Tele-Immersion Concepts

Stephan Ohl

Abstract—Tele-immersive systems development is always driven as well as restricted by the available immersive technology. Hence, existing such systems are described mainly from a technological point of view; their conceptual description is usually limited to the description of a scenario that is implementable with or circumvents the limitations of the chosen technology. This focus on technology makes it difficult to compare systems' concepts; moreover, it has led to different views on tele-immersion in different fields, such as remotely controlled robots, immersive video conferencing, and tele-collaboration. In this work, we give a general, structured principle to describe the conceptual part of any tele-immersion system. This principle naturally unifies the different views on tele-immersion. Our idea is based on the insight that, in order to be general, immersion must be described separately for each direction of communication. We characterize communication between locations using a graph; for each directed edge of this graph, we describe immersion as operations on volumes. Using this principle, we define a typology, which enables the comparison and enumeration of tele-immersion concepts. We apply this typology to survey the concepts of existing tele-immersion systems and thereby demonstrate how three well-known tele-immersive scenarios—Marvin Minsky's tele-operated robot, the Office of the Future, and the asymmetric Beaming scenario—integrate naturally. We show how the general principle can be utilized conveniently to grasp conceptual ideas in tele-immersion, such as direct interaction, locational presence, spatial consistency, symmetries, and self-inclusion.

Index Terms—Tele-immersion, telepresence, immersive tele-collaboration, virtual reality, augmented reality, mixed reality, communication notation, immersive extent, joint social attention

1 Introduction

Tele-immersion aims to immerse a user in a way that blurs the borders between her or his local environment and a physically distant location. Marvin Minsky [1] coined the term telepresence for a user immersed in the world of a remote-controlled robot; the user feels the robot's distant world as if 'being there.' Later, the term was used by the authors of the 'Office-of-the-Future' [2] for joining multiple physically distributed rooms virtually. Each user stays and perceives his own location; however, at the same time, all users are immersed in one shared workspace.

Both scenarios, the tele-operated robot and the Office-of-the-Future, immerse a user so that she or he can perceive the remote physical location. To put it another way, the remote physical location is telecommunicated immersively to the user to alter her or his sense of presence. Therefore, a technology to telecommunicate a physical location immersively is a basic component of both scenarios. Many different such technologies from the fields of virtual reality, computer vision, and robotics have been proposed.

The development of immersive technologies as well as their integration into a tele-immersive system is non-trivial. High-quality data representations of our three-dimensional world have to be acquired, compressed, transmitted, decompressed, and rendered in real-time and with low latency. These high technological demands have led to mainly technology-focused description of systems. If a concept description is stated, it is often given as an application-centered scenario (such as a shared virtual table environment), which is mixed with descriptions of immersive technology and does not follow a general principle.

Technology-based scenario descriptions are very valuable in their own right. However, we think that a clear separation between concept and technology has a number of benefits. A clear and concise general principle to describe tele-immersion concepts not only would help to analyze systems but also would be beneficial to structure existing literature, make concepts comparable, reason about their pros and cons, help in choosing a concept for a certain application, and explore new concepts and applications. This work gives such a general principle to describe tele-immersion concepts. We make the following contributions:

(i) a general principle and typology to describe tele-immersion concepts, which unifies the view on tele-immersion concepts and makes them comparable and enumerable;
(ii) a discussion on how this general principle can be utilized conveniently to grasp conceptual ideas such as direct interaction, locational presence, spatial consistency, symmetries, and self-inclusion;
(iii) a review of concepts of existing tele-immersive systems.

2 Related Work

Concepts for tele-immersive systems have emerged from different areas of research. The most important fields are robotics, video conferencing, and tele-collaboration. All

- *The author is with the Computer Science, University of Rostock, Rostock 18051, Germany. E-mail: stephan.ohl@posteo.net.*

Manuscript received 16 Oct. 2016; revised 29 July 2017; accepted 16 Aug. 2017. Date of publication 1 Nov. 2017; date of current version 31 Aug. 2018. (Corresponding author: Stephan Ohl.)
Recommended for acceptance by D. Schmalstieg.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TVCG.2017.2767590

fields use similar virtual reality technology to immerse users. However, traditionally, each field focused on different technologies to represent users, namely, robots, light fields, and virtual avatars.

In robotics, immersion into the world of a remote-controlled robot was proposed by Marvin Minsky [1]. Paulos and Canny coined the term 'tele-embodiment' for the concept of a physical mobile robot platform with an integrated display and camera for communication [3]. Recent systems that follow this concept are, for instance, the humanoid EDGAR robot [4] and 'robotic humanoid surrogates' [5].

In the field of video conferencing, a number of shared virtual table concepts have been proposed (see, e.g., [6], [7], [8], [9]). The seating geometry is arranged virtually such that eye-contact is established and multiple rooms are joined. The 'Office-of-the-Future' [2] generalized this concept by introducing an additional shared workspace that virtually overlays the joined rooms. A similar concept of combining person spaces with a joint task space was introduced by Buxton [10].

Virtual avatars are used traditionally in fully immersive tele-collaboration systems. Even though virtual avatars evolved from simple polygonal models to advanced representations based on real-time acquisition, the underlying concept of complete immersion in a virtual world, however, stays the same. Possibly the most prominent example of systems using this concept is blue-c [11].

Earlier, group-to-group communication was restricted to the domain of shared virtual table conferencing concepts with a primary focus on eye-contact [12], [13]. More recently, however, a group-to-group concept was implemented in the fully immersive C1x6 system [14], [15]. Additionally, to ease navigation, the concept introduced a world-in-miniature, which reincludes virtual avatars in a down-scaled version. A similar concept of self-inclusion of avatars, however in life-size, can be found in the 'third-person' and 'mirror' interaction modes described in [16].

Major leaps in computer vision and display technology led to borders between different representations vanishing. Humanoid robots are enhanced with projected faces [4] and immersive transparent displays are free-standing in rooms [17]. The advances in optical see-through head-mounted displays in combination with real-time depth-sensing devices allow numerous possibilities to mix the virtual and the real world [18]. This mix of different technologies is reflected in recent conceptual tele-immersion scenarios (see, e.g., [19], [20], [21]).

Not all proposed tele-immersion concepts are symmetric. Interestingly, the Beaming concept puts emphasis on this asymmetry [20]. A fully immersed user visits a group of lighter immersed users at another location, whereby the concept aims to counter-balance the social asymmetry of the visitor scenario by the technical asymmetry of immersion. We will show how this asymmetry can be expressed within our general principle. A more thorough review of tele-immersion concepts of existing systems, which is based on our typology, is given in Section 8.

3 Overview

In order to define a general description principle for tele-immersion concepts, we divide tele-immersion into parts that can be described individually. The two main parts of tele-immersion are communication over distance—tele—and immersion. First, we characterize communication between conversational partners who are geographically distributed in Section 4. Then, in Section 5, for one direction of communication between two different geographical locations, we characterize immersion. This isolated directional description of immersion is at the very heart of our principle and typology. A complete tele-immersion concept is defined by its communication and its immersion for each directed communication. We consider different aspects of complete tele-immersion concepts in Section 6 and discuss enhancements of our principle in Section 7. We develop a tele-immersion concept typology along with the definition of the general principle. This typology is applied in Section 8 to review the tele-immersion concepts of existing systems.

4 Communication

We term a physical place a *location*. A *location* may be, for instance, an office, a football stadium, or even a module of the International Space Station. A non-physical place is termed a *virtual location*. Virtual locations are part of many tele-immersive collaboration scenarios and are ubiquitous in online virtual worlds.

At each location, there are zero, one, or more *users*. A location with zero users is an *empty location*. An empty location may be, for instance, an underwater structure that is explored immersively using a remotely operated vehicle. At a virtual location, there can not be any physical users. Therefore, a location is either empty, non-empty, or virtual.

If we are referring to a digital representation of a location and its users, we use the term *scene*. There can be multiple scenes of the same location. Virtual scenes are digital representations of virtual locations, and, unlike non-virtual scenes, they do not correspond to any physical places.

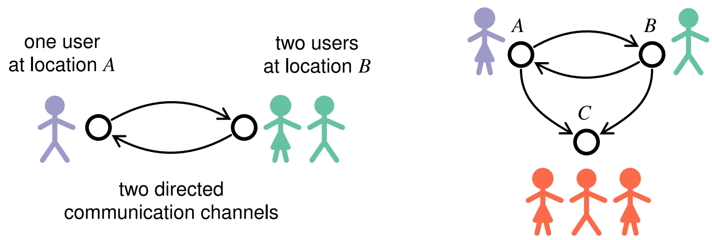
The communication within a tele-immersion concept can be abstracted by locations, the number of users present at each location, and the remote scenes integrated at each location.

4.1 Communication Graph

Communication over distance is described by a simple directed graph that we term *communication graph* (CG). The vertices of a CG correspond to locations and are labeled with the number of users at each location or they are marked virtual. There is a directed edge from location A to location B if and only if location A can be perceived at location B. We term these directed edges *communication channels*.

A CG describes communication at the perceptual level. This means—with regard to tele-immersion—that communication channels convey representations of locations. A communication channel between A and B says that a scene of A is integrated within the physical space of B. In most cases, to allow mutual communication, there is a forward and a backward communication channel. Fig. 1 illustrates the description of communication by CGs.

CGs describe partitions of the set of users and thereby restrict the possible *communication patterns*. An example is a lecture that is streamed to a remote audience, a one-to-many tele-immersion concept. If, for instance, the remote



(a) three users at two locations

(b) five users at three locations

A communication graph is a triple $G = (P, E, u)$ of a finite set of locations P, a set of directed communication channels $E = \{(p, p') \mid p, p' \in P \wedge p \neq p'\}$, and a function $u : P \mapsto \{0, 1, \ldots\} \cup \{V\}$ that assigns a non-negative number of users or V to each location. The two example graphs are given as follows.

$$G_a = (\{A, B\}, \{(A, B), (B, A)\}, \{(A, 1), (B, 2)\})$$

$$G_b = (\{A, B, C\}, \{(A, B), (B, A), (A, C), (B, C)\}, \{(A, 1), (B, 1), (C, 3)\})$$

(c) communication graphs noted mathematically

Fig. 1. Communication graphs. A communication graph consists of locations, the number of users at each location, and directed communication channels between locations. Illustration (a) shows a complete communication graph between two locations. In this graph, the two users who are at one location can communicate naturally with one another; the communication with the user at the other location, however, is always restricted by the communication technology that is utilized. In illustration (b), an incomplete communication between three different locations is shown. This type of communication is found, for instance, in television news broadcasts. In illustration (c), a more mathematical notation of simple communication graphs is given along with the example communication graphs G_a and G_b.

audience is distributed over the world and can be seen by the lecturer, the corresponding CG is different from a CG where the audience is in one room and there is no back-channel. If there are multiple users at one location of the CG, they can communicate without restriction. This is the latter example of the audience being in the same room.

If there is a communication channel between every pair of locations, then the CG is said to be *complete*. Most desktop-sized conferencing concepts have a complete CG. Incomplete CGs can be found in one-to-many broadcast concepts.

4.2 Communication Notation

A CG is a graphical representation of communication. To have a short textual description of a CG, we propose *communication notation* (CN). CN cannot describe all possible CGs, but we think that its simplicity makes up for this drawback. For instance, the Chinese whispers game played by children cannot be described in CN.

