A virtual community communication system where two or more technicians carry or access an Augmented Reality (AR)-enhanced apparatus to communicate and exchange, over a LAN or the Internet, information regarding assembly or servicing or maintenance operations performed on complex machinery. Data streams exchange between the peers of the virtual community is performed by means of a centralised server. Various arrangements are presented that can be selected based on the needs of the operation to be performed, such as the number of members of the community and the type of communication equipment. The system is applicable to any application of the virtual community communication system and is optimized for application to industrial machinery. An explicit mechanism for the reduction and compensation of end-to-end communication latency is provided.
Published:
— without international search report and to be republished upon receipt of that report (Rule 48.2(g))
FIELD OF THE INVENTION

The present invention relates to an information processing method for remote assistance during assembly or maintenance operations. Moreover, the invention relates to an apparatus that carries out such method.

DESCRIPTION OF THE PRIOR ART

Present-day industrial machinery have a very complex design, and as a consequence maintenance or repairing operations need the intervention of a variety of specialists, often only available as members of the manufacturer’s technical staff.

When machinery needs intervention, it is often required for the or each expert to travel to the site where the machinery is hosted. Travelling is expensive and takes a long time, both factors influencing the total cost of a servicing operation. Also, while the expert is in transit additional costs are due to the reduced efficiency of the faulty machine.

On the other hand, a large part of the servicing operations would not really require in-deep knowledge of a machine working principle or internal structure, and could be performed by an in-situ technician who should have at least some step-by-step basic notions on the manual tasks to be performed on the machine.

Several attempts have been made to put the in-situ technician in touch with a remote expert using a variety of communication means, ranging from voice communication to more complex data transmission technology.

In particular, for the in-situ technician Augmented Reality (AR) has been proposed for man-machine interaction, since it presents a major potential for supporting industrial operational processes. It overlays computer-generated graphical information onto a physical (real) world, by means of a see-trough near-eye display controlled by a computer.
The field of view of the observer is enriched with the computer-generated images.

For example EP1157314 discloses an AR system for transmitting first information data from a technician at a first location to a remote expert at a second location. A sensor system is provided for data acquisition at the technician's site, and for evaluating them at the expert's site, assigning then real objects to stored object data, which are provided at the technician's site.

US2002010734 disclose an internetworked augmented reality (AR) system, which is mainly dedicated to entertainment and consists of one or more local stations and one or more remote stations networked together. The remote stations can provide resources not available at a local AR Station such as databases, high performance computing (HPC), and methods by which a human can interact with the person(s) at the local station.

The above known systems cannot assure that a remote expert has a clear and updated view on the actions to do, since an exchange of a number of high-quality video and data streams is required. Moreover, the above knows system cannot handle the case of multiple experts concurrently supporting the in-situ technician(s) from different remote locations. Finally, no special technique is employed to compensate for end-to-end communication latency, fact that has a severe impact on effective communication.

**Summary of the invention**

It is therefore a feature of the invention to provide an advanced communication system used to create a virtual team of technicians performing assembly or servicing operations of complex machinery in a collaborative way forming a virtual community of geographically distributed experts and technicians.

In particular, one or more of the participating members can be located at different geographical sites with respect to the machinery, providing remote help to the in-situ technicians.

It is another feature of the invention to specify a system and a
method which, in concrete operational situations, permits an effective virtual co-participation of the remote experts to the in-situ technicians actions and decisions, in order to enhance their technical ability up to perform some of the servicing operation without the need of the physical presence of the expert.

It is a further feature of the invention to provide an information processing apparatus and method for remote technical assistance in which resolution, frame rate, and network latency allow a practical, effective and fast operation.

It is a further feature of the invention to provide a system architecture organised around a centralised communication server that allows multiple experts to assist the in-situ technician(s), each of them with the freedom to participate from different geographical locations.

It is a further feature of the invention to include an explicit mechanism for the reduction and compensation of end-to-end communication latency.

