Complexity Issues in Virtual Environments

Sudhanshu K Semwal

Department of Computer Science
University of Colorado, Colorado Springs
CO, 80933-7150, USA
semwal@redcloud.uccs.edu

Abstract

Virtual environments are expected to provide a complete range of human interactions and emotions to participants. Holistic virtual environments are to include both the reductionist and the non-reductionist approaches for modeling complex human interactions. Implications and benefits of holistic virtual environments are discussed in the context of avatar research. There is a correlation between observable motion of the participants and their emotions. Therefore, full body unencumbered tracking, without any wires and gadgets, is beneficial in virtual environments. For unencumbered tracking, camera-based systems area ideal, however, these systems require satisfactory resolution of correspondence across multiple camera-images. We have been developing the Scan&Track virtual environment with a goal of designing a holistic virtual environment. The Scan&Track system is to provide full body unencumbered tracking using the active-space indexing method. We will discuss the use of the active-space indexing method in resolving correspondence across multiple camera-images, and present results of our implementations.

Keywords: Camera based virtual environments, hyper-realistic avatars, R and NR complexity.

1. Motivation

Virtual Environments (VEs) are expected to facilitate real-time interaction with human participants by using five major senses [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11]. These major channels provide means of interaction as well as tools for communication. The existence of several other senses, such as sense of fear, sense of well being, are also aspects of human interaction. The present technological focus of Virtual Reality (VR) is mainly on the five primary channels. Such a reductionist approach has been only partial successful, and can be considered flawed [12, 13]. Thus there is a need to systematically study and create *Holistic* Virtual Environments which are sensitive to the emotional, psychological needs of the participants connected through a virtual environment.

Even in such modern times, a completely different message can be communicated due to language barriers, social, and cultural differences. Gestures and styles in interaction also vary with where you are. These interaction-styles have been developed through ages geared towards optimizing the efficiency of interaction with the surrounding circumstances or nature. So Holistic Environments must work with these differences, and address verbal, non-verbal, and other multi-modal communication channels. Human participants would be modeled as dynamic, complex multi-dimensional systems, capable of chaotic behavior.

Some theoretical questions arise and must be addressed: (a) How can man-made (Virtual) Environment pose, or at least pretend to be warm and fuzzy, like humans, and have emotions. (b) Can a Virtual Environment (VE) really simulate a human being? (c) Is it necessary for a Virtual Environment to interact like a human being? (d) What do we want to achieve in a VE? (e) What are our motivations and goals? (f) What would be the criteria to determine if the virtual environment facilitated ***human-like*** interaction? (g) Would a Virtual Environment ever be able to replace another human's touch, presence, words, or body?

Although these questions can not be fully answered at this time, and *maybe* seems to be the best answer to these above questions. We believe that systematic study and seeking answers to the above questions would be an important application of our efforts in the future.

It is generally accepted that the sense of smell is useful in providing mood changes. In addition, smell may also change the taste of what we are eating. As the sense of smell and taste are related, both psychological and functional effects of smell on the participants are to be also studied in Holistic Virtual Environments.

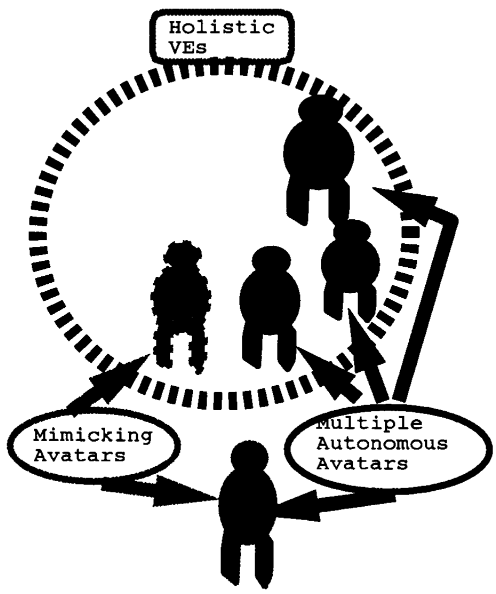

Figure 1: Virtual Environment with hyper-realistic Avatars.

