

(19)日本国特許庁(JP)

(12)公表特許公報(A)

(11)公表番号
特表2025-506378
(P2025-506378A)

(43)公表日 令和7年3月11日(2025.3.11)

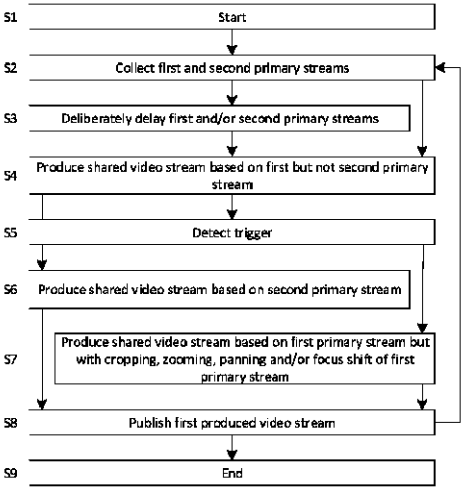
(51)国際特許分類		F I		テーマコード (参考)	
H 0 4 N	21/234 (2011.01)	H 0 4 N	21/234	5 C 1 6 4	
H 0 4 N	7/15 (2006.01)	H 0 4 N	7/15	5 E 5 5 5	
G 0 6 F	3/01 (2006.01)	G 0 6 F	3/01	5 1 0	
		G 0 6 F	3/01	5 7 0	
審査請求 未請求 予備審査請求 有 (全57頁)					
(21)出願番号	特願2024-545896(P2024-545896)	(71)出願人	524135738	最終頁に続く	
(86)(22)出願日	令和5年2月3日(2023.2.3)		ライブアリーナ テクノロジーズ アクチ		
(85)翻訳文提出日	令和6年10月1日(2024.10.1)		エボラグ		
(86)国際出願番号	PCT/SE2023/050090		L I V E A R E N A T E C H N O L O		
(87)国際公開番号	WO2023/149835		G I E S A B		
(87)国際公開日	令和5年8月10日(2023.8.10)		スウェーデン国, スtockホルム, 1 0		
(31)優先権主張番号	2250113-4		0 6 1, ボックス 1 1 2 0 8		
(32)優先日	令和4年2月4日(2022.2.4)	(74)代理人	100109380		
(33)優先権主張国・地域又は機関	スウェーデン(SE)		弁理士 小西 恵		
(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	(74)代理人	100109036		
	最終頁に続く		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, 1 8 7 7 7		
			, ステンキルスヴァーゲン 1 3		
		(72)発明者	ニルソン, アンダース		

(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（S 2）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（S 4）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不可視であるように、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；トリガ検出ステップ（S 5）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、前記トリガは、検出された注視またはジェスチャであり；第2の生成ステップ（S 6，S 7）において、前記第2のストリームおよび／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；公開ステップ（S 8）において、前記出力ストリームを消費者に提供する。本発明はまた、コンピュータソフトウェア製品およびシステムに関する。

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（S 2）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（S 4）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；

トリガ検出ステップ（S 5）において、以下の少なくとも 1 つを実行してトリガを自動的に検出し：

a) 前記第 1 のデジタルビデオストリームをデジタル的に分析して、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）に向かって注視している、または前記オブジェクト（440）に関連して所定のジェスチャを行うという形式で、前記トリガを自動的に検出すること、ここで、前記検出は、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おり、前記検出はさらに、当該参加者（430）の身体方向、頭部方向、視線方向およびジェスチャ方向のうちの少なくとも 1 つのデジタル画像ベースの決定に基づく、および、

b) 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、前記第 2 のデジタルビデオストリームに描かれている参加者（430）の視線方向が、第 2 のカメラ（420）を注視していないことから前記第 2 のカメラ（420）を注視することに遷移する形式で、トリガを自動的に検出すること、ここで、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（S 6，S 7）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（S 4）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（S 8）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 1 または 2 に記載の方法であって、

前記第 2 のオブジェクト（440）は、人間、人間のグループ、機械、ホワイトボードまたはプレゼンテーション画面である、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、
ジェスチャは、前記オブジェクト（440）を指し示すことである、方法。

【請求項 5】

請求項 1 から 4 のいずれか 1 項に記載の方法であって、
前記トリガの前記検出は、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所定の画像パターンを検出することを含む、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、以下をさらに含む方法：
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くデジタル画像処理を使用して、前記第 1 のカメラ（410）および／または前記第 2 のカメラ（420）に対して、前記参加者（430）および前記オブジェクト（440）の位置を決定する。

10

【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、
前記第 2 の生成ステップ（S6，S7）は、前記オブジェクト（440）を現在示している複数の可能な第 2 のカメラの中から特定の 1 つの第 2 のカメラ（420）を決定し、および、当該第 2 のカメラ（420）を当該第 2 のビデオストリームを提供するものとして選択すること、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、
前記第 2 の生成ステップ（S6，S7）における前記生成は、前記オブジェクト（440）への視覚的フォーカスを増加させるように、前記出力デジタルビデオストリームを強調またはフォーカスをシフトするように構成される、方法。

