

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

**DECLARATION OF HENRY HOUH
UNDER 37 C.F.R. § 1.68 IN SUPPORT OF
PETITION FOR *INTER PARTES* REVIEW**

TABLE OF CONTENTS

	Page
I. INTRODUCTION	5
II. QUALIFICATIONS AND PROFESSIONAL EXPERIENCE	7
III. LEVEL OF ORDINARY SKILL IN THE ART	13
IV. RELEVANT LEGAL STANDARDS	14
V. OVERVIEW OF THE 654 PATENT	17
A. The Disclosures of the 654 Patent.....	18
B. The Prosecution History of the 654 Patent	22
VI. CLAIM CONSTRUCTION	23
VII. IDENTIFICATION OF HOW THE CLAIMS ARE UNPATENTABLE.....	23
VIII. THE PRINCIPAL PRIOR ART	24
1. Summary of Rosenthal.....	24
2. Summary of Vesely	25
3. The Combination of Rosenthal and Vesely.....	26
IX. THE CHALLENGED CLAIMS ARE UNPATENTABLE	32
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	32
1. Independent Claim 32	32
2. Dependent Claims 33–36, 43–45, 48–50, 53	61
3. Claims 1, 12–16, 21–23, 26-27, 29–30.....	79
B. Ground 2: Claims 14, 17, 19, 21, 43, and 46–48 Are Obvious in View of Rosenthal and Vesely.....	80
1. Dependent Claims 43 and 46–48	80
2. Claims 14, 17, 19, 21	87
X. OBJECTIVE INDICIA OF NON-OBVIOUSNESS	88
XI. CONCLUSION.....	88

EXHIBIT LIST

No.	Exhibit Description
1001	U.S. Patent No. 10,200,654
1002	File History of U.S. Patent No. 10,200,654
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), https://www.hanselman.com/blog/2006-is-the-year-of-video-chat
1009	Jonathan Roubini, <i>Yahoo! Messenger</i> , PC Mag UK (Jan. 18, 2006), https://uk.pcmag.com/smb/28728/yahoo-messenger
1010	Microsoft Computer Dictionary, definitions of “MPEG,” “data compression,” and “latency”

TABLE OF ABBREVIATIONS AND CONVENTIONS

Abbreviation	Meaning
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, Henry Houh, declare as follows:

I. INTRODUCTION

1. I am making this declaration in support of Petitioner’s request for an *Inter Partes* Review of U.S. Patent No. 10,200,654 (“the 654 Patent”) to Szymczyk et al.

2. I am being compensated for my work in this matter at my standard hourly rate. I am also being reimbursed for reasonable and customary expenses associated with my work and testimony in this proceeding. My compensation is not contingent on the outcome of this matter or the specifics of my testimony.

3. I have been asked to provide my opinions regarding whether the subject matter of claims 1, 12–17, 19, 21–23, 26–27, 29–30, 32–36, 43–50, and 53 (“the Challenged Claims”) of the 654 Patent are obvious to a person having ordinary skill in the art (“POSITA”) at the time of the alleged invention, in light of the prior art. It is my opinion that the Challenged Claims would have been obvious to a POSITA.

4. In the preparation of this declaration, I have studied:

- the 654 Patent (Ex.1001),
- the prosecution history of the 654 Patent (Ex.1002),
- Preliminary Infringement Contentions (Ex.1006)
- the additional references and publications I refer to below and/or identified in the exhibit list.

5. In forming the opinions expressed below, I have considered: the documents listed above; the relevant legal standards, including the standard for obviousness; and my own knowledge and experience based upon my work in the field of network communications and security as described below.

6. Unless otherwise noted, all **emphasis** in any quoted material has been added.

7. Since I was retained, I have been in frequent contact with counsel while working on this matter to communicate my opinions. My opinions are presented both in this Declaration and in the Petition for Inter Partes Review, which it supports. In many, if not most, cases the language used to express my opinions in both documents is substantially the same. This is because I worked with counsel to capture my opinions, and generally speaking we captured them in the Petition itself as opposed to developing both the Petition and Declaration in tandem. My opinions, as reflected in the Petition we prepared, were then incorporated into my Declaration. I make this statement to explain the work process and underscore that the opinions in this Declaration are my own, were formed by me and were documented with the assistance of counsel in the manner described. As detailed herein, it is my opinion that each of the challenged claims is rendered obvious by prior art references that predate the earliest priority date of the 654 Patent. If requested, I am prepared to testify about my opinions expressed herein.

II. QUALIFICATIONS AND PROFESSIONAL EXPERIENCE

8. I am over the age of 18 and am competent to write this declaration. I have personal knowledge, or have developed knowledge, of these technologies based upon education, training, or experience, of the matters set forth herein.

9. My complete qualifications and professional experience are described in my Curriculum Vitae, a copy of which can be found in Ex.1004. The following is a brief summary of my relevant qualifications and professional experience.

10. I received a Ph.D. in Electrical Engineering and Computer Science from the Massachusetts Institute of Technology (“MIT”) in 1998. I also received a Master of Science degree in Electrical Engineering and Computer Science in 1991, a Bachelor of Science degree in Electrical Engineering and Computer Science in 1989, and a Bachelor of Science Degree in Physics in 1990, all from MIT.

11. I am currently self-employed as an independent technical consultant. I am also president of a company that provides supplemental science, technology, engineering, and mathematics (“STEM”) education to children of all ages.

12. During the course of my education at MIT, I was a teaching assistant (“TA”) in the Electrical Engineering and Computer Science Department’s core Computer Architectures course. I first was a TA as a senior for a role typically reserved for graduate students. I later became head TA. The course covered various topics in computer architectures, including caching concepts. As a TA, I helped

write homework assignments, lab assignments, and exams. I also taught in the recitation sections.

13. Later, as part of my doctoral research at MIT from 1991-1998, I was a research assistant in the Telemedia Network Systems (“TNS”) group at the Laboratory for Computer Science. The TNS group built a high-speed gigabit network and created applications that ran over the network. Example applications included ones for remote video capture, processing, display of video on computer terminals, and video conferencing. Our system was designed to process video in real time in order to create live effects such as green screening, motion detection, motion extraction, motion tracking, scene change detection, and effects such as a live video puzzle game.

14. I worked on the design of core network components, designing and building the high speed links, and designing and writing the device drivers for the interface cards. I also set up the group’s web server at a time when there may have been just hundreds of web servers in existence. I also studied web caches. I helped to build the web pages that initiated video sessions via a web interface. Vice-President Al Gore visited our group in 1996 and received a demonstration of—and remotely drove—a radio controlled toy car with a wireless video camera mounted on it; the video was encoded by TNS-designed hardware, streamed over the TNS-designed network and displayed using TNS-designed software. I have also edited,

fixed bugs in, compiled, and installed linux device drivers for a video capture card to be used for video capture and streaming.

15. I defended and submitted my Ph.D. thesis, titled “Designing Networks for Tomorrow’s Traffic,” in January 1998. As part of my thesis research, I analyzed local-area and wide-area flows to show a more efficient method for routing packets in a network, based on traffic patterns at the time.