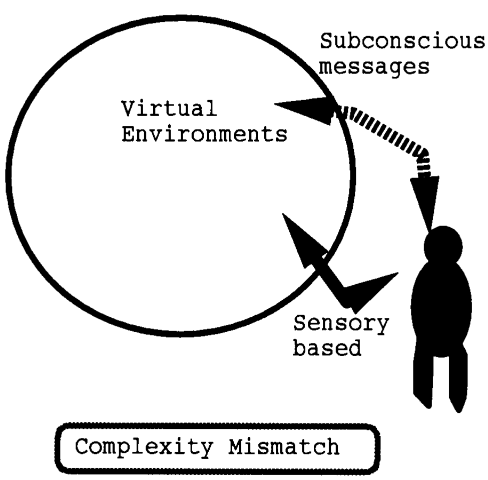

Figure 2: Complex communication in a Virtual Environment.

2. Possible Applications of Holistic Virtual Environments

As mentioned earlier, Holistic environments would create a new dimension to existing research by addressing human-sensitivity issues, especially the emotional and non-verbal aspects of human-communication. Sub-conscious cues and indirect message are also to be handled in these environments. This will provide new richness of experiences to participants. Emotional evaluation, emotion generation, emotion synthesis, and control of emotional contents are new dimensions and options available to participants in holistic virtual environments.

Due to psychological reasons, we would like to create an avatar which would be almost exactly-similar to us (Figure 1). This leads to new possibilities and expansion of human-expression and communication, that of autonomous-avatar and mimicking-avatars, both existing at the same time in a Holistic virtual environment. Autonomous-avatars may represent some specific set of motions, emotions, and capabilities in a particular domain. Multiple autonomous-avatars and their interactions with mimicking-avatars would create new paradigms of human interaction in virtual environments (Figure 1). This indeed then would bridge the gap between space and time as full body avatars can create a feasibility that *virtual* and *physical* presence are indeed identical. This creates a variety of possibly new social interactions. A new type of medium would have to be created to deal with the holistic (or intrinsic) nature of participants instead of dealing only with their senses [12].

Although 2D-images and 3D-graphical models, representing avatars of the participants, have provided interesting applications, we further raise our expectations of avatars [9] in holistic communications to another level. Avatars in holistic communication are to go beyond mimicking the actions of the participants. In holistic virtual environment we expect to address several new possibilities: (i) can avatars be autonomous enough to complete an interaction with other autonomous-avatars, and form an agreement acceptable with *their* human participants. They can also set up introductions before the *real* (busy) participants interact.

One important issue which holistic virtual environments must address is to assimilate a variety of social, psychological, cultural, and complexity issues. Large complex virtual environments provide both clutter and useful information, the issue then would be *what* information to provide to the participants, and *when*. The context based presentation is important. Change of domain could change the avatar's knowledge model. For example, multiple domain avatars can assist the participant to deal with other avatars and participants. These domain specific avatars will fine-tune their communication channels based upon the audience the intended message is targeted for.

In holistic virtual environments, we are interested in providing with new tools and modalities for human-to-human communication. So holistic virtual environments are expected to grow as the computation, knowledge, and time spent by the participants in these environments grows. This opens possibilities of a variety of holistic communications: personal, professional, and a variety of group activities.

There is also a need to systematically study long-standing questions: (i) Since human participants and different forms of their avatars, are present in holistic virtual environments, how can we measure the *intelligence* in these environments. (ii) Is there an intelligence in these environments which could be compared to ***human*** intelligence (iii) Can a virtual environment ever understand human participants to predict the participant's next action, (iv) what is the relationship between the knowledge of human participants and that of their avatars.

