LVC Training in Urban Operation Skills

Tijmen Joppe Muller
Karel van den Bosch
Philip Kerbusch
TNO Defence, Security and Safety
PO Box 23
3769 ZG Soesterberg, The Netherlands
tijmen.muller@tno.nl, karel.vandenbosch@tno.nl, philip.kerbusch@tno.nl

Lkol. J.H. Freulings
Netherlands Army, Simulation and Training Center (SIMCEN)
Postbus 3003, 3800 DA Amersfoort, the Netherlands
jh.freulings@mindef.nl

Keywords:
Live/Virtual/Constructive, Training, Simulation

ABSTRACT: *Urban Operations are broadly recognized as an important part of current military expeditions. A key aspect in these operations is the interaction with the local population; soldiers need to handle ubiquitous threats while winning the hearts & minds of the civilians. Sensemaking and decision making in such situations is a real challenge. Highly trained and experienced soldiers are essential for successful urban operations.*

Current training programs include live field exercises, but often lack sufficient role players to populate the training. In this paper we report efforts on integrating Live, Virtual, and Constructive (LVC) components to create a flexible and contextually rich environment for training in Urban Short Range Interaction (USRI). In this concept, trainees enter a live, physical environment (a room), enriched with virtual characters (projected on the wall). Constructive models allow these virtual characters to act and respond intelligently to the trainees and to the events taking place in the room. Achieving such a training facility requires (a) tracking of trainee behavior, (b) modeling behavior of virtual characters, and (c) rendering behavior of virtual players in the training environment. The paper addresses the principal design issues, experiences and conclusion in the development of an LVC integrated USRI trainer.

1. Introduction

World events of the last two decades demand a reorientation of military missions. Military deployment has shifted away from preparing for major large-scale conflicts towards coalitional, peace keeping, and other forms of newer types of operations. An important part of this current military reality are the close encounters with opponent forces, neutrals and civilians. They often take place in small and confined areas (e.g. streets, houses, markets, shops, etc). The presence of locals in such *Urban Short Range Interaction (USRI)* operations is an important factor, because it is often unclear whether the local people are friendly or hostile. Additionally, winning the hearts and minds of the population is considered a critical success factor of operations in political terms. This faces modern soldiers often with a dilemma. When approaching a small crowd in a street or when confronted with a group of people in a house, the commander and his team have to consider the likelihood of hostile intentions. Not acting appropriately to a threat may endanger the own group. On the other hand, acting with violence against people or groups that have no evil intentions harms the respect and goodwill of the local population. This, in turn, endangers the mission as a whole and puts the safety of own forces at risk on

Figure 1: Operation in urban environment

subsequent occasions. It is therefore clear that USRI-operations demand highly trained and experienced soldiers that can assess the nature of situations adequately and rapidly, even in complex and ambiguous circumstances (Lussier, 2003).

The Netherlands Army realizes that the required skills go beyond the traditional military training, and has introduced new facilities and methods for training USRI-operations. In addition, they initiate research into advanced forms of training that open up new opportunities.

Current military training programs for urban operations focus on exercises in live training facilities, like shoot-houses. These are high-fidelity replica of villages and houses. There is no glass in the windows, but apart from that, the interior (number and size of rooms, doors, hallways, stairs, et cetera) is similar to ordinary houses. This allows proper training in the procedures for approaching, entering and clearing a house, generally labeled as training TTPs (Tactics, Techniques, Procedures). However, as already pointed out, the critical element in such operations is not only the question of how to get into the house, but also in how to deal with the local people present. This involves: how to assess the intentions of the local people, how to evaluate whether they impose a threat, how to identify immediate danger, how to make our own intentions clear, etc. At this point, the Netherlands Army recognizes a bottleneck. In order to learn and practice these sensemaking and decision making skills, the training facilities need to be populated with 'locals'. On few occasions, staff members are asked as role players, acting as 'local' people. However, staff members are often not available. When they are available, they often turn out not to behave in a representative fashion, or they fail to act in the best interest of the trainee. As a result, situations with realistic locals are very rarely practiced. They often acquire their skills and experience once they are in the mission area, through "on-the-job" training under supervision of an experienced commander.

