

Generating Semantic Information for Virtual Environments

J. Timothy Balint¹ and Jan M. Allbeck²

Abstract—In order to build virtual humans with pleasing interactions in virtual scenarios, a consistent representation of all relevant relations and information is required. For virtual scenarios, this equates to having a meaningful (semantic) understanding of the objects and virtual agent performable actions in the scenario. The process of creating semantics for rich virtual environments is a tedious and time consuming endeavor, such that only the most pertinent pieces of information are added to the most used objects in the scene. This has the effect of decreasing realism as only a few objects can be acted upon. Furthermore, authoring relations by hand does not guarantee consistent semantics, which can also erode realism. We discuss the need and progress in automating semantic generation for objects and actions in virtual scenarios. We provide special consideration for semantics connecting objects to actions, known as operational information. We also examine the need for automated methods to generate domain specific knowledge such that a virtual environment is not overburdened by semantic information but instead finds a balance of relations.

I. INTRODUCTION

Virtual humans have played an impressive role in creating more realistic virtual scenarios. They provide ambiance as Non-Player Characters in games [1], and are essential for testing out new scenarios before physical humans work in them [2], [3]. For a virtual human to have a meaningful impact on the environment, it requires the ability to move and interact with objects and other agents. This can be as simple as *walking in a straight line in a room* or as complex as *ordering a drink*, which would require several animations and connections to objects and other agents in the environment. One of the grand challenges to increased levels of presence in virtual reality is taking into account the expectation of participants as those expectations relate to real human behaviors [4]. While this is generally the goal of virtual human simulation, having uncommon or atypical behavior is sometimes required, especially when virtual humans are used for training simulations [5]. Uncommon behavior may also be emergent, which provides plausible action combinations that the simulation author does not intend, but is appropriate and pleasing in the simulation. Behaviors portrayed through animations performed by a virtual human can be executed either sequentially or in parallel as long as an understanding is in place for the proper order of the actions. Animations can be controlled by scripting actions, through state machines or decision networks [6], and is a common method in game

design to piece together animations and ground them in a 3-D virtual environment.

Rich virtual environments are complex, containing several dozens or even hundreds of objects. Creating agent behavior scripts for each action and object desired in a scenario can be a time-consuming and tedious process. A mistake in one script can have immersion breaking impact on the scenario. It also makes it more difficult to have emergent actions occur in a simulation. Modulization of domain knowledge on graphical models and animations can mitigate errors, as the scripts can then use known information parameters, (i.e. semantics), such as *density* and *hue*. Note that these properties do not have to be directly related to the graphical objects, but instead can provide precise knowledge that can be used in other systems of the scenario, such as qualitative physics engines [7]. Animated actions can also have semantics attached to them in a similar way graphical objects do, such as the *goal* or *type* of animation. Action semantics are generally used to modularize scripts themselves, allowing high level agent decision making processes to determine which action should be performed.

Development of semantics for virtual humans is generally an author-driven process, in that it still requires a user to provide grounded values for all semantics. Even with templates, as the number of unique objects and actions grows it becomes prohibitive to author large amounts of semantic data. Furthermore, semantic relations which require several relationships in their representation, such as operational information which connects virtual human actions to all participant objects, must have knowledge of the purpose and effect of each part of the relationship. These requirements cause semantic development to be ad-hoc in nature, crippling the intended purpose of creating domain focused knowledge; namely transferability from one scenario to a similar one. This work creates the foundation for the claim that automated inclusion of action and object semantics is critical to increasing the utility of virtual environments. Consistent, automated methods of semantic generation would allow transferability of relationships between scenarios, as they should in theory generate similar semantics for similar scenarios.

In order to determine the usefulness of automated methods for semantics in virtual scenarios, it is important to understand the type of semantics required for virtual agents and environments. Then, requirements for consistent automated techniques for semantic generation for virtual agents and environments can be determined. This paper is organized in the following manner: Section II lays the foundation for semantic information of virtual objects and describes how semantics are used by other processes in a virtual simulation. Section III

¹J. Balint is Graduate Student in the Department of Computer Science, George Mason University, 4400 University Drive, Fairfax, VA jbalint2@gmu.edu

²J. Allbeck is Associate Professor in the Department of Computer Science, George Mason University, 4400 University Drive, Fairfax, VA jallbeck@gmu.edu

describes semantics attached to animations, mainly for use by virtual humans. It also goes into detail on *interactional* or *operational* information, which is used by virtual characters to determine how to combine and interact with objects. We use these two sections as a basis for dividing an ontology for virtual humans into an action and object component. Finally Section IV describes methods to automate the generation of semantics, and argues the need for methods that generate pertinent semantics for agent actions. It also argues why ontologies connected to a virtual scenario aid in the creation of these semantics.

