２０、３００の接続品質の変化であって、画像品質の変化、画像データの損失、または画像データの再取得をもたらす接続品質の変化；および、ビデオ内の人物またはオブジェクトの動き、ビデオ内の照明の変化、オーディオ内の突然の鋭いノイズ、またはオーディオ品質の変化などの当該一次ビデオストリーム２１０、３０１内で検出された動きの物理的イベント等の、任意の数のタイプのイベントが含まれ得る。これは、網羅的なリストであることを意図するものではなく、これらの例は、現在説明されている原理の適用可能性を理解するために提供されるものであることを理解されたい。
【００８２】
同期機能１３３によって実行される後続の同期ステップにおいて、複数の一次デジタルビデオストリーム２１０は、時間同期される。この時間同期は、共通の時間基準２６０に対して実行されてよい。図６ｄに示されるように、この時間同期は、例えば共通時間基準２６０を使用して、複数の一次ビデオストリーム２１０、３０１を互いに対して整列させることを含んでよく、これにより、それらが組み合わされて、時間同期されたコンテキストを形成することができる。共通時間基準２６０は、データのストリーム、ハートビート信号または他のパルスデータ、または個別の複数の一次ビデオストリーム２１０、３０１のそれぞれに適用可能な時間アンカであってもよい。共通時間基準を個別の複数の一次ビデオストリーム２１０、３０１のそれぞれに適用できるようにすることによって、共通の時間軸に関して、当該一次ビデオストリーム２１０、３０１の情報コンテンツを共通時間基準に対して一義的に関連付けることができる。言い換えれば、共通時間基準により、複数の一次ビデオストリーム２１０、３０１は、タイムシフトを介して、現在の感覚(present sense)において時間同期されるように整列させられる。他の実施形態では、時間同期は、測定値など、当該一次ビデオストリーム２１０、３０１間の時間差に関する既知の情報に基づいてもよい。
【００８３】
図６ｄに示されるように、時間同期は、複数の一次ビデオストリーム２１０、３０１のそれぞれについて、例えば、共通時間基準２６０に関連して、またはビデオストリーム２１０、３０１のそれぞれについて、他方のビデオストリーム２１０、３０１に関連してまたは他の複数のビデオストリーム２１０、３０１に関連して、１つまたは複数のタイムスタンプ２６１を決定することを含んでよい。
【００８４】
パターン検出機能１３４によって実行される後続のパターン検出ステップにおいて、時間同期された複数の一次デジタルビデオストリーム２１０、３０１は、第１のパターンセットから選択された少なくとも１つのパターン２１２を検出するために分析される。これを図６ｅに示す。
【００８５】
イベント検出ステップとは対照的に、パターン検出ステップは、好ましくは、時間同期された複数の一次ビデオストリーム２１０、３０１の少なくとも２つの一部として含まれるビデオおよび／またはオーディオ情報に基づいて実行される。
【００８６】
上記第１のパターンのセットは、複数の参加者が入れ替わり立ち替わりまたは同時に発話する；またはプレゼンテーションスライドの変更が別の参加者が発話するなどの別のイベントとして同時に発生するなど、任意の数のタイプのパターンを含んでよい。このリストは網羅的なものではなく、例示的なものである。
【００８７】
代替的実施形態では、検出されたパターン２１２は、複数の一次ビデオストリーム２１０、３０１のうちの複数に含まれる情報ではなく、複数の一次ビデオストリーム２１０、

３０１のうちの１つのみに含まれる情報に関連してよい。そのような場合、そのようなパターン２１２は、少なくとも２つの検出されたイベント２１１、例えば２つ以上の連続して検出されたプレゼンテーションスライド変更または接続品質変更に跨る、その単一の一次ビデオストリーム２１０、３０１に含まれるビデオおよび／またはオーディオ情報に基づいて検出されることが好ましい。一例として、検出されたスライド変更イベントごとの１つの個別のスライド変更パターンとは対照的に、時間の経過とともに急速に互いに続く複数の連続したスライド変更が、１つの単一のスライド変更パターンとして検出される場合がある。

【００８８】

第１のイベントセットおよび第１のパターンセットは、パラメータおよびパラメータ間隔のそれぞれのセットを使用して定義された、所定のタイプのイベント／パターンを備えてよいことが理解される。以下に説明するように、上記セットのイベント／パターンは、様々なＡＩツールを使用して定義および検出することもできる。

【００８９】

生成機能１３５によって実行される後続の生成ステップにおいて、共有デジタルビデオストリームは、時間同期された複数の一次デジタルビデオストリーム２１０、３０１の連続的に考慮された複数のフレーム２１３と、検出されたパターン２１２とに基づいて、出力デジタルビデオストリーム２３０として生成される。

【００９０】

以下に説明し詳述するように、本発明により、例えば出力デジタルビデオストリーム２３０のようなビデオストリームを完全に自動生成することができる。

【００９１】

例えば、このような生成は、どの一次ビデオストリーム２１０、３０１からどのようなビデオおよび／またはオーディオ情報を、出力ビデオストリーム２３０でどの程度使用するかの選択；出力ビデオストリーム２３０のビデオスクリーンレイアウト；時間に亘る異なるそのような使用またはレイアウト間の切り替えパターンなどを含んでよい。

【００９２】

これは、図６ｆにも示されており、この図６ｆには、出力ビデオストリーム２３０の生成において、（例えば共通時間基準２６０に対して）時間同期され、時間同期された複数の一次ビデオストリーム２１０、３０１と協働して使用され得る、追加デジタルビデオ情報ストリームなどの、時間に関連する（共通時間基準２６０に関連してよい）デジタルビデオ情報２２０の１つまたは複数の追加部分が示されている。例えば、追加ストリーム２２０は、検出されたパターンに動的に基づくような、使用する任意のビデオおよび／またはオーディオ特殊効果に関する情報；ビデオ通信のための計画されたタイムスケジュールなどを含んでよい。

【００９３】

公開機能１３６によって実行される後続の公開ステップにおいて、生成された出力デジタルビデオストリーム２３０は、上述のように、共有デジタルビデオストリームの消費者１１０、１５０に連続的に提供される。生成されたデジタルビデオストリームは、例えばビデオ通信サービス１１０を介して、１つまたは複数の参加者クライアント１２１に提供され得る。

【００９４】

後続するステップで、本方法は終了する。しかしながら、最初に、本方法は、図５に示されるように、出力ビデオストリーム２３０を連続的に提供されるストリームとして生成するために、任意の回数、反復されてもよい。好ましくは、出力ビデオストリーム２３０は、リアルタイムまたはほぼリアルタイム（途中のすべてのステップによって付加される待ち時間の合計を考慮する）で、かつ連続的に（より多くの情報が利用可能になったときに直ちに公開されるが、ただし、以下に説明する意図的に付加される待ち時間はカウントしない）消費されるように生成される。このようにして、出力ビデオストリーム２３０は、双方向（インタラクティブ）方式で消費されてよく、これにより、ビデオ通信サービス

１１０に、または、閉じたフィードバックループを形成するように収集機能１３１に再び供給される一次ビデオストリーム２１０の生成の基礎を形成する他のコンテキストに、出力ビデオストリーム２３０がフィードバックされ；または、異なる（システム１００の外部、または、少なくとも中央サーバ１３０の外部の）コンテキストに、出力ビデオストリーム２３０が消費され、そこで、リアルタイムの双方向ビデオ通信の基礎を形成することができる。

【００９５】

上述したように、いくつかの実施形態では、複数の一次デジタルビデオストリーム２１０、３０１の少なくとも２つ、例えば少なくとも３つ、例えば少なくとも４つまたは少なくとも５つは、ビデオ通信サービス１１０によって提供されるような共有デジタルビデオ通信の一部として提供され、このビデオ通信は、当該一次デジタルビデオストリーム２１０を提供するそれぞれの遠隔接続された参加者クライアント１２１を含む。このような場合、収集ステップは、共有デジタルビデオ通信サービス１１０自体から、当該ビデオ通信サービス１１０内からビデオおよび／またはオーディオストリームデータへのアクセスを順番に付与された自動参加者クライアント１４０を介して、および／またはビデオ通信サービス１１０のＡＰＩ１１２を介して、当該一次デジタルビデオストリーム２１０の少なくとも１つを収集することから構成され得る。

【００９６】

さらに、この場合および他の場合において、収集ステップは、共有デジタルビデオ通信サービス１１０の外部である情報ソース３００から収集されたそれぞれの外部デジタルビデオストリーム３０１として、上記複数の一次デジタルビデオストリーム２１０、３０１の少なくとも１つを収集することを備えてよい。このような外部ビデオソース３００の１つまたは複数が、中央サーバ１３０の外部にあってもよいことに留意されたい。

【００９７】

いくつかの実施形態では、複数の一次ビデオストリーム２１０、３０１は、同じ方法でフォーマットされていない。そのような異なるフォーマットは、それらが異なるタイプのデータコンテナ（ＡＶＩまたはＭＰＥＧなど）で収集機能１３１に供給される形式であり得るが、好ましい実施形態では、複数の一次ビデオストリーム２１０、３０１の少なくとも１つは、（一次ビデオストリーム２１０、３０１の少なくとも１つの他のものに対して）偏位する(deviating)フォーマットに従ってフォーマットされ、これは、偏位する一次デジタルビデオストリーム２１０、３０１が偏位するビデオ符号化；偏位する固定または可変フレームレート；偏位するアスペクト比；偏位するビデオ解像度；および／または偏位するオーディオサンプルレートを有するという点においてである。

【００９８】

収集機能１３１は、収集されたすべての一次ビデオストリーム２１０、３０１で発生するすべての符号化フォーマット、コンテナ規格などを読み取り、解釈するように事前に構成されていることが好ましい。これにより、本明細書で説明するような処理を実行することが可能になり、これらの処理の比較的後の段階まで（当該一次ストリームがそれぞれのバッファに入れられるまで；イベント検出ステップの後まで；あるいはイベント検出ステップの後までなど）、デコーディングを必要としない。しかしながら、一次ビデオフィード２１０、３０１のうちの１つまたは複数が、収集機能１３１がデコードせずには解釈できないコーデックを使用して符号化されているという稀なケースでは、収集機能１３１は、そのような一次ビデオストリーム２１０、３０１のデコードおよび分析を実行し、その後、例えばイベント検出機能によって処理できるフォーマットへの変換を実行するように構成されてよい。この場合でも、この段階で再エンコードを実行しないことが好ましいことに留意されたい。

【００９９】

例えば、ビデオ通信サービス１１０によって提供されるような、複数当事者間のビデオイベントからフェッチされる一次ビデオストリーム２２０は、典型的には、低遅延に対する要求があり、したがって、典型的には、参加者１２２が効果的な通信を行うことを可能

にするために、可変フレームレートおよび可変ピクセル解像度と関連付けられる。言い換えれば、全体的なビデオおよびオーディオの品質は、低遅延のために必要に応じて低下する。

【０１００】

一方、外部ビデオフィード３０１は、典型的には、より安定したフレームレートを持ち、高画質であるが、それゆえ遅延が大きくなる可能性がある。

【０１０１】

したがって、ビデオ通信サービス１１０は、各時点で、外部ビデオソース３００とは異なる符号化（エンコーディング）および／またはコンテナを使用し得る。したがって、本明細書で説明する分析およびビデオ生成プロセスは、この場合、異なるフォーマットのこれら複数のストリーム２１０、３０１を、結合された経験のために新たな１つのストリームに結合する必要がある。

【０１０２】

上述したように、収集機能１３１は、フォーマット固有の収集機能１３１ａのセットを備えてよく、それぞれが、特定のタイプのフォーマットの一次ビデオストリーム２１０、３０１を処理するように構成される。たとえば、これらのフォーマット固有の収集機能１３１ａの各１つは、Windows（登録商標）Media（登録商標）またはDivX（登録商標）などの異なるビデオそれぞれのエンコード方法／コーデックを使用してエンコードされた複数の一次ビデオストリーム２１０、３０１を処理するように構成されてよい。

【０１０３】

しかしながら、好ましい実施形態では、収集ステップは、複数の一次デジタルビデオストリーム２１０、３０１の少なくとも２つ、例えばすべて、を共通プロトコル２４０に変換することを含む。

【０１０４】

この文脈で使用されるように、用語「プロトコル」は、デジタルビデオ／オーディオストリームに含まれる情報をどのように格納かを特定する情報構造化規格またはデータ構造を示す。しかしながら、共通プロトコルは、好ましくは、デジタルビデオおよび／またはオーディオ情報を、例えばバイナリレベルなど（すなわち、音および画像自体を指示する符号化／圧縮されたデータ）、どのように格納するかを規定するのではなく、代わりに、そのようなデータを格納するための所定のフォーマットの構造を形成する。言い換えれば、共通プロトコルは、そのような格納に関連してデジタルビデオデコーディングまたはデジタルビデオエンコーディングを何ら実行することなく、場合によってはバイナリ形式のバイト列を連結および／または分割することとは別に既存のバイナリ形式を全く修正することなく、生データのバイナリ形式でデジタルビデオデータを格納することを規定している。代わりに、当該一次ビデオストリーム２１０、３０１の生データの（符号化／圧縮された）バイナリデータコンテンツは、プロトコルによって定義されたデータ構造でこの生データのバイナリデータを再パック(repacking)しながら、保持される。いくつかの実施形態では、共通プロトコルは、ビデオファイルコンテナフォーマットを定義する。

【０１０５】

図７は、一例として、それぞれのフォーマット固有収集機能１３１ａによって再構築され、上記共通プロトコル２４０を使用する、図６ａに示す複数の一次ビデオストリーム２１０、３０１を示す。

【０１０６】

したがって、共通プロトコル２４０は、デジタルビデオおよび／またはオーディオデータを、好ましくは、当該一次ビデオストリーム２１０、３０１に関連する時間軸に沿って、離散的で連続したデータのセットに分割されたデータセット２４１に格納することを規定する。このような各データセットは、１つまたはいくつかのビデオフレーム、および関連するオーディオデータを含むことができる。

【０１０７】

共通プロトコル２４０はまた、格納されたデジタルビデオおよび／またはオーディオデ

ータセット２４１に関連して、指定された時点に関連付けられるメタデータ２４２を格納することを規定してよい。

【０１０８】

メタデータ２４２は、当該生データのバイナリデータを生成するために使用されるデジタルビデオエンコーディング方法またはコーデック；ビデオデータの解像度；ビデオフレームレート；フレームレート変動フラグ；ビデオ解像度；ビデオアスペクト比；オーディオ圧縮アルゴリズム；またはオーディオサンプリングレートに関する等の、当該一次デジタルビデオストリーム２１０の生データのバイナリフォーマットに関する情報を含んでよい。また、メタデータ２４２は、例えば当該一次ビデオストリーム２１０、３０１自体の時間基準に関連する、または上述の異なるビデオストリームに関連する、格納されたデータのタイムスタンプに関する情報を含んでよい。

【０１０９】

フォーマット固有の収集機能１３１ａを共通プロトコル２４０と組み合わせて使用することにより、受信したビデオ／オーディオデータをデコーディング／再エンコーディングすることによる待ち時間（遅延）を追加することなく、一次ビデオストリーム２１０、３０１の情報コンテンツを迅速に収集することが可能になる。

【０１１０】

したがって、収集ステップは、当該一次ビデオストリーム２１０、３０１を解析し、解析された生データのバイナリデータを、任意の関連するメタデータとともに、共通プロトコルを使用してデータ構造に格納するために、複数のフォーマット固有収集機能１３１ａのうち異なる収集機能１３１ａを使用して、異なるバイナリビデオおよび／またはオーディオエンコーディングフォーマットを使用してエンコードされた複数の一次デジタルビデオストリーム２１０、３０１を収集することを備えてよい。自明であるが、どの一次ビデオストリーム２１０、３０１に対してどのフォーマット固有の収集機能１３１ａを使用するかについての決定は、当該一次ビデオストリーム２１０、３０１のそれぞれの所定のおよび／または動的に検出された特性に基づいて、収集機能１３１によって実行されてよい。

【０１１１】

このように収集された各一次ビデオストリーム２１０、３０１は、中央サーバ１３０内のＲＡＭメモリバッファのような、それ自身の別個のメモリバッファに格納されてよい。

【０１１２】

各フォーマット固有の収集機能１３１ａによって実行される一次ビデオストリーム２１０、３０１の変換は、したがって、こうして変換された各一次デジタルビデオストリーム２１０、３０１の生データのバイナリデータを、順序付けられたセットのより小さいデータセット２４１に分割することを備えてよい。

【０１１３】

さらに、変換は、より小さいセット２４１のそれぞれ（またはサブセット、例えば当該一次ストリーム２１０、３０１のそれぞれの時間軸に沿って規則的に分布するサブセット）を、例えば共通時間基準２６０に関連して、共有の時間軸に沿ったそれぞれの時間に関連付けることも備えてよい。この関連付けは、以下に説明する原理方法のいずれか、または他の方法で、生データのバイナリビデオおよび／またはオーディオデータの分析によって実行されてもよく、一次ビデオストリーム２１０、３０１の後続する時間同期を実行できるようにするために実行されてもよい。使用される共通時間基準２６０のタイプに応じて、各データセット２４１のこの関連付けの少なくとも一部は、同期機能１３３によっても、またはその代わりに実行されてもよい。後者の場合、収集ステップは、代わりに、より小さいセット２４１の各々またはそのサブセットを、当該一次ストリーム２１０、３０１に固有の時間軸のそれぞれの時間に関連付けることを含んでよい。