These and other features are accomplished with one exemplary method for remote assistance during assembly or maintenance operations, the method comprising the steps of:

- providing at least one technician at a first location and at least one expert at a second location,
- exchanging information data via high-efficiency video compression means between said at least one technician and said at least one expert, through a set of communication channels, including audio and video streams, interactive 2D and 3D data,
- wherein the information data are selected among video images, graphics and speech signals of the technician and wherein additional information data in the form of augmented-reality information are transmitted from the remote expert at the second location to the in situ technician at the first location, highlighting specific objects in the field of view of the technician,
- said expert being equipped with a computer and videoconferencing devices;
- said technician being equipped with
a wearable computer having a radio antenna associated to said wearable computer for data transmission;
a headset connected to said computer including headphones, a noise-suppressing microphone, one near-eye see-through AR display, and a miniature camera mounted on the display itself used to capture what is in the front of view of the technician; characterised in that said at least one technician and said at least one expert are arranged respectively at an in-situ-node and at a remote node of a network, said nodes communicating and exchanging data through the internet via a centralised communication server, and in that the following steps are provided of:
sampling in real-time the position of said headset, providing position data of said headset at a predetermined sampling time.
streaming video images, graphics and speech signals of the technician from the in-situ-node to the remote node;
creating by said expert said additional information data in the form of augmented-reality at least on a determined position of said streamed video images and sending back said additional information data referred to said position from said remote node to said in-situ-node
calculating a shifted position of said additional information data according to movements of said headset occurred between two determined sampling times;
displaying said additional information data on said see-through AR display in said shifted position.
Preferably, said position data are in the form of a 3DOF or 6DOF transformation matrix, wherein at each sampling time a transformation matrix is generated.
Preferably, said images are in the form of a succession of frames, and to each transformation matrix a frame index is associated, each transformation matrix being responsive to position changes between an actual frame and a immediately previous frame.
Preferably, said shifted position is determined by transforming the position determined by said expert by a transformation matrix
corresponding all the changes occurred between a starting frame with a starting frame index and an actual frame with an actual frame index.

Advantageously, a step is provided of sending at said sampling time additional numerical data adapted to reduce end-to-end latency effects from said in-situ-node to the remote node, said additional numerical data comprising position data corresponding to movements of said headset measured at said sampling time, in particular said position data are in the form of said transformation matrix.

In a preferred embodiment, to said remote node a plurality of further experts are connected that look at said images at an expert display, said further experts displaying said additional information data in an actual shifted position customized for each further expert, said shifted position being determined by each further expert on the basis of a transformation matrix available at said remote node and corresponding to the index frame of the frame actually seen by each further expert.

According to another aspect of the invention, an apparatus for remote assistance during assembly or maintenance operations comprises:

means for exchanging information data between at least one technician at a first location and at least one expert at a second location through a set of communication channels, including audio, voice and interactive graphics, as well as 3D data,

wherein the information data are a collection of video images, graphics and speech signals of the technician and wherein additional information data in the form of augmented-reality information are transmitted from a remote expert at the second location to an in situ technician at the first location highlighting specific objects in the field of view of the technician,

a computer and videoconferencing devices to be used by said expert;

a unit to be used by said technician comprising a wearable computer having a radio antenna associated to said wearable computer for data transmission; a headset connected to said computer including headphones, a noise-suppressing microphone, one near-eye see-through
AR display, and a miniature camera mounted on the display itself used to capture what is in the front of view of the technician;

characterised in that

means are provided for communicating and exchanging data between at least an in-situ-node and a remote node of a network, said nodes communicating and exchanging data through the internet via a centralised communication server,

means for sampling in real-time the position of said headset and for providing position data of said headset at a predetermined sampling time.

means for streaming video images, graphics and speech signals of the technician from the in-situ-node to the remote node;

means for creating by said expert said additional information data in the form of augmented-reality at least on one determined position of said streamed video images and sending back said additional information data referred to said position from said remote node to said in-situ-node

means for calculating a shifted position of said additional information data according to movements of said headset occurred between two determined sampling times;

means for displaying said additional information data on said see-through AR display in said shifted position.

Preferably, said position data are in the form of a 3DOF or 6DOF transformation matrix, and said means for sampling are adapted to generate at each sampling time a transformation matrix.

Preferably, video compression means to reduce streaming bandwidth of said data are provided.

Advantageously, a hand held camera is provided connected to said computer equipped with a light source for lighting desired targets,

Advantageously, an RFID sensor is mounted on said camera to allow for the detection of parts code and associated information

Advantageously, additional automated remote computing nodes are provided to create additional video feeds, in particular auxiliary fixed cameras that are positioned by the technicians and that can be controlled by the remote experts for pan, zoom and tilt movements.
Preferably, the organisation of a multitude of in-situ technicians and remote experts situated at different geographical locations is established in a distributed virtual community for the exchange of knowledge, wherein at least one on-situ technician at one node and at least one remote expert at another nodes are provided communicating and exchanging data with each other.

