

**UNITED STATES PATENT AND TRADEMARK OFFICE**

---

**BEFORE THE PATENT TRIAL AND APPEAL BOARD**

---

Cisco Systems, Inc.,  
Petitioner,  
v.  
Zugara, Inc.,  
Patent Owner.

---

IPR2026-00266

U.S. Patent No. 10,200,654

---

**PETITION FOR *INTER PARTES* REVIEW OF**

**U.S. PATENT NO. 10,200,654**

**TABLE OF CONTENTS**

|                                                                                                                         | <b>Page</b> |
|-------------------------------------------------------------------------------------------------------------------------|-------------|
| I. INTRODUCTION .....                                                                                                   | 1           |
| II. MANDATORY NOTICES .....                                                                                             | 1           |
| A. Real Party-in-Interest .....                                                                                         | 1           |
| B. Related Matters.....                                                                                                 | 1           |
| C. Lead and Back-Up Counsel and Service Information .....                                                               | 1           |
| III. PAYMENT OF FEES .....                                                                                              | 2           |
| IV. REQUIREMENTS FOR <i>INTER PARTES</i> REVIEW .....                                                                   | 2           |
| A. Grounds for Standing .....                                                                                           | 2           |
| B. Identification of Challenged Claims .....                                                                            | 3           |
| V. THE 654 PATENT .....                                                                                                 | 3           |
| A. Effective Filing Date .....                                                                                          | 3           |
| B. Person of Ordinary Skill in the Art.....                                                                             | 3           |
| C. Overview of the 654 Patent.....                                                                                      | 4           |
| VI. CLAIM CONSTRUCTION .....                                                                                            | 5           |
| VII. PRINCIPAL PRIOR ART .....                                                                                          | 6           |
| A. Summary of Rosenthal .....                                                                                           | 6           |
| 1. Priority Date.....                                                                                                   | 6           |
| 2. Overview of Rosenthal .....                                                                                          | 6           |
| B. Summary of Vesely .....                                                                                              | 7           |
| 1. Priority Date.....                                                                                                   | 7           |
| 2. Overview of Vesely.....                                                                                              | 8           |
| C. The Combination of Rosenthal and Vesely .....                                                                        | 9           |
| VIII. THE CHALLENGED CLAIMS ARE UNPATENTABLE .....                                                                      | 15          |
| A. Ground 1: Claims 1, 12–16, 21–23, 26–27, 29–30, 32–36, 43–45,<br>48–50, and 53 Are Obvious in View of Rosenthal..... | 15          |
| 1. Independent Claim 32:.....                                                                                           | 15          |

|     |                                                                                                     |    |
|-----|-----------------------------------------------------------------------------------------------------|----|
| 2.  | Dependent Claims 33–36, 43–45, 48–50, 53 .....                                                      | 40 |
| 3.  | Claims 1, 12–16, 21–23, 26–27, 29–30 .....                                                          | 57 |
| B.  | Ground 2: Claims 14, 17, 19, 21, 43, and 46–48 Are Obvious in<br>View of Rosenthal and Vesely ..... | 59 |
| 1.  | Dependent Claims 43 and 46–48 .....                                                                 | 59 |
| 2.  | Claims 14, 17, 19, 21 .....                                                                         | 66 |
| IX. | CONCLUSION .....                                                                                    | 67 |
|     | <b>CERTIFICATE OF COMPLIANCE</b> .....                                                              | 68 |
|     | <b>CERTIFICATE OF SERVICE</b> .....                                                                 | 69 |
|     | <b>CLAIM LISTING</b> .....                                                                          | 71 |

**TABLE OF AUTHORITIES****Page(s)****Cases**

|                                                                                                       |    |
|-------------------------------------------------------------------------------------------------------|----|
| <i>AC Techs. S.A. v. Amazon.com, Inc.</i> ,<br>912 F.3d 1358 (Fed. Cir. 2019) .....                   | 54 |
| <i>Nidec Motor Corp. v. Zhongshan Broad Ocean Motor Co.</i> ,<br>868 F.3d 1013 (Fed. Cir. 2017) ..... | 5  |
| <i>United Servs. Auto. Ass’n v. PNC Bank N.A.</i> ,<br>2025 WL 706080 (Fed. Cir. Mar. 5, 2025) .....  | 54 |
| <i>Vivid Techs., Inc. v. Am. Sci. &amp; Eng’g Inc.</i> ,<br>200 F.3d 795 (Fed. Cir. 1999) .....       | 5  |

**Regulations**

|                             |   |
|-----------------------------|---|
| 37 C.F.R. § 42.100(b) ..... | 5 |
|-----------------------------|---|

**EXHIBIT LIST**

| <b>No.</b> | <b>Exhibit Description</b>                                                                                                                                                                                                   |
|------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1001       | U.S. Patent No. 10,200,654                                                                                                                                                                                                   |
| 1002       | File History of U.S. Patent No. 10,200,654                                                                                                                                                                                   |
| 1003       | Declaration of Dr. Houh                                                                                                                                                                                                      |
| 1004       | CV of Dr. Houh                                                                                                                                                                                                               |
| 1005       | U.S. Pub. No. 2007/0242066 (“Rosenthal”)                                                                                                                                                                                     |
| 1006       | Preliminary Infringement Contentions (“PIC”)                                                                                                                                                                                 |
| 1007       | U.S. Pub. No. 2012/0162384 (“Vesely”)                                                                                                                                                                                        |
| 1008       | Scott Hanselman, <i>2006 is the Year of Video Chat</i> , Hanselman (Jan. 7, 2006), <a href="https://www.hanselman.com/blog/2006-is-the-year-of-video-chat">https://www.hanselman.com/blog/2006-is-the-year-of-video-chat</a> |
| 1009       | Jonathan Roubini, <i>Yahoo! Messenger</i> , PC Mag UK (Jan. 18, 2006), <a href="https://uk.pcmag.com/smb/28728/yahoo-messenger">https://uk.pcmag.com/smb/28728/yahoo-messenger</a>                                           |
| 1010       | Microsoft Computer Dictionary, definition of “MPEG,” “data compression” and “latency”                                                                                                                                        |

**TABLE OF ABBREVIATIONS AND CONVENTIONS**

| <b>Abbreviation</b> | <b>Meaning</b>                                                       |
|---------------------|----------------------------------------------------------------------|
| 654 Patent          | Ex.1001: U.S. Patent No. 10,200,654                                  |
| IPR                 | <i>inter partes</i> review                                           |
| Petitioner          | Cisco Systems, Inc. (“Cisco”)                                        |
| Patent Owner or PO  | Zugara, Inc. (“Zugara”)                                              |
| <i>xx:yy–zz</i>     | column <i>xx</i> , lines <i>yy</i> to <i>zz</i>                      |
| PIC                 | Ex.1006: Patent Owner’s Preliminary Infringement Contentions (“PIC”) |

## I. INTRODUCTION

Petitioner submits this Petition for *Inter Partes* Review of claims 1, 12–17, 19, 21–23, 26–27, 29–30, 32–36, 43–50, and 53 (the “Challenged Claims”) of U.S. Patent No. 10,200,654 (“the 654 Patent” (Ex.1001)), assigned to Patent Owner (“PO” or “Zugara, Inc.”). Petitioner respectfully submits that the Challenged Claims of the 654 Patent are unpatentable under 35 U.S.C. § 103 in view of the prior art references discussed herein. Accordingly, it is respectfully requested that the Board institute an *inter partes* review of the 654 Patent.

## II. MANDATORY NOTICES

### A. Real Party-in-Interest

The real party-in-interest for this Petition is Cisco Systems, Inc.

### B. Related Matters

The 654 Patent was asserted by PO in *Zugara, Inc. v. Cisco Systems, Inc.*, 2-25-cv-00437 (EDTX). On February 24, 2026, the case was dismissed without prejudice.

### C. Lead and Back-Up Counsel and Service Information

| Lead Counsel                                                                                                                                                    | Back-Up Counsel                                                                                                                                                                           |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Scott M. Border</b><br>(Reg. #77,744)<br>WINSTON & STRAWN LLP<br>1901 L Street, N.W.<br>Washington, D.C. 20036<br>Tel: (202) 282-5054<br>sborder@winston.com | <b>Krishnan Padmanabhan</b><br>( <i>pro hac</i> vice to be submitted)<br>WINSTON & STRAWN LLP<br>200 Park Avenue<br>New York, NY 10166<br>Tel: (212) 294-6700<br>kpadmanabhan@winston.com |

|  |                                                                                                                                                                                                                                                                                                                                                                                              |
|--|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|  | <p><b>Christopher T. Gresalfi</b><br/>(Reg. # 78,949)<br/>WINSTON &amp; STRAWN LLP<br/>200 Park Avenue<br/>New York, NY 10166<br/>Tel: (212) 294-6700<br/>cgresalfi@winston.com</p> <p><b>Brady Mills</b><br/>(<i>pro hac vice</i> to be submitted)<br/>WINSTON &amp; STRAWN LLP<br/>800 Capital Street, Suite 2400<br/>Houston, TX 77002<br/>Tel: (713) 651-2600<br/>bmills@winston.com</p> |
|--|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

Petitioner consents to electronic service by email at [WS-Cisco-Zugara@winston.com](mailto:WS-Cisco-Zugara@winston.com) and the e-mail addresses listed above.

### III. PAYMENT OF FEES

Petitioner authorizes the Office to charge the filing fee and any other necessary fee to Deposit Account No. 501814.

### IV. REQUIREMENTS FOR *INTER PARTES* REVIEW

#### A. Grounds for Standing

Petitioner certifies that the 654 Patent is available for *inter partes* review. Petitioner is not barred or estopped from requesting an *inter partes* review challenging the claims on the identified grounds herein. Petitioner has not filed a civil action challenging the validity of a claim of the 654 Patent. This petition is being filed no more than 1 year after the date on which Petitioner was served with a complaint alleging infringement of the 654 Patent.

**B. Identification of Challenged Claims**

**Ground 1:** Claims 1, 12–16, 21–23, 26–27, 29–30, 32–36, 43–45, 48–50, and 53 are unpatentable as obvious under 35 U.S.C. § 103 based on Rosenthal.

**Ground 2:** Claims 14, 17, 19, 21, 43, and 46–48 are unpatentable as obvious under 35 U.S.C. § 103 based on Rosenthal and Vesely.

**V. THE 654 PATENT**

**A. Effective Filing Date**

Petitioner assumes for the purposes of this Petition that **February 27, 2013**, is the effective filing date.

**B. Person of Ordinary Skill in the Art**

A Person of Ordinary Skill in The Art (“POSITA”) in February 2013 would have had a working knowledge of the electronic communications and networking art that is pertinent to the 654 Patent. A POSITA would have had a bachelor’s degree in electrical engineering, computer science, computer engineering, or an equivalent, and two years of relevant experience in digital video processing, video communications systems, computer networking, and interactive media systems. Such experience would have included familiarity with real-time video streaming, feature or object tracking within video streams (including, for example, three-dimensional coordinate tracking), and techniques for augmenting or compositing

content into video data. Lack of professional experience can be remedied by additional education, and vice versa. Ex.1003, ¶24.

### **C. Overview of the 654 Patent**

The application underlying the 654 Patent was filed on February 27, 2013. The 654 Patent issued February 5, 2019.

The 654 Patent addresses the synchronization of augmented reality (AR) systems. Ex.1001, Abstract, 2:5-10. Specifically, the claims of the 654 Patent are directed at “allow[ing] real time manipulation of objects within multiple dynamic and augmented video streams/feeds synchronized in an augmented reality (AR) or multi-dimensional space.” Ex.1001, Abstract. This synchronization is intended to allow multiple AR users to participate in one augmented video stream. Ex.1001, Abstract; Ex.1003, ¶34.

The 654 Patent generally describes that synchronizing the augmented video streams of users creates an interactive AR environment in a video chat or conference platform, enabling multiple users to simultaneously interact with the virtual objects in the AR video stream. Ex.1001, 2:24-30. By combining the augmented video streams of each user, the users “continually see and interact with augmented objects in the synchronized augmented video streams across multiple AR systems or platforms.” Ex.1001, 2:35-40. The method claim recites the combination of a real-world video stream of a real-world 3D environment including a participant with an

augmented video stream from another participant which includes 3D virtual objects. This combination results in the display of a single video stream that comprises the video of the participant's real-world 3D environment and the other participant's virtual 3D virtual objects. Ex.1001, cl. 32; Ex.1003, ¶¶35–39.

The 654 Patent touts the benefit of the above methodology in the context of virtual shopping, “allow[ing] the initial/first user to connect his/her augmented video stream with other users so that all users could see what (virtual) clothes other people are trying on.” Ex.1001, 2:45-47.

## **VI. CLAIM CONSTRUCTION**

Claims are given their “ordinary and customary meaning” as understood by a POSITA and the prosecution history pertaining to the patent. 37 C.F.R. § 42.100(b). Because a POSITA would find the challenged claims unpatentable under any interpretation consistent with their plain and ordinary meaning in the context of the 654 Patent, the Board need not expressly construe the claim terms. *See, e.g., Nidec Motor Corp. v. Zhongshan Broad Ocean Motor Co.*, 868 F.3d 1013, 1017 (Fed. Cir. 2017) (quoting *Vivid Techs., Inc. v. Am. Sci. & Eng’g, Inc.*, 200 F.3d 795, 803 (Fed. Cir. 1999)) (“[W]e need only construe terms ‘that are in controversy, and only to the extent necessary to resolve the controversy.’”). Petitioner applies the plain and ordinary meaning of the Challenged Claims to the asserted prior art references in support of each ground below.