Each location in CN is denoted by its number of users $(0, 1, 2, \ldots)$ or capital V (for virtual location). The binary communication operators $\leftarrow$, $\leftrightarrow$, and $|$ can be used to describe the communication between two sets of locations. The symbol $\leftarrow$ means directed communication, the symbol $\leftrightarrow$ means mutual communication, and the symbol $|$ means no communication. These operators take two CGs and result in a combined CG. Communication is defined between the two sets of locations from the right and left CG operand, respectively. The type of communication is given by the operator.

The operators can be used to build terms. If the same operator repeats a number of times, then a power notation can be used, whereby the operator precedes the exponent (e.g., $(1 \leftrightarrow 1 \leftrightarrow 1) \equiv 1^{\leftrightarrow 3}$). To better illustrate the notation, a number of example CGs with corresponding CN terms are given in Fig. 2. The precedence of CN operators is defined in Appendix A.2. However, in order to ease the understanding of CN terms, we encourage the usage of explicit parentheses.

At first, this special notation for communication may seem pedantic and its usage considered tedious. However, we will show in the following two sections that CN resolves ambiguities and greatly simplifies enumerating certain classes of communication. This, in turn, helps developers to give concise descriptions and to explore the possible communication designs of tele-immersion systems in a systematic way.

4.3 The One-to-Many Ambiguity

For tele-immersion concepts, there is a confusion about the meaning of *one-to-many*. Does one-to-many refer to users or locations and is communication unidirectional? We term this confusion the *one-to-many ambiguity*. A clear definition is especially important in tele-immersion because the communication design influences technical and spatial considerations. For instance, the immersion of multiple users at the same location requires display and tracking technology to simultaneously generate and present a different set of perspectives to each user, which in turn may influence the choice of light field representations. CN resolves the one-to-many ambiguity.

In CN, the term $3 \leftrightarrow 1$ describes a communication design with a group of three users (e.g., an audience) at the first location and one user (e.g., a lecturer) at the second location. On the contrary, a communication design with all three audience users and the lecturer at four different locations is noted as $1^{\leftrightarrow 3} \leftrightarrow 1 \equiv 1^{\leftrightarrow 4}$ in CN. In addition, the CN terms for both cases state that each user can communicate with each other user. The two cases have complete communication.

If the lecturer cannot see the audience—if there is only unidirectional communication—then the CN term for the first case becomes $3 \leftarrow 1$. In the second case of unidirectional communication to three locations having one user each, the CN term becomes $1^{|3} \leftarrow 1$. This is a typical broadcast scenario where there is also no communication within the audience.

For example, in order to resolve the discussed ambiguity, the authors of [15] describe their system as a group-to-group tele-immersion system. Using CN, we can say concisely that the usability experiments in [15] were conducted with a $3 \leftrightarrow 2$ communication design whereas the system itself allows a higher $6 \leftrightarrow 2$ communication design. The authors can present simultaneously up to twelve perspectives (stereo) to six users at the first location [14].

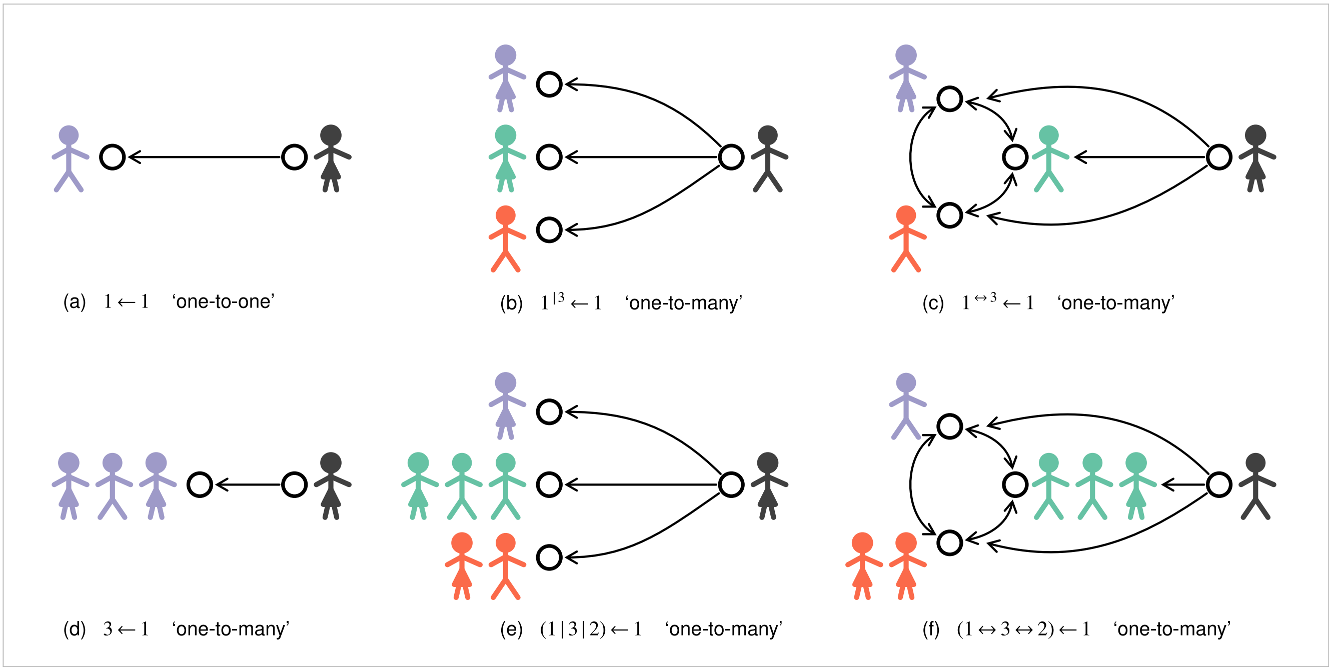


Fig. 2. Communication notation. Each of the illustrations (a-f) shows a communication graph and the corresponding term in communication notation. The top row shows communication with one user per location and in the bottom row there are one or multiple users per location. The illustrations (b-f) show the one-to-many ambiguity. For tele-immersion systems, the communication shown in (d-f) implies the immersion of multiple users simultaneously at one location.

4.4 Communication Types

A specific CG defines one concrete *communication type*. Communication types are enumerable because CGs are enumerable. CN can not be used to enumerate all possible communication types because not all CGs are expressible within CN. However, CN can be used conveniently to enumerate certain *classes of communication types*, which can help developers to explore the possible communication designs of tele-immersion systems systematically.

By using variables like m or n in CN terms, one can define easily a set of CGs and, thereby, class similar communication types. Variables can be used to denote a variable number of users or a variable number of locations. For the latter one, the variable is used in the exponent of the power notation.

For instance, there are six classes for unidirectional one-to-many communication types in CN: (a) $1 \leftarrow 1$ (b) $1^{|m} \leftarrow 1$, (c) $1^{\leftrightarrow m} \leftarrow 1$, (d) $n \leftarrow 1$, (e) $n_i^{|m} \leftarrow 1$, and (f) $n_i^{\leftrightarrow m} \leftarrow 1$. In order for these classes not to overlap, we assume $m, n > 1$ and $\sum n_i > m$. A concrete communication type from each class is given in Figs. 2a, 2b, 2c, 2d, 2e, and 2f, respectively.

To illustrate the expressiveness of CN, we give a more complex class of communication types, which is defined by the CN term $n_i^{|m} \leftarrow (2 \leftrightarrow 1)$. Interestingly, this example class is not of a purely academic nature but exists in everyday life, albeit in the context of television. Imagine two news presenters on television talking to one reporter who is at the scene of the action. The scene is transmitted to a number m of television sets and watched by a different number n_i of users at each location i.

For a tele-immersion concept, the channels of its communication type define which locations extend virtually to which other locations. Communication over these channels takes place via scenes that are extracted at some location and sent to other locations. For instance, we know that scenes of the reporter in the last example are sent to the television studio as well as to each television set. What can neither be expressed in CN nor by a CG is how these scenes of the reporter are integrated. A structured principle to describe how scenes can be extracted and integrated is the subject of the next section.

5 Immersion

Marvin Minsky described telepresence coming from a teleoperated robot scenario [1]. He envisioned the remote presence—the robot—to be able to physically change a remote location. At the same time, the user who operates the robot is immersed with all senses such that he feels 'being there.' Minsky's description of telepresence is actually two-fold: the robot 'is present' and the user 'feels present'. Additionally, it is asymmetric; there is a human at one location and a robot at the other. Interestingly, this is at the very heart of our general principle to describe the integration of two locations with one another. Putting it in Minsky's scenario, we describe the integration of the robot and the immersion of the user each in isolation albeit with the same idea of what we term *immersive extent*.

5.1 Immersive Extent

We use the term *immersive extent* to refer to the immersive link between two locations. Figuratively, a location extends immersively to another location. Immersive extent is directional and defined per communication channel of a CG. Hence, for each communication channel a different type of immersive extent can be defined.

We take an approach that is based on spatial volumes to characterize immersive extent. Although this characterization of immersive extent is static, we find it more intuitive to explain it as being the result of a three-stage process. The three stages are (i) *scene extraction*, (ii) *scene warping*, and

