Internet-Based Exploration of Remote Sites

Robert Laganière

School of Information Technology and Engineering, University of Ottawa
150 Louis Pasteur, Ottawa, Ontario, K1N 6N5
laganier@site.uottawa.ca

ABSTRACT

This paper describes a telepresence system for remote site exploration allowing both real-time unrestricted navigation and live information visualization. The proposed strategy is based on the use two distinct and complementary models: a structural model and an appearance model. These two models are combined together in order to obtain a complete 3D representation.

1. INTRODUCTION

Telepresence can be defined as the art of enabling physical proximity despite geographical distances. Although truly immersive telepresence is difficult to achieve, applications of this kind should soon play an important role in interactive content delivery. This paper presents a system where a viewer has the opportunity to interactively explore an existing remote site. The two main characteristics of this system are that it provides live visual information about the site remotely explored and gives a complete freedom of navigation across the remote site that can thus be seen from any virtual point of view. In addition, this system can be used under a low bandwidth communication channel, making it suitable for the Internet.

Several applications can benefit from this approach, such as teleoperation, telesurveillance, teleinspection, e-commerce and entertainment. In particular, in the present project, we have targeted telerobotic as one the main sector of application for our system. The ability to manipulate and update 3D models from a remotely accessible bank of video cameras looking at a robotic work site is considered to be strategic in teleoperation applications. In a supervisory control mode, this system will provide the visual feedback required to adequately supervise some robot activities.

2. TELEPRESENCE OVERVIEW

Any telepresence system is made of three components: the local unit, a communication link and the remote site equipment [8].

The local unit includes the devices with which the actor in the telepresence experience interacts. A computational device is also present primarily to interface with the communication link. It is also used to make some local computation such as the manipulation and the rendering of the 3D models.

The communication link is the bridge between the user and the environment of immersion. Costly systems can envisage a dedicated fast connection, but more realistically, solutions that can give good performance in spite of a low bandwidth link are more attractive. Communication is often considered as the main responsible for the delay and the saccadic motion that are perceived by the user.

The choice of the equipment to be installed at the remote site has a great impact on the global performance of the system. However, the way the information acquired by these sensors is exploited is also of importance. Remotely controlled pan-tilt cameras, or cameras installed on mobile robotic equipments can be used but they are generally more costly and less reliable. Because of that, fixed cameras might represent a better choice, although their static nature requires more sophisticated integration and interpolation processes.

Currently, most telepresence systems are, more or less, based on one the following two opposite strategies. The first approach relies on the use of multiple cameras disseminated across the remote environment in order to obtain a large visual coverage of the site [4][10]. Image sequences are then sent to the user, giving him a live access to the site. However, the exploration is, in this case, limited to the available viewpoints. In some cases, the camera can move (rotating cameras or camera mounted on mobile robot). In other cases, virtual motions are possible through image interpolation or image warping over omni views. But still serious constraints are present in the continuum of available views. The second approach consists in the prior generation of a virtual 3D model of the site that is to be used during the telepresence experience [1][2][3]. Laser range technologie is now becoming an affordable technology, making possible the acquisition of complex 3D scenes [13][14]. Other approaches use multiple images [12] or video sequences [11] to obtain 3D structure. Unrestricted real-time navigation across the virtualized environment becomes then possible. However, in this case, the virtual 3D model represents the environment at the time it was acquired and constructed; it therefore does not necessarily reflect the current reality.

Although these solutions are of interest, we believe that to provide a high fidelity telepresence experience in evolving environments a good system should be able to offer both *real-time unrestricted navigation* and *live information* that reflects the current site conditions. The strategy proposed here is to use a 3D model of the remote site that has to be acquired before the telepresence session starts. The local availability of this model makes possible real-time unrestricted navigation in the environment. Visualization of live information is achieved by letting the appearance of this 3D model be regularly refreshed. This updating is based on the information coming from a limited set of cameras strategically located at the remote site.

3. SYSTEM DESCRIPTION

The system relies on a client/server architecture where all the computational aspects are distributed over 3 units as illustrated in Figure 1. To initiate a telepresence session, a user must connect to the server using a web browser. Java technology is used here to ensure portability and accessibility of the application via the Internet.

3.1 The two models

The system is characterized by the used of two distinct and complementary models: the *structural model*, the *visual appearance model*.

The structural model contains information about the 3D structure of the static elements of the scene. This model can be obtained from images of the site, taken from different points of view. Since the structural information contained in this model does not change, it can be computed before the telepresence system becomes operational. This is important since the automatic reconstruction of complex scenes from multiple views is a very difficult task [5]. The 3D model is therefore obtained during an offline session where a user assists the reconstruction process.