【０１１４】

いくつかの実施形態では、収集ステップはまた、複数の一次ビデオストリーム２１０、３０１から収集された生データのバイナリビデオおよび／またはオーディオデータを、均

一な品質に変換、および／または周波数を更新することを含む。これは、必要に応じて、複数の一次デジタルビデオストリーム２１０、３０１の生データのバイナリデジタルビデオおよび／またはオーディオデータを、共通のビデオフレームレート；共通のビデオ解像度；または共通のオーディオサンプリングレートにダウンサンプリングまたはアップサンプリングすることを含んでよい。当該フォーマット固有収集機能１３１ａが正しいバイナリ符号化ターゲットフォーマットに従って生データのバイナリデータを直接処理することができるため、このような再サンプリングは、完全なデコーディング／再エンコーディングを実行することなく、あるいは、デコーディングを全く実行することすらなく、実行することができることに留意されたい。

【０１１５】

好ましくは、複数の一次デジタルビデオストリーム２１０、３０１の各々は、上述したように、個々のフレーム２１３またはフレーム２１３のシーケンスとして、個々のデータ格納バッファ２５０に格納され、また、各々は、共通時間基準２６０に順に関連付けられた対応するタイムスタンプに関連付けられる。

【０１１６】

例示のために提供される具体例では、ビデオ通信サービス１１０は、Microsoft（登録商標） Teams（登録商標）であり、同時接続の複数参加者１２２が関与するビデオ会議を実行している。自動参加者クライアント１４０は、Teams（登録商標）会議の会議参加者として登録されている。

【０１１７】

次に、一次ビデオ入力信号２１０が、自動参加者クライアント１４０を介して収集機能１３０に提供され、収集機能１３０によって取得される。これらはＨ２６４フォーマットの生データ信号であり、各ビデオフレームのタイムスタンプ情報を含む。

【０１１８】

関連するフォーマット固有の収集機能１３１ａは、設定可能な事前定義されたＴＣＰポート上で、ＩＰ（クラウドのＬＡＮネットワーク）経由で生データをピックアップする。すべてのTeams（登録商標）会議参加者と関連する音声データは、個別のポートに関連付けられる。次に、収集機能１３１は、オーディオ信号（５０Ｈｚ）からのタイムスタンプを使用し、ビデオデータを２５Ｈｚの固定出力信号にダウンサンプリングしてから、ビデオストリーム２２０をそれぞれの個別のバッファ２５０に格納する。

【０１１９】

上述のように、共通プロトコル２４０は、データを生データのバイナリ形式で保存する。これは、非常に低レベルで、ビデオ／オーディオデータの生データのビットおよびバイトを処理するように設計することができる。好ましい実施形態では、データは、単純なバイト配列または対応するデータ構造（スライスなど）として共通プロトコル２４０に格納される。これは、データが従来のビデオコンテナに入れられる必要が全くないことを意味する（共通プロトコル２４０は、この文脈ではそのような従来のコンテナを構成しない）。また、ビデオのエンコードおよびデコードは、計算量的に重いので、遅延を引き起こし、高価なハードウエアを必要とする。さらに、この問題は参加者の数によってスケーリングされる。

【０１２０】

共通プロトコル２４０を使用すると、Teams（登録商標）の各会議参加者１２２に関連付けられた一次ビデオストリーム２１０と、任意の外部ビデオソース３００のために、収集機能１３１内のメモリを確保し、プロセス中にオンザフライで割り当てメモリ量を変更することが可能になる。このようにして、入力ストリームの数を変更することが可能になり、その結果、各バッファを有効に保つことができる。例えば、解像度、フレームレートなどの情報は可変であるが、共通プロトコル２４０にメタデータとして格納されるため、この情報を使用して、必要に応じて各バッファのサイズを迅速に変更することができる。

【０１２１】

以下は、本タイプの共通プロトコル２４０の仕様の一例である。

【０１２２】

【表１】

Bytes	Example	Description
1 byte	1	0=video; 1=audio
4 bytes	1234567	Buffer Length (int)
8 bytes	424234234	Timestamp from the incoming audio/video buffer
		Measured in ticks, 1tick = 100ns. (long int)
1 byte	0	VideoColorFormat
		{
		NV12 = 0,
		Rgb24 = 1,
		Yuy2 = 2,
		H264 = 3
		}
4 bytes	720	Video frame pixel height (int)
4 bytes	640	Video frame pixel width (int)
4 bytes	25.0	Video frame rate
		Number of frames per second (float)
1 byte	0	Is audio silence? 1 = true; 0 = false
1 byte	0	AudioFormat
		{
		0 = Pcm16K
		1 = Pcm44KStereo
		}
1 byte	0	Detected event in, if any
		0 = no event
		1, 2, 3, etc. = event of specified type detected
30 bytes		Reserved for future use
8 bytes	1000000	Length of binary data in bytes (long int)
Variable	0x87A879...	Raw binary video/audio data of this frame(s)
4 bytes	1234567	Dominant speaker Port
4 bytes	1234567	Active speaker

【０１２３】

上記表では、「もしあれば、検出されたイベント（Detected event in, if any）」データは、共通プロトコル２６０の仕様の一部として含まれている。しかしながら、いくつかの実施形態では、この情報（検出されたイベントに関する）は、代わりに別のメモリバッファに入れてもよい。

【０１２４】

いくつかの実施形態では、オーバーレイまたはエフェクトであり得るデジタルビデオ情報２２０の上記少なくとも１つの追加部分もまた、共通時間基準２６０に順に関連付けら

れた対応するタイムスタンプにそれぞれ関連付けられた個々のフレームまたはフレームのシーケンスとして、それぞれの個別バッファ２５０に格納される。

【０１２５】

上記に例示したように、イベント検出ステップは、共通プロトコル２４０を使用して、検出されたイベント２１１を記述するメタデータ２４２を、当該イベント２１１が検出された一次デジタルビデオストリーム２１０、３０１に関連付けて記憶することを含んでよい。

【０１２６】

イベント検出は、異なる方法で実行することができる。ＡＩコンポーネント１３２ａによって実行されるいくつかの実施形態では、イベント検出ステップは、第１の訓練されたニューラルネットワークまたは他の機械学習コンポーネントが、前記イベント２１１のいずれかを自動的に検出するために、複数の一次デジタルビデオストリーム２１０、３０１の少なくとも１つ、例えばいくつか、または全部を個別に分析することを含む。これは、ＡＩコンポーネント１３２ａが、管理された分類において、一次ビデオストリーム２１０、３０１のデータを、予め定義されたイベントのセットに分類すること、および／または、管理されていない分類において、動的に決定されたイベントのセットに分類することを含んでよい。

【０１２７】

いくつかの実施形態では、検出されたイベント２１１は、当該一次ビデオストリーム２１０、３０１である、またはこれに含まれるプレゼンテーションのプレゼンテーションスライドの変更である。

【０１２８】

例えば、プレゼンテーションの発表者が、聴衆に対してその時点で行っているプレゼンテーションのスライドを変更することを決定した場合、これは、所定の視聴者にとって興味深いものが変化し得ることを意味する。新しく表示されるスライドは、いわゆる「バタフライ」モード（例えば、出力ビデオストリーム２３０でスライドを発表者のビデオと並べて表示する）で短時間見るのに最適な全体レベルの画像に過ぎないかもしれない。あるいは、スライドには、詳細が多く、フォントサイズが小さいテキストなどが含まれる場合もある。後者の場合、スライドはフルスクリーンで表示され、通常の場合よりも多少長い時間表示される。この場合のスライドは、発表者の顔よりもプレゼンテーションを見る人の興味を引くかもしれないので、バタフライモードはそれほど適切ではないかもしれない。

【０１２９】

実際には、イベント検出ステップは、以下の少なくとも１つで構成される。

【０１３０】

まず第１に、イベント２１１は、検出されたスライドの第１の画像と、それに後続する検出されたスライドの第２の画像との差の画像解析に基づいて検出することができる。一次ビデオストリーム２２０、３０１の性質がスライドを示すものであることは、ＯＣＲ（光学式文字認識）と組み合わせた動き検出を使用するなど、それ自体は従来のデジタル画像処理を使用して自動的に決定することができる。

【０１３１】

これは、自動コンピュータ画像処理技術を使用して、検出されたスライドが、実際にスライドの変更として分類するのに十分なほど大きく変化したかどうかをチェックすることを含んでよい。これは、ＲＧＢカラー値に関して、現在のスライドと前のスライドとの間のデルタ(delta)をチェックすることによって行うことができる。例えば、問題のスライドがカバーする画面領域で、ＲＧＢ値がどの程度グローバルに変化したかを評価し、同時に、これに協調して変化する隣接するピクセルのグループを見つけることが可能かどうかを評価することができる。こうすることで、関連するスライドの変化を検出することができ、同時に、例えば、画面全体にわたるコンピュータのマウスの動きのような無関係な変化をフィルタリングすることができる。このアプローチで完全な構成可能性が実現される

。例えば、発表者がコンピュータのマウスを使って異なるものを指し示しながら何かを詳細に発表したい場合など、コンピュータのマウスの動きをキャプチャできることが望まれることがある。

【０１３２】

第２に、イベント２１１は、より高い特異性でイベントの種類を決定するために、第２の画像自体の情報複雑性の画像解析に基づいて検出されてもよい。

【０１３３】

これは例えば、問題のスライド上のテキスト情報の総量や、関連するフォントサイズを評価することを含むかもしれない。これは、ディープラーニングに基づく文字認識技術など、従来のＯＣＲ手法を使用して行うことができる。

【０１３４】

評価されたビデオストリーム２１０、３０１の生データのバイナリフォーマットが既知であるため、これは、最初にビデオデータをデコードまたは再エンコードすることなく、バイナリドメインで直接実行されてもよいことに留意されたい。例えば、イベント検出機能１３２は、画像解釈サービスのための関連するフォーマット固有の収集機能を呼び出してもよく、またはイベント検出機能１３２自体が、多数の異なるサポートされる生データのバイナリビデオデータフォーマットに対して、個々のピクセルレベルなどへの画像情報を評価するための機能を含んでもよい。

【０１３５】

別の例では、検出されたイベント２１１は、デジタルビデオ通信サービス１１０に対する参加者クライアント１２１の通信接続の喪失である。この場合、検出ステップは、当該参加者クライアント１２１に対応する一次デジタルビデオストリーム２１０の一連の後続ビデオフレーム２１３の画像解析に基づいて、当該参加者クライアント121が通信接続を喪失したことを検出することを含んでよい。

【０１３６】

参加者クライアント１２１は、異なる物理的な場所や異なるインターネット接続に関連付けられているため、誰かがビデオ通信サービス１１０または中央サーバ１３０への接続を喪失することが起こり得る。そのような状況では、生成された出力ビデオストリーム２３０に黒または空の画面が表示されないようにすることが望ましい。

【０１３７】

これに替えて、このような接続の喪失は、イベント検出機能１３２によってイベントとして検出することができ、例えば、使用される２つのクラスが接続されている／接続されていない（データなし）である２クラス分類アルゴリズムを適用することによって検出することができる。この場合、「データなし」は、発表者が意図的に黒いスクリーンを送信することとは異なると理解される。わずか１フレームまたは２フレームのような短時間の黒いスクリーンは、最終的に生成されるストリーム２３０では目立たない可能性があるため、２クラス分類アルゴリズムを経時的に適用して時系列を作成することができる。次に、接続中断の最小長さを指定する閾値を使用して、接続が喪失したかどうかを決定することができる。

【０１３８】

以下で説明されるように、上記で例示されたタイプの検出されたイベントは、パターン検出機能１３４によって、適切かつ所望のように、様々な対応をするために使用され得る。

【０１３９】

上記のように、個々の一次ビデオストリーム２１０、３０１は、それぞれ共通の時間基準２６０に関連しており、同期機能１３３が互いに対して時間同期させることができる。

【０１４０】

いくつかの実施形態では、共通時間基準２６０は、共通オーディオ信号１１１（図１から図３参照）に基づくか、または共通オーディオ信号１１１を備え、共通オーディオ信号１１１は、上述したように、少なくとも２つの遠隔接続された参加者クライアント１２１

が参加する共有デジタルビデオ通信サービス１１０に共通であり、各々が、一次デジタルビデオストリーム２１０のそれぞれの１つを提供する。

【０１４１】

上述したMicrosoft（登録商標）Teams（登録商標）の例では、共通オーディオ信号が生成され、自動参加者クライアント１４０を介して、および／またはＡＰＩ　１１２を介して、中央サーバ１３０によってキャプチャされ得る。この例および他の例では、このような共通オーディオ信号をハートビート信号として使用して、このハートビート信号に基づいて個々の一次ビデオストリーム２２０を特定の時点に結合することによって、個々の複数の一次ビデオストリーム２２０を時間同期させることができる。このような共通オーディオ信号は、別個の（他の一次ビデオストリーム２１０の各々との関係において）信号として提供されてもよく、これにより、他の一次ビデオストリーム２１０は各々、当該他の一次ビデオストリーム２１０に含まれるオーディオに基づいて、または（自動画像処理ベースのリップシンク技術を使用するなど）そこに含まれる画像情報に基づいて、共通オーディオ信号に個別に時間相関(time -correlated)されてもよい。

【０１４２】

言い換えると、個々の一次ビデオストリーム２１０に関連する可変の待ち時間および／または異なる待ち時間を処理し、結合されたビデオ出力ストリーム２３０の時間同期を実現するために、このような共通オーディオ信号が、中央サーバ１３０内のすべての一次ビデオストリーム２１０のためのハートビートとして使用される（ただし、おそらく外部の一次ビデオストリーム３０１ではない）。言い換えれば、他のすべての信号は、すべてが時間同期していることを確認するために、この共通のオーディオ時間のハートビートにマッピングされる。

【０１４３】

別の例では、時間同期は、出力デジタルビデオストリーム２３０に導入され、参加者クライアント１２１の１つまたはいくつかの個々のものの一部として提供されるそれぞれのローカル時間同期ソフトウエア機能１２５によって検出される時間同期要素２３１を使用して達成される。ローカルソフトウエア機能１２５は、出力ビデオストリーム２３０における時間同期要素２３１の到着時間(time of arrival)を検出するように構成される。理解されるように、このような実施形態では、出力ビデオストリーム２３０は、ビデオ通信サービス１１０にフィードバックされるか、そうでない場合、各参加者クライアント１２１および当該ローカルソフトウエア機能１２５に利用可能となる。

【０１４４】

例えば、時間同期要素２３１は、一定の時間間隔で出力ビデオ２３０に配置または更新される、所定の順序または方法で色が変化する画素などの視覚的マーカ；出力ビデオ２３０に更新および表示される視覚的時計；（例えば、十分に低い振幅および／または十分に高い周波数を有することによって、参加者１２２に聞こえないように設計され得る）音声信号であって、出力ビデオストリーム２３０の一部を形成する音声に追加される音声信号、であってよい。ローカルソフトウエア機能１２５は、適切な画像処理および／または音声処理を使用して、時間同期要素（のそれぞれ）２３１のそれぞれの到着時間を自動的に検出するように構成される。

【０１４５】

次に、共通時間基準２６０は、少なくとも部分的には、検出された到着時間に基づいて決定され得る。例えば、ローカルソフトウエア機能１２５の各々は、検出された到着時間を示すそれぞれの情報を中央サーバ１３０に通信してもよい。

【０１４６】

このような通信は、当該参加者クライアント１２１と中央サーバ１３０との間の直接通信リンクを介して行われてよい。ただし、通信は、当該参加クライアント１２１に関連付けられた一次ビデオストリーム２１０を介して行われてもよい。例えば、参加者クライアント１２１は、中央サーバ１３０による自動検出のために、当該参加者クライアント１２１によって生成される一次ビデオストリーム２１０に、上述したタイプのような視覚的ま

たは可聴的コードを導入し、共通時間基準２６０を決定するために使用してもよい。
【０１４７】
さらに追加的な実施例では、各参加者クライアント１２１は、ビデオ通信サービス１１０に対して、すべての参加者クライアント１２１が視聴可能な共通ビデオストリームにおいて画像検出を実行し、そのような画像検出の結果を、上述したものに対応する方法で中央サーバ１３０に中継し、そこで経時的に、互いに対する各参加者クライアント１２１のそれぞれのオフセットを決定するために使用されてもよい。このようにして、共通時間基準２６０は、個々の相対オフセットのセットとして決定され得る。例えば、共通に利用可能なビデオストリームの選択された基準ピクセルは、ローカルソフトウエア機能１２５などによって、いくつかの、またはすべての参加クライアント１２１によってモニタリングされ、そのピクセルの現在の色が中央サーバ１３０に伝達されてもよい。中央サーバ１３０は、多数の（またはすべての）参加クライアント１２１の各々から連続して受信されたそのような色値に基づいてそれぞれの時系列を計算し、相互相関(cross-correlation)を実行することにより、異なる参加クライアント１２１間に亘る相対的な時間オフセットの推定されるセットを生成してもよい。
【０１４８】
実際には、ビデオ通信サービス１１０に供給された出力ビデオストリーム２３０は、当該ビデオ通信のすべての参加者クライアントの共有画面の一部として含まれてよく、したがって、参加者クライアント１２１に関連付けられるそのような時間オフセットを評価するために使用されてよい。特に、ビデオ通信サービス１１０に供給された出力ビデオストリーム２３０は、自動参加者クライアント１４０および／またはＡＰＩ１１２を介して、中央サーバに再び利用可能になってよい。
【０１４９】
いくつかの実施形態では、共通時間基準２６０は、複数の一次デジタルビデオストリーム２１０、３０１のうちの第１のもののオーディオ部分２１４と、複数の一次デジタルビデオストリーム２１０、３０１のうちの上記第１のものの画像部分２１５との間の検出された不一致(discrepancy)に少なくとも部分的に基づいて決定され得る。このような不一致は、例えば、当該第１の一次デジタルビデオストリーム２１０、３０１において視聴される発話中の参加者１２２のデジタルリップシンクビデオ画像分析に基づくものでよい。このようなリップシンク解析はそれ自体従来のものであり、例えば訓練されたニューラルネットワークを使用してもよい。分析は、利用可能な共通のオーディオ情報に関連して、各一次ビデオストリーム２１０、３０１について同期機能１３３によって実行されてもよく、個々の一次ビデオストリーム２１０、３０１に亘る相対的なオフセットは、この情報に基づいて決定されてもよい。
【０１５０】
いくつかの実施形態では、同期ステップは、最大３０秒、例えば最大５秒、例えば最大１秒、例えば最大０．５秒、ただし０秒よりも長い遅延（この文脈において「遅延」および「待ち時間」は同じものを意味することが意図される）を意図的に導入することを含み、これにより、出力デジタルビデオストリーム２３０には少なくともその遅延が提供される。いずれの長さであっても、意図的に導入される遅延は、収集ステップにおける任意の再サンプリング後に格納されるこの数のフレーム（または個別の画像）のような、少なくとも３つ、あるいは少なくとも５つ、さらには１０つといった、少なくとも複数のビデオフレーム分である。本明細書で使用する場合、「意図的に」という用語は、同期の問題などに基づいてそのような遅延を導入する必要性とは無関係に遅延が導入されることを意味する。言い換えれば、意図的に導入される遅延は、複数の一次ビデオストリーム２１０、３０１間の互いの時間同期をとるために、一次ビデオストリーム２１０、３０１の同期の一部として導入される遅延に加えて導入される。意図的に導入される遅延は、共通時間基準２６０との関係で、予め決められたもの、固定されたもの、または可変のものであってもよい。遅延時間は、複数の一次ビデオストリーム２１０、３０１のうち最も潜在的(latent)でない１つに関連して測定されてよく、上記時間同期の結果、これらのストリーム