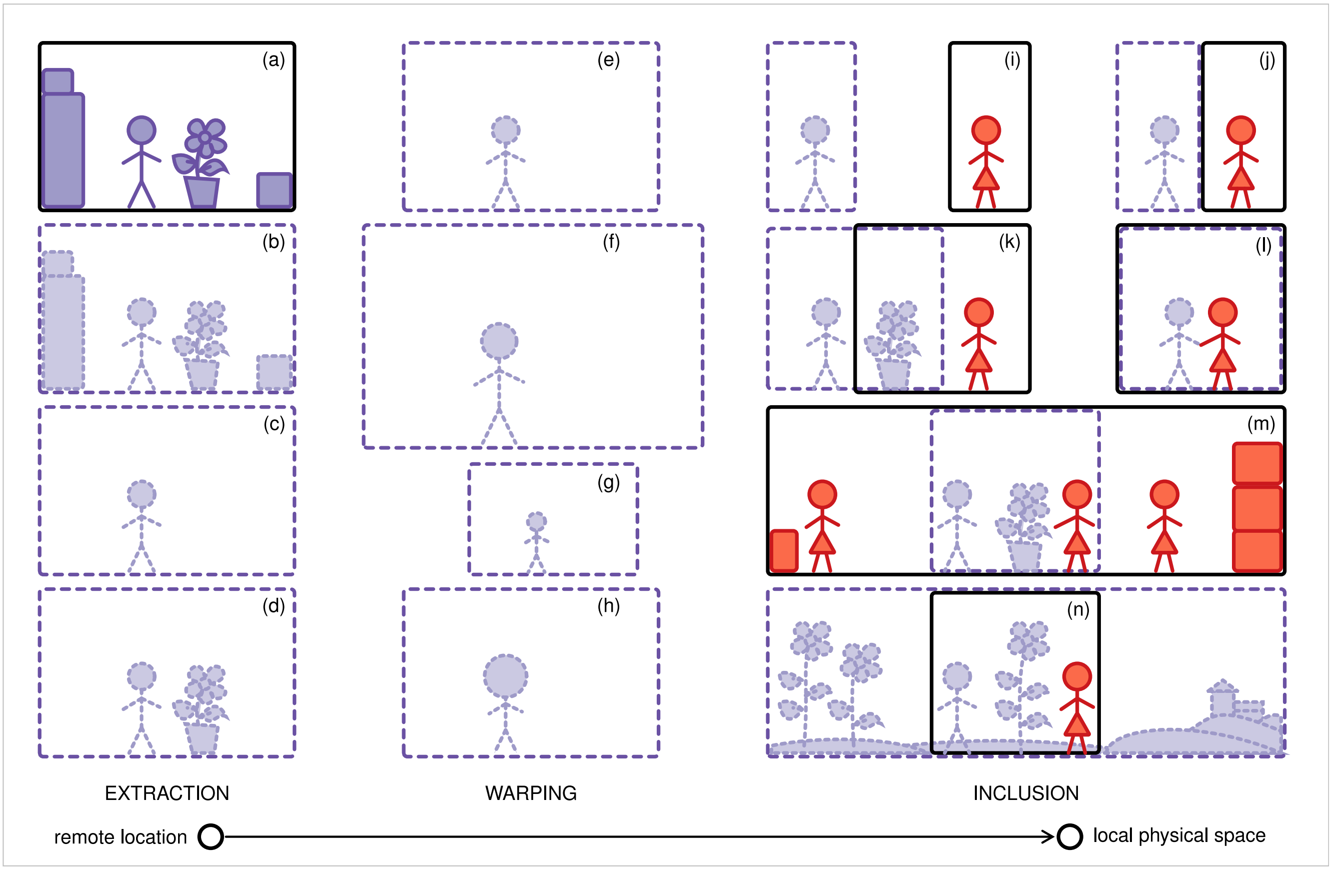


Fig. 3. Immersive extent. Directed immersion is characterized conceptually by the notion of immersive extent. Immersive extent is the result of the extraction, warping, and inclusion of a remote scene into another location, which is illustrated in the corresponding three columns. Scene extraction is shown in the illustrations (a-d); (a) is a remote physical location, (b,c) show the location extraction type (all segments) and user extraction type (user segments), and (d) is an arbitrary extraction example (user and plant segments). The second column of illustrations shows the four different types of scene warpings: (e) identity, (f) upscale, (g) downscale, and (h) non-similarity. The six different types of scene inclusion are: (i) disjoint, (j) adjacent, (k) partial, (l) perfect, (m) subset, and (n) superset. Remote scene and remote location volumes are depicted non-solidly in blue; physical location volumes (remote and local) have a solid black boundary. Note, immersive extent is always defined directionally for one communication channel. Hence, the female stick figures' locations in illustrations (i-n) may have completely different types of immersive extent in the opposite direction.

(iii) *scene inclusion*. The stages are depicted in the illustrations in the three columns of Fig. 3. In the following paragraphs, we describe the principle of each stage along with a semantical differentiation between types. These types combined constitute our systematic typification of immersive extent.

5.1.1 Scene Extraction

The scene extraction stage digitalizes a scene from the remote location. In terms of spatial volumes, this stage defines a *scene volume*, which must be a sub-volume of the *location volume*. The location volume is usually defined as bounded by, for instance, the dimensions of the room or the camera frustum. However, the location volume can also be defined as (partially) unbounded, which is useful for outdoor locations. A user can move and interact physically within the boundaries of the location volume.

In order to define types for the extraction stage, we assume a semantic segmentation of the location volume. That means the physical volume is partitioned in meaningful sub-volumes. In more technical terms, we define a labeled segmentation of the location volume; the set of possible labels must be finite. The labeling semantics can be chosen application specific. Possible labels for sub-volumes are, for instance, user, head, left hand, a specific bone, right input device, table, tree, or a voxel position.

With regard to this work, we decided to differentiate semantically only between head, body, and background segments. We found that this coarse and application-agnostic differentiation is sufficient to explain the principle and review the existing literature on tele-immersive systems. However, the aforementioned examples show that other more fine-grained semantic segmentations can be defined easily.

The definition of types for the scene extraction stage is based on the possible combinations of labels. In our case, we name three scene extraction types explicitly: location {head, body, background}, user {head, body}, and head {head}. We omit the scene extraction types that correspond to the remaining combinations like (head, background) because they are less relevant. The scene extraction stage is depicted in Figs. 3a, 3b, 3c, and 3d.

5.1.2 Scene Warping

The scene warping stage performs a spatial distortion of the complete location volume along with the contained extracted scene volume. Note that by defining the same distortion for the extracted scene as well as for the volume that defines the physical space, we keep a notion of the remote location's boundaries. If, for instance, we shrink the extracted user, we shrink the room as well. The spatial distortion is defined by a *warp mapping*.

The most common choice is to keep scenes in life-size and not to distort space. In this case, the warp mapping is the identity. To enlarge or shrink scene and location volume are other common choices. Other distortions exist as well; however, they are less common. Hence, we differentiate the following four types of scene warping: identity, upscale, downscale, and non-similarity. The scene warping types are depicted in Figs. 3e, 3f, 3g, and 3h. Identity, upscale, and downscale scene warping are similarity transformations. All other possible spatial distortions of the scene are grouped under the non-similarity scene warping type.

5.1.3 Scene Inclusion

The scene inclusion stage places the warped remote scene volume relative to the local location volume. This placement is defined by a position and an orientation. Thus, the *include mapping* is an euclidean transformation between warp frame and local frame.

Implicitly, the include mapping defines as well where the boundaries of the remote location volume are placed. An interesting question that arises is what happens with respect to the boundaries of the local location volume? Do the two volumes share parts of their boundaries? Do they overlap? Completely? These kinds of differentiations are conceptually important because they may constrain the ways in which users interact with each other and with the remote location.

In order to differentiate systematically, we define the following six types of scene inclusion: disjoint, adjacent, partial, perfect, subset, and superset. Figs. 3i, 3j, 3k, 3l, 3m, and 3n depicts all six scene inclusion types, respectively. These types resemble possible spatial set operations including the explicit special case of adjacent inclusion, which means that only parts of the two boundaries are in the intersection. A mathematical definition is given in Appendix A.3. Note that this typification makes scene warping and scene inclusion independent with respect to the important conceptual notion of boundaries.

The three stages—scene extraction, scene warping, and scene inclusion—in combination conceptually define directed spatial immersion. We introduce a semantic segmentation of the remote location volume and then we introduce warp and include mapping as a factorization of a *joint mapping* from remote to local space. This allows us to define types for each of the three stages. The different types are complete for each stage. Hence, in combination, we can typify each immersion concept that is formulated with our general principle. The conceptual description principle for immersive extent and the typification are general enough to be used for indoor and outdoor tele-immersion scenarios. We illustrate each part of the description of immersive extent in Fig. 3.

A user can touch an object of a remote scene if and only if the space of the physical location and the perceived remote scene volume overlap. We term this intersection volume *directly shared space*. A directly shared space exists for partial, perfect, subset, and superset scene inclusion. It is degenerated for adjacent scene inclusion (part of boundaries) and it does not exist for disjoint scene inclusion. The *direct manipulation* of objects in the perceived remote scene is possible only within the *directly shared space*. Outside of the directly shared space an indirect manipulation technique has to be applied (see, for example, [22] for such an indirect manipulation technique).

5.2 Quality of Immersive Extent

Immersive extent introduces a causal relationship. A remote scene that is included at a location manipulates that location. For instance, the display of a remote scene influences the thoughts of the user who perceives it. Our description principle for immersive extent describes immersion conceptually using a volumetric approach. We did not specify how the volume is represented.

Typically, in VR technology-based tele-immersion systems, we display a light field; whereas, in robotic tele-immersion, a 'matter display'—the robot—represents the remote scene volume. Both types of systems introduce the same causal relationship. Unarguably, however, the quality of that relationship is different because, unlike a light field, a robot can manipulate a location mechanically. We can differentiate this quality of the scene representation by saying that a location or user has either *virtual extent* or *material extent* with respect to another location.

A physical location can also manipulate the representation of an included scene. For instance, using VR input devices, a user can manipulate a displayed three-dimensional scene or cause a robot to fall over. In order to look at this aspect of tele-immersion systematically, we can use our qualitative differentiation between *virtual* (V) and *material* (M) not only for scene representations but for locations as well. Thus, a physical location is also considered to be a *material location*; virtual locations have already been introduced in Section 4.

Combinatorially, there are four cases of manipulations between virtual (V) and material (M): (i) V ⇢ V, (ii) V ⇢ M, (iii) M ⇢ V, and (iv) M ⇢ M. We used dashed arrows to differentiate from CN. We first look at the cases that become relevant when a scene representations manipulates a location: cases (i), (ii), and (iv). Note that virtual locations cannot contain material representations and, therefore, case (iii) is irrelevant. A virtual representation can easily manipulate a virtual scene (done in software) and the same can be said about a material representation manipulating a material (physical) location (e.g., a robot lifting an object)—cases (i) and (iv). The (mechanical) manipulation of the physical world with a virtual representation, case (ii), is a problem well known from classical VR. Though, there are, for example, all sorts of haptic and tactile devices, the possible manipulations are weaker. Second are the cases that correspond to the opposite direction, that means a location manipulating the remote scene representation: cases (i), (iii), and (iv). Here, for the same reason as above, case (ii) is irrelevant. In this direction, the manipulation of the virtual representation by a material scene—case (iii)—is easier to implement (for instance, by using depth-sensing cameras). Again, cases (i) and (iv) are easy.

Manipulations, as described so far, take place at one location between a remote scene representation and the location. If the remote scene representation can be manipulated, it can be considered detached from the originating remote location. Naturally, we expect a manipulation of a remote scene, however, to extend back to the remote location. Then,

(a) VR-type causal extent loop (MVVM)

(b) robot-type causal extent loop (MVMM)

Fig. 4. Causal extent loops. We differentiate scene representations coarsely into virtual (V) and material (M) and do the same for locations; physical locations are material locations. This qualification of scene representations and locations can be used to typify the *causal extent loop* as depicted in illustrations (a) and (b). A causal extent qualifies the interaction between locations and scenes. For instance, in the robot-type causal extent cycle MVMM in illustration (b), the interaction between robot and location is a real physical interaction. The qualification of representations can be refined application-specific.

at the remote location the manipulation is performed by a scene representation of the local location. Consequently, the manipulations of the remote location extend back to the local location and so forth. We term this *causal extent loop*. Causal extent loops are illustrated in Fig. 4. Note that a causal extent loop is not a loop (or cycle) in the underlying communication graph but grasps the concept of manipulations between scenes and locations.

For mutual extent we can use a simple four-letter notation for causal extent loops. The first and last letter qualify the location and the two middle letters qualify the scene representations. Hence, VR-technology-based tele-immersion systems' causal extent loops are noted MVVM and robot-based systems can be noted by MVMM or MMMM. A classical fully immersive $1 \leftrightarrow V$ VR system has an MVVV causal extent loop. The notation omits the backchannel half of the loop. For directional extent we can use an underscore for the missing representation. Hence, MV_M denotes the causal extent of a virtual reproduction of a (real) remote scene. This can be, for instance, an immersive underwater monitoring system with $n \leftarrow 1$ communication type.