The appearance model conveys visual information from the remote site to the local unit. Contrary to the structural model which is static, the appearance model is dynamic, i.e. it must be continuously updated. The model is built from the data coming from the cameras. The combination of this model with the structural model produces a complete 3D representation of the environment. The fact that the transmission link between the on-site cameras and the local workstation is a simple TCP/IP connection imposes severe limitations in terms of transmission rate. However, it is important to note that the appearance model is actually made of textural information extracted from the available images. All the processing required to build this model can be done at the remote site before transmission. To do so, the remote computational unit uses the knowledge available concerning the scene geometry in order to extract the required textural information. The segmentation of the appearance model into small textural elements can be advantageously exploited to reduce the transmission rate. It first ensures that no redundant information will be sent. It also allows transmitting new visual information only when significant changes occur in the visual aspect of the scene and when these changes affect the currently observed portion of the 3D model.

In order to obtain a complete 3D representation of the scene, these two models must be combined at the local unit. This means that an accurate registration between the 3D model and its appearance model must be established. This is to say that the relative position of the cameras (used to extract the appearance model) with respect to the environment (modelized by the structural model) must be accurately determined. This is a classical pose estimation problem [6]. However, the problem is again facilitated by the particular configuration of the proposed system. Indeed, since fixed cameras are used, good initial camera pose estimates are available. From these initial positions, a more accurate estimation of each camera pose can be obtained by minimizing the difference between the projected 3D model and the observed images. Once an accurate registration established, any discrepancy between the two models (that might be caused by small changes in camera positions) can be detected and a new pose estimation process can be reinitiated.

When a telepresence session starts, the structural model is downloaded into the local 3D visualization tool. Using the facilities offered by this tool, the user can then freely navigate inside the remote environment During the session, the application regularly receives new updates of the appearance model that are integrated to the 3D model and that are thus made visible to the user.

3.2 Benefits

The existence of the two independent models constitutes the central aspect of the system and has the following two important benefits:

1) *Because a 3D model of the scene under observation is locally available, the frequency at which visual information obtained from the cameras is refreshed can be kept low*. Indeed, since all navigation operations are based on a local 3D model, realistic real-time exploration of the site does not necessitate obtaining new visual information from the remote site each time a request for a new virtual viewpoint is issued. The required visual information is rather obtained from the integration of the images so far received. These are transmitted only at rate sufficient to ensure that the visual appearance of the model realistically reflects the conditions that currently prevails at the remote site. Therefore, in spite of the large amount of information generated by this multi-cameras system, it is possible to keep low the amount of data transmitted and to avoid any needs for a high-speed dedicated communication link.

2) *Because real textures (extracted from the images of the site) will be mapped on the 3D model, this latter does not have to be very accurate*. It is indeed extremely difficult to build very accurate models of existing 3D environments. However, it has been observed that the use of natural textures mapped on a 3D model can provide the human visual system with several visual cues that compensate for the inaccuracies in the 3D representation [9]. This means that the finest details in a real structure can be ignored in the corresponding 3D model and be represented only in the appearance model as captured by the cameras. This releases an important obstacle that often limits the applicability of modeling systems in a real context.

4. AN EXPERIMENTAL SETUP

The feasibility of the system has been demonstrated on a simple indoor scene. This scene is observed by two cameras. In order to build the structural model, the PhotoModeler software tool has been used. This 3D model has then been exported into our Java3D applet.

When the system is in operation, the two cameras continuously acquire images of the remote site. These images are used to build the appearance model. The fact that the scene is made of planes and that the cameras location is known greatly simplify the texture extraction process. A homographic transformation can indeed be computed and applied such that a fronto-parallel view of any of the observed plane can be obtained [7]. The thus extracted textural information is then regularly uploaded to the server. When a new appearance model is available, the applet automatically incorporate it into the Java 3D model manipulated by the user. In our experiment, the refresh rate of the appearance model has been set to 20 sec. It is however important to note that this frequency does not affect the exploration activity. The user can indeed freely move across the scene because the navigation process is strictly based on the structural model. The specified refresh rate of the appearance model simply guarantees that what the user observed is at most 20 seconds old. Right side of Figure 1 shows snapshots of a remote site exploration session. During this session, a document has been deposited on the table. Immediately, this one becomes locally visible and can thus be examined by the user.

5. CONCLUSION

The objective of this paper was to describe an approach for a telepresence system for remote site exploration that can offer at the same time both *real-time unrestricted navigation* and *live information*. The strategy proposed here is to use two distinct and complementary models that are combined at the local unit.

An experimental version was built to demonstrate the feasibility and the potential of the concept. We are now working on an extension of the system that uses a more complex indoor scene. We plan also to work with outdoor environments and with moving elements such as robots.

6. ACKNOWLEDGMENTS

The author wishes to thank Pascal Bonnevay and Vincent Brun for their help in the implementation of this system.

7. REFERENCES

[1] Cao, Z.L., Oh, S.J., Hall, E.L., “Dynamic Omni Vision for Mobile Robots”, *J. Robotics Systems* 3(1):5-17, 1986.