２１０、３０１のより多くの潜在的なものが、相対的に小さい意図的に追加された遅延に関連付けられるようにしてもよい。

【０１５１】

いくつかの実施形態では、０．５秒以下のような比較的小さな遅延が導入される。この遅延は、出力ビデオストリーム２３０を使用するビデオ通信サービス１１０の参加者にはほとんど気づかれない。他の実施形態では、出力ビデオストリーム２３０が対話型コンテキストでは使用されず、代わりに外部消費者１５０への一方向通信で公開される場合などに、より大きな遅延が導入される可能性がある。

【０１５２】

この意図的に導入された遅延は、同期機能１３３が、収集された個別のプライマリストリーム２１０、３０１のビデオフレームを正しい共通時間基準２６０のタイムスタンプ２６１にマッピングするのに十分な時間を達成するのに十分なものであってよい。また、失われた一次ストリーム２１０、３０１の信号、スライドの変更、解像度の変更などを検出するために、上述のイベント検出を実行するのに十分な時間をもたらすのに十分なものであってよい。さらに、遅延を意図的に導入することにより、以下に説明するように、パターン検出機能１３４を改善するのに十分なものであってよい。

【０１５３】

遅延の導入は、当該バッファリングされたフレーム２１３を使用して出力ビデオストリーム２３０を公開する前に、収集され時間同期された複数の一次ビデオストリーム２１０、３０１の各々をバッファリング２５０することを含むことが理解される。言い換えれば、複数の一次ビデオストリーム２１０、３０１の少なくとも１つ、いくつか、またはすべてのビデオデータおよび／またはオーディオデータは、キャッシュのように、しかし（従来のキャッシュバッファのように）帯域幅が変化する状況に対処できることを意図して使用されるのではなく、上記の理由のために、特にパターン検出機能１３４によって使用されるように、バッファリングされた方法で中央サーバ１３０に存在することになってよい。

【０１５４】

そのため、いくつかの実施形態では、パターン検出ステップは、複数の一次デジタルビデオストリーム２１０、３０１の少なくとも１つ、例えばいくつか、例えば少なくとも４つ、またはすべて、の特定の情報を考慮することを含み、この特定の情報は、出力デジタルビデオストリーム２３０の生成においてまだ使用されていない、時間同期された一次デジタルビデオストリーム２１０のフレームよりも後のフレーム２１３に存在する。したがって、新たに追加されたフレーム２１３は、出力ビデオストリーム２３０の一部（またはその基礎）を形成する前の特定の待ち時間の間、当該バッファ２５０に存在する。この期間中、当該フレーム２１３の情報は、出力ビデオストリーム２３０の現在のフレームを生成するために現在使用されているフレームとの関係において、「未来」の情報を構成する。出力ビデオストリーム２３０のタイムラインが当該フレーム２１３に達すると、当該フレームは出力ビデオストリーム２３０の対応するフレームの生成に使用され、その後は破棄されてもよい。

【０１５５】

言い換えれば、パターン検出機能１３４は、出力ビデオストリーム２３０を生成するためにまだ使用されていないビデオ／オーディオフレーム２１３のセットを自由に利用でき、このデータを使用して上記パターンを検出する。

【０１５６】

パターン検出は、異なる方法で実行することができる。ＡＩコンポーネント１３４ａによって実行されるいくつかの実施形態では、パターン検出ステップは、第２の訓練されたニューラルネットワークまたは他の機械学習コンポーネントが、上記パターン２１２を自動的に検出するために、複数の一次デジタルビデオストリーム１２０、３０１の少なくとも２つ、例えば少なくとも３つ、例えば少なくとも４つ、あるいはすべて、を協調して(in concert)分析することを含む。

【０１５７】
いくつかの実施形態において、検出されたパターン２１２は、共有ビデオ通信サービス１１０に対する、各々がそれぞれの参加者クライアント１２１に関連付けられた、少なくとも２人、例えば少なくとも３人、例えば少なくとも４人の異なる発話中の参加者１２２を含む発話パターンを備え、これら発話中の参加者１２２の各々は、複数の一次デジタルビデオストリーム２１０、３０１のそれぞれの１つにおいて視覚的に視聴される。
【０１５８】
好ましくは、生成ステップは、出力ビデオストリーム２３０の現在の生成状態を決定し、追跡し、更新することを含む。例えば、そのような状態は、どのような参加者１２２（もしあれば）が出力ビデオストリーム２３０において可視であるか、および画面上のどこで可視であるか；いずれかの外部ビデオストリーム３００が出力ビデオストリーム２３０において可視であるか、および画面上のどこで可視であるか；いずれかのスライドまたは共有画面が、フルスクリーンモードで表示されるか、またはいずれかのライブビデオストリームと組み合わせて表示されるか、などを決定付けることができる。したがって、生成機能１３５は、生成された出力ビデオストリーム２３０に関するステートマシン(state machine)として見ることができる。
【０１５９】
出力ビデオストリーム２３０を、例えば最終消費者１５０によって視聴される結合されたビデオ体験として生成するために、中央サーバ１３０は、個々の一次ビデオストリーム２１０、３０１に関連付けられる個々のイベントを単に検出するよりも深いレベルで何が起こるかを理解できることが有利である。
【０１６０】
第１の例では、プレゼンテーションしている参加者クライアント１２１が、現在表示されているスライドを変更している。このスライド変更は、上述したようにイベント検出機能１３２によって検出され、スライド変更が起こったことを示すメタデータ２４２が当該フレームに追加される。プレゼンテーションしている参加者クライアント１２１が多数のスライドを前方に急速に連続してスキップしていることが判明し、その結果、一連の「スライド変更」のイベントが、検出機能１３２によっても検出され、当該一次ビデオストリーム２１０の個別のバッファ２５０に対応するメタデータ２４２とともに格納されるので、これは何回も起こる。実際には、そのような急速に前方にスキップされた各スライドは、ほんの数秒の間しか表示されないことがある。
【０１６１】
パターン検出機能１３４は、これらの検出されたスライド変更の複数にまたがる、当該バッファ２５０内の情報を参照して、多数のまたは急速に実行されたスライド変更ではなく、１つの単一のスライド変更に対応するパターンを検出する（すなわち、前方スキップにおける最後のスライドへの単一のスライド変更であり、一旦急速なスキップが終了すると、その最後のスライドは可視のままである）。言い換えれば、パターン検出機能１３４は、例えば、非常に短時間に１０回のスライド変更があったこと、なぜそれらが１つの単一のスライド変更を意味する検出されたパターンとして扱われるのかに注目する。その結果、生成機能１３５は、パターン検出機能１３４によって検出されたパターンへのアクセスを持ち、この最後のスライドが上記ステートマシンにおいて潜在的に重要であると判断するので、出力ビデオストリーム２３０において数秒間フルスクリーンモードでその最後のスライドを表示することを選択することができる。また、出力ストリーム２３０において、中間的に閲覧されるスライドを全く表示しないことを選択することもできる。
【０１６２】
複数の急速なスライド変更を有するパターンの検出は、単純なルールベースのアルゴリズムによって検出されてもよいが、これに替えて、分類によって動画像のそのようなパターンを検出するように設計され訓練されたニューラルネットワークを使用して検出されてもよい。
【０１６３】

別の例では、例えば、ビデオ通信がトークショー、パネル討論、または同様のものである場合に有用である可能性があるが、一方では、穏やかで滑らかな出力ビデオストリーム２３０を生成して公開することによって消費者１５０に関連する視聴体験を与えながら、一方では、現在の話者の間で視覚的注目を素早く切り替えることが望ましい可能性がある。この場合、イベント検出機能１３２は、各一次ビデオストリーム２１０、３０１を連続的に分析して、その特定の一次ビデオストリーム２１０、３０１において視聴されている人物が現在話しているか否かを常に判定することができる。これは、例えば、それ自体従来の画像処理ツールを使用して、上述のように実行され得る。次に、パターン検出機能１３４は、複数の一次ビデオストリーム２１０、３０１を含む、特定の全体的なパターン(overall patterns)を検出するように動作可能であってもよく、このパターンは、滑らかな出力ビデオストリーム２３０を生成するのに有用である。例えば、パターン検出機能１３４は、現在の話者間の非常に頻繁なスイッチのパターンおよび／または複数の同時話者を含むパターンを検出することができる。

【０１６４】

次に、生成機能１３５は、例えば、半秒しか話さずに再び無音になる話者に視覚的フォーカスを自動的に切り替えないようにしたり、両者が入れ替わりながら又は同時に話しているある時間帯に複数の話者が並んで表示される状態に切り替えるようにしたりするなど、生成状態に関連して自動化された決定を行う際に、このような検出されたパターンを考慮することができる。この状態決定プロセスは、それ自体、時系列パターン認識技術を使用して、または訓練されたニューラルネットワークを使用して実行することができるが、少なくとも部分的に所定のルールセットに基づくこともできる。

【０１６５】

いくつかの実施形態では、並行して検出され、生成機能１３５のステートマシンへの入力を形成する複数のパターンがあってもよい。このような複数のパターンは、生成機能１３５によって、異なるＡＩコンポーネント、コンピュータビジョン検出アルゴリズムなどで使用され得る。一例として、一部の参加者クライアント１２１の不安定な接続を同時に検出しながら、恒久的なスライドの変更を検出することができ、他のパターンは現在の主な発話中の参加者１２２を検出する。このような利用可能なパターンデータをすべて使用して、このようなパターンデータの時系列を分析するために、分類器ニューラルネットワークを訓練し、および／またはルールのセットを開発することができる。このような分類は、少なくとも部分的に、例えば完全に、上記の生成で使用される決定された所望の状態変化をもたらすように、教師付きにすることができる。例えば、様々な異なる生成スタイルおよび要望に従って出力ビデオストリーム２３０を自動的に生成するように特に構成された、異なるそのような所定の分類器を生成することができる。訓練は、所望の出力としての既知の生成状態変化シーケンス、および訓練データとしての既知のパターン時系列データに基づいて行うことができる。いくつかの実施形態では、ベイズモデルを使用して、そのような分類器を生成することができる。具体的な例では、経験豊富なプロデューサから先験的に情報を得ることができ、「トークショーでは、話者Ａから話者Ｂに直接切り替わることはしないが、他の話者が非常に支配的で大きな声で話す場合を除き、他の話者に焦点を当てる前に、常に最初に概要を示す」といった入力を提供することができる。この生成ロジックは、「もしＸが真であれば｜Ｙが真であるという事実が与えられれば｜Ｚを実行する」という一般的な形式のベイズモデルとして表現される。実際の検出（誰かが大きな声で話しているかどうかなど）は、分類器や閾値ベースのルールを使って行うことができる。

【０１６６】

（パターン時系列データの）大規模なデータセットがあれば、ディープラーニングの手法を用いて、ビデオストリームの自動生成に使用するための、正しく魅力的な生成フォーマットを開発することができる。

【０１６７】

要約すると、個別の複数の一次ビデオストリーム２１０、３０１に基づくイベント検出

；意図的に導入された遅延；複数の時間同期された一次ビデオストリーム２１０、３０１および検出されたイベントに基づくパターン検出；および検出されたパターンに基づく生成プロセスの組み合わせを使用することにより、テイストおよびスタイルの幅広い可能な選択に従って、出力デジタルビデオストリーム２３０の自動生成を実現することが可能になる。この結果は、イベント検出機能１３２、パターン検出機能１３４、および生成機能１３５によって使用される、可能なニューラルネットワークおよび／またはルールに基づく分析技術の広い範囲に亘って有効である。特に、第１の生成ビデオストリームが第２の生成ビデオストリームの自動生成に使用されること；および参加者クライアントの異なるグループに対して異なる意図的に追加された遅延を使用すること、を特徴とする以下に説明する実施形態においてこれは有効である。特に、検出されたトリガが、生成された出力ビデオストリームにおいてどのビデオストリームが使用されるかの切り替えをもたらすか、または出力ビデオストリームにおいて使用されるビデオストリームの自動クロップまたはズームをもたらす、後述の実施形態においても有効である。

【０１６８】

上記に例示したように、生成ステップは、出力デジタルビデオストリーム２３０における複数の一次デジタルビデオストリーム２１０、３０１の個々のものの可視性；視覚的および／または聴覚的ビデオコンテンツの配置；使用される視覚的または聴覚的効果；および／または出力デジタルビデオストリーム２３０の出力モードに関する、予め決定されたおよび／または動的に可変なパラメータのセットに基づいて、出力デジタルビデオストリーム２３０を生成することを含んでよい。このようなパラメータは、生成機能１３５のステートマシンによって自動的に決定され、および／または生成を制御するオペレータによって設定され（半自動にされ）、および／またはある先験的な構成上の要望（出力ビデオストリーム２３０のレイアウト変更または上記例示されたタイプの状態変更の間の最短時間など）に基づいて予め決定されてよい。

【０１６９】

実際の例では、ステートマシンは、フルスクリーンの発表者ビュー（現在発話中の参加者１２２をフルスクリーンで示す）；スライドビュー（現在共有されているプレゼンテーションスライドをフルスクリーンで示す）；「バタフライビュー」（現在発話中の参加者１２２と現在共有されているプレゼンテーションスライドの両方を並べたビューで示す）；マルチスピーカビュー（参加者１２２のすべてまたは選択されたサブセットを並べてまたはマトリックスレイアウトで示す）など、出力ビデオストリーム２３０に適用され得る所定の標準レイアウトのセットをサポートしてよい。様々な利用可能な生成フォーマットは、利用可能な状態のセット（上記標準レイアウトのセットなど）と共に、（上記に例示したような）ステートマシン状態変更ルールのセットによって定義することができる。例えば、そのような生成フォーマットの１つは「パネルディスカッション」であり、別の生成フォーマットは「プレゼンテーション」などである。中央サーバ１３０へのＧＵＩまたは他のインタフェースを介して特定の生成フォーマットを選択することによって、システム１００のオペレータは、予め定義されたそのような制作フォーマットのセットのうちの１つを迅速に選択することができ、次いで、中央サーバ１３０が、完全に自動的に、上記のような利用可能な情報に基づいて、当該生成フォーマットに従って出力ビデオストリーム２３０を生成することを可能にする。

【０１７０】

さらに、生成中に、上述のように、各会議参加者クライアント１２１または外部ビデオソース３００に対して、それぞれのメモリ内バッファが作成され、維持される。これらのバッファは、その場で容易に削除、追加、および変更することができる。次に、中央サーバ１３０は、出力ビデオストリーム２３０の生成中に、追加／削除(dropped-off)された参加者クライアント１２１およびスピーチを行う予定の参加者１２２；プレゼンテーションの予定されたまたは予期せぬ一時停止／再開；現在使用されている生成フォーマットに対する所望の変更などに関する情報を受信するように構成されてよい。このような情報は、例えば、上述したように、オペレータＧＵＩまたはインタフェースを介して中央サー