16. From 1997 to 1999, I was a Senior Scientist and Engineer at NBX Corporation, a start-up that made business telephone systems for streaming packetized audio over data networks instead of using traditional telephone lines. NBX was later acquired by 3Com Corporation, and the phone system is still used today by numerous businesses. As part of my work at NBX, I designed the core audio reconstruction algorithms for the telephones, as well as the packet transmission algorithms. I wrote software that programmed the processor to interact directly with the ethernet chip for transmitting packets as well as setup and configuration of chip parameters, which is a role that device drivers in an operating system typically have. I also designed and validated the core packet transport protocol used by the phone system. The protocol was used for all signaling in the phone system, including for the setup of conference calls. The NBX system also featured a computer interface for initiating phone calls, which could also initiate conference calls. The NBX system also supported TAPI, the Telephony Application

Programming Interface, thus allowing other computer programs to integrate with our system telephony features. We obtained U.S. Patent No. 6,967,963, entitled “Telecommunication method for ensuring on-time delivery of packets containing time-sensitive data,” as a result of part of this work.

17. From 1999-2004, I was employed by Empirix or its predecessor company, Teradyne. Empirix was a leader in test tools for telecommunications protocols and systems, providing functional testing tools as well as load testing tools. From 2000-2001, I conceived and built a test platform for testing Voice-over-IP (VoIP). The first application on this new test platform was a cloud emulator for simulating the effects of transmitting VoIP over a busy network.

18. In 2006, as part of my role at BBN Technologies, I helped found PodZinger Inc., now known as RAMP Inc. PodZinger utilized BBN’s speech recognition algorithms to search through the spoken words in audio and video segments. While I was Vice President of Operations and Technology, PodZinger followed its initial prototype with a full streaming audio and video search solution.

19. In 2012, I founded Einstein’s Workshop, a science, technology, engineering, and math enrichment program for kids of all ages. The Workshop teaches programming, robotics, “making” (akin to creating, building, and tinkering), engineering, 3D CAD, mobile application development and many other subjects. Einstein’s Workshop teaches a variety of programming classes from Scratch to Java.

We also teach game development for mobile devices, for which I evaluated a number of mobile application programming environments. I also developed a robotics programming class in which robots could be programmed to be autonomous or under direct wireless control.

20. Einstein's Workshop also created and developed BlocksCAD, a 3D CAD system for kids and adults. BlocksCAD is used world-wide and has more than 110,000 registered users and 350,000 total users. BlocksCAD is being used in maker spaces worldwide and is integrated into school curriculum in various school systems in the U.S. at elementary, middle and high school levels. BlocksCAD was spun out into a separate company that was part of the LearnLaunch business accelerator program for educational technologies, MIT Playlabs 2018 and the 2018 MassChallenge Boston Cohort. BlocksCAD also received Phase 1 SBIR funding. BlocksCAD was acquired in 2025. Also, at Einstein's Workshop, I helped create a 3D photogrammetry system, in which we captured and created full 3D models of people using a camera system with 104 cameras. The 3D model could be rendered in 3D and a user could rotate the model and zoom in and out using 3D model rendering software. We also made full color 3D printouts of people.

21. I have been awarded several United States patents, and I have several patent applications pending including the following examples:

- U.S. Patent No. 7,975,296, “Automated security threat testing of web pages”,
- U.S. Patent No. 7,877,736, “Computer language interpretation and optimization for server testing”,
- U.S. Patent No. 7,801,910, “Method and apparatus for timed tagging of media content”,
- U.S. Patent No. 7,590,542, “Method of Generating Test Scripts Using a Voice-Capable Markup Language”,
- U.S. Patent No. 6,967,963, “Telecommunication method for ensuring on-time delivery of packets containing time-sensitive data”,
- U.S. Patent Publication No. 20070106685, “Method and apparatus for updating speech recognition databases and reindexing audio and video content using the same”,
- U.S. Patent Publication No. 20070106693, “Methods and apparatus for providing virtual media channels based on media search”,
- U.S. Patent Publication No. 20070106760, “Methods and apparatus for dynamic presentation of advertising, factual, and informational content using enhanced metadata in search-driven media applications”,

- U.S. Patent Publication No. 20070112837, “Method and apparatus for timed tagging of media content”,
- U.S. Patent Publication No. 20070118873, “Methods and apparatus for merging media content”,
- U.S. Patent Publication No. 20090222442, “User-directed navigation of multimedia search results”.

III. LEVEL OF ORDINARY SKILL IN THE ART

22. My opinions are provided based on what a person of ordinary skill in the art (“POSITA”) in the technical field of the invention would have understood at the time of the purported invention of the 654 Patent. I have been asked to assume the date of the purported invention for the 654 Patent is February 27, 2013.

23. I understand there are multiple factors relevant to determining the level of ordinary skill in the pertinent art, including (1) the levels of education and experience of persons working in the field at the time of the invention; (2) the sophistication of the technology; (3) the types of problems encountered in the field; and (4) the prior art solutions to those problems.

24. Based on the materials and information I have reviewed, and on my experience in the technical areas relevant to the 654 Patent, a POSITA with respect to the 654 Patent as of the time of the invention, i.e., the earliest possible priority date of the 654 Patent, February 27, 2013, 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.

25. For purposes of this Declaration, in general, and unless otherwise noted, my statements and opinions, such as those regarding my own experience and what a POSITA would have understood or known generally (and specifically related to the references I consulted herein), reflect the knowledge that existed in the relevant field as of the priority date of the 654 Patent.

IV. RELEVANT LEGAL STANDARDS

26. I am not an attorney. In preparing and expressing my opinions and considering the subject matter of the 654 Patent, I am relying on certain basic legal principles that counsel has explained to me.

27. I understand that prior art to the 654 Patent includes patents and printed publications in the relevant art that predate the priority date of the 654 Patent. For

purposes of this Declaration, I am applying February 27, 2013, as the priority date of the 654 Patent, but I take no position on that issue.

28. I have been informed and understand that subject matter claimed in a patent is obvious under 35 U.S.C. § 103 if a person of ordinary skill in the art at the time the alleged invention was made would have had reason to combine or modify the disclosures of one or more prior art references to arrive at the claimed subject matter. A claim may be unpatentable even if each claim limitation is not present or disclosed in a single prior-art item.

29. I have been informed and understand that, under the doctrine of obviousness, a claim is unpatentable if the differences between the invention and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which the subject matter pertains. A person of ordinary skill in the art is presumed to have knowledge of the relevant prior art at the time of the claimed invention.

30. I have been informed and understand that obviousness is based on the scope and content of the prior art, the differences between the prior art and the claim, the level of ordinary skill in the art and secondary indicia of obviousness and nonobviousness to the extent such indicia exist. I have been advised that the following evidence, when present, must be considered before forming a conclusion that a claimed invention is obvious: (1) the invention's commercial success, (2) long

felt but unresolved needs, (3) the failure of others, (4) skepticism by experts, (5) praise by others, (6) teaching away by others, (7) recognition of a problem, and (8) copying of the invention by competitors. In this instance, however, I am not aware of any evidence of any of these factors. I will of course consider any evidence on these issues which is presented to me.