Therefore, a virtual community of skilled specialists is created where members communicate by means of internetworked computers and several input/output devices. The virtual community can therefore be conceptualised as a group of technicians each of them equipped with a computer (computing node) plus some automated remote computing node used to provide additional video feeds. Each computing node exchanges data over a wide-area communication network. Some of these nodes can share the same physical space while other can be located at multiple geographical locations.

Preferably, the use of Augmented Reality is provided to overlap special visual markers on the objects falling inside the field of view of the operator.

Advantageously, said headset can also be equipped with a 3DOF tracking system, used to compute head moments of the in-situ technician, in such a way to compensate such movements in terms of visual displacement of the computer generated graphical markers that are overlapped on the field of view of the technician.

Advantageously, said 3DOF tracking capability is used to compensate for end-to-end communication delay.

Advantageously, video streaming is associated to VoiceOverIP technology.

Preferably, said video compression means comprise H.264 Compression Technology.

Preferably, video compression means are arranged in such a way that the video streams and audio streams are compressed and combined, preferably 384 Kbit/s uplink and 384 Kbit/s down link.

Brief description of the drawings
The invention will be now shown with the following description of an
exemplary embodiment thereof, exemplifying but not limitative, with reference to the attached drawings, in which:

- figure 1 shows an architecture of a virtual community communication system for remote technical assistance;

- figure 2 shows a particular embodiment of an architecture of a virtual community communication system for remote technical assistance where the nodes are arranged as sub-communities according to affinity criteria;

- figure 3 shows the architecture of figure 1 where at the computing nodes in-situ technicians, remote experts and remotely controlled video-cameras looking at the machinery are indicated;

figures 4 to 6 show an on-field technician equipped with a wearable computing system and a special head set integrating an Augmented Reality see-trough display;

figures 7 to 8 show an on-field technician equipped also with a hand held camera;

- figure 9 shows an an in-situ fixed node, composed by a remotely controlled pan-tilt-zoom camera mounted on a tripod;

- figure 10 shows a Graphic Technician Interface of the application running at a technician's node where to a technician three different streaming video data are presented.

- figure 11 shows a block diagram of a preferred working unit of an apparatus according to the invention;

- figure 12 shows a data communication scheme applied to the architecture of the virtual community communication system for remote technical assistance of figure 1 using the preferred working units of figure 11;

- figure 13 shows a data communication scheme applied to a different embodiment of a virtual community communication system for remote technical assistance, using a peer to peer architecture, and using the preferred working units of figure 11.

- Figure 14 shows the direction of the video data stream in the virtual community, traversing the internet from the in-situ technician headset camera towards the centralised communication server. The server
retransmit the signal towards one or more expert technician(s).

- Figure 15 shows the direction of the audio data streams in the virtual community, traversing the internet between the various computing nodes and the centralised communication server.

- Figure 16 shows the direction of the data streams associated with the tracking functionalities of the invention, traversing the internet between the various computing nodes and the centralised communication server.

- Figure 17 shows the working principle of the object-tracking feature of the invention, as well as the flow of data between the in-situ technician and the remote expert concerning the use of tracking to highlight objects falling in the field of view of the in-situ technician.

- Figure 18 shows the same tracking data flow as figure 17 when in the virtual community there is more than one expert assisting the in-situ technician;

- Figure 19 shows the working principle of the object highlighting feature of the invention when multiple experts in the virtual community are assisting the in-situ technician, each expert highlighting independent objects falling in the field of view of the technician.

- Figure 20 shows a block diagram of the main steps of the method according to the invention.

Description of the preferred embodiments

With reference to figure 1, an information processing apparatus and method are provided to establish a virtual community of geographically distributed experts and technicians for remote assistance during assembly or servicing operations of complex devices. The technician(s) and the expert(s) are arranged at nodes 1-N. Nodes 1-N communicate with one another and exchange data through the internet via a centralised communication server 8.

In addition, the centralised Communication Server 8 is used for monitoring the data, checking the data traffic, controlling the access rights and storing usage statistics.
In figure 2, the capability of the system is shown to group technicians in sub-communities which can be created according to various criteria, such as affinity in terms of servicing scenario, physical contiguity etc.

In particular, the presence of a centralised server allows for a dynamic management of how the technicians are grouped in sub-teams. In fact, multiple virtual teams composed by some in-situ technicians, some automated cameras and some remote experts can operate at the same time on multiple locations. Members of one team can be dynamically be allocated to another team even for a limited amount of time: this maximises the possibility that experts with specific know-how can quickly be contacted and involved in the assembly/servicing operation. In addition, in-situ technicians in a particular operation can quickly be transformed in remote experts for another particular operation, changing their roles amongst the teams. This dynamical architecture assures that even the skills and knowledge of the highly trained technicians is at disposal of the collectivity.