20

【請求項 9】

請求項 1 から 9 のいずれか 1 項に記載の方法であって、以下をさらに含む方法：
遅延ステップ（A3）であって、少なくとも前記第 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が意図的に導入され、前記待ち時間は、前記共有デジタルビデオストリームに存在するステップ、ここで、
前記トリガ検出ステップは、前記待ち時間を導入する前に、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る。

【請求項 10】

請求項 9 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。

30

【請求項 11】

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ェア製品であって、該コンピュータソフトウェ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ェア製品：

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；

40

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；

トリガ検出ステップ（S5）において、以下の少なくとも 1 つを実行してトリガを自動的に検出し：

a) 前記第 1 のデジタルビデオストリームをデジタル的に分析して、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）に向かって注視している、または前記オブジェク

50

ト（４４０）に関連して所定のジェスチャを行うという形式で、トリガを自動的に検出すること、ここで、前記検出は、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おり、前記検出はさらに、当該参加者（４３０）の身体方向、頭部方向、視線方向およびジェスチャ方向のうちの少なくとも１つのデジタル画像ベースの決定に基づく、および、

ｂ）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、前記第２のデジタルビデオストリームに描かれている参加者（４３０）の視線方向が、第２のカメラ（４２０）を注視していないことから前記第２のカメラ（４２０）を注視することに遷移する形式で、トリガを自動的に検出すること、
ここで、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する。

【請求項１２】

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'''）；

以下の少なくとも１つを実行してトリガを自動的に検出するよう構成されるトリガ検出機能：

ａ）前記第１のデジタルビデオストリームをデジタル的に分析して、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）に向かって注視している、または前記オブジェクト（４４０）に関連して所定のジェスチャを行うという形式で、トリガを自動的に検出すること、ここで、前記検出は、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おり、前記検出はさらに、当該参加者（４３０）の身体方向、頭部方向、視線方向およびジェスチャ方向の少なくとも１つのデジタル画像ベースの決定に基づく、および、

ｂ）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、前記第２のデジタルビデオストリームに描かれている参加者（４３０）の視線方向が、第２のカメラ（４２０）を注視していないことから前記第２のカメラ（４２０）を注視することに遷移する形式で、トリガを自動的に検出すること、
ここで、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、

10

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。

【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ェ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。

20

【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。

30

【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ンネルを通じて見せるかといったことが重要である。

【0005】

40

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

50

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ェ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へ供給されるため、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、ハードウェ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ェアおよび／またはソフトウェ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。

10

20

30

40

50

【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。

10

【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、共有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のデジタルビデオ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；トリガ検出ステップにおいて、以下の少なくとも1つを実行してトリガを自動的に検出し：a) 前記第1のデジタルビデオストリームをデジタル的に分析して、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に向かって注視している、または前記オブジェクトに関連して所定のジェスチャを行うという形式で、トリガを自動的に検出すること、ここで、前記検出は、第1のカメラ、前記参加者、および前記オブジェクトの相対的な向きに関する情報に基づいており、前記検出はさらに、当該参加者の身体方向、頭部方向、視線方向およびジェスチャ方向のうちの少なくとも1つのデジタル画像ベースの決定に基づく、および、b) 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、前記第2のデジタルビデオストリームに描かれている参加者の視線方向が、第2のカメラを注視していないことから前記第2のカメラを注視することに遷移する形式で、トリガを自動的に検出すること、ここで、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、異なるズームング、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する。

20

30

40

【0021】

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ェア製品に関し、該コンピュータソフトウェア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；

50

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；トリガ検出ステップにおいて、以下の少なくとも1つを実行してトリガを自動的に検出し：a) 前記第1のデジタルビデオストリームをデジタル的に分析して、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に向かって注視している、または前記オブジェクトに関連して所定のジェスチャを行うという形式で、トリガを自動的に検出すること、ここで、前記検出は、第1のカメラ、前記参加者、および前記オブジェクトの相対的な向きに関する情報に基づいており、前記検出はさらに、当該参加者の身体方向、頭部方向、視線方向およびジェスチャ方向のうちの少なくとも1つのデジタル画像ベースの決定に基づく、および、b) 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、前記第2のデジタルビデオストリームに描かれている参加者の視線方向が、第2のカメラを注視していないことから前記第2のカメラを注視することに遷移する形式で、トリガを自動的に検出すること、ここで、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する。

10

20

30

40

50

【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；以下の少なくとも1つを実行してトリガを自動的に検出するよう構成されるトリガ検出機能：a) 前記第1のデジタルビデオストリームをデジタル的に分析して、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に向かって注視している、または前記オブジェクトに関連して所定のジェスチャを行うという形式で、トリガを自動的に検出すること、ここで、前記検出は、第1のカメラ、前記参加者、および前記オブジェクトの相対的な向きに関する情報に基づいており、前記検出はさらに、当該参加者の身体方向、頭部方向、視線方向およびジェスチャ方向のうちの少なくとも1つのデジタル画像ベースの決定に基づく、および、b) 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、前記第2のデジタルビデオストリームに描かれている参加者の視線方向が、第2のカメラを注視していないことから前記第2のカメラを注視することに遷移する形式で、トリガを自動的に検出すること、ここで、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しよう構成される公開機能。