By qualifying representations and their relation, we can describe what is going to happen on (qualified) causal extent loops. The authors of [23] use the term 'mechanical presence' for manipulations of type M $\dashrightarrow$ V. We are aware that the described qualification of scene representations and locations is very coarse and becomes easily blurred when considering, for instance, different VR technologies for different sensory modalities. We argue similarly like in the case of coarse scene extraction types. The given differentiation is sufficient to describe the principle and to review the existing literature. More fine-grained semantic differentiations—for instance between acoustic, optic, (passive) haptic, and material representations—can be defined easily depending on the concept application. Note that the qualification of representations can be used as well without spatial context.

5.3 The Complete Picture

In order to show the complete picture, we will summarize the abstractions presented so far. Remember that our main intent is the systematic description of possible tele-immersion concepts. Whereby changes in one technological component of a concrete system influence a variety of properties, in a conceptual description, ideally, changes in one conceptual part do not affect other parts of the concept. Thus, we also want the systematic description principle to be orthogonal.

Each part of the presented tele-immersion concept description principle—communication (communication graph structure, number of users), spatial immersion (extraction, warping, inclusion), and qualification of representations (scenes, locations)—can be changed without (or with only minimally) affecting other parts. The parts can be used in isolation to concentrate on a specific aspect of tele-immersion or they can be used in combination to describe and typify a complete tele-immersion concept.

The complete type of a tele-immersion concept is given by its communication graph (possibly noted in CN), the spatial immersive extent type (a combination of extraction, warping, and inclusion types) for—and this is important—*each* directed edge (communication channel) of the communication graph, and a qualification of the immersive extent of each directed edge. We propose a set of types for each part; however, we think that for specific applications more fine-grained and different, more specific types can be beneficial and easily defined.

We chose to describe the spatial immersion part of the description principle using the analogy of a pipeline of stages. Note, however, that the result is a static description of immersive extent. It can be thought of as being a description of one specific moment in time. For some tele-immersion concepts the type of extent stays the same over time and for some concepts it changes. We will come back to this in Section 7.

We use an extended version of Minsky's robotic tele-immersion scenario [1] to give a complete example in Fig. 5. Minsky envisioned the user of the tele-operated robot to be fully immersed in the remote scene; thus, this remote scene of the robot's location has superset inclusion. The user can act in the directly shared space, and his interactions are issued to the robot. He has direct mechanical presence. Two

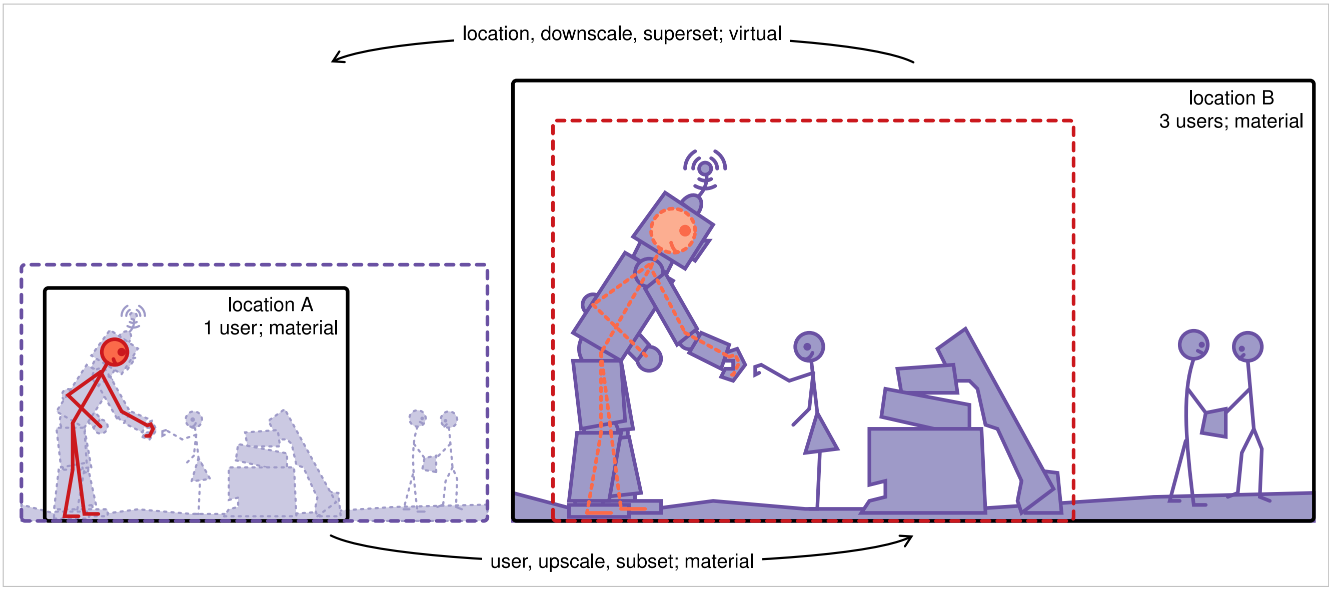


Fig. 5. Minsky's telepresence scenario. This illustration shows an extended version of Minsky's telepresence scenario and our typification of the underlying tele-immersion concept. The one user at location A is immersed in the surrounding of the tele-operated robot that is at location B. The three users at location B can communicate with the robot; thus, we have a $1 \leftrightarrow 3$ communication type. The immersive extent from B-to-A has location extraction, downscale warping, and superset inclusion (the dotted blue remote scene volume is a superset of the black location volume). The reverse communication channel A-to-B has user extraction, upscale warping, and subset inclusion (dotted red user within the robot). The causal extent loop is of type MVMM. The Beaming system of [20] has a very similar asymmetric concept. In the concept of Beaming, both channels have identity warping. Additionally, at each location, a scene of a virtual location is integrated. Note that the example in this figure combines the concept parts illustrated in Figs. 1, 2, 3, and 4.

examples for immersive extent in science fiction can be found in Appendix B.

6 Properties

The general tele-immersion description principle enables the definition and analysis of concept properties. Such properties are the subject of this section. We describe how conceptual immersion relates to different conceptualizations of presence, we define spatial consistency, and we give different definitions of symmetry.

6.1 Tele, Immersion, and Presence

Tele-immersion systems change the sense of presence. Mel Slater's note [24] defines presence in VR as being a response to immersion. He gives a powerful analogy from color theory: a specific visible light spectrum is the stimulus and color perception is a (user-specific) response. The term *presence* itself is characterized as a user's feeling to be in a coherent 'place'. This corresponds to the conceptualization of '*presence as transportation*'—one of the six different conceptualizations of *presence*, which are given in the thorough analysis of Matthew Lombard and Theresa Ditton [25]. We will characterize the conceptual immersion stimulus with regard to the provoked *location of presence*.

6.1.1 Immersive Extent Numbers

Immersive extent defines the proportions of (possibly multiple) remote scenes and a user's physical location. For instance, superset inclusion of a complete remote scene will lead to other proportions than the subset inclusion of only one remote user. Conceptually, we term the perceivable proportion of a particular remote scene the *extent number* of that scene. If this number is close to 1, then, conceptually, the local user is immersed almost completely in the corresponding remote scene, which is, for instance, the case for a classical $1 \leftrightarrow V$ fully immersive VR system. If the extent number is close to 0, then the corresponding remote scene constitutes only marginally to the combined location. We can also define an extent number for the location—it is one minus the extent numbers of all included scenes. Even though here we use extent numbers as an abstraction, note that it is possible to define actual numerical values for concrete systems. In Fig. 6, concrete extent numbers can be derived, for instance, from the red, green, and blue areas that illustrate different remote scenes.

6.1.2 A Color Perception Analogy for Tele-Immersion

We can give a Color Perception Analogy similar to that of Mel Slater [24]. Perceived locations are analogous to colors. In classical fully-immersive VR, there are two main location stimuli—real and virtual world, two colors—and one question is, which one is perceived by the user. Tele-immersion is like mixing colors. Your physical location has and every remote location has a color spectrum. By mixing parts of those color spectra you get a new spectrum. Extent numbers correspond to the proportions of the mixed spectra. Figuratively, one may say that a tele-immersion concept description principle defines a color space and a typification defines a color palette.

6.1.3 Immersive Location Shift

Depending on the extent numbers, the local user will feel more likely either that the remote scene is part of her or his location or that she or he is present in the remote scene. Figuratively, the difference in provoked presence consists of the local user 'staying here' or 'being there.' Or, one may

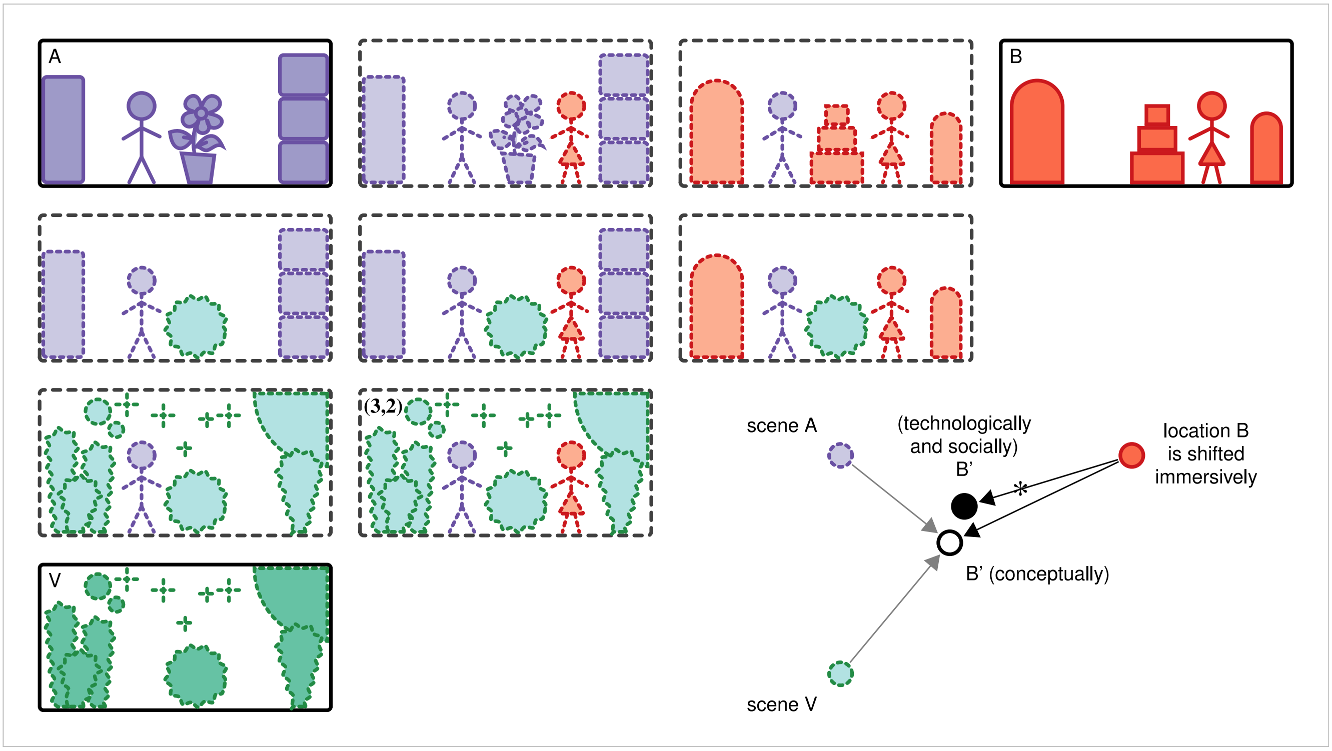


Fig. 6. Immersive location shift. By extracting, warping, and including remote scenes into a user's physical location, tele-immersion changes the perceived location. In this figure, there are three locations: Two physical locations (A and B) and one virtual location (V). By extracting different spatial segments from each location and combining them, new locations are created. We use extent numbers to refer to the perceivable proportions of locations with respect to each other. For instance, the combined location **(3,2)**—row 3, column 2—has a high immersive extent number for V and equally low immersive extent numbers for A and B. Conceptually, the users are at a different location. We term the difference between a user's location and the immersively altered location *immersive location shift*, which is illustrated on the bottom right using a '(locational) immersion-presence diagram'. We illustrate that an actual technological and social immersive location shift of location B will differ from a conceptual immersive location shift. The presence of the user is depicted by the $*$. The position of the $*$ should not be interpreted as a new location but as a likelihood of acceptance 'to be there' between the physical location and the immersively shifted location.