One opportunity to develop a contextually rich and flexible environment for training urban operations is the combination and integration of *live*, *virtual*, and *constructive* tools (Frank, Helms & Voor, 2000), abbreviated as LVC. In this paper we present a study in which we explore the opportunities of LVC for training USRI-operations. First we shortly discuss the concept of LVC. Then we will discuss the functional and technical requirements for a LVC-USRI training environment. Opportunities and issues to be solved are discussed in the final section of this paper.

Figure 2: Live training facility Marnehuizen

2. LVC

'Live' simulations are the traditional method of training. In 'Live' simulations, human operators use real (usually their own) hardware (rifle, vehicle, etc). Thanks to achievements of technology in the past decades, we can now create high-fidelity virtual simulations, that enable people to train in virtual worlds. In a 'Virtual' simulation, human operators use simulated systems. Integrating Live with Virtual simulations provide new possibilities for training and instruction. The main advantages of using a Virtual world for training and instruction are (1) the amount of control one has over the virtual world, (2) the possibility to simulate assets that are scarce or not available at all in the live world, and (3) efficient delivery of training.

An important control feature of virtual worlds is the possibility for designing the training context: aspects such as terrain and objects can be designed at will. Thus, scarce or not available assets can be simulated. Secondly, the instructor has far more control over events and situations than in a live world. Finally, virtual worlds offer the opportunity to present trainees with a wide variety of training situations in a relative short time.

In an integrated live-virtual training environment, staff and/or instructor is still very much needed to control the course of events. The instructor decides whether an event takes place or not, instructs entities (such as teammates, enemies or civilians) how to act and respond, declares whether equipment fails or not, etc. He needs this control to create a situation that fits the current training needs of the trainee as good as possible. Self-evidently, realizing this control requires a lot of preparations and instructor capacity, and for many cases it is very hard to realize the desired control at all. It would be very convenient if such intelligent control could be added to the training environment automatically. Here comes the element of constructive simulation into play. 'Constructive' simulation entails the artificial intelligence producing simulated human entities. For example, models of human behavior can be used to control how virtual players will or should act and react in a training scenario (whether defined beforehand or controlled in real-time).

By integrating 'Live', 'Virtual', and 'Constructive' simulation assets, strengths of the tools can be combined, or weaknesses canceled out (van den Berg, de Reus, & Voogd, 2011). This integration requires several challenges to be overcome. A main issue is how to represent live entities and their behavior in a virtual environment and vice versa. In the next sections we will discuss our approach for achieving this in an USRI case.

3. USRI

For the training of Urban Short Range Interaction (USRI) we intend to implement a system for situations in which the trainee needs to engage with one or more persons in an urban environment. The scenarios to be trained involve making split-second decisions, but also situations requiring some sort of interaction with 'local' people to make sense of the encountered situation. The training system is composed of live, virtual, and constructive components. The live component consists of a (physical) room, the trainee and his equipment. The virtual world consists of the extension of that same room, containing one or more entities and objects, with which the trainee can interact. The constructive component are the models underlying the behavior of the virtual characters in the room. For creating realistic and immersive training it is considered important that trainees can move freely through their environment and interact naturally with the system. This means we do not want to use mouse or keyboard for interaction nor do we want to use intrusive sensors that may hinder the user. The training system described above provides a natural environment in which the trainee can be offered a wide variety of scenarios in which he can train his sensemaking and decision making skills. For the moment the facility under development is to be used in a stand-alone fashion, but the long-term goal is to integrate the training system in an existing live training facility of the Netherlands Army.

3.1 Functional requirements

In order to allow interaction between the human trainee acting 'live' in a real training environment, and the virtual role player acting in the form of an avatar in the same environment, aspects of the live world need to be represented virtually and vice versa.