バ１３０に供給されてよい。
【０１７１】
上記に例示したように、いくつかの実施形態では、複数の一次デジタルビデオストリーム２１０、３０１の少なくとも１つは、デジタルビデオ通信サービス１１０に提供され、次に、公開ステップは、出力デジタルビデオストリーム２３０をその同じ通信サービス１１０に提供することを含んでよい。例えば、出力ビデオストリーム２３０は、ビデオ通信サービス１１０の参加者クライアント１２１に提供されてもよく、または、ＡＰＩ１１２を介して、ビデオ通信サービス１１０への外部ビデオストリームとして提供されてもよい。このようにして、出力ビデオストリーム２３０は、ビデオ通信サービス１１０によって現在実現されているビデオコミュニケーションイベントの参加者の複数または全員が利用できるようにすることができる。
【０１７２】
上述もしたように、追加的または代替的に、出力ビデオストリーム２３０は、１つまたは複数の外部消費者１５０に提供されてよい。
【０１７３】
一般に、生成ステップは、中央サーバ１３０によって実行され、出力デジタルビデオストリーム２３０を、ＡＰＩ１３７を介してライブビデオストリームとして１つまたは複数の同時消費者に提供することができる。
【０１７４】
上述したように、参加者クライアント１２１は、２つ以上の参加者クライアント１２１のグループに編成することができる。図８は、そのようなグループが存在する場合に出力ビデオストリームの自動生成を実行する構成におけるシステム１００の簡略図である。
【０１７５】
この図８において、中央サーバ１３０は、上記のような収集機能１３１を備えている。
【０１７６】
中央サーバ１３０はまた、第１の生成機能１３５'、第２の生成機能１３５''、および第３の生成機能１３５'''を備える。このような各生成機能１３５'、１３５''、１３５'''は、生成機能１３５に対応し、生成機能１３５に関連して上記で説明したことは、生成機能１３５'、１３５''、１３５'''にも同様に適用される。生成機能１３５'、１３５''、１３５'''は、中央サーバ１３０の詳細な構成に応じて、別個のものであってもよいし、複数の機能を１つの単一の論理関数に共有配置してもよく、３つを超える数の生成機能が存在してもよい。生成機能１３５'、１３５''、１３５'''は、場合によっては、１つの同じ生成機能１３５の異なる機能的側面であってもよい。生成機能１３５'、１３５''、１３５'''と他のエンティティとの間の様々な通信は、適切なＡＰＩを介して行われてよい。
【０１７７】
さらに、生成機能１３５'、１３５''、１３５'''またはそのような生成機能のグループのそれぞれに対して別個の収集機能１３１が存在してもよく、詳細な構成に応じて、各々がそれぞれの収集機能１３１を有する、論理的に分離された複数の中央サーバ１３０が存在してもよいことが理解される。
【０１７８】
さらに、中央サーバ１３０は、第１の公開機能１３６'、第２の公開機能１３６''、および第３の公開機能１３６'''を備える。このような各公開機能１３６'、１３６''、１３６'''は、公開機能１３６に対応し、公開機能１３６に関連して上記で説明したことは、公開機能１３６'、１３６''、１３６'''にも同様に適用される。公開機能１３６'、１３６''、１３６'''は、中央サーバ１３０の詳細な構成に応じて、別個の機能であってもよく、複数の機能を備えた１つの単一の論理機能に共有配置されてもよく、また、３つを超える数の公開機能が存在してもよい。公開機能１３６'、１３６''、１３６'''は、場合によっては、１つの同じ公開機能１３６の異なる機能的側面であってもよい。
【０１７９】
図８では、例示の目的で、参加者クライアントの３つのセットまたはグループが示され

ており、それぞれが上述の参加者クライアント１２１に対応している。したがって、このような参加者クライアント１２１の第１のグループ１２１'、このような参加者クライアントの第２のグループ１２１''、およびこのような参加者クライアントの第３のグループ１２１'''が存在する。これらのグループの各々は、１つまたは好ましくは少なくとも２つの、参加者クライアントから構成され得る。このようなグループは、詳細な構成に応じて、２つだけであってもよいし、３つ以上であってもよい。グループ１２１'、１２１''、１２１'''間の割り当ては、各参加者クライアント１２１が最大１つのグループ１２１'、１２１''、１２１'''に割り当てられるという意味で、排他的であってもよい。代替的構成では、少なくとも１つの参加者クライアント１２１が、同時に複数のそのようなグループ１２１'、１２１''、１２１'''に割り当てられてもよい。

【０１８０】

図８はまた、外部消費者１５０を示しており、上述のように、そのような外部消費者１５０が複数存在してもよいことが理解される。

【０１８１】

図８は、単純化の目的のためビデオ通信サービス１１０を示していないが、上述の一般的なタイプのビデオ通信サービスを中央サーバ１３０と共に使用してもよく、例えば、上述の方法で中央サーバ１３０を使用して各参加者クライアント１２１に共有ビデオ通信サービスを提供してもよいことが理解される。

【０１８２】

それぞれの一次ビデオストリームは、収集機能１３１によって、上記グループ１２１'、１２１''、１２１'''の参加者クライアントなどのそれぞれの参加者クライアント１２１から収集されてよい。提供された一次ビデオストリームに基づいて、生成機能１３５'、１３５''、１３５'''は、それぞれのデジタルビデオ出力ストリームを生成してよい。

【０１８３】

図８に示されるように、１つまたは複数のそのような生成された出力ストリームは、１つまたは複数のそれぞれの生成機能１３５'、１３５'''から別の生成機能１３５''にそれぞれの入力デジタルビデオストリームとして供給され、当該生成機能１３５''は、順に、公開機能１３６''による公開のための二次デジタル出力ビデオストリームを生成してよい。この二次デジタル出力ビデオストリームは、こうして、１つまたは複数の入力一次デジタルビデオストリームだけでなく、１つまたは複数の事前に生成されたデジタル入力デジタルビデオストリームに基づいて生成される。

【０１８４】

２つ以上の異なる生成ステップ１３５'、１３５''、１３５'''は、それぞれの時間遅延の導入を含んでよい。いくつかの実施形態では、これらの生成ステップ１３５'、１３５''、１３５'''からのそれぞれの生成された出力デジタルビデオストリームのうちの１つまたは複数は、上記時間遅延の導入により、上記公開ステップにおいて他の参加者クライアントに供給され得るビデオストリームのうちの任意の他のものと時間的に同期していなくてよい。このような時間遅延は、本明細書で説明される方法のいずれかにおいて、意図的に追加されてもよく、および／または、当該生成されたデジタルビデオストリームの生成の直接的な結果であってもよい。その結果、上記時間非同期の生成された出力デジタルビデオストリームを消費する任意の参加者クライアントは、上記他の参加者クライアントのビデオストリーム消費「タイムゾーン」に対して（時間的に）わずかにオフセットされた「タイムゾーン」で消費することになる。

【０１８５】

例えば、参加者クライアント１２１のグループ１２１'、１２１''、１２１'''のうちの１つは、第１のそうした「タイムゾーン」でそれぞれの生成ビデオストリームを消費してよく、一方、当該グループ１２１'、１２１''、１２１'''のうちの別の１つの参加者クライアント１２１は、第２のそうした「タイムゾーン」でそれぞれの生成ビデオストリームを消費してよい。これらのそれぞれの生成ビデオストリームは両方とも、少なくとも部分的に同じ一次ビデオストリームに基づいて生成される可能性があるため、そのような参加

者クライアント１２１はすべて、同じビデオ通信においてアクティブであるが、互いに対して異なる「タイムゾーン」にある。換言すれば、生成ビデオストリーム消費のためのそれぞれのタイムラインは、異なるグループ１２１'、１２１''、１２１'''の間で時間的にオフセットされてよい。

【０１８６】

例えば、いくつかの生成ステップ（１３５'、１３５'''等）は、（意図的に導入された時間遅延を使用することなく）直接的であり得、および／または、公開のための供給の前に計算上比較的軽量な処理のみを含み得、一方、他の生成ステップ（１３５''等）は、意図的に導入された時間遅延、および／または、比較的重量級の処理を含み得、これは、当該生成デジタルビデオストリームが、前者の生成ステップ１３５'、１３５'''のそれぞれのデジタルビデオストリームの公開のための最も早い遅延に関連する遅延で、最も早い公開のために生成されることにつながる。

【０１８７】

そのため、上記グループ１２１'、１２１''、１２１'''の１つまたは複数における各参加者クライアント１２１は、同じ知覚された時間遅延で互いに対話することができる場合がある。同時に、より大きなそれぞれの時間遅延に関連付けられるグループは、より大きな時間遅延を有するグループが当該グループの後の「タイムゾーン」で視聴する出力ビデオストリームを生成するときに、より小さな時間遅延を有するグループからの生成ビデオストリームを使用すべき入力ビデオストリームとして使用することができる。

【０１８８】

（ステップ１３５''における）この第１のより大きな時間遅延の生成の結果は、したがって、例えば、処理されたまたは処理されていない形態で、サブパートとして、１つまたは複数の当該一次ビデオストリームを視覚的に含み得る、上述のタイプの生成デジタルビデオストリームである。この当該生成ビデオストリームは、中央サーバ１３０によって生成されたビデオ出力ストリームに関連して一般的に上述されたように、ライブキャプチャされたビデオストリーム、スライド、外部から供給されたビデオまたは画像などを含み得る。当該生成ビデオストリームはまた、上述の一般的な方法で、参加者クライアント１２１によって供給される、意図的に遅延されたまたはリアルタイムの入力一次ビデオストリームの検出されたイベントおよび／またはパターンに基づいて生成されてもよい。

【０１８９】

例示的な実施形態では、第１のグループ１２１'の参加者クライアントは討論パネルの一部であり、相対的に低いレイテンシ（待ち時間）でビデオ通信サービス１１０を使用して通信し、これらの参加者クライアントの各々は、公開ステップ１３６'からの生成ビデオストリーム（または互いのそれぞれの一次ビデオストリーム）を継続的に供給される。討論パネルの聴衆は、第２のグループ１２１''の参加者クライアントによって構成され、生成ステップ１３５''からの生成ビデオストリームが連続的に供給され、ここでは僅かに高いレイテンシ（待ち時間）に関連付けられる。生成ステップ１３５''からの生成ビデオストリームは、上述の一般的な方法で、個々の討論パネルの話者のビュー（第１のグループ１２１'に割り当てられた参加者クライアント、このようなビューは、収集機能１３１から直接供給される）と、すべての討論パネルの話者を示す生成されたビュー（このビューは、第１の生成ビデオストリームである）との間で自動的にシフトするように、自動的に生成され得る。したがって、聴衆は、パネルの話者が最小の待ち時間で互いに対話することができる一方で、十分に演出された経験を享受することができる。

【０１９０】

生成ステップ１３６''で使用されるそれぞれの一次ビデオストリームに意図的に追加される遅延は、少なくとも０．１秒、例えば少なくとも０．２秒、例えば少なくとも０．５秒であってよく、最大でも５秒、例えば最大でも２秒、例えば最大でも１秒であってよい。また、意図的に追加される遅延は、使用される一次ビデオストリームのそれぞれと、生成ステップ１３５'から生成ステップ１３５''に入力される生成ビデオストリームとの間の完全な時間同期を達成するように、それぞれの一次ビデオストリームに関連付けられた

継承(inherit)待ち時間に依存してもよい。

【０１９１】

生成ステップ１３５'からの生成ビデオストリームと同様に、当該一次ビデオストリームはすべて、上述の一般的な方法で第２の生成機能１３５''で使用するためのパターン検出を改善するように、追加的に意図的に遅延させることができることが理解される。

【０１９２】

図８は、さらに、中央サーバ１３０によって生成された様々な生成ビデオストリームを公開する多数の代替または同時の方法を示す。

【０１９３】

一般に、第１の生成機能１３５'から第１の生成ビデオストリームを受信するように構成される第１の公開機能１３６'によって実行される公開ステップにおいて、上記第１の生成ビデオストリームは、第１の参加者クライアント１２１および第２の参加者クライアント１２１の少なくとも一方に連続的に供給され得る。例えば、この第１の参加者クライアントは、第１の生成機能１３５'にそれぞれの一次デジタルビデオストリームを供給するグループ１２１'からの参加者クライアントであってよい。

【０１９４】

いくつかの実施形態では、グループ１２１'の参加者クライアントの１つまたは複数は、第２の生成機能１３５''から第２の生成ビデオストリームを受信するように構成される第２の公開機能１３６''によって、第２の生成ビデオストリームを受信してもよい。

【０１９５】

したがって、第１のグループ１２１'に割り当てられている一次ビデオストリーム供給参加者クライアントの各々は、互いの一次デジタルビデオストリームが直接供給されない場合、第１の生成ビデオストリームを供給されてよく、これは、上記複数の一次ビデオストリーム間の同期に起因する一定の遅延または待ち時間を伴い、さらに、上述したように、イベントおよび／またはパターン検出のための十分な時間を確保するために、意図的に追加され得る遅延または待ち時間を伴う。

【０１９６】

これに対応して、第２のグループ１２１''に割り当てられている参加クライアントの各々は、第２の生成ビデオストリームを供給されてよく、これは、第１の生成ビデオストリームを上記第１および第２の一次ビデオストリームと時間同期させる目的で追加された、第２の生成ステップに関連して上記意図的に追加された遅延も含む。この追加的な遅延は、例えば、第２のグループ１２１''の参加者クライアントが第１のグループ１２１'の参加者クライアントとは異なる方法でビデオ通信サービス１１０と対話（相互作用）することにより、第２のグループ１２１''の参加者クライアント間のコミュニケーションを困難にする可能性があり、またはその可能性がない。

【０１９７】

したがって、第１のグループ１２１'の参加者クライアントは、当該ビデオ通信サービス１１０に現在参加しているすべての参加者クライアント１２１のサブグループを形成し、第１の生成ビデオストリームまたは第２の生成ビデオストリームなどの生成ビデオストリームが代わりに連続的に供給されるどの参加者クライアントよりもわずかに先（１～３秒先など）の「タイムゾーン」において当該サービスに存在して当該サービスを使用している。それでもなお、他の参加者クライアント（第１のグループ１２１'に割り当てられておらず、代わりに第２のグループ１２１''に割り当てられている）は、第２の生成ビデオストリームを連続的に供給されることになる。この第２の生成ビデオストリームは、第２の生成ビデオストリームがこれに基づいて生成される一次ビデオストリームの少なくとも１つに基づいて生成されるが（そして、各時点で、複数の一次ビデオストリームのどちらか一方または両方を含み得る）、わずかに後の「タイムゾーン」にある。第１の生成ビデオストリームは、これらの少なくとも１つの一次ビデオストリームに基づいて直接生成されるため、一次ビデオストリーム自体に基づいて既に生成されたビデオストリームに時間同期させるために追加される遅延や待ち時間はなく、より直接的で低レイテンシ（待ち

時間）のビデオ通信サービス１１０の経験が、これらの参加者クライアント１２１に提供される。

【０１９８】

この場合も、第１のグループ１２１'に割り当てられた参加者クライアント１２１には、第２の生成ビデオストリームへのアクセスが提供されないことを意味し得る。

【０１９９】

図８に示されるように、第２の生成ビデオストリームはまた、第３の生成ステップ１３５'''の生成出力ビデオ（第３の生成出力デジタルビデオストリーム）に追加的に基づくデジタルビデオストリームとして生成されてもよい。したがって、生成ステップ１３５'、１３５''、１３５'''の各々からの生成ビデオストリームは、少なくとも部分的に重複する一次入力デジタルビデオストリームに基づくものの上記一次ビデオストリームに対して異なるそれぞれの意図的に追加された待ち時間（「タイムゾーン」）を使用して生成され、参加者クライアント１２１の異なるそれぞれのグループ１２１'、１２１''、１２１'''に供給される。

【０２００】

第３のグループ１２１'''に割り当てられた参加者クライアント１２１は、第１のグループ１２１'に割り当てられた参加者クライアント１２１よりも、より厳密でない待ち時間要件を有してよい。例えば、第１のグループ１２１'の参加者クライアント１２１は、上述の討論パネルのメンバーであってよく（リアルタイムで互いに対話するため、低レイテンシ（待ち時間）を必要とする）、一方、第３のグループ１２１'''の参加者クライアント１２１は、パネルと対話するものの、より構造化された方法（明確な質問／回答を使用するなど）で対話する専門家パネルまたは同様のパネルを構成してよく、したがって、第１のグループ１２１'よりも高いレイテンシ（待ち時間）を許容してよい。

【０２０１】

第１の生成ビデオストリームおよび第３の生成ビデオストリームの両方は、場合によっては、第２の生成ビデオストリームの生成の基礎とするために、第２の生成機能１３５''に供給されてよい。

【０２０２】

したがって、第１の生成ステップ１３５'は、第１および第２の一次ビデオストリームの同期の一部として導入される遅延に加えて導入される、上述したタイプの意図的な遅延または待ち時間を導入することを含んでよく、これにより、例えば、効率的なイベントおよび／またはパターン検出を実行するために十分な時間が達成されてよい。このような意図的な遅延または待ち時間の導入は、上記同期機能１３３によって実行される同期の一部として行われ得る（簡略化の理由から図８には示されていない）。第３の生成ステップ１３５'''についても同様であるが、第１の生成ステップ１３５'のために意図的に導入された遅延または待ち時間とは異なる意図的な遅延または待ち時間を導入してよい。

【０２０３】

特に、意図的に導入された遅延または待ち時間は、第１の生成ビデオストリームと第３の生成ビデオストリームとの間の時間非同期をもたらす。これは、第１および第３の生成ビデオストリームが、各個別のフレームの生成時に即時かつ連続的に公開される場合に、共通のタイムラインに従わないことを意味する。

【０２０４】

したがって、３つの別個に生成されたビデオストリームが同時に生成され、消費／公開されてよいが、「タイムゾーン」は異なる。これらは少なくとも部分的に同じ一次ビデオ素材に基づいているにもかかわらず、複数の生成ビデオストリームは異なる待ち時間で公開される。最も低いレイテンシ（待ち時間）を必要とする第１のグループ１２１'は、非常に低いレイテンシ（待ち時間）を提供する第１の生成ビデオストリームを使用して対話することができる。一方、わずかに大きいレイテンシ（待ち時間）を受け入れることをいとわない第３のグループ１２１'''は、より大きいレイテンシ（待ち時間）を提供するものの他方で意図的に追加された遅延の点でより大きな柔軟性を提供する第２の生成ビデオ

ストリームを使用して対話することができ、これにより、本明細書の他の箇所に記載されているようにより良い自動生成を達成することができる。一方、待ち時間にそれほど敏感でない第２のグループ１２１''は、第１のグループ１２１'および第３のグループ１２１'''からの素材の双方を組み込むことができ、また非常に柔軟な方法で自動的に生成される第２の生成ビデオストリームを使用して対話を享受することができる。参加者ユーザのこれら全てのグループ１２１'、１２１''、１２１'''は、上記の様々な待ち時間を使用し、それゆえ異なる「タイムゾーン」で行動しているにもかかわらず、上記ビデオ通信サービス１１０を使用して互いに対話することに特に留意されたい。しかしながら、各生成機能における個々の入力ビデオストリームの同期化により、参加者ユーザ１２１は、それぞれの観点から、異なる待ち時間に気付くことはない。