II. SEMANTIC INFORMATION OF OBJECTS

Semantic information of objects has been a widely studied topic in virtual environments and a survey on the subject can be found in [8]. One interesting topic in the semantics of objects is that of Smart Objects [9], [10], which have also gone by the name of environmental object models [11], with the difference being the use of an ontology in the later to manage domain knowledge. Both of these models allow simulation authors to create templates with specific properties that can be filled in. For example, a *door* object can have a *wooden* or *metal* property. From examination, properties can be set in a templated slot either through binary, set, or continuous semantics. Binary semantics are either true or false meanings that are applied to every object, and differ from continuous semantics (such as temperature) and set semantics (such as RGB color) in that binary semantics do not need to be grounded by an author at run-time (in the general case). Other aspects of a virtual simulation can take advantage of these known slots, such as having all *wooden* items burn. This allows an author to add a structured sense of realism to the scenario.

Using semantics has had a profound impact on the types of background interactions that take place in a virtual environment. Lugin and Cavazza used an ontology of semantic relations and a differentiation of recognition and execution of actions in order to create more emergent simulations [12]. Semantics of virtual objects have also been used to create virtual environments, either as a tool aiding simulation authors to create more plausible environments [13] or through automated generation of environments [14]. More emergent simulations in both creation and effect provide a more immersive experience for a physical human, but unless it is described to a virtual human, will still go unnoticed. Therefore, any semantic system for objects must also capture the changing information, and provide that to virtual humans.

Virtual humans capture object semantics through the use of perception and attention [15], [11], [16]. Perception and attention methods focus a virtual human by only allowing the character to know of objects in their environment that a physical human would notice. This changes and simplifies the objects that a virtual human would consider when deciding on actions. The modular nature of object semantics allows comparisons to be made between different objects and quickly determine what is important for the virtual agent to know in the environment. Salient methods that

process the environment by generating true graphical or false color images require creating and then processing the image. For groups of virtual agents, this means generating several frames from different perspectives (one from each agent) and then running computer vision techniques on each. While this scales well for the number of objects in the scene, it becomes prohibitive as the number of agents increase. Semantic methods, especially ones that allow for a large amount of overlap in processing, allow for more virtual agents to have more human-like behaviors.

III. SEMANTIC INFORMATION OF ACTIONS

Semantic information of actions affords virtual humans the ability to compare and decide on which action to perform. In many cases, action semantics consist of goal states [17], [18], and a few formal languages designed for agent decision use goal states [19], [20]. Goal oriented decision making is a straight-forward approach to determine the agent's next action and determining those goals is the approach of Belief-Desire-Intention (BDI) agents. When two similar actions with the same end goal are considered, other semantics pertinent to the scenario, such as the duration or necessary pre-conditions of the action, allow for one action to be chosen over another. This shows that action semantics originate in the artificial intelligence community and are used in agent decision making.

In the virtual agent's community there have been several proposals to connect actions to objects, mainly following the methodology of Smart Objects [9]. The terminology from these proposals is not consistent and can be compared in Table I, found in the appendix. As there are slight differences in the definitions, this work will use *interactional* information when referring to the general case of action semantics attached to objects, such that the interaction is known to the virtual character through the object. Interaction information is a form of action semantics that connects 3-D graphical models with the animations a virtual human would perform, with the term being coined in Farenc et al. [21]. Peters et al. [22] introduced the idea of interactional information as slots, allowing multiple agents to interact using an object. This consisted of having a simulation author create a site-action pair and link connections with actions through a script as seen in Figure 1a. Donikian and Paris [23] used the concept of *affordances* to attach virtual objects to animations. Heckel and Youngblood [24] further removed affordance connections from the virtual agent's consideration if the smart object was not in a correct state to be used in the action. Finally, Kraayenbrink et al. [2] used interactional information to control action selection for populations of virtual humans. The progress made in these works is meant to ease the cognitive decision process that virtual characters must make when selecting actions to perform given available resources in the scenario.

Interactional information connects smart objects to virtual human animation through the use of action-site combinations. This relationship between actions and objects is in almost all cases connected to the object, providing the scripts

able to generate a handle for a grasp site or find large flat surfaces for a sit action from a labeled set of parts. For complex actions involving several objects, or for objects that use actions not based off of affordances, other methods to generate semantics are needed.