Tracking behavior of trainee(s): Relevant actions of the trainee need to be recognized by the virtual player. Of course, not all behavioral elements of the human trainee needs to be conveyed to the virtual player. For example, the eye blink frequency of the trainee is considered not relevant, hence the virtual player needs not to be informed on this. However, when the trainee issues a command or when he points his rifle, then the virtual player do needs to know this, as this will affects his response. We identified the following behaviors of the trainee as relevant for USRI-operations:

- position in the room and relative position to virtual player;
- pose;
- direction of gaze;
- use of equipment (e.g. radio);
- direction of weapon;
- use of weapon;
- verbal communication;
- physical communication (e.g. gestures).

As we want our LVC-USRI training system to track players non-intrusively, the player should not need to be equipped with dedicated sensors. The tracking system should recognize position, orientation, pose and gestures of multiple persons (house searching is conducted in teams). Furthermore, it should identify position and orientation of rifles. Ideally, the system should be able to read and recognize facial expressions of the trainees.

Modeling behavior of virtual player(s): The (re)actions of the virtual role player to the events taking place (including the actions taken by the trainee) need to be modeled and represented in the live environment. This calls for a reasoning framework (AI model) that incorporates both relevant information and events from the virtual environment, and information on relevant behaviors from the live human players. There are many ways to model the behavior of virtual players, for example Finite State Machines (FSM) and Belief-Desire-Intention (BDI) modeling. FSM defines behavior of a character as a function of *states* (a specific constellation of circumstances in the scenario), *state transitions,* and *actions* that bring about these state changes (Gill, 1962). In BDI models (Bratman, 1987), a character is not instructed to act upon a certain state in the scenario, but rather upon the interpretation of that state (not physical states, but mental states, so to say). An event brings about "a belief" in the mind of a character (e.g. hearing rifle fire creates the belief that


Figure 3: Microsoft Kinect camera

there is an attack). The belief triggers a goal. What goal is triggered depends upon the context and role of the character: a combatant, for instance, may adopt the goal to fire back; an innocent civilian may adopt the goal to hide himself; a child may adopt the goal to run away, etc. Each approach to modeling behavior has its strength and weaknesses. FSM is, for example, particularly suited for well-defined tasks under relatively tight control. BDI-modeling is particularly suited for modeling goal-directed behavior.

A potential strength of an LVC-USRI trainer is the opportunity to create a wide variety of scenarios. In order to realize this strength, a variety of entities needs to be modeled (male, female, children of different ethnicities), clothing, objects (e.g. weapons, IEDs), and anything that is needed by the scenario and relevant for the sense- and decision making of the trainee. Optionally, the reasoning framework should take into account higher-order factors influencing the behavior of virtual players in an USRI-like setting. For example, cultural values or religious views contribute to how an individual interprets and responds to certain actions of another individual (e.g. reaching one’s hand, or gaze at a woman).

Rendering virtual players’ behavior in the training environment: The AI-framework sends the behavior descriptions of the virtual characters to the simulation system. This system renders all elements of the behavior description in the simulation (e.g. facial expression, movements, verbal utterances) in such a fashion that the behavior can be correctly recognized and interpreted by the human player. A wide range of animations (e.g. walking, smiling, looking angry, squatting down, ducking away, reaching one’s hand, etc) is needed to adequately visualize the behavior of the virtual role players.

Clearly, achieving an LVC USRI training environment requires coordinated and integrated interplay of tracking, modeling and rendering components.