## VII. PRINCIPAL PRIOR ART

### A. Summary of Rosenthal

#### 1. Priority Date

U.S. Publication No. 2007/0242066 (“Rosenthal”), published October 18, 2007, was based on App. No. 11/783,995, claims priority to Prov. App. 60/791,894, filed April 14, 2006, and expressly incorporates it by reference. Rosenthal, [0001]. Rosenthal is therefore prior art at least under pre-AIA 35 U.S.C. § 102(a), (b) and (e) because the earliest possible priority date of the 654 Patent is February 27, 2013, several years after Rosenthal was published.

#### 2. Overview of Rosenthal

Rosenthal generally relates to “the real-time manipulation of video images,” describing a method and apparatus that “provide a hardware-independent virtual camera that may be seamlessly integrated with existing video camera and computer system equipment.” Rosenthal, [0003], Abstract. This enables the ability “to track a defined set of three-dimensional coordinates within a video stream” and also provides the capability to “dynamically insert rendered 3-D objects within the video stream on a real-time basis.” Rosenthal, Abstract; Ex.1003, ¶45.

Rosenthal identifies the desire to manipulate video streams in a variety of ways which are compatible with a wide range of user video-based applications and done in real time. Rosenthal, [0005]. For example, “applications may include real-

time manipulation of a video stream generated by a video camera associated with, for example, a real-time video conference.” Rosenthal, [0007]. As another example, “a bi-directional data interface may support a wide range of interactive capabilities that allow injected content and characteristics to be controllably passed across virtual camera video streams in support of interactive gaming.” Rosenthal, [0013]. Accordingly, “the described approach may be applied to the visualization of geospatial imaging data...and/or any number of user applications.” Rosenthal, [0014]. Moreover, as explained by Dr. Houh, Rosenthal provides a level of detail sufficient to have enabled a person of ordinary skill in the art to have practiced the techniques described therein and to create and deploy the created 3-D virtual and bi-directional system, without undue experimentation, as of at least the priority date of the challenged claims. Ex.1003, ¶46.

## **B. Summary of Vesely**

### **1. Priority Date**

U.S. Publication No. 2012/0162384 (“Vesely”), published June 28, 2012, was based on App. No. 13/333,339, claims priority to Prov. App. 61/426,451, filed December 22, 2010, and expressly incorporates it by reference. Vesely, [0001]. Vesely is therefore prior art at least under pre-AIA 35 U.S.C. § 102(a) and (e) because the earliest possible priority date of the 654 Patent is February 27, 2013.

## 2. Overview of Vesely

Vesely generally relates to “presenting a three-dimensional (3D) collaboration integrating a graphics object and a subject into the same view of a 3D scene.” Vesely, [0010]. More specifically, Vesely discloses interacting with objects in real time through virtual space by hand or another input device, and displaying this 3D scene comprising a participant and objects to another participant. See Vesely, [0041], [0044], [0048], [0010]; Ex.1003, ¶47.

Vesely describes the determination of the user’s viewpoint or eyepoint to render a 3D scene with the subject and objects together which are within the user’s viewpoint. See Vesely, [0025]–[0026], [0039]–[0040]. Users can interact with the virtual objects within the 3D scene through user control devices, such as their own hands. Vesely, [0041]. This interaction with virtual objects includes manipulating or controlling the object in a variety of ways. Vesely, [0041], [0044]. Furthermore, as the user’s eyes move and their eyepoint shifts, the 3D scene can be changed, along with the objects therein. Vesely, [0047]. Moreover, as explained by Dr. Houh, Vesely provides a level of detail sufficient to have enabled a person of ordinary skill in the art to have practiced the techniques described therein and to create and deploy the created 3-D virtual system and manipulation techniques, without undue experimentation, as of at least the priority date of the challenged claims. Ex.1003, ¶48.

### C. The Combination of Rosenthal and Vesely

It would have been obvious to include Vesely's techniques related to interacting with virtual objects in a three-dimensional scene in real time and the use of input devices or a hand to manipulate the objects in the method and system of Rosenthal for the reasons explained below. In particular, Rosenthal already discloses “the ability to track one or more defined sets of three-dimensional coordinates associated with image content within a video stream” and a “virtual camera allow[ing] virtual 3-D objects to be inserted within the video stream” in real time. Rosenthal, [0006]. Specifically, Rosenthal provides the spatial alignment and scene-compositing infrastructure necessary to create and manipulate dynamic video streams within a multi-dimensional space. Rosenthal, [0040]–[0048]; Ex.1003, ¶49.

Dr. Houh explains that manipulating and interacting with virtual objects in a collaborative multi-dimensional space would merely be an additional step in the process disclosed by Rosenthal. Rosenthal, [0084]–[0099] (describing a process of synchronizing video streams from users containing virtual objects); Ex.1003, ¶50. Similar to Rosenthal, Vesely discloses “a three-dimensional (3D) collaboration integrating a graphics object and a subject into the same view of a 3D scene.” Vesely, [0010]. Within this 3D scene, “digital object content” and “images of other collaborating users” may be captured through the use of “tracking sensors” and displayed within the scene. Vesely, [0010]–[0011]. Vesely expressly states that these

virtual objects in the 3D scene may be interacted with or manipulated through various methods, and the scene updates in the displays according to the interaction.

Vesely, [0041]–[0044], [0047]–[0048]; Ex.1003, ¶50.

Dr. Houh further explains that a POSITA would be motivated to modify the teachings of Rosenthal to include the interaction and manipulation of the 3D scene and the objects therein as taught by Vesely for several reasons. Ex.1003, ¶51. As explained, Rosenthal already describes the ability to track three-dimensional coordinates and insert virtual objects within a video stream. Rosenthal also provides teaching of synchronizing video streams for interaction between users on separate systems. Vesely expressly discloses the tools needed to interact and manipulate virtual objects and further discloses the ability to synchronize 3D scenes into a shared virtual space for collaboration between users. Vesely, [0010]–[0011], [0041]. Thus, as described by Dr. Houh, a POSITA would be motivated to modify the teachings of Rosenthal to include the interaction and manipulation taught in Vesely as well as the 3D collaboration taught therein because developers sought to improve remote interaction and collaboration for real-time, 3D, virtual interfacing in a variety of industries and for a variety of purposes. Ex.1003, ¶51; Rosenthal, [0005], [0014]; Vesely, [0009], [0052]; Ex.1001, 1:25-28, 1:37-45. For example:

In one embodiment, the first 3D collaboration system may receive input (e.g., from a user) to specify a perspective

view....This viewpoint may allow the user of the first system to see digital object content 180, such as one or more projected objects, as the other user may see it. Such a viewpoint may be useful in a collaborative environment where two or more users participate in a video conference. In the same shared workspace, i.e., 3D scene, users may be able to have a face-to-face conversation while one or more users describe, manipulate, or alter some digital object content 180, with the other users being able to see those manipulations from the manipulating user's viewpoint.

Vesely, [0081].

The 654 Patent itself confirms that a POSITA would recognize the need and advantages for a system such as Rosenthal to have real-time interaction capabilities within a 3D scene. In the background section of the 654 Patent, it explains that the then existing augmented reality systems only support a single user interacting with virtual objects, and users are unable to share video streams or augmented objects with others. Ex.1001, 1:22-36. It further explains that current video chat systems do not support the synchronization of augmented objects. Ex.1001, 1:37-45. Thus, the 654 Patent recognized that there existed a need for multiple users to collaborate with one another and interact with objects in real-time across video chat systems. *See* Ex.1001, 2:35-43; Ex.1003, ¶52.

Both Rosenthal and Vesely are analogous art to the 654 Patent for at least two reasons. *First*, Rosenthal and Vesely are both directed towards the same field of endeavor of the 654 Patent—collaborative augmented reality. Ex.1001, 2:5-20; Rosenthal, Abstract, [0006]; Vesely, Abstract, [0010]; Ex.1003, ¶53.

*Second*, both Rosenthal and Vesely are reasonably pertinent to the problem described by the 654 Patent: the need for augmented reality systems that “allow real time manipulation of objects within multiple dynamic and augmented video streams/feeds synchronized in an augmented reality (AR) or multi-dimensional space.” Ex.1001, 2:5-9; Ex.1003, ¶54.

For example, Rosenthal discloses the insertion of virtual objects into a shared video stream. For instance, “the virtual camera control software allows **both chatters** in a video conference to see **virtual objects embedded within the video.**” Rosenthal, [0012].<sup>1</sup> Rosenthal further discloses that the system supporting “a bi-directional data interface may support a wide range of interactive capabilities that **allow injected content and characteristics** to be controllably passed across virtual camera video streams.” Rosenthal, [0082]; Ex.1003, ¶55.

Rosenthal also discloses a virtual camera that “support[s] the ability to track one or more defined sets of three-dimensional coordinates associated with image

---

<sup>1</sup> Emphasis added throughout when bolded.

content within a video stream and may allow virtual 3-D objects to be inserted within the video stream, on a real-time basis, based upon the tracked sets of 3-D coordinates.” Rosenthal, [0039]; Ex.1003, ¶56.

Similarly, Vesely discloses that it allows for a “realistic interactive video conferencing system between two 3D collaboration systems.” Vesely, [0076]. “[E]ach user may see each other user and each user may be able to **interact with one or more projected objects** in the same shared workspace” which “may include a 3D scene.” Vesely, [0077]; Ex.1003, ¶57.

Vesely also further discloses the ability to manipulate objects in real time within augmented video streams. For example, “users may be able to have a face-to-face conversation while **one or more users describe, manipulate, or alter some digital object content** 180, with the other users being able to see those manipulations from the manipulating user’s viewpoint.” Vesely, [0081]; *see also* Vesely, [0047] (“[T]he 3D scene generator may update the displayed 3D scene based on changes in the user’s eyepoint, manipulations via the user input devices, etc. Such changes may be performed dynamically, at run-time.”). Vesely discloses that the manipulation of these virtual objects may be physical or may be virtual, and “the actions of the manipulating device in the spatial space may apply to a virtual object in the graphic space.” Vesely, [0044]; Ex.1003, ¶58.

It would have been further obvious to combine Vesely's teachings of 3D collaboration with the ability to insert virtual objects and synchronize video streams of Rosenthal, as each utilizes means for tracking coordinates in order to inject virtual objects into a combined stream, and each discusses the ability to dynamically alter the video stream in real time with the virtual objects. See Rosenthal, [0019], [0036], [0039]; Vesely, [0047], [0054]; Ex.1003, ¶59.

In addition, a POSITA would have been motivated to combine Vesely with Rosenthal based on the teachings of Rosenthal. Dr. Houh notes that Rosenthal explains that its system and methods enable real-time manipulation of video streams shared between users including the insertion of 3D objects. Rosenthal, [0006]–[0007]. This manipulation “allow[s] new types of interactive video chat games,” enables efficient sharing of user interests or other details through visually displaying interactive hyperlinks in the video stream, provides for remote simultaneous control of computer systems, and may be applied in any number of industries including vehicle maintenance, astronomy, or medical surgeries. Rosenthal, [0070]–[0071], [0013]–[0014]. Likewise, Vesely explains that it “enable[s] real-time collaboration allowing computer-generated 3D simulations, between multiple users, that coexist in physical space and time with the users interacting with the projected objects.” Vesely, [0052]. Thus, Dr. Houh concludes, a POSITA would have recognized that Vesely's teachings related to collaborative interaction and manipulation of 3D

objects would be complementary to Rosenthal given their focus on 3D simulations and Vesely's explanation that the systems and methods disclosed "allow for a realistic interactive video conferencing system between two 3D collaboration systems." Vesely, [0076]; Ex.1003, ¶60.

Dr. Houh also explains that a POSITA would have also had a reasonable expectation of success in combining Rosenthal and Vesely because both utilize well-known components and techniques and could be seamlessly integrated. A combined Rosenthal/Vesely system would involve minimal changes to the existing Rosenthal system. Ex.1003, ¶61.

## **VIII. THE CHALLENGED CLAIMS ARE UNPATENTABLE**

### **A. Ground 1: Claims 1, 12–16, 21–23, 26–27, 29–30, 32–36, 43–45, 48–50, and 53 Are Obvious in View of Rosenthal**

#### **1. Independent Claim 32:**

##### **a. Preamble**

##### **i. [32.0] A method comprising:**

Rosenthal's invention "pertains to the real-time manipulation of video images" and describes methods and apparatus that "provide a hardware-independent virtual camera that may be seamlessly integrated with existing video camera and computer system equipment." Rosenthal, Abstract, [0002], [0006]. Particularly, Rosenthal explains that the virtual camera is capable of "track[ing] one or more defined sets of three-dimensional coordinates associated with image content within

a video stream,” and “the virtual camera allows virtual 3-D objects to be inserted within the video stream, on a real-time basis, based upon the tracked sets of 3-D coordinates.” Rosenthal, [0006]; Ex.1003, ¶62.

For example, Rosenthal describes methods and apparatus that “may be used to manipulate any sort of incoming video signal regardless of the source of the video.” Rosenthal, [0007]. Referring to figures, Rosenthal illustrates that “FIGS. 14 and 15 are a flow diagram that present an exemplary process that may be performed by a virtual camera embodiment that supports a bi-directional data interface, as described above with respect to FIG. 8-13.” Rosenthal, [0082]; Ex.1003, ¶63.