【０２０５】

上述したように、前記グループ１２１'、１２１''、１２１'''のそれぞれに割り当てられたそれぞれの参加者クライアント１２１は、第２の生成ビデオストリームが連続的に公開される１つの同じビデオ通信サービス１１０に参加することができる。

【０２０６】

そして、上記グループ１２１'、１２１''、１２１'''の異なるものは、上記ビデオ通信サービス１１０における異なる参加者対話権限に関連付けられ得る。これらの実施形態および他の実施形態において、上記グループ１２１'、１２１''、１２１'''の異なるものは、当該グループ１２１'、１２１''、１２１'''に割り当てられた参加者クライアント１２１に対して公開されるそれぞれの生成ビデオストリームを生成するために使用される異なる最大時間遅延（待ち時間）に関連付けられてもよい。

【０２０７】

例えば、パネル討論の参加者クライアントの第１のグループ１２１'は、完全な対話権限に関連付けられてもよく、いつでも発言することができる。参加者クライアントの第３のグループ１２１'''は、例えば、ビデオ通信サービス１１０がマイクのミュートを解除することによって発言できるようになる前にフロアを要求する必要があるなど、わずかに制限された対話権限に関連付けられてもよい。聴衆参加者ユーザの第２のグループ１２１''は、例えば、共通のチャットルームにおいて文章で質問を投げかけることができるだけで、発言することはできないなど、さらに制限された対話権限に関連付けられてよい。

【０２０８】

したがって、参加者ユーザの様々なグループは、待ち時間が対話権限の減少の増加関数となるように、異なる対話権限と、それらに公開されるそれぞれの生成ビデオストリームに対する異なる待ち時間とに関連付けられてよい。当該参加者ユーザ１２１が、ビデオ通信サービス１１０によって、他のユーザとより自由に対話することが許可されるほど、許容されるレイテンシ（待ち時間）は低くなる。許容されるレイテンシ（待ち時間）が低いほど、対応する自動生成機能が検出されたイベントやパターンなどを考慮する可能性は小さくなる。

【０２０９】

最大の待ち時間を持つグループは、ビデオ通信サービスへの受動的な参加以外に対話権限を持たない、視聴者のみのグループであってよい。

【０２１０】

特に、上記グループ１２１'、１２１''、１２１'''の各々についてのそれぞれの最大時間遅延（待ち時間）は、すべての一次ビデオストリームと、当該グループの参加者クライアントに連続的に公開されるあらゆる生成ビデオストリームとに亘る最大の待ち時間差として決定され得る。この合計には、上述したように、イベントおよび／またはパターンを検出する目的で意図的に追加された時間遅延が追加されてもよい。

【０２１１】

本明細書で使用される場合、「生成」および「生成（された）デジタルビデオストリーム」という用語は、異なるタイプの生成を参照する場合がある。一例では、単一の、明確に定義されたデジタルビデオストリームが、中央サーバ１３０などの中央エンティティに

よって生成され、当該生成されたデジタルビデオストリームを消費することになる特定の参加者クライアント１２１の各セットに供給および公開するために、当該生成デジタルビデオストリームを形成する。他の例では、異なる個々のそのような参加者クライアント１２１は、当該生成デジタルビデオストリームのわずかに異なるバージョンを視聴することができる。例えば、生成デジタルビデオストリームは、参加者クライアント１２１のローカルソフトウエア機能１２５が当該ユーザ１２２に切り替えさせ、スクリーン１２４上に配置させ、またはその他の方法で構成もしくは処理させることができる、複数の別個または結合されたデジタルビデオストリームを含んでよい。多くの場合、重要なのは、時間同期されたサブコンポーネントを含む生成デジタルビデオストリームが、どの「タイムゾーン」で（すなわち、どの待ち時間で）供給されるかということである。したがって、図８に関連して上述した、第１のグループ１２１'の異なる参加者クライアント１２０が互いの一次ビデオストリームを供給されるケースは、第１の生成デジタルビデオストリームがこれらの参加者クライアントに供給されると見なすことができる（生データまたは処理された第１および第２の一次デジタルビデオストリームの時間同期されたセットが、第１および第２の参加者クライアントの両方で利用可能になるという意味で）。

【０２１２】

上述した参加者クライアントグループ１２１'、１２１''、１２１'''の使用をさらに明確にし、例示するために、３つの異なる同時並行「タイムゾーン」を含むビデオ通信サービス会議の形式で、以下の例を提供する：

【０２１３】

参加者クライアントの第１のグループ１２１'は、リアルタイムで、または少なくともほぼリアルタイムで（不可避のハードウエアおよびソフトウエアの待ち時間に依存するが）対話を経験している。これらの参加者クライアントは、当該ユーザ１２２間のこのような対話およびコミュニケーションを実現するために、互いから音声を含むビデオを供給される。第１のグループ１２１'は、他の（第１のグループ１２１'以外の）参加者クライアントが参加するために関心を持つ可能性のある対話を、会議の中核をなすユーザ１２２にサービス提供してよい。

【０２１４】

そのような他の参加者クライアントの第２のグループ１２１''は、同じ会議に参加するが、第１のグループの参加者クライアント１２１'よりもリアルタイムから離れた、異なる「タイムゾーン」にいる。例えば、第２のグループ１２１''は、第１のグループ１２１'に質問を投げかけることができる等の対話権限を有する聴衆であってよい。第２のグループ１２１''の「タイムゾーン」は、第１のグループ１２１'の「タイムゾーン」との関係において、提起された質問と回答が、気付くことができるが短い遅延を伴うように遅延を有してよい。一方、このわずかに大きい遅延により、この第２のグループ１２１''の参加者クライアントは、より複雑な方法で自動的に生成された、生成デジタルビデオストリームを体験することができ、より快適なユーザ体験を提供することができる。

【０２１５】

そのような他の参加者クライアントの第３のグループ１２１'''も、同じ会議に参加するが、視聴者としてのみ参加する。この第３のグループ１２１'''は、第２の「タイムゾーン」よりもさらに遅延を有する第３の「タイムゾーン」で消費される、さらに精巧かつ複雑な方法で自動的に生成され得る生成デジタルビデオストリームを消費する。しかしながら、第３のグループ１２１'''は、第１のグループ１２１'および第２のグループ１２１''に影響を与えるような方法で通信サービスに入力を供給することができないので、第３のグループ１２１'''は、会議が「リアルタイム」で行われ、納得のいく演出で行われることを体験する。

【０２１６】

当然、このような参加者クライアントのグループは４つ以上あってもよく、本明細書で説明する原則を用いて、時間的な遅延が次第に大きくなり、生成の複雑さが増す、それぞれの会議「タイムゾーン」に関連付けられる。

【０２１７】
図９は、本発明による方法を示す図であり、この方法は、共有デジタルビデオストリームを供給するためのものである。
【０２１８】
最初のステップＳ１において、方法が開始される。
【０２１９】
続く収集ステップＳ２において、第１のデジタルビデオストリームが第１のデジタルビデオソースから収集され、第２のデジタルストリームが第２のデジタルビデオソースから収集され、これらは一般的に上述した方法で行われる。したがって、第１および／または第２のデジタルビデオストリームはそれぞれ、それぞれの参加者クライアント１２１から収集されてもよく、または外部ソース３００から収集されてもよく、中央サーバ１３０の収集機能１３１によって実行されてもよい。
【０２２０】
続く第１の生成ステップＳ４において、上記共有デジタルビデオストリームが出力デジタルビデオストリームとして生成される。この生成は、生成ステップ１３５、１３５'、１３５''、１３５'''などによって、一般的に上述したように実行され得る。
【０２２１】
第１の生成ステップＳ４において、共有デジタルビデオストリームは、上記第１のデジタルビデオソースからの画像情報が共有デジタルビデオストリームにおいて可視であるが、上記第２のデジタルビデオソースからの画像情報が共有デジタルビデオストリームにおいて可視でないように、上記第１のデジタルビデオストリームの連続的に考慮された複数フレームに基づいて生成される。換言すれば、共有ビデオストリームは、少なくともある程度、第１のデジタルビデオストリームに由来する視覚的素材を含むが、第２のデジタルビデオストリームに由来する視覚的素材を含まない。
【０２２２】
続くトリガ検出ステップＳ５において、上記第１および／または第２のデジタルビデオストリームは、少なくとも１つのトリガを検出するようにデジタル的に分析される。
【０２２３】
この分析および検出は、当該共有ビデオストリームを生成するのと同じ生成ステップ１３５、１３５'、１３５''、１３５'''によって実行されてよく、所定のタイプの画像および／または音声パターンの自動検出に基づいている。
【０２２４】
トリガは、（それぞれイベント検出機能１３２およびパターン検出機能１３４によって実行される）上記で説明および例示されたタイプのイベントまたはパターンであってもよく、その検出は、典型的には、当該デジタルビデオストリーム（複数可）に含まれる音声データおよび／または画像／ビデオデータの自動デジタル処理を使用して行われる。例えば、上記で例示したように、訓練されたニューラルネットワークまたは他の機械学習ツールを使用するような自動画像処理アルゴリズムが、上記第１および／または第２のビデオフィードに含まれる画像に基づいて特定のトリガの存在を自動的に検出するために採用され得る。これに対応して、対応するタイプのそれ自体従来の自動音声処理アルゴリズムを使用して、上記第１および／または第２のビデオフィードに含まれる音声に基づいて、特定のトリガの存在を検出することができる。
【０２２５】
画像および／または音声パターンが「所定のタイプ」であるということは、問題のパターンが、当該検出に先立って定義される１つまたは複数の絶対的または相対的なパラメータ値のセットの観点から特徴付けられることを意味する。これは以下に例示される。
【０２２６】
一般に、上記音声または画像パターンの存在が、当該対応するトリガを構成する。さらに、トリガは、共有ビデオストリームを生成する自動生成ステップ１３５、１３５'、１３５''、１３５'''に対して、所定の生成ルールに従って、当該トリガの検出時に共有デジ

タルビデオストリームの生成モード（ルール）を変更するように指示するために、特定に予め定められている。

【０２２７】

したがって、生成ステップ１３５、１３５'、１３５''、１３５'''は、対応する画像および／または音声パターンを特徴付ける対応するパラメータ値に関して、即時にまたは経時的に、１つまたは複数のトリガを定義するデータベースを備えるか、または当該データベースへのアクセスを有してよい。

【０２２８】

上記トリガの検出に応答して開始される後続の第２の生成ステップＳ６またはＳ７において、共有デジタルビデオストリームは、その後、同じ（または異なる）生成ステップ１３５、１３５'、１３５''、１３５'''によって、再び生成されるが、第１の生成ステップＳ４と同じ方法ではない。

【０２２９】

第１の代替的な第２の生成ステップＳ６において、共有デジタルビデオストリームは、上記第２のデジタルビデオソースからの画像情報が共有デジタルビデオストリームにおいて可視であるように、上記第２のデジタルビデオストリームの連続して考慮される複数フレームに基づいて出力デジタルビデオストリームとして生成される。この場合、出力デジタルビデオストリームはさらに、第１のデジタルビデオソースからの画像情報が共有デジタルビデオストリームにおいて可視であるように、第１のデジタルビデオストリームの連続的に考慮される複数フレームに基づいて生成されるか、または、基づかずに生成されてもよいことに留意されたい。換言すれば、第１の生成ステップＳ４から第２の生成ステップＳ６に切り替わるとき、共有ビデオストリームは、第１のビデオストリームからのコンテンツを表示するが第２のビデオストリームからのコンテンツを表示しない状態から、第２のビデオストリームからのコンテンツを表示するが第１のビデオストリームからのコンテンツを表示しない状態になるか、または第１のビデオストリームと第２のビデオストリームの両方からのコンテンツを表示する状態になってよい。

【０２３０】

第２の代替的な第２の生成ステップＳ７において、共有デジタルビデオストリームは、上記第１のデジタルビデオソースの連続的に考慮された複数フレームに基づいて、出力デジタルビデオストリームとして生成されるが、第１の生成ステップＳ４と比較して、上記第１のデジタルビデオストリームの異なるクロッピング、異なるズーミング、異なるパンニング、および異なるフォーカスプレーンの選択のうちの少なくとも１つが行われる。換言すれば、共有ビデオストリームにおいて表示されるビデオストリームのコンテンツは、第１の生成ステップＳ４で使用されるようなビデオストリームの現在のクロップ／ズーム／パン／フォーカスプレーンの状態に対して、クロップされ、クロップ解除され、ズームインされ、ズームアウトされ、垂直方向および／または水平方向にパンされ、および／または当該ビデオストリームのフォーカスプレーンがシフトされる。

【０２３１】

このようなクロップ／パン／ズーム／フォーカスプレーンは、当該生成ステップ１３５、１３５'、１３５''、１３５'''により、既存のビデオストリーム（それ自体、ピクセルレベルで異なる画像情報を有する複数の可能なフォーカスプレーンを含む可能性がある）に基づいて、および／または、当該生成ステップ１３５、１３５'、１３５''、１３５'''が、対応するキャプチャパラメータをそれに応じて修正するために、当該ビデオストリームをキャプチャするビデオソース（デジタルビデオカメラなど）にコマンドを通信することによって、実行され得ることが理解される。例えば、これは、次に、生成ステップ１３５、１３５'、１３５''、１３５'''によって提供される指示に従って、問題のビデオストリームをキャプチャする対応するカメラがズーム、パン、および／またはそのフォーカスプレーンをシフトすることを伴ってよい。

【０２３２】

続く公開ステップＳ８において、上記出力デジタルビデオストリームは、上述の一般的

な方法で、参加者クライアント１２１および／または外部消費者１５０などの、共有デジタルビデオストリームの消費者に連続的に供給される。

【０２３３】

その後、図９に示されるように、この方法はステップＳ２に戻って繰り返すことができる。

【０２３４】

続くステップＳ９で、本方法は終了する。

【０２３５】

第１のデジタルビデオストリームは、第１のデジタルカメラによって連続的にキャプチャされてもよく、第２のデジタルビデオストリームは、第２の異なるデジタルカメラによって連続的にキャプチャされてもよい（したがって、これらは本明細書で既に使用された用語を使用して一次ビデオストリームを構成する）。あるいは、上述した異なるそれぞれの待ち時間（「タイムゾーン」）に関連付けられる参加者クライアント１２１の複数の異なるグループ１２１'、１２１''、１２１'''を使用する場合のように、第１および／または第２のデジタルビデオストリームは、それぞれの以前に生成されたデジタルビデオストリームを構成してもよく、これら以前に生成されたデジタルビデオストリームは、順に、現在生成されている共有ビデオストリームと比較して異なる待ち時間（「タイムゾーン」）で生成される（このような「タイムゾーン」に関する詳細については上記を参照）。

【０２３６】

第１のビデオストリームが既に生成されたビデオストリームである場合、クロップ／ズーム／パン設定を変更するように上流のカメラに指示するのではなく、既に存在する第１のビデオストリームに基づいて、クロップ／ズーム／パン設定が実行されることが好ましい（必須ではないが）。

【０２３７】

本方法を使用すると、それぞれのビデオストリームの実際の音声／ビデオコンテンツが、自動生成がある自動生成フォーマットから異なるそのようなフォーマットに移行することになるトリガを検出するために使用されるため、生成された共有ビデオの消費者に対してより直感的で自然な経験を提供することができる共有ビデオストリームの自動生成が達成され得る。

【０２３８】

トリガは、様々なニーズに対応するために、適切なパラメータを用いて予め定義することができる。例えば、第１および／または第２のビデオストリームに描かれている個々の人物の行動は、そのようなトリガに関して自動的に評価され、そのような行動がどのように行われるかに応じて生成（演出）フォーマットを変更することが許可されてよい。他の例では、特定の事前定義されたトリガは、生成が進行している間に、当該生成フォーマットをその場(on the fly)で変更するために、第１および／または第２のビデオストリームに描かれている人間によって与えられる、手動キューとして使用され得る。

【０２３９】

以下では、このようなトリガとそれに対応する生成フォーマットの変更の例をいくつか説明する。

【０２４０】

図１０ａおよび図１０ｂに示される第１の例では、所定のパターンは、第１のデジタルビデオカメラ４１０によってキャプチャされたものとして図１０ａに図示される、上記第１のデジタルビデオストリームに描かれた第１の（人間または例えば機械の）参加者４３０を含む。第１の参加者４３０は、図１０ａの例では、第２の（人間または機械の）参加者であるオブジェクト４４０の方を注視し、第２の参加者４４０は、順に、第２のデジタルビデオストリームに描写される。第２のデジタルビデオストリームは、この例では、第２のデジタルビデオカメラ４２０によってキャプチャされる。

【０２４１】

第２のオブジェクト４４０は、人間の参加者のグループまたは進行中の通信のための一

般的な関心対象の任意の物理的オブジェクトなどの他の何かであってもよいことが理解される。例えば、共有ビデオストリームは、医療処置の資料であってもよく、それにより、オブジェクト４４０は、患者の一部であってもよい。他の例では、オブジェクト４４０は、教育セッションまたは販売プレゼンテーションのオブジェクトであってもよい。また、オブジェクト４４０は、例えば、ホワイトボードやスライドプレゼンテーションの画面であってもよい。

【０２４２】

第１および第２のビデオストリームは、本明細書で上述したタイプのいずれであってもよい。

【０２４３】

図１０ａおよび図１０ｂに示されるケースでは、所定の画像および／または音声パターンは、上記第１のカメラ４１０、上記参加者４３０、および上記オブジェクト４４０の相対的な向きに関する情報に基づいて検出される。この情報は、セットアップ／構成中に事前に供給されたもの、および／または、生成が進行している間に自動的に検出されたものなど、システム１００内（特に中央サーバ１３０内）に存在してよい。