describe the perceived presence of the remote user as 'being here' or 'staying there.' The corresponding remote scene extent numbers are low and high, respectively. In both cases, the sense of presence in the location is altered. Remote objects or users are made present locally or the user feels she or he has traveled to another location. We term the stimulus that leads to this alteration *immersive location shift*. The immersive location shift is directional from the user's location toward the remote scene's location. The length of the immersive location shift depends on the extent number ratio, how different the involved locations are, and how they are combined. Immersive location shift is illustrated in Fig. 6. The joint location of a robot-type MVMM tele-immersion system, for instance, will always be shifted toward the (physical) location of the robot. Now imagine that the two distant rooms are actually identical. Then, the length of the immersive location shift will be small. For identity scene warping and perfect scene inclusion, the immersive location shift may become (conceptually) unnoticeable. A mathematical analogy in terms of vectors is illustrated in the lower right half of Fig. 6. Keep in mind that an actual location shift depends not only on spatial stimuli but on technological and social factors as well.

6.1.4 Quality of Immersion

The immersive qualification of scene representations adds extra dimensions to the space of possible immersive stimuli. A material scene representation, for instance, will create a different stimulus than a scene representation using only spatial audio. Jonathan Steuer [26] introduced the term *vividness* for a 'representational richness' axis. Conceptually, a (close to perfect) material representation (M) is more vivid than a virtual representation (V). Think of VR-enhanced robot versus VR instead of robots versus VR. Certainly, the conceptual quality of immersion will influence the level of user response. However, it does not influence the conceptual spatial location that is formed by combining the spaces. In other words, we differentiate the spatial part of the stimulus from the quality part. These parts correspond, respectively, to the conceptualizations of 'presence as transportation' and 'presence as social richness' (of a medium) in [25]. The quality of immersion that is delivered using a concrete technology is described as 'presence as realism' in [25]. In tele-immersion, one way of thinking about 'presence as realism' is in terms of the perceptual difference between a scene extracted from a physical location and the location itself.

6.1.5 Joint Social Attention

An important aspect of tele-immersion is that it also conveys spatial non-verbal cues like eye contact, deictic gaze, and deictic gestures, which create a more natural communication. For instance, Stephen Shepherd points out that the deictic gaze is a strong (and partially reflexive) stimulus for joint social attention and provides a 'window into the mind' [27]. In our concept description principle, the transmission of these spatial non-verbal cues is given inherently by the

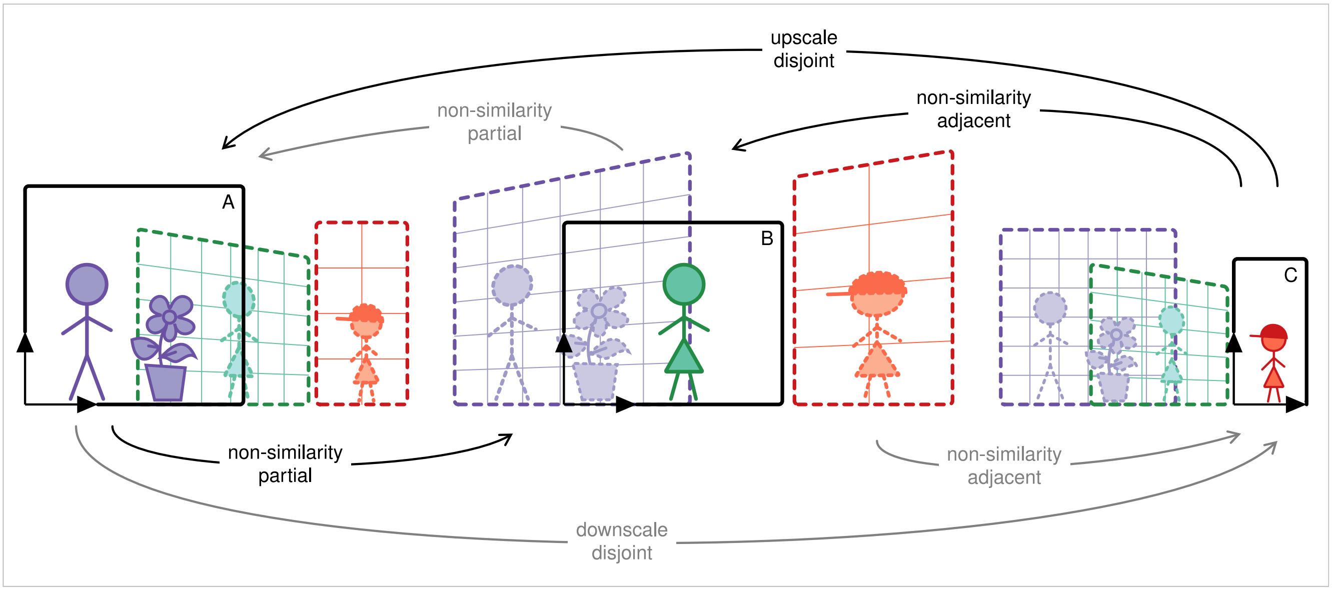


Fig. 7. Spatial consistency. This example illustrates spatial consistency between three locations A, B, and C within a $1^{\leftrightarrow 3}$ tele-immersion concept. Spatial consistency between two locations means that the spatial transformation from one location to another and back must be the identity. For instance, the scene of location A (man and plant) is downscaled in location C (child) and a scene of C is upscaled in location A again. To demonstrate non-linear warpings, we use homographies in projective 2-space $\mathbb{P}^2$ [28]; thus, the distorted scenes in the illustration are to be interpreted as transformations in the plane and not as perspective drawings. To get a better insight, we can think about what happens if the woman at location B walks in the direction of the virtual plant. In order to keep spatial consistency, the woman gets taller in location A and B. Or, what happens, if the child in location C decides to downscale herself? Scenes A and B at location C are upscaled and the scenes of C at locations A and B are downscaled. Breaking spatial consistency may sometimes have advantages. Shared task spaces, however, should always be spatially consistent.

presentation of a remote spatial volume. The differentiation between different (spatial) gestures, however, is not explicit.

We can change this by giving a different semantic segmentation. Instead of extracting head and body segments, we extract, for instance, an (abstract) skeleton that is parametrized by line segments, (the eye-ball orientation can be parameterized by lines, too). Note that this is still a volume segmentation (though degenerated). Based on this semantic segmentation, we can define different types of supported (spatial) cues like head orientation, arm configuration, body posture, or eye orientation. These types can be used to delineate stimuli that are known to affect *social attention*. Essentially, this low-rank spatial parametrization of the human body is the one used in classical avatar-based VR-research. In addition, in the inclusion stage, we can differentiate between, for instance, placing users side by side, mirrored, or opposite one another. Seating geometry is known to influence social attention.

Our tele-immersion concept description principle has a strong focus on the spatial aspects of tele-immersion. Spatial immersion is a strong stimulus; however, forms of presence are induced by a number of factors. Though our description principle is not focused on delineating those factors, we think that it can form a good basis for discussion because it contains the necessary entities. For example, for the conceptualization of 'presence as social actor within medium' [25], we can attribute virtual users at virtual locations. A thorough discussion of all factors that influence social awareness is beyond the scope of this work.

6.2 Spatial Consistency

By considering extent for each direction in isolation, we can delineate a great variety of tele-immersion concepts within our general principle. However, not all such concepts enforce spatially consistent virtual worlds.

Consider, for instance, the illustration of Minsky's telepresence scenario in Fig. 5 and imagine that instead of shrinking the scale of scene B in location A, we would keep scene B the same size. This would clearly break consistency because the robot at the physical location B is about twice as tall as the female stick figure. Breaking spatial consistency is clearly not useful in this case because the user's interactions with scene B at location A will not match the robot's interactions at location B.

For warping and inclusion to be *spatially consistent*, the joint transformation $f_{A\to B}$ from A to B must be the inverse of the joint transformation $f_{B\to A}$ from B to A: $f_{A\to B} = f^{-1}_{B\to A}$. This consistency can be defined as well for communication graphs with more than two locations. In short, the combined transformations of each cycle must form the identity. Spatial consistency is illustrated in Fig. 7.

We can transfer the idea of consistency for scene warping and scene inclusion to the concept typology. Scene warping and scene inclusion types, however, are defined independently with respect to the notion of boundaries, and we do not consider explicit transformations but classes of transformations. Hence, our notion of consistency must differ and be weaker. It is reasonable to require that the type of scene inclusion is the same for opposing extents except for superset and subset scene inclusion. Superset scene inclusion must be opposite to subset scene inclusion and vice versa.

The consistency definitions for warping types follow from our notion of consistency for transformations. The inverse of downscale is upscale warping and the inverse of a non-similarity is a non-similarity. Note that for more than two locations these types must cancel out on every (undirected) cycle in the communication graph (e.g., downscale, non-similarity, non-similarity). Spatial consistency with regard to warping and inclusion types is illustrated in Fig. 7 as well.

Though many tele-immersion concepts are spatially consistent, there are cases where breaking spatial consistency can be beneficial. One such example is video communication via mobile phones. Because of the relatively small displays, real-life faces are usually downscaled in both communication directions. We can clearly 'touch' the face, however, we do not expect our face to be reached by the miniature representation. This is spatially inconsistent because the user at the other location will say the opposite. A spatially more immersive example is easy to imagine using a technology that corrects for mutual eye contact (see, e.g., [19]) and an autostereoscopic mobile phone display.

In general, spatial consistency is needed if the concept requires a joint task space. When such a joint task space is not needed, we can break spatial consistency. Note that even if we break spatial consistency, we may decide to correct computationally for eye contact or deictic gestures—we can keep spatial consistency partially.

6.3 Symmetries

Tele-immersion concepts are often described as symmetric or asymmetric. Within the tele-immersion concept description principle and the typology, symmetry can be analyzed with regard to different parts. We restrict our considerations to two locations.