31. I have been informed and understand the determination of whether the asserted claims would have been obvious to a person of ordinary skill in the art and, therefore, invalid, is not governed by any rigid test or formula. A determination that a claim is obvious is, instead, based on a common-sense determination that the claimed invention is merely a combination of known limitations to achieve predictable results. Any of the following rationales are acceptable justifications to conclude that a claim would have been obvious: (1) the claimed invention is simply a combination of known prior art methods to yield predictable results; (2) the claimed invention is a simple substitution of one known element for another to obtain predictable results; (3) the claimed invention uses known techniques to improve similar devices (methods, or products) in the same way; (4) the claimed invention applies a known technique to a known device (method, or product) ready for improvement to yield predictable results; (5) the claimed invention was “obvious to try”—choosing from a finite number of identified, predictable solutions, with a reasonable expectation of success; (6) there is known work in one field of endeavor

that may prompt variations of it for use in either the same field or a different one based on design incentives or other market forces if the variations would have been predictable to one of ordinary skill in the art; or, (7) there is some teaching, suggestion, or motivation in the prior art that would have led one of ordinary skill in the art to modify the prior art reference to combine prior art teachings to arrive at the claimed inventions.

32. I have been informed and understand that an analysis of whether a claimed invention is obvious must not rely on a hindsight combination of prior art. The analysis must proceed in the context of the time of the invention or claimed priority date and consider whether the invention as a whole would have been obvious to a person of ordinary skill in the art, taking into consideration any interrelated teachings of the prior art, the effects of demands known to the design community or present in the marketplace, and the background knowledge possessed by a person having ordinary skill in the art, all in order to determine whether there was an apparent reason to combine any known elements in the fashion claimed by the patent at issue.

V. OVERVIEW OF THE 654 PATENT

33. The 654 Patent was filed February 27, 2013. The 654 Patent issued February 5, 2019. I have been asked to assume that February 27, 2013 is the

effective filing date, but, for purposes of this Declaration, I take no position on that issue.

A. The Disclosures of the 654 Patent

34. 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.

35. Fig. 1 of the 654 Patent, shown below, purports to disclose a “system diagram 100 to support real time synchronization of multiple augmented video streams in an AR environment.” Ex.1001, 2:56-59. The 654 Patent says that “the components portrayed in this figure can be arbitrarily combined or divided into separate software, firmware, and/or hardware components,” and “can execute on the same host or multiple hosts.” *Id.*, 2:58-67.

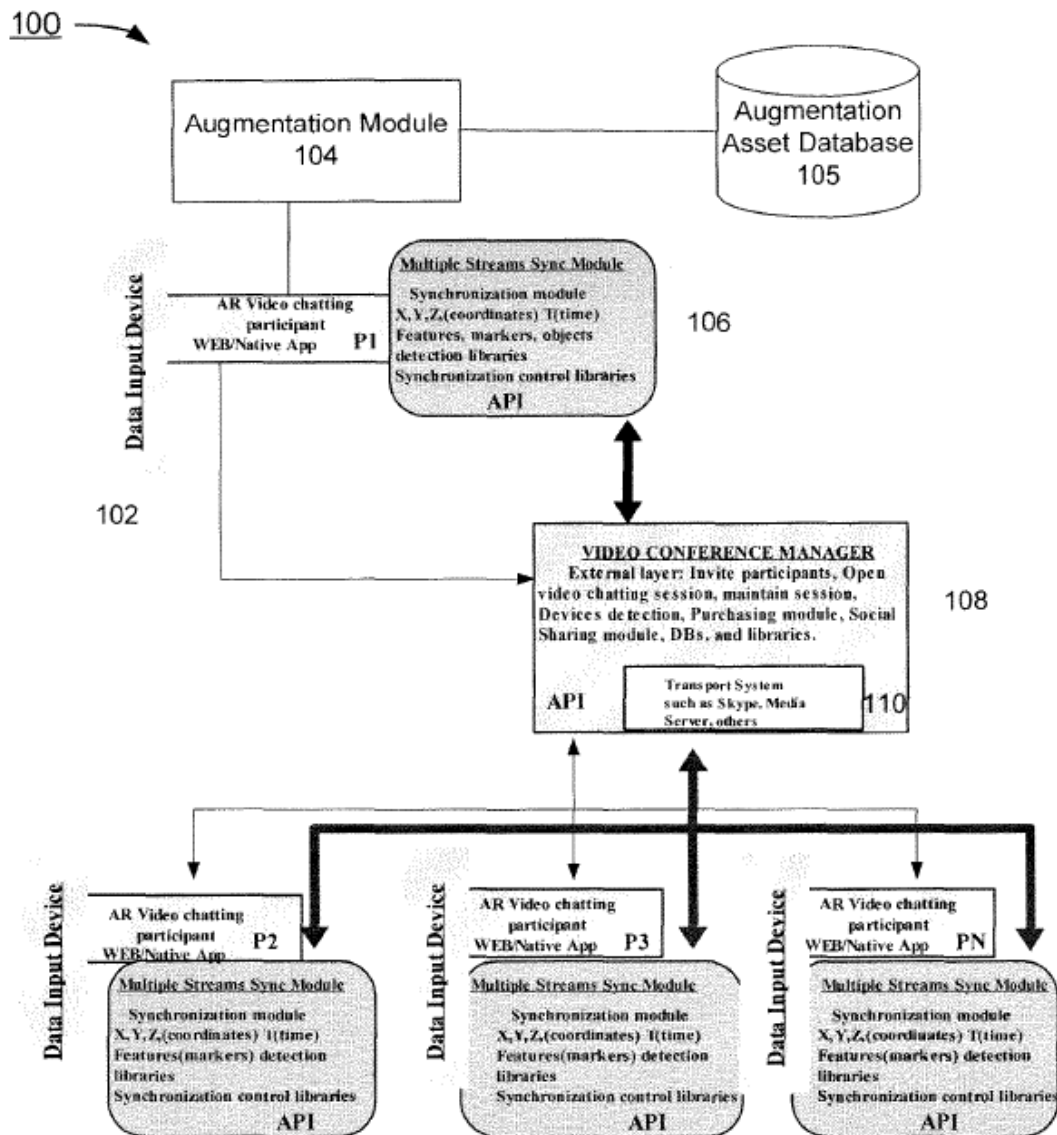


FIG. 1

36. With respect to Fig. 1, shown above, the 654 Patent explains that the “system 100 includes at least one data input device 102, an augmentation module

104, and a multiple streams sync module 106 associated with each of the multiple AP video chatting participants as well as a video conferencing manager 108, and a transport system 110.” Ex.1001, 3:1-5.

37. The 654 Patent explains that “each augmentation module 104 synchronizes each AR video chatting participant's initial video stream originated by an application running on the hosting device of the participant via data input device 102 with augmented objects and/or streams requested by the participant to create an augmented video stream in real time.” Ex.1001, 4:1-6. The augmentation module does this by “synchronizing the augmented objects with the initial video stream based on relative (X, Y, Z) coordinates and time (T) of the objects.” *Id.*, 4:9-13. It also “retrieves” augmented objects from a “augmented asset database 105” which can be “at the same location as the hosting device...or at a different location in the ‘cloud.’” *Id.*, 4:13-20. The 654 Patent further explains that the “augmentation module 104 may 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.” *Id.*, 4:21-30.