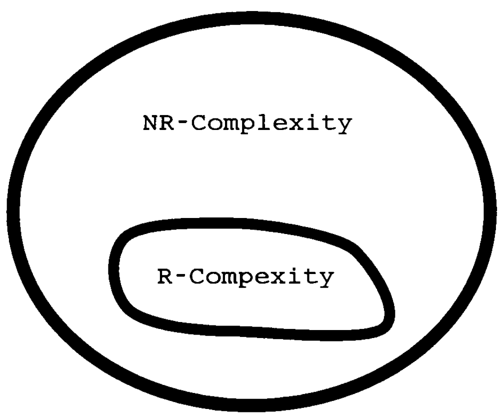

Figure 3: The Relationship between R and NR complexity (Note the similarity with P and NP completeness [14].

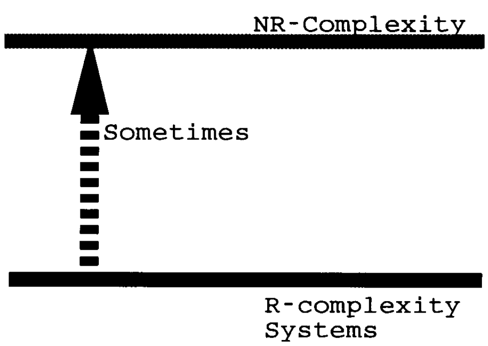

Figure 4: Sometimes R-complex virtual environment might provide NR-complexity, but not consistently.

3. R and NR Complexity

Movies and arts provide powerful emotional experiences. However, there are only handful of virtual environments where powerful emotions can be created, and felt. Emotions deal with subconscious minds as well, so there is a need in virtual environments to generate subconscious messages (Figure 2) [15, 16, 17, 18]. We also must acknowledge the duality of mind and body. Subconscious subtle cues and emotions have been extensively used in movies, novels, and arts. We need to develop similar methods to directly appeal to mind along with the five major senses.

When we consider the realistic emotions displayed in animated movies, a large amount of information is presented in a short time. For example, a full length 90 minute movie may consist of over two million frames to be generated [19]. Much higher performance is expected in Holistic Virtual environments. So we at least would need hyper-realistic avatars so that there is absolute similarity between the participant and its avatar in appearance. If there is no noticeable difference between the participant and avatar then experiences of the participants would be more profound to the participants at the sub-conscious level.

There are two sets of complexities which a Holistic environment can provide: R-complexity and NR-Complexity (Figure 3). The virtual environments provide R-complexity by using structured, domain specific knowledge where certain things happen when specific conditions are met (reductionist approach). The virtual environments providing NR-complexity are those where simple interactions are capable of generating interactions of arbitrarily large complexity. A Virtual environment, that interprets facial features using Ekman's [20] models for classification of emotions, can be considered an example of R-complex Holistic virtual environment, as it is based on the reductionist approach for understanding facial expressions. When the VE is capable of creating complex emotions from simple local expressions and actions, they are said to provide NR-Complexity. The idea of NR-complexity has been already explored in the efforts in [21] for procedural animations. The NR-complexity relates to subconscious models which are not rule based [12]. Both recognition and generation of emotions in virtual environments, and providing NR-complexity, is an open question. However some work has already been done. Sometimes, under certain circumstances systems exhibiting R-complexity may also provide NR-complexity by establishing some serenity or connection with the participants for example, the IamAScope experiments [22] or experiments conducted for dealing with the fear of flying [23]. Sometimes, there is an indirect message of calm and peace. These messages are subtle and can be lost (hence the dotted line in Figure 4). What is the best way to provide indirect, subtle, subconscious messages *consistently* in virtual environments is still an open question. We also must address the following question: Can Virtual Environments providing R-complexity also provide NR-complexity? In other words, Is R=NR?