10

【0023】

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】

【0024】

【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。

20

【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。

30

【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。

40

【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】

【0025】

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。

【0026】

50

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。

【0027】

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。

【0028】

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。

【0029】

システム 100 は中央サーバ 130 を備える。

【0030】

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ェアに実装されてもよいし、ソフトウェアと仮想ハードウェアおよび／または物理ハードウェ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。

【0031】

中央サーバ 130 が実行する物理的または仮想的ハードウェア、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ェアは、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。

【0032】

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。

【0033】

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ェア、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ェアが、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。

【0034】

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。

【0035】

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

10

20

30

40

50

／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。

【0036】

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。

10

【0037】

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。

20

【0038】

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ェアを実行するか、リモートで実行されるコンピュータソフトウェアにアクセスして、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。

【0039】

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。

30

【0040】

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。

【0041】

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。

40

【0042】

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。

【0043】

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。

50

【0044】

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。

【0045】

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。

10

【0046】

これは以下に例示的に説明される。

【0047】

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。

【0048】

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。

20

【0049】

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ェア機能125を備えており、その詳細は後述する。

30

【0050】

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。

【0051】

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。

【0052】

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

40

50

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。

【0053】

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。

【0054】

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。

【0055】

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。

【0056】

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ェア上で実行されるコンピュータソフトウェ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。

【0057】

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。

【0058】

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。

10

20

30

40

50

【0059】

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。

【0060】

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。

【0061】

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。

【0062】

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。

【0063】

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。

【0064】

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。

【0065】

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

10

20

30

40

50

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。

【0066】

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つまたは複数を用意す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。

【0067】

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）。

【0068】

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。

【0069】

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。

【0070】

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。

【0071】

最初のステップで、本方法は始まる。

【0072】

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。

10

20

30

40

50

【0073】

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。

【0074】

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。

10

【0075】

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。

20

【0076】

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。

【0077】

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。

30

【0078】

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。

40

【0079】

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。

【0080】

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

50

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。

【0081】

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 6 d 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。

【0082】

図 6 d 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。

【0083】

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 6 e に示す。

【0084】

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。

【0085】

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。

【0086】

代替的实施形態では、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 のうちの複数の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。

【0087】

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A Iツールを使用して定義および検出することもできる。

【0088】

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。

【0089】

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。

【0090】

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。

【0091】

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。

【0092】

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。

【0093】

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。

10

20

30

40

50

【0094】

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。

10

【0095】

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、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。

【0096】

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MP EG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。

20

【0097】

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、一次ビデオフィールド210、301のうちの1つまたは複数、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。

30

40

【0098】

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。

【0099】

一方、外部ビデオフィールド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。

50

【0100】

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。

【0101】

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。

【0102】

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。

【0103】

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。

【0104】

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。

【0105】

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。

【0106】

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。

【0107】

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。

【0108】

フォーマット固有の収集機能 131 a 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。

10

【0109】

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 131 a のうち異なる収集機能 131 a 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 131 a 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。

20

【0110】

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。

【0111】

各フォーマット固有の収集機能 131 a 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。

【0112】

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。

30

40

【0113】

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 131 a 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

50

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。

【0114】

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。

【0115】

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。

10

【0116】

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。

【0117】

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。

20

【0118】

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、高価なハードウェアを必要とする。さらに、この問題は参加者の数によってスケールアップされる。

30

【0119】

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。

40

【0120】

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。

【0121】

【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 Licks, 1tick = 100ns. (long int)	
1 byte	0	VideoColorFormat { NV12 = 0, Rgb24 = 1, Yuy2 = 2, I1264 = 3 }	10
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)	20
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
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	40

【0 1 2 2】

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 2 6 0 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。

【0 1 2 3】

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 2 2 0 の上記少なくとも 1 つの追加部分もまた、共通時間基準 2 6 0 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 2 5 0 に格納される。

【0124】

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。

【0125】

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。

【0126】

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。

【0127】

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。

【0128】

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。

【0129】

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。

【0130】

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。

【0131】

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。

【0132】

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。

【0133】

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呼び出しでもよく、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。

10

【0134】

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。

20

【0135】

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。

【0136】

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。

30

【0137】

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。

40

【0138】

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。

【0139】

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。

50

【0140】

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。

10

【0141】

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。

20

【0142】

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ェア機能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。ローカルソフトウェア機能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ェア機能125に利用可能となる。

【0143】

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。ローカルソフトウェア機能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。

30

【0144】

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、ローカルソフトウェア機能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。