38. Fig. 1 also includes a “multiple stream sync module 106 associated with the participant” which “synchronize[s] the augmented video stream of the participant with the initial or augmented video stream of the other participants.” *Id.*, 4:31-35.

“[W]hen the participant of an AR system connects with another

participant...multiple streams sync module 106 will determine the X, Y, Z, and T coordinates of the objects in the augmented video stream of the participant.” *Id.*, 4:31-42. “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, and T coordinates dynamically modified in real time.” *Id.*, 4:42-50. Dependent claims 29–30 and 35–36 of the 654 Patent specify that “multiplexing the augmented video stream with the real-world video stream” can be accomplished by “overlaying the one or more virtual objects onto the video of the real-world environment” or “inserting the one or more virtual objects into the video of the real-world environment.” *Id.*, 11:8-16, 11:62-12:2.

39. The challenged claims all require 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 objects. *See, e.g.*, Ex.1001, cl. 32.

B. The Prosecution History of the 654 Patent

40. I have reviewed the prosecution history of the application that issued at the 654 Patent—U.S. Patent Appl. No. 13/778,430 (the “’430 App.”). In a December 29, 2017, office action the pending claims stood rejected as invalid over the prior art—the applicant’s own YouTube video (the “Zugara video”),¹ U.S. Patent Publication No. 2012/0320158 (“Junuzovic”), and a Journal Article.² Ex.1002 at 269–83. In a subsequent June 29, 2018, response, the applicant declined to amend the claims, and instead argued that the Zugara video did not teach or suggest the subject matter of the independent claims. *Id.* at 301–05. In particular, the applicant argued that “[n]othing in the Zugara video teaches or suggests how such augmented reality/ video conferencing technology might be implemented” for example, “there is simply nothing in the Zugara video that teaches or suggests that video V2 is generated by a module that ‘multiplexes the augmented video stream received from the other participant with the real-world video stream of the real-world 3D environment . . . to generate’ such a video stream as recited in claim 1.” *Id.* at 303.

¹ Zugara, *ZugSTAR - Streaming Augmented Reality*, (YouTube, Feb 4, 2010), <https://www.youtube.com/watch?v=eSFhCHkZgkA>.

² S. Nilsson et al., *Hands Free Interaction with Virtual Information in a Real Environment: Eye Gaze as an Interaction Tool in an Augmented Reality System*, *PsychNology Journal*, 7(2), 175-196 (2009) (“Nilsson”).

A notice of allowability was subsequently issued without identifying any particular reason for allowance. *Id.* at 313–22.

VI. CLAIM CONSTRUCTION

41. It is my understanding that for the purposes of this *inter partes* review, the claims are to be construed under the so-called *Phillips* standard, under which claim terms are given their ordinary and customary meaning as would have been understood by a POSITA in light of the specification and prosecution history, unless the inventor has set forth a special meaning for a term. For the purpose of this proceeding, no terms need be expressly construed—they are instead to be given their ordinary and customary meaning.

VII. IDENTIFICATION OF HOW THE CLAIMS ARE UNPATENTABLE

42. The discussion in this Declaration provides a detailed analysis of how the asserted prior art teaches each limitation of the Challenged Claims.

43. As part of my analysis, I have considered, and discuss in detail, the scope and content of the prior art and any differences between the alleged invention and the prior art.

44. It is my opinion that the alleged invention recited in the Challenged Claims would have been obvious in view of the teachings of the asserted prior art and the knowledge of a POSITA, as I demonstrate in detail below.

VIII. THE PRINCIPAL PRIOR ART

1. Summary of Rosenthal

45. 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.

46. 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]. 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.

2. Summary of Vesely

47. Vesely is directed 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].

48. 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, 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.

3. The Combination of Rosenthal and Vesely

49. To the extent one asserts that Rosenthal does not expressly disclose the Challenged Claims, 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].

50. A POSITA would recognize 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). 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].

51. 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. 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, 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. 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].

52. 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.

53. 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].

54. *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.

55. 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]. 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].

56. 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 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].

57. 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].

58. 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].

59. 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].

60. In addition, a POSITA would have been motivated to combine Vesely with Rosenthal based on the teachings of Rosenthal. For example, 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, 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].

61. 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.

IX. 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: A method

i. [32.0] A method comprising:

62. 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].

63. 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].

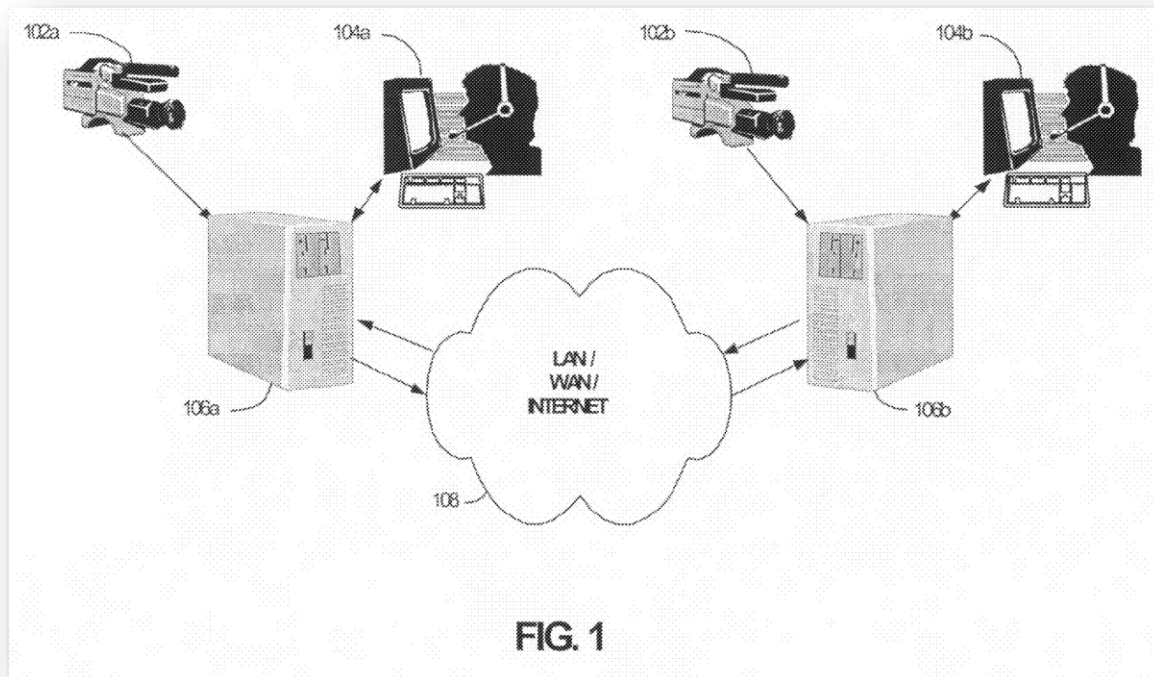
64. Accordingly, Rosenthal discloses “[a] method.”