4. Unencumbered Methodologies

There have been several attempts to track participants in a virtual environment. These can be broadly divided into two main categories: encumbering and non-encumbering [24]. A variety of encumbering tracking devices have been developed, for example acoustic [24], optical [25], mechanical [1, 26, 27], bio-controlled, and magnetic trackers [4]. For Holistic applications, it is essential that participants are not tethered by wires or gadgets, or the work area could be large, and may contain magnetic interference so encumbered methods are impractical or undesirable. Study of human-emotions related to motion patterns, facial and gesture applications in virtual environments are some other examples where unencumbered 3D tracking has been useful.

Camera-based techniques are well suited for developing unencumbering 3D tracking. For determining the 3D position of the participants from multiple cameras, both occlusion and correspondence across multiple cameras are inherent in camera-based systems [24]. The occlusion problem is more severe when multiple participants are present in a camera-based virtual environment. To date, there are no camera-based virtual environments where the occlusion-problem has been solved completely to provide real-time interaction.

The well known correspondence problem is also ever-present when purely camera-based systems are used, and require several compromised methods. Extraction of only a few key-features or significant points from an image, has been investigated in great detail for this purpose [6]. The idea behind significant point extraction is that a small collection of points may provide enough information about the intent or pose of the participant. Blob models [3] are used in the Pfinder. Recently, there has been an interest in pursuing 3D tracking using a relatively larger set of spatially distributed data points [5]. We also use a 3D grid of points for implementing the Scan&Track system.

5. The Scan&Track Virtual Environment

We are developing the Scan&Track virtual environment using multiple cameras. The main difference between the proposed Scan&Track system and other existing camera-based virtual environments is that correspondence and occlusion solution is being resolved for a much larger set of points in the Scan&Track system. We are attempting to directly estimate the 3D locations of points for creating hyper-realistic avatars. The number of such points can be large (thousands) instead of only a few (tens of) significant points. So we expect much realistic 3D surfaces for representing an avatar and a truer resemblance of an avatar to the participant. This is different than using synthetic human forms as in [28, 29]. Although we are far from hyper-realistic avatars at this time, we have obtained good results in 3D position estimation and resolving correspondence.

The active-space access used in the Scan&Track system provides a framework for unencumbered 3D tracking based upon multiple video sequences. The active-space access method uses points on a regular 3D grid for estimating precise 3D position [7].

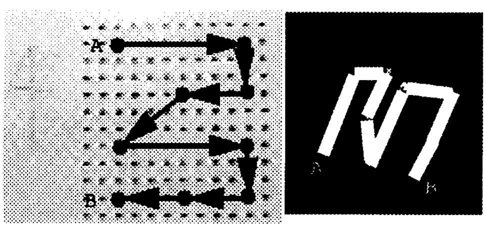

Figure 5: Extraction of 3D positions given nine imprint-sets.

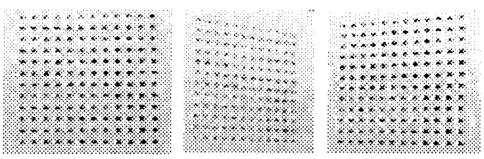

Figure 6: Three camera images with dot patterns.

Initial results of our implementation are shown in Figures 5-8. Precise 3D position estimation using the active space accessing method is possible (Figure 5) given a set of corresponding points. Correspondence solutions are being developed and our feasibility studies are quite encouraging. The three images from left, center and right cameras are shown in Figure 6. The candidates for corresponding points are identified in Figure 7. If there are n1, n2, and n3 candidate points identified in left, center, and right camera-images, then using a brute force approach $n1 \times n2 \times n3$ unique corresponding sets will have to be compared. This is a rather large number. For example, typically, n1, n2 and n3 are around 400 to 500, so brute force must consider 64 to 125 million imprint-set. This was quite time consuming for even the multi-processor Onyx machine, taking several hours. In the Scan&Track system, the active-space indexing allows an efficient filtering algorithm so that correspondence was established in one to two minutes. This is around 1000 times or more faster algorithm. Corresponded points are shown in Figure 8.