Within the general principle, the joint transformation that describes warping and inclusion can be analyzed regarding spatially symmetric properties. An existing categorization based on this approach are the interaction modes given in [16]. The 'mirror' interaction mode is an improper euclidean transformation, whereas the 'first-person' and 'third-person' interaction modes are proper euclidean transformations. We will see in Section 7.1 how to differentiate between the latter two using the notion of *self-inclusion*. Typologically, we require identity warping for both extents and their inclusion types must be equal. This definition implies consistency between the two locations and, regarding the general principle, restricts symmetry considerations to symmetries of euclidean transformations.

For extraction, it is possible to define symmetry on a semantic level, which means that objects of the same types must be extracted from each location. Hence, we need the same type of semantic segmentation at both locations. For instance, if furniture is extracted at one location, it also needs to be extracted at the other location. With regard to our typology, the extraction types must be equal. Another definition for symmetry of extraction can be given on a quantitative level in terms of extent numbers. Then, for opposing extractions to be symmetric, the corresponding extent numbers must be the same. In other words, the 'portion of extracted scene' must be the same. This definition relates *quantitative extraction symmetry* to conceptual immersion. Interestingly, the 'technological asymmetry' of the Beaming concept [20] creates an *immersion asymmetry*, which in turn relates directly to the quantitative definition of extraction symmetry.

The second asymmetry in the Beaming concept is 'social asymmetry'; one fully-immersed user is 'beamed' to a group of users. This *communication asymmetry* can be formulated between two locations easily. Depending on the number n of users at a location, we label the location with either none ($n = 0$), one ($n = 1$), or group ($n \geq 2$). Then, the communication between two locations is symmetric if the corresponding labels are equal; otherwise, the partition is asymmetric. These labels can serve as well to partially dissolve the one-to-many ambiguity, which was used in [15]. Note that communication symmetry is only defined for bidirectional communication.

In summary, we have shown that the proposed description principle can be used to analyze concepts with regard to space, immersion, and communication symmetries.

7 Enhancements

In this section, we discuss briefly a number of possibilities to enhance the general principle. The enhancements are based on modifications and attributions of the general principle's parts.

7.1 Multi and Self-Inclusion

Some tele-immersion concepts include representations of users into their own physical space, which we term *self-inclusion*. For instance, *self-inclusion* is used for interaction [16] and is known to positively affect presence [29] and cognitive load [30]. The C1x6 tele-immersion system [14], [15] not only self-includes user representations but also includes the virtual scene twice. The second included scene presents a world-in-miniature (WIM) to ease orientation.

To allow self-inclusion and *multi-inclusion* in our model, the partition graph must be defined as a multigraph with self-loops, whereby the extent is defined for each edge in the multigraph separately. Self-loops and multiple edges between two locations make it possible to describe segment-wise inclusion and break consistency. Interestingly, this kind of consistency breaking is used in philosophy to understand fundamental questions about our conscious self [31]. Self-inclusion is important as well for mixed reality, where the local scene needs to be extracted in order to place virtual objects correctly.

We have described how the three interaction modes in [16] can be differentiated partially by considering spatial symmetries. By additionally considering self-inclusion, we can differentiate completely: 'first-person' (proper euclidean, no self-inclusion), 'third-person' (proper euclidean, self-inclusion), and 'mirror' (improper euclidean, self-inclusion). The analysis shows that one interaction mode is missing to complete the enumeration: 'first-person mirrored' (improper euclidean, no self-inclusion). Conceptually, the interaction modes in [16] are given by considering spatial symmetry and self-inclusion.

7.2 Time Shift

So far, our tele-immersion concept description principle ignores time. Time can be considered with regard to communication and immersion. For instance, we can define a time shift per edge in the partition graph. Then, we have the possibility to discuss the different options to merge time-shifted remote scenes locally. The same problems arise in multiplayer online games, which have a $1^{\leftrightarrow m} \leftrightarrow V$ communication (with m being the number of players and V a virtual scene provided by the game server). Each connection between users has a certain latency, which makes keeping consistency challenging, especially for fast action. Although there are solutions for purely virtual worlds from the field of video games, in tele-immersion the need for consistency between virtual and physical worlds poses further restrictions.

7.3 Dynamic Immersive Extent

Without considering time with respect to extent, it is not possible to differentiate a concept that allows the user to walk around physically from a concept that implements walking virtually. In both cases, the type of extent stays the same; however, in the latter case, the remote scene translates with respect to the user's location, whereas in the case of physical walking, it stays the same—extent is relative to location.

A basic typification for movement can be defined by considering the transformations from user to location $f_{U \to L}$ and from location to scene $f_{L \to S}$. The user's position is either fixed (F) or it changes (C) while the user is walking. The same differentiation can be made for the location's position and orientation with respect to the scene. This results in the four basic cases of movements: FF, FC, CF, and CC.

In the first case, the user stays at a fixed position in the room and in the remote scene. The FC case corresponds to the user "floating" in the remote scene. Often, some kind of vehicle that is fixed with respect to the location is used to make this movement more plausible and to reduce cognitive load. The vehicle can be a material part of the location or included virtually. The CF case allows the user to walk around in the physical location and thereby to move in the included scene. The CC case corresponds to continuous movement redirection technics, such as redirected walking [32]. For this particular case, a good taxonomy that also differentiates between rotational and translational movements can be found in [33].

One possibility to refine this basic typification of movement is to differentiate changes—similar to [33]—into continuous changes (C) and discrete changes (D). This allows, for instance, the differentiation of redirected walking (CC) from the concept of flexible spaces (CD) [34], [35]. A movement concept with discrete changes in user position, typically, first 'freezes' the immersion—the user is indicated to take a new position and orientation with respect to the room—and then it 'unfreezes' the immersion and applies a continuous or discrete change that reflects the new user position in the included scene (see, e.g., [36]).

7.4 Warping, Morphing, and Mixing

In the warping stage of the immersive extent definition, we allow the scene to be spatially distorted. We can enhance this concept by also considering a change of appearance. Borrowing from image processing terminology, we can name such a combined stage *morphing stage*. In the morphing stage, extracted segments can be morphed into completely different objects.

When remote scenes are integrated into the local physical space, we implicitly assume that the remote scene content overlays the physical space. Usually, the local physical space will be empty within the areas where remote scene content is placed. In some cases, however, a more precise concept of mixing scenes and physical space might be interesting. For instance, in mixed reality, a blending of colors or materials could be used to alter the appearance of existing physical objects; the modified objects are self-included into the location.

7.5 Other Senses

Finally, we focus on the visual part of tele-immersion; the concept typology was designed with the visual sense in mind. For instance, we did not include a discussion of spatial audio, which is another important sense. A possible conceptual integration could be implemented by considering vibrations of extracted objects.

8 Concepts of Existing Systems

We use the tele-immersion typology to analyze the concepts of existing tele-immersion systems. Following our concept typology, we split the analysis into a first part for communication and a second part for extent. Whenever possible, we differentiate between the proposed concept and the actual implemented system prototype. Our analysis of existing tele-immersion system concepts is summarized in Table 1.

8.1 Communication of Existing Systems

The list of systems with one user per location is long, for it is technically easier to immerse only one user per location. An early system to do this was MAJIC [6], which placed three users at three different locations. In the Virtual Auditorium [38] system, for instance, there are many communication channels between students and instructor but each user is at his own location. Systems with multi-perspective renderings are the Teleport system [7], the im.point system [9], and the Coliseum system [8]. A more recent system is the Viewport system [51].

If multiple users are present at one location, their number is limited mainly by the immersive display technology. A simple workaround is to assign a separate screen to each user. An early example using the translucent mirror principle is the MPEC system [12]. Multi-display solutions are MultiView [13] and Informal [41]. Systems using autostereoscopic displays or glasses are 3DPresence [55], Encumbrance Free [49], and Beaming [20]. A room-sized implementation of a group-to-group system is C1x6 [15].

A number of tele-immersion systems include virtual locations. Virtual locations either resemble room geometry [7], [8] as a replacement, or they are purely virtual worlds [11], [15]. In the former case, users typically do not interact with the virtual scene, which is denoted by using the $\leftarrow$ communication operator in the table. The GrImage system [52] implements a complete physical interaction with the virtual world.

Most systems are scalable with respect to the number of locations. However, in shared virtual table systems, seating geometry becomes an additional limiting factor. Because of time and effort, the prototypic implementations seldom have more than two locations. For the same reasons, often only a communication channel for one direction is implemented.

8.2 Immersive Extent of Existing Systems

In the literature, the vast majority of tele-immersion systems has the same type of immersive extent for all

TABLE 1
Tele-Immersion Concepts of Existing Systems

System	Concept															
	Communication		Extent													
			Extraction			Warping				Inclusion						
			location	user	head	identity	upscale	downscale	non-similarity	disjoint	adjacent	partial	perfect	subset	superset
ClearBoard [37]	$1 \leftrightarrow 1$		•			•					•				
MAJIC [6]	$1^{\leftrightarrow 3}$		•			•					•				
MPEC [12]	$2 \leftrightarrow 2$		•			•					•				
Virtual Auditorium [38]	$1 \leftrightarrow 1^{\leftrightarrow n}$		•			•					•				
Holoport [39]	$1 \leftrightarrow 1$		•			•					•				
MultiView [13]	$3 \leftrightarrow 3$		•			•					•				
i2i [40]	$1 \leftrightarrow 1$		•			•					•				
Informal [41]	[a] $m \leftrightarrow n$		•						•			•			
3D TV [42]	$m \leftarrow n$		•			•						•			
Holovizio [43]	$m \leftarrow n$		•			•						•			
HeadSPIN [44]	$n \leftrightarrow 1$	$n \leftarrow 1$			•	•								•	
Situated Multiview [45]	$n \leftrightarrow 1$	$n \leftarrow 1$			•	•								•	
Office of the Future [2]	$1^{\leftrightarrow n} \leftrightarrow V$		•			•					•				
Teleport [7]	$1 \leftrightarrow 1 \leftarrow V$			•		•					•				
im.point [46]	$1^{\leftrightarrow 3}$		•			•					•				
3DPresence [46]	$2^{\leftrightarrow 3}$		•			•					•				
Coliseum [47]	$1^{\leftrightarrow 3} \leftarrow V$			•		•					•				
Terrascale [48]	$1 \leftarrow n$		•			•					[b] •				
TEEVE [16]	$1^{\leftrightarrow 2} \leftrightarrow V$		•			•								•	
Beaming [20]	$n \leftrightarrow 1 \leftrightarrow V$	$n \leftarrow 1$		•		•								•	
		$1 \leftarrow n$	•			•									•
Encumbrance Free [49]	$2 \leftrightarrow 2$		•			•					•				
FreeCam [50]	$1 \leftarrow n$			•		•						—			
Viewport [51]	$1^{\leftrightarrow 3}$			•		•					•				
Transparent [17]	$1 \leftrightarrow 1$			•		•								•	
GrImage [52]	$1 \leftrightarrow V$	$1 \leftarrow V$	•			•									•
TI Grid [53]	$1^{\leftrightarrow 3} \leftrightarrow V$	$1 \leftarrow 1$		•		•								•	
blue-c [11]	$1 \leftrightarrow 1 \leftrightarrow V$	$1 \leftarrow 1$		•		•							[c] •		
EWM [54]	$1 \leftrightarrow 1 \leftrightarrow V$	$1 \leftarrow 1$	•			•				•					
C1x6 [15]	$2 \leftrightarrow 3 \leftrightarrow V$	$2 \leftarrow 3$		•		•						•			
		$2 \leftarrow_{\text{WIM}} V$	•					•						•	
EDGAR [4]	$n \leftrightarrow 1$	$n \leftarrow 1$		•		•								•	
Surrogates [5]	$1 \leftrightarrow 1$			•		•								•	

The implemented tele-immersion concepts are analyzed regarding communication and immersive extent. For communication, we considered the communication as implemented in the prototype system. We do not describe extent for every channel of every system because this would go beyond the scope of this overview. Instead, we focus on the significant channels. [a] *We assume a personal camera for every display tile.* [b] *We concluded that from the images in the publication.* [c] *One possibility is if portals overlay perfectly.*

communication channels. However, often only one location has a fully implemented extent in prototypes.