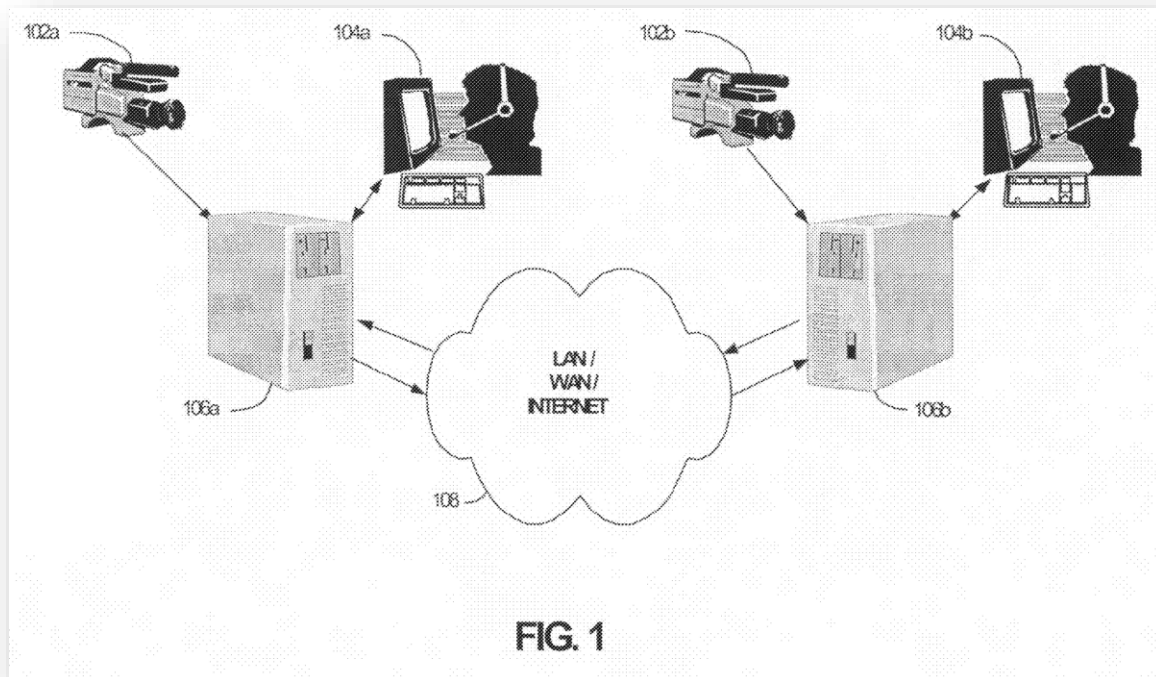
Accordingly, Rosenthal discloses “[a] *method*.” Ex.1003, ¶64.

**b. Real-world 3D Environment**

- i. [32.1] providing, using a video camera, a real-world video stream comprising video of a real-world 3D environment including a participant in a video chat or conference;

Rosenthal explains that “a physical camera may be connected to a computer [“*providing, using a video camera*”] and used to capture a stream of video images associated with a video conference to a video storage buffer [“*a real-world video stream comprising video*.”].” Rosenthal, [0008]. Rosenthal further explains that the “*video camera*” may include a “seamlessly integrated virtual camera” that “supports the ability” to provide the “*real-world video stream*” techniques of Rosenthal with

“three-dimensional” functionality. Rosenthal, [0006]. Rosenthal states that “[e]xemplary applications” of his invention “may include **real-time manipulation of a video stream generated by a video camera** associated with, for example, **a real-time video conference....**” Rosenthal, [0007]; Ex.1003, ¶65. The “*real-world video stream*” conferencing system is generally depicted in Figure 1:

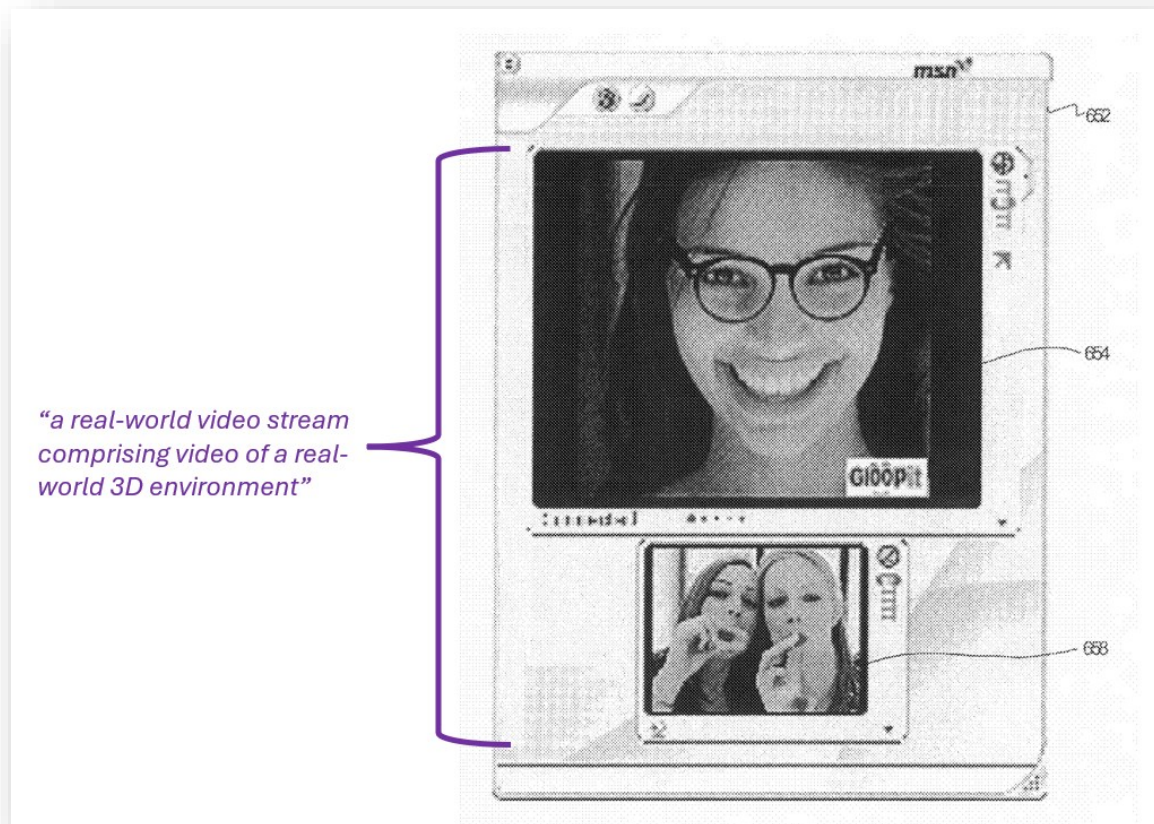


Rosenthal also explains that “*real-world video stream*” captured by the “*video camera*” and the associated virtual camera functionality permits “track[ing] one or more **defined sets of three-dimensional coordinates associated with image content within a video stream,**” and “the virtual camera allows **virtual 3-D objects** to be inserted within the video stream, on a real-time basis, based upon the

tracked sets of 3-D coordinates,” [*“a real-world 3D environment”*]. Rosenthal, [0006]; Ex.1003, ¶66.

Rosenthal discloses that the “*real-world 3D environment*” includes the “*participant in a video chat or conference.*” For example, Rosenthal explains that “[t]he virtual camera controller may send the resulting stream of video data to a virtual camera device defined within the operating system of a computer system that may be accessed by user applications such as ICQ Messenger, Skype, ICQ, AOL Messenger, Yahoo Messenger, or any other software capable of connecting to a physical video camera or other video input stream.” Rosenthal, [0011]. As Dr. Houh explains, applications like “ICQ Messenger, Skype, ICQ, AOL Messenger, Yahoo Messenger” were all well-known video chat and/or conferencing applications that display real-world video streams including the chat participants in a real-world environment. Ex.1003, ¶67 (citing Ex.1008, Scott Hanselman, *2006 is the Year of Video Chat*, HANSELMAN (Jan. 7, 2006), <https://www.hanselman.com/blog/2006-is-the-year-of-video-chat> and Ex.1009, Jonathan Roubini, *Yahoo! Messenger*, PC MAG UK (Jan. 18, 2006), <https://uk.pcmag.com/smb/28728/yahoo-messenger>); *see also* Rosenthal, [0012] (explaining that a user can use Rosenthal’s virtual camera functionality in Messenger to “send to his or her friends a video of **his or her face** including inserted virtual objects”). Ex.1003, ¶67.

Referencing Figure 6 (below), Rosenthal explains that it depicts “an exemplary user screen 652 presented by an exemplary video telephone user application (e.g., Microsoft Messenger) that includes a remote party video display 654 received from a remote party (e.g., via an Internet connection) as well as a local display 658 that displays a copy of the image sent to the remote party.” Rosenthal, [0038].



Rosenthal, Figure 6 (annotated). As shown above, Figure 6 illustrates the display of real-world video stream(s) including chat participants with a real-world environment. Ex.1003, ¶68.

Accordingly, Rosenthal discloses “*providing, using a video camera, a real-world video stream comprising video of a real-world 3D environment including a participant in a video chat or conference.*” Ex.1003, ¶69.

**c. Augmented Video Stream**

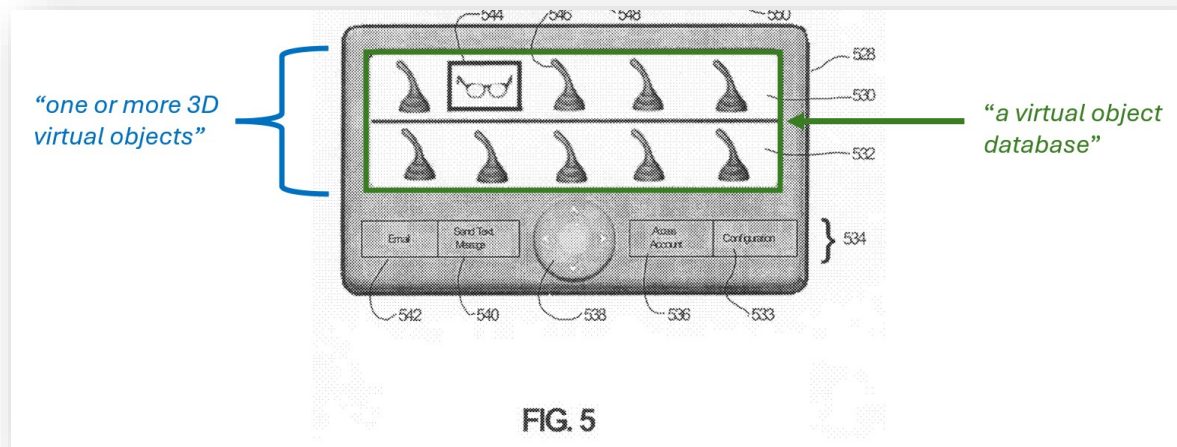
- i. [32.2] receiving, from another participant in the video chat or conference, an augmented video stream comprising augmentation data representing one or more 3D virtual objects retrieved from a virtual object database;

Rosenthal discloses that video chat participants can insert “*one or more 3D virtual objects*” into a real-world video stream by using its virtual camera functionality. Rosenthal, Abstract (“[T]he virtual camera supports the ability to track a defined set of three-dimensional coordinates within a video stream and to **dynamically insert rendered 3-D objects within the video stream on a real-time basis.**”). The virtual camera functionality in Rosenthal includes a “3-D engine” that provides the following functionality:

The 3D engine may **insert** virtual 3-D video objects [*“one or more 3D virtual objects”*] into the video stream [*“an augmented video stream comprising augmentation data representing one or more 3D virtual objects”*] based upon the received facial features data matrix and a list of virtual objects identified by a user for insertion into the data stream. These virtual objects may be selected from a gallery tool bar [*“retrieved from a virtual object*

*database*”] and may include, for example, a virtual mouth, glasses, hat or more specifically small 3-D objects.

Rosenthal, [0010]. Further, referencing Figure 5, Rosenthal describes a “[v]irtual camera control user interface” [“*a virtual object database*”<sup>2</sup>] “that allows a user” (“*another participant in a video chat or conference*”) to choose “one or more virtual items (e.g., glasses 544 and hat 546)” (“*one or more 3D virtual objects*”). Rosenthal, [0036]; Ex.1003, ¶70.



Rosenthal, Figure 5 (annotated).

<sup>2</sup> Rosenthal explains that “[t]he described hardware independent virtual camera and real-time video coordinate tracking and content insertion device may accommodate any quantity and any type of data set files and/or databases or other structures containing stored data sets, measured data sets and/or residual data sets in any desired format (e.g., ASCII, plain text, any word processor or other application format, etc.).” Rosenthal, [0111].

In addition, Dr. Houh explains that a POSITA would have understood that Rosenthal's selectable 3-D virtual objects (e.g., glasses and a hat)—such as those depicted in the controls of Figure 5—would be held electronically in a traditional “*database*” underlying the graphical user interface. Ex.1003, ¶71. Indeed, Rosenthal discloses that the “[c]ontrol software, or firmware, for the described hardware independent virtual camera and real-time video coordinate tracking and content insertion approach may be implemented in any desired computer language, and may be developed by one of ordinary skill in the computer and/or programming arts based on the functional description contained herein and illustrated in the drawings.” Rosenthal, [0108]. For example, Rosenthal explains that “[t]he various modules and data sets may be **stored in any quantity or types of file, data or database structures.**” Rosenthal, [0108]; Ex.1003, ¶71.

Rosenthal also explains that the 3-D engine may “send back a buffer to the virtual camera control software with the original image augmented with the objects [*“an augmented video stream comprising augmentation data representing one or more 3D virtual objects retrieved from a virtual object database”*] that rotate and move with the user face in 3-D.” Rosenthal, [0011]. Specifically, Rosenthal states that the virtual object is **encoded in a video stream** (*“augmented video stream”*) which is then transmitted from one computer to another. Rosenthal, [0069] (“[E]ncoding/decoding engine may encode information related to the virtual

object...into the individual frames of a **generated virtual camera video stream**...[which is] transmitted by the virtual camera to another virtual camera via a data communication.”); [0073] (“Decoding module 1004 may be used to retrieve information **encoded within a virtual camera video stream** [*“augmented video stream”*] and may pass the retrieved information to virtual camera controller 816. Virtual camera controller 816 may use the information to insert corresponding injected content into a generated virtual camera video stream.”). Ex.1003, ¶72.