40

【0145】

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。

【0146】

50

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、ローカルソフトウェア機能 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。

10

【0147】

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 API 112 を介して、中央サーバに再び利用可能になってよい。

20

【0148】

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。

30

【0149】

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。

40

50

【0150】

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。

【0151】

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。

10

【0152】

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。

20

【0153】

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。

30

【0154】

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る、このデータを使用して上記パターンを検出する。

40

【0155】

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。

【0156】

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

50

1 1 0 に対する、各々がそれぞれの参加者クライアント 1 2 1 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 1 2 2 を含む発話パターンを備え、これら発話中の参加者 1 2 2 の各々は、複数の一次デジタルビデオストリーム 2 1 0、3 0 1 のそれぞれの 1 つにおいて視覚的に視聴される。

【0 1 5 7】

好ましくは、生成ステップは、出力ビデオストリーム 2 3 0 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 1 2 2 (もしあれば) が出力ビデオストリーム 2 3 0 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 3 0 0 が出力ビデオストリーム 2 3 0 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 1 3 5 は、生成された出力ビデオストリーム 2 3 0 に関するステートマシン(state machine)として見ることができる。

10

【0 1 5 8】

出力ビデオストリーム 2 3 0 を、例えば最終消費者 1 5 0 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 1 3 0 は、個々の一次ビデオストリーム 2 1 0、3 0 1 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。

20

【0 1 5 9】

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 1 2 1 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 1 3 2 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 2 4 2 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 1 2 1 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 1 3 2 によっても検出され、当該一次ビデオストリーム 2 1 0 の個別のバッファ 2 5 0 に対応するメタデータ 2 4 2 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。

30

【0 1 6 0】

パターン検出機能 1 3 4 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 2 5 0 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 1 3 4 は、例えば、非常に短時間に 1 0 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 1 3 5 は、パターン検出機能 1 3 4 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 2 3 0 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 2 3 0 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。

40

【0 1 6 1】

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。

【0 1 6 2】

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

50

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。

10

【0163】

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。

【0164】

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。

20

30

40

【0165】

(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。

【0166】

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

50

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。

10

【0167】

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效果；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。

20

【0168】

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。

30

【0169】

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。

40

【0170】

50

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、API 112 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。

【0171】

10

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。

【0172】

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、API 137 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。

【0173】

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。

20

【0174】

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。

【0175】

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。

30

【0176】

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。

【0177】

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。

40

【0178】

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

50

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。

【0179】

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。

10

【0180】

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。

【0181】

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。

20

【0182】

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''、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。

【0183】

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。

30

40

【0184】

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

50

れぞれのタイムラインは、異なるグループ 1 2 1'、1 2 1''、1 2 1'''の間で時間的にオフセットされてよい。

【0185】

例えば、いくつかの生成ステップ（1 3 5'、1 3 5'''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の処理のみを含み得、一方、他の生成ステップ（1 3 5''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 1 3 5'、1 3 5'''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。

10

【0186】

そのため、上記グループ 1 2 1'、1 2 1''、1 2 1'''の1つまたは複数における各参加者クライアント 1 2 1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。

【0187】

（ステップ 1 3 5''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 1 3 0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 1 2 1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。

20

【0188】

例示的な実施形態では、第1のグループ 1 2 1'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 1 1 0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 1 3 6'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 1 2 1''の参加者クライアントによって構成され、生成ステップ 1 3 5''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 1 3 5''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 1 2 1'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 1 3 1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。

30

40

【0189】

生成ステップ 1 3 6''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 1 3 5'から生成ステップ 1 3 5''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。

【0190】

50

生成ステップ 1 3 5'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 1 3 5''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。

【0191】

図 8 は、さらに、中央サーバ 1 3 0 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。

【0192】

一般に、第 1 の生成機能 1 3 5'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 1 3 6'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 1 2 1 および第 2 の参加者クライアント 1 2 1 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 1 3 5'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 1 2 1'からの参加者クライアントであってよい。

10

【0193】

いくつかの実施形態では、グループ 1 2 1'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 1 3 5''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 1 3 6''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。

【0194】

したがって、第 1 のグループ 1 2 1'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。

20

【0195】

これに対応して、第 2 のグループ 1 2 1''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 1 2 1''の参加者クライアントが第 1 のグループ 1 2 1'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 1 1 0 と対話（相互作用）することにより、第 2 のグループ 1 2 1''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。

30

【0196】

したがって、第 1 のグループ 1 2 1'の参加者クライアントは、当該ビデオ通信サービス 1 1 0 に現在参加しているすべての参加者クライアント 1 2 1 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 1 2 1'に割り当てられておらず、代わりに第 2 のグループ 1 2 1''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 1 1 0 の経験が、これらの参加者クライアント 1 2 1 に提供される。

40

50

【0197】

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。

【0198】

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。

10

【0199】

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。

20

【0200】

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。

【0201】

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。

30

【0202】

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。

【0203】

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

40

50

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。

【0204】

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。

【0205】

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。

【0206】

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。

【0207】

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。

【0208】

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。

【0209】

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。

【0210】

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

10

20

30

40

50

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ェア機能 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）。