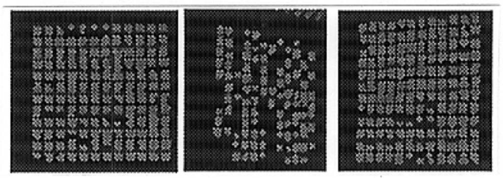

Figure 7: Extracting imprint-sets for camera-images.

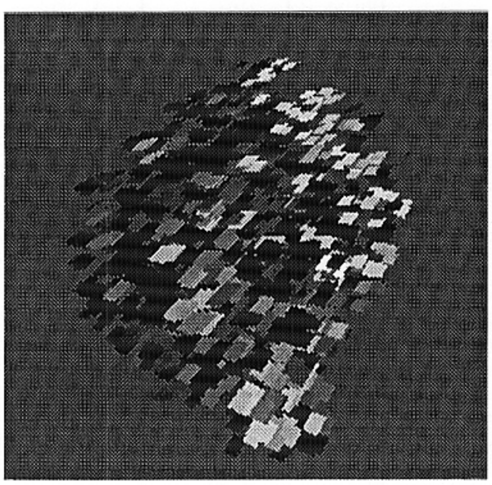

Figure 8: 3D points corresponding to every imprint-set identified in Figure 7.

6. Concluding Remarks

The Scan&Track system uses a robust method of localized estimation and correspondence. We hope to develop hyper-realistic avatars as a first step towards developing the Holistic Scan&Track virtual environment. We have also presented our motivation for two classes of virtual environments providing R and NR complexity. A variety of open questions and methodologies were also discussed.

We expect Holistic Virtual Environments to have a major impact on the computer-human interaction in near future, and would provide new technology and products. Holistic virtual environments would facilitate a forum for a new type of interaction, and provide a better medium of expression to pursue and expand human-activity and emotions.

Acknowledgments

Part of this research was conducted when the author was at the ATR Media Integration and Communication Research Laboratories, at Kyoto, Japan. Special thanks to Dr. R. Nakatsu, President of ATR Media Integration & Communications (MIC) Research Laboratories, and Dr. Jun Ohya, Manager, Department 1, for providing all the necessary support during my stay at ATR.

References

[1] I.E. Sutherland. A head-mounted three dimensional display *ACM Joint Computer Conference* 33(1), pp 757-764 (1968).

[2] M.W. Krueger. *Artificial Reality II.* Addison Wesley Publishing Company, Reading, MA, pp 1-277 (1991).

[3] C Wren, A Azarbayejani, T Darrell, A Pentland. Pfinder: Real-Time Tracking of the Human Body, International Conference on Automatic Face and Gesture Recognition, 1996, pp 51-56 (1996).

[4] NI Badler, MJ Hollick, and JP Granieri. Real-Time Control of a Virtual Human using Minimal Sensors, *PRESENCE*, 2(1), pp 82-86 (1993).

[5] PJ Narayan, PW Rander, and T Kanade. Constructing Virtual Worlds using dense stereo, Proceedings ICCV, pp. 3-10 (1998).

[6] I Haritaoglu, D Harwood, LS Davis. W4: Who? When? Where? What?A Real Time System for Detecting and Tracking People, accepted for publication at the Third IEEE International Conference on Automatic Face and Gesture Recognition April 14-16, Nara, Japan, IEEE Computer Society (1998).

[7] Sudhanshu Kumar Semwal and Jun Ohya. The Scan&Track Virtual Environment. The first International Virtual Worlds (VW98) conference,

Paris, July 1-4, 1998, Springer Verlag. Lecture Notes in Artificial Intelligence 1434, pp. 63-80 (1998).