An example of remote technical assistance through the invention is shown in figure 3, where a network is illustrated managed by centralised communication server 8. Industrial machinery 11, for example large machinery located in an industrial plant, has to be serviced or assembled or inspected by technicians 9 and by auxiliary fixed video cameras 10. The experts 9 advise the technicians on how to operate.

In particular, the architecture is shown of the virtual community communication system, where the computing nodes, such as one or more remote nodes where experts 12 are present, one or more in-situ mobile nodes where a technician 9 is present, and fixed nodes 10 with remotely controllable video cameras, communicate and exchange data through the internet via centralised communication server 8.

In-situ technicians 9 use wearable equipment and move freely around the machinery 11. One or more among auxiliary remotely controlled video-cameras 10 can also be placed around the machinery 11 to provide extra video streams of the operations being performed by the technicians. Pan, zoom and tilt of these auxiliary cameras 10 can be
controlled by the remote experts 12, who can adjust them in order to obtain desired images of the machine. Remote experts 12 are connected to the internet from one or more remote locations and are equipped with standard laptop computers 14 and videoconferencing devices, such as voice communication headphones 13.

The remote experts 12 receive and examine all the information coming from the technicians 9 and the cameras 10 and can consequently send back manipulation instructions by means of voice or by remotely controlling the display of special dynamic graphical markers (described hereafter with reference to figure 10) that appear on the field of view of the in-situ technicians by means of the Augmented Reality display.

With reference to figures 4-6, a on-field technician a wearable computing system 1 and a special head set 4 integrating an Augmented Reality see-trough display. The wearable AR-based apparatus is composed of a backpack 3 containing a portable computer and a helmet 4 where a video camera 5, headphones 6 with a microphone 6A and a see-through display 7 are mounted.

With reference to figures 7 and 8 an in-situ technician 12 wearing the AR-based apparatus 1 can hold an additional hand-held camera 2, having a lighting system, preferably with white LEDs, connected to the computer and that can be used to show to the remote experts 12 portions of the real scene that would be impractical to show using the video camera mounted on the headset or the fixed video cameras.

With reference to figure 9, and as previously indicated in figure 4, additionally to the computing nodes associated with in-situ technicians and remote experts, a third additional kind of computing node can be inserted in the community, comprising a remote controlled high-quality video-camera 10. It is mounted on a tripod 15 that can be placed around the machinery 11 (see fig. 4) to provide additional view-points on the operations. For example, in figure 9 instead of machinery a computer station 30 is shown, for example for remotely instructing technicians on how to assemble or service the station, or for training purposes. Each camera 10 is equipped with motorised Pan, Zoom and Tilt support that
can be controlled by the remote experts 12.

The camera 10 can either be a stand-alone network camera, equipped with video compression and network streaming capabilities, or a device connected to a computer 20, capable of acquiring, compressing and transmitting video data over the network and to the centralized communication server.

Figure 10 shows what the remote expert sees on the screen of its laptop, as seen by fixed camera 10 of figure 9, as well as by the micro camera on the headset or the hand held camera, and what kind of visual feedback can produce that will be overlapped on the field of view of the in-situ technician. Figure 10 can be the Graphic Technician Interface of the application running at the expert site. Figure 10 can be however also the Graphic Technician Interface of the application running at the technicians site.

To the expert the following is presented using three different streaming video data: in 31 video data is displayed coming from the fixed camera of a fixed in-situ node; in 32 video data are shown coming from the hand-held camera operated by the in-situ operator; in 33 video data coming from the helmet camera worn by the in-situ operator. The settings of each of these views can be customized using a system of slider and buttons 34. In particular, the in-situ fixed camera can be remotely operated modifying its orientation and its zooming. The technician at the technician's site or the expert can select which of these views is currently the active view 5 and have an audio/textual chat 36 with the other operators of the community. Moreover, the expert can draw enhancing symbols and markers, 37 or 38, using a selected technician interface mouse, pen, touch-screen etc. (not shown) on the active view, causing this information to appear on the see-through display worn by the in-situ operator. The latter, in this way, can be guided with extreme precision in actions, since the guidance is contextualized in the physical space on the field of view. In the same way the expert can send other kind of useful graphical information that is superimposed on the field of view of the in-situ operator, like cad drawings, text, 3D data, animations etc.
It is advantageous that the technician at the technician's site has a see-through monitor, so that the technician can see contemporaneously and on a same screen the images of the site and the images sent by the remote expert.