Semantic information for objects and actions to be stored is in an ontology. Ontologies contain relationships and taxonomies, such that information can be attached to an abstract concept, and links to that abstract concept can be assumed to have relationships between them. Taxonomies also allow for semantic data to be placed at more general abstractions of an object or action, reducing the redundancy between objects. Ontologies have been used by several virtual simulations [29], [33], [12], [34] to connect and provide meaning to the objects and actions virtual humans must interact with. Ontology research such as Drumond and Girardi [35] makes a distinction between concepts with a taxonomic relationship (*IsA* or *IsPartOf* relationships) and non-taxonomic relationships (such as operational information). This break-down of relationships becomes a key concept in ontological learning methods.

If we assume that the 3-D models and animation information is connected together in an ontology, then techniques used in ontology generation become available for semantic generation. The field of ontology generation and learning from large data-sets is a well studied topic, with several books [36], [37] and survey papers written on the topic [35]. Ontology learning from structured and semi-structured data sources generates consistent hierarchies over different parameters, as the learning methods are essentially pruning away un-necessary data. Techniques to generate ontologies for virtual humans have made some progress using semi-structured data sources such as WordNet[38].

A. Ontology Generation for Virtual Humans

Throughout this paper, it has been made clear that there are two important pieces of virtual humans interacting in virtual environments, the 3-D graphical models (objects) and animations (actions). These pieces are fundamentally distinct and their creation and use, connected strictly through interactional or operational information. Therefore, we can divide an ontology into two distinct hierarchies, tackling each separately. Dividing the ontology into objects and actions can focus the information extraction process, and remove the confusion attempting to process both at the same time entails. For example, graphical models and animations are related to nouns and verbs from natural language, and there are several words that can be both, such as a *cook*(model) who *cooks*(verb). If the names of the action or object are used to retrieve semantic information about it, then the system must determine its proper meaning from a larger set (the set of verbs and nouns). A smaller set would increase the accuracy of generation, and since objects and actions are disjoint, the author can propagate this distinction before semantic generation begins.

Information retrieval for virtual humans has seen progress in each of these two sub-problems. Pelkey and Allbeck [39]

used the names of 3-D models to build up a taxonomy of objects, and attach binary semantics to each object. In their paper, objects such as a *shrimp platter* are connected with a taxonomic relation to *food* and have a non-taxonomic relation to *edible*. Both of these semantics are attached to all shrimp-platters in the scene, and therefore, all *shrimp platter* are considered *edible*. For action semantics, Balint and Allbeck [27] created a taxonomy of actions, and populated operational information from an already generated object hierarchy, considering only the objects in question. They provide an example of *cooking*, which connect a *cook* action to a *food* object in a pre-built object ontology. Using these two methods together would connect *cooking* to a *shrimp platter* object. This method was tested against a ground truth, with a maximum recall achievable around 60%. It is discussed in the conclusion of this work that the other 40% of the operational information was contained through binary or set semantics that would be attached to objects.

Each of these techniques have been able to generate taxonomies and retrieve semantics for separate pieces of a virtual simulation. However, pieces of a virtual simulation should co-exist, that is, the actions a virtual human can perform should fit in with the objects available in an environment. For planning systems, having operational information is useful for determining what actions can be performed (at run-time or pre-planned), but operational connections can be decided during a specification phase, when the ontologies are generated. Provided a set of actions and object are known, the important domain knowledge for virtual humans in that system should exist before any planning system is necessary. Having an understanding of the actions and objects in a scenario should also allow a system to know what properties are not needed in the simulation. For example, if a simulation author does not have objects that have the property of *containing* a liquid, then it does not make sense to have an action require *containing* a liquid. If a simulation author then added an object with a *containing* property, the automated methods should propagate this change by rebuilding the ontology. From Pelkey and Allbeck, adding objects does not seriously impact the time to create a hierarchy, and so the generation systems could stay consistent, even with this change. Therefore, while progress has been made due to separating out the semantics, methods that will inevitably combine the relationships are still needed.

V. CONCLUSIONS

According to VR experts, one of the grand challenges to increased levels of presence is taking into account the expectation of participants as those expectation relate to real human behaviors, hence a need for additional domain knowledge [4]. In this paper we have discussed automated methods for assembling such semantics. While the use of semantics is constrained to what a simulation author can produce, methods to automate the generation and population of semantics for virtual scenarios are starting to be developed, especially methods to populate ontologies for virtual environments. These techniques benefit from dividing

the ontology into disjoint sets of objects and actions and determine semantics for each separately. The work on automatically populating ontologies for virtual humans is scarce compared to other information retrieval areas, but should be given considerable attention provided their utility in the area. One possible avenue of future work should be to determine if techniques can be developed that symbiotically build the ontology through each component, using the semantics of one to improve the retrieval of the other.