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;

65. 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]. The “*real-world video stream*” conferencing system is generally depicted in Figure 1:

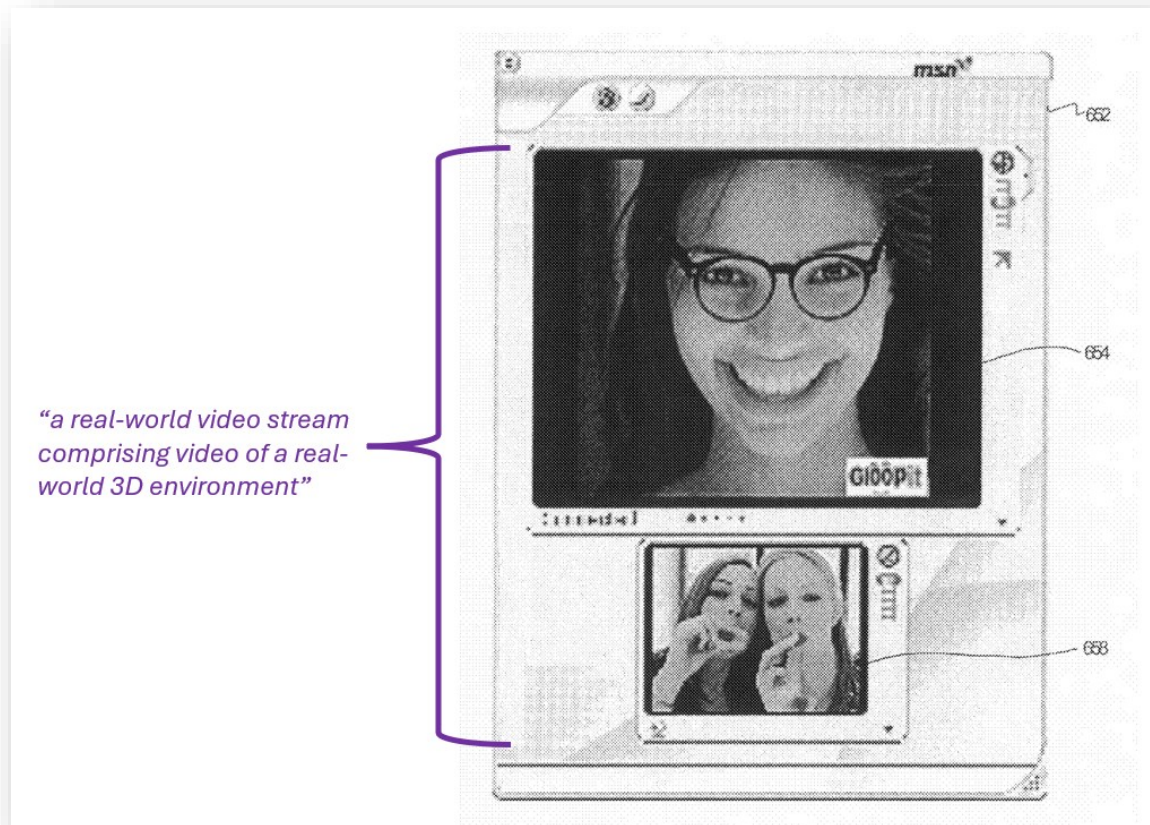


66. Rosenthal also explains that “*real-world video stream*” captured by the “*video camera*” and the associated virtual camera functionality permits “track[ing] one or more **defined sets of three-dimensional coordinates associated with image content within a video stream,**” and “the virtual camera 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].

67. Rosenthal discloses that the *“real-world 3D environment”* includes the *“participant in a video chat or conference.”* Rosenthal explains that its invention can be used with commonly used video conferencing systems as of the time of Rosenthal’s invention: “[t]he virtual camera controller may send the resulting stream of video data . . . 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]. Applications such as “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 well before the priority date of the 654 Patent. *See, e.g.*, 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>; 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.”).

68. Consistent with Rosenthal's implementation with commonly used video conferencing applications, as I describe above, Rosenthal discloses that the participant himself is included in the real-world video captured that is utilized by Rosenthal's system. This is illustrated in Figure 6 (below), and in associated disclosure, Rosenthal explains that Figure 6 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 the chat participants within a real-world environment.

69. 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.”

c. Augmented Video Stream

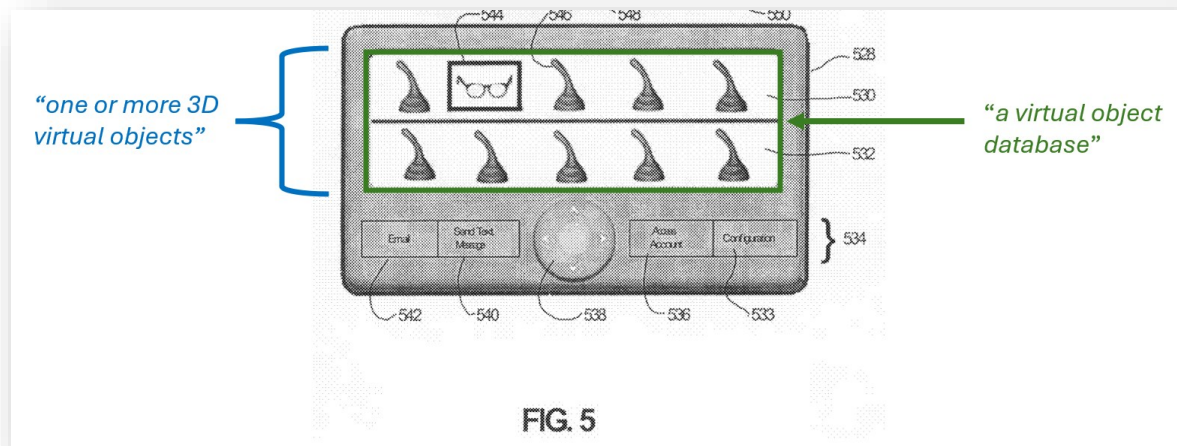
- i. [32.2] receiving, from another participant in the video chat or conference, an augmented video stream comprising augmentation data representing one or more 3D virtual objects retrieved from a virtual object database;

70. 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 can insert virtual 3-D video objects as described below:

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].

Moreover, referencing Figure 5, Rosenthal describes a “[v]irtual camera control user interface” [“*a virtual object database*”³] “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].



Rosenthal, Figure 5 (annotated).

71. 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 stored electronically in a traditional “*database*” underlying the graphical user interface used to select the 3-D virtual objects. Indeed, Rosenthal

³ 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].

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]. Likewise, Rosenthal discloses that “[t]he format and structure of internal information structures used to hold intermediate information in support of the described hardware independent virtual camera and real-time video coordinate tracking and content insertion approach, may include any and all structures and fields and are not limited to files, arrays, matrices, status and control booleans/variables.” Rosenthal, [0109]. Accordingly, a POSITA would have understood from Rosenthal’s disclosure that the selectable 3-D virtual objects would be stored in a database.

72. 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.”).

73. Rosenthal explains that the augmented video stream is transmitted via the bi-directional interface. Rosenthal states that its invention supports “a direct data communication channel to pass data between virtual camera embodiments executing on different communication systems connected via a network,⁴” (Rosenthal, [0066]), which permits a “user operating a first virtual camera executing on a first computer

⁴ 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]. This further demonstrates the transmission of virtual camera output (including virtual objects) between video conference participants.

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 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.”).