10

【0211】

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：

【0212】

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（不可避のハードウェアおよびソフトウェ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。

20

【0213】

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。

30

【0214】

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。

40

【0215】

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。

【0216】

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

50

ムを供給するためのものである。

【0217】

最初のステップS1において、方法が開始される。

【0218】

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。

10

【0219】

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。

【0220】

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。

20

【0221】

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。

【0222】

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。

【0223】

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のビデオフィール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のビデオフィール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。

30

40

【0224】

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。

【0225】

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。

50

【0226】

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。

【0227】

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。

10

【0228】

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。

20

【0229】

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、異なるズーム、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。

30

【0230】

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。

40

【0231】

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。

50

【0232】

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。

【0233】

続くステップS9で、本方法は終了する。

【0234】

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）。

10

【0235】

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）。

20

【0236】

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。

【0237】

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。

30

【0238】

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。

【0239】

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。

40

【0240】

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、

50

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。

【0241】

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。

【0242】

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。

【0243】

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。

【0244】

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。

【0245】

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。

【0246】

図10aに示されるように、第1の参加者430は、第2の参加者440の方ではなく、図の下方を注視している。

【0247】

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。

【0248】

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）。

【0249】

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44

10

20

30

40

50

0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。

【0250】

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。

【0251】

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。

【0252】

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ということが理解される。

【0253】

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。

【0254】

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。

【0255】

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。

【0256】

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るが、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

ラ 4 2 0（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 S 6 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 4 2 0 を選択することを含んでよい。

【0257】

図 1 0 d に示す例では、カメラは 1 つ、すなわち第 1 のカメラ 4 1 0 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 4 1 0 は、第 1 の参加者 4 3 0 とオブジェクト 4 4 0 の両方をキャプチャする。第 1 の参加者 4 3 0 がオブジェクト 4 4 0 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において、第 1 のカメラ 4 1 0 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。

10

【0258】

実際の例では、第 1 の参加者 4 3 0 の検出された注意（上記に例示したように、第 1 の参加者 4 3 0 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 4 4 0 への視覚的なフォーカスを増加させる。

20

【0259】

図 1 1 a および図 1 1 b 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 4 3 0 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 4 2 0 の方を注視することを含む。具体的には、図 1 1 a では第 1 の参加者 4 3 0 は第 2 のカメラ 4 2 0 を注視していないのに対し、図 1 1 b の第 1 の参加者 4 3 0 は第 2 のカメラ 4 2 0 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。

【0260】

30

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变化を含む。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 4 1 0 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 4 2 0 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 4 3 0 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 4 3 0 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ない、その逆もあるかもしれない。

40

【0261】

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 4 1 0 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

50

、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。

【0262】

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。

10

【0263】

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。

【0264】

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。

【0265】

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。

20

30

【0266】

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。

【0267】

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：

【0268】

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。

40

【0269】

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。

【0270】

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、

50

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。

【0271】

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)で動きを検知したとき、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。

【0272】

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。

10

【0273】

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、自動生成機能はそのカメラに切り替わるように設定することができる。

【0274】

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。

【0275】

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。

20

【0276】

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。

【0277】

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ェア機能に関する。そして、このようなコンピュータソフトウェア機能は、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ェア機能は、上述したように、中央サーバ130の物理ハードウェアまたは仮想ハードウェア上で実行されるように構成されてもよい。

30

【0278】

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ェ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。

40

【0279】

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

50

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。

【0280】

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。

10

【0281】

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。

【0282】

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。

【0283】

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。

20

【0284】

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。

30

【0285】

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。

【0286】

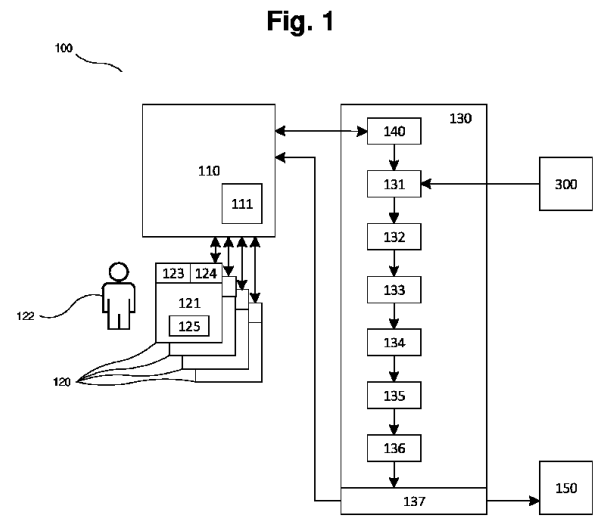
一般に、本方法に関して述べたことはすべて、本システムおよびコンピュータソフトウェア製品に適用可能であり、その逆もまた同様である。