In a preferred embodiment, summarized by the block diagram of Fig. 11, the apparatus according to the invention has a computing system worn by the user that, in an advantageous embodiment of the invention, controls: a see-trough near eye display or a standard display; an auxiliary standard display; a RFID or barcode reader; two or more video cameras; a H.264 compression technology; input devices (like keyboard, mouse, etc).

Finally, for an effective communication over narrow-band links to the internet the system make explicit use of video and audio compression technology. In particular, the video streams and audio streams are compressed and combined in order to stay within the limits of the standard UMTS data plans (384 Kbit/s uplink and 384 Kbit/s down link). The system is also equipped with adaptive algorithms that increase the quality of the video-audio-data streams when the availability of larger bandwidth is detected. In particular, H.264 Compression Technology can be used.

The various data streams of the virtual community are managed centrally through the internet by the centralised communication server; figure 14 illustrates the data communication scheme for the video data applied to the architecture of the virtual community communication system for remote technical assistance of figure 1; video data is flowing from the in-situ technician towards the centralised server that is used to generate in return multiple identical streams, to feed the computing node of each expert in the community - each expert can join the virtual community from a different physical location. Figure 15 illustrates the flow of data related to audio feedback: each member can speak into his/her microphone, and his/her voice will be received (with a minimal latency) by every member of the community. Each voice stream generated by a member is first sent to the centralised communication server, that is used to replicate and stream the data towards every other
node.

 Optionally, the headset can also be equipped with a 3DOF tracking system, used to measure rotational head moments of the skilled technician. This is used to compensate such movements in terms of visual displacement of the computer generated graphical markers that are overlapped on the field of view of the technician. Say for example that the technician is looking at a complex control panel populated by a variety of controls: the remote expert is pointing the attention to a specific object overlapping graphical markers around it. This correspondence is obviously valid only as long as the in-situ technician does not translate or rotate the head. While translational movements are not very frequent in a typical maintenance operation, small rotational movements can occur frequently with a consequent loss of the correspondence between the objects and the overlapped markers. The presence of a 3DOF or a 6DOF tracking system on the headset allows to compensate for such rotational movements, helping to track the correct object-marker correspondence. The system also takes in account the inevitable delays occurring in the communication between the in-situ technician and the remote expert.

 Several solutions can be used for 3DOF or 6DOF tracking: as an example a combination of miniaturised accelerometers and gyroscopes can be mounted on the headset of the in-situ technician to achieve 6DOF tracking []. Another exemplary option is provided by applying computer vision algorithms to the video feed captured by the headset camera [].

 In any case, the computation needed for 3DOF tracking is advantageously performed on the computing node of the in-situ technician: if accelerometers, gyroscopes or other sensor are used, these needs to be mounted on the headset of the in-situ technician to detect head movements. If computer vision techniques are used, video analysis is better performed on the in-situ technician computing node, as video data here is of pure quality, not being affected by the quantisation errors nor time latencies introduced by the video-compression apparatus.
It is therefore an important aspect of this invention to perform tracking computation on the in-situ technician computing node, and, as tracking data is available, precisely associate this data to each frame of the video stream and distributed the result to all the others members of the virtual community.

Figure 16 illustrates the data communication scheme for the tracking data applied to the architecture of the virtual community communication system for remote technical assistance of figure 1; tracking data is flowing from the in-situ technician towards the centralised server that is used to generate in return multiple identical streams, to feed the computing node of each expert in the community.

An essential aspect of the invention is how tracking data is advantageously used to allow the highlighting of specific objects and to compensate for data communication delays inside the virtual community. An introduction of the problem is necessary: end-to-end communication delay is a fundamental factor for the usability of any communication apparatus. In the present apparatus, the remote expert(s) perceive the environment in front of the in-situ technician with a certain delay; this is mainly due to the following factors: video acquisition time delay -the time needed to the computing node to sample and store in digital form the image coming from the camera(s)-, video compression time delay -the time necessary to execute the complex compression algorithms used to reduce bandwidth requirements, - network traversal delay -the time needed for the data to traverse the internet, going through the communication server and reaching the expert(s)computing node, decompression and display time delay - the time needed for the stream to get converted back in a digital image and visualised on the monitor of the expert(S). Moreover, the expert will look at this moving images and will need some time to decide what object in the image should be highlighted by the invention in the field of view of the in-situ-technician; depending on the reactivity of the expert, this delay can actually be quite large. Once a decision is taken, the expert will use mouse and keyboard to specify the selected object. Information about this object will then have to traverse the network back to the computing node of the in-situ
technician. Even if this kind of data is lightweight and therefore not affected by significant compression/decompression delays, the latency due to network traversal will still be significant.