74. Rosenthal provides further 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].

75. 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.*”

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

76. 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.*”

77. 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*.”).

78. Rosenthal describes virtual camera software functionality that supports a bi-directional data interface. Rosenthal, [0082]. In particular, Rosenthal explains that after “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 then 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**.”).

79. 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.”).

80. 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].

81. 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, Rosenthal's insertion of remotely received virtual objects into a local virtual camera video stream results in a single combined video stream (“*a multiplexed video stream*”).

82. 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].

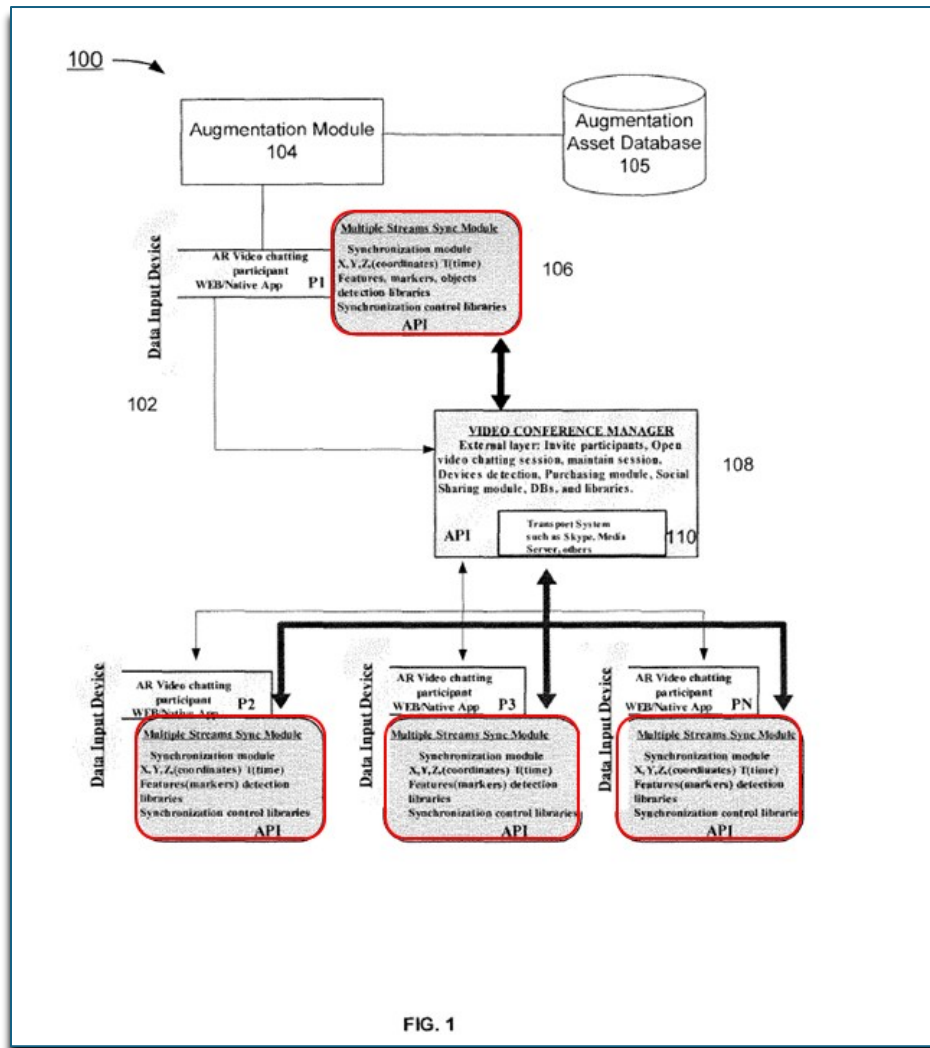
83. Accordingly, Rosenthal discloses “*multiplexing the augmented video stream received from the other participant with the real-world video stream of the real-world 3D environment in real time to generate a multiplexed video stream comprising the video of the real-world 3D environment and the one or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

e. Extracting 3D Virtual Objects

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

84. Important to my analysis is the manner in which “*extract[ion]*” is used within the 654 Patent. As illustrated in annotated Fig. 1 of the 654 Patent shown

below, the 654 Patent explains that each participant in a video conference has an associated “multiple streams sync module 106” (illustrated with red annotations) that each “synchronizes the augmented video stream of the participant with the initial or augmented video streams from other participants of the video chat or conference in both time (T) and in three-dimensional coordinates (X, Y, Z) to generate a synchronized augmented video stream.” Ex.1001, 4:29-37, *see also id.*, 3:1-7.



Ex.1001, Fig. 1.

85. The 654 Patent then explains “specifically” what this generation of a synchronized augmented video stream entails:

Specifically, when the participant of an AR system connects with another participant of the same, similar or different AR system during a video chat or conference session, multiple streams sync module 106 will determine the X,Y,Z, and T coordinates of the objects in the augmented video stream of the participant. **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.

86. The 654 Patent only provides high-level cursory description of what extraction entails—*extraction* is 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, 4:42-57. There are no technical details in the 654 Patent about what *extraction* otherwise entails. *See, e.g.*, Ex.1001, 5:3-7 (“...multiple streams sync module 106 allows display and extraction of one or more virtual/augmented objects from the synchronized augmented video stream, which could then in turn be manipulated by the participants.”); Fig. 3 (“Enable each of the participants to extract and manipulate the one or more objects in the synchronized augmented video stream in real-time.”).

87. In addition, the 654 Patent at column 4 lines 31 to 57 does not describe *extraction* as a discrete, post-processing step performed only after multiplexing is complete. Rather, 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.1001, 4:42-54.

88. Rosenthal discloses this technique. For example, Rosenthal’s techniques utilize a “bi-directional data interface” that “support[s] 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 the 3D objects in real-time. *Id.*; *see also id.*, [0082].

89. These interactive features in Rosenthal are enabled through tracking functionality (including Rosenthal’s feature tracking and 3-D engine), which “*extracts 3D virtual objects in 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. 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].

90. These disclosures make clear that Rosenthal's tracking functionality permits "*extraction*" through its identification of coordinates associated with virtual objects. *See also* Rosenthal, [0060] ("Frame 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."), Rosenthal, [0054]. This identification provided by the tracking functionality is then used by the 3D-engine for and during the real-

time manipulation of the coordinates of those virtual objects within the shared virtual camera video streams. *See, e.g., Rosenthal*, at [0010]–[0011] (“[T]he 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.”), [0070]. *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].

91. *Rosenthal* explains that using the “feature coordinates matrix data, 3-D engine 820 may **generate appropriate 3-D views of the selected injected content location** within the respective video frame images received from physical camera video buffer 810,” and then the “virtual camera controller may **next pass instructions** to encoding/decoding engine 817 resulting . . . in an encoded data stream that may be transmitted between virtual camera embodiments executing on different computer systems connected by a network.” *Rosenthal*, [0054]; *see generally Rosenthal*, [0066]–[0070].