【0287】

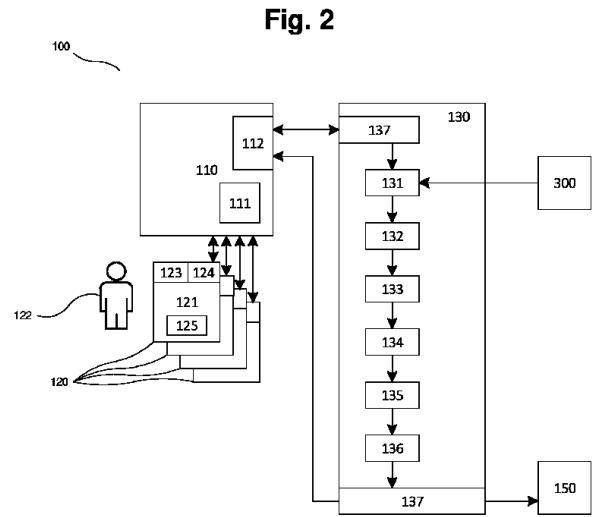
したがって、本発明は、記載された実施形態に限定されるものではなく、添付の特許請求の範囲内で種々の変更が可能である。

40

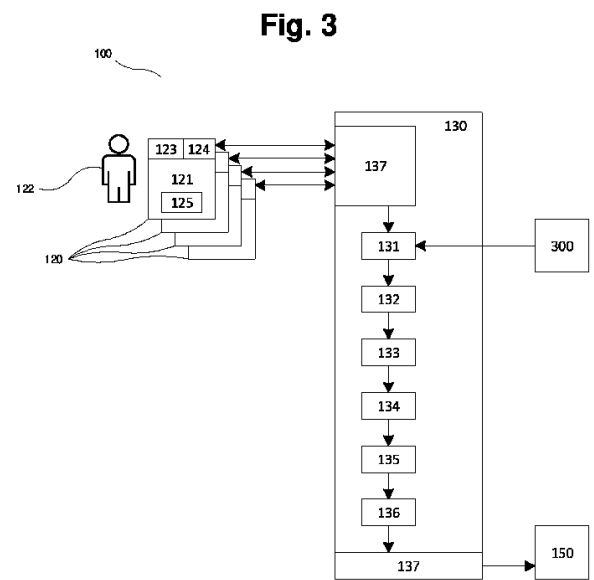
【図面】
【図 1】



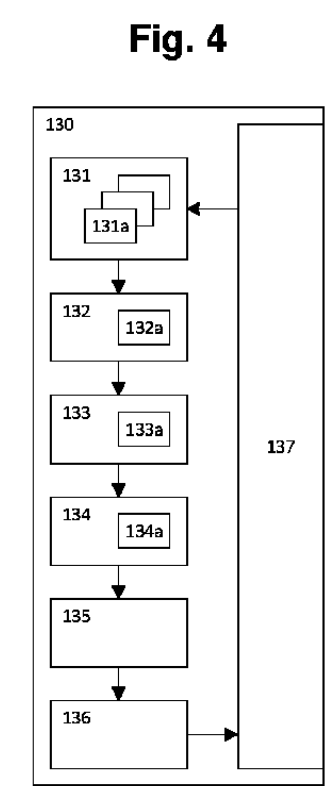
【図 2】



【図 3】



【図 4】



10

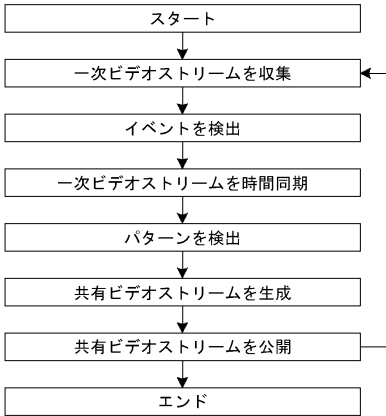
20

30

40

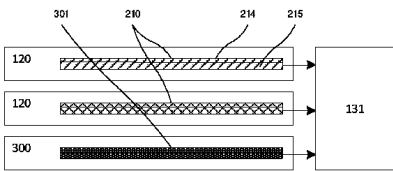
50

【 図 5 】



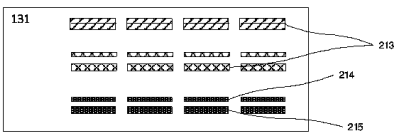
【 図 6 a 】

Fig. 6a



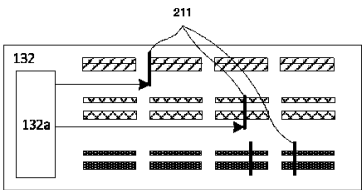
【 図 6 b 】

Fig. 6b



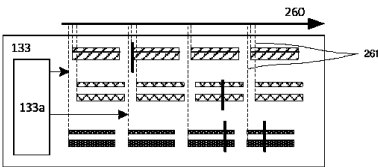
【 図 6 c 】

Fig. 6c



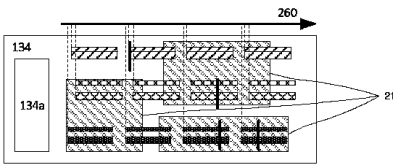
【 図 6 d 】

Fig. 6d



【 図 6 e 】

Fig. 6e



10

20

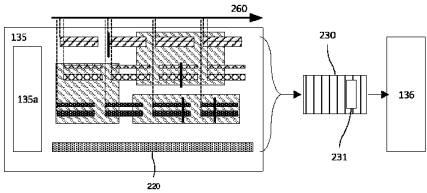
30

40

50

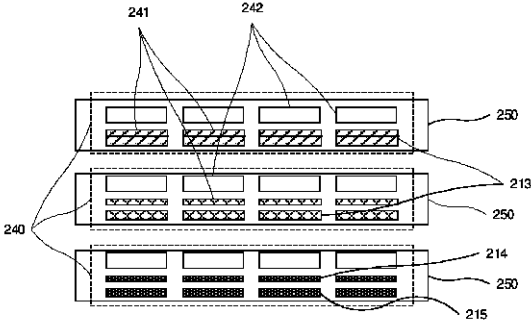
【図 6 f】

Fig. 6f



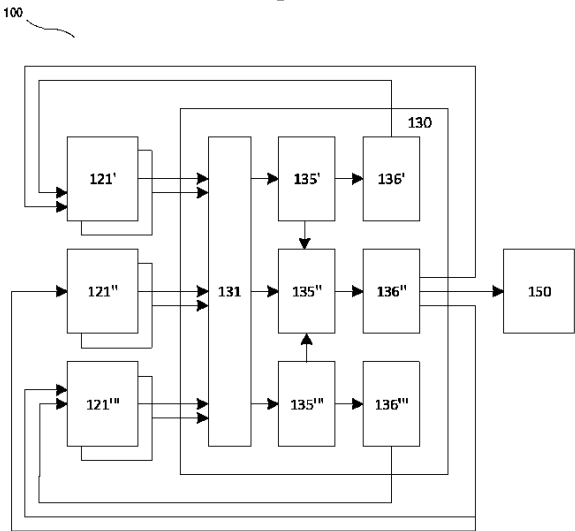
【図 7】

Fig. 7

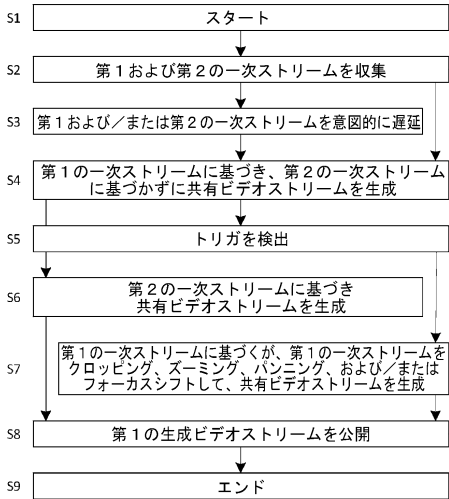


【図 8】

Fig. 8



【図 9】



10

20

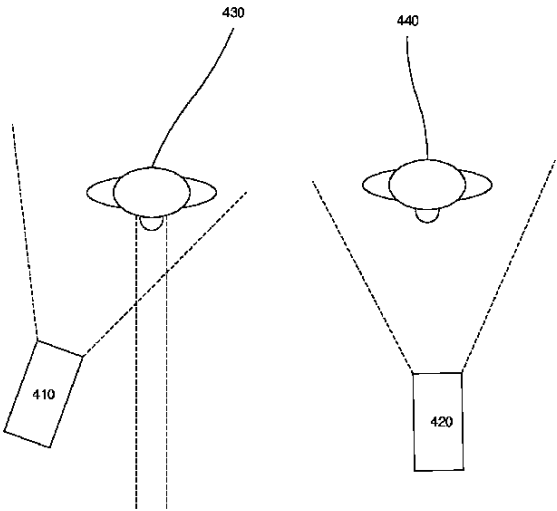
30

40

50

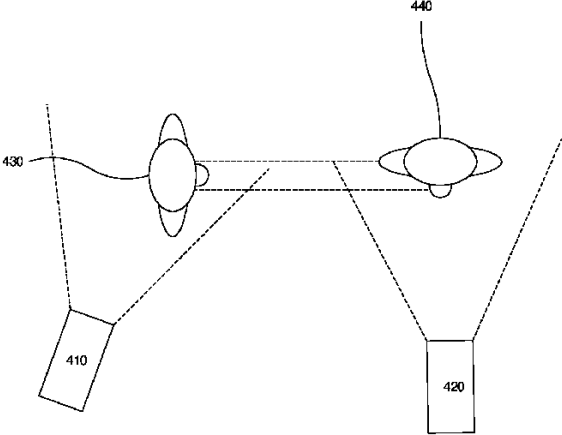
【図 1 0 a】

Fig. 10a



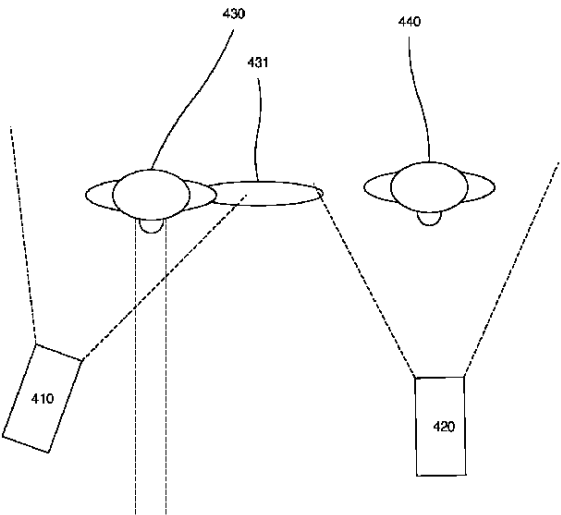