Rosenthal also discloses “a direct data communication channel to pass data between virtual camera embodiments executing on different communication systems connected via a network,<sup>3</sup>” (Rosenthal, [0066]), which permits a “user operating a first virtual camera executing on a first computer system 801a **[to] send a virtual object**, such as a ball, cream pie, etc., over a network connection 830” to a second computer system—i.e., “*receiving, from another participant in the video chat or conference, an augmented video stream.*” Rosenthal, [0068]; *see also id.*, [0070] (“**[A] local user may send, in response to something said by the remote user**

---

<sup>3</sup> Rosenthal also refers to this direct communication path as a “bi-directional data interface,” which Rosenthal explains allows the system to “support a wide range of interactive capabilities that allow injected content and characteristics to be controllably passed across virtual camera video streams in support of interactive gaming.” Rosenthal, [0013]; *see also id.*, [0056]–[0057]. Dr. Houh explains that this further demonstrates the transmission of virtual camera output (including virtual objects) between video conference participants. Ex.1003, ¶73.

during a chat session, an electronic cream pie that may be inserted into the virtual video stream produced by the remote user, sticks to the remote user's face and slowly slides down. In a similar manner, other virtual objects may be exchanged between both user's virtual worlds, thus allowing new types of interactive video chat games."); Ex.1003, ¶73.

Rosenthal provides detail as to the generation of an "*augmented video stream*" from "*another participant*":

[An] encoding/decoding engine may **encode information related to the virtual object...into the individual frames of a generated virtual camera video stream** and...may encode information related to the virtual object **within a data stream transmitted by the virtual camera to another virtual camera** via a data communication channel, e.g., using TCP/IP or other communication protocol.

Rosenthal, [0069]. This "virtual object may [then] be included **within the virtual camera video stream produced by the second virtual camera** executing on the second computer system...." Rosenthal, [0070]; Ex.1003, ¶74.

Accordingly, Rosenthal discloses "*receiving, from another participant in the video chat or conference, an augmented video stream comprising augmentation data representing one or more 3D virtual objects.*" Ex.1003, ¶75.

**d. Multiplexing the Augmented Video Stream**

- i. [32.3] multiplexing the augmented video stream received from the other participant with the real-world video stream of the real-world 3D environment in real time to generate a multiplexed video stream comprising the video of the real-world 3D environment and the one or more 3D virtual objects represented by the augmentation data received from the other participant

As explained above in [32.2], Rosenthal discloses that the “*participant’s*” computer “*receiv[es], from another participant in the video chat or conference, an augmented video stream comprising augmentation data representing one or more 3D virtual objects.*”

Rosenthal further demonstrates that the “*real-world video stream of the real-world 3D environment*” captured by a physical camera of the “*participant*” is combined (or, “*multiplexed*”) with “*one or more 3D virtual objects*” received from the “*another/other participant,*” thereby producing a single composite video stream (“*a multiplexed video stream.*”). Ex.1003, ¶77.

Rosenthal describes virtual camera software functionality that supports a bi-directional data interface. Rosenthal, [0082]. In particular, Rosenthal explains that at “the virtual camera controller determines that data has been received on a bi-directional communication channel”—i.e., “*the one or more 3D virtual objects represented by the augmentation data received from the other participant,*” it proceeds to “decode” this information. Rosenthal, [0088]–[0089]; *see also*

Rosenthal, [0073] (the “decoding module 1004 may be **used to retrieve information encoded within a virtual camera video stream** and may pass the retrieved information to virtual camera 816. Virtual camera controller 816 may **use the information to insert corresponding injected content into a generated virtual camera video stream.**”). Ex.1003, ¶78.

Rosenthal explains that if “one or more injected content features has been selected for insertion within the physical video stream,” the selected feature is “injected” (or inserted) into the real-world video stream. *See* Rosenthal, Fig. 14, [0090]–[0094]; *see also* Rosenthal, [0080] (“Decoded information may be passed to the virtual camera controller for use in controlling injected 3-D content and/or text inserted into a locally generated virtual camera video stream.”). Ex.1003, ¶79.

Rosenthal further confirms this operation in the context of interactive video chat:

For example, using such a feature, a local user may send, in response to something said by the remote user during a chat session, **an electronic cream pie that may be inserted into the virtual video stream produced by the remote user**, sticks to the remote user’s face and slowly slides down. In a similar manner, **other virtual objects may be exchanged between both user’s virtual worlds**, thus allowing new types of interactive video chat games. For example, a local user could throw, e.g., head, a virtual

soccer ball to a remote user and the remote user may either head the ball back, or if the ball is missed, a goal may be scored by the local user.

Rosenthal, [0070]; Ex.1003, ¶80.

Rosenthal also explains that the virtual camera controller generates a single “virtual camera video stream by combining a physical camera video stream with injected content.” Rosenthal, [0033]–[0035], [0080]. Thus, Dr. Houh confirms that Rosenthal’s insertion of remotely received virtual objects into local video streams results in a single combined video stream (“*a multiplexed video stream*”). Ex.1003, ¶81.

Rosenthal further confirms that “*multiplexing the augmented video stream*” occurs in “*real-time*,” as the virtual camera controller copies physical camera frames from a video buffer and inserts injected content into those frames as they are processed, producing the composite video stream on a frame-by-frame basis during live video chat. Rosenthal, [0082] (“As described above...a virtual camera may support the ability to track one or more defined sets of three-dimensional coordinates associated with image content within a video stream and may allow virtual 3-D objects to be inserted within the video stream, **on a real-time basis**, based upon the tracked sets of 3-D coordinates.”), [0035] (“The virtual camera controller may copy

frames from a physical camera video buffer and **modify those frames in real time...**"); *see also* Rosenthal, [0088]–[0094]; Ex.1003, ¶82.

Accordingly, Rosenthal discloses “*multiplexing the augmented video stream received from the other participant with the real-world video stream of the real-world 3D environment in real time to generate a multiplexed video stream comprising the video of the real-world 3D environment and the one or more 3D virtual objects represented by the augmentation data received from the other participant.*” Ex.1003, ¶83.

**e. Extracting 3D Virtual Objects**

- i.** [32.4] extracting at least one of the 3D virtual objects from the multiplexed video stream,

Initially, the 654 Patent explains the following with respect to the role of “*extraction*”:

As the connection occurs, multiple streams sync module 106 multiplexes the augmented video stream of the participant with the initial or augmented video stream from the other participant and **allows real time extraction of the augmented object(s) from the multiplexed video stream, manipulates the extracted object(s), and synchronize them back into the synchronized augmented video stream according to new X,Y, Z,T coordinates dynamically modified in real time.** Once synchronized, the X,Y, Z and T coordinates of the object

in the 3D or multi-dimensional space of the synchronized augmented video streams **are shared/displayed by the multiple streams sync module 106 to each of the participants, who will be able to see, interact with, and manipulate the objects** according to the interaction and manipulation of the X, Y, Z and T coordinates of the objects in real time.

Ex.1001, 4:42-57; Ex.1003, ¶¶84–85. Thus, as Dr. Houh explains, the 654 Patent is describing the “*extract[ion]*” of a virtual object as a continuous process (as the “connection occurs”) required for “dynamically modifying” in “real time” the coordinates of the virtual object within the multi-dimensional space of the synchronized augmented video stream. Ex.1001, 4:47-57; Ex.1003, ¶86. Both multiplexing and extraction are functionally described as contemporaneous, real-time capabilities of the multiple streams sync module that both occur “as the connection occurs.” Ex.1003, ¶87.

Rosenthal discloses this technique. For example, Rosenthal techniques “support a bi-directional data interface” that “support a wide range of interactive capabilities that allow injected content and characteristics [*“3D virtual objects”*] to be controllably passed across virtual camera video streams in support of **interactive gaming.**” Rosenthal, [0013]. Rosenthal further explains that “**via the interactive exchange of information**, a user operating a first virtual camera executing on a first

computer system 801a, **may send a virtual object, such as a ball, cream pie, [“3D virtual objects”] etc., over a network connection 830**, and a second virtual camera executing on a second computer system 801b may be informed of the direction, speed and nature of the object.” Rosenthal, [0068]. Using such an “interactive” feature, Rosenthal explains the system permits the following:

For example, a local user could throw, e.g., head, **a virtual soccer ball to a remote user and the remote user may either head the ball back**, or if the ball is missed, a goal may be scored by the local user.

Rosenthal, [0070]. Thus, like the 654 Patent, Rosenthal’s techniques allow the participants in the video stream to dynamically manipulate 3D objects in real-time. *Id.*; *see also id.*, [0082]; Ex.1003, ¶88.

The interactive features in Rosenthal are enabled through tracking functionality (including Rosenthal’s feature tracking and 3-D engine), which Dr. Houh explains “*extracts*” the “*3D virtual objects from the multiplexed video stream*” through its identification of virtual object specific data within shared virtual camera video streams so that virtual objects may be manipulated by the participants and re-rendered thereafter in real-time. Ex.1003, ¶89. As Rosenthal details:

[A] virtual camera may support the ability **to track one or more defined sets of three-dimensional coordinates associated with image content within a video stream**

**and may allow virtual 3-D objects to be inserted within the video stream, on a real-time basis, based upon the tracked sets of 3-D coordinates.** As described above with respect to FIGS. 8-13, virtual camera embodiments that support a bi-directional data interface may **support a wide range of interactive capabilities that allow injected content and characteristics to be controllably passed** across virtual camera video streams in support of interactive gaming.

Rosenthal, [0082].

Dr. Houh explains that this disclosure (as well as other disclosures in Rosenthal described herein) demonstrate that Rosenthal's tracking functionality permits "*extract[ion]*" through its identification of 3-D coordinates associated with virtual objects. Ex.1003, ¶90 (citing Rosenthal, [0060], [0054]). Indeed, Rosenthal explains that "[f]eature tracking engine 818 may scan a video frame image for visual features related to a set of selected injected content" and "may return to virtual camera controller 816 **a matrix of coordinates** that may include coordinates within the video image frame for use in generating and inserting content into video frame images." Rosenthal, [0054]. Ex.1003, ¶90.

Rosenthal's 3-D engine uses this identification (of coordinates associated with virtual objects) to enable real-time manipulation of the coordinates of those virtual objects within the shared virtual camera video streams. Ex.1003, ¶90. For example,

Rosenthal explains that “the 3-D engine may insert virtual 3-D video objects into the video stream based upon the received facial features data matrix and a list of virtual objects identified by a user for insertion into the data stream.... The 3-D engine may send back a buffer to the virtual camera control software with the original image augmented with the objects that rotate and move with the user face in 3-D.”

Rosenthal, [0010]–[0011]; *see also id.*, [0054], [0066]–[0070]. Ex.1003, ¶¶90–91.

Dr. Houh also explains how Rosenthal’s decoding engine discloses “*extract[ion]*” through its identification of virtual object specific data within shared virtual camera video streams so that virtual objects may be manipulated by the participants and re-rendered thereafter in real-time. Ex.1003, ¶92. For example, Rosenthal explains that:

[E]ncoding/decoding engine 817 may encode information related to the virtual object, e.g., object type, speed, direction, etc., into the individual frames of a generated virtual camera video stream” and this “information related to the virtual object” may then be encoded “within a data stream transmitted by the virtual camera to another virtual camera via a data communication channel.

Rosenthal, [0069]. Rosenthal then discloses that on the receiving side, “[d]ecoding may be a continuous procedure that may detect and decode information from the respective video frames images” which can be “use[d] in controlling injected 3-D

content and/or text inserted into a local generated virtual camera video stream.”

Ex.1003, ¶92, (citing Rosenthal, [0075]).

Accordingly, Rosenthal teaches or suggests “*extracting at least one of the 3D virtual objects from the multiplexed video stream.*” Ex.1003, ¶93.

**f. Translating Coordinates**

- i. [32.5] based on an action of the participant, translating coordinates of the extracted virtual object in at least three dimensions within a multi-dimensional space of the multiplexed video stream, wherein the three dimensions of the multi-dimensional space are defined by the video of the real-world 3D environment; and

Initially, Rosenthal’s “field of invention” is explicitly characterized as pertaining to “the real-time **manipulation** of video images.” Rosenthal, [0003]. Rosenthal’s tracking functionality provides this ability. Rosenthal explains that its system anchors virtual objects (or injected content) to the tracked coordinates of visual features (e.g., a user’s face, nose, or lips) in the real-world 3D environment captured by the physical camera. Rosenthal, [0055] (“Based on the features selected for insertion within the video...and the received feature coordinates matrix data, 3-D engine 820 may generate appropriate 3-D views of the selected injected content and may insert the generated content at an appropriate location within the respective video frame images received from physical camera video buffer 810.”). Ex.1003, ¶95.

As Dr. Houh explains, a POSITA would have understood that this tracking functionality and dynamic manipulation of the 3D objects by participants demonstrates **translating coordinates of the extracted virtual object in three dimensions within a multi-dimensional space of the multiplexed video stream based on an action of the participant**. Ex.1003, ¶99. By anchoring virtual objects to visual features in the video, Rosenthal's system allows "the [virtual] objects [to, for example] rotate and move with the user face in 3-D" in the video stream. Rosenthal, [0011]; Ex.1003, ¶96.

As another example, Rosenthal discloses that its system "update[s] the generated set of three-dimensional coordinates based on changes in image content [e.g., a user his or her face]...thereby tracking a position of the one or more located features," that the "three-dimensional image of the selected injected content feature" ("*extracted virtual object*") is likewise "update[ed]...based on the updated set of three-dimensional coordinates," ("*translating coordinates of the virtual object in at least three dimensions*") and the "updated three-dimension image" is thereafter "insert[ed]...into the next received video frame image" ("*translating coordinates...within the multi-dimensional space of the multiplexed video stream*"). Rosenthal, cl. 1, 2. Ex.1003, ¶96.