By the time the in-situ technician receive this data, he/she might have moved considerably in front of the machine: very commonly any correspondence of what is in front of him and what was selected on the screen of the expert is lost. Without tracking, icons and markers placed by the expert in the field of view of the technician might be completely misplaced, impeding effective visual guidance from the expert to the technician. Also, as multiple experts can be at different locations, each of them will experience a different amount of network latency. In this scenario, communication can be extremely difficult without mechanism that allows some form of time-spatial synchronisation.

It is therefore a fundamental aspect of our invention to compensate for this latency effects embedding tracking data in each frame of the video stream sent from the technician to the expert(s). This tracking data is used by the various computing nodes in the virtual community to insure that markers placed by the expert(s) on their image space are then properly converted in to the image space of each other community member, including the in-situ technician and the other remote experts. The working principle of our method is depicted in Figure 17: the left column shows what is seen by the in-situ technician trough the headset display during a given interval of time. He/she has in front a compound object (table with objects). Every 100 milliseconds the video camera samples the image in front of the technician. The image changes over time due to the movements of the technician, and each image is compressed and sent to the internet. Such sequence of images are then received by the expert(s) of the virtual community. The right column of Figure 17 shows what is seen by a remote expert on the display of his/her computer node. As discussed, each image arrives at the expert with a certain amount of delay. In order to better study the images and select and highlight a specific object in the image, the expert can "freeze" the video sequence on his side, introducing some additional delay between what is in front of the technician and what is seen by him. When
the selection of the object is complete, the video sequence is released and the data of the selection is sent from the expert computing node to the technician computing node. Obviously, a screen-to-screen correspondence would be lost due to the round-trip delay of this process; here is when the tracking information embedded in each image is used to reconstruct such correspondence: the actual tracking data of the in-situ technician gets confronted with the tracking data embedded in the image "freezed" by technician. Using this information, the image markers position sent by the remote expert gets transformed into the current image-space of the image-technician, and the marker displayed at the proper position. In this way, the correspondence between the object selected by the expert and the object seen by the in-situ technician is maintained regardless of the latency due to compression/transmission of data.

It is a fundamental aspect of the present invention to replicate the transformation process not only for the in-situ technician, but on each of the computing node of the virtual community. Figure 18 illustrate the process and streams of tracking data when there are multiple experts assisting the in-situ technician. As each expert can experience a different latency in the communication, the before mentioned process of image-space transformation using embedded tracking data is repeated for each of them. In this way every expert can see the markers placed by the other experts on the same video stream - each expert has a specific marker screen colour (for instance BLUE for expert 1, GREEN for expert 2, RED for expert 3 and so on). This form of visual coordination between the whole community, associated to the unified conference voice communication between the members allows for an innovative and very effective collaborative work, an essential advantage of the present invention.

Figure 19 details what is seen by every member of the community providing an example of three experts assisting the in-situ technician, with two experts drawing markers on the video streams. The diagram, composed by 16 successive steps (from left to right and from top to bottom) shows what appears on the monitor and on the headset of each
member of the community.

Figure 14 illustrates the data communication scheme for the video data applied to the architecture of the virtual community communication system for remote technical assistance of figure 1. Video data is flowing from the in-situ technician 9 towards the centralised server 8 that is used to generate in return multiple identical streams 17, to feed the computing node of each expert 12 in the community; each expert 12 can join the virtual community 16 from a different physical location.

Figure 15 illustrates the flow of data related to audio feedback: each member 9 or 12 can speak into his/her microphone (not indicated), and his/her voice will be received, with a minimal latency, by every member 9-12 of the community 16. Each voice stream 18 generated by a member 9 or 12 is first sent to the centralised communication server 8, that is used to replicate and stream the data towards every other node.

Figure 16 illustrates the data communication scheme for the tracking data applied to the architecture of the virtual community 16 communication system for remote technical assistance of figure 1. Tracking data 19 is flowing from the in-situ technician 9 towards the centralised server 8 that is used to generate in return multiple identical streams 21, to feed the computing node of each expert 12 in the community 16.