[8] Sudhanshu Kumar Semwal and Jun Ohya. Geometric-Imprints: A Significant Points Extraction Method for the Scan&Track Virtual Environment, IEEE Computer Society Third International Face and Gesture Recognition Conference, Nagaoya, Japan, pp. 480-485, (1998).

[9] Sudhanshu Kumar Semwal, Ron Hightower, and Sharon Stansfield. Constant Time Mapping Algorithms for Real-Time Control of an Avatar using Eight Sensors, PRESENCE journal, vol 6, no 1, pp 1-21, MIT Press (1998).

[10] Proceedings of the Second International Conference on Automatic Face and Gesture Recognition, October 14-16 1996,Killington, Vermont, USA, IEEE Computer Society Press (1996).

[11] Proceedings of the Second International Workshop on Automatic Face and Gesture Recognition, June 26-28, 1995, Zurich, Switzerland, IEEE Computer Society Press (1996).

[12] Against Cognitivism Alternative Foundations for Cognitive Psychology edited by Arthur Still and Alan Costall, Harvester Wheatsheaf, Great Britain (1997).

[13] R Rivlin and K Gravelle. Deciphering the Senses, The expanding world of human perception. Chapter 1, pp. 9-28. Simon & Schuster, Inc. New York (1984).

[14] Michael R. Garey and Davis S. Johnson. Computers and Intractability: A guide to the Theory of NP-Completeness. pp. 34, WH Freeman and Company (1978).

[15] H. Haken. Principles of Brain Functioning, A Synergetic Approach to Brain Activity, Behavior, and Cognition, Springer Verlag (1995).

[16] David J. Depew and Bruce H. Weber. Darwinism Evolving, The MIT Press, Cambridge, MA (1995).

[17] Rudolf Groner, Gary dYdewalle, and R Parham. From Eye to Mind, North-Holland, Amsterdam (1990).

[18] John R. Searle. Intentionality: An Essay in the Philosophy of Mind, Cambridge University Press (1983).

[19] Frank Thomas and Ollie Johnston. The Illusion of Life Disney Animation. pp. 317, Walt Disney Productions (1981).

[20] Paul Ekman (editor). Darwin and Facial Expressions: A century of Research in Review, Academic Press (1973).

[21] Karl Sims. Interactive evolution of equations for procedural models. The Visual Computer, vol. 9, pp. 466-476, springer-Verlag (1993).

[22] Sydney Fels. IamAScope exhibit, ATR MIC Research Laboratories (1997).

[23] L.F. Hodges, B.O. Rothbaum, B. Watson, G.D. Kessler, and D. Opdyke. A Virtual Airplane for Fear of Flying Therapy, IEEE VRAIS 1996 conference Proceedings, pp. 86-93, IEEE Computer Society (1996).

[24] K. Meyer, H.L. Applewhite, and F.A. Biocca. A Survey of Position Trackers. *PRESENCE*, 1(2), pp 173-200, MIT Press (1992).

[25] RF Rashid. Towards a system for the Interpretation of Moving Light Displays, *IEEE Transaction on PAMI*, 2(6) pp 574-581 (1980).

[26] F.P. Brooks, M.J. Ouh-Young, J.J. Batter, and P.J. Kilpatrick. Project GROPE – Haptic Displays for Scientific Visualization. Proceedings of *SIGGRAPH 1990* 24(4), pp 177-185 (1990).

[27] H. Iwata. Artificial Reality with Force-Feedback: Development of Desktop Virtual Space with Compact Master Manipulator, Proceedings of *SIGGRAPH 1990* 24(4), pp 165-170 (1990).

[28] TK Capin, H Noser, D Thalmann, IS Pandzic, NM Thalmann. *Virtual Human Representation and Communication in VLNet*, vol. 17, No. 2, pp. 42-53 (1997).

[29] JK Hodgins, WL Wooten, DC Brogan, JF O'Brien. Animating Human Athletes, In *Computer Graphics* Proceedings, Annual Conference Series, ACM SIGGRAPH, 71-78 (1995).