Further, Rosenthal also discloses this limitation through its audio-based object manipulation feature. Ex.1003, ¶100. Rosenthal discloses an "audio tracking module

904” is adapted to “scan an audio track associated with an incoming video stream for any number of phrases.” Rosenthal, [0062]. “[E]ach phase [*sic*] may be associated with a predetermined action that results in an insertion of injected content into the video stream.” Rosenthal, [0062]. For example, “on detecting a certain phrase, such as ‘I love you’ audio track feature tracking module 904 may add, within the matrix of feature coordinates, data that results in the insertion of injected features such as hearts beaming from an individuals eyes.” Rosenthal, [0061]; Ex.1003, ¶100.

Accordingly, when the participant acts (e.g., moves/rotates/translates their head/face, in the captured scene or speaks certain phrases), Rosenthal’s system **translates the coordinates** in real-time to place/render the injected object **in three dimensions (X, Y, Z)** within the video stream’s coordinate space defined by the real-world scene content captured by the camera. Rosenthal, [0011]; Ex.1003, ¶96. Rosenthal describes, for example, that a “a local user could throw, e.g., head, a virtual soccer ball to a remote user and the remote user may either head the ball back, or if the ball is missed, a goal may be scored by the local user.” Rosenthal, [0070]; Ex.1003, ¶97.

Accordingly, Rosenthal teaches or suggests “*based on an action of the participant, translating coordinates of the extracted virtual object in at least three dimensions within a multi-dimensional space of the multiplexed video stream,*

*wherein the three dimensions of the multi-dimensional space are defined by the video of the real-world 3D environment.”* Ex.1003, ¶102.

**g. Displaying Multiplexed Stream and Each Participant Translating Coordinates**

- i.** [32.6] causing a display device associated with the participant to display the multiplexed video stream comprising the video of the real-world 3D environment and the one or more 3D virtual objects, wherein both the participant and the other participant can translate the coordinates of the one or more 3D virtual objects within the three dimensions of the multi-dimensional space.

As described in limitation [32.3], Rosenthal discloses generating a “*multiplexed video stream comprising the video of the real-world 3D environment and the one or more 3D virtual objects.*” Rosenthal discloses that the generated “*multiplexed video stream*” is displayed on a “*display device associated with the participant.*” For example, Rosenthal explains that “[t]he locally produced virtual camera video stream may be sent by a locally executing user application 812 **to the local user’s display**...and may be sent over network 830 to a user application executing on a remote computer which may **display the virtual camera image stream to a remote user via a display.**” Rosenthal, [0069]; *see also* [0012], [0034]. Thus, as Dr. Houh explains, a POSITA would understand Rosenthal discloses both participants’ display devices simultaneously displaying virtual objects added to the real-world video stream of both participants within their respective real-world

environments—such as displaying video showing the first “*participant*” heading a soccer ball to “*another*” or the “*other*” “*participant*”. Ex.1003, ¶103 (citing Rosenthal, [0070]).

Indeed, Rosenthal describes a “bi-directional data interface” that supports the ability to “controllably pass[] across virtual camera video streams” the virtual objects embedded within each video stream, enabling “interactive gaming.” Rosenthal, [0082]. As Dr. Houh explains, this form of interactive gaming is made possible through the local user (“*the participant*”) seeing both the video stream from the remote party, and their own video stream (“*the real-world 3D environment*”). Ex.1003, ¶104; *see also* Rosenthal, [0038] (“FIG. 6 is an exemplary user screen...that includes a **remote party video display** 654 received from a remote party...**as well as a local display** 658 that displays a copy of the image sent to the remote party.”). As an example illustrating the simultaneous display and control of “*one or more virtual objects*” in the “*multiplexed stream*” by both “*the participant*” and the “*other participant*,” Rosenthal explains that “a **local user could throw, e.g., head, a virtual soccer ball** to a remote user and the **remote user may either head the ball back**, or if the ball is missed, a goal may be scored by the local user.” Rosenthal, [0070]; Ex.1003, ¶104.

Rosenthal disclosure also demonstrates the ability for “*both the participant and the other participant*” to “*translate the coordinates of the one or more 3D virtual objects within the three dimensions of the multi-dimensional space.*” Ex.1003, ¶105.

As explained above in [32.4] and [32.5], the interactive features in Rosenthal are enabled through tracking functionality (including Rosenthal’s feature tracking and 3-D engine), which Dr. Houh explains permits the “*translation*” of three-dimensional coordinates in real-time so that they may be **dynamically manipulated and synchronized** between participants. Ex.1003, ¶106. As Rosenthal details:

[A] virtual camera may support the ability **to track one or more defined sets of three-dimensional coordinates associated with image content within a video stream and may allow virtual 3-D objects to be inserted within the video stream, on a real-time basis, based upon the tracked sets of 3-D coordinates.** As described above with respect to FIGS. 8-13, virtual camera embodiments that support a bi-directional data interface may **support a wide range of interactive capabilities that allow injected content and characteristics to be controllably passed** across virtual camera video streams in support of interactive gaming.

Rosenthal, [0082].

This ability for participants to translate the coordinates of virtual objects injected into the real world display of other participants is provided through Rosenthal's "bi-directional interface." Rosenthal discloses that the "bi-directional interface" "support[s] the **simultaneous control** and operation of local hardware devices by computer systems at remote locations...based on the relative position of objects dynamically tracked features." Rosenthal, [0082]; *see also id.*, [0065] ("Such an exchange of information may be used to pass peer-to-peer information for any purpose, such as, for example, to pass control information that implements interactivity within an Interactive Video Space (IVS)."); [0070] (as detailed above, describing interactive gaming with a virtual soccer ball that permits both participants in the video to manipulate the soccer ball). Ex.1003, ¶107.

Accordingly, Rosenthal teaches or suggests "*causing a display device associated with the participant to display the multiplexed video stream comprising the video of the real-world 3D environment and the one or more 3D virtual objects, wherein both the participant and the other participant can translate the coordinates of the one or more 3D virtual objects within the three dimensions of the multi-dimensional space.*" Ex.1003, ¶108.

**2. Dependent Claims 33–36, 43–45, 48–50, 53****a. Claim 33**

- i. [33] The method of claim 32, wherein multiplexing the augmented video stream with the real-world video stream comprises synchronizing the augmented video stream with the real-world video stream in both time (T) and three-dimensional coordinates (X, Y, Z).

In summarizing the invention, Rosenthal describes, “[t]he virtual camera supports the ability to track one or more defined sets of three-dimensional coordinates associated with image content within a video stream.” Rosenthal, [0006]. It continues, “[f]urther, the virtual camera allows virtual 3-D objects to be inserted within the video stream, on a real-time basis, based upon the tracked sets of 3-D coordinates,” (Rosenthal, [0006]; *see also* [0039], [0082]), which Dr. Houh explains provides an example of the disclosure of synchronizing the time coordinate of the multiplexed video stream. Ex.1003, ¶109; *see also* Rosenthal, [0009] (“The first subprocess may be a human facial feature tracking process that is able to detect a position and generate 3-D (e.g., X, Y and Z) coordinates for several points associated with a feature of a user’s face positioned in front of the video camera.”).

Rosenthal further discloses the synchronization of the augmented and real-world video streams. “A single version of the virtual camera control software allows both chatters in a video conference to see virtual objects embedded within the video....” Rosenthal, [0012]. Another example discusses the local computer system

that “support[s] a bi-directional data interface with a virtual camera embodiment executing on a remote computer system....Further, such a bi-directional interface may also be used to support the simultaneous control an[d] operation of local hardware devices by computer systems at remote locations...based on the relative position of objects dynamically tracked features within virtual camera and/or physical camera video streams.” Rosenthal, [0013]; Ex.1003, ¶110.

Further, Dr. Houh explains that a POSITA would have understood that Rosenthal is describing multiplexing the augmented video stream with the real-world video stream comprises synchronizing the augmented video stream with the real-world video stream in time (T). For example, Rosenthal (just like the specification of the 654 Patent) describes that its invention allows the insertion of “virtual 3-D objects” to the video stream on a “real-time basis.” Rosenthal, [0082]; *see also* Ex.1001, 2:20-23; Ex.1003, ¶111.

Accordingly, Rosenthal teaches or suggests claim 33. Ex.1003, ¶112.

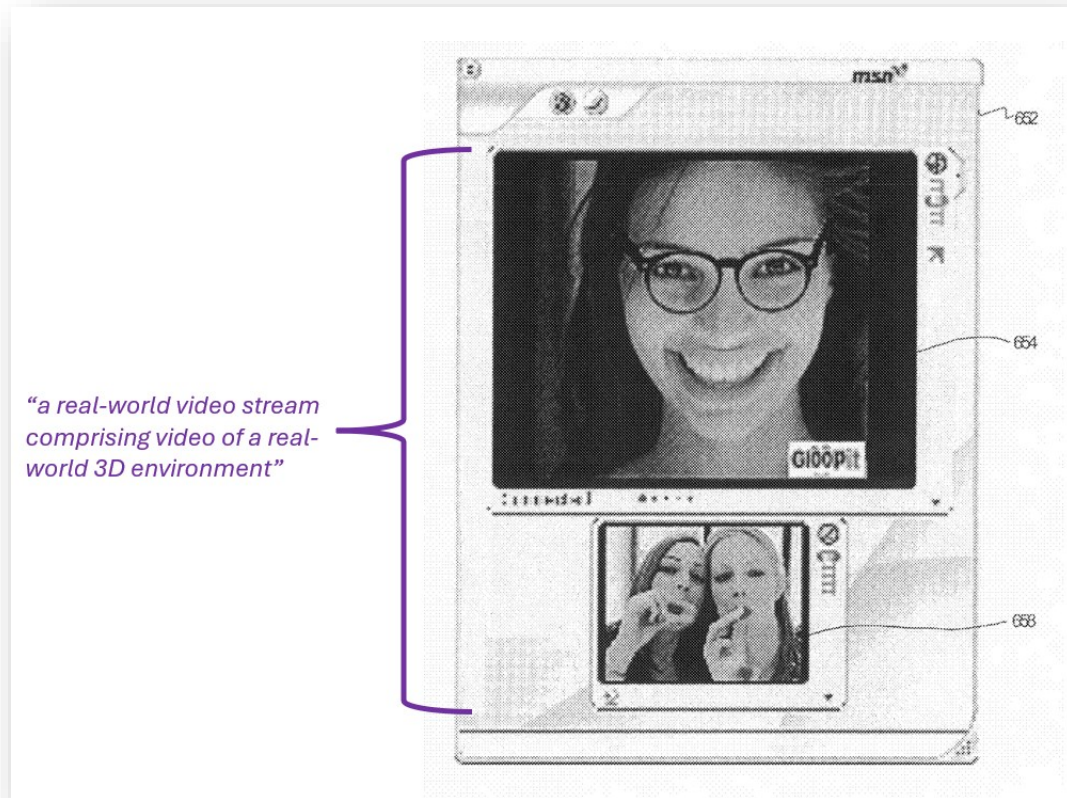
**b. Claims 34, 35, and 36**

- i. [34] The method of claim 32, wherein the real-world video stream comprises a first real-world video stream, and wherein the augmented video stream is synchronized with a second real-world video stream comprising video of a real-world environment including the other participant.
- ii. [35] The method of claim 32, wherein multiplexing the augmented video stream with the real-world video stream comprises overlaying the one or more virtual objects onto the video of the real-world environment.
- iii. [36] The method of claim 32, wherein multiplexing the augmented video stream with the real-world video stream comprises inserting the one or more virtual objects into the video of the real-world environment.

Rosenthal discloses this claim for the reasons set forth above in [32.3] and [32.6]. For example, annotated Figure 6 below depicts a “*real-world video stream*” which is described as “a local display 658 that displays a copy of the image sent to the remote party” and a second real-world video stream that includes the other participant which is described as “a remote party video display 654 received from a remote party,” which satisfies claim 34. Rosenthal, [0038]. Similarly, Rosenthal explains that “the described approach allows implementation of virtual cameras **as overlays** to existing physical camera and/or other video input devices...” which satisfies claim 35. Rosenthal, [0012]. And Rosenthal further explains that using this technique “the 3-D engine may **insert** virtual 3-D video objects into the video stream

based upon the received facial features data matrix and a list of virtual objects identified by a user for insertion into the data stream,” which satisfies claim 36.

Rosenthal, [0010]. Ex.1003, ¶113.



Rosenthal, Figure 6 (annotated).

Accordingly, Rosenthal teaches or suggests claims 34, 35, and 36. Ex.1003, ¶114.

**c. Claim 43**

- i. [43] The method of claim 32, wherein the multiplexing of the augmented video stream and the real-world video stream to generate the multiplexed video stream is performed based on active control on synchronization of HB (human behavior) of the participants in addition to X, Y, Z, and T coordinates.

Rosenthal discloses this claim for the reasons set forth above in [32.3], [32.6], and [33]. For example, “[t]he 3-D engine may send back a buffer to the virtual camera control software with the original image augmented with the objects that rotate and move with the user face in 3-D.” Rosenthal, [0011]. As the user moves (“*active control*”) their face (“*human behavior*”), the corresponding virtual objects follow and the “*multiplexed video stream*” is updated accordingly. Similarly, “special effect sounds” may be added to augment the voice of a user, and the 3-D engine “may synchronize the lips of a virtual mouth [(“*human behavior*”)] with the real user mouth based on the audio stream.” Rosenthal, [0011]. This occurs with both the local user and the remote user, and their video streams are simultaneously displayed. See Rosenthal, [0038] (“an output from a virtual camera, in use by a remote party, may be received and displayed in the same manner as an image received from a user transmitting output from a physical video camera; however, the image received from the virtual camera may include virtual items”). Ex.1003, ¶115.

Another example of Rosenthal disclosing this claim states “[v]ia the interactive exchange of information, a user operating a first virtual camera executing on a first computer system 801a, may send a virtual object...and a second virtual camera executing on a second computer system 801b may be informed of the direction, speed and nature of the object.” Rosenthal, [0068]. This could be done, for example, by the local user virtually throwing a cream pie, which “sticks to the remote user’s face and slowly slides down” or by heading a virtual soccer ball to the remote user “and the remote user may either head the ball back [(“*human behavior*”)], or if the ball is missed, a goal may be scored by the local user.” Rosenthal, [0070]. As Dr. Houh explains, the decision by a user (“*active control*”) to send a virtual object to the remote user by their physical action (“*human behavior*”) causes the multiplexing to occur and the response by the remote user “*synchroniz[es]*” the human behavior. Ex.1003, ¶116.

Accordingly, Rosenthal teaches or suggests claim 43. Ex.1003, ¶117.

**d. Claim 44**

- i.** [44] The method of claim 32, wherein the multiplexing of the augmented video stream and the real-world video stream to generate the multiplexed video stream comprises synchronizing the augmented and real-world video streams in such a way as to reduce latency of the multiplexed video stream.

As explained in [32.3] and [34]–[36], and for the reasons below, Rosenthal discloses this claim. The 654 Patent explains that “augmentation module 104 may engine update coordinates for the positions of the retrieved augmented objects in real-time according to changing tracking data and timing within each session of video chat or conference in order to reduce latency of the augmented video stream. Ex.1001, 4:21-26. Accordingly, as discussed in [32.3], and as explained by Dr. Houh, the insertion of virtual content and the multiplexing of the augmented video stream with the real-world video stream happens **in real-time**—doing so according to the 654 Patent “*reduce[s] latency.*” Ex.1003, ¶118.

In addition, Rosenthal states that “the encoded data included within a selected image format, e.g., **MPEG, or other image format....**” Dr. Houh explains that at the time of the 654 Patent, MPEG was a well-known standard for audio and video compression. Ex.1003, ¶119 (citing Ex.1010, Microsoft Computer Dictionary, definition of “MPEG”). Data compression technologies like MPEG work to “reduc[e] the amount of space and bandwidth needed to store or transmit a block of

data....” Ex.1003, ¶119 (citing Ex.1010, Microsoft Computer Dictionary, definition of “data compression”). Dr. Houh explains that a POSITA would understand data compression technologies like MPEG would reduce latency—“the time required for a signal to travel from one point to another”—in a video transmission system because it would reduce the amount of bandwidth (and therefore time) necessary to transmit video. Ex.1003, ¶119 (citing Ex.1010, Microsoft Computer Dictionary, definition of “latency”).

Accordingly, Rosenthal teaches or suggests claim 44. Ex.1003, ¶120.

**e. Claim 45**

- i.** [45] The method of claim 32, further comprising:  
broadcasting the multiplexed video stream to all  
participants in the video chat or conference.

For the same reasons demonstrated in [32.1], [32.2], and [32.6], Rosenthal teaches or suggests this claim. For example, “[a] single version of the virtual camera control software allows both chatters in a video conference to see virtual objects embedded within the video, thereby providing exposure of the product to a potential base of new users....” Rosenthal, [0012]; *see also* [0007] (“Exemplary applications may include real-time manipulation of a video stream generated by a video camera associated with, for example, a real-time video conference....”). Ex.1003, ¶121.

Another example of disclosure is the following:

For example, both application 812 executing on computer system 801a and application 812 executing on computer system 801b may be configured to present a virtual camera stream to a user via local display/keyboard/other hardware device 813, **transmit** the virtual camera stream to user application 812 executing on a remote computer system via network 830 and to receive a virtual camera stream from a remote computer system via network 830. In such a configuration, virtual camera 816 executing on each of the respective network connected computer systems may generate a virtual camera video stream that includes injected content based on selections made by a local user via display/keyboard/other hardware device 813.

Rosenthal, [0057]; Ex.1003, ¶122.

Accordingly, Rosenthal teaches or suggests claim 45. Ex.1003, ¶123.

**f. Claim 48**

- i.** [48] The method of claim 32, further comprising: controlling and manipulating the one or more virtual objects within the multiplexed video stream by using pupil tracking to track movement of the participant's eyes.

Through Rosenthal's description of the feature tracking engine, "image enhancement features" may be applied to a video stream, which includes changing the eye color of an individual. Rosenthal, [0059]. Within the feature tracking

engine, the “[f]rame image feature tracking module 902 may scan a video frame image to locate key pixel coordinates required to appropriately insert each selected injected content feature and may store the key pixel coordinates in a matrix of feature coordinates that may be passed by virtual camera controller 816 to 3-D engine 820.” Rosenthal, [0060]. This can include tracking and storing the coordinates associated with an individual’s eyes—then, as the individual’s eyes move, “[e]ach alteration of a video frame, such as correcting an individual’s eye color...may each require a set of image coordinates and/or other parameters that may be collected...to support the selected injected content.” Rosenthal, [0060]. Thus, as the individual’s color-altered eyes move, the feature tracking engine tracks the eyes and controls the selected injected content/virtual object (eye color) to follow the eyes (“*using pupil tracking to track movement of the participant’s eyes.*”). Rosenthal, [0059]–[0060]; Ex.1003, ¶124.

Patent Owner’s Preliminary Infringement Contentions accuse “Eye Tracking” technology that is used to “estimate the location and orientation of user’s eyes” as infringing this claim. Ex.1006, pg. 125. Similarly, as discussed above, Rosenthal discloses tracking the position and orientation of a user’s eyes throughout. *See, e.g.,* Rosenthal, [0009] (“human facial feature tracking process” can track coordinates for a user’s eyes), [0060] (“frame image feature tracking module 902 may obtain and

store coordinates associated with the top of an individual's head, eyes and ears").

Furthermore:

The facial feature tracking process may send facial feature matrix data to a 3-D feature visualization subprocess (or 3-D engine). The 3-D engine may receive the facial feature matrix data in near real-time, and may track the user facial features, head rotation and translation in the video based upon the received facial feature matrix data. Further, the 3-D engine may insert virtual 3-D video objects into the video stream **based upon the received facial features** data matrix and a list of virtual objects identified by a user for insertion into the data stream.

Rosenthal, [0010]. Thus, based at least in part on the facial feature tracking, virtual objects in the video stream may be controlled and manipulated by being inserted into the video stream and by being moved according to the matrix data. Ex.1003, ¶125.

However, the facial features discussed are not the only ones disclosed in Rosenthal. The "video feature tracking system" can be "configured to detect a wide range of features." "Such features are not necessarily limited to facial features." Rosenthal, [0014]; Ex.1003, ¶126.

Accordingly, Rosenthal teaches or suggests claim 48. Ex.1003, ¶127.

**g. Claim 49**

- i.** [49] The method of claim 32, further comprising: synchronizing the augmented and real-world video streams across multiple platforms so that the augmented video stream provided by the other participant from one platform is shared with the participant using a different platform.

For the same reasons demonstrated in [32.2], Rosenthal teaches or suggests this claim. In addition, Rosenthal states that an example includes, “[t]he virtual camera control software may be independent of device hardware and operating system software and thus fully compatible with all of them.” Rosenthal, [0012]. Another example states that “[t]he described hardware independent virtual camera and real-time video coordinate tracking and content insertion approach may be integrated within a stand-alone system or may execute separately and be coupled to any number of devices, computer systems, server computers or data storage devices via any communication medium....” Rosenthal, [0106]; *see also* [0103] (“The present invention is not limited to use within any specific network, but may be applied to any deployed network infrastructure that supports video based user applications.”). Throughout Rosenthal, it discusses hardware-independent (or platform-independent) embodiments of virtual cameras, capable of being integrated with any platform or equipment. *See, e.g.*, Rosenthal, [0006]. Ex.1003, ¶128.

Accordingly, Rosenthal teaches or suggests claim 49. Ex.1003, ¶129.

## h. Claim 50

- i. [50] The method of claim 32, further comprising: creating a chat room or conference session for the video chat or conference, inviting the participants to join the video chat or conference, and maintaining the session during the video chat or conference.

Rosenthal states that “[a] single version of the virtual camera control software allows **both chatters** in a video conference to see virtual objects embedded within the video....” Rosenthal, [0012]. As Dr. Houh explains, in order to facilitate the video chat or conference sessions in Rosenthal, it would be required that the chat or session be “*created*,” that participants be “*invited...to join*,” and that the session be “*maintain[ed]*” during the chat or conference. Ex.1003, ¶130. This is also seen in another example:

For example, using such a feature, a local user may send, in response to something said by the remote user during a chat session, an electronic cream pie that may be inserted into the virtual video stream produced by the remote user, sticks to the remote user’s face and slowly slides down. In a similar manner, other virtual objects may be exchanged between both user’s virtual worlds, thus allowing new types of interactive video chat games. For example, a local user could throw, e.g., head, a virtual soccer ball to a remote user and the remote user may either head the ball back, or if the ball is missed, a goal may be scored by the local user. Such a virtual game could be

played amongst a plurality of users connected via the network.

Rosenthal, [0070]. Again, this discloses the elements of this claim because in order to play “such a virtual game” or for virtual objects to be “exchanged between both user’s virtual worlds,” the chat or conference session must have been created, participants must have been invited to join in some manner, and the session must be maintained to keep it open for these interactions. Ex.1003, ¶130.

Accordingly, Rosenthal teaches or suggests claim 50. Ex.1003, ¶131.

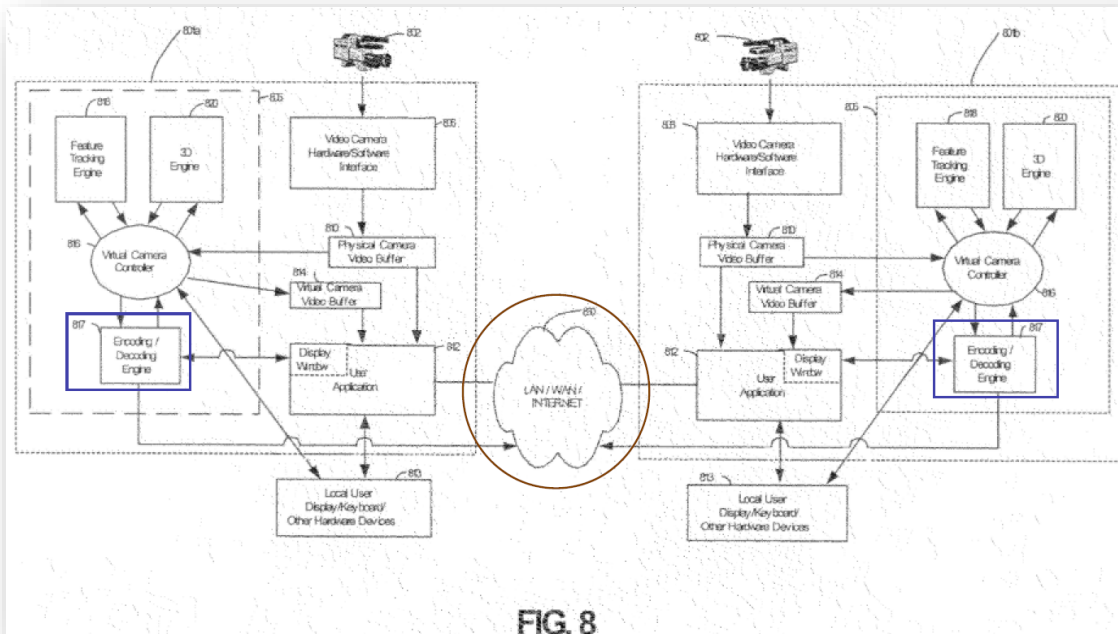
**i. Claim 53**

- i. [53] The method of claim 32, further comprising: encoding and distributing the video streams among the participants without any limitation on one or more of: the number of the participants in the video chat/conference, the number or type of objects controlled or accessed by the participants, the locations of the objects, and the types of real-world or augmented video streams.

“[A] reference need not state a feature’s absence in order to disclose a negative limitation.” *United Servs. Auto. Ass’n v. PNC Bank N.A.*, 2025 WL 706080, at \*3 (Fed. Cir. Mar. 5, 2025) (quoting *AC Techs. S.A. v. Amazon.com, Inc.*, 912 F.3d 1358, 1367 (Fed. Cir. 2019)). To the extent that this claim contains a negative limitation (“*without any limitation on one or more of...*”), Rosenthal discloses it.

In addition, Rosenthal discloses “*encoding and distributing the video streams among the participants.*” “[V]irtual camera controller may next pass instructions to

encoding/decoding engine 817, resulting in either a group of encoded pixels...or in an encoded data stream that may be transmitted between virtual camera embodiments executing on different computer systems connected by a network.” Rosenthal, [0055]. The video streams are encoded by the encoding/decoding engine and then distributed between the other computers connected to the network. This is further exemplified by Figure 8. Ex.1003, ¶133.



Rosenthal, Figure 8 (annotated in blue to show “encoding” and in brown to show how “the video streams among the participants” are “distribut[ed]”—i.e., via a network connection). The encoding/decoding engines of each participant’s system “may be used to encode information within the pixels...and/or may be used to encode information within a data stream passed data between virtual camera

embodiments executing on different communication systems connected via a network.” Rosenthal, [0072]. Additionally, for distributing the video streams, “a user application 812 may transmit the virtual camera video stream via network 830 to a corresponding user application 812 executing on a remote computer system 801b.” Rosenthal, [0056]. Ex.1003, ¶133.

Rosenthal also discloses “*without any limitation on one or more of: the number of the participants in the video chat/conference.*” For example, in describing various interactive chat games that could be played, Rosenthal states “[s]uch a virtual game could be played amongst a **plurality of users** connected via the network.” Rosenthal, [0070]. Similarly, Rosenthal describes “[s]uch user applications 812 may present one or more virtual camera video streams and/or real and/or virtual video streams received from other sources....” Rosenthal, [0053]. These support that the disclosure of Rosenthal is not limited to any number of participants in the video chat or conference. Ex.1003, ¶134.

Also disclosed is “*without any limitation on one or more of...the number or type of objects controlled or accessed by the participants, the locations of the objects.*” For example, “the 3-D engine may insert virtual 3-D video objects into the video stream based upon the received facial features data matrix and a list of virtual objects identified by a user for insertion into the data stream.” Rosenthal, [0010]. The list of virtual objects is not limited, and the objects are accessed by the

participants to be inserted into the video stream. This is further supported throughout Rosenthal. See, e.g., Rosenthal, [0037] (“Virtual camera control user interface 500 may also include a controls display 528 that allows a user to select from one or more virtual item display bars...one or more virtual items...that may be applied to the image 524.”). Ex.1003, ¶135.

Additionally, the element “*without any limitation on one or more of...the locations of the objects*,” is taught or suggested throughout Rosenthal, such as a “virtual camera may support the ability to track one or more defined sets of three-dimensional coordinates associated with image content within a video stream and may allow virtual 3-D objects to be inserted within the video stream, on a real-time basis, based upon the tracked sets of 3-D coordinates.” Rosenthal, [0039]; [0035] (“Based upon the features selected for insertion within the video...and the received facial feature coordinates matrix data, the 3-D engine may generate appropriate 3-D views of the selected injected content and may insert the generated content at an appropriate location within the received video stream.”). Furthermore, another example includes, “transferred data used to control operation of a local hardware device may include data related to the relative position of virtual objects, and/or real objects, tracked within the respective real camera video image streams and/or virtual camera video image streams exchanged between remote computer systems via network 830.” Rosenthal, [0066]. These each support that Rosenthal discloses an

invention that is not limited by the location of the objects to be inserted into the video streams. Ex.1003, ¶136.

Together, the number and types of objects, and the location of these objects are again disclosed in the following: “The described hardware independent virtual camera and real-time video coordinate tracking and content insertion device may accommodate any quantity and any type of data set files and/or databases or other structures containing stored data sets, measured data sets and/or residual data sets in any desired format....” Rosenthal, [0111]. Thus, not only are unlimited number or types of objects available for insertion disclosed by Rosenthal, the unlimited location of these objects is also disclosed. Ex.1003, ¶137.

Rosenthal further discloses “*without any limitation on one or more of...the types of real-world or augmented video streams.*” As discussed above, Rosenthal discusses hardware-independent embodiments of virtual cameras, capable of being integrated with any platform or equipment. See Rosenthal, [0006]. This discloses that any type of real-world or augmented video streams may be used. Ex.1003, ¶138.

Accordingly, Rosenthal teaches or suggests claim 53. Ex.1003, ¶139.

### **3. Claims 1, 12–16, 21–23, 26–27, 29–30**

Claims 1, 12–16, 21–23, 26–27, and 29–30 recite elements that are identical or analogous in scope to the claims above and are thus rendered obvious for the same reasons, as incorporated below. Ex.1003, ¶140.

| <b>Claims 32–36, 43–45,<br/>48–50, 53</b> | <b>Analogous Claim<br/>Limitations</b> | <b>Analogous Sections</b> |
|-------------------------------------------|----------------------------------------|---------------------------|
| [32.0]                                    | [1.0]                                  | §VIII.A.1.a               |
| [32.1]                                    | [1.1]                                  | §VIII.A.1.b               |
| [32.2]                                    | [1.2], [1.3]                           | §VIII.A.1.c               |
| [32.3]                                    | [1.4]                                  | §VIII.A.1.d               |
| [32.4]                                    | [1.5]                                  | §VIII.A.1.e               |
| [32.5]                                    | [1.6]                                  | §VIII.A.1.f               |
| [32.6]                                    | [1.7]                                  | §VIII.A.1.g               |
| [33]                                      | [12]                                   | §VIII.A.2.a               |
| [34]                                      | [13]                                   | §VIII.A.2.b               |
| [35]                                      | [29]                                   | §VIII.A.2.b               |
| [36]                                      | [30]                                   | §VIII.A.2.b               |
| [43]                                      | [14]                                   | §VIII.A.2.c               |
| [44]                                      | [15]                                   | §VIII.A.2.d               |
| [45]                                      | [16]                                   | §VIII.A.2.e               |
| [48]                                      | [21]                                   | §VIII.A.2.f               |
| [49]                                      | [22]                                   | §VIII.A.2.g               |
| [50]                                      | [23]                                   | §VIII.A.2.h               |
| [53]                                      | [26], [27]                             | §VIII.A.2.i               |

**B. Ground 2: Claims 14, 17, 19, 21, 43, and 46–48 Are Obvious in View of Rosenthal and Vesely**

**1. Dependent Claims 43 and 46–48**

**a. Claim 43**

- i. [43] The method of claim 32, wherein the multiplexing of the augmented video stream and the real-world video stream to generate the multiplexed video stream is performed based on active control on synchronization of HB (human behavior) of the participants in addition to X, Y, Z, and T coordinates.

To the extent one asserts that Rosenthal does not expressly disclose “[t]he method of claim 32, wherein the multiplexing of the augmented video stream and the real-world video stream to generate the multiplexed video stream is performed based on active control on synchronization of HB (human behavior) of the participants in addition to X, Y, Z, and T coordinates,” it would have been obvious in view of Vesely. A POSITA would be motivated to modify Rosenthal to include Vesely’s disclosure of manipulating virtual objects through hand gestures, motions, and touches for the reasons demonstrated above. See § VII.C; Ex.1003, ¶141.

In Patent Owner’s Preliminary Infringement Contentions, the “*active control and synchronization* are done by the participant using **hand gestures** to control the size and orientation of the objects in use.” Ex.1006, pg. 99. Vesely discloses that hand gestures may be used to manipulate virtual objects in the 3D scene. Vesely,

[0041]–[0042] (“[T]he user’s hands may be used to directly interact with virtual objects of the 3D scene....Thus, an accurate and tangible physical interaction is provided by allowing a user to touch and manipulate projected objects with his hands or hand held tools....”). The manipulation of virtual objects includes “allow[ing] the object to be **positioned, oriented, altered, or affected** as if the object actually was within the spatial space and acted upon by the manipulating device.” Vesely, [0044]; *see also* [0048] (“The system 100 may also allow the displayed image to be **magnified, zoomed, rotated, and moved.**”). As these manipulations occur, the “3D scene generator may update the displayed 3D scene based on changes in the...manipulations via the user input devices.” Vesely, [0047]. Ex.1003, ¶142.

Accordingly, Rosenthal and Vesely render claim 43 obvious. Ex.1003, ¶143.

**b. Claim 46**

- i. [46.1] The method of claim 32, further comprising: accepting one or more inputs selected from the group consisting of hand gestures, motions, and touches of the participant through a video-based computer interface; and

Although Rosenthal does not expressly disclose “*accepting one or more inputs selected from the group consisting of hand gestures, motions, and touches of the participant through a video-based computer interface,*” it would have been obvious in view of Vesely. As explained, a POSITA would be motivated to modify Rosenthal to include Vesely’s disclosure of manipulating virtual objects through

hand gestures, motions, and touches for the reasons demonstrated above. *See* § VII.C; Ex.1003, ¶144.

For example, Vesely discloses that hand gestures, motions, or touches may be used to manipulate virtual objects in the 3D scene. Vesely, [0041]–[0042] (“[T]he user’s hands may be used to directly interact with virtual objects of the 3D scene....Thus, an accurate and tangible physical interaction is provided by allowing a user to touch and manipulate projected objects with his hands or hand held tools....”). The manipulation described in Vesely includes “handling, managing, or otherwise using in some process,” encompassing manipulations of objects such as “picking it up, rotating it in any direction, moving it forward or backward, deconstructing it, etc.” Vesely, [0044]. Each of these forms of manipulating a virtual object can be done by hand in the virtual space, “allow[ing] the object to be positioned, oriented, altered, or affected as if the object actually was within the spatial space and acted upon by the manipulating device.” Vesely, [0044]; *see also id.* (“[T]he system may recognize a hand, apply it to a skeletal model, and use feature extraction to allow for such [virtual] manipulations.”). Ex.1003, ¶145.

- ii. [46.2] using the one or more inputs to perform one or more acts selected from the group consisting of: creating or controlling the one or more virtual objects in the multiplexed video stream, selecting various commands from an object submenu, and drawing three dimensional curves, surfaces, and/or ribbons in the multiplexed video stream.

Vesely discloses that the system may include “a peripheral device 130 that may allow the user to interact with the 3D scene with hand controlled or hand-held tools.” Vesely, [0048]. Additionally, as discussed in the prior limitation, Vesely details various means of creating or controlling the virtual objects in the video stream. *See* Vesely, [0044]. Ex.1003, ¶146.

Rosenthal teaches or suggests that the interface for the virtual camera control can include a “controls display 528 that allows a user to select from one or more virtual item display bars...one or more virtual items (e.g., glasses 544 and hat 546) that may be applied to image 524.” Rosenthal, [0037]. This display also includes features like to “set a variety of configuration parameters” and allows other commands to be performed such as “browse through the virtual items,” “highlight an individual virtual item,” and “activate/deactivate special features associated with a virtual item.” Rosenthal, [0037]. These functions disclose the selection of “*various commands from an object submenu.*” Ex.1003, ¶147.

In Rosenthal, a control display interface is physically interacted with to select various commands and also to create or control virtual objects. *See* Rosenthal,

[0037]. Similarly, Vesely explains the use of hand gestures, motions, and touches to interact with objects and the 3D space as a whole. *See* Vesely, [0042], [0044]. Together, Dr. Houh explains that a POSITA would be motivated to combine the creation/control of virtual objects and the selection of various commands on a physical control interface of Rosenthal with Vesely's use of hand motions, gestures, and touches to manipulate the 3D space for the reasons already demonstrated above. *See* § VII.C; Ex.1003, ¶148.

Accordingly, Rosenthal and Vesely render claim 46 obvious. Ex.1003, ¶149.

**c. Claim 47**

- i. [47] The method of claim 46, further comprising: limiting the controlling hand gestures of the participant to his/her first finger, second finger, and arm movements.

For the same reasons demonstrated in [46.1] and [46.2], claim 47 is obvious over Rosenthal and Vesely.

As described in Rosenthal, “a virtual camera may support the ability to track one or more defined sets of three-dimensional coordinates associated with image content within a video stream.” Rosenthal, [0082]. Based on these coordinates, the virtual camera “may allow virtual 3-D objects to be inserted within the video stream, on a real-time basis.” *Id.* Further, the “3-D engine may receive the facial feature matrix data in near real-time, and may track the user facial features, head rotation and translation in the video based upon the received facial feature matrix data.”

Rosenthal, [0010]; *see also id.* [0014] (“Such features are not necessarily limited to facial features.”). Thus, the hand gestures and other actions can be registered and updated accordingly. Ex.1003, ¶151.

Vesely describes that a user’s hands (or a glove, stylus, arm, or other user control device or tool) may be used to manipulate or control virtual objects. *See Vesely*, [0010], [0041]–[0042], [0079]. Specifically, “[t]he first 3D collaboration system may receive information regarding other tools or physical objects, **like a collaborator/user’s hand or arm**, in addition to or other than stylus 130.” Vesely, [0079]. Furthermore, “[i]n other embodiments, objects may be manipulated without a physical manipulator, user control device, or gloves.” Vesely, [0044]. By way of example, but by no means limited to such, Vesely describes how the system may “recognize a hand, apply it to a skeletal model, and use feature extraction to allow for such manipulations.” Vesely, [0044]. This is equally applicable to a “*first finger, second finger, and arm movements*.” *See, e.g., Vesely*, [0079], [0054] (“In the illustrated embodiment, two cameras 160 may track **any** physical object, and/or **body part or extension of user** 310.”). Rosenthal with Vesely’s use of hand motions, gestures, and touches to manipulate the 3D space for the reasons already demonstrated above. *See* § VII.C; Ex.1003, ¶152.

Accordingly, Rosenthal and Vesely render claim 47 obvious. Ex.1003, ¶153.

**d. Claim 48**

- i.** [48] The method of claim 32, further comprising: controlling and manipulating the one or more virtual objects within the multiplexed video stream by using pupil tracking to track movement of the participant's eyes.

To the extent Rosenthal does not expressly disclose “*controlling and manipulating the one or more virtual objects within the multiplexed video stream by using pupil tracking to track movement of the participant's eyes,*” it would have been obvious in view of Vesely. A POSITA would be motivated to modify Rosenthal to include Vesely's disclosure of manipulating virtual objects through hand gestures, motions, and touches for the reasons demonstrated above. *See* § VII.C; Ex.1003, ¶154.

For example, Vesely discloses the use of eyewear, or any other pupil tracking method, “to track the eyepoint of a user viewing a 3D scene presented by the system 100.” Vesely, [0039]; *see also id.*, [0026] (eyepoint is defined as “the physical viewpoint of a single eye or a pair of eyes”). Furthermore, “[c]ameras 160 may be used to image a user of system...track a user's movement, or **track a user's head or eyes.**” Vesely, [0045]; Ex.1003, ¶155.

Specifically, Vesely discloses that through tracking the movement of the participant's eyes, “the 3D scene generator may update the displayed 3D scene based on changes in the user's eyepoint.” Vesely, [0047]. This allows the “two or more

users participat[ing] in a video conference” to be able to see the “manipulations from the manipulating user’s viewpoint.” Vesely, [0081]. By changing the viewpoint of the virtual object, and the 3D scene updating correspondingly for all users, the “*one or more virtual objects within the multiplexed video stream*” are “*controll[ed] and manipul[at]ed*” through the “*track[ed] movement of the participant’s eyes*.” Ex.1003, ¶156.

Accordingly, Rosenthal and Vesely render obvious claim 48. Ex.1003, ¶157.

## 2. Claims 14, 17, 19, 21

Claims 14, 17, 19, and 21 recite elements that are identical or analogous in scope to the claims above and are thus rendered obvious for the same reasons, as incorporated below. Ex.1003, ¶158.

| Claims 43, 46–48 | Analogous Claim Limitations | Analogous Sections |
|------------------|-----------------------------|--------------------|
| [43]             | [14]                        | §VIII.B.1.a        |
| [46.1]           | [17]                        | §VIII.B.1.b.i      |
| [46.2]           | [17]                        | §VIII.B.1.b.ii     |
| [47]             | [19]                        | §VIII.B.1.c        |
| [48]             | [21]                        | §VIII.B.1.d        |

## **IX. CONCLUSION**

Claims 1, 12–17, 19, 21–23, 26–27, 29–30, 32–36, 43–50, and 53 of the 654 Patent are unpatentable for the reasons discussed above.

Dated: March 4, 2026

Respectfully Submitted,

/Scott Border/

Scott M. Border

Reg. No. 77,744

Winston & Strawn LLP

1901 L Street, NW

Washington, D.C. 20036

202.282.5054

sborder@winston.com

**CERTIFICATE OF COMPLIANCE**

I hereby certify that this petition complies with the type-volume limitations of 37 C.F.R. § 42.24, because it contains 13,258 words (as determined by the Microsoft Word word-processing system used to prepare the petition and including annotated figures), excluding the parts of the petition exempted by 37 C.F.R. § 42.24.

Dated: March 4, 2026

Respectfully Submitted,

/Scott Border/

Scott M. Border

Reg. No. 77,744

Winston & Strawn LLP

1901 L Street, NW

Washington, D.C. 20036

T: 202.282.5054

sborder@winston.com

**CERTIFICATE OF SERVICE**

Pursuant to 37 C.F.R. § 42.6(e), I hereby certify that on this 4th day of March, 2026, I caused to be served a true and correct copy of the foregoing and any accompanying exhibits by priority mail on the following:

Goodwin Procter LLP  
Attn: Patent Administrator  
100 Northern Avenue  
Boston, MA 02210

Zugara, Inc.  
13101 Washington Blvd. Suite 403,  
Los Angeles, California 90066

Dated: March 4, 2026

Respectfully Submitted,

/Scott Border/

Scott M. Border  
Reg. No. 77,744  
Winston & Strawn LLP  
1901 L Street, NW  
Washington, D.C. 20036

IPR2026-00266

T: 202.282.5054  
sborder@winston.com

## CLAIM LISTING

| <b>Claim 1</b>                                                                                                                                                                                                                                                                                                                                                                             |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| [1.0] A system, comprising:                                                                                                                                                                                                                                                                                                                                                                |
| [1.1] a data input device comprising a video camera, which in operation, provides a real-world video stream comprising video of a real-world 3D environment including a participant in a video chat or conference; and                                                                                                                                                                     |
| [1.2] a multiple stream sync module, which in operation:                                                                                                                                                                                                                                                                                                                                   |
| [1.3] receives, from another participant in the video chat or conference, an augmented video stream comprising augmentation data representing one or more 3D virtual objects retrieved from a virtual object database, and                                                                                                                                                                 |
| [1.4] multiplexes the augmented video stream received from the other participant with the real-world video stream of the real-world 3D environment in real time to generate a multiplexed video stream comprising the video of the real-world 3D environment and the one or more 3D virtual objects represented by the augmentation data received from the other participant,              |
| [1.5] extracts at least one of the 3D virtual objects from the multiplexed video stream,                                                                                                                                                                                                                                                                                                   |
| [1.6] based on an action of the participant, translates coordinates of the extracted virtual object in at least three dimensions within a multi-dimensional space of the multiplexed video stream, wherein the three dimensions of the multi-dimensional space are defined by the video of the real-world 3D environment, and                                                              |
| [1.7] causes a display device associated with the participant to display the multiplexed video stream comprising the video of the real-world 3D environment and the one or more 3D virtual objects, wherein both the participant and the other participant can translate the coordinates of the one or more 3D virtual objects within the three dimensions of the multi-dimensional space. |
| <b>Claim 12</b>                                                                                                                                                                                                                                                                                                                                                                            |
| [12] The system of claim 1, wherein the multiple stream sync module, in further operation, synchronizes the augmented video stream with the real-world video stream in both time (T) and three-dimensional coordinates (X, Y, Z).                                                                                                                                                          |
| <b>Claim 13</b>                                                                                                                                                                                                                                                                                                                                                                            |
| [13] The system of claim 1, wherein the real-world video stream comprises a first real-world video stream, and wherein the augmented video stream is synchronized with a second real-world video stream comprising video of a real-world environment including the other participant.                                                                                                      |

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Claim 14</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| [14] The system of claim 1, wherein: the multiple stream sync module, in operation, multiplexes video streams from two or more participants into a multiplexed video stream via active control on synchronization of HB (human behavior) of the two or more participants in addition to X, Y, Z, and T coordinates.                                                                                                                                                                                                                                                                                                        |
| <b>Claim 15</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| [15] The system of claim 1, wherein: the multiple stream sync module synchronizes augmented video streams from two or more participants in such a way as to reduce latency of multiplexed video streams shared among the two or more participants.                                                                                                                                                                                                                                                                                                                                                                         |
| <b>Claim 16</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| [16] The system of claim 1, wherein: the multiple stream sync module broadcasts the multiplexed video stream to all participants in the video chat or conference.                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| <b>Claim 17</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| [17] The system of claim 1, wherein: the multiple stream sync module accepts one or more inputs selected from the group consisting of hand gestures, motions, and touches of the participant through a video-based computer interface of the data input device, and wherein the one or more inputs are used to perform one or more acts selected from the group consisting of: creating or controlling the one or more virtual objects in the multiplexed video stream, selecting various commands from an object submenu, and drawing three dimensional curves, surfaces, and/or ribbons in the multiplexed video stream. |
| <b>Claim 19</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| [19] The system of claim 17, wherein: the multiple stream sync module limits the controlling hand gestures of the participant to his/her first finger, second finger, and arm movements.                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| <b>Claim 21</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| [21] The system of claim 1, wherein: the multiple stream sync module controls and manipulates the one or more virtual objects within the multiplexed video stream by using pupil tracking to track movement of the participant's eyes.                                                                                                                                                                                                                                                                                                                                                                                     |
| <b>Claim 22</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| [22] The system of claim 1, wherein: the multiple stream sync module works across multiple platforms so that the augmented video stream provided by the other participant from one platform is synchronized and shared with the participant using a different platform.                                                                                                                                                                                                                                                                                                                                                    |
| <b>Claim 23</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |

|                                                                                                                                                                                                                                                                                                                                                                                  |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| [23] The system of claim 1, further comprising: a video conference manager, which in operation, creates a chat room or conference session for the video chat or conference, invites the participants to join the video chat or conference, and maintains the session during the video chat or conference.                                                                        |
| <b>Claim 26</b>                                                                                                                                                                                                                                                                                                                                                                  |
| [26] The system of claim 1, further comprising: a transport system, which in operation, encodes and distributes the video streams among the participants.                                                                                                                                                                                                                        |
| <b>Claim 27</b>                                                                                                                                                                                                                                                                                                                                                                  |
| [27] The system of claim 26, wherein: the transport system supports the distribution of the video streams without any limitation on one or more of: the number of the participants in the video chat/conference, the number or type of objects controlled or accessed by the participants, the locations of the objects, and the types of real-world or augmented video streams. |
| <b>Claim 29</b>                                                                                                                                                                                                                                                                                                                                                                  |
| [29] The system of claim 1, wherein multiplexing the augmented video stream with the real-world video stream comprises overlaying the one or more virtual objects onto the video of the real-world environment.                                                                                                                                                                  |
| <b>Claim 30</b>                                                                                                                                                                                                                                                                                                                                                                  |
| [30] The system of claim 1, wherein multiplexing the augmented video stream with the real-world video stream comprises inserting the one or more virtual objects into the video of the real-world environment.                                                                                                                                                                   |
| <b>Claim 32</b>                                                                                                                                                                                                                                                                                                                                                                  |
| [32.0] A method, comprising:                                                                                                                                                                                                                                                                                                                                                     |
| [32.1] providing, using a video camera, a real-world video stream comprising video of a real-world 3D environment including a participant in a video chat or conference;                                                                                                                                                                                                         |
| [32.2] receiving, from another participant in the video chat or conference, an augmented video stream comprising augmentation data representing one or more 3D virtual objects retrieved from a virtual object database;                                                                                                                                                         |
| [32.3] multiplexing the augmented video stream received from the other participant with the real-world video stream of the real-world 3D environment in real time to generate a multiplexed video stream comprising the video of the real-world 3D environment and the one or more 3D virtual objects represented by the augmentation data received from the other participant;  |

|                                                                                                                                                                                                                                                                                                                                                                                              |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| [32.4] extracting at least one of the 3D virtual objects from the multiplexed video stream,                                                                                                                                                                                                                                                                                                  |
| [32.5] based on an action of the participant, translating coordinates of the extracted virtual object in at least three dimensions within a multi-dimensional space of the multiplexed video stream, wherein the three dimensions of the multi-dimensional space are defined by the video of the real-world 3D environment; and                                                              |
| [32.6] causing a display device associated with the participant to display the multiplexed video stream comprising the video of the real-world 3D environment and the one or more 3D virtual objects, wherein both the participant and the other participant can translate the coordinates of the one or more 3D virtual objects within the three dimensions of the multi-dimensional space. |
| <b>Claim 33</b>                                                                                                                                                                                                                                                                                                                                                                              |
| [33] The method of claim 32, wherein multiplexing the augmented video stream with the real-world video stream comprises synchronizing the augmented video stream with the real-world video stream in both time (T) and three-dimensional coordinates (X, Y, Z).                                                                                                                              |
| <b>Claim 34</b>                                                                                                                                                                                                                                                                                                                                                                              |
| [34] The method of claim 32, wherein the real-world video stream comprises a first real-world video stream, and wherein the augmented video stream is synchronized with a second real-world video stream comprising video of a real-world environment including the other participant.                                                                                                       |
| <b>Claim 35</b>                                                                                                                                                                                                                                                                                                                                                                              |
| [35] The method of claim 32, wherein multiplexing the augmented video stream with the real-world video stream comprises overlaying the one or more virtual objects onto the video of the real-world environment.                                                                                                                                                                             |
| <b>Claim 36</b>                                                                                                                                                                                                                                                                                                                                                                              |
| [36] The method of claim 32, wherein multiplexing the augmented video stream with the real-world video stream comprises inserting the one or more virtual objects into the video of the real-world environment.                                                                                                                                                                              |
| <b>Claim 43</b>                                                                                                                                                                                                                                                                                                                                                                              |
| [43] The method of claim 32, wherein the multiplexing of the augmented video stream and the real-world video stream to generate the multiplexed video stream is performed based on active control on synchronization of HB (human behavior) of the participants in addition to X, Y, Z, and T coordinates.                                                                                   |
| <b>Claim 44</b>                                                                                                                                                                                                                                                                                                                                                                              |

|                                                                                                                                                                                                                                                                                                                                                    |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| [44] The method of claim 32, wherein the multiplexing of the augmented video stream and the real-world video stream to generate the multiplexed video stream comprises synchronizing the augmented and real-world video streams in such a way as to reduce latency of the multiplexed video stream.                                                |
| <b>Claim 45</b>                                                                                                                                                                                                                                                                                                                                    |
| [45] The method of claim 32, further comprising: broadcasting the multiplexed video stream to all participants in the video chat or conference.                                                                                                                                                                                                    |
| <b>Claim 46</b>                                                                                                                                                                                                                                                                                                                                    |
| [46.1] The method of claim 32, further comprising: accepting one or more inputs selected from the group consisting of hand gestures, motions, and touches of the participant through a video-based computer interface; and                                                                                                                         |
| [46.2] using the one or more inputs to perform one or more acts selected from the group consisting of: creating or controlling the one or more virtual objects in the multiplexed video stream, selecting various commands from an object submenu, and drawing three dimensional curves, surfaces, and/or ribbons in the multiplexed video stream. |
| <b>Claim 47</b>                                                                                                                                                                                                                                                                                                                                    |
| [47] The method of claim 46, further comprising: limiting the controlling hand gestures of the participant to his/her first finger, second finger, and arm movements.                                                                                                                                                                              |
| <b>Claim 48</b>                                                                                                                                                                                                                                                                                                                                    |
| [48] The method of claim 32, further comprising: controlling and manipulating the one or more virtual objects within the multiplexed video stream by using pupil tracking to track movement of the participant's eyes.                                                                                                                             |
| <b>Claim 49</b>                                                                                                                                                                                                                                                                                                                                    |
| [49] The method of claim 32, further comprising: synchronizing the augmented and real-world video streams across multiple platforms so that the augmented video stream provided by the other participant from one platform is shared with the participant using a different platform.                                                              |
| <b>Claim 50</b>                                                                                                                                                                                                                                                                                                                                    |
| [50] The method of claim 32, further comprising: creating a chat room or conference session for the video chat or conference, inviting the participants to join the video chat or conference, and maintaining the session during the video chat or conference.                                                                                     |
| <b>Claim 53</b>                                                                                                                                                                                                                                                                                                                                    |

[53] The method of claim 32, further comprising: encoding and distributing the video streams among the participants without any limitation on one or more of: the number of the participants in the video chat/conference, the number or type of objects controlled or accessed by the participants, the locations of the objects, and the types of real-world or augmented video streams.

IPR2026-00266

[[**BLANK PAGE**]]