Systems in the literature extract either only the user or the complete location. If only the user is extracted, a segmentation technique is applied. There are three main reasons for user extraction: (i) composition with another scene, (ii) data reduction, or (iii) estimation of physical boundaries. Corresponding example systems are (i) Teleport [7] and Transparent [17], where the user cut-out is rendered as impostor; (ii) Viewport [51], point-cloud size is reduced for processing; and (iii) GrImage [52], user boundaries are used for physically based user-scene interaction. The reasons for full location extraction are mainly (i) ease (no extra segmentation technique is needed) or (ii) the location itself is of interest. Multiview [13] and Terrascale [48] are corresponding example systems. Explicit head extraction can be found in the HEADSpin [44] and Situated Multiview [45] systems.

Most shared virtual table systems have adjacent inclusion. The remote participant is placed virtually on a shared table. Seating geometry is discussed, for instance, in [6], [12], [55], and [51]. The display separates the two rooms even if a non-immersive display technique is used, such as in the i2i system [40]. The shared boundary can be used to display and interact with two-dimensional content like in the ClearBoard system [37]. In a similar fashion, the space between disjoint rooms can be used to display three-dimension virtual content such as that proposed in the Extended Window Metaphor (EWM) Scenario [54].

Overlap of remote scene and location is typically implemented by systems with an immersive display technique. Prominent example systems are blue-c [11] and 3D-TV [42].

The former one uses shutter glasses and the latter one an autostereoscopic display. The HeadSPIN system [44] uses a special turning mirror-based autostereoscopic display. The system creates a subset inclusion, thereby allowing users to walk around the rendered remote head. The Transparent system [17] creates subset inclusion by projecting the user onto a transparent foil. If users are immersed completely in a virtual world, superset inclusion is typically used. The users meet in a virtual environment that extends over room boundaries.

We believe that the extraction of objects will become more widely used in future immersive telecommunication systems where collaboration becomes a major aspect. All of the tabularized systems have identity warping, except the C1x6 system [15], which has downscale warping for its world-in-a-miniature, and the Informal system [41], which has non-similarity warping. We expect future systems to apply a variety of different warping techniques to overcome technical limitations and to implement overview and detail techniques.

The majority of systems has a symmetric immersion concept. Immersive asymmetry is mainly found in systems that use robotic avatars such as EDGAR [4]. The asymmetry of the Beaming concept [20] is discussed in detail in Section 6.3.

9 Conclusion

We give a general principle and typology to characterize tele-immersion conceptually. We show how this principle and typology allow us to compare, enumerate, and analyze concepts of tele-immersion systems. Furthermore, we demonstrate how the principle helps to describe various aspects of tele-immersion concisely.

Based on the definitions of Mel Slater [24], Matthew Lombard and Theresa Ditton [25], and our general principle, we give a conceptual view of the link between spatial immersion and presence for tele-immersion. We think that this view can help others to reason about the various facets of presence. Ultimately, the feeling of 'being together' is a complex response to a variety of stimuli. This work is focussed on the conceptual spatial aspects of tele-immersion and shows how the principle can be used to systematically enumerate spatial stimuli for joint social attention.

Recent big leaps in virtual reality, computer vision, and robotics will most presumably bring many high-quality immersive hardware components to the consumer market within the next decade. We hope that this concept description principle will not only help to explore tele-immersion concepts systematically but as well lead to new systematic approaches to describe and advance research in mixed reality.

Appendix A
Definitions

A.1 Communication Graph

A communication graph is a triple $G = (P, E, u)$ of a finite set of locations P, a set of directed communication channels $E = \{(p, p') \mid p, p' \in P \wedge p \neq p'\}$, and a function $u : P \mapsto \{0, 1, \ldots\} \cup \{V\}$ that assigns a non-negative number of users or V to each location.

A.2 Communication Notation

Communication notation is a compact textual notation for a subclass of CGs. Let $x \in \{0, 1, \ldots\} \cup \{V\}$, then x is the communication graph $(\{p\}, \{\}, \{(p, x)\})$ with some single location p. Let $X = (P_X, E_X, u_X)$ and $Y = (P_Y, E_Y, u_Y)$ be two communication graph operands, then the binary communication operators $|, \leftarrow, \leftrightarrow$ result in the communication graphs formed by the following sets, respectively: $P_{|,\leftarrow,\leftrightarrow} = P_X \cup P_Y$, $u_{|,\leftarrow,\leftrightarrow} = u_X \cup u_Y$, $E_| = E_X \cup E_Y$, $E_\leftarrow = E_X \cup E_Y \cup \{(y, x) \mid x \in P_X \wedge y \in P_Y\}$, $E_\leftrightarrow = E_X \cup E_Y \cup \{(x, y) \mid x,\ y \in P_X \cup P_Y\}$. The communication operators can be used to build terms. The operator precedence from high to low is $\leftrightarrow$, $|$, $\leftarrow$. Parentheses can be used to change the precedence. The operator associativity for $\leftarrow$ is from right to left. A power notation can be used, whereby the operator precedes the exponent: $(1 \mid \cdots \mid 1) \equiv 1^{|n}$, $(1 \leftarrow \cdots \leftarrow 1) \equiv 1^{\leftarrow n}$, $(1 \leftrightarrow \cdots \leftrightarrow 1) \equiv 1^{\leftrightarrow n}$. Variables can be used to define sets of communication graphs (e.g., $m \leftarrow 1^{\leftrightarrow n}$). One may say that the 'ultimate goal of tele-immersion' is to create highly immersive $n_i^{\leftrightarrow m}$ systems, which induce a strong sense of presence between many users (n_i) at numerous locations (m).

A.3 Inclusion Types

The inclusion types that are described in Section 5 and Fig. 3 can be defined formally using set notation. Let R and L denote the remote scene volume and local location volume exclusive of their boundaries; the volumes' boundaries are denoted by R_B and L_B, respectively. The inclusion types are defined as follows: disjoint—$R \cap L = \emptyset \wedge R_B \cap L_B = \emptyset$, adjacent—$R \cap L = \emptyset \wedge R_B \cap L_B \neq \emptyset$, partial—$R \cap L \neq \emptyset \wedge R \neq L$, perfect—$R = L$, subset—$R \subset L$, superset—$R \supset L$.

Appendix B
Immersive Extent in Science Fiction

We believe the following two prominent examples from science fiction to be a good illustration of the generality of our ideas. In the Star Wars galaxy, a device named holoprojector is used to transmit and display a three-dimensional translucent representation of a remote person; the scene overlaps the physical location. The technology provides user extraction and subset scene inclusion. Often, the holoprojector shows an upscaled or downscaled scene. In the Star Trek universe created by Gene Roddenberry, the holodeck is used to simulate scenery including virtual characters (see, for instance, [56] for the physics of Star Trek). The immersive extent of the simulated scenery usually has superset inclusion because the scenery extends over the borders of the holodeck. The causal extent loop types for holoprojector and holodeck are MVVM and MMMV, respectively.

Acknowledgments

The idea of a tele-immersion concept description principle was sparked by an early discussion about the field with Malte Willert. Putting emphasis on orthogonality and completeness has been inspired by the work of Hans-Jörg Schulz and his colleagues on design space approaches in visualization. The anonymous colleagues who thoroughly reviewed the manuscript did a great job to help improve

terms and notations, description, and the focus of this paper. Finally, this work would not have been possible without all those people who dedicate themselves to build actual tele-immersion system prototypes.

References

[1] M. Minsky, "Telepresence," *Omni*, vol. 4, pp. 45–51, 1980.

[2] R. Raskar, G. Welch, M. Cutts, A. Lake, L. Stesin, and H. Fuchs, "The office of the future: A unified approach to image-based modeling and spatially immersive displays," in *Proc. 25th Annu. Conf. Comput. Graph. Interactive Techn.*, 1998, pp. 179–188. [Online]. Available: http://doi.acm.org/10.1145/280814.280861

[3] E. Paulos and J. Canny, "Social tele-embodiment: Understanding presence," *Auton. Robots*, vol. 11, no. 1, pp. 87–95, 2001. [Online]. Available: http://dx.doi.org/10.1023/A:1011264330469

[4] P. W. Ching, W. C. Yue, and G. S. G. Lee, "Design and development of edgar—A telepresence humanoid for robot-mediated communication and social applications," in *Proc. IEEE Int. Conf. Control Robot. Eng.*, Apr. 2016, pp. 1–4.

[5] A. Nagendran, A. Steed, B. Kelly, and Y. Pan, "Symmetric telepresence using robotic humanoid surrogates," *Comput. Animat. Virtual Worlds*, vol. 26, no. 3/4, pp. 271–280, May 2015. [Online]. Available: http://dx.doi.org/10.1002/cav.1638

[6] K.-I. Okada, F. Maeda, Y. Ichikawaa, and Y. Matsushita, "Multiparty videoconferencing at virtual social distance: Majic design," in *Proc. ACM Conf. Comput. Supported Cooperative Work*, 1994, pp. 385–393. [Online]. Available: http://doi.acm.org/10.1145/192844.193054

[7] S. J. Gibbs, C. Arapis, and C. J. Breiteneder, "TELEPORT - Towards immersive copresence," *Multimedia Syst.*, vol. 7, no. 3, pp. 214–221, May 1999. [Online]. Available: http://dx.doi.org/10.1007/s005300050123

[8] H. H. Baker, et al., "The coliseum immersive teleconferencing system," in *Proc. Int. Workshop Immersive Telepresence*, 2002, pp. 5–8.

[9] R. Tanger, P. Kauff, and O. Schreer, "Immersive meeting point," in *Proc. 5th Pacific Rim Conf. Advances Multimedia Inf. Process. - Volume Part I*, 2004, pp. 89–96.

[10] W. A. S. Buxton, "Telepresence," in *Proc. Conf. Graph. Interface*, 1992, pp. 123–129. [Online]. Available: http://dl.acm.org/citation.cfm?id=155294.155309

[11] M. Gross, et al., "blue-c: A spatially immersive display and 3D video portal for telepresence," *ACM Trans. Graph.*, vol. 22, no. 3, pp. 819–827, Jul. 2003. [Online]. Available: http://doi.acm.org/10.1145/882262.882350

[12] L. De Silva, M. Tahara, K. Aizawa, and M. Hatori, "A teleconferencing system capable of multiple person eye contact (MPEC) using half mirrors and cameras placed at common points of extended lines of gaze," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, no. 4, pp. 268–277, Aug. 1995.