REFERENCES

- [1] B. Sunshine-Hill and N. I. Badler, "Perceptually realistic behavior through alibi generation," *AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment; Tenth Artificial Intelligence and Interactive Digital Entertainment Conference*, 2010.
- [2] N. Kraayenbrink, J. Kessing, T. TuteneL, G. de Haan, and R. Bidarra, "Semantic crowds," *Entertainment Computing*, vol. 5, no. 4, pp. 297–312, 2014.
- [3] A. Normoyle, M. Likhachev, and A. Safonova, "Stochastic Activity Authoring with Direct User Control," in *Proceedings of the 18th Meeting of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, ser. i3D '14. New York, NY, USA: ACM, 2014, pp. 31–38.
- [4] M. Slater, "Grand challenges in virtual environments," *Frontiers in Robotics and AI*, vol. 1, p. 3, 2014.
- [5] J. Linssen, T. de Groot, M. Theune, and M. Bruijnes, "Beyond simulations: serious games for training interpersonal skills in law enforcement," in *Proceedings of European Social Simulation Association*, 2014, pp. 604–607.
- [6] I. Millington and J. Funge, *Artificial Intelligence for Games, Second Edition*, 2nd ed. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2009.
- [7] M. Cavazza, S. Hartley, J.-L. Lugin, and M. Le Bras, "Qualitative physics in virtual environments," in *Proceedings of the 9th international conference on Intelligent user interfaces*. ACM, 2004, pp. 54–61.
- [8] T. TuteneL, R. Bidarra, R. M. Smelik, and K. J. D. Kraker, "The Role of Semantics in Games and Simulations," *Comput. Entertain.*, vol. 6, no. 4, pp. 57:1–57:35, Dec. 2008.
- [9] M. Kallmann and D. Thalmann, "Direct 3d Interaction with Smart Objects," in *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, ser. VRST '99. New York, NY, USA: ACM, 1999, pp. 124–130.
- [10] M. Gutierrez, F. Vexo, and D. Thalmann, "Semantics-based representation of virtual environments," *International journal of computer applications in technology*, vol. 23, no. 2-4, pp. 229–238, 2005.
- [11] J. van Oijen and F. Dignum, "Scalable Perception for BDI-Agents Embodied in Virtual Environments," in *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2011 IEEE/WIC/ACM International Conference on*, vol. 2, Aug. 2011, pp. 46–53.
- [12] J.-L. Lugin and M. Cavazza, "Making Sense of Virtual Environments: Action Representation, Grounding and Common Sense," in *Proceedings of the 12th International Conference on Intelligent User Interfaces*, ser. IUI '07. New York, NY, USA: ACM, 2007, pp. 225–234.
- [13] B. Pellens, O. De Troyer, W. Bille, and F. Kleinermann, "Conceptual Modeling of Object Behavior in a Virtual Environment," in *Virtual Concept*, Biarritz, France, 2005, p. Paper 69.
- [14] T. TuteneL, R. Bidarra, R. M. Smelik, and K. J. De Kraker, "Rule-based layout solving and its application to procedural interior generation," in *Proceedings of the CASA Workshop on 3D Advanced Media In Gaming And Simulation*, 2009.
- [15] P. Sequeira, M. Vala, and A. Paiva, "What Can I Do with This?: Finding Possible Interactions Between Characters and Objects," in *Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems*, ser. AAMAS '07. New York, NY, USA: ACM, 2007, pp. 5:1–5:7.
- [16] T. Balint and J. M. Allbeck, "Whats going on? multi-sense attention for virtual agents," in *Intelligent Virtual Agents*. Springer, 2013, pp. 349–357.
- [17] T. Knik and J. Laird, "Learning goal hierarchies from structured observations and expert annotations," in *Inductive Logic Programming*. Springer, 2004, pp. 198–215.
- [18] M. Dastani, M. B. van Riemsdijk, and M. Winikoff, "Rich Goal Types in Agent Programming," in *The 10th International Conference on Autonomous Agents and Multiagent Systems - Volume 1*, ser. AAMAS '11. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, 2011, pp. 405–412.
- [19] K. V. Hindriks, F. S. De Boer, W. Van Der Hoek, and J.-J. Ch. Meyer, "Agent Programming in 3apl," *Autonomous Agents and Multi-Agent Systems*, vol. 2, no. 4, pp. 357–401, Nov. 1999.