92. Rosenthal's decoding engine also discloses *extraction* 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. Rosenthal explains that its system encodes virtual object-specific information (e.g., object type, speed, direction) into the virtual camera video streams exchanged between participants. Rosenthal, [0069] (“[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.”) On the receiving side, Rosenthal discloses that “[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.” Rosenthal, [0075] (emphasis added). Rosenthal explicitly describes that encoded information pertaining to virtual objects is thus “*extract*[ed]” in order “to add corresponding injected content to the virtual video stream produced by the remote computer and/or to control locally connected hardware devices based on the embedded information.” Rosenthal, [0069].

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

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

94. Rosenthal discloses this limitation. As an initial matter Rosenthal’s “field of invention” is explicitly characterized as pertaining to “the real-time **manipulation** of video images.” Rosenthal, [0003].

95. Rosenthal discloses this limitation is through Rosenthal’s tracking functionality. 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.”). For example, Rosenthal explains that a “frame image feature tracking module” 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, [0059], [0060]. For example, for “insertion of a virtual hat, frame image feature tracking module 902 may obtain and store coordinates associated with the top of an individuals head, eyes and ears; for insertion of virtual glasses [it] may obtain and store coordinates associated with eyes and ears; and for insertion of virtual lipstick, [and it] may obtain and store coordinates associated with an individuals lips.” Rosenthal, [0060].

96. 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]. For example, Rosenthal states that its invention’s “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]. 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 moving 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.

97. Further, as set forth in detail above with respect to claim limitation 32.4, Rosenthal makes clear that the virtual objects received from other participants are continuously decoded and extracted (*extracted virtual object*), and can thereafter being manipulated by the participant. Rosenthal illustrates this functionality through its description 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].

98. Just like Rosenthal’s disclosure described above, the 654 Patent also describes that participants can manipulate virtual objects through their body movement. *See, e.g.*, Ex.1001, 5:3-27 (The “multiple streams sync module 106 accepts a combination of one or more...**motions**...of a participant through a video-based computer interface of data input device 102.”).

99. Accordingly, a POSITA would have understood that Rosenthal’s tracking functionality 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.

100. Further, Rosenthal also discloses this limitation through its audio-based object manipulation. In particular, Rosenthal discloses that its system includes an “audio tracking module 904” that is “configured 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]. As discussed in limitation 32.4 above, this object-specific information is continuously decoded from the multiplexed virtual camera video stream during real-time processing, enabling audio-based actions (e.g., spoken phrases) to trigger coordinate translation and manipulation.

101. The 654 Patent analogously describes that participants can manipulate virtual objects through “voice commands.” *See, e.g.*, Ex.1001, 5:27-32.

102. 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.”

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.

103. 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].

104. A POSITA would have understood that Rosenthal’s disclosure contemplates that both participants’ display devices simultaneously display virtual objects added to real-world video streams of both participants. Indeed, Rosenthal

explains that it's "bi-directional data interface" supports the ability to "controllably pass[] across virtual camera video streams" the virtual objects embedded within each video stream, which enables "interactive gaming." Rosenthal, [0082]. 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*"). See 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].

105. Rosenthal also discloses that "*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.*"

106. For example, 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), permitting the "*translation*" of three-

dimensional coordinates in real-time so that they may be **dynamically manipulated and synchronized** between participants. 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].

107. This ability for both 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).

108. 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.*”

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).

109. 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]), providing an example of the disclosure of synchronizing the time coordinate of the multiplexed video stream. *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.”).

110. 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].

111. Further, a POSITA would have understood that Rosenthal discloses that 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). As a technical matter, a POSITA would have understood that

inserting or compositing virtual 3-D objects into a live video stream in “real time” necessarily requires temporal alignment between the augmentation data and the underlying video frames. This understanding is consistent with the 654 Patent’s own description of multiplexing augmented content with a real-world video stream in real time. Ex.1001 at 2:20-23 (“Here, the initial and/or augmented video streams from the participants are synchronized **in real time, not only in time** but also in three-dimensional space/coordinates (X, Y, Z).”) (emphasis added).

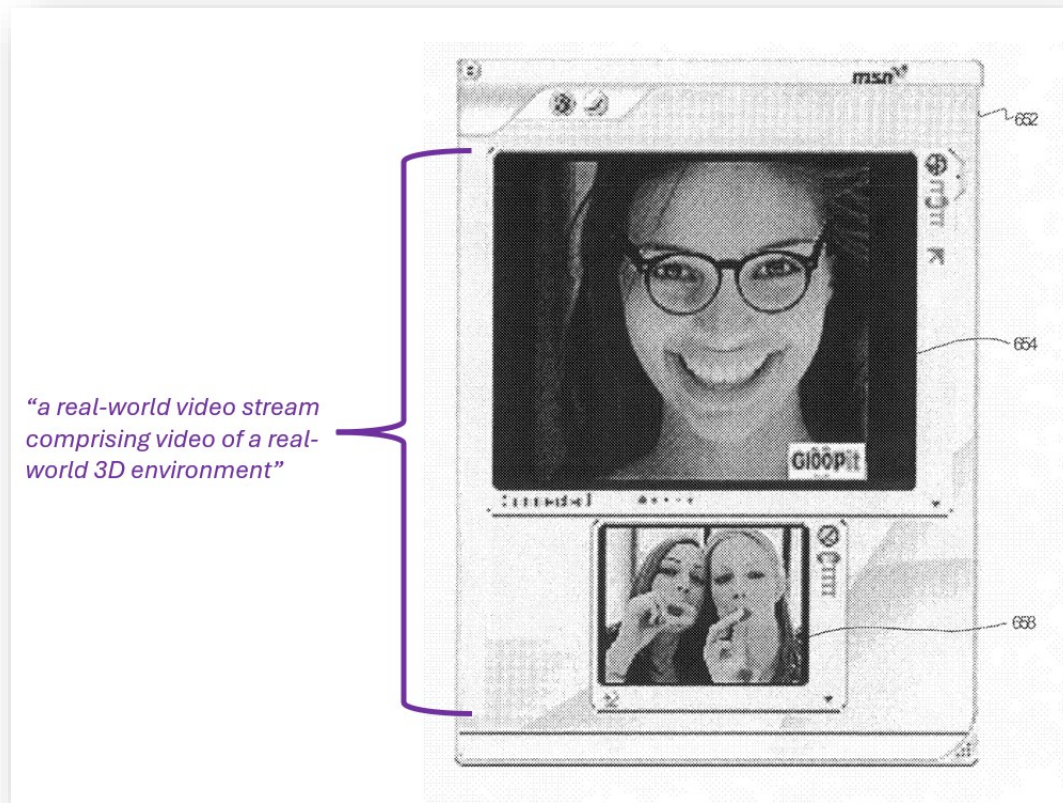
112. Accordingly, Rosenthal teaches or suggests claim 33.

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.

113. 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].



Rosenthal, Figure 6 (annotated).

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

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.

115. 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”).

116. 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]. 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.

117. Accordingly, Rosenthal teaches or suggests claim 43.

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.

118. 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], 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.*”

119. In addition, Rosenthal states that “the encoded data included within a selected image format, e.g., **MPEG, or other image format . . .**” At the time of the 654 Patent, MPEG was a well-known standard for audio and video compression. 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.1010, Microsoft Computer Dictionary, definition of “data compression”. 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.1010, Microsoft Computer Dictionary, definition of “latency”.

120. Accordingly, Rosenthal teaches or suggests claim 44.

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.