【図 1 0 b】

Fig. 10b



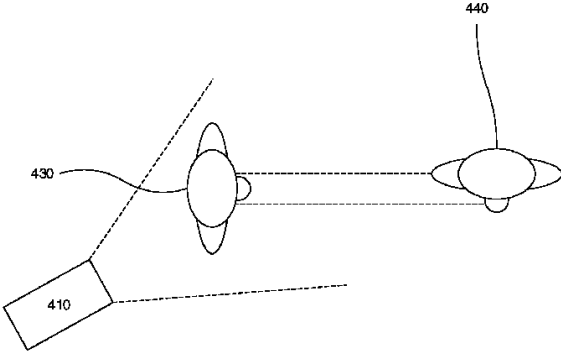
【図 1 0 c】

Fig. 10c



【図 1 1 a】

Fig. 11a



10

20

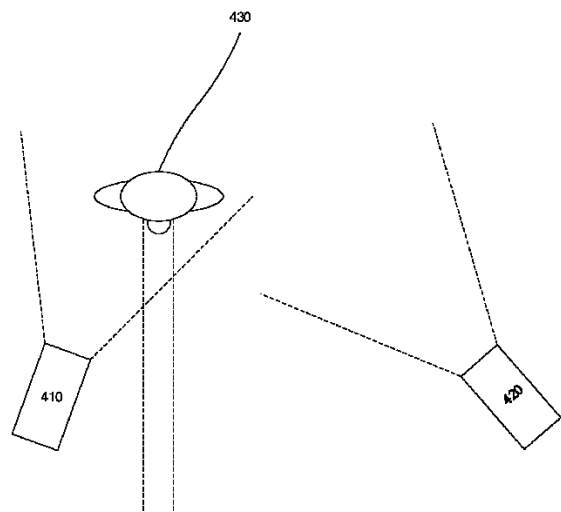
30

40

50

【 図 1 1 b 】

Fig. 11b



10

20

30

40

50

【国際調査報告】

INTERNATIONAL SEARCH REPORT

International application No.
PCT/SE2023/050090

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 --	1-12
A	US 20080297587 A1 (KURTZ ANDREW F ET AL), 4 December 2008 (2008-12-04); paragraphs [0143]-[0150] --	1-12
A	US 20100238262 A1 (KURTZ ANDREW F ET AL), 23 September 2010 (2010-09-23); whole document --	1-12
☒ 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 14-02-2023		Date of mailing of the international search report 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/SE2023/050090

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 --	1-12
A	US 20160148649 A1 (REHN JOHN), 26 May 2016 (2016-05-26); whole document --	1-12
A	US 20150281586 A1 (MATE SUJEET SHYAMSUNDAR ET AL), 1 October 2015 (2015-10-01); whole document --	1-12
A	US 20190215464 A1 (KUMAR NAVNEET ET AL), 11 July 2019 (2019-07-11); whole document -- -----	1-12

10

20

30

40

INTERNATIONAL SEARCH REPORT

International application No.
PCT/SE2023/050090

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)

10

20

30

40

50

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/SE2023/050090

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

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,C
O,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,I
R,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, 5 2 1 4 7, サンクト オロフスガタン 2 2

(72)発明者 エルマン, ラーズ

スウェーデン国, レクサンド, 7 9 3 3 5, ピプブラサルヴァーゲン 1 2

(72)発明者 ブランケンス, パー

スウェーデン国, エネビーバーク, 1 8 2 4 6, エンマンズヴァーグ 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, 7 4 6 3 6, 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, 4 1 7 1 7, オスタンヴィンズガタン 1 3 シー

F ターム (参考) 5C164 FA06 FA07 FA10 SA25S SB01P SB41S VA23P YA12

5E555 AA64 BA02 BA13 BB02 BB13 CA42 CB65 CB66 CC03 DB57

DC07 EA22 FA00

【要約の続き】

【選択図】 図 9