The method 100 for remote assistance during assembly or maintenance operations, according to the invention, is described with reference to Figures 17 an 20. The step of providing at least one technician 9 at a first location and at least one expert 12 at a second location is followed by a step of exchanging information data 101 via high-efficiency video compression means between said at least one technician 9 and said at least one expert 12, through a set of communication channels, including audio and video streams, interactive 2D and 3D data. The information data 42 are selected among video images, graphics and speech signals of the technician 9 and additional information data in the form of augmented-reality information, in particular a marker M, are transmitted from the remote expert 22 at the second location to the in situ technician 9 at the first location, highlighting
a specific object 26 in the field of view of the technician 9. The expert 12 is equipped with a computer 25 and videoconferencing devices, not displayed, while the technician 9 is equipped as described with a wearable computer 3' having a radio antenna 3" (figure 4) associated to said wearable computer for data transmission, and wears a headset 22 connected to said computer including headphones 6, a noise-suppressing microphone 6A, one near-eye see-through AR display 7, and a miniature camera 5 mounted on the display itself used to capture what is in the front of view of the technician. Peculiar to the method 100 is that the technician 9 and the expert 12 are arranged respectively at an in-situ-node 28 and at a remote node 29 of a network, in particular an end-to-end network 200 having; the nodes communicate and exchange data through the internet via a centralised communication server 8; still peculiar to the method 100 are the is a step 102 of sampling in real-time the position of said headset 22, providing position data at a predetermined sampling time, as well as a step 103 of streaming video images 24, graphics and speech signals of the technician 9 from the in-situ-node 29 to the remote node 28. A step 104 of creating additional information data is then carried out by the expert 12 in the form of augmented-reality, in particular a marker M, at least on a determined position, for instance the position of an object 26', of the streamed video images 24' and sending back the additional information data referred to the position of an object 26', from the remote node 28 to the in-situ-node 29. A subsequent step 106 of calculating a shifted position of the additional information data according to movements of the headset between two determined sampling times is finally followed by a step 107 of displaying the additional information data on the see-through AR display 7 (figure 4) in said shifted position.

Figure 18 illustrate the process and streams of tracking data when there are multiple experts 12 assisting the in-situ technician 9. As each expert 12 can experience a different latency in the communication, the before mentioned process of image-space transformation using embedded tracking data 41 is repeated for each of them. In this way, every expert 12 can see the markers M placed by the other experts 12
on the same video stream 45.

Figure 19 details what is seen by every member of the community providing an example of three experts 12, 12’, 12” assisting the in-situ technician 9, with two experts drawing markers M1 e M2 on the video streams. The diagram, composed by sixteen successive steps (from left to right and from top to bottom) shows what appears on the monitor 25, 25’, 25” and on the headset 22 of each member 9-12 of the community.

The foregoing description of a specific embodiment will so fully reveal the invention according to the conceptual point of view, so that others, by applying current knowledge, will be able to modify and/or adapt for various applications such an embodiment without further research and without parting from the invention, and it is therefore to be understood that such adaptations and modifications will have to be considered as equivalent to the specific embodiment. The means and the materials to realise the different functions described herein could have a different nature without, for this reason, departing from the field of the invention. It is to be understood that the phraseology or terminology employed herein is for the purpose of description and not of limitation.
1. A method for remote assistance during assembly or maintenance operations, the method comprising the steps of:
   providing at least one technician at a first location and at least one expert at a second location,
   exchanging information data via high-efficiency video compression means between said at least one technician and said at least one expert, through a set of communication channels, including audio and video streams, interactive 2D and 3D data,
wherein the information data are selected among video images, graphics and speech signals of the technician and wherein additional information data in the form of augmented-reality information are transmitted from the remote expert at the second location to the in situ technician at the first location, highlighting specific objects in the field of view of the technician,
   said expert being equipped with a computer and videoconferencing devices;
   said technician being equipped with
   a wearable computer having a radio antenna associated to said wearable computer for data transmission;
   a headset connected to said computer including headphones, a noise-suppressing microphone, one near-eye see-through AR display, and a miniature camera mounted on the display itself used to capture what is in the front of view of the technician;
characterised in that
   said at least one technician and said at least one expert are arranged respectively at an in-situ-node and at a remote node of a network, said nodes communicating and exchanging data through the internet via a centralised communication server,
and in that the following steps are provided of:
   sampling in real-time the position of said headset, providing position data of said headset at a predetermined sampling time.
   streaming video images, graphics and speech signals of the
technician from the in-situ-node to the remote node;
creating by said expert said additional information data in the form of augmented-reality at least on a determined position of said streamed video images and sending back said additional information data referred to said position from said remote node to said in-situ-node

calculating a shifted position of said additional information data according to movements of said headset occurred between two determined sampling times;

displaying said additional information data on said see-through AR display in said shifted position.