[2] Fuchs, H., Bishop, G. McMillan, A.L., Bajscv, R., Wook, S., Farid, H., “Virtual Space Teleconferencing using a Sea of Cameras”, in *Proc. of Int. Symp. on Medical Robotics and Computer Assisted Surgery* Sept. 1994.

[3] Gagalowicz, A., "Tools for Advanced Telepresence Systems". *Computer & Graphics*, 19(1):73-88, 1995.

[4] Halme, A., Suomela, J., "Applying telepresence and augmented reality to teleoperate field robots". *Robotics and Autonomous Systems*, 26:117-125, 1999.

[5] Hartley, R., Zisserman, A., "Multiple View Geometry", *Cambridge University Press*, 2000.

[6] Joseph, S.H., "Optimal Pose estimation in Two and Three Dimensions". *Computer Vision and Image Understanding*, 73(2) :215-231, 1999.

[7] Laganière, R. "Compositing a Bird's Eye View Mosaic", in *Vision Interface'2000*, pp. 382-387, 2000.

[8] Mair, G.M., "Telepresence - The Technology And Its Economic And Social Implications", in *Proceedings of the IEEE International Symposium on Technology and Society*, pp. 118 – 124, 1997.

[9] Nyland, L., "The Impact of Dense Range Data on Graphics", in *Workshop on Multi-View Modeling*, pp. 3-10, 1999.

[10] Onoe, Y., Yamazawa, K., "Telepresence by Real-Time View-Dependent Image Generation from Omnidirectional Video Stream". *Computer Vision and Image Understanding*, 71(2):154-165, 1998.

[11] Pollefeys, M., Koch, R., Vergauwen, M., Van Gool, L., "Metric 3D Surface Reconstruction from Image Sequences", in *SMILE workshop*, pp. 139-154, Freiburg, Germany, 1998.

[12] Rander, P.W., Kanade, T., "Constructing Virtual Worlds Using Dense Stereo". in *Proc. of ICCV97*, pp. 3-10, 1997.

[13] Sequeira, V., Ng, K., Wolfart, E., Gonçalves, J.G.M., Hogg, D.C., "Automated Reconstruction of 3D Models from Real Environments". *ISPRS Journal of Photogrammetry and Remote Sensing*, 54:1-22, 1999.

[14] Whitaker, R., Gregor, J., Chen, P., "Indoor Scene Reconstruction from Range Images", in *Proc. Conf. on 3D Digital Imaging and Modeling*, pp. 348-357, Canada, 1999.

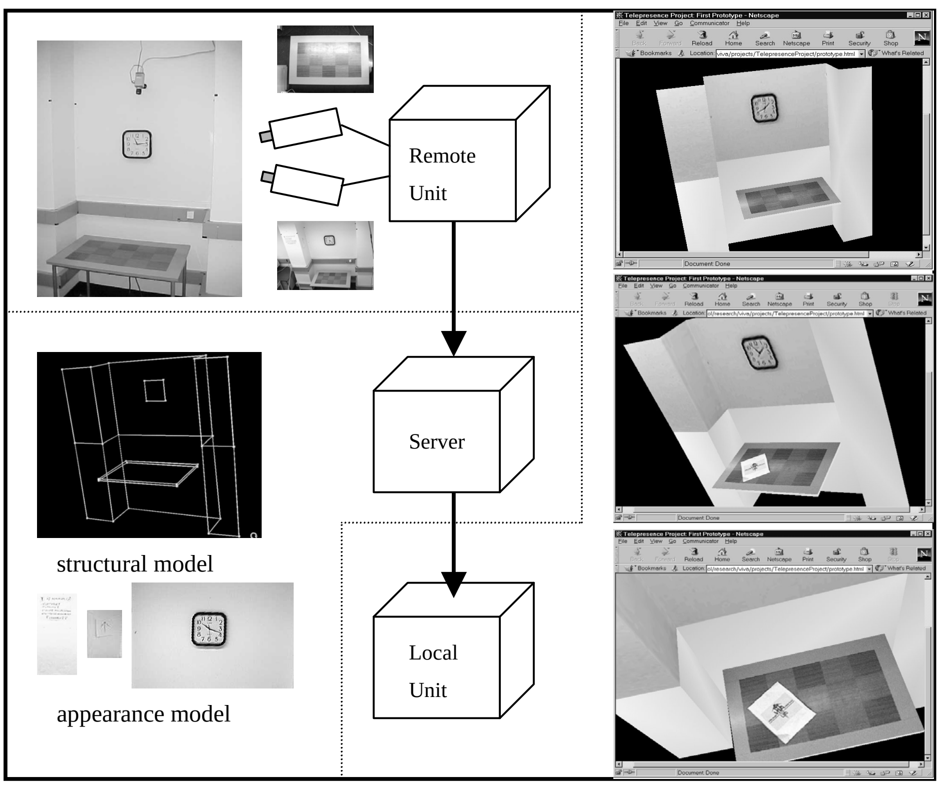

Figure 1. The system architecture.