【０２４４】

例えば、第１のカメラ４１０および第２のカメラ４２０のそれぞれの位置は、加速度計およびジャイロを備えるＭＥＭＳ回路などの測定手段、または何らかの適切な幾何学的基準に基づいて当該カメラ４１０、４２０の現在位置を連続的または断続的に測定するように配置されたステッピングモータなどのそれ自体は従来の位置測定手段、を備えるカメラによって、当該それぞれのカメラ４１０、４２０から中央サーバ１３０に供給されてもよい。他の例では、当該カメラ４１０、４２０の向きは、適切な自動デジタル画像処理アルゴリズムを使用して、第３のカメラ（図示せず）によって検出されてもよく、第３のカメラは、カメラ４１０、４２０の少なくとも１つを画像内にキャプチャし、デジタル画像処理を使用して、このキャプチャされた画像情報に基づいて、当該相対的な向きを決定する。

【０２４５】

この文脈では、「方位(orientation)」は位置と方位の両方のコンポーネントを包含し得ることに留意されたい。

【０２４６】

何らかの適切な基準系（第１のカメラ４１０および／または第２のカメラ４２０に対してなど）に対する第１の参加者４３０およびオブジェクト４４０の位置は、第１のカメラ４１０および／または第２のカメラ４２０によってキャプチャされたビデオストリーム（複数可）に基づいて、デジタル画像処理を使用して決定することができる。

【０２４７】

図１０ａに示されるように、第１の参加者４３０は、第２の参加者４３０の方ではなく、図の下方を注視している。

【０２４８】

この例では、所定の画像および／または音声パターンは、第１のデジタルビデオストリームに基づく、第１の参加者４３０の身体方向、頭方向および視線方向の少なくとも１つのデジタル画像ベースの決定に基づいてさらに検出される。

【０２４９】

図１０ｂに示されるように、第１の参加者４３０は、第２の参加者４４０の方に向きを変えて、第２の参加者４４０の方を向いている（注視している）。

【０２５０】

第１の参加者４３０の身体および頭部の方向は、第１のカメラ４１０によってキャプチャされた、第１のビデオストリームのデジタル画像処理に基づいて決定され得る。そのようなアルゴリズムは、それ自体従来のものであり、例えば、様々な方向に向きを変えたときに第１のカメラ４１０によってキャプチャされたビデオストリームにおける第１の参加者４３０の予想される形状に関する先験的知識を使用することができる。これは、訓練さ

れたニューラルネットワークまたは他の機械学習コンポーネントを使用して達成することができる。第１のカメラ４１０の相対的な向き、第１の参加者４３０とオブジェクト４４０との相対的な位置、および第１の参加者４３０の決定された身体または頭の向きとともに、中央サーバ１３０は、第１の参加者４３０が（頭または身体を）問題のオブジェクト４４０の方に向いたかどうかを決定することができる。

【０２５１】

第１の参加者４３０の視線方向は、第１のビデオカメラ４１０によってキャプチャされた画像に基づくなど、同様の方法で達成することができる。このような視線追跡技術はそれ自体公知であり、例えば、第１の参加者４３０の目の中で視認できる瞳孔と光の反射の位置を識別することに基づいてもよい。

【０２５２】

当該トリガは、第１の参加者４３０による、オブジェクト４４０に向けられていない、またはオブジェクト４４０を注視していない状態から、第１の参加者４３０が実際にオブジェクト４４０に向いている、および／またはオブジェクト４４０を注視している状態への遷移など、遷移パターンの検出として定義することができる。したがって、中央サーバ１３０は、適切に設定された対応する絶対パラメータ値または相対パラメータ値に基づいて、そのような遷移を継続的に監視してもよく、そのような遷移が発生すると、トリガが検出されてもよい。

【０２５３】

この実施形態および他の実施形態において、複数の異なる所定の画像および／または音声パターンが同時に監視されてもよく、そのような検出された所定の画像および／または音声パターンは、次に、検出された対応するトリガを構成し、この検出された対応するトリガは、順に、所定のパラメータ化された生成ロジックに従って、異なるそれぞれの対応するモードに自動生成を切り替えることにつながることが理解される。

【０２５４】

図１０ｃは第２の例を示し、図１０ｂに示すものと類似しているが、問題のトリガに対応する所定の画像および／または音声パターンは、第１の参加者４３０がオブジェクト４４０に振り向いたり、注視したりすることではない。その代わりに、所定の画像および／または音声パターンは、（第１のカメラ４１０などによる）第１のデジタルビデオストリームに描かれた参加者（第１の参加者４３０など）が所定のジェスチャを行うことを含む。

【０２５５】

ジェスチャは、上記第２のデジタルビデオストリームに描かれたオブジェクト４４０に幾何学的に関連するジェスチャなど、所定のパラメータ化されたタイプの任意のジェスチャであってよい。具体的には、ジェスチャは、第１の参加者４３０が上記オブジェクト４４０の方を指し示すことであってもよい（図１０ｃに描かれているように、第１の参加者４３０の腕４３１がオブジェクト４４０の方を指し示している）。しかしながら、ジェスチャは、例えば、第１の参加者４３０の手または指のみに基づいてもよい。

【０２５６】

このジェスチャタイプの所定のパターン（および、特に、もしあれば、空間におけるその方向）は、図１０ｂに関連して説明した状況に対応する方法で検出されてもよく、したがって、上記第１のカメラ４１０、上記第１の参加者４３０および上記オブジェクト４４０の相対的な向きに関する情報に基づいて検出され、さらに、第１のデジタルビデオストリームに基づく第１の参加者４３０のジェスチャ方向のデジタル画像ベースの決定に基づいて検出される。

【０２５７】

図１０ｂに示される場合と図１０ｃに示される場合の両方について、第２のカメラ４２０の向きも検出され、トリガが検出されるか否かの判定に使用され得る。例えば、オブジェクト４４０が第２のカメラ４２０において視認可能であることを判定するために使用することができ、このことは、当該検出されたトリガが第２のカメラ４２０への切り替えに

関連する場合に、トリガが検出されるための条件を構成し得る。いくつかの実施形態では、第２の生成ステップＳ６は、オブジェクト４４０を現在表示している１つの第２のカメラ４２０（複数の可能な第２のカメラのうち）を決定することと、第２の生成ステップＳ６において当該第２のビデオストリームを供給するものとして当該第２のカメラ４２０を選択することとを含んでよい。

【０２５８】

図１０ｄに示す例では、カメラは１つ、すなわち第１のカメラ４１０のみである（もちろん、様々な実施形態において、より多くのカメラおよび他のビデオストリーム生成コンポーネントが存在してもよいことは理解される）。第１のカメラ４１０は、第１の参加者４３０とオブジェクト４４０の両方をキャプチャする。第１の参加者４３０がオブジェクト４４０に向かって身体、頭、または視線を向けると、この検出された画像パターンが検出されたトリガを構成する。この場合、第２の生成ステップは、当該生成出力ビデオストリームの視聴者の注意をオブジェクト４４０に集中させるように、生成出力ビデオストリームにおいて、第１のカメラ４１０によってキャプチャされたビデオストリームがパンされ、および／またはズームされ、および／またはクロッピング（トリミング）されることを含んでよい。

【０２５９】

実際の例では、第１の参加者４３０の検出された注意（上記に例示したように、第１の参加者４３０の身体、頭部、または視線方向で具体化される）は、（上記第２の生成ステップにおいて）自動生成を引き起こし、既存のビデオストリームに基づいて、または当該一次ビデオストリームを生成するユニットに指令することによって、何らかの方法で強調またはフォーカスをシフトさせ、これにより、問題のオブジェクト４４０への視覚的なフォーカスを増加させる。

【０２６０】

図１１ａおよび図１１ｂに示される別の例では、所定の画像および／または音声パターンは、第２のデジタルビデオストリームに描かれた第１の参加者４３０が第２のデジタルビデオストリームを連続的にキャプチャしている第２のカメラ４２０の方を注視することを含む。具体的には、図１１ａでは第１の参加者４３０は第２のカメラ４２０を注視していないのに対し、図１１ｂの第１の参加者４３０は第２のカメラ４２０の方向を注視している。したがって、この検出された視線方向の切り替えは、問題の対応するトリガの検出を構成し得る。

【０２６１】

他の例では、所定の画像及び／又は音声パターンは、第１のデジタルビデオストリームおよび／または第２のデジタルビデオストリームにおける動きの相対的変化を含む。例えば、第１のビデオストリームに示される一般的な動きの量がパラメータ化され、第１のビデオストリームおよび／または第２のビデオストリームのズームが、一般的な動きの減少の関数として、またはその逆の関数として、増加されてもよい。あるいは、第２の生成ステップは、検出された一般的な動きの増加の場合に、第２のビデオストリームに切り替えてもよく、第２のビデオストリームは、第１のカメラ４１０によって描かれたシーンのより広角またはより遠景のビューを示す第２のカメラ４２０によってキャプチャされる。これに対応して、記録された参加者４３０の発話音声を使用して、このようなズームイン／ズームアウト／カメラ切り替えのパフォーマンスを決定することができる。例えば、参加者４３０がより大きな声で話すことが記録されている場合、出力される共有ビデオストリームにおいて、第１のビデオストリームのズームアウトがあるかもしれず、その逆もあるかもしれない。

【０２６２】

さらに他の例では、所定の画像および／または音声パターンは、例えば特定の音捕捉マイクロフォンの相対マイクロフォン音量によって決定される、その周波数および／または振幅および／または振幅時間微分値および／または絶対振幅変化および／または音位置によって特徴付けられる所定の音声パターンを順に含んでいる。このようなマイクロフォン

は、例えば第１のカメラ４１０の一部であってもよいし、独立したマイクロフォンであってもよい。このようなマイクロフォンは、第１のビデオストリームに表示されるシーンで、またはシーンに直接関連して生じる音声を記録するように配置される。

【０２６３】

例えば、所定の音声パターンは、少なくとも１つの口頭で発声された単語を含む所定のフレーズで構成されてもよい。音声は、記録され、第１の生成ステップに供給されるビデオストリームの一部として提供されてもよく、所定のパターンは、中央サーバ１３０によって検出されてもよく、その後、対応するトリガが検出されると、生成は、第２の生成ステップに切り替わってもよい。音声解析は、所定のサウンドパターンが検出されたか否かを判定するために、様々な音声情報（ピッチ、振幅、パターンマッチングなど）を使用するルールベースの判定エンジン、訓練されたニューラルネットワークなど、適切なデジタル音声処理アルゴリズムを使用することができる。

【０２６４】

上記で詳細に説明したように、本方法はまた、遅延ステップ（図９参照）を含んでもよく、ここで、待ち時間は、少なくとも第１および第２のデジタルビデオストリームに対して意図的に導入され、この待ち時間は、共有デジタルビデオストリームに存在する。そして、トリガ検出ステップは、上記待ち時間を導入する前に、第１および／または第２のデジタルビデオストリームに基づいて実行されてもよい。

【０２６５】

待ち時間は最大３０秒、例えば最大５秒、例えば最大１秒、例えば最大０．５秒である。

【０２６６】

このような意図的に追加された遅延を使用して、自動生成は、例えば、検出されたトリガ（参加者４３０が第２のカメラ４２０を覗き込むなど）に基づいて、第１のカメラ４１０から第２のカメラ４２０への自動切り替えを計画することができ、この計画は、このイベントが生成された共有ビデオストリームに参加する一定時間（０．５秒など）前に行われる。このようにして、このような切り替えは、トリガが実際に発生する時間に正確に、または、問題の参加者４３０の発話リズムとタイミングを合わせるなど、他のパラメータ化された生成パラメータと最適に適合する時間に実行することができる。上述したタイプのグループ１２１'、１２１''、１２１'''が複数ある場合、そのような計画は、そのようなグループ１２１'、１２１''、１２１'''の異なるもののそれぞれの参加者クライアント１２１に関して（による消費のために）生成された生成出力ビデオストリームに関して、異なる時間軸(time horizon)で行われ得る。

【０２６７】

所定の画像および／または音声パターンは、イベント検出機能１３２およびパターン検出機能１３２に関連して上述したタイプの「イベント」および／または「パターン」を構成（として決定され）してよい。

【０２６８】

以下は、本発明が実際にどのように具現化され得るかを示すいくつかの例である：

【０２６９】

マルチカムプレゼンテーション、トークショー、パネルディスカッションでは、異なるカメラで発表者の異なるカメラアングルを提示することができる。自動生成機能は、発表者がどのカメラを見ているかに応じて、および／または、ジェスチャや音声キューをトリガとして、異なるカメラアングルを自動的に選択するように構成できる。

【０２７０】

ビデオポッドキャストやトーキングヘッドビデオでは、自動生成機能を、現在の話者がどのカメラに向いているかによって、異なる複数カメラの間を自動的に切り替えるよう構成できる。

【０２７１】

タウンホールミーティングでは、自動生成機能は、参加者からのインプットや質問を追

加するために、聴衆に向けたカメラに切り替えるよう構成できる。これは、当該カメラに関連する音声フィードを監視し、そのカメラがライブになる一定レベルに達したときか、または「聴衆からの質問」などの音声コマンドによって、トリガすることができる。

【０２７２】

製品のプレゼンテーションやレビューでは、自動生成機能は、そのソース(source)で動きを感知したとき、または別のトリガに基づいて、製品に向けられたカメラに自動的に切り替えるように構成することができる。

【０２７３】

ロボット手術では、手術を撮影しているロボットカメラによってキャプチャされたビデオストリームは、所定のジェスチャ、音声キューを検出したとき、または外科医が外科医コンソールを使用していないこと、または外科医コンソールから顔を上げていることを認識したときに、通常の情報提示に置き換えることができる。

【０２７４】

教育的な文脈では、カメラを通常のホワイトボードや黒板に向けるように設定し、教師がボードに向かってジェスチャをしたり、音声コマンドで指示を出したりしたときに、自動生成機能がそのカメラに切り替わるように設定することができる。

【０２７５】

コンサートなどの文化イベントでは、歌手、バンドメンバー、オーケストラに向けられた複数カメラの間を切り替えるように自動生成機能を構成することができる。これは、ジェスチャか、タレントがどのカメラを見ているかによってトリガされてよい。

【０２７６】

劇場公演では、自動生成機能は、顔追跡、音声キュー、ジェスチャによって、または所定のスケジュール（ランダウン）に従って、誰が話しているかに応じて異なるカメラアングルの間でカットするように構成することができる。

【０２７７】

したがって、本明細書で説明したタイプのトリガ検出に加えて、自動生成は、所定のスケジュール（ランダウン）に基づいて、あるフォーマット（生成ルール）から別のフォーマットに切り替えることもでき、もちろん、場合によっては手動でオーバーライドすることもできる。

【０２７８】

本発明はまた、上述の内容に従って共有デジタルビデオストリームを提供するためのコンピュータソフトウエア機能に関する。そして、このようなコンピュータソフトウエア機能は、実行時に、上述した収集、遅延、第１の生成、トリガ検出、第２の生成、および公開ステップの少なくともいくつかを実行するように構成され得る。コンピュータソフトウエア機能は、上述したように、中央サーバ１３０の物理ハードウエアまたは仮想ハードウエア上で実行されるように構成されてもよい。

【０２７９】

本発明はまた、共有デジタルビデオストリームを提供するシステムであり、順に中央サーバ１３０を構成する、そのようなシステム１００に関する。中央サーバ１３０は、順に、上述した収集、遅延、第１の生成、トリガ検出、第２の生成、および公開ステップの少なくとも一部を実行するように構成され得る。例えば、これらのステップは、中央サーバ１３０が、上述のステップを実行するための上記コンピュータソフトウエア機能を実行することによって実行されてもよい。収集は、収集機能１３１によって実行されてもよい。上述した所定の画像および／または音声パターンおよびトリガの検出は、中央サーバ１３０のイベント・パターン検出機能１３２、１３４、または生成機能１３５によって実行されてもよい。任意の意図的な遅延は、収集機能１３１などによって、一般的に上述した方法で実行されてもよい。公開は、公開機能１３６によって実行されてよい。

【０２８０】

このような入力ビデオストリームの時間同期、イベントおよび／またはパターン検出、ならびにトリガ検出などを含む、上述した入力ビデオストリームの利用可能なセットに基

づく自動生成の原理は、異なるレベルで同時並行的に適用されてもよいことが理解される。したがって、このような自動生成されたビデオストリームの１つが、ビデオストリームを順に生成する下流の自動生成機能に対する利用可能な入力ビデオストリームを形成してよい。

【０２８１】

中央サーバ１３０は、個々の参加者クライアント１２１に対するグループ１２１'、１２１''、１２１'''の割り当てを制御するように構成されてよい。例えば、ライブビデオ通信サービスセッションの過程で、特定のそのような参加者クライアントに対するグループ割り当てを動的に変更することは、中央サーバ１３０による当該ビデオ通信サービスの自動生成の一部であってもよい。このような再割当は、所定のタイムテーブルに基づいて、または、例えば、個々の参加者クライアントユーザ１２２の（当該クライアント１２１を介して提供される）要求に応じて、例えば、時間とともに動的に変化し得るパラメータデータの関数として動的にトリガされてもよい。

【０２８２】

これに対応して、中央サーバ１３０は、所定の時間枠（例えば、予定されたパネル討論の間）だけ特定のグループを使用するなど、ビデオ通信サービスの過程でグループ構成を動的に変更するように構成されてよい。