121. 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 . . .”).

122. 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].

123. Accordingly, Rosenthal teaches or suggests claim 45.

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.

124. 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].

125. 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.

126. 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].

127. Accordingly, Rosenthal teaches or suggests claim 48.

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.

128. 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].

129. Accordingly, Rosenthal teaches or suggests claim 49.

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.

130. 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]. 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. 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.

131. Accordingly, Rosenthal teaches or suggests claim 50.

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.

132. I have been informed that a reference doesn't need to state a feature's absence in order to disclose a negative limitation. To the extent that this claim

contains a negative limitation (“without any limitation on one or more of . . .”), Rosenthal discloses it.

133. 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.

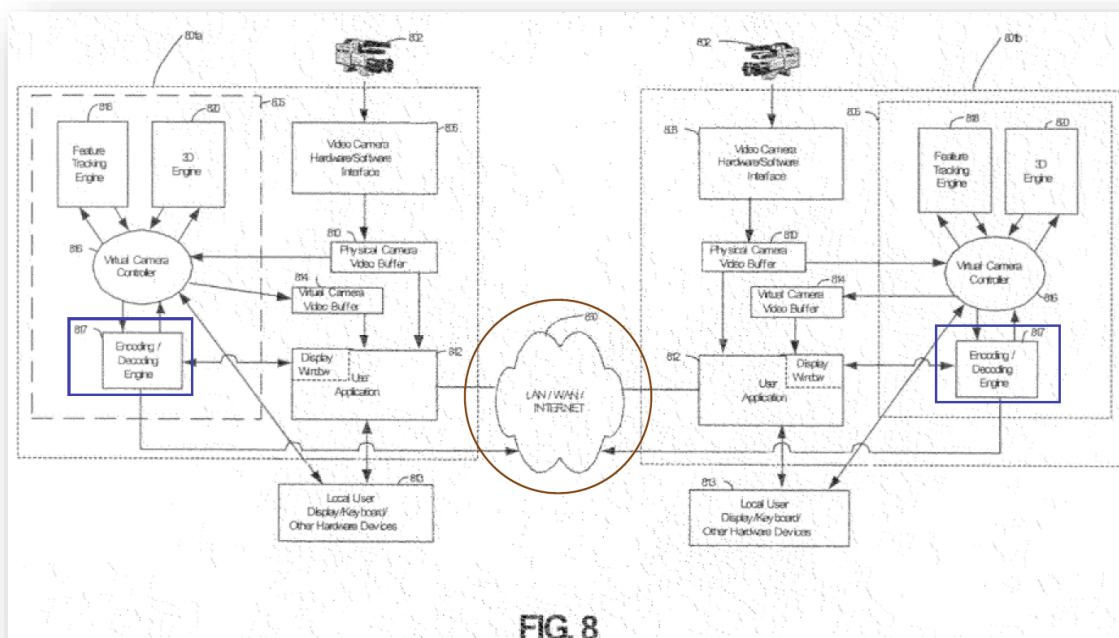


FIG. 8

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].

134. 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.

135. 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.”).

136. 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.

137. 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.

138. 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.

139. Accordingly, Rosenthal teaches or suggests claim 53.

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

140. As part of my analysis, I have considered the rest of the claims of the 654 Patent. It is my opinion that 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.

Claims 32–36, 43–45, 48–50, 53	Analogous Claim Limitations	Analogous Sections
[32.0]	[1.0]	§IX.A.1.a
[32.1]	[1.1]	§IX.A.1.b
[32.2]	[1.2], [1.3]	§IX.A.1.c
[32.3]	[1.4]	§IX.A.1.d
[32.4]	[1.5]	§IX.A.1.e
[32.5]	[1.6]	§IX.A.1.f
[32.6]	[1.7]	§IX.A.1.g
[33]	[12]	§IX.A.2.a
[34]	[13]	§IX.A.2.b
[35]	[29]	§IX.A.2.b

Claims 32–36, 43–45, 48–50, 53	Analogous Claim Limitations	Analogous Sections
[36]	[30]	§IX.A.2.b
[43]	[14]	§IX.A.2.c
[44]	[15]	§IX.A.2.d
[45]	[16]	§IX.A.2.e
[48]	[21]	§IX.A.2.f
[49]	[22]	§IX.A.2.g
[50]	[23]	§IX.A.2.h
[53]	[26], [27]	§IX.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.

141. To the extent 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* § VIII.3.

142. 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].

143. Accordingly, Rosenthal and Vesely render claim 43 obvious.

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

144. 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. 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* § VIII.3.

145. 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.”).

- 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.

146. 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].

147. 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*.”

148. 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, 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* § VIII.3.

149. Accordingly, Rosenthal and Vesely render claim 46 obvious.

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.

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

151. 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.

152. 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 § VIII.3.*

153. Accordingly, Rosenthal and Vesely render claim 47 obvious.

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.

154. 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 § VIII.3.

155. 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].

156. 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[ed]*” through the “*track[ed] movement of the participant’s eyes.*”

157. Accordingly, Rosenthal and Vesely render obvious claim 48.

2. Claims 14, 17, 19, 21

158. As part of my analysis, I have considered the rest of the claims of the 654 Patent. It is my opinion that 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.

Claims 43, 46–48	Analogous Claim Limitations	Analogous Sections
[43]	[14]	§IX.B.1.a

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

X. OBJECTIVE INDICIA OF NON-OBVIOUSNESS

159. I am not aware of any objective indicia of non-obviousness that would change my opinion that the challenged claims of the 654 Patent are obvious in view of the prior art discussed above in this declaration. It is my opinion that any potential objective indicia of non-obviousness would not overcome the clear disclosures in the prior art showing that the subject matter of these claims would have been obvious to a POSITA as of February 27, 2013. However, I reserve the right to supplement my opinions if Patent Owner alleges or offers any purported evidence of objective indicia of non-obviousness.

XI. CONCLUSION

160. I declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true, and that these statements were made with knowledge that willful false statements and

the like so made are punishable by fine or imprisonment, or both, under section 1001 of Title 18 of the United States Code.

A handwritten signature in black ink that reads "Henry H. Houh." The signature is written in a cursive, slightly slanted style.

Henry H. Houh, PhD

Dated: February 20, 2026

CLAIM LISTING

Claim 1
[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.
Claim 12
[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).
Claim 13
[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.
Claim 14
[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.
Claim 15
[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.
Claim 16
[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.
Claim 17
[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.
Claim 19
[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.
Claim 21
[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.
Claim 22
[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.
Claim 23
[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.
Claim 26
[26] The system of claim 1, further comprising: a transport system, which in operation, encodes and distributes the video streams among the participants.
Claim 27
[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.
Claim 29
[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.
Claim 30
[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.
Claim 32
[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.
Claim 33
[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).
Claim 34
[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.
Claim 35
[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.
Claim 36
[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.
Claim 43
[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.
Claim 44
[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.
Claim 45
[45] The method of claim 32, further comprising: broadcasting the multiplexed video stream to all participants in the video chat or conference.
Claim 46
[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.
Claim 47
[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.
Claim 48
[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.
Claim 49
[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.
Claim 50
[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.

Claim 53

[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.