3.2 Technical specifications

The LVC-USRI training system has the following components:

- A live environment (a small room), built up from mobile elements, so the floor plan of the room can be easily changed. Additionally, furniture or other elements that may influence the trainee’s actions may be added
- Live sensors:
    - a tracking system to record information on actions of the trainee (e.g. position movements, gaze direction, et cetera)
    - a speech recognition system to record utterances of the trainee
- A model of a virtual world to be superimposed on the live environment, including various objects (e.g. paintings, doors or windows to be projected onto the walls of the live environment)
- Behavior models of the virtual entities present in the scenario (e.g. enemy combatants, civilians, etc)
- Animation models for rendering the behavior of the virtual players
- Speech synthesis system for producing verbal utterances of the virtual players
- Hardware for rendering the virtual environment onto the live environment: this includes one or more projector and sound systems
- Middleware to achieve the required interoperability between the live and the virtual environment

We have been experimenting with two tracking systems: (a) the iisu motion tracking system by SoftKinetic and (b) the OpenNI/NITE tracking system, which makes use of the Microsoft Kinect camera (see Figure 3). Both support non-intrusive tracking of multiple persons. SoftKinetic is commercial software. OpenNI/NITE is freely available.

We found the iisu system having trouble recognizing the orientation of the trainee: it only worked well if the trainee was front-facing the camera. When the trainee rotated around his vertical axis, the system did not recognize this motion.

The OpenNI/NITE system does not recognize posture. However, it is relatively easy to add external functionalities to the OpenNI/NITE system, so we expect this to be a solvable issue.

Both tracking systems failed to recognize the direction of head (i.e. gazing view); they only were successful in determining body orientation. The recognition of position and orientation of equipment proved

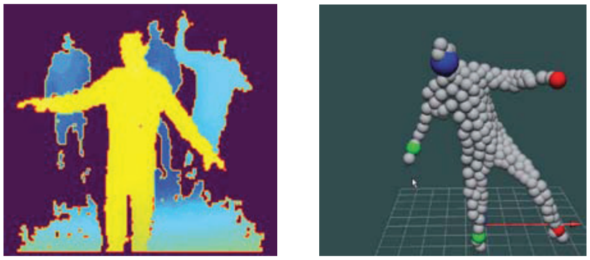

Figure 4: Stances recognized by tracking system

troublesome: for example, a weapon was typically 'recognized' as a very long arm. Also, gesture recognition is limited to a few predefined gestures for both systems.

A solution for both limitations would be to equip a helmet and a weapon with a wireless orientation sensor and deduct position of head and weapon from the position of the trainee itself.

We have chosen to use the military-off-the-shelf game VBS2 as the tool for generating our virtual environment (see Figure 5). VBS2 has already a large amount of USRI-relevant content. It also has a large user group, providing continuous development of new content (entities, objects, and animations). The game is relatively open, as it supports a scripting interface (ASI), allowing the creation of extension components that add new or improve existing functionality of VBS2. VBS2 also supports HLA, a standard for simulation interoperability. Both the ASI and HLA can be used to integrate other functionality (e.g. an AI framework or speech recognition).

For constructing behavior models, we as yet use Finite State Modeling. For the more complex scenarios that lie ahead of us, we will use the concept of Belief-Desire-Intention to be implemented in JADEX. Over all, we will probably not rely on one single modeling paradigm. A hybrid approach is taken, where some components of behavior are modeled using one particular paradigm, and other components are modeled using another paradigm.

Figure 5: Virtual role player in VBS2

Scenarios for the LVC-USRI trainer have been developed with the support of military subject matter experts from the Netherlands Army. SMEs were from the USRI training and simulation departments. The scenarios are military relevant, and designed in such a way to be suitable for use in a shoot-house training setting. The scenarios vary across several variables in order to offer a diverse interaction; e.g. visual 'target' representation, 'target' intention and personality, number of 'targets', and environment can be different.

4. Discussion

Urban Operations are an increasingly important part of current military expeditions. Making sense of a situation and assessing the intentions of people from their behavior are critical skills for USRI-operations. New facilities and methods are needed to train such skills effectively and efficiently. In this paper we have presented issues in the design of an advanced training USRI facility, integrating live, virtual and constructive elements.