【０２８３】

所定の方法でグループ割り当てを変更することは、図９から図１１ｂに関連して上述したことに対応する方法で、特定のトリガの検出の自動的な結果であってもよい。

【０２８４】

上述したすべての態様において、本発明はさらに、対話ステップを含んでよく、この対話ステップにおいて、第１のグループの少なくとも１つの参加者クライアント（第１のグループは第１の待ち時間に関連付けられる）は、同じ第１のグループまたは第２のグループの少なくとも１つの参加者クライアント（第２のグループは第２の待ち時間に関連付けられる）と双方向で対話し、第２の待ち時間は第１の待ち時間とは異なる。これらの参加クライアントはすべて、上述のタイプの１つの同じ通信サービスの参加者であってもよいことが理解される。

【０２８５】

以上、好ましい実施形態について説明した。しかしながら、本発明の基本的な思想から逸脱することなく、開示された実施形態に多くの変更を加えることができることは、当業者には明らかである。

【０２８６】

例えば、本明細書で説明するシステム１００の一部として、本明細書では説明しない多くの追加機能を提供することができる。一般に、現在説明されている解決手段は、ビデオデータのストリームが通信に使用される多種多様な具体的アプリケーションに対応するために、詳細な機能性および特徴を構築することができるフレームワークを提供する。

【０２８７】

一般に、本方法に関して述べたことはすべて、本システムおよびコンピュータソフトウエア製品に適用可能であり、その逆もまた同様である。

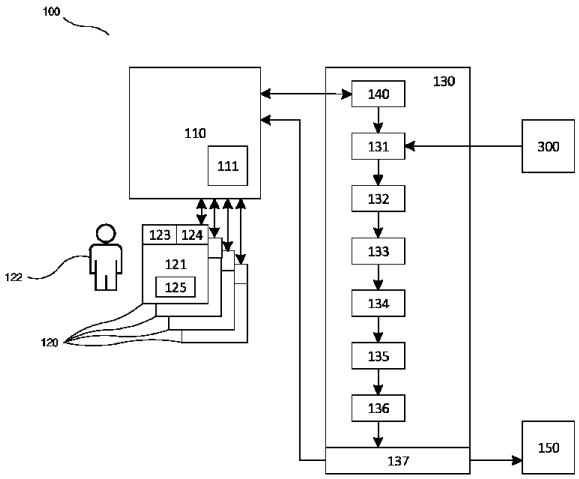
【０２８８】

したがって、本発明は、記載された実施形態に限定されるものではなく、添付の特許請求の範囲内で種々の変更が可能である。

【図面】

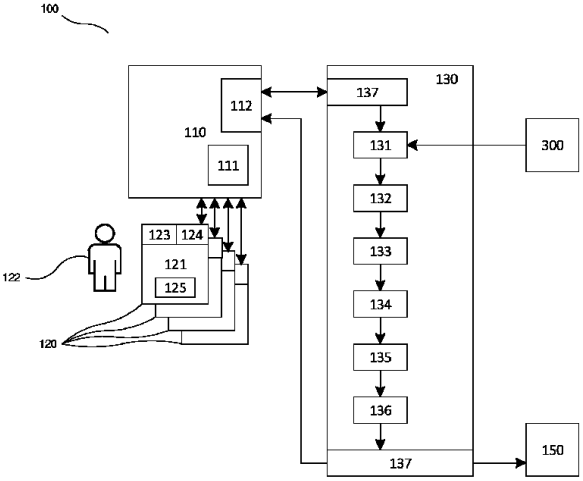
【図１】

Fig. 1



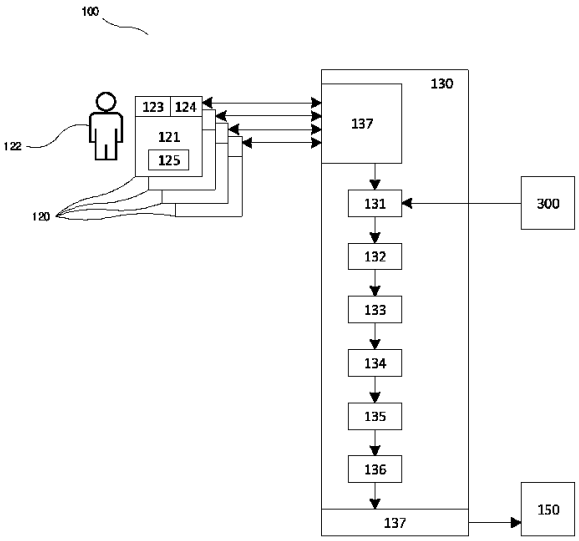
【図２】

Fig. 2



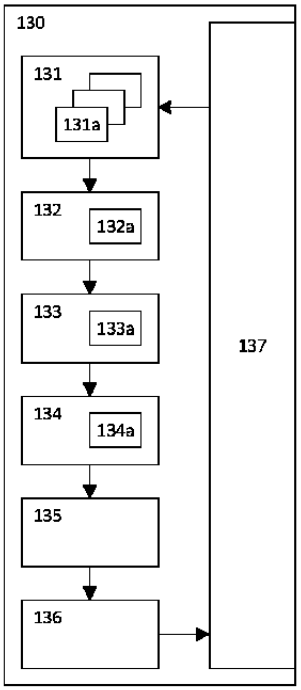
【図３】

Fig. 3

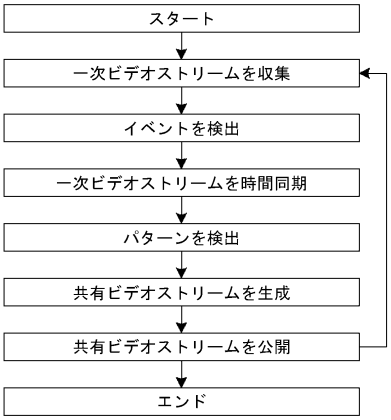


【図４】

Fig. 4

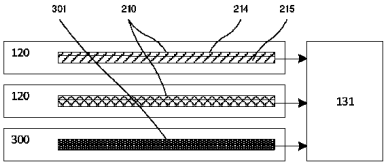


【図５】



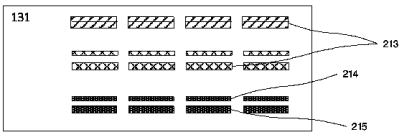
【図６ａ】

Fig. 6a



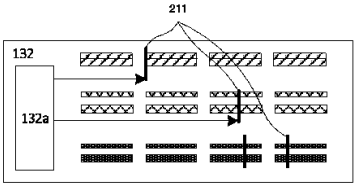
【図６ｂ】

Fig. 6b



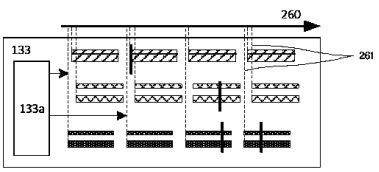
【図６ｃ】

Fig. 6c



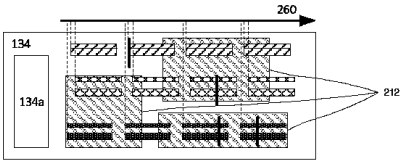
【図６ｄ】

Fig. 6d



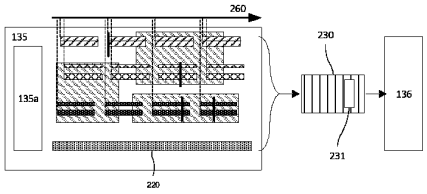
【図６ｅ】

Fig. 6e



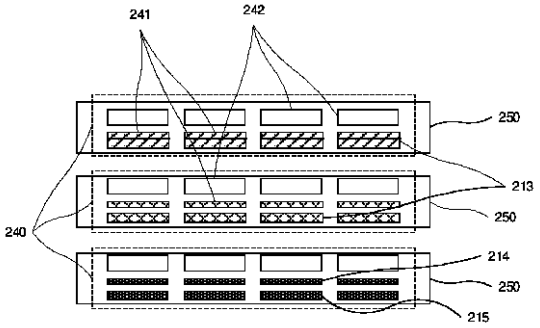
【図６ｆ】

Fig. 6f



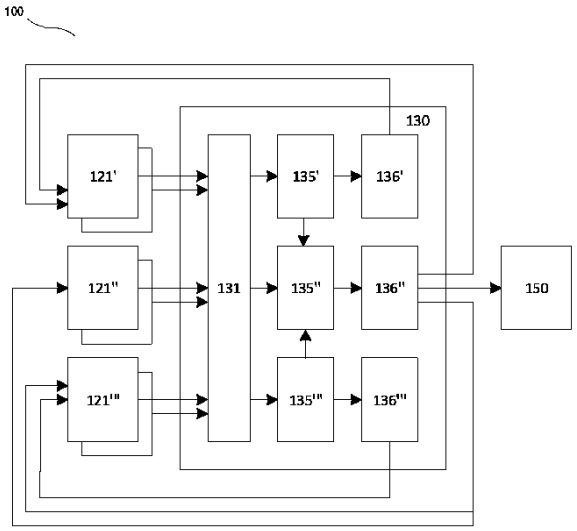
【図７】

Fig. 7

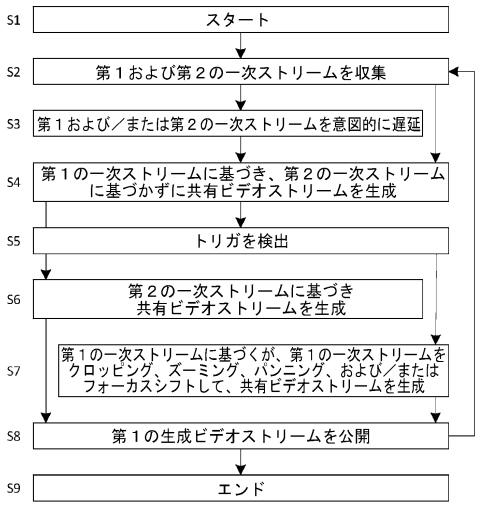


【図８】

Fig. 8

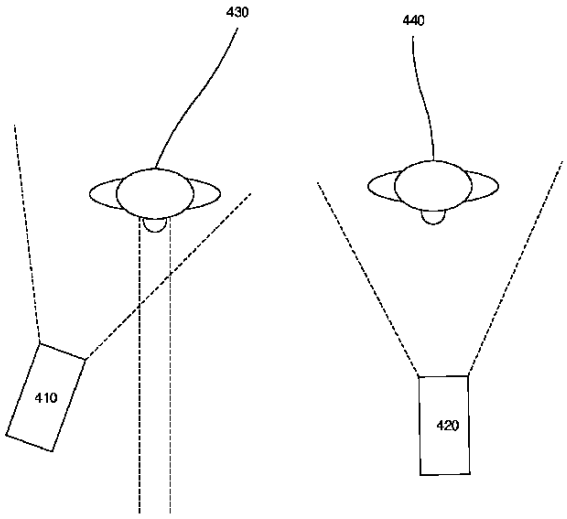


【図９】



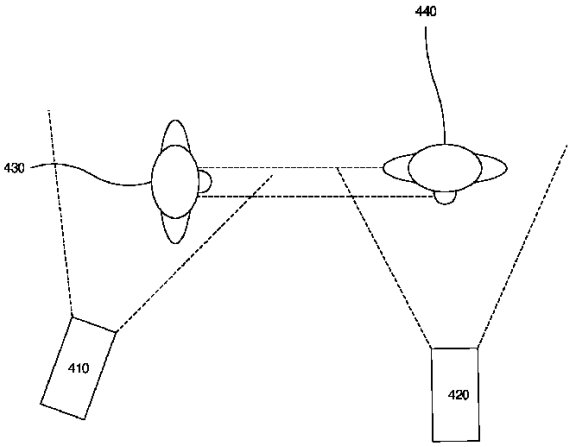
【図１０ａ】

Fig. 10a



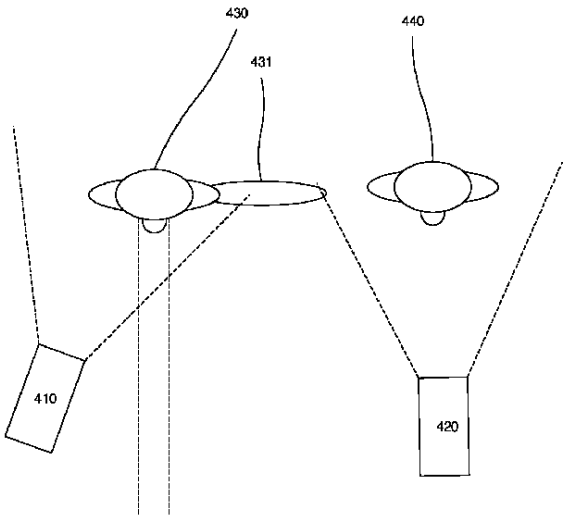
【図１０ｂ】

Fig. 10b



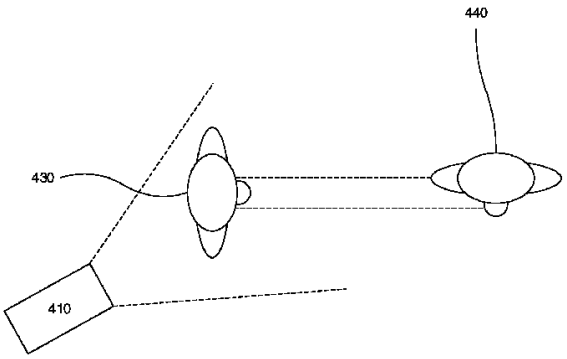
【図１０ｃ】

Fig. 10c



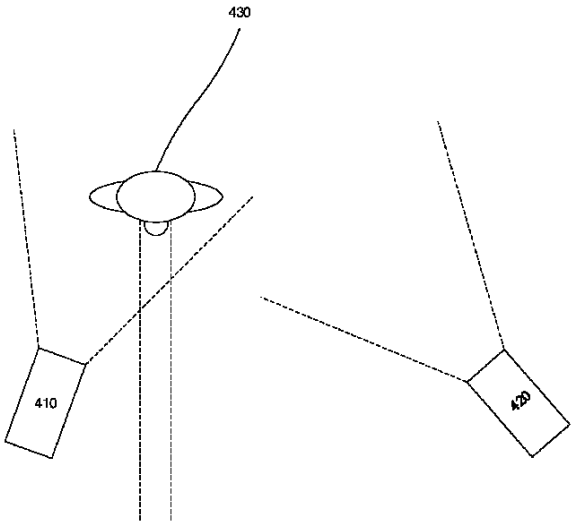
【図１１ａ】

Fig. 11a



【図１１ｂ】

Fig. 11b



【手続補正書】
【提出日】令和6年2月15日(2024.2.15)
【手続補正１】
【補正対象書類名】特許請求の範囲
【補正対象項目名】全文
【補正方法】変更
【補正の内容】
【特許請求の範囲】
【請求項１】
共有デジタルビデオストリームを提供する方法であって、該方法は以下のステップを含む方法：
収集ステップ（Ｓ２）において、第１のデジタルビデオソースから第１のデジタルビデオストリームを収集し、第２のデジタルビデオソースから第２のデジタルビデオストリームを収集し；
第１の生成ステップ（Ｓ４）において、前記第１のデジタルビデオソースからの画像情報が前記共有デジタルビデオストリームにおいて可視であり、前記第２のデジタルビデオソースからの画像情報が前記共有デジタルビデオストリームにおいて可視でないように、前記第１のデジタルビデオストリームの連続的に考慮される複数フレームに基づいて、前記共有デジタルビデオストリームを出力デジタルビデオストリームとして生成し；
遅延ステップ（Ａ３）において、前記第１および第２のデジタルビデオストリームに対して、待ち時間を意図的に導入し；
トリガ検出ステップ（Ｓ５）において、前記第１および／または第２のデジタルビデオストリームをデジタル的に分析して、少なくとも１つのトリガを検出し、ここで、前記分析は、所定のタイプの画像および／または音声パターンの自動検出に基づいており、前記トリガ検出ステップ（Ｓ５）は、前記待ち時間を導入する前の前記第１および／または第２のデジタルビデオストリームに基づいて実行され、前記パターンの存在が前記トリガを構成し、前記トリガは、所定の生成ルールに従って、前記共有デジタルビデオストリームの生成モードを変更するよう指示するものであり；
前記トリガの検出に応答して開始される第２の生成ステップ（Ｓ６，Ｓ７）において、前記第２のデジタルビデオソースからの画像情報が前記共有デジタルビデオストリームにおいて可視であるように、前記第２のデジタルビデオストリームの連続して考慮される複数フレームに基づいて、前記共有デジタルビデオストリームを出力デジタルビデオストリームとして生成し、および／または、前記第１のデジタルビデオソースの連続的に考慮される複数フレームに基づいて、前記共有ビデオストリームを出力デジタルビデオストリームとして生成するが、前記第１の生成ステップ（Ｓ４）と比較して、前記第１のデジタルビデオストリームの異なるクロッピング、異なるズーミング、異なるパンニングまたは異なるフォーカスプレーン選択のうちの少なくとも１つを行い；、
公開ステップ（Ｓ８）において、前記出力デジタルビデオストリームを、前記共有デジタルビデオストリームの消費者に継続的に提供し、ここで、前記待ち時間が前記共有デジタルビデオストリームに存在し；および、
対話ステップにおいて、前記第2の生成ステップ（Ｓ６，Ｓ７）において生成された前記共有デジタルビデオストリームに基づいて、第１の参加者ユーザと第２の参加者ユーザとの間の対話を促進し、ここで、
前記第２の生成ステップ（Ｓ６，Ｓ７）において、前記出力ビデオストリームの一部または基礎を形成する前に、導入された前記待ち時間に対応する時間の間、前記第1および／または第2のデジタルビデオストリームの新たに追加されたフレーム（２１３）がバッファ（２５０）に存在し、ここで
前記第２の生成ステップ（Ｓ６，Ｓ７）は、前記新たに追加されたフレーム（２１３）の情報を使用して前記共有ビデオストリームのフレームを生成し、前記共有ビデオストリームの前記フレームは、前記新たに追加されたフレーム（２１３）に関連して後にある前記