[13] D. T. Nguyen and J. Canny, "MultiView: Improving trust in group video conferencing through spatial faithfulness," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2007, pp. 1465–1474. [Online]. Available: http://doi.acm.org/10.1145/1240624.1240846

[14] A. Kulik, et al., "C1x6: A stereoscopic six-user display for co-located collaboration in shared virtual environments," *ACM Trans. Graph.*, vol. 30, no. 6, pp. 188:1–188:12, Dec. 2011. [Online]. Available: http://doi.acm.org/10.1145/2070781.2024222

[15] S. Beck, A. Kunert, A. Kulik, and B. Froehlich, "Immersive group-to-group telepresence," *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 4, pp. 616–625, Apr. 2013.

[16] G. Kurillo and R. Bajcsy, "3D teleimmersion for collaboration and interaction of geographically distributed users," *Virtual Reality*, vol. 17, no. 1, pp. 29–43, 2013. [Online]. Available: http://dx.doi.org/10.1007/s10055-012-0217-2

[17] C. Plüss, et al., "An immersive bidirectional system for life-size 3D communication," in *Proc. 29th Int. Conf. Comput. Animation Social Agents*, 2016, pp. 89–96. [Online]. Available: http://doi.acm.org/10.1145/2915926.2915931

[18] A. Maimone, X. Yang, N. Dierk, A. State, M. Dou, and H. Fuchs, "General-purpose telepresence with head-worn optical see-through displays and projector-based lighting," in *Proc. IEEE Virtual Reality*, Mar. 2013, pp. 23–26.

[19] C. Kuster, et al., "Towards next generation 3D teleconferencing systems," in *Proc. 3DTV-Conf. True Vis. Capture Transmiss. Display 3D Video*, Oct. 2012, pp. 1–4.

[20] A. Steed, et al., "Beaming: An asymmetric telepresence system," *IEEE Comput. Graph. Appl.*, vol. 32, no. 6, pp. 10–17, Nov. 2012.

[21] H. Fuchs, A. State, and J.-C. Bazin, "Immersive 3D telepresence," *Comput.*, vol. 47, no. 7, pp. 46–52, 2014.

[22] I. Poupyrev, M. Billinghurst, S. Weghorst, and T. Ichikawa, "The go-go interaction technique: Non-linear mapping for direct manipulation in VR," in *Proc. 9th Annu. ACM Symp. User Interface Softw. Technol.*, 1996, pp. 79–80. [Online]. Available: http://doi.acm.org/10.1145/237091.237102

[23] B. Petit, et al., "Multi-camera real-time 3D modeling for telepresence and remote collaboration," *Int. J. Digit. Multimedia Broadcast.*, vol. 2010, 2010, Art. no. 12.

[24] M. Slater, "A note on presence terminology," *Presence Connect*, vol. 3, pp. 1–5, Jan. 2003.

[25] M. Lombard and T. Ditton, "At the heart of it all: The concept of presence," *J. Comput.-Mediated Commun.*, vol. 3, no. 2, 1997. [Online]. Available: http://dx.doi.org/10.1111/j.1083–6101.1997.tb00072.x

[26] J. Steuer, "Defining virtual reality: Dimensions determining telepresence," *J. Commun.*, vol. 42, no. 4, pp. 73–93, 1992. [Online]. Available: http://dx.doi.org/10.1111/j.1460-2466.1992.tb00812.x

[27] S. Shepherd, "Following gaze: Gaze-following behavior as a window into social cognition," *Frontiers Integrative Neurosci.*, vol. 4, 2010, Art. no. 5. [Online]. Available: http://journal.frontiersin.org/article/10.3389/fnint.2010.00005

[28] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, ISBN: 0521540518, 2004.

[29] A. Steed, S. Frlston, M. M. Lopez, J. Drummond, Y. Pan, and D. Swapp, "An 'in the wild' experiment on presence and embodiment using consumer virtual reality equipment," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, no. 4, pp. 1406–1414, Apr. 2016.

[30] A. Steed, Y. Pan, F. Zisch, and W. Steptoe, "The impact of a self-avatar on cognitive load in immersive virtual reality," in *Proc. IEEE Virtual Reality*, Mar. 2016, pp. 67–76.

[31] B. Lenggenhager, T. Tadi, T. Metzinger, and O. Blanke, "Video ergo sum: Manipulating bodily self-consciousness," *Sci.*, vol. 317, no. 5841, pp. 1096–1099, 2007. [Online]. Available: http://science.sciencemag.org/content/317/5841/1096

[32] S. Razzaque, D. Swapp, M. Slater, M. C. Whitton, and A. Steed, "Redirected walking in place," in *Proc. Workshop Virtual Environ.*, 2002, pp. 123–130. [Online]. Available: http://dl.acm.org/citation.cfm?id=509709.509729

[33] E. A. Suma, G. Bruder, F. Steinicke, D. M. Krum, and M. Bolas, "A taxonomy for deploying redirection techniques in immersive virtual environments," in *Proc. IEEE Virtual Reality Workshops*, Mar. 2012, pp. 43–46.

[34] E. A. Suma, Z. Lipps, S. Finkelstein, D. M. Krum, and M. Bolas, "Impossible spaces: Maximizing natural walking in virtual environments with self-overlapping architecture," *IEEE Trans. Vis. Comput. Graph.*, vol. 18, no. 4, pp. 555–564, Apr. 2012.

[35] K. Vasylevska, H. Kaufmann, M. Bolas, and E. A. Suma, "Flexible spaces: Dynamic layout generation for infinite walking in virtual environments," in *Proc. IEEE Symp. 3D User Interfaces*, Mar. 2013, pp. 39–42.

[36] B. Williams, et al., "Exploring large virtual environments with an HMD when physical space is limited," in *Proc. 4th Symp. Appl. Perception Graph. Vis.*, 2007, pp. 41–48. [Online]. Available: http://doi.acm.org/10.1145/1272582.1272590

[37] H. Ishii and M. Kobayashi, "ClearBoard: A seamless medium for shared drawing and conversation with eye contact," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 1992, pp. 525–532. [Online]. Available: http://doi.acm.org/10.1145/142750.142977

[38] M. Chen, "Design of a virtual auditorium," in *Proc. 9th ACM Int. Conf. Multimedia*, 2001, pp. 19–28. [Online]. Available: http://doi.acm.org/10.1145/500141.500147

[39] M. Kuechler and A. Kunz, "HoloPort - A device for simultaneous video and data conferencing featuring gaze awareness," in *Proc. IEEE Conf. Virtual Reality*, 2006, pp. 81–88. [Online]. Available: http://dx.doi.org/10.1109/VR.2006.71

[40] H. Regenbrecht, L. Müller, S. Hoermann, T. Langlotz, and A. Duenser, "Implementing eye-to-eye contact in life-sized videoconferencing," Inf. Sci., HCI, Univ. Otago, Dunedin, New Zealand, 2012.

[41] M. Dou, Y. Shi, J.-M. Frahm, H. Fuchs, B. Mauchly, and M. Marathe, "Room-sized Informal telepresence system," in *Proc. IEEE Virtual Reality*, 2012, pp. 15–18. [Online]. Available: http://dx.doi.org/10.1109/VR.2012.6180869

[42] W. Matusik and H. Pfister, "3D TV: A scalable system for real-time acquisition, transmission, and autostereoscopic display of dynamic scenes," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 814–824, Aug. 2004. [Online]. Available: http://doi.acm.org/10.1145/1015706.1015805
[43] T. Balogh and P. T. Kovács, "Real-time 3D light field transmission," in *Proc. Soc. Photo-Opt. Instrum. Eng. Conf. Series*, May 2010, Art. no. 6.
[44] A. Jones, et al., "Achieving eye contact in a one-to-many 3D video teleconferencing system," *ACM Trans. Graph.*, vol. 28, no. 3, 2009, Art. no. 64.
[45] Y. Pan and A. Steed, "A gaze-preserving situated multiview telepresence system," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2014, pp. 2173–2176. [Online]. Available: http://doi.acm.org/10.1145/2556288.2557320
[46] P. Kauff and O. Schreer, "An immersive 3D video-conferencing system using shared virtual team user environments," in *Proc. 4th Int. Conf. Collaborative Virtual Environ.*, 2002, pp. 105–112. [Online]. Available: http://doi.acm.org/10.1145/571878.571895
[47] H. H. Baker, et al., "Understanding performance in coliseum, an immersive videoconferencing system," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 1, no. 2, pp. 190–210, May 2005. [Online]. Available: http://doi.acm.org/10.1145/1062253.1062258
[48] N. Kelshikar, et al., "Real-time terascale implementation of tele-immersion," in *Proc. Int. Conf. Comput. Sci.*, 2003, pp. 33–42. [Online]. Available: http://dl.acm.org/citation.cfm?id=1757599.1757604
[49] A. Maimone and H. Fuchs, "A first look at a telepresence system with room-sized real-time 3D capture and life-sized tracked display wall," in *Proc. Int. Symp. Inf. Commun. Autom. Technol.*, Nov. 2011, pp. 4–9.
[50] C. Kuster, T. Popa, C. Zach, C. Gotsman, and M. H. Gross, "FreeCam: A hybrid camera system for interactive free-viewpoint video," in *Proc. Vis. Model. Vis.*, 2011, pp. 17–24.
[51] C. Zhang, Q. Cai, P. Chou, Z. Zhang, and R. Martin-Brualla, "Viewport: A distributed, immersive teleconferencing system with infrared dot pattern," *IEEE MultiMedia*, vol. 20, no. 1, pp. 17–27, Jan.–Mar. 2013.
[52] J. Allard, J. Allard, J.-S. Franco, C. Ménier, E. Boyer, and B. Raffin, "The GrImage platform: A mixed reality environment for interactions," in *Proc. 4th IEEE Int. Conf. Comput. Vis. Syst.*, 2006, Art. no. 46.
[53] B. Petit, et al., "A 3D data intensive tele-immersive grid," in *Proc. Int. Conf. Multimedia*, 2010, pp. 1315–1318. [Online]. Available: http://doi.acm.org/10.1145/1873951.1874210
[54] M. Willert, S. Ohl, A. Lehmann, and O. Staadt, "The extended window metaphor for large high-resolution displays," in *Proc. 16th Eurographics Conf. Virtual Environ. 2nd Joint Virtual Reality*, 2010, pp. 69–76. [Online]. Available: http://dx.doi.org/10.2312/EGVE/JVRC10/069–076
[55] O. Schreer, I. Feldmann, N. Atzpadin, P. Eisert, P. Kauff, and H. J. W. Belt, "3DPresence-A system concept for multi-user and multi-party immersive 3D videoconferencing," in *Proc. 5th Eur. Conf. Visual Media Prod.*, 2008, pp. 1–8.
[56] L. M. Krauss, *The physics of Star Trek*, 2nd ed., Basic Books, 2007.

Stephan Ohl received the German diploma degree in computer science and the PhD degree from the University of Rostock. He worked as a researcher at the University of Rostock and the Fraunhofer IGD. His research interests include tele-immersion theory and systems, 3D reconstruction, and analyzable data flow models for distributed systems modeling and implementation.