2. Method according to claim 1, wherein said position data are in the form of a 3DOF or 6DOF transformation matrix, wherein at each sampling time a transformation matrix is generated.

3. Method according to claim 1, wherein said images are in the form of a succession of frames, and to each transformation matrix a frame index is associated, each transformation matrix being responsive to position changes between an actual frame and an immediately previous frame.

4. Method according to claim 1, wherein said shifted position is determined by transforming the position determined by said expert by a transformation matrix corresponding all the changes occurred between a starting frame with a starting frame index and an actual frame with an actual frame index.

5. Method according to claim 1, wherein a step is provided of sending at said sampling time additional numerical data adapted to reduce end-to-end latency effects from said in-situ-node to the remote node, said additional numerical data comprising position data corresponding to movements of said headset measured at said sampling time, in particular said position data are in the form of said transformation matrix.

6. Method according to claim 1, wherein to said remote node a plurality of further experts are connected that look at said images at an expert display, said further experts displaying said additional
information data in an actual shifted position customized for each further expert, said shifted position being determined by each further expert on the basis of a transformation matrix available at said remote node and corresponding to the index frame of the frame actually seen by each further expert.

7. An apparatus for remote assistance during assembly or maintenance operations, comprises:

   means for exchanging information data between at least one technician at a first location and at least one expert at a second location through a set of communication channels, selected among audio, voice and interactive graphics, as well as 3D data,

   wherein the information data are a collection of video images, graphics and speech signals of the technician and wherein additional information data in the form of augmented-reality information are transmitted from a remote expert at the second location to an in situ technician at the first location highlighting specific objects in the field of view of the technician,

   a computer and videoconferencing devices to be used by said expert;

   a unit to be used by said technician comprising a wearable computer having a wi-fi antenna associated to said wearable computer for data transmission; a headset connected to said computer including headphones, a noise-suppressing microphone, one near-eye see-trough AR display, and a miniature camera mounted on the display itself used to capture what is in the front of view of the technician;

   characterised in that it further comprises

   - means for communicating and exchanging data between at least an in-situ-node and a remote node of a network, said nodes communicating and exchanging data through the internet via a centralised communication server,

   - means for sampling in real-time the position of said headset and for providing position data of said headset at a predetermined sampling time.
- means for streaming video images, graphics and speech signals of the technician from the in-situ-node to the remote node;
- means for creating by said expert said additional information data in the form of augmented-reality at least on one determined position of said streamed video images and sending back said additional information data referred to said position from said remote node to said in-situ-node
- means for calculating a shifted position of said additional information data according to movements of said headset occurred between two determined sampling times;
- means for displaying said additional information data on said see-through AR display in said shifted position.

8. The apparatus according to claim 7, wherein said position data are in the form of a 3DOF or 6DOF transformation matrix, and said means for sampling are adapted to generate at each sampling time a transformation matrix.

9. The apparatus according to claim 7, wherein video compression means to reduce streaming bandwidth of said data are provided.

10. The apparatus according to claim 7, wherein a hand held camera is provided connected to said computer equipped with a light source for lighting desired targets,

11. Advantageously, an RFID sensor is mounted on said camera to allow for the detection of parts code and associated information

12. Apparatus according to claim 2, wherein additional automated remote computing nodes are provided to create additional video feeds, in particular auxiliary fixed cameras that are positioned by the technicians and that can be controlled by the remote experts for pan, zoom and tilt movements.

13. Method according to claim 1, wherein the organisation of a multitude of in-situ technicians and remote experts situated at different geographical locations is established in a distributed virtual community for the exchange of knowledge, wherein at least one on-
situ technician at one node and at least one remote expert at another nodes are provided communicating and exchanging data with each other.

14. Apparatus according to claim 2, wherein Augmented Reality means are provided to overlap special visual markers on the objects falling inside the field of view of the operator.

15. Apparatus according to claim 7, wherein said video streaming is associated to VoiceOverIP technology.

16. Apparatus according to claim 7, wherein said video compression means comprise H.264 Compression Technology.

17. Apparatus according to claim 7, wherein video compression means are arranged in such a way that the video streams and audio streams are compressed and combined, preferably 384 Kbit/s uplink and 384 Kbit/s down link.