The USRI training system consists of a physical room, extended with virtual objects and virtual entities (people). The trainee is physically present in the room and interacts with the virtual entities in a natural fashion. We identified three main requirements for such an LVC-USRI trainer to be realized: (1) tracking the human trainee(s), (2) modeling the behavior of virtual players, and (3) rendering virtual players' behavior in the training environment.

From the reported experiences with tracking devices we conclude that the open source community has picked up the developments very adequately. Especially the freedom to add new functionalities and the low cost of the system makes open source variants at this moment very attractive from a research perspective. With respect to speech recognition, our current scenarios require only a small set of utterances to be recognized (e.g. "show me your hands", "open the door", "are you alone?"). This can be achieved by standard low- to middle-end speech recognition systems.

For modeling the behavior of virtual players we conclude that for the rudimentary scenarios we are currently working with, Finite State Machines (e.g. predefined scripts of behavior for each anticipated events in the scenario) is yet the most appropriate approach. However, with more realistic and subtle scenarios in the foreseeable future, Belief-Desire-Intentions model are most likely the best approach to model behavior of virtual characters. BDI-modeling is particularly suited for modeling goal-directed behavior. The advantage of BDI over FSM is that behavior models are more flexible and reusable (Van den Bosch & Van Doesburg, 2005).

With respect to rendering the virtual worlds, we consider VBS2 as the best option so far. VBS2 is widely used by the international military community. It has already a large amount of USRI-relevant content and new content is developed by many parties. The game is relatively open and can be connected to middleware relatively easy. However, there are also downsides. It has only a limited set of animations. For example, squatting down, gesturing, averting an attacker, hand shaking and many more animations are not in the standard repertoire of VBS2. Although the software includes tools to create custom animation, this takes a lot of time and efforts. For synthesizing speech expressed by the virtual characters, any middle-end system offers sufficient functionality.

Considering our progress with the LVC-USRI facility, we feel subsequent scenarios to require more, and longer, interaction (e.g. dialogues). We think these lie within reach, given the developments of technology.

5. References

Berg, T.W. van den, Reus, N.M. de, Voogd, J.M. (2011). LVC Architecture study *Simulation Interoperability Workshop*, Spring 2011.

Bosch, K. van den, Doesburg, W.A. van (2005). Training Tactical Decision Making Using Cognitive Models. In: *Proceedings of the Seventh International NDM Conference*. Amsterdam, the Netherlands.

Bratman, M. E. (1987). *Intention, Plans, and Practical Reason*. Cambridge, Mass., Harvard University Press.

Frank, G.A., Helms, R. & Voor, D. (2000). Determining the Right Mix of Live, Virtual, and Constructive Training, *Proceedings of the 21st Interservice/Industry Training Systems and Education Conference*.

Gill, A. (1962). Introduction to the Theory of Finite-state Machines. McGraw-Hill.

Lussier, J. (2003). Intelligent tutoring systems for command thinking skills. ARI (Army Research Institute) Newsletter, 13, 7-9.

Author Biographies

TIJMEN JOPPE MULLER is a researcher at the Training Innovations department at TNO. He graduated at the University of Twente in Enschede with a degree in computer science and a teaching degree in mathematics. He has a professional interest in both serious gaming and artificial intelligence. His work focuses on the use of games and simulations in training design for the military and safety domains.

KAREL VAN DEN BOSCH is a Senior Researcher at the Training Innovations department at TNO. Karel investigates how intelligent agents can successfully support simulation-based training, thus making training more systematic (uniform behaviour of agents), more effective (agents consistently eliciting intended behaviour of trainee), and more efficient (role players need no longer be present during training).

PHILIP KERBUSCH, MSc is a research scientist at TNO Defense, Security and Safety. Philip holds a MSc. degree cum laude in Artificial Intelligence from Maastricht University (2007). He has a strong background in computer science and knowledge engineering. His current work involves various applications of artificial intelligence in the defense and safety domain, specializing in behavior modeling for serious games and simulation.