- [20] M. J. Wooldridge, *Reasoning about rational agents*. MIT press, 2000.
- [21] N. Farenc, S. R. Musse, E. Schweiss, M. Kallmann, O. Aune, R. Boulic, and D. Thalmann, "A paradigm for controlling virtual humans in urban environment simulations," *Applied Artificial Intelligence*, vol. 14, no. 1, pp. 69–91, 2000.
- [22] C. Peters, S. Dobbyn, B. MacNamee, and C. O'Sullivan, "Smart Objects for Attentive Agents," in *Proceedings of 11th International Conference in Central Europe on Computer Graphics*, 2003.
- [23] S. Donikian and S. Paris, "Towards Embodied and Situated Virtual Humans," in *Motion in Games*, ser. Lecture Notes in Computer Science, A. Egges, A. Kamphuis, and M. Overmars, Eds. Springer Berlin Heidelberg, 2008, vol. 5277, pp. 51–62.
- [24] F. Heckel and G. Youngblood, "Contextual Affordances for Intelligent Virtual Characters," in *Intelligent Virtual Agents*, ser. Lecture Notes in Computer Science, H. Vilhjmsson, S. Kopp, S. Marsella, and K. Thrisson, Eds. Springer Berlin Heidelberg, 2011, vol. 6895, pp. 202–208.
- [25] M. Kapadia, J. Falk, F. Znd, M. Marti, R. W. Sumner, and M. Gross, "Computer-assisted Authoring of Interactive Narratives," in *Proceedings of the 19th Symposium on Interactive 3D Graphics and Games*, ser. i3D '15. New York, NY, USA: ACM, 2015, pp. 85–92.
- [26] J. Gibson, "Perceiving, Acting and Knowing," R. Shaw and J. Bransford, Eds. Lawrence Erlbaum, 1977.
- [27] T. Balint and J. Allbeck, "Automated Generation of Plausible Agent Object Interactions," in *Intelligent Virtual Agents*, ser. Lecture Notes in Computer Science, W.-P. Brinkman, J. Broekens, and D. Heylen, Eds. Springer International Publishing, 2015, vol. 9238, pp. 295–309.
- [28] R. E. Fikes and N. J. Nilsson, "STRIPS: A new approach to the application of theorem proving to problem solving," *Artificial intelligence*, vol. 2, no. 3, pp. 189–208, 1972.
- [29] R. Bindiganavale, W. Schuler, J. M. Allbeck, N. I. Badler, A. K. Joshi, and M. Palmer, "Dynamically Altering Agent Behaviors Using Natural Language Instructions," in *Proceedings of the Fourth International Conference on Autonomous Agents*, ser. AGENTS '00. New York, NY, USA: ACM, 2000, pp. 293–300.
- [30] O. Van Kaick, A. Tagliasacchi, O. Sidi, H. Zhang, D. CohenOr, L. Wolf, and G. Hamarneh, "Prior knowledge for part correspondence," vol. 30. Wiley Online Library, 2011, pp. 553–562.
- [31] Y. Wang, S. Asafi, O. van Kaick, H. Zhang, D. Cohen-Or, and B. Chen, "Active Co-analysis of a Set of Shapes," *ACM Trans. Graph.*, vol. 31, no. 6, pp. 165:1–165:10, Nov. 2012.
- [32] X. Zhang, T. TuteneL, R. Mo, R. Bidarra, and W. F. Bronsvooort, "A Method for Specifying Semantics of Large Sets of 3d Models," in *GRAPP/IVAPP*, 2012, pp. 97–106.
- [33] M. Gutierrez, D. Thalmann, and F. Vexo, "Semantic virtual environments with adaptive multimodal interfaces," in *Proceedings of the 11th International Multimedia Modelling Conference, 2005. MMM 2005*. IEEE, 2005, pp. 277–283.
- [34] F. Worgotter, E. Aksoy, N. Kruger, J. Piater, A. Ude, and M. Tamosiunaite, "A Simple Ontology of Manipulation Actions Based on Hand-Object Relations," *Autonomous Mental Development, IEEE Transactions on*, vol. 5, no. 2, pp. 117–134, Jun. 2013.
- [35] L. Drumond and R. Girardi, "A Survey of Ontology Learning Procedures," *WONTO*, vol. 427, pp. 1–13, 2008.
- [36] P. Buitelaar, P. Cimiano, and B. Magnini, *Ontology learning from text: methods, evaluation and applications*. IOS press, 2005, vol. 123.
- [37] P. Cimiano, *Ontology Learning and Population from Text: Algorithms, Evaluation and Applications*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.
- [38] P. University, "About WordNet," Princeton University, Tech. Rep., 2010. [Online]. Available: <http://wordnet.princeton.edu>
- [39] C. Pelkey and J. M. Allbeck, "Populating Virtual Semantic Environments," *Computer Animation and Virtual Worlds*, vol. 24, no. 3, pp. 405–414, May 2014.