第1および／または第2のデジタルビデオストリームのフレームに基づいて生成される。
【請求項２】
請求項１に記載の方法であって、
前記第１のデジタルビデオストリームは、第１のデジタルカメラ（４１０）によって連続的にキャプチャされ、
第２のデジタルビデオストリームは、第２のデジタルカメラ（４２０）によって連続的にキャプチャされる、方法。
【請求項３】
請求項２に記載の方法であって、
前記所定の画像および／または音声パターンは、前記第１のデジタルビデオストリームに描かれた参加者（４３０）が、前記第２のデジタルビデオストリームに描かれたオブジェクト（４４０）に向かって注視することを含み、ここで、前記所定の画像および／または音声パターンは、前記第１のカメラ（４１０）、前記参加者（４３０）、および前記オブジェクト（４４０）の相対的な向きに関する情報に基づいて検出され、前記所定の画像および／または音声パターンは、さらに、前記第１のデジタルビデオストリームに基づく、当該参加者（４３０）の身体方向、頭部方向、および視線方向の少なくとも１つのデジタル画像ベースの決定に基づいて検出される、方法。
【請求項４】
請求項１から３のいずれか１項に記載の方法であって、
前記所定の画像および／または音声パターンは、前記第１のデジタルビデオストリームに描かれた参加者（４３０）が所定のジェスチャをすることを含む、方法。
【請求項５】
請求項４に記載の方法であって、
前記ジェスチャは、前記第２のデジタルビデオストリームに描かれたオブジェクト（４４０）に関連する、例えば、前記オブジェクト（４４０）を指し示すジェスチャであり、前記所定の画像および／または音声パターンは、前記第１のカメラ（４１０）、前記参加者（４３０）、および前記オブジェクト（４４０）の相対的な向きに関する情報に基づいて検出され、前記所定の画像および／または音声パターンは、さらに、前記第１のデジタルビデオストリームに基づく当該参加者（４３０）のジェスチャ方向のデジタル画像ベースの決定に基づいて検出される、方法。
【請求項６】
請求項１から５のいずれか１項に記載の方法であって、
前記所定の画像および／または音声パターンは、前記第２のデジタルビデオストリームに描かれた参加者（４３０）が、前記第２のデジタルビデオストリームを連続的にキャプチャしている第２のカメラ（４２０）に向かって注視していることを含む、方法。
【請求項７】
請求項１から６のいずれか１項に記載の方法であって、
前記所定の画像および／または音声パターンは、前記第１および／または第２のデジタルビデオストリームにおける相対的な動きの変化を含む、方法。
【請求項８】
請求項１から７のいずれか１項に記載の方法であって、
前記所定の画像および／または音声パターンは、例えば特定の音捕捉マイクロフォンの相対的なマイクロフォン音量によって決定される、その周波数および／または振幅および／または振幅の時間微分値および／または絶対的な振幅の変化および／または音位置によって特徴付けられる所定の音声パターンを含む、方法。
【請求項９】
請求項８に記載の方法であって、
前記所定の音声パターンは、少なくとも１つの口頭で発声された単語を含む所定のフレーズを含む、方法。
【請求項１０】

請求項１から９のいずれか１項に記載の方法であって、
前記待ち時間は、少なくとも複数ビデオフレーム、例えば少なくとも３ビデオフレーム、例えば少なくとも５ビデオフレーム、例えば少なくとも１０ビデオフレームである、方法。
【請求項１１】
請求項１から１０のいずれか１項に記載の方法であって、
前記待ち時間は、最大３０秒、例えば最大５秒、例えば最大１秒、例えば最大０．５秒である、方法。
【請求項１２】
請求項１から１１のいずれか１項に記載の方法であって、
前記画像および／または音声パターンの前記検出は、前記第１および／または第２の一次デジタルビデオストリームの特定の情報を考慮することを含み、ここで、前記特定の情報は、前記出力デジタルビデオストリームの生成にまだ使用されていないフレームよりも後のフレームに存在する、方法。
【請求項１３】
共有デジタルビデオストリームを提供するコンピュータソフトウエア製品であって、該コンピュータソフトウエアの機能は実行される際に、以下のステップを実行するコンピュータソフトウエア製品：
収集ステップ（Ｓ２）において、第１のデジタルビデオソースから第１のデジタルビデオストリームを収集し、第２のデジタルビデオソースから第２のデジタルビデオストリームを収集し；
第１の生成ステップ（Ｓ４）において、前記第１のデジタルビデオソースからの画像情報が前記共有デジタルビデオストリームにおいて可視であり、前記第２のデジタルビデオソースからの画像情報が前記共有デジタルビデオストリームにおいて可視でないように、前記第１のデジタルビデオストリームの連続的に考慮される複数フレームに基づいて、前記共有デジタルビデオストリームを出力デジタルビデオストリームとして生成し；
遅延ステップ（Ａ３）において、第１および第２のデジタルビデオストリームに対して、待ち時間を意図的に導入し；
トリガ検出ステップ（Ｓ５）において、前記第１および／または第２のデジタルビデオストリームをデジタル的に分析して、少なくとも１つのトリガを検出し、ここで、前記分析は、所定のタイプの画像および／または音声パターンの自動検出に基づいており、前記トリガ検出ステップ（Ｓ５）は、前記待ち時間を導入する前の第１および／または第２のデジタルビデオストリームに基づいて実行され、前記パターンの存在が前記トリガを構成し、前記トリガは、所定の生成ルールに従って、前記共有デジタルビデオストリームの生成モードを変更するよう指示するものであり；
前記トリガの検出に応答して開始される第２の生成ステップ（Ｓ６，Ｓ７）において、前記第２のデジタルビデオソースからの画像情報が前記共有デジタルビデオストリームにおいて可視であるように、前記第２のデジタルビデオストリームの連続して考慮される複数フレームに基づいて、前記共有デジタルビデオストリームを出力デジタルビデオストリームとして生成し、および／または、前記第１のデジタルビデオソースの連続的に考慮される複数フレームに基づいて、前記共有ビデオストリームを出力デジタルビデオストリームとして生成するが、前記第１の生成ステップ（Ｓ４）と比較して、前記第１のデジタルビデオストリームの異なるクロッピング、異なるズーミング、異なるパンニング又は異なるフォーカスプレーン選択のうちの少なくとも１つを行い；
公開ステップ（Ｓ８）において、前記出力デジタルビデオストリームを、前記共有デジタルビデオストリームの消費者に継続的に提供し、ここで、前記待ち時間が前記共有デジタルビデオストリームに存在し；および、
対話ステップにおいて、前記第2の生成ステップ（Ｓ６，Ｓ７）において生成された前記共有デジタルビデオストリームに基づいて、第１の参加者ユーザと第２の参加者ユーザとの間の対話を促進し、ここで、

前記第２の生成ステップ（Ｓ６，Ｓ７）において、前記出力ビデオストリームの一部または基礎を形成する前に、導入された前記待ち時間に対応する時間の間、前記第1および／または第2のデジタルビデオストリームの新たに追加されたフレーム（２１３）がバッファ（２５０）に存在し、ここで
前記第２の生成ステップ（Ｓ６，Ｓ７）は、前記新たに追加されたフレーム（２１３）の情報を使用して前記共有ビデオストリームのフレームを生成し、前記共有ビデオストリームの前記フレームは、前記新たに追加されたフレーム（２１３）に関連して後にある前記第1および／または第2のデジタルビデオストリームのフレームに基づいて生成される。

【請求項１４】

共有デジタルビデオストリームを提供するシステム（１００）であって、該システム（１００）は中央サーバ（１３０）を備え、該中央サーバ（１３０）は以下の機能を備えるシステム：

第１のデジタルビデオソースから第１のデジタルビデオストリームを収集し、第２のデジタルビデオソースから第２のデジタルビデオストリームを収集するよう構成される収集機能；

前記第１のデジタルビデオソースからの画像情報が前記共有デジタルビデオストリームにおいて可視であり、前記第２のデジタルビデオソースからの画像情報が前記共有デジタルビデオストリームにおいて可視でないように、前記第１のデジタルビデオストリームの連続的に考慮される複数フレームに基づいて、前記共有デジタルビデオストリームを出力デジタルビデオストリームとして生成するよう構成される第１の生成機能（１３５，１３５'，１３５''，１３５'''）；

第１および第２のデジタルビデオストリームに対して、待ち時間を意図的に導入するよう構成される遅延機能；

前記第１および／または第２のデジタルビデオストリームをデジタル的に分析して、少なくとも１つのトリガを検出するよう構成されるトリガ検出機能、ここで、前記分析は、所定のタイプの画像および／または音声パターンの自動検出に基づいており、前記分析は、前記待ち時間を導入する前の第１および／または第２のデジタルビデオストリームに基づいて実行され、前記パターンの存在が前記トリガを構成し、前記トリガは、所定の生成ルールに従って、前記共有デジタルビデオストリームの生成モードを変更するよう指示するものであり；

前記トリガの検出に応答して開始されるよう構成され、および、前記第２のデジタルビデオソースからの画像情報が前記共有デジタルビデオストリームにおいて可視であるように、前記第２のデジタルビデオストリームの連続して考慮される複数フレームに基づいて、前記共有デジタルビデオストリームを出力デジタルビデオストリームとして生成し、および／または、前記第１のデジタルビデオソースの連続的に考慮される複数フレームに基づいて、前記共有ビデオストリームを出力デジタルビデオストリームとして生成するが、前記第１の生成機能（１３５，１３５'，１３５''，１３５'''）と比較して、前記第１のデジタルビデオストリームの異なるクロッピング、異なるズーミング、異なるパンニング又は異なるフォーカスプレーン選択のうちの少なくとも１つを行うよう構成される第２の生成機能（１３５，１３５'，１３５''，１３５'''）；および、

前記出力デジタルビデオストリームを、前記共有デジタルビデオストリームの消費者に継続的に提供するよう構成される公開機能、ここで、前記待ち時間が前記共有デジタルビデオストリームに存在し；ここで、
前記中央サーバ（１３０）は、前記第2の生成機能（１３５，１３５'，１３５''，１３５'''）において生成された前記共有デジタルビデオストリームに基づいて、第１の参加者ユーザと第２の参加者ユーザとの間の対話を促進するよう構成され、ここで、
前記第２の生成機能（１３５，１３５'，１３５''，１３５'''）において、前記出力ビデオストリームの一部または基礎を形成する前に、導入された前記待ち時間に対応する時間の間、前記第1および／または第2のデジタルビデオストリームの新たに追加されたフレーム（２１３）がバッファ（２５０）に存在し、ここで

前記第２の生成機能（１３５，１３５'，１３５''，１３５'''）は、前記新たに追加されたフレーム（２１３）の情報を使用して前記共有ビデオストリームのフレームを生成し、前記共有ビデオストリームの前記フレームは、前記新たに追加されたフレーム（２１３）に関連して後にある前記第1および／または第2のデジタルビデオストリームのフレームに基づいて生成される。

【国際調査報告】

PATENT COOPERATION TREATY

PCT

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference		FOR FURTHER ACTION
220008PC2		see Form PCT/ISA/220 as well as, where applicable, item 5 below.
International application No.	International filing date *(day/month/year)*	(Earliest) Priority Date *(day/month/year)*
PCT/SE2023/050091	03-02-2023	04-02-2022
Applicant		
LIVEARENA TECHNOLOGIES AB		

This international search report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This international search report consists of a total of 5 sheets.

☐ It is also accompanied by a copy of each prior art document cited in this report.

1. **Basis of the report**

a. With regard to the **language**, the international search was carried out on the basis of:

☒ the international application in the language in which it was filed.

☐ a translation of the international application into _____ which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1(b)).

b. ☐ This international search report has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43.6*bis*(a)).

c. ☐ With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, see Box No. I.

2. ☐ **Certain claims were found unsearchable** (See Box No. II).

3. ☐ **Unity of invention is lacking** (See Box No. III).

4. With regard to the **title**,

☒ the text is approved as submitted by the applicant.

☐ the text has been established by this Authority to read as follows:

5. With regard to the **abstract**,

☒ the text is approved as submitted by the applicant.

☐ the text has been established, according to Rule 38.2, by this Authority as it appears in Box No. IV. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority.

6. With regard to the **drawings**,

a. the figure of the **drawings** to be published with the abstract is Figure No. 9

☒ as suggested by the applicant.

☐ as selected by this Authority, because the applicant failed to suggest a figure.

☐ as selected by this Authority, because this figure better characterizes the invention.

b. ☐ none of the figures is to be published with the abstract.

Form PCT/ISA/210 (first sheet) (July 2022)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/SE2023/050091

A. CLASSIFICATION OF SUBJECT MATTER

IPC: see extra sheet

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC: H04L, H04N

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

SE, DK, FI, NO classes as above

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

EPO-Internal, PAJ

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 20060251382 A1 (VRONAY DAVID ET AL), 9 November 2006 (2006-11-09); paragraphs [0012], [0054]-[0056], [0071]-[0072], [0077], [0092]-[0122], [0125], [0127]-[0128], [0148]-[0170], [0173]; figures 4,5,11-14; claims 1,3	1-14
A	US 20080297587 A1 (KURTZ ANDREW F ET AL), 4 December 2008 (2008-12-04); paragraphs [0143]-[0150]; figure 1	1-14
A	US 20100238262 A1 (KURTZ ANDREW F ET AL), 23 September 2010 (2010-09-23); whole document	1-14

☒ Further documents are listed in the continuation of Box C. ☒ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"D" document cited by the applicant in the international application
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
14-02-2023	15-02-2023
Name and mailing address of the ISA/SE Patent- och registreringsverket Box 5055 S-102 42 STOCKHOLM Facsimile No. + 46 8 666 02 86	Authorized officer Henrik Andersson Telephone No. + 46 8 782 28 00

Form PCT/ISA/210 (second sheet) (July 2022)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/SE2023/050091

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 9438937 B1 (BUCHHEIT BRIAN K), 6 September 2016 (2016-09-06); whole document --	1-14
A	US 20160148649 A1 (REHN JOHN), 26 May 2016 (2016-05-26); whole document --	1-14
A	US 20150281586 A1 (MATE SUJEET SHYAMSUNDAR ET AL), 1 October 2015 (2015-10-01); whole document --	1-14
A	US 20190215464 A1 (KUMAR NAVNEET ET AL), 11 July 2019 (2019-07-11); whole document --	1-14
A	EP 3908006 A2 (TVU NETWORKS CORP), 10 November 2021 (2021-11-10); whole document --	1-14
A	US 11082467 B1 (HARTNETT ALLISON ET AL), 3 August 2021 (2021-08-03); whole document --	1-14
A	US 20130216206 A1 (DUBIN JAY ET AL), 22 August 2013 (2013-08-22); whole document -- --------	1-14

Form PCT/ISA/210 (continuation of second sheet) (July 2022)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/SE2023/050091

Continuation of: second sheet
International Patent Classification (IPC)
H04N 21/234 (2011.01)
H04L 12/18 (2006.01)
H04N 7/15 (2006.01)
H04N 21/2187 (2011.01)

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/SE2023/050091

US	20060251382	A1	09/11/2006	US	20060251383	A1 09/11/2006
				US	20060251384	A1 09/11/2006
US	20080297587	A1	04/12/2008	US	8154578	B2 10/04/2012
				WO	2008150427	A1 11/12/2008
US	20100238262	A1	23/09/2010	US	8274544	B2 25/09/2012
US	9438937	B1	06/09/2016	US	9930416	B1 27/03/2018
US	20160148649	A1	26/05/2016	CN	105635667	B 24/08/2018
				EP	3023987	A1 25/05/2016
				JP	2016105587	A 09/06/2016
				JP	6291464	B2 14/03/2018
				KR	20160060553	A 30/05/2016
				KR	101847590	B1 10/04/2018
				TW	I684360	B 01/02/2020
				TW	201625010	A 01/07/2016
				US	10878850	B2 29/12/2020
US	20150281586	A1	01/10/2015	EP	2936794	A4 10/08/2016
				WO	2014096502	A1 26/06/2014
US	20190215464	A1	11/07/2019	WO	2019140161	A1 18/07/2019
EP	3908006	A2	10/11/2021	CN	113596484	A 02/11/2021
US	11082467	B1	03/08/2021	US	20220116437	A1 14/04/2022
				US	20220116436	A1 14/04/2022
				US	20220116435	A1 14/04/2022
				US	20220070243	A1 03/03/2022
				WO	2022051083	A1 10/03/2022
US	20130216206	A1	22/08/2013	US	8818175	B2 26/08/2014

Form PCT/ISA/210 (patent family annex) (July 2022)

フロントページの続き

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

スウェーデン国，ファルシェーピング，５２１　４７，サンクト オロフスガタン　２２

(72)発明者　エルマン，ラーズ

スウェーデン国，レクサンド，７９３　３５，ピップブラサルヴァーゲン　１２

(72)発明者　ブランケンス，パー

スウェーデン国，エネビーバーグ，１８２　４６，エンマンスヴァーグ　６ビー

(72)発明者　ソルクヴィスト，マックス

スウェーデン国，バルスタ，７４６　３６，ヴァルヴァーゲン　７

(72)発明者　ラグナーッソン，マーカス

スウェーデン国，ヨーテボリ，４１７　１７，オスタンヴィンズガタン　１３　シー

Ｆターム（参考）　5C122　DA08 EA66 EA67 FA18 FE05 GC53 GD06 HB01

5C164　FA10 SA41S SB01P SB41S VA23S YA12 YA21

【要約の続き】

る。本発明はまた、コンピュータソフトウエア製品およびシステムに関する。

